earPod: Efficient Hierarchical Eyes-free Menu Selection

by

Shengdong Zhao

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Department of Computer Science
University of Toronto

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-58989-2
Our file Notre référence
ISBN: 978-0-494-58989-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.
.
The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

earPod: Efficient Hierarchical Eyes-free Menu Selection

Shengdong Zhao

Doctor of Philosophy

Department of Computer Science
University of Toronto

2009

Abstract

The research in this dissertation developed and evaluated a new method for menuing interaction that is intended to be better suited than current methods with respect to mobile eyes-free scenarios. The *earPod* prototype was developed and then evaluated in a series of four experiments. In the first two experiments, *earPod* was first compared against an iPod-like (visual) interface and then against a fuller set of competitive techniques that included dual vs. single modality presentations, audio vs. visual modalities, and radial vs. linear mappings. The third experiment consisted of a longitudinal study designed to understand the learning patterns that occurred with these techniques. The fourth experiment examined performance in a conventional (single task) desktop setting and in a driving simulator (i.e., a dual task situation where participants carried out the driving task while interacting with the mobile device).

The results of these experiments, comparing *earPod* with an iPod-like visual linear menu technique on fixed-sized static menus, indicated that *earPod* is comparable both in terms of speed and accuracy. Thus it seems likely that *earPod* should be an effective and efficient eyes-free menu selection technique. The comprehensive 3x2 study implemented in Experiment 2 showed that the benefit of *earPod* was largely due to the radial menu style design. While performance using it was comparable in both speed and accuracy with the visual linear menus, its performance was slower than for a visual radial style menu. In the multi-task simulated driving condition in Experiment 4, where concurrent tasks competed for visual attention, the eyes-free *earPod* interface was found to be beneficial in improving performance with respect to the safety related driving parameters of following distance and lateral movement in the lane. Thus auditory feedback appears to mitigate some of the risk associated with menu selection while driving. Overall, the results indicated that not only should *earPod* menuing be able to provide safer interaction in dual task settings, but also that, with sufficient training, audio only menu selection using innovative techniques such as those employed by *earPod* can be competitive with visual menuing systems even in desktop settings.

Acknowledgments

When I graduated with my Masters degree from the University of California, Berkeley in 2001, I had little idea what to do next. My father encouraged me to pursue a PhD and his inspiration, together with the opportunity to work with Monica Schraefel, then at the University of Toronto (now a professor at the University of Southampton, UK), led me to undertake this memorable and invaluable six-year experience as a step towards a career of academic research and teaching. I wholeheartedly thank Dad and Monica for guiding me to find the career of my choice.

As a young graduate student, influenced by the software industry mindset from Silicon Valley, I struggled in finding a suitable research topic and was without a research advisor after Monica's move to the UK. That situation changed after taking Ravin Balakrishnan's Topics of Interactive Computing class in 2004. Under Ravin's guidance, I published my first major publication at the ACM UIST conference. I sincerely thank Ravin for helping me launch my research career.

I was also fortunate at that time to find my current supervisor, Mark Chignell. In addition to his intellectual guidance, Mark has a warm, flexible personal touch with his students. The Interactive Media Lab (IML) that Mark directs has the atmosphere of a loving family. In many ways, Mark is the caring father of the lab. It's truly a blessing to have fulfilled my PhD dream in the caring and warm environment of IML. Through personal demonstration, Mark not only taught us the way to conduct scientific research, but also showed us the right attitude to take in our lives.

The UIST paper significantly boosted my research career. In addition to receive recognition from reviewers, I had the opportunity to get to know Ken Hinckley, who later introduced me to two stimulating and productive internships at Microsoft Research in the summers of 2006 and 2007. At Microsoft, I was fortunate to work under the direct guidance of Maneesh Agrawala and Ken. Both of them taught me so much about how to conduct research and face challenges in life. I also had the opportunity to learn from many great minds in the field, including Patrick Baudisch, Ed Cutrell, Michael Shilman, and Desney Tan. Special thanks are due to Mary Czerwinski, who was my external examiner, and who provided me with constructive feedback on my dissertation.

The thesis topic, *earPod* and eyes-free menu selection, was not determined until 2006 -- the 4th year of my PhD program. The interest to auditory interaction can be traced back to Mark's interest in auditory interaction and the Vocal Village, a voice conferencing tool using spatial

audio. Mark inspired me to carry out a literature review in auditory interface and interaction techniques and helped me to establish a rich knowledge base which later led to the invention of *earPod*.

The initial idea of developing an efficient auditory menu came from a discussion of my research idea with Shumin Zhai at UIST 2004. Shumin is a prominent researcher in HCI whom I have tremendous respect for. He gave me initial encouragement on my idea of an auditory acronym menu. The idea was later explored and transformed into auditory marking menu under the ARISE project where Mark and Ravin were co-principle investigators (with Professor Charlie Clarke of the University of Waterloo).

The initial auditory menu design relied on sequential playback of audio messages like the Interactive Voice Response system and used a pen as its input device was inefficient for novice users and impractical for the mobile environment. Refinement of the idea, which used circular gestures to combine both relative and absolute access to menu items, happened when I was riding the subway from home to school in early 2006.

The earPod project got a real boost from Pierre Dragicevic, a visiting Post-doc fellow of the Dynamic Graphic Project Lab at the University of Toronto (now a scientist at INRIA, France). Pierre and I went through many design explorations in early 2006. He also replaced the slow Java Sound API with an audio implementation using the Jazz library which made *earPod* truly reactive. The technique was initially called Sonic Glide, but was later changed to *earPod* due to Patrick Baudisch's excellent suggestion.

I am very grateful to the opportunity to collaborate with Dario Salvucci and Duncan Brumby at the Drexel University to extend the study of *earPod* to a dual task driving setting. Many thanks are also due to members of my PhD committee. In addition to Mark, Ravin, and Mary, I was also indebted for the constructive feedback provided by Steve Easterbrook and Khai Truong. I also want to thank David Modjeska for his constructive feedback on my literature review chapter, and for his invaluable advice for my PhD presentation.

I benefited tremendously from vigorous discussions and moral support from many colleagues, including members of the IML Lab: Alvin Chin, Danielle Lottridge, Annie Xu, Flora Wang, Jamy Li, Anna Malandrino, Rick Bodner, Silvia Gonzalez, and many others, and members of the

DGP lab: Michael McGuffin, Tomer Moscovich, Koji Yatani, Xiaojun Bi, and many others. Mark's son, Andrew Chignell, also helped me greatly in conducting the driving experiment.

I especially want to thank my mother Jenny for her generous and unconditional support from all these years. Without her, this thesis will never be possible. Special thanks to my wife Zoey, my brother Weidong, and my sister Yun. Together with Dad and Mom, you formed the special connection among us and created a spiritual land called home that made our individual efforts worthwhile and meaningful. I also want to acknowledge the encouragement and moral support of my dear friends: Hua Jiafeng, Zhao Zhifang, Kevin Chen, Joe Zhang, Susy Wu, Amy Zhu, and many others.

Finally, this thesis is dedicated to my Grandpa, Yuting, who unfortunately left us this past October. Grandpa, you are the foundation of our family. Your love and support for us is rooted deeply in our hearts. We love you forever.

Table of Contents

List of Figures

List of Appendices

Chapter 1
Introduction

1.1 Overview

As of this writing, billions of mobile phones, hundreds of millions of MP3 players, and many millions of other mobile devices (including various types of PDAs) are being used around the world. The mobile phone has become the most widely used computer platform, far outdistancing desktop and laptop computers in terms of the number of units in use. Mobile phones have capabilities that desktop computers don't have and that laptops are only clumsy substitutes for, namely the ability to carry out computing functions anytime, anywhere (provided that service is available) with the possibility for a whole new set of services based on where the person is at a particular point in time. The possibilities for new forms of mobile computing seem almost endless. As one example, the combination of social networking software and locational awareness plus local business directories leads to whole new ways of meeting people, interacting with services and sharing experiences.

In a decade or so the mobile phone went from being a curiosity to a necessity and it seems likely that in the next decade the mobile phone and similar mobile devices will emerge as fully functioned computing platforms. The mobile platform differs from the traditional desktop-centric computing environment in many ways. Not only are devices redesigned and packaged into much smaller form factors to achieve portability, they are increasingly used in a wide variety of different contexts and settings, ranged from offices to airports, subway trains, cars, streets, treadmills, etc. Mobile interface design has become increasingly challenging due to the different cognitive demands, physical constraints, and social constraints imposed by such diverse usage scenarios. Many of the mobile computing scenarios require multi-tasking where the primary task demands visual attention. Interacting with a visual interface in these scenarios creates competition for limited visual resources. For example, interaction with an iPod while driving may be distracting and constitutes a potential safety hazard (Salvucci et al. 2007). In addition to safety concerns and the problem of competition for people's visual attention, there is also the issue that mobile devices tend to have small screens that are difficult to read in bright sunlight or for people with somewhat impaired vision, including the elderly. Thus there is a strong motivation to create new kinds of interface, and auditory interfaces in particular, that can meet the special requirements of mobile computing.

Today's mobile devices tend to use similar interface widgets to those found in desktop interfaces. In many respects, mobile computer interfaces can be characterized as "shrunk-down" desktop interfaces. However, it seems unlikely that mobile computing interfaces that were carefully designed from the ground up would look like the desktop-legacy systems that are currently in use. Can mobile computing be re-invented with a more mobile-friendly, safe, and efficient interface that can work for a broad range of users? It is the belief of this author that not only is radical re-thinking of the mobile interface possible, but necessary. However, the re-invention of mobile computing is a huge undertaking and far beyond the scope of a single Ph.D. dissertation. Thus the research reported below will look at the re-invention of a particular functionality within mobile interfaces, namely menu selection.

One of the most fundamental tasks in interacting with computers is selecting from a set of alternatives (Foley et al. 1984), typically presented as some form of menu (Norman 1991). While current menuing techniques have been developed for use on desktop computers, new mobile computing platforms are now in general use, and menuing systems need to be adapted to the properties and requirements of these devices.

Thus, it is important to find methods of menu interaction that are efficient and safe to use while mobile. This thesis presents a touch-based eyes-free menu selection technique called *earPod*. *earPod* is designed to minimized the delay caused by using audio as the communication media. It uses interruptible and non-speech audio to shorten the playback time, allowing users to skip ahead to items of interest. The system also follows a reactive interaction model, so that auditory items are played back as soon as the finger reaches their locations.

1.2 Research Strategy

The research carried out in this dissertation focused specifically on the design and testing of an appropriate menu selection technique for eyes-free interaction that can function effectively when utilized on a mobile device in the context of a visually-demanding primary task. In order not to compete with the primary task for the visual channel, the design activity focused on a combination of haptic (touch) and auditory modalities assigned to the input and output aspects, respectively, of menu interaction.

The design was jointly influenced both by the properties of the menu selection task and also by the properties of the auditory and haptic modalities. For instance, sound has a number of properties that makes its use as a sole source of feedback challenging. The temporal nature of sound requires users to rely on their short-term memory to compare and select objects. The temporal nature of sound (and speech) also means that it generally takes longer to provide feedback, as compared to vision. An additional property of sound is that we cannot close our ears if we are not interested in the information it provides (although we can wear earplugs) which can be beneficial in some situations (being able to hear alarms) but not in others (e.g., being distracted or annoyed by noise). Special care is needed in designing sound feedback so that it can be perceived relatively unobtrusively. All these factors make sound harder to control than vision.

The eyes-free menu selection method that resulted from the design activity in this dissertation is a touch-input auditory feedback technique called *earPod* (Zhao et al. 2007). *earPod* provides users with audio feedback that is synchronously linked to touch input. *earPod* is intended to allow users to discover menus at their own pace. Seamless transition from novice to expert use is facilitated with an absolute radial menu layout that allows direct access to known items. Spatialized audio feedback additionally reinforces the mappings of items to a circular touch-sensitive surface.

Even given the constraints cited above there is a huge potential design space for eyes-free interfaces that use a combination of auditory and haptic modalities. The research strategy adopted in this dissertation was not to explore that design space broadly but rather to choose a promising region within that design space, then to iteratively design a specific prototype which could then be evaluated in a series of experiments.

The research was carried out in the following steps:

1. Define the research problem
2. Define a general design concept
3. Iteratively refine the concept through pilot testing with a series of prototypes until a workable prototype emerges that can be used in formal testing

4. Carry out a series of experiments to explore the properties of the prototype and how well it compares to competitor techniques
5. Construct a set of design recommendations, based on the experimental results, that can be used in designing future eyes-free menu selection systems

In this dissertation, the motivation for the design of the *earPod* is provided along with a description of the prototype that was used in this research. A detailed evaluation of the *earPod* prototype is reported as a series of experiments that were carried out in this dissertation to compare the effectiveness of *earPod* menu interactions to existing methods for visual and auditory menu selection. The main dependent measures considered in the experiments were speed, accuracy, and the rate and amount of learning. Since performance with an interaction technique may vary greatly with the skill of the user, learning effects, and the transition from novice to expertise performance was of particular interest in these studies.

The first experiment compared the new *earPod* method against a dominant existing mobile menu selection method (as implemented on the iPod). It established that the new method could be competitive with existing techniques.

The second experiment compared all six possible combinations of menuing methods constructed using three types of modality (audio, visual, dual) and two types of menu presentation (linear vs. radial).

Due to the learning effects observed in the second experiment it became clear that a lengthier longitudinal study was required to see how learning stabilized over time and to further trace the transfer to expertise when using a novel menu selection method. This longitudinal study was carried out in the third experiment.

In order to examine the properties of *earPod* interaction in a dual task setting, the fourth experiment looked at menu selection carried out during a simulated driving task. In this study, menu selection during simulated driving was compared with a desktop menu selection condition where menu selection was the primary and only task.

Statistical analyses were used in this thesis to highlight interesting differences that provide insight into the design of eyes free menu selection. By convention a significance level of .05 was used in the ANOVAs and t-tests reported in this thesis to signal potentially interesting differences. Since there were a number of statistical effects examined in each study, readers who are concerned about possible inflation of the family-wise alpha level (i.e., the likelihood that some of the tests reported as significant may be due to chance since the probability that at least one test will be significant by chance rises as an increasing number of tests are carried out) may choose to focus on effects that have a p-value of less than .01 in this dissertation.

In conducting this research, many decisions had to be made about what research questions should be addressed in the experiments and how the experiments should be conducted. Given the novelty of the *earPod* method that was used, there was little in the way of theory to allow predictions to be made concerning the results that would be obtained under different conditions. Thus an explicit hypothesis testing approach to the analysis of the data has not been used. The intent of this research is to open up a new area of design and research for auditory based menu selection, and to provide preliminary findings concerning the viability of *earPod* menu selection under different conditions.

One other technical issue was that three of the four experiments in this dissertation involved the use of repeated measures designs which are known to be susceptible to the effects of asymmetric transfer between conditions across different orderings (Poulton and Freeman 1966). Analyses of order effects were carried out to ensure that the results reported were not tainted by the effects of asymmetric transfer. The detailed analyses of order affects are reported in Appendix 1 and in cases where they cast doubt on the results using repeated measures analyses, their implications are also discussed in the corresponding experimental chapters in the body of the dissertation.

1.3 Contributions

This thesis has a number of contributions which are briefly noted here. A more detailed discussion of these contributions is provided in Chapter 8 of this dissertation. Contributions 1-3 are on *earPod* methodology; contributions 4-7 relate to empirical results, and contributions 8-9 are concerned with design recommendations.

Contribution 1: Development of a method for unifying relative (gliding) and absolute (tapping) menu access in a circular touch input, whilst provide a smooth and seamless transition between those two methods of access.

Contribution 2: Development of an innovative eyes-free menu selection method with touch input and reactive audio feedback.

Contribution 3. Development of a method for using continuous spatialized audio feedback during menu selection to reinforce the user's cognitive mapping between menu items and spatial locations on the touchpad.

Contribution 4. Demonstration, through experimental results, that performance with the earPod method was generally comparable with performance obtained using a visual linear menu.

Contribution 5.Demonstration, through experimental results, that learning was greater in the earPod condition than in the visual menu selection conditions.

Contribution 6. Demonstration, through experimental results, that earPod menu selection outperforms other techniques in the context of a visually demanding primary task.

Contribution 7. Demonstration, through experimental results, that transition from novice to expert performance in earPod menu selection can be relatively fast, but is dependent on the number of menu items to be learned.

Contribution 8. Derivation of a design recommendation to use visual menus under single task settings, and to use auditory menus in a dual task setting when a visually demanding primary task is involved.

Contribution 9. Derivation of a design recommendation where earPod or visual radial menus are recommended for menus with a static structure and a maximum breadth of 12 or less.

1.4 Roadmap

The remainder of this document is organized as follows. Chapter 1 introduces the motivation for the research, reviews the state of the art, and discusses the problems and challenges for mobile interaction. Chapter 2 reviews the past literature. Chapter 3 introduced the *earPod* technique and its design rationales. Chapter 4 presents the first study that was carried out (published at the CHI 2007 conference). Chapter 5 introduces the findings and results for the 3x2 user (second) study. Chapter 6 investigates the learning behavior of *earPod* and a set of related techniques in a longitudinal study (third study). Chapter 7 extends the study of these techniques in dual task simulated driving environment (fourth study). Chapter 8 concludes the thesis with contributions, design recommendations, limitations, and future work.

Chapter 2
Background and Literature Review

2.1 Introduction

While selection in hierarchical menus has been extensively studied, little if any of that research has involved eyes-free menu design; however, the respective research literatures on eyes-free interaction and on hierarchical menu selection provide some guidance for design of eyes-free menu interactions. In this chapter, section 2.2 reviews the relevant literature on menu design; section 2.3 discusses research on eyes-free interactions; and section 2.4 concludes the chapter by summarizing the lessons learned from previous research.

2.2 Menus

Menus are commonplace in today's computer interfaces. Menus facilitate *selection* (indicating objects from a set of alternatives), which is one of the elemental tasks[1] in Graphical User Interface (GUI) (Foley et al. 1984). They are so fundamental that they become an integrated part of today's GUI standard, signified by the letter "M" in WIMP[2].

Menus as a linking construct between users and system commands have three essential properties that distinguish them from command language: *guided interaction*, *recognition-based memorization, straight-forward yet constrained user input* (Norman 1991). These properties help prevent errors that would otherwise be caused by guessing the syntax and spelling of command languages, and users are spared the trouble of having to read documentation so as to learn the terms and syntax for invoking functions. As a result, menus score much higher than command languages in terms of ease of use (Norman 1986).

---

[1] *Text* (entering symbolic data), *select* (indicating objects from a set of alternatives), *position* (pointing to a screen coordinate), and *quantify* (specifying an exact numeric value) are the elemental tasks for GUI, as defined by Foley et al.

[2] WIMP stands for window, icon, menu, and pointing device

The structure of a menu can be as simple as a list of items (single menu), or as complex as a multilevel hierarchy where leaf items are selected by traversing a path through the tree (hierarchical menus). This section focuses on hierarchical (multi-level) menus because: 1) single-level menus are rare in today's increasingly complex systems. 2) Discussion of hierarchical menus will implicitly include the single-level menus nested within them.

Many factors contribute to effective design for hierarchical menus, including: style of presentation, organization of structure and content, learnability, and design for novice and expert behaviors. In the following discussion, Section 2.2.1 describes the two major styles of menu presentation and lists some of their variants; section 2.2.2 reviews strategies to optimize menu hierarchies; section 2.2.3 discusses aspects related to menu learning and approaches to accommodate novice and expert behaviors.

2.2.1 Menu styles

Many menus have been developed for diverse applications and platforms. They can be classified under different systems but the research in this dissertation focused on two contrasting menu styles, i.e., linear style menus vs. radial style menus. *linear style menus* lay out their items linearly (or relative to each other) where the accessing cost to each item is different (Figure 2.1 left); *radial style menus* lay out the items radially in a polar coordinate system where there is a constant distance from each item to the center of the circle in which the menu is embedded (Figure 2.1 right). Items in linear menus are also *relative* to each other in the sense that they have to be traversed sequentially in order to reach the target item. In contrast, items in radial menus have *absolute* locations in the sense that, with sufficient skill and knowledge, users can go directly to the target item without having to traverse through other items on the way. In this dissertation, the linear vs. radial terminology will be used throughout for distinguishing between menu types. However, it should be kept in mind that linear menus are also relative, and that radial menus are also absolute.

Callahan et al. (1988) summarized the pros and cons for each style. Linear style menus are easier for arranging items and they are more flexible in the number of choices allowed in a single menu/submenu. They are also more familiar to users (Sears and Shneiderman 1994). However, because items are lay out sequentially, access time to each item is uneven: depending on the initial placement of the cursor, items closer to the cursor are quicker to select than items further away. Radial style menus, on the other hand, lay out items equal-distance from the center and require constant access time. On average, radial style menus have better performance than linear style menus (Callahan et al. 1988; Kurtenbach and Buxton 1994). However, placing labels in a circular layout requires more space (Figure 2.1 right), and the number of items allowed in one circular array is typically limited to no more than 12 items due to performance concerns (Kurtenbach and Buxton 1993; Zhao and Balakrishnan 2004; Zhao, Agrawala, and Hinckley 2006).

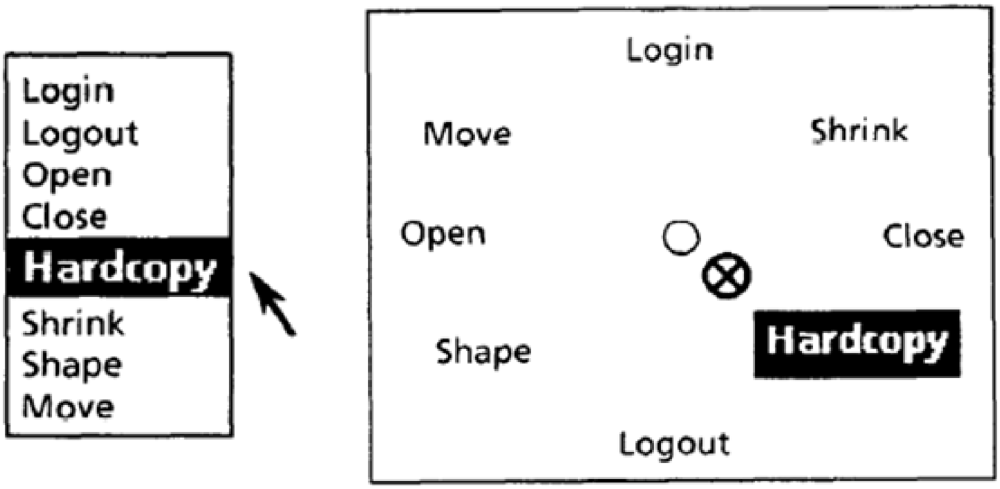


Figure 2.1: A typical linear menu (left), and a crude radial menu (right) (After Callahan et al. 1988).

Variants of linear style menus are commonly used in commercial applications (e.g., the start/system menus in Windows or Apple's series of operating systems, application menus in Microsoft Office, etc.). A considerable amount of research has been carried out on specialized techniques for improving the efficiency of linear menu selections: e.g., distributing items according to their frequency of use and dividing menu item lists into two areas with easier access for high frequency items (Sears and Shneiderman 1994); Cockburn and Gin (2006) introduced a method that made items easier to access by enlarging the active areas in cascading menus; Ahlstroem et al. (Ahlstroem, Alexandrowicz, and Hitz 2006) shortened users' access time by predicting users' intentions and jumping ahead.

Variants of radial style menus are popular in the research community. For example, Pie Menus (Callahan et al. 1988) have been used in many research prototypes (e.g., Lin et al. 2000), as well as some open source applications (e.g., the easy gesture plug-in in Mozilla Firefox). Tracking

Menus (Fitzmaurice et al. 2003) and Trailing Menus (Forlines et al. 2006) are radial style menus suitable for large display applications. Tracking Menus also appear in Alias's product Sketch Book. Flow Menus (Guimbretiere and Winograd 2000) and Control Menus (Pook et al. 2000) combine command selection with parameterization. Compound-stroke Marking Menus (Kurtenbach 1993) is part of the commercial 3D modeling application – Maya. Marking menus are also used in research prototypes such as InkSeine (Hinckley et al. 2007).

Most of the prior research on menu design has focused on visual menus. A comparison of linear and radial styles of menus in the auditory domain had yet to be investigated prior to this dissertation.

2.2.2 Menu architecture: breadth vs. depth tradeoff

Menu designers need to choose the internal hierarchical structure in addition to the overall presentation style. This section introduces related research about choosing the optimal menu architecture.

The number of items allowed in branches (or sub-branches) is the menu *breadth* while the total number of levels within a hierarchy is the menu *depth*. Breadth and depth can be adjusted; for instance, the same menu hierarchy can be either broad and shallow or narrow and deep according to designers' needs.

Researchers are interested in a variety of questions concerning the breadth vs. depth tradeoff, such as the following. Which is more preferable, breadth or depth? Are there any limits in breadth or depth? Are there optimal combinations of breadth and depth, and if so, what are they? For example, a menu hierarchy with 64 terminal nodes may be arranged in the forms of 2^6, 4^3, 8^2, or 64^1; which configuration yields the best performance and is the most preferred by users? Numerous other menu architectures may be compared such as 4x16 vs. 8^2, and 16x4, or 16x32 vs. 32x16. How do modalities of feedback and menu styles affect the breadth vs. depth tradeoff?

Researchers have found that the effect of menu architecture on menu selection performance varies between the auditory and visual modalities. Thus each of these modalities will be considered separately in the following subsections.

2.2.2.1 Visual menus

For visual menus, it is generally believed that breadth is preferred over depth. Landauer and Nachbar (1985) used a simplified model that combines Hick-Hyman Law (Hick 1952; Hyman 1953) and Fitts Law (Fitts 1954) to analyze the relationship between breadth and depth on menu selection using touch screens. They concluded that increasing breadth generally results in better overall performance than increasing depth.

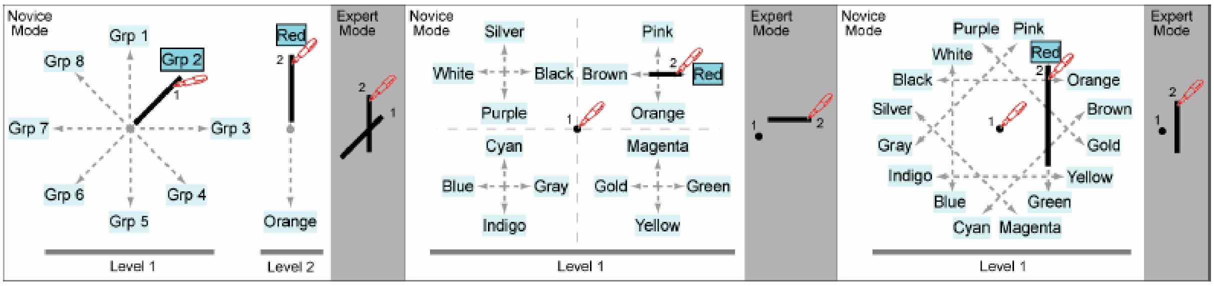
Landauer and Nachbar tested their theory via an experiment with 4096 items arranged in breadth 2, 4, 8, and 16, respectively. Their results showed that a breadth of 16, the maximum breadth they tested, had the best overall performance. Other studies have drawn similar conclusions. Snowberry, Parkinson, and Sisson’s experiment (1983), Schultz and Curran’s study (1986), and Larson and Czerwinsky’s research (1998) all obtained results favoring breadth over depth. Although increasing breadth yields theoretical benefits, maximum breadth is constrained by screen real estate as well as system power. When the hierarchy is large, the limited screen space prevents displaying all items legibly. Similarly, items may take too long to be drawn in a slow system. Under these cases, maximum breadth needs be adapted to the available resources.

The balance of the menu structure appears to be useful in determining the optimal configuration of a menu. Kiger (1984) comparing 8^2, 16x4, and 4x16 menus, finding that a more balanced design, 8^2, was preferable to the less balanced ones (4x16, 16x4). In the case of 16x32 vs. 32x16 structures, when both design are equally balanced, Larson and Czerwinsky (1998) suggested that the quality of categorization may be an important factor in determining the performance of different menu structures.

Sometimes, the breadth vs. depth tradeoff can also be affected by the ability to acquire the target efficiently and accurately. For a radial style menu, increasing the menu breadth reduces the angular space allotted to each item and makes each item more difficult to select. In the case of Compound-stroke Marking Menus (Kurtenbach 1993), it has been found that the breadth for each level is preferably 8 or less. For a breadth-8 menu, *depth* is preferably no more than 2 levels; otherwise, menu selection accuracy drops to below 90%.

The depth limitation can be alleviated by altering the input method for making the selection. Zhao and Balakrishnan (2004) have proposed breaking compound selection strokes into a sequence of inflection-free strokes with pen lifts between each straight line stroke (Figure 2.2 a). Such Multi-stroke Marking Menus allow users to work with breadth-8 menus up to depth-3 (512 items), at an accuracy rate of 93%.

In both cases, the amount of breadth that can be used is limited by users' ability to perform the gesture efficiently and free of errors. By altering the gestures, Polygon Menus (Figure 2.2 c) extend the breadth limit to 16, and Zone Menus (Figure 2.2 b) further pushes the limit to 32 (Zhao et al., 2006). Both types of menu expand the breadth while still providing good speed and accuracy.



(a) Orientation-based menu (b) Zone menu (c) Polygon menu

Figure 2.2: A number of variants of marking menus are illustrated here. a) Orientation-based multi-stroke marking menu, b) Zone menu, c) Polygon menu (After Zhao, Agarwala, and Hinckley 2006)

2.2.2.2 Audio menus

When a visual interface is available, visual menus are typically used. Audio menus are generally used when a visual display is not available, or in an eyes-free setting. Guidelines for visual menu design often cannot be applied to auditory menus since rapid scanning of items "at a glance" is no longer feasible. It has been suggested that the breadth of auditory menus is constrained by working memory (Marics and Engelbeck 1997; Balentine and Morgan 1999; Cohen, Giangola, and Balogh 2004; Schumacher, Hardzinski, and Schwartz 1995).

A number of researchers (e.g., Marics and Engelbeck 1997; Balentine and Morgan 1999; Cohen, Giangola, and Balogh 2004; Schumacher, Hardzinski, and Schwartz 1995) have proposed that the optimal breadth for auditory menu is around 4-5 or less, mainly based on the lower bound of the 7 +/- 2 working memory span (Miller 1956). It is assumed in these analyses that users need to remember all the items in their working memory in order to make an auditory menu selection, and that more than 5 items may over-tax users' memory capability.

However, Commarford et al. (2008) have recently proposed that users do not need to remember all the items in order to make a selection. In their model, instead of remembering all the items, users only need to keep the best matching candidate in the working memory, and choose to hold or discard later choices as they are compared to the best one. Their model suggests that users only need to place 1 to 2 items in their working memory to perform a selection. Based on their analysis, the breadth of auditory menu may be unlimited. However, the experiment carried out only used 11 highly familiar terminal nodes, which were relatively short and which could be "chunked" into higher level units. Further research is needed to investigate the relationship between breadth and depth for auditory menus over a range of menu content types and presentation conditions.

2.2.2.3 Summary

For visual menus, it is generally believed that breadth is preferable over depth. However, no consensus has been made for that of auditory menus. Issues such as the quality of menu categorization, order of the menu items, balance of breadth and depth in the hierarchy, availability of system resources such as screen space and computational power, and nature of the interaction techniques all take part in determining the optimal arrangement of menu items. In addition, general issues such as human perceptual, memory, motor capabilities, application specific issues such as the quality of categorization labels, as well as personal issues such as individual differences are relevant to the relationship between menu architecture and selection performance, but are outside the scope of the dissertation research reported in subsequent chapters.

2.2.3 Menu experience

When users first encounter a new menu technique, selection performance is often slow and cautious. Through time and usage, users gradually learn the menu structure and the acquisition technique, and accumulate confidence. Over time they tend to become more expert in terms of how they interact with the menu. Once they have completely learned the structure of a menu, their efficiency will increase. The following section reviews research about the process relevant to the acquisition of expertise in menu selection. Section 2.2.3.1 discusses the learning curve which can be used to model a wide variety of learning behavior (Newell and Rosenbloom 1981). Second 2.2.3.2 talks about the importance of novice and expert behavior and how can we design interfaces to accommodate these behaviors and provide a smooth transition to expertise.

2.2.3.1 Power law of practice

Mathematically, the power law of practice states that the relationship between a performance measure (e.g., speed) and a measure of experience (e.g., number of trials) is a power function. Its formulas are summarized below.

$$Time = \ MinTime + B * (N + E)^{-\beta} \ (Full\,power\ law\ formula)$$

$$Time = B * N^{-\beta}(Simple\ power\ law\ formula)$$

$$Tim = B * e^{-\alpha N}(Simple\ exponential\ formula)$$

where B is the range of learning, N is the trial number, E is the number of previous practice trials, and α, β are the learning rate parameters (after Ritter and Schooler 2001). The most accurate among these formulas is the full power law formula, but several terms in it are difficult to calculate, and the minimum reaction time (asymptote) is typically only observed after more than 1000 trials (Newell and Rosenbloom 1981); therefore, the simple power law formulas (that do not include the previous practices) are widely used instead as a good enough approximation (Ritter and Schooler 2001).

The power law has been observed in a wide variety of contexts. The implication of the power law is that performance on most tasks improves with practice, and that the rate of improvement starts high and then decreases over time to a point where learning is no longer observed and an asymptote is reached. The learning of almost all HCI tasks (e.g., Wigdor and Balakrishnan 2003; Castellucci and MacKenzie 2008) seems to follow this pattern, and this type of learning curve has also been observed for menus (Parkinson, Sisson, and Snowberry 1985).

With practice, many aspects of a menuing system can be learned, including: association of menu items with commands, learning where a submenu is nested or learning the entire hierarchical structure, association of menu item with response codes, and physical actions required to make responses (Norman 1991). Menus with different learning components tend to exhibit different learning curves (McDonald, Stone, and Liebelt 1983).

One can accelerate the learning process by increasing the mental effort required to use an interface, especially when learning spatial locations. Cockburn et al. (2007) showed that a more effortful interface can help users to learn spatial locations faster.

Various ways can make the interface more effortful to use, such as demanding extra physical movement (e.g., frost browsing, Cockburn et al. 2007), introducing waiting time (e.g., Marking Menu, Kurtenbach 1993). However, efforts need to be carefully controlled since difficulty can induce frustration and discourage usage. Certain techniques, although they increase the overall efforts, have been found to be ineffective in facilitating learning, such as delaying the visual feedback of menu appearance for learning the hotkeys (Grossman, Dragicevic, and Balakrishnan 2007).

Real world learning curves are not always smooth and continuous. Most users typically face many tasks, and usage of a particular technique may be inconsistent. In such cases, the recorded learning curves may be interrupted. For instance, Kurtenbach and Buxton (1994) observed that users often fall back to novice behavior after a period of interruption.

2.2.3.2 Design for novice and expert behaviors

Ideally, user interfaces need to be both easy to use for novice users and efficient for expert users, and they should provide a quick and smooth transition between the two types of behavior. However, in reality, there may sometimes be a tradeoff between simplicity and power (Nielsen 1992).

One solution for accommodating both experts and novices is to provide alternative strategies for novice and expert behaviors (Shneiderman 2004). An example of this strategy is demonstrated by menus and hotkeys. Menus are easy to use and novice friendly, but less efficient; hotkeys, on the other hand, are fast to access and suitable for expert users. However, since alternative strategies can be quite different, if not designed carefully, transition from novice to expert behavior can be difficult. Because hotkeys are designed to use completely different physical actions on a different input device, they are often neglected and not remembered by users (Grossman, Dragicevic, and Balakrishnan 2007).

One exception that combines ease of use with efficient expert access is demonstrated by Compound-stroke Marking Menus (Kurtenbach and Buxton 1993; Kurtenbach and Buxton 1994; Kurtenbach, Sellen, and Buxton 1993), a variant of pie or radial menu. Compound-stroke Marking Menus integrate the novice and expert behaviors into a single system. Menu selection can be made by either selecting an item from a popup radial menu, or by making a straight mark in the direction of the desired menu item without popping-up the menu. Novice users unfamiliar with the menu structure can see choices via the popup radial menu after dwelling and holding the pen for 1/3 of a second. As users gain more experience, they can select the menu items using marks only, without waiting for the menu to popup, and significantly increase the speed of selection (up to 3.5 times faster than both linear or pie menu). Compound-stroke Marking menus support hierarchies where novice users select items like a hierarchical pie menu and expert users make “zig-zag” compound marks to select from multiple levels of submenus (Figure 2.3).

Instead of using two different approaches, the fast and efficient expert behavior of Compound-stroke Marking Menu uses the same motion as is used in novice behavior. Each time a menu item is selected in the novice mode, the physical motion for selecting the same item using expert behavior is rehearsed. As a result, the novice-to-expert transition happens seamlessly.

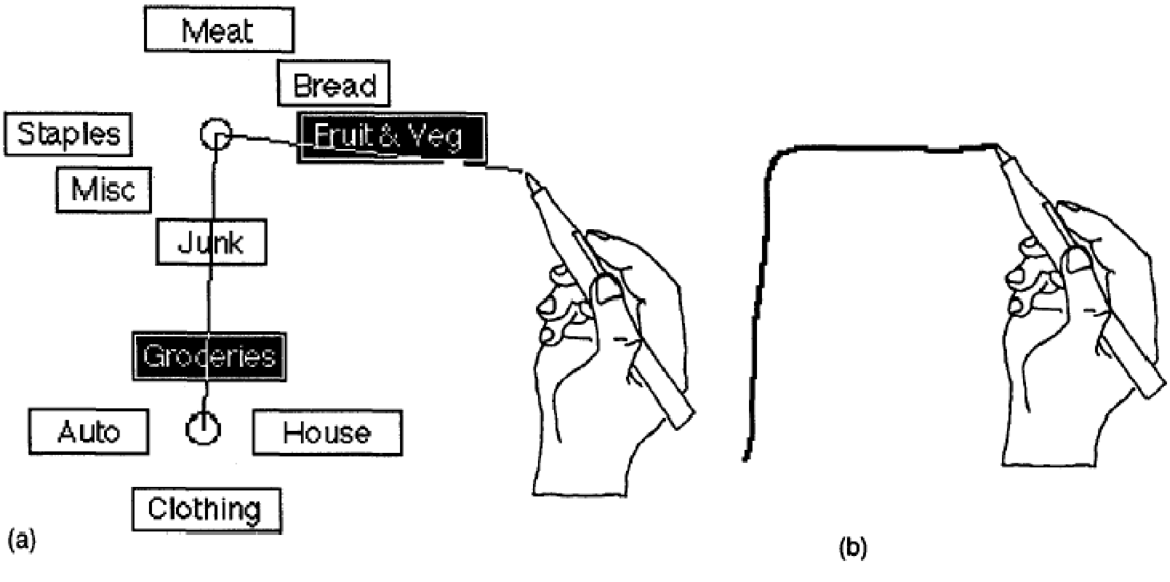


Figure 2.3: Demonstration of the novice and expert behavior in Compound-stroke Marking Menu a) select a second level item “Fruit & Veg” using a popup hierarchical radial menu. b) the same selection using a mark. (After Kurtenbach and Buxton 1993)

A similar design principle has been adopted in Multi-stroke Marking Menus (Zhao and Balakrishnan 2004), Zone and Polygon Menus (Zhao, Agrawala, and Hinckley 2006), Wave Menus (Bailly, Lecolinet, and Nigay 2007), Flower Menu (Bailly, Lecolinet, and Nigay 2008) and ShapeWriter (Kristensson and Zhai 2004).

However, it can be overly restrictive to force the novice and expert behavior to use the same actions. As long as novice and expert behavior is carefully designed, and a smooth and seamless transition is provided between those two types of behavior, using two separate strategies can be effective, as demonstrated later in this dissertation.

The above section summarized the related research in menu design. In the following section, I will discuss relevant research in eyes-free interaction.

2.3 Eyes-free interaction

Eyes-free interaction refers to interaction with computing devices/interfaces without visual attention. The research literature in non-visual input and output is reviewed in section 2.3.1, after which eyes-free applications are discussed in section 2.3.2.

2.3.1 Eyes-free I/O

While research in non-visual input and output is a large topic and covers numerous possibilities, this dissertation chooses to emphasize the specific combination of touch-based gesture input with auditory output. Combining gesture input with auditory output is a relatively new research area; nevertheless, initial exploration by Pirhonen et al. (Pirhonen, Brewster, and Holguin 2002) has shown its effectiveness and potential. In this section, relevant literature on this topic is reviewed. Section 2.3.1.1 discusses gesture input; section 2.3.1.2 talks about touch input; section 2.3.1.3 reviews transfer functions; and section 2.3.1.4 covers auditory output.

2.3.1.1 Gesture input

Internal sensations of body posture, motion, and muscle tension (Burdea 1996; Gibson 1962) allow users to sense how they are moving using kinesthetic feedback (i.e., direct awareness of motions being performed). Consequently, some gestures can be reliably performed without visual feedback after practice. Examples of such gestures include Unistroke -- a stylus-based touch typing technique (Goldberg and Richardson 1993), Marking Menus and its variants (Kurtenbach 1993; Zhao and Balakrishnan 2004; Zhao, Agrawala, and Hinckley 2006; Oakley and Park 2007) , EdgeWrite (Wobbrock et al. 2003), and ShapeWriter (Kristensson and Zhai 2004).

However, since gestures are not self-revealing (Hinckley et al. 2005; Kurtenbach and Buxton 1991; Hinckley et al. 2007) and need recognition, computing systems are required to provide feedback in order to teach users about their existence and train users to perform them correctly. All previously mentioned techniques use visual feedback to inform and guide users; therefore, these gesture techniques are not “purely” eyes-free. Eyes-free use only happens when users have completely learned these gestures. One reason that most interfaces train users visually is because visual feedback is often more effective in providing spatial information as compared to audio (Cohen et al. 1989). It has been shown that relatively complex gestures such as signatures can be difficult to learn without visual guidance (Plimmer et al. 2008). Nevertheless users are capable of performing certain gestures using only non-speech audio (Pirhonen, Brewster, and Holguin 2002). Thus true eyes-free interaction using gesture input is feasible and will be utilized in the research reported later in this dissertation.

2.3.1.2 Touch-sensitive Input

Gestures can be modeled using position, velocity, and acceleration information of their trajectories (Hong and Landay 2000). Reliable recognition of gestures often requires robust sensing techniques. While techniques such as computer vision (e.g., Cao and Balakrishnan 2003) have shown great potential, manual input devices are often preferred due to their availability, cost, and robustness. Commonly used position and motion-sensing devices include mice, trackballs, joysticks, and touch-sensitive devices (Hinckley 2001).

Although mice, trackballs, and joysticks each have their advantages in specific contexts, they are relative pointing devices that can only sense motion and acceleration, and this limits their potential to recognize gestures. Touch-sensitive devices (tablets or touch screens), on the other hand, support both relative and absolute pointing (Hinckley 2001), enabling them to be used in creating a variety of gestures. In the following section, properties and related research of touch-sensing devices are reviewed.

Touch-sensitive devices can be defined as horizontal or near horizontally positioned flat digital surfaces that can sense the position of fingers. A number of properties make them particularly attractive for eyes-free interaction (Buxton et al. 1985).

- They require no intermediate devices such as pucks or stylus to operate, which simplifies interaction;
- Their operation is motion and vibration resistant, which is ideal for all type of stable or dynamic environments;
- They offer multi-touch capability, which allows them to support complex gestures;
- They are light-weight and can be easily integrated with smooth-surface devices.

The ability to sense both position and motion allows touch-sensitive devices to support both continuous gliding gestures and discrete clicks via tapping (Zhao et al. 2007). Touch sensing devices can be combined with other input devices. Hinckley and Sinclair (1999) combine touch-sensing capability with mice and keyboard to create touch-sensing input devices that are capable of a variety of new tasks. A similar approach has also been demonstrated by Rekimoto et al. (2003) in their PreSense technique. Touchpad can be modified to support active tactile feedback (Enns and MacKenzie 1998), which can be potentially useful to complement or even replace audio feedback for eyes-free operations.

Touch sensing has been used in many mobile research prototypes. Innovative usage of touchpad including placed them into the back of cell phones in the Behind Touch prototype (Shigeo, Isshin, and Kiyoshi 2003), and to the back of a bendable computer for 2D positioning by Schwesig and colleagues (Schwesig, Poupyrev, and Mori 2004). Touch sensing devices can also be used on both the front and back side of a device, such as the hybrid touch system (Sugimoto and Hiroki 2006)

Touch-sensing devices have also been widely used in commercial systems. Motorola has begun to support touch-based gestures with the capacitive phone keypads in the Motorola A668 phone (http://direct.motorola.com). Synaptics' also demonstrated similar features in their MobileTouch technologies (http://www.synaptics.com). As mentioned earlier, ClickWheel is a hybrid keypad/touchpad input device that has become an integrated part of the iPod success story (Figure 2.4).

Figure 2.4: Apple's ClickWheel is a hybrid input device that combines touchpad with keypad.

Despite its advantages, a touch-sensitive tablet has the limitation that clicking via tapping or double tapping can be accidentally triggered (MacKenzie and Oniszczak 1997). Like trackballs, the small size of touchpads necessitates frequent clutching, and touchpads can be awkward to use while holding down a button, unless the user employs his/her other hand (Hinckley 2001). However, the many unique advantages of touch sensitive input devices counteract some of these issues, making those devices promising platforms for implementing eyes-free interfaces.

2.3.1.3 Transfer functions

Regardless of which input devices are used, the data directly obtained from those devices is typically transformed in some way so that the device can be more reliably and intuitively controlled (Hinckley 2001). The dynamic relation between displacement of a control device and the resulting behavior of the system being controlled can be described as a transformation functions, or order of control. As described in detail by Jagacinski and Flach (2002), the control order refers to the number of integrations between the control input to a device and the resulting output of the device. Depending on the number of integrations between the input and output, the system can be classified as zero-order, first-order, second-order, or higher orders. Most input devices use zero-order, or first-order control, which are typically referred to as position control and rate control, respectively.

Position control directly maps the movement of the human operator to the movement of the object. Rate control, on the other hand, maps human action to the velocity of the resulting object movement. In most tracking tasks, position control and rate control are typically recommended in preference to higher order control (such as acceleration control) (Zhai and Milgram 1998).

Position control is generally regarded as more isomorphic, or direct, and more "intuitive" for humans to operate. The majority of studies comparing position and rate control have concluded that position control is superior. However, rate control provides the advantages of better filtering of involuntary noise, smoother movement, and an unlimited operating range (Zhai 1995).

Position control is often the preferred transfer function when circumstances permit. However, when facing footprint constraints, or large data size requirements, or when it becomes difficult to map application parameters directly to sensory properties, then a rate control technique may be preferred.

2.3.1.4 Auditory output

In addition to input techniques, interfaces need some form of output feedback to guide and inform users. Haptic (or tactile) output is one possibility. Haptic output is by nature a “private” display and can communicate information even in noisy environments (Wagner et al. 1999; Luk et al. 2006). However, most users are not familiar with haptic-based languages such as the Braille alphabets, making it difficult to act as an independent output modality.

Auditory output, on the other hand, may utilize both speech and non-speech audio, allowing it to communicate a rich set of information to users. Sound can travel in space and is ommidirectional, making it particularly suitable for delivering important messages like alarms and alerts.

Despite having these advantages, sound also has a number of intrinsic properties that raise difficulties for using it as an independent feedback modality in computer interfaces as noted earlier in Chapter 1 of this dissertation. In particular, the serial and temporal nature of information presentation makes audio-based interfaces reliant on, and constrained by, human short-term memory capabilities. Short term memory is limited both in capacity and duration (G Miller 1956; L. Peterson and M. Peterson 1959). To avoid rapid forgetting, auditory messages need to be constantly rehearsed (Klatzky 1980). Moreover, sound can be annoying since we cannot close our ears in the same way we can close our eyes (Gaver 1997).

While screen real estate is the constraining factor for visual interfaces, for auditory interfaces, time is a limiting factor (Stifelman et al. 1993) because audio messages take time to play back. Various strategies can be used to increase the time efficiency of audio messages. Two of them, interruption and substitution, will be discussed below.

2.3.1.4.1 Interruption

Interruption can be applied at sentence or paragraph level, for partial sentences or paragraphs often provide enough information for users to decide whether or not to keep listening. In Interactive Voice Response systems (IVR), this feature is called *barge-in* (or "*cut-through*") (Mane et al. 1996). Since audio messages are typically long, barge-in saves time by allowing experienced users to interrupt the current message and proceed to the next. Barge-in requires systems to keep listening to users' input even they are busy. Most IVR systems (Walker et al. 1998; Yankelovich, Levow, and Marx 1995) respond to either key presses or speech interruptions.

Interruption can also be applied at the word level. This is possible because humans often infer words from incremental phonetic information. It has been shown that listeners can identify words without listening to the entire word (Marslen-Wilson 1987; McClelland and Elman 2002). Words with more distinctive phonetics in the first few syllables are easier to distinguish and have a higher probability to be understood when they are partially heard.

The ability for humans to understand partial sentences or words has implications in designing both the content and the interaction for auditory interfaces. When composing auditory messages, it is beneficial to place the distinctive and important words earlier in a sentence. At the word level, preference should be given to phonetically more distinctive words. When designing interaction techniques, it is important to allow interruption at both the sentence/paragraph level and the word level.

2.3.1.4.2 Substitution

Another time reduction strategy is substitution: using shorter auditory messages to replace longer ones. A common strategy involves using much shorter non-speech audio segments to represent equivalent messages using speech (e.g., Brewster 2003). This strategy is similar to the use of space efficient icons to represent labels in graphical interfaces. This approach was first introduced by Gaver and colleagues (Gaver and Smith 1991; Gaver 1989), who designed an auditory system called Sonic Finder for the Apple Macintosh computer using everyday sound to represent objects, tasks, and events. These symbolic sounds, which are analogous to what they represent, are called auditory icons. Since their introduction, auditory icons have been widely used. However, one problem with auditory icons is that it can be difficult to find suitable iconic sounds for events in an interface since they might not correspond to a sound-producing event in the real world that can be easily mapped to corresponding interface actions or states.

To overcome this limitation, Blattner et al. developed earcons (Blattner, Sumikawa, and Greenberg 1989), and the earcon approach was later extended by Brewster et al (1993). Earcons are auditory messages consisting of music or synthetic sounds that can provide navigational cues. Earcons were found to help navigation even for hierarchies (those with twenty-seven nodes) using auditory interfaces, and could be effective in a lower-quality audio environment (Brewster 1998), as well as for PDAs in the mobile environment (Brewster and Cryer 1999).

However, by themselves, auditory icons or earcons are limited in the amount of information they can convey to users. They are typically intended as information that is complementary to a primary source of information presented in written language, or as spoken words. In an audio menu system, non-speech audio may be used together with speech to balance understandability and efficiency for the system.

2.3.1.4.3 Spatialization

In addition to increase time efficiency, spatialization is another property of sound that can be used to enhance the effectiveness of auditory based interfaces.

The fact that humans have two ears allows them to detect the direction and location of auditory signals to varying extents (based on a variety of factors including the hearing ability of the person and the type of sound). Spatialization can be simply achieved by manipulating mono voice streams to incorporate Interaural Time Differences (ITDs) and Interaural Intensity Differences (IIDs), the two major binaural cues for localizing sounds on the left-right axis (Bernstein 1997). Our ability to locate sounds on the two other axes uses more complex mechanisms and supporting it would require sophisticated signal filtering methods (such as Head Related Transfer Functions) as well as individual calibration (Bernstein 1997).

Audio as a feedback channel has been used widely prior to this writing. However, audio has a number of properties that makes it as a sole source of feedback challenging (Gaver 1997). Auditory interfaces need to carefully manage the duration of their messages, for time is a scarce commodity. In the following section I consider applications that primarily use auditory feedback.

2.3.2 Eyes-free applications

Audio-based interfaces are predominantly used in two niche areas: interfaces for the visually impaired (section 2.3.2.1) and applications for mobile computing (section 2.3.2.2). Each area is briefly reviewed below.

2.3.2.1 Interfaces for the visually impaired

Research has long been carried out for creating auditory interfaces for the blind. These interfaces aim to help blind people with a number of tasks such as the Soundtrack system (Edwards 1989), which assists blind users in editing, and Emacsspeak (Raman 1996), which uses speech to provide content and feedback to support use of the EMACS editor. Other research has addressed the problem of communicating graphical information to the blind. Kennel developed the AUDIOGRAF system (1996) to display 2D graphs using music. IC2D (Kamel and Landay 2002) is a grid-based auditory interface to allow the visual impaired to perform drawing tasks. Other systems were designed to assist blind people to navigate through Windows systems. For instance, the Mercator project (Mynatt and Weber 1994) used a hierarchical tree to organize the desktop widgets, and allow blind users to navigate the widgets without any spatial pointing devices.

Commercial screen readers such as Jaws[3] and Windows Eyes[4] combine window navigation, web browsing, and content reading to allow blind users to experience most features out of a computer.

Audio feedback can be combined with haptic output since simple haptic constrains such as raised edges or tactile landmarks can help people to orient in a fixed space (e.g., EdgeWrite). The Textual and Graphical User Interfaces for Blind People (GUIB) project directly mapped the two-

---

[3] http://www.freedomscientific.com/fs_products/software_jaws.asp

[4] http://www.gwmicro.com/Window-Eyes/

dimensional graphic user interface into new input/output devices using spatial audio and a haptic tablet (Mynatt and Weber 1994). Rassmus-Grohn et al. (2007) designed an audio-haptic drawing program for the visually impaired. “McSig” (Plimmer et al. 2008) is a multimodal environment that combines haptic and audio output to teach visually impaired on how to handwrite characters.

Accessible interfaces are typically built to “enable” the visually impaired to use functionalities and features available for sighted people. They are designed for a specific user group and often not intended for the general population unlike the mobile audio-based interfaces as described below.

2.3.2.2 Mobile audio applications

Recently, mobile computing has become increasingly popular. Mobile computing may be defined as “using computers while moving or being moved”[5] or “using a computing device while in transit”[6]. Many things change when we are moving in a dynamic environment; instead of having dedicated attention, we have to distribute our cognitive and physical resources among a number of time-sharing tasks (Pascoe, Ryan, and Morse 2000; Oulasvirta et al. 2005). Visual attention is often not available to operate the computing devices while mobile. Audio-based eyes-free interaction becomes attractive in mobile environments. Multiple Resource Theory (Wickens 2002) suggests that using auditory output for the secondary task may alleviate interference in a dual task setting where the primary task is visually demanding.

Much of the earlier work in this area has considered task-specific audio-based mobile applications. VoiceNotes used a hand-held portable audio note-taking device to capture and organize short spoken notes. NewsComm, designed by Roy and Schmandt (1996), allowed users to select and navigate structured audio in a mobile environment. In NewsComm, annotations, created using automatic pitch analysis on the data, were added to the structured audio for easier navigation and selection. Although primarily audio-based, these systems did not allow 100% eyes-free interactions. Operating both VoiceNote and NewsComm required some kind of visual feedback and was not strictly eyes-free.

---

[5] by Merriam Webster Dictionary

[6] by PC Magazine, http://www.pcmag.com/encyclopedia_term/0,2542,t=mobile+computing&i=47137,00.asp

Korhonen (2005) developed a cell phone dialing technique that used both speech and non-speech audio to aid users’ interaction with the buttons on mobile phones. Their technique allowed complete eyes-free operation in dialing phone numbers. Although system-wise, not all the operations were eyes-free, they certainly moved a step closer to realizing complete eyes-free operation.

A number of mobile auditory applications that are completely eyes-free have been developed. For instance, Nomadic radio (Sawhney and Schmandt, 1999) is an audio-only wearable computing device for the mobile environment. Nomadic Radio uses speech input and spatial auditory output. The highlight of the system is its seven level notification system based on the priority of the message, and its context awareness ability. However, instead of using a menu, it uses a speech based dialog model to retrieve information.

Pirhonen et al. (2002) investigated the use of simple gestures and audio-only feedback to play music in PDAs. They developed a number of intuitive gestures to control a set of basic music playing functions: play, pause, volume up, volume down, and skip. Through empirical studies, they demonstrated that when gesture input is combined with auditory feedback, users can successfully control music players eyes-free.

BlindSight (Li et al. 2008) is an auditory calendar lookup interface for the mobile phones. By replacing visual feedback with audio, BlindSight allows users to browse their calendar without interrupting the ongoing conversation.

Eyes-free interfaces should be particularly useful in mobile environments. However, most of the current solutions are application and scenario-specific. In order to build more complex and all-encompassing auditory interfaces, general purpose interface components, such as menus, need to be investigated and developed. In the next section, several attempts to create auditory menus are summarized.

2.3.3 Eyes-free menus

Compared to visual menus, audio-based menus are rare. Some menus may be primarily auditory; however, using them is not completely eyes-free. One example of such a menu is the Interactive

Voice Response System (IVR). Although IVR is widely used, designing efficient IVR systems has long been recognized as a difficult problem (Marics and Engelbeck 1997; Roberts and Engelbeck 1989).

In addition to the problems associated with sound in general, IVR systems follow a *sequential* or *asynchronous dialog* model. A sequential dialog model is useful when the user needs assistance for a complex task. However, it is not appropriate for simple and repetitive tasks such as menu selection and is particularly irritating when combined with speech output. Since speech is slow, the user may spend considerable time listening to system prompts. After users press a key, they often wait several seconds to hear the entire audio feedback before initiating the next action. This has led to an unpleasant user experience for IVR menuing systems, commonly described as “touchtone hell” (Yin and Zhai 2006).

Various improvements to IVR menus have been proposed to ease frustration (Brewster 1998; Marics and Engelbeck, 1997; Paap and Cooke, 1997; Resnick and Virzi, 1992; Suhm et al., 2001), but despite these improvements, conventional auditory menus are still harder to use than visual menus.

Two completely eyes-free menuing systems have been developed. The Bullseye (Friedlander et al. 1998) is designed for the visually impaired. Bullseye menu consists of a series of concentric circles divided into sectors. Selection is determined by the direction and distance of the stroke performed by the user (Figure 2.5).

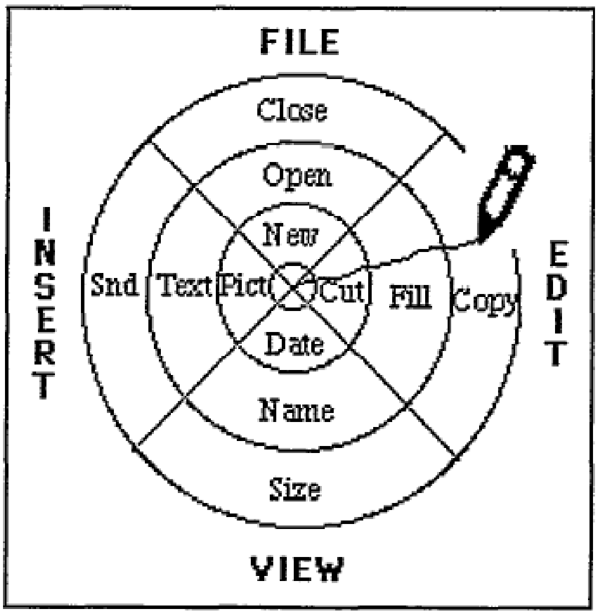


Figure 2.5: The Bullseye menu is a series of concentric circles divided into sectors. Auditory feedback is provided when user moves the cursor. (After Friedlander et al. 1998)

Using a Bullseye menu is much slower than using visual menus. In addition, Bullseye menus require a large input space if the breadth of the menu is large. However, the idea of spatially laying out the items in a polar coordinate system on a touch-sensitive surface is promising and highly relevant to the research reported in this dissertation.

Brewster et al (2003) developed another eyes-free menu technique that used multimodal interaction to allow users to select items in mobile situations. Their system used a 3D audio radial pie menu with head gestures being used to selecting items (Figure 2.6). This arrangement was found to reduce task completion time and perceived annoyance, and to lessen the impact of device interactions or walking speed. Their design frees both users' visual attention and hand operation for interacting with computers. Their study showed the effectiveness of their technique in mobile situations such as walking.

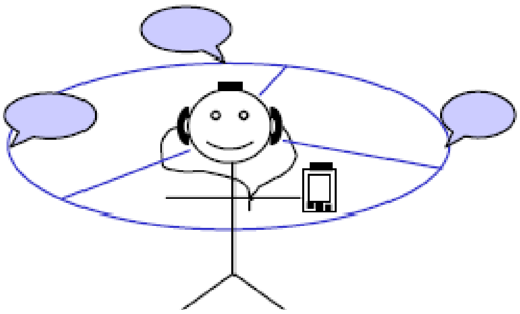


Figure 2.6: An illustration of the 3D auditory pie menu. (After Brewster et al. 2003)

However, head gestures are awkward to use in real life. It is not natural to see users nod and shake their heads seemingly at random when using this type of interface on the street. In addition, the head gesture menu created by Brewster et al. used only four options, which is insufficient for the wide range of functionality that exists in today's devices.

Despite the previous research that has been carried out, designing efficient audio-based hierarchical menu remains a difficult task, and the need to address this problem motivates the research carried out in this dissertation.

2.4 Summary

This chapter summarized the relevant literature for effective menu design, reviewed the various techniques for eyes-free I/O, and discussed the areas where eyes-free menus and applications are typically used. Several observations are made. Eyes-free menus (or interfaces) intended for people with normal vision instead of visually impaired are rare. Many so called "eyes-free" interfaces are not "purely" eyes-free. They often cannot be operated eyes-free interactions from the beginning. However, there are many mobile scenarios where eyes-free interfaces are attractive for people with normal vision. There appears to be considerable potential for efficient and scalable hierarchical eyes-free menus[7].

Designing eyes-free menu selection methods is challenging. The remainder of this thesis focuses on an innovative hierarchical auditory eyes-free menu technique called *earPod* that was

---

[7] Such menus, although not designed for the blind users, do not prevent them from using it and allow users to completely free the visual channel for other tasks.

developed in this dissertation. It is designed to take advantage of lessons learned from this literature review and to overcome many of the above-mentioned challenges of auditory interfaces and menus.

Chapter 3
earPod: Efficient Hierarchical Eyes-free Menu Selection

3.1 Introduction

As previously mentioned, time is an important limiting factor for auditory interfaces. Dealing with the serial and temporal nature of sound is a major challenge for auditory-based menus. *earPod* aims to improve efficiency by incorporating a number of design features such as gesture-based touch input, reactive, spatialized, and interruptible audio feedback, a combination of speech and non-speech audio, sequential and direct access to menu items, and a seamless path for transition to expert behavior. This chapter first provides a walkthrough for the *earPod* technique, a discussion of the associated design rationales, and a description of the *earPod* implementation in section 3.2.and a discussion of related issues on mobile usage of *earPod* in section 3.3, follows by a summary in section 3.4.

3.2 The *earPod* technique

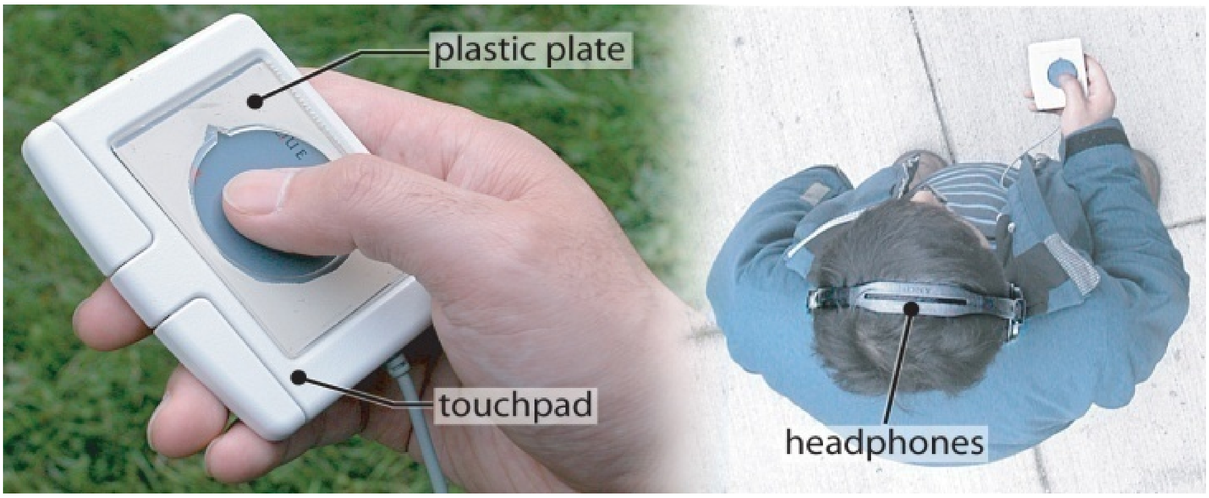


Figure 3.1: *earPod* prototype uses a headset and a modified touchpad

earPod is an eyes-free menu technique using touch input and reactive auditory feedback. The *earPod* technique is designed for an auditory device controlled by a circular touchpad whose output is heard via a headset (Figure 3.1 left), as is found, for example, on an Apple iPod shows how the touchpad area is functionally divided into an inner disc and an outer track called the *dial*. The dial is divided evenly into sectors, similar to a Pie (Callahan et al., 1988) or Marking Menu (Zhao and Balakrishnan, 2004; Kurtenbach, 1993; Zhao et al., 2006).

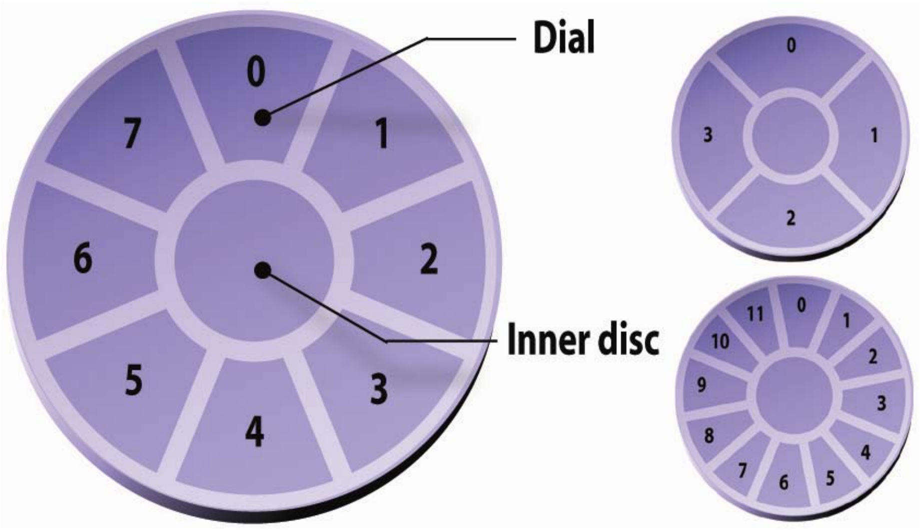


Figure 3.2: The functional areas of *earPod*'s touchpad. Up to 12 menu items can be mapped to the track. The inner disc is used for canceling a selection.

The interactive sequence that occurs when using *earPod* is illustrated in Figure 3.3. The user starts by touching the dial, and the audio menu then responds by saying the name of the menu item located under the finger (Figure 3.3 a). Users may continue to press their finger on the touch surface, or initiate an exploratory gesture on the dial (Figure 3.3 b). Whenever the finger enters a new sector on the dial, playback of the previous menu item is aborted. Boundary crossing is reinforced by a click sound, after which the auditory feedback for the new menu item is played. Once a desired menu item has been reached, users select it by lifting the operating finger, which is confirmed by a "camera-shutter" sound (Figure 3.3 c). Users can abort item selections by moving their finger to the center of the touchpad and releasing it. If a selected item has submenus, users repeat the above process to drill down the hierarchy, until they reach a desired menu (leaf) item. Users can skip items rapidly using fast dialing gestures (Figure 3.3 d).

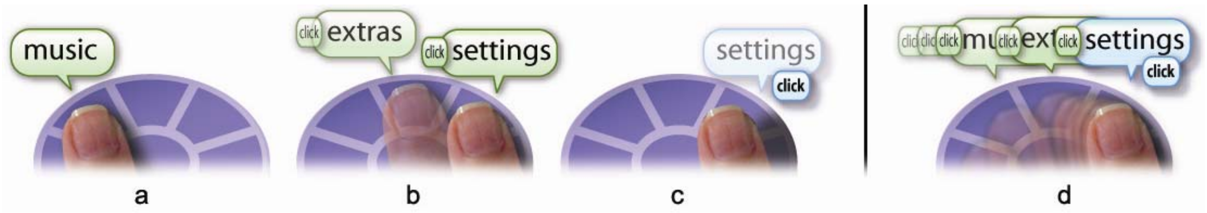


Figure 3.3: Using *earPod*. (a, b) Sliding the thumb on the circular touchpad allows discovery of menu items; (c) the desired item is selected by lifting the thumb; (d) faster finger motions cause partial playback of audio.

earPod is designed to allow fast expert usage. As users gain knowledge of the menu configuration through practice, they tend to use brief corrective gestures (Figure 3.3 b) instead of large exploratory ones (Figure 3.3 d). Eventually, as users remember the exact locations of desired menu items, they select these items by directly tapping on them. *earPod* is motivated by the fact that existing audio menu selection methods (such as implemented in IVR systems) are inconvenient, not only due to the serial and temporal nature of sound, but also because of the sequential dialog between the system and users that is required.

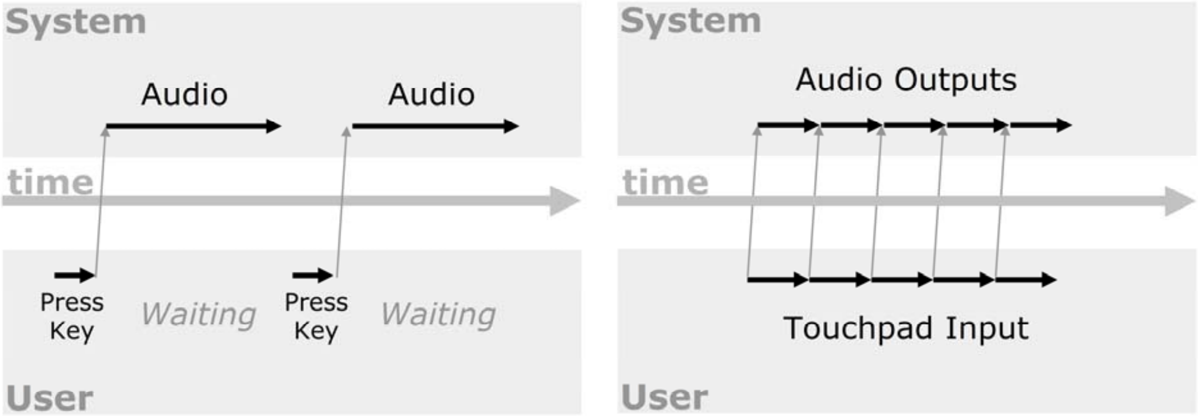


Figure 3.4: (left) The asynchronous interactive model of standard voice menus (IVR). (right) The synchronous interactive model of *earPod*.

3.2.1 Lessons learned from previous design

The limitations of the audio modality compared to the visual modality are well understood. The visual modality allows a list of choices to be displayed instantaneously and persistently. As a result, users can easily scan and compare visual menu items back and forth at their own pace, without having to commit any to memory (Yin and Zhai 2006). In contrast, using audio to convey a list of choices requires users to adapt to an imposed rate of presentation and to rely on their short-term memory. The presentation rate will often be either too slow (e.g., when information is irrelevant) or too fast (e.g., when information is critical) for the user.

Many of today's IVR systems allow quick access to specific menu items provided that the user knows the exact code. However, because recovering from errors in sequential dialog systems is a costly process, using these features requires a great deal of self-confidence. As a result, most intermediate users will listen to the whole list of options repetitively rather than take the risk of hitting the wrong key.

Traditional IVR systems typically suggest a list of options, and then prompt the user to choose an option (typically by pressing a number or other key). They follow a *sequential* or *asynchronous dialog* model (Figure 3.4, left). A sequential dialog model is useful when the user needs assistance for a complex task. However, it is not appropriate for simple and repetitive tasks such as menu selection, and is particularly irritating when combined with speech output. The user often has to wait, with little or no control over the system. As Figure 3.4 (left) illustrates, after users press a key, they often wait several seconds to hear the entire audio feedback before initiating the next action.

Berry (1997) contrasted the asynchronous interaction paradigm with the notion of a *reactive system*. In a reactive system, the user always has the initiative and never waits. Ideally, the system waits for user actions and reacts promptly whenever they occur. Such an approach has been widely adopted in modern direct manipulation interfaces. This approach gives the user a positive *feeling of control* (Shneiderman 2004). In addition, *proactive discovery* as well as *progressive acquisition of expert knowledge* are emphasized (Kurtenbach 1993; Zhao, Agrawala, and Hinckley 2006; Zhao and Balakrishnan 2004; Shneiderman 2004).

It seems that eyes-free mobile menu selection can benefit from reactive audio feedback to improve the efficiency of the selection technique, and to enhance user satisfaction.

3.2.2 *earPod* approach

In order to address the difficulties with IVR menu selection noted earlier, a number of design modifications were made.

Touch input: A touchpad was used instead of a keypad for input. Touchpads arguably have a richer input vocabulary than keypads because they allow gliding gestures in addition to discrete taps. These gliding gestures allow browsing of menu items before confirming the selection.

Karlson et al (2006) reported the result of thumb usage on the touch screens of PDA and Cell phones, and found that movement in the NW-SE directions is difficult for right-handed users. The “sweet spot” of thumb movement is located in the center of the device. The corners of the devices are either “too close” or too “too far” to allow comfortable movement.

Based on these findings, it seems reasonable for the touch input area to have a circular shape and to be placed in the center of the graspable portion of the device, so it can be easily reached by the thumb. Such a design approach is very similar to the iPods’ ClickWheel.

During exploration, the user’s finger is guided by the raised edges of the circular-shaped touchpad. Unlike a keypad with a fixed number of physical buttons, a touchpad allows flexible division of input area into arbitrary numbers of subsections (Figure 3.2), which can then be assigned to different commands. In addition, the circular touchpad used is compatible with existing devices such as the iPod, potentially allowing our technique to be installed simply as a software update.

Reactivity: instead of relying on an asynchronous communication model a synchronous approach is adopted in which the system only reacts to the user’s actions (Figure 3.4, right). Menu items are mapped to physical areas on the touchpad and are played (as a voice cue) upon invocation. Thus users have the ability to proactively discover available options at their own pace. When the information is not of interest, they can skip to an adjacent area. Alternatively, they can listen to the entire message and repeat it if needed. By moving the finger back and forth, the user achieves the effect of "scanning and comparing" the menu items without having to commit them to memory. Such benefits were previously only provided by visual menus.

Interruptible audio: following the reactive interaction model, an auditory item is played back as soon as the finger reaches its location, regardless of whether the previous item has finished playing or not. During early evaluation of the design concept, simultaneous playback (all items are played back entirely), interrupted playback (each new playback stops the previous one) and mixed approaches (previous playbacks fade out) were tested. Preliminary tests indicated that simultaneous playback was confusing for users. To provide a much stronger feeling of reactivity, interrupted playback was adopted.

It was found to be especially useful to be able to promptly switch to a new item before the previous one finishes playing because users often understand partial audio messages[8]. Moreover, the amount of information needed to identify an option can be further reduced as the user gains more information about the menu options. This allows the user to find an item faster than if all of the menu items had to be fully presented.

Non-speech audio: in addition to speech playback of menu items, non-speech audio was used to provide rapid navigational cues to the user. Non-speech audio has been shown to be effective in enhancing the graphical user interface (Gaver 1989; Gaver and Smith 1991), and in improving navigation of non-visual information on mobile (Brewster and Cryer 1999; Sawhney and Schmandt 2000) and IVR systems (Brewster 1998). Short mechanical click sounds are played each time a boundary is crossed on the touchpad, in a way similar to the iPod's ClickWheel. These sounds are very helpful for separating the playback of new items from the playback of previous items. And even when the finger slides too fast for the speech audio to be heard, this mechanical sound gives a rough idea of the number of items or boundaries crossed. A "camera-shutter" sound is also used to confirm item selection.

Direct item access: even though the content of the entire menu can be played back sequentially, items can also be accessed directly. This is inspired by the design of Pie or Marking Menus, which lay out items radially. All items can thus be theoretically accessed in an equal amount of time. In the system developed in this research, direct access can also be achieved by simply tapping the touchpad at the appropriate location whenever the user remembers an item's location. Such support for direct invocation is particularly beneficial in audio menus due to the slow rate of speech output. It allows users to completely bypass the audio playback, saving time.

Transition to expert use: in contrast to IVR systems, the *earPod* interface provides a smooth transition from novice to expert in a way similar to Marking Menus. Each time users select an

---

[8] For example, if the user is looking for the item "apple", "banana" can be rejected as soon as the syllable "ba" is heard.

item using a dial gesture, they gain experience which should facilitate their transition to expert usage. In the beginning, the user tends to glide a longer distance to reach the desirable item (Figure 3.3 d). As the user learns the absolute position of the target, the navigation path on the touchpad will be shortened (Figure 3.3 b) and eventually approach direct invocation by tapping (Figure 3.3 c). To distinguish between novice and expert usage, Marking Menus are typically implemented with a ~300 ms time-out before transitioning from marking to popup menu mode. However, the length of the timeout can change with different systems (Liao, Guimbretiere, and Loeckenhoff 2006), and should also be adjusted to the needs of different users (Zhao, Agrawala, and Hinckley 2006). Instead of using a timeout, *earPod* eliminates the need for an artificially determined threshold by using interruptible audio and a self-discoverable touchpad layout. This transition is self discoverable, seamless, and arguably “natural”.

Input / output mapping: Spatialized audio is used to reinforce the user's cognitive mapping between menu items and spatial locations on the touchpad. For example, if the finger touches an item on the right side of the touchpad, the audio will be played back on the right side of the user (using binaural spatial cues). Consistency between the spatial knowledge obtained through finger exploration and the audio feedback is designed to help the user memorize the spatial layout of the items of interest.

3.2.3 Implementation details

To make the *earPod* more accessible, our design only requires a circular touchpad as the input device and an ordinary stereo headset for output. Particular attention was paid to audio design as even minor details can have significant consequences on the usability of an auditory interface.

3.2.3.1 Input

The input device I used was a Cirque EasyCat USB external touchpad covered by a thin plastic overlay with a circular cutout, resulting in a circular touchpad. The radius of the touch-sensitive area is 19 mm and the overlay is 2 mm thick. The cutout includes a 3 mm notch on top, providing a tactile orientation cue to the user (Figure 3.1 right). *earPod* was implemented in Java 1.5 and uses JNI to read raw absolute (x, y, pressure) touchpad data via a special driver supplied by Cirque.

Noise and inaccuracy were handled in two ways. First, involuntary movements occurring when the finger is lifted were filtered out. Second, a spatial hysteresis algorithm was applied to avoid inadvertently releasing the finger on an adjacent area. The algorithm works as follows: every time the finger crosses a boundary separating two functional areas of the touchpad, this boundary is slightly moved to the opposite direction. For example, if the finger moves to an area situated on the right, it will require a slightly larger movement to go back to the area on the left[9]. The same filtering methods were applied to the visual menu used in the experiment.

3.2.3.2 Output

All speech sounds used in the *earPod* interface are human voices recorded in CD quality (16 bits, 44KhZ) using professional equipment. Since Java's native sound library adds a small but perceptible lag when playing sounds, I extended a real-time sound physics simulation library for use in *earPod* (Doel and Pai 2001). Sound files were post-processed after loading them to remove leading silences, and the sound signals were normalized to avoid the unpleasant effect of non-uniform playback volumes.

The third and final post-processing step was spatialization. In the current implementation, spatialization is simply achieved by manipulating mono voice streams to incorporate Interaural Time Differences (ITDs) and Interaural Intensity Differences (IIDs), the two major binaural cues for localizing sounds on the left-right axis (Bernstein 1997). Positioning sounds on the two other axes (front-back and up-down) it is more difficult and would require sophisticated signal filtering methods (such as Head Related Transfer Functions) as well as individual calibration (Bernstein 1997). The *earPod* technique does not preclude use of such methods, but simple left-right spatialization is probably sufficient for menu selection and is well suited for mobile devices with limited computing power.

3.3 Design for Mobile Environment

Although the *earPod* technique can be used in many situations, it is particularly suited to mobile use. An important lesson learned from Pirhonen et al.'s study (Pirhonen, Brewster, and Holguin

---

[9] The accompanying video provides a visual demonstration of the hysteresis algorithm.

2002) is that gestures must be robust enough to be used when moving and that simple taps are easily triggered by accident.

To test the viability of our technique in a mobile environment, several people were asked to perform a few selections while standing, walking, or jogging informally. Participants had no problem performing the glide and tap gestures to select items when standing or walking. However, tapping became more difficult and less accurate while running, when the touchpad started shaking in the participants' hands.

To overcome this problem the touchpad was secured by attaching it to a hand wrap using a Velcro fastener as shown in Figure 3.5. With this addition, participants found it much easier to perform the taps while in motion. The fact that participants are able to use our technique (with the addition of the wristband-mounted touchpad) while running gives us more confidence that the tapping selections can be made under a variety of mobile situations, especially since most mobile scenarios, such as walking or traveling on the train, are less challenging than running.

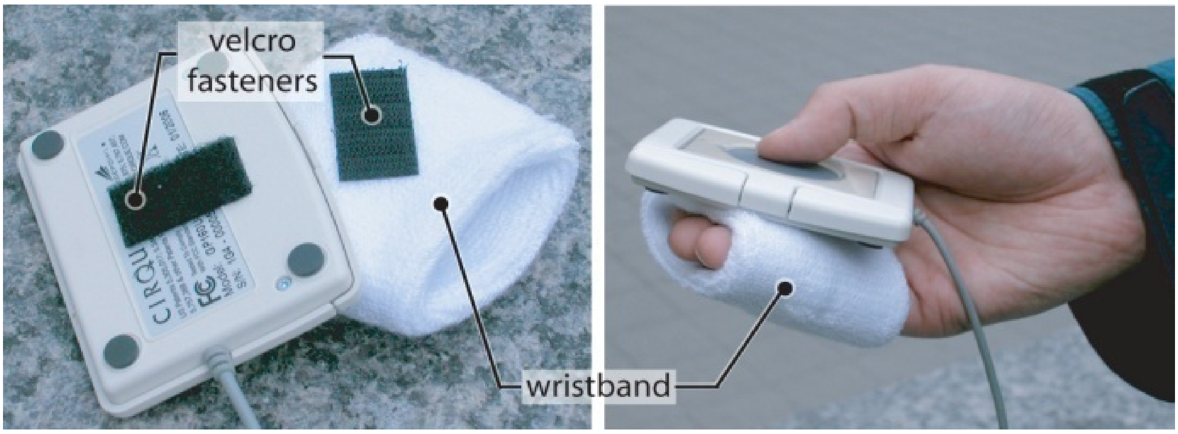


Figure 3.5: Using a wrist band and Velcro fasteners (left) to secure touchpad in hand (right) while the user is in motion.

The circular touchpad could be embedded into devices of many different form factors or it could be implemented as a separate component, like a wireless remote control. Touchpads are typically very light, allowing a touchpad remote to be carried easily on a neck lanyard or a key chain. While some custom design will undoubtedly be needed to adapt *earPod* to such individual devices, the effort required to execute the designs should be relatively straightforward.

3.4 Conclusion

The *earPod* menu selection method is a promising technique that should address the problems associated with auditory menu selection. However, research is needed to demonstrate that the theoretical benefits of the *earPod* approach actually work in practice. How efficient is *earPod* selection and can it compete with visual menu selection methods? How much learning is required to use *earPod* menu selection efficitively? These and other questions are addressed in four experiments that are reported in the following chapters.

Chapter 4
Evaluation 1: *earPod* vs. iPod-like Linear Menu

4.1 Introduction

As explained in the preceding chapters, the *earPod* includes a number of design features that should facilitate the task of menu selection using auditory feedback. However, the expected benefit of *earPod* menu selection needs to be validated through experimental analysis. Preliminary testing in informal pilot studies indicated that menu selection using *earPod* was surprisingly easy and fast. Given the apparent speed of *earPod* selection and the lack of a comparably fast menu selection method that used auditory feedback, it was decided to compare the *earPod* approach with the commonly used visual selection approach implemented in many millions of iPod devices (Figure 4.1 left). The iPod is a mature product that has been embraced by consumers worldwide and that has undergone several generations of iterative design. Similar to the *earPod* hardware, the iPod uses a circular touchpad for input, and allows navigation of alternatives by gliding the finger along the outer ring of the touchpad. As users glide the finger along the outer track of the ClickWheel (Figure 4.1 right), the corresponding items are highlighted. Users select the currently highlighted item by pressing the center button.

Can the eyes-free method of menu selection in *earPod* be a viable alternative to the visual interface of the iPod? Would the expected transition from novice to expert performance happen quickly enough to enable its use by a wide range of people. Experts were expected to tap directly on target items rather than move the finger around the touchpad's circumference and it was this kind of tapping behavior that was expected to produce selection performance fast enough to match iPod visual menu selection. To answer these questions, Experiment 1 compared an *earPod* with an iPod-like visual menu.

4.2 Experiment

Extensive research has been carried out on the configuration of radial menus (Kurtenbach 1993; Zhao and Balakrishnan 2004; Zhao, Agrawala, and Hinckley 2006). In order to achieve acceptable speed and accuracy, it has been recommended that the circle used in such menus be divided up into 8 regions per menu level. In this dissertation, pilot studies were conducted with menus having breadths of 4, 8, and 12 items, and with depths of 1 or 2 levels, and the results of those pilot studies were consistent with the earlier findings obtained by other researchers. Although 12 items are usable, 8 items or less per menu level tend to work best in terms of speed and accuracy.

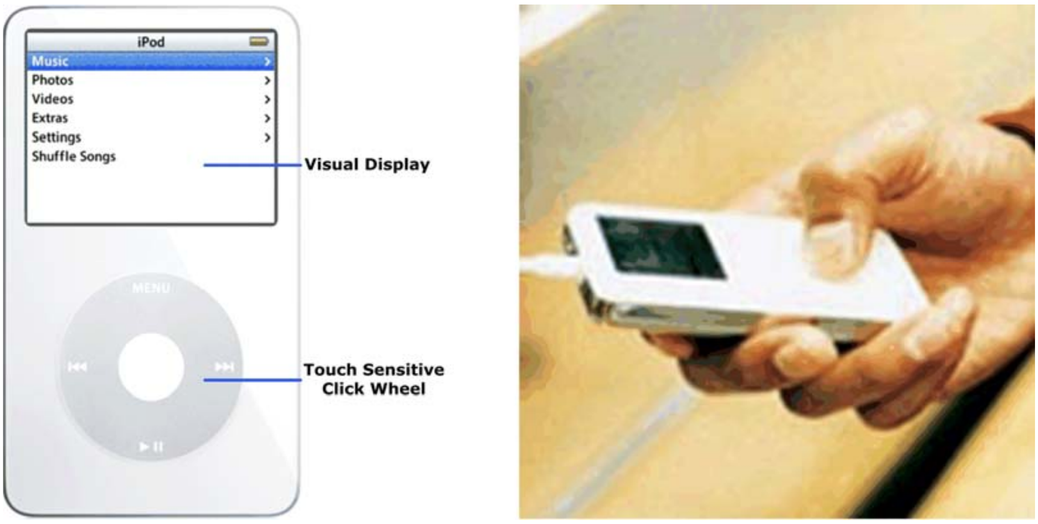


Figure 4.1: The iPod visual menu (left) and its interaction technique (right)

Since iPod implementation details were not available from Apple, the iPod linear menu was simulated using a circular touchpad connected to a notebook computer. The implementation was nearly identical to the recently released *iPod Nano* Media Player, apart from the fact that items were selected when the finger was lifted from the input surface.

4.2.1 Participants

Twelve right-handed participants (3 female) ranging in age from 17 to 30 years (mean 24), recruited within the university community, participated in the experiment. Subjects received $10 dollars per hour for their participation in the study. None had previously used an iPod.

4.2.2 Apparatus

The experiment was conducted on a Dell Inspiron 8000 laptop running Microsoft Windows XP, with a 19" external LCD display. Input and output were handled as explained earlier in section 3.2.3.

4.2.3 Task and stimuli

To make the experiment more realistic, the *earPod* technique was tested with a 2-level (8x8) hierarchy where there were 8 items nested within each of the 8 top-level categories, resulting in 64 items in total. 64 items would likely be sufficient to capture the common commands for many of today's audio-based devices.

Following Miller (1981), who used real world hierarchies as stimuli in his experiment, the categories used in this study were the following familiar words selected from hierarchies developed by KidsClick! (http://sunsite.berkeley.edu/KidsClick!/) and Wikipedia (http://www.wikipedia.org/):

Clothing: Apron, Brief, Cloak, Coat, Dress, Hat, Jacket
Fish: Carp, Cod, Eel, Haddock, Pollock, Redfish, Salmon, Sardine
Instrument: Bassoon, Cello, Clarinet, Drums, Flute, Guitar, Organ, Piano
Job: Actor, Cook, Doctor, Driver, Farmer, Hunter, Lawyer, Soldier
Animal: Ants, Apes, Bats, Bears, Eagles, Zebras, Elephants, Horses
Color: Black, Blue, Grey, Green, Lime, Navy, Olive, Purple
Country: Brazil, China, Denmark, Egypt, England, Finland, France, Greece
Fruit: Apple, Banana, Cherry, Grape, Guava, Kiwi, Lemon, Mango

All words had one to three syllables and an audio duration of about 1 second.

To ensure that the audio interface was not favoured, visual presentation of stimuli was used in both conditions.

4.2.4 Design

A within-participants design was used. Participants were randomly assigned to two six-person groups. The first group performed the experiment with the *earPod* technique first, while the second group used the iPod-like visual menu technique first.

For each technique, participants made selections from 2 menu layouts: a single level menu containing the 8 categories, and a two-level menu with 8 items per level (8x8), containing all 64 items organized into the same 8 categories. The use of the *earPod* vs. iPod technique was counter-balanced, but a single ordering of menu depth, from easy to difficult (first the 8-item menu, then the 8x8-item menu) was used. The menu content and item orderings were chosen in advance and were identical for both techniques and for all participants:

Condition 8: all 8 possible stimuli were used.

Condition 8x8: as previously discussed in the "task and stimuli" section, to keep the experiment manageable, a subset of 16 of the possible 64 stimuli were chosen (two sub-items per top-level menu item, pre-selected before the experiment). Menu item presentation order was randomized across participants.

Participants were allowed to take breaks between trials. Breaks were enforced between different techniques and layouts. Before the experiment, participants received 5 minutes of training per interface using a different set of stimuli than the one used in the study. Each participant performed the entire experiment in one sitting which took approximately 90 minutes.

In summary, the design was as follows (excluding training):

12 participants x
2 techniques (audio and visual) x
(8+16) items for the 2 menu configurations (8 and 8x8) x
20 blocks
= 11520 menu selections in total.

4.2.5 Design rationale

The above design was chosen due to a number of reasons. First, the within-subject design was chosen (as oppose to between-subject design) due to its great statistical power (Box, Hunter, and Hunter 2005). Subjects are inevitably different from each other. In between-subject designs, the differences between subjects are treated as errors. In within-subjects designs, since all subjects go through all conditions, the difference within individuals is known and can be measured separately from other sources of error, generally reducing the size of the error term and increasing the sensitivity of the F test. A counter-balanced design was used to reduce the likelihood that order effects tainted the results obtained. Additional statistical analyses checking for the presence of asymmetric transfer between conditions were also carried out (Appendix 1) and will also be considered in the discussion of the results later in this chapter.

Participants were recruited from the University community because iPod and mobile devices are largely used by a younger and tech-savvy demographic. The sample size of 12 participants is typical for this type of HCI experiment and enabled the statistical tests to be carried out with sufficient power. Following Miller, real world hierarchies are used as testing content. In order to save time and collect more data, only one-word menu items with one to three syllables were used.

Pilot testing suggested that learning the locations for a 64 item menu hierarchy would likely require several hours. To keep the duration of the experiment within reasonable limits, only 16 of the 64 possible stimuli – 2 items per top-level menu, were used. This still required participants to search the 8x8 item menu options and therefore made the selection process more demanding than a single 16-item menu. Limiting the number of possible stimuli, however, increased the frequency with which participants encountered each item – consistent with power law distributions observed for the frequency of use of command menus (Witten, Cleary, and Greenberg 1984).

The focus of the study was to find the intrinsic properties of the techniques in general, instead of a particular implementation; therefore, this experiment used research prototypes implemented using the same software platform.

4.2.6 Procedure

Before their first trial, participants were instructed to put on the headphones and to hold the touchpad with the right hand as shown in Figure 3.1, leaving the thumb off the touchpad. Participants then pressed the spacebar using their left hand to start the trial. A visual stimulus (the item to select) was then displayed in the center of the screen as shown in Figure 4.2. Participants responded by bringing their thumb in contact with the touchpad and dragging the thumb in search of the target.

In the audio condition, participants heard the spoken names of each traversed menu item through their headphones. In the visual condition, menu items were displayed on the screen (see Figure 4.2). The selection process remained active as long as the participant's thumb remained in contact with the touchpad surface. Participants completed selections (either a menu or submenu item) by lifting the thumb off the touchpad. In both conditions, a short click sound was played whenever a selection was made. If a stimulus had two levels, participants went through the thumb-down → thumb-drag → thumb-up[10] process twice to finish the task. A trial was considered erroneous if any of the selected targets did not match the stimuli. In this case, participants were notified by a "mismatch" visual message. No additional feedback was given for successful trials. After each trial, a visual message in the center of the screen instructed participants to press the spacebar to proceed to the next trial.

A general note before discussing results in the next section: For the analyses of variance reported in this chapter and subsequent chapters, whenever a within subjects effect was tested, the assumption of sphericity in the data was first assessed using Mauchly's test of sphericity as implemented in SPSS. In cases where the sphericity assumption was violated (where sphericity may be thought of as the multivariate analogue of heteroscedasticity) degrees of freedom in the corresponding ANOVA test of the effect were adjusted using the Huyn-Feldt criterion. In the majority of within subjects effects tested in this way the assumption of sphericity was not violated, and when it was, the adjustment of degrees of freedom did not affect whether or not the effect tested was statistically significant. Thus for the sake of simplicity, the ANOVA results are reported in this dissertation without any adjustment to the degrees of freedom.

---

[10] When users become experts in the audio condition, they skip the thumb-drag and only perform a tap to select an item.

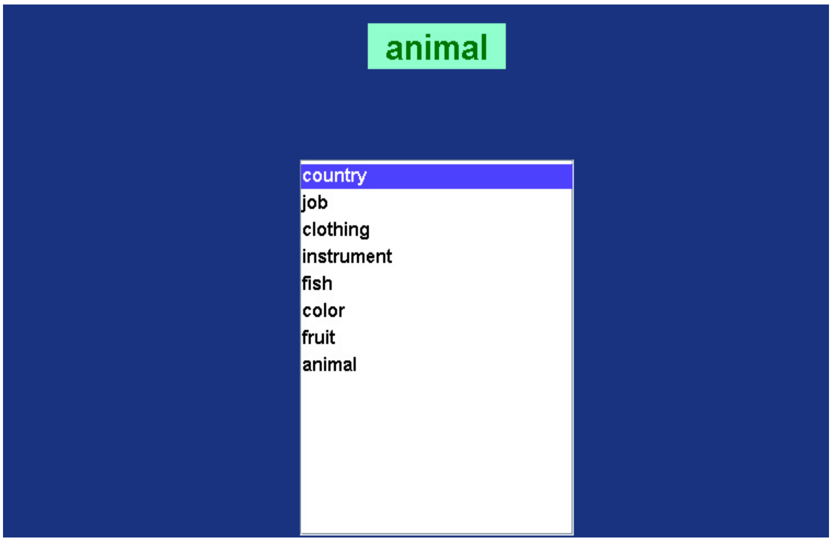


Figure 4.2: The lower half of the screen: the item to be selected from the menu is displayed at the center of the screen. The menu at the bottom is displayed only in the visual condition.

4.2.7 Results

Repeated measures analyses of variance were used to assess the effects of interface (audio vs. visual) on accuracy and selection time. For these analyses, the learning effects were assessed by grouping the 20 blocks of trials within each session into four groups of five contiguous blocks, which will be referred to as the four “time periods” below. Analyses were carried out across all four of the time periods to assess learning effects, and across the last time period (last five blocks) to assess the participants’ performance with each technique after some training had taken place.

4.2.7.1 Accuracy

Overall, the audio technique had an accuracy of 92.1% in this study, while the visual technique yielded 93.9% accuracy. This difference was not statistically significant ($p>.05$). As might be expected, there was a significant effect for the number of menu levels ($F_{1,11}=21.16$, $p<.001$); accuracy with single level menus (94.2%) being higher than the corresponding level of accuracy for two level menus (91.8%). None of the other main effects or interactions involving accuracy were significant when all four time periods were considered, nor when only the last time periods (last five blocks) were considered.

4.2.7.2 Response time

On each trial, selection (response) time was measured as the duration from the appearance of the stimulus to the completion of the selection. As is typical with response time data, the distribution of raw times was positively skewed. During aggregation of the data (within participants and within each cell of the design), medians were used as measures of central tendency in order to reduce the effect of potential outliers (Rosenberger and Gasco 1983). Means of these median values were then used to estimate average selection times across the participants.

There was no significant overall difference in speed between the audio and the visual technique ($F<1$). As expected, however, it took significantly longer ($F_{1,10}=82.84$, $p<.001$) to complete selections in two level menus (averaging 3.4 seconds) than in one-level menus (1.9 seconds). There were also significant learning effects on response time across the experimental sessions. There was a main effect of overall learning ($p<.001$), and significant interactions between learning and number of menu levels ($p<.001$), and learning and technique ($p<.001$).

The learning effect with respect to the significant three-way interaction that was observed between time period, technique and number of menu levels ($F_{3,30}=7.75$, $p<.01$), is shown in Figure 4.3. The learning rate was faster with the audio menu than with the visual menu as illustrated in Figure 4.3 by the crossover in the curves that occurred for both menu levels. The audio menu was initially slower than the visual menu, but with experience, performance on the audio menu became faster. In the final time period (last five blocks of the twenty blocks) in the experimental session, the audio menu, at an average of 2.1 seconds, was significantly faster than the visual menu at an average of 2.5 seconds ($F_{1,10}=6.03$, $p<.05$).

An order effects analysis (Appendix 1) was also carried out by examining the interaction between order and the experimental factor for both accuracy and response time. No significant interaction was found and thus the experimental results are unlikely to have been influenced by the effects of asymmetric transfer.

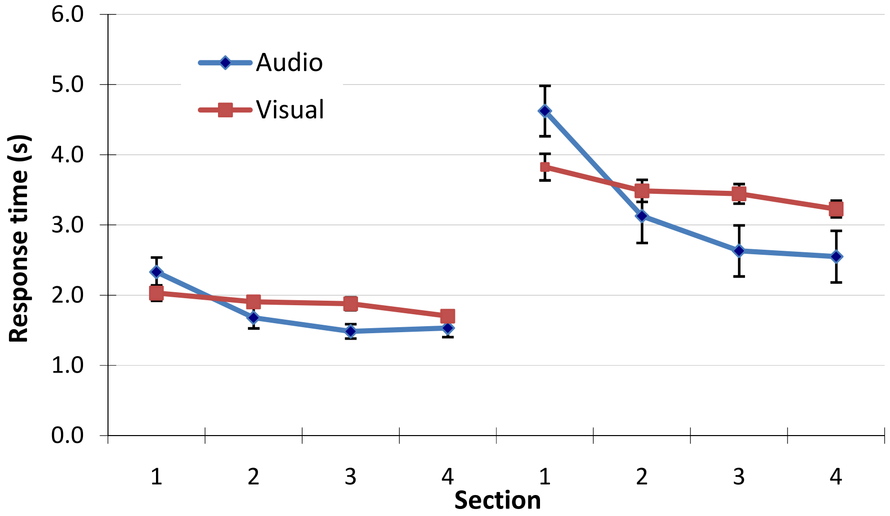


Figure 4.3: Selection times for the two techniques by number of menu levels and time period (1 time period = 5 contiguous blocks of trials).

4.2.7.3 Detailed log analysis

In addition to the accuracy and selection time data, a detailed log was maintained of the exact movements made during the sessions. Using this log it was possible to distinguish between glide selections, where participants traversed to the target region around the perimeter of the circular touchpad, and tap selections, where participants jumped to the target region and tapped it.

At the beginning of the experiment, all participants used the glide technique, with considerable variability in the details of the gliding traversals employed. The touchpad was divided into 8 zones as shown in Figure 3.2. From the beginning of the experiment, participants typically started by gliding to and around the target because they did not have the expertise to tap on the target immediately. For example, participant 1 went through the following sequence to find target item 5 on his first trial: zone 6 was touched first, and then he traversed (dragged) through zones 7, 1, 2, 3, 4, before finally settling on zone 5, and releasing his finger to select it. However, by the end of the session the participant was tapping on the target zones directly without any drags, indicating that expert performance had been reached. Similarly, for a 2-level selection, expert performance was indicated by using only 2 taps, avoiding the need to drag. However, not all the participants were able to reach expert performance in the experiment. Overall, 9 out of 12 participants reached expert performance for the single level menu, and 8 out of 12 participants reached expert performance in the two level menu. It should be noted that the only technique taught to the participants was gliding, thus 9 out of 12 participants independently discovered the tapping technique. Even for the 3 participants who did not reach expert behavior by the end of the session, examination of their detailed logs indicated that they used tapping in some of the experimental trials.

4.2.7.4 Subjective preference

In terms of subjective preference, four participants expressed a preference for the audio menu, four preferred the visual menu (including the three participants who did not achieve expert performance), and four had no preference between the two techniques. This result seems promising, as the participants had no incentive to use an eyes-free technique under the given experimental settings (i.e., they were not required to walk or asked to perform a visually distracting task while making their menu selections). All participants except one either agreed or strongly agreed that it would be desirable to *combine* the audio technique with the visual technique.

Some participants in our study commented that they suspected tapping to be more likely to be triggered accidentally. This could be avoided by requiring users to hold a button or perform a squeeze gesture (Harrison et al. 1998) to activate a command.

4.3 Summary and discussion

Compared with the iPod-like visual linear menu, the audio technique that was used has comparable speed and accuracy overall. Although initially slower than the visual technique, the audio technique was significantly faster after only 30 minutes of practice. At the start of each session, the audio technique was half a second slower for one-level menu selections and one second slower for two-level menu selections.

Analysis of the experimental logs indicated that the transition from novice to expert usage was spontaneous. Although the learning rate differed between participants, most participants were able to learn the expert behavior (tapping) quickly.

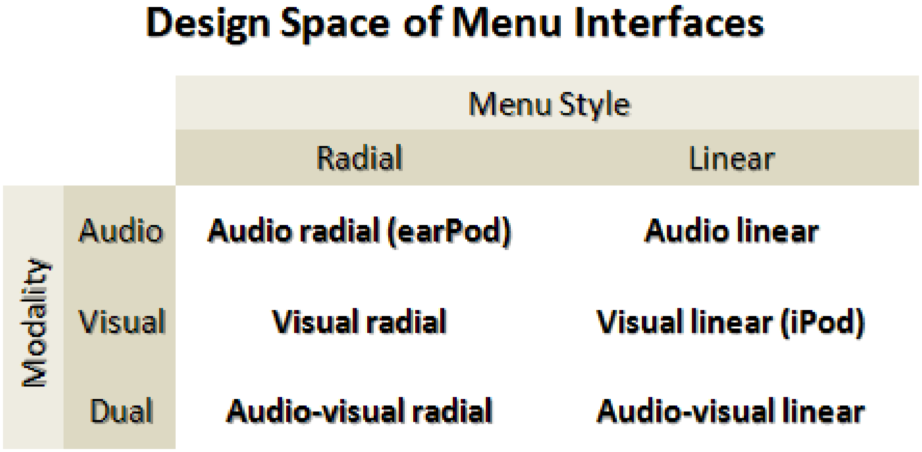
With its combination of synchronous communication model, reactive audio feedback, and intuitive input output mapping, the *earPod* audio menu selection technique was found to provide acceptable novice performance, fast learning rate, and quick transition to expert usage. Although this experiment has encouraging results, it only compared *earPod* with the visual linear technique. In the next chapter, a systematic exploration of menu techniques using different modalities of feedback and menu styles will be presented.

Chapter 5
Experiment 2: 3 X 2 Study

5.1 Introduction

In the previous study, which was a direct comparison of *earPod* with an iPod-like visual menu, the performance of the two interfaces was found to be comparable in terms of speed and accuracy on fixed-sized static menus on the single task desktop setting. While the first study demonstrates the efficiency of the *earPod* design with an iPod-like visual linear menu, The precise reason for those results is unclear, as explained in the following paragraphs.

The two interfaces tested (*earPod* and iPod-like menu) differed in two aspects, namely, the *modality* of feedback, and the *menu style* for presenting and navigating the menu items. For *modality*, the iPod-like menu primarily relied on a visual display to present the menu options and navigational cues, while *earPod* carried out these functions using audio only. In terms of *menu style*, the iPod used the *linear* menus (Figure 5.2, right) where items are placed relative to each other, and there is no one-to-one mapping between specific input areas to menu items; *earPod*, on the other hand, used a *radial* menu layout where each menu item is directly mapped to a physical location on the touchpad. Figure 5.2 (left) shows an example of radial layout for the visual interface. The radial layout allows expert users to access any item in the list in constant time. In the previous experiment, *menu style* was correlated with the distinction between auditory and visual menus. Which of these factors influenced the observed results?



Design Space of Menu Interfaces

		Menu Style	
		Radial	Linear
Modality	Audio	**Audio radial (earPod)**	**Audio linear**
	Visual	**Visual radial**	**Visual linear (iPod)**
	Dual	**Audio-visual radial**	**Audio-visual linear**

Figure 5.1: The 3x2 design space of modality vs. menu style. Both previously studied interfaces - *earPod* and iPod

The above question can be addressed by looking at a larger portion of the design space spanned by *modality* and *menu style*. If *modality* and *menu style* are considered to be two dimensions in a design space as shown in Figure 5.1, it can be seen that there are a number of different designs that can be implemented. The popular iPod interface fits within the "*visual linear*" category, while the *earPod* interface resides in the "*audio radial*" cell. Two additional interface possibilities are "*audio linear*" and "*visual radial*". Since audio and visual feedback can co-exist and are not mutually exclusive (unlike *menu style*), there is also a third possible choice in the *modality* dimension, the *audio-visual* or *dual* modality, which can then be implemented with each of the two menu styles. This yields a 3x2 matrix of six design possibilities. These alternative designs cover a variety of interesting properties and thus warrant investigation.

In the following discussion, the alternative design choices within the design space are briefly described, after which an experiment is described which evaluated all six interfaces implied in the design space.

5.1.1 Audio linear

The *audio linear* option shown in Figure 5.1 provides spoken-word auditory feedback to users as they scroll up or down a menu list. In some respects the interface is similar to that used in the popular Apple iPod digital music player, except that in the absence of a visual display, auditory feedback is provided. In principle, such an *audio linear* interface could be easily integrated with the existing iPod interface, which is currently available through an open source solution from rockbox.org.

5.1.2 Visual radial

As discussed above, the *earPod* interface has a radial input area that supports a *radial* menu layout where specific spatial regions on the input device have a one-to-one mapping with items in the menu. Figure 5.2 (left) shows the design of the *visual radial* interface. Although the interaction method differs, its appearance looks similar to that of a marking menu. Notice that this is in contrast to the *linear* menu layout, where the input device supports a vertical scroll of a focus point through the menu (see, Figure 5.2, right). It is possible that the performance advantages of the *earPod* interface discussed earlier, may, in part, be due to the *radial* menu layout used.

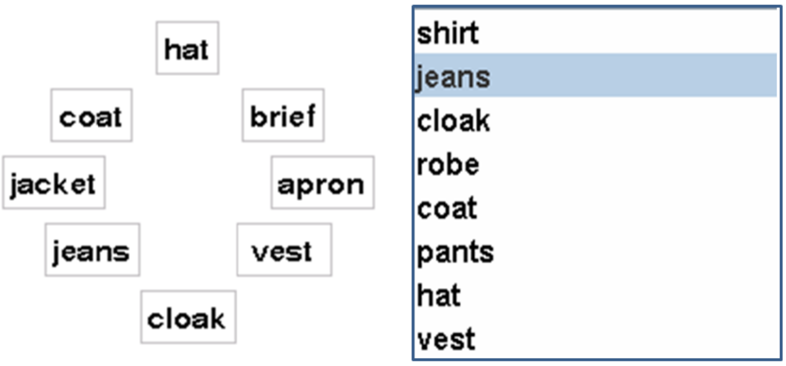


Figure 5.2: Screenshots of the visual radial interface (left) and visual linear interface (right)

5.1.3 Dual linear and dual radial

By providing both *audio visual* feedbacks simultaneously, the *audio-visual* interfaces may combine the best of both worlds — namely, they can be operated using either *modality*, thereby giving users a choice of which *modality* to attend to in different situations. For example, if the device is operated inside one's pocket, the *visual* feedback can be ignored. If the device is in a noisy environment, the *visual* feedback will be more useful. Since both channels of feedback use the same *menu style*, the training received in either *modality* will be beneficial to the other. However, simultaneously providing both modalities might waste resources (such as battery power), and one source of feedback has the potential to interfere with the other if the user primarily uses that type of feedback in a given scenario: for example, a user who prefers *visual* feedback could be annoyed by simultaneous *audio* feedback.

5.2 Experiment

To disentangle the individual effects of the two design dimensions, and further explore the properties of the other four design alternatives relative to the iPod and *earPod* interfaces in the baseline desktop conditions, this study used a 3x2 experimental design that employed all six interfaces from Figure 5.1.

5.2.1 Participants

Twelve right-handed participants (3 females) ranging in age from 18 to 29 years (mean 22), recruited within the University of Toronto community, participated in the experiment. Subjects received $10 dollars per hour for their participation in the study.

5.2.2 Apparatus

The experiment was conducted on a Compaq Presario V2000 laptop running Microsoft Windows XP. The input and output setup and the experiment and prototype software were exactly the same as Experiment 1.

5.2.3 Task and stimuli

In most aspects the experimental setup was identical to that used in Experiment 1. However, this experiment differed in the following ways. In particular, while the previous experiment studied both 1-level and 2-level hierarchies (menu depth) with 8 items per category Experiment 2 used only a 1-level hierarchy with 8 items. While this made the experiment more efficient it was also expected that it would not affect the results greatly, since the performance of 1-level and 2-level selection in the preceding experiment were generally similar, with no significant interaction effect being found between interface and menu depth.

Six previously used categories of menu items were used (*clothing, fish, instrument, job, animal, color*). To minimize cross-condition learning, the ordering of the six experimental conditions, and the assignment of the six menu categories to those conditions, were counterbalanced across subjects to the extent possible given the number of participants to be used.

5.2.4 Design

A within-participants design was used. The order of *modality* was totally counterbalanced, while the ordering of *menu style* was randomized within each *modality*. As previously mentioned, participants made selections from a 1-level menu of 8 items for each condition. The menu content and item orderings were chosen in advance. All 8 possible stimuli were used for each of the six menu categories.

Participants were allowed to take breaks between trials. Breaks were enforced between different techniques. Before each condition, participants received 8 practice trials (1 block) of that particular interface. After the experiment, a questionnaire was used to access subjective feedbacks. Each participant performed the entire experiment in one session which took approximately 60-90 minutes. In summary, the design was as follows:

12 participants x
6 techniques x
8 menu items x
13 blocks (12 blocks + 1 practice block)
= 7488 menu selections in total.

5.2.5 Design rationale

This experiment implemented a 3x2 multi-variant factorial design. A within-subjects design was again chosen due to its greater statistical power and efficiency. However, increasing the number of factors also raises the risk of crossover learning (asymmetric transfer). Since fully counter-balancing the six different conditions in the experiment would have required 6! (720) subjects it was decided to only fully counter-balance modality, the more important factor. Additional statistical analysis on order effects was also carried out (Appendix 1) to test for the presence of asymmetric transfer in the data.

University students were chosen as participants for the same reasons mentioned in the previous chapter with respect to Experiment 1. The software and menu structure of this experiment were kept the same in order to allow between-experiment comparisons. In order to keep the length of the experiment within a 60-90 minute session, 13 blocks of trials were used (this requirement was derived from the time used by participants in the pilot studies that were run prior to the experiment).

5.2.6 Procedure

The experimental procedure followed that used in Experiment 1, except for the *dual* modality conditions where participants both heard the menu items through the headphones and watched them on the screen.

5.2.7 Results

5.2.7.1 Accuracy

There was no significant differences ($p>.05$) among either *modalities* (*audio* 95.2%, *visual* 94.6%, *dual* 95.1%) or *menu styles* (*radial* 94.7%, *linear* 95.3%), nor was there a statistically significant interaction between modality and menu style. The mean accuracy for each of the different techniques was: *audio radial* (94.4%), *audio linear* (96.1%), *visual radial* (94.9%), *visual linear* (94.1%), *audio-visual radial* (94.7%), and *audio-visual linear* (95.3%). Furthermore, there was no indication of learning of accuracy across the 12 blocks (*n.s.*) indicating that users were able to reach a fairly consistent level of accuracy at the end of the practice trials and did not become more accurate thereafter.

5.2.7.2 Response time

There was a significant main effect of *modality*, ($F_{2,22}=36.79$, $p<.001$). The average response time for *audio* was 2.06 seconds, for *visual* was 1.56 seconds, and for *dual* was 1.62 seconds. Pairwise t-tests (with Bonferroni correction) showed that *audio* was significantly slower than both *visual* and *dual*, while there was no significant difference between the *visual* and *dual modalities*.

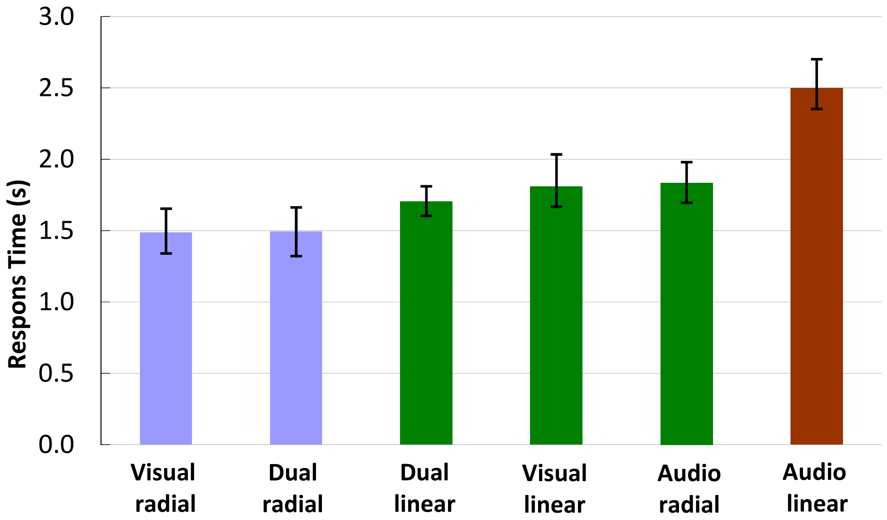


Figure 5.3: Response time for all 6 interfaces, sorted by speed, with fastest interface on the left, and the slowest one on the right. Bars with the same colors are not statistical different while bars with different color are significantly different.

There was also a significant main effect for *menu style*, ($F_{1,11}$=30.55, p<.001), with *radial* (1.56 s) performing significantly faster than *linear* (1.94 s). The average selection time for the six interfaces was: *audio radial* (1.77 s), *audio linear* (2.47 s), *audio-visual radial* (1.50 s), *audio-visual linear* (1.83 s), *visual radial* (1.48 s), and *visual linear* (1.70 s) (Figure 5.3). Pairwise t-tests (with Bonferroni correction) showed that *audio linear* was the slowest of all the techniques (p < .01), followed by *audio radial*, *visual linear*, and *audio-visual linear* techniques, which did not significantly different among themselves (*n.s.*), and also showed that the *visual radial* and *audio-visual radial* were the two fastest interfaces.

There was a significant *modality* x *menu style* interaction, ($F_{2,22}$=8.69, p<.01). This effect was caused by greater improvement between the *radial* and *linear* interfaces for the *audio modality* as compared to either the *visual* or *dual modalities*.

There was a significant main effect of *block*, ($F_{11,121}$=16.91, p<.001), indicating that learning occurred during the experiment. Pairwise t-tests (with Bonferroni correction) indicated that most of the learning happened in the first 4 *blocks* of trials, and that performance stabilized after that point, with no significant differences found from block 5 through block 12, $p>.05$.

There was also a significant *modality* x *block* interaction, ($F_{22,242}$=3.32, p<.001), indicating that learning rates differed across different *modalities*. A closer examination of the data showed that *audio* had a greater learning rate compared with *visual* or *dual*, while there was no significant difference between *visual* or *dual* in terms of learning rate.

Order analysis (Appendix 1) was again carried out to test for if asymmetric transfer may have affect the results obtained in this study. In order to avoid asymmetric transfer in the analysis only the first modality that each participant was exposed to was used in the analysis. The only difference found in this order analysis was that the interaction between modality and menu style was no longer significant. Since there appeared to be sufficient statistical power to test the interaction effect, which only appeared in the repeated measures analysis, it seems that the interaction affect obtained in the repeated measures analysis may have been due to the presence of asymmetric transfer.

5.2.7.3 Observations and subjective preference

Feedback from the post-experimental interviews indicated that the *visual radial* and *audio-visual radial* interfaces were the most promising, and that the *audio linear* interface had the lowest user satisfaction score. Ten of the participants reported that they preferred visual feedback over audio feedback in this experiment. The remaining two participants, who claimed to prefer audio feedback, actually performed better, in terms of selection time, using the visual interface.

5.3 Summary and discussion

The results from experiment 2 showed that techniques using visual feedback generally outperform those using audio feedback and that radial layouts generally have a performance advantage over linear layouts. The expectation that dual-channel feedback techniques would outperform single-channel feedback was not supported by the experimental results. The dual-channel feedback techniques were found to be comparable to the visual techniques. Subjective feedback from participants also suggested that users are mostly using visual feedback when both auditory and visual feedbacks are available. While this experiment was helpful in providing an overview of how the techniques within the design space performed, questions regarding the long term usage and learning behavior of these techniques are not answered. In the next chapter, a longitudinal study is carried out to assess longer term learning effects in more detail.

Chapter 6
Experiment 3: Longitudinal Study

6.1 Introduction

The studies reported in the two preceding chapters compared the performance of *earPod* menu selection against that of other techniques over the course of a single experimental session. In actual computer use, menu selection techniques are typically used over many sessions and long-term learning may occur. Thus it is interesting to see how the relative advantages of different menu selection techniques may change over time, and this requires the use of a longitudinal study. In addition, longitudinal studies are also useful in providing a more detailed account of the transition from novice to expert behavior. This chapter reports on a longitudinal study that was carried out. The study was designed to address a number of issues, as discussed in the following paragraphs.

Analysis of menu selection performance with a larger amount of content The previous studies used single sessions lasting from one to two hours, with no more than 16 items in the menus. Since expert level performance in the audio radial (*earPod*) condition requires memorization of the positions of the menu items, transition to expert performance would be expected to be much more difficult with menus containing more items that have to be learned. In the study reported in this chapter, an *earPod* longitudinal study (involving five daily sessions over the course of a week) with a full 2-level hierarchy containing 64 items allowed assessment of how relative performance between the techniques varied when a greater number of menu items was used.

Learning patterns between auditory vs. visual modalities Previous studies suggested a stronger learning effect for audio feedback conditions: The auditory techniques started slower but showed greater improvements over time. As a result, the response time curves for the auditory and visual modalities tended to converge with time. Would this pattern also hold over the course of a week containing many more trials, and with a more difficult task that used 64 menu items instead of 16? If participants were able to learn the positions of all the items in the two level menu hierarchy, they should eventually be able to perform the task in the auditory radial condition without needing to hear the auditory feedback, creating a situation where the initial advantage of the visual condition might disappear. However, it should be noted that in order to perform the task with little or no feedback, the user must not only know the spatial position of the target items within the menu structure, but must also acquire the physical skill to move efficiently to each of the spatial positions at will. The longitudinal study allowed a detailed analysis of relative learning across the experimental conditions over a sustained period of time and with a menu of more realistic size.

Learning patterns between linear vs. radial menu styles Previous studies indicated that accuracy was comparable between linear and radial menus, but that learning rates differed with respect to performance efficiency (speed). As with the modality effect discussed in the previous point, it was of interest to see if shifting to a more demanding task performed over a longer period of time would alter the learning rates associated with each type of menu.

Transition from novice to expert behavior In the previous studies, some participants appeared to transition to an expert style of performance within the single experimental session whereas others did not. One of the expected benefits of using a more demanding longitudinal study was that the transition to expert performance could be examined in more detail and over a longer period of time.

To address the above issues, I carried out a longitudinal study on the two dimensions: *modality* of feedback and *menu style*.

6.2 Experiment

6.2.1 Participants

Eight right-handed volunteers recruited from the University of Toronto community ranging in age from 18 to 30 years old participated in this experiment. To reduce the variability, all the participants chosen for this study were male undergraduate students. Participants received 10 dollars for each daily one hour session with sessions being carried out on a daily basis over the course of a week (i.e., 5 sessions per participant). A 20 dollar bonus was given to participants after they completed all 5 days. To motivate the participants to perform well in this long experiment, the best performer of each technique was entered in a randomized draw to win a 100 dollar prize[11].

6.2.2 Apparatus

The experiment was conducted using four desktop computers running Microsoft Windows^{XP} with service pack 2. Input and output were handled as explained earlier for Experiment 1 (Chapter 4). The experiment was conducted over a two-week period. Four people participated in the first week of the study and the other four people participated in the second week of the study.

6.2.3 Task and stimuli

All possible combinations (64 different choices) from a 2-level hierarchy (as presented in section 4.2.3) were tested with each participant.

6.2.4 Design

To avoid asymmetrical transfer effects (e.g., Poulton and Freeman 1966), I chose a between-subject design for the experiment. The 2 x 2 design examined two levels of menu type (linear vs. radial) and two types of feedback (visual vs. auditory). The eight participants were divided into four two-person groups where each group was exclusively associated with one of the four combinations of menu type and type of feedback. Each person in a group used the corresponding technique for his group for one hour a day and for five consecutive days. Within each week of the experiment, each of the four participants for that week had a different combination of the two factors (i.e., the experimental conditions were balanced across the two weeks).

The design for the entire experiment was as follows:

2 participants (per technique) x

4 techniques (audio radial, visual radial, audio linear, visual linear) x

64 items for 1 menu configuration (8x8) x

7 blocks per day x

---

[11] A participant had a chance to win the 100 dollar prize if s/he was the faster person for a particular technique. Since there were four techniques, four out of the eight participants qualified for the final draw with the 100 dollar prize being awarded to one of those four participants.

5 days

= 17920 menu selections in total

Aside from specific details noted above, in all other aspects the apparatus and procedures for this experiment corresponded to those used in Experiment 2 (as described in Chapter 5).

6.2.5 Design rationale

Contrary to the first two experiments, this experiment utilized a between-subject design, so that the adaptation of each participant to the particular technique assigned to him could be examined in detail. To obtain an accurate measure of the learning behavior of each technique, transfer effects that might have arisen in a within-subject design had to be avoided. As a consequence of choosing a between-subject design, extra care had to be taken to minimize individual difference among subjects. To minimize variance between subjects, only technology-savvy undergraduate male students were chosen as participants. To keep aspects of learning similar (aside from using the different techniques), the experiment was carried out around the same time each day (around 9-10 am), and sessions were carried out on consecutive weekdays. To keep the subjects motivated during this longer experiment, subjects were informed about the 100 dollar prize prior to the first day of the experiment. In addition, the hardware and software setup were kept identical to the previous two experiments in order to allow cross-experimental comparisons.

6.2.6 Results

Separate analyses were carried out to assess 1) improvement from one day to another (between-day analysis) and 2) improvement within each day (within-day analysis).

For between-day learning, mixed analyses of variance (ANOVA) was used to test the between subjects effects of interface (audio vs. visual), menu style (linear vs. radial) over successive days (within subjects) for both accuracy and response time.

For the within-day analysis, linear regression analysis on the natural log transformed data was used to inspect learning within each day. Beta values (standardized slopes) fitted in the linear regression analyses were used as indicators of how much learning occurred within each day (Figure 6.4 and 6.8).

In the following presentation, for each dependent measure (accuracy and response time), the results for between-day analyses are presented first, followed by the results for the within-day analyses.

6.2.6.1 Accuracy

6.2.6.1.1 Between-day analysis

There was a significant main effect for the within-subject factor: *day* ($F_{4,16} = 3.18$, $p < .05$), indicating that accuracy differed between days (Figure 6.1). Within-subject contrasts showed that the accuracy of day 4 (96.6%) was significantly higher than that of day 1 (94.7%) ($F_{1,4} = 5.30$, $p < .01$). Figure 6.1 shows the modest increase in accuracy that occurred during the course of the experiment.

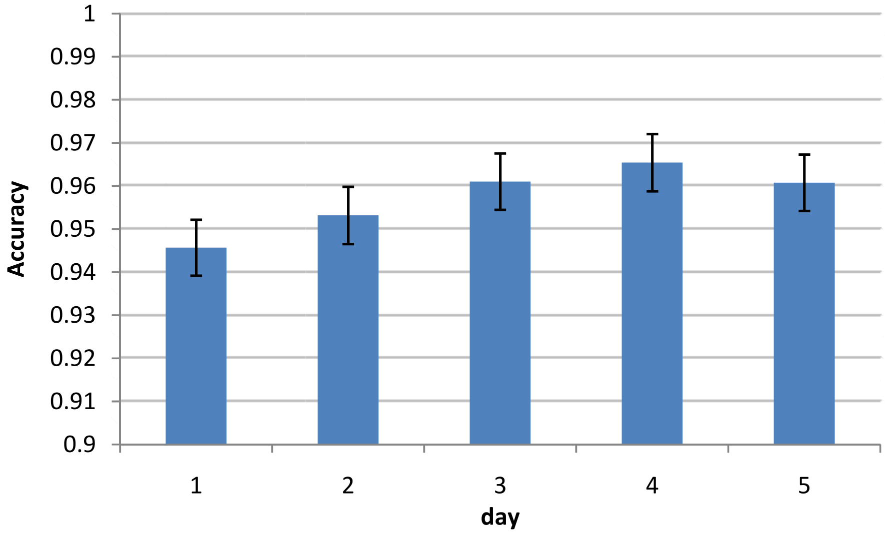


Figure 6.1: The overall accuracy for all the techniques. The categories on X axis represent different days.

There was also a borderline significant *day* x *modality* interaction effect ($F_{4, 16} = 2.73$, $p < .1$), suggesting the possibility that learning patterns may have differed between the audio and visual modalities across the five days. The visual techniques exhibited a learning curve in accuracy while the no such effect is observed for the auditory ones, which had comparatively high accuracy throughout the study. Figure 6.2, left panel illustrates this possible interaction. The corresponding main effect of type of feedback visible as a separation of the two lines in the left panel of Figure 6.2 was also significant ($F_{1,4} = 10.75$, $p < .05$), although the accuracy of the two types of feedback seems to converge by days four and five of the study. There was a significant main effect of type of menu ($F_{1,4} = 47.74$, $p < .05$) on accuracy with linear style menus being more accurate than radial style menus throughout the study. The mean accuracy for audio techniques (96.6%) was significantly higher than the corresponding mean accuracy for visual ones (95%), and mean accuracy for linear style menus (97.4%) was higher than the corresponding mean accuracy for radial style menus (94.2%).

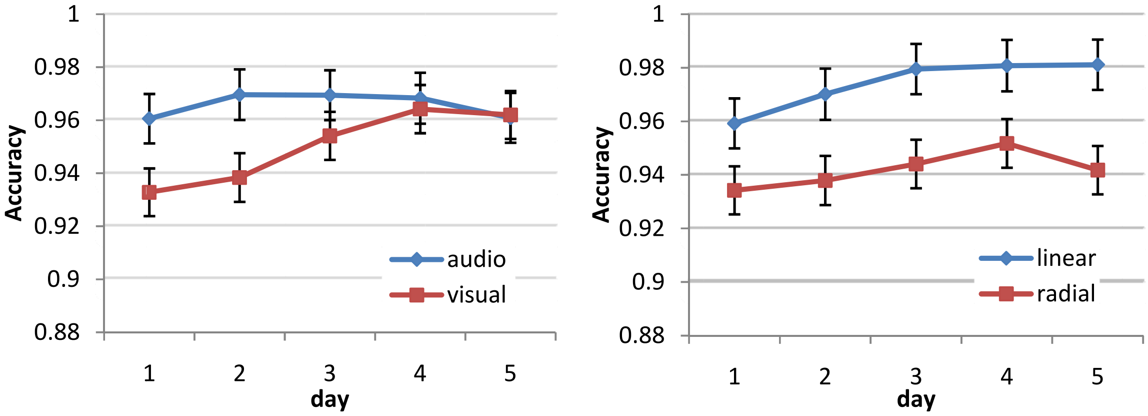


Figure 6.2: Accuracy change according to the modality of feedback (left panel) and menu style (right panel) for the five days.

6.2.6.1.2 Within-day analysis

Separate learning curves were fitted for each combination of day and experimental condition (i.e., 20 learning curves in all), with the data for fitting each learning curve being pooled across the two participants who were exposed to the particular condition. As described earlier, the learning curves were fitted using linear regression analysis carried out on natural log transformed data. The standardized beta coefficients for the slopes of the fitted lines were then used as indicators of strength of learning, and are plotted across days and experimental conditions in Figure 6.3. Learning within days was variable, and statistically significant evidence of learning occurred on only two of the days within the visual linear condition, where there was a significant improvement in accuracy on the first day and a significant decrease in accuracy on the last day.

Thus there was, in general, no steady improvement within each day, although average performance across days tended to improve for some conditions as noted in the previous section.

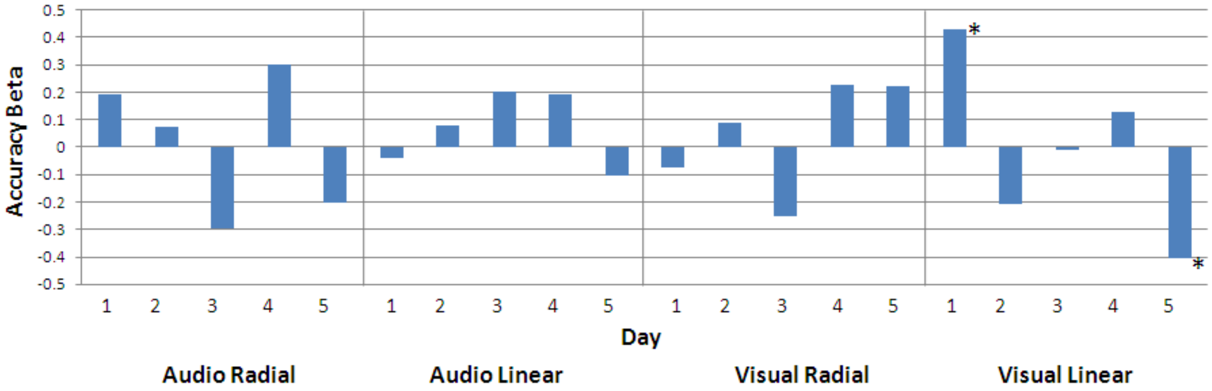


Figure 6.3: Within-day analysis for accuracy. The beta values per day are graphed separately for the four techniques. Each statistically significant beta value is indicated with a "*" mark.

6.2.6.2 Response time

In contrast to accuracy, there was a strong learning effect visible in the response time data for both between-day analysis and within-day analysis.

6.2.6.2.1 Between-day analysis

There was a significant speed-up in performance over the five days ($F_{4,16} = 42.26$, $p < .01$), as shown in Figure 6.4. Within-subjects contrasts indicated that mean performances on days 2, 3, 4, and 5 were all significantly faster than corresponding mean performance on day 1 ($p < .01$). (Mean response times steadily improved from 3.94 seconds on day 1 to 2.44 seconds on day 5). The performance difference over the five days in Figure 6.4 suggests that the rate of this learning slowed during the course of the week.

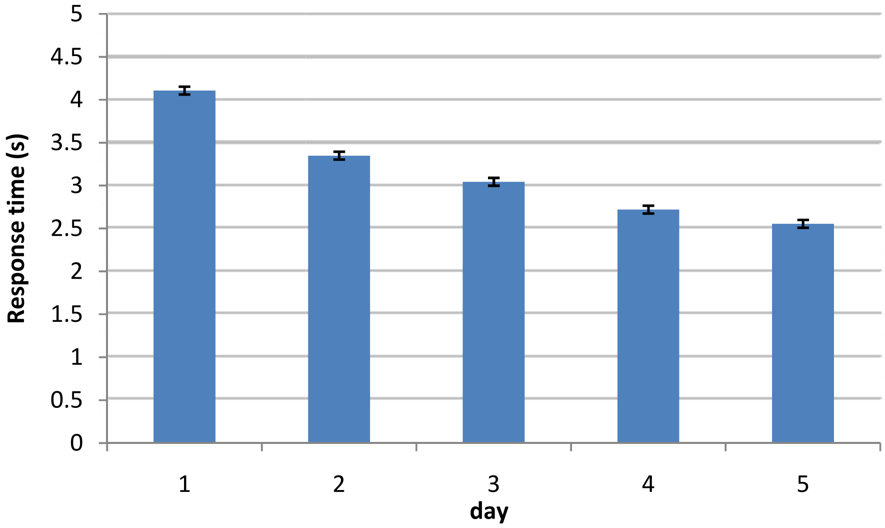


Figure 6.4: The overall response time across each of the five days.

There was a significant *day* x *modality* interaction ($F_{4,16} = 10.82$, $p < .01$) as shown in the left panel of Figure 6.5. The visual modality started out with much faster response times, but due to the greater learning effect for the audio menus, the difference in response times between the two types of modalities was negligible by the end of the week ($t_6 = .28$, $p > .05$). However, there was no corresponding *day* x *menu type* interaction (Figure 6.5, right panel), indicating that the speed advantage of radial menus was maintained throughout the course of the study, suggesting that radial menus may be inherently faster than linear menus.

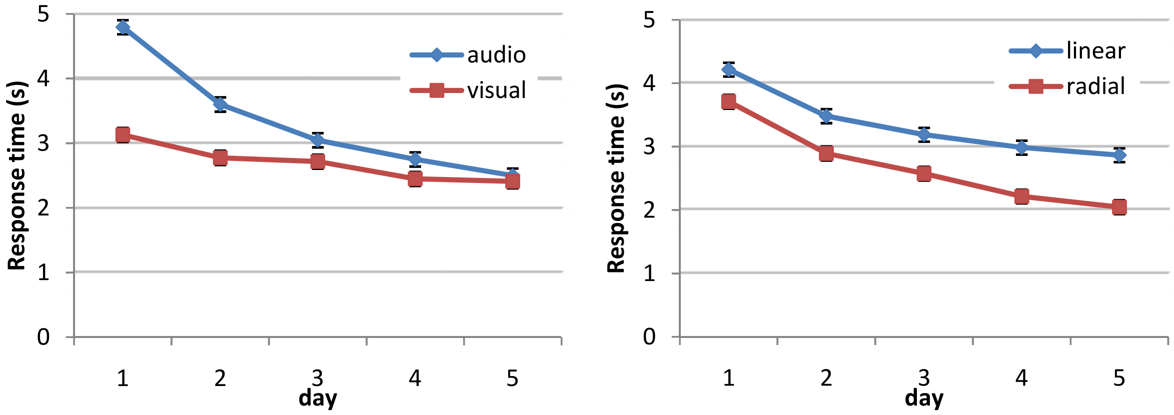


Figure 6.5: Response time change according to the modality of feedback (left) and menu style (right) for the five days.

Figure 6.6 shows the between day learning curves for all four of the experimental conditions. It can be seen that performance in the visual radial condition is fastest throughout the experiment, whereas learning is greatest for the audio radial condition. Although independent t-tests show that the difference between these two techniques is significant both on day 1 ($t_{110} = 12.18$, $p < .01$) and day 5 ($t_{110} = 7.89$, $p < .01$), the magnitude reduces over the course of a week. While visual radial was over two seconds faster than audio radial on day one, by day five that difference had declined to under half a second.

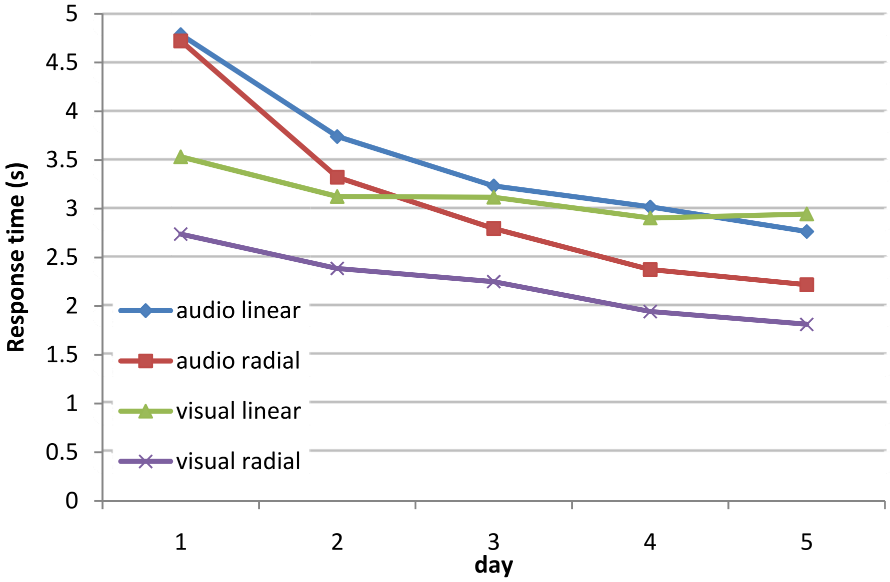


Figure 6.6: Response time is graphed separately for the four techniques for five days.

6.2.6.2.2 Within-day analysis

Learning effects were then examined within days with regression analyses being carried out for each day and participant. The fitted beta coefficients were then averaged across participants to obtain the mean beta values plotted in Figure 6.7. The beta values per day are graphed separately for the four techniques. Since learning is indicated by a lower response time, the slopes of the fitted regression lines, and thus the beta coefficients are expected to be negative. Each statistically significant beta value is indicated with a “*” mark.

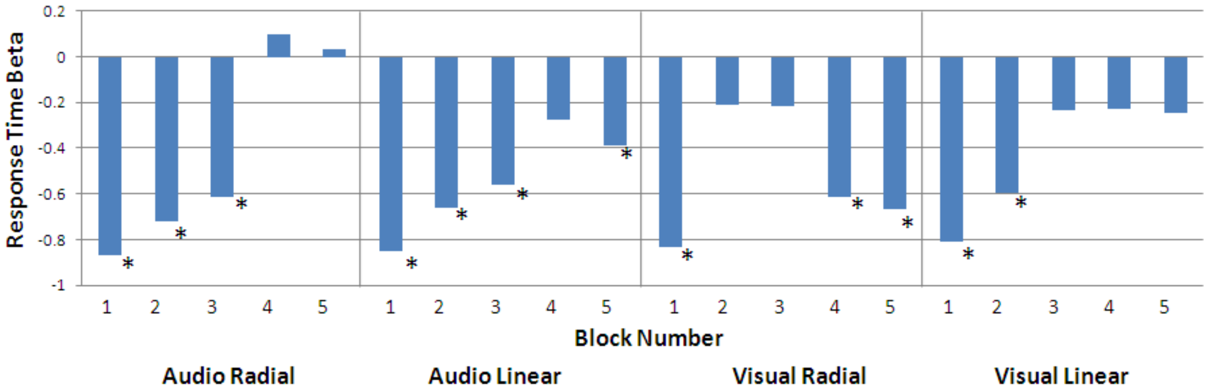


Figure 6.7: Within-day analysis for response time. The beta values per day are graphed separately for the four techniques. Each statistically significant beta value is indicated with a "*" mark.

It can be seen that with the exception of visual radial menus, the speed up in performance ($F_{4, 24} = 11.57$, $p < .01$) across the course of the experiment was much more consistent than the improvement in accuracy that was observed during the study. Using the significance of the corresponding beta coefficients as indicators of significant learning within days it can be seen that significant learning occurs only in the first two or three days for the audio radial and visual linear conditions, whereas except for the fourth day, significant learning continues into the fifth day for the audio linear condition.

6.2.6.3 Detailed log analysis

In addition to the normal measures of speed and accuracy, several additional measures were logged in order to assess other aspects of performance and behavior in this task. The variation in these measures across experimental conditions will now be assessed.

Items explored for each selection this measure records the number of items users traverse before committing to a final selection (include duplicates).

Figure 6.8 shows how this measure varied over different days. Results of repeated measures analysis revealed that there was a significant effect of day ($F_{4,16} = 7.14$, $p < .01$). A test of within subjects contrasts shows that the number of items explored on the first day (4.36 items explored per selection) was significantly higher than the other days (day 2-5 varies from 3.60 – 3.84 items explored per selection) ($p < .05$), while no other significant differences were found ($p > .05$). This indicates that the biggest improvement happened on the first day.

There was a significant effect on condition ($F_{3,4} = 140.67$, $p<.01$), indicating that the number of items explored differed across experimental conditions. Post-hoc Tukey HSD tests show the relationship among the condition being:

visual radial (1.32) < *audio radial* (2.55) < *audio linear* (5.65) ~ *visual linear* (5.80))

where the conditions connected with '<' are significantly different from each other ($p < .05$), and conditions connected with '~' are not ($p >.05$).

There was also a significant day x condition interaction effect ($F_{12, 16} = 7.84$, $p < .01$). After performing separate repeated measures analysis on each condition, it was found that only results in the audio radial condition differed across days ($F_{4,4} = 13.91$, $p < .05$), indicating significant improvement over days, while all other conditions showed no significant improvement on the number of items explored ($p > .05$). Visual radial menus performance was consistently close to the minimum number of one item that reflects direct selection of (tapping on) the target. Linear style menus showed little improvement in the number of items traveled with an average of 5.73 items being explored per selection. This number is higher than the average number of items that are needed for efficient selection in an 8 item menu (4.5 items), suggesting a certain degree of inefficiency. The logs were then examined in more detail to assess what led to this inefficient performance. It was found that much of the extra travel in making a selection was attributable to over-shooting the item.

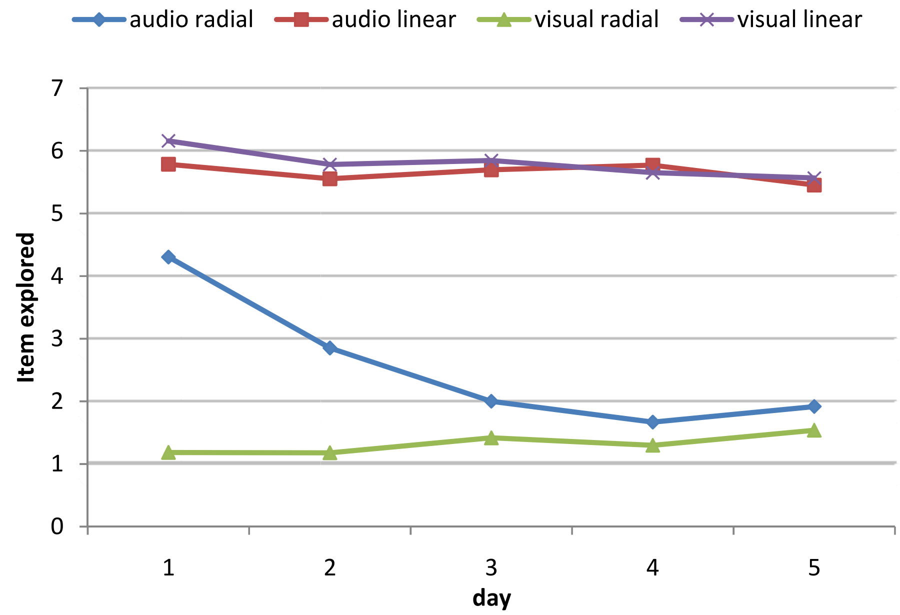


Figure 6.8: The average number of items explored for each selection is graphed for each technique.

Tapping percentage The second measure logged in this study was the percentage of tapping behavior for each day. There was a statistically significant main effect of day ($F_{4,16} = 9.64$, $p < .01$) as well as a significant day X condition interaction ($F_{12,16} = 15.08$, $p < .01$), indicating that tapping percentage changed over the course of the study and that the amount of this change varied between the different experimental conditions. Since linear menus used very little tapping, Figure 6.9 only illustrates the difference that occurred between the visual and auditory radial conditions. Among all the conditions, only audio radial menus showed a steady improvement (increase) in terms of tapping percentage across days ($F_{4,4} = 30.48$, $p < .01$), while the other conditions showed no significant change across days ($p > .05$).

An independent t-test carried out on the data for the first two days comparing the audio and visual conditions showed that the visual condition had significantly more tapping than the audio condition ($t_{10} = 4.36$, $p < .05$) on those days, but a t-test on data for the final three days showed no corresponding difference in tapping percentage between the audio and visual conditions ($p > .05$).

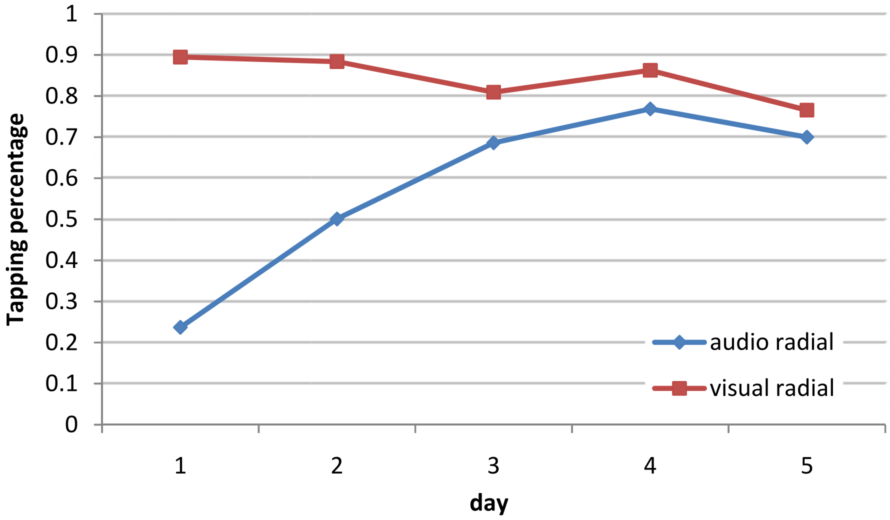


Figure 6.9: The percentage of tapping for audio and visual radial menus. The tapping percentage remains in a high level from day 1 for the visual radial technique while it improves steadily for the audio radial technique.

Gliding speed Gliding speed was calculated using the finger travel distance divided by the time used for each selection. This speed is different between gliding and tapping, which are two different types of behavior. Tapping typically happens in the radial techniques while gliding is the dominant behavior for the linear conditions. Figure 6.10 shows the gliding speed change between the audio and visual linear conditions. A repeated measures analysis for only those two conditions (audio and visual linear) revealed significant main effects for day ($F_{4,8} = 3.92$, $p< .05$) and for the day x condition ($F_{4,8} = 5.27$, $p< .05$) interaction. Figure 6.10 shows that the increase in the gliding speed is much stronger in the audio linear condition than in the visual linear condition. This increase in speed also translates into a greater amount of learning for the audio linear techniques.

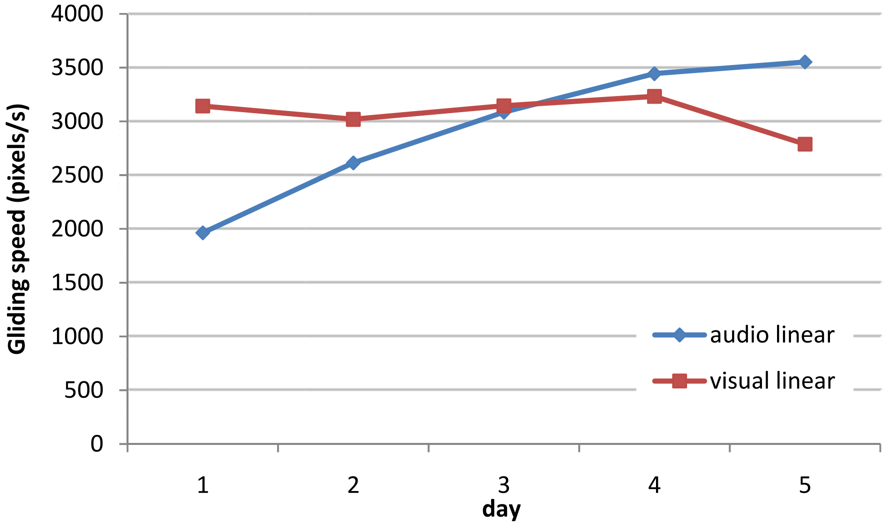


Figure 6.10: The gliding speed for linear techniques is graphed. Audio linear technique has a significant improvement in gliding speed over time while the gliding speed of visual linear technique remains relatively constant.

6.3 Discussion

6.3.1 Overall results and learning pattern

Overall, the results are comparable with earlier results from other researchers (e.g., Kurtenbach and Buxton 1993) and that of Experiment 1 and 2 with one notable exception. The earlier results did not find a significant difference in accuracy for auditory vs. visual modality, nor for linear vs. radial menu styles. In the present study there was a small, but statistically significant difference among these techniques where the auditory techniques were more accurate than the visual techniques, and where the linear menu style was more accurate than the radial style menu. Because a between-subject design is used, it is possible that the differences found in this study may be attributable to individual difference. However, this result may also due to a speed vs. accuracy tradeoff, since the more accurate techniques in this study tended to also be performed more slowly.

6.3.2 Selection performance with a larger amount of content

Since this was a longitudinal study, it was possible to increase both the number of menu items that were used (64 instead of the 16 used in Experiment 1) and the number of times (repetitions) that each item was shown to each participant. The speed of performance at the start of the *earPod* condition in Experiment 1 was roughly equivalent (at just over four and a half sections) to performance in the *earPod* condition of this experiment after 15 to 20 repetitions. However, the response time for *earPod* performance after 15 to 20 repetitions in Experiment 1 had already dropped to an average of 2.5 s, while in this experiment, from block 14-21, the performance was 2.8 s, which is slightly higher, but not much higher. This indicates that users are able to select from a larger amount of content, and the learning curve for each item is not significantly different from testing a smaller sample from the same menu.

6.3.3 Audio vs. visual

One interesting aspect of learning not observed previously is the difference in *accuracy* between the audio and visual modalities. The auditory techniques exhibited consistency in accuracy from day one: no major improvements were observed from either the between-day and within-day analysis. This was in contrast with the visual techniques, where a negatively accelerated power curve typical of learning was observed.

One possible explanation for the relatively high accuracy and lack of learning in the audio conditions is that they were more novel and thus people were more careful in their use from the outset. When users deal with a less familiar interface, they tend to concentrate more on the interaction techniques (Atkins 1993), and resulting in more careful handling of each trial and therefore higher accuracy. The visual techniques, which were more familiar to users, would be less likely to be treated with such caution.

In terms of speed, learning was observed in both the auditory and visual techniques. As expected, the auditory techniques demonstrated stronger learning effects than the visual techniques. However, the convergence of speed performance across the five days between the audio and visual conditions was not anticipated.

6.3.4 Linear vs. radial

Callahan et al. (1988), found no indication of accuracy difference between linear style and radial style menus, consistent with the results of our earlier experiments. In the present study, performance with the linear style menus were significantly more accurate (2% more accurate on average) than the radial style menu. Given that only four participants used each menu style in this study, it is possible that the difference in accuracy observed in this study may be due to individual differences.

6.3.5 Interaction between modality and menu style

Looking at the overall pattern of results in this experiment, it can be seen that the initial speed is more or less determined by the modality of feedback, while the final speed (expert performance) is more or less determined by the menu style (Figure 6.6). This can be explained as a diminished influence of modality of feedback as the user's behavior evolves from a close-loop behavior to an open-loop behavior where feedback is no longer as important.

One unexpected finding from this experiment was that there were learning pattern differences between the auditory linear menu and the visual linear menu. Since tapping behavior is not possible for selecting items for linear menus, and feedback plays an important role even if the users have improved their skills, why does speed using the auditory linear menu catch up with the corresponding speed for visual linear menus by the end of the study? Further analysis shows that as users gain more knowledge of the menu, their listening time for the audio messages is reduced. The average time they spend on each item was reduced from 370 ms per item to 191 ms per item from day 1 to day 5. The last day performance is comparable with visual linear menus, which is 192 ms per item on day 5. This surprising result can be regarded as a strong demonstration of the effectiveness of using interruptible audio in audio interfaces. Users are capable of successfully finding the target item by quickly scanning through partial auditory messages, in a speed even comparable to browsing the same menu visually.

6.4 Summary and implication for interface design

In summary, the results of this experiment further confirmed the viability of auditory based radial menus, indicating that users are capable of remembering relatively large number of menu items (64) with enough repetitions of each item. Under the 80-20 rule where a small proportion of menu items will be accessed most of the time, 64 items represent one fifth of a menu hierarchy of 320 items, and many users might need fewer than 64 items for many applications and for operating mobile devices. As users remember the menu structure and improve their selection skill, the performance difference between auditory based menus and visual radial menus should diminish, largely due to feedback requirements being reduced for expert behavior.

Due to the intrinsic properties of sound, it is difficult for auditory interfaces to compete with visual ones if a significant amount of feedback is required for their operation. However, long term usage performance depends on how well a technique supports expert behavior. If expert behavior can be performed with allow minimum feedback; it can largely reduce the impact of the modality of feedback. When designing auditory-based menus or interfaces, it is recommended that the design should allow expert behavior to be performed using minimum feedback.

Performance with auditory linear menus might not be as bad as was originally assumed. In Experiment 2, auditory linear menus had the worst performance among the techniques; however, Figure 6.6 of this experiment shows that its performance is comparable to visual linear technique on day 5 (audio linear 2.77 s vs. visual linear 2.95 s, $p > .05$). The audio linear technique also improved faster than did the visual linear technique, likely due to the improvement of the gliding speed as shown in Figure 6.10. Post experimental interview indicated that as users started to remember the relative position of each item and got more familiar with the content of the menu, they were able to move faster to the point where the speed of movement in auditory menus was comparable to the speed of movement for visual menu.

The longitudinal study was performed to see how the relative advantages of different menu selection techniques may change over time, and to provide a more detailed account of the transition from novice to expert behavior. The findings are consistent with the previous experiments except the specific findings highlighted above. It is encouraging to see the *earPod* can scale to a full two level menu and still maintain the theoretical advantages intended by its design under a longer testing period.

Chapter 7
Experiment 4: Driving Simulation Study

7.1 Introduction

Although the desktop evaluation results presented in the preceding chapters strongly favored visual or dual interfaces, the desktop condition is not the only place where menu selection occurs. Today's mobile devices are often used while a person is performing another task. In particular, cars are often driven while performing other tasks or dealing with various distractions, and this multitasking environment has attracted considerable research attention.

According to a recent survey of American drivers[12], menus on mobile devices (specifically the iPod) are commonly used while driving, especially by young drivers ages 18-24. Although the use of mobile devices while driving is considered to be dangerous, and its use is or will be prohibited in many jurisdictions (legislation has been introduced in many countries including Australia, France, Germany, Japan, Russia, Singapore, and the UK; for a complete list see [13]), people continue to use their cell phones while driving. For instance, compliance with the UK ban has slipped from 90% at its introduction in 2003 to around 75% in 2007. Today there are some 10 million UK motorists who admit to using a phone while driving, even though this activity is against the law[14].

Previous work on driver distraction resulting from cell phone use suggests that it competes for limited visual attentional resources, thus harming driving performance (e.g., Alm and Nilsson 1994; McKnight and McKnight 1993). Other research suggests that cognitive load alone, separate from perceptual/motor load, is sufficient to produce distraction effects. For instance, Strayer and colleagues in a series of studies (Strayer, Drews, and Johnston 2003; Strayer and Johnston 2001) indicated that the cognitive act of generating a word is sufficient to cause

---

[12] "2006 GMAC Insurance National Drivers Test," http://www.gmacinsurance.com/SafeDriving/2006/

[13] "Countries that ban cell phones while driving", available at http://www.cellular-news.com/car_bans/

[14] "Careless talk", available at http://news.bbc.co.uk/2/hi/uk_news/magazine/6382077.stm

noticeable distraction effects. It is unclear then whether designing a mobile device so that it does not place additional demands on visual attentional resources would mitigate the harmful effects of distraction. The increased cognitive load of interacting, even with an eyes-free device such as the *earPod*, may be sufficient to result in adverse effects for driving performance.

Given that it is difficult to make people stop engaging in secondary tasks while driving, there may be substantial value in directing efforts to better designing mobile devices to make their use by the driver of a car less egregious. That is, a user-centered design approach that is sensitive to the environmental constraints imposed by using a mobile device in the context of an on-going dynamic task.

Experiment 2 provided empirical results for various menu interaction techniques under the desktop setting. It strongly suggested using the *visual modality* as the primary mean of feedback in desktop environment. However, it is an open empirical question whether *visual* interfaces also offer performance gains when the user is concurrently engaged in an ongoing dynamic task, such as driving a car. In particular, because *visual* interfaces demand visual attention, we might assume that this might lead to greater driver distraction than using *audio*. In the next section of this paper, I describe an experiment that tests the use of several menu techniques in a simulated driving environment to address this question.

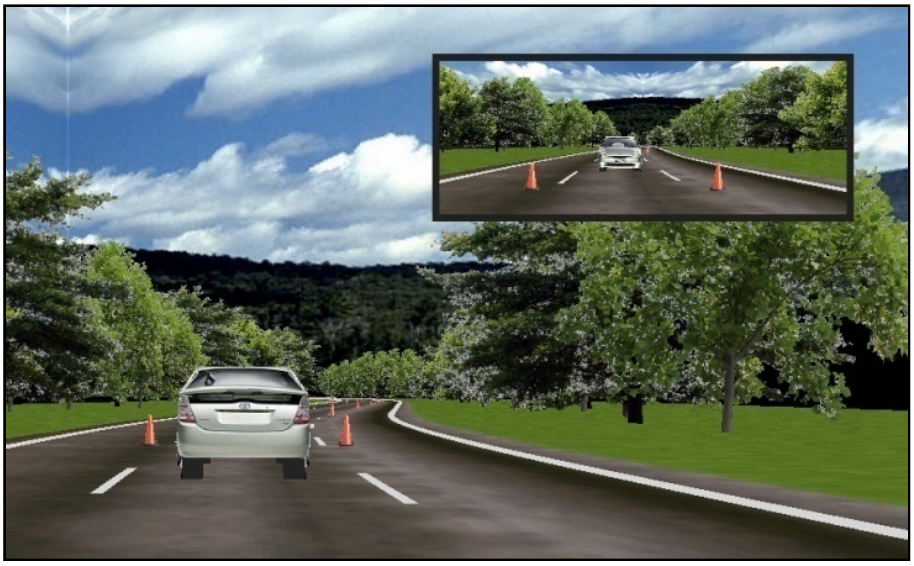


Figure 7.1: Driving simulation environment.

7.2 Experiment

7.2.1 Participants

Twelve participants (1 female) ranging in age from 20 to 35 years (mean 27), recruited within the university community, participated in the experiment. Subjects received $10 dollars per hour for their participation in the study.

7.2.2 Procedure

After completing the desktop trials, participants were asked to answer a set of questions regarding their experience with the desktop conditions. They then moved to the driving simulator and complete the menu selection tasks while driving. During the driving task at least 10 seconds elapsed between the end of one menu-selection trial and the start of the next trial; this time allowed participants to perform any necessary corrective steering after each trial and re-center the vehicle to a normal driving state. (Note that this constraint reduced the number of trials possible in the driving context, but was necessary to maintain the integrity of the driver performance data.) Participants were allowed to take breaks between trials. Breaks were enforced after a maximum of 15 minutes of driving to avoid fatigue. Before each of the desktop and driving conditions, participants received 8 practice trials (1 block) for that particular interface. Each participant performed the entire experiment in one sitting which took approximately 90 minutes (the desktop condition typically finished within 20 minutes, and the driving condition typically lasted 40 minutes, with the extra time being used for questionnaires and breaks). After completion of the driving session, the same set of questions was asked again regarding user experience during the trials.

7.2.3 Task and stimuli

7.2.3.1 Driving setup and task

The driving experiment was conducted using a desktop driving simulator. The simulation environment, coded in Java with OpenGL graphics, incorporates a three-lane highway with the driver's vehicle in the center lane, as shown in Figure 19. The highway includes alternating straight segments and curved segments with varying curvatures, all of which can be driven at normal highway speeds. Navigating in this environment, drivers were asked to follow an automated lead car that runs at a constant speed of 65 miles/hour (~105 km/h) and to maintain a reasonable, realistic following distance. A second automated vehicle, visible in the rear-view mirror, followed behind the driver's car at a distance of roughly 50 feet (15 m), to provide an incentive for the driver not to fall too far behind the lead car. Construction cones were placed on each side of the driver's lane to motivate as accurate lane keeping as possible. Previous versions of a very similar environment have been used to study various aspects of driver behavior (e.g., Salvucci 2005; Salvucci and Macuga 2002; Salvucci 2001).

The hardware setup comprised of a desktop computer controlled by a Logitech MOMO® steering wheel with force feedback. The simulation was run on an Apple desktop computer with an Intel Xeon CPU running at 2.00 GHz, 2 GB RAM, and an NVIDIA GeForce 7300 GT graphics card. The environment was displayed on a 30' (69 cm) monitor at a distance of roughly 33' (85 cm) from the driver. For added realism, a soundtrack of real driving noise was run continuously during the driving portions of the study.

7.2.3.2 Mobile device setup and task

The experiment was divided into two parts. The first part replicated Experiment 2 except it was much shorter. This setting allowed users to get familiar with the menu and the interaction techniques before they moved to the second and arguably more difficult part: driving while carrying out menu selection at the same time. The experiment was designed to simulate a realistic usage scenario.

For the desktop conditions, the setup was exactly the same as for Experiment 2 except the following exception.

1). Instead of using visual feedback for menus only, both audio and visual feedback were provided simultaneously. This allowed users to pick which form of feedback they preferred to attend to in different interfaces. Since driving is very different from desktop interaction, the visual feedback could have been a possible source of distraction, thus complicating the interpretation of results. However, using audio feedback alone would not have permitted analysis of modality effects. To address these issues, I decided to provide both types of feedback and allow users to pick the one they like to attend to during the experiment. In order to facilitate comparison between the desktop and driving conditions, both audio and visual feedback were also provided in the desktop condition.

For the mobile task setup in the driving conditions, an additional change was introduced.

7.2.4 Design

A within-participants design was used. The exact design is summarized below.

Desktop condition:

12 participants x

6 techniques x

8 items of 1 menu configurations: (condition 8) x

4 blocks (3 blocks + 1 practice block for desktop conditions)

\+

Driving condition

12 participants x

6 techniques x

8 selection of menu items: (condition 8) x

2 block (1 block+1 practice block for driving conditions)

= 4032 menu selections in total (2880 + 1152).

7.2.5 Design rationale

The overall experiment design, especially for the conditions carried out under desktop settings, follows closely that of Experiment 2, and thus it is subject to similar tradeoffs. In the following section, only the differences between the two experiments are highlighted. The dual task simulated driving condition is the focus of this experiment, but the desktop condition is also essential since it provides the necessary training for users to get familiar with techniques. This closely simulates real world scenarios where users typically already have some experience with their devices before using them inside vehicles.

Due to the extra time requirement imposed by the driving condition, only a one-level menu instead of two-level menus is used due to less time requirement for each trial and because a one-level menu has shown to be a fairly good indicator of the general properties of a technique. The number of trials was adjusted in both the desktop and driving conditions to fit a 90 minute testing limit. To make the driving task more realistic, curvy roads were used.

7.2.6 Results

7.2.6.1 Desktop Setting

Using repeated measures analyses, both accuracy and response time results for the desktop setting were consistent with results obtained in experiment 1. Actual numbers varied, but no change was found concerning significance of effects. In order to control for possible effects of asymmetric transfer between modalities, a mixed analysis of variance was also run with only the data for the first modality experienced by each participant being included in the analysis. These analyses also yielded similar results, demonstrating that the observed effects were not attributable to the effects of asymmetric transfer between modalities when different modalities were used in sequence. The results of these analyses are summarized in the following paragraph.

There was a significant main effect of modality ($F_{2,9} = 28.26$, $p < .01$), where performance in the audio modality (4.18 s) was significantly slower on average than both dual (2.12 s) and visual (2.35 s) modality performance. There was also a significant main effect of menu style ($F_{1,9} = 13.07$, $p < .01$). Linear menu style (3.28 s) was significantly slower on average than radial menu style (2.49 s). A significant interaction effect for modality x menu style was also found ($F_{2,9} = 7.72$, $p<.05$). Figure 7.2 shows the effect of the interaction which is consistent with the pattern of results obtained in the repeated measures analysis of the full set of data.

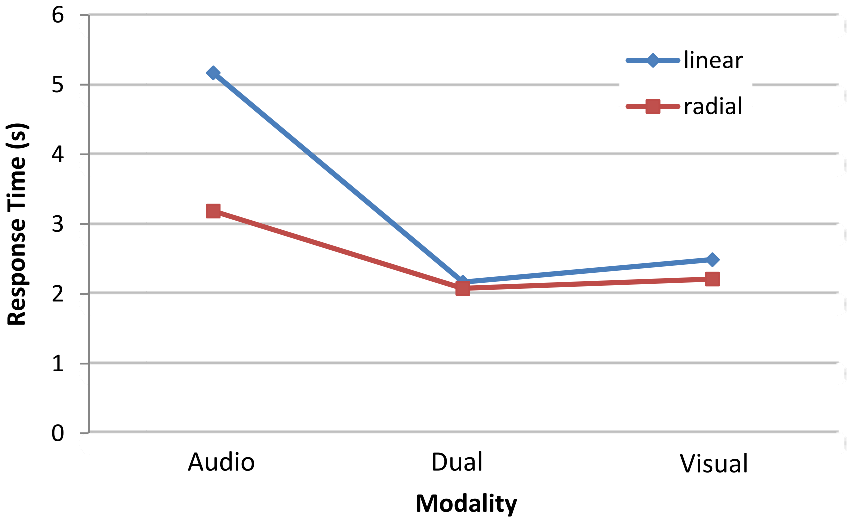


Figure 7.2: Interaction effect between modality and menu style in the between-subject analysis for Experiment 4.

7.2.6.2 Observations & subjective preference

Although this experiment differed little from Experiment 2, a set of additional questions in the post experimental questionnaire allowed us to gain more insights into users' experience. Since I used both *visual* and *audio* stimuli in this experiment, users were asked "Which stimuli did you attend to during the experiment?" Most participants reported using *visual* (11/12 subjects), with the remaining participant reporting the use of both types of feedback. For the question "Which feedback modality did you use under the dual modality conditions?" The answers were again consistently "visual" or "primarily visual". For the question: "If you only used one type of feedback or primarily used only one type of feedback, did you found the other kind of feedback (audio or visual) distracting?", several users answered "Yes, I found the audio feedback a bit distracting", while most subjects said "No". Based on this feedback, it's clear that the visual modality is preferred under the single task desktop environment.

7.2.6.3 Results in the Driving Condition

7.2.6.3.1 Accuracy

There were no significant differences in accuracy for either *modality* (*audio* 88.9%, *visual* 88.3%, and *dual* 88.8%) or *menu style* (*radial* 88.3%, and *linear* 89.0%). This is consistent with the findings from the desktop settings.

7.2.6.3.2 Response Time

There was a significant main effect of *menu style*, ($F_{1,11}$=32.86, p<.001). *Radial* (3.34 s) was significantly faster than *linear* (4.12 s), which is also consistent with the findings from the desktop conditions. The average selection time for the six interfaces was: *audio radial* (3.27 s), *audio linear* (4.09 s), *audio visual radial* (3.53 s), *audio visual linear* (4.15 s), *visual radial* (3.27 s), and *visual linear* (4.10 s). There was no significant main effect of modality on response time while driving. This is somewhat surprising since response time for *audio* was significantly slower than for *visual* in the desktop conditions.

Between-subject analysis was also performed on the driving condition data (i.e., where only the first modality experienced was used, so that modality became a between subjects factor) to ensure that the results obtained using repeated measures analysis were not attributable to the effects of asymmetric transfer between modalities when experienced in different orders. The results were consistent with the within-subject analysis. No significant effects were found for accuracy. For response times, there was a significant effect of menu style ($F_{1,9} = 11.55$, $p < .01$), with the linear menu style (4.35 s) being significantly slower on average than the radial menu style (3.56 s). The comparison among different modalities was again not significant.

7.2.6.3.3 Lateral Velocity

In testing interaction in the driving context, arguably the most important aspect of this interaction is the effect on driver performance. One common way to measure performance involves analysis of the vehicle's lateral (side-to-side) velocity as an indicator of vehicle stability. I computed the average lateral velocity over a time window that included the interaction time with the device plus a period of 5 seconds after the completion of the interaction; this latter period accounts for vehicle "correction" that typically takes place after distraction — during which the driver corrects the lateral position of the vehicle — which is best attributed to the immediately preceding interaction trial.

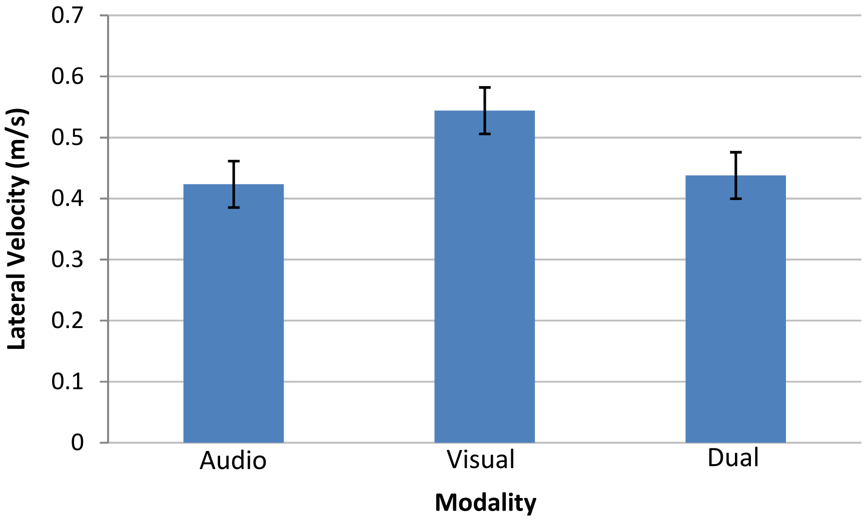


Figure 7.3: Lateral velocity by modality.

For this experiment, I found a significant effect of *modality* ($F_{2,22}$=5.63, p<.05) but no significant effect of *menu style* (p=.35) and no significant interaction between modality and menu style (p=.82). The effect of modality is shown in Figure 7.3. Pairwise comparisons showed no significant differences between the *audio* and the *dual* modalities, but both of these modalities differed significantly from the *visual* modality (p<.05). The lower lateral velocity (i.e., higher stability) for the *audio* versus the *visual* condition indicates, not surprisingly, that the visual attention needed for the *visual* condition caused additional distraction, leading to worse performance. Interestingly, the *dual* condition produced essentially the same reduction in the distraction effect as the *audio* condition, suggesting that drivers relied on the audio portion of the *dual* interaction while driving (which is supported by the drivers' post-experiment reports as discussed below).

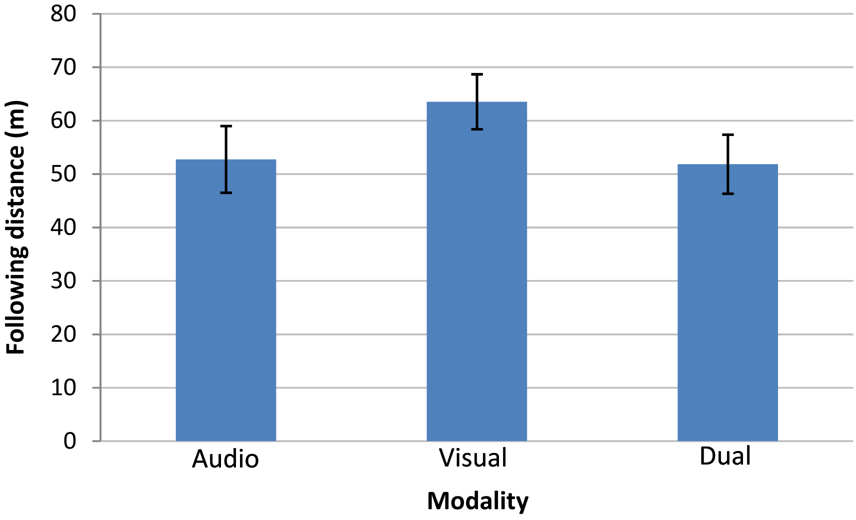


Figure 7.4: Following distance by modality.

7.2.6.3.4 Following Distance

Lateral velocity is a measure of the results of distraction on driver performance. Another measure of distraction is the following distance to the lead car: in essence, as drivers feel themselves being distracted, they tend to back away from the car in front of them for safety reasons. I computed the average following distance using the same time window around a particular trial as used for the analysis of lateral velocity.

Overall, as for lateral velocity, drivers exhibited a significant effect of *modality* ($F_{2,22}$=7.34, p<.01) but there was no significant main effect of *menu style* (p=.75) nor was there a significant interaction between modality and menu style (p=.33). The average following distances by modality are shown in Figure 7.4. It can be seen that both following distance and lateral velocity

(Figure 7.3) are affected by modality in a similar way, suggesting that the effects of visual distraction affect both lateral velocity and following distance in a similar way. In terms of following distance, drivers responded to the increased distraction by backing off from the lead car, giving themselves more room for error.

7.2.6.3.5 Observations & subjective preference

When the setting shifted from single task desktop to dual task in the driving simulator, the responses to the post-experiment questions also changed. When questioned after the driving condition, for the question, “Which stimuli did you attend to during the experiment?” most participants said *audio* (10/12 subjects), with two participants saying both audio and visual. For the question “Which feedback modality did you use under the dual modality conditions?” The answers were again consistently “audio”. All participants felt that *audio* was much safer to use than *visual* while driving. For the question: “If you only used one type of feedback or primarily used only one type of feedback, did you find the other kind of feedback (audio or visual) distracting?”, most users reported that they totally ignored the visual feedback thus turning the dual modality interface into an audio only interface. However, users who occasionally glanced at the visual interface found the visual feedback not just distracting, but dangerous.

7.2.6.4 Desktop vs. driving

In the analyses reported in this section the results for the desktop conditions were compared with those for the driving condition. A new factor called *experiment type* was introduced into the analysis. The *experiment type* had two possible values: *desktop* condition and *driving* condition.

7.2.6.4.1 Accuracy

There was a significant main effect of experiment type, ($F_{1,11}$=27.26, p<.001). The mean accuracy for *desktop* conditions (94.4%) was significantly higher than for the *driving* conditions (88.6%). This is not surprising since the user had to perform a more difficult task (2 selections in a row), and also had to deal with a secondary driving task.

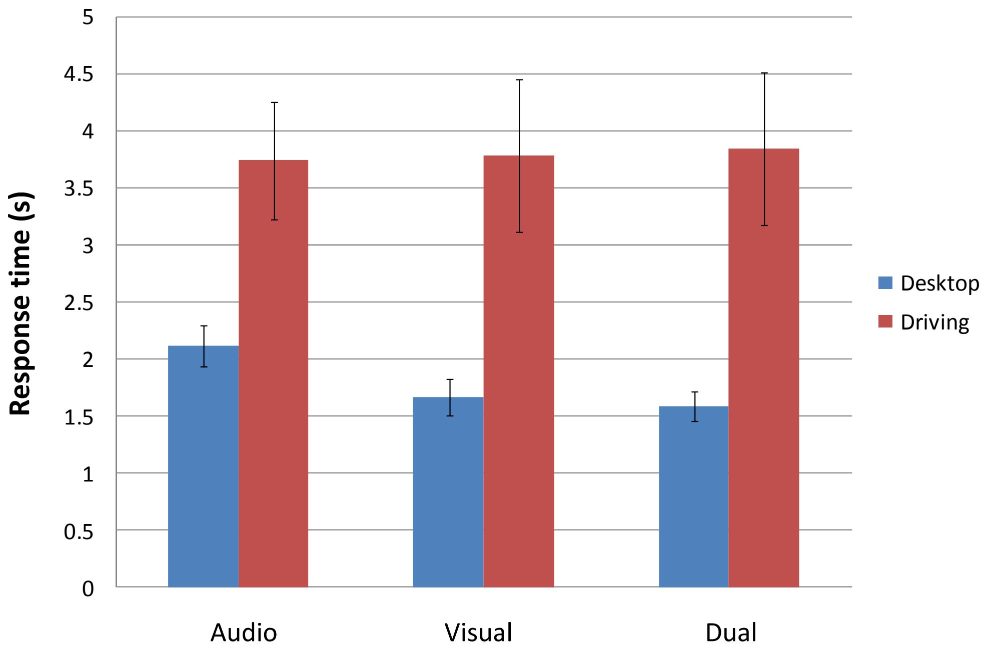


Figure 7.5: Experiment type x modality interaction

7.2.6.4.2 Response Time

There was a significant main effect of *experiment type* on response time ($F_{1,11}$=133.74, p<.001). The mean selection performance in the *driving* conditions (3.73 s) was significantly slower than the corresponding performance in the *desktop* conditions (2.63 s). This delay is likely due to the secondary driving task.

There was a significant *experiment type* x *modality* interaction, ($F_{2,22}$=12.13, p<.001). While response time was significantly slower for the audio conditions in the desktop settings, it was no slower than the other conditions while driving (Figure 7.5). This finding along with the empirical data obtained on lateral velocity and following distance all strongly suggest that the audio modality may be useful in driving, since it may increase safety without harming performance when interacting with a device.

There was also a significant *experiment type* x *menu style* interaction, ($F_{1,11}$=32.98, p<.001). A closer examination indicates that the *radial* menu style has a larger advantage in terms of response time than the *linear* menu style for the *desktop* setting (Figure 7.6).

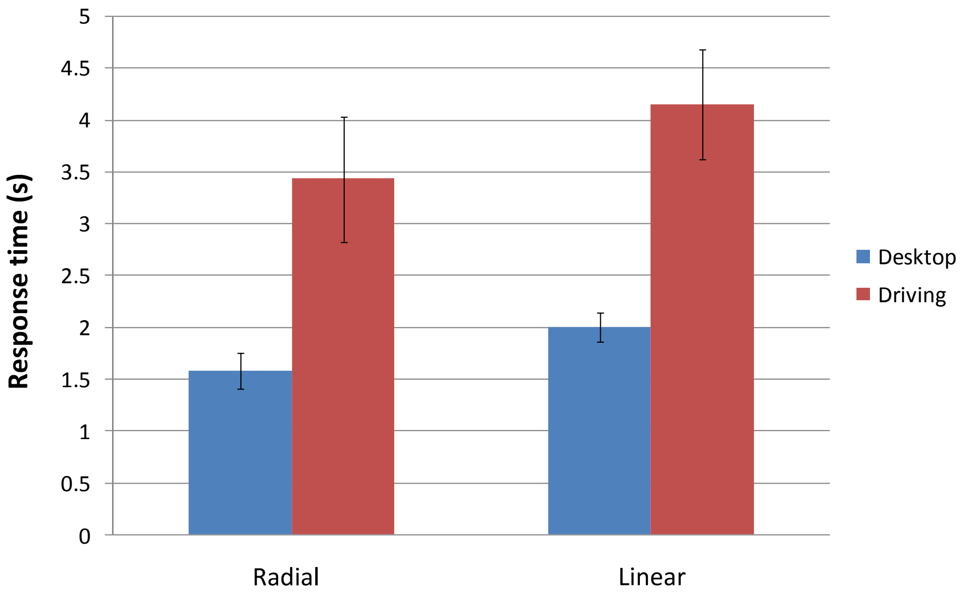


Figure 7.6: Experiment type x menu style interaction

7.3 Discussion and design recommendations

The results obtained differed substantially between the desktop and driving conditions. While the visual modality was the dominant and favored method of feedback in the desktop conditions and provided the best overall performance, audio feedback was found to be safer (to the extent that lateral velocity and following distance predict safety) and preferred while driving. There were also interesting and distinct usage patterns of the dual modality interfaces under different scenarios. All these effects, along with design recommendations for both desktop and driving scenarios, and general discussions of interface design for mobile and ubiquitous computing, are provided in detail below.

7.3.1 Differences between Modalities

The most dramatic differences were found for the *modality* of feedback. In fact, the effect of modality on performance and subjective feedback differed markedly between the desktop and driving conditions.

7.3.1.1 Desktop condition

Visual feedback was generally preferred by users, although 2/12 participants told us they preferred audio even in the desktop settings. However even for them, performance on the visual and dual interface was much faster than the audio interfaces. Thus visual feedback is advantageous in this setting. Users' experience for the dual modality interface was interesting. These interfaces received the highest overall ranking, and were ranked either as the favorite or as the second favorite interface. However, for users who strictly preferred a single kind of feedback their reaction toward the other modality differed. Users who preferred visual interfaces tended to find the audio slightly annoying. However, people who preferred audio feedback were not affected by the presentation of the visual interfaces and tended to rank them as acceptable alternatives. This is perhaps due to ability for people to close their eyes or not pay attention to the screen if they are tired of looking (Gaver 1997).

Overall, visual and dual interfaces were the favorites for desktop settings. Perhaps the best strategy for the desktop setting is the dual interface but having the ability to turn off the audio or visual feedback if they wish to do so.

7.3.1.2 Driving setting

The change in user reaction between the desktop and driving settings was dramatic. In contrast to the desktop condition, audio was consistently judged to be much better than visual for driving. This was true even for users who strongly preferred visual feedback in the desktop settings. One such user said after completing the driving conditions: “Although I prefer visual feedback for desktop, I found it completely useless while driving, where audio is much better.” For the dual interfaces in the desktop setting, users tend to use both modalities of feedback when performing trials. While driving, most users completely ignored the visual feedback. Even for users who occasionally glanced at the screen for extra information, they felt negative about it. As one subject put it, “having the option to look at the visual information while driving is a potential safety hazard. I found myself tending to look at it while I was having difficulty finding the desirable item through audio, but it felt very dangerous, and I prefer an audio-only interface since it doesn't allow me to look at all.”

7.3.2 Linear vs. radial

Compared to modality, *menu style* had a less dramatic effect, but still generated some interesting findings. Under both desktop and driving conditions, *radial menu style* yielded better performance than *linear*, and was more preferred by users. However, compared to the desktop settings, the radial menu style did relatively better in terms of speed in the driving conditions, as indicated by the *experiment type* x *menu style* interaction effect shown in Figure 7.6. This indicates that in the more difficult driving environment, there is actually more incentive to switch to the radial menu style if possible.

7.4 Summary and implications for design

Based on the results of the experiments reported in this paper, I offer the following recommendation for the design of menu selection interfaces for use in single-task (i.e., desktop-like) conditions or while the user is engaged with an on-going dynamic task (such as while driving a car).

For desktop settings, *radial* and *linear menu style* each have their advantages — *radial* style is quicker to access, while *linear* style is more flexible and easier to design the structure and content of the menu. If the designer has a fixed-sized static menu, using visual or dual radial layout would likely yield better performance. Otherwise, visual or dual linear menu styles are also usable and are perhaps more suitable for menus that are longer or that have dynamic content.

For driving conditions, I recommend audio radial menu interaction for fixed-sized static menu. If the menu size is longer or contains dynamic content, audio linear is probably more suitable, and visual interfaces are not recommended because of their interference with the primary driving task. Although audio interfaces are likely safer under driving conditions, they still impose a cognitive load which could may affect driving performance and reduce overall safety.

Overall, if such devices need to be used in both conditions, I would recommend dual interfaces, but providing users with the option to “turn-off” either modality of feedback if desired. One particularly interesting design idea for driving conditions is that the system be augmented with an intelligent sensing technology, which could automatically turn-off the visual display in order to prevent any temptation for the driver to look at the display while driving, or could disable interaction with the mobile device in situations where the driving task becomes difficult (e.g., when driving through a busy intersection).

Chapter 8
Conclusion

8.1 Introduction

This thesis introduced the *earPod* eyes-free menu selection technique. *earPod* combines gesture-based touch input with reactive, interruptible audio feedback, and was shown in the research reported above to be a relatively effective and efficient method of menu selection. Experiments 1 and 2 investigated *earPod* together with a number of related techniques varying in *modality of feedback* and *menu style*. Experiment 3 was a longitudinal study that compared the learning curves for *earPod* and related techniques. All the techniques considered in this dissertation were then tested in Experiment 4 in a dual task simulated driving scenario.

This concluding chapter summarizes the contributions made in this dissertation, and points out limitations and opportunities. Section 8.2 focuses contributions in the following three categories: *earPod* methodology (section 8.2.1), empirical results (section 8.2.2), and design guidelines (8.2.3). Section 8.3 discusses some limitations of this research. Section 8.4 points out some opportunities for future research. Section 8.5 concludes with some final remarks.

8.2 Contributions

The contributions made in this research fall into three categories: a) designing and fine tuning the *earPod* technique, b) conducting experiments and analyzing the empirical results, c) forming design recommendations inferred from the results and observations obtained in the four experiments that were carried out.

8.2.1 *earPod* methodology

earPod combines gesture-based touch input with reactive and interruptible audio feedback. The two main contributions of its design concern the innovative use of reactive and interruptible audio feedback, and the use of both gliding and tapping user interaction that facilitates the transition from novice to expert performance.

Contribution 1: Development of a method for unifying relative (gliding) and absolute (tapping) menu access in a circular touch input, whilst provide a smooth and seamless transition between those two methods of access.

The intent of this design is to facilitate novice-friendly relative auditory scanning using gliding movements and permit rapid and natural transition to an expert mode of behavior that utilizes tapping gestures. Both behaviors use the same input circular touch-based input device, and practice obtained in using gliding naturally tends to lead to tapping.

Contribution 2: Development of an innovative eyes-free menu selection method with touch input and reactive audio feedback.

Reactive and interruptible audio allows users to proactively discover available options at their own pace. They can control whether or not they hear (by touching or airing or by moving to or avoiding menu items using the finger), how much they want to hear about each item by control the speed of gliding. By moving the finger back and forth, the user achieves the effect of "scanning and comparing" the menu items without having to commit them to memory.

Contribution 3: Development of a method for using continuous spatialized audio feedback during menu selection to reinforce the user's cognitive mapping between menu items and spatial locations on the touchpad.

Spatialized audio feedback provides an additional cue as to the user’s current location in the menu selection space. Since *earPod* is an absolute technique where menu items are assigned to spatial positions, learning the mapping between items and positions is a necessary step for transitioning to expert (tapping) performance, which may be facilitated using spatialized audio feedback.

8.2.2 Empirical results

Contributions 4-7 summarize the major empirical results.

Contribution 4: Demonstration, through experimental results, that performance with the earPod method was generally comparable with performance obtained using a visual linear menu.

When compared with an iPod-like visual linear menu in both one and two level menus in experiment 1, *earPod* was comparable in both speed and accuracy. With a small number of content items to be learned (16) *earPod* was significantly faster than iPod-like menu selection by the end of the session in experiment 1, but balanced against this the accuracy for iPod like selection was greater at the end of the longitudinal study (Experiment 3). Results from experiment 2 and 3 for the *earPod* and iPod-like techniques were generally comparable.

Contribution 5: Demonstration, through experimental results, that learning was greater in the earPod condition than in the visual menu selection conditions.

earPod menu selection tended to start slower but end up either faster than, or as fast as, visual menu selection. The higher learning rate for response time was also shown for audio selection in linear menus. However, the generalizability of this result will need to be tested in future research since there was some evidence of a speed accuracy tradeoff operating in the studies reported in this dissertation. There was a tendency in some conditions for *earPod* performance to be faster, but less accurate than the other menu selection techniques.

Contribution 6: Demonstration, through experimental results, that earPod menu selection outperforms other techniques in the context of a visually demanding primary task.

When used as a secondary task, *earPod*'s performance is comparable to that of visual radial menus, and is faster than that of visual linear menus. It is also safer to perform menu selection using *earPod* than either of the visual techniques in a simulated driving setting. These findings strongly suggest that the audio modality may be useful in driving, since it may increase safety without harming performance when interacting with a device.

Contribution 7: Demonstration, through experimental results, that transition from novice to expert performance in earPod menu selection can be relatively fast, but is dependent on the number of menu items to be learned.

The results of experiment 3 indicate that about 70-80% of the 64 items can be remembered after 20 times (repetitions) of use. This number of repetition reduces to below 20 when only 16 items need to be remembered.

8.2.3 Design recommendations for using auditory interfaces in mobile devices.

Contribution 8: Derivation of a design recommendation to use visual menus under single task settings, and to use auditory menus in a dual task setting when a visually demanding primary task is involved.

Empirical results from this research show that it is beneficial to allow audio and visual interfaces to co-exist in the same device and to let them be used according to different scenarios and users' specific needs. The methods for menu selection should be invariant to the interaction technique used, as demonstrated by the *earPod* and visual radial menus, where both visual and auditory interactions work in a consistent manner, meaning that the user need learn only a single mental model of the methods required to select an item from the menu.

Contribution 9: Derivation of a design recommendation where earPod or visual radial menus are recommended for menus with a static structure and a maximum breadth of 12 or less.

One of the conclusions drawn from this research is that the decision about when to use absolute vs. relative (list) menus in audio selection should depend on the amount, structure, and dynamism of the content. When their use is indicated, menus with absolute and direct access methods have better performance than menus that only allow relative (linear) access, especially for expert users.

8.3 Limitations

This research tested static menus with fixed breadth. The radial layout permits a breadth of up to 12 items. The expert tapping mode of *earPod* will not work well with dynamic menus, nor with menus with breadth over 12 items. In these cases, alternative techniques need to be used.

The menu items used in this research were limited to one-word items. It would be desirable for future studies to examine performance on real world menus where items are not limited to one word and where the items have a practical meaning (e.g., as functions or settings within a particular applications). Interruptible audio is most effective for menus consisting of one-word items. When menu items consist of multiple words or sentences, interruption may cause users to miss important information, resulting in selection errors. This is especially problematic for novice users who are not familiar with menu content. If users listen to the entire message, performance will be penalized. However, this is less of a problem for users already familiar with the menu content. More investigation is needed to reveal the tradeoff between interruption and item length.

The longitudinal study that was conducted had a limited number of participants. It is possible that some of the observed significant difference in accuracy between audio and visual modality as well as between linear and radial menu styles may be due to individual differences. It would be useful in future research to look at larger sample sizes and also to perform a longer study (e.g., more than one week long, or more than 35 repetitions per menu item) to determine if the audio radial technique will reach the performance of visual radial menu at asymptote, and to observe whether or not the audio linear technique can outperform the visual linear technique after sufficient learning has taken place.

Experiment 4 was a cross-sectional study and did not look at long term learning effects. Given more time, it is possible that some of the users in our experiment might have learned to drive more “safely”, even in the visual condition. In general, performance improves following a power law of practice (Newell and Rosenbloom 1981). Further research is required to investigate asymptotic performance for menu selection while driving. However, since driving is considered a high-risk task, the potential cost associated with in-car training is extremely high. Even if an interface could eventually be used safely during driving (after extensive practice), any mistakes during practice could potentially be catastrophic. Thus the current experimental setting may have practical value in guiding safe vehicle interface design.

8.4 Future work

As previously mentioned, the design and evaluation of *earPod* are only one step towards the grand vision of establishing eyes-free interface as a new interaction paradigm. Many directions are possible for future work.

earPod was shown to lead to relatively efficient performance, but there is still room for further technological refinement of the technique. Several directions are possible in this regard.

Further improvement for the hardware and software is one way to fine tune the interface. In the current approach, finger lifts are treated as button clicks to trigger menu selections. User experience may be improved if a clickable touch-sensitive device, such as Apple's ClickWheel, is used. As a research prototype, the hardware and software used by *earPod* has not gone through the typical extensive optimization phase experienced by most industrial products. The method of selection detection can also be further tuned to improve accuracy.

earPod currently uses only left to right spatial positioning based on time and intensity differences between the ears. Head Related Transfer Function (HRTF) may be used to enhance the spatialization effect. In certain environments, such as cars, sophisticated spatialization without wearing a headset, using external speakers.

In addition to hardware refinement and optimization of software, two strategies may further improve the audio interaction efficiency: audio compression and dichotic listening. It is known that compression can speed up audio playback up to 2 times without losing clarity (Arons 1997). Future studies should explore the effect of audio compression on *earPod* performance. Dichotic listening is another possibility to increase information throughput using audio. Ranjan et al. (Ranjan, Balakrishnan, and Chignell 2006) found that participants were able to search audio content faster if two different audio tracks were presented simultaneously to each ear. Although our preliminary tests showed that overlapping audio signals might confuse users, this limitation might be overcome with additional design and research. However, both compression and dichotic listening may increase the cognitive load when interacting with *earPod*, which may not be desirable for mobile multi-task interactions.

To further understand the impact of a particular design feature on cognitive load, methods are needed to quantify the amount of cognitive load required for techniques under different scenarios. Designing methods to empirically measure cognitive load is a promising future research topic for human-computer interaction in general. Accurate measurement of cognitive load may lead to improved theoretical models of menu selection using auditory feedback.

Compared to the extensive research literature that already exists concerning the breadth vs. depth tradeoff in visual menus, with its consensus view that breadth is generally preferable to depth, the picture is much less clear for auditory menu interaction. For auditory interaction, much of the prior work has focused on designing IVR, but the results from those studies have been mixed, with some results favoring breadth over depth while others have favored depth over breadth. While selection in visual menus can be modeled fairly well with the Hick-Hyman and Fitts' Laws, there is no comparable predictive model for the time or effort associated with auditory menu selection. Thus it would be useful in future to develop both theoretical models of auditory menu selection as well as better understanding of the breadth vs. depth tradeoff in audio menu interaction.

The results obtained concerning menu interactions in simulated driving should be validated in in-vehicle studies before being applied to the design of menu interactions whilst driving. In addition to driving, there are other common usage scenarios that were not evaluated in this dissertation. One possible direction for future work is to evaluate these interfaces in other mobile scenarios such as walking, running, etc. Future research could also compare the mental load of auditory-based techniques with visual techniques and examine the cognitive mechanisms that drive differences between audio and visual interactions in dual task settings.

earPod was shown to be effective for static menu hierarchies of reasonable size in this dissertation. Strategies for traversing dynamic menus and menus of arbitrary length, such as a song list are worth exploring in future studies. Menu selection is only one of many possible tasks in an auditory interface. Menu selection was chosen as the focus for this research since command selection is a fundamental building block for more complex applications. However, development of an all-encompassing interface for eyes-free operations on auditory devices is a task for future research. Other tasks, such as drag and drop for the auditory interface are also worth exploring. Various interaction tasks can have different associated access models and therefore different properties for design. Further investigation may lead to a core set of auditory interaction techniques and may establish eyes-free interaction as a new interaction paradigm.

Finally, it will be worthwhile to study eyes-free interaction in the context of real world applications. Many popular applications, such as audio conferencing and instant messaging, are starting to migrate to the mobile arena. If navigating the environment also requires close visual attention, using these applications simultaneously becomes difficult. The emergence of eyes-free interfaces may provide more freedom for users to use these applications in scenarios where they are currently difficult or impossible to use.

8.5 Final remarks

Desktop computing has been dominated by visual interfaces. Auditory interfaces have generally been restricted to specialized devices such as telephones and MP3 players. However, as mobile devices become more widely deployed and used in a wider range of settings, there is an increasing need to develop truly eyes-free interfaces. As a first step in the process of developing innovative and comprehensive auditory interfaces, the research in this dissertation examined the problem of auditory menu selection. Touch input and reactive auditory feedback were combined in a synergistic way to create an innovative method of eyes-free menu selection that was shown to have favorable performance characteristics in a series of experiments.

There seems to be no inherent reason why the current approach could not be extended to other aspects of the auditory interface. For instance, using spatialized audio, an audio space could be constructed as an analog to the “desktop and windows of the currently widely used graphical user interface. Mobile phones represent a particularly exciting platform on which to build futuristic auditory interfaces.

As is typically in research, this dissertation may have raised more questions than it answered, Individual differences, the applied psychology of using *earPod* given the availability or lack thereof of various cognitive resources, patterns of learning for different menu architectures and types of menu (e.g., dynamic menus) are examples of many issues that could be addressed in future.

However, specific research results aside, perhaps the most important point of this research was to show that innovative design can overcome some of the apparent limitations of the auditory modality. It took years of innovative research and design to construct the modern visual interface that we now take so much for granted, so why should the level of effort be any different for designing the auditory interface of the future?

References

Ahlstroem, D., R. Alexandrowicz, and M. Hitz. 2006. Improving menu interaction: a comparison of standard, force enhanced and jumping menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 1067-1076. Montréal, QC, Canada.

Alm, H., and L. Nilsson. 1994. Changes in driver behaviour as a function of handsfree mobile phones--a simulator study. *Accident; Analysis and Prevention* 26, no. 4 (August), pp. 441-451.

Arons, B. 1997. SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transaction on Computer-Human Interaction* 4, no. 1, pp. 3-38.

Atkins, M. J. 1993. Evaluating interactive technologies for learning. *Journal of Curriculum Studies* 25, no. 4, pp. 333.

Bailly, G., E. Lecolinet, and L. Nigay. 2007. Wave menus: Improving the novice mode of marking menus. *Proceedings of International Conference on Human Computer Interaction (INTERACT)*, pp. 475-88. Rio de Janeiro, Brazil.

Bailly, G., E. Lecolinet, and L. Nigay. 2008. Flower menus: A new type of marking menus with large menu breadth, within groups and efficient expert mode memorization. *Proceedings of advanced visual interfaces (AVI)*, pp. 15-22. Napoli, Italy.

Balentine, B., and D. P. Morgan. 1999. *How to build a speech recognition application: a style guide for telephony dialogues*. San Ramon: Enterprise Integration Group.

Bernstein, L. R. 1997. Detection and discrimination of interaural disparities: Modern earphone-based studies. *Binaural and Spatial Hearing in Real and Virtual Environments*, pp. 117-138. NJ, USA.

Berry, G. 1997. The esterel primer. Technical report, Ecole des Mines de Paris and INRIA.

Blattner, M, D. Sumikawa, and R. Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 4, no. 1, pp. 11-44.

Box, G. E. P., J. S. Hunter, and W. G. Hunter. 2005. *Statistics for Experimenters: Design, Innovation, and Discovery. 2nd. sl.* John Wiley & Sons, Inc.

Brewster, S. A. 2003. Nonspeech auditory output. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, pp. 220-239. Lawrence Erlbaum Associates, Inc.

Brewster, S. A., J. Lumsden, M. Bell, M. Hall, and S. Tasker. 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 473-480. Ft. Lauderdale, FL, USA.

Brewster, S. A. 1998. Using nonspeech sounds to provide navigation cues. *ACM Transaction on Computer-Humam Interaction (TOCHI)* 5, no. 3, pp. 224-259.

Brewster, S. A., and P. G. Cryer. 1999. Maximising screenspace on mobile computing devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*, pp. 224-225. Pittsburgh, PA, USA.

Brewster, S. A., P. C. Wright, and A. D. N. Edwards. 1993. An evaluation of earcons for use in auditory human-computer interfaces. *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 222-227. Amsterdam, The Netherlands.

Burdea, G. 1996. *Force and touch feedback for virtual reality*. New York, NY, USA: John Wiley & Sons.

Buxton, W., R. Hill, and P. Rowley. 1985. Issues and techniques in touch-sensitive tablet input. *SIGGRAPH Computer Graphics* 19, no. 3, pp. 215-224.

Callahan, J., D. Hopkins, M. Weiser, and B. Shneiderman. 1988. An empirical comparison of pie vs. linear menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 95-100. Washington, D.C., USA.

Cao, X., and R. Balakrishnan. 2003. VisionWand: interaction techniques for large displays using a passive wand tracked in 3D. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 173-182. Vancouver, BC, Canada.

Castellucci, S. J., and I. S. MacKenzie. 2008. Graffiti vs. unistrokes: an empirical comparison. *Proceeding of ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 305-308. Florence, Italy.

Cockburn, A., and A. Gin. 2006. Faster cascading menu selections with enlarged activation areas. *Proceedings of Graphics Interface (GI)*, pp. 65-71. Quebec, Canada.

Cockburn, A., P. O. Kristensson, J. Alexander, and S. Zhai. 2007. Hard lessons: effort-inducing interfaces benefit spatial learning. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 1571-1580. San Jose, CA, USA.

Cohen, M. H., J. P. Giangola, and J. Balogh. 2004. *Voice User Interface Design*. Addison-Wesley Professional.

Cohen, P. R., M. Dalrymple, D. B. Moran, F. C. Pereira, and J. W. Sullivan. 1989. Synergistic use of direct manipulation and natural language. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 227-233.

Commarford, P. M., J. R. Lewis, J. A. Smither, and M. D. Gentzler. 2008. A comparison of broad versus deep auditory menu structures. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50, *(February)*, pp. 77-89.

Doel, K. v. d., and D. K. Pai. 2001. A Java audio synthesis system for programmers. *Proceedings of the International Conference on Auditory Display*. Helsinki, Finland.

Edwards, A. D. N. 1989. Soundtrack: An auditory interface for blind users. *Human-Computer Interaction* 4, no. 1, pp. 45-66.

Enns, N. R. N., and I. S. MacKenzie. 1998. Touchpad-based remote control devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 229-230. Los Angeles, CA, USA.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, pp. 381-391.

Fitzmaurice, G., A. Khan, R. Pieke, B. Buxton, and G. Kurtenbach. 2003. Tracking menus. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 71-79. New York, NY, USA.

Foley, J. D., V. L. Wallace, and P. Chan. 1984. The human factors of computer graphics interaction techniques. *IEEE Computer Graphics Applications* 4, no. 11, pp. 13-48.

Forlines, C., D. Vogel, and R. Balakrishnan. 2006. HybridPointing: Fluid switching between absolute and relative pointing with a direct input device. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. pp. 211-220. Montreux, Switzerland.

Friedlander, N., K. Schlueter, and M. Mantei. 1998. Bullseye! when Fitts' law doesn't fit. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 257-264. Los Angeles, CA, USA.

Gaver, W. W. 1989. The sonic finder: An interface that uses auditory icons. The use of non-speech audio at the interface. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. Austin, TX, USA.

Gaver, W. W., and R. B. Smith. 1991. Auditory icons in large-scale collaborative environments. *SIGCHI Bulleten* 23, no. 1: 96.

Gaver, W. W. 1997. Auditory interfaces. *Handbook of Human-Computer Interaction*, pp. 1003-1041.

Gibson, J. J. 1962. Observations on active touch. *Psychological Review* 69, pp. 477-491.

Goldberg, D., and C. Richardson. 1993. Touch-typing with a stylus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 80-87. Amsterdam, The Netherlands.

Grossman, T., P. Dragicevic, and R. Balakrishnan. 2007. Strategies for accelerating on-line learning of hotkeys. *Proceedings of the ACM Conference on Human Factors in Computing Systems*, pp. 1591-1600. San Jose, CA, USA.

Guimbretiere, F., and T. Winograd. 2000. FlowMenu: combining command, text, and data entry. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. pp. 213-216. San Diego, CA, USA.

Harrison, B., K. Fishkin, A. Gujar, C. Mochon, and R. Want. 1998. Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, pp. 17-24. New York, NY, USA.

Hick, W. E. 1952. On the rate of gain of information. *The Quarterly Journal of Experimental Psychology* 4, no. 1: 11.

Hinckley, K. 2001. Input technologies and techniques. *Handbook of Human-Computer Interaction*. Lawrence Erlbaum & Associates.

Hinckley, K., P. Baudisch, G. Ramos, and F. Guimbretiere. 2005. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 451-460. Portland, OR, USA.

Hinckley, K., and M. Sinclair. 1999. Touch sensing input devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 223-230. New York, NY, USA.

Hinckley, K., S. Zhao, R. Sarin, P. Baudisch, E. Cutrell, M. Shilman, and D. Tan. 2007. InkSeine: *in situ* search for active note taking. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 251-260. San Jose, CA, USA.

Hong, J. I., and J. A. Landay. 2000. SATIN: A toolkit for informal ink-based applications. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 63-72.

Hyman, R. 1953. Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology* 45, pp. 188-196.

Jagacinski, R. J., and J. M. Flach. 2002. *Control Theory for Humans CL: Quantitative Approaches to Modeling Performance*. Lawrence Erlbaum Associates.

Kamel, H. M., and J. A. Landay. 2002. Sketching images eyes-free: A grid-based dynamic drawing tool for the blind. *Proceedings of the ACM Conference on Assistive Technologies (ASSET)*, pp. 33-40. Edinburgh, Scotland.

Karlson, A. K., B. B. Bederson, and J. L. Contreras-Vidal. 2006. *Understanding Single-Handed Mobile Device Interaction.* Tech Report HCIL-2006-02.

Kennel, A. R. 1996. Audiograf: A diagram-reader for the blind. *Proceedings of the ACM Conference on Assistive Technologies (ASSET)*, pp. 51-56.

Kiger, J. I. 1984. The depth/breadth tradeoff in the design of menu-driven interfaces. . *International Journal of Man-Machine Studies,* 20, pp. 201-213.

Klatzky, R. L. 1980. *Human memory: structures and processes*. San Francisco: Freeman.

Korhonen, H. 2005. Audio feedback on a mobile phone for number dialing. *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI)*, pp. 345-346. Salzburg, Austria.

Kristensson, P., and S. Zhai. 2004. SHARK2: A large vocabulary shorthand writing system for pen-based computers. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 43-52. Santa Fe, NM, USA.

Kurtenbach, G., and W. Buxton. 1991. Issues in combining marking and direct manipulation techniques. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST),* pp. 137-144.

Kurtenbach, G. 1993. The design and evaluation of marking menus. PhD thesis, University of Toronto.

Kurtenbach, G., and W. S. Buxton. 1993. The limits of expert performance using hierarchic marking menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 482-287. Amsterdam, The Netherlands.

Kurtenbach, G., and W. S. Buxton. 1994. User learning and performance with marking menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 258-264. Boston, MA, USA.

Kurtenbach, G., A. Sellen, and W. Buxton. 1993. An empirical evaluation of some articulatory and cognitive aspects of "marking menus". *Human Computer Interaction* 8, no. 1.

Landauer, T. K., and D. W. Nachbar. 1985. Selection from alphabetic and numeric menu trees using a touch screen: breadth, depth, and width. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 73-78. San Francisco, CA, USA.

Larson, K., and M. Czerwinski. 1998. Web page design: implications of memory, structure and scent for information retrieval. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 25-32. Los Angeles, CA, USA.

Li, K. A., P. Baudisch, and K. Hinckley. 2008. Blindsight: eyes-free access to mobile phones. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 1389-1398. Florence, Italy.

Liao, C., F. Guimbretiere, and C. E. Loeckenhoff. 2006. Pen-top feedback for paper-based interfaces. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 201-210.

Lin, J., M. W. Newman, J. I. Hong, and J. A. Landay. 2000. DENIM: Finding a tighter fit between tools and practice for Web site design. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 510-517. The Hague, The Netherlands.

Luk, J., J. Pasquero, S. Little, K. E. MacLean, V. Levesque, and V. Hayward. 2006. A role for haptics in mobile interaction: Initial design using a handheld tactile display prototype.

Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI). pp. 171-180. Montreal, QC, Canada.

MacKenzie, I. S., and A. Oniszczak. 1997. The tactile touchpad. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*, pp. 309-310. Atlanta, GA, USA.

Mane, A., S. Boyce, D. Karis, and N. Yankelovich. 1996. Designing the user interface for speech recognition applications. *SIGCHI Bulletin,* 28, no. 4, pp. 29-34.

Marics, M.A., and G. Engelbeck. 1997. Designing voice menu applications for telephones. *Handbook of Human-Computer Interaction*, pp. 1085-1102. Amsterdam: Elsevier.

Marslen-Wilson, W. D. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25, no. 1-2 (March), pp. 71-102.

McClelland, J. L., and J. L. Elman. 2002. The TRACE Model of Speech Perception. *Psycholinguistics: Critical Concepts in Psychology* 18, pp. 1-86.

McKnight, A. J., and A. S. McKnight. 1993. The effect of cellular phone use upon driver attention. *Accident; analysis and prevention* 25, no. 3 (June), pp. 259-65.

Miller, D. P. 1981. The depth/breadth tradeoff in hierarchical computer menus. *Human Factors Society*, pp. 296-300.

Miller, G. 1956. The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, pp. 81-97.

Mynatt, E. D., and G. Weber. 1994. Nonvisual presentation of graphical user interfaces: contrasting two approaches. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 166-172. Boston, MA, USA.

Newell, A., and P. S. Rosenbloom. 1981. Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, pp. 1-55.

Nielsen, J. 1992. Finding usability problems through heuristic evaluation. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 373-380.

Norman, K. L. 1991. *The psychology of menu selection: Designing cognitive control at the human/computer interface*. Ablex Publishing Corporation.

Oakley, I., and J. Park. 2007. A motion-based marking menu system. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*, pp. 2597-2602. San Jose, CA, USA.

Oulasvirta, A., S. Tamminen, V. Roto, and J. Kuorelahti. 2005. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 919-928.

Paap, K.R., and N.J. Cooke. 1997. Designing menus. In *Handbook of Human-Computer Intereaction*, pp. 533-572. Amsterdam: Elsevier.

Pascoe, J., N. Ryan, and D. Morse. 2000. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction* 7, no. 3, pp. 417-437.

Peterson, L., and M. Peterson. 1959. Short-term retention of individual verbal items. *Journal of Experimental Psychology* 58, pp. 193-198.

Pirhonen, A., S. A. Brewster, and C. Holguin. 2002. Gestural and audio metaphors as a means of control for mobile devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 291-298. Minneapolis, MN, USA.

Plimmer, B., A. Crossan, S. A. Brewster, and R. Blagojevic. 2008. Multimodal collaborative handwriting training for visually-impaired people. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*,pp. 393-402. Florence, Italy.

Pook, S., E. Lecolinet, G. Vaysseix, and E. Barillot. 2000. Control menus: excecution and control in a single interactor. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts),* pp. 263-264. The Hague, The Netherlands.

Poulton, E. C., and P. R. Freeman. 1966. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin* 66, no. 1, pp. 1-8.

Raman, T. V. 1996. Emacspeak—a speech interface. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp 66-71.

Ranjan, A., R. Balakrishnan, and M. Chignell. 2006. Searching in audio: The utility of transcripts, dichotic presentation, and time-compression. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 721-730. Montreal, QC, Canada.

Rassmus-Gröhn, K., C. Magnusson, and H. E. Eftring. 2007. Iterative design of an audio-haptic drawing application. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*, pp. 2627-2632. San Jose, CA, USA.

Rekimoto, J., T. Ishizawa, C. Schwesig, and H. Oba. 2003. PreSense: interaction techniques for finger sensing input devices. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 203-212. Vancouver, Canada.

Resnick, P., and R. A. Virzi. 1992. Skip and scan: Cleaning up telephone interface. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).*pp. 419-426. Monterey, CA, USA.

Ritter, F.E., and L.J. Schooler. 2001. The learning curve. *International encyclopedia of the social and behavioral science.* Oxford, UK.

Roberts, T. L., and G. Engelbeck. 1989. The effects of device technology on the usability of advanced telephone functions. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 331-337.

Rosenberger, J. L., and M. Gasco. 1983. Comparing location estimators: Trimmed means, medians, and trimean. *Understanding robust and exploratory data analysis*. New York: John Wiley.

Roy, D. K., and C. Schmandt. 1996. NewsComm: a hand-held interface for interactive access to structured audio. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 173-180. Vancouver, BC, Canada.

Salvucci, D. D. 2001. Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies* 55, no. 1, pp. 85-107.

Salvucci, D. D. 2005. A Multitasking General Executive for Compound Continuous Tasks. *Cognitive Science: A Multidisciplinary Journal* 29, no. 3: 457.

Salvucci, D. D., and K. L. Macuga. 2002. Predicting the effects of cellular-phone dialing on driver performance. *Cognitive Systems Research* 3, no. 1 (March), pp. 95-102.

Salvucci, D. D., D. Markley, M. Zuber, and D. P. Brumby. 2007. iPod distraction: effects of portable music-player use on driver performance. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 243-250. San Jose, CA, USA.

Sawhney, N., and C. Schmandt. 1999. Nomadic radio: Scaleable and contextual notification for wearable audio messaging. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 96-103. Pittsburgh, PA, USA.

Sawhney, N., and C. Schmandt. 2000. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions Computer-Human Interactaction* 7, no. 3, pp. 353-383.

Schumacher, R. M., M. L. Hardzinski, and A. L. Schwartz. 1995. Increasing the Usability of Interactive Voice Response Systems: Research and Guidelines for Phone-Based Interfaces. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (June), pp. 251-264.

Schwesig, C., I. Poupyrev, and E. Mori. 2004. Gummi: a bendable computer. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 263-270. Vienna, Austria.

Sears, A., and B. Shneiderman. 1994. Split menus: Effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction* 1, no. 1, pp. 27-51.

Shigeo, H., M. Isshin, and T. Kiyoshi. 2003. Behind Touch: A text input method for mobile phone by the back and tactile sense interface. *Joho Shori Gakkai Shinpojiumu Ronbunshu* 2003, no. 7, pp. 131-138.

Shneiderman, B. 2004. *Design the user interface: strategies for effective human-computer-interaction*. Addison-Wesley.

Stifelman, L. J., B. Arons, C. Schmandt, and E. A. Hulteen. 1993. VoiceNotes: A speech interface for a hand-held voice notetaker. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 179-186. Amsterdam, The Netherlands.

Strayer, D. L., F. A. Drews, and W. A. Johnston. 2003. Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied* 9, no. 1, pp. 23-32.

Strayer, D. L., and W. Johnston. 2001. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science* 12, no. 6 (November), pp. 462-466.

Sugimoto, M., and K. Hiroki. 2006. HybridTouch: an intuitive manipulation technique for PDAs using their front and rear surfaces. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pp. 137-140. Helsinki, Finland.

Suhm, B., B. Freeman, and D. Getty. 2001. Curing the menu blues in touch-tone voice interfaces. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI Extended Abstracts)*. pp. 131-132. Seattle, WA, USA.

Wagner, C.R., S.J. Lederman, and R.D. Howe. 1999. Design and performance of a tactile display using RC servomotors. *Haptics-e Electronic Journal of Haptics Research (www.haptics-e.org)* 3, no. 4.

Walker, M. A., J. Fromer, G. D. Fabbrizio, C. Mestel, and D. Hindle. 1998. What can I say?: evaluating a spoken language interface to email. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 582-589. Los Angeles, CA, USA.

Wickens, C. D. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3, no. 2, pp. 159-177.

Wigdor, D., and R. Balakrishnan. 2003. TiltText: using tilt for text input to mobile phones. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 81-90. Vancouver, BC, Canada.

Witten, I. H., J. G. Cleary, and S. Greenberg. 1984. On frequency-based menu-splitting algorithms. *International Journal of Man-Machine Studies* 21, no. 2, pp. 135-148.

Wobbrock, Jacob O., Brad A. Myers, and John A. Kembel. 2003. EdgeWrite: A stylus-based text entry method designed for high accuracy and stability of motion. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. pp. 61-70. Vancouver, BC, Canada.

Yankelovich, N., G. Levow, and M. Marx. 1995. Designing SpeechActs: Issues in speech user interfaces. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 369-376. Denver, CO, USA.

Yin, M., and S. Zhai. 2006. The benefits of augmenting telephone voice menu navigation with visual browsing and search. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 319-328. Montreal, QC, Canada.

Zhai, S. 1995. Human performance in six degree-of-freedom input control. PhD thesis, University of Toronto, Department of Industrial Engineering.

Zhai, S., and P. Milgram. 1998. Quantifying coordination in multiple DOF movement and ints application to evaluating 6 DOF input devices. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 320-327. New York, NY, USA.

Zhao, S., M. Agrawala, and K. Hinckley. 2006. Zone and polygon menus: Using relative position to increase the breadth of multi-stroke marking menus. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 1077-1086. Montreal, QC, Canada.

Zhao, S., and R. Balakrishnan. 2004. Simple vs. compound mark hierarchical marking menus. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, pp. 33-42. Santa Fe, NM, USA.

Zhao, S., P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch. 2007. earPod: Eyes-free Menu Selection with Touch Input and Reactive Audio Feedback. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI).* pp. 1395-1404.. San Jose, CA, USA.

Appendices 1: Order effects analysis

Within-subject designs tend to have greater statistical power than corresponding between-subjects designs because variability within individuals tends to be lower than variability between individuals (e.g., Box, Hunter, and Hunter 2005). Thus statistically significant results can typically be obtained in within-subject designs with fewer subjects. Between subject comparison require more subjects because each participant experiences only one cell of the design, and many additional participants are required in a between subjects study if there are a larger number of cells in the design matrix (i.e., combinations of factors). In addition, creating matched samples of subjects with similar skills and backgrounds can be a difficult and subjective task. However, crossover effects can happen between conditions in within-subjects design and order effects need to be examined to ensure that they do not provide a competing explanation for the results obtained.

If learning effects are similar no matter which ordering of conditions is used, the problem of learning can generally be solved by counterbalancing the orders of presenting conditions to each subject so that there is no bias where some conditions benefit more from learning than others. However, counterbalanced designs cannot solve problems where learning effects are asymmetrical so that some conditions benefit more than others from being preceded by other conditions (Poulton and Freeman, 1966). For instance, if performance in condition B results in a learning benefit when it is done after condition A, but condition A's performance receives no similar benefit when done after Condition B, then asymmetric transfer of learning is said to have occurred. This type of asymmetric transferring of learning can undermine the validity of results obtained, since a condition may appear to be better than another when a within subjects is used, but that advantage might evaporate if the conditions were subsequently compared in a between subjects design. There is no certain way to avoid possible asymmetric transfer effects except to use a between subjects design, but as noted above between subjects designs have their only problems and they would be wasteful and inefficient if used in situations where asymmetric transfer is not a problem.

In realistic experimental contexts, and particularly in cases such as this dissertation where novel research is being carried out, it is generally not possibly to know in advance if asymmetric transfer will occur or not. In such cases a prudent strategy may be to conduct the experiment and then carry out diagnostic analyses on the data to see if there is evidence of asymmetric transfer. In order to test for asymmetric transfer, the data can be censored so that only the initial conditions are considered in the analysis, essentially converting the experiment into a between subjects design. In this case the problem of asymmetric transfer is solved but at the cost of throwing away data, which may result in further data needing to be collected in order to obtain sufficient power for the statistical analyses.

In the remained of this appendix, the diagnostic method used to test for the presence of asymmetric transfer in the experiments carried out in this dissertation is presented. Only possible asymmetric transfer effects are considered here, since counter-balanced designs were used for data collection throughout the dissertation research in order to ensure that no symmetrical learning biases occurred.

The following strategy was used to protect against order and transfer effects.

1. An efficient (incomplete) balanced design was used to implement a representative and balanced set of orders.

2. In the case of experiment 1, where only one experimental factor was used, the possible order effect was testing using the interaction between the order factor and the experimental factor.

3. For Experiments 2 and 4, the data were re-analyzed using only the portion of the data that represented a between-subjects design for the modality factor. Thus only the data of the first condition was kept in the analysis (with the data for the remaining two modalities per participant being discarded for this analysis). Since asymmetric transfer could not have affected the resulting between subjects analysis, comparison of the repeated measures and between subjects results could then be used to determine if when an asymmetric transfer occurred.

Order Analysis of Experiment 1

Experiment 1 involved only two conditions. Possible influence of order effects was investigated using a mixed analysis (using order as a between subject factor) on both accuracy and response time. Results revealed no significant main effect for order, and there were no significant interactions between order and the experimental condition. These results rule out the possibility that the results obtained in experiment 1 were tainted either by a basic order effect, or by asymmetric transfer.

Order Analysis of Experiment 2

Experiment 2 involved two factors: modality of feedback and menu style, with 3 and 2 levels, respectively. The order of the main effects was balanced, where modality of feedback had 6 different variations, and menu style had 2. The interaction between these two dimensions was a 3x2 design involving a total of six conditions. A complete design would have involved all possible orders of the six conditions, requiring 720 (6!) subjects. Thus a full balanced design for the interaction effects was not used.

Between-subjects analysis of variance was performed on the interaction between modality and menu type to assess what would happen if asymmetric transfer was entirely removed from the data. There was a significant main effect of menu style ($F_{1,13} = 25.60$, $p < .01$), where linear menu style (1.97 s) was significantly slower than radial menu style (1.51 s). There was also a significant main effect on modality ($F_{2,13}=5.24$, $p < .05$). Audio modality (2.05 s) was significantly slower than both dual modality (1.68 s) and visual modality (1.5 s), while visual and dual are not significantly different from each other ($p > .05$). These results were consistent with the within-subject analysis. In contrast to the earlier repeated measures analysis, no significant interaction effect was found for modality x menu style. Since this negative result was not likely due to lack of statistical power it seems likely that the significant interaction found between menu type and modality in the repeated measures analysis may actually have been due to the effects of asymmetric transfer.

Order Analysis of Experiment 4

Since the longitudinal study conducted in Experiment 3 used a between subjects design, there were no possible order effects and thus no order analysis was necessary. Thus the final order analysis carried out was for Experiment 4. The design of Experiment 4 used different groups of participants for the desktop condition and the driving condition respectively. Aside from the addition of the driving condition, the design of Experiment 4 was identical to that of, the Experiment 2, so the same strategy used to analyze Experiment 2 was also used for the data from

Experiment 4. For the order analysis of Experiment 4, the results for the desktop condition and the driving conditions are reported separately below.

For the between-subjects analysis of the data, the following results were obtained. There was a significant main effect of modality ($F_{2,9} = 28.26$, $p < .01$), where performance in the audio modality (4.18 s) was significantly slower on average than both dual (2.12 s) and visual (2.35 s) modality performance. There was also a significant main effect of menu style ($F_{1,9} = 13.07$, $p < .01$). Linear menu style (3.28 s) was significantly slower on average than radial menu style (2.49 s). A significant interaction effect for modality x menu style was also found ($F_{2,9} = 7.72$, $p<.05$). Figure 1 shows the effect of the interaction which is consistent with the pattern of results obtained in the repeated measures analysis of the full set of data.

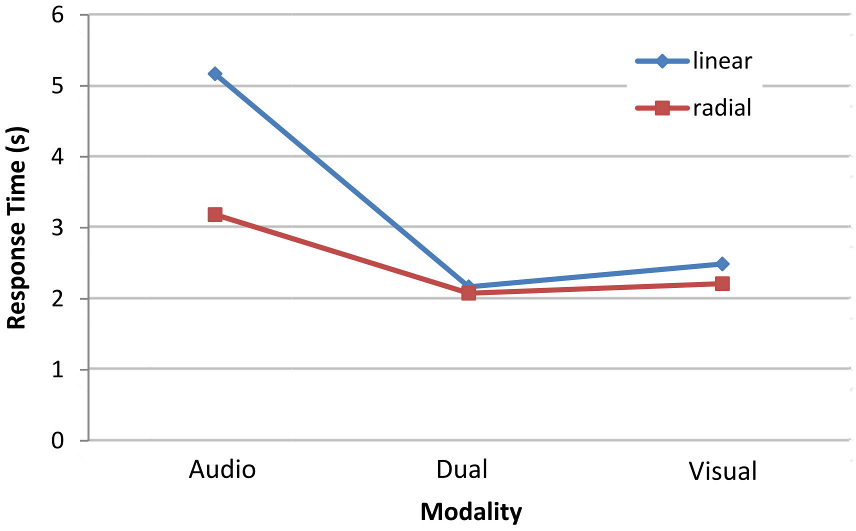


Figure A1.1: Interaction effect between modality and menu style in the between-subject analysis for Experiment 4.

Between-subject analysis was also performed on the driving condition data. The results were generally consistent with the within-subject analysis. No significant effects are found on accuracy. For response times, there was a significant effect of menu style ($F_{1,9} = 11.55$, $p < .01$), with the linear menu style (8.69 s) being significantly slower on average than the radial menu style (7.01 s). Neither the effect of modality nor the interaction between menu type and modality were significant, consistent with the results of the repeated measures analysis. Thus there was no evidence that asymmetric transfer affected the results of Experiment 4.

Appendices 2: Threats to validity

All experimental designs represent a choice concerning tradeoffs that exist within the potential experimental design space. In this appendix the threats to validity of experimental results from this dissertation are noted. This appendix also attempts to justify, to the extent possible, the experimental design decisions that were made by highlighting the tradeoffs that governed and constrained the decisions that were made.

For the four experiments carried out in this dissertation work, a number of limitations are common across experiments, while others are localized to particular experiments.

Limitations Affecting Multiple Experiments

The participants selected for all of the experimental studies were young and tech-savvy university students. As a result, the obtained experiment results may not be generalizable to older demographics or none tech-savvy populations. In addition, due to the limited number of subject used in each experiment, it is possible that the sample participants may not be representative even in the chosen demographic groups. This problem is particularly acute in the case of the longitudinal study (Experiment 3) where 8 participants were used. Although sample sizes were relatively small, the relative homogeneity of participants used also reduced error variation due to individual differences, leading to statistically significant, and interpretable, effects being identified. Collecting larger, or more diverse subject groups would have required considerably more effort and may have shifted the focus of this thesis more in the direction of confirmatory evaluation than design exploration. In addition, obtaining larger samples of university students would not have addressed the issues of generalizing to broader populations.

Experiment 1, 2, and 4 ran fewer trials as compared to Experiment 3, so results obtained from these three experiments are more likely to reflect novice performance with each technique. It is possible that the relative advantages of each technique may change as people gain more experience with them. However, comparing the results of Experiment 1, 2, and 4 with those of Experiment 3, there is a general consistency in the results that makes it less likely that they will change dramatically when each menu method is used more extensively.

When comparing the results of Experiment 3 and others, it is found that the average accuracy for the longitudinal experiment is noticeably higher than the rest. Since the subjects are randomly selected from university community, it is unlikely that the sample population of Experiment 3 is different from the other experiments. A more plausible explanation is that Experiment 3 had an additional reward of 100 dollars that people could earn if they performed well and were fortunate in the ensuing raffle.

Thus participants were likely more motivated to maintain high accuracy in Experiment 3. Thus it is possible that differences in results between the experiments may have been affected by differences in motivation. While this difference in motivation may be justifiable due to the need to provide participants with better motivation in a longer study differences in motivation stand as an additional threat to the validity of the results. It could be argued that motivation of participants is a problem in many human-computer interaction studies and that future research should address this issue more seriously.

In all the experiments, only one word menu items were used. Menu with average length of more than one word per item may exhibit different characteristics, especially for menus with long items. Under these conditions, the advantage of the visual modality may be greater because of the greater load on auditory memory and the longer listening time that would likely result in the auditory conditions.

The focus in the evaluation carried out in this dissertation was on performance with *earPod* and other menu selection methods. A choice was made to focus on design for the general population and there were no analyses of individual differences. However individual differences (such as audio/visual learner, left/right hemisphere thinker, and holistic/analytic users) are likely to influence and mediate menu selection performance and thus deserve to be considered in future research.

Limitations for Specific experiments

Limitations for Experiment 1

One design tradeoff made in first experiment is that a subset of stimuli (16/64 choices) was chosen instead of all the 64 possible stimuli for a 2-level 8x8 menu. There is a possible risk that using only 25% of the possible menu items may make the results difficult to generalize to a full 2-level menu. However, this risk is somewhat mitigated by the results from Experiment 3 which showed similar characteristics to those of Experiment 1 even when all 64 items in an 8 x 8 menu were used.

Limitations for Experiment 2

Experiment 2 only used a 1 level menu, thus possible difference in performance and learning for deeper menu is not accessed. With 13 blocks of data, it is possible that the number of blocks was not large enough to access the full learning pattern. However, this deficiency was addressed in Experiment 3 where a detailed analysis of learning within a longitudinal study was carried out.

Limitations for Experiment 3

Because only male subjects were used, the results may not generalize to females. While the error variance may have been reduced by selecting a more homogeneous set of participants, the number of subjects was small and the experimental results may have been affected by individual differences. The experiment was run for the course of a week, but learning would likely have continued after that period and the collected data was likely part of a longer learning curve. This experiment used a 2-level menu. However, real world menus can be more complex and further research is needed to see if the present results generalize to more complex menu hierarchies.

Limitations for Experiment 4

Due to the limited trials used in the driving simulator, long term learning was not assessed in the driving condition. Despite the efforts made to make the simulated driving realistic, real world driving is necessarily more complex. Thus the results obtained from simulated driving may differ from the results that would have been obtained with real world driving. The menu used was a one-level menu, so research with larger and more complex menu may deserve exploration in future, although safety concerns about the effect on the primary driving task will tend to increase as the complexity of the menu selection task increases.

Appendices 3: Ethics documents

Ethics application form

1. Background, Purpose, Objectives

Input techniques are essential to interactive computing. The diversifying of devices has raised challenges to the mouse and keyboard input model. In particular, due to the limited space for input and output, mobile devices require new input interaction methodologies.

Previous research demonstrated pen based input techniques and audio cues, which require very limited screen space, are particularly suitable for mobile devices. Establishing interaction paradigms for such devices using a mixture of gestures, speech, and physical button for inputs, and visual and spatial audio cues for feedbacks, is an active research topic. Past research has show that gestural menus are a promising alternative to status-quo linear menus.

The purpose of this study is to examine the performance of different types of gestural menu. The ultimate goal of this research is to provide design recommendations for mobile device interfaces.

2. Research Methodology

Participants will perform a set of menu commands using a pen on a tablet. Upon conclusion of the experiment, participants will answer a series of questions concerning the interface used in the experiment (please see attached questionnaire).

A multi-factor, between-subject design will be used for the experiment, with one independent variable being the type of gestural interaction used. The second independent variable will be Chinese writing knowledge (novice vs. expert). This factor is included to assess whether or not experience with the greater variety of gestures used in Chinese writing will affect performance on gesture-based menus.

Performance and preference data will be logged using software and they will be analyzed using analysis of variance (ANOVA). A power analysis will be used to determine the required sample size for this experiment. We anticipate that the necessary sample size will be between 12 and 20 participants.

3. Participants

Participants in this experiment will be University of Toronto students and their friends. We believe them to be representative of users of mobile devices. Participants will be 35 years old or younger and will have some experience in using a mobile device.

1. Recruitment

Participants will be recruited through email solicitation. A copy of the email message to be used for recruitment is attached.

2. Risks and benefits

There is minimal risk undertaken by participants in this experiment. Only summaries of the data will be presented in the final project report and no such summary will contain data that in any way identifies individual participants. Participants will be informed (on the attached consent form) that they may leave the study at any time for any reason and that they may request that their data be destroyed.

Subjects will be paid $10 per hour for their participation in the study. Each participant will spend no more than two hours carrying out the experiment.

3. Privacy and confidentiality

Before participants agree to participate in the experiment, they will be given a complete description of the study and will be told how data the collected from them will be treated. Participant will sign a consent form (attached) agreeing in writing to participate in the experiment. Participants may at any time request that their data be removed from consideration (and destroyed) as indicated in the consent form that they sign.

Participants will be identified by an assigned code and all data collected will be kept confidential. Information such as participant’s name, email addresses and telephone numbers will be kept separately from the participant’s questionnaire and performance data. This personal information will not be used in any internal or external reports without the participant’s explicit consent. No personal or identifying information will remain in the coded data. Participants will be provided with information about their own results upon request.

Only summaries of performance data will be presented in the report and no such summary will contain data that in any way identifies individual participants.

4. Compensation

Participants in the experiment will be paid $10 per hour for their participation in the experiment.

5. Conflicts of interest

There are no conflicts of interest. None of the participants have invested any resources in the development of the questionnaires being evaluated.

6. Informed Consent Process

Please see attached consent forms.

7. Scholarly review

The research poses minimal risk to participants.

8. Additional ethics reviews

Not Applicable

9. Contracts

Not Applicable

10. Clinical Trials

Not Applicable

Consent form for participation in the experiment "Interaction techniques for Mobile Devices"

Investigators: Shengdong Zhao and Professor Mark Chignell
Department of Computer Science, University of Toronto
10 King's College Road, Toronto, ON M5S 3G4, 416.978.6025

Introductory Information: I have been invited to participate in the experiment "Interaction techniques for Mobile Devices". The purpose of this study is to examine the performance of different types of gestural menu. The ultimate goal of this research is to provide design recommendations for mobile device interfaces.

What Will I be Asked to Do: I understand that, as a participant in this experiment, I will be required to perform a set of menu commands using a pen on a tablet. Upon conclusion of the experiment, I will answer some questions concerning the interface used in the experiment (please see attached questionnaire).

Risks and Benefits: I understand there is minimal risk undertaken by participants in this experiment. I also understand that my participation in this experiment will provide me with minimal direct benefit other than the opportunity to experience interactive techniques for mobile devices, and a small payment.

Compensation: I understand that I will be paid $10 per hour for participating and that there will be no academic consideration for participation in the experiment. To receive compensation, I will be asked to meet with the experimenter at a specified time and place. During this face-to-face meeting, the experimenter will explain the details of the study to me and I will have the opportunity to ask any questions I have regarding the experiment (formal debriefing). The experiment should take about 2 hours or less to complete.

Access to Information: The research team, Shengdong Zhao and Professor Mark Chignell, will have access to the data in its raw and coded forms. Questionnaires and experimental data will be kept for the period of approximately 1 year. All retained information will be coded.

Confidentiality and Publication of Results: I understand that the results of my participation in this experiment will be kept confidential and that there will be no possibility that I can be identified from or associated with any results or summaries of results, as presented or published subsequent to this experiment.

Contact Information: Please contact Shengdong Zhao for further information, phone: (416) 978-7581, email: sszhao@dgp.toronto.edu.

I have read the information provided to me on this experiment and I hereby consent to participate in the experiment “Interaction techniques for Mobile Devices”. The objectives, methods, tasks and procedures have been thoroughly explained to me and all of my questions and concerns of the experiment have been answered completely to my satisfaction. I have the right to withdraw from this experiment at any point in the experiment without penalty, and to request that my data be destroyed.

Participant’s name:____

Participant’s signature:______________________________________

Participant’s preferred email:__________________________________

Date:____