Spatial Auditory Maps for Blind Travellers

by

Martin Talbot

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2011

Abstract

Empirical research shows that blind persons who have the ability and opportunity to access geographic map information tactually, benefit in their mobility. Unfortunately, tangible maps are not found in large numbers. Economics is the leading explanation: tangible maps are expensive to build, duplicate and distribute.

SAM, short for Spatial Auditory Map, is a prototype created to address the unavailability of tangible maps. SAM presents geographic information to a blind person encoded in sound. A blind person receives maps electronically and accesses them using a small inexpensive digitalizing tablet connected to a PC. The interface provides location-dependent sound as a stylus is manipulated by the user, plus a schematic visual representation for users with residual vision.

The assessment of SAM on a group of blind participants suggests that blind users can learn unknown environments as complex as the ones represented by tactile maps - in the same amount of reading time. This research opens new avenues in visualization techniques, promotes alternative communication methods, and proposes a human-computer interaction framework for conveying map information to a blind person.

Acknowledgements

I am thankful to my supervisor, Bill Cowan, whose encouragements, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

I offer my regards and blessings to (in alphabetic order) Robin Cohen, Richard Ladner, Edward Lank, Deborah Lashbrook, Richard Mann, Ken Salem, Graham Strong, Mike Terry and Margaret Towell. I am thankful to my parents Monique et Louis and colleagues & friends Mark Belcarz, Laurent Charlin, Claude-Guy Quimper, Tyrel Russel, the CGL and HCI lab members, the anonymous participants, and all of those who supported me in any respect during the completion of the project.

Lastly, I would like to thank my life-partner Caroline Charest for her constant support and for endlessly listening to me on the subject of my research. Without her, this would have never been possible.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Blind persons have difficulty keeping up with a sighted population that is increasingly vision-centered. Emiliani *et al.* [1] explain that 'one of the major obstacles to the socio-economic and cultural integration of blind people into society is the problem of making graphic information available to them in the various application environments they face in everyday life'.

Not long ago, in the era of the command line, blind persons were as effective as sighted persons on computers because both groups interacted with the keyboard and because lines of characters were straightforward to read with a screen reader [2]. But the advent of the Graphical User Interface (GUI) in computers and the Internet with the services it offers, which adhere to the WIMP paradigm[1], create challenges for a blind person who is compelled to access visual information by alternative means.

Internet services like Google Maps™ have changed how sighted people think and address their navigation problems. They offer street maps, urban business locators, and route planners for travel by foot, car, or public transport, in countries around the world. The blind population would like to take advantage of these oportunities, but the information conveyed in these maps is difficult to translate into non-visual forms of communication.

Blind people who access map information tactually benefit in their mobility. A body of evidence suggests that congenitally blind (CB) and adventitiously blind (AB) children and adults learn an environment more accurately from a tangible map than from direct exploration [3, 4]. A *tangible map* (TM) is a 3D relief map, which may be examined by touch and is useful to a blind reader. Good TMs eliminate distortion and confusion of environmental concepts for blind travellers [5][2]. Bentzen was the first to demonstrate that blind persons can use TMs to plan routes and follow them [5]. Ungar points out that

[1] WIMP is the acronym for Window, Icon, Menu, Pointing device.

[2] Compared to having the traveller navigate through the space in order to learn a new route.

TMs provide blind persons with an impression of the geography of an environment and therefore the construction of accurate, well integrated cognitive maps [3]. A *cognitive map* is an internal representation of a set of geographic locations. After acquiring a cognitive map, a person has a good understanding of the environment because he[3] can visualize images of the surroundings, which reduces his cognitive load and enhances recall.

Alas, direct translation of a visual map to a tactile one is, in most cases, not sufficient to provide usable information to a blind reader. Touch has lower resolution and a smaller periphery than vision and touch lacks capabilities like zooming that are important for placing local details within a spatial context [6]. Hence, the layout of the information must be adapted, usually by eliminating information, increasing the scale and enlarging symbols to make them discriminable by touch.

To help map designers, heuristics and practical solutions for constructing TMs have been published. Much of this literature, usually in handbook form, is based on experience and intuition, and is unsupported by empirical evidence [7, 8, 9, 10, 11].

There does exist evidence-based research in the perception and cognition of tactile information that demonstrates that the design of TMs is complex, requiring specialized knowledge in tactile perception and cognition [12, 13, 14, 15]. For this reason, TMs are constructed by *Orientation and Mobility Specialists* (OMSs), expert professionals who train persons with low vision to move about safely in their homes and to travel by themselves.

Unfortunately, not every OMS has the skills of a cartographer, and the construction quality of TMs varies significantly. To illustrate this, Figure 1.1 shows a TM of Conestoga Mall™, as currently used by OMSs to explain the layout of the mall in the region of Waterloo. There have been national and international efforts to improve the quality of TMs via standardization, such as the Nottingham Kit [9]. Regrettably, they have not been widely accepted. OMSs who tried adopting standards reported finding it very difficult [16]. A recent international survey of approximately one hundred and fifty OMSs indicated that 'the overriding need is just to get the map done [which] is the main driving force behind the design' [16, p.107].

The time that an OMS spends designing and producing maps is time unavailable for other client needs. To address this problem, mass production has been attempted [18]. Micro-capsule paper, thermoform and crafted models are the most popular methods [19]. These have different strengths and weaknesses, but all have significant production overhead, which must be amortized over many copies. However, since each blind traveller visits a different set of destinations, economies of scale are rarely available [19]. Therefore, most TMs end up being individually hand-crafted to accommodate a blind traveller's unique navigation needs. They are often built from scratch, taking into account the specific

[3] The male gender is used to refer to both male and female persons for ease of reference throughout the document.

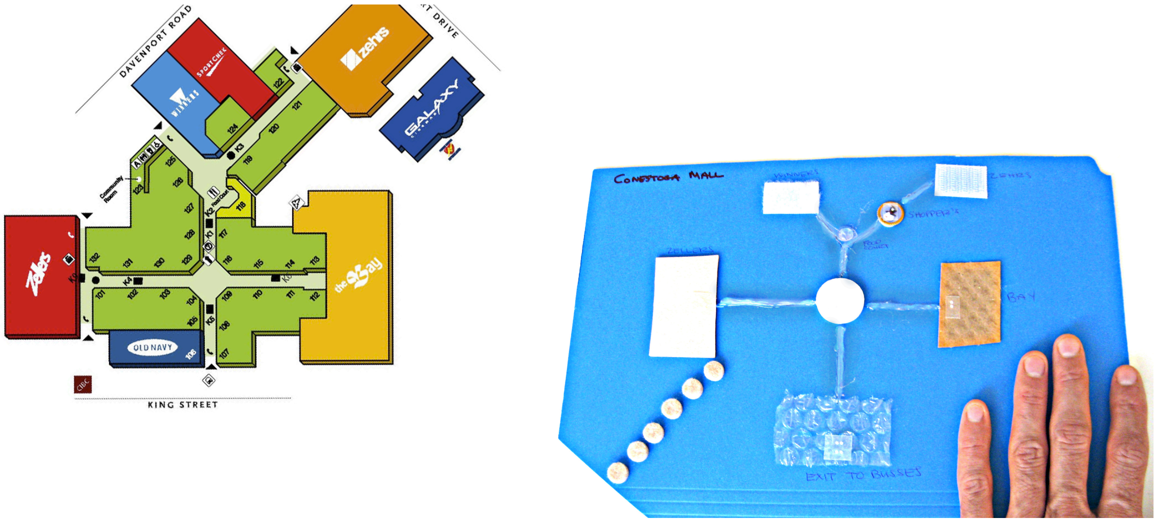

Figure 1.1: A map of the Conestoga Mall in Waterloo Ontario is shown on the left and a TM of the same mall on the right (the alignment of dots represents an outside path to access the mall). The corridors on the TM are outlined with a glue gun and the stores with buttons, velcro, bubble wrap and carton pieces of different textures. One Braille letter per store identifies them. Labelling is a common problem with TMs because Braille labels are inflexible - one size only, no colour or font types. In fact, when enough labels are added to a map to ease understanding, the map often becomes cluttered and illegible [17]. Only the stores at the extremities of the corridors and the food court are shown on the TM, so a blind reader needs assistance to find a boutique. We note the diversity of material used here for the construction of this TM. The same material may not be available to another cartographer, which forces a blind reader to relearn the mapping of symbols and textures every time a new map is used.

cognitive and perceptual abilities of a client and the low resolution of TMs. Then Braille text is added, as shown in Figure 1.1, even though less than 10% of the blind population in Canada and US reads Braille [20]. The final result is in physical form and is delivered to the blind client. The cartographer normally familiarizes the client with the map and the depicted environment.

Unfortunately, TMs are rare. Economics is the explanation: TMs are expensive to build, duplicate and distribute [19]. The Canadian National Institute for the Blind (CNIB) has recently closed its tactile department owing to budget cutbacks, which leaves blind Canadians with very few resources to learn unknown environments and limits them to asking sighted persons for assistance. The ability to travel with confidence is important for helping a blind person to develop independence and self-esteem [21]. Jacobson [22] points out that the ability to make informed spatial decisions has a direct effect on a human's quality of life.

Third world respondents in a recent international survey of TM cartographers from around the world expressed a need in their countries for a computerized solution, but added that they would depend on the developed world to provide it [19]. Of course, the solution needs to be inexpensive [23], something Edwards [24] realized twenty-five years ago when he pointed out the small potential market of blind computer users. Regrettably his message did not reach the community. Thus, to respond to the needs of a large population of blind persons around the world, a computerized solution that stores maps and provides spatial information to a blind user should be low in price to acquire and to operate.

Without a doubt, computer science has enhanced education and employment opportunities, along with overall quality of life, for the blind population. Recent advances in Human-Computer Interaction (HCI) have had much to offer. For example, auditory displays that speak the text in a printed document make digitized reading material available to blind users [25]; multi-modal displays, auditory and visual, allow visually impaired users who have residual vision to comprehend and interact with the complex geometry of web pages [26]; and window interfaces allow them to access metadata encoded in screen layout–like image descriptions, so as to direct their interaction to specific information with which they wish to interact [27].

In this thesis, displaying the physical characteristics of TMs, which is essential for their usability, is identified as central to the central challenge that needs to be addressed. The investigation of alternative forms of representation and interaction supporting effective non-visual access and manipulation of environmental information could lead to new methods for communicating spatial information to blind users (or sighted users who cannot rely on their vision for the task), which would build knowledge and contribute to the field of HCI. This open problem was the focus of my doctoral research.

1.1 Doctoral Thesis

This thesis hypothesizes that a state-of-the-art sound interface implemented on an off-the-shelf computer system can assist blind persons to build cognitive maps of unknown environments. The feasibility of this interface is demonstrated through a prototype referred to as *Spatial Auditory Map* (SAM), and is validated by an empirical study.

My PhD research builds a theory and practice of HCI for the elaboration of a computerized map system for blind persons, designed by understanding the cognitive and perceptual abilities of blind users faced with the task of becoming familiar with unknown environments. This information was used to determine a set of design rules, and to discover how to build an effective computerized solution, which is presented in this document.

A functional prototype is explained in terms of data representation and interconnectivity, how user outputs and inputs meet and interact. SAM is a framework for the design of an auditory interface that translates environmental information into sound messages a blind user can understand. The efficacy of the prototype is compared to TMs in an empirical assessment detailed in Section 7.

TMs are the state-of-the-art technology available to blind travellers for learning unknown environments. The assessment consists in measuring the cognitive maps of seven blind participants interacting with the prototype and a TM. The assessment of the cognitive maps built using these two technologies is measured by numerical methods, by post-interviews with the participants, and by two OMSs. The prototype and the assessment results serve to corroborate my thesis.

1.2 Prototype Overview

SAM is the prototype I created to address the unavailability of TMs. It presents geographic information to a blind person encoded in sound. The usual encoding method is *sensory substitution*: what the blind traveller is missing is vision, and audition is substituted for vision by encoding visual information using sound, in effect providing 'synthetic vision'. Doing so is problematic because the two senses differ greatly: vision is spatial and parallel, audition temporal and sequential; vision is directed and audition is not because a perceiver cannot focus the ears on details as with the eyes; visual information is persistent and auditory information is transitory; information bandwidth is much greater in vision than in audition; and a CB person has no memory of visual experience to assist interpretation. These different characteristics of sensory modalities need to be considered in developing effective non-visual interfaces for blind persons.

SAM addresses these problems by a 'synthetic touch' technique: audition extends touch, the primary source of spatial information in blind persons. Synthetic touch has important interface advantages. Both modalities are sequential, and the bandwidth of audition is two orders of magnitude superior to that of touch [28], which makes redundant coding possible. Blind persons have experience with both modalities, touch for nearby spatial information, audition, via echo location, for more distant information. Touch outperforms audition for fine detail, while touch is limited by body size, complementary roles to which blind persons are accustomed.

In a nutshell, SAM is made up of two distinct tools: a drawing application used by OMSs to create maps, and a portable 'reading' application, used by blind persons to access them. The OMS uses a stylus and a digital tablet to draw and label routes and points of interest[4]. Labels, spoken to the blind user, can be arbitrarily elaborate. Routes can have

[4] Points of interest are fiducial points, the locations of which are well known by the user.

any geometry. Maps can be previewed and edited at will. A blind person receives maps electronically and accesses them using a small inexpensive tablet for input. The reader interface provides location-dependent sound as a stylus is manipulated, plus a schematic visual representation for users with residual vision. The current prototype runs on an Intel Core Duo 1.86 GHz with three Gbytes of RAM and a stock sound card. Maps have a low memory size, requiring a few seconds to download over a slow connection. Figure 1.2 depicts a user interacting with SAM in order to access a stored map.

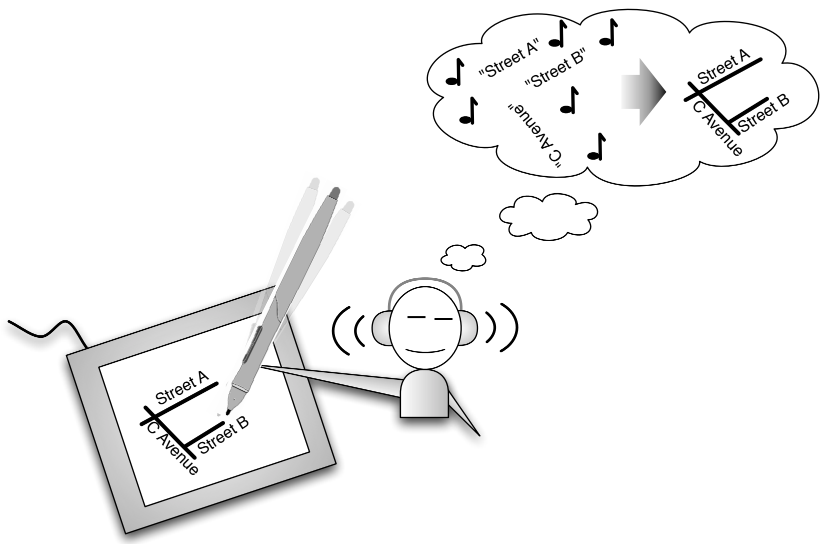

Figure 1.2: This figure shows a user interacting with SAM. The user wears a pair of headphones and listens to sound information describing the geometry and names of landmarks and routes, which he can trace by sliding a stylus over a flat digitizing tablet connected to a standard PC or a tablet-PC. The spatial relationships of streets and landmarks found in a map are also maintained by the system. *Inside the thought-bubble*: the user hears non-speech audio and synthetic speech (left), and constructs a corresponding mental image (right).

1.3 The User

'A person is diagnosed blind when the smallest detail that can be resolved visually in the better eye with refractive errors corrected is ten minutes of arc or greater, or the horizontal extent of the visual field with both eyes open is less than or equal to twenty degrees' [29].

Vision loss has many causes, including eye disease, damage caused by stroke, diabetes, premature birth, and trauma to the eye. The World Health Organization [30] estimated in

2003 that 40 to 45 million people are blind and that 135 million have low vision[5]. Diabetes is the leading cause of new cases of vision loss in people aged 20 to 74 [31]. Epidemiologists predict a rise of 27% in diabetes cases in the United States within the next five years [32]. More than 4 million Canadians suffer from age related, blinding ocular disease [33]. With the aging demographics that number will double. Totally blind persons are rather rare and most blind persons have some residual vision. Diseases affect vision in different ways including loss of central vision, blurred vision, tunnel vision, or depleted light perception, many of which spare perception that helps a blind person in orientation tasks [34].

Hearing is a blind person's primary modality for perceiving the environment beyond reach. A skilful blind traveller can approach an intersection and, by listening to the traffic, perceive the number of lanes in each direction, the traffic density, the vehicle types, the width of lanes, the presence of pedestrian medians, and the type of traffic signalization [35, 36]. As expected, not every blind traveller demonstrates these skills [37], but enough do to support the theory that blind people have supernormal auditory capabilities. Paying greater attention plays an important role in auditory perception, and a blind person has more experience at relying on auditory information to interpret the environment than a sighted person [38].

Spatial orientation is a major mobility problem encountered by all individuals with profound vision loss [39, 40], but the blind population is highly heterogeneous, especially in terms of spatial skills. For mobility and orientation tasks, the ability of blind individuals depends on factors like the age of onset, the number of years of experience without sight, history of movement, suitable training, IQ[6], and the mobility tool of choice (long cane or guide dog).

There is a considerable literature comparing the conceptual and spatial abilities of the CB, the AB and the sighted. Many authors have repeatedly reported the deficit in spatial abilities of the CB [43]. In contrast, the AB generally have similar abilities to blindfolded sighted persons, though factors like age of onset and the number of years of blindness have an effect on spatial performance [44, 45, 46, 47, 48, 49, 50]. Although the inherent factors behind the deficit remain unknown, psychologists point out that a CB person never experiences visual concepts such as perspective, horizon, occlusion, shading, and parallax from egomotion. Jones [51] observed that history of movement deprivation[7] often results in a limited understanding of space. Conversely, training[8] can help a blind person to gain a better understanding of his body-image and to learn mobility and orientation techniques [52, 53].

[5] Low vision is when corrective eyeglasses are unable to bring a person's sight up to normal sharpness.

[6] IQ score is correlated with performance in mobility tasks [41]: to navigate independently, a blind pedestrian must memorize the layout of a given area, learn path segments and angles between them, and has to recognize these characteristics when travelling [42].

[7] E.g., over-protected children.

[8] E.g., schools for the blind.

Braille was once the most popular method of communication for the blind, but has lost much popularity with the advent of new technologies. Less than ten percent of the visually impaired population in Canada and USA reads Braille, and many wonder if the Braille is dying [20]. 'Today in America only 10 percent of blind children are learning to read Braille in school. This continues despite the fact that studies have shown that 80 percent of all employed blind people read and write Braille fluently' [54]. John Rafferty [55], the head of the Canadian National Institute for the Blind (CNIB), says even twenty years ago, adults who lost their sight would learn Braille to read and write, but today, most prefer to make use of screen reading software. Dr. Lee Hamilton [56], President and CEO of Freedom Scientific, a software company developing tools for blind persons, argues that technology should be supporting Braille literacy, not replacing it; he notes that the combination of a screen reader and a Braille display makes learning and using Braille easy, but Braille displays have always been expensive, an entry price of $1700 USD, which prevents many blind persons from acquiring them.

The technologies expressly designed to help blind persons tend to be expensive to purchase, which conflicts with the economic reality of the blind population, since many blind individuals have low incomes. Fortunately, blind Canadians receive financial support from governmental programs to update their home computers at regular intervals. In Ontario, for example, a blind person has his PC renewed by the Assistive Devices Program every five years. Thanks to these provincial programs, the large majority of blind Canadians own decent home computers, which can be exploited by SAM.

1.4 Sensory Substitution

Like James Gibson [57], Suzanna Millar [34], a psychologist expert on blindness, argues that human perceptual experience is *amodal*, that is, different senses yield different sensations, but the same information about the world. The contributions of different modalities overlap sufficiently to be felt as the same rather than different.

Vision, audition and touch have much more in common than was initially thought. Research in neurology suggests that there is a functional junction of modalities. Doubt has been cast on the traditional view that separate brain regions respond to vision, audition, touch, and olfaction. Representations seem to be shared across modality regions [58].

In fact, recent research suggests considerable modality overlap in the brain, which supports the theory of amodal perception proposed by Gibson and Millar, and provides the scientific basis for the phenomenon of sensory substitution upon which my research rests. For instance, Sathian & Prather [59] report visual cortex activation during tactile orientation tasks. Similarly, James [60] reports that haptic[9] exploration of solid objects

[9] Relating to the sense of touch [61, p.416].

produces activation in both the somatosensory regions and the visual cortex, more precisely the areas normally involved in visual processing of forms. This observation is consistent with evidence presented by Pascual-Leone *et al.* [62], who found visual cortex activation while reading Braille. Similarly, James *et al.* [60] hypothesize that brain regions that have been considered visual may also be tactile. Similar evidence against the separateness of brain regions is also reported for audition and vision [63].

1.5 Contributions of this Thesis

The designing of SAM has allowed me to identify several problems that must be studied to build an effective system. This section describes the set of design principles that I established for the integration of sound into a map interface for blind persons.

I identified curved lines and points of interests as modality independent information on the map and discarded the rest, which reduced the amount of information relayed to the auditory sense. As a result, I developed a data visualization framework for transforming visual maps into SAM maps.

I designed a synthetic touch encoding technique based on sensory substitution, replacing touch feedback by non-speech sound feedback. The problem of encoding synthetic touch feedback is examined here in two ways, namely, how to implement a virtual sensor and how to present that data to the user. The solution first requires a non-symmetrical sensor of optimal resolution. Second, a presentation protocol encodes rules for displaying spatial characteristics in a form that respects the culture of blind users.

Two empirical studies conducted on blind persons provided the basis of the curve tracing and 'locatone' cues, both of which direct a user in 2D space with non-speech sounds. Inherent to SAM is the concept of tonal interfacing, which relies on musical tonality to enhance, by way of non-speech sound, the user's accuracy when tracing curves (*e.g.*, streets).

Last, I developed a series of reading tools, all constrained by reading strategies motivated by visuospatial cognition theories validated by pilot studies and formal experiments. These strategies support free map exploration, but impact on user interactions in so far as they enforce reading rules developed to produce superior results.

1.6 Summary

This research seeks to establish a set of design principles for its integration of sound feedback communication into the HCI of map for blind persons. The focus of this endeavour is on the design and assessment of a non-visual user-computer interface for the construction of cognitive maps in blind users. My work draws on visuospatial theory[10] [64], perception [58] and ethnographic methods[11] [65] to explore technological and cultural issues in the development of an effective and efficient alternative to physical TMs. The contributions of this research are a better understanding of comprehensive auditory feedback mechanisms and of user-interaction methods for real-time curve-tracing in a multi-curve and multi-landmark environment. Beyond this specific application, the contribution of this research is a framework for communicating non-trivial spatial information to a perceiver who cannot see using alternative communication channels.

While accessibility for blind individuals is the driving force in my research, sound feedback can be beneficial for a broader range of users, especially when environmental conditions prevent the use of traditional visual-only displays. For example line of sight may be obscured or unstable, or visual attention may be required elsewhere. Audition, on the other hand, does not require a physical or stable line-of-sight [66, 67]. I hope that my research will open new avenues in visualization techniques and promote alternative communication and user interface technologies for all. The following chapter surveys previous research dedicated to conveying spatial information to blind persons. Chapter 3 describes the ethnographic study I conducted on blind persons interacting with TMs. Chapter 4 and 5 develop the presentation and directional encoding frameworks and Chapter 6 covers system design and implementation details. Chapter 7 assesses the prototype against TMs with blind participant, and, finally, Chapter 8 summarizes the findings, elaborates design recommendations, and discusses future research.

[10] Study the ability of humans to process and interpret perceived information about where objects are in space.

[11] Employed for gathering empirical data on users through participant observation, interviews, questionnaires.

Chapter 2

Previous Work on Map Systems

A good understanding of a user's ethnographic and goal is viewed by the HCI community as key to improve the chances of success of new user interfaces [65]. Chapter 1 set out three fundamental problems of current systems, user-centred improvement of which could improve user acceptance.

First improvement: the new system must, like TMs, be able to handle the complex geography of the real world. Sub-urban and urban environments have landmarks and streets of many geometries intersecting together. Second improvement: the new system must help blind users to construct cognitive maps. Third improvement: the acquisition and operation cost of the new system should be low and maps should be inexpensive, easy and fast to construct.

In addition, it is desirable to eliminate the need for sighted assistance, which is required in TMs. Finally, the technology must be accessible to those users who are neither Braille nor computer literate readers.

As shown below, no current system satisfies all these requirements. Previous research, which examines interface design for computerized maps, virtual curve tracing and computerized navigation systems, is diversified in its methods for capturing input and displaying output. In this chapter existing technologies are weighed against the requirements listed above, which provide a taxonomy for grouping and contrasting different solutions.

This chapter is divided into three sections that describe tabletop, *in situ* and miscellaneous solutions. The table top section discusses the usability of small surface systems for conveying spatial information to a blind person; the real-environment section focuses on systems that direct blind pedestrians in real environments and the miscellaneous section examines relevant research that does not fit into the other two categories.

In the remainder of this chapter and in following chapters, the direction of transmission of a device is labelled from a user's point of view. For example, the keyboard and mouse are output devices because a user transmits data to the system, while headphones are an input device because the user receives input from them.

2.1 Research on Tabletop systems

This section describes four types of systems: physical, tour-based and virtual systems with and without haptic cues. Physical systems include TMs: they encode spatial information for tactile input. Tour-based systems are technologies that provide information coded as music to a listener. Virtual systems with or without haptic cues give the user a physical interaction with impalpable spatial information. We begin with physical systems.

2.1.1 Research on Physical Systems

The canonical example of a physical system was created by Parkes in 1988 [68], and known as the Nomad Mentor. Audio-Touch [69], TTT [70], IVEO from ViewPlus[1] and the Talking Tactile Pen[2] are similar systems developed later. This family of hybrid technology, which layers a TM over a digitizing surface, allows a user to feel an embossed TM while hearing names and descriptions via synthesized speech. Although the benefits of these systems have been demonstrated empirically, they have not been popular, owing to the high cost of the device itself and of producing two sets of maps, tangible and auditory.

Seeking to improve the economic and practical properties of TMs, Holmes & Jansson [73] compared fixed grid overlays of 8x8 cells to TMs . They broke the content of a TM into sixty-four square cells, replicating the TM, but at larger granularity. The user receives the information by touching a touch pad located beneath the grid, hearing the names of streets(s) and building(s) within the cell. Empirical results obtained from four blind participants showed that exploration time was faster with the fixed grid overlays of 8x8 cells technique, but that the TMs provided a superior mental representation.

Miele & Gilden [74] developed T-MAP to increase the availability of TMs and address the distribution problem discussed in Chapter 1. The system generates TMs from Google's™geospatial database. The TMs can be downloaded from the T-MAP website and printed on a Braille embosser for reading. The data source limits the system to outdoor environments, and the automated construction approach offers a homogeneous solution to a heterogeneous population of blind persons, with no opportunity for customization. Users must read Braille to understand the maps, which limits possible users to a small fraction

[1] Commercially available from [71]

[2] This system is slightly different because it uses a pen cameral to read out points of interests on a tactile map. Commercially available from [72]

of blind persons. To overcome this problem, T-MAP [74] and TTT [70] researchers have integrated the two systems. However, the acquisition cost of the TTT's hardware plus the cost of a Braille embosser (∼$3000), is a significant obstacle to most blind persons.

Schneider & Jochen [75] developed a graspable interface[3] that allows a blind user to interact with map components. Routes and landmarks are signified by blocks assembled on a flat surface by the user. A camera positioned above the map captures the layout of the block fed to a computer vision algorithm, which identifies each block and speaks the name of the associated feature using synthesized speech. Block by block directions are given by the technology to help a user construct a new map from scratch. Informal testing suggest that unknown environments could be learned with this system. Unfortunately, resolution is much lower in this technology than in TMs, because the blocks need to be large enough to be manipulated by hand. Also, the camera requires fine adjustments that a blind person cannot perform without sighted assistance. Lastly, a user needs to assemble a new map before reading it, which is inefficient and time-consuming.

The body of work presented in this section suggests a trade-off between information density and cost or convenience. Systems that provide adequate detail are either too expensive and less expensive ones display too little information for the user to build a satisfactory cognitive map, as observed by Holmes & Jansson [73] . However, the digital map format used by Schneider & Jochen [75] and Miele & Gilden [74] is a possible solution to the distribution cost problem mentioned earlier. The research described below is focussed on digital maps.

2.1.2 Research on Tour-Based Systems

A tour-based system is a non-interactive sound interface that presents location-dependent information to a listener. It resembles listening to music, because the user has little opportunity to control the presentation [76]. The research described in this section shows that tour-based systems are cost effective because sound cards are standard in PCs and headphones are inexpensive, but ineffective at building spatial relations among features in a listener, which impairs the quality of cognitive maps. The most important problem is intersection, which is not represented explicitly. For example, tour-based systems present a world of non-overlapping shapes and curves, omitting much that is important to a user dealing with a real environment. Let's consider Figure 1.1 to exemplify how tour-based systems work: the horizontal corridor would be represented by a melody varying in stereo balance moving between the ears with constant pitch[4], then the vertical corridor would be played by a melody varying in pitch centred between the ears, until reaching the 'Y'

[3] An interface a user can grip to interact with a computer.
[4] The frequency of a sound.

intersection where the melody would shift towards the right ear; then, the left part of the corridor would be played starting from the branching point, increasing in pitch and shifting towards the left ear. The sequential characteristic of the display and sound perception in humans, which is discussed in Section 4.2, make it hard for a listener to determine where the corridors intersect in space. The remainder of this section examines tour-based in order of increasing user interaction.

Alty & Rigas [77] created the Audiograph to present the spatial locations of objects using sound. A variety of scanning techniques for communicating the position, type, size and shape of objects were compared. Overall, sound encodes position or pitch[5] with piano timbre[6] for elevation and organ timbre for azimuth. High frequencies correspond to the rightmost and topmost coordinates. Six blindfolded participants were presented lines, circles and squares. They could get a general idea of the diagrams, but perceiving spatial relations between objects was poor and remembering non-meaningful graphical information was difficult.

Biederman's Geon Theory [78] is applied for the sonification of two-dimensional images. Cronly-Dillon *et al.* [79] deconstruct simple images into primitive parts called geons[7], and encode their outlines into polyphonic melodies. To recognize an image a user needs to perceive the geons and their relative positions, and mentally to assemble them into a representation. A user can listen one sequence at a time (one geon) or all sequences for context (all geons). Horizontal lines are sonified as repetition of the same note, vertical lines as sets of musical notes of different frequencies played simultaneously, i.e. chords, and oblique lines as ascending and descending scales. Results from a series of experiments conducted on thirty-six sighted and six blind participants show that participants could associate complex sound patterns with geons and could perceive relatively complex images as arrangements of geons. However, most responses were spatially distorted owing to misperceived spatial relationships between geons.

Zhao *et al.* [80] built a 'choropleth map' that sonifies data collected by the US Census. It provides information to a user who can navigate the fifty states in eight directions using a numeric keypad or listen to a spatial sweep of data in adjacent states. The sound varies in timbre, frequency, stereo balance, azimuth and elevation. A percussive sound occurs when the user enters a new state. Seven blind participants used the system successfully, but post-interviews revealed that participants had difficulty knowing where they were on the map.

Kamel *et al.* [81] enhanced tour-based systems by adding interaction that provides sonic feedback. A participant finds objects in a diagram using a stylus on a digitalizing tablet.

[5] The frequency of a sound.

[6] The character or quality of a sound independent of its pitch and volume. For example, the characteristics that differentiate one instrument, voice or sound from another.

[7] Geons are basic 3D shapes such as cubes, spheres, cylinders, cones or wedges

Object specific nonspeech sounds were localized using Head Related Transfer Functions (HRTF), a 3D sound technique explained in Section 4.2.2. Frequency redundantly encoded vertical direction because participants had difficulty perceiving in elevation from HRTF. Twelve blind participants successfully tracked sound position to hear the shape of simple objects, but the HRTF encoding was reported as marginally effective for the task. On the other hand, the spatial relations of features, which was acquired with the stylus, was good.

Roth & Kamel [82] were interested in three encoding methods: auditory feedback only (tour-based), kinaesthetic[8] interaction with force-feedback mouse, and a combination of the two. Auditory feedback encodes elevation by frequency, mapping high frequencies with high coordinates & low frequencies with low coordinates, and azimuth with stereo balance, a common technique I will refer to as *frequency-balance mapping*. A percussive sound signifies hovering over a curve with the stylus. The mouse force-feedback method dragged the hand towards the curve. The study conducted on twelve blind participants puts the bimodal method as most effective for curve tracing.

The work of Kamel *et al.* [81] and Roth & Kamel [82] suggest that combining haptic feedback with sound improves spatial accuracy, the primary shortfall of tour-based systems. The adoption of a perception-action loop, with physical interaction leading to continuous kinaesthetic feedback, helped users to coordinate activities in space [83], especially in tasks requiring exploration of higher dimensional spaces [66]. Cognitive psychologists have observed that much inference is required when constructing a two-dimensional mental representation from a one-dimensional transmission [84], which probably explains the low memorability issues reported by Rigas and Alty [77]. Systems examined in the rest of this chapter adopt perception-action techniques. The next section examines of systems like that of Roth & Kamel [82], classified as virtual systems because no physical representation of space is present.

2.1.3 Research on Virtual Systems with Haptic Cues

Roth & Kamel [82] shows that providing kinaesthetic feedback improves spatial accuracy. Zheng *et al.* [85], who studied navigation devices and the development of spatial representations, recommend absolute pointing over relative pointing. Research in this section shows that spatial precision from relative mode[9] is poor and also shows that spatial precision from absolute mode is disappointing without sonic feedback. These results suggest that absolute displacement in space provides greater precision, but that adding sound to this improves performance significantly. The remainder of this section examines research that uses haptic feedback as the primary modality for spatial information, but sometimes augments it with sound that carries complementary or redundant information.

[8] The feeling of motion. Relating to sensations originating in muscles, tendons and joints [61, p.416].

[9] Relative mode devices report differences between positions, not positions themselves.

Lahav & Mioduser [86] designed a multi-sensory virtual environment for indoor areas, to teach users the layout of a remote room. A force feedback joystick simulate friction near walls and objects during navigation. Synthetic speech is also provided on request. An assessment conducted on one blind participant showed that the locations of objects near walls were learned, but not the locations of objects in the centre of the room. Neither were the shape of objects perceived by the participants. It appears that force feedback joysticks do not deliver enough spatial accuracy for perceiving the shape of geometric objects.

Patel & Vij [87] did pioneering work on route navigation with fully digital maps on an inexpensive output device. Participants were directed by synthetic speech while travelling along virtual streets using a joystick that vibrates when the cursor is near a map feature, like a landmark or the junction of two routes. Two blind and two blindfolded participants could navigate and learn virtual routes. Unfortunately, the map chosen for the assessment was very simple with little detail and only three streets, which were straight and wide. Indeed, the larger the features the less detail can be provided: a user can read a map with a joystick, but the utility of the system is debatable.

Parente [88] developed Bats, a PC-Based sound tactile map system. A blind user accesses a large scale map using a force feedback joystick that vibrates when the cursor enters a region of interest. OpenAL, an sonic 3D open source package, presents sounds and synthetic speech callouts. The user navigates a map region by region, following the intensity and direction of sounds. For example, over the ocean there is the sound of crashing waves, clicking the joystick button over a city announces its name. In a pilot study a blind participant found the auditory scene too dense and had difficulty sensing the direction of the 3D sounds.

Noble & Martin [89] examined several tactile cueing strategies for guiding a blind computer-user around geometric shapes. The system provides information through a tactile mouse indicating the direction to move by patterns in two Braille cells on the mouse. Among thirteen blind participants, four successfully recognized all the shapes, seven recognize nothing and the rest had mixed results. Apparently, the relative coordinate system of the mouse[10], prevented participants from developing accurate mental representations.

Pakkanen & Raisamo [90] created a stick interface that converts graphical information into haptic and melodic feedback. A horizontal line in the scene is represented by a physical stick attached to a computer mouse. When the stick intersects a geometric shape there is vibration in the mouse, strong at the centre of the stick, weaker towards its ends. Sound reinforces the mouse location by adopting the frequency-balance mapping technique. Nine sighted participants tested the system with little success recognizing basic geometric shapes. Spatial precision was the predominant problem.

[10] A mouse detects movement relative to an underlying surface and spatial correspondence is lost when it is lifted and no visual feedback is available.

Ramloll *et al.* [91] examined how to represent curves to the auditory and haptic senses, using a MIDI[11] synthesizer and the PHANToM[12]. Virtual grooves represented curves. Leaving a curve produced a physical force in the direction of the curve. Sound feedback redundantly provides the x and y coordinates of a curve using the frequency-balance mapping technique. Grid guidelines underneath the curves are given less friction to differentiate them from the target curve. The assessment of the system on blind participants revealed that although relatively straight curves could be traced effectively, steep curvatures were difficult to trace. Also, the participants confused curves with guidelines at intersection points, because friction levels were hard to differentiate.

Crossan & Brewster [92] studied directional cueing a non-visual approach to teach geometric shapes to blind participants. The PHANToM pulls a user through a trajectory playback sequence while sound encodes the frequency-balance mapping technique for feedback. Nine blind and six sighted participants remembered better the outline of a shape with haptic-sound feedback than with haptic playback alone. The same results were obtained by Plimmer *et al.* [93] who conducted a character shape study using the same methodology and apparatus.

Petrie *et al.* [94] developed TEDUB, to give blind users access to technical drawings, such as UML diagrams. When the user points a force feedback joystick toward a neighbouring node, the node's label is automatically spoken. 3D sound cues, of which the encoding method adopted is unspecified, reinforce the spoken output and the force-effects. Participants (neither their number nor their nature is specified) were able to explore diagrams and search for information. Although each graph has many edges, inter-edge interactions are not significant to the graph reader. Hence, spatial cues are local to nodes, allowing a user to understand a diagram without building a mental model of its geography.

Like Roth & Kamel [82] , both Crossan & Brewster [92] and Plimmer *et al.* [93] observed a performance improvement when sound is redundantly encoded with kinaesthetic cues, such as hand position. Similarly, Yu and Brewster [95], comparing the PHANToM to the Logitech WingMan, an inexpensive force feedback mouse, on curve tracing task, reported no performance improvement with the expensive device when sound supplemented the display. The high price scale of the PHANToM or custom I/O devices, as used by Pakkanen & Raisamo [90], burdens a system with high acquisition cost, something to be avoided.

Unfortunately, only the work of Patel & Vij [87] explicitly examines streets intersections, but their maps are low resolution with simple geometry , which does not occur in practice. Ramloll *et al.* [91] did have curves intersecting grid guidelines and the result was confusion. These projects are the only studies curve-intersection examined in this subsection. Except the physical systems described in Section 2.1.1 and a system developed by Jacobson [22], to be examined in the next section, no other systems addresses curve intersection.

[11] Musical Instrument Digital Interface.

[12] The PHANToM™is an expensive haptic device that delivers force feedback in 3D.

This completes the review of virtual research with haptic cues. We are now ready to examining the virtual systems that convey spatial information to a blind user without haptic feedback.

2.1.4 Research on Virtual Systems without Haptic Cues

The previous section suggested that redundant sound encoding and kinaesthetic cues provide more precise judgement. This section examines sound-only systems with no haptic feedback. They support one of two capabilities, either collision detection or curve tracing. The former provide simple feedback when the user is on or off the curve; the latter offer elaborate cues to help guide a user along a curve with non-speech sounds.

Sharmin *et al.* [96] compared auditory and haptic feedback on a curve tracing task. Participants gave output using a stylus on a digitizing tablet. Haptic feedback vibrates the stylus. The curve is widened into three parallel regions. The stylus vibrates at different rates in different regions, the central region vibrates at 167 Hz, the side regions at 250 Hz. With auditory feedback the system maps unique sounds to unique locations. The frequency-balance mapping is used, but does not help the user to stay on a curve. A percussive collision sound is added to notify a user that the stylus is over a curve. The multilayer feedback of the haptic condition is richer than the sonic feedback, explaining why the blindfolded participants traced the curve faster in the tactile condition than in the auditory one.

Following Sharmin *et al.* [96] , Evreinova *et al.* [97] compared directional-predictive sounds to vibration encodings for tracing curves with a stylus. Each curve was encoded as a series of points, each emitting a sound that diminishes in volume with distance. The collection of sound emitting points defines the curve. Different backward, forward and crossing cues are conveyed by sounds and vibrations of different frequencies, which inform the user of leaving, returning towards and being over the curve respectively. The two encodings were compared using with eight blindfolded participants. The sonic encoding showed time and precision advantages compared to the tactile encoding. These results contradict Sharmin *et al.*'s [96] observations.

Jacobson [22] studied an atlas of maps to increase visible features in a low resolution map. Normally an atlas puts each region on a separate page, thus hiding the global view from the user. In Jacobson's system a user touches the boundaries of the system's touch-screen to navigate map content north, south, east or west, which has the effect of panning a small window across big maps. Map features are drawn large for easy finding by touch, so that few map features are visible at a time. Touching an object plays a sound that represents it. Mapping object to sounds is problematic: what is the sound of a drugstore? To discover a route, its width and direction a user must repeatedly touch the screen at random locations to narrow down his search.

Jacobson [22] understood the importance of stable spatial relationships between input and output, but his atlas system fails at building cognitive maps. Atlas maps increase the number of errors and response time in sighted readers [98], thus blind users are likely to have more problems. As mentioned above, Jacobson's system is the only system in this class to support intersections, but it has the same elementary geometry and low resolution problems observed in Patel & Vij's system [87]. Touch screen navigation makes street tracing and layout understanding difficult because the small window through which the environment is shown support cognitive integration poorly. The user must move a long distance to trace a street with an adequate level of detail. Nonetheless, ten blind and visually impaired participants who assessed the system in a pilot study were able to reconstruct maps of low resolution.

Cohen *et al.* [99, 100] developed PLUMB, a software package that uses sound feedback and a stylus-based interface to present relational diagrams to a blind user. Sound frequency variations indicate progress between two nodes along an edge and loudness variations indicate the stylus's orthogonal distance from an edge. The symmetrical characteristic of this encoding is problematic in the sense that feedback does not differentiate the possible directions of errors: a user knows he is leaving the edge, but is unsure from which side. Intersecting edges are also not supported by the system. Unfortunately, no formal assessment of the software is available.

The work presented in this subsection chooses non-expensive absolute pointing devices to present spatial information to a blind user. The user perceives information positioned by kinaesthetic feedback from hand displacement. Jacobson [22] opted for a random approach to detect streets, which works as long as the level of detail is low and map features are large. Cohen *et al.* [99, 100] chose a stylus, but developed a symmetrical encoding for tracing straight edges that never intersect. Evreinova *et al.* [97] elaborated a non-symmetrical encoding that supports curvature tracing with a stylus, but her system sonically guides the user along a trajectory, preventing free user exploration. Her approach allows one sonic curve per environment.

A map technology should take into consideration the cartographer's experience; as recommended at the beginning of this chapter, a map should be faster and easier to draw than a TM. Unfortunately, Jacobson's system, the only technology that addresses map making, requires a cartographer to represent map features with sonic and visual icons[13], both of which are difficult to build or find.

Results from the work covered in this section suggest that augmenting non-speech sound with kinaesthetic feedback provide high spatial precision at a low cost. The work on real environment systems discussed in the next section examines the use of non-speech

[13] Visual feedback benefits a user with residual vision and is needed for OMSs, who rely on their vision to construct maps.

sound and kinaesthetic feedback from body displacement for supporting a blind navigating pedestrian.

2.2 Research on Real Environment Systems

Real environment systems are navigation devices that improve the mobility of blind travellers by guiding them through space. A map reader moves much faster through a city block than a pedestrian does through the environment, thus providing a slower presentation rate is fine. Lower precision is also enough because the disposition of environmental features like sidewalks and corridors help enclosing a blind pedestrians's path during navigation.

Although the systems in this section only comply to the first recommendation raised at the beginning of this chapter, that is being able to convey ecological environments, this body of work reinforces the theory that spatial representation can be precise and well understood by a blind person when kinaesthetic and sound information are joined together.

Real environment systems are divided into three classes of technologies, Electronic Travel Aid (ETA), beacon, and electronic orientation aid systems.

2.2.1 Research on ETA Systems

In the last half-century, technical advance has produced a variety of range-finding technologies for blind pedestrians, which can extend the range of the long cane. An ETA is a form of assistive technology that detects obstacles not present in a user's mental map, such as other pedestrians. ETAs have relied on users' kinaesthetic, auditory and haptic senses to compensate for vision.

A variety of user-directed pointing telemeters are shown in Table 2.1. Blind users point these devices in space to perceive measures of distance between their position and obstacles in their vicinity. Empirical assessments of these technologies suggest that the mobility of blind travellers who learn to use them improves due to their greater distance range. Indeed, ETAs allow preemptive actions to be taken during navigation; unlike a cane user, an ETA user avoids obstacles without colliding into them[14].

The twenty-two commercial systems shown in Table 2.1 substantiate the industry's widespread acceptance that kinaesthesia and sound can substitute for spatial vision in conditions of blindness. Because audition is the primary modality in blind persons to perceive space and avoid danger, blind travellers prefer to listen to the environment and get

[14] I should point out that ETAs are never used without the long cane, which serves to detect low obstacles or holes in the ground.

ETAs feedback by touch. However, haptic and sound feedback are not equal in the quality of information they provide. For example, the Télétact™ developed by Farcy [101, 102] supports two levels of vibrations for touch feedback, and twenty-eight levels of sound frequencies for sound feedback, yielding a much higher resolution with sound than with touch.

Table 2.1: This table shows a non-exhaustive list of commercial ETAs that rely on a user's kinaesthesia to determine the direction of obstacles in the environment. Kinaesthesia conveys a precise sense of direction in azimuth & elevation; sound in distance [103]. The information shown in the table is taken from [104, 105, 106, 107, 108, 109]

System	Sound Cues	Vibration Cues	Interaction
Cane C5™	✓	✓	Attached to a cane
Canterbury Child's Aid™	✓		Head Mounted
CDM-90™	✓		Head mounted
FOA™	✓		Attached to a cane
Guidecane™		✓	Attached to a cane on wheel
Handguide™	✓	✓	Handheld
K-Sonar™	✓		Attached to a cane
Lasercand N-200™	✓	✓	Attached to a cane
Miniguide™	✓	✓	Handheld
Mowat™	✓	✓	Handheld
NOD™	✓		Handheld
Pathsounder™	✓	✓	Head or neck mounted
Pilot Light™& Mini radar™	✓	✓	Handheld
Polaron™	✓	✓	Handheld
Sensory 6™	✓		Head mounted
Sonic Pathfinder™	✓	✓	Head mounted
Télétact™ I & II	✓	✓	Handheld or attached to a cane
Tom Pouce™		✓	Handheld
Ultracane™		✓	Attached to cane
Walkmate™	✓	✓	Handheld or chessmounted

2.2.2 Research on Beacon Systems

Beacon systems display trajectories from multiple emitting waypoints a blind pedestrian follows from departure to destination. Emitting stations can be physically located in space or virtually located using Global Positioning Systems (GPS) and Geographical Information Systems (GIS). The concept of beacon was introduced earlier by the work of Evreinova *et* *al.* [97], who relied on a series of successive emitting points to display the trajectory of a curve on a digitalizing tablet.

Lutz [110] examined two sound interfaces to relay wayfinding information to blind pedestrians. One method consists of providing non-speech sound that falls off in frequency and timbral quality when the user orients away from destination. The second method remains silent as long as the traveller is within the path towards destination. Distance from a landmark is given as a variation of clicks per second or in generated speech. The author mentions that pilot tests on blind participants showed potential for both methods, but no further details are provided.

SWAN developed by Walker & Lindsay [111] emits 3D non-speech sounds distributed along a given path through several waypoints towards destination. Reaching a waypoint mutes its transmission and starts the emitting of the next waypoint along the path. Object sounds of unique timbre are used to indicate walking surface transitions and the nature of the things on a traveller's path. Speech is used to give instructions and absolute values and nonspeech sounds for rapid feedback on actions. The authors observed that performance for navigation were negatively impacted when concurrent speech information was added to 3D beacons sounds, but reported that blind participants could navigate space with SWAN.

Beacon systems guides a user along a path to destination, which eliminates the problem of route intersection because one trajectory is disclosed at one time, but prevents a user from exploring the environment freely. Navigating a predetermined path by following distributed beacons is a cognitively simple way of travelling independently, but leaves a blind traveller unaware of nearby or distant on and off route features, a prerequisite to build solid cognitive maps [112].

2.2.3 Research on Electronic Orientation Aid (EOA) Systems

This class of system is divided into two groups of technologies, the position locator systems and the narrative map systems.

Unlike beacon systems, which rely on non-speech cues to guide a user in space, position locator systems use context-sensitive verbal information, messages that are determined by a user's position and orientation in the environment.

Narrative map systems are spoken instructions like traveling recipes recorded on tape or on a portable media player a blind traveller listens to, while walking in the environment.

Not unlike beacons systems, position locator and narrative map systems provide step-by-step instructions from start to end, leaving a blind traveller unaware of features outside his path.

Position Locator Systems

Position locator devices are a family of prosthesis systems designed to enhance spatial relationships and refresh the memory of routes and areas in blind travellers [9]. Recent advancements in GPS/GIS technologies offer modern alternatives to sighted guidance for outdoor navigation. Adapted GPS/GIS products are now commercially available for blind persons: the Trekker Breeze™, the MoBIC Travel Aid™, the Strider™, the Talking Radar Cap™, the Navitact™and the BrailleNote GPS™ [105, 113]. Other directional sensors like electronic compasses are also offered in stores, such as the C2 Talking Compass™. These talking technologies are limited in that they can only be used outdoors, but a recent assessment of GPS/GIS technologies on route guidance has shown that blind pedestrians travel better with them than without them [114].

There are significant limitations to these systems. Their acquisition cost is high[15] (as much as $6000) and new maps must be periodically purchased because commercial points of interests change regularly. Also, accuracy in GPS/GIS technologies is too low for precise localization, unless enhanced with a bulky and expensive differential correction system[16]. Lastly, the maps themselves are sometimes inexact, the signal can be disrupted by tall buildings or trees, and indoor usage is not possible.

Several technologies are available for indoor orientation. Commercial systems like the Talking Signs and Marco [105] use infrared light transmitters with handheld receivers to inform travellers of their location in the interior of a building. Short-range transmitters distributed through the building allow a blind traveller with a receiver to hear spoken information about his current location (as walking near a transmitter that broadcasts recorded messages). Crandall and Bentzen [116] developed a device based on light signals that does the same thing. A handheld receiver decodes frequency-modulated light in the environment to infer location in public space and communicate it to the traveller. Conventional light systems, e.g. conventional neon lighting tubes used as a light source in buildings, need to be replaced for this technology to work. These two approaches require considerable investments for their installation, and entail ongoing maintenance costs for organizations that purchase them.

Lui *et al.* [117] examined the use of Internet access points and triangulation methods to infer indoor locations at low cost. Because wireless access is pervasive, this technology has great potential because any generic WI-FI handheld device can be used as a receiver.

[15] I should point out that it is likely to see in the next few years numerous GPS applications move to the smart phone. For example, WalkyTalky is an accessible navigation aid on Android, which periodically updates the status bar with a traveller's current location to the nearest street address. See [115].

[16] Differential correction positioning technology uses ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions. The correction signal is normally broadcasted over UHF radio modem.

In contrast, a technology that relies on WI-FI triangulation cannot provide the level of accuracy a blind traveller needs because hotspots come and go [118].

Indoor technologies are limited in that the indoor environments must be altered and maintained, which diminishes the availability and reliability of these services, and in that they are restricted to providing fixed messages about the immediate local environment. Narrative maps systems, which I discuss next, work in both environments and do not require any spatial alterations.

Narrative Maps Systems

Narrative Maps Systems hold recordings of textual descriptions of the environment on tape or in a digital sound file. This technology did not allow a user to travel from/to any location on a map, or reverse a route direction, until the arrival of Click-and-Go™ [119], a state-of-the-art narrative map technology.

Click-and-Go renders public facilities accessible to blind travellers from different parts and directions of an environment by creating a variety of sound files a user can choose from. The verbal maps, which are built by professional OMSs, are free to blind travellers, but agencies or organizations must pay for their production. Accessing the service requires no special equipments or purchases besides an internet connection and a telephone or cell phone. To access a map, a user is required to browse $\frac{n(n-1)}{2}$ files to find a narrative map that matches a desired starting and ending location, n being the size of the set of destinations in an environment.

Unfortunately, narrative maps deny blind persons the possibilities of exploring the environment from different perspectives, answering their own questions and building up their own mental models of the environment, some requirements for building solid cognitive maps [3].

ETA, Beacon and position locator systems display one trajectory per environment at any given time at low resolution and slow presentation rate. Except for narrative map system, all systems require manufacturing portable computing devices customized for the user and the task, which increases their cost of acquisition. Narrative map systems can be read with a consumer-grade device, but the maps cannot be modified without rebuilding them, and need to be built from scratch every time the departure and destination change, a daunting process which increases their fabrication cost. The aim of these systems, to travel a blind pedestrian from point A to point B, differs from my research goal of constructing cognitive maps of holistic areas.

The following section examines methodological principles for readapting the visualization of assistive systems from one modality to another, discusses a space reduction method, and positions my work with respect to research on drawing systems for blind users.

2.3 Miscellaneous Research

Indirectly connected to the field of maps is the work on sonification of flow charts for blind persons [120] and relational diagrams [121, 122, 123, 124, 99, 100]. Unlike maps, the spatial layout of nodes and edges is unimportant in these graphs, which is why all the aforementioned systems, with the exception of Cohen's [99, 100] and Petrie *et al.*'s work [94] discussed earlier, readapt their visualization to preserve modality independent information. *Modality independent information* are the characteristics that must be captured in order to build a usable system when translating graphically represented information from the visual to an alternative sense [125, 126, 127].

For example, Metatla & Stockman [121] adapted the layout of complex relational diagrams to allow blind users to read and navigate graphs with a computer keyboard and synthetic speech feedback. Their system abides by three modality-independent principles: first, the translation from one modality to another must discard the dimensional representations that are specific to the original modality and only preserve the ones that are indispensable for understanding the key information to the problem that is addressed; second, information associated with each item must be preserved and represented separately from other items, including the ones of same type; third, the information must be overlaid with special navigation capabilities that allow its flexible access beyond basic hierarchical navigation. The implementation of these principles produced a multiple perspectives system, from which blind participants could gain structured access to information and reason about every aspect of a diagram.

Metatla & Stockman's modality independent principles can be interpreted for the design of a new map technology. Interpreting the first principle: colours and font attributes in printed maps are helpful characteristics for the visual sense but not indispensable information for learning the content of a map, thus can be discarded. In contrast, the spatial emplacement of landmarks, routes, their connectivity and labels are essential for learning an environment and should be preserved. The second principle can be interpreted in terms of interaction affordances, for example users should be able to 'read' a route in any direction, access any POI from any route connected to it and any route from any POIs connected to it, and so on; the third principle suggests a user should be capable of navigating without restriction from any location on a map to any landmark or route, like sighted do with their vision.

Although drawing does not require a user to trace a pre-determined trajectory, some research on adapted drawing interfaces for blind persons can inspire the design of a map interface. For example, Kamel & Landay [128] with their system IC2D, developed a grid-based interface to help blind artists relocate points in their drawings. The drawing surface is divided into a 3x3 grid arranged like a telephone's numeric keypad. Each cell is dividable into a 3x3 grid to allow finer-grained point selections. A computer keyboard is used to navigate the grid, select points, and draw shapes. Voice feedback communicates the current position in a drawing and the system's status to a user. Assessment of the system has shown that blind participants could develop meaningful drawings within four weeks of training, which suggest that this divide-and-conquer approach affords a user to effectively relocate points in space. This technique could service access to unknown landmarks or a points of interest on a map, and meet the third principle advanced by Metatla & Stockman [121].

Hong *et al.* [129] developed an interface that allows blind individuals to drive a vehicle with sound feedback. The blind driver follows spoken steering commands, e.g., 'two clicks to the right', and the steering wheel emits an audible click feedback at every five degrees it is turned. Slow or stop is signalled by a vibrating vest. Blind participants successfully drove a dune buggy across a parking lot. Unfortunately, post-evaluations of the drivers' cognitive maps was not carried out.

This elementary technique that moves a user along a 2D trajectory with simple left/right directives could be used to drive along or trace a route with an inexpensive USB steering wheel device on a map. However, this driving paradigm makes quick map reallocation to distant landmarks difficult, violating the third principle articulated by Metatla & Stockman [121]. Also, a measure of driving distance is missing and is needed to instil a mental trajectory in the driver. Lastly, research suggests poor cognitive map construction in drivers when low visual feedback is available [130], a reality blind drivers will face.

The work covered in this section wraps up my thesis' related work on map systems. Metatla & Stockman's [121] modality independent principles had an impact on SAM's design decisions, described in Chapter 3. Kamel & Landay's [131] work influenced the design of the Grid tool in SAM, as discussed in Chapter 6, and the left/right directives of Hong *et al.*'s [129] inspired a family of sound encodings I examine in Chapter 5.

2.4 Summary

This survey shows that very few systems strive to enhance blind persons' orientation: most systems fail to represent streets of complex geometry, and often restrict space to singular curves, which had researchers develop approaches that cannot be adapted easily for multi-curve environments. Other systems lack the precision needed for a user to understand spatial relationships between features, or support too low a resolution to show environments of reasonable size and complexity, all of which are needed for building cognitive maps of real environments. Table 2.2 shows that none of the related work covered in this chapter comply with the three conditions presented at the beginning of the chapter, conditions this research is aiming to meet.

Clearly, multi-curve interface research is in need, but the real HCI challenge is to study the problem while thinking about the ethnicity of the blind population, including their ability to understand space, their economic condition and the convoluted environments an interface must relay. Fortunately, we can limit the research to man-made environments, the type of areas a blind person navigates through independently. This type of habitat has streets, boulevards, roads, sidewalks or corridors along which landmarks or points of interests are distributed. Hence to be practical, paths, which take the physical form of a curve and are constituted by an aggregate of streets or corridors, must be represented and be perceived by a blind user in order to learn a new territory and navigate through it.

Table 2.2: This table shows an overview of the related work measured against the three conditions presented at the beginning of the chapter. The table can be read as follows (second row): physical systems can display ecological environments and help a user to build cognitive maps but are not cost effective. N/Y and Y/N symbolize the cognitive-map/cost trade-off we observed in this chapter. For example, the third row inside the table can be read as follows: none of the virtual-with-haptic systems can display ecological environments, however, some can build cognitive maps but are not cost effective, others are cost effective but cannot build cognitive maps. Please note that no systems satisfy the three conditions.

Tabletop	**Ecological Env.**	**Cognitive Maps**	**Cost**
Physical	Y	Y	N
Tour-Based	N	N	Y
Virtual with Haptic	N	N/Y	Y/N
Virtual without Haptic	N	Y	Y

Real Env.	**Ecological Env.**	**Cognitive Maps**	**Cost**
ETA	Y	N	N
Beacon	Y	N	N
EOA	Y	N	N

Chapter 3

Contextual Design

A contextual design methodology is adopted for the design of SAM [65]. The design process is broken down into five activities: understanding the user, understanding the task, innovating from data, designing and prototyping. This chapter describes these activities.

3.1 Understanding the User

Systems that conflict with users' self-images, or do not respect their values, do not succeed [65]. Fortunately, the blind population has been studied by ethnographers, yielding a rich literature in which I found much insight. Also, my affiliation as a volunteer with the Canadian National Institute for the Blind (CNIB) gave me the opportunity to develop friendships with blind persons and gain practical insights into their social characteristics.

Blind persons regard socialization as an important aspect of their lives. They like to fit into the sighted world as seamlessly as possible and dislike drawing attention to themselves [132, 133]. Technological inventions must be portable, effective, visually pleasing (aesthetics), perceived as 'cool' devices by blind and sighted companions. Cyborg-like solutions are resisted and should be avoided by technologists [134]. Blind persons want full autonomy from an assistive technology, *i.e*, an assistive technology should not require the need for external help, like from sighted persons, to use it [135].

The blind population is heterogeneous in terms of capabilities. Individual difference variables include cognitive abilities like attention and memory, learning approaches, perceptual abilities, spatial abilities, listening skills, musical abilities and level of independence [66]. Most of the blind population has low incomes [136].

3.2 Understanding the Task

Edwards [24] points out the importance of involving blind persons in the design and evaluation of innovations intended for their use: 'Learning to use a device is an investment, and any person will only use something if they perceive that the payoff in terms of usefulness to them is worth their investment of effort in learning to use it and using it'. This section describes research I conducted using the contextual design framework, developed by Beyer and Holtzblatt [65], to understand the task of reading tactile maps (TMs).

3.2.1 Method

The goal of this study is to discover the common structures in the behaviour of blind users reading TMs and to examine system design ideas drawn from observation data [65]. A small group, five blind adults and two Orientation and Mobility Specialists (OMSs), was chosen for the field study.

The study on blind participants was divided into two sessions conducted days apart. Session one: the blind participants were interviewed and observed when interacting with TMs. An expert-novice relationship was adopted with the researcher (the author of this thesis) playing the novice role. Participants were asked to think aloud during the observations. Session two: blind participants were presented with low-fidelity prototypes inspired by observations from the first session. The Wizard of Oz (WoZ) protocol was used during the low-fidelity sessions. The OMSs were interviewed, shown the low-fidelity prototypes and asked to comment on them; I met with the OMSs several times.

The sessions were conducted on one participant at a time. All participants knew that the study was carried out to guide the design of a new computerized map system for blind persons and as part of my PhD research in Computer Science. The blind participants were paid $15 per session. The OMS were paid $15, which they donated to charities. The study received ethics clearance from the University of Waterloo Office of Research Ethics.

Details about the participants follows, then the interviews and observations are summarized and, finally, the resulting data is consolidated for the design of the SAM technology.

Participants

Table 5.1 provides details about the blind adults who participated in the study. They are all Braille literate and have excellent computer skills. CBF1 recently graduated with a bachelors degree in science, CBF2 worked as a rehabilitation teacher in the past but is now working in an office, CBM is looking for work in communication and technology, ABF is

Table 3.1: This table shows a profile of the blind participants. The left column shows the participants' identification: CB and AB stand for congenitally and adventitiously blind, and F and M stand for female and male respectively. In the topmost row, Age is rounded, Onset is the age of onset of blindness, Training is TM Training (school). Inside the table, L stands for light perception, N for no residual vision, D for guide dog, C for cane, Y for yes, and S for self-trained.

Id	Age	Onset	Residual Vision	Mobility Aid	Training
CBF1	30	0	L	D	Y
CBF2	55	0	N	C,D	Y
CBM	30	0	L	C	Y
ABF	30	10	L	C,D	S
ABM	43	6	N	C,D	Y

completing a PhD, and ABM has a managerial position in a leading communication and technology company.

Each OMS has more than twenty-five years of work experience. One is working as an OMS instructor in a school for blind children, the other as an OMS instructor for a non-profit organization helping independent blind adults. Because the working environments differ significantly, I identified the participants with acronyms, which are useful for correlating answers. SFB is the specialist who works in the school for the blind, while NPO works in the non-profit organization. Similar naming is adopted with the blind participants, using the identification codes in Table 5.1.

3.2.2 Interviews with Blind Participants

A series of questions addressed design and accessibility issues. This section summarizes the points I found important for the design of SAM.

First I addressed the problem of readability. All the participants agreed that too much information makes TMs difficult to read. They also reported that details often obscure the big picture, and that aesthetic ornament was a nuisance. They indicated that ornament often requires extraneous explanations. Low standards of construction were reported as impairing readability, as illustrated in a colourful way by CBM.

> `Some people make maps out of anything they have laying around. One time I was reading a map and asked the person who built it, 'Why is there a parking lot there?' The person replied, 'Oh no, this is a street, I ran out of this stuff so I used that instead'.`

All the participants valued good tactile differentiation and simplicity which, they indicated, requires preserving only essential information.

> `Maps are easy when they display where the intersections are, what are the different streets, the traffic lights and the stops, that's it. I do not need to know there is a hospital on the left, I would not know anyway because I don't see it... these are details that don't need to be there, and when they are not, reading a map becomes fairly easy (ABF).`

Another important point that came out was the concept of natural mapping [137]. Four participants mentioned a preference for the symbols on a map to match the appearance of what they represented, such as representing a patch of grass with material that feels like grass. Standardization of symbols was also important. The participants considered it pointless to use different shapes to illustrate different buildings. Reducing the number of distinct symbols reduced the cognitive workload when reading a new map. To indicate particular stores they preferred Braille labels to unique symbols plus a key.

Blind adults do not use TMs frequently. Two participants reported using TMs about once every five years and the others once a year. When asked why, non-availability was always the reason.

The participants reported using alternative but sub-optimal solutions to find their way around.

> `I use MapQuest to get an overview of distance and directions. But it is not ideal. For example, it would sometimes tell me, 'Bear left,' but I am walking not driving (CBM).`

I asked the participants to recount their most recent experience with a TM: how they requested the map, how they obtained it, who built it for them and the time it took them to get it. Three said they had never requested a map, two that the maps they get are constructed on the spot using a Picture Book[1] owned by the instructor, which helps them understand what they have encountered in the environment during their training.

I also asked about the importance of system portability: how useful would they find an application that allowed them to read maps at home. Three participants said they would use it, but preferred a system that provided maps readable in the environments they portrayed. The other interviewees would use only a home system.

[1] The Picture Book consists of a Velcro surface onto which plastic geometric shapes and lines can be assembled to produce quick illustrations of spatial relationships

I asked for the participants' opinions about map customization. Two said that a map customized for their specific needs was easier to read than a standard map. Another said that maps should be constructed to accommodate individual needs. She explained that a reader who does not like details should be able to remove them leaving the bare minimum, while someone who wants to know an environment with many details should be able to put them back in (ABF).

All the participants indicated that spatial sounds moving around in space were preferred over a spoken Cartesian coordinate system for spatial direction in auditory games.

> `I can never associate the x-y coordinates properly. I cannot assimilate them in my brain [...] instead I use the left/right echo, which tells me when I am in a narrow corridor, when there is an intersection or an opening (CBF1)`[2].

3.2.3 Interviews with OMS Instructors

The two OMS participants reported constructing and using TMs. Both viewed TMs as excellent tools for teaching spatial concepts to their clients. When asked why maps are not used by blind adults, NPO said there was a lack of time and money to construct them, a universal reality [19, 16]. She started to ask clients with whom she conducted orientation and mobility sessions to type the navigation information into a text file after receiving their training session. Then, using their computer's screen reader, her clients can listen to the travelling descriptions before leaving home the next time they need to travel that path.

When asked what artifact they use the most to explain spatial concepts to their clients, NPO mentioned the Picture Book, and SFB indicated that financial constraints limited her to home made things like 'crafty kind of maps, quick on the go'. She indicated she would prefer using a computer.

When asked which characteristics of TMs should be preserved in a new technology, NPO recommended preserving points of interest (POIs). She explained that blind travellers use POIs to locate themselves. She added that TMs provide orientation cues and allow a blind traveller to understand spatial relationships in the environment. A new technology, she said, should do all that.

Both experts stated that too much information on a map reduces their clients' ability to understand the information so that TMs must be kept as simple as possible. Both indicated that maps are conventional and training is needed [5], and that blind persons vary in their ability to understand maps [138]. SFB pointed out a difference between

[2] CBF1 is referring to the auditory version of Pacman™

AB and CB children, the latter group having more difficulties with spatial concepts in general [34]. She indicated that most of her clients benefit from maps, but some lack the ability to read them well. Often, she added, it depends on how simple the map is.

I asked which environments are most frequent in TMs. Bus terminals, malls, subway stations, campuses and university buildings, public building, new workspace, difficult street intersections and bus routes were the most common answers. Note the predominance of indoor environments. Both preferred portable systems to home systems because carrying maps makes their clients feel more secure. However, NPO pointed out that when her clients have a good understanding of the environment, they no longer need to carry a map.

> `Like a sighted person that looks at a map to prepare a trip, a blind traveller can get a clear idea of where she wants to go by looking at a map at home (NPO).`

SFB pointed out that whether or not a system that is fun to use is a determining factor for technology adoption.

3.2.4 Observations of Blind Participants

The participants were asked to read simple TMs I built from the design recommendations found in [139, 9, 5]. These maps were shown to NPO to validate their legibility before using them with blind participants. The summary of my observations is reported in this section. Figure 3.1 shows the material I used during the sessions.

The focus of the sessions was to isolate the key hand movements that form a motor-copy image of the map, while trying to group the strategies that govern focal and contextual exploration. Although excellent research on bimanual interaction exists (e.g., [140, 141, 142]), and research on tangible graphics is not new (e.g., [143, 144, 145]), neither has been applied to navigation technology *per se*. Thus, I could not rely on existing literature to determine user intentions and needed to obtain original data. However, relevant observations from the literature are mentioned in the text. Braille labels were omitted because only a small percentage of the blind population reads Braille and I did not know in advance who would participate in the study. I decided to say the name of the things out loud instead, a Nomad-like system response [68], as explained in Section 2.1.1.

Exploration Strategies: When presented with a map for the first time, a reader must explore its features and their spatial relationships in order to build a cognitive map of the environment. Motions of the explorative hands are categorized into micro- and macro-movements [146]: the former investigates the texture of objects and seems to play a minor role in the perception of spatial qualities (roughness detection, Braille symbols), the latter

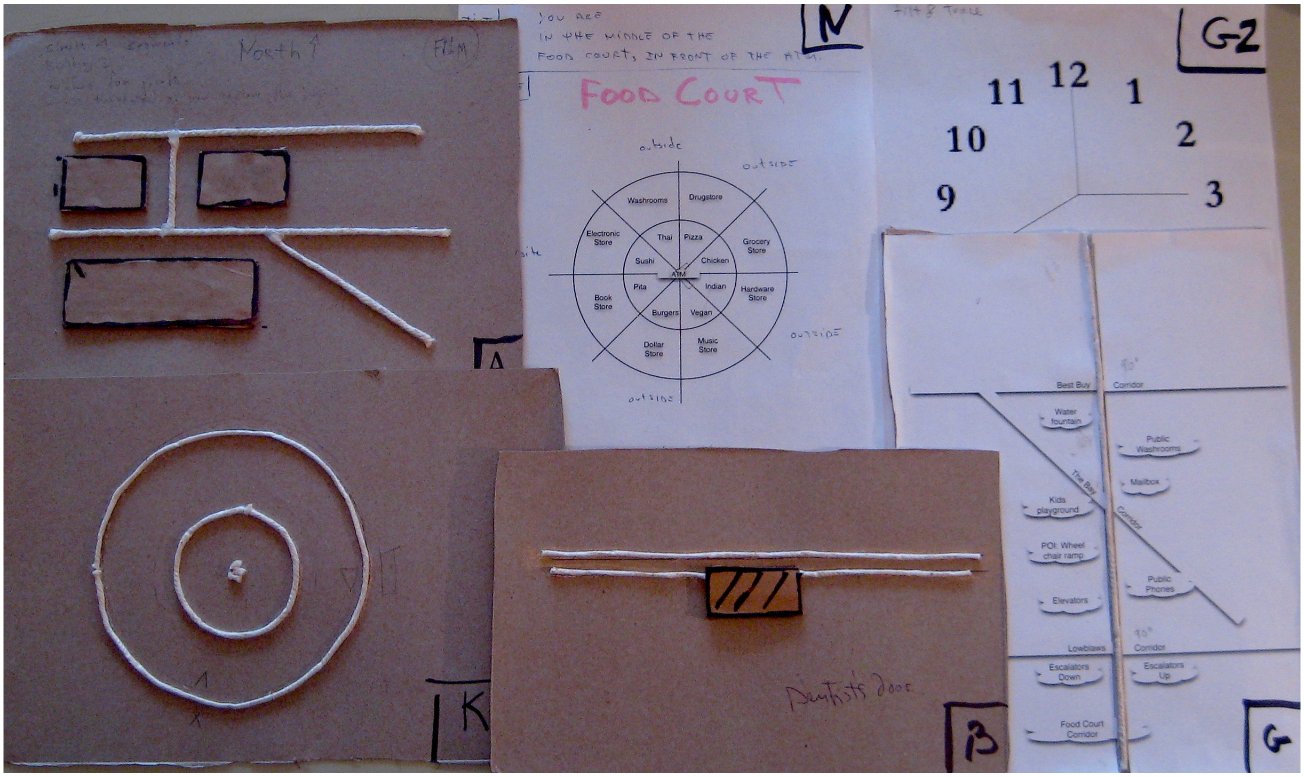

Figure 3.1: This figure shows material I used during the study. All material had a two-hand-span maximum width, as recommended by experts [9]. White cords of about one-quarter centimetre in diameter were glued on carton boards to represent streets or paths. Sandpaper rectangles raised about half a centimetre above the surface represented buildings. The non-tangible material on the top-right served for the WoZ sessions.

constructs a cognitive map from motor copies of feature locations. Macro-movements are the focus of this investigation.

Two tactics were observed. CBM, CBF1 and CBF2 began by tracing the boundaries of the geometric figures (i.e., buildings). They then explored each building successively, and returned to their starting points. ABM and ABF did not pay as much attention to building edges and searched for the largest, referring to it as a potential POI. They looked for spatial relationships between buildings from this point, while alternating between newly discovered buildings and familiar ones. Similar behaviour has been observed in earlier studies [147, 148]. All the participants were systematic in their exploration, adopting either a horizontal or vertical search strategy, a behaviour also observed by Berla [149].

Interestingly, attempts to examine the streets and how they connected the buildings occurred later in the exploration. When asked why, the participants who were able to answer the question indicated that buildings are larger than the streets and their attention is directed towards them.

Focussed Exploration: All participants used their Braille finger[3] to trace routes and outlines of objects. The dominant hand was kept open but relaxed in a position that allowed the other fingers to detect nearby features (mostly buildings). Sometimes the palm of the dominant hand rested lightly on the map, which increased its stability and confirmed the presence of things underneath the Braille finger. When an object was identified by the Braille finger, the index finger of the non-dominant hand was moved to the location of the Braille finger, staying on the newly identified feature while the Braille finger examined new objects near the feature just identified. Presumably, the non-dominant index finger provided a spatial reference for the Braille finger [142]. One participant (ABM) occasionally used his two index fingers to trace routes, starting at a central point on a line and moving his fingers horizontally apart on horizontal routes and vertically apart on vertical routes. The other participants traced with their Braille fingers. The two-hand activity I observed was similar to Braille reading technique, in which the index finger of the non-dominant hand marks the return position for the reading finger [150].

All participants explained that they imagined themselves walking in the environment when tracing routes. The participants reported perceiving their finger displacement directly in environmental terms, maintaining their awareness of their displacements in terms of changing distances and directions relative to the features in the surroundings. They all explained how features on their left when going in one direction shifted to their right when going the other. This suggests that body position relative to the map representation was perceived during reading.

> `In the same way that a map is a scaled down version of the world, my finger is a scaled down version of me (CBM).`

Hand Teleportation: As participants familiarized themselves with a map, their dominant hand started jumping to previously discovered features in order to refine spatial relationships among POIs. The jumps, in which the finger lost contact with the map at one location and regained contact at another, hence teleportation, were precise in the sense that the readers knew where the target was, owing to kinaesthetic feedback.

Frames of Reference: As the participants built their cognitive maps, I observed that the thumb of the non-dominant hand often contacted a feature at the edge of the map, mostly corners of buildings, but also extremities of streets located near the edge of the map. Again, the non-dominant hand established a spatial frame of reference within which the dominant hand could move [140].

[3] The index finger of the dominant hand.

3.3 Innovation from data

The analysis that follows presents a synthesis of the data gathered from the interview and observation sessions. The section is divided into three parts: map representation, user-interaction and user-output device. Information in each part is taken into account for designing the map technology, which I interchangeably refer to as device and system.

Implications for Map Representation

The interview data I obtained and the literature on TM design [9] agree that TM layouts must be simple. The participants were unanimous that details obscure the big picture, and that ornament is a nuisance. They reported no benefit in knowing the size of buildings and were indifferent to knowing the name of irrelevant buildings along the route, which are invisible to them. They wanted to know where to enter a destination building and to have confirmatory details they could perceive, like traffic direction, traffic signals, ramps, stairs, smells, and so on.

TMs are hard to design and produce. They require design by human specialists who rely on heuristics that stipulate pitfalls to avoid [9]. To help, unnecessary information should be eliminated, something Metatla & Stockman [121] suggested in Chapter 2, an opinion shared by other TM researchers [151, 152, 153]. Indeed, Golledge [153] recommends two basic rules for TM design, regardless of their type: include information that is absolutely necessary, and err on the side of providing too little. I could find three reasons for uniformity and simplicity in the design of orientation maps:

1. People tend to impose symmetry on cognitive maps because memory retains distance less well than area, alignment or orientation [64]. Hence, enforcing a reasonable symmetry, at the risk of distorting an environment, helps learnability. For example, smooth curves that are hard to detect can be straightened. Empirical studies show that sighted and blind pedestrians veer to one side or the other when walking, which adds variability and error to location judgement [154].

2. There is variability in the spatial ability of the blind population. Designing maps for the least capable user increases the number of maps that can be read by the population of blind persons. Removing unnecessary ornament and keeping what is absolutely needed makes a map readable by more potential readers.

3. Man-built space is predictable, allowing a blind traveller to generalize about the environment, and eliminating the need for redundant characteristics of spatial structure and arrangement [7]. Evidence shows that blind persons can build mental maps

from simplified and schematic topological maps [152, 7]. We note that most visual maps simplify reality without losing utility for navigation [155, 64].

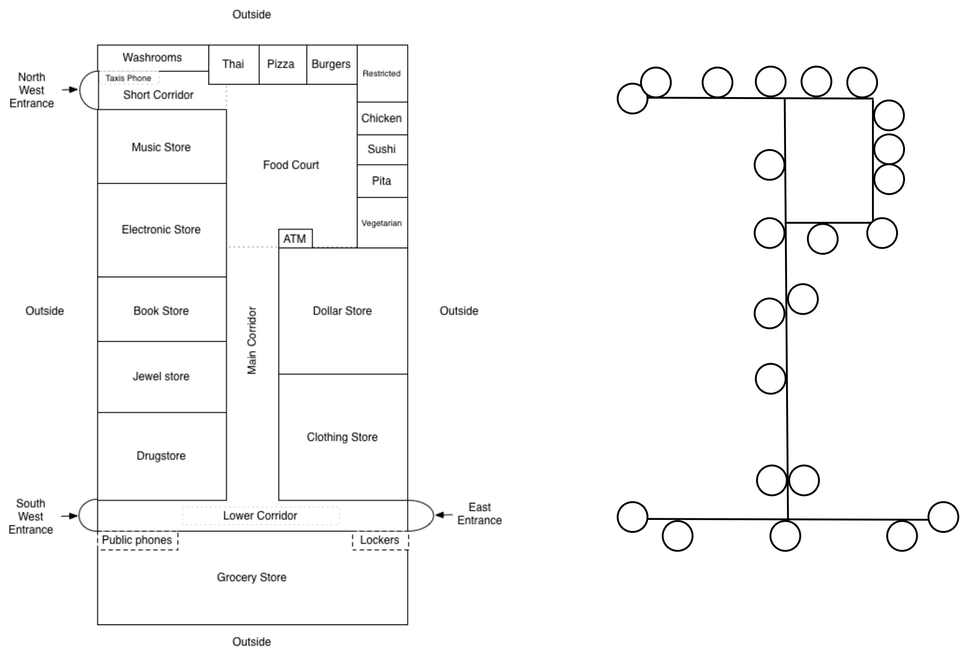

Figure 3.2: The *left figure* shows the map of a hypothetical mall. The *right figure* shows the same mall but redrawn based on feedback from the blind participants and influenced by the work of Preiser on TMs [152]: the circles represent entrance doors and important points of interest such as ATMs or telephones, and the straight curves represent the corridors.

The OMSs agreed maps should be easy and fast to prepare, update and distribute. Looking at Figure 3.2 for example, we note that the map on the right is much easier and faster to draw than the map on the left.

The forthcoming system should be effective and efficient without being expensive to acquire and feed with environmental data: the blind participants indicated that they were willing to spend time learning a new system, but only on the assumption that it provides useful information at a reasonable cost.

The resulting technology should be minimalist, supporting geometric figures like curves, which may or may not be straight, and points which have fuzzy non-zero radii. Curves would symbolize environmental features that afford linear motion, such as sidewalks and streets. Points would symbolize POIs, such as the entrance to a building, a bus stop, or confirmatory details a blind traveller can perceive. A prototype layout, which was presented to the expert during the study, is shown in Figure 3.2. Figure 3.3 on page 47 shows a similar approach adopted by the Royal Museum of Ontario in Toronto to help blind visitors navigate their exhibits.

Evidence shows that while map rotation reduces the mental workload in navigation tasks, it hinders development of a cognitive map in the reader. Cognitive map formation benefits from a stable orientation provided by a north-up map [156]. Maps in the resulting technology should be kept aligned north-up.

Implications for User Interfaces

In the observation sessions the participants began their exploration by finding POIs (buildings) first and curves (streets) later. This strategy is viewed as sub-optimal by several researchers. Indeed, their research suggests that path structures are the most important features for learning an environment, not POIs [157, 158, 159, 160].

Because blind pedestrians navigate streets and path, along which POIs are distributed [161], the resulting technology should encourage discovery of POIs via route exploration. This decision is supported by evidence that humans understand space by integrating disparate pieces of spatial information (POIs) through common object reference and physical interconnectivity [64]. Indeed, Thorndyke [162] argues that cognitive acquisition from path exploration extends the knowledge contained in unordered acquisition by including sequential information in which the sequence will be used.

Research shows that self-directed exploration, i.e., people moving through and perceiving their surroundings, leads to greater environmental learning [163] and that active learning outperforms passive learning [164]. The resulting technology should leave the user free to go anywhere on the map and to explore space freely.

We recall the precision with which the blind participants could teleport their fingers on the map. This activity is viewed as an effective approach for strengthening the construction of cognitive maps in a blind reader [3]. Kinaesthesia is well developed in blind users and the resulting technology should exploit their ability.

Implications for User Output Device

Pointing devices based on relative position, like mice and joysticks, are sub-optimal for helping blind users to build cognitive maps, whether or not they are augmented with tactile or force feedback. These devices report differences between positions, and not individual positions, to the visually disabled user. Previous work covered in Chapter 2 shows low accuracy in the mental map of the participants and the difficulty of knowing where they were on the map [86, 87, 82, 89, 90, 88, 94]. Blind persons, like sighted ones, must know their absolute position with respect to the frame of a map [3]. Blind users, who lack visual feedback and cannot see the cursor, need strong cues in another modality.

Force feedback devices like the PHANToM, as used by Ramloll *et al.* [91] in Chapter 2 are unaffordable to the majority of blind individuals and not necessarily the best device for the task: we recall Crossman & Brewster's observation of blind participants losing of control of the PHANToM's end effector [92]. Low cost and user autonomy should prevail.

Relative position devices do not support random access - like teleportation - but force a sequential exploration of space, infringing the third modality-independent principle proposed by Metatla & Stockman [121]. Among the sighted, printed maps, which can be read non-sequentially are more effective than navigation for constructing cognitive maps [64]. Blind readers can only hold positions with their fingers or use kinaesthetic feedback.

A multi-touch surface could provide the affordances of TMs, but raises precision issues: human finger as a pointing device has much lower 'resolution' than a stylus tip, which 1) makes it hard to point at targets that are smaller than the finger width [165] and 2) difficult to know the cursor location beneath the fingertip when no visual feedback is available [166]. A stylus interacting with a digitizing tablet is precise, dependable and light (touch screens like used by Jacobson [22] in Chapter 2 are not), and affords controlled teleportation. Using an unmodified consumer-grade digitizing tablet keeps the price low. Adding a vibrator to a stylus, like proposed by Sharmin *et al.* [96] or Evreinova *et al.* [97] in Chapter 2 should be avoided because of acquisition cost. Modelling the resulting technology on a surface device allows blind users to exploit their knowledge of TMs, which most have acquired through training [9]. The form-factor of digitizing tablets is comparable to a TM: both are planar 2D surfaces analogous to the ground plane on which most travelling occurs [5].

Although multi-finger touch input cannot be replicated with the stylus, other stylus degrees of freedom can be used with digital input analysis to obtain richer input. For example, even inexpensive tablets capture pressure, tilt, rotation and hover from the stylus. The computer can maintain a complete interaction history, which can vary the interpretation of the input. Similarly, the computer can have complete information about all features of the map, such as points of interest, extremity of curves, intersections, sharp curvatures and direction[4]. The rich computational environment can compensate for the loss of multi-touch input.

The tip of the stylus can be used to engrave TMs, something that cannot be done with fingers on a multitouch surface. An extra layer of Styrofoam or tactile drawing film can be laid on top of the tablet's surface. The drawing film reacts with friction creating raised lines, which are readable by touch. This way, blind users can build their own TMs and

[4] Sharp curvature and extremity of curve areas are important because they break curve tracing flow. Intersection and point of interest areas are important as they are focal points with respect to which blind travellers make decisions about bearing.

drop cognitive traces to mark discovered features on a map at low cost [98, 167]; Styrofoam panels and tactile drawing film layers cost 25¢ each [168].

Blind readers rely on their Braille finger when exploring TMs. Blind travellers are taught to extend their index finger along the hand grip of the mobility cane, so that the cane becomes an extension of the finger, which improves localization of obstacles [169]. By analogy, the stylus is an extension of the Braille finger.

The resulting technology should eliminate Braille labels and symbols, and providing spoken detail-on-demand. Sound is much more flexible than touch because computerized tools for creating and editing sound are far more advanced that ones for creating physical artifacts.

Mode switching[5] must occur without keyboard interaction to allow a reader to keep both hands on the surface, a common practice for building reference frames for the dominant hand.

Portability is beneficial but not essential. Conversely, all users wanted implementations for home computers, because off-line learning is essential to them. Thus, portability should be a secondary goal.

The discoveries described above motivated the initial design of SAM. A Digitizing tablet and a stylus, which report location and not just velocity to a user and which are inexpensive to acquire, are chosen. With a stylus, hands and arms never obscure the map, hence interference issues are eliminated, allowing a user to keep map locations in reference by leaving their fingers on the surface while exploring new areas with the stylus.

3.4 System Design

SAM is built into two modes, writing and reading. The cartographer uses the writing mode to draw curves, POIs and annotate them. A simple drawing interface is implemented to support SAM's writing affordances.

Reading mode, the focus of my research, provides acoustic feedback, but includes a visual representation for users with residual vision. Blind users explore a map by moving a stylus on the tablet while receiving auditory feedback that describes the spatial environment of the stylus. In this mode, information inflow comes exclusively from the user, *i.e.*, by feeding the system with stylus displacement input. Information outflow is system response, *i.e.*, how the system responds to a user who moves the stylus at different locations. Outflow to blind users is the primary focus of my research.

[5] A user performs the system mode switch, or changes the system's state, when he needs access to new system resources like a new tool.

Digitizing tablets are flat and do not provide tangible information, hence users cannot rely on the cutaneous mechanoreceptors of their fingers to detect edges and discriminate roughnesses. Touch, which is normally used to direct hand movements on a tactile map, is substituted by sound. User's kinaesthesia remains the sense a blind reader uses to extract geographical information: a listener moves the stylus in space to trace non-tangible curves and discover POIs. This approach raises interesting outflow challenges because curves on a map have unique labels a user needs to know, curves can intersect, turn, end, and may have POIs distributed along them, sometimes over intersections, turns and/or ends, and sometimes by themselves.

I drew upon data and knowledge acquired from visuospatial cognition theory [64, 98, 155, 170, 171] to design a series of tools and interactions I thought would cast information in ways that make sense to a blind user and help in building cognitive maps. Low fidelity prototyping sessions were conducted to assess these ideas and collect new data to improve them.

3.5 Low-Fidelity Prototyping

The sessions with blind participants observed them working with mock-ups of possible designs. The sessions with the OMS examined data presentation methodologies.

3.5.1 WoZ on Blind Participants

The Wizard of Oz (WoZ) protocol [65] consists of having participants interacting with a fake system they 'believe' to be autonomous, but that is being operated by the researcher. The WoZ protocol is frequently used for prototyping user interface designs and usability testing before having actual application software in place. The benefits are twofold: timely and honest feedback from participants. The participants tend to critique more openly a low-fidelity design than a design they know the researcher spent time and effort implementing.

The blind participants played user roles in low-fidelity conceptual design sessions. A series of tangible and printed maps were constructed for the sessions. Figure 3.1 on page 35 shows some of them. I read text printed on the maps and personified the system in response to user output during WoZ simulations.

3.5.2 Data Presentation Mock-Ups on Blind Participants

The concept of notification sounds was examined. Winn [172] suggests that perceptual precedence of map features (e.g., pop-out features) receive early attentional resources affecting the hierarchical organization of features in the perceiver's cognitive map [173]. Notification sounds served that purpose.

Sliding the stylus on the maps shown in Figure 3.1 'played' notification sounds to alert the participants to salient areas like POIs, intersections, turns and end of curves. A POI was represented by a three-note arpeggio of a G major chord like the sound heard in public spaces before an announcement. An intersection was represented by the sound of a cuckoo, which is widely used in North America at traffic light intersections. A curve was represented by a rotary sound characterized by phase shift variations, which suggests spinning, rotating, or in the case of curve-tracing, turning. An extremity was represented by a synthetic sound gliding down in pitch that semantically suggests the end of something, like 'losing a life' in sonic games designed for blind persons. The participants found the sounds easy to distinguish from one another and commented their meaning was easy to remember.

Maps are normally made of multiple curves and POIs distributed in space, hence being at a location on a map involves being surrounded by features. The system must be able to display features in the vicinity of the stylus, but space must be described sequentially because audition is temporal and weakly spatial. The names of the features can be enumerated by angle or distance to a user. Reporting the features by distance consists in presenting the closest features first, regardless of directions. Reporting the features by angle consists in presenting features circularly, regardless of their distance. The former was preferred to the latter because things at closer proximity were viewed as more important. However, both methods were viewed as too slow. No participant liked waiting for information, a chronic problem with sound.

Filtering out unwanted data items helps trim large data collections to a manipulable size, which lets the user focus on the items of interest; the tools explained in Section 3.5.4 are characterized by their unique filtering methods. These tools however, need to be selected by the user, hence mode switching is discussed before.

3.5.3 Mode Switching on Blind Participants

As observed in Section 3.3, mode switching should not require a user to move either hand from the surface of the digitalizing tablet, which is why I examined the possibility of enabling a user switch modes using the stylus.

The participants, some of which had very little experience with handling a pen, were handed a stylus and asked to move it in different directions and angles. All the participants could differentiate straight from tilted, could increase & decrease pressure on the stylus, could multi-tap the stylus while staying centred at a location on the tablet, could press the stylus buttons, and could hover the stylus while keeping hand-jerk under control. Three participants indicated it would take them a little bit of practice to become comfortable with holding a stylus. These actions will serve to select different tools, explained in the following section.

3.5.4 Tools and Stylus Interactions on Blind Participants

Three issues were examined in each session. I presented the participants with families of novel tools and interactions to measure their ability to understand the tools, to use them effectively, and to identify ergonomic issues that could result from stylus manipulations. A list of the tools, interactions, and brief descriptions are shown in Table 3.2. The same common structure is used for each tool: I define a tool, explain its rational, affordance, and report user responses.

Table 3.2: This table shows an overview of the tools and stylus interactions I examined during the low-fidelity sessions.

Tool Name	Stylus Interaction	Tool Description
Front Tool	Tilt & Rotate	Displays features in the pointing direction
Grid Tool	Hold Button & Move	Displays features inside the current cell
Radius Tool	Increase Pressure	Displays features in the periphery
Radius Tool	Multi-Tap	(Same as above)
Alignment Tool	Tilt & Slide	Displays features orthogonal to the tilt angle
Details-on-Demand	Hoover	Displays information of features
Details-on-Demand	Multi-Tap	(Same as above)
Details-on-Demand	Press Button	(Same as above)
Index Tool	Hit a Keyboard Key	Displays a feature's name and information
Index Tool	Slide Stylus	(Same as above)

Front Tool - Tilt & Rotate: Like vision, the front tool has a limited field-of-view and only sees the things that are in front of the user.

This vision-like functionality, which allows a user to look around while staying at a location on the map, seeks to reinforce feature-to-feature encoding. Blind users who adopt a strategy of relating objects to each other and/or to the frame of a map are less affected by map rotation [174], reduce systematic error in estimation of distance along road segments [175] and build superior cognitive maps [176]. Furthermore, spatial memory is said to be orientation specific: it is argued that the orientation in which spatial stimuli are viewed is preserved in memory and serves to organize other non-viewed spatial relationships [177].

To use the tool, the user tilts and rotates the stylus to get information in the pointing direction of the tilt. For example, a user puts the stylus on a POI and rotates it to find the features around that POI.

Two interpretations of directionality were possible, towards the sharp end and towards the eraser end. The participants reported the former to be more intuitive. All the participants understood the tool and could rotate the stylus, some with more dexterity than others. I observed the participants who experienced vision in their lifetime understanding better and appreciating more the tool than the CB participants.

Grid Tool - Hold Stylus Button & Move: The tool is derived from the work of Kamel & Landay [128] in Chapter 2, which divides the map into nine non-intersecting square regions, like a tic-tac-toe board, to reduce search space. Features inside the region that is in contact with the stylus emit their presence (beacons). Features outside the contact region are filtered out and remain silent.

Humans learn unknown environments better when a map is partitioned into smaller areas [162], which allows them to use a local to global strategy while learning the content [178]. McNamara *et al.,* [179, 180] have found that people impose clusters on non-partitioned maps with locally defined frames of reference, which are later related to each other in higher-order frames of reference. There is a body of evidence that suggests that imposed divisions facilitate the memorization of landmark locations [98].

To use the tool, the user holds the stylus's upper button (near the eraser) and moves the stylus into one of the nine regions.

The tool was well understood by all the participants. Stylus manipulations associated with the tool, which were characterized as easy and relaxed, were preferred over the front tool. One participant suggested having the system support columns and rows as well.

Periphery Tool - Pressure or Multi-Tap: The periphery tool filters out everything that is outside a peripheral distance range of the stylus, which is adjustable by the user. Like audition, the tool has a field of view of 360° and a limited perceivable distance range.

Pressure and multi-tap interactions were examined as a means to increase and decrease distance range. Higher pressure or greater number of successive taps increased distance.

The tool was conceptually well understood, but the participants experienced difficulties mapping pressure or multi-taps interactions to a distance range. Both interactions were found confusing. Increasing stylus pressure was found difficult by one participant who had limited physical strength, especially after repeated tries.

Alignment Tool - Tilt & Slide: The alignment tool detects alignments of features in the vertical or horizontal direction on a map.

The alignment tool was designed in response to the horizontal and vertical exploration strategies observed by blind users of TMs in Section 3.2.4. Also, horizontal and vertical directions are in alignment with the cardinal directions. Studies of spatial reference in human memory show that egocentric orientation aligned with salient external references, like the cardinal directions, are learned more easily than others [181].

To use the tool, a participant slides the tilted stylus across the tablet horizontally or vertically with the eraser-end down. The analogy of slicing a map in two with a blade was used to explain the tool. Vertical movement speaks the names of the features located to the left or the right of the stylus in the left or right ear, respectively. Horizontal movement speaks the name of the features above and below the stylus at different pitches, with high pitch higher on the map. Fast motion transforms the features as short percussive sounds, in which clusters are easily detectable.

The participants understood the tool well and liked its simplicity. Stylus manipulations felt comfortable for all the participants.

Index Tool - Hit Keys or Slide Stylus: The index tool (or map key tool) similar to the map legend found on printed maps and TMs. The tool allows a user to find a given feature on the map without needing to scrutinize the entire map to find it. Hence the motivation for this tool is to support map index functionality.

I examined two possible interactions to access information: associating every feature on the map with its unique keyboard key and distributing every feature on the map along the bottom edge of the tablet. Features are accessible with the stylus in the latter case. To use the tool, a participant needs to hit keys on the keyboard or slide the stylus. With the keyboard, a key is mapped to one feature or nothing. Searching a feature on a map index is normally the first step performed by a map reader, hence keyboard interaction is acceptable with this tool.

The tool and the two interactions were well received. Two participants preferred the keyboard interaction because they could hit physical buttons (keys) they could feel by touch to find map features.

Details-on-Demand - Hover/ Multitap / Button: Details-on-demand provides further information to the user who requests it. The motivation for this tool is to empower the user with the possibility of obtaining rich details if needed, while keeping the amount of sonic information displayed by the interface to the user as low as possible.

To use the tool, the user above a feature can hover the stylus, multi-tap the stylus or press the button near the stylus tip to have the system speak the feature's name and further details, if available. Hovering the stylus was the favourite option because it requires less stylus manipulation and the least effort.

Curve Tracing: Low-fidelity prototyping methods for directional and curve tracing cues were found ineffective because of the high density of information a system needs to convey during stylus displacement. They were abandoned in favour of computerized prototypes described in Chapter 5.

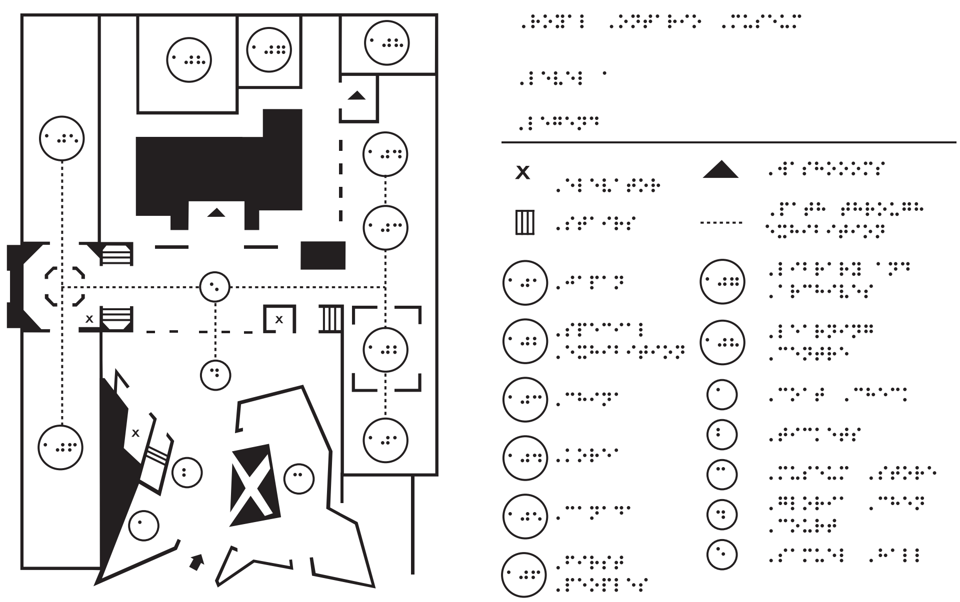

Figure 3.3: The relationship between representations and cognition behaviours has been the focus of much research on graphic displays, e.g., [182, 183]. The map shown here is a TM of the Royal Ontario Museum in Toronto, as proposed in 2010 by Gottschalk + Ash International, after months of research. This design is analogous to the one in SAM: the trajectories are reduced to straight curves and rooms to circular points. (I was granted permission to display this map in my thesis, but any further reproduction requires written consent from the design firm and the Royal Ontario Museum).

3.5.5 Data Presentation Sessions on OMS Participants

Data obtained from the low-fidelity sessions with blind participants provided many insights for the design of functionalities and human-computer interactions in SAM, but did not examine methods of data presentation for blind users. Thus, I lacked an effective framework to describe environmental concepts, and an effective order of presentation to address simultaneous conditions, such as a POI overlaying an intersection.

I examined these issues with NPO who has near three decades of experience describing spatial environments to blind persons. During the session, she was shown environments, like those in Figure 3.4 on page 50, and asked to describe how she would explain these concepts to her clients.

She responded that spatial concepts are difficult to explain without two-dimensional references, because speech is sequential. She uses the Picture Book as often as possible. She is often forced to borrow visual symbols to describe common intersections, 'this is a T intersection', 'this is a + intersection' [139, 5], and finds that unusual environmental conditions are difficult to explain; see [9]. She considers visual symbols to be inappropriate for CB persons and speech alone inadequate for spatial description. To describe the visual world, the visual stimulus must be put in a form compatible with the information provided in language. It is unclear how to translate the information derived from the visual system into speech information to a person who has never seen [184]. She reported observing the facial expressions of her clients to verify their understanding, something SAM cannot do.

She added that even order of presentation is difficult to determine, depending as it does on the situation and life experience of the client. The POIs she places on a map for her clients are important to them; presenting POI information first makes sense, but is not always the best strategy. A possible solution to address those issues could be letting the cartographer select the order of presentation at each POI, but this would increase the complexity of the building of a map. Another possibility could involve picking a consistent order–POIs first–and building concise messages that allow the remaining information (*i.e.,* intersection and curvature information) to follow quickly. Because consistency in data presentation is a good thing, and because reducing the difficulty for building spatial auditory maps is judged to be important to the adoption of the technology, the latter design option has been chosen. That presentation framework is described in Chapter 4.

3.6 Summary

The results of this chapter provide insights useful for the design of a new map technology and identifies aspects of TMs that should be included or omitted to shape information in ways that make sense to a blind user and help in building cognitive maps.

Section 3.2.4 describes how participants perceive finger displacement on a TM in environmental terms, indicating they understand TMs as a scaled down version of the world. Integrating gesture with real time sound feedback will probably enhance this impression and boost causal associations between the sound and the gesticulation in a user's mind, which most likely will create a virtual reality experience [185]. A digitizing tablet has a similar form factor and supports a level of interactions similar to TMs, like random access for example. Opting for a digitizing tablet complies with Gigante's [186] view that a virtual reality technology should be compliant with the real system's form-factor (TMs in occurrence). What's more, a digitizing tablet and stylus are inexpensive to purchase.

Admittedly, a stylus is more precise than a finger, allows several degrees of freedom to obtain rich input and can be used to engrave TMs in styrofoam or tactile drawing film layers, making it possible for a blind person to build his own tactile maps.

This chapter also establishes the importance of building simple maps. On one hand, OMSs stress that complex maps are hard and lengthy to build and on the other hand, blind participants indicate that details obscure the big picture. Paths and POIs are identified as the key features a blind reader needs to find to learn unknown space on a map. This format also benefits the cartographers because curves and points are quick & easy to draw.

Blind participants show a preference to natural mapping design, which is the motivation behind the design of the notification sounds I presented them. Experts mention how challenging it can be to translate the information derived from the visual system into speech information to a person who has never seen. Order of presentation is also viewed as difficult to determine. Chapter 4 revisits these points in detail and presents a solution derived from the information obtained here, visuospatial and psychoacoustic theory, and participants feedback.

These findings constitute a starting point for the implementation of SAM, at which point it becomes imperative to capture the lived experience of the user, which involves implementing a working prototype. 'It is very difficult to accurately anticipate the actual experience of the user; and the lack of knowledge concerning the final user is particularly drastic in the case of high-technology applications, where the potentialities of the technology are hugely superior to the acceptability of the final user' [187, p.35].

In conclusion, the sessions left many open questions, which were revisited in pilot studies with blind and sighted participants. The results are presented in Chapter 4.

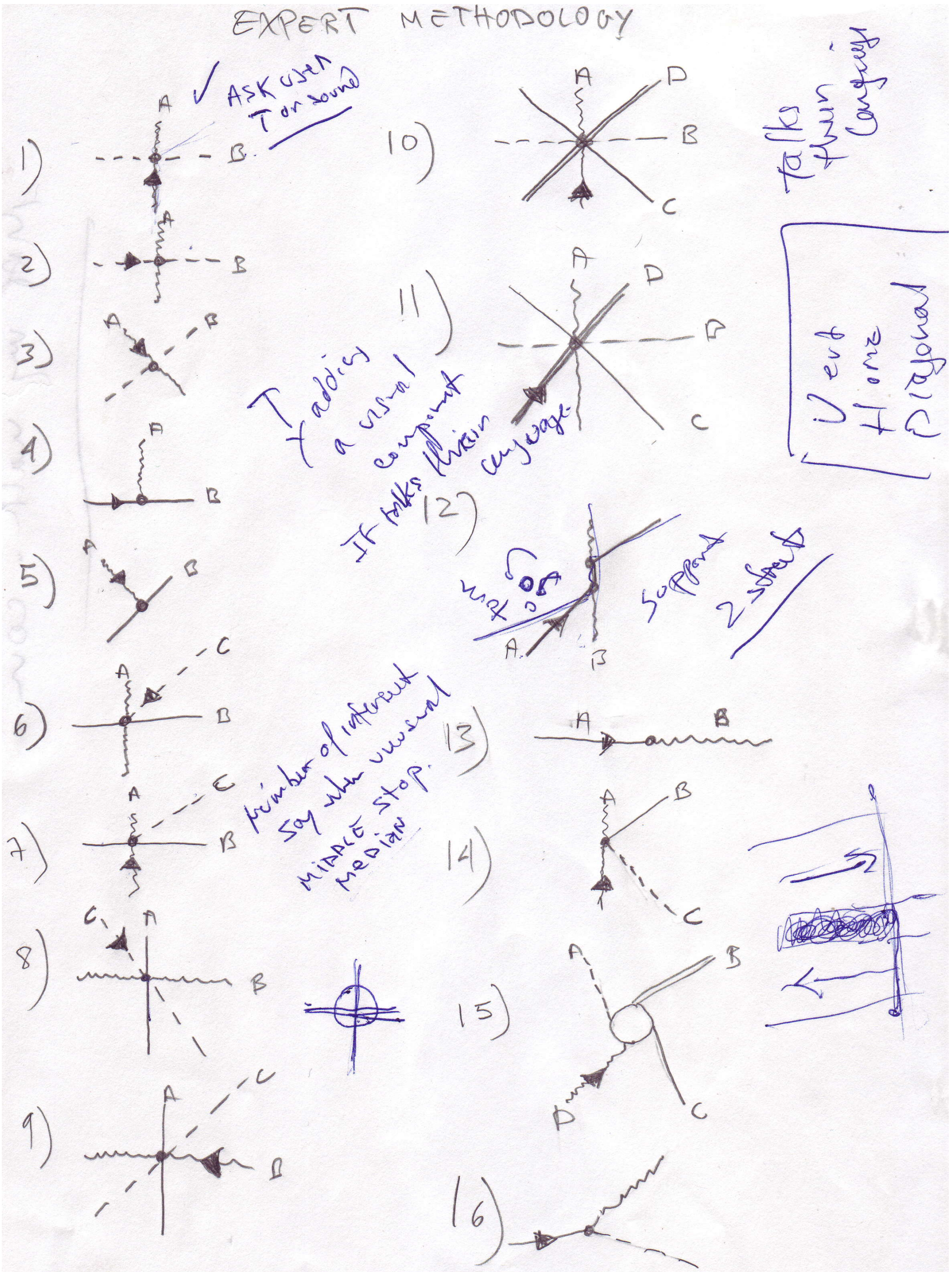

Figure 3.4: A sample of the material I used to determine how to explain complex intersections to a blind user.

Chapter 4

Sound Display

Flowers [188] points out that auditory display designers are influenced by visual graphic design practices, and fail to consider on how differently vision and audition operate. He blames designers for overoptimism with respect to auditory bandwidth in humans. Researchers tend to forget that the design of a sound display depends on the specific task to which it is to be applied.

Understanding of psychoacoustics is required for building effective auditory displays. Thus, this chapter begins with a discussion of psychoacoustics, then describes the individual components of the auditory display and how they are integrated. In effect this is a bottom-up approach, examining the sound objects and linking them into a presentation structure, which is in turn to user interactions in Chapters 5 and 6.

Sections 4.1, 4.2 and 4.3 discuss the psychoacoustic theory that influenced the design of SAM. Section 4.4 overviews the system, which is presented to help the reader link the individual parts to higher functionalities. Finally, Chapters 4.5 & 4.6 detail the design of the sound objects and the presentation framework.

4.1 Speech Sounds for Spatial Directions

Screen readers effectively transform written text into synthetic speech for blind users [189], but they have poor usability when the content goes beyond text, to illustrations and maps for example [190]. Sequential speech, which is one-dimensional, is not optimal for communicating higher dimensional information [191]. This section shows how we know that speech performs badly when providing spatial direction to a user.

Mowbray and Gebhard [192, 193] observed that it is difficult to attend to several simultaneous speech sounds because humans speech-processing capacity is limited, but two or more simultaneous non-speech sounds can easily be perceived [194].

Stokes *et al.* [195] observed that spoken messages generally take longer than non-speech sound to communicate information: as long as the computer talks, the user must wait through the entire message.

This problem is important when direction needs to be conveyed to a user. Angular or clock directions can shorten the message, but hearing repeatedly *turn 330 degrees* or *turn to 7 o'clock* raises aesthetic and cognitive issues. Speech directions are more irritating to listen to than non-speech ones [196, 197]. I also observed that the blind persons with whom I worked were slow at processing angular and clock directions. For these reasons, SAM uses non-speech sounds to guide a user in space.

4.2 Sound Perception

Characterizing human hearing is essential for representing auditory data. This section introduces concepts from the theory of human sound perception, which influenced the design of SAM. It includes sound dimensions, direction, separation, grouping and their interactions.

4.2.1 Sound Dimensions

Acoustic stimuli are described in term of pitch, loudness, duration and timbre. *Pitch* is the property of a sound that varies with variation in the frequency of vibration, *loudness* is the perceived intensity of a sound, *duration* is the time interval of a sound from start to end and *timbre* is the character or quality of a sound. Further details about these dimensions follow.

Pitch

Humans scale pitch logarithmically to frequency. In the Western music system pitch variations are ordered on a musical scale composed of 96 half-step pitch intervals perceived as equally distant one from another. Almost all humans hear pitch increasing or decreasing relatively, not with reference to an external standard [198].

Pitch intervals can be arranged into a *diatonic scale*, a sequence of seven notes comprising five whole steps and two half steps, which can give a sense of tonality to a listener when played successively. *Tonality* is a system of music characterized by specific hierarchical pitch relationships based on a tonic. For example, the pitches of the white keys on a piano form a diatonic scale. Playing the white keys from C3 to C4 immerses a listener in the tonality of C major, establishing C as the tonic (or key). Westerners - both musicians and non-musicians - easily detect the tonic in tonal musical segments [199].

Pitch encoding has been used successfully by many technologies designed for blind people, as illustrated in Chapter 2. SAM makes extensive use of pitch and tonality.

Loudness

Loudness (L) and intensity (I) are related by the Power Law [200], that is, the perceived loudness of a sound is doubled at a 10-dB intensity increase. Sound *intensity* is the amount of energy transmitted by sound and is expressed in decibels. Although humans can perceive a wide range of loudnesses, absolute loudness judgments are difficult and relative judgments are limited to a scale of four or five different levels [201]. Judgments are also affected by dynamic range limitations in sound equipment. As a result, loudness variation is often ineffective in auditory displays [202]. Furthermore, the use of loudness variations to convey data often annoys users [203].

Duration

Duration is the time duration of a sound. Long continuous sounds are difficult to localize [204] and fade into the background of consciousness after a short period of time [205].

Sequential sounds of different durations creates rhythmic structures and the succession of sounds of equal durations creates *beats*. The human auditory system is highly sensitive to temporal aspects of sound [206, 67]; response time for auditory stimuli is faster than for visual stimuli [67, 207, 208]. The sounds are kept short in SAM and information is often encoded in temporal repetitions of sounds.

Timbre

Timbre is the sound quality or the colour that distinguishes one sound from another, for example, one musical instrument from another or a human voice from another. Although the salient dimensions of timbre are poorly understood [209, 210], timbre is one of the most immediate and easily recognizable characteristics of sound [211].

Envelope

The envelope of a sound is its overall shape in intensity over time, an important physical characteristic that mediates the perception of timbre. Attack, decay, sustain and release are the four basic stages of the envelope. *Attack* is the time taken for initial run-up of level from 0 dB to peak. *Decay* is subsequent run down time interval from the attack peak to the specified sustain level. *Sustain* is the constant intensity of the sound prior release. *Release* time is the time taken for the sound to decay from the sustain level to 0 dB.

Much effort was put forth into the design and processing of aesthetic, rich and distinguishable sounds in SAM.

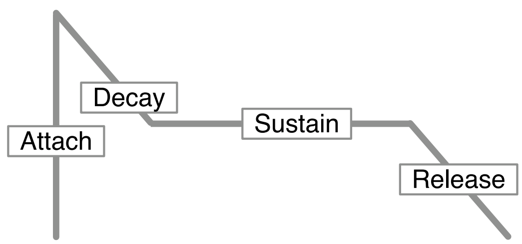

Figure 4.1: The four basic envelop parameters of a sound: attack, decay, sustain, release. The x-axis is time and the y-axis intensity.

.

4.2.2 Sound Directionality

Directionality of a sound refers to the spatial origin of the sound with respect to the listener. The ability of humans to localize directionality of sound sources is poor. We are best at judging azimuth, poor at judging elevation, and worst at judging distance. This discrepancy occurs because different auditory mechanisms are used for each dimension [212].

Azimuth & Elevation

The auditory system uses interaural differences to localize sounds in azimuth, specifically differences in time and intensity. Sounds from the side reach the ears at different times and with different intensities [213]. Although humans can hear sounds coming from all around them, lateral hearing resolution varies. For example, Strybel *et al.* [214] reported resolutions of 1° with median sound sources (facing the listener) and 40° with sounds 90° in azimuth (from the side). Hence, sounds on either side must move a greater angular distance than sound in front to be perceived as moving by a listener.

Perceived elevation of a sound source is a monaural experience. The cavities and bumps in the pinnae affect the audio spectrum of sound sources. It is possible to measure the transformation that maps spectrum at the sound source to the spectrum to the eardrum, and synthesize filters called head-related transfer functions (HRTF) to reproduce it. A computer can then use the HRTF to position sound in space. Sounds that have richer acoustic timbral characteristics are easier to perceive than sounds composed of a pure tone because filtering wide band sounds is more effective than filtering sounds that only have fundamental frequency characteristics [204].

Unfortunately, HRTF are both expensive to measure and listener specific. Playing a sound source through HRTF not customized for the listener, like generic or best-fit HRTF [215, 216], is rarely successful. For example, listeners may confuse front and back or up and down [204], may hear the sound origin as inside the head, not outside [204], are slow at making localization judgments [131]. In addition, the acuity of audition is much lower in elevation than azimuth, 4° in an anechoic laboratory [217], but between 15° and 30° in the environment [218, 219]. For example, a sound object in the environment can move as much as 30° in elevation without being perceived as moving. Hence, vertical positioning is difficult and not dependable. SAM resolves the problem by limiting spatial sounds to the transversal plane passing through the head at the level of the ears. Elevation is encoded with pitch, using the Pratt effect.

Pratt hypothesized a pitch-elevation correspondence in 1930 [220], after observing that successions of tones in a musical phrase can induce a sensation of apparent movement. The *Pratt effect* originates in associations, developed early in life, between elevation and pitch: short wavelength sound being described as high, long wavelength sound as low. A large body of empirical research examined the Pratt effect and corroborates the pitch-elevation correspondence [221].

Distance

Humans' ability to assess distance from sound has been extensively measured. Three types of cues are found to be useful to a listener: intensity, spectral filtering of sound with distance and the ratio of direct to reverberant sound.

Intensity. Listeners judge less intense sources to be further away than more intense ones [222, 223]. Intensity is the primary auditory cue used to judge the distance of sounds in the environment [224, 222].

Spectral Filtering. Distance judgments are affected by the frequency spectrum of a sound source, which is attenuated at high frequencies by about 3 to 4 dB per 100 meters at 4 kHz [225, 226].

Reverberant Sound. Sound can take many paths from the source to the ear; the fraction of direct sound is stronger for near sources [227, 228].

Familiarity. This characteristic is an observation, not a cue. Distance is better perceived when the sound is familiar to the listener. Knowing the nature of a sound source and its spectrum at different intensities helps a listener in judging its distance [229, 230, 231].

Because reverberation and spectral filtering are secondary cues and are computationally intensive, SAM uses intensity to convey distance to users. However, intensity variations, loudness variations as perceived by humans, have limited dynamic range. A recent study of methods for communicating distance to blind travellers using non-ecological sound cues showed a consistent association between distance and beat-rate [103], with proximal sources represented by fast beat rates and distal sources by slow beat rates. The combination of coarse loudness variations and beat-rates increases distance resolution in SAM.

Echos

Echolocation is another binaural cue used by blind pedestrians to hear sound reflecting on the surface of obstacles around them. Blind travellers listen to a room's echoes, from tongue clicks or cane taps, which help them perceive nearby surfaces. They hear subtle timbral changes caused by small time delays [232], by which they stay centred in corridors or hear lateral landmarks along the way.

4.2.3 Sound Streaming

The auditory system naturally groups related sound events into *streams*, segregating unrelated events from one another. This ability is demonstrated by the cocktail party effect, where an individual follows one voice ignoring others. The *segregation* mechanisms separate different voices while the *grouping* mechanism integrates words into sentences.

Segregation and grouping are influenced by timbre [233, 234], pitch [235], temporal proximity [206], spatial location [236], rhythm regularity [237] and prior familiarity with the pattern of sounds [206]. For example, successive footsteps, which are similar in timbre, group perceptually into a single stream. Their regular rhythm, continuous location, comparable pitch, and temporal proximity reinforce the grouping.

Grouping mechanisms in streaming, which are governed by the Gestalt principles of perceptual organization [206], are fundamental in the design of effective auditory displays. Sound elements must be sufficiently separable to avoid becoming a sound soup, and sound elements that belong to the same group must naturally stream together. For example, listeners can measure timing more precisely between sound events that belong to the same stream than between events in separate streams [206]. Also, listeners perceive a melodic motif with much greater ease when its notes are grouped into one stream [206].

SAM provides simultaneous sound events to a user, grouped according to tasks faced by the user. Existing theory provides solid guidelines for the designer, but providing experience requires musical skill that goes beyond the guidelines. Theory tells the designer the principles that govern grouping and separation, thus knowledge about ecological sound perception is used to guide the selection of non-ecological sound cues.

4.2.4 Interaction of Sound Dimensions

An auditory display designer must deal with a myriad of interactions that limit the design space. This section examines the interaction of pitch, loudness, duration, timbre, envelope and beat rate on perception, all critical to the design of SAM. Additional details are available here [206, 204, 238, 200, 239, 240].

Pitch and loudness interaction. Changing the pitch of a sound often affects its perceived loudness and vice versa. Intensity increase at less than 2 kHz slightly raises the perceived pitch, while intensity increase above 3 kHz does the opposite. Also, stimuli of the same intensity at frequencies below 1 kHz and above 5 kHz are perceived as quieter than stimuli inside that range [204]. SAM uses a lookup table to adjust intensity levels according to pitch change values.

Perception of pitch and loudness depend on duration. The greater the duration of a sound, the better humans are at perceiving its pitch. The *musician's uncertainty principle* explains that pitch differences between two sounds are perceivable when the product of the difference between the two sound frequencies, measured in steps of the scale (Hz), and the duration, in milliseconds, is greater than one [241]. Similarly, loudness judgments are affected by the length of a sound; sounds shorter than 200 ms are perceived as softer. Because sound duration is not used by SAM to encode information relevant to the user, the duration of the stimuli was adjusted so that participants could perceive quarter-tone intervals or greater. Intensity of short sounds was pre-adjusted in ProTools™, a digital audio workstation platform.

Onset of a sound affects its timbre and its localization. The onset of a sound changes its perceived timbre [242] and a human's ability to localize it in space [204], because a short attack creates click-transients with a complex spread of high frequencies, even for sounds with few high frequencies in their body. The human auditory system relies on this short duration of high frequencies to localize sounds in elevation and in azimuth [204]. Consequently, repeated short sounds with strong attack are preferable to lengthy continuous ones.

Beat rate perception depends on training. Perception of beat rate in auditory sequences depends on factors like musical training, tempo change direction, and musical versus isochronous sequences [103]. SAM uses four levels of isochronous beat rate variations (75, 90, 105 and 120 beats per minute), which were adjusted and validated by short pilot studies on naive participants, which ensures that the difference between beat-rates is easily distinguishable.

The pilot studies followed the same procedure. The slowest and fastest beat rates were predetermined by the researcher (author of this thesis) and remained the same throughout the pilots (75 and 120 beats per minute). Slower than 75 beats per minutes required a listener to wait for too long to make a judgment. Faster that 120 beats per minute raised aesthetic issues. The intermediate rates were determined by concatenating them into a sequence built by increasing and decreasing speed: every sequence had four beat rates join together without silence–except for the silence between the beats. The sequences were played at once and the participants were asked to tell when they perceived a rate change. A total of six variations were tested, 90 and 105 were the two intermediate beat rates that were best perceived.

Absolute Information by Tonality

Vision presents many things simultaneously, some of which provide references, like the ground plane, that make absolute judgments possible. Because the perceiver directs the order in which objects are perceived, looking back and forth makes precise judgments possible. The same is not possible in audition because perception is fleeting, the order is controlled by the creator of the sound display, and memory must be used for comparison [243]. The dimensions of sound either offer dynamic range too compressed for conveying a practical range of values, such as loudness and sound directionality, or are not perceived absolutely by the listener, such as pitch. Pitch is an effective cue for non-categorical information in auditory displays. For example, pitch is used to sonify quantitative data like time durations, length measurements and even amounts of money [111].

SAM provides a tonal interface, a novel idea which allows the user to identify a certain pitch value, the tonic, as a reference at the centre of a tonal scale. A strong tonality is established by governing every nonspeech sound by tonal laws: all non-speech sounds in SAM are in G major, a tender and joyful tonality [244]. Uniform tonality makes it possible for a user to recognize centres of dimensions, which can be applied, for example, to the stylus sitting on a curve. Tonality is the key to constructing a cohesive and aesthetic display.

4.3 Classes of Auditory Displays Used in SAM

SAM implements three types of auditory display in one system: auditory icons, spatial sonification and speech. This section explains the rationale for choosing these modalities.

4.3.1 Auditory Icons

Data from the field study shows that participants preferred not to learn the meaning of arbitrary symbols. This information has repercussions for the acceptability of previous systems: requiring extensive learning entails rejection by blind users [245]. For this, SAM avoids *earcons*, which are structured musical patterns used to represent specific items, because they lack intuitive connections to the objects they symbolize [246] and *spearcons*, which consists in accelerated speaking of the information [247]. Spearcons are easier to learn than earcons but are hard to segregate when superposed, a characteristic SAM requires.

Instead, short related sounds having a semantic connection with the event or process they represent, known as auditory icons, were chosen. *Auditory icons* are easier to remember [248] because of a user's association with the objects they represent. For example the sound of a car honk can be used to represent a car. However, sounds and events are not always naturally linkable, which can create ambiguity. What is the sound of a drugstore? The stronger the semantic connection, the easier auditory icons are to learn and remember [249].

Maynatt [250] has undertaken a formal study of best practice for auditory icons, which provided guidelines for SAM. He proposed the following heuristics: use short sounds of wide bandwidth with roughly equal length, intensity, and sound quality; test that auditory icons can be identified using free-form answers; measure the learning required by auditory cues that are not readily identified; test possible conceptual mappings for the auditory cues using a repeated-measures design in which the independent variable is the concept the cue will represent; evaluate sets of auditory icons for malignant interactions including masking, discriminability, and conflicting mappings; conduct usability experiments with interfaces using the auditory icons.

In SAM, notification sounds are played to signal that a landmark is near. Like alarms, they provide one bit indicating an event starts or finishes, telling the user that action may be required [251]. SAM implements four notification sounds at points of interests (POIs), curve intersections, tight curvatures on curves, and extremities of curves. In addition to the notification sounds, SAM implements auditory icons that structure information when successive sounds are presented or to indicate the current state. When natural associations were unavailable, SAM relies on short combinations of words, or tokens, which are spoken to the user. All auditory icons and speech tokens were evaluated in pilot studies with blind participants.

The procedure followed by the pilot studies is described in this paragraph. A set of sounds was prepared by the researcher and played to a participant, one at a time. Every sound assessment followed the same course of action: a sound was played to a participant through a pair of headphones and then, after listening to it, the participant was asked to determine its natural association, tell if he judged it pleasant, provide any comment and propose sonic ideas he might have from listening to it. A free-form answer format was used.

4.3.2 Spatial Sonification

SAM implements four families of directional sounds: curve tracing, locatone, vocoder-integer and pulse-presence cues. These directional sounds are not auditory icons because they sonify spatial information. These are perceptually different from one another but have pitch and stereo balance information in common, like existing systems [82, 90, 91, 92, 96], and have a fast attack for superior localization.

4.3.3 Speech

Like several existing systems [68, 69, 70, 75, 86, 88, 111] and the electronic orientation aid systems described in Chapter 2, SAM synthetically speaks the names of streets (curves) and landmarks (POIs), or any annotations entered by the cartographer.

4.4 Design Overview

We now examine the design of SAM in greater depth. Figure 4.2 shows the overall organization of the system. The tree shows the complete set of tools implemented by SAM and their interdependencies.

SAM's reader interface is organized into two sets of tools: the *foveal tools* with which a user traces a curve and discovers POIs along it, and the *peripheral tools* with which a user perceives features far from the current location on the map. Tools in the middle row of Figure 4.2 are independent and can be activated at any time by the user. Tools in the lower row enhance the tools above them, but offer different affordances, depending on the tool it enhances. For example, details-on-demand (DoD) for the Curve Tracer are different than DoD for the Front Tool. The sound objects used by these tools are common to all, but matched differently to build sounds that encode different information.

The syntax of the encoding is examined in Section 4.5, the semantics in Section 4.6.

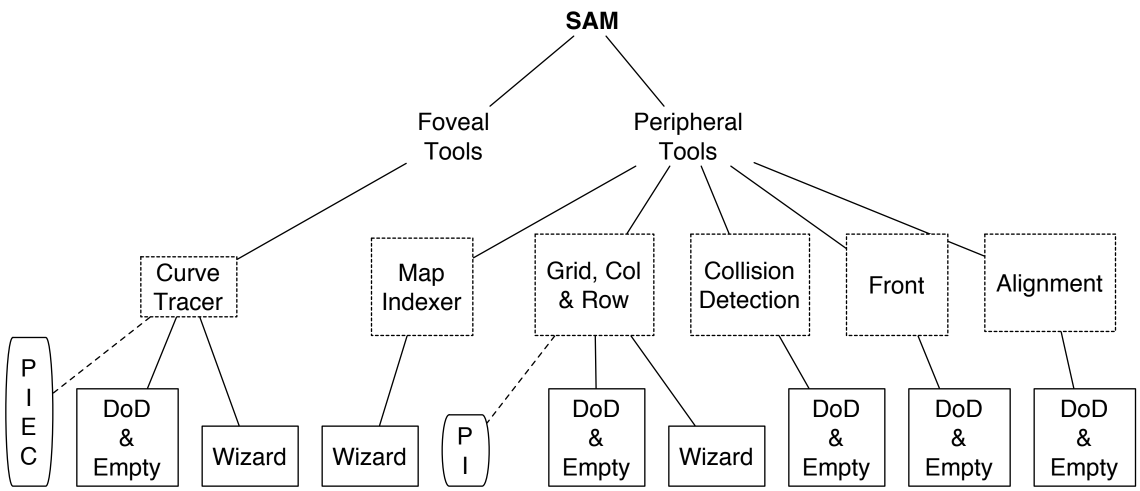

Figure 4.2: This figure shows a high-level representation of a User Environment Design diagram (EUD) of SAM. The tools in the lower row are multi-mode tools (MMT) as they can be activated by more than one tool. (DoD stands for Details on demand and Empty for empty tool.) These tools share a rectangle to save space, but are in fact independent to one another. Notifications tools are denoted by the rounded rectangles that groups them: P stands for POI, I for Intersection, E for Extremity and C for Curvature. More information about MMT is available in Section 6.2.2

.

4.5 Sound Design

This section provides a description of the sounds given in the chapter. Five characteristics serve to support the author's judgements in the creation of these sound:

1. general considerations from psychoacoustic studies,
2. prior research proving that sounds like the ones used by SAM are effective,
3. intuitive descriptions of the sounds,
4. sound samples the reader can listen to and
5. evidence from participants, formal and informal.

The first half of this chapter and Chapter 2 address the two first points. The other points are examined in the remainder of this chapter and in Chapter 7, which provides formal evidence that the holistic system works. All the sound waves are available online for listening[1].

4.5.1 Design Process

Natural sounds are complex, stochastic and rich in idiosyncratic details. SAM's sound icons are generic without details that are unnecessary for recognition. This design principle enhances comprehension and audibility [252].

There are many methods for synthesizing sound with computers [203]. SAM's auditory icons are the result of sample-based synthesis, a pragmatic approach for building sounds. It starts with recording of real sounds, either environmental or synthetic. They are processed by a series of sound engineering techniques, like truncating, pasting, frequency modulating, filtering, compressing, phase shifting, and so on. The construction process is complex and differs icon to icon. Its detail, which is lengthy and of little scientific value is omitted. All sound icons used by SAM are available online[2]. Further details about these techniques, which have been used by sound recording engineers for decades, are generally available [253].

The sounds are built from pre-recorded sources: the intersection notification sound is a recording of cuckoo song available on the Freesound Project web site [254]; the curvature and extremity notification sounds are from the Digidesign SampleCell™ sound library; the other sounds originate in Digidesign xPand™.

4.5.2 Sound Description

This section describes the unique characteristics of every sound in SAM. Visual representations of the sounds are available in Figure 4.3 and Figure 4.4.

CBM and CBF from Chapter 5 (see their description in Section 5.2.2) agreed to participate in a short pilot that assessed a set of new auditory icons. The pilot, which followed the procedure explained in Section 4.3.1, also re-examined the notification sounds described in Section 3.2.1 against new icons I created. The rationale for a repeated-measure was to verify that the semantic connections in the original sounds (described in Section 3.5.2) were most favourable, as recommended by Maynatt [250]. Note that the design of some of the original notification sounds was improved for the pilot. For example, I appended a new note to the POI notification sound to enhance tonality perception. Details are found below. Because I met with NPO on a regular basis, I could present her with sound ideas.

[1] http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/SoundMapping.html
[2] http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/SoundMapping.html

Although she did not take part of the pilot, her comments are provided in the text when available.

Notification Sounds

Notification icons are like radio emitters: the nearer the user is to a notification location, the louder the signal; outside the range of an emitter the notification sound is silent. Ideally, sound notifications are immediately recognizable but may be repeated a few times when the user slowly approaches a notification area. All notification icons share this characteristic and support extended replays. There is however a trade-off between functionality and aesthetics with short sounds: while icons with long developments can make beautiful loops, they take longer to be recognized.

In addition, the number of sound notification layers is variable. Scenarios E, H, I, J, and K in figure 4.7 on page 71 show occurrences of superposed sounds where POI, intersection, curvature and extremity notifications are overlaid, causing segregation and streaming challenges in terms of design. Short pilot studies on sighted and blind participants, which consisted in having participants test an early version of SAM, were conducted to verify that a user could segregate stream combinations of notification sounds, while being superposed by curve tracing sounds in the foreground. Aesthetics was also a primary concern. The rest of this section describes the notification sounds individually.

The *POI notification* is a four-note tone played sequentially, as seen in the top left graphs of Figure 4.3, the arpeggio of a G major chord (G3, B3, D3 and G4). Similar sounds are often heard in public transportation areas like airports to alert listeners before an announcement. A POI notification plays a similar role for a blind user. The blind participants and the orientation and mobility specialist (OMS) preferred this sound over the sound of a doorbell, and the sound of an arrow hitting a target. The doorbell was rejected because it confused the participants thinking someone was at their door step. The arrow hit was judged irritating.

The *intersection notification* is the sound of a traffic light cuckoo. The sound plays two consecutive major thirds downward from B3 to G3. The sound of a cuckoo is known to blind travellers; it is a common sound at traffic light intersection indicating which crossing is currently safe. The blind participants were presented the muffled metallic sound of an electrical switch, but felt the cuckoo was more natural to them. One pointed out, however, that the metallic sound reminded him passing over something, like driving over a train track with a car.

The *curvature notification* is a rotary[3] sound, which semantically suggests some circular movement, or turning. The sound rotates in phase, which gives it a unique timbral colour,

[3] A sound that perceptually moves in space on a closed orbit.

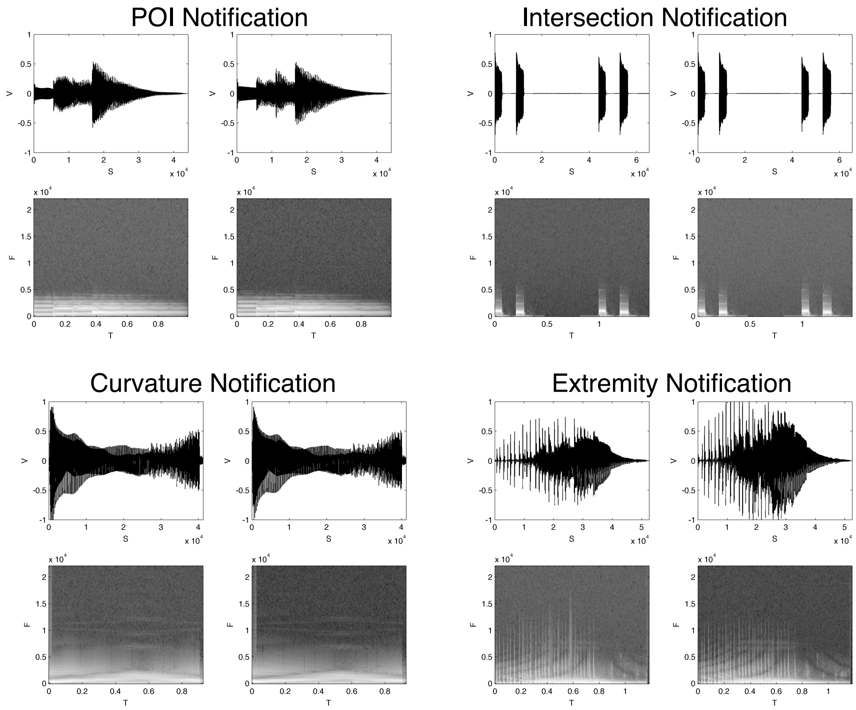

Figure 4.3: The four notification sounds. The upper graph shows the envelope of the sound (S = Samples, V = Voltage). The dark region corresponds to sound intensity, a vertically wide region is louder than as narrow one. The lower graph shows the spectrogram of the sound (T = Time [seconds] and F = Frequency [Hz]). Spectrograms can be interpreted as a measure of the energy of the signal contained in the time-frequency domain (spectrum), brightness maps to energy. The sounds on the left side map the left ear and the sounds on the right side map the right ear.

and alters its pitch producing a minor seventh from G3 to F4, an interval used by composers to lead to a new tonality, foreshadowing a change [255]. Similar sounds with stereo balance modulations were rejected because they suggested explicit turning directions (either left or right) to the listeners.

The *extremity notification* is the sound of friction on a rubber surface that suggests stopping to a blind listener, or 'game over' when playing a computer audio game. The granular sound glides from G4 to G3. The sound of a car braking abruptly was also found effective but rejected because of its association with traffic accidents.

Metadata Nonspeech

Another set of auditory icons structures data presentation, personifies map features and tells the user the current state of the system. Again, the envelope and spectrogram of the sound cues discussed in this section can be found online at the address provided above.

The *emptiness* sounds tells a user that SAM is functional, and there is nothing underneath the stylus on the map. Emptiness cues also communicate stylus state information to the user. The timbre of the sound, which is perceived as a rich, bassy percussive sound, is transformed to communicate stylus states to the user, that is, the sound is altered using filters and intensity according to the stylus tip in use (eraser = low-pass filter, sharp = high-pass filter), stylus elevation (down = high intensity, hovering = low intensity) and stylus inclination (tilted = no filter). The cues are repeated as the stylus is moved by the user. Sounds of the wind were suggested but eliminated because of their continuous characteristic, which had the effect of filling the display even when there was nothing underneath the stylus. The bassy percussive sound on the other hand suggested nothing to the participants, but was bandwidth efficient and easily distinguishable from all the other sounds in SAM.

The *beginning- and end-of-list* cues structure the information when a list of features is spoken to the user: a beginning-of-list icon informs the user that a list is about to be itemized and an end-of-list icon indicates closure. The open- and close-list sound cues are 12-step portamenti up (G3 to G4) and down (G4 to G3), respectively. The participants found the semantic connection of the sounds with their meaning highly intuitive, specially since they are always used in a specific context.

The *item separator* icon is a very short sound that structures list information. This high pitched sound (G5) is between every feature in a list. The sound only needed to be distinctive and smooth to the listener. It was instantly perceived as a 'comma' by the participants when presented in its context.

The *right on* icon is a soothing G major chord played by an 'ah' choir synthetic pad, which evokes the first chord of the Hallelujah Chorus in Handel's Messiah; the same voicing is used. The sound is triggered in the background whenever the user lands perfectly on a curve during tracing. This icon is designed to reinforce the G major tonality in SAM and to confirm the user's position on the curve. Other timbres of the same chord were also suggested, such as a Hammond organ, a jazz guitar and string pizzicatos, but none of them were judge as soothing as the 'ah' choir sound, a prime design goal for this cue.

The *grid frontier* icon is the metallic sound (G5) of passing over a train track. It signals a user passing over the boundary of a cell, column or row when using the grid, column or row tool, respectively. The electrical switch sound from the intersection notification is reused.

Spatial Sonification

Curve tracing cues. These sounds carry the information a user needs to acquire when tracing a curve. The cues are built into two interchangeable sound sets that differ in their timbral signatures, one with a woody sound, the other with a glassy sound. Timbre toggles every time a new curve (street) is touched, which allows a user to detect the change.

Locatones. These cues are used by the wizard tool to walk a user to any given location in the absence of curve tracing. They are also used by the DoD tools to provide direction cues at notification areas. Their timbral colour is similar but richer in frequencies to the sound of a square wave. SAM implements two families of locatones, one to be perceived as distant, the other as nearby. A full description of line tracing and locatones cues is given in Section 5.5.2.

Vocoder-integers. These tokens are spoken numbers treated by a vocoder effect, which gives speech information a tonal quality (G3 in pitch) and a richer spectral composition for superior sound localization [256]. These numbers are assigned to map features as an alternative to speaking their names when a fast overview is needed. The tool that uses them relies on users' ability to recognize them, not memorize them, which addresses the potential problem of memory load, as explained further in Section 4.6.4.

Pulse-presence. This sound has a short pulsing transient at pitch G3 that, like the vocoder-integers, personifies features for fast overview. Like the item separator, the sound needed to be distinctive and smooth, but also long enough for a listener to hear its pitch. The context in which the sound is used made the sound/feature association unambiguous to the participants.

Speech

SAM uses two different male voices, chosen for their intelligibility; the available female voices were difficult to comprehend, a shortcoming of the speech synthesizers.

One voice speaks the names and annotations of features on a map, as entered by the cartographer such as street and POI names and their annotation; the other speaks the metadata information, a small set of predetermined tokens explained in this section. The former is done by the FreeTTS voice synthesizer, an open source synthesizer [257], and the latter by ALEX, a Mac OS X Leopard™ synthetic voice [258]. The spectrograms and envelopes of the word 'intersecting', spoken by the two voices, is shown in Figure 4.5. The remainder of this chapter explains words spoken by ALEX.

Tokens associated with the foveal tools. The tokens 'intersecting' and 'is continuing' tell the user that the intersection tool is in use and that a curve continues beyond the intersection. The token 'and' concatenates map features together. The tokens 'location',

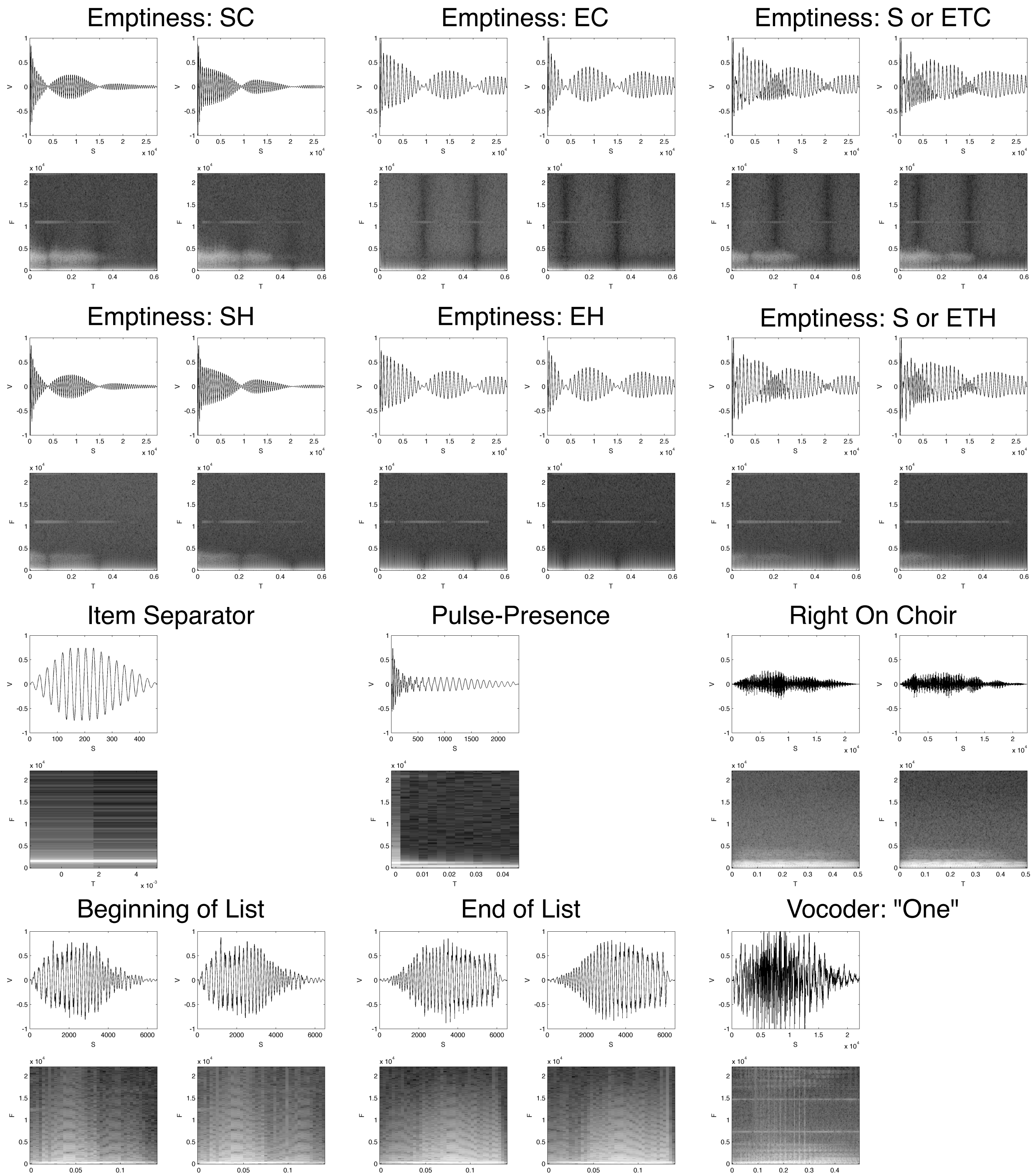

Figure 4.4: Miscellaneous Sounds. Acronyms: S = sharp tip, E = eraser tip, T = tilted stylus, C = in contact with tablet, H = hoover tablet. The item separator, pulse-presence and vocoder sounds are mono.

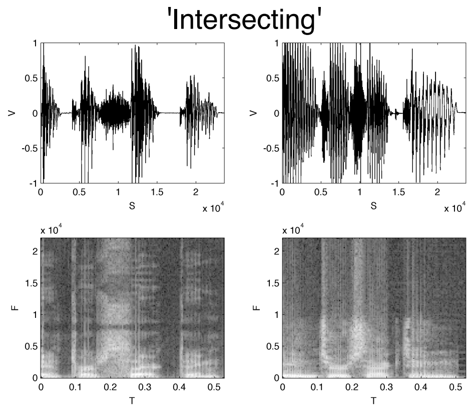

Figure 4.5: The word 'intersecting' spoken by the FreeTTS voice synthesizer on the left, and by ALEX on the right.

'is turning' and 'extremity of' inform the user that the POI, curvature and extremity tools, respectively, are in use.

Tokens associated with the peripheral tools. The tokens 'left col', 'mid col', 'right col', 'top row', 'mid row', 'base row', 'cell one', 'cell two', 'cell three', and so on up to 'cell nine' are associated with the grid tool family. They tell the user the map area in which the stylus is. The token 'wizard tool, walking towards' informs the user that the wizard is in use. The token 'no keyboard mapping' tells the user that this key on the keyboard is not associated with anything. Finally, the tokens 'grid tool', 'col tool', 'row tool' 'front tool' and 'alignment tool' inform the user that the grid, col, row, front or alignment tools are in use.

4.6 Data Presentation

SAM preserves auditory bandwidth by minimizing ongoing information, while allowing a user to obtain rich details on demand. Data presentation for these two levels of granularity is governed by a rule-based presentation protocol.

The presentation framework is designed to speak a language that respects the culture of its users. SAM eliminates the visual imagery typical of spatial descriptions by building sentences consisting of speech tokens, auditory icons and spatial sounds. The result uses bandwidth efficiently: it displays rich and structured information at a rate that is comfortable for a user.

A per-tool approach is adopted to describe the presentation protocol because of the fine distinctions found in every tool. The actions that need to be performed by a user to invoke these tools are described in Chapter 6.

4.6.1 Foveal Tools

While tracing a curve a user is notified at curve intersections, high curvatures, extremities and POIs. At such notification areas, the user can obtain details on demand. The flow chart in Figure 4.6 describes how the protocol orders presentation at notification areas, including when several notification types occur together. In addition, Figure 4.7 gives concrete examples of the results. The reader can also hear a sound example of a sonified intersection[4].

Details on demand are also available from the curve tracing tool. On request, SAM speaks the label of the curve being traced, per the cartographer's annotations.

4.6.2 Wizard Tool

The wizard tool uses locatones to walk the user to a destination. Details are displayed as a destination is reached: the locatones are muted and the name of the feature is spoken.

4.6.3 Map Index Tool

The computer keyboard can identify features on a map; the character keys are mapped one-to-one with the features on a map. Full details about a feature are obtained by a keystroke: the system either speaks the feature's annotation or informs the user that no keyboard mapping exists. The user can then call up the wizard to reach a selected feature.

4.6.4 Grid, Column & Row Tools

This family of tools allows a user to find POI and intersection areas on the map without tracing curves or using the map index. Intersection and POI areas are important as they are focal points with respect to which blind travellers make decisions about bearing.

A map is divided into a grid of nine cells labelled using the touch-tone layout convention. The divisions are consistently symmetrical and equal sized, as shown in Figure 6.5 on page 112.

[4]http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/Interactions.html

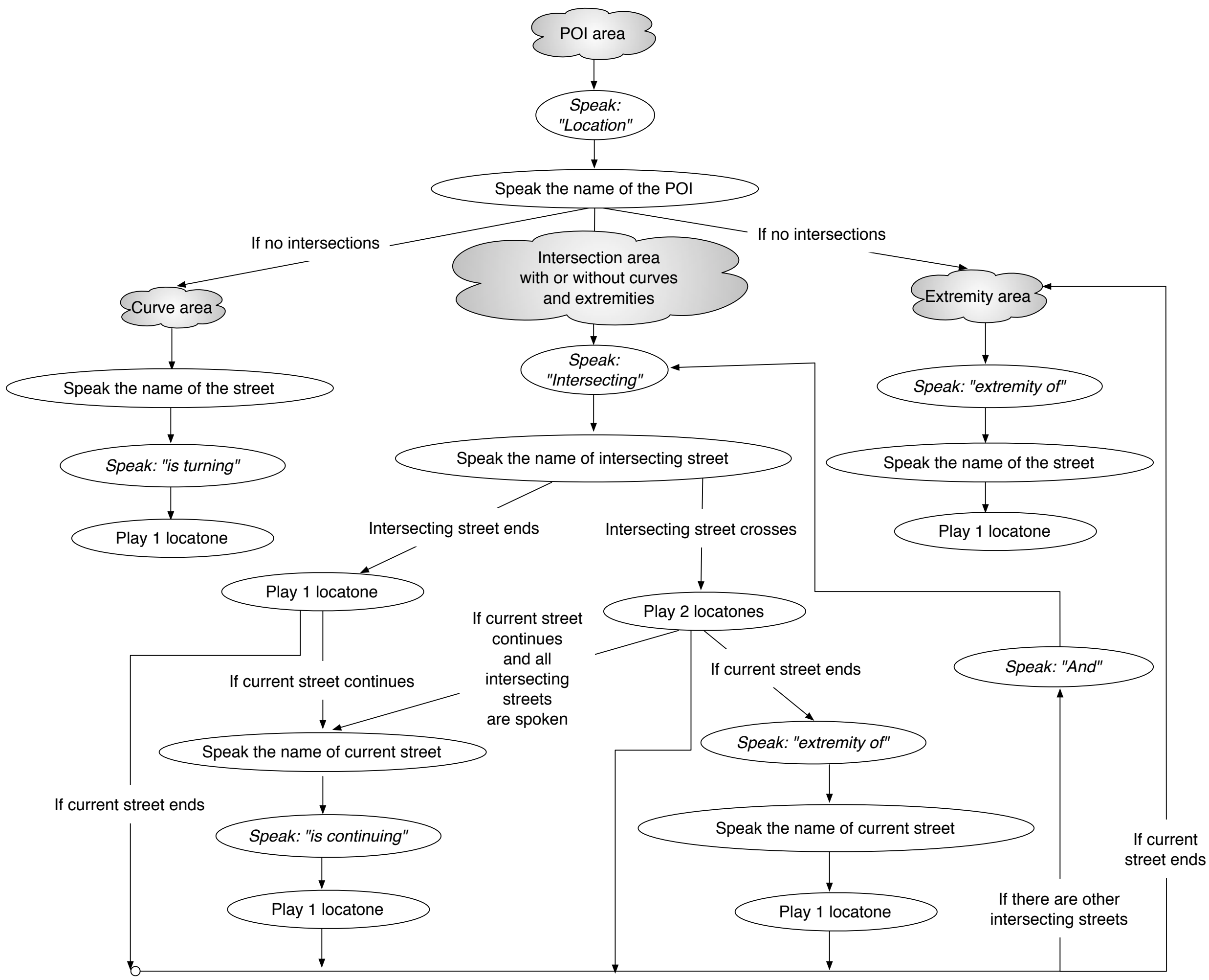

Figure 4.6: This figure shows the presentation protocol for the foveal tools, upon a detail-on-demand request by a user. The clouds represent notification areas. For example, at an area that overlays an intersection and a POI, SAM would begin by speaking the word 'location' then speak the POI's annotation, then it would speak 'intersecting' followed by the name of the intersecting curve, play the appropriate locatone, and continue branching down according to the context: SAM knows where the user is on the map, which curve he is tracing and the tracing direction.

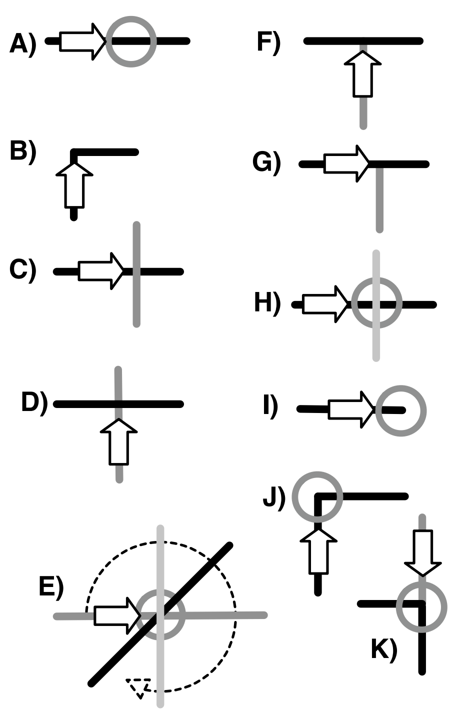

Figure 4.7: These are details-on-demand scenarios for the notification tools. The empty arrow shows the arriving direction of the user. The tip of the empty arrow represents the current position of the stylus. *Caption details*: the cardinal directions (N, S, W, E) are used to indicate the directional cues conveyed by the locatones. The diagonal directions are indicated by two letters, e.g., NE for north-east. The words in italic are spoken by ALEX. The IDs are curve and landmark annotations spoken by FreeTTS. The concatenation symbol '+' chains everything together into a phrase. In **scenario A** the phrase would be constructed as follows: *Location* + POI ID; **scenario B:** current curve ID + *is turning* + Locatone E; **scenario C:** *intersecting* + vertical curve ID + locatone N + locatone S + horizontal curve ID + *is continuing* + locatone E; **scenario D:** *intersecting* + horizontal curve ID + locatone E + locatone W + vertical curve ID + *is continuing* + locatone N; **scenario E:** *intersecting* + vertical curve ID + locatone N + locatone S + *and* + *intersecting* + diagonal curve ID + locatone NE + locatone SW + current curve ID + *is continuing* + locatone E. The dashed arrow indicates the consistent direction of presentation when three or four curves intersect; **scenario F:** *intersecting* + horizontal curve ID + locatone E + locatone W + *extremity of* + vertical curve ID + locatone S; **scenario G:** *intersecting* + vertical curve ID + locatone S + vertical curve ID + *is continuing* + locatone E; **scenario H:** scenario A + scenario C; **scenario I:** *Location* + POI ID + *Extremity of* + horizontal curve ID + Locatone W; **scenario J:** scenario A + scenario B; **scenario K:** scenario A + *intersecting* + square angled curve ID + locatone S + locatone W + *Extremity of* + vertical curve ID + locatone N.

The intersections or POIs in the region containing the stylus make their sounds, the rest are mute. Loudness, tonal pitch and stereo balance convey the location of the features. The individual tools are discussed below.

Grid tool. Because there can be many features in a region, the name of a feature is replaced by an arbitrary number, which is linked to a feature. The numbers are assigned in increasing order. For example, a cell that contains two features matches them with numbers one and two. Vocoder-integers are used instead of spoken names because it would take too long to speak the names of features of a large set, and speaking the names of features does not carry tonal pitch information to the listener as the vocoder-integers do. The vocoder-integers are enumerated in increasing order, and looped. A short silence separates each iteration. The spatial locations of the features are refreshed in real time (pitch, stereo balance and loudness) to readjust for a stylus displacement. When the stylus is on a feature, a POI or/and intersection notification sound tells the user its nature, implicitly offering details on demand. The details-on-demand flowchart in Figure 4.8 describes the presentation structure. The system allows a user to have SAM speak the name of all the map features inside an active grid. Every feature in a grid is mapped to a num key on the keyboard, which allows a user to 'consult' the menu of features. Upon recognition of a given feature, the user memorizes the num key value and begins searching for it with the stylus on the map: only one integer value needs to be memorized. Finding that vocoder-integer value is finding the feature.

Row and Column tools. Appending three vertical grid regions makes a column and three horizontal grid regions, a row. There can be many features per columns or rows: enumerating everything is cumbersome. Thus, pulse-presence sounds replace the vocoder-integers to save time. A short silence separates each loop iteration. The spatial locations of the features are refreshed in real time to readjust to stylus displacement. These tools are designed to detect clusters of features along rows and columns and for this reason, details-on-demand is not supported. The sounds of the grid tool are available online[5].

4.6.5 Collision Detection Tool

Curves/stylus collisions are detected by the collision detection tool [259]. Crossing a curve triggers a pulse-presence sound. A details-on-demand gives the name of the curve and available annotations.

[5] http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/Interactions.html

4.6.6 Front Tool

A triangular beam from the location of the stylus tip silences features outside its 'field of view'. The user controls the direction of the beam by angling the stylus. Loudness, tonal pitch and stereo balance provide the locations of features in the beam.

The front tool allows a user to go through map content hastily and obtain details. In skimming mode, the names of features are replaced by pulse-presence sounds, ordered by increasing distance. In details-on-demand mode, the names of features are spoken in the same order. The details-on-demand flowchart in Figure 4.8 describes the presentation structure. A sound example of the front tool is available online[5].

4.6.7 Alignment Tools

With the alignment tools, a POI or a curve signals its presence when it comes in contact with a horizontal or vertical scan line centred on the stylus tip, as shown in Figure 6.5 on page 112. Features along the horizontal scan line are encoded by pitch, since they range vertically, and features along the vertical scan line are encoded binaurally, because they range horizontally. Loudness encodes distance in both cases.

A user can use the alignment tool to scan space or get rich details. In skimming mode, features are indicated by pulse-presence sounds, ordered by increasing distance. In details-on-demand mode, the names of the features are spoken.

4.7 Summary

Memorability was the paramount concern in the presentation design. In addition to eliminating visual references and adopting natural mappings, the syntax of details-on-demand stresses concision to accommodate the limited capacity of short term memory. Likewise, the repetition of information by the peripheral tools [260] and the bundling of map features into chunks, reduce the number of meaningful units a user must retain [261].

The design preserves users' cognitive resources. For example, notification sounds encode the generic quality of an omitting details. In practice, simply indicating the presence of an intersection is usually enough for a knowledgeable user. Attention is conserved for tracing and learning new features. Similarly, at a lower level, encoding elevation by pitch and tonality eases tracing curves with sound, again reducing cognitive overhead.

Finally, ambiguities in the interface, which should be minimized, was examined in the pilot studies. Alternative interpretations noted by participants were disambiguated. For example consider the token 'is continuing', which tells the user that a curve continues

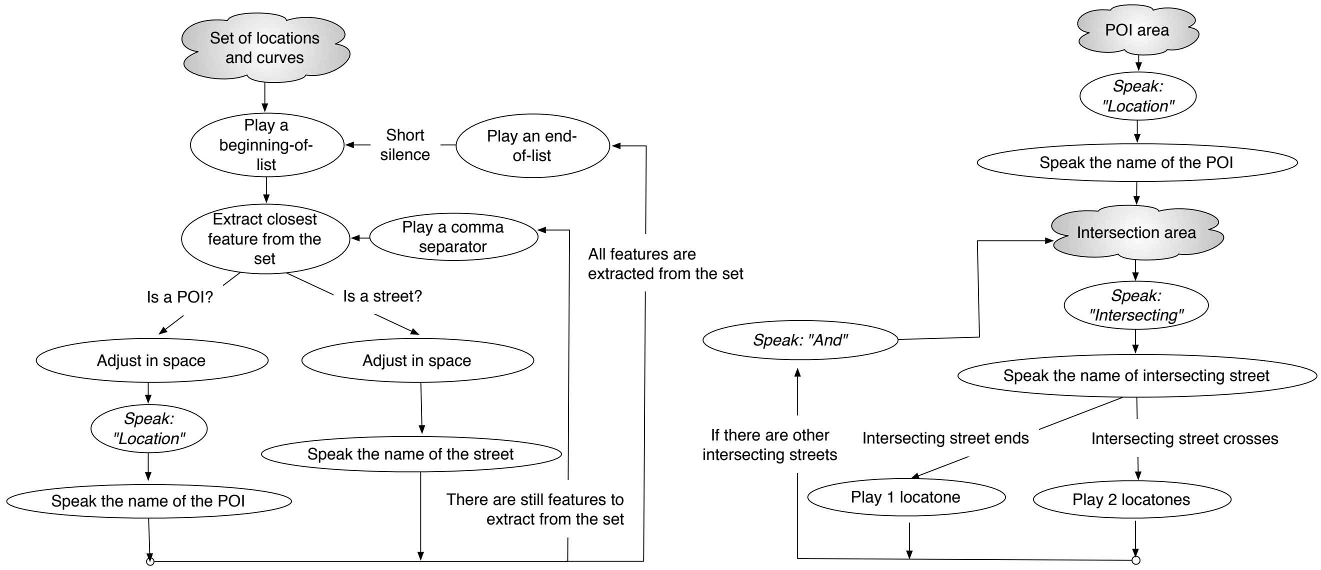

Figure 4.8: The *left* figure shows the presentation protocol for the alignment and front tools upon a details-on-demand request by a user. The list of locations and curves is sorted by increasing distance and itemized. Distance affects loudness. The *right* figure shows the presentation protocol for the grid tool; because curves are not traced with this tool, all the intersecting curves are treated in the same manner.

beyond intersection. At a lower level are toggle cues, which notify a user when a new curve is touched by the stylus.

Pilot tests on the data presentation protocol with blind participants suggested that the information could be heard, differentiated, understood and remembered, the goal of the design.

Chapter 5

Directional Encodings

5.1 Introduction

The goal of the studies described in this chapter is to understand spatial interpretation of nonspeech sounds in blind persons. Thus, this chapter investigates sound design and data presentation issues further. Two empirical studies are discussed: the first focuses on nonspeech sound interpretation and has minimal user interaction. The second integrates user interaction with nonspeech sound interpretation, by having participants follow trajectories with a stylus. The results of these studies assisted me in the design of locatones and curve tracing cues.

5.1.1 Background Theory on Visuospatial Cognition

Acting in an environment is only possible if the direction and distance of objects in the environment are known. Location is understood in terms of a reference system. Research on spatial cognition identifies three types of reference systems: exocentric, egocentric and sequential. An *exocentric* reference system is independent of the perceiver using it, north/south east/west, for example. An *egocentric* reference system is fixed in the perceiver, ahead/behind left/right, for example. A *sequential* reference system is fixed by a sequence of movements.

Haptic perception, the most precise spatial modality available to a blind person, is inherently egocentric [34]. Congenially blind (CB) persons tend to understand space in egocentric and sequential terms, since they are more immediate and reliable to them [138]. Exocentric coding is common in adventitiously blind (AB) persons, but only occasionally in CB persons [262]. Blind persons who can understand space in exocentric terms build more accurate cognitive maps than those who do not [34].

Susanna Millar [34] observed that kinaesthetic feedback in blind persons, particularly from hand movements, are important for understanding external relationships like the location of objects in space. She suggests that spatial information is best given to blind persons in body-centred terms and that exocentric references can be constructed by adopting effective movement strategies. This process requires a cognitive transfer from egocentric experience to exocentric experience, towards environmental representation of space in memory [263].

Many other researchers view movement strategies as key to building accurate cognitive maps in blind persons [264, 147, 265]. The word *strategy*, is used in two different but related senses: in one context, it refers to the actions carried out to explore a volume of space; in the other, it refers to the types of cues that are used to learn spatial information (spatial coding). As pointed out by Ungar *et al.* [174], a particular search suggests a particular form of coding, and the particular form of coding pushes the perceiver toward a particular kind of search.

The results reviewed in Chapter 3 suggest effective methods of searching in maps, but leave interpretation of abstract sounds untouched. It is necessary to validate encoding strategies in terms of their effects on exocentric representation in CB and AB users. Thorndyke and Stasz [266] warn, however, not to expect universal success as they observed blind participants with lower ability who made negligible progress at adopting exocentric representations.

5.2 Experiment 1

Speech can provide precise directions to a person attempting to reach an unseen destination. *Move straight ahead* or *turn left* are clear and unambiguous, but a moving pedestrian needs frequent precisely timed updates, and the longer the message, the slower the refresh rate. Map technology requires frequent refresh because the environment is highly condensed.

Nonspeech sounds, encoded spatially and short in duration, can provide the same quality of information, but may be misinterpreted by a blind user. Indeed, there are two possibilities: the listener might be moving in a fixed world, or the listener might be fixed while the world moves. The interpretations send the user in different directions [267]. These interpretations are known as the Ecological Model and the Tutored Model [170] and are illustrated in Figure **??**. Which interpretation is natural has been studied in airplaine pilots [268], but not in blind persons. Airplane pilots assume an ecological model[1],but blind perceivers, who are egocentric, might well differ. This ambiguity is examined in experiment 1.

[1] However, the attitude indicator in aircrafts adopts the tutored model to show the orientation of an aircraft relative to earth, which creates much confusion in novice pilots [268].

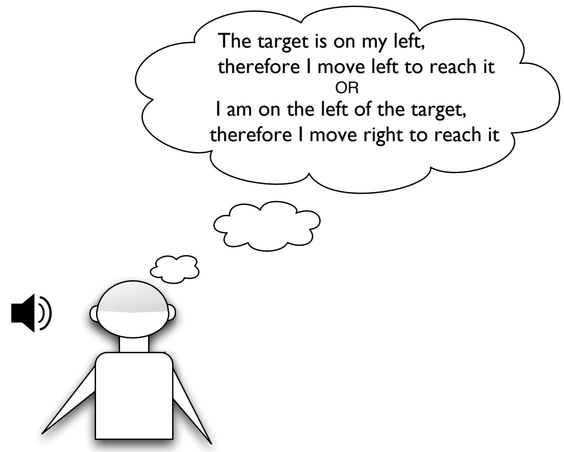

Figure 5.1: A blind user hears a nonspeech sound in his left ear. How should he interpret that cue? The first interpretation is ecological, the second is tutored.

Reference systems are closely related to the two interpretations. Experiments show that most CB persons understand space egocentrically. In fact, the field study in Chapter 3 showed similar egocentric behaviour in all participants, including the two AB ones. All reported perceiving themselves walking on the map during haptic exploration and, all described changes of their lateral axes when their fingers changed direction on a tangible map. For example, when moving northward on a map, east and west are reversed compared to moving southward. There is no consensus about this navigation-like strategy, which felt natural to the participants. Some consider it sub-optimal for building cognitive maps [138]; others disagree [269].

Exocentric coding and 'in alignment with the body axis' coincide because the reader faces the map so that body-axis directions are the same as exocentric direction regardless of finger orientation. Thus egocentric coding means that spatial directions are provided with respect to the index finger's frame of reference: recall that the index finger is a smaller representation of the reader on a map.

In summary, there are two fundamental issues to test. The first asks which direction model comes naturally to blind persons, and the second determines which spatial guidance is most effective. Because pitch encodes elevation, the experiment also examines an unusual model: egocentric encoding in azimuth and exocentric encoding in elevation. This model is based on the stable association between pitch and elevation (the Pratt effect), which may dominate the shift of lateral axis during displacement. Indeed, the up/down axis is very stable in humans because of its unique asymmetrical quality [64].

Directional Stimuli

The directional sounds are binaurally manipulated for horizontal directions, and in pitch for vertical directions. The cues must be short in duration and have a fast attack for superior localization, and must satisfy the musician's uncertainty principle, which constrains their duration.

The directional sounds originate in electronic toms from the Digidesign xPand™ library. Perceived pitch of the toms was calibrated in a short pilot experiment. The sound was reworked in Pro Tools™ to increase its transients in the attack, flatten its frequencies and dynamics, and enhance its warmness. To avoid *inside-the-head localization*, which gives listeners the illusion that the sound image is located from inside the head, a subtle reverberation was added to the binaural stimuli [204].

The resulting sound, which is characterized by a sharp attack (< 1ms), short sustain (≈ 5ms) and relatively short decay (≈ 30ms), is reproduced into 13 binaural sounds at pitch G3, generated at 24-bit resolution with a data rate of 44.1kHz: stimuli 1 to 7 are gradually panned in stereo from extreme left to the centre, and stimuli 7 to 13 from centre to extreme right. The binaural changes are just above the noticeable difference as measured in pilot experiments. The stimuli are available online for listening[2].

Each binaural sound is post-modulated into pitch variation for elevation. Because a quarter-tone is the smallest interval easily perceivable by humans [270] and small pitch variations are preferable for streaming [206], pitch is modulated by quarter-tone steps, creating a series of intervals ranging between $\frac{1}{4}$ and 3 whole tones in each direction. These intervals are grouped into three classes: quarter-tone, half-tone, and whole tone or greater. The classes are cross-compared in the study.

5.2.1 Method

Two orthogonal tracks of equal length intersecting each other to form an 'L' are carved into a thin surface board, which is laid on top of a digital tablet. The width of the tracks is adjusted for the stylus tip, allowing the stylus to slide along them. A raised dot is glued at their intersection. See Figure 5.2 for a visual representation.

Hidden to the participants are lines stored in the computer, which are centred on the track for about one centimetre from the intersection, and then shift on either side, staying parallel to it, the dashed line in figure 5.2.

[2] http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/DirectionalSounds.html

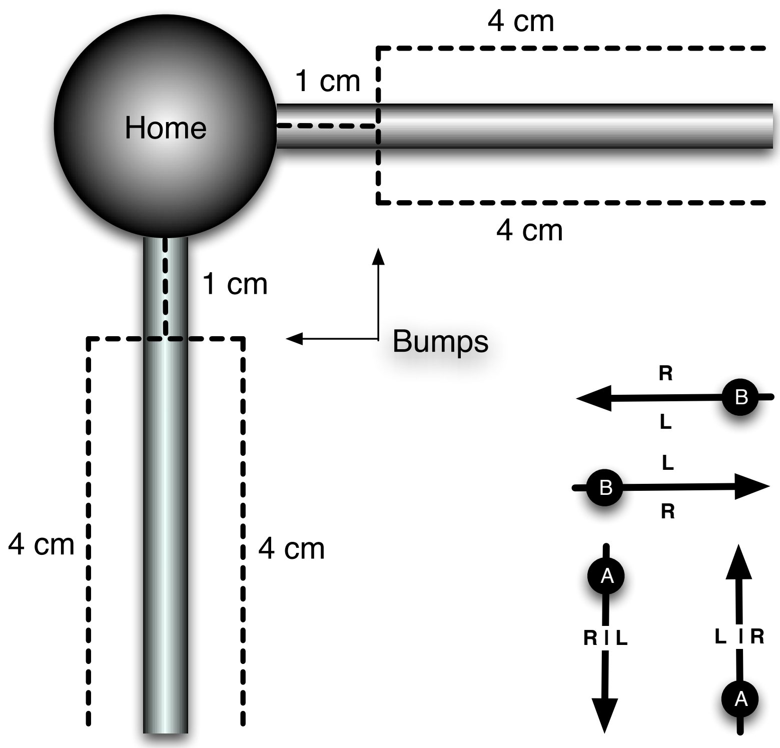

Figure 5.2: *The upper-left region of the diagram* shows a representation of the two orthogonal tracks, the central dot, and the shifted lines inside the computer (dashed lines), which could be closer or further away from the tracks. Also, but not mentioned in the text, two small bumps, felt tactually when moving the stylus over them, are placed at the shifting location of the virtual lines, to redundantly bring the shifts to the attention of the participants. *The lower-right region of the diagram* shows two pairs or arrows labelled A and B. The arrows show how the axis of control in the lateral direction shifts depending on the direction of the stylus. For example, the left arrow labelled A shows that leaving home and moving towards the end of the track reverses the left-right directions (because the perceiver faces the map). The right vertical arrow shows that the left-right direction is consistent with the perceiver when moving towards home. The same concept applies with the horizontal directions, as shown by the arrows labelled B.

Sensor

A sensor detects the computer lines. (Details about the sensor are covered in Experiment 2, as they are unimportant here.) When the stylus moves, stimuli are repeated at a constant rate of 6.7 Hz for streaming coherence; otherwise, the system is mute.

Spatial Instructions

Details about the three families of spatial instructions are provided in this section. The first family displays egocentric cues in the lateral and longitudinal dimensions (GG), the second family presents egocentric cues in the lateral dimension and exocentric cues in the longitudinal dimension (GX), and the last family provides exocentric cues in both lateral and longitudinal dimensions (XX).

GG, with the finger's frame as reference model, needs a bit of explanation. With this encoding, elevation information is never displayed. When on a curve, regardless of direction, the finger is either centred on it, or leaving it to the right or the left. Picture a curve as a corridor. Regardless of direction a traveller leaving the centre of the corridor translates to the right or the left. With GG, the horizontal axis changes in polarity with direction, as explained in Figure 5.2. Thus, this encoding inspired by Hong's approach [129], is in effect a hybrid between navigating and surveying perspective [271].

Because up is always up and down is always down, as discussed earlier, GX provides egocentric information in the lateral direction and uniform pitch polarity in the longitudinal direction (exocentric).

The XX encoding guides the participants with directions referred to the frame of the map, or in alignment with their body axis. Up always means north, left always means west.

5.2.2 Experimental Procedure

The same action was performed in every trial: participants slid the stylus from the raised dot to the end of the track and came back to their starting position without lifting the stylus, and were then asked, based on the sound they heard, in which direction they thought they would need to move the stylus to reach the shifted target.

One hidden target line was presented per trial. The participants responded by saying 'right' or 'left' for the vertical track trials and 'up' or 'down' for the horizontal track trials. Their responses were recorded by the experimenter (and author of this thesis) so they could have both hands on the surface of the tablet.

The participants were then asked to reposition their stylus inside one of the two tracks and perform another trial.

Details on a Trial

The stimulus is binaurally centred and pitch is set to G3 near the raised dot. Moving the stylus on a track beyond the bump, as shown in Figure 5.2, alters stereo balance or modulates pitch. The distance between the shifted line and the stylus track determines the amount of change, i.e., the greater the shift, the more off-centre the stereo balance and the greater the interval variation.

Participants

Four sighted and four blind participants took part in the study. The sighted group consisted of graduate students, two males and two females in their mid-twenties. As explained earlier, research suggests that sighted persons understand space in exocentric terms and prefer the ecological model: sighted participants are used to validate the design of the experiment and to compare the performance of blind participants.

Table 5.1 provides further details about the blind participants. All the participants reported normal hearing and demonstrated the ability to hear relative pitch.

The study, which received ethics clearance through the Office of Research Ethics, was conducted one subject at a time, each session taking about 40 minutes. The participants received $15 in compensation for their time.

Table 5.1: This table shows the profile of the blind participants. The left column shows the participants' identification: CB and AB stand for congenitally and adventitiously blind; F and M for female and male. In the topmost row, A stands for age (rounded up), O for age of Onset, R for Residual vision, MA for mobility aid and T for tangible map training (at school). Inside the table, L stands for light perception, N for no residual vision, D for guide dog, C for cane, Y for yes and S for self-trained.

Id	A	O	R	MA	T
CBF	30	0	L	D	Y
CBM	30	0	L	C	Y
ABF	30	10	L	C,D	S
ABM	45	6	N	C,D	Y

Apparatus

Trials were created and responses recorded using an Intel Core 2 Duo PC clocked at 1.86 GHz with 3GB of RAM running Vista and hosting a RealTek™ high definition sound card. The OpenAL sound engine (v.1.1) was used in the implementation to play back stereo sounds. The participants provided input using a Wacom™ Intuos 2 graphics tablet, model XD-0912-U. This tablet has a coordinate resolution of 100 lines per mm, a pen accuracy of ± 0.25 mm and a maximum report rate of 200 points per second. Sounds were played through a pair of Sony™ MDR-7506 headphones.

Experimental Design

A between subjects design was employed. A total of 168 trials were randomized by the system with no variables blocked, including encodings, then presented in sequence to the participants. The study trials were presented in eight blocks of twenty-one questions so that participants had many opportunities to take many breaks. Few requested a break, indicating that the task was not unduly onerous.

A short tutorial, which contained trials similar to the study trials and which could be repeated *ad libitum*, preceded the experiment.

Briefing the Participants

A tactual representation, similar to but simpler than the one shown in Figure 5.2, was built to brief the blind participants. A visual representation was used for sighted participants. The participants were told that their responses should take their movement directions into consideration, and were asked to avoid taking too long with any single trial because there were many of them.

Measures

Response time and direction are used as measures. An encoding that takes shorter deliberation time than the others is preferable. The direction responses examine the participants' interpretation of directions.

5.2.3 Results

Participants interpretation (ecological vs. tutored) and frame of reference (egocentric vs. exocentric) are encoded together in their responses. The arrows labelled A in Figure 5.2

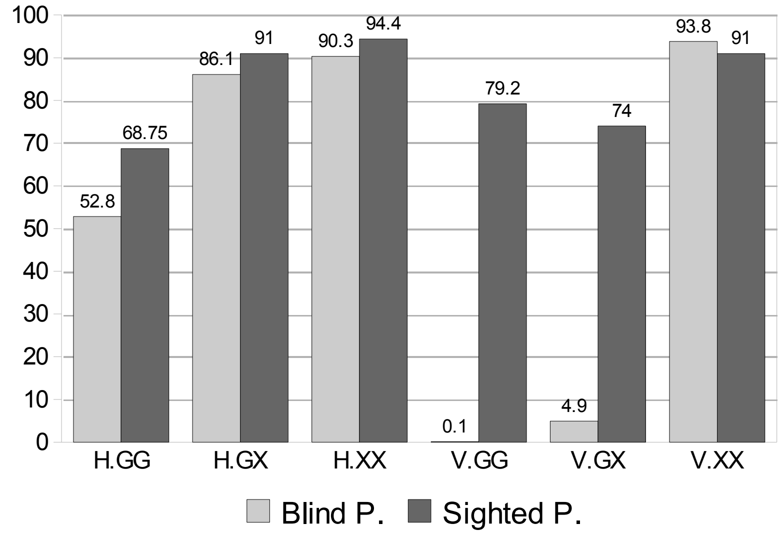

Figure 5.3: This figure shows the scores for direction. H stands for horizontal and V for vertical. The data should normally be interpreted as follows: a direction score near 100% indicates that participants adopted the ecological model, and near 0%, the tutored model. The results are interpreted and explained in Section 5.2.3

show that sliding the stylus along the vertical track inverts the body axis position of the participant who sees his finger as a smaller representation of himself. The axis is re-aligned on the way back. Tracing along the horizontal track aligns the lateral direction when leaving home, but inverts it when coming back, as illustrated by the arrows labelled B in Figure 5.2. A participant who understands space egocentrically should perform the rotation when necessary. The blind participants did not.

As we can see in Figure 5.3, the blind participants shifted their polarity in V.GG and V.GX but less in H.GG and not at all in H.GX. In post-interviews, all the blind participants mentioned paying attention to the first directional cues and ignoring the second, which reveals that they did not understand their locality egocentrically. Indeed, a close look at the data confirms their strategy: vertical questions always consisted in leaving home downward, which inverted the axis on the way down and re-established it on the way up. Horizontal questions always consisted in leaving home towards east, which aligned the axis when leaving home and inverted it when coming back. This is what we find in V.GG & V.GX and H.GG & H.GX.

On the other hand, all the sighted participants performed the proper mental rotation, as shown in Figure 5.3. A closer look at the data reveals a three-way interaction between the two groups of participants, the encoding, and direction ($F(2, 1332) = 50.675$, $p \leq$

0.0001); two Sheffe post hoc tests on the GG and GX encodings in the vertical direction suggest a significant score difference between the two groups in both cases ($p<0.0001$). Details of the analysis are in Appendix A.2.

The high scores found in H.XX and V.XX in Figure 5.3 indicate that the blind and sighted participants adopted the ecological model. This interpretation is corroborated by answers from the participants in post-interviews.

We now examine frequency intervals. Frequency intervals had no significant effect on time but had an effect on direction judgements ($F(2, 1338) = 7.218$, $p = 0.0008$), where the intervals greater or equal to half tones scored significantly higher than the quarter tone intervals ($p = 0.0033$).

Binaural variations were well perceived, because there were no significant interactions between encodings and directions on time and direction judgements. I conclude that small variations in the azimuth direction were well perceived.

5.3 Discussion

The results suggest that sighted participants deciphered the GG encoding by mentally rotating their body axis when directing the stylus towards them, but the blind participants did not. When in doubt, the blind participants relied on the orientation of their trunk to respond, a phenomenon known as the trunk dominance principle [170]. The blind participants failed to personify the stylus as a representation of themselves during the experiment.

Experiment 2 measures the prevailing encoding, XX, against an echolocation encoding. Like GG and Hong's encoding [129] discussed in Chapter 2, the echolocation encoding provides 2D information by left/right sonic asymmetry. A key difference is that echolocation is used by blind travellers.

Both encodings are explained to the participants before running Experiment 2. Because the participants judge direction ecologically, only the ecological model is implemented. Quarter-tone intervals are eliminated and half-tone intervals are chosen as the smallest pitch variation SAM displays to its users.

5.4 Experiment 2

Experiment 2 followed Experiment 1 by a few weeks.

5.4.1 Method

Experiment 2 examines sensor design characteristics and sound encodings for curve tracing tasks. Two encodings are also compared, XX and echolocation. The stimuli in XX, which I will refer to from now on as the binaural/pitch encoding, are treated as described in Experiment 1. The echolocation encoding is explained in the next section.

Stimuli

Echolocation is a natural method of understanding a volume that is learned early in life by blind persons. It enables them to detect objects by reflected sound. The distance to an obstacle is determined by the time an echo takes to return. Kish [232] tested whether echolocation could enable blind travellers to walk in the middle of a long corridor, observing straighter trajectories when echoes were present. Corridors and curves are similar: could a blind person trace a curve using echolocation? If echolocation keeps a blind pedestrian equidistant between two walls, might it be used to keep them on a curve? These questions motivated examining echolocation.

I relied on two resources to build the encoding: I looked at the theory of echolocation [232] summarized below, and examined echoes recorded using a Sennheiser MKE 2002™ binaural head.

Timbre. De l'Aune *et al.* [272] analyzed spectrograms of recordings taken through artificial ears, observing that the cross-section at the corridor shapes the spectrum of echoes. Cotzin & Dallenbach [273] reported that reliable distance estimation was difficult with pure tones but less so with complex sounds, because a pure tone carries a single frequency, which may or may not reverberate well. Complex sounds, in contrast, are composed of many frequencies, some of which are bound to resonate well. Indeed, mouth clicks, which are used by blind travellers performing echolocation, have energy throughout the frequencies from 2 to 6 kHz. Sounds with significant energy between 500 Hz to 8 kHz were observed to perform well when detecting walls and doors at a distance [274, 275].

Pitch. As the distance between an obstacle and a perceiver decreases, the low frequencies of a sound source have too little space to expand and reach to the listener, which changes the spectral composition of echoes. The effect is perceived as pitch variation.

Envelope. Sound localization is important in echolocation: a sharp attack outperforms a slow-rise onset.

Duration. The duration of a sound should be between 6.6 and 20 ms, the average duration of a mouth-click [232].

Sound Design

The same electronic toms were chosen as the base sound for echolocation. The same transformations of Experiment 1, slightly adapted to better match the wall reflections captures by the MKE 2002 binaural head, were used to adjust the sound parameters. The transformed stimuli complied with the timbre, pitch, envelope and echo requirements discussed above. The echolocation stimuli are available online[3].

Local vs. Global References

Space on a map can be presented globally or locally. Global representation consists in building a bijective correspondence between the set of positions on a map and the set of sound parameters in the display. For example, for a stereo balance and pitch encoding, each position on a map corresponds to a unique combination of pitch and stereo balance: high pitch in the left ear would indicate upper left the map, low pitch in the right ear, lower right. Researchers who attempted to use a global reference scheme to display curves with sound reported limited success [82, 87, 90, 91, 92, 96].

Local reference, as commonly implemented in ETAs and beacons systems, depends on kinaesthesia for position, giving more freedom in the sound guiding a user following a trajectory. Like the input provided by the mechano-receptors of a Braille finger, the sonic cues stay local to the stylus tip. Thus, the stylus tip remains at the centre of the auditory scene. This solution finesses the low spatial resolution of audition because the auditory range is confined to an area the size of a finger tip. Also, it places the perceptual experience at the tip of the stylus, allowing the user to scrutinize any area. Finally, it takes advantage of a higher information bandwidth, as the bit rate of audition is estimated to be one hundred times that of a finger tip [28].

Sensor Design

The sensor design enlarges sensitive area of the stylus, allowing the detection of near features. Three sensor shapes were implemented and tested in short pilots preceding the experiment: circular, semi-circular and linear. The three sensors were equal in performance but not in precision:

> *Performance.* Nearby features were detected with the same reliability and time delay.
>
> *Precision.* The linear sensor has a smaller sensitive area than the circular and semi-circular by factors of π and $\pi/2$ respectively, and is therefore more precise. For these reasons, the linear sensor was chosen.

[3] http://www.cgl.uwaterloo.ca/∼m2talbot/thesisAppendix/DirectionalSounds.html

The sensor has the following characteristics: it is centred at the stylus position; it is orthogonal to the tracing direction; and it is divided into thirteen sensitive regions distributed along its width[4], as shown is figure 5.5. As a region comes in contact with a curve, the sensor computes the angle between its centre and that region to direct the user spatially.

Equations 5.1, 5.1 and 5.3 explain how the sensor's end points are computed in closed-form by the system. Once the end points are known, a line is drawn using Java's API and subdivided into regions using pre-determined Euclidean distance values measured from the centre. Then the system monitors this line for intersection with a curve.

In equation 5.1, w is half the width of the sensor, x_o and y_o are the current position of the stylus on the tablet, θ is the angle between the curve and the sensor. Equation 5.1 shows the distance form corollary [276] for finding (x_a, y_a) and (x_b, y_b). The two end points of the sensor as solved using Equations 5.2 and 5.3. Please refer to the illustration in Figure 5.4 for a visual representation.

$$\frac{(x - x_o)}{cos\ \theta} = \frac{(y - y_o)}{sin\ \theta} = w \tag{5.1}$$

$$\begin{aligned} x_a &= w\ cos\ \theta + x_o \\ y_a &= w\ sin\ \theta + y_o \end{aligned} \tag{5.2}$$

$$\begin{aligned} x_b &= x_o - w\ cos\ \theta \\ y_b &= y_o - w\ sin\ \theta \end{aligned} \tag{5.3}$$

Sensor Widths

The width of a sensor impacts its coverage range and precision. A wide sensor is less likely to lose contact with a curve, but because the number of contact regions on the sensor remains the same, a wide sensor is less responsive than a narrow one.

Five sensors of different widths were tested: 0.05, 0.9, 1.4, 1.9 and 2.4 cm. Evreinova *et al.* [97] examined this problem on a low resolution sensor[5], but provided measures in pixel per inch, which are difficult to interpret because the mapping of screen pixels varies with the size of the tablet. A length unit like cm is preferable because it is independent to tablet size. Here, the width of the thirteen individual regions is the product of the sensor

[4] I use width and not length to describe the sensor's size from end to end, because the sensor stays orthogonal to a stylus direction and scans space from side to side.

[5] Their sensor had only three sensitive regions.

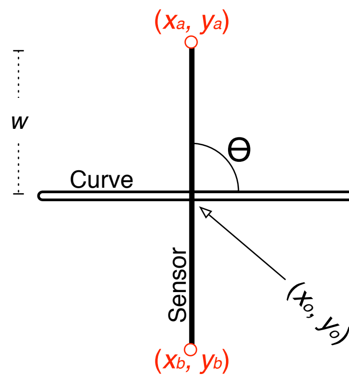

Figure 5.4: (x_o, y_o) is the stylus coordinate. (x_a, y_a) and (x_b, y_b) are the two end points of the sensor and w half its width.

width and the normalized width values shown in Figure 5.5. The sensor of width 0.05 cm is in effect a collision detector because its width is negligible. It is a base case providing no off curve direction guidance.

Mapping Regions to Directions

The system provides guidance in two distinct directions: horizontal and vertical. The two are presented individually; with the stream integrated within the user. For example, a user tracing a diagonal curve receives a series of orthogonal directions *move up, move left, move up, move left, move left, move up, and so on* in the form of non-speech sounds.

When a sensor contacts a diagonal curve, the system determines the best direction between horizontal and vertical. The system projects the stylus position to the curve along the x and y axes and picks the direction with the shortest Euclidean distance, as illustrated in Figure 5.5.

5.4.2 Experimental Procedure

The apparatus described in Experiment 1 was reused.

Participants

Sighted participants were eliminated because they have little to no experience with echolocation. The same blind participants were the only subjects.

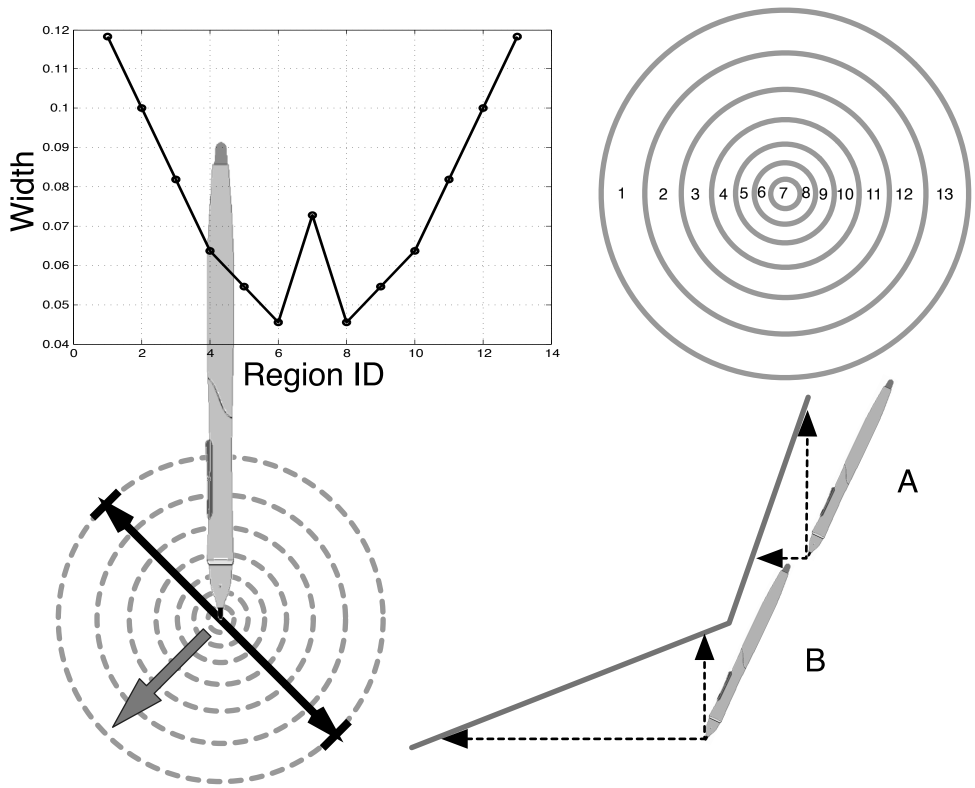

Figure 5.5: *Top-left:* this graph shows the width of the 13 sensor regions, which were adjusted in pilots by trial and error. The x-axis values identify the regions: region ID 1 and 13 being the two extremities and 7 the centre. The y-axis holds normalized width values. The central region is wider than its adjacent regions to prevent the system from giving a new and unnecessary direction to a user already centred on a curve. *Top-right:* the figure shows the span of the linear sensor and its 13 regions (spun about its centre); the numbers inside the circles correspond to the region's ID shown in the graph. *Bottom-left:* The diagram shows how the sensor and its subdivided regions stay orthogonal to stylus direction. The double-headed arrow line represents the active sensor, the grey arrow the current stylus direction, the dashed circles represent its potential range and regions (if the stylus was to change direction). *Bottom-right:* the diagram shows the greedy algorithm at work. In example A, the system sends a user left; in B, the system sends a user up.

The study, which received ethics clearance through the Office of Research Ethics, was conducted one subject at a time, each session taking about 40 minutes. The participants received $15 in compensation for their time.

Details of the Trials

Without the tracks used in Experiment 1, participants can move the stylus freely on the tablet within a 16 by 24 cm frame. A raised dot labelled 'Home' at the centre of the frame, was the starting position of every trial, as shown in Figure 5.6.

The procedure was the same for every trial: the participant positions the stylus near the raised dot and is told the starting direction of the curve: north, south, east or west. Every curve runs straight for a short distance before changing direction to give the participant a precise orientation.

Every trajectory is recorded as a sequence of time position triplets. Recording stops as the stylus arrives within 0.3 cm of the curve's end point.

Measures

Each trial is assessed by three measures: the time taken to trace a curve; error, which is the distance between a participant's trace and the underlying curve; and the length of a trace[6]. The distance error is computed off-line in Matlab™ using an Edit Distance algorithm explained below.

The original *Edit Distance* algorithm calculates the level of similarity between two sequences of characters, that is, the amount of error is the minimum number of operations required to transform one string into the other, mapping the common subsequences together at no cost and penalizing the parts that are different [277]. Here, a modified version of the algorithm measures similarity between two curves: a user response and a solution curve. Two points are the same if they are within a threshold distance, which I adjusted and kept constant to 5 pixels throughout the evaluation. The algorithm is shown in Section A.1.

Experimental Design

A within-subject randomized design was employed to gather data. Every participant did 40 trials: 4 different headings (north, south, west and east), 5 sensor widths, and 2 sound encodings.

[6] The measures are normalized for cross comparisons because the trials are not all equal in length. The normalization is done by dividing a result by the length of its corresponding curve trial.

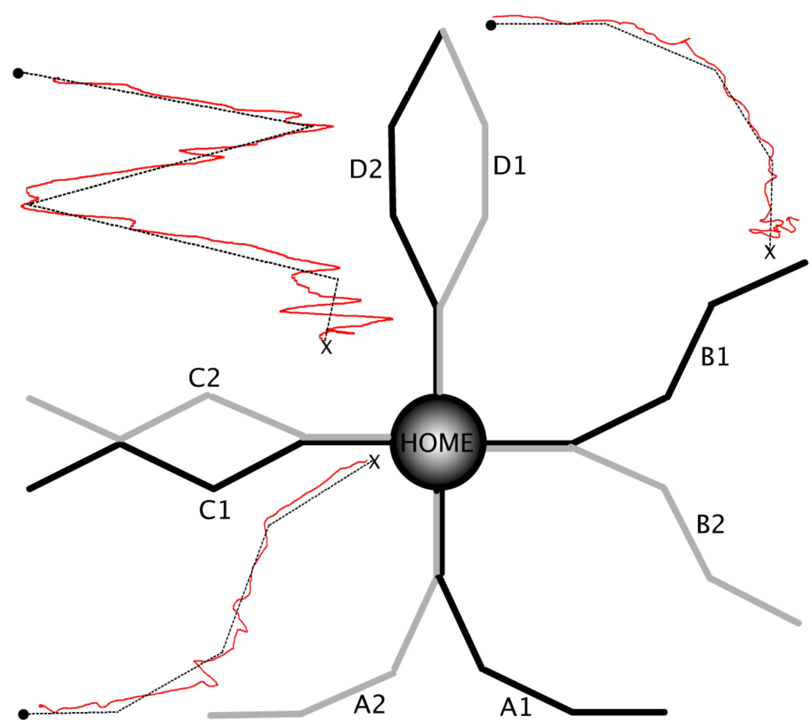

Figure 5.6: *At the centre*, leaving from the large dot labelled 'Home', are shown the four families of curve that could be presented to a participant. Every family could start in any cardinal direction, for a total of 32 different trial questions. *The wiggly lines* are responses by the participants; the solution is shown underneath. The top-left example is taken from the tutorial set; the trace was performed by a participant getting acquainted with the encoding. Each curve is labelled by a letter and a number.

The curves were sampled randomly from the curves shown in Figure 5.6 with the constraint that every curve was presented at least once.

A short tutorial, with trials similar to the study trials and which could be repeated *ad libitum*, preceded the experiment. One of the tutorial curves is shown in Figure 5.6.

The study trials were presented in four blocks of ten questions, giving the opportunity for three breaks. Only one break to rest was requested.

Briefing the Participants

The encodings and the randomized structure of the trials were explained to the participants, who were asked to try keeping the stylus centred on curves. They were informed that their movements and response times would be recorded by the system and used to determine the best encoding.

5.4.3 Results

Distance error, displacement and time were subjected to an analysis of variance. To start, an analysis of variance with width of sensor and encodings as factors, and including two-way interactions, was performed. In all three cases, interactions were not significant.

Sensor Width. Three consecutive analysis of variances were performed without interaction on distance error, displacement and time with only the sensor width as a factor. The analysis revealed that changing the width of the sensor affected precision ($F(4, 151) = 4.3336$, $p = 0.0024$). A Sheffe post hoc test indicated that 1.4 cm produced more precise traces than 0.05 cm ($p = 0.0044$), and also more precise than other sensor widths, but not significantly so. The analysis of variance on displacement also reveals a significant effect ($F(4, 151) = 6.1438$, $p \leq 0.0001$). Post hoc tests again show the worst performance for the 0.05 cm ($p<0.007$) and the best performance for 1.4 cm, but not significantly so. Finally, the analysis of variance on time also shows a sensor width effect ($F(4, 151) = 4.8239$, $p = 0.0011$). Width 0.05 cm took marginally more time that other widths (0.9 cm $p = 0.0789$, 1.4 cm $p = 0.0012$, 1.9 cm $p = 0.0553$, 2.4 cm $p = 0.0217$). 1.4 cm took the least time, but not significantly so. The reader may refer to Figure 5.7 for an overview of the results.

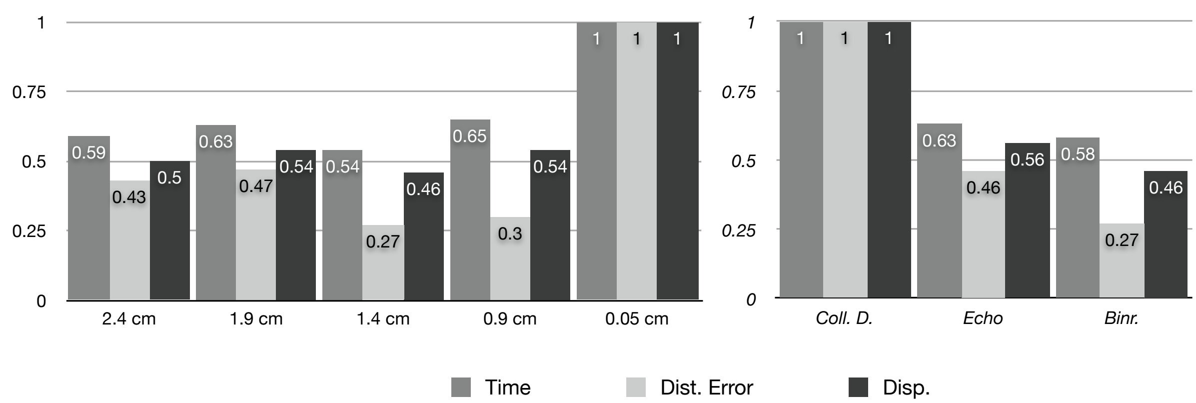

Figure 5.7: *The left graph* assembles the results for the five sensor width treatments. Response time, distance error and displacement are normalized against the highest performance values. Similarly, *the right graph* compares the three encodings. Coll. D. stands for Collision Detection, Echo for echolocation, Binr. for binaural/pitch encoding. Less is best for each of these measures.

Encodings. As above, I performed three consecutive analyses of variances without interaction on distance error, displacement and time, but this time encoding method was used as factor. Encoding had a significant effect on distance error ($F(3, 152) = 6.0126$, $p = 0.0007$). Post hoc tests indicated that collision detection produced greater distance error than binaural/pitch ($p = 0.0011$), and marginally more distance error than echolocation ($p = 0.0305$). Binaural/pitch encoding was the most precise, but not statistically so. The analysis of variance on displacement also showed a significant encoding effect ($F(3, 152) = 8.7899$, $p \leq 0.0001$). Collision produced larger displacements than binaural/pitch or echolocation ($p \leq 0.0001$ and $p = 0.0015$, respectively). Binaural/pitch was more efficient than echolocation, but not significantly so. The analysis of variance on time shows similar results ($F(3, 152) = 6.3727$, $p = 0.0004$). Collision took greater time than binaural/pitch and echolocation ($p = 0.0025$ and $p = 0.0099$ respectively). Binaural/pitch took less time than echolocation, but not significantly so.

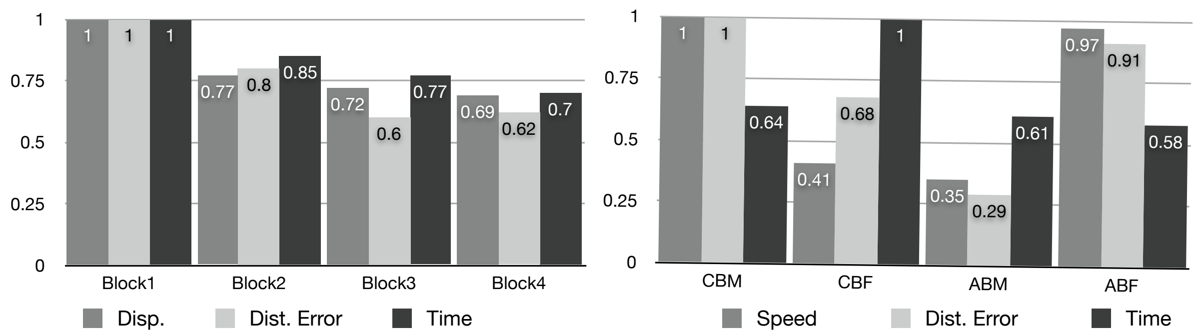

Figure 5.8: *The left graph* shows the participants' performance progression as they proceeded throughout the experiment. The participants moved from Block 1 to Block 4, but block questions differ for every participant. Time, distance error, and displacement are normalized against the highest performance values. Again, less is best for each of these measures. *The right graph* shows the participants' speed, distance error and time. A trade-off between speed and distance error is apparent: CBM and ABF moved fast and could not trace a curve with precision while ABM and CBF, who moved slow, could. Interestingly, CBM and ABM took approximately the same tracing time.

Blocks. The trials were randomly assigned to four sequential blocks to examine learning and fatigue for time, distance error and displacement. With block order as the only factor, performance progression from block to block was examined. Figure 5.8 shows a gradual, but non-significant, improvement over time.

Participants. Three analyses of variances on average speed (displacement divided by time), distance error and time with participants as a factor were performed to compare the performance of the four participants. Average speed varied among participants ($F(3, 152) = 12.381$, $p \leq 0.0001$). CBM and ABF traced slower than CBF and ABM ($p \leq$

0.001). Not surprisingly, time also varied among the participants: ($F(3, 152) = 8.2292$, $p \leq 0.0001$), CBF took significantly more time than the others ($p < 0.0001$, $p < 0.0013$, $p < 0.0053$). Time response for the rest of the group was not statistically different. Distance error varied among the participants, CBF and ABM being more precise than CBM and ABF, but not significantly so. Figure 5.8 shows an overview of the results.

I performed a Pearson Product-Moment correlation on average speed and distance error, grouping echolocation and binaural/pitch encodings together, and obtained a correlation of $r = 0.657$. The same analysis, but this time on binaural/pitch and echolocation separately, returned $r = 0.498$ and $r = 0.846$ respectively, confirming an expected speed/accuracy trade-off, but also suggesting the possibility that high speed had a greater impact on distance error with echolocation than binaural/pitch.

5.4.4 Qualitative Data

The participants were post-interviewed. They were also encouraged to think-aloud while doing the task, the results of which were recorded. This section examines this data.

Participants disliked the collision detection encoding: they complained of losing contact with curves and of having to zigzag. One participant was unable to visualize the shape of a curve because of performing too many displacements in too many directions.

Echolocation and binaural/pitch were equally liked: ABF and ABM preferred binaural/pitch and CBF and CBM preferred echolocation. Binaural/pitch was considered as less confusing. Echolocation was considered to be less complicated because only one cue, echo, is used which is easier than of two, sound balance and pitch.

Sensor width variations were noticed during the study: the participants pointed out that finding end points in easy trials was hard. As indicated earlier, a narrow sensor yields tighter traces, which keeps the stylus tip near a curve so that the stylus arrives close to the end point. All participants disliked spending time searching for end points, possibly because of a competitive spirit I observed.

All the participants tried to outperform themselves as they progressed throughout the study, but increasing speed did not help: CBM, the participant who exhibited the highest average tracing-speed, realized that reducing speed eased the task. He mentioned: 'Once you get yourself calibrated with the system, this gets easy'. A closer look at his performance reveals a gradual decrease in speed over time: Block1 = 1.82, Block2 = 1.78, Block3 = 1.20, Block4 = 0.78 (cm/s).

The trunk dominance I observed in Experiment 1 vanished, possibly because of the explicit briefing given to the participants. CBF commented a few times when tracing a curve downward, 'I am going down, left and right are reversed'.

5.5 Discussion

The second experiment of this chapter helped determining sensor design characteristics, but also examined further the problem of encoding tracing directions with non-speech sound. The data from this experiment shows the importance of telling the user the direction to the curve. Both echolocation and binaural/pitch encodings out-performed the collision detection encoding for the task of tracing curves with sound. Although the egocentric representation of echolocation was well understood by the participants, binaural/pitch out-performed it: the participants were more precise, performed fewer unnecessary displacements and traced at faster speed with the binaural/pitch encoding than with echolocation.

Echolocation provides a navigation-like experience on a map because the direction of features is signalled by the viewpoint of the perceiver in the display. This makes it appealing for trip rehearsing as it immerses a blind user in a navigation-like experience on a map. However, an immediate problem with echolocation is the keyhole-cost induced by its egocentric cues, which increases demand on spatial working memory, as it requires a perceiver to integrate multiple narrow views over time [170].

The experimental results do not provide significant evidence for or against a particular approach, suggesting that both encodings are worth further investigation. But because binaural/pitch was more efficient than echolocation, and because Binaural/pitch encoding was the most precise, the binaural/pitch encoding is chosen for SAM. Echolocation should be revisited in future work.

A wide sensor may touch more than one curve when several curves are present on a map, which creates the problem of determining which one should be displayed and which one should be hidden from a user. Admittedly, building a reliable user prediction model is not an easy task. For example, selecting the level of aggregation at which to make predictions is challenging [278]. However the most difficult task is clarifying user intentions, which is often unclear. A system can present a clarification dialogue to discover intentions [279], which distracts the user who is tracing a curve with sound. Alternatively, the system can present everything and let the user segregate it. How then, does a system best display many layers of sound simultaneously? An answer that demands careful consideration of perceptual stream segregation.

Fortunately, this issue can be addressed by adjusting the system's resolution. The results suggest that a sensor width of 1.4 cm is best. Setting the system resolution to half the sensor's width keeps the sensor from intersecting more than one curve at a time since two curves are never closer than 0.7 cm, except near intersection areas.

Interestingly, overall resolution of 0.7 cm is close to tactile resolution, which is of 0.6 cm [7]; curves separated by less than 0.6 cm are perceived by touch as one wide curve and not two separate ones. Indeed, Berlá & Butterfield [280] observed that lines of width 0.63 cm were traced with greater speed and accuracy than narrower ones, and with less distraction at intersections.

With the binaural/pitch encoding, vertical curves are easier to trace than horizontal ones, because perfect binaural balance is easy to perceive and can be used as a reference for alignment. Evidently, pitch reference is missing for horizontal curves because humans do not perceive pitch absolutely. The musical key of a tonality gives users the reference they need, as demonstrated in Chapter 7. The tuned interface, which is described in Sections 5.5.1 and 5.5.2, is the version SAM implements for its final assessment in Chapter 7.

5.5.1 Curve Tracing Cues: Final Design

The most important improvement is establishing a reference system to improve absolute elevation perception. Both the sensor and the sound display must be changed. The number of sensor divisions is increased to 15 regions from 13, improving resolution and, most important, extending the pitch range. With two extra tones the sensor spans a G major diatonic scale of two full octaves. With this improvement the sensor presents the central tone (tonic/key) three times, G3 at one end, G4 at the centre and G5 at the other end, which strengthens the tonality [199]. Azimuth is concomitantly divided into 15 stereo balance-steps for consistency.

In addition, as the pitch leaves the centre note of the sensor the loudness of the tracing cues decreases because the user is moving away from the curve. A choir sound signals a perfect sensor alignment with a curve. These new characteristics are shown in Figure 5.9. Finally, the subtle reverberation that was protected against inside-the-head localization is removed, as it smeared the display. The stimuli are available online for listening[7].

5.5.2 Locatones: Final Design

Locatones attract a user to a target destination. They encode azimuth and elevation using binaural and pitch, and distance to the target by loudness variations, beat-rate and reverberation.

Locatones consist of two parts. The first is the position of the user/stylus on the map; the second is the direction to the target.

As in real environments, where a perceiver stays centred in the perceived world, the first part is presented as centred, at pitch G4 and in stereo balance. The second part of a locatone connects the user to map features giving the direction towards the target

[7] http://www.cgl.uwaterloo.ca/∼m2talbot/SAMsDemos/

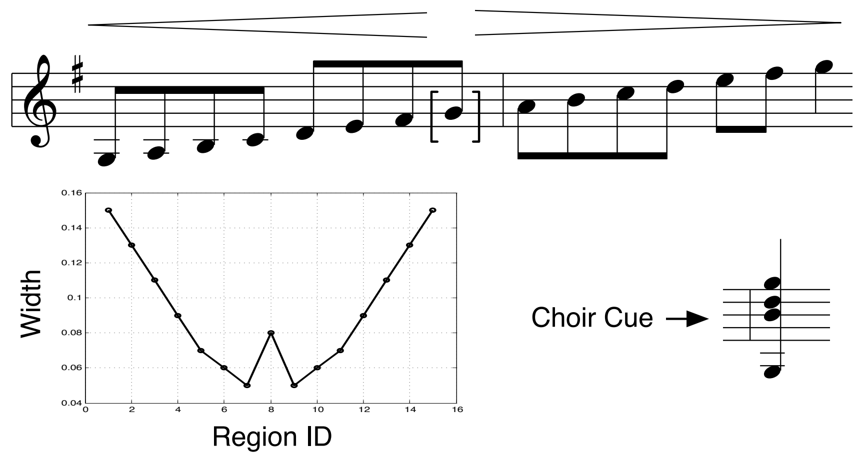

Figure 5.9: *Top:* elevation cues are divided into 15 pitch variations ranging from G3 to G5. The bracketed note, G4, corresponds to a perfect alignment in elevation. Loudness increases near the centre of the sensor. Loudness changes are represented by the crescendo symbols over the music staff. *Button left:* the graph shows the resulting width ratio of the 15 sensor regions for a sensor of width 1.4 cm. *Bottom right:* shows the voicing of the soothing choir cue played in the background when the sensor is well aligned with a curve in elevation or azimuth.

destination. The directions are repeated, separated by a short silence for segregation, and are updated as the user moves. The eight directions supported by the system are shown in Figure 5.10.

Locatones consists of three or four notes with distinctive rhythms, for easy discrimination of cardinal and diagonal directions. Ascending and descending melodies played at equal loudness in the two ears, point north and south respectively; and repeated notes played in the left or right ear point west and east respectively. Mixtures of pitch and stereo balance point in diagonal directions. Quiet reverberated locatones at a slower tempo indicate that the target is far. Increasing loudness and tempo occur as the distance to the target decreases. The locatone distance function is explained in Figure 5.10.

Two things occur upon reaching a locatone-defined target: the name of the target is spoken and the locatones are muted or, the locatones lose their reverberation and the repeat rate is relaxed, which signals to the user that the stylus is aligned with a curve. The calm locatones point out the local direction(s) of the curve so the user can begin tracing it without guessing. Further details about user interaction are provided in Chapter 6. Locatones, far and near, are available online for listening[8].

[8]http://www.cgl.uwaterloo.ca/∼m2talbot/SAMsDemos/

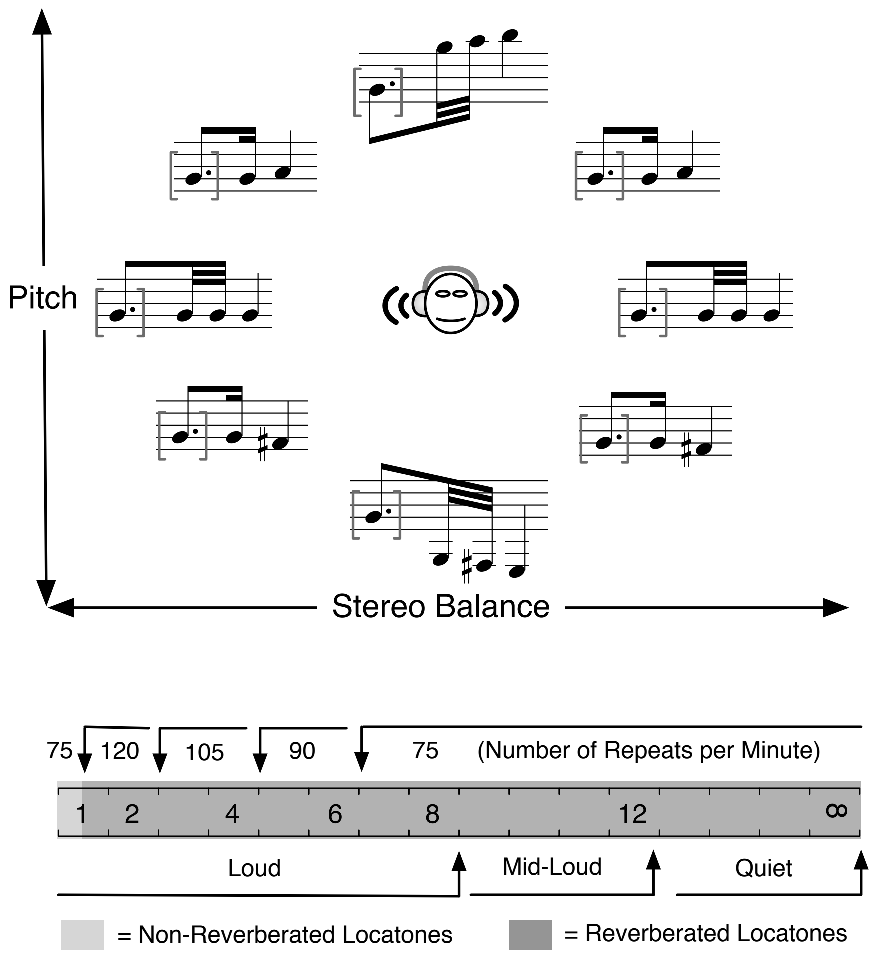

Figure 5.10: *The upper graph:* shows the set of locatones pointing in eight directions: 0 ° (east), 45 ° (n-e), 90 ° (north), and so on. The note G4 surrounded by square brackets forms the first part of every locatone. This part is always binaurally centred. The notes that follow, which form the second part, are either panned left-most (west), mid-left (northwest, southwest), mid-right (northeast, southeast), right-most (east) or remain centred (north, south) depending on a target direction. *The lower graph:* the values above and below the ruler represent repeat rate and loudness. The product of the sensor's width (1.4 cm) and a distance unit [1, 2, 4, 6, 8, 12, ∞] determines which locatone is played to the user. SAM implements two types of locatones, one treated with reverberation and the other without. For example, being at 6 distance units from a target plays a loud reverberated sequence at a repeat rate of 90 rpm. Being at less than 1 distance unit from a target plays a loud non-reverberated sequence at 75 rpm.

5.6 Summary

SAM implements two types of directional cues: open and constrained. Open cues give the user total freedom as how to reach a destination in open space. Locatones are an open cue. Constrained cues constrain the user to follow a well-defined trajectory to reach a destination. Curve tracing cues are a constrained cue.

After Experiment 2 the curve tracing cues were tuned to address the points discussed in Section 5.5. Both curve tracing and locatone cues adopt the same model, mapping elevation to pitch and azimuth to stereo balance.

Chapter 6

System Integration

As might be expected, the implementation of the system includes a final refinement of its design. This chapter describes how the SAM prototype was implemented, including the preprocessing algorithms that convert hand-drawn maps into spatial auditory maps, the architecture, the user interaction and the system response. It also points out features introduced or changed during the implementation.

6.1 Implementation

The SAM prototype is a Java application using open source components for sound playback, user output and synthetic speech. A sonic version of OpenGL, JoAL [281] (a Java API that shares the style and conventions of OpenGL), manages all aspects of sound playback. Stylus input is captured by the Java Pen tablet access library JPen [282],and speech is synthesized by the text-to-speech API FreeTTS [283].

The SAM prototype has the cartographer draw maps by hand using the stylus, annotating POIs and curves with a computer keyboard. The reading interface preprocesses the cartographer input, removing noise in the hand-drawn input and reducing computational resources at run-time, as explained in the next section. Then follows a discussion of the architecture of the reading interface.

6.1.1 Preprocessing

Preprocessing steps transform all curves to a standard resolution, with noise filtered out, extracts points of interest (POIs), intersections, curvature and extremity areas; transforms annotation strings into synthesized speech; tags intersection and POI areas on a grid; stores all the transformed curves and extracted points into curve and point lists; and builds a data structure for the curve collision algorithm. These processing steps are detailed below.

Curve Interpolation

First, all duplicates of points on the curve are removed. Then the curve is smoothed using a moving window. A smoothed point p_k^* is the average of an odd number of consecutive $2w+1$ ($w = 1, 2, 3, ...$) points of the original data points p_k, p_{k+1}, p_{k+2}, $\ldots$, as shown in Equation 6.1. The first two and last points of a curve of n points are preserved, which makes the domain k an element of $(3, n-2)$. The odd number $2w+1$ is the width of the filter; the greater the width, the smoother the result. A filter of width $w = 5$ was used to remove noise but preserve a visual resemblance between the original curve and the resulting one.

$$p_k^* = \frac{1}{2w+1} \sum_{i=w}^{i=-w} p_{k+i} \tag{6.1}$$

Curvatures

Areas of curvature are extracted from every curve in the map using an algorithm similar to the Gaussian Curvature algorithm [284]. The Gaussian Curvature algorithm finds the Gaussian curvature κ at any point on a curve by iterating over a distance parameter until results stabilize. The algorithm implemented in SAM, which solves for a relaxed problem that does not need to find κ, extracts closed-form solution of curvature that are tighter than a predefined κ value and ignores the rest.

The algorithm consists of moving a circle of diameter κ (the width of the sensor) on a curve. The circle remains centred on the curve, which makes the distance between the two contact points of a straight curve and the circle diameter equal. Intersecting a curvature makes the distance between the two contact points shorter than the circle diameter, which indicates a curvature. The shortest distance in a curvature area reveals the curve locus[1], the central point of a curvature, as shown in Figure 6.1. The contact points, which I call angular points, are used by SAM to provide directions to a user at tight curvatures.

Extremities, POIs and Intersections

The first and last points of a curve are its extremities. They are obtained directly, as are POIs.

[1] Wiggles at too high frequency, like shown in Figure 6.1, are not detected correctly by the algorithm.

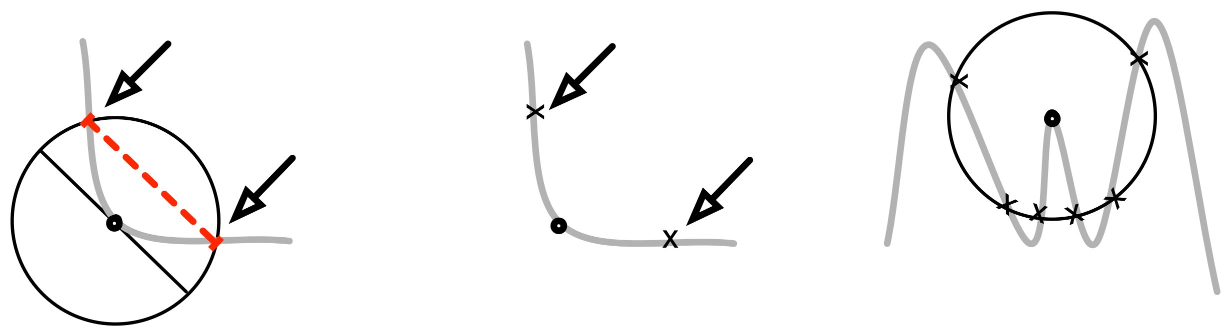

Figure 6.1: *The left diagram* shows the circle detector over a curve (grey), the circle diameter, the curvature locus, the two curvature angular points (the contact points shown by the two arrows) and the transversal line passing through the angular points (dashed), which is shorter than the circle diameter. *The centred diagram* shows the results for this area of curvature: a locus (dot) and two angular points (X). *The right diagram* shows an example of a case where the algorithm fails to detect a curvature correctly. In this example, the resolution of the map is too high and would need to be lowered.

For n curves on a map, $\frac{n(n-1)}{2}$ pairwise comparisons are needed to detect all intersecting areas. Two curves intersect if they touch one another. Intersecting areas that are separated by less than half the width of the sensor are merged[2].

Intersection detection is performed using methods of the Java Classes PathIterator and Shape. PathIterator provides a curve segment by segment. Pair of segments from different curves are tested using the intersect method of the Shape interface. The intersection locus and two angular points at equidistance from the locus of each detected intersection area are preserved to provide directions for locatones, as explained in Figures 4.6 & 4.8.

Synthesized String Information into Speech

Annotations of curves and POIs are transformed to synthesized speech sound files and stored in memory for fast access and playback control. The FreeTTS API creates a wave file containing the spoken version of the string.

[2] This heuristic does not handle all the possible cases: it is possible to find pairs of intersection points separated by less than half a sensor, e.g., Intersection A & intersection B, intersection B & intersection C, and have intersection A & intersection C separated by more than half a sensor width. The problem could be resolved by realigning the curves to meet at a unique intersection point or by increasing the resolution of the map, which would distant the intersection points and make them distinct.

Sound speech files are created when a map is loaded and stored on disk. Creating sound files on loading reduces the size of the typical maps to about 20 kB.

Grid Tag

The POIs and intersection areas are tagged with a grid identification number for the grid, column and row tools.

Lists of Features

Loading a map instantiates two lists: a curve-list and a point-list.

The curve-list holds Curve objects, each of which includes a geometric description of a street plus whatever sound files it uses.

The point-list is a collection of POI, Intersection, Curvature, and Extremity objects. A POI object is a location on the map plus the speech file that describes it. An Intersection is a location and a pair of angular points for every intersecting curve. A Curvature is a location, a pair of angular points, and a reference to the curve. An Extremity object is a pair of end-point locations with an angular point for each.

After initialization, SAM uses the curve-list and point-list to find curve/sensor and area/sensor intersections. Curve detection is performed by Java AWT, point detection by measuring the Euclidean distance between the listed points and the position of the stylus tip.

Data Structure for Collision Detection

Using the collision detection tool a user can move the stylus along the surface of the tablet and hear curves colliding with the stylus. Computation is minimized by partitioning the curves into buckets.

The map is partitioned into a matrix of small cells. The cell structure and its use is explained in Figure 6.2. In reading mode, the collision detection reads through the resulting list and compares it with stylus coordinates. This implementation cannot provide curve direction, only contact.

6.1.2 The MVC Design Pattern

Problems may arise when an interface application mixes data access, domain-logic, and presentation code. For example, application maintenance may become difficult owing to

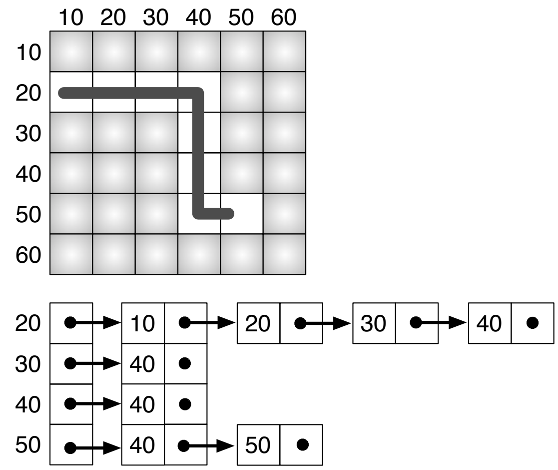

Figure 6.2: *Top:* shows the upper left area of a map with a curve. Like [73] in Chapter 2, but at a higher resolution, the map area is divided into a matrix of cells ten pixels square. A cell is preserved when a curve intersects it. *Bottom:* a list containing four keys of preserved rows (i.e., the y-coordinates) is shown in the left column. Each key points to a bucket containing the list of preserved columns.

interdependence between components, causing ripple effects when code is changed. Also, strong coupling impedes class reuse when classes depend on other classes. The Model-View-Controller (MVC) design pattern addresses such problems by decoupling data access, domain-logic, and system response from user interaction.

However, two key design constraints, the type of display and the support of mode-switching in the interface, allowing the user to keep both hands on the tablet, forced a rethink of the pattern's architectural details[3].

Controller

The SAM prototype implements the Controller as a finite state machine, which holds stylus modes, tracing context and keyboard events. Like OpenGL, a state in SAM is the exact configuration of the system at any particular time. Any tool activation in SAM

[3] The transformation of the MVC is inspired by the Immediate Mode GUI architectural pattern, an alternative pattern used by game developers to decentralize the flow (*i.e.,* to eliminate callbacks and synchronization). The pattern has the GUI drawn at every frame from the event loop and encapsulates application state, logic and behaviour into the model. Unfortunately, the pattern, which is discussed in blogs by game programmers, is poorly documented [285]. The adapted MVC is explained in the following subsections and illustrated in Figure 6.3

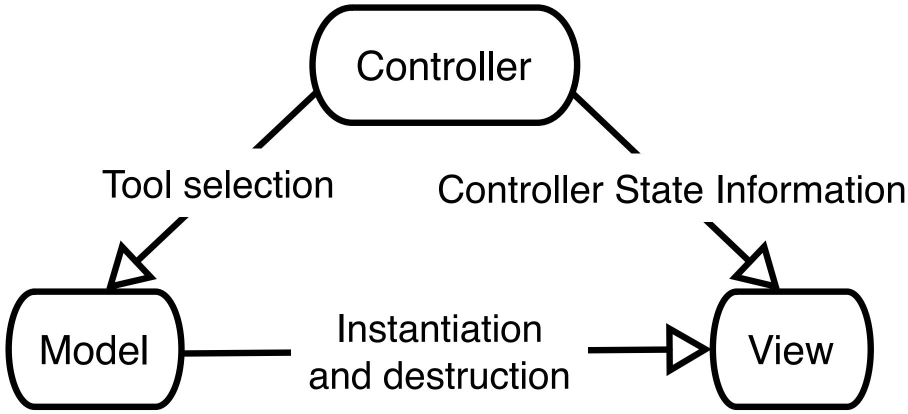

Figure 6.3: The Controller holds the system state, which it relies upon to call appropriate tools from the Model, which in turn instantiates/destroys appropriate Views, which adapt themselves to user interactions by accessing Controller state information.

is determined by a unique Controller state status. A Controller object is made of seven instance variables:

1. Stylus/Tablet contact.
2. Tracing direction.
3. Current stylus tip.
4. Stylus angle.
5. Stylus buttons.
6. Last key stroke.
7. Next-to-last and last curve contact ID.

Model

The Model is the domain-specific representation of the application logic, which describes the functional processes that mediate between user output and user input. The model is mode independent.

Tools are implemented as methods of the Model, called by the Controller when a unique set of Controller-state condition is satisfied. Once activated, a tool is responsible for instantiating a View and interrupting/destroying it when it is inactivated.

Each tool is called from the event loop, which is explicitly implemented in SAM, and can remain active throughout many iterations. A singleton design pattern is adopted to avoid creating a new instance of a view at each iteration. The SAM prototype implements singleton objects by having a single instance of the class as a static field, and limiting the scope of the constructor to private. Objects are created by calling the getMySingletonView method, as shown in the code sample below:

```
public class SingletonView {
    private static mySingletonView = new SingletonView();
    private SingletonView() {
        // construct object here. . .
    }
    public static synchronized SingletonView getMySingletonView() {
        if (mySingletonView == null)
            mySingletonView = new MySingleton();
        return mySingletonView;
    }
   // Remainder of class definition here. . .
 }
```

View

The designer of a visual interface assumes that a visible representation, once refreshed, remains visible to the user. But an auditory presentation must be repeated to remain before the user because sounds do not persist. In addition, a user can perceive only one stream at a time. Thus, sequencing the display elements is very important.

In a graphical interfaces the state of a widget or a tool can be updated by the main process, but in an auditory display, sound information takes a long time to convey and multiple processes are required to play and stop concurrent sound events. Playing a sound is extended over time: separate process play and stop concurrent sound events. The View instantiate and manages these processes.

Upon instantiation of a View process, a collection of child processes are instantiated. These processes actually play sounds provided by their parent. The parent manages the child processes as a pool of workers, obtaining information from the Controller's finite state machine, and providing it to workers for display, according to the presentation protocol explained in Section 4.6.

6.2 User Interaction & System Response

The SAM prototype has evolved into a series of individual tools, each tool governing a well-defined domain-logic function associated with a well-defined user interaction and system response. This section explains how the tools are activated and used.

6.2.1 Prevention of Mode Errors

Mode errors occur when a user misjudges the system state, resulting in user actions that are appropriate for the user judgement, but inappropriate for the factual situation [137]. To avoid users having to diagnose and correct mode errors, SAM announces mode changes and confirms user requests. For example, speaking the name of a tool before its activation confirms (or not) to the blind user that the correct tool is activated. These auditory messages attempt to play the same role buttons pressed and checked checkboxes do for visual interfaces. Examples are shown in Table 6.1

Table 6.1: Examples of SAM announcements at tools activation.

Tools Activation	**Preceding Spoken Messages**
Wizard Tool	'Wizard tool: walking towards *X*', (*X* is the feature destination)
POI	'Location: *X*' (*X* is the POI name)
Grid Tool	'Grid Tool'
Column Tool	'Column Tool'
Row Tool	'Row Tool'
Alignment Tool	'Alignment Tool'
Front Tool	'Front Tool'

The curve tracer, notification, Details-on-demand (DoD), empty space and collision tools are exceptions because they provide immediate, distinctive and on going feedback upon activation. This distinction was made to improve the system's interaction flow [286].

6.2.2 Multiple Mode Tools

SAM prototype implements three multi-mode tools (MMTs): the wizard tool, the DoD tool, and the empty space tool, all of which have internal modes that are chosen depending on the context within which the tool was activated. For example, launching the wizard in a context were the curve trace is active will have the wizard behave differently than in a context where the front tool is active. Thus, the behaviour of a MMT is adapted to serve the tool that launches it. The system's User Environment Design diagram shown in Figure 4.2 on page 61 gives an overview of the relationship between MMTs and their respective launching tools.

Empty Space Tool This MMT sonifies 'nothing is beneath the stylus'. Stylus angle, tip and contact (touching the tablet or not) alter the sound feedback, reminding the user that the system is working but sensing nothing. Motion activates the tool.

Wizard Tool This MMT provides assistance to the user trying to regain contact with a curve or to reach a feature displayed by the grid or the index tool. Calling the wizard deactivates the sensor so that locatones are free from sound interference.

Details-On-Demand Tool This MMT provides further information to a user about a curve or a particular area.

The MMTs are further explained together with the tools that launch them in Sections 6.2.3, 6.2.4 and 6.2.5.

6.2.3 Foveal Tools

The curve tracer and toggle tools with the POI, intersection, curvature and extremity notification tools support curve tracing, letting the user know about approaching areas. They are activated under the conditions shown in Table 6.2 on page 117, with the relationship between the stylus and features on the maps making the final disambiguation: the curve tracer when the stylus contacts a curve[4]; the notification tool when a notification area is nearby; and the empty space tool otherwise.

DoD at POI, intersections, curvature and extremity areas is activated under the conditions shown in Table 6.2 with the relationship between the stylus and features on the maps making the final disambiguation: DoD at POI when a POI is underneath the stylus; DoD at intersections when an intersection area is underneath the stylus, and so on. The stylus must stay motionless and not be in contact with the tablet in all cases.

The wizard is activated by the conditions shown in Table 6.2, with the relationship between stylus and keyboard interaction making the final disambiguation: If the last user interaction was a keystroke, the wizard walks the user to that location. If the last user interaction was tracing a curve, the wizard walks the user to the last contact point with the curve to help him regain contact.

[4] The timbre of the tracing tool toggles when the cursor touches a new curve.

6.2.4 Index & Key Tools

The key tool is activated by positioning the stylus over the top left corner (hot corner) of the tablet, as shown in Figure 6.4 on page 111: it speaks the name of the map, its scale, its creation date and the cartographer's name. Placing the stylus tip along the left edge (hot edge) provides a content overview by speaking the name of each feature on a map. Presentation speed is determined by the stylus position on the edge: lowest is fastest. The presentation order of features is consistent: top-left first, bottom-right last. A curve topmost and left extremity determines its location. Lifting the stylus or moving it away from the hot region mutes the key tool.

The index tool allows a user to find a map feature with the computer keyboard and reach its location with the stylus. Every POI and curve is linked to a letter key on the keyboard. Typing a key speaks the feature's annotated information and activates the index tool. As mentioned in Chapter 4, the words "nothing's assigned" are spoken when no feature is mapped. The mapping is explained in Figure 6.4.

Calling the wizard, by pressing the side button on the stylus near its sharp tip, walks the user to the feature linked with the last keystroke. The wizard does so using the framework explained in Section 5.5.2.

6.2.5 Peripheral Tools

Unlike the foveal tools, which share a single compact sensor, the peripheral tools implement sensors that vary in shape and range. In general, peripheral tools see wide swatches of the map, which makes them better for detecting geometric patterns like clusters, alignments of features, and feature-to-feature relationships.

Grid Tool The grid tool divides the map into nine rectangular cells. A single cell is active at a time, the cell beneath the stylus, which filters out about 90% of the map easing detection and localization of clusters of POIs and intersections. The controller state for activating the tool is shown in Table 6.2.

Figure 6.5 provides a scenario showing how the grid tool works. Two POIs are in a cell, and two vocoder-integers displayed at equal but low pitch, with one vocoder-integer in each ear: the POIs are south and on each side of the stylus. Moving the stylus inside the cell changes the pitch, stereo balance and loudness to correspond to each new stylus position. Moving to a new cell triggers a grid-frontier sound at the border and speaks the ID number of the new cell.

A notification sound is played when the stylus encounters a feature, which tells the user that DoD is available for that feature. If there are no intersections or POIs inside a cell,

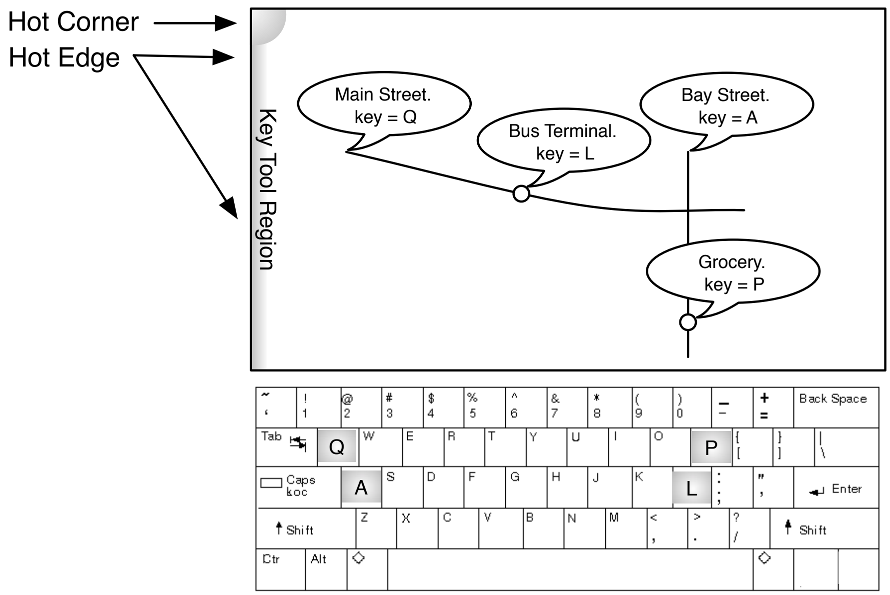

Figure 6.4: The activation regions (Hot Corner & Hot Edge) for the *key tool* are shown on the left side of the figure. Keyboard mapping for the *index tool* is explained: the left keys on the keyboard are linked to the left curves on the map and the right keys to the right POIs, hence the curves, which are distributed on the left, grow towards the right. The POIs do the opposite. The current system can link a maximum of 26 features. In situations where two different features share the exact same location in azimuth, the highest one has priority. Looking at the example in the figure, a user can expect the POI in L (bus terminal) to be on the west side of the POI in P (grocery). Furthermore, the user knows that these features are POIs because of their location on the keyboard and because the word 'location' is prefixed to every spoken POI annotation, e.g. 'location: grocery'.

then the empty space tool is activated. The state conditions for the DoD and the empty space tools are shown in Table 6.2.

Activating the wizard from the grid tool is like calling the index tool on a POI or intersection area but the numeric keypad is used instead of the character area. In response to a keystroke on the numeric keypad the wizard says the key ID and then the name of the feature linked to that number. If no feature is assigned, SAM says "nothing's assigned". Pressing the stylus side button near the sharp tip walks the user to that feature. At an intersection, a full description of the intersecting curves is provided using the presentation framework explained in Section 4.6.4. Annotated information is spoken upon reaching a POI destination.

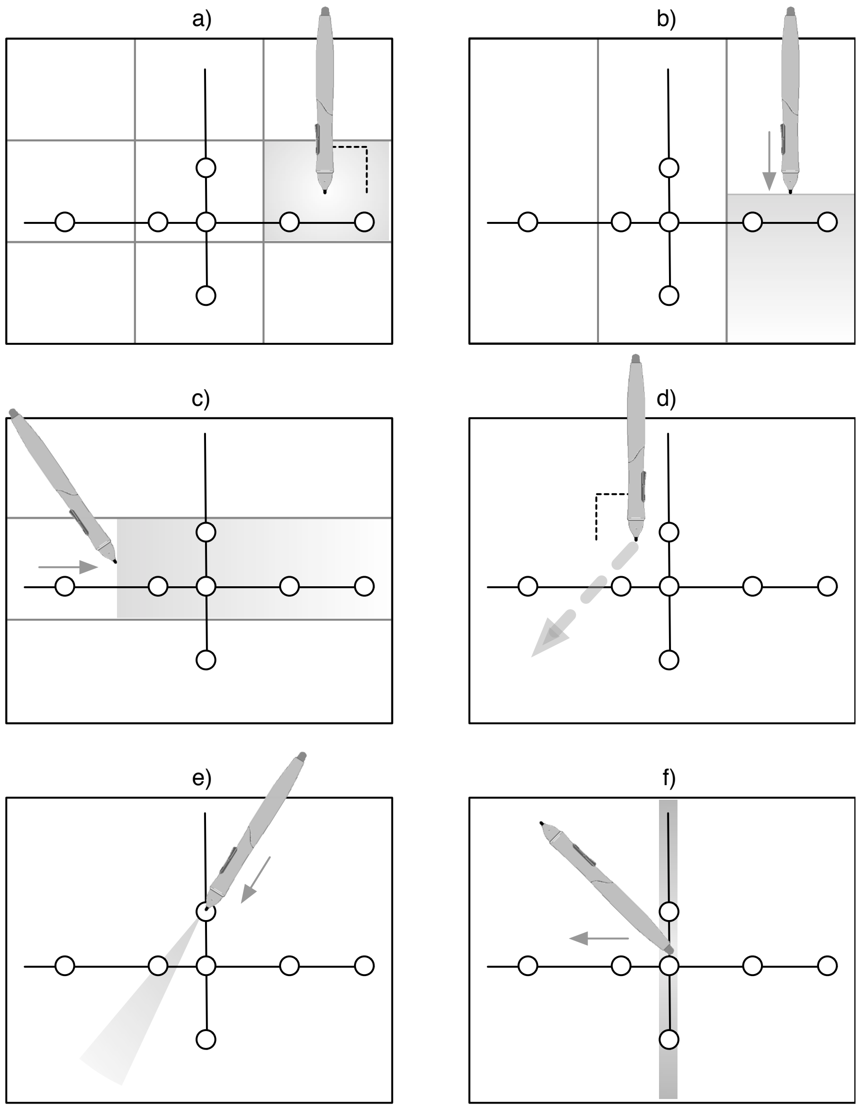

Figure 6.5: The layout of two intersecting curves and seven POIs are used to explain the peripheral tools. The grey areas represent the active regions of the sensor. The dashed corners in a) and d) indicate that the stylus is orthogonal to the tablet; otherwise the stylus is tilted. Example a) shows the grid tool; b) shows the column tool; c) shows the row tool. In a), b), and c) the stylus button near the eraser tip must be held to activate the tools. In b) and c) the sensor starts from the stylus location and span towards the outer edge of a map. Example d) shows the collision detector. The dashed arrow indicates that the sensor's active region follows the stylus displacement; e) shows the front tool; f) shows the alignment tool, which is activated by turning the stylus upside down and tilting it. Leaving the stylus orthogonal to the tablet (not shown in the figure) launches the POI alignment tool, which only displays POI and filters out everything else.

Colomn & Row Tools The column and row tools divide the content of a map into three rectangular regions, as illustrated in Figure 6.5. The controller state conditions shown in Table 6.2 activate one or the other tool, disambiguated by tilt direction of the stylus: a vertical tilt activates the column tool, a horizontal tilt the row tool.

The vertical bounds of the map are invariable for the row tool and the horizontal bounds for the column tool, as shown in Figure 6.5; the active region runs from the stylus tip to the edge of the map in the direction of the stylus tilt, so as to coincide with the pointing direction of the stylus. In example c), the stylus points to the right and the active region runs from the stylus position up to the right edge of the map.

Example b) of Figure 6.5 shows how the column tool responds to stylus displacement: moving the stylus from downwards shrinks the active region downward. As the stylus approaches the two POIs inside the column, both their loudness and pitch increase. When the stylus is at the elevation of POIs, their pitch is G4 and loudness maximized. When the stylus passes them, they disappear from the display. Thus the user can find the location of any feature on a map by alternating between the column and row tools, alternating between vertical and horizontal tilts. The wizard and DoD are accessible from the grid tool by making the stylus orthogonal to the tablet. An empty column or row activates the empty space tool as shown in Table 6.2.

Zooming with large perimeter Neither audition nor touch naturally affords 'zooming in' on a region to obtain more detailed geometry or 'zooming out' to get an overview of the map. The grid, column and row tools attempt to compensate for this lacuna by showing features located outside a zoom-in region, while providing richer information about features inside the region. This approach, which is similar to data visualization methods developed for vision, e.g. [287], delimits a region where the features are magnified, with features outside the region projected to its edge, as shown in Figure 6.6. Like the curve sensor, explained in Chapter 5, the zoom-in region is divided into fifteen azimuth and elevation levels, making it sensitive to stylus displacement. The diameter of the zoom-in region is set to seven centimetres, five times the diameter of the curve tracing sensor[5].

Collision Detection Tool The collision detection tool allows a user to find a curve on the map by moving the stylus across the tablet. Example d) of Figure 6.5 shows how the collision tool responds to stylus displacement.

SAM tries to match the y-coordinate of the current location of the stylus with a y-coordinate key in the table, as explained in Section 6.1.1. When it finds a match, the list

[5] I determined this value by testing the tool on myself. An empirical study on blind participants is needed to examine different values, which I keep for future work.

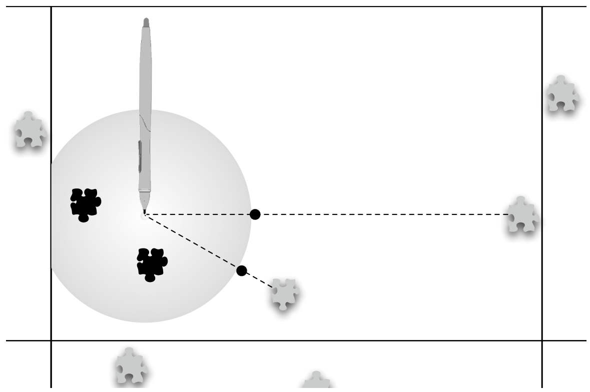

Figure 6.6: This figure explains the zooming technique applied on the grid, column and row tools, using the grid tool to illustrate the idea. The stylus is inside cell two (top row middle column position), represented by the rectangular frame in the figure. The large circle, centred at the stylus tip, is the zoom-in region. *Outside cell two:* the four features outside the active cell are filtered out by the grid tool. The left area of the zoom-in region is truncated for the same reason. *Inside cell two:* the two features outside the zoom-in region (grey) are projected (dashed lines) to the edge of the circle (black dots). The two features inside the zoom-in region (black) are displayed without transformation. *System response:* pitch, balance and loudness vary with small stylus displacements for features inside the zoom-in region, i.e., the closer a feature to the centre of the zoom-in region, the more it changes sonically. The loudness of features outside the zoom-in region is set to a constant intensity, determined by the radius of the zoom-in region, which guarantees that every feature inside an active cell can be heard. Because distal features vary less than closer ones, they draw less attention during displacement. Participants wanted this feature because things far away are less urgent but should still be perceived.

of items linked to the key is compared to the x-coordinate of the stylus position. A pulse-presence sound is heard on a match, informing the user that a curve is underneath. This contact point is preserved by the system, allowing the user to call the wizard, on which the local directions of the curve are displayed by non-reverberated locatones, as explained in Section 5.5.2.

The controller state conditions that calls the collision detector is shown in Table 6.2. Stopping on a curve provides the DoD, as shown in Table 6.2. The empty space tool is activated when the stylus is not intersecting a curve in the state condition shown in Table 6.2.

Front Tool The front tool senses space through a 'field of vision' technique, which allows a user to look ahead and far away from any location on a map. The opening angle of the sensor beam is fixed[6] to 15°. The controller state that activates the front tool is shown in Table 6.2. Figure 6.5 shows that tilting the stylus sends the sensor beam in the direction of the tilt with the beam widening, but not in term of angle, as distance increases. Pulse-presence sounds encoded in pitch, stereo balance and loudness indicated the presence of features in the beam. Lifting the stylus, as shown in Table 6.2, calls the DoD tool, which speaks the annotation of features sequentially as explained in Figure 4.8. The empty space tool is activated when the beam contains no features and the state agrees with Table 6.2. This tool allows a user to perform a circular scan by rotating the tilted stylus while remaining at a location.

Alignment Tool The alignment tool senses space vertically or horizontally, as shown in Figure 6.5. Stylus tilt determines the direction of the beam; east-west and north-south tilts set the beam in column or row mode respectively. The beam has the width of the curve tracer. Pulse-presence sounds inform the user when features are aligned with the beam; lifting the stylus as shown in Table 6.2 activates DoD, which presents the information using the framework explained in Figure 4.8. The empty space tool is activated when the beam intersects no features in the state condition shown in Table 6.2.

POI Alignment Tool This tool differs from the Alignment Tool in two respects. Perceptually, it displays POIs filtering out curves. To use it, the user holds the stylus orthogonal to the tablet and the displacement direction is used to determine the angle of the beam; horizontal or vertical directions set the beam in column or row mode respectively. See Table 6.2 for further details.

6.3 SAM's Affordances

The SAM prototype follows the second and third modality-independent principles proposed by Metatla & Stockman [121]; the user is able to 'read' a route in either directions and access any landmark from any route connected to it. In addition, the peripheral tools allow a user to jump to any landmark or route on a map.

[6] I determined this value by testing the tool on myself. A larger angle captured too many features. An empirical study on blind participants is needed to examine different values, which I keep for future work.

6.3.1 Foveal Tools

The foveal tools encourage a reading strategy in which a reader learns streets and discovers POIs simultaneously while passing along a route. The nature of curve tracing requires the user pass through intermediate points before reaching the destination. To make a sense of the information flow, the user must methodically follow a street, as occurs when navigating space, but more safely and quickly than walking in the environment.

6.3.2 Peripheral Tools

The peripheral tools allow the user to explore distant space by jumping to any intersection, POI, or street. In addition to supporting teleportation, the tools encourage the formation of exocentric geometry by conveying feature-to-feature relationships via alignment, proximity and perimetry.

6.4 Summary

Modularity and inheritance were used to enhance maintainability and code reuse. The adaptation of the MVC pattern lowered dependencies and eased the design process, allowing the addition and removal of layers of sounds, and the reworking of presentation details without having to consider interdependencies in the display.

Table 6.2 shows the rich palette of outputs a stylus device offers to a system such as SAM. Stylus pressure output is not monitored by SAM so that blind users can carve maps on styrofoam, by applying pressure on the stylus when tracing curves and encountering POIs. In its current version, the SAM prototype uses less than 28% of possible stylus-output attributes, because most tools are activated by placing the stylus tip on the tablet, to avoid having to flip the stylus while reading a map.

As pointed out by Nielsen [288], avoiding recall in favour of recognition is a key usability principle in interface design. Unfortunately, the transience of sound makes it challenging to comply with this principle in auditory displays, which probably explains why software technology for the blind relies so heavily on recall, e.g., [27, 25], something blind people are accustomed to.

Stylus interactions in SAM must be memorized. I observed participants confusing the actions needed to activate a tool as the number of tools increased. Post-interviews with the participants involved in the design of the prototype suggested that a short training session could resolve the problem. However, the amount of training remains to be determined, a point examined in Chapter 7 where seven naive participants are taught to use the system, as a method for learning unknown environments.

Table 6.2: This table shows the necessary state conditions for activating tools. 'Button State 0' means that no buttons need to be pressed, 'Button State 1' corresponds to pressing the button near the sharp tip, 'Button State 2' corresponds to pressing the button near the eraser tip. N/A are unused states and leave SAM silent. Wizard1 allows a user to reach a target selected from keyboard input like the index or the grid tool. Wizard2 allows a user to hear the direction of the last contact point with a curve. In column 'Stylus Tip': 'S' stands for Sharp tip and 'E' stands for eraser tip. In column 'Stylus Angle': 'I' stands for straight up (orthogonal to the tablet surface) and '/' stands for tilted. In column 'Stylus Displacement': 'Mv' stands for Moving and 'St' for stationary. In column 'Tablet Contact': 'T' stands for Touched and 'NT' stands for not touched.

Stylus			Tablet	Button State		
Tip	Angle	Displacement	Contact	0	1	2
S	I	Mv	T	Curve/Notification/Empty	Wizard1	Grid/Empty
S	I	St	T	N/A	Wizard2	Grid
S	I	Mv	NT	Collision/Empty	N/A	N/A
S	I	St	NT	DoD: Curve/Collision/Notification	N/A	DoD: Grid
S	/	Mv	T	Front/Empty	N/A	Col/Row/Empty
S	/	St	T	Front	N/A	Col/Row
S	/	Mv	NT	Front/Empty	N/A	N/A
S	/	St	NT	DoD: Front	N/A	DoD: Col/Row
E	I	Mv	T	Alignment(POI)/Empty	N/A	N/A
E	I	St	T	Alignment(POI)	N/A	N/A
E	I	Mv	NT	N/A	N/A	N/A
E	I	St	NT	DoD: Alignment(POI)	N/A	N/A
E	/	Mv	T	Alignment/Empty	N/A	N/A
E	/	St	T	Alignment	N/A	N/A
E	/	Mv	NT	N/A	N/A	N/A
E	/	St	NT	DoD: Alignment	N/A	N/A

Chapter 7

Experiment

This chapter experimentally compares SAM to tangible maps (TMs), the best technology available to teach unknown areas to a blind traveller. The assumption in this experiment is that the method used to convey spatial information has an impact on the quality of the participants' mental maps. A physical reproduction of the participants cognitive maps, which was rebuilt by every participant using self-adhesive strings covered with wax, is examined to measure the systems. The acquisition rate of spatial information and the level of autonomy of both systems is also examined.

7.1 Method

Most critical to the assessment are the exact methods by which the maps are created, which are described below, followed by details about the participants, the experimental procedure, the results and discussions.

7.1.1 Tangible Map

I did not have the resources to build a TM and judge its legibility value, so I looked for an existing one. I focused on three points to make a selection: the expertise of the cartographer, the site the map described, and the map's construction date, which could not be too recent so I knew the map had already been read by blind persons.

On the advice of Lesley MacDonald, national coordinator on mobility & orientation for the Canadian National Institute for the Blind (CNIB) and a leader in accessibility for visual impairment, I chose a plastic map[1] of the National Gallery of Canada [289].

Had I had access to the CNIB tactile department, which had though been closed due to budget cutbacks, I could have asked for the construction of a new TM to display the SAM environment, which would have allowed me to cross-examine the two environments with the two systems–four maps, two environments per system. Unfortunately this was not possible. Instead, I was compelled to inspire the design of the spatial auditory map (SAM) based on the TM I was given, which compounds the study because it is hard to make sure the level of difficulty between the two environments is equal.

7.1.2 Map Transformations

The rooms on the ground level floor of the museum, which is the area I used for the experiment, were relabelled with generic store names in an attempt to reduce memory load, transforming the museum into a mall. This eliminated uncommon names found in the museum and replaced them with store names participants were familiar with. Only few original names were preserved, those that were fully written in Braille on the map, namely: elevator, ramp, lobby, cafe, auditorium, bookstore, and washrooms. All the other rooms labelled with numbers were renamed. This modification was unnoticed by participants.

7.1.3 Spatial Auditory Map

The map of another mall, which would be read with SAM, needed to be built. Both maps would be read by the participants, so their environments needed to be distinct but similar in complexity.

The construction was achieved by altering the spatial distribution of the rooms and the overall shape of the building, while reusing the same geometric outlines, preserving the same overall surface area, and rebuilding the same number of rooms. The resulting transformation in the top right, which would be read with SAM, and the original map in the top left, which would be read tactually, are shown side-by-side in Figure 7.1. The lower part of the figure shows the same malls redrawn in SAM's format, as explained in Chapter 3.

Different store labels were chosen to reduce confusion between the TM and the spatial auditory map, but store-to-store connotations were developed to balance them. The resulting relationships are shown in Table 7.1.

[1] Plastic is, from a perceptual standpoint, the best medium available for TMs. A plastic map affords greater variations in relief and sharp edges for maximum stimulation and legibility [146, p.374].

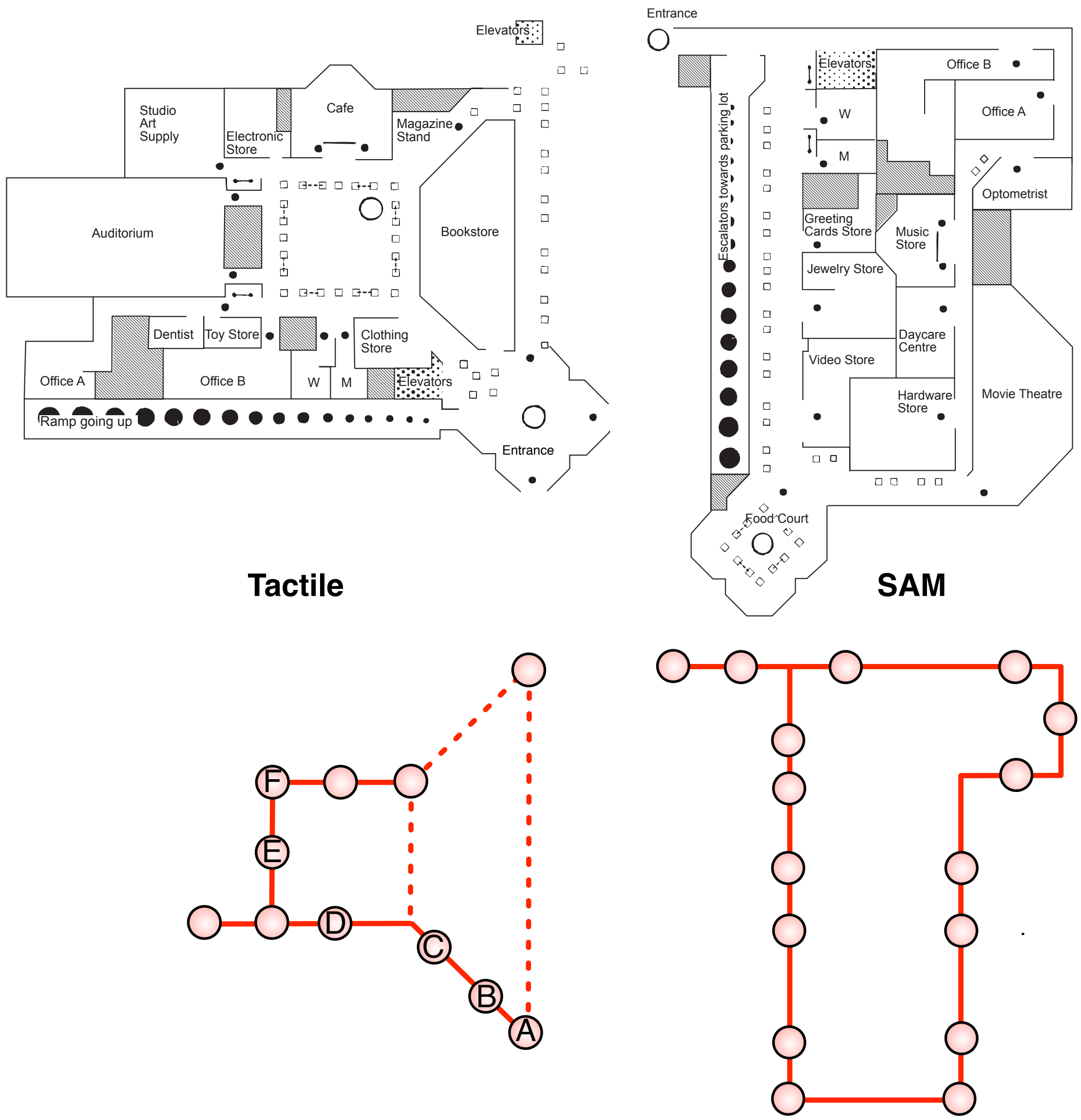

Figure 7.1: The layout of the two malls are shown side-by-side. The mall represented in the TM is on the left, the mall represented by SAM on the right. *Upper layouts*: the greyed regions are restricted areas, the small squares are pillars, the small dots are doors and the large circles are reference points. *Lower Layouts*: the edges represent the corridors a blind traveller has to take to visit every landmark. The dashed edges are optional corridors because not all of them need to be taken to visit all the landmarks. The nodes are entrance locations. Letters inside the nodes designate tuples of proximal landmarks, as shown in the original map above. The mall entrance and the ramp are near 'A'; the clothing store entrance and the elevator are near 'B'; the women's and men's washrooms are near 'C'; the toy store, dentist and office A entrances are near 'D'; the two auditorium entrances are merged together at location 'E'; and the art store and electronic store entrances are near 'F'.

Table 7.1: This table shows the store names chosen for the tactile (left column) and spatial auditory maps (mid column). The right column describes the relationship between them.

Tactile	SAM	Relationship
Art Supply	Music Store	Arts
Auditorium	Movie Theatre	Entertainment
Bookstore	Video Store	Entertainment
Electronic Store	Hardware Store	Functional
Clothing Store	Jewellery	Aesthetics
Toy Store	Day Care Centre	Children
Cafe	Food Court	Restaurants
Magazine Store	Greeting cards	Stands
Dentist	Optometrist	Health
Ramp	Escalator	Inclined Stairways
Office A	Office A	No Difference
Office B	Office B	No Difference
Washrooms	Washrooms	No Difference
Elevator	Elevator	No Difference
Entrance	Entrance	No Difference

7.1.4 Participants

Like most disabled persons, blind persons are vulnerable to abuse and are rightfully hesitant to participate in empirical studies, hence the difficulty of finding disabled participants [290]. Fortunately, the CNIB Kitchener/Waterloo office could provide me a list of twelve blind participants who agreed to meet with me or be called home to discuss the study. Among them, a total of seven participants met the study requirements and agreed to participate.

The participants needed to be adults, have no functional vision for ten years or more, have previous experience with TMs, have the ability to hear pitch, report normal hearing, have no cognitive disability, and agree to participate in a multi-session study. The latter point was problematic: the multi-session format was viewed by the candidates who refused to get involved in the study as too much commitment. In an attempt to entice participants to participate, they were offered the choice of meeting in the usability lab of the HCI department, in the conference room of the CNIB or at their homes, where I would drive with the material. Among the group of participants who agreed to participate, some could not commit themselves to more than five sessions, which constrained the number of possible meetings for the entire group. Six participants preferred to meet at their homes and one preferred the CNIB. Each session was conducted one subject at a time, which means that thirty-five sessions needed to be scheduled and attended by the researcher.

The group of participants was composed of one unemployed person (B1), one student (K1), one retired programmer (K3), one retired office employee (K2), one office employee (W1), and two professional consultants on accessibility and technology (M1 and D1). Details about the participant are shown in Table 7.2.

The study received ethics clearance through the Office of Research Ethics at the University of Waterloo. The participants were paid a total $85 for their participation in the study and for filling in a post-study questionnaire.

Table 7.2: This table shows the participants' profile. All the participants had attended a specialized school for blind persons where they learned to read TMs and Braille. Onset stands for age of onset - the age the participants became blind, RP stands for Retinitis Pigmentosa, RoP for Retinopathy of Prematurity and the star symbol (*) identifies the participants that play a musical instrument.

ID	Gender	Onset	Age	Cause	Travel Skills	Mobility Aid	Residual Vision
B1*	Female	2	50	Cancer	Weak	Guide Dog	None
D1	Female	1	55	RP	Good	Guide Dog	Light
K1*	Female	0	30	RoP	Good	Guide Dog	None
K2	Female	2	60	Cancer	Good	Guide Dog	None
K3	Male	7	65	RP	Good	Long Cane	Light
M1	Male	4	40	Cancer	Good	Guide Dog	None
W1	Female	0	55	RoP	Weak	Guide Dog	None

7.2 Experimental Procedure

The multi-session is divided into four sixty-minute tutorial sessions preceding a ninety-minute experiment session, every session interleaved by a day or two of rest: evidence suggests that resting brain activity help trainees consolidate what they have learnt [291]. Table 7.3 shows the meeting schedule.

Table 7.3: Five meetings distributed over two weeks.

Week 1			Week 2	
Monday	Wednesday	Friday	Monday	Wednesday
Tutorial 1	Tutorial 2	Tutorial 3	Tutorial 4	Experiment

Tutorial instructions are written into text files and read by a speech synthesizer to the participants. The text files are available in Section A.3 and summarized below.

The experimenter (and author of the thesis) takes the experiment to the living place of the participants (or a destination of their choice), a necessity to attract participants in the study. The sessions are conducted in a quiet environment, commonly in the participant's kitchen. The experimenter and the participant sit around the kitchen table. A typical training session consists of driving to meet the participant, paying the participant, assembling the apparatus on a table, reading the instructions and asking the participant to carry out a series of practice trials. The two latter activities are carried simultaneously, as explained in Section A.3. The participants are authorized to ask questions at any time.

The experiment procedure is identical to the tutorial sessions, with the exceptions that a video camera pointing towards the hands of the participant is installed and experiment procedure is explained. When finished, the participants are offered a financial incentive ($10) for filling a post-questionnaire.

Two orientation & mobility specialists (OMSs) and employees of the CNIB in Toronto post-grade the participants' responses. The primary function of an OMS is to instruct blind and visually impaired individuals with safe and effective travel, which makes them ideal candidates for judging the participants' cognitive maps because they are familiar with spatial distortion and common errors found in the cognitive maps of blind travellers.

The remaining of this section provides further details about the apparatus, the tutorial and experimental sessions and measures.

7.2.1 Apparatus

Researchers in spatial cognition have proposed a variety of tabletop methodologies to examine the properties of mental maps in subjects [34, 64, 98]. Thus, there was no need to send the participants into the environment to test their spatial understanding. The participants stayed seated throughout the entire study.

The same tablet, stylus, computer, and headphones as described in Chapter 5 were used for the study. Half-centimetre thick Styrofoam panels, which were laid on top of the tablet's surface and which did not affect communication between stylus and tablet, were engraved by the tip of the stylus during the trials. The panels served two purposes: first, the participants could review with touch the curves they engraved and second, the panel were post-examined by the researcher to identify the difficulties encountered by the participants, which helped adjusting the tutorial sessions along the way. All the spatial auditory maps had a surface area of 22 x 28, the same area as the TM, to preserve scale.

7.2.2 Tutorial Sessions: Preparation for the Study

The participants are naturally biased in favour of TMs because of their earlier experience with them. I tried to remove the bias by training the participants to use SAM, but four hours of training is a short time to teach a technology like SAM to a naive user. I had to take into consideration that an ecological environment is complex, that the experiment reading time is short and that all the participants had a lifetime of experience with TMs but none with SAM. Hence, it was essential to prepare the participants for the type material they would be tested on.

In practical terms, the participants needed to learn how to trace corridors that turn in any direction, begin and end, intersect, and are overlaid by POIs and conjunctions of things like POIs, extremity of curves and intersections. The foveal exploration tools were taught first and the peripheral exploration tools later, because the foveal tools must be understood to allow a participant to trace a corridor. Every tutorial session needed to cover new material but also reuse what was taught *hitherto* to make sure the participants preserved and reinforced the skills they learned in previous sessions. The four tutorials are summarized below and available in Section A.3.

Tutorial 1

The participants were told they were attending a party and were standing in the middle of a room filled with famous persons who have/had visual impairment, past and present (e.g., Louis Braille, Stevie Wonder, Claude Monet, Galileo Galilei and so on). The surface of the tablet was a scaled down representation of the room. Their task was to meet everyone in the room. Their starting position is represented by the star symbol in the upper left of Figure 7.2 on page 126.

To succeed, the participants needed to use the index tool to find a person, and then use the wizard to reach that person's location. Once there, the participants were asked to punch a hole in the Styrofoam sheet with the stylus and then begin a new search from the starting position, which they could perceive tactually. On completion of the tutorial the participants could read a Braille message they have written themselves saying 'you rock'.

The aim of this tutorial was to familiarize participants with the use of the key tool, the detail-on-demand tool, the wizard tool and the emptiness detection tool.

Tutorial 2

The participants were presented with a map filled with curves they had to trace. Their task consisted in finding a particular curve using the index tool, reaching it with the wizard

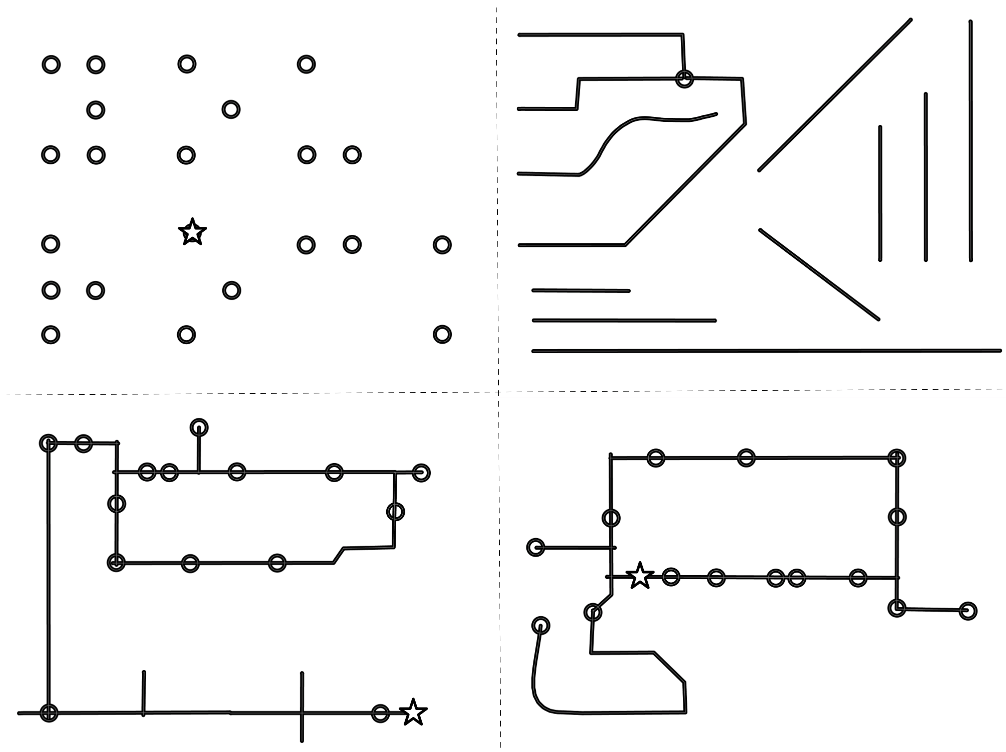

Figure 7.2: This figure shows a visual representation of the tutorial sessions. The dashed lines separate the tutorial maps. Tutorial 1 is in the upper left corner, tutorial 2 is in the upper right corner, tutorial 3 is in the lower left corner and tutorial 4 is in the lower right corner.

tool and tracing it in both directions. They were asked to punch a hole at each extremity and verify by touch that the curve's orientation and length matched their impression. A special meaning was assigned to every curve to make the task less abstract. For example, the three vertical curves shown in the upper right of Figure 7.2 represented the scaled height of the Empire State Building, the CN tower and the Burj Tower in Dubai, which they could compare by touching the holes they punched at the extremity of each curve.

The aim of this tutorial was to familiarize participants with the use of the tracing tool, the wizard tool (to regain contact with a lost curve), the curve detection tool, and the POI, curvature and extremity notification tools.

Tutorial 2 needed to be interrupted to customize polarity of curve tracing cues in elevation for some participants. This unanticipated adjustment, which was effective immediately and preserved throughout the remaining of the study, is discussed in Section 7.2.2.

Tutorial 3

The participants were presented a map of a hotel and five streets. They were asked to find the starting location on the map and then trace their way up to the hotel and visit its inside. This part of the tutorial was completed upon a full visit of the hotel. The star in the lower left of Figure 7.2 represents the starting location. The second part of the tutorial consisted of using the participants' knowledge of the environment to teach them the alignment tool. To practice, the participants were asked to detect the alignment of POIs along corridors.

The aim of this tutorial was to familiarize participants with the use of the intersection, toggle notification and alignment tools. Unfortunately, the participants indicated not having enough time to sufficiently practice the material covered in this tutorial, and consequently found it difficult. These comments had an impact on the design of Tutorial 4.

Tutorial 4

A map similar to Tutorial 3 was constructed to address the difficulties mentioned in the previous tutorial. To save time, the participants were asked to start at a given position they could perceive by touch, represented by the star in the lower right of Figure 7.2, and then visit every location. The participants were interrupted at random times and asked to describe the shape of the routes, the location of POIs in space, and so on. These interruptions were put in place to motivate them to pay close attention to the things they perceived. The second part of the tutorial consisted of using participants' knowledge of the environment to teach them the grid, column and cell tools. To practice, participants were asked to detect and then walk towards POIs and intersections without following any curves or using the index tool.

The aim of this tutorial was to reinforce the material covered in tutorial 3 and familiarize participants with the use of the grid tool family. The front tool was not taught to the participants due to lack of time.

Customized Polarity for Curve Tracing

Probably the biggest surprise in this study was the contradiction in the elevation polarities I observed in participants during the tutorial sessions.

Initially, SAM's directional sound cues were adjusted as indicated in Chapter 5, that is, high pitch was meant to inform the user that a target was higher and low pitch that a target was lower. The same polarity was consistently encoded for curve tracing and locatone cues. This setting worked fine for the locatone encodings for all participants, but five participants preferred an inverse pitch polarity for curve tracing during the tutorial sessions.

This phenomenon struck me during the second tutorial when I realized that a participant who had a good musical ear, and who was excellent at reaching targets with locatone cues, experienced great difficulty tracing horizontal curves. Inverting the pitch polarity of curve tracing cues resolved the problem. I also tried doing the same thing with the locatone cues, but reaching targets in elevation became counterintuitive for that participant. Inverting pitch for curve tracing and leaving locatones untouched made elevation cues inconsistent, which is what the five participants wanted.

Four other participants, all of whom also had musical backgrounds, benefited from the same inversion. Tutorials 2, 3 and 4 and the experiment were adjusted to match these participants' preferences.

This phenomenon was not observed in Chapter 5, but only one person among the group of participants reported musical experience (a singer). I leave further investigation of this phenomenon for future work.

7.2.3 Experiment Details

A within-subject experimental design was chosen to address error variance associated with individual spatial abilities, which vary significantly among blind persons [43] and because blind participants are hard to find, as mentioned in Section 7.1.4.

The experiment was divided into two blocks of forty five minutes each, separated by a short break. Learning new environments from TMs and SAM takes time, which is why only four maps were presented to the participants: two small practice maps and two assessment maps.

Each block was assigned a map type, TM or SAM. A block began with a brief introduction explaining the task, followed by one practice trial and one experiment trial. The order of the blocks was randomly assigned to participants. An accelerated movie of the experiment shows participant # 6 reading a tactile map and then a spatial auditory map [292].

Practice Trials

The participants were presented with one of the two maps shown in Figure 7.3, given five minutes of reading time, and then asked to reconstruct the environment from memory using pre-built objects explained in Section 7.2.4. No help was provided during the reconstruction besides that described in Section 7.2.4. The same procedure was adopted for the second map thereafter.

The practice trials were designed to warm-up the participants and thus, to avoid data irregularities as they were getting acclimatized to the task.

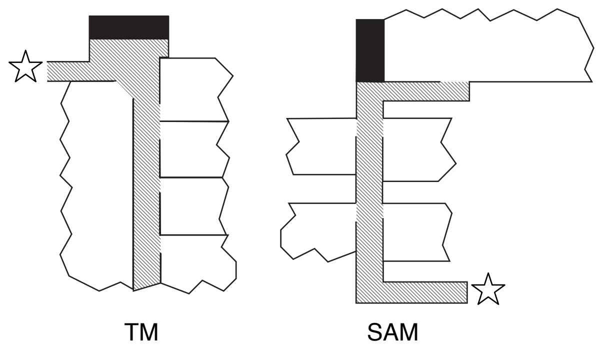

Figure 7.3: This figure shows a visual representation of the two practice maps. The black areas are elevators, the grey areas are corridors, and the white areas are rooms. The star represents users' starting positions. The rooms are not labelled because the participants were asked to recreate the layout only. The practice map is constructed from the same material and by the same cartographer as the study map. The SAM practice map is also a derivation of the TM. Both maps shown in this figure are approximately 12 cm high.

Experiment Trials

The participants were presented with one of the two maps shown in Figure 7.1 on page 121, given fifteen minutes of reading time in total, which was divided into three periods of five minutes each. The participants were stopped from reading at the end of a period and asked to reconstruct the environment from memory using pre-built objects, which they would lay on top of another board handed over to them for their responses.

They were asked to try reconstructing as many landmarks and pathways as they could, while being careful to connect landmarks to pathways, when applicable. Every reconstruction was done from scratch. No help was provided during the reconstruction besides that described in Section 7.2.4. The same procedure was used for the second map thereafter.

7.2.4 Responses

The participants were handed over a flat panel whose dimensions were the same as the TMs to avoid the need for scaling, and a set of pre-built features to reconstruct the environment. The features were built into four categories: small room entrances, large room entrances, elevators/escalators/ramps and corridors. The rooms were built into circular surfaces (small or large), the elevators/escalators/ramps into unified rectangular surfaces, and the corridors into elongated but bendable threads. Initially, larger circular surfaces were meant to represent larger stores, but the idea was dropped before the study; a last minute decision to alleviate short term memory, which was burden enough with recalling the mall layout.

The pre-built features, which can be seen in Figure A.13 & A.14 on pages 195 & 196, were built from a set of flexible self-adhesive strings covered with wax, recommended by Sabaon Weera, a CNIB specialist on independent living skills. These sticks, he explained, are often used by independent living specialists and OMSs to teach spacial concepts to their clients, who can feel them with touch, and bend them in any direction to interchange knowledge with their instructor. All participants reported having interacted with these sticks before the study.

Map reconstruction always followed the same procedure: a participant requested a specific feature (*e.g.,* a store name or a corridor), the experimenter handed over the feature requested to the participant, who positioned it on the panel and spoke its name. The reconstruction was self-paced and participants decided when they were done. A picture of the response was taken and the panel cleared for the next round.

7.2.5 Measures

The aim of the experiment is to evaluate the accuracy of internal maps participants built from interacting with the systems, to measure the speed at which they could learn spatial information and to determine how much external help they needed.

Cognitive Maps

Cognitive maps are internal entities that cannot be seen. This is why researchers in spatial cognition have developed a variety of methodologies to examine them [98]. Three accepted methods were considered: additive similarity trees (AST), multidimensional scaling (MDS) and sketching. Two seemed unsatisfactory because they indicate how landmarks are clustered in memory but not so much about their layout (AST) [98], or required participants to produce Euclidean distance judgments (MDS), which were observed as difficult to make from tactile maps [293] .

I chose the Sketch map methodology, which requires participants to draw the environment [157]. Although blind persons can draw [294, 295, 128], I adopted an approach similar to Lahav & Mioduser [86], which requires participants to reconstruct a model of the environment using tangible objects. Blind persons can manipulate objects in space with great dexterity, for example when eating. The reconstructed models would show the properties of the participants' mental maps [98].

There is a danger to have the experimenter interpret the data. Even an honest researcher may be subjective in his judgment because bias is hard to control. This problem was addressed by having the maps interpreted by OMS experts, completely unaware of where the maps came from and what I was trying to study. Both experts worked with specific written instructions, a methodology similar to double-blind studies in medical science where similar problems exist.

The experts were informed that the responses were reconstructed from memory by a group of blind participants, received a digital copy of the participants' responses by email, which are shown in Figure A.13 & A.14, the layout of the malls shown in Figure 7.1, and a set of instructions asking them to grade the data according to four dimensions: the spatial relationships between landmarks, the layout of the corridors, the topology and the overall quality of the responses.

The dimension landmark-to-landmark (LL) examines the spatial relationships between landmarks and ignore the corridors; the dimension corridor layout (CL) examines bearings, shape and position of corridors; the dimension topology (T) examines the sequential arrangement of landmarks along corridors and their reachability; and the overall quality dimension (Q) examines the pragmatic sense of the maps.

The markers, who knew each other, were specifically asked to work individually. They developed their own marking scheme based on the instructions they received, which are reproduced in Section A.4, and then graded the participants' responses. They were asked to look at all the maps and give one perfect score per dimension and one lowest score per dimension, and use these bounds to grade the responses left. This request forced the markers to examine every map before starting to grade them, but also prevented having all the scores clustered in the same range. When finished, the markers e-mailed their marking schemes and scores, and received $50 in compensation for their time. The schemes and scores are in Section 7.3.

Acquisition Rate

I examined the rate at which participants could remember the location of landmarks and their names by counting the number of correctly assembled landmarks at five-, ten- and fifteen-minute intervals. I ignored corridors because too many parameters needed to be examined to judge their quality (*e.g.,* starting & ending points, shape, interconnectivity), which seemed difficult to measure with objectivity. A landmark was considered correct if its spatial relationships with the other landmarks were preserved under continuous deformations, as described in Figure 7.4. A misplaced landmark was not counted.

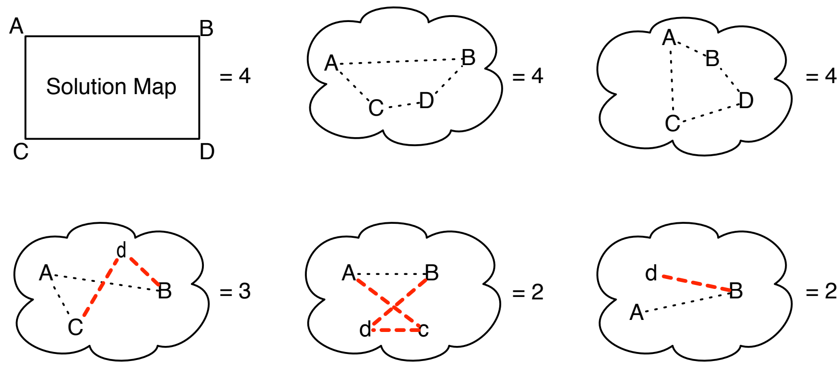

Figure 7.4: This figure exemplifies how acquisition rate score was counted. Each of the four letters at the corner of the solution-map represents a landmark. As shown, a perfect reconstruction gets a score of 4. The letters inside the clouds are mock-up reconstructions with their score written on their side. *Clouds in the top row:* these map received a score of 4 because spatial relationship of landmarks is preserved under continuous transformations, which include magnification, shrinking/stretching and translation but exclude rotation and reflection. *Clouds in the lower row*: the left most response receives a score of 3 because landmark D is misplaced (misplaced landmarks are written as lower case letters). Also, a quick look at the edges (which are only shown to ease the visualization of the transformation) shows that C & d crosses A & B, indicating a non-continuous transformation. The remaining relationships are preserve (A, B and C). The two remaining maps get a score 2 because only the relationship of landmarks A and B is preserved under the continuous deformation rules described above and because a landmark is missing (right).

The acquisition rate can be helpful to determine the number of landmarks participants could learn over time with a system, which consists in finding the slope of the regression lines drawn from the number of correctly assembled landmarks at each interval. See Figure 7.6 for more details.

Autonomy

Even the best TMs are not read without assistance, that is, tactual perception must be combined with verbal descriptions [296], which is why I followed the recommendations from NPO[2] and offered my assistance on demand to the participants. Assistance requirement is also supported by Bentzen [9], who warns us that raised-curve tracing on a surface is not a easy task, and is often accompanied by a good deal of off-line searching, especially if the shape is complex.

[2] The OMS participant introduced in Section 3.2.1

I distinguished two types of assistance: landmark identification and technical clarification. *Landmark identification* consisted in telling a user the name of landmarks on the map. Space restriction prevented the cartographer from writing the names of landmarks on the map and forced her to use identification numbers written in Braille. The full names could be found in the map key. Every time a participant requested the name of a place, I spoke it, like the Nomad [68], Audio-Touch [69] or TTT [70]. *Technical clarifications* consisted in telling a user what the tactual symbols represented or explaining the layout. Although the former could have been spoken by systems like the Nomad, the latter sometimes needed physical assistance: almost every participant asked for finger guidance in areas they found difficult to understand on the TM. Such technical assistance could be provided upon request, but also be offered by the researcher when the participant was getting stuck losing precious time, become impatient or discouraged, as captured by Table 7.4.

However, the majority of the technical assistances came from questions asked by the participants: 'what does this square represent?' 'what is this line, a wall or a door?' 'what does this surface mean?' 'why are these lines touching themselves?' 'why can't I figure out how to enter this store?' 'Can you tell me what this thing is again?' These are examples that counted as technical help.

Table 7.4: This table shows the criteria that elicited help from the researcher. The delay of 5 seconds was estimated by the researcher. The two latter points under Technical Assistance were offered to avoid lost of interest or discouragement in participants. The goal was to keep the participants keen about the task and make them feel they are doing just fine even if certain aspects of the task were difficult to them.

Technical Assistance	Identification Assistance
Upon request. Participants stuck for ∼5 seconds or more. Participants became impatient. Participants sighed.	Upon request.

My assistance reduced TM reading time because the participants did not need to consult the map key, which is a time consuming process. I provided the same assistance with SAM when needed, although landmark identification was not necessary because SAM provides this information on user request.

Instances of assistance in the two categories were recorded and used as a measure of autonomy for both systems. All assistance was counted, including when the same question was asked more than once.

7.3 Results

The quantitative data is analyzed in a three-step process: first, the assessment of the reconstructions is examined, then acquisition rate and autonomy are discussed. The qualitative data from participants is discussed in Section 7.3.5.

7.3.1 Cognitive Map Results

This subsection describes the experts' marking schemes, presents their scores and data analysis.

The experts were asked to provide their individual marking schemes, one per dimension. The following paragraphs group their comments. One of experts also commented upon every map individually. His statements are not shown in this chapter due to space constraints, but can be found in Section A.5 on page 192.

Experts' comments for the landmark-to-landmark dimension: all landmarks needed to be present; the landmarks needed to include entrances and exits because these can be hidden; open spaces needed to be preserved, i.e. space between landmarks, and not only the sequence of landmarks, needed to be shown. The experts added that assumptions can often be made about where one is and what one will find next in a closed indoor space; having all the landmarks present and accounted for is a crucial factor as it gives a blind traveller an automatic idea of what is in the mall.

Experts' comments for the corridor layout dimension: all corridors relevant to locating landmarks in the solution map needed to be present in the reconstructions; accurate beginning and ending points for all corridors needed to be shown; corridor shape needed to be replicated. The experts added that accurately-shaped corridors in the mental map of a blind person provide a valuable overview of the layout of an area.

Experts' comments for the topology dimension: accessibility of every landmark from corridors was needed - no floating landmarks; an appropriate sequence of landmarks was needed; landmarks needed to be on appropriate side of corridors. Even if a reconstruction is not an exact replica of the original, some features must remain consistent. The experts stressed the importance of knowing which side of a corridor to trail to locate a specific store, because many mall corridors are too wide to scan both sides at once.

Experts' comments for the overall quality dimension: the ability to locate any store in the mall was needed; the ability to locate stores on the appropriate side of the corridor was needed - because this increases the functionality of the map; the ability to enter and exit the mall is needed. The experts pointed out that to be functional, a reconstruction must allow a person unfamiliar with the environment to travel around independently and locate specific landmarks. However, they added that certain assumptions can be made by a more experienced traveller, to help fill in the gaps. Thus, a map can still be somewhat functional even if an individual dimension scored poorly.

Table 7.5: The reconstruction scores attributed by the two experts are shown in this table. Every column holds a perfect score of 100 and a lowest score of 0, which is what I requested. The map IDs in the left column of this table match the IDs of the reconstructions shown on page 195 & 196. The acronym LL stands for landmark-to-landmark, CL for curve layout, T for topology and Q for overall quality. The overall mean scores for the two methods are shown in the right column.

Map ID	Marker Male				Marker Female				Mean
	LL	CL	T	Q	LL	CL	T	Q	
TM B1	100	30	50	100	100	40	60	100	
TM D1	20	60	20	20	10	50	30	30	
TM K1	55	0	50	65	25	0	30	40	
TM K2	90	40	100	70	70	40	100	60	60.2
TM K3	80	70	60	85	60	60	80	75	
TM M1	70	60	50	70	60	70	40	60	
TM W1	80	100	70	90	90	100	60	75	
Mean	70.7	51.4	57.1	71.4	59.3	51.4	57.1	63.0	
SAM B1	35	60	40	40	25	65	50	50	
SAM D1	0	20	0	0	0	10	0	0	
SAM K1	70	70	70	70	60	75	60	65	
SAM K2	70	80	50	70	60	65	60	75	58.6
SAM K3	90	80	75	85	85	70	60	95	
SAM M1	80	70	70	80	85	75	65	75	
SAM W1	80	70	60	70	95	65	65	70	
Mean	60.7	64.3	52.1	59.3	59.0	60.7	51.4	61.4	

Data Analysis

The data analysis is performed on the scores shown in table 7.5. The means of both systems are close, 60.2% & 58.6% for TM and SAM, respectively. The lowest coefficient of variation (the standard deviation as a percentage of the mean - a measure of dispersion) is found in the corridor layout (CL) dimension for SAM while the highest coefficient of variations is found in the CL dimension for TM. Please refer to Figure 7.5. The analysis of variance on technology with dimensions for factor revealed no significant difference between the systems.

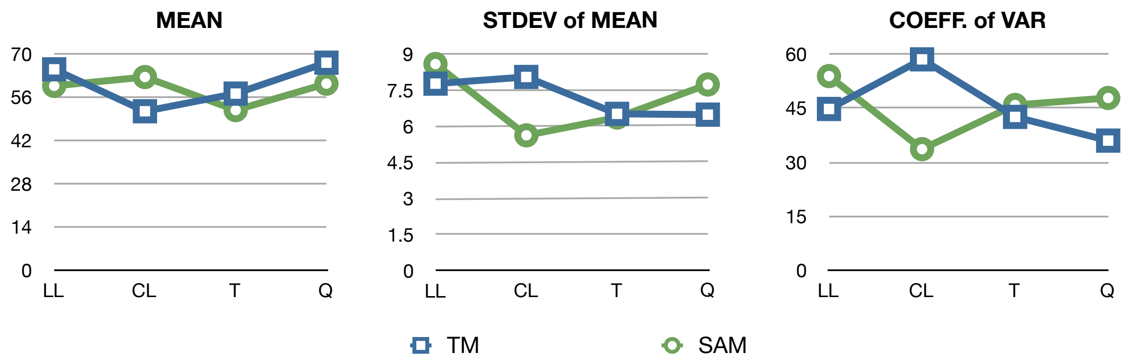

Figure 7.5: The left graph shows the means of the four dimensions in SAM and TM. The standard deviations of the means are shown at the centre. The right graph shows the coefficient of variations; their absolute differences $|TM - SAM|$ are 9.124, 24.90, 3.30, 11.83 for LL, CL, Q and T respectively.

The experts were requested to grade highest score to one hundred and lowest to zero, which resulted in data not obviously normally distributed. A uniform distribution between zero and one hundred seems logical, so parametric and non-parametric test on data was performed.

The two systems were compared using three within subjects pair-wise tests on the four dimensions: a distribution-free sign test, a Wilcoxon test for paired observations and a two-sample t-test. None of the tests indicated a significant difference between the dimension scores. Correlation tests on landmark-to-landmark (LL), CL, overall map quality (Q) and topology (T) for SAM and TM were computed, rejecting the null hypothesis of no correlation between three dimensions LL, T and Q, but not for CL. The correlation values are tabulated in Table 7.6.

Then, a distribution-free Spearman rank correlation coefficient was calculated to determine if a systematic difference between the scores given by the two markers could be found, which rejected the null hypothesis of no correlation at $p = 0.01$ for the four dimensions, suggesting the two experts produced very similar results.

A measure of power was computed to determine the sensitivity of the experiment, *i.e.*, the smallest significant effect detectable at 5 % with seven participants and two markers, which revealed the experiment was powerful enough to find difference as small as 20%.

The same measure of power was used to determine the minimum number of participants required to make significant effects at the 5 % level. Calculations determine that the power needs to increase by at least 20%, which means five to ten times as many participants

Table 7.6: The correlation for SAM and TM on LL, CL, Q and T. These values reject the null hypothesis of no relationship for all dimensions, but CL; the critical value of Pearson's correlation coefficient at $p = 0.1$ is 0.458 and at $p = 0.05$ is 0.532. Also, t-test results (two-tail) in the second row reject the null hypothesis that the results are a chance occurrence for every dimensions but CL (critical value: $t_{0.1,\ df(12)} = 1.782$ and $t_{0.05,\ df(12)} = 2.179$).

	LL	CL	T	Q
r	0.455	-0.043	0.449	0.531
t	1.768	-0.148	1.742	2.173

would have been needed. Table 7.7 shows the smallest detectable effect and the number of participants needed.

Table 7.7: This table shows the smallest detectable effect and the number of participants needed to obtain a level of difference that would make a paired-wise t-test on dimensions significant.

	LL	CL	T	Q
Smallest Detectable Effect	18.5	21.6	14.6	15.0
Number of Participants needed	138	44	85	57

A sequence of correlation analyses were computed to investigate the differential between the four dimensions. The correlation matrix r in Table 7.8 shows the correlation of dimensions when the difference in performance is calculated within participants and the technologies and markers are collapsed. The correlation coefficients shown in the table are proportional to the ability to predict values from linear models on other values: the results show that we can get a good prediction of LL, T, and Q, but not one of CL, which is quite independent of the other dimensions.

7.3.2 Acquisition Rate Results

The regression lines of the acquisition rate for TM and SAM are shown in Figure 7.6. Their slopes are 0.9824 and 0.9114 for TM and SAM respectively, which indicates participants learned about one feature per minute with the systems. The positive intercepts at zero minute indicate the overhead, something the participants needed to do before learning landmarks. We observe diminishing returns at fifteen minutes: participants learned fewer things as they explored the map during this time interval.

Table 7.8: The correlation matrix in this table shows that LL, T and Q are measuring something similar about the subjects and the task, and that CL is measuring something different.

	LL	CL	T	Q
LL	1	0.47	0.81	0.95
CL	0.47	1	0.44	0.49
T	0.81	0.44	1	0.76
Q	0.95	0.49	0.76	1

7.3.3 Autonomy Results

I provided a total of 39 technical assistances for SAM and 205 technical assistances for TM. In addition to the technical assistances, I provided a total of 110 identification assistances for TM and one for SAM. Figure 7.6 shows these numbers averaged across all participants at different time intervals.

An analysis of variance with participants, technology and time-intervals on number-of-identification-assistance and then on number-of-technical-assistance were performed including two-way interactions. A significant interaction between time-interval and technology on number-of-indentification-assistance was found ($F(2,36) = 67.378$, $p \leq 0.0001$). The same interaction, but this time on number-of-technical-assistance, was found marginally significant ($F(2, 36) = 2.857$, $p = 0.0705$). Details of the analysis are in Appendix A.2.

7.3.4 Discussion

This experiment assessed the usability of SAM at three levels: its capability to convey spatial information to blind users, the speed at which blind users could learn an environment from interacting with it and its level of autonomy. A well-crafted TM showing a different environment, but one similar in complexity to the one represented by SAM, was chosen as metric because blind persons have reading experience with TMs and because TMs are well studied by psychologists, who judge them as effective tools to teach spacial concepts to blind travellers.

Two experts graded the tactile responses of seven blind participants. Each participant reconstructed two final maps of two malls they have read apart, one with TM and one with SAM. Each map was evaluated into four different dimensions, namely, landmark layout, corridor layout, topology and general quality of a map, yielding a total of 56 scores per expert. A Spearman rank correlation computed on these dimensions suggested a marking consistency among the two experts, such that the variance in one marker can be predicted

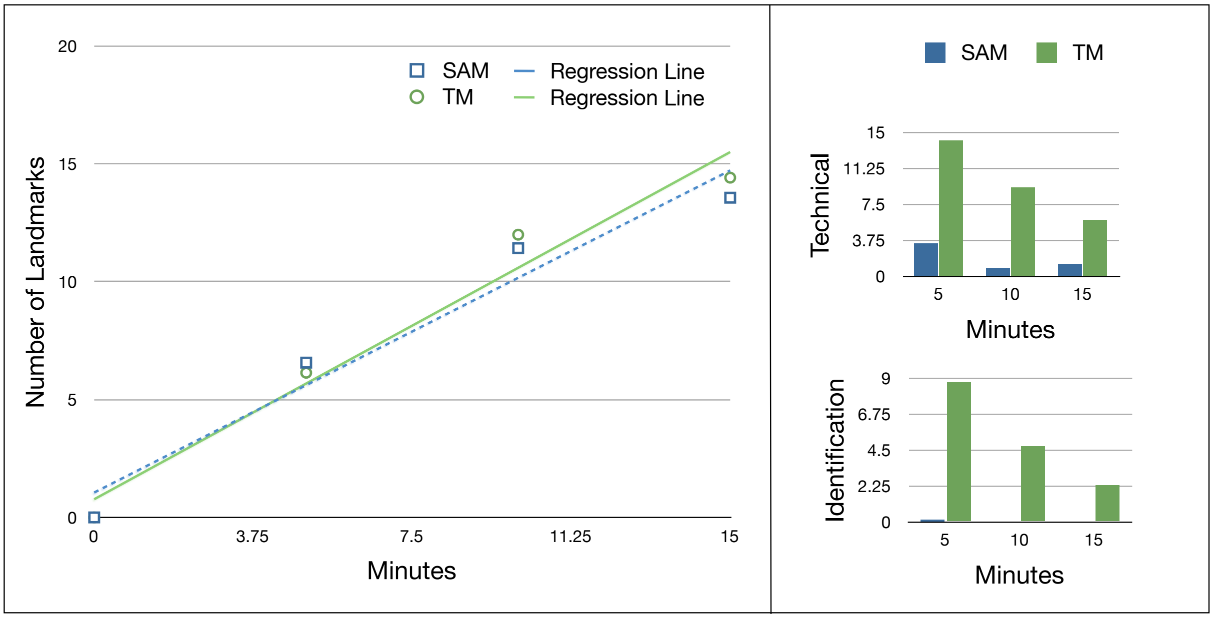

Figure 7.6: *The left graph* shows a regression line drawn from four measures: the number of landmarks remembered at zero-, five-, ten- and fifteen-minute time intervals. *The bar charts* show the amount (averaged across all participants) of technical and identification assistance participants received during the first, second and third reading intervals when interacting with the systems.

using a linear function of the other marker. This result proposes solid information in the data and made it reasonable to aggregate across experts.

The cross examination tests performed on the dimensions (a distribution-free sign test, a Wilcoxon test for paired observations and a two-tailed t-test) failed to statistically accept the alternative hypothesis that the technologies are different. Conversely, a correlation test on these dimensions, shown in Table 7.6, could reject the null hypothesis that no relationship between the systems exists in the Q dimension, suggesting the systems are equivalent in that dimension. Thus, the LL, CL and T dimensions could neither be shown different, nor the same, which incited measuring the power of the experiment and which revealed that five to ten as many participants would be needed to find significant results, assuming the same markers, same marking directives and participant of similar quality are used. An experiment of that scope is of course too expensive to conduct in the academic realm, and would be quite challenging to organize by anyone because of the difficulty of finding so many blind participants [290].

The differential between the four dimensions was further examined by a series of correlation tests that brought experts and systems together: the correlation matrix is shown in

Table 7.8, which indicates a reciprocal relationship between LL, T and Q, one dimension being the linear prediction of the others, and which gives CL a status of independence.

The correlation between LL, T and Q suggests that similar capabilities are used to do these aspects of the task but CL, by contrast, probably uses quite different mechanisms. Could this be explained in terms of spatial coding qualities? The experts' marking schemes describe LL, T and Q in terms of landmarks and landmarks along routes and CL in terms of shape, end-points and intersections. It is likely that learning landmarks relationships necessitate a greater ability of exocentric coding - recall the participants teleporting their hands in Section 3.2.4. Accordingly, learning corridors, which forces a perceiver to discover space through continuous exploration, probably requires a greater ability of egocentric and sequential coding - in fact, curve tracing in SAM rests upon this basis. Unfortunately the literature of psychology on blind persons examines coding strategies as a whole and does not break strategies in these terms[3].

A closer look at the means in Figure 7.5 suggests that SAM outperforms TM in the CL dimension. Although the difference in the means are not statistically significant, it is interesting to find CL for SAM with the lowest coefficient of variation, and find CL for TM in the opposite range. These observations suggest a pattern: participants have learned corridors better with SAM than with TM.

With SAM, the foveal tools enforced a reading strategy that has a user discover landmarks along corridors, making corridor tracing a requirement for reading a map, something TM does not oblige.

CL measures learning of route-oriented geography: with SAM, the task required the user to carefully trace routes using sound directions. It is possible that the brain has redundantly integrated sonic streams to kinaesthetic ones, which reinforced the knowledge of curve layout. Tracing curves with SAM is a well structured task, an egocentric one[4] that requires a user to immerse himself in a virtual world, demanding full attention on the task. TM on the other hand leaves the user to choose his own strategy, which may or may not be optimal to learn curve layout (see Section 5.1.1), and which may or may not require the same level of attention. Many psychologists judge choice of movement strategies as key to building accurate cognitive maps in blind persons, as discussed in Section 5.1.

LL measures learning of landmarks geography: recall the participant strategies observed when reading TM (page 36). Indeed landmarks are larger than corridors, which have readers attend to them naturally. Because attention is directed there, these items are better learned, at the detriment of others, it appears. With SAM, landmarks are integrated to corridors, a design decision I made to build holistic knowledge in the reader: these things

[3] I was unsuccessful at finding work that compares coding strategies for learning continuous and discrete features in space.

[4] Recall the participant's statement: '[my] finger is a scaled down version of me' on page 36.

are learned together because they should be linked together. For example: 'I am at the bus terminal, therefore I am on Charles Street'.

The experiment allowed participants to use the tools they found most appropriate for the task in both systems. I observed the participants relying mostly on the foveal tools with SAM, because they felt they were better trained and more comfortable with these tools, as discussed further in Section 7.4. Their choice of tools most likely affected the spatial knowledge they built: would the results be the same if they had used a greater fraction of peripheral tools and a lower one of foveal tools ? With other tools, the user's attention would be shifted on other spatial characteristics, which would probably have an effect on performance. Different aspects of a task, which require focused attention and mental effort on different characteristics, produce different memories. There is no doubt that attention and memory are intertwined, and the understanding of the complex nature of these interactions is still current research in neuroscience and psychology [297].

The counting protocol I relied upon to measure learning-rate imposed fewer restrictions than the one adopted by the two experts. Recall that experts indicated that sequence order of landmarks and open spaces needed to be preserved to get high scores in the LL dimension. I considered asking them to grade the participants intermediate responses at the five- and ten-minute intervals but abandoned the idea because it took them more than two months to produce the scores shown in Table 7.5 due to too much work at the CNIB. Going back to grading these maps, which would have required them to do twice the work they did, was inadmissible because of time constraints.

The open space restriction puts SAM a little behind TM in the LL dimension, which suggests that participants may have had more difficulties perceiving distance between landmarks with SAM than with TM, even though the two systems rely on user's kinaesthesia for distance perception. The inherent reasons are difficult to identify and the question needs further investigation: maybe the malls represented by SAM and TM were not equally hard in that dimension[5], maybe the participants needed more practice time calibrating themselves with the stylus, or maybe the display is responsible? Recall the curve tracing encoding kept a stable presentation rate of 6.7 Hz (Section 5.2.1) notwithstanding the tracing speed. Would adjusting the presentation rate to stylus speed sensitize the user to distance displacement? This question is worth exploring in future work.

The measures of autonomy show the most striking results. Although the data on technical assistance is taxed with subjective interpretation, as explained in Table 7.4, I was careful to be consistent in how I provided and counted assistance for both systems. The results indicate that much fewer interventions were needed with SAM; however, I should mention that some of the assistance I provided had high impact. For example, the area below the optometrist was the most difficult one, where some participants needed to

[5] As explicitly pointed out by the experts and described in Section 7.4.

be reminded to listen to the curvature notification cue because they were missing the turn, and ended up searching for a straight continuing curve that did not exist; these participants reported being surprised finding two turns so close in proximity from one another. However, I never guided the participants' hands with SAM, something I was compelled to do with TM, especially near the dentist area because several participants could not understand the corridor layout.

Identification assistance data puts SAM much ahead of TM in terms of autonomy. This should not surprise anyone because SAM provides details-on-demand to the user, which eliminates the need of having someone speaking the name of features. This characteristic is especially beneficial to Braille illiterate users, who normally have to commit the names of features in memory to be able to use a TM on their own.

In the end, it is difficult to argue against SAM's capability for communicating non-trivial geometry to blind users. The resemblance between the reconstructions shown in Figures A.14 and the original map shown in Figure 7.1 is apparent. Having had increased the experiment's training time, which some participants saw as quite a commitment already[6], would have probably improved the quality of the reconstructions, which would have increased the performance gap between SAM and TM in favour of SAM. Likewise, increasing the number of participants would have improved the reliability of the findings, but I doubt that a larger group of experts would have had much effect on the results because of the similarity in their scores.

The following section presents subjective responses from the participants, which are discussed in line for better readability. Then a Final Discussion about the holistic results ends the chapter.

7.3.5 Subjective Responses and Discussion

I e-mailed a post-study questionnaire to the participants, which they filled out and e-mailed back to me. This section describes the participants' impression of SAM and the experiment. The same participants' ID shown in Table 7.2 on page 123 are used in this section to link quotes to participants. The seventy-two questions and original answers are available online[7].

[6] Participants liked learning SAM, but did not like the rigid schedule imposed by the experiment. Section 7.3.5 discusses this further.

[7] http://www.cgl.uwaterloo.ca/~m2talbot/thesisAppendix/Questionnaire.pdf

Map Equivalence

The reading difficulty for the two maps was the same for three participants; two thought that the TM was harder to read and two thought the opposite. However, all the participants thought the mall layout was more complex to understand in the TM than in SAM. Indeed, I observed participants experiencing great difficulties in the Dentist area. For example, several participants found it difficult to see that Office B was accessible from an entrance near the Toy Store, and did not know how to enter.

Participants reported losing time trying to find corridors and entrance doors and criticized the TM for being too busy in certain areas, making it difficult to figure out how to travel through the mall. In fact, six participants thought the details on the TM were unnecessary information that made the map less legible. For example, K1 and K3 commented that the TM had 'way too much needless detail' and was 'too busy'. Only one participant found the details helpful for her understanding of the layout.

Four participants indicated that they did not want SAM to convey more detail, and two said that they would like to see an option where they could choose the level of details. All the participants thought that curve and POI representations were sufficient for the task at hand. POIs and curves, as suggested, could carry richer information on request, like explaining the geometry of the pillars or the facing wall at the entrance of the cafe. Overall, four participants thought they understood the two malls equally well and two thought they understood the mall they explored with SAM better. Only one thought the opposite.

Four participants perceived the mall in the TM as smaller and three thought the malls were equal in size, but the sum of the length of the corridors was approximately 1% greater in SAM and the total surface area was approximately 1% larger in TM, making the two malls almost identical in size. This impression may be attributed to different reading strategies. With SAM, the participants traced the corridors and discovered landmarks along the way. With TM, they searched for landmarks and then tried to figure out corridor position. Hence exploring the map with SAM had them travelling long corridors, something they did not necessarily do with TM.

I did observe the same approach during the reconstruction periods. With SAM, the participants began by drawing the corridors, and then dropped the landmarks along them, whereas with TM, they began by dropping the landmarks and then struggled to fit the corridors in between.

Usability

All the participants felt confident in their ability to use SAM. All thought the exploration of spatial auditory maps felt natural. They all indicated they understood the routes and

POI representations and their properties and relationships.

All the participants found the directional sound encodings effective, but one participant indicated she would have preferred speech guidance over locatones; the others all reported the opposite. M1 commented: ‘Sound is faster. Faster is better. Perhaps the speech option could be used for tutorial mode when someone is learning how to use the product.’ All the participants rated the stereo balance encoding for curve tracing as good, none as fine and none as poor; six participants rated the pitch encoding for curve tracing as good and one as fine; four participants rated the effectiveness of the encoding when curves are neither vertical nor horizontal as good and three as fine. All found the choir sound helpful, especially with horizontal curves. All indicated they felt confident in their ability to trace curves with SAM.

The notification sounds were distinguishable by everyone, and information at intersections was well understood by everyone, but two participants reported experiencing some difficulties with the curve direction cues. One of them, D1, commented: ‘more tweaking might be good here but in general they were fine’. This remark is interesting because D1 judged the locatone cues and presentation methodology at street intersections as good; Figure 4.6 on page 70 shows that both convey direction the same way.

The participants indicated they could adapt to a stylus, but two participants would prefer if SAM did not need one. One reported experiencing hand cramps during the sessions and the other commented: ‘only because you stand less chance of misplacing the stylus’ (D1). Selecting tools from stylus manipulations was describe as effective by all but one participant who reported sometimes being confused.

All the participants thought the peripheral tools were straightforward to understand, but all indicated not getting sufficient practice time with them. All the participants thought the index tool was effective for finding the location of things on the map.

Five participants indicated knowing exactly in which circumstances they could call the wizard, and two participants reported being less sure. These two participants reported knowing well how to use it from the index and the curve tracing tool, but less from the grid and the collision detection tools. One of these two participants commented: ‘with more practice that would become clearer’ (K3).

Aesthetics

The questionnaire assessed the aesthetic quality of every feature individually. All the participants viewed SAM as an aesthetically pleasing system. However, K2 suggested making the curve tracing encoding ‘less monotonous and hypnotic’. But like sonars, SAM conveys a series of short discrete sounds rich in transients for superior localization that need to be refreshed to communicate realtime feedback to a user.

Alternative to TMs

All the participants viewed SAM as a possible alternative to TMs. M1 added: 'Not only tactile maps, but a whole slew of different applications: learning to write, game play, spreadsheet navigation, an add-on to screen reader access, simple drawing creations for Braille embossers, math graphics.'

All the participants indicated they would be willing to invest eighty dollars for a digital tablet to use SAM. All the participants but one indicated they would build their own TMs using a Styrofoam or a raising paper layer, which they would preserve and refer to when needed. The possibility of building a physical representation of a virtual map addresses the concern raised by Scaife *et al.* [298], who think virtual reality systems should allow a user to leave traces.

Learnability

TMs could not be used to measure SAM's learnability because the participants have been using them their entire life, which makes it hard for them to remember the difficulties they had learning them. Instead, I chose JAWS as a measure [27]. JAWS is a screen reader application used by all the participants to read the content of their computer screen.

Six participants thought SAM was much easier to learn, and one thought they were the same. D1, one of the two blind expert consultants, commented: 'It takes less time to learn [SAM], and there is less to remember [...] I'd say 10 hours of working with SAM would give you the grounding you need to be successful using it.' M1, the other blind expert, commented: 'JAWS has a very steep learning curve, plus you need to also learn an operating system. SAM is self-voicing and could easily be used by all users without too much effort.'

Acceptability

All the participants indicated they would use SAM in public spaces like cafes or restaurants. All indicated they would recommend SAM to a close friend. All indicated they would be willing to practice using SAM on their own without any financial compensation, if SAM and maps of different environments they cared about were installed on their home computers.

The questionnaire ended with an open question where participants could write additional comments. K1 wrote: 'I believe that this is a great tool for mobility instructors to use with kids who are just learning their way around a new place, such as school. Other people would also benefit, but introducing SAM to kids would make it very easy for them to get around since it would help them build maps at a very early age [...] I had fun playing with the software and I think that you have developed a terrific tool which can be used by different types of people not just visually impaired, and also for different purposes.' M1 wrote: 'I'm a big fan of SAM and hope that development continues.' Overall, all the participants liked the system, and hoped research and development will persist.

7.4 Final Discussion

Normally, a blind pedestrian travels through the environment by walking on sidewalks along streets, lanes or boulevards, and tries to avoid open space as much as possible. Yet the data from the field study discussed in Chapter 3 revealed that blind readers prioritized buildings over streets when reading TMs, because buildings stand out and attention is directed towards them.

Although landmark-to-landmark knowledge is desirable, research shows that congenitally blind persons rarely build these relations and rely on egocentric and sequential coding to understand space. Indeed, all the participants, including the adventitiously blind, imagined themselves walking in the environment as they traced embossed streets with their index fingers (Section 3.2.4).

This breakdown inspired me to enforce a reading strategy, one that has a reader learn streets and discover POIs simultaneously along the way. After all, every human-made landmark is accessible by a pathway and blind travellers are more interested in locating the entrances of buildings than their centroids. Furthermore, learning landmarks and curves separately builds unconnected relationships that need to be reconciled later. This transformation follows user-centric methodologies, which recommend reinventing work practice based on the foundation of user data [65].

The significant correlation in the Q dimension suggests that the overall quality of the mental maps were similar with SAM and TM, despite the fact that knowledge could not be shown neither different nor the same in the CL, LL and T dimensions. Indeed, both systems could *allow blind participants unfamiliar with the environments to travel around independently and locate specific landmarks* (taken from the experts' definition for the Q dimension in Section 7.3.1). This confirms that effective learning of landmark relations from sequential exploration is possible. These results are in accord with Thorndyke and Hayes-Roth [269].

SAM is a novel assistive technology that enforces a map-reading strategy, *i.e.*, one that imposes constraints on exploration, but leaves the user free to go anywhere he wants on a map. For example, considering the foveal tools, there is not much information that can be drawn from a flat surface if curves are not followed, but a user can trace any curve and start from any feature he wants. SAM also implements a set of peripheral tools that support free-landmark exploration that do not require a user to trace a curve to discover landmarks. These Peripheral tools, however, are not absolutely needed to learn a map. Using them gives quick access to remote locations and may help building exocentric representations in the user.

In that regard, I observed participants relying mostly on the foveal tools to read the maps during the experiment. Post interviews revealed that, except for the map indexer, the participants, eager to perform well, felt they were not sufficiently trained with the peripheral tools to use them during the assessment. Given the limited reading time they had, they preferred using the tools they knew best. One participant recommended ten hours of training instead of four: two-and-a-half times what they were administered during the study. We note, however, that ten hours is much less than the four weeks of training needed to learn the drawing application IC2D[8] [128]. The participants could trace curves and simultaneously attend to notification cues. This suggests that the perceptual demand for tracing curves was not unduly heavy, even with insufficient training[9].

In fact, the participants discovered and memorized landmarks at a comparable rate with both systems[10], which suggests that the perceptual demands of reading the tangible map and the spatial auditory map used in the experiment were not significantly different. By extension, these results suggest that SAM's interactive design and data presentation methodologies, at least the ones tested, were effective. The results suggest that SAM could successfully circumvent the difficulties entailed by the transient nature of auditory displays [302, 66].

Autonomy and acquisition rates were measured by counting the instances of assistance and the number of landmarks the participants could reconstruct from memory. These measures, however, need to be interpreted with caution: having let the participants retrieve the information themselves from the map key would have reduced the number of requests for assistance, and increased the level of autonomy in TMs, but would have lowered the acquisition rate. This measure suggests an autonomy/acquisition rate trade-off, which would probably vanish if uniform standards were adopted in TMs, if a participant was familiar with the symbols used by the cartographer, or if a purified representation, like one made exclusively of curves and points, was used.

Interestingly, the OMS experts, who commented upon the two malls shown in Figure 7.1, pointed out that the curved corridors in the mall represented by SAM made it

[8] As reported in Section 2.3.

[9] Perceptual load is a major determinant for selective attention [299]. A user can process lower-priority stimuli (e.g., attending to a notification cue like a POI) when the high-priority stimuli (e.g., tracing a curve) do not demand all attentional capacity [300]: too-high perceptual load conditions necessitate selection or filter [301], which cause a perceiver to miss low-priority stimuli.

[10] With the caveat that open space was better preserved with TM than SAM, as discussed in Section 7.3.4.

difficult to maintain appropriate distance between landmarks – creating the curve too soon or too late or/and creating open space and crowding respectively. The corridors in the mall represented by TM were drawn in straight lines, which alleviated this problem. On the other hand, they indicated that the ramp and landmarks located behind other landmarks in the TM representation were problematic as it could make it difficult for a blind person to figure out how to access certain stores. Overall, they commented that each map had its own difficulties.

The question of polarity inversion discussed in Section 7.2.2 and musical background needs to be examined further. Musical training and auditory display performance have a long history of contradictory results [66]. Some researchers have observed superior performance in musicians using auditory displays, and others have reported weak to nonexistent differences in performance. Unfortunately, I could not find any research reporting evidence of polarity inversion in musicians. Investigating the causes, and more importantly to HCI, measuring the ratio of users (blind and sighted) who share this pitch inversion bias, could contribute to build a body of knowledge that could serve the community of auditory display researchers and designers. I leave this research for future work.

Finally, the results in this study suggest that SAM is equivalent to TM in terms of usability, but it should be clear that more studies involving multiple TMs of different construction styles and environments with many more participants are needed to provide further evidence to support my observations. The take home message from this experiment is that SAM was successful at representing a non-trivial environment that blind persons would normally learn about with a TM.

Chapter 8

Conclusion

With tangible maps (TMs), blind persons can learn the layout of novel workplaces, campuses, malls, bus terminals or neighbourhoods, and thus prepare journeys to unknown areas in the comfort of their home. Unfortunately, a shortage of TMs prevents most blind people from accessing them. Jacobson, who views space as fundamental to human existence, points out the great influence space has on human thinking [22]; blind persons are cognitively sound individuals with a perceptual impairment and should, like sighted persons, be given the tools they need to access physical space.

The solutions described in Section 2 are either financially unrealistic, offer too low accuracy to understand space with enough precision, or fail to examine scenarios complex enough to offer ecologically valid solutions. Table 8.1 puts every previous research into one or more of these categories.

The research that compares auditory-only encoding to haptic-only ones reports better performance with the auditory-only approach [97]. Studies that examined systems with auditory and haptic feedback have observed that a suboptimal auditory encoding could improve a sophisticated haptic one [82, 92, 93]. These results, along with evidence that audition has a higher information bandwidth than touch [28], and that auditory interfaces offer technical advantages over haptic ones[1], make substituting sound for touch appealing.

Unfortunately, current curve tracing research is ineffective when addressing the problem of displaying multiple curves in space. A single curve environment eases the design of any interface because the encoding can make full use of auditory bandwidth. Multiple curves intersecting, turning and cohabiting with landmarks share the bandwidth, leaving fewer resources for guidance cues. Thus, the success of this research cannot be measured solely

[1] E.g., computerized tools for creating and editing sound are far more advanced that ones for creating physical artifacts. Distribution of digitized sound is effectively free compared to distributing physical tactile maps. Finally, recent advances in small format sound devices makes them easy to carry and inconspicuous in use.

Table 8.1: This table shows the previous research weighted against the pre-conditions discussed in Chapter 2. *Cost* includes the hybrid systems that require the purchase of expensive equipment or have high operation cost; *Precision* includes the systems that are unable to display spatial information with enough accuracy for a user to understand the spatial relations of features; *Ecology* includes the systems that are unable to convey landmarks and streets of many geometries intersecting together to a user. ETAs and Position Locator Systems technologies are not shown in the table, but all pertain to the cost category. As we can see, some systems appear more than once, as they belong to more than one category.

Cost	**Precision**	**Ecology**
[68, 69, 70, 74]	[73, 75, 77, 79]	[77, 79, 80, 81]
[91, 92, 96, 97]	[80, 81, 86, 87]	[82, 87, 88, 89]
	[88, 89, 90, 91]	[90, 91, 92, 94]
	[22, 110, 111]	[96, 97, 99, 100]
		[93]

on the basis of curve tracing. Ambiguous guidance, as occurs in symmetrical encodings [96, 99, 100, 110], demands a longer attention span and consumes cognitive resources needed for constructing cognitive maps, but the same is true for ineffective data presentation protocols [129] or poorly designed human-computer interaction[77, 79, 80].

The design and development of a technology that covers the same environments as TMs while fostering the same quality of cognitive map in the user is a real challenge. Moreover, to be ecologically valid, the technology must be affordable to acquire and operate. The design of SAM is based on user-centric methodologies in HCI [65], from which decisions about user-input/user-output devices and user-computer interactions were drawn. The latest technologies in computer science were chosen to implement a set of innovative tools (multi-touch tablets not being available at a low cost back then). Studies and pilots were conducted on blind participants and experts throughout the design of SAM, the proposed technology.

Recruiting participants from a heterogeneous population was a persistent challenge throughout my research, a problem encountered by other researchers working with special needs populations [290]. Because gender difference have also been reported to affect performance in orientation tasks [171], I have put much effort throughout my research into building diversified groups of participants composed of congenitally blind and adventitiously blind participants, males and females.

Sections 3.4 and 3.3 describe why SAM preserves the form factor and the resolution of

TMs. Streets and POIs are laid out as they are in TMs; they are located by a user with a stylus and sound feedback, an economical way to substitute information normally acquired by touching the embossed features of TMs. Like TMs, SAM supports free exploration, relying on the readers ability to locate streets, intersections, and points of interest. Thus, a SAM user deploys skills acquired from TMs to read spatial auditory maps.

Because SAM is the same as TMs with respect to resolution, and because streets and landmarks may have any geometry in SAM, it can provide an environment a TM can. SAM implements two families of tools, foveal and peripheral. Foveal tools afford high acuity for tracing; peripheral tools allow a user to reach remote locations and encourage the formation of exocentric knowledge. The digital format of the maps in SAM makes the distribution almost costless and eases the creation and adaptation of environmental maps, because they can be drawn by hand with a stylus.

Maps would usually be read on download, like ebooks. A tactile representation of a spatial auditory map can also be built by placing an inexpensive styrofoam or tactile drawing film [303] on the tablet. The resulting TM can then be brought into the environment for reading on location or replaced on the tablet for a multimodal experience.

Maps of malls and bus terminal maps are frequently requested[2], so SAM was evaluated using the maps of two malls. These malls, however, required more complexity than the TM representation in Figure 1.1. They needed many landmarks and corridors intersecting, turning, and ending in areas overlaid by landmarks. Post interviews with participants suggested that the reading difficulty of the two maps was balanced: three viewed them as equal, two thought TM was harder and two thought the opposite. However, all the participants thought the mall layout was more complex to understand in the TM than in SAM. In fact, the two OMS who assessed the participants' reconstructions identified challenges with both maps. According to them, the map represented by TM made it difficult for a blind person to figure out how to access certain stores, while the map represented by SAM made it difficult to maintain appropriate distance between landmarks.

The results of the assessment in Chapter 7 suggest that, after four hours of training, SAM equalled TM in performance for a blind reader building a cognitive maps of an unknown environment. SAM also equalled TM in terms of acquisition rate and exceeded TM in terms of autonomy. However, a closer look at the dimension scores, explained in Chapter 7: landmark-to-landmark (LL), corridor layout (CL), topology (T) and overall quality (Q), suggests that SAM outperformed TM in the CL dimension. Why is this?

I argued in Section 7.3.4 that attention could be the reason. The task of tracing curves with sound is a structured one, which enforces a rigid strategy: with SAM, stores are integrated to corridors which forces a user to pay a great attention to corridor characteristics to learn the mall. With TM, a user, who naturally has his attention directed to stores,

[2] According to the two OMSs interviewed in Chapter 3

is on his own in terms of strategy, which may or may not be optimal, and which has a direct effect on how he directs his attention towards map characteristics. Because different aspects of a task require focussed attention and mental effort on different characteristics, different memories are produced.

I also questioned the meaning of the reciprocal relationship I observed between LL, T and Q, one dimension being the linear prediction of the others and the status of independence in CL. It is possible that similar capabilities are used to do the aspects of the task attached to the LL, T and Q dimensions, but quite different mechanisms are needed for CL. I suggested that building store relationships, the commonality between these three dimensions, necessitates a greater ability of exocentric coding (Section 7.3.4). Accordingly, learning corridors requires a greater ability of egocentric and sequential coding, something SAM makes full use of in its design.

8.1 Doctoral Contributions

My thesis makes two types of contribution to Computer Science: ethnographic and technological. The ethnographic contribution is provided by empirical data on blind persons and OMS experts, collected through observation, interviews and a questionnaire. This data, found in Chapters 3 and 7, will benefit other HCI researchers interested in providing spatial information to blind users.

The technological contributions are a comprehensive set of design principles for the integration of sound feedback communication into the HCI of map for blind persons, with novel user-interaction methods I created for real-time curve-tracing in a multi-curve and multi-landmark environment.

The review of previous work in Chapter 2 establishes two points: a lack of a design solution that satisfies the ecological constraints specified in this research, and confirms that several methods I adopted are widespread. Like many researchers, I encode elevation with pitch and azimuth with stereo balance[3]; use synthetic speech to utter the name of the features on a map[4]; access information from stylus interactions[5], and follow a user-centric methodology to elicit the users' requirements and design a prototype.

However, SAM, which is built into a set of modular tools integrated together and accessible through stylus interactions, contributes to an ongoing action research dealing with innovative map systems design and evaluation. Much of my contribution stands out in the attention and integration of details:

[3] [77, 80, 81, 82, 82, 90, 91, 92, 93, 96, 97, 99, 100, 110]
[4] [68, 69, 70, 75, 86, 87, 88, 111, 121]
[5] [81, 82, 96, 97, 99, 100]

1. I developed a sensory substitution system for tracing curves with sound, which consists in compressing the bandwidth of audition to the size of the fingertip (Section 1.4 and 5.4.1).

2. I adapted a purified representation of map content for audition, by preserving the modality independent information and discarding the rest, thus reducing the amount of information relayed to the auditory sense (Section 3.5.2).

3. I built a tonal interface, thus enhancing curve tracing accuracy and aesthetics (Section 4.2.4).

4. I developed a presentation protocol that describes spatial characteristics in a language that respects the culture of its users (Section 4.6).

5. I examined directional models in blind persons: blind persons naturally choose the ecological model - as opposed to the tutored model (Section 5.2.2).

6. I examined spatial guidance in blind persons: blind persons naturally interpret spatial cues in terms of map reference. (Section 5.2.2).

7. I designed an effective scanline sensor and determined its optimal resolution. Scanline sensors were found ineffective in previous research (Section 5.4.1).

8. I adapted a presentation model built in real environment navigation systems: SAM, like beacon systems, represents space relative to the user and not absolute to the map (Section 5.4.1).

9. I designed a non-symmetrical encoding for effective curves tracing with non-speech sound feedback (Section 5.5.1).

10. I designed Locatones, which are spatial sounds that give the user total freedom in how to reach a destination in open space (Section 5.5.2).

11. I elaborated new reading strategies that allow free map exploration but structure reading rules (Section 6.3).

While current research is still struggling to communicate singular trajectories to blind users, SAM succeeds at conveying multi-curve and multi-landmark information, allowing a blind user to acquire knowledge and understand the geometry of space displayed by the system. No other work has approached this level of performance until now.

The functional prototype that resulted from my research and the empirical results I described in Chapter 7 provide significant evidence that *a state-of-the-art sound interface implemented on an off-the-shelf computer system can assist blind persons in building cognitive maps of unknown environments.*[6]

8.1.1 Other Benefits

In the same way Emacs [304] is a text editor, an email client, a web browser, a numerical calculator and more, SAM could potentially be used for 'a whole slew of different applications like learning to write [signature], game play, spreadsheet navigation, an add-on to screen reader access, simple drawing creations for braille embossers, and a math graphics reader' (M1 in Chapter 7). Indeed, mathematics can prove very challenging to younger blind children as they are required to learn about geometry, graphs and formulas that rely on the spatial layout of data to provide important information.

From a broader perspective, there are times when haptic output devices are not available due to space constraints, an operator's visual system is overloaded, visual attention is diverted or there is not enough bandwidth to send visual feedback to an operator, e.g., over air waves. In these situations, providing sonic directional information may be necessary either to replace or augment vision. This research offers an alternative communication method to visualize space with a modality other than vision.

8.2 Three Factors that Contributed to the Success of SAM

This research seeks to establish design principles for the integration of sound into the map interfaces used by blind persons. I sum up three factors that are responsible for the success of this research.

> **First design principle.** *Determine the appropriate modalities, their roles and output device for the task.*
>
> Absolute space is perceived by kinaesthetic feedback in SAM, because audition is an inferior modality for visualizing space. Indeed, blind persons rely on kinaesthesia to place and retrieve things with great precision. A digitalizing tablet (uni- or multi-touch) is effective because it encompasses the form factor of TMs and interact with

[6] I am referring the reader to Section1.1 on page 4.

the user via absolute position, which allow a blind person interacting with a digitizing tablet to rely on spatial aptitude when exploring two-dimensional space.

Previous work in Chapter 2 showed that redundant encoding of cursor position with sound fails to assist map reading and curve tracing tasks. Instead, I use sound to encode the position of features near the user's location. Because humans do not hear things that are far away, the bandwidth of audition is used to display items near the user, with distal items suppressed. Hence, absolute position in space is mapped to hand position, while items near a location, which are normally found by touch, are displayed sonically.

Second design principle. *Use knowledge from psychoacoustics to build a task-dependent auditory display.*

The ability of humans to localize directionality of sound sources is poor. In addition, synthetic methods for communicating the location of objects in space are expensive and listener specific. Non-expensive alternatives exist, like generic binaural head related transfer functions, but are unreliable, especially when elevation judgements are needed (Section 4.2.2).

Fortunately, empirical research on sound localization examined and corroborated a pitch-elevation correspondence in humans, known as the Pratt effect. Unfortunately, humans hear pitch increasing or decreasing relatively, but not absolutely (Section 4.2.1), which makes precise judgments difficult. Tonality, on the other hand, is perceived by the vast majority of Westerners and can turn relative pitch information into absolute one, because it allows a listener to identify a certain pitch value, the tonic, as a reference at centre of a tonal scale (Section 4.2.4). Uniform tonality makes it possible for a user to recognize a centre of something, which is applied to inform the user when the stylus is sitting on a curve.

SAM implements a tonal interface which is established by governing every nonspeech sound by tonal laws: all non-speech sounds in SAM share the same tonality. This design decision also benefits the aesthetics of the interface, which displays a series of different 'musical segments', sometimes overlaid, all conform to the same tonality, which is how Westerners' music is composed.

Also, sound is obtrusive and temporal, which is why SAM displays the least information possible, but offers the choice of obtaining a rich set of details on demand. SAM adopts a non-visual language composed of speech and non-speech sounds to give concise instructions to its user and rapid feedback on actions.

Third design principle. *Allow free map exploration but provide structured reading rules.*

In interaction design, a *forcing function* is a behaviour-shaping restriction that prevents undesirable user input [137]. In the same way a word processor leaves a user free to type anywhere in a page but forces text to remain within margins, SAM supports free map exploration but enforces reading rules meant to produce superior results. The rules must be determined from observation, motivated by visuospatial cognition theory and validated by pilots and experiments.

I provide three concrete examples of reading rules to demonstrate this: the foveal tools require a user to trace streets to discover landmarks along them; the alignment tool requires a user to move the stylus across space to discover alignment of features; the front tool requires the user to point the stylus in the direction of features to perceive them. These tools enforce behaviour-shaping restrictions as they attempt to synchronize kinaesthetic input with sound feedback to display spatial information. However, the spatial information that is provided to the user is given piece-by-piece, each portion spatially connected to one another and in compliance to specific reading rules.

For example, a user makes full use of egocentric coding by discovering landmarks along streets with the tracing tool, which helps him building structures of things pertaining to one another. Then sitting on a landmark, which is connected to a street he already knows, the user can look around with the front tool, building exocentric relationships using a tool that enforces its own exploration strategies based on visuospatial cognition theory, but also affords free exploration.

8.3 Future Work

I now discuss possible future research and propose two directions I would like to examine further: research to expand SAM on a wider scale, and research that would refine and improve the curve tracing and directional cues.

8.3.1 Expansion of SAM

The design of an effective writer interface is central to the success of SAM. The writer interface, which was implemented to feed data to the reader interface during the study, needs to be developed further. The same user-centric process should be adopted to capture detailed information about how orientation and mobility specialists (OMS) interact with TMs in their work environments.

SAM users could then evolve into a community analogous to *YouTube.com*. Blind users could access a comprehensive database of maps stored at a central location, they could discuss and rate. OMS, and volunteers supervised by OMS, could access the writer interface to draw, store, duplicate, modify and comment maps. Every map would contain metadata informing readers about the location represented, its scale, the creator's name and title, date of creation, the name of the persons who modified it, when, and how. Again, contextual design methods would be used to gather data from blind travellers and OMS cartographers to guide the building of an effective server application.

Possibly, SAM could use Google's geo database to offer automated map creation. To do that, a Keyhole Markup Language (KML) translation layer is needed to read geographic data from Google's servers. *KML* is an XML-based language schema for formulating geographic annotation and for visualizing two-dimensional maps [305].

Fortunately, SAM's data structure is compatible with KML's *placemarks*, which are the same as POIs in SAM, and *paths*, which, like curves in SAM, are strings of point-coordinates. The translation layer would also be responsible for optimizing the resolution of the map. Navigation mechanisms are needed to move through the content of a map in different cardinal directions, something Jacobson has already examined [22]. This feature requires particular attention because the user must be able to maintain his current location mentally during translation.

The recent advent of affordable multi-touch GPS-enabled systems like Apple's iPad™, open up many new possibilities. Clearly, multi-touch output sonification should be studied comprehensively; however, a multi-touch version of SAM would use the knowledge developed in my research as a starting point. I provide four plausible examples using multi-contact points to trigger and feed the foveal and peripheral tools.

With one finger: the sensor and sound encoding would sonify the curve tracing finger.

With two fingers: sliding the index and the major fingers of the dominant hand on the sensitive surface would launch the alignment tool and delimit its bounds.

With three fingers: three contacts could launch the front tool. The Braille finger would represent the location of the observer while the index and middle fingers of the non-dominant hand, spread apart, could hold the two other corner locations of the triangle, giving a user full control on distance range and FoV angle.

With four fingers: four contacts could launch the row/column/cell tools. The four contact locations would determine a row/column/cell region's bounds. In addition, the embedded-GPS could determine the current outdoor position of a traveller on a Google map, so a SAM user could navigate the neighbourhood by tracing the streets around his current location.

8.3.2 Directional Cues

To start, an individual assessment of the peripheral tools should be conducted. I need to measure how effective these tools are for building cognitive maps. I also need to verify how successful the tools are for encouraging formation of exocentric knowledge.

In the current design, the sensor is a scan line orthogonal to the stylus direction. Novice users find the design difficult when the stylus moves perpendicular to a straight curve and towards it because the sensor remains parallel to the curve. Reshaping the sensor into a $\perp$ shape would resolve the problem: the horizontal line is then orthogonal to movement direction (as in the current design) and the vertical bar, a lower resolution sensor, points in the direction of movement.

The benefits and drawback of a dynamic cursor should be investigated. The spatial distribution of sensitive regions along the sensor could vary according to the local geometry of a curve. For example, a linear distribution may be preferable in curvature areas, while the distribution adopted in this research may be preferable in straight areas.

The experiment in Chapter 7 reported difficulties in participants with distance perception between landmarks. I propose adjusting SAM's presentation rate with stylus speed in order to sensitize users to distance displacement. I would like to verify if this redundant treatment to the current kinaesthetic feedback can improve judgement.

A close investigation of the inverse polarity phenomenon I observed in Chapter 5 could benefit the HCI and psychoacoustic communities. Measuring the ratio of users (blind and sighted persons, musicians and non-musicians) who manifest pitch inversion could contribute to a greater body of knowledge about spatial encoding and provide insights into the Pratt effect.

Recently, I realized that Fitt's law [306] could be used to compare the performance of different encodings and identify bottlenecks in improving their design [24]. For example it could be interesting to assess the effectiveness of different locatone encodings.

Echolocation should be revisited, but with more blind participants. It scored marginally worse than the encoding I adopted. Echolocation provides a navigation-like experience on a map, with the direction of features signalled by the viewpoint of the perceiver in the display.

Finally, the study of the MVC pattern for auditory displays and alternative interfaces, as explained in Section 8.1.1, deserves further attention. Although the core principle of decoupling data access, domain-logic, and system response from user interaction remains the same for all systems, implementation details vary because a design cannot assume persistence in the display modality. Sequencing sound elements raises interface design challenges because sound is perishable and extended in time: auditory presentation must be repeated to remain before the user, who perceives only one stream at a time. The adaptation of the MVC pattern as proposed in this thesis is a starting point. I would like to revisit my solution, probably refine it, situate it in context, and propose it as a possible architecture for auditory display design. This work could create new knowledge in HCI and could serve as a starting point for other auditory display researchers.

8.3.3 Addendum

Research concurrent with mine, of which I was unaware, came to my attention when the thesis editing was well advanced. Su et al. [307] implemented Timbremap, a sonic map system for handheld devices. The system allows users to use their fingertip to trace curves on a flat surface. It provides binaural cues in the lateral direction and spearcon cues modulated in pitch in the vertical direction ('up' & 'down'), plus an area hinting tool that helps a user regain contact with curves. Like [22], Timbremap implements a panning interaction technology, which is atlas-like, a concept explained in Section 2.1.4, to display maps larger than the screen. Unfortunately the participants reported the feature difficult to use.

Same features are shared by both systems: like SAM, Timbremap implements a non-symmetrical encoding that allows a user to trace streets effectively - the six blind participants who assessed Timbremap could identify 85% of the curves they traced using a multiple choice response format; Timbremap represents the environment simply: with curves and points; Timbremap uses notification cues to signal intersections and POIs; Timbremap uses a speech synthesizer to speak the names of POIs and streets; and Timbremap relies on a user's kinaesthesia to understand space.

However, SAM differs from Timbremap in several ways. Unlike SAM, Timbremap interprets space statically, which requires users to perform mental rotations to understand directions – SAM dynamically adapts direction with a user's orientation and hand movements on the map, which reduces user's cognitive load; Timbremap implements many fewer tools than SAM and is missing important ones: for example, the authors reported that some participants had to search for as long as three minutes to find a landmark – SAM's index tool allows a user to walk directly to any landmark on a map; Timbremap only displays curve tracing cues and features' name information – SAM speaks an elaborate language developed for the user and the task; Timbremap speaks 'up' & 'down' cues to guide the user in space, which probably explains the low aesthetic score it received during its assessment (also, recall from Section 2.2.2 the negative impact on navigation performance reported by Walker and Lindsay [111] when concurrent speech information is added to non-speech sounds) – SAM innovates with a tonal interface, which was reported as aesthetically pleasing and unambiguous by the participants; the Timbremap paper does not address the question of map/sensor resolution – its maps are at a much lower resolution than SAM, probably the result of using fingertip input; the Timbremap paper does not discuss perceptual grouping of sound, so it is unclear how much consideration was put into sound streaming issues: this could explain the criticism raised by participants who reported sounds being too different from one another; finally, Timbremap's assessment is qualitative – SAM is assessed both qualitatively and quantitatively (Chapters 5 & 7).

Nonetheless, the results reported for Timbremap provide further evidence that tracing curves with sound is feasible when the encoding is not ambiguous and kinaesthesia is used to perceive the geometry of space. These results with the ones reported in my thesis suggest that the task is not onerous because two independent groups of blind participants could understand non-trivial geometry without prior knowledge of their shape, with relatively short training.

Future work could take advantage of both studies. I would be interested in re-using the methods discussed in Chapter 5, along with SAM's curve tracing encoding, to investigate the resolution of fingertip input on a handheld device for the task of tracing curves with sound. This would allow me to learn two important things: first, to determine the minimum map resolution task/fingertip and compare it with the stylus and second, to determine if the low resolution found in Timbremap is the consequence of using the fingertip as an input device - it could be its sound encoding (e.g. low refresh rate). If this input device is rejected, then there is also the possibility the authors simply settled for a suboptimal resolution.

APPENDICES

Appendix A

Glossary

Adventitiously Blind is a person that became blind later in life.

Amodal is when different senses yield different sensations, but the same information about the world.

Attack is the time taken for initial run-up of level from 0 dB to peak in a sound. The onset of a sound.

Audition is the faculty of hearing.

Auditory Display is a computer interface that uses sound to communicate information to a user.

Beacon is a physical or virtual emitting sound station strategically positioned in space a blind person can refer to position himself.

Beat is the rhythmic structure and the succession of sounds of equal durations.

Binaural Head is a binaural recording mannequin head that has a mic in each ear.

Blind Person is a person lacking visual perception due to physiological or neurological factor. The National Research Council defines blindness as follows: 'a person is diagnosed blind when the smallest detail that can be resolved visually in the better eye with refractive errors corrected is ten minutes of arc or greater, or the horizontal extent of the visual field with both eyes open is less than or equal to twenty degrees' [29].

Braille Embosser is a printer that renders text as Braille.

Canadian National Institute for the Blind (CNIB) is a volunteer agency and charitable organization dedicated to assisting Canadians who are blind or living with vision loss.

Cognitive Map is an internal representation, the schema of a set of geographic locations in the human mind.

Congenitally Blind (CB) A person blind from birth.

Decay is a sound subsequent run down time interval from the attack peak to the specified sustain level.

Digitizing Tablet is a computer tablet meant to provide a one-to-one relationship between its surface and the computer's screen real estate.

Echolocation is the action of determining the location of objects by measuring the time it takes for an echo to return from it.

Ecological Model is moving around in the environment that remains fixed.

Egocentric Coding is understanding space in relation to oneself.

Electronic Travel Aid (ETA) is a form of assistive technology designed to enhance mobility for the blind pedestrian.

Envelop of a Sound is the overall shape in intensity over time of a sound. It is often described into four basic stages: attack, decay, sustain and release.

Exocentric Coding is understanding space in relation to external cues.

Foveal Tools are tools developed in SAM to allow a user to trace a curve effectively and discover points of interest along a curve.

Frequency-Balance Mapping is a method that sonifies elevation by frequency, mapping high frequencies with high coordinates & low frequencies with low coordinates, and azimuth with stereo balance.

Graphical User Interface (GUI) is a user interface based on icons, pictures and menus.

Haptic is relating to the sense of touch.

Head Related Transfer Function (HRTF) is a set of mathematical transformations that can be applied to filter sounds and simulate the physiological characteristics of the outer ear in humans, which allows a listener to perceive sounds in elevation.

Intensity (Sound Intensity) is the amount of energy transmitted by sound and is expressed in decibels.

Jacques Bertin is a French cartographer and theorist, known from his book Semiology of Graphics edited in 1967.

JAWS is an application for visually impaired users, produced by the Blind and Low Vision Group at Freedom Scientific of St. Petersburg, Florida, USA.

Kinaesthesia is the feeling of motion. Relating to sensations originating in muscles, tendons and joints.

Loudness is the perceived intensity of a sound.

Low Fidelity Prototype is sketchy and incomplete interface that is used to quickly test broad concepts with user participants.

Micro-Capsule Paper is a swell paper with a special coating of chemical reactive, which produces a relief upon pressure or heat.

MIDI is a protocol to transfer note and system information to or from a computer to or from other midi devices (e.g. a music synthesizers).

Modality Independent Information is the characteristics that must be captured and preserved when transforming graphically represented information from the visual to an alternative sense, like audition.

Mode Switching A user performs a system mode switch, or changes the system's state, when he needs access to new system resources like a new tool.

Model-View-Controller (MVC) is an architecture programming pattern for separating data, logic and presentation.

Multi State Tools (MTS) are tools that can be activated by multiple tools in SAM but behave differently depending on the tools that activate it.

Musician's Uncertainty Principle When two frequencies differ by Δf, then humans need a time of $\frac{1}{\Delta f}$ to notice.

Notification Sounds are sounds emitted by SAM to inform the user the proximity of a curvature, curve extremity, point of interest or curve intersection. They are implemented as beacons, i.e., being at closer proximity to a notification icon produces a louder signal and being out of range produces no signal.

Nottingham Kit is a standard set of symbols and material developed in 1970s by J.D. Armstrong and G.A James to make tactile maps using plastic-formed technology.

OpenAL (JoAL for Java) is a free software cross-platform audio API. Its style and conventions resemble that of OpenGL for graphics.

Orientation and Mobility Specialist (OMS) is an expert professional who train persons with low vision to move about safely in their homes and travel by themselves.

Peripheral Tools are tools developed in SAM to allow a user to perceive features that are far from their current location on the map. The tools were also developed to encourage exocentric coding in the user.

PHANToM is an expensive haptic device manufactured by SensAble Technologies that delivers force feedback in 3D. (For more, see: http://www.sensable.com/)

Picture Book consists of a Velcro surface onto which plastic geometric shapes and lines can be assembled to produce quick illustrations of spatial relationships.

Pitch is the property of a sound that varies with variation in the frequency of vibration.

Point of Interest (POI) is a fiducial point, the location of which is well known by the user.

Pratt Effect originates in associations, developed early in life, between elevation and pitch: short wavelength sound being described as high, long wavelength sound as low.

Pro Tools is a Digital Audio Workstation platform.

Release is the time taken for the sound to decay from the sustain level to 0 dB.

Reverberation is the repetition of a sound resulting from reflection of the sound waves on object near the emitting source and the listener.

Screen Real Estate is the amount of space available on a computer screen for an application to provide output to a user.

Sensory Substitution is the action of transforming the characteristics of one sensory modality into stimuli of another sensory modality.

Sequential Coding is understanding space in relation to a sequential collection of movements.

Sighted Person is a person that has the ability to see, i.e, not blind and not visually impaired.

Sound is audible vibrations that travel through the air.

Streaming is a sense of connectedness making successive sounds appear to arise from the same sound source.

Stylus is a pen-shaped device that is used to input commands to a digitizing tablet.

Sustain is the constant intensity of the sound prior release.

Synthetic Speech a computer model that converts normal language text into audible speech.

Tangible or Tactile Map (TM) is a 3D relief model of a geographical map that can be felt by touch and interpreted by a blind reader.

Teleportation (hand) is the action of voluntarily loosing contact with the tactile map at a location and then regaining contact at another.

Thermoform is the the process of heating a plastic sheet to forming it into a finished 3D shape by means of heat or pressure.

Timbre is the character or quality of a sound.

Tonality is a system of music based on hierarchical pitch relationships organized into major and minor keys or tonics.

Tonic is the keynote of a major or minor scale, that is to say, the first note of a diatonic scale.

Tour-Based System is a non-interactive sound interface that presents location-dependent information to a listener.

Tutored Model is moving the world around while remaining in a fixed location.

Visually Impaired Person is a person who has difficulty seeing even when wearing corrective glasses.

Window, Icon, Menu, Pointing device (WIMP) denotes a style of user-computer interaction using these elements.

Wireless Fidelity (WI-FI) is a broadband service that provides wireless access to computers.

Wizard of Oz (WoZ) is a technique that has participants interact with a computer system they 'believe' is autonomous, but that is actually being operated the researcher.

A.1 Edit Distance Algorithm

Explanations: the EditDist function is inspired by the Levenshtein distance algorithm [308], but modified to compare two curves. The function takes two traces of n x 2, n being the number of points, which can be different for each curve. The threshold_of_tolerence parameter is the tolerated Euclidean distance between two points to make them equal. The buildMtrxDist(trace1, trace2) is a simple function (not shown here) that pre-computes pairwise Euclidean distances between every point in trace1 and trace 2.

```
function [ d, D, threshCount] = EditDist(trace1, trace2,  threshold_of_tolerence)

  distMtrx = buildMtrxDist(trace1, trace2);
  [n1, n2] = size(distMtrx);

  %Initialize dynamic matrix D with appropriate size:
  D = zeros(n1+1,n2+1);

  %Initialization of D
  for i = 1:n1
      D(i+1,1) = D(i,1) + distMtrx(i,1);
  end;
  for j = 1:n2
      D(1,j+1) = D(1,j) + distMtrx(1, j);
  end;
  threshCount = 0;

  % Algorithm on D
  for i = 1:n1
      for j = 1:n2
          if distMtrx(i, j) <= threshold_of_tolerence
              cost = 0;
              threshCount = threshCount + 1;
          else
              cost = distMtrx(i,j) ;
          end
          [D(i+1, j+1), index] = ...
                   min([D(i, j)+cost  D(i+1, j)+cost  D(i, j+1)+cost ]);
      end
  end
  d = D(n1+1, n2+1);
```

A.2 Analysis of Variance

DESIGN

Dependent variables

Name	Code
Score	Scr

Type of analysis: OLS ANOVA

Factors

Name	Code	Nested in	F/R	Kind
Dir	Dir	()	Fix	Disc
Encoding	Eng	()	Fix	Disc
SsCat	SCt	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 3 - way

No Modifications

No Selector

No variance variable

RESULTS

General Results

1344 total cases

ANOVA

Analysis of Variance For **Score**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	738.107	738.107	7318	≤ 0.0001
Dir	1	14.1681	14.1681	140.47	≤ 0.0001
Eng	2	36.3651	18.1825	180.27	≤ 0.0001
Dir*Eng	2	17.5635	8.78175	87.068	≤ 0.0001
SCt	1	21.0125	21.0125	208.33	≤ 0.0001
Dir*SCt	1	10.2722	10.2722	101.85	≤ 0.0001
Eng*SCt	2	12.8889	6.44444	63.894	≤ 0.0001
Dir*Eng*SCt	2	10.2222	5.11111	50.675	≤ 0.0001
Error	1332	134.347	0.100861		
Total	1343	257.893			

Results for factor Dir*Eng*SCt

Coefficients

Coefficients of: Score on Dir*Eng*SCt

Level of Dir*Eng*SCt	Coefficient	std. err.	t Ratio	prob
"B","GG","H"	0.05671	0.0165	3.436	0.0006
"B","GG","V"	-0.05671	0.0165	-3.436	0.0006
"B","GX","H"	0.06019	0.01248	4.824	≤ 0.0001
"B","GX","V"	-0.06019	0.01248	-4.824	≤ 0.0001
"B","XX","H"	-0.1169	0.01248	-9.37	≤ 0.0001
"B","XX","V"	0.1169	0.01248	9.37	≤ 0.0001
"S","GG","H"	-0.05671	0.0165	-3.436	0.0006
"S","GG","V"	0.05671	0.0165	3.436	0.0006
"S","GX","H"	-0.06019	0.01248	-4.824	≤ 0.0001
"S","GX","V"	0.06019	0.01248	4.824	≤ 0.0001
"S","XX","H"	0.1169	0.01248	9.37	≤ 0.0001
"S","XX","V"	-0.1169	0.01248	-9.37	≤ 0.0001

Expected Cell Means

Expected Cell Means of: Score on Dir*Eng*SCt

Level of Dir*Eng*SCt	Expected Cell Mean	Cell Count
"B","GG","H"	0.5208	48
"B","GG","V"	166.5e-18	48
"B","GX","H"	0.8611	144
"B","GX","V"	0.04861	144
"B","XX","H"	0.9028	144
"B","XX","V"	0.9375	144
"S","GG","H"	0.6875	48
"S","GG","V"	0.7917	48
"S","GX","H"	0.9097	144
"S","GX","V"	0.7361	144
"S","XX","H"	0.9444	144
"S","XX","V"	0.9097	144

Scheffe Post Hoc Tests

	Difference	std. err.	Prob
"B","GG","V" - "B","GG","H"	-0.520833	0.06483	20.5391e-15
"B","GX","H" - "B","GG","H"	0.340278	0.05293	1.45255e-9
"B","GX","H" - "B","GG","V"	0.861111	0.05293	0
"B","GX","V" - "B","GG","H"	-0.472222	0.05293	0
"B","GX","V" - "B","GG","V"	0.0486111	0.05293	0.656008
"B","GX","V" - "B","GX","H"	-0.8125	0.03743	0
"B","XX","H" - "B","GG","H"	0.381944	0.05293	8.10552e-12
"B","XX","H" - "B","GG","V"	0.902778	0.05293	0
"B","XX","H" - "B","GX","H"	0.0416667	0.03743	0.538281
"B","XX","H" - "B","GX","V"	0.854167	0.03743	0
"B","XX","V" - "B","GG","H"	0.416667	0.05293	70.3881e-15
"B","XX","V" - "B","GG","V"	0.9375	0.05293	0
"B","XX","V" - "B","GX","H"	0.0763889	0.03743	0.124992
"B","XX","V" - "B","GX","V"	0.888889	0.03743	0
"B","XX","V" - "B","XX","H"	0.0347222	0.03743	0.65039
"S","GG","H" - "B","GG","H"	0.166667	0.06483	0.0370052
"S","GG","H" - "B","GG","V"	0.6875	0.06483	0
"S","GG","H" - "B","GX","H"	-0.173611	0.05293	0.00471308
"S","GG","H" - "B","GX","V"	0.638889	0.05293	0
"S","GG","H" - "B","XX","H"	-0.215278	0.05293	269.257e-6
"S","GG","H" - "B","XX","V"	-0.25	0.05293	15.7045e-6
"S","GG","V" - "B","GG","H"	0.270833	0.06483	171.618e-6
"S","GG","V" - "B","GG","V"	0.791667	0.06483	0
"S","GG","V" - "B","GX","H"	-0.0694444	0.05293	0.423125
"S","GG","V" - "B","GX","V"	0.743056	0.05293	0
"S","GG","V" - "B","XX","H"	-0.111111	0.05293	0.110846
"S","GG","V" - "B","XX","V"	-0.145833	0.05293	0.0227165
"S","GG","V" - "S","GG","H"	0.104167	0.06483	0.275349
"S","GX","H" - "B","GG","H"	0.388889	0.05293	3.2343e-12
"S","GX","H" - "B","GG","V"	0.909722	0.05293	0
"S","GX","H" - "B","GX","H"	0.0486111	0.03743	0.430462
"S","GX","H" - "B","GX","V"	0.861111	0.03743	0
"S","GX","H" - "B","XX","H"	0.00694444	0.03743	0.982935
"S","GX","H" - "B","XX","V"	-0.0277778	0.03743	0.759307
"S","GX","H" - "S","GG","H"	0.222222	0.05293	157.642e-6
"S","GX","H" - "S","GG","V"	0.118056	0.05293	0.0835237
"S","GX","V" - "B","GG","H"	0.215278	0.05293	269.257e-6
"S","GX","V" - "B","GG","V"	0.736111	0.05293	0
"S","GX","V" - "B","GX","H"	-0.125	0.03743	0.00387292
"S","GX","V" - "B","GX","V"	0.6875	0.03743	0
"S","GX","V" - "B","XX","H"	-0.166667	0.03743	53.1957e-6
"S","GX","V" - "B","XX","V"	-0.201389	0.03743	603.249e-9
"S","GX","V" - "S","GG","H"	0.0486111	0.05293	0.656008
"S","GX","V" - "S","GG","V"	-0.0555556	0.05293	0.576613
"S","GX","V" - "S","GX","H"	-0.173611	0.03743	23.1834e-6
"S","XX","H" - "B","GG","H"	0.423611	0.05293	26.0902e-15
"S","XX","H" - "B","GG","V"	0.944444	0.05293	0
"S","XX","H" - "B","GX","H"	0.0833333	0.03743	0.0842429
"S","XX","H" - "B","GX","V"	0.895833	0.03743	0
"S","XX","H" - "B","XX","H"	0.0416667	0.03743	0.538281
"S","XX","H" - "B","XX","V"	0.00694444	0.03743	0.982935
"S","XX","H" - "S","GG","H"	0.256944	0.05293	8.46817e-6
"S","XX","H" - "S","GG","V"	0.152778	0.05293	0.0157244
"S","XX","H" - "S","GX","H"	0.0347222	0.03743	0.65039
"S","XX","H" - "S","GX","V"	0.208333	0.03743	223.434e-9
"S","XX","V" - "B","GG","H"	0.388889	0.05293	3.2343e-12
"S","XX","V" - "B","GG","V"	0.909722	0.05293	0
"S","XX","V" - "B","GX","H"	0.0486111	0.03743	0.430462
"S","XX","V" - "B","GX","V"	0.861111	0.03743	0
"S","XX","V" - "B","XX","H"	0.00694444	0.03743	0.982935
"S","XX","V" - "B","XX","V"	-0.0277778	0.03743	0.759307
"S","XX","V" - "S","GG","H"	0.222222	0.05293	157.642e-6
"S","XX","V" - "S","GG","V"	0.118056	0.05293	0.0835237
"S","XX","V" - "S","GX","H"	-111.022e-18	0.03743	1
"S","XX","V" - "S","GX","V"	0.173611	0.03743	23.1834e-6
"S","XX","V" - "S","XX","H"	-0.0347222	0.03743	0.65039

Figure A.1: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
Score	Scr

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Sensitivity	Sny	()	Fix	Disc
frq	frq	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 2 - way

No Modifications

RESULTS

General Results

1344 total cases

ANOVA

Analysis of Variance For **Score**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	738.107	738.107	3887.5	≤ 0.0001
Sny	1	0.859484	0.859484	4.5267	0.0336
frq	2	2.74095	1.37047	7.218	0.0008
Sny*frq	2	0.144221	0.0721106	0.37979	0.6841
Error	1338	254.043	0.189868		
Total	1343	257.893			

Results for factor Sny*frq

Coefficients

Coefficients of: Score on Sny*frq

Level of Sny*frq	**Coefficient**	**std. err.**	**t Ratio**	**prob**
"c","l"	0.01534	0.01764	0.8692	0.3849
"c","n"	-0.01534	0.01764	-0.8692	0.3849
"d","l"	-0.008102	0.01764	-0.4592	0.6462
"d","n"	0.008102	0.01764	0.4592	0.6462
"q","l"	-0.007234	0.01601	-0.4518	0.6515
"q","n"	0.007234	0.01601	0.4518	0.6515

Expected Cell Means

Expected Cell Means of: Score on Sny*frq

Level of Sny*frq	**Expected Cell Mean**	**Cell Count**
"c","l"	0.776	192
"c","n"	0.7969	192
"d","l"	0.7396	192
"d","n"	0.8073	192
"q","l"	0.6563	288
"q","n"	0.7222	288

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
"c","n" - "c","l"	0.0208333	0.04447	0.896088
"d","l" - "c","l"	-0.0364583	0.04447	0.714659
"d","l" - "c","n"	-0.0572917	0.04447	0.436362
"d","n" - "c","l"	0.03125	0.04447	0.781268
"d","n" - "c","n"	0.0104167	0.04447	0.972942
"d","n" - "d","l"	0.0677083	0.04447	0.314123
"q","l" - "c","l"	-0.119792	0.0406	0.0130463
"q","l" - "c","n"	-0.140625	0.0406	0.00254788
"q","l" - "d","l"	-0.0833333	0.0406	0.122039
"q","l" - "d","n"	-0.151042	0.0406	0.00102264
"q","n" - "c","l"	-0.0538194	0.0406	0.415555
"q","n" - "c","n"	-0.0746528	0.0406	0.184787
"q","n" - "d","l"	-0.0173611	0.0406	0.912624
"q","n" - "d","n"	-0.0850694	0.0406	0.111712
"q","n" - "q","l"	0.0659722	0.03631	0.192354

Figure A.2: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
Score	Scr

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Sensitivity	Sny	()	Fix	Disc
frq	frq	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 2 - way

No Modifications

RESULTS

General Results

1344 total cases

ANOVA

Analysis of Variance For **Score**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	738.107	738.107	3887.5	≤ 0.0001
Sny	1	0.859484	0.859484	4.5267	0.0336
frq	2	2.74095	1.37047	7.218	0.0008
Sny*frq	2	0.144221	0.0721106	0.37979	0.6841
Error	1338	254.043	0.189868		
Total	1343	257.893			

Results for factor frq

Coefficients

Coefficients of: Score on frq

Level of frq	**Coefficient**	**std. err.**	**t Ratio**	**prob**
"c"	0.03675	0.01764	2.083	0.0375
"d"	0.02373	0.01764	1.345	0.1789
"q"	-0.06047	0.01601	-3.777	0.0002

Expected Cell Means

Expected Cell Means of: Score on frq

Level of frq	**Expected Cell Mean**	**Cell Count**
"c"	0.7865	384
"d"	0.7734	384
"q"	0.6892	576

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
"d" - "c"	-0.0130208	0.03145	0.917854
"q" - "c"	-0.0972222	0.02871	0.0033109
"q" - "d"	-0.0842014	0.02871	0.0137335

Figure A.3: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
DistErN	DsN

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Wdth	Wdh	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **DistErN**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	4.17928e6	4.17928e6	69.351	≤ 0.0001
Wdh	4	1.04461e6	261153	4.3336	0.0024
Error	151	9.09967e6	60262.7		
Total	155	10.1443e6			

Results for factor Wdh

Coefficients

Coefficients of: DistErN on Wdh

Level of Wdh	**Coefficient**	**std. err.**	**t Ratio**	**prob**
3	195.5	51.92	3.766	0.0002
15	-75.65	40.5	-1.868	0.0638
20	-87.2	34.71	-2.512	0.0130
30	-8.307	40.02	-0.2075	0.8359
40	-24.37	39.56	-0.6161	0.5388

Expected Cell Means

Expected Cell Means of: DistErN on Wdh

Level of Wdh	**Expected Cell Mean**	**Cell Count**
3	386.6	16
15	115.4	30
20	103.9	47
30	182.8	31
40	166.7	32

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
15 - 3	-271.172	75.99	0.0152247
20 - 3	-282.722	71.05	0.00439073
20 - 15	-11.5498	57.37	0.999795
30 - 3	-203.832	75.57	0.128039
30 - 15	67.3405	62.87	0.886187
30 - 20	78.8903	56.8	0.748713
40 - 3	-219.9	75.16	0.0786072
40 - 15	51.2722	62.39	0.953967
40 - 20	62.822	56.26	0.869822
40 - 30	-16.0683	61.86	0.999437

Figure A.4: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
LenRatio	LRo

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Wdth	Wdh	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **LenRatio**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	8100.29	8100.29	316.47	≤ 0.0001
Wdh	4	629.019	157.255	6.1438	0.0001
Error	151	3864.98	25.5959		
Total	155	4494			

Results for factor Wdh

Coefficients

Coefficients of: LenRatio on Wdh

Level of Wdh	**Coefficient**	**std. err.**	**t Ratio**	**prob**
3	5.103	1.07	4.77	≤ 0.0001
15	-0.928	0.8348	-1.112	0.2680
20	-1.949	0.7153	-2.725	0.0072
30	-0.85	0.8248	-1.031	0.3044
40	-1.376	0.8154	-1.688	0.0935

Expected Cell Means

Expected Cell Means of: LenRatio on Wdh

Level of Wdh	**Expected Cell Mean**	**Cell Count**
3	13	16
15	6.971	30
20	5.95	47
30	7.049	31
40	6.523	32

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
15 - 3	-6.03117	1.566	0.00657286
20 - 3	-7.05207	1.464	228.324e-6
20 - 15	-1.02091	1.182	0.945188
30 - 3	-5.95315	1.557	0.00717224
30 - 15	0.0780215	1.296	0.999998
30 - 20	1.09893	1.171	0.926766
40 - 3	-6.47938	1.549	0.00224791
40 - 15	-0.448208	1.286	0.998205
40 - 20	0.572699	1.16	0.993063
40 - 30	-0.52623	1.275	0.996532

Figure A.5: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
Time	Tim

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Wdth	Wdh	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **Time**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	296382	296382	401.34	≤ 0.0001
Wdh	4	14249.4	3562.36	4.8239	0.0011
Error	151	111510	738.477		
Total	155	125759			

Results for factor Wdh

Coefficients

Coefficients of: Time on Wdh

Level of Wdh	**Coefficient**	**std. err.**	**t Ratio**	**prob**
3	22.63	5.747	3.938	0.0001
15	-1.964	4.484	-0.4381	0.6620
20	-11.56	3.842	-3.008	0.0031
30	-3.106	4.43	-0.7011	0.4843
40	-6.001	4.38	-1.37	0.1727

Expected Cell Means

Expected Cell Means of: Time on Wdh

Level of Wdh	**Expected Cell Mean**	**Cell Count**
3	69.6	16
15	45.01	30
20	35.42	47
30	43.87	31
40	40.97	32

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
15 - 3	-24.5937	8.413	0.0789868
20 - 3	-34.1874	7.866	0.00128138
20 - 15	-9.59365	6.35	0.684397
30 - 3	-25.7357	8.365	0.0553631
30 - 15	-1.14196	6.96	0.999909
30 - 20	8.45169	6.288	0.77103
40 - 3	-28.6306	8.321	0.0217272
40 - 15	-4.03692	6.906	0.986831
40 - 20	5.55673	6.228	0.93856
40 - 30	-2.89496	6.848	0.996194

Figure A.6: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	Code
DistErN	DsN

Type of analysis: OLS ANOVA

Factors

Name	Code	Nested in	F/R	Kind
Enc	Enc	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **DistErN**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	4.17928e6	4.17928e6	70.053	≤ 0.0001
Enc	3	1.07612e6	358707	6.0126	0.0007
Error	152	9.06816e6	59659		
Total	155	10.1443e6			

Results for factor Enc

Coefficients

Coefficients of: DistErN on Enc

Level of Enc	Coefficient	std. err.	t Ratio	prob
"Bin"	-88.33	32.67	-2.704	0.0076
"Coll"	192.6	49.5	3.891	0.0001
"Echo"	-14.48	32.67	-0.4431	0.6583
"Flat"	-89.81	49.5	-1.814	0.0716

Expected Cell Means

Expected Cell Means of: DistErN on Enc

Level of Enc	Expected Cell Mean	Cell Count
"Bin"	105.7	62
"Coll"	386.6	16
"Echo"	179.5	62
"Flat"	104.2	16

Scheffe Post Hoc Tests

	Difference	std. err.	Prob
"Coll" - "Bin"	280.958	68.49	0.00112779
"Echo" - "Bin"	73.8561	43.87	0.42063
"Echo" - "Coll"	-207.102	68.49	0.0305397
"Flat" - "Bin"	-1.47623	68.49	0.999997
"Flat" - "Coll"	-282.434	86.36	0.0156634
"Flat" - "Echo"	-75.3324	68.49	0.750862

Figure A.7: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
LenRatio	LRo

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
Enc	Enc	( )	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **LenRatio**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	8100.29	8100.29	321.51	≤ 0.0001
Enc	3	664.384	221.461	8.7899	≤ 0.0001
Error	152	3829.62	25.1948		
Total	155	4494			

Results for factor Enc

Coefficients

Coefficients of: LenRatio on Enc

Level of Enc	**Coefficient**	**std. err.**	**t Ratio**	**prob**
"Bin"	-2.078	0.6714	-3.095	0.0023
"Coll"	4.971	1.017	4.886	≤ 0.0001
"Echo"	-0.7229	0.6714	-1.077	0.2834
"Flat"	-2.169	1.017	-2.132	0.0346

Expected Cell Means

Expected Cell Means of: LenRatio on Enc

Level of Enc	**Expected Cell Mean**	**Cell Count**
"Bin"	5.954	62
"Coll"	13	16
"Echo"	7.309	62
"Flat"	5.863	16

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
"Coll" - "Bin"	7.04895	1.407	35.1327e-6
"Echo" - "Bin"	1.35548	0.9015	0.521898
"Echo" - "Coll"	-5.69347	1.407	0.0013752
"Flat" - "Bin"	-0.0910484	1.407	0.999928
"Flat" - "Coll"	-7.14	1.775	0.00148226
"Flat" - "Echo"	-1.44653	1.407	0.787696

Figure A.8: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	Code
Time	Tim

Type of analysis: OLS ANOVA

Factors

Name	Code	Nested in	F/R	Kind
Enc	Enc	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **Time**
No Selector

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	296382	296382	403.28	≤ 0.0001
Enc	3	14050.5	4683.5	6.3727	0.0004
Error	152	111709	734.928		
Total	155	125759			

Results for factor Enc

Coefficients

Coefficients of: Time on Enc

Level of Enc	Coefficient	std. err.	t Ratio	prob
"Bin"	-5.937	3.626	-1.637	0.1037
"Coll"	23.51	5.495	4.278	≤ 0.0001
"Echo"	-2.577	3.626	-0.7108	0.4783
"Flat"	-14.99	5.495	-2.729	0.0071

Expected Cell Means

Expected Cell Means of: Time on Enc

Level of Enc	Expected Cell Mean	Cell Count
"Bin"	40.16	62
"Coll"	69.6	16
"Echo"	43.52	62
"Flat"	31.11	16

Scheffe Post Hoc Tests

	Difference	std. err.	Prob
"Coll" - "Bin"	29.4428	7.602	0.002462
"Echo" - "Bin"	3.35919	4.869	0.923959
"Echo" - "Coll"	-26.0836	7.602	0.00985358
"Flat" - "Bin"	-9.05536	7.602	0.701542
"Flat" - "Coll"	-38.4981	9.585	0.00151683
"Flat" - "Echo"	-12.4146	7.602	0.448342

Figure A.9: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
SpeedN	SpN

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
User	Usr	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **SpeedN**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	1694.68	1694.68	185.47	≤ 0.0001
Usr	3	339.366	113.122	12.381	≤ 0.0001
Error	152	1388.84	9.1371		
Total	155	1728.21			

Results for factor Usr

Coefficients

Coefficients of: SpeedN on Usr

Level of Usr	**Coefficient**	**std. err.**	**t Ratio**	**prob**
"C"	-1.318	0.4158	-3.17	0.0018
"G"	-1.627	0.4158	-3.912	0.0001
"K"	1.39	0.4158	3.343	0.0010
"S"	1.554	0.4308	3.608	0.0004

Expected Cell Means

Expected Cell Means of: SpeedN on Usr

Level of Usr	**Expected Cell Mean**	**Cell Count**
"C"	2.018	40
"G"	1.709	40
"K"	4.726	40
"S"	4.89	36

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
"G" - "C"	-0.3085	0.6759	0.976146
"K" - "C"	2.70825	0.6759	0.00156875
"K" - "G"	3.01675	0.6759	303.606e-6
"S" - "C"	2.87253	0.6944	999.615e-6
"S" - "G"	3.18103	0.6944	194.031e-6
"S" - "K"	0.164278	0.6944	0.996522

Figure A.10: ANOVA for Chapter 5

DESIGN

Dependent variables

Name	**Code**
Time	Tim

Type of analysis: OLS ANOVA

Factors

Name	**Code**	**Nested in**	**F/R**	**Kind**
User	Usr	()	Fix	Disc

Partial (Type 3) Sums of Squares

Interactions up to 1 - way

No Modifications

RESULTS

General Results

156 total cases

ANOVA

Analysis of Variance For **Time**
No Selector

Source	**df**	**Sums of Squares**	**Mean Square**	**F-ratio**	**Prob**
Const	1	296382	296382	416.41	≤ 0.0001
Usr	3	17571.7	5857.25	8.2292	≤ 0.0001
Error	152	108188	711.761		
Total	155	125759			

Results for factor Usr

Coefficients

Coefficients of: Time on Usr

Level of Usr	**Coefficient**	**std. err.**	**t Ratio**	**prob**
"C"	18.06	3.67	4.922	≤ 0.0001
"G"	-6.099	3.67	-1.662	0.0986
"K"	-7.732	3.67	-2.107	0.0368
"S"	-4.232	3.802	-1.113	0.2675

Expected Cell Means

Expected Cell Means of: Time on Usr

Level of Usr	**Expected Cell Mean**	**Cell Count**
"C"	61.54	40
"G"	37.38	40
"K"	35.75	40
"S"	39.25	36

Scheffe Post Hoc Tests

	Difference	**std. err.**	**Prob**
"G" - "C"	-24.1612	5.966	0.00135159
"K" - "C"	-25.794	5.966	509.765e-6
"K" - "G"	-1.63275	5.966	0.994644
"S" - "C"	-22.294	6.129	0.00526383
"S" - "G"	1.86725	6.129	0.992653
"S" - "K"	3.5	6.129	0.954908

Figure A.11: ANOVA for Chapter 5

Dependent variables

Factors

Name	Code	Nested in	F/R	Kind
Time	Tim	‹›	Fix	Disc
Technology	Tcy	‹›	Fix	Disc

Partial (Type 3) Sums of Squares   Design Help

Interactions up to 2 - way

Modifications

cases selected according to TechFilter

No variance variable

RESULTS

General Results

84 total cases of which 42 are missing

ANOVA

Analysis of Variance For   #
cases selected according to   **TechFilter**
84 total cases of which 42 are missing

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	1417.52	1417.52	120.03	≤ 0.0001
Tim	2	201.762	100.881	8.5423	0.0009
Tcy	1	656.095	656.095	55.556	≤ 0.0001
Tim*Tcy	2	67.4762	33.7381	2.8569	0.0705
Error	36	425.143	11.8095		
Total	41	1350.48			

Results for factor Tim*Tcy

Coefficients

Expected Cell Means

Expected Cell Means of: # on Tim*Tcy

Level of Tim*Tcy	Expected Cell Mean	Cell Count
SAM,t1	3.429	7
SAM,t2	0.8571	7
SAM,t3	1.286	7
TM,t1	14.14	7
TM,t2	9.286	7
TM,t3	5.857	7

Scheffe Post Hoc Tests

	Difference	std. err.	Prob
SAM,t2 - SAM,t1	-2.57143	1.837	0.385159
SAM,t3 - SAM,t1	-2.14286	1.837	0.512786
SAM,t3 - SAM,t2	0.428571	1.837	0.973169
TM,t1 - SAM,t1	10.7143	1.837	6.29802e-6
TM,t1 - SAM,t2	13.2857	1.837	96.6153e-9
TM,t1 - SAM,t3	12.8571	1.837	192.595e-9
TM,t2 - SAM,t1	5.85714	1.837	0.011361
TM,t2 - SAM,t2	8.42857	1.837	251.325e-6
TM,t2 - SAM,t3	8	1.837	491.477e-6
TM,t2 - TM,t1	-4.85714	1.837	0.0409712
TM,t3 - SAM,t1	2.42857	1.837	0.425949
TM,t3 - SAM,t2	5	1.837	0.0344318
TM,t3 - SAM,t3	4.57143	1.837	0.0574138
TM,t3 - TM,t1	-8.28571	1.837	314.631e-6
TM,t3 - TM,t2	-3.42857	1.837	0.189623

Dependent variables

Factors

Name	Code	Nested in	F/R	Kind
Time	Tim	‹›	Fix	Disc
Technology	Tcy	‹›	Fix	Disc

Partial (Type 3) Sums of Squares   Design Help

Interactions up to 2 - way

Modifications

cases selected according to IdFilter

No variance variable

RESULTS

General Results

84 total cases of which 42 are missing

ANOVA

Analysis of Variance For   #
cases selected according to   **IdFilter**
84 total cases of which 42 are missing

Source	df	Sums of Squares	Mean Square	F-ratio	Prob
Const	1	293.357	293.357	69.873	≤ 0.0001
Tim	2	77.2857	38.6429	9.2042	0.0006
Tcy	1	282.881	282.881	67.378	≤ 0.0001
Tim*Tcy	2	70.3333	35.1667	8.3762	0.0010
Error	36	151.143	4.19841		
Total	41	581.643			

Results for factor Tim*Tcy

Coefficients

Expected Cell Means

Expected Cell Means of: # on Tim*Tcy

Level of Tim*Tcy	Expected Cell Mean	Cell Count
SAM,t1	0.1429	7
SAM,t2	55.51e-18	7
SAM,t3	0	7
TM,t1	8.714	7
TM,t2	4.714	7
TM,t3	2.286	7

Scheffe Post Hoc Tests

	Difference	std. err.	Prob
SAM,t2 - SAM,t1	-0.142857	1.095	0.991531
SAM,t3 - SAM,t1	-0.142857	1.095	0.991531
SAM,t3 - SAM,t2	-222.045e-18	1.095	1
TM,t1 - SAM,t1	8.57143	1.095	17.046e-9
TM,t1 - SAM,t2	8.71429	1.095	11.6914e-9
TM,t1 - SAM,t3	8.71429	1.095	11.6914e-9
TM,t2 - SAM,t1	4.57143	1.095	821.388e-6
TM,t2 - SAM,t2	4.71429	1.095	568.038e-6
TM,t2 - SAM,t3	4.71429	1.095	568.038e-6
TM,t2 - TM,t1	-4	1.095	0.00343651
TM,t3 - SAM,t1	2.14286	1.095	0.162202
TM,t3 - SAM,t2	2.28571	1.095	0.128002
TM,t3 - SAM,t3	2.28571	1.095	0.128002
TM,t3 - TM,t1	-6.42857	1.095	5.64141e-6
TM,t3 - TM,t2	-2.42857	1.095	0.0998184

Figure A.12: ANOVA for Chapter 7

A.3 Tutorial Files

This section holds the four tutorial text (Chapter 7). Italics is used to describe the participants' actions.

A.3.1 Tutorial Session 1: Introduction

SAM represents the world simply: things on a map are either curves or points of interest, POIs for short. A curve can be used to represent a street, a sidewalk, a corridor or a route. POIs are locations.

Keyboard

There is very little keyboard interactions with SAM. The keyboard is only used to find features on a map: every POI or route is associated with a keyboard key. Upon a key stroke SAM speaks the name of the street associated with that key. Note that there is no alphabetic connections between the key and the name of a POI or route.

Routes and POIs are grouped separately. SAM splits the keyboard into two regions, left and right. Some maps have more routes than POIs and the opposite can be true for other maps, so the split point varies from map to map. The keyboard has only 26 keys (the alphabetic keys), the other keys are removed to reduce confusion during the study.

Participants touch the keyboard.

Routes are on the left of the keyboard and POIs on the right. If a map contains less than 26 features then the keys at the centre remain unassigned. Typing an unassigned key has SAM speak 'nothing's assigned'. Typing a key associated with a route speaks the name of the route. Typing a key associated with a POI speaks the word 'Location', which informs this feature is a POI, and then SAM speaks the name of that POI.

Participants strike keys and listen.

Locatones

Locatones are sounds SAM uses to help a user reach different locations on a map. The best way to explain them is by example, but before, I explain how space is represented with sounds.

Because maps have two dimensions, horizontal and vertical, SAM needs to provide the user with a way to navigate in both dimensions. SAM represents horizontal directions with stereo balance. For example, a sound in the left ear points in the left direction on the map. Similarly, a sound in the right ear points right. Accordingly, a sound heard at the centre can be interpreted as centred.

Pitch, also known as sound frequencies, represents the vertical directions. The concept is similar to the horizontal directions. For example a high pitch sound points north on the map. A low pitch sound points south. Accordingly, a middle pitch sound can be interpreted as being centred.

Locatones are spatial sounds made of two parts. The first part represents your current position and the second part indicates the direction to go.

Locatones can point in eight different directions: north, south, east, west, north-west, north-east, south-west and south-east. Lets listen to them individually and try to guess where they are pointing.

Participants listen to sound examples.

Wizard

The wizard tool uses locatones to walk the user to destination. There is a side button along the stylus. The button has two positions: near the sharp tip and near the eraser tip (the rounded end). Pressing the button near the sharp tip calls the wizard. Pressing the button near the eraser end activates another tool we will learn later.

Task: you are attending a party and are standing in the middle of a room filled with famous persons who have/had visual impairment, past and present (e.g., Louis Braille, Stevie Wonder, Claude Monet and so on). The surface of the tablet is a scaled down representation of the room. Your task is to meet everyone in the room.

The steps to follow are described below. You will see there is many little things to do. Do not worry, you do not have to remember the order of the steps because I will help you. Work at a pace you are comfortable with.

Task:

Step 1: lift the stylus above or move it on the side of the tablet to avoid any communication between the stylus and the tablet.

Step 2: hit a POI key on the keyboard. SAM will speak the name of the selected POI. In this context, a POI is a famous person in the room.

Step 3: press and hold the stylus button to activate the wizard tool.

Step 4: while holding the stylus button, put the stylus tip in contact with the surface of the tablet. Drop it near the starting position.

Step 5: start moving the stylus in the direction pointed out by the locatones.

Step 6: once you found the person, punch a hole with the stylus on the carton board using the tip of the stylus. Doing this will mark the persons position.

Step 7: can you tell me the direction and distance you travelled to reach the person?

Note that I am not recording your answers. A last thing before you begin, you will notice that being at closer proximity of a target destination increases the sound intensity of the locatones and the speed at which they are repeated. These changes indicate that you are approaching the target. When you get closer, try reducing your speed.

Participants perform the task.

A.3.2 Tutorial 2: Curve Tracing and Notifications

A person tracing a street with their fingers on a tactile map notice when they are loosing contact with it. Without even thinking about it, the person's nervous system redirects the hand to correct its trajectory, which re-centres the fingers on the street. With SAM, tracing a street on a digital tablet also involves moving the stylus along the street, but the user cannot feel the street because the surface of the digital tablet is flat. SAM substitutes touch with sound: instead of perceiving the street by touch, the user perceives it with sound.

This is something you never did before and practice is needed. Fortunately, the concepts you learned with Locatones, namely the concepts of using pitch and stereo information, is reused.

Vertical Curves

Tracing vertical curves: having the stylus sitting perfectly on a curve plays sounds centred in stereo, plus a soothing 'Ah' sound. Gradually leaving the curve from the right moves the sound cues in the left ear, telling the user to direct his movement towards the left. The sound intensity also drops when moving away from the curve. Similarly, leaving the curve from the left shifts the sound cues in the right hear. Moving too far from the curve mutes the tracing sound and plays a different sound indicating contact with the curve is lost. This sound indicates that nothing is underneath the stylus. Let's listen to these sounds before practicing.

Participants listen to sound examples.

Task:

Step 1: use the keyboard to find the curve 'Empire State Building'.

Step 2: use the wizard to reach the curve. SAM brings you on top of the building. Punch a hole, release the side button and then start tracing the curve down. Once you reached the bottom, punch another hole and move back to the top. Note that a new sound is played when you reach the extremity of the curve, this is the curve extremity notification sound.

Step 3: follow the same procedure for the curve 'CN Tower' and the 'Burj Tower' in Dubai. Note that lifting the stylus over a curve, without motion, speaks the name of the curve. Lifting the stylus over a curve extremity notification speaks the name of the street, plus a locatone pointing the direction to return on the curve. Heres how the extremity notification sound sounds like. Which tower is the tallest?

Participants have their hand controlled by the researcher and then are asked to try by themselves.

Horizontal Curves

The same concept is reused to trace horizontal curves, but with pitch: high pitch and low pitch feedback tell the user to redirect their movement up and down. Like with vertical curves, intensity decreases when going away and a soothing 'ah' sound is played when the user is right on the curve. SAM uses a musical scale of fifteen notes to guide the user. These notes correspond to the white keys on the piano. The notes goes like this: 'do', 're', 'mi', 'fa', 'sol', 'la', 'ti', 'do', 're', 'mi', 'fa', 'sol', 'la', 'ti', 'do'.

Did you notice that the musical note 'do' is repeated three times? 'Do' is the lowest, the central, and the highest note in the scale. The central 'do' tells the user that the stylus is centred on a curve. The other notes indicate gradual progression towards or away the centre. You do not need to memorize any of these notes. Let me show you what I mean.

Participants listen to sound examples.

Task:

Step 1: use the keyboard to find the curve 'Vancouver Toronto'.

Step 2: use the wizard to reach the curve. SAM brings you on the left hand side of the curve. Punch a hole, release the side button and then start moving towards the right. Once you reached the end, punch another hole and move back to the left.

Step 3: Follow the same procedure for 'Paris Toronto' and 'Sidney Toronto'. When done, compare the distances.

Participants have their hand controlled by the researcher and then are asked to try by themselves.

Diagonal Curves

When a curve is neither vertical nor horizontal, SAM provides several pitch and stereo cues. SAM finds the direction that best characterizes the local direction of the curve, and provides the feedback in that direction. SAM updates this information constantly many times per second. As a result, tracing a diagonal curve that gradually goes downward towards the right may require a user to listen to pitch cues telling to go down, then stereo cues telling to go right, then pitch cues telling to go down again, and so on.

Task:

Step 1: use the keyboard to find the curve 'Two O'clock'.

Step 2: use the wizard to reach the curve. Punch a hole, release the side button and then start moving in diagonal. Once you reached the end, punch another hole and move back to the beginning.

Step 3: follow the same procedure for the 'Four O'clock' curve. Use the collision detector to find curves you lost contact with: lift the stylus over the tablet and move it: SAM plays a short percussive sound when the stylus passes over a curve. Stop moving the stylus when aligned with a curve and SAM speaks the name of that curve.

Participants have their hand controlled by the researcher and then are asked to try by themselves.

Non-Linear Curves

Lets put in practice everything you learned. Before you do so, I would like to tell you a little bit more about the curve notification sound and tell you how to use the wizard tool when tracing a curve.

Like the extremity notification sound, the curve notification sound does not require you to do anything special. Its job is to notify you that a steep curve is coming. If you want, you can hear the direction of the curve by lifting the stylus. Locatones point the direction.

Try using the wizard tool to regain contact with a curve. Pressing the side button on the stylus activates the wizard, which walks you back to your last contact point.

Task: there are four curve starting on the lefthand side of the tablet. The dots represent their starting position. Try tracing them. Can you tell me how many of these curves end at the same location? Can you describe the directions you took? Which one the four curves is the longest? Punch a hole at the end of every curve to ease the task.

Participants are asked to try by themselves.

A.3.3 Tutorial 3: Map Reading

You know how to use the keyboard with the wizard tool to find things on a map. You also know how to use the wizard tool to regain contact with a curve, but we will revisit this functionality. This will give you a little bit more practice with the tool. Two new notification sounds and two tools are explained.

Wizard Revisited

Press and hold the side button on the stylus when you loose contact with a curve: the wizard tool walks you back to your last contact-point. A clearer set of locatones point the local directions of a curve when destination is reached: on a curve you hear two directions and at curve extremities you hear one. Begin tracing in these directions.

Similarly, you can press the side button when you are already on a curve. SAM looks around and use locatones to tell you the local directions of the curve.

Participants are asked to try by themselves.

POI and Intersection Notification Sounds

Like the curve extremity and curve notification sounds, the POI notification sound does not require you to do anything special; it simply notifies you that a POI is near. Lifting the stylus speaks the name of the POI. The word 'Location' is always spoken first.

Similarly, an intersection notification is automatically heard when you reach an area where two or more curves intersect. Lifting the stylus speaks the name of the curves and plays locatones showing the directions of the intersecting curves. Here's how POI and intersection notifications sound like.

Participants listen to sound examples.

Toggle

It is frequent to see streets changing name. For example, King street South changing to King street North. SAM notifies the user by alternating the timbre of the sound tracing tool. Lifting the stylus tells the user the new name of the street. Let's listen to the two tracing sounds.

Participants listen to sound examples.

Task: find King Street West using the keyboard and the wizard. Once you reached King Street West, listen to its name and direction and then begin tracing. You will encounter POIs and intersections along your way. Explore every routes, intersections and POIs.

Can you tell me where King Street West changes to King Street East? Can you tell me the directions of the intersecting streets? Describe the things you discover along the way.

Participants are asked to try by themselves.

Task: find the hotel entrance using the keyboard, walk there using the wizard and start exploring. Pay attention to the things you discover along the way because I will question you about them. For example I may ask you: 'You are standing at the lobby, can you point the direction to the washrooms?' I want to see how hard it is for a user to read maps with SAM. I am not recording your answers and do not worry if you do not know the answers. Dont forget that you can count on me to assist you.

Participants are asked to try by themselves.

Alignment

The map of the hotel is reused to learn the alignment tool. Turn the stylus upside down, with the eraser end in contact with the tablet. There are two possible stylus angles: inclined or straight up.

If you incline the stylus towards you and slowly slide it in your direction (from north to south), you hear popping noises that personify curves and POIs aligned on the left and right side of the stylus, a little bit like if you were cutting the map in two. Stop moving and SAM speaks the POIs and corridor names. Similarly, you can hear the features above and below the stylus by inclining the stylus on the left or right side: pitch is used instead of stereo balance. Things that are closer are louder and spoken first.

Leave the stylus straight up to hear POIs only. The direction of the stylus determines the alignment direction. Moving up/down displays the POIs on each side, moving left/right displays the POIs above and below.

Participants listen to sound examples, have their hand movement controlled by the researcher and then are asked to try by themselves.

A.3.4 Tutorial 4: Recapitulation

Today we recap the material you learned and also learn another tool. There is a new map in front of you, you may start exploring it using any tool you like. Begin by finding the entrance. In need, do not hesitate to request my assistance.

Participants are asked to try by themselves.

Grid Tool

The grid tool allows you to easily locate intersections and POIs with the stylus.

The tool divides the map into a 3x3 grid arranged like a telephones numeric keypad. Place the stylus inside a cell region to examine the features inside. The features outside that regions are not exposed.

Hold the stylus straight in contact with the tablet over a cell, then press and hold the upper side button (near the eraser end) on the stylus. A click sound, which sounds like driving over a train track, is heard when a new region is entered; the region ID is spoken; and then the map information is displayed.

Regions that contain POIs and/or intersections: SAM speaks identification numbers that personify them. These numbers are arbitrarily assigned.

Regions that do not contain POIs and/or intersections: SAM plays the empty sound.

You can hear the names of the features inside a region by typing numb keys on the keyboard. Once you found the feature that interests you, use the wizard tool to reach it. SAM plays a notification sound upon destination. Lift the stylus to hear its name. Stereo panning, pitch and intensity cues are used again.

Participants listen to sound examples, have their hand movement controlled by the researcher and then are asked to try by themselves.

Column and Row Tools

Tilt the stylus left or right to explore rows of cells or tilt the stylus up or down to explore columns of cells. The column and row tools break the map in three columns or rows. This action concatenates cells that are aligned together. In either mode, number are replaced by percussive sounds. SAM speaks the name of the activated tool (grid, column or row) before displaying map information.

To activate either tool the user needs to press and hold the upper side button near the eraser end on the stylus. The stylus must remain in contact with the tablet.

Participants listen to sound examples, have their hand movement controlled by the researcher and then are asked to try by themselves.

A.4 OMS Instructions

Maps of two different malls were reconstructed from memory by a group of blind participants. Your task consists in grading the correctness of these reconstructions.

A.4.1 Measures

The value of every reconstruction is determined by four measures: the spatial relationships between landmarks, the layout of corridors, the topology, and the overall functionality of a map. You will be asked to grade every map four times, once per measure. Details about the measures are provided below:

Grading Landmark-to-Landmark locations: you are asked to assess the spatial relationships between landmarks (e.g., stores, elevators, ramp, escalators and washrooms) and ignore the corridors, that is, only verify that the spatial relationship between landmarks is preserved under continuous deformations (reasonable stretching or shrinking is expected and fine).

Grading Corridor Layout: you are asked to assess the bearings, connectivity, curvatures, and beginning & ending positions of corridors, but ignore the spatial relationships between landmarks. Again, continuous deformation is ok.

Grading the Topology: you are asked to verify the sequential arrangement of landmarks along corridors and their reachability (e.g., are the landmarks aligned with the corridors and are the landmarks reachable by travelling along corridors). Again, only the spatial relationships need to be preserved under continuous deformations.

Grading the Overall Functionality: you are asked to verify that the reconstructed maps are accurate in a pragmatic sense, allowing someone to operate within the environment to the extent desired. The maps, whether exact or not, must be good enough to get a traveller around the entire mall.

A.4.2 Solution Maps

Because two different environments were used during the study, two solution maps are provided for grading the reconstructed maps. The two maps are identified by the letters A and B. Seven reconstructed maps rebuild the environment shown in map A and seven in map B.

A.4.3 Grading Procedure

Start by saving this file on your computer disk (if this is not already done). This is necessary because you will enter your grades on it and will send it back to me when done.

I would like you to follow these six steps when grading the maps:

1. Select one of the four measures.
2. Take the time to examine the fourteen reconstructed maps and develop your own marking scheme for this measure. Your grading scheme should be consistent with the explanations I provided above for the measure. This examination step is important because no reconstruction is perfect and the variance is high among the responses.
3. Find the best map among the fourteen maps and give it a perfect score of 100% for this measure.
4. Note that a perfect score may be attributed to a reconstruction of type Map A or type Map B, but does not need to be attributed to maps of both types; only one map out of the fourteen maps should get a perfect score.
5. Find the worst reconstruction and give it the lowest score.
6. Then grade the twelve remaining maps accordingly, by taking in consideration the best and worst maps, and by applying your marking scheme consistently. Once done grading the maps for this measure, move on to the next measure and repeat these steps.

You should produce a total of 56 grades (14 maps x 4 measures). I am expecting to find four 100% scores in total, one per measure. Please enter the grades on this file. You will find a reserved area for this purpose underneath every map. Finally, tell me using single words or short sentences the leading factors that influenced your grading for a measure. Space is available at the end of this file. This information will help me understand your grading strategy.

When finished, please fill up the invoice below and resend me this file, which should contain your grades. You will receive a check by mail (if you prefer I can send you the money by email - please let me know). I am the one paying you, but I will be refunded by my advisor, Dr. Bill Cowan, professor at the University of Waterloo.

If you have any question, please contact me by e-mail: talbotm@acm.org

A.5 OMS Comments

I copied Marker A's comments for every reconstruction, which can be found in Section A.6. Marker B did no provide this level of details in her assessments.

TM Reconstructions Comments

MAP M1

+ Overall good layout, shows entrance and elevator not near the main store areas.

- Got toy store misplaced, clothing store location not aligned with other major landmarks (bookstore, bathroom).

- Only has one elevator, no clear path to auditorium.

- Bathrooms transposed (an important error, as both rooms contain the same kinds of equipment, and very difficult to determine without being informed).

MAP K2

+ All landmarks in proper order with only minor distance issues.

+ Wall between toy store and office B very good, shows relative location of Office B hidden behind toy store.

- Doesn't indicate clear routes, only store locations relative to each other and walls.

- Missing second elevator at upper right.

MAP W1

+ Shows walls and store locations, so implies where corridors would be located.

+ Tried to indicate large area for entrance.

- Got offices mixed up, no clear route to get to Offices.

MAP B1

+ great work indicating locations of stores and landmarks.

+ shows area and wall of bookstore, as well as ramp from entrance.

- lacks walls (and implied corridors) for left and upper side of mall.

- no clear method of reaching offices or bathroom.

MAP K3

+ Good sense of layout of stores and area.

+ Shows entrance has a large area, not just one doorway.

+ Shows clear ramp goes along side of mall, starting at entrance area straight to left side of mall.

- Office A displaced, no clear path to bathrooms.

MAP K1

+ Has stores and landmarks in correct order, but proportion to other side is noticeably off.

+ Relative locations of Offices to dentist/toy store is good, lacks routes to get there.

- Lacking corridors or walls to show where person can go.

MAP D1

+ Shows overall mall layout, how to travel within the mall.

- Auditorium, bookstore, electronic store, clothing store and Office A misplaced (bit problem for functionality).

- Will get lost easily as the auditorium and bookstore are situation on the wrong wall.

SAM Reconstruction Comments

MAP M1

+ Good overall store layout.

+ Good route layout.

- Lacks left/right distinction in corridors (theatre looks like it would be on the same side as the hardware store).

MAP W1

+ Good overall store layout, although that extra store will be confusing.

+ Good route layout, shows basic shape of mall.

- Lacks left/right distinction in corridors (theatre looks like it would be on the same side as the hardware store).

MAP K2

+ Good route layout , although slightly distorted on the right side.

- Bathrooms reversed.

- Store order correct, but placement on right side distorted compared to left. Spacing between offices and optometrist very far. Also missing music store.

MAP B1

+ Left side well done, only right side has issues.

- Optometrist, music store, daycare and hardware store is out of order.

- Route defined, but no clear turns, only vague sense of mall shape.

- Distance between offices distorted.

MAP K1

+ Good overall store layout.

+ Good route layout (wall consistently inside, stores outside and go around).

- Missing entrance, offices and optometrist mixed up.

- No angles on corridor at upper right, strange corridor. coming off at theatre.

- Food court and theatre seem much closer than reality

MAP D1

- Lacks clear routes and has numerous mistakes in landmark placement.

- Because it doesn't show many useable routes or proper store order, I have to rate it very low.

MAP K3

+ Good overall store layout, with proper order.

+ Good route layout, shows basic shape of mall with clearly defined turns.

- Lacks left/right distinction in corridors (theatre looks like it would be on the same side as the hardware store).

A.6 Participants' Reconstructions

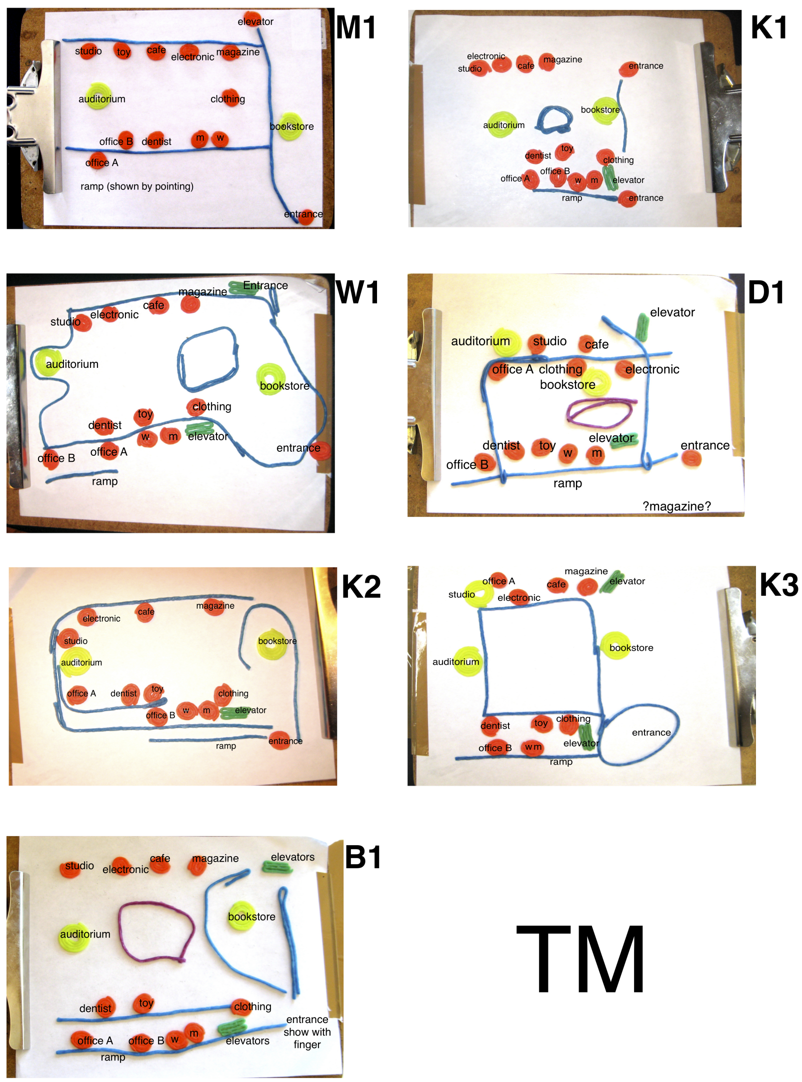

Figure A.13: The participants' responses for the TM (Chapter 7)

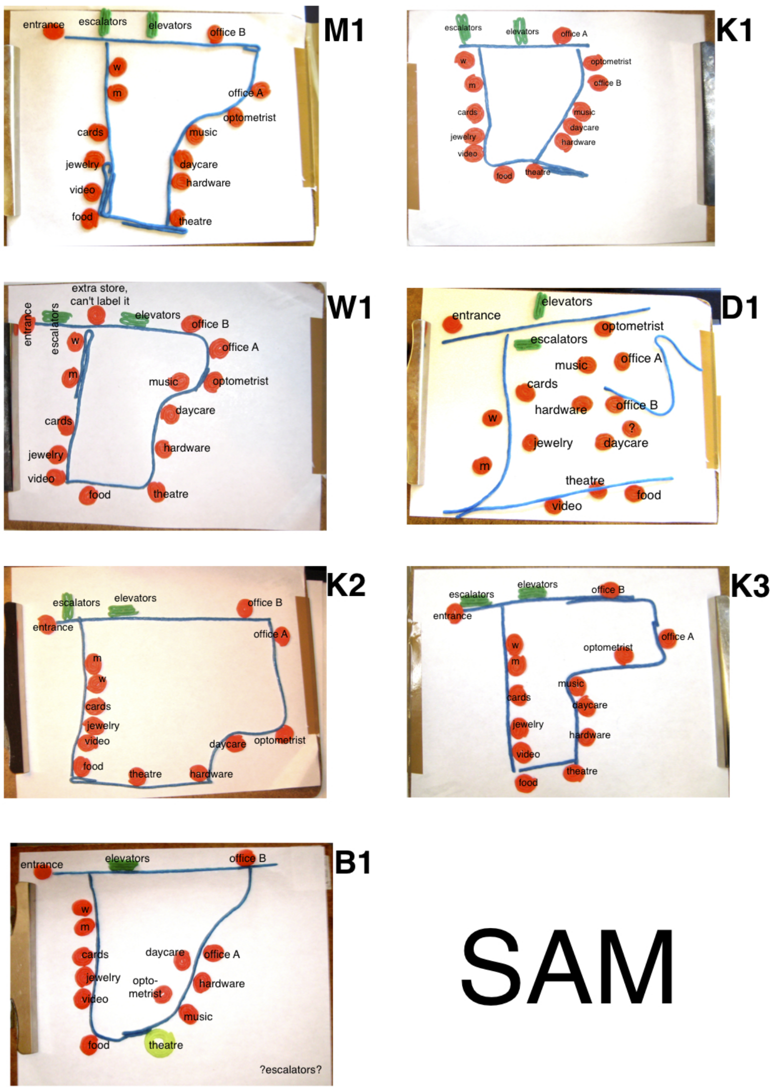

Figure A.14: The participants' responses for the spatial auditory map (Chapter 7)

References

[1] Emiliani, P.L. Access to pictorial information issues in telecommunication and disability. In S. Von Tetzchner, editor, *Issues in telecommunication and disability*, pages 396–416. Commission of the European Communities, 1991.

[2] Kamel, H.M., Landay, J.A. A study of blind drawing practice: creating graphical information without the visual channel. In *Assets '00: Proceedings of the fourth international ACM conference on Assistive technologies*, pages 34–41, New York, NY, USA, 2000. ACM.

[3] Ungar, S., Blades, M., Spencer, C. *The construction of cognitive maps by children with visual impairments*, pages 247–273. Kluwer Academic Publishers, The Netherlands, 1996.

[4] Espinosa, M.A., Ochaita,E. Using tactile maps to improve the practical spatial knowledge of adults who are blind. *Journal of Visual Impairment and Blindness*, 92:338–345, 1998.

[5] James G.A. *Mobility Maps*. Cambridge University Press, USA, 1982. 334-363.

[6] Erikson, C.W., St. James, J.D. Visual attention within and around the field of focal attention: a zoom lens model. *Perception and Pyschophysics*, 40:225–240, 1986.

[7] Edman, P. *Tactile Graphics*. American Foundation for the Blind, NY, USA, 1992.

[8] Amick, N., Corcoran, J. Guidelines for the design of tactile graphics. *American Printing House for the Blind*, 1997.

[9] Bentzen, B.L. Orientation Aids. In R.L. Welsh, B.B. Blasch, editor, *Foundations of Orientation and Mobility*. American Foundation for the Blind, New York, NY, 1997.

[10] Eriksson, Y., Jansson, G., Strucel, M. Tactile maps: guidelines for the production of maps for the visually impaired. *The Swedish Braille Authority*, 2003.

[11] Gardiner, A., Perkins, C. Best practice guidelines for the design, production and presentation of vacuum formed tactile maps, 2003. www.tactilebooks.org/tactileguidelines/page1.htm.

[12] Nolan, C.Y., Morris, J.E. Improvement of tactual symbols for blind children: final report. *American Printing House for the Blind*, 1971.

[13] Berla, E.P., Murr, J.M. The effects of noise on the location of point symbols and tracking a line on a tactile pseudomap. *The Journal of Special Education*, 9(2):183–190, 1975.

[14] Bentzen, B.L., Peck, A.F. Factors affecting traceability of lines for tactile graphics. *Journal of Visual Impairment and Blindness*, 73:264–269, 1979.

[15] Vasconcellos, R. Representing the geographical space for visually handicapped students : a case study on map use. *International Cartographic Association*, 1993.

[16] Rowell, J., Ungar, S. The world of touch: an international survey of tactile maps. Part 2: design. *British Journal of Visual Impairment*, 21(3):105–110, 2006.

[17] Tatham, A.F. The design of tactile maps: Theoretical and practical considerations. *Proceedings of international cartographic association: Mapping the nations*, pages 157–166, 1991.

[18] Royal National Institute of Blind People. RNIB National Centre for Tactile Diagrams, 2009. http://www.nctd.org.uk/MakingTG/index.asp.

[19] Rowell, J., Ungar, S. The world of touch: an international survey of tactile maps. Part 1: production. *British Journal of Visual Impairment*, 21(3):98–104, 2003.

[20] National Federation for the Blind. General Braille Information Center, 2007. http://www.nfb.org/nfb/Braille_general.asp.

[21] Hamzat, T.K., Seyi-Adeyemo, O.R. Health status, self-esteem and participation: a comparison between young adults with mobility disability and those without disability in a Nigerian community. *Scandinavian Journal of Disability Research*, 10(3).

[22] Jacobson, D.R. Navigating maps with little or no sight: a novel audio-tactile approach. In *Proceedings of Content Visualization and Intermedia Representations*, 1998.

[23] Vincent, T. Computing and the blind. *(Tech. Rep. No. 52). Milton Keynes, UK: Open University, Computer-Assisted Leaning Research Group.*, 1986.

[24] Edwards, A.D.N. Soundtrack: an auditory interface for blind users. *Hum.-Comput. Interact.*, 4(1):45–66, 1989.

[25] KNFB Reading Technology Inc. Kurzweil-National Federation of the Blind: Readers, 2011. http://www.knfbreader.com/.

[26] Jacko, J.A., Scott, I.U., Sainfort, F., Barnard, L., Edwards, P.J., Emery, V.K., Kongnakorn, T., Moloney, K.P., Zorich, B.S. Older adults and visual impairment: what do exposure times and accuracy tell us about performance gains associated with multimodal feedback? *Proc. CHI 2003, ACM Press*, pages 33–40, 2003.

[27] Freedom Scientific Inc. JAWS for Windows Screen Reading Software, 2011. http://www.freedomscientific.com/products/fs/jaws-product-page.asp.

[28] Kokjer, K.J. The information capacity of the human figertips. *IEEE Transactions on Systems, Man and Cybernetics*, 1:100–102, 1987.

[29] National Research Council. Visual Impairments: Determining Eligibility for Social Security Benefits. Committee on Disability Determination for Individuals with Visual Impairments. In P. Lennie, S.B. Van Hemel, editor, *Division of Behavioral and Social Sciences and Education*. National Academy Press, 2002.

[30] World Health Organization. Up to 45 million blind people globally - and growing, 2003. http://www.who.int/mediacentre/news/releases/2003/pr73/en/.

[31] Begg, I.S., *et al.* In H.C. Gerstein, R.B. Haynes, editor, *Evidence-Based Diabetes Care*. Lawrence Erlbaum Associates, 2001.

[32] Medical News Today. USA - diabetes/obesity rises by 27% in 5 years., 2011. http://www.medicalnewstoday.com/articles/4428.php.

[33] The National Coalition for Vision Health. Vision Loss in Canada, 2011. http://eyesite.ca/activities/NCVH/VisionLossinCanada_e.pdf.

[34] Millar S. *Understanding and representing space: theory and evidence from studies with blind an sighted children*. Oxford University Press, NY, USA, 1994.

[35] Sauerburger, D. Safety awareness for crossing streets with no traffic control. *J. Vis. Impair. Blind*, 89:423–431, 1991.

[36] Wiener, W.R., Lawson, G., Maghshineh, K., Brown, J., Bischoff, A., Toth, A. The use of traffic sounds to make street crossings by persons who are visually impaired. *J. Vis. Impair. Blind.*, 91:435–445, 1997.

[37] Crandall, W., Bentzen, B.L., Myers, L. Smith-Kettlewell research on the use of talking signs at light controlled street crossings. *Report to NIDRR, Smith-Kettlewell Rehabilitation Engineering Research Center Report, San Francisco, CA*, 1998.

[38] Zwiers, M.P., Van Opstal, A.J., Cruysberg, J.R. Two-dimensional sound localization behavior of early-blind humans. *Exp. Brain Res.*, 140:206–222, 2001.

[39] LaGrow, S., Weessies, M. *Orientation and mobility: Techniques for independence.* Royal New Zealand Foundation for the Blind, Dunmore Press, New Zealand, 1994.

[40] Welsh, R., Blasch, B. *Foundations of orientation and mobility (2nd ed.).* American Foundation for the Blind, New York, 1997.

[41] Gomulicki, B.R. The development of perception and learning in blind children. Technical report, Cambridge University, 1961.

[42] Golledge, R.G., Klatzky, R.L., Loomis, J.M. Cognitive mapping and wayfinding by adults without vision. In J. Portugali, editor, *The construction of cognitive maps*, volume 60, pages 215–246. Springer, Netherlands, 1996.

[43] Warren, D.H. *Blindness and early childhood development.* American Foundation for the Blind, New York, NY, 2nd edition, 1984.

[44] Cassey, S.M. Cognitive mapping by blind persons. *Journal of Visual Impairment and Blindness*, 72:297–301, 1978.

[45] Byrne, R.W., Salter, E. Distance and directions in cognitive maps of blind persons. *Canadian Journal of Psychology*, 37:293–299, 1983.

[46] Herman J.F., Herman T.G., Chatman, S.P. Constructing cognitive maps from partial information: a demonstrative study with congenitally blind subjects. *Journal of Visual Impairment and Blindness*, 77:195–198, 1983.

[47] Hollyfield, R.L., Foulke, E. The spatial cognition of blind pedestrians. *Journal of Visual Impairment and Blindness*, 77:204–210, 1983.

[48] Millar S. Utilization of shape and movement cues in simple spatial tasks by blind and sighted children. *Perception*, 8:11–20, 1979.

[49] Reiser J., Guth D.A. Sensitivity to perspective structure while walking without vision. *Perception*, 15:137–188, 1986.

[50] Warren, D.H. Early vs. Late vision: the role of early vision in spatial reference systems. *New Outlook for the Blind*, pages 157–162, 1974.

[51] Jones B. Spatial perception in blind person. *British Journal of Psychology*, 66:461–472, 1975.

[52] Welsh, R.L., Blasch B.B. *Foundations of orientation and mobility*. American Foundation for the Blind, New York, NY, 1980.

[53] Cratty, B.J.S. *The body Image of blind children*. The American Foundation for the Blind, New York, NY, 1968.

[54] National Federation for the Blind. Braille Readers are Leaders, 2000. http://www.nfb.org/nfb/Braille_coin.asp.

[55] Biba, P. Technology overtakes braille in Canada, 2010. http://www.teleread.com/2010/04/30/technology-overtakes-braille-in-canada/.

[56] Freedom Scientific. Freedom Scientific Reduces Price of Focus 40 Blue Braille Display by 38% to Support Drive for Braille Literacy, 2010. http://www.freedomscientific.com/news/pressroom/2010/Braille-Prices-Support-Braille-Literacy.asp.

[57] Gibson J.J. *Perception of the Visual world*. Houghton Mifflin, MA,USA, 1966.

[58] Heller M.A., Ballesteros, S. *Touch and blindness: psychology and neuroscience*. Lawrence Erlbaum Associates, New York, NY, 2004.

[59] Sathian, K., Prather, S.C. Cerebral cortical processing of tactile form: evidence from functional neuroimaging. In M.A. Heller, S. Ballesteross, editor, *Touch and blindness: psychology in neuroscience*. Lawrence Erlbaum Associates, NJ, 2006.

[60] James T.W., Harman, J.K., Humphrey, K.G., Goodale, M.A. Do visual and tactile objects representations share the same neural substrate? In S. Ballesteross M.A. Heller, editor, *Touch and blindness: psychology in neuroscience*. Lawrence Erlbaum Associates, NJ, 2006.

[61] Oakey, I., McGee, M.R., Brewster, S., Gray, P. Putting the Feel in 'Look and Feel'. *CHI*, pages 415–422, 2000.

[62] Pascual-Leone, A., Theoret, H., Merabet, L., Kauffmann, T., Schlaug, G. The role of visual cortex in tactile processing: a metamodal brain. In S. Ballesteross M.A. Heller, editor, *Touch and blindness: psychology in neuroscience*. Lawrence Erlbaum Associates, NJ, 2006.

[63] Amedi, A., Stern, W.M., Camprodon, J.A., Bermpoh, F., Marabet, L., Rotman, S., Hemond, C., Meijer, P., Pascual-Leone, A. Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, 10:687–689, 2007.

[64] Tversky, B. Functional Significance of Visuospatial Representations. In P. Shah, A. Miyake, editor, *The Cambridge Handbook of Visuospatial Thinking*, pages 1–34. Cambridge University Press, USA, 2005.

[65] Beyer, H., Holtzblatt, K. *Contextual design: defining customer-centered systems*. Morgan Kaufmann, San Francisco, CA, 1998.

[66] Nees, M.A., Walker, B.N. Auditory Interfaces and Sonification. In C. Stephanidis, editor, *The Universal Access Handbook*, pages 507–521. CRC Press Taylor & Francis, New York, 2008.

[67] Karmer, G. *Audification and Auditory Interfaces*. Addison-Wesley, USA, MA, 1994.

[68] Parkes, D. Nomad, an audio-tactile tool for the acquisition, use and management of spatially distributed information by visually impaired people. *Second International Symposium on Maps and Graphics for Visually Handicapped People*, 1988.

[69] Lötzsch, J. Computer-Aided Access to Tactile Graphics for the Blind. *Proceedings of ICCHP*, pages 575–581, 1994.

[70] Landau, S., Wells, L. Merging tactile sensory input and audio data by means of the talking tactile tablet. *Euro Haptics*, 2003.

[71] ViewPlus Technologies Inc. IVEO Hands-on Learning, 2011. http://www.viewplus.com/products/software/hands-on-learning/.

[72] Touch Graphics Inc. News, 2011. http://www.touchgraphics.com/.

[73] Holmes, E., Jansson, G. A touch tablet enhanced with synthetic speech as a display for visually impaired people's reading for virtual maps. *CSUN*, 1991.

[74] Miele, J.A. *et al.* Tactile Map Automated Production, 2009. http://www.ski.org/Rehab/TMAP/.

[75] Schneider, J., Strothotte, T. Constructive exploration of spatial information by blind users. In *Assets '00: Proceedings of the fourth international ACM conference on Assistive technologies*, pages 188–192, New York, NY, USA, 2000. ACM.

[76] Franklin, K.M., Roberts, J.C. A path based model for sonification. *Proceedings of the Eighth International Conference on Information Visualization (IV 04)*, pages 865–870, 2004.

[77] Rigas, D.I., Alty, J.L. The Use of Music in a Graphical Interface for the Visually Impaired. In *INTERACT '97: Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction*, pages 228–235, London, UK, UK, 1997. Chapman & Hall, Ltd.

[78] Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115–147, 1994.

[79] Cronly-Dillon, J., Persaud, K.C., Blore, R. Blind subjects construct conscious mental images of visual scenes encoded in musical form. *Biological sciences, The Royal Society*, pages 2231–2238, 2000.

[80] Zhao, H., Plaisant, C., Shneiderman, B., Lazar, J. Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data. *ACM Trans. Comput.-Hum. Interact.*, 15(1):1–28, 2008.

[81] Kamel, H.M., Roth, P. Graphics and User's Exploration via Simple Sonics (GUESS): Providing Inter-relational Representation of Objects in a Non-visual Environment. In *In Proceedings of Seventh International Conference on Auditory Display*, pages 261–266, 2001.

[82] Roth, P., Kamel, H., Petrucci, L., Pun, T. A comparison of three nonvisual methods for presenting scientific graphs. *Journal of Visual Impairment and Blindness*, 96(6):(non–paginated), 2002.

[83] Hermann, T., Hunt, A. An Introduction to Interactive Sonification. *Multimedia, IEEE*, 12(2):20– 24, 2005.

[84] Larkin, J.H., Simon, H.A. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1):65–100, 1987.

[85] Zheng, X.S., McConkie, G.W., Schaeffer, B. Navigational control effect on representing virtual environments. *Proceedings of the Human Factors and Ergonomics Society*, pages 2113–2117, 2003.

[86] Lahav, O., Mioduser, D. Multisensory Virtual Environment for Supporting Blind Persons' Acquisition of Spatial Cognitive Mapping - a Case Study. *World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 95–102, 2001.

[87] Patel, K.K., Vij, S.K. Externalizing Virtually Perceived Spatial Cognitive Maps. *Systems Conference, 2008 2nd Annual IEEE*, 2008.

[88] Parente, P. BATS: the blind audio tactile mapping system. In *Proceedings of ACM South Eastern Conference*, 2003.

[89] Noble, N., Martin, B. Shape discovering using tactile guidance. *In Proceeding of the 6th. International Conference EuroHaptics*, 2006.

[90] Pakkanen, T., Raisamo, R. Perception Strategies in Modal-Redistributed Interaction. *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2005.

[91] Ramloll, R., Yu, W., Brewster, S., Riedel, B., Burton, M., Bimigen, G. Constructing sonified haptic line graphs for the blind student: first steps. *ACM ASSETS*, pages 17–25, 2000.

[92] Crossan, A., Brewster, S. Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users. *ACM Trans. Access. Comput.*, 1(2):1–34, 2008.

[93] Plimmer, B., Crossan, A., Brewster, S.A., Blagojevic, R. Multimodal collaborative handwriting training for visually-impaired people. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 393–402, New York, NY, USA, 2008. ACM.

[94] Petrie, H., Schlieder, C., Blenkhorn, P., Evans, D.G., King, A., O'Neill, A.-M., Ioannidis, G.T., Gallagher, B., Crombie, D., Mager, R., Alafaci, M. Tedub: A system for presenting and exploring technical drawings for blind people. In *ICCHP '02: Proceedings of the 8th International Conference on Computers Helping People with Special Needs*, pages 537–539, London, UK, 2002. Springer-Verlag.

[95] Brewster S. Yu, W. Comparing Two Haptic Interfaces for Multimodal Graph Rendering. *Haptic Interfaces for Virtual Environment and Teleoperator Systems, International Symposium on*, page 3, 2002.

[96] Sharmin, S., Evreinov, G., Raisamo, R. Non-visual feedback cues for pen computing. *IEEE Press, First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2006.

[97] Evreinova, T.G., Evreinov, G., Raisamo, R., Vesterinen, L.K. Non-visual interaction with graphs assisted with directional-predictive sounds and vibrations: a comparative study. *Universal Access in the Information Society*, 7(1):93–102, 2007.

[98] Taylor, H.A. Mapping the understanding of understanding maps. In P. Shah, A. Miyake, editor, *The Cambridge Handbook of Visuospatial Thinking*. Cambridge University Press, USA, 2005.

[99] Cohen, R.F., Yu,, R., Meacham, A., Skaff, J. PLUMB: displaying graphs to the blind using an active auditory interface. In *ACM Assets*, pages 182–183, New York, NY, USA, 2005. ACM.

[100] Cohen, R.F., Haven, V., Lanzoni, J.A., Meacham, A., Skaff, J., Wissell, M. Using an audio interface to assist users Who are visually impaired with steering tasks. In *ACM Assets*, pages 119–124. ACM, 2006.

[101] Farcy, R., Bellik, Y. Locomotion Assistance for the Blind. *White Paper*, 2002.

[102] Farcy, R., Ledoux, R., Damaschini, R., Legras, R. Laser Telemetry to improve the mobility of blind people : report of the 6 month training course. *White Paper*, pages 1–7, 2003.

[103] Talbot, M., Cowan, W. On the audio representation of distance for blind users. In *ACM CHI '09*, pages 1839–1848. ACM, 2009.

[104] Farmer L. W. Mobility devices. In R.L. Welsh, B.B. Blasch, editor, *Foundations of Orientation and Mobility*. American Foundation for the Blind, NY, USA, 1980.

[105] Farmer L.W., Smith D.L. Adaptive technology. In R.L. Welsh, B.B. Blasch, editor, *Foundations of Orientation and Mobility*. American Foundation for the Blind, NY, USA, 1997.

[106] Martinez, C. Orientation and Mobility Training: The Way to Go, 1998. http://www.tsbvi.edu/Outreach/seehear/fall98/waytogo.htm.

[107] Damaschini, R., Legras, R., Ledoux, R., Farcy R. Electronic Travel Aid for blind people. *Association for the Advancement of Assistive Technology in Europe*, 2005.

[108] Yen, D.H. Currently Available Electronic Travel Aids for the Blind, 2005. http://www.noogenesis.com/eta/current.html.

[109] Brennan, T. ETA Manuals, 2005. http://titan.sfasu.edu/ g_brennantg/etas.html.

[110] Lutz, R.J. Prototyping and evaluation of landcons: auditory objects that support wayfinding for blind travelers. *Accessibility and Computing*, 86:8–11, 2006.

[111] Walker, B.N., Lindsay, J. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors*, 48(2):286–299, 2006.

[112] Golledge, R.G., Klatzky, R.L., Loomis, Spiegle, J., Tietz, J. A geographical information system for a GPS based personal guidance system. *International Journal of Geographical Information Science*, 12(7):727–749, 1998.

[113] Roentgen U., Gelderblom G.J. State of the art in electronic mobility aids for visually impaired Persons. *International Conference on Low Vision: Research and Rehabilitation Partnerships*, 2008.

[114] Marston, J.R., Loomis, J.M., Klatzky, R.L., Golledge, R.G., Smith, E.L. Evaluation of spatial displays for navigation without sight. *ACM Trans. Appl. Percept.*, 3(2):110–124, 2006.

[115] Android Zoom. WalkyTalky, 2011. http://www.androidzoom.com/android_applications/travel/walkytalky_ktmx.html.

[116] Crandall F.W., Bentzen B.L. Passenger Conveyance Vehicles and systems utilizing the same, Pattent number 754240, 1999. http://www.patentstorm.us/patents/5986786-description.html.

[117] Lui, A.L., Hile, H., Kautz, H., Borriello, G., Brown, P.A., Harniss, M., Johnson, K. Indoor wayfinding: Developing a functional interface for indifidual with cognitive impairments. *Proceeding of the ACM Conference on Assistive Technology*, pages 95–102, 2006.

[118] Pfeifer, T. Redundancy vs. Imperfect Positioning for Context-dependent Services. *Modeling, Reasoning, and Management (UbiComp 2004)*, 2004.

[119] InTouch Graphics. Click and go: wayfinding maps, 2010. http://www.clickandgomaps.com/.

[120] Bennett, J.D. Effects of navigation and position on task when presenting diagrams to blind people using sound. In *The second international conference on Diagrammatic representation and inference - Lecture Notes in Computer Science.*, volume 2317, pages 307–319, 2002.

[121] Metatla, O., Bryan-Kinns, N., Stockman, T. Constructing relational diagrams in audio: the multiple perspective hierarchical approach. In *Assets '08: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 97–104. ACM, 2008.

[122] Metatla, O., Bryan-Kinns, N., Stockman, T. Using hierarchies to support non-visual access to relational diagrams. *BCS HCI'07*, pages 215–225, 2007.

[123] Blenkhorn, P., Evans, D.G. Using speech and touch to enable blind people to access schematic diagrams. *J. Netw. Comput. Appl.*, 21(1):17–29, 1998.

[124] Kennel, A.,. AudioGraf: A diagram-reader for the Blind. *ASSETS, ACM press*, pages 51–56, 1996.

[125] Scaife, M., Rogers, Y. External cognition: how do graphical representations work. *International Journal of Human-Computer Studies*, 45:185–213, 1996.

[126] Zhang, J. The nature of external: Representations in problem solving. *Cognitive science*, 21:179–217, 1997.

[127] Zhang, J. A representational analysis of relational information displays. *International Journal of Human-Computer Studies*, 45:59–74, 1996.

[128] Kamel, H.M., Landay, J.A. IC2D. Sketching images eyes-free: a grid-based dynamic drawing tool for the blind. In *Assets '02: Proceedings of the fifth international ACM conference on Assistive technologies*, pages 33–40, New York, NY, USA, 2002. ACM.

[129] Dassault Systèmes SolidWorks Corp. Why shouldnt the blind drive? *CBS Money Watch.com*, 2009. http://investor.news.com/bnet/?GUID=10787906&Page=MEDIAVIEWER.

[130] Oliver, K. Cognitive map development and driver distraction: the role of vehicle navigation systems. *M.Sc.Thesis, School of Computer Science and Information Technology University of Nottingham*, 2007.

[131] Kamel H., Roth, P., Sinha R. Graphics and user's exploration via simple sonics (GUESS): Providing inter-relational representation of objects in a non-visual environment. *Proceedings of the 9th International Conference on Auditory Display*, pages 261–266, 2001.

[132] Cray C. CCLVI Convention Reports. *Vision Access*, 12:3, 2006.

[133] Shinohara K. Designing assistive technology for blind users. *Proceeding of the ACM Conference on Assistive Technology*, pages 293–294, 2006.

[134] Knightonn R. *Cockeyed: a memoir.* PublicAffairs, USA, 2006.

[135] Shinohara, K. Designing assistive technology for blind user. *Proceeding of the ACM Conference on Assistive Technology*, pages 293–294, 2006.

[136] The Canadian National Institute for the Blind. The Status of Canadian Youth Who are Blind or Visually Impaired: A Study of Lifestyles, Quality of Life, and Employment, 2003. http://www.cnib.ca/en/about/publications/research/youth-study-executive-summary.aspx.

[137] Norman D.A. *The design of everyday things*. Basic Books, New York, NY, 1988.

[138] Ungar S. Cognitive mapping without visual experience. In S. Freundschuh R. M. Kitchin, editor, *Cognitive Mapping: Past, Present and Future*. Routledge, London, 2000.

[139] Bentzen, B.L. Production and testing of an orientation and travel map or visually handicapped persons. *The new Outlook for the Blind*, 66:249–255, 1972.

[140] Guiard, Y. Asymmetric division of labour in human skilled bimanual action: the kinematic chain as a model. *Journal of Motor Behaviour*, pages 486–517, 1987.

[141] Latulipe, C. A Symmetric Interaction Model for Bimanual Input. *PhD Thesis, University of Waterloo*, 2006.

[142] Digby, E., Roy, E.A. *Manual asymmetries in motor performance*. CRC Press, USA, 1996.

[143] Pick H.L.Jr, Klein, R.E., Pick, A.D. Visual and tactual identification of form orientation. *Journal of Experimental Child Psychology*, 4:391–397, 1966.

[144] Zemtsova, M. *Characteristics of perceptual activity in the blind*. Basic Books, New York, 1969.

[145] Lederman, S.J., Campbell, J.I. Tangible line graphs: An evaluation and some systematic strategies for exploration. *Journal of Visual Impairment & Blindness*, 77(3).

[146] Berla, P.E. *Haptic perception of tangible graphic displays*. Cambridge University Press, USA, 1982. 334-363.

[147] Hill E.W., Rieser J.J., Hill M.M., Halpin J. How persons with visual impairments explore novel spaces: strategies of good and poor performers. *Journal of Visual Impairment and Blindness*, 87(8):295–301, 1993.

[148] Gaunet, F., Thinus-Blanc, C. Early-blind subjects' spatial abilities in the locomotor space: exploratory strategies and reaction-to-change performance. *Perception*, 25:967–981, 1997.

[149] Berla, E. Behavioral strategies and problems in scanning and interpreting tactual displays. *The new Outlook for the Blind*, 66:277–286, 1972.

[150] Bertelson, P., Mouosty, P., D'Alimonte, G. A study of braille reading: 2. Patterns of hand activity in one-handed and two-handed reading. *The Quarterly Journal of Experimental Psychology*, 37(2):235–256, 1985.

[151] Preiser, W.F.E. A combined tactile/electronic guidance system for the visually handicapped. In J.W. Wiedel, editor, *Symposium on Maps and Graphics for the Visually handicapped*, pages 86–98. Association of American Geographers, USA, 1983.

[152] Preiser, W.F.E. A combined tactile/electronic guidance system for visually impaired persons in indoor and outdoor spaces. *International Conference on Building Use and Safety Technology*, 1985.

[153] Golledge, R.G. Tactual strip maps as navigational aids. *Journal of Visually Impairment and Blindness*, 85(7).

[154] Millar S. Veering re-visited: noise and posture cues in walking without sight. *Perception*, 28:765–780, 1999.

[155] Montello, D.R. Navigation. In P. Shah, A. Miyake, editor, *The Cambridge Handbook of Visuospatial Thinking*. Cambridge University Press, USA, 2005.

[156] Aretz, A.J. The design of eletronic map displays. *Human Factors*, 3(1):85–101, 1991.

[157] Lynch, K. *The image of the city*. MIT Press, Cambridge, MA, 1960.

[158] Appleyard, D. Styles and methods of structuring a city. *Environment and Behavior*, 2:100–116, 1970.

[159] de Jonge, D. Images of urban areas: their structure and psychological foundations. *Journal of American Institute of Planners*, 28:266–267, 1962.

[160] Shemyakin, F.N. General problems of orientation in space and space representations. In B.G. Ananyev, editor, *Psychological Science in USSR*, volume 1. Office of Technical Reports (NTIS No. TT62-11083), Arlington, VA: US, 1962.

[161] Fletcher, J. Spatial Representation in Blind Children. 1: Development Compared to Sighted Children. *Visual Impairment and Blindness*, 74(10):318–385, 1980.

[162] Thorndyke, P.W., Stasz, C. Performance models for spatial and locational cognition. *Rand*, 1980.

[163] Feldman, A., Acredolo, L.P. The effect of active versus passive exploration on memory for spatial location in children. *Child Development*, 50:698–704, 1979.

[164] Soloway, E. Log on education: Quick, where do the computers go? *Commun. ACM*, 34(2):29–33, 1991.

[165] Shumin Z. Albinsson, P. High precision touch screen interaction. pages 105–112, 2003.

[166] Olwal, A., Feiner, S., Heyman, S. Rubbing and tapping for precise and rapid selection on touch-screen displays. pages 295–304, 2008.

[167] Schneider, J., Strothotte, T. Constructive exploration of spatial information by blind users. In *ACM Assets*, pages 188–192, New York, NY, USA, 2000. ACM.

[168] Association for Healthcare Philanthropy.org. Overlays, 2011. https://shop.aph.org /webapp/wcs/stores/servlet/Product_parts%20communication%20modes %20tactile%20graphics%20family%20raised%20lines%20students%20teachers%20parents%20 adults%20graphics%20mathematics%20science%20handwriting%20games%20tracing %20art%20drawing%20film%20stylus%20art%20artwork%20games_1-08858-00P_100 01_10001.

[169] The National Federation of the Blind. Cane Travel. In B. Cheadle, editor, *Future Reflections*, volume 3. Baltimore, MD, 1984.

[170] Wickens, C.D., Vincow, M., Yeh, M. Design Application of Visual Spatial Thinking. In P. Shah, A. Miyake, editor, *The Cambridge Handbook of Visuospatial Thinking.* Cambridge University Press, USA, 2005.

[171] Hegarty, M., Waller, D.A. Individual Differences in Spatial Abilities. In P. Shah, A. Miyake, editor, *The Cambridge Handbook of Visuospatial Thinking.* Cambridge University Press, USA, 2005.

[172] Winn, W.D. Learning from maps and diagrams. *Educational Psychology*, 3.

[173] Hirtle, S.C., Jonides, J. Evidence of hierarchies in cognitive maps. *Memory and Cognition*, 13.

[174] Ungar, S., Simpson, A. Blades, M. Strategies for organizing information while learning a map by blind and sighted people. *Touch, Blindness and Neuroscience*, 2004.

[175] Millar, S., Al-Attar, Z. Illusions in reading maps by touch: Reducing distance errors. *British Journal of Psychology*, pages 643–657, 2001.

[176] Gaunet, F., Martinez, J.-L., Thinus-Blanc, C. Early-blind subjects' spatial representation of manipulatory space: exploratory strategies and reaction to change. *Perception*, 26:345–366, 1997.

[177] Waller, D., Montello, D.R., Richardson, A.E., Hegarty, M. Orientation specificity and spatial updating of memories for layouts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6):1051–1063, 2002.

[178] Rossano, M.J., Hodgson, S.L. The process of learning from small-scale maps. *Applied Cognitive Psychology*, 8:565–582, 1994.

[179] McNamara, T.P., Hardy, J.K., Hirtle, S.C. Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15:211–227, 1989.

[180] Mou, W., McNamara, T.P. Intrinsic frames of references in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1):262–170, 2002.

[181] Shelton, A.L., McNamara, T.P. Systems of Spatial Reference in Human Memory. *Cognitive Psychology*, 43:274–310, 2001.

[182] Cleveland, W.S. *The elements of graphing data*. Wadsworth, Monterey, CA, 1985.

[183] Bertin J. *Semiology of graphics*. University of Wisconsin Press, Madison, WI, 1983.

[184] Jackendoff, R. On Beyond Zebra: The relation of linguistic and visual information. *Cognition*, 26:89–114, 1987.

[185] Vickers, P. Haptic input + auditory display = musical instrument? *Haptic and Audio Interaction Design (HAID)*, pages 56–67, 2006.

[186] Gigante, M.A. Virtual Reality: Definitions, History and Applications. In *Virtual Reality Systems*.

[187] Thouvenin I. Guén A. Gapenne O. Stewart J. Maillet B. Lenay, C. Designing the ground for pleasurable experience. In *DPPI '07: Proceedings of the 2007 conference on Designing pleasurable products and interfaces*, pages 35–58. ACM, 2007.

[188] Flowers, J.H. Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. pages 406–409, 2005.

[189] Tobias, J. Information technology and universal design: An agenda for accessible technology. *Journal of Visual Impairment and Blindness*, 97:592–601, 2003.

[190] Mankoff, J., Fait, H., Tran, T. Is your web page accessible? A comparative study of methods for assessing web page accessibility for the blind. *In the Proceedings of the ACM CHI Conference on Human Factors Computing Systems (CHI 05)*, pages 41–50, 2005.

[191] Walker, B.N., Lindsay, J. The effect of a speech discrimination task on navigation in a virtual environment. *Proceedings of the Human Factors and Ergonomics Society*, pages 1538–1541, 2006.

[192] Mowbray, G.H. Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *Journal of Experimental Psychology*, 46:365–372, 1953.

[193] Mowbray, G.H., Gebhard, J.W. Mans senses as informational channels. *Human factors in the design and use of control systems*, 46:115–149, 1961.

[194] Talbot M., Cowan B. Trajectory Capture in Frontal Plane Geometry For Visually Impaired. *Proceeding of the International Conference on Auditory Displays*, 2006.

[195] Stokes, A., Wickens, C. D., Kite, K. *Display technology: Human factors concepts.* Society of Automotive Engineers, Warrendale, PA, 1988.

[196] Björk, E.A. Laboratory annoyance of some natural sounds. *Acustica*, 59:282–285, 1986.

[197] Tran, T.V., Letowski, T., Abouchacra, K.S. Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics*, 43(1):807–827, 2000.

[198] Parncutt, R., Levitin, D.J. Absolute pitch. *New Grove Dictionary of Music and Musicians*, 1:37–39, 2001.

[199] Taylor, J.A. Perception of Tonality in Short Melodies. *Journal or Research in Music Education*, 24(4):197–208, 1976.

[200] Gelfand, S.A. *Hearing: An introduction to psychological and physiological accoustics.* Marcel Dekker, New York, 1981.

[201] Buxton, W., Gaver, W., Bly, S. Tutorial number 8: the use of non-speech audio at the interface. *Association for Computing Machinery Computer Human Interaction*, 1991.

[202] Walker, B.N., Kramer, G. Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. G. Neuhoff, editor, *Ecological psychoacoustics*, pages 150–175. Academic Press, New York, 2004.

[203] Brewster, S. Nonspeech auditory output. In A. Sears, J.A. Jacko, editor, *The human-computer interaction handbook : fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum, USA, 2008.

[204] Begault D. R. *3D sound: for virtual reality and multimedia.* Academic Press, Cambridge, MA, 1994.

[205] Buxton, W. Introduction to this special issue on nonspeech audio. *Human-Computer Interaction*, 4:1–9, 1989.

[206] Bregman, A.S. *Auditory scene analysis: the perceptual organization of sound.* MIT Press, Cambridge, MA, USA, 1990.

[207] Spence, C., Driver, J. *Audiovisual links in attention: Implications for interface design.* Ashgate Publishing, Hampshire, U.K., 1997.

[208] Pierpond W.G. The art and skill of radio-telegraphy, 3rd edition, 2002. http://www.qsl.net/n9bor/n0hff.htm.

[209] Terasawa, H., Slaney M., Berger J. Perceptual Distance In Timbre Space. *Proceeding of the International Conference on Auditory Displays*, 2005.

[210] Schouten, J.F. The Perception of Timbre. *In Reports of the 6th International Congress on Acoustics (Paper GP-6)*, 1968.

[211] Blattner, M.M, Sumikawa, D.A., Greenberg, R.M. Earcons and icons: their structure and common design principles. *Human-Computer Interaction*, 4(1):11–44, 1989.

[212] Wenzel E.M., Arruda M., Kistler D.J., Wightman F.L. Localization using nonindividualized head-related transfer functions. *Journal of Acoustical Society of America*, 94:111–123, 1993.

[213] Diracdelta. Duplex Theory of Localization, 2001. http://www.diracdelta.co.uk/science/source/d/u/duplex%20theory%20of %20localization/source.html.

[214] Strybel, T.Z., Manligas, C.L., Perrott, D.R. Minimum audible movement angle as a function of the azimuth and elevation of the source. *Human Factors*, 34:267–275, 1992.

[215] Zotkin, D.N., Duraiswami, R., Davis, L., Mohan, A. Virtual audio system customization using visual matching of ear parameters. *In International Conference on Pattern Recognition*, 2002.

[216] Algazi, V.R., Duda, R.O., Morrison, R.P., Thompson, D.M. Structural composition and decomposition of HRTF's. *In IEEE ASSP Workshop on Applications of Signal Process- ing to Audio and Acoustics*, 2001.

[217] Perrott D.R., Saberi K. Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of Acoustical Society of America*, 87(4):1728–1731, 1990.

[218] Weiner W.R. Audition. In R.L. Welsh, B.B. Blasch, editor, *Foundations of Orientation and Mobility*. American Foundation for the Blind, New York, NY, 1980.

[219] Jack C.E., Thurlow W.R. Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Percept. Mot. Skills*, 37:967–979, 1973.

[220] Pratt, C.C. The spatial character of high and low tones. *Journal of Experimental Psychology*, 24(4):178–185, 1930.

[221] Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C. Butterworth, B. Spatial representation of pitch height: the SMARC effect. *Cognition*, 99(2):113–129, 2006.

[222] Gamble, E.A. Intensity as a criterion in estimating the distance of sounds. *Psychol. Rev.*, 16:416–426, 1909.

[223] Békésy, G.V. Uber die Entstehung der entfernungsempfindingbeim horen. *Akust. Z.*, 3:21–31, 1938. (Available in English in: Experiments in hearing, Wever, E.G., editor, Wiley, NY, USA, 1960).

[224] Brungart, D. Auditory localization of nearby sources III: Stimulus effects. *J. Acoustical Soc. Am.*, 106:3589–3602, 1999.

[225] Ingard, U. A review of the influence of meteorological conditions on sound propagation. *J. Acoust. Soc. Am.*, 25:405–411, 1953.

[226] Nyborg, W., Mintzer, D. Review of sound propagation in the lower atmosphere. *USAF WADC tech. Rep*, 54:602, 1955.

[227] Mershon, D.H., King, E. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Percept. Psychophys.*, 18:409–415, 1975.

[228] Nielsen, S.H. Auditory distance perception in different rooms. *J. Audio Eng. Soc.*, 41:755–770, 1993.

[229] Coleman, P.D. Failure to localize the source distance of an unfamiliar sound. *J. Acoustical Soc. Am.*, 34:345–346, 1962.

[230] Gardner, M.B. Distance estimation of 0 degree of apparent 0 degree-oriented speech signals in anechoic space. *J. Acoust. Soc. Am.*, 45:47–53, 1969.

[231] McGregor, P., Horn, A.G., Todd, M.A. Are familiar sounds ranged more accurately? *Percept. Mot. Skills.*, 61:1082, 1985.

[232] Kitch, D. Echolocation: How humans can "see" without sight, 1982. http://www.martinnaef.ch/index.php?menuid=39&reporeid=66.

[233] Terasawa H., Slaney M., Berger J. Perceptual Distance In Timbre Space. *Proceeding of the International Conference on Auditory Displays*, 2005.

[234] Grey J.M., Moorer J.A. Perceptual evaluation of synthesized musical instrument tones. *Journal of Acoustical Society of America*, 62:454–462, 1977.

[235] Deutsch D. Octave generalization and tune recognition. *Percept. Psychophys.*, 11:411–412, 1972.

[236] Brant H. *Space as an essential aspect of music composition. Contemporary Composers on Contemporary Music.* Holt, Rinehart and Winston, New York, NY, 1967.

[237] Jones, M.R. Time, our lost dimension: toward a new theory of perception, attention and memory. *Psychological Review*, 83:323–355, 1976.

[238] Blauert, J. *Spatial Hearing.* MIT Press, Cambridge, 1983.

[239] Durlach, N.L., Colburn, H.S. Binaural Phenomena. In Carterette, E.C., Friedman, M.P., editor, *Handbook of Perception: Hearing*, volume 4. Academic Press, NY, USA, 1978.

[240] Friberg, A., Sundberg, J. Handbook of Perception: Hearing. In Kungl. Tekniska Hoegskolan, editor, *Speech Transmission Laboratory*, volume 34, pages 49–56. Academic Press, 1993.

[241] Wolfe, J. Uncertainty principle and the musician's uncertainty principle, 2011. http://www.phys.unsw.edu.au/ jw/uncertainty.html.

[242] Charles D., Thomas, A.J. *Computer Music.* Schirmer Books, New York, 1997.

[243] Walker, B.N., Lindsay, J., Godfrey, J. The audio abacus: representing numerical values with nonspeech sound for the visually impaired. *SIGACCESS Access. Comput.*, (77-78):9–15, 2004.

[244] Galeyev B.M., Galeyev I.L. Color Hearing and affect theory (with respect to tonality semantics study), 2009. http://prometheus.kai.ru/affect_e.htm.

[245] Massof, R.W. Auditory assistive devices for the blind. pages 271–275, 2003.

[246] Blattner, M.M., Sumikawa, D.A., Greenberg, R.M. Earcons and Icons: Their Structure and Common Design Principles (Abstract only). *SIGCHI Bull.*, 21(1):123–124, 1989.

[247] Walker, B.N., Nance, A., Lindsay, J. Spearcons: Speech- based earcons improve navigation performance in auditory menus. *Proceedings of the International Conference on Auditory Display (ICAD2006)*, 2006.

[248] Bonebright,T.L., Nees,M.A. Memory for auditory icons and earcons with localization cues. *Proceedings of the International Conference on Auditory Display (ICAD2007)*, 2007.

[249] Keller, P., Stevens, C. Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *Journal of Experimental Psychology: Applied*, 10:3–12, 2004.

[250] Maynatt, E.D. Designing with auditory icons: how well do we identify auditory cues. *Association for COmputer Machinery Computer Human Interaction Conference (CHI 94)*, pages 269–270, 1994.

[251] Edworthy, J., Loxley, S., Dennis, I. Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors*, 33:205–231, 1991.

[252] Gaver W.W. How do we hear in the world: explorations in ecological acoustics. *Ecological Psychology*, 5:285–313, 1993.

[253] Owsinski, B. *The Mixing Engineer's Handbook, Second Edition*. Mix Book, Vallejo California, USA, 1999.

[254] FreeSound Project, 'acclivity'. Cockoo Sound, 2011. http://www.freesound.org/samplesViewSingle.php?id=19561.

[255] Dahlhaus, C., Gjerdingen, R.O. *Studies in the Origin of Harmonic Tonality*. Princeton University Press., USA, 1990.

[256] Aikin, J. How to Make Your Sound Sing with Vocoders, 2011. http://digitalmedia.oreilly.com/2006/03/29/vocoder-tutorial-and-tips.html.

[257] Sun Microsystems, Inc. FreeTTS 1.2 - A speech synthesizer written entirely in the JavaTM programming language, 2011. http://freetts.sourceforge.net/docs/index.php.

[258] Apple. Accessibility, 2011. http://www.apple.com/accessibility/macosx/vision.html.

[259] Van den Doel, K., Smilek, D., Bodnar, A., Chita, C., Corbelt, R., Nekrasouski, D., McGrenere, J. Geometric shape detection with SoundView. *Proceeding of the International Conference on Auditory Displays*, 2004.

[260] Baddeley, A.D., Thomson, N., Buchanan, M. Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior*, 14:575–589, 1975.

[261] Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

[262] Avraamides M.N., Loomis J. M., Klantzky R. L., Golledge R. G. Functional equivalence of spatial representation derived from vision and language: evidence from allocentric judgements. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(4):801–814, 2004.

[263] Waller, D. Egocentric and non-egocentric coding in memory for spatial layout: evidence from scene recognition. *Memory, and Cognition*, 34:491–504, 2006.

[264] Tellevik J. M. Influence of spatial exploration patterns on cognitive mapping by blindfolded sighted persons. *Journal of Visual Impairment and Blindness*, 86:221–224, 1992.

[265] Thinus-Blanc C., Gaunet F. Representation of space in blind persons: vision as a spatial sense? *Psychological Bulletin*, 12(1).

[266] Thorndyke, P.W., Stasz, C. Training Procedure for Map Learning. *The Rand Corporation N-1018-ONR*, pages Non–Paginated, 1979.

[267] Roscoe, S.N. Ergavionics: Designing the job of flying an airplane. *International Journal of Aviation Psychology*, 12(4):331–339, 2002.

[268] Kovalenko, P.A. Psychological aspects of pilot spatial orientation. *ICAO Journal*, 46:18–23, 1991.

[269] Thorndyke, P.W., Hayes-Roth, B. Difference in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 4:560–589, 1982.

[270] Krumhansi, C.L. *Cognitive foundations of musical Pitch.* Oxford University Press, USA, 2001.

[271] Shelton, A.L., Pippitt, H.A. Fixed versus dynamic orientations in environmental learning from ground-level and aerial perspectives. *Psychological Research*, pages 333–346, 2006.

[272] De l'Aune, W., Gillespie, G.M., Carney, J., Needham, W. Frequencies used by blind patients in the detection of open and closed spaces. *45th Annual Meeting of the Eastern Psychological Association Philadelphia*, 1974.

[273] Cotzin, M., Dallenbach, K.M. Facial Vision: the role of pitch and loudness in the perception of obstacles by the blind. *American Journal of Psychology*, 63(4):485–515, 1950.

[274] Riley, L.H., Luterman, D.M., Cohen, M.F. Relationship between hearing ability and mobility and a blinded adult population. *The New Outlook for the Blind*, 58:139–141, 1964.

[275] Carlson-Smith, C., Wiener, W.R. The auditory skills necessary for echolocation: a new explanation. *Journal of Visual Impairment & Blindness*, 90:21–35, 1996.

[276] TutorNext. The Distance Form or Symmetric Form or Parametric Form of a Line, 2011. http://www.tutornext.com/lesson/the-distance-form-or-symmetric-form-or-parametric-form-of-a-line/34799.

[277] Gonzalo N. A Guided Tour to Approximate String Matching. *ACM Computing Surveys*, 33:2001, 1999.

[278] Langley, P. User modeling in adaptive interfaces. In *UM '99: Proceedings of the seventh international conference on User modeling*, pages 357–370, 1999.

[279] Van Beek, P., Cohen, R. Resolving plan ambiguity for cooperative response generation. In *IJCAI'91: Proceedings of the 12th international joint conference on Artificial intelligence*, pages 938–944, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.

[280] Berla, E.P., Butterfield, L.H. Tactile Political maps: Two exploratory designs. *J. Visual Impairment & Blindness*, 71:262–264, 1977.

[281] JOAL. JOAL, 2011. https://joal.dev.Java.net/.

[282] SourceForge. JPEN, 2011. http://sourceforge.net/apps/mediawiki/jpen/index.php.

[283] SoundForge. FreeTTS 1.2, 2011. http://freetts.sourceforge.net/docs/index.php.

[284] Casey, J. *Exploring Curvature*. Friedr. Vieweg & Sohn Verlagsgesellschart, Braunschweig/Wiesbaden, Germany, 1996.

[285] Immediate Mode Gui, 2011. http://c2.com/cgi/wiki?ImmediateModeGui.

[286] Pace, S. A grounded theory of the flow experiences of Web users. *International Journal of Human-Computer Studies*, 60(3):327–363, 2004.

[287] Gustafson, S., Baudisch, P., Gutwin, C., Irani P. Wedge: Clutter-Free Visualization of Off-Screen Locations. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pages 787–796, 2008.

[288] J. Nielsen. Usability inspection methods. In R.L. Mack J. Nielsen, editor, *Heuristic evaluation*. John Wiley & Sons, NY, 1994.

[289] National Gallery of Canada. National Gallery of Canada, 2010. http://cybermuse.gallery.ca/cybermuse/showcases/virtualTour/flash/en/index_e.jsp.

[290] Moffatt, K., McGrenere, J., Purves, B., Klawe, M. The participatory design of a sound and image enhanced daily planner for people with aphasia, booktitle = CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, year = 2004, pages = 407-414, publisher = ACM,.

[291] Vincent, J.L. Learning and Memory: While You Rest, Your Brain Keeps Working. *Current biology*, 19(12):484–486, 2009.

[292] Martin Talbot. SAM Demo: Experiment, 2011. http://www.youtube.com/watch?v=F1HeUUiDVCk.

[293] Ungar S., Blades, M., Spencer, C. Teaching visually impaired children to make distance judgments from a tactile map. *Journal of Visual Impairment and Blindness*, 91(2):164–174, 1997.

[294] Kurze, M., Berlin, F.U. TDraw: A Computer-Based Tactile Drawing Tool for Blind People. In *In Proceedings of Second Annual ACM Conference on Assistive Technologies: Assets '96*, pages 131–138, 1996.

[295] Rassmus-Gröhn, K., Magnusson, C., Eftring, H.E. Iterative design of an audio-haptic drawing application. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, pages 2627–2632, New York, NY, USA, 2007. ACM.

[296] Schiff, W., Foulke, E. *Tactual perception: a sourcebook*. Cambridge University Press, USA, 1982.

[297] Awh, E., Vogel, E.K., Oh, S.-H. Interactions between attention and working memory. *Neuroscience*, 139(1):201–208, 2006.

[298] Scaife, M., Rogers, Y. External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45(2):185–213, 1996.

[299] Lavie, N. Perceptual Load as a Necessary Condition for Selective Attention. *Journal of Experimental Psychology*, 21(3):451–468, 1995.

[300] Lavie, N., Cox, S. On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection. *Psychological Science*, 8:395–398, 1997.

[301] Broadbent, D.E. *Perception and communication*. Pergamon Press, London, 1958.

[302] Frauenberger, C., Stockman, T., Bourguet, M.-L. Pattern design in the context space: A methodological framework for auditory display design. pages 513–518, 2007.

[303] Martin Talbot. SAM Demo: System, 2011. http://bit.ly/cVym1s.

[304] GNU Operating System. Emacs, 2011. http://www.gnu.org/software/emacs/.

[305] Google. KML Tutorial, 2010. http://code.google.com/apis/kml/documentation/kml_tut.html.

[306] Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–393, 1954. (Reprinted in Journal of Experimental Psychology: General, 121(3):262269, 1992).

[307] Su, J., Rosenzweig, A., Goel, A., de Lara, E., Truong, K.N.

[308] Gusfield, D. *Algorithms on strings, trees, and sequences: computer science and computational biology*. Cambridge University Press, Cambridge, UK, 1997.