



Review

Multimodal Navigation Systems for Users with Visual Impairments—A Review and Analysis

Bineeth Kuriakose * , Raju Shrestha and Frode Eika Sandnes

Department of Computer Science, Oslo Metropolitan University, 0167 Oslo, Norway;
raju.shrestha@oslomet.no (R.S.); frodes@oslomet.no (F.E.S.)

* Correspondence: bineethk@oslomet.no

Received: 27 August 2020; Accepted: 12 October 2020; Published: 16 October 2020



Abstract: Multimodal interaction refers to situations where users are provided with multiple modes for interacting with systems. Researchers are working on multimodality solutions in several domains. The focus of this paper is within the domain of navigation systems for supporting users with visual impairments. Although several literature reviews have covered this domain, none have gone through the research synthesis of multimodal navigation systems. This paper provides a review and analysis of multimodal navigation solutions aimed at people with visual impairments. This review also puts forward recommendations for effective multimodal navigation systems. Moreover, this review also presents the challenges faced during the design, implementation and use of multimodal navigation systems. We call for more research to better understand the users' evolving modality preferences during navigation.

Keywords: navigation; multimodal; blind; visual impairments; accessibility

1. Introduction

Navigation is an essential activity in human life. Montello [1] describes navigation as "coordinated and goal-directed movement through the environment by a living entity or intelligent machines." Navigation requires both planning and execution of movements. Several works [2–5], divide navigation into two components: orientation and mobility. Orientation refers to the process of keeping track of position and wayfinding, while mobility refers to obstacle detection and avoidance. Hence, effective navigation involves both mobility and orientation skills.

Several studies have documented that people with visual impairments often find navigation challenging [6–8]. These challenges may include issues with cognitive mapping, lack of access, spatial inference and updating [3,5,6], to mention a few. More complex spatial behaviors, such as integrating local information into a global understanding of layout configuration (e.g., a cognitive map); determining detours or shortcuts; and re-orienting if lost, can be performed only by a person with good mobility and orientation skills. These skills are critical for accurate wayfinding, which involves planning and determining routes through an environment. People with visual impairments may lack those skills and consequently may struggle to navigate successfully [5]. Obstacles can be avoided effectively using conventional navigation aids, such as a white cane or a guide dog [9]. However, these aids do not provide vital information about the surrounding environment. Giudice [6] describes that it is difficult to gain access to environmental information without vision, yet it is essential for effective decision making, environmental learning, spatial updating and cognitive map development. Moreover, visual experiences play a critical role in accurate spatial learning, for the development of spatial representations and for guiding spatial behaviors [3]. Giudice [6] claims that spatial knowledge acquisition is slower and less accurate without visual experience.

The studies conducted by Al-Ammar et al. [10] show that to improve the navigation accessibility of users with visual impairment, one needs to enable navigation as an independent and safe activity. Conventional navigation aids such as white canes and guide dogs have a long history [11,12]. Studies have also shown that there are limitations associated with such conventional tools [11]. To improve upon conventional navigation aids, several navigation systems have been proposed that use different technologies [13–15]. They are designed to work indoors, outdoors or both [15] and rely on certain technologies [2,16]. The World Health Organization (WHO) defines such tools collectively as “assistive technology” [17]. WHO further points out that assistive technology products maintain or improve an individual’s functioning and independence by nurturing their well-being. Hersh and Johnson [11] elaborated that assistive navigation tools for users with visual impairments have the potential to describe the environment such that obstacles can be avoided. Different devices and systems have been proposed for navigation support to users with visual impairments. These devices and techniques can be divided into three categories [9]: electronic travel aids (ETAs), electronic orientation aids (EOAs) and position locator devices (PLDs) [18]. ETAs are general devices to help people with visual impairments avoid obstacles. ETAs may have sensing inputs such as depth cameras, general cameras, radio frequency identification (RFID), ultrasonic sensors and infrared sensors. EOAs help visually impaired people navigate in unknown environments. These systems provide guiding directions and obstacle warnings. PLDs help determine the precise position of a device, and use technologies such as the Global Positioning System (GPS) and geographic information systems (GISs). Lin et al. [18] gives a detailed explanation of these categories. In this paper, the term “navigation system” is used to denote any tool, aid or device that provide navigation support to users with visual impairments. In addition, we use the term “users” or “target users” to denote the term “users with visual impairments”.

Researchers have been exploring the navigation support applications of emerging technologies for several decades [19]. Advancements in computer vision, wearable technology, multisensory research and medicine have led to the design and development of various assistive technology solutions, particularly in the domain of navigation systems for supporting users with visual impairments [14,20]. Ton et al. [21] observed that the research had explored a wide range of technology-mediated sensory-substitution to compensate for vision loss. The developments in artificial intelligence (AI)—in object detection using machine learning algorithms, location identification using sensors, etc.—can be exploited to understand the environment during navigation. The developments in smartphone technologies have also opened up new possibilities in navigation system design [22–24]. One key challenge is how to communicate the information in a simple and understandable form to the user, especially as other senses (touch, hearing, smell and taste) have lower bandwidths than vision [13]. Therefore, effective communication of relevant information to the users is a major requirement for such navigation systems.

Bernsen and Dybkjær [25] defines the term modality in the human–computer interaction (HCI) domain as a way of representing or communicating information in some medium. The term “multimodality” refers to the use of different modalities together to perform a task or a function [26,27]. Modalities are typically visual, aural or haptic [28]. Navigation systems that use different modes to communicate with the user are called multimodal navigation systems [29]. Several multimodal systems were proposed to assist users for navigation [30–32]. Many prototypes have been reported without much practical evaluation involving target users [30]. It is therefore uncertain whether these proposals offer any actual benefits to users. A few studies have also been published with convincing validation involving the users [33,34].

Several surveys have addressed navigation systems designed for users with visual impairments [13–15,35]. Some focused on the types of devices or technology, while others on the environments of use. To the best of our knowledge, no systematic surveys have addressed multimodal navigation systems. This paper, therefore, provides an overview of the major advances in multimodal navigation systems for supporting users with visual impairments.

This paper is organized as follows. Section 2 presents the general theory of multimodality. Section 3 gives a brief overview of the application of multimodality in the navigation system and also describes the methodology we used for this review. Section 4 discusses the multimodal navigation systems and their affiliated studies. Section 5 summarizes the challenges and presents a set of recommendations for the design of a multimodal navigation system for people with visual impairments. The paper concludes in Section 6.

2. Multimodality in Human–Computer Interaction

In the context of HCI, a modality can be considered as a single sensory channel of input and output between a computer and a human [28,36]. A unimodal system uses one modality, whereas a multimodal system relies on several modalities [36]. Studies have shown that multimodal systems can provide more flexibility and reliability compared to unimodal systems [37,38]. Oviatt et al. [39] elaborates on the possible advantages of a multimodal interaction system, such as freedom to use a combination of modalities or to switch to a more-suited modality. Designers and developers working with HCI have also tried to utilize different modalities to provide complementary solutions to a task that may be redundant in function but convey information more robustly to the user [40,41]. Based on the perception of information, modalities can be generally defined in two forms: human–computer (input) and computer–human (output) [27]. During the interaction, the available input modalities are utilized by the user to communicate with the system, and the system uses several output modalities to communicate back to the user [42].

Computers utilize multiple modalities to communicate and send information to users [43]. Vision is the most frequently used modality, followed by audio and haptic. Haptic communication occurs through vibrations or other tactile sensations. Examples of touch-based (haptic) modality channels include smartphone vibrations. The other modalities such as smell, taste and heat are less used in interactive systems [44]. Audio offers the benefits of rich interaction experiences depending on the context of use and helps provide more robust systems when used in combination with other modalities [45]. Such redundancies are used when a user wants to communicate with a system via voice while driving a car without taking the hands off the steering wheel.

Epstein [46] and Grifoni et al. [47] have shown that with the increasing use of smartphones and other mobile devices, users are becoming more comfortable in experimenting with different new modalities. After the introduction of voice assistants such as Siri, Alexa, Cortana and Google Home, some users began to use voice assistants as an alternative way to communicate with computers and other digital devices [48,49]. This epitomizes how certain modalities with contrasting strengths are useful in various situations [50]. Some other modalities such as computer vision can be utilized to capture three-dimensional gesticulations using depth cameras, such as Microsoft Kinect [44].

Multimodal systems have the potential to increase accessibility to users by relying on different modalities. Due to the benefits of using multimodal inputs and outputs, multimodal fusion is also used in various applications to support user needs [51]. The process of integrating information from multiple input modalities and combining them into a specific format for further processing is termed multimodal fusion [52,53]. To allow their interpretation, a multimodal system must recognize different input modalities and combine them according to temporal and contextual constraints [47,54,55]. An example of a multimodal human–computer interaction system is illustrated in Figure 1. This two-level flow of modalities (action and perception) explains how a user and a system interact with each other and also the different steps involving in the process [56].

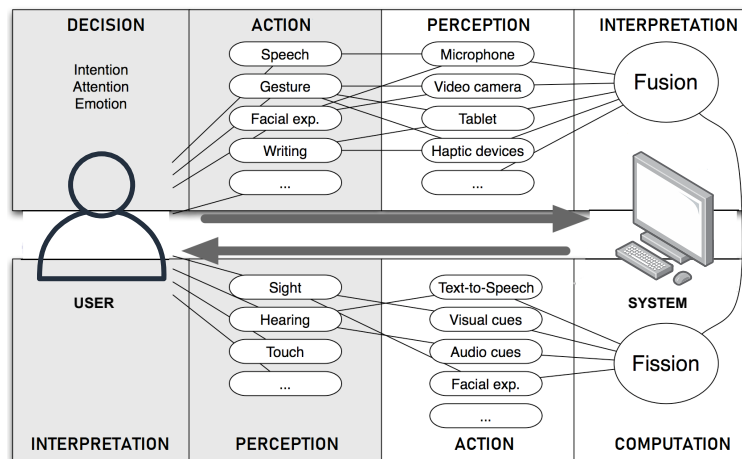


Figure 1. Multimodal human–computer interaction (adapted from [56]).

3. Multimodality in Navigation Systems

Multimodal navigation systems have several advantages compared to unimodal navigation systems. Vainio [57] explained that multimodal navigation systems allow the user the flexibility to give inputs or receive outputs, in a preferred modality. He also emphasized the need for developing multimodal navigation systems to assist mobile users. Brock et al. [58] also showed that navigation systems proposed for users with visual impairments could not be considered as effective if the inputs and outputs depend upon only a single mode of interaction. Multimodalities help improve the system robustness [45]. This is helpful in situations where one of the modalities fails, and a different modality can be used instead [55]. This is mostly applicable in a navigation system with different redundant modalities which serves a similar function in the system [59]. Jacko [43] argued that the multimodal navigation systems allow for greater accessibility and flexibility for users who can perform tasks much better with unimodal systems. Sears and Jacko [60] affirmed that different combinations of modalities have the possibility to enhance user comfort in human–computer interactions. For example, in a noisy environment, vibratory feedback may be more effective than aural feedback when receiving directions. Alternatively, audio may be a more suitable choice, if the user wants to get more details about the environment of navigation such as landmarks and traffic signs.

Although there are a vast number of documented studies on assistive navigation systems for people with visual impairments, we were unable to find many studies exploring multimodality. We used the major publication databases ACM Digital Library, IEEEExplore, ScienceDirect and Google Scholar to find the relevant publications matching with the inclusion criteria. We used the keywords “navigation systems + visually impaired” and “navigation systems + blind”. After reviewing the abstracts, we excluded those that were outside the scope. The papers selected for this review were not limited to those documented as complete and functioning systems, but also at prototyping stages.

The reviewed papers have been categorized and discussed in the sub-sections based on how the multimodality concepts were utilized. Papers describing navigation systems which use the multimodal interaction were placed in one group. Next, papers describing interactive map-based multimodal systems constituted the second category. The third category included papers that document multimodal interfaces. Papers which focus on virtual environments for training the users for using multimodal navigation systems belonged to the fourth category.

3.1. Multimodal Navigation Systems

Multimodal navigation systems hold the potential for enhancing the accessibility for users with visual impairments. In a multimodal system, the user has the flexibility to give instructions and to receive the guidance in their most preferred modality.

The EyeBeacons system [61] is a framework for multimodal wayfinding communication using wearable devices. The framework uses three different modalities for passing navigation instructions: aural, tactile and visual. The system has three main components: a bone conduction headset, a smartphone and a smartwatch. Bone-conduction headphones rely on the sound being transmitted through vibrations on the bones of the head and jaw, instead of eardrums as in traditional headsets. This is particularly useful for improved situational awareness [62]. A smartwatch was used to sense the wayfinding messages in the form of vibrations. The participants who tested the system reported that both vibrations and audio tones were difficult to distinguish. The Assistive Sensor Solutions for Independent and Safe Travel (ASSIST) indoor navigation system [33] was also designed to give three types of sensory feedback to the users similar to those reported in [61], namely, visual, aural and tactile. The system's usability testing was carried out with users, and they expressed favorable opinions about the system. The participants also suggested offering options to turn on or off certain features.

Tyflos [30] is a multimodal assistive piece of technology designed for reading and navigation. A stochastic Petri-net model is used to drive its multimodal interaction. A camera captures visual information from the environment. This visual information is transformed into either vibratory or aural feedback. The user communicates with the system via a speech recognition interface. The feedback information is communicated to the user through a vibration array vest attached to the abdomen. The authors did not report any user evaluation of the system.

Gallo et al. [63] proposed a system which can be integrated with a conventional white cane and thus provides multimodal augmented haptic feedback. The multimodal feedback system consisted of a shock system to simulate the behavior of a long cane, a vibrotactile interface to display obstacle distance information and an auditory alarm system for head level obstacles. The device is triggered when a distant obstacle is detected, and the user experiences a sensation in the cane handle. The auditory feedback is mainly used as an emergency handler to alert the users. User evaluation showed that object detection and distance information was helpful and easy to understand. However, they expressed that to get better estimations of the distances to the obstacles, users needed training using the device. The Range-IT system [64] is a similar system which uses a white cane. After detecting the obstacle using a 3D depth camera, Range-IT provides information such as type of object, distance and direction in relation to the user, using an aural-vibrotactile interface. The output from the vibrotactile belt and the sonification messages, along with the verbal messages from a bone conduction headset, helped participants to perceive multimodal feedback during the navigation in a laboratory setup. The weight was an issue with the prototype as the user had to carry a laptop and 3D cameras.

HapAR [31] is a mobile augmented reality application which was introduced to guide users around a university campus. The user can activate the application by giving a Siri voice command. The system processes the request and tries to find the location of interest. When the user is close to the destination or any point of interest, both aural feedback and haptic feedback are triggered. User feedback showed that sound feedback was masked by outdoor environment noises such as wind and people talking. Additionally, the intensity of the haptic feedback varied with different smartphone models, negatively affecting the system's performance. Another similar system which provided both aural and tactile feedback is Personal Radar [65]. This indoor system performs obstacle detection, provides the current location and gives directions.

NavCog [66] is a smartphone-based navigation system for blind users. The system uses a network of Bluetooth low energy (BLE) beacons. The NavCog interaction was designed to avoid overloading the user with cognitively demanding messages. NavCog uses simple sounds and verbal cues to give turn-by-turn instructions. Users interact with the system through a simple touch interface. NavCog also informs users about nearby points of interest (POI) and possible accessibility issues. The system needs to be improved in terms of localization accuracy to avoid confusion when making small turns.

iASSIST [67] is an iOS-based indoor navigation application for both sighted and visually impaired users. Hybrid indoor models were created with Wi-Fi/cellular data connectivity, beacon signals and a 3D spatial model. During the navigation stage, the user with the mobile application is localized within the floor plan using the connected data network to give an optimal route to the destination. The system uses visual, aural and haptic feedback to provide turn-by-turn navigation instructions to the user. The limitations of the system include dependability on data connectivity in delivering services and the absence of obstacle and scene understanding features. The authors did not report any user evaluation results.

Fusiello et al. [32] proposed a navigation system which used a combination of stereo vision and sonification. The user would hear the sound in the environment with a stereo headset. Visual processing includes the segmentation of objects detected and corresponding three-dimensional (3D) reconstruction. The aural processing includes the experiential enhancement of the 3D scene through artificially created sounds. The system provides auditory cues to help the user to identify the position and distance of the pointed object or surface. The system is strenuous to use, as the user has to continuously listen to audio signals. Similarly, Sound of Vision [34] also provides a three-dimensional representation of the environment through sound and tactile modalities.

The Personal Guidance System [68,69] consists of different components, such as a module for determining the traveler's position and orientation in space, a GIS comprising a detailed database for route planning and a user interface. The system has different display modes such as spatialized sound from a virtual acoustic display, and verbal commands issued by a synthetic speech display. Compared to verbal commands, the virtual display showed the highest effect in terms of both guidance performance and user preferences. Disadvantages of this system include partially occluded external sounds which are essential in echolocation, and a high system weight, making it impractical to carry around. There is also an additional cost and complexity associated with virtual acoustic hardware.

The system proposed by Wang et al. [70] included a camera and an embedded computer with three feedback modes, vibration, braille and audio. The system used techniques from computer vision and motion planning to identify walkable space, and recognize and locate specific types of objects such as chairs. These descriptions are communicated through vibrations. The user also receives feedback via a braille display and audio that is synthesized using text-to-speech. The evaluation of the system was conducted by blind participants. The user evaluations showed that the haptic obstacle feedback was more comfortable. Braille displays offered richer high-level feedback but had longer reaction times due to sweeping of the fingers on the braille cells. Audio feedback was considered undesirable because of the low refresh rate and long latency, and due to the potential obstruction of other sounds from the environment.

None of the systems reviewed here fully utilized the multimodality concept. Moreover, only a small fraction of the studies conducted convincing user evaluations. A consolidated summary of the different multimodal navigation systems is given in Table 1. The table categorizes the reviewed systems with the main software and hardware components, localization technologies and modalities involved.

Table 1. Multimodal navigation systems.

System	Main Software/Hardware Components	Localisation Technologies	Modalities Involved
<i>Tested with target users</i>			
EyeBeacons [61]	Smartphone, bone conduction headset, smartwatch, SideWalk wayfinding framework.	IMU	Aural, Visual and Tactile
ASSIST [33]	Smartphone, BLE beacons, Google Tango.	IMU, BLE beacons	Aural, Visual and Vibration Alerts
Sound of Vision [34]	IR-based depth sensor, stereo cameras and IMU device.	IMU	Aural and Tactile
NavCog [66]	Smartphone, BLE beacons.	BLE beacons	Sound alerts and Verbal Cues
Personal Guidance System [68]	GPS receiver, GIS, keypad, earphones, speech synthesizer, acoustic display hardware.	GPS	Aural and Verbal Commands
Wearable Vision-based System [70]	Embedded computer, depth sensor camera, haptic array, braille display.	Vision-based	Haptic, Braille and aural
<i>Tested with blindfolded users</i>			
HapAR [31]	Smartphone, voice recognizer.	IMU	Voice, Audio and Vibration
Personal Radar [65]	Ultrasonic sensors, tactile actuators, and Arduino ATmega2560 microcontroller.	Ultrasonic-based	Aural and Vibrotactile
<i>Trail Evaluation</i>			
Augmented White Cane [63]	White cane, distance and obstacle sensors, shock device electronics, vibrating motors.	Ultrasonic-based	Shock Simulations, Vibrotactile, Audio Alerts
Electronic Travel Aid [32]	Earphones, sunglasses fitted with two micro cameras and palmtop computer.	Vision-based	Aural and Visual
Range-IT [64]	3D depth camera, 9 Degrees of Freedom IMU, bone conduction headphone, vibrotactile belt, smartphone.	Vision-based	Aural and Vibrotactile
<i>No evaluations reported</i>			
Tyflos [30]	Stereo cameras, microphone, ear speakers, portable computer and vibration array vest.	Vision-based	Aural and Vibration
iASSIST [67]	Smartphone, ARKit, Bluetooth beacons, 2D/3D models.	Beacons, Wi-Fi/cellular	Voice and Vibrations

3.2. Interfaces

Several computer interfaces have been designed and developed to enhance the interaction between humans and computers. The usage of multimodal interfaces in navigational systems allows users to interact with systems using several communication modes. Diaz and Payandeh [71] argue that multimodal interfaces enable powerful, flexible and feature-rich interactive experiences.

ActiVis [72] was implemented with the main objective of giving necessary directions to the users by perceiving its surroundings. By creating a multimodal user interface, ActiVis is designed to help users receive navigational information in the form of aural and vibration cues in a more effective manner. This multimodal interface was implemented on Google's Project Tango device developed using Android and Tango SDKs. Their multimodal user interface includes a co-adaptive module to help users learn user behavior over time and also adapt the feedback parameters to improve user performance.

Bellotto et al. [73] proposed a concept for a multimodal interface for an active vision system to control a smartphone camera orientation, using a combination of verbal messages, 3D sounds and vibrations. It was implemented as a smartphone application. Usability tests were conducted with several blindfolded users to identify the accuracy, success rate and user response times. Users reported difficulty in interpreting the sound signals correctly.

TravelMan [74] was introduced as a multimodal mobile application for serving public transport information in Finland. It also provides pedestrian guidance for users with visual impairments. The application supports several output modalities, including synthesized speech, small display-based graphical elements using fisheye techniques, non-speech sounds and haptics. The input modalities of the system consist of text input, speech recognition, physical gestures and positioning information. The camera-based movement detection was reported to be less robust, and the physical gestures feature needed to be expanded. Another drawback of the system was with the graphical interface, which is language-dependent and thus requires much display space.

The systems reviewed in this section gives an overview of how multimodality can be used in navigation application interfaces. Table 2 provides summarized information about the papers on multimodal interfaces.

Table 2. Multimodal interfaces.

System	Main Software/Hardware Components	Modalities Involved
<i>No evaluations reported</i>		
ActiVis [72]	Google Tango, bone conducting headset.	Aural and Vibration Cues
Human-in-the-Loop [73]	Smartphone, headset, IVONA TTS engine, OpenAL.	Vocal Messages, Aural and Vibrations
TravelMan [74]	Smartphone with camera, Bluetooth GPS device.	Aural, Visual and Tactile

3.3. Maps

Visual maps have several advantages as a tool for navigation, as they can give an overview of an environment and possess high information density. Over the last decades, several promising technologies have emerged that replace visual maps, such as point and sweep gestures, spatial sound, tactile information and other multimodal options. Tactile maps give users access to geographical representations. Although those maps serve as useful tools for the acquisition of spatial knowledge, they have some limitations, such as the need to read braille. Ducasse et al. [75] did an exhaustive review of interactive map prototypes. The authors compared the maps based on cost, availability, technological limitations, content, comprehension and interactivity. They suggested improving the accessibility of digital maps using wearable technologies and designing interaction techniques that provide users with more interactive functions, such as zooming and panning, for map exploration. In addition to several interactive digital maps, different multimodal maps have been proposed.

Brock et al. [76] proposed an interactive multimodal map prototype, which relies on a tactile paper map, a multi-touch screen and aural output. Four steps were involved in the design of the interactive map. The first step involved drawing and printing the tactile paper map. The second step concerned the choice of multi-touch technology. The third step included the selection of output interaction technology. The final step dealt with the selection of the software architecture for the prototype. The prototype was made with different software modules interconnected with middleware. The authors claimed that the prototype could be used as a platform for advanced interactions in spatial learning. User evaluations showed that some users found multi-touch and double-tapping difficult.

An instant tactile-aural map prototype was proposed by Wang et al. [77] that automatically created interactive tactile-aural maps from the local visual maps. The multimodal maps generated by the system could be used for navigation. The first step in the system is to extract text from local map images. The second step involves the recreation of tactile graphics. The third step comprises of multimodal integration and rendering. The final results are multimodal tactile-aural representations of the original map images. The users get instant aural annotations associated with the map graphics by pressing certain symbols in the generated map. Some of the shortcomings reported with the system include issues with graphics conversion, which may lead to broken navigation paths in the tactile map.

Talking TMAP [78] was a system which was designed to help with the automated generation of aural-tactile maps using Smith-Kettlewell's TMAP software. It combines Internet content, a geographic information system, braille embossers and a touch tablet to create aural-tactile street maps of neighborhood areas. There is an extra device called Talking Tactile Tablet (TTT) connected to the system which acts as a tactile graphics viewer.

The Vibro-Audio map (VAM) proposed by [79] supported environmental learning, cognitive map development and wayfinding behavior. VAM used a low-cost touchscreen-based multimodal interface of a commercial tablet. VAM was an example of a digital interactive map (DIM) that was rendered using vibrotactile and auditory information. The built-in vibration motor of the tablet device was used to provide haptic (vibrotactile) output. Evaluations conducted with target users showed that VAM performed similarly to the traditional tactile map overlays. The findings from the study were limited to indoor building environments only.

The TouchOver map study [80] investigated whether vibration and speech feedback can be used to make a digital map on a touchscreen device. The prototype consisted of an android map navigation application. When the user touched the map where there were underlying roads, the device vibrated and read the name of the road. Their results indicated that it is indeed possible to get a basic overview of the map layout, even if a person does not have access to the visual presentation. Shortcomings include the inability to detect whether roads are close and whether they cross. It is also hard to determine the directions of short roads.

The audio-tactile you-are-here (YAH) map system [81] presented map elements and updated location on a mobile pin-matrix display. The system consists of a set of tactile map symbols with raised and lowered pins representing varying map elements. Users can input map operation commands (panning, zooming, etc.) via either a mobile phone or an electronic cane. A field test was conducted with both visually impaired and blindfolded users who did not have experience with tactile maps and braille. Conclusions were that the system needed higher location accuracy, improved portability and a one-hand map exploration method.

Touch It, Key It, Speak It (Tikisi) [82] was a software framework for the accessible exploration of graphical information. Tikisi facilitated multimodal input through multi-touch gestures, keystrokes and spoken commands; and aural output. The system was used by moving a finger across a geographical map and issuing commands to go to specific locations such as cities or states. The testing of the Tikisi was done with target users. Feedback was positive. However, Tikisi used a standard tablet, so the shape recognition was not possible; c.f. tactile displays. Additionally, because of the lack of tactile feedback, it was not easy to estimate the relative size of the two objects.

SpaceSense [83] was a map application that ran on an iPhone. It was used for representing geographical information and also included custom spatial tactile feedback hardware. SpaceSense uses multiple vibration motors attached to different locations on the mobile touchscreen device. It offers high-level details on the distance and direction towards a destination and bookmarked locations. Through vibrotactile and sound feedback, the application helps users to maintain the spatial relationships between points. However, the system was only tested in one neighborhood. More work is needed to understand how the number of places and the route instructions affect the spatial relationship learning capability of users.

This section discussed how digital tactile multimodal maps could enhance navigation accessibility. We observed that only a few works mentioned here had explored the developments in sensor-based technologies on their research. A summary of the works is shown in Table 3.

Table 3. Multimodal maps.

System	Main Software/Hardware Components	Modalities Involved
<i>Tested with target users</i>		
Interactive Map [76]	Multi-touch screen, inkscape editor, TTS engine, middleware.	Aural and Tactile
Instant Tactile-Audio Map [77]	Tactile touchpad, SVG, tactile embosser.	Aural and Tactile
Vibro-Audio Map (VAM) [79]	Tablet	Aural and Tactile
TouchOver Map [80]	Android device, OpenStreetMap	Aural and Tactile
You-Are-Here (YAH) Map [81]	Touch-Sensitive Pin-Matrix Display, Mobile Phone, Wiimote Cane, Computer, OpenStreetMap	Aural and Tactile
Tikisi [82]	TikiSi Framework	Multitouch Gestures, Voice Commands, Speech
SpaceSense [83]	iPhone device, vibration motors, FliteTTS1 package	Aural and Tactile
<i>No evaluations reported</i>		
Talking TMAP [78]	Braille embossers, SVG, tactile tablet, TMAP software, macromedia director.	Aural and Tactile

3.4. Virtual Learning Environments

Virtual navigation environments allow users to experience navigating unknown locations in safety. The device setup can be used to simulate a real navigation experience by receiving feedback through different modalities. Moreover, the virtual environment can be controlled with parameters such as complexity level, and users can analyze the effects of various modalities [84]. Different multimodal virtual environments have been proposed to meet users' navigation needs.

Haptic and audio multimodality to explore and recognise the environment (HOMERE) [85] was a virtual reality (VR)-based multimodal system for exploring and navigating inside virtual environments. The system provided four types of feedback to stimulate feedback in a real environment. The force feedback was complementary to the cane simulation, thermal feedback complementary to the sun simulation and auditory feedback to the ambient atmosphere. Visual feedback was implemented for partially sighted people or sighted people to follow the navigation guidance from the virtual environment.

NAV-VIR [86] was a multimodal virtual environment designed to help discovering and exploring unknown areas. The system used aural–tactile feedback. The NAV-VIR system comprised two parts: (1) An interactive tactile interface called the Force Feedback Tablet (F2T). It output spatial information about possible paths for navigation. (2) A dynamic audio environment that provided a realistic, orientation-aware 3D simulation of the audio cues during the actual journey of a user with visual impairments. A preliminary evaluation done with target users showed that NAV-VIR was capable of generating convincing tactile stimuli.

A virtual environment platform for the development of assistive devices was proposed by Khoo et al. [87]. The environment was designed to evaluate multimodal sensors to be used in navigation and orientation tasks. The main focus of the work was to help in the design of the sensor interfaces and simulators in the virtual environment for future experimentation.

Canetrroller [88] was designed to simulate white cane interactions to help transfer cane skills into the virtual world. A VR headset was used for 3D sound. Three types of multimodal feedback could be experienced by the user. First, physical resistance was generated by the controller when the virtual cane came in contact with virtual objects. Second, vibrotactile feedback simulated when the white cane touched objects. Third, spatial 3D audio simulated sound from the real world. The system was designed to work in both indoor and outdoor virtual environments.

The BlindAid system [89] was equipped with a haptic device and stereo headphones which provided multimodal feedback during the interaction. The system assisted the users with exploring the

virtual environment based on their prior real space orientation skills. Additionally, the system provided spatial landmarks through haptic and auditory cues. Three modalities of operation (visual, aural and haptic) provided spatial information.

The system proposed by Kunz et al. [84] helped users build a cognitive navigation map of the surroundings. The virtual environment was controllable and could map objects such as walls and stairs to real-world entities. The user received acoustic and/or haptic feedback when an obstacle was detected in the environment.

Most of the papers discussed here employed recent technological advancements such as VR and 3D sound. A summary of the works on virtual learning environments for navigation is given in Table 4.

Table 4. Multimodal virtual navigation environments.

System	Main Software/Hardware Components	Modalities Involved
<i>Tested with target users</i>		
NAV-VIR [86]	Tablet, 2 servomotors moving a flat joystick, Arduino single-board microcontroller, and immersive HRTF-based 3D audio simulation, Google VR Audio.	Force, 3D Audio Cues
HOMERE [85]	VIRTUOSE 3D, Haptic device, infrared lamps, speakers, gamepad, Sense8 for graphics rendering and VORTEX 1.5 for collision detection speakers.	Force, Thermal, Aural, Haptic and Visual
Canetroller [88]	Folding canes, magnetic particle brake, voice coil, Vive tracker, VR headset for 3D audio, IMU with gyro and accelerometer, Unity game engine.	Breaking feedback, Vibrotactile and 3D Audio
BlindAid [89]	Computer, desktop phantom.	Aural and Haptic
<i>Tested with blindfolded users</i>		
Virtual Navigation Environment [84]	Intersense DOF tracking system, Oculus Rift DK, headphones, Arduino Uno microcontroller, Unity game engine.	Aural and Haptic
<i>No evaluations reported</i>		
Multimodal Sensors VE [87]	XBox controller, Microsoft Kinect, head mounted electrodes, Brainport's vision technology.	Aural, Vibration and Haptics

4. Discussion

Behavioral and cognitive neuroscience studies conducted by Ho et al. [90], Stanney et al. [91] and Calvert et al. [92] show that systems with multiple modalities can maximize user information processing. Moreover, systems designed with multimodal preferences can provide various combinations of signals from different sensory modalities and subsequently have beneficial effects on user performance with a particular system. Lee and Spence [93] argued that the presentation of multimodal feedback outputs to users was found to have enhanced performance and more pronounced benefits over unimodal systems. However, other studies [94–96] claim that multimodal feedback modes can confuse users. Confusion may occur when the user has several multimodal options for one function with a similar purpose but can enter into a dilemma on what to choose and which one is better. In terms of design considerations of a multimodal navigation system, it is important to consider how effectively and easily each of the multimodal feedback methods can be utilized by the users.

Common modalities used in almost all multimodal navigation systems are aural and tactile (see Table 1). Some systems enhance the audio with spatial or 3D audio. Many systems tested their prototypes with the target users, while some reported tests with blindfolded users. Some authors did not document any user evaluation of their systems.

The multimodal interfaces discussed in this review mostly utilized mobile systems such as tablets and smartphones (see Table 2). Different modalities such as aural (in the form of messages and

non-speech alerts) and vibration cues were used in the systems for interaction. The multimodal maps mainly used the aural and tactile modalities (see Table 3). In cases in which systems have not been evaluated with users, it is impossible to conclude whether the intentions of the systems were met.

The different hardware components in the virtual-navigation laboratory environment can simulate different modalities in the real-time environments (see Table 4). The common modalities—such as audio, and haptic and its variants, such as vibration and force feedback—could be simulated in a virtual environment setup. These virtual training platforms claimed to help the users to experience the multimodal navigation systems before proceeding to the navigation in real environments.

It is interesting to note that almost all multimodal systems reviewed here employed the aural modality. Some user-interaction studies validate this perspective by pointing out that speech cues can be useful in providing mobility feedback for users [97,98]. This may be because audio modality can be a suitable choice when the users want to hear information about the environment during navigation. Moreover, empirical results show that users are uncomfortable when using audio feedback in public environments [99]. In noisy environments such as public places, it can be challenging to hear audio feedback. Additionally, users might have a social stigma when they think that auditory feedback is audible to the public as well. There are also privacy and security concerns in using audio feedback in public environments.

The advantage of haptic feedback is that the users can use it anywhere, anytime, without interrupting others. At the same time, vibrations are often more similar to each other, and not easy to distinguish compared to auditory feedback, which might create confusion among users [99].

5. Recommendations and Challenges in the Design of Multimodal Navigation Systems

5.1. Recommendations

Multimodal technology is a promising candidate in human–machine interfaces which may improve the accessibility within user environments such as mobile devices and navigation systems. The study conducted by Giudice [6] showed the importance of developing useful learning strategies to remedy travel-related problems faced by the users and argued that the focus of the research must be redirected to consider spatial information from all sensory modalities. Wentzel et al. [100] also supported the fact that multimodality is the key necessity for accessibility for a broad audience. Moreover, the authors confirmed that in an accessibility system, different modalities of interaction should be available and should be equivalent to each other. The European Telecommunications Standards Institute (ETSI) guidelines [101] prioritize the use of multimodal presentation and interaction for accessible systems. Two multimodality principles are mentioned in the guidelines. First, the use of multimodal presentation of information, which allows users with different preferences and abilities to use information in their desired manner. Second, the use of multimodal interaction to allow users to interact with a system, which follows individual needs and preferences. In addition, Wentzel et al. [100] suggested that a system or an application should be able to provide relevant multimodal feedback on user behavior. Since different users have different preferences, it is more appropriate to have customizable system settings for input/output modalities and frequency of feedback.

Based on the review, analysis and discussion presented herein, we put forward the following recommendations.

- **Multimodality**—multiple modalities should be available, and among them, audio feedback is always expected.
- **Customisability**—flexible customisation option should be available for user-preferred settings.
- **Extendibility**—it should be possible to extend a new feature or a new modality at a later stage.
- **Portability**—the whole system should be portable and should not create an extra burden to the user with many devices.
- **Simplicity**—adding additional modalities should not make the users feel that the system is complex or create confusion in selecting them.

- **Dynamic mode selection**—it should allow users to dynamically select the most appropriate mode of interaction for their current needs/environments.
- **Adaptability**—using machine learning techniques, multimodal systems can be designed to be adaptable based on varying environments.
- **Privacy and security**—it should address both the privacy and the security of the user.

5.2. Challenges

Adoption of multimodality in navigation system design also introduces some challenges. There can be system implementation level challenges and challenges associated with user adaptability to a new system. System-level challenges may occur during the stages such as data acquisition, transfer, fusion, processing and in the final delivery in a suitable form and an environment. Limited availability of multimodal datasets for navigation purposes is a barrier in the related research works [102]. Even though some multimodal datasets have been reported [103–105], the contributions in this area are scarce. Developers can face challenges in training and implementing the system with these limited options.

Vainio [57] showed that when users interact with a multimodal navigation system, they exhibit different patterns. For instance, either the users interact with a system simultaneously, or they integrate their interaction sequentially. Designing a system with varying user preferences could be challenging for developers. The processing stage in a multimodal system involves additional hurdles. Finding a suitable fusion level and fusion algorithm for the multimodal data is one of them. Caspo et al. [106] pointed out the trade-offs between the storage and the generation of different multimodalities while designing a multimodal system. Even though some flexibility in generation in terms of real-time parametrization exists, more complex processing is required. This review does not go into detail about the technical aspects of multimodal fusion and fission, but the complexity involved in the implementation is high [102].

Another challenge in the multimodal system design is the delivery of information in a user-preferred form. Different users have different interests and preferences, and these can change depending on their navigation environments. Implementing a system based on the adaptation to user preferences and learning them according to the situations and environmental conditions could be difficult [106]. Developers cannot decide which the most appropriate modality is and what is more favorable for a particular user. Moreover, if developers integrate too many multimodal feedback options, the user can get confused and distracted. As stated by Liljedahl et al. [107], good user experiences do not require cutting edge technology, but careful design of a multimodal user-centered system could provide better results. Determining the appropriate multimodal feedback methods based on the changing environmental conditions should be done before designing a system. One possible approach to address this would be to use machine learning to understand user preferences and make suitable recommendations. Research shows that it is possible to implement a system with self-learning to enable adaptive settings based on user preferences in varying environments [108,109].

Yet another challenge is how to make a multimodal system comfortable for users. Any user with visual impairment should not experience any difficulty in using the system. Moreover, adding multiple modalities is analogous to adding an extra layer of complexity to the system. Experiences of difficulties may lead to abandoning the technology [110].

6. Conclusions

Multimodal technology can be considered as a promising option that can be utilized in the design of effective and accessible navigation systems for users with visual impairments. Multimodal navigation solutions proposed for people with visual impairments were reviewed. The primary modalities that are utilized in almost every multimodal navigation system discussed in this review are aural and tactile. Even though many multimodal navigation solutions have been proposed, there is little evidence of what degree the target users continued to use these technologies in practice.

Studies concerning the extent to which multimodal systems are helpful for people with visual impairments in real-life navigation contexts are an important avenue to consider. Challenges are associated with designing, implementing and using multimodal navigation systems. Exploring the effectiveness of recent advancements in artificial intelligence and related technologies to help with tackling the different challenges in multimodality is an important area of future research. Moreover, we argue that more studies are needed to better understand the evolving preferences in modalities among users with visual impairments.

Author Contributions: Conceptualization, B.K.; methodology, B.K.; investigation, B.K.; writing—original draft preparation, B.K.; writing—review and editing, R.S. and F.E.S.; visualization, B.K.; supervision, R.S. and F.E.S.; project administration, R.S.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the anonymous reviewers for their valuable comments throughout the manuscript revision process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Montello, D.R. Navigation. In *The Cambridge Handbook of Visuospatial Thinking*; Cambridge University Press: Cambridge, UK, 2005; pp. 257–294.
2. Giudice, N.A.; Legge, G.E. Blind navigation and the role of technology. In *The Engineering Handbook of Smart Technology for Aging, Disability, Additionally, Independence*; Wiley: Hoboken, NJ, USA, 2008; Volume 8, pp. 479–500.
3. Schinazi, V.R.; Thrash, T.; Chebat, D.R. Spatial navigation by congenitally blind individuals. *WIREs Cogn. Sci.* **2016**, *7*, 37–58. [[CrossRef](#)] [[PubMed](#)]
4. Thinus-Blanc, C.; Gaunet, F. Representation of space in blind persons: vision as a spatial sense? *Psychol. Bull.* **1997**, *121*, 20. [[CrossRef](#)] [[PubMed](#)]
5. Long, R.G.; Hill, E. Establishing and maintaining orientation for mobility. In *Foundations of Orientation and Mobility*; American Foundation for the Blind: Arlington County, VA, USA, 1997; Volume 1.
6. Giudice, N.A. Navigating without vision: Principles of blind spatial cognition. In *Handbook of Behavioral and Cognitive Geography*; Edward Elgar Publishing: Cheltenham, UK, 2018.
7. Riazi, A.; Riazi, F.; Yoosfi, R.; Bahmehi, F. Outdoor difficulties experienced by a group of visually impaired Iranian people. *J. Curr. Ophthalmol.* **2016**, *28*, 85–90. [[CrossRef](#)] [[PubMed](#)]
8. Manduchi, R.; Kurniawan, S. Mobility-related accidents experienced by people with visual impairment. *AER J. Res. Pract. Vis. Impair. Blind.* **2011**, *4*, 44–54.
9. dos Santos, A.D.P.; Medola, F.O.; Cinelli, M.J.; Ramirez, A.R.G.; Sandnes, F.E. Are electronic white canes better than traditional canes? A comparative study with blind and blindfolded participants. In *Universal Access in the Information Society*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 1–11.
10. Al-Ammar, M.A.; Al-Khalifa, H.S.; Al-Salman, A.S. A proposed indoor navigation system for blind individuals. In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services, Ho Chi Minh City, Vitenam, 5–7 December 2011; pp. 527–530.
11. Hersh, M.; Johnson, M.A. *Assistive Technology for Visually Impaired and Blind People*; Springer Science & Business Media: Berlin, Germany, 2010.
12. Wendt, O. *Assistive Technology: Principles and Applications For Communication Disorders and Special Education*; Brill: Leiden, The Netherlands, 2011.
13. Chanana, P.; Paul, R.; Balakrishnan, M.; Rao, P. Assistive technology solutions for aiding travel of pedestrians with visual impairment. *J. Rehabil. Assist. Technol. Eng.* **2017**, *4*. [[CrossRef](#)]
14. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. *J. Multimodal User Interfaces* **2017**, *11*, 149–172. [[CrossRef](#)]
15. Real, S.; Araujo, A. Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems. *Sensors* **2019**, *19*, 3404. [[CrossRef](#)]

16. Hersh, M. The design and evaluation of assistive technology products and devices Part 1: Design. In *International Encyclopedia of Rehabilitation*; (Blouin, M. and Stone, J); Center for International Rehabilitation Research Information and Exchange (CIRRIE), University at Buffalo; New York, NY, USA, 2010. Available online: <http://sphhp.buffalo.edu/rehabilitation-science/research-and-facilities/funded-research-archive/center-for-international-rehab-research-info-exchange.html> (accessed on 14 August 2020).
17. Assistive Technology. 2018. Available online: <https://www.who.int/news-room/fact-sheets/detail/assistive-technology> (accessed on 14 August 2020).
18. Lin, B.S.; Lee, C.C.; Chiang, P.Y. Simple smartphone-based guiding system for visually impaired people. *Sensors* **2017**, *17*, 1371. [[CrossRef](#)]
19. Khan, I.; Khusro, S.; Ullah, I. Technology-assisted white cane: evaluation and future directions. *PeerJ* **2018**, *6*, e6058. [[CrossRef](#)]
20. Manduchi, R.; Coughlan, J. (Computer) vision without sight. *Commun. ACM* **2012**, *55*, 96–104. [[CrossRef](#)]
21. Ton, C.; Omar, A.; Szedenko, V.; Tran, V.H.; Aftab, A.; Perla, F.; Bernstein, M.J.; Yang, Y. LIDAR Assist spatial sensing for the visually impaired and performance analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 1727–1734. [[CrossRef](#)] [[PubMed](#)]
22. Croce, D.; Giarré, L.; Pascucci, F.; Tinnirello, I.; Galioto, G.E.; Garlisi, D.; Valvo, A.L. An indoor and outdoor navigation system for visually impaired people. *IEEE Access* **2019**, *7*, 170406–170418. [[CrossRef](#)]
23. Galioto, G.; Tinnirello, I.; Croce, D.; Inderst, F.; Pascucci, F.; Giarré, L. Sensor fusion localization and navigation for visually impaired people. In Proceedings of the 2018 European Control Conference (ECC), Limassol, Cyprus, 12–15 June 2018; pp. 3191–3196.
24. Kuriakose, B.; Shrestha, R.; Sandnes, F.E. Smartphone Navigation Support for Blind and Visually Impaired People-A Comprehensive Analysis of Potentials and Opportunities. In *International Conference on Human-Computer Interaction*; Springer: Berlin, Germany, 2020; pp. 568–583.
25. Bernsen, N.O.; Dybkjær, L. *Modalities and Devices*; Springer: London, UK, 2010; pp. 67–111.
26. What Is Multimodality. 2013. Available online: <https://www.igi-global.com/dictionary/new-telerehabilitation-services-elderly/19644> (accessed on 14 August 2020).
27. Mittal, S.; Mittal, A. Versatile question answering systems: seeing in synthesis. *Int. J. Intell. Inf. Database Syst.* **2011**, *5*, 119–142. [[CrossRef](#)]
28. Jaimes, A.; Sebe, N. Multimodal human–computer interaction: A survey. *Comput. Vis. Image Understand.* **2007**, *108*, 116–134. [[CrossRef](#)]
29. Bourguet, M.L. *Designing and Prototyping Multimodal Commands*; IOS Press: Amsterdam, The Netherlands, 2003; Volume 3, pp. 717–720.
30. Bourbakis, N.; Keefer, R.; Dakopoulos, D.; Esposito, A. A multimodal interaction scheme between a blind user and the tyflos assistive prototype. In Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence, Dayton, OH, USA, 3–5 November 2008; Volume 2, pp. 487–494.
31. Basori, A.H. HapAR: Handy Intelligent Multimodal Haptic and Audio-Based Mobile AR Navigation for the Visually Impaired. In *Technological Trends in Improved Mobility of the Visually Impaired*; Springer: Cham, Switzerland, 2020; pp. 319–334.
32. Fusiello, A.; Panuccio, A.; Murino, V.; Fontana, F.; Rocchesso, D. A multimodal electronic travel aid device. In Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces, Pittsburgh, PA, USA, 16 October 2002; pp. 39–44.
33. Nair, V.; Budhai, M.; Olmschenk, G.; Seiple, W.H.; Zhu, Z. ASSIST: personalized indoor navigation via multimodal sensors and high-level semantic information. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer International Publishing: Cham, Switzerland, 2018; pp. 128–143.
34. Caraiman, S.; Morar, A.; Owczarek, M.; Burlacu, A.; Rzeszotarski, D.; Botezatu, N.; Herghelegiu, P.; Moldoveanu, F.; Strumillo, P.; Moldoveanu, A. Computer vision for the visually impaired: The sound of vision system. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1480–1489.
35. Kuriakose, B.; Shrestha, R.; Sandnes, F.E. Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. *IETE Tech. Rev.* **2020**, 1–16. [[CrossRef](#)]
36. Karray, F.; Alemzadeh, M.; Saleh, J.A.; Arab, M.N. Human-computer interaction: Overview on state of the art. *Int. J. Smart Sens. Intell. Syst.* **2008**, *1*, 137–159. [[CrossRef](#)]

37. Oviatt, S.; Lunsford, R.; Coulston, R. Individual differences in multimodal integration patterns: What are they and why do they exist? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; pp. 241–249.
38. Bohus, D.; Horvitz, E. Facilitating multiparty dialog with gaze, gesture, and speech. In Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, Beijing, China, 2–12 November 2010; pp. 1–8.
39. Oviatt, S.; Cohen, P.; Wu, L.; Duncan, L.; Suhm, B.; Bers, J.; Holzman, T.; Winograd, T.; Landay, J.; Larson, J.; et al. Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. *Hum.-Comput. Interact.* **2000**, *15*, 263–322. [[CrossRef](#)]
40. Huang, D.S.; Jo, K.H.; Figueroa-García, J.C. Intelligent Computing Theories and Application. In Proceedings of the 13th International Conference, ICIC 2017, Liverpool, UK, 7–10 August 2017; Springer: Berlin, Germany, 2017; Volume 10362.
41. Palanque, P.; Graham, T.C.N. Interactive Systems. Design, Specification, and Verification. In Proceedings of the 7th International Workshop, DSV-IS 2000, Limerick, Ireland, 5–6 June 2000; Revised Papers, Number 1946; Springer Science & Business Media: Berlin, Germany, 2001.
42. Bernsen, N.O. Multimodality theory. In *Multimodal User Interfaces*; Springer: Berlin, Germany, 2008; pp. 5–29.
43. Jacko, J.A. *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*; CRC Press: Boca Raton, FL, USA, 2012.
44. Kurosu, M. Human-Computer Interaction: Interaction Modalities and Techniques. In Proceedings of the 15th International Conference, HCI International 2013, Las Vegas, NV, USA, 12–26 July 2013; Springer: Berlin, Germany, 2013; Volume 8007.
45. Bainbridge, W.S. *Berkshire Encyclopedia of Human-Computer Interaction*; Berkshire Publishing Group LLC: Great Barrington, MA, USA, 2004; Volume 1.
46. Epstein, Z. Siri Said to Be Driving Force behind Huge iPhone 4S Sales. 2011. Available online: <https://bgr.com/2011/11/02/siri-said-to-be-driving-force-behind-huge-iphone-4s-sales/> (accessed on 14 August 2020).
47. Grifoni, P.; Ferri, F.; Caschera, M.C.; D’Ulizia, A.; Mazzei, M. MIS: Multimodal Interaction Services in a cloud perspective. *arXiv* **2014**, arXiv:1704.00972.
48. Hoy, M.B. Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Med. Ref. Serv. Q.* **2018**, *37*, 81–88. [[CrossRef](#)]
49. Kepuska, V.; Bohouta, G. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 99–103.
50. Kurkovsky, S. *Multimodality in Mobile Computing and Mobile Devices: Methods for Adaptable Usability*; IGI Global: Hershey, PA, USA, 2010.
51. Djaid, N.T.; Saadia, N.; Ramdane-Cherif, A. Multimodal Fusion Engine for an Intelligent Assistance Robot Using Ontology. *Procedia Comput. Sci.* **2015**, *52*, 129–136. [[CrossRef](#)]
52. Corradini, A.; Mehta, M.; Bernsen, N.O.; Martin, J.; Abrilian, S. Multimodal input fusion in human-computer interaction. *NATO Sci. Ser. Sub Ser. III Comput. Syst. Sci.* **2005**, *198*, 223.
53. D’Ulizia, A. Exploring multimodal input fusion strategies. In *Multimodal Human Computer Interaction and Pervasive Services*; IGI Global: Hershey, PA, USA, 2009; pp. 34–57.
54. Caschera, M.C.; Ferri, F.; Grifoni, P. Multimodal interaction systems: Information and time features. *Int. J. Web Grid Serv.* **2007**, *3*, 82–99. [[CrossRef](#)]
55. Grifoni, P. *Multimodal Human Computer Interaction and Pervasive Services*; IGI Global: Hershey, PA, USA, 2009.
56. Dumas, B.; Lalanne, D.; Oviatt, S. Multimodal interfaces: A survey of principles, models and frameworks. In *Human Machine Interaction*; Springer: Berlin, Germany, 2009; pp. 3–26.
57. Vainio, T. Exploring cues and rhythm for designing multimodal tools to support mobile users in wayfinding. In *CHI’09 Extended Abstracts on Human Factors in Computing Systems*; ACM: New York, NY, USA, 2009; pp. 3715–3720.
58. Brock, A.M.; Truillet, P.; Oriola, B.; Picard, D.; Jouffrais, C. Interactivity improves usability of geographic maps for visually impaired people. *Hum.-Comput. Interact.* **2015**, *30*, 156–194. [[CrossRef](#)]
59. Paternó, F. Interactive Systems: Design, Specification, and Verification. In Proceedings of the 1st Eurographics Workshop, Bocca Di Magra, Italy, 8–10 June 1994; Springer Science & Business Media: Berlin, Germany, 2012.

60. Sears, A.; Jacko, J.A. *Human-Computer Interaction: Designing for Diverse Users and Domains*; CRC Press: Boca Raton, FL, USA, 2009.
61. van der Bie, J.; Ben Allouch, S.; Jaschinski, C. Communicating Multimodal Wayfinding Messages For Visually Impaired People Via Wearables. In Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, Taipei Taiwan, 1–4 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–7.
62. Tang, M. Benefits of Bone Conduction and Bone Conduction Headphones. 2019. Available online: <https://www.soundguys.com/bone-conduction-headphones-20580/> (accessed on 14 August 2020).
63. Gallo, S.; Chapuis, D.; Santos-Carreras, L.; Kim, Y.; Retornaz, P.; Bleuler, H.; Gassert, R. Augmented white cane with multimodal haptic feedback. In Proceedings of the 2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, Tokyo, Japan, 26–29 September 2010; pp. 149–155.
64. Zeng, L.; Weber, G.; Simros, M.; Conradie, P.; Saldien, J.; Ravyse, I.; van Erp, J.; Mioch, T. Range-IT: detection and multimodal presentation of indoor objects for visually impaired people. In Proceedings of the MobileHCI '17: 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria, 4–7 September 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1–6.
65. Hosseini, S.M.F.; Riener, A.; Bose, R.; Jeon, M. “Listen2dRoom”: Helping Visually Impaired People Navigate Indoor Environments Using an Ultrasonic Sensor-Based Orientation Aid; Georgia Institute of Technology: Atlanta, GA, USA, 2014.
66. Ahmetovic, D.; Gleason, C.; Ruan, C.; Kitani, K.; Takagi, H.; Asakawa, C. NavCog: A navigational cognitive assistant for the blind. In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, Florence, Italy, 6–9 September 2016; pp. 90–99.
67. Chang, Y.; Chen, J.; Franklin, T.; Zhang, L.; Ruci, A.; Tang, H.; Zhu, Z. Multimodal Information Integration for Indoor Navigation Using a Smartphone. In Proceedings of the 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), Las Vegas, NV, USA, 11–13 August 2020; pp. 59–66.
68. Loomis, J.M.; Golledge, R.G.; Klatzky, R.L. Navigation system for the blind: Auditory display modes and guidance. *Presence* **1998**, *7*, 193–203. [[CrossRef](#)]
69. Loomis, J.M.; Marston, J.R.; Golledge, R.G.; Klatzky, R.L. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *J. Vis. Impair. Blind.* **2005**, *99*, 219–232. [[CrossRef](#)]
70. Wang, H.C.; Katzschmann, R.K.; Teng, S.; Araki, B.; Giarré, L.; Rus, D. Enabling independent navigation for visually impaired people through a wearable vision-based feedback system. In Proceedings of the 2017 IEEE international conference on robotics and automation (ICRA), Singapore, 29 May–3 June 2017; pp. 6533–6540.
71. Diaz, C.; Payandeh, S. Multimodal Sensing Interface for Haptic Interaction. *J. Sens.* **2017**, *2017*. [[CrossRef](#)]
72. Lock, J.C.; Cielniak, G.; Bellotto, N. *A Portable Navigation System with an Adaptive Multimodal Interface for the Blind*; 2017 AAAI Spring Symposium Series; Stanford, CA, USA, 27–29 March 2017; AAAI: Palo Alto, CA, USA, 2017.
73. Bellotto, N. A multimodal smartphone interface for active perception by visually impaired. In *IEEE SMC International Workshop on Human-Machine Systems, Cyborgs and Enhancing Devices (HUMASCEND)*; IEEE: Manchester, UK, 2013.
74. Turunen, M.; Hakulinen, J.; Kainulainen, A.; Melto, A.; Hurtig, T. Design of a rich multimodal interface for mobile spoken route guidance. In Proceedings of the Eighth Annual Conference of the International Speech Communication Association, Antwerp, Belgium, 27–31 August 2007.
75. Ducasse, J.; Brock, A.M.; Jouffrais, C. Accessible interactive maps for visually impaired users. In *Mobility of Visually Impaired People*; Springer: Berlin, Germany, 2018; pp. 537–584.
76. Brock, A.; Truillet, P.; Oriola, B.; Picard, D.; Jouffrais, C. Design and user satisfaction of interactive maps for visually impaired people. In *International Conference on Computers for Handicapped Persons*; Springer: Berlin, Germany, 2012; pp. 544–551.
77. Wang, Z.; Li, B.; Hedgpeth, T.; Haven, T. Instant tactile-audio map: enabling access to digital maps for people with visual impairment. In Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, Pittsburgh, PA, USA, 25–28 October 2009; pp. 43–50.
78. Miele, J.A.; Landau, S.; Gilden, D. Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell’s TMAP software. *Br. J. Vis. Impair.* **2006**, *24*, 93–100. [[CrossRef](#)]

79. Giudice, N.A.; Guenther, B.A.; Jensen, N.A.; Haase, K.N. Cognitive mapping without vision: Comparing wayfinding performance after learning from digital touchscreen-based multimodal maps vs. embossed tactile overlays. *Front. Hum. Neurosci.* **2020**, *14*, 87. [[CrossRef](#)]
80. Poppinga, B.; Magnusson, C.; Pielot, M.; Rasmus-Gröhn, K. TouchOver map: Audio-tactile exploration of interactive maps. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, Stockholm, Sweden, 30 August–2 September 2011; pp. 545–550.
81. Zeng, L.; Weber, G. Exploration of location-aware you-are-here maps on a pin-matrix display. *IEEE Trans. Hum.-Mach. Syst.* **2015**, *46*, 88–100. [[CrossRef](#)]
82. Bahram, S. Multimodal eyes-free exploration of maps: TIKISI for maps. *ACM SIGACCESS Access. Comput.* **2013**, 3–11. [[CrossRef](#)]
83. Yatani, K.; Banovic, N.; Truong, K. SpaceSense: Representing geographical information to visually impaired people using spatial tactile feedback. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012, pp. 415–424.
84. Kunz, A.; Miesenberger, K.; Zeng, L.; Weber, G. Virtual navigation environment for blind and low vision people. In *International Conference on Computers Helping People with Special Needs*; Springer: Berlin, Germany, 2018; pp. 114–122.
85. Lécuyer, A.; Mobuchon, P.; Mégard, C.; Perret, J.; Andriot, C.; Colinot, J.P. HOMERE: A multimodal system for visually impaired people to explore virtual environments. In Proceedings of the IEEE Virtual Reality, Los Angeles, CA, USA, 22–26 March 2003; pp. 251–258.
86. Rivière, M.A.; Gay, S.; Romeo, K.; Pissaloux, E.; Bujacz, M.; Skulimowski, P.; Strumillo, P. NAV-VIR: An audio-tactile virtual environment to assist visually impaired people. In Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, 20–23 March 2019; pp. 1038–1041.
87. Khoo, W.L.; Seidel, E.L.; Zhu, Z. Designing a virtual environment to evaluate multimodal sensors for assisting the visually impaired. In *International Conference on Computers for Handicapped Persons*; Springer: Berlin, Germany, 2012; pp. 573–580.
88. Zhao, Y.; Bennett, C.L.; Benko, H.; Cutrell, E.; Holz, C.; Morris, M.R.; Sinclair, M. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–14.
89. Lahav, O.; Schloerb, D.; Kumar, S.; Srinivasan, M. A virtual environment for people who are blind—a usability study. *J. Assist. Technol.* **2012**, *6*. [[CrossRef](#)] [[PubMed](#)]
90. Ho, C.; Reed, N.; Spence, C. Multisensory in-car warning signals for collision avoidance. *Hum. Factors* **2007**, *49*, 1107–1114. [[CrossRef](#)] [[PubMed](#)]
91. Stanney, K.; Samman, S.; Reeves, L.; Hale, K.; Buff, W.; Bowers, C.; Goldiez, B.; Nicholson, D.; Lackey, S. A paradigm shift in interactive computing: Deriving multimodal design principles from behavioral and neurological foundations. *Int. J. Hum.-Comput. Interact.* **2004**, *17*, 229–257. [[CrossRef](#)]
92. Calvert, G.; Spence, C.; Stein, B.E. *The Handbook of Multisensory Processes*; MIT Press: Cambridge, MA, USA, 2004.
93. Lee, J.H.; Spence, C. Assessing the benefits of multimodal feedback on dual-task performance under demanding conditions. In Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction (HCI), Liverpool, UK, 1–5 September 2008; pp. 185–192.
94. Oviatt, S.; Schuller, B.; Cohen, P.; Sonntag, D.; Potamianos, G. *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*; Morgan & Claypool: San Rafael, CA, USA, 2017.
95. Rodrigues, J.; Cardoso, P.; Monteiro, J.; Figueiredo, M. *Handbook of Research on Human-Computer Interfaces, Developments, and Applications*; IGI Global: Hershey, PA, USA, 2016.
96. Common Sense Suggestions for Developing Multimodal User Interfaces. 2016. Available online: <https://www.w3.org/TR/mmi-suggestions/> (accessed on 14 August 2020).
97. Havik, E.M.; Kooijman, A.C.; Steyvers, F.J. The effectiveness of verbal information provided by electronic travel aids for visually impaired persons. *J. Vis. Impair. Blind.* **2011**, *105*, 624–637. [[CrossRef](#)]

98. Adebisi, A.; Sorrentino, P.; Bohlool, S.; Zhang, C.; Arditti, M.; Goodrich, G.; Weiland, J.D. Assessment of feedback modalities for wearable visual aids in blind mobility. *PLoS ONE* **2017**, *12*, e0170531. [[CrossRef](#)] [[PubMed](#)]
99. Jacob, S.V.; MacKenzie, I.S. Comparison of Feedback Modes for the Visually Impaired: Vibration vs. Audio. In *International Conference on Universal Access in Human-Computer Interaction*; Springer: Berlin, Germany, 2018; pp. 420–432.
100. Wentzel, J.; Velleman, E.; van der Geest, T. Developing accessibility design guidelines for wearables: Accessibility standards for multimodal wearable devices. In *International Conference on Universal Access in Human-Computer Interaction*; Springer International Publishing: Cham, Switzerland, 2016; pp. 109–119.
101. Schneider-Hufschmidt, M. Human factors (hf): Multimodal interaction, communication and navigation guidelines. In *Proceedings of the 19th International Symposium on Human Factors in Telecommunication*, Berlin/Heidelberg, Germany, 1–4 December 2003; European Telecommunications Standards Institute: Sophia Antipolis, France, 2003; Volume 1, pp. 1–53.
102. Lahat, D.; Adali, T.; Jutten, C. Multimodal data fusion: An overview of methods, challenges, and prospects. *Proc. IEEE* **2015**, *103*, 1449–1477. [[CrossRef](#)]
103. Gjoreski, H.; Ciliberto, M.; Wang, L.; Ordonez Morales, F.J.; Mekki, S.; Valentin, S.; Roggen, D. The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics With Mobile Devices. *IEEE Access* **2018**, *6*, 42592–42604. [[CrossRef](#)]
104. Rouat, S.B.S.C.J. CREATE: Multimodal Dataset for Unsupervised Learning and Generative Modeling of Sensory Data from a Mobile Robot. *IEEE Dataport* **2018**. [[CrossRef](#)]
105. Cheng, R.; Wang, K.; Bai, J.; Xu, Z. OpenMPR: Recognize places using multimodal data for people with visual impairments. *Meas. Sci. Technol.* **2019**, *30*, 124004. [[CrossRef](#)]
106. Caspo, A.; Wersényi, G.; Jeon, M. A survey on hardware and software solutions for multimodal wearable assistive devices targeting the visually impaired. *Acta Polytech. Hung.* **2016**, *13*, 39.
107. Liljedahl, M.; Lindberg, S.; Delsing, K.; Polojärvi, M.; Saloranta, T.; Alakärppä, I. Testing two tools for multimodal navigation. *Adv. Hum.-Comput. Interact.* **2012**, *2012*. [[CrossRef](#)]
108. Gallacher, S.; Papadopoulou, E.; Taylor, N.K.; Williams, M.H. Learning user preferences for adaptive pervasive environments: An incremental and temporal approach. *ACM Trans. Auton. Adapt. Syst. TAAS* **2013**, *8*, 1–26. [[CrossRef](#)]
109. Yao, Y.; Zhao, Y.; Wang, J.; Han, S. A model of machine learning based on user preference of attributes. In *International Conference on Rough Sets and Current Trends in Computing*; Springer: Berlin, Germany, 2006; pp. 587–596.
110. Phillips, B.; Zhao, H. Predictors of assistive technology abandonment. *Assist. Technol.* **1993**, *5*, 36–45. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).