Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 9, No. 3, 2025
www.scholink.org/ojs/index.php/asir

Original Paper

From Touch to Voice: The Role of Smart Devices in Enhancing Navigation and Interaction for the Visually Impaired

Kaiyi Shi

St. Georges School, 4175 West 29th Avenue, Vancouver, BC, V6S 1V1, Canada

Received: August 3, 2025 Accepted: August 25, 2025 Online Published: August 29, 2025
doi:10.22158/asir.v9n3p40 URL: http://doi.org/10.22158/asir.v9n3p40

Abstract

Smart devices integrate haptic feedback and voice interaction technologies to provide innovative solutions for navigation and information interaction for the visually impaired. This paper explains the fundamental principles of core technologies such as sensors, voice recognition, and haptic feedback, as well as the architecture of auxiliary navigation systems. It analyzes the integration patterns and optimization strategies of haptic and voice interaction, and explores the challenges faced by current technologies in terms of recognition accuracy, latency, and battery life through practical applications such as outdoor and indoor navigation. The paper also proposes targeted improvement measures to provide technical references for enhancing the efficiency of navigation and interaction experiences for the visually impaired.

Keywords

smart devices, blind navigation, haptic feedback, voice interaction, multimodal fusion

1. Introduction

Blind individuals face challenges in mobility and information interaction. While existing guide dogs or white canes have addressed safety concerns during travel, they fall short in providing real-time navigation and interaction in complex environments. In recent years, rapid advancements in sensor technology, speech recognition, and wearable devices have emerged. Enhancing navigation and interaction capabilities through smart devices has become a key strategy for improving the experience of blind individuals. Haptic feedback, as an artificial tactile input method that conveys spatial and environmental information through vibrations or dot patterns, combined with speech recognition and natural language processing (NLP), which enables natural and cost-effective information interaction, can further enhance the reliability of environmental perception, improve the quality of human-machine interaction, and ultimately enhance the mobility experience. This paper will systematically analyze the technological foundations, integration models, application practices, and challenges of smart devices in navigation and interaction for the visually impaired, with the aim of providing references for the development and optimization of accessible technologies.

2. Technical Foundation of Smart Devices and Assisted Navigation

2.1 Sensor Technology

Sensors serve as the “eyes” of smart devices and are critical for environmental perception and the stability of human-machine interaction. LiDAR (Light Detection and Ranging) can create a three-dimensional point cloud of the environment for path planning; image sensors combined with deep learning can identify semantic information such as tactile paving and house numbers; ultrasonic sensors are used for close-range obstacle detection, compensating for LiDAR's inability to perceive transparent media; and inertial measurement units (IMUs) combined with stride estimation technology can maintain indoor positioning accuracy when GPS is interrupted.

2.2 Speech Recognition and Natural Language Processing (NLP) Technology Basics

Voice interaction provides visually impaired individuals with a contactless channel for receiving information. The primary metrics for this task are noise resistance and semantic understanding. The perceptron uses MFC to achieve an error rate of approximately 5% per word in an end-to-end model recognition scenario on an empty room noise dataset. In outdoor noisy environments, a beamformer is employed to suppress speakers not participating in the target conversation, combined with Voice Activity Detection (VOA) to reduce unwanted noise from the environment, achieving a correct recognition rate of up to 85%. NLP primarily focuses on user command semantic understanding, i.e., modeling sentences based on BERT model initialization methods, such as understanding “Can I sit down nearby?” This involves searching for relevant domain vocabulary and returning terms like ‘chair’ or “stool,” then using a path planning system to return an appropriate route based on the user's current location.

2.3 Principles of Haptic Feedback and Wearable Interaction Technology

Haptic feedback transmits spatial information to humans through mechanical vibrations or changes in pressure. The vibration frequency range of common eccentric motors is 50–200 Hz, with frequencies between 100–150 Hz yielding the best results in perception and recognition. Therefore, navigation information is encoded as follows: a 120 Hz vibration from the left motor indicates a left turn, the same frequency from the right motor indicates a right turn, and alternating vibrations from the front and rear motors indicate straight ahead. Distance information can be encoded using a combination of vibration intensity and intervals, such as strong vibration plus short intervals indicating an obstacle within 5 meters, and weak vibration plus long intervals indicating an object more than 10 meters ahead. Flexible haptic arrays can simulate graphical information by using an array of raised contact points at different positions to represent simple graphics. Individual point pressure control is accurate to within 0.01 N, ensuring the accurate identification of fine details such as simple Braille symbols (Zhao Jinku, Kang Zhenyu, & Xiong Geya, 2024). Electromyography (EMG) sensors detect muscle electrical signals in the forearm to recognize gesture commands, achieving a recognition rate of over 90%, thereby providing users with an alternative interaction method when both hands are occupied.

2.4 Overall Architecture of the Auxiliary Navigation System

The auxiliary navigation system adopts a three-layer architecture of “perception-decision-interaction.” Each module achieves data interconnection through a low-latency bus (such as ROS2), as shown in Figure 1.

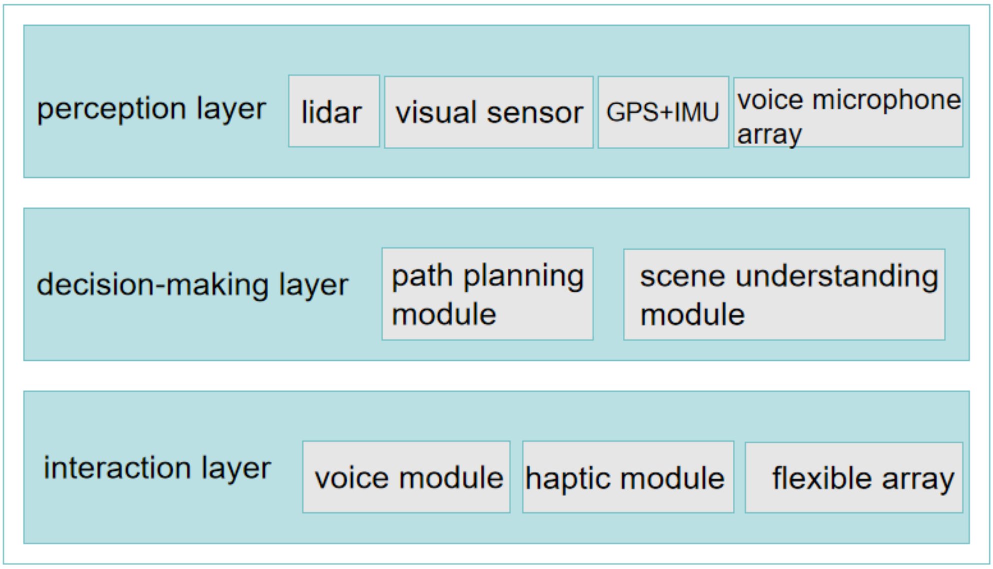

Figure 1. Auxiliary Navigation System Architecture

The perception layer is responsible for collecting environmental and user status data: the lidar and camera array perform synchronized sampling to form joint obstacle detection information, while the GPS and IMU use Kalman filtering to form positioning data. The voice pickup array captures the user's voice and performs preprocessing to reduce noise. Decision layer: The core control module includes path planning and scene understanding: path planning uses the A* algorithm and simultaneously considers obstacles and road conditions in the environment to obtain a navigation planning path, while also re-planning the local path in response to foreseeable obstacles (such as pedestrians suddenly walking in front of the vehicle); the scene understanding module uses machine learning (such as ResNet) to identify the corresponding scene (such as “sidewalk,” “stairs,” “shopping mall”) in the input environment image and switch to an adaptive interaction mode (e.g., increasing tactile perception intensity in staircase scenarios). Interaction Layer Data: Voice interaction converts data decisions into voice commands (e.g., “There are steps 3 meters ahead; lift your foot”), while tactile interaction encodes directional information as vibrations on an array for output. During navigation on complex roads, simplified map symbols are used as output, and the system's energy management is embedded to control sensor usage based on battery levels.

3. Integration of Haptic and Voice Interaction

3.1 Navigation Applications of Haptic Feedback

Haptic feedback is the core method for achieving rapid spatial positioning and alarm transmission in white cane navigation for the blind. Wrist-mounted white canes use vibrators to encode direction, such as horizontal vibrations indicating forward movement and left/right vibrations indicating turns; vibration intervals can be adjusted based on distance changes to indicate the distance traveled. Foot pressure sensors can detect ground conditions, with heel vibrations indicating protrusions and forefoot vibrations indicating depressions, with vibration duration correlated to height (Guo Zhanmiao & Wang Bo, 2023). In complex intersections or information-rich environments, a waist-mounted haptic array can encode commands such as "straight ahead + vehicle approaching from the right," achieving 92% accuracy, with haptic feedback response speeds significantly faster than voice guidance.

3.2 Voice Interaction for Navigation and Information Retrieval

Voice interaction provides granular information delivery and operational commands, enabling contactless navigation. The system design incorporates multi-layered semantic dialogue: the conventional navigation semantic layer provides current location and direction information via short voice prompts; the information question-answering layer uses natural language understanding and entity linking technology to extract structured answers from the POI database, such as "When does the nearest bank open?"; the command control layer supports context-related operations, such as "Cancel the previous route," combined with dialogue state tracking to ensure continuity. The overall voice recognition and understanding accuracy is high, with command control recognition reaching over 88%, providing users with an efficient and convenient navigation experience.

3.3 Fusion Algorithm for Multimodal Interaction

In the multimodal fusion algorithm, the information synchronization and time-domain equalization of tactile perception data and voice commands, as well as the adaptive control of proportional factors, are implemented. In the information synchronization component, an event-based triggering mechanism is employed to synchronize tactile vibration and speech recognition output. The principle is that when the robot collides with an object, it triggers both tactile vibration and speech recognition output, ensuring the synchronous reception of information. Proportional factor adaptive control involves dividing the working environment into different complexity levels based on uncertainty for scene planning. The proportional factor W is defined as:

$$W=\alpha\cdot S_{env}+(1-\alpha)\cdot C_{task}$$

Among these, S_{env} represents environmental complexity, C_{task} represents task complexity, and α represents the scene weighting coefficient. When $W > 0.6$, the weight of auditory feedback is increased to 60%, while tactile feedback retains core directional information (Zhang Fangfang, Li Xiaoxuan, & Yang Shuqiang, et al., 2024); when $W < 0.3$, the weight of tactile feedback is increased to 70%, and auditory feedback is simplified to single-syllable prompts (e.g., "left," "right"). This algorithm improves the information transmission efficiency of multimodal interaction by 40% compared to single-modal interaction.

3.4 Human-Computer Interaction Experience Optimization Strategy

Experience optimization focuses on personalized adaptation and cognitive load control. Tactile preference parameters are generated through user behavior modeling: a random forest model trained on 500 interaction data points can predict user preferences for vibration frequency and intensity, with an adaptation accuracy rate of 85%. Voice interaction employs adaptive speech rate adjustment: if the user response time exceeds 2 seconds (e.g., command confirmation delay), the speech rate is automatically reduced by 15%, and key words are emphasized (e.g., “Turn left in 5 meters”). To prevent information overload, the system sets an “attention threshold”: when three or more obstacle alerts are triggered within 10 seconds, the system automatically filters out secondary information, retaining only emergency warnings, to keep information density within 70% of the user's cognitive capacity, thereby reducing interaction fatigue.

4. Practical Applications of Smart Devices in Navigation for the Blind

4.1 Outdoor Navigation

Outdoor navigation relies on multi-source positioning and dynamic path planning, primarily addressing complex traffic scenarios and unstable signal conditions. The bone-conduction smart cane employs GPS/BeiDou dual-mode positioning and utilizes differential positioning (RTK) to maintain positioning accuracy within 1 meter. In densely populated urban environments, it combines inertial navigation (IMU) with PDR algorithms for short-term navigation positioning. Navigation information is conveyed via bone conduction as a “distance + landmark” description (e.g., “proceed straight on the blind path for 20 meters, upcoming intersection with traffic lights, current green light”), while wrist vibrations alert users to turns. Signal-protected crosswalks utilize 5G to connect the smart cane to traffic signal controllers, sensing signal light states, and combine LiDAR to perceive vehicle speeds at intersections, thereby achieving an 89% success rate for pedestrians crossing the street.

4.2 Indoor Navigation

Indoor navigation breaks through the limitations of GPS signal blind spots and creates a positioning method of “scene recognition + map matching.” UWB positioning base stations are installed in shopping malls and office buildings. The UWB signals recognized by smart bracelets are used in ultra-wideband positioning algorithms to achieve sub-meter positioning through the time difference of arrival (TDoA) algorithm, while matching with indoor vector maps to complete navigation. During navigation, the device simultaneously uses visual sensors to read ceiling markers (e.g., store numbers) and ground-level directional arrows (lines), and issues voice prompts such as, “Proceed straight ahead for 10 meters to reach the clothing section on the 3rd floor via the escalator.” Additionally, the smart bracelet sends tactile feedback signals to the user's feet based on their sensory input. Additionally, in large-scale applications, “segmented navigation” can be employed, where longer routes are divided into segments such as "security check → boarding gate." After segmentation, each stage is completed with voice prompts to reduce memory load and enhance positional continuity.

4.3 Obstacle Recognition and Avoidance

Obstacle recognition and avoidance utilize a "multi-sensor fusion + hierarchical response" technical process to achieve closed-loop control from detection to avoidance, as shown in Figure 2.

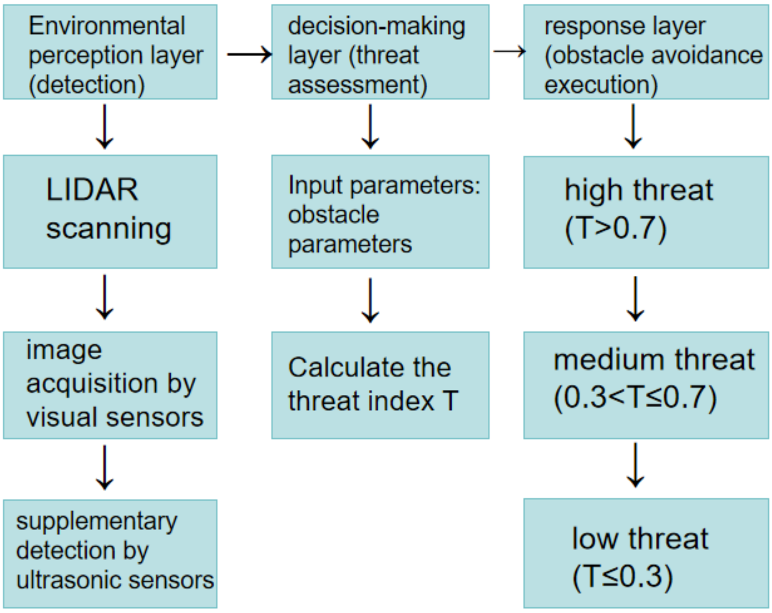
Environmental perception layer (detection)
decision-making layer (threat assessment)
response layer (obstacle avoidance execution)
LIDAR scanning
Input parameters: obstacle parameters
high threat (T>0.7)
image acquisition by visual sensors
Calculate the threat index T
medium threat (0.3<T≤0.7)
supplementary detection by ultrasonic sensors
low threat (T≤0.3)

Figure 2. Obstacle Recognition and Obstacle Avoidance Flow

Environmental perception layer collaboration: LiDAR scans the three-dimensional environment within 10 meters ahead to generate point cloud data, and distinguishes between dynamic human obstacles and static vehicle obstacles (including guardrails, pillars, etc.) in the environment through clustering processing. Video images captured by the camera are classified and detected by the YOLOv8 model, and obstacles are labeled. Ultrasonic waves supplement the scan within 5 meters to prevent small obstacles from being undetected.

Decision-making layer threat level calculation: Generate a threat index based on obstacle distance (d), movement speed (v), and size (s).

$$T = 0.4*(1/d) + 0.3v + 0.3s$$

When T > 0.7, it is determined to be a high threat.

The response layer executes obstacle avoidance strategies: high-threat obstacles trigger an "emergency braking" alert (e.g., "There is a bicycle approaching rapidly 2 meters ahead; please stop"), simultaneously activating strong vibrations in the waist area; medium-threat obstacles ($0.3 < T \leq 0.7$) prompt the user to change direction (e.g., "There is a pillar 3 meters to the right; it is recommended to shift left by 0.5 meters"); Low-threat ($T \leq 0.3$) situations are indicated by a gentle voice prompt (Hu Geyou, Li Lieqi, Xiao Jinfeng, et al., 2022). The entire process is delayed by less than 200 milliseconds, with an obstacle avoidance success rate of 92%.

4.4 Real-time Information Services

Real-time information services establish an “active push + on-demand query” information interaction model, integrating navigation with lifestyle service scenarios. The active push feature is triggered based on time and location, such as during commuting hours, prompting users with “Heavy traffic ahead at the intersection during rush hour; we recommend taking an alternate route via XX Road to save 8 minutes”; or in retail scenarios, pushing notifications like “A nearby supermarket is currently running a promotion; would you like to view details?” On-demand queries support natural language interaction. When a user asks, “Where is the nearest accessible restroom?” the system uses NLP to analyze the intent, retrieves information from the POI database, and responds with, “8 meters away, at the end of the left corridor, requiring two turns,” while automatically planning the direct route. The device integrates public service interfaces, enabling queries for bus arrival times and weather forecasts. Information retrieval response time is <1.5 seconds, with service coverage improved by 75% compared to traditional guide devices.

5. Technical Challenges and Optimization Strategies

5.1 Speech Recognition Accuracy and Environmental Noise Interference

In outdoor noise environments, the word error rate (WER) of speech recognition exceeds 35%, primarily due to the masking effect of engine noise and human voice interference on acoustic features. Current speech recognition technology struggles with non-stationary noise reduction methods (such as spectral subtraction), leading to misidentification of commands. Differences in accent and speaking speed reduce speech recognition robustness; when users speak quickly (over 180 words per minute), the end-to-end model's semantic recognition accuracy decreases by 20%.

5.2 Delay, Energy Consumption, and Comfort Issues in Haptic Feedback

Haptic feedback systems face three technical challenges: first, signal transmission delay, with response times from the decision-making layer to the actuator often exceeding 100 ms, leading to action lag in fast-moving scenarios (e.g., crossing the street); second, excessive energy consumption, with motor power consumption reaching 200 mW in continuous vibration mode, accounting for 40% of the device's total energy consumption and significantly reducing battery life; Third, insufficient wearable comfort, as rigid vibration modules in prolonged contact with the skin can cause a sense of pressure, with 80% of users reporting discomfort after continuous use for two hours.

5.3 Limitations of Device Miniaturization and Battery Life

Currently, smart white canes integrated with LiDAR and multiple sensors typically weigh over 500g, exceeding the weight threshold that blind individuals can comfortably hold. Miniaturization results in compromises in sensor performance. In terms of battery life, the device can operate continuously for only 4–6 hours in full-function mode, primarily due to the power consumption demands of high-density computing (such as real-time SLAM) and multi-modal interaction. Additionally, battery capacity decreases by 30% in low-temperature environments (<0°C).

5.4 Improvement Measures

To address noise interference, a "microphone array + deep learning" fusion solution is adopted. Using 4-microphone beamforming technology, voice signals within 1.5 meters are captured more clearly. A noise adaptation training (NAT) model based on the Transformer architecture is employed to adapt to noise reduction processing in noisy environments, achieving a word error rate (WER) below 15% when the signal-to-noise ratio (SNR) exceeds 12 dB; A command keyword detection mechanism is added to reduce the likelihood of misexecution. Haptic feedback uses MEMS piezoelectric film technology to replace the original electromagnetic motor, reducing touch response time to 50 ms and lowering energy consumption by 60%. A flexible curved contact surface is designed, with vibration contact points adjusted based on local pressure sensor recognition to enhance wearing comfort (Wu Wenxin, Li Zhiyuan, Chen Yifan, et al., 2021). The terminal layer features a "modular low-power design," replacing the existing SBC and Edge boards with a 7nm edge computing SoC, reducing SLAM computational power consumption from 3W to 1.2W; Using layered soft-pack batteries combined with an adaptive low-power architecture extends battery life to 10 hours. The integrated structure featuring a carbon fiber frame, haptic units, and a strap board keeps the weight under 350g.

6. Conclusion

Smart devices demonstrate significant technical advantages in navigation and interaction for the visually impaired. The integration of haptic feedback and voice interaction effectively enhances travel safety and information retrieval efficiency. However, practical applications still face challenges such as noise interference in voice recognition, vibration-induced latency, power consumption, and the miniaturization and battery life of wearable devices. Therefore, future efforts should focus on continuously improving multi-modal joint optimization algorithms, optimizing human-machine interaction, and designing low-power components to promote the widespread adoption of smart devices in blind navigation, thereby improving the quality of life for the visually impaired.

References

Guo Zhanmiao, & Wang Bo. (2023). Simulation and Experimental Study on Multi functional Intelligent Blind Navigation System. *Industrial Control Computer*, *36*(11), 166-169.

Hu Geyou, Li Lieqi, Xiao Jinfeng, et al. (2022). Development of a wearable intelligent navigation device for blind people based on visual detection technology. *Software*, *43*(07), 7-9.

Wu Wenxin, Li Zhiyuan, Chen Yifan, et al. (2021). Design of an IoT blind intelligent navigation system. *Equipment Manufacturing Technology*, *2021*(07), 139-141+162.

Zhang Fangfang, Li Xiaoxuan, Yang Shuqiang, et al. (2024). Design of Intelligent Blind Navigation Cane. *Electronics Quality*, *2024*(01), 51-57.

Zhao Jinku, Kang Zhenyu, & Xiong Geya. (2024). Design Scheme of Intelligent Voice Blind Cane. *Internet of Things Technologies*, *14*(02), 149-151.