IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 26, 2020, accepted September 30, 2020. Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.3029426

Sight-to-Sound Human-Machine Interface for Guiding and Navigating Visually Impaired People

GUOJUN YANG, (Member, IEEE), AND JAFAR SANIIE, (Life Fellow, IEEE)
Illinois Institute of Technology, Chicago, IL 60616, USA
Corresponding author: Jafar Saniie (saniie@iit.edu)

ABSTRACT Visually impaired people often find it hard to navigate efficiently in complex environments. Moreover, helping them to navigate intuitively is not a trivial task. In sighted people, cognitive maps derived from visual cues play a pivotal role in navigation. In this paper, we present a sight-to-sound human-machine interface (STS-HMI), a novel machine vision guidance system that enables visually impaired people to navigate with instantaneous and intuitive responses. The proposed STS-HMI system extracts visual context from scenes and converts them into binaural acoustic cues for users to establish cognitive maps. A series of experiments were conducted to evaluate the performance of the STS-HMI system in a complex environment with difficult navigation paths. The experimental results confirm that the STS-HMI system improves visually impaired people's mobility with minimal effort.

INDEX TERMS Acoustic cues, human-machine interface, navigation, visually impaired.

I. INTRODUCTION

People rely on cognitive maps to navigate. A cognitive map is the knowledge and understanding of an environment for navigation [1]–[3]. Over the past decade, numerous systems for assisting visually impaired people have been developed. Previous studies focused on building an assisting device or system that determines the position of the user and then generates instructions for navigation using voice prompts or haptic cues. The proposed sight-to-sound human-machine interface (STS-HMI) employs a camera to perform scene analysis and generates novel binaural acoustic cues as feedback. Compared with existing solutions, the proposed STS-HMI system has the following advantages: (i) versatile and adaptive, (ii) easy to use, (iii) rich in visual context, and (iv) simple to realize with common technology.

Indoor positioning can be achieved through triangulation [4], [5], pattern matching [6]–[8], direct sensing [9]–[12], and dead-reckoning [13]–[16]. Triangulation requires a device to measure the distance to at least three reference objects in order to determine the location of the device accurately [4]. Triangulation limits the type of indoor environments in which the method can be used since it is challenging to maintain a line-of-sight connection with reference objects at all times [4]. In contrast, wireless signal pattern matching does not require line-of-sight connections. Instead, wireless signals (i.e., Wi-Fi) are used as fingerprints to achieve localization [6]–[8]. Wireless signal pattern matching requires an intensive survey of the signal pattern within an environment before usage. Similarly, visually based pattern matching requires an intensive survey of the scenes within an environment in advance [17]–[19]. Direct sensing involves associating tags with objects or locations, which allows determining the user's location within an environment. Tags used for direct sensing can be created using RFID (radio-frequency identification) [9]–[12], infrared signals [20], [21], ultrasound identification [18], [22], Bluetooth beacons [23], and barcodes [24], [25]. Like localization with pattern matching methods, creating labels requires extensive studies of the environment. Dead-reckoning estimates a user's location by recording the user's cumulative steps [13]–[16]. Such a method requires no modification or survey of the environment. However, the measurement error of the dead-reckoning method will accumulate over time, resulting in reduced accuracy.

Visual cues are critical for navigating in complex and dynamic indoor environments [3]. Sighted people can navigate while determining their location by observing and memorizing key objects in the scenes. Furthermore, they can also locate themselves by comprehending the context of surrounding objects. For example, a unique combination of furniture can help sighted people to identify a room and determine their location in a building. Visual perception is also critical for finding a navigation path. For example, when exiting

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

a building, one may look for a revolving door. Therefore, if visually impaired people can perceive visual cues from the environment, it will be much easier for them to navigate.

There have been many attempts to help visually impaired people to gain awareness of their surroundings. Existing systems assisting visually impaired people are often vision substitution systems. These systems use sensors or cameras to convert scenes into nonvisual feedbacks [26]. Moreover, these systems provide the following functions: ETA (electronic travel aid), EOA (electronic orientation aid), and PLD (position locator device) [26]–[29]. Conventional guiding systems designed for visually impaired people lack the adaptability to avoid obstacles and engage the users with their environments [30]–[33]. As shown in Fig. 1, the conventional guiding systems need to employ different subsystems to navigate the users. Each subsystem requires sensors to gauge the environment and then interacts with the users through different interfaces. Consequently, each subsystem requires specific software and hardware. Object detection, localization, sign interpretation, and guidance performed by multiple independent subsystems may greatly fatigue visually impaired users.

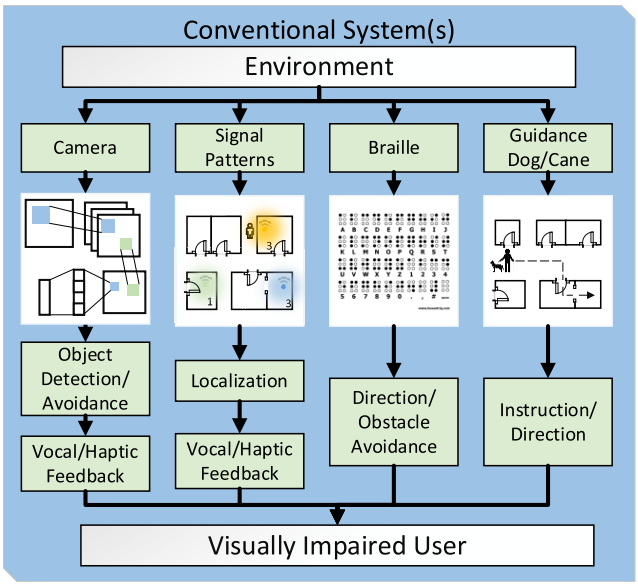


FIGURE 1. Design of a conventional guiding system for assisting visually impaired people.

To combat the shortcomings associated with conventional guiding systems, this paper proposes the STS-HMI system. The STS-HMI system informs visually impaired users by extracting visual cues from scenes and translating them into acoustic cues for guidance. As a result, complex information can be sent to the user instantly. Moreover, by manipulating the amplitude (loudness) of the binaural acoustic cues, the system enables users to infer a scene's complex information efficiently. By mitigating the limitations of conventional human-machine interfaces, the new interface enables visually impaired people to comprehend their surrounding environments with a richness similar to that experienced by sighted people.

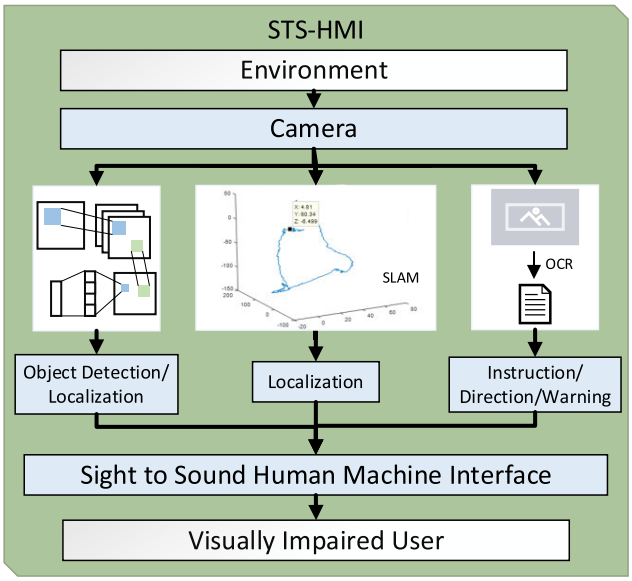


FIGURE 2. Proposed camera-based STS-HMI system.

Fig. 2 shows the proposed STS-HMI system with integrated multiple functions using computer vision. Moreover, computer vision systems are becoming increasingly capable of comprehending scenes through object detection, localization, and classification. Furthermore, with more powerful computing hardware, computationally demanding machine vision algorithms can be executed in real time on mobile devices for assisting visually impaired people.

Besides comprehending the environment for navigation, existing methods interact with users inefficiently in terms of intuitiveness, easiness to learn, instantaneousness, constrains on hardware, versatility, affordability, and/or the ability to interpret complex scenes. Some methods use voice prompts or haptic feedback to interact with users. Voice prompts and haptic feedback are unnatural and may cause fatigue. Fig. 3 functionally compares conventional navigation solutions with the proposed STS-HMI system. These conventional solutions include guide canes, guide dogs, electronic guiding devices, and voice assistants. We scored each type of solution from 0 to 3 based on seven attributes (intuitiveness, easiness to learn, instantaneousness, constrains on hardware, versatility, affordability, and the ability to interpret complex scenes). A higher score indicates better performance. Fig. 3 shows that conventional solutions such as guide canes are hard to use, while guide dogs are expensive to train and cost prohibitive. Electronic guiding devices are vision substitution systems that perform tasks such as ETA, EOA, or PLD. These systems typically require cameras, ultrasonic sensors, and lidar to perform scene analysis, the results of which are conveyed to the user by tactile, vibration, or voice feedback. Systems using sensors other than cameras usually require a customized device, in addition to a smartphone. These systems also lack versatility. Systems using voice prompts

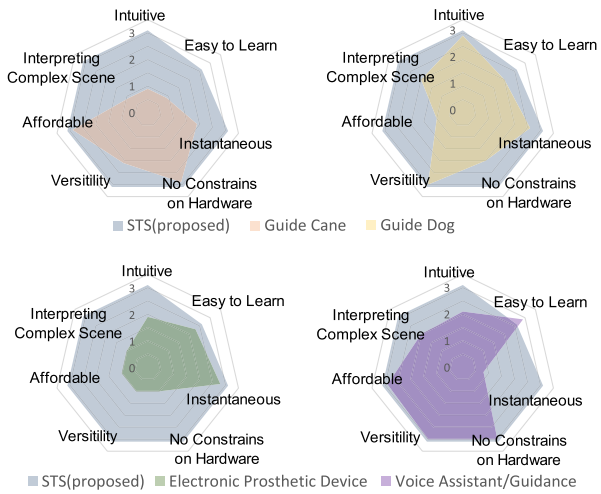


FIGURE 3. Radar chart comparing the proposed STS-HMI system with various other techniques: guide canes, guide dogs, electronic guiding devices, and voice assistants and guidance.

can convey complex information and are easy to comprehend; however, they are not fast enough for real-time navigation and are not intuitive [24], [34]. Other solutions such as wearable devices using haptic feedback and ultrasonic sensors set limitations on the system hardware [17], [35]–[38] or have limited applications [39].

Section II presents methods for identifying and localizing objects within the field of view of the camera. Section III presents the novel STS-HMI system for navigating visually impaired people. In Section IV, the experimental navigation results are presented.

II. OBJECT IDENTIFICATION AND LOCALIZATION

The STS-HMI system comprises two major components: (i) scene analysis for object detection, classification, and localization and (ii) a human-machine interface for guiding the visually impaired user. This section will explain how the STS-HMI system detects and locates various objects to build cognitive maps for navigation.

A. EXTRACTING VISUAL CUES FOR BUILDING COGNITIVE MAPS

Constructing cognitive maps is feasible using scene analysis, object detection, localization, and identification. In the STS-HMI system, YOLO (you only look once) [40] is used as the object detection engine for extracting visual cues from a scene. YOLO [41] is a state-of-the-art, real-time, all-purpose neural network for detecting a vast variety of objects in real time. The COCO (Common Objects in Context) data set [42] was used to train the YOLO network. The COCO data set contains objects from 80 categories. These objects are common in indoor environments such as homes, offices, and hospitals. YOLO can be retrained with an expanded COCO data set to identify additional objects if necessary. In practice, depending on the indoor environment, only a fraction of the object categories is relevant for scene analysis. Each object detected and localized by YOLO represents a visual cue. In the STS-HMI system, the object's distance and aspect are used to generate binaural acoustic cues for navigation. A cognitive map representing the physical distribution of objects within an environment can be constructed by comprehending and deciphering the binaural acoustic cues.

The STS-HMI system was designed on a smartphone platform to analyze scenes and to identify objects for navigation. Fig. 4 illustrates the necessary steps for a user to navigate. Navigation can be achieved by recognizing key objects, following the predetermined waypoints, and spotting signs. Key objects help the user to construct a cognitive map. Waypoints help the user to navigate by staying on a predetermined path. Signs provide essential guidance information regarding safe and reliable paths in complex spaces. In scene analysis, when an unsafe situation arises, such as encountering stairs or being on course to a collision, the STS-HMI system can be designed to generate a voice prompt to warn the user to proceed cautiously. Complimenting acoustic cues with voice prompts enables users to interact with the environment for efficient and safe navigation.

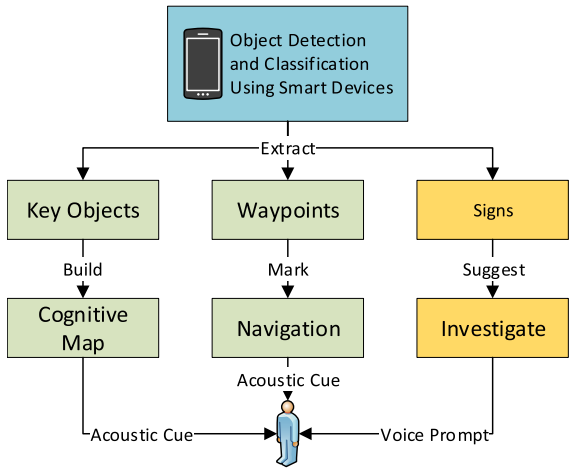


FIGURE 4. Block diagram for the STS-HMI navigation system.

B. LOCATING OBJECTS USING PHOTOGEOMETRY

Typical cameras are not designed to measure the distance of an object in captured images. The projection of an object onto the image plane is determined by the intrinsic matrix of the camera and the spatial relationship between the camera and the object. More specifically, the intrinsic matrix describes the optical characteristics of the camera including distortion, focal length (shown as f), and the resolution of the image sensor. The intrinsic matrix can be determined through calibration [43] and remains constant in a fixed-focal-length camera. Consequently, the intrinsic matrix needs to be measured only once for a particular camera. The projection of an object can be approximated using the pinhole model. In the pinhole model, the size of an object on the image plane is determined by the principle of similarity. Hence, a system cannot calculate the distance and the size of

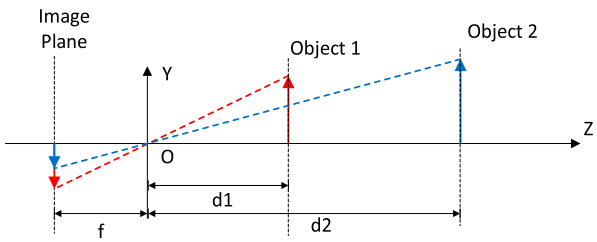


FIGURE 5. Illustration of the pinhole camera model.

an object simultaneously. Fig. 5 illustrates the pinhole camera model; this figure displays only the projection onto the $Y - Z$ plane.

As shown in Fig. 5, the projection of two objects is determined by their size and distance to the camera. Object 1 appears bigger on the image plane even though in reality it is smaller than Object 2. In general, the size and distance estimation with a regular camera remains challenging. The ambiguity of size and distance is intrinsic in vision systems, as defined by the following equation:

$$d = \frac{s}{\hat{s}} \times f \tag{1}$$

The distance between the object and the camera d can be inferred by the size of the projection $\hat{s}$, the camera's focal length f, and the actual object size s. In (1), s is the only unknown variable. The system can infer the size s of any detected object based on its label given by YOLO. For example, when the system detects a person, it can estimate the distance between the person and the camera based on the expected height of a person. Hence, the system can infer the distance between the person and the camera with reasonable accuracy. The system can also determine the aspect using the law of similarity. Fig. 6 shows the aspect of an object relative to the user, θ, and this can be calculated using the offset of the object $\hat{x}$ on the image plane and the focal length f:

$$\theta = \tan^{-1}\left(\hat{x}/f\right) \tag{2}$$

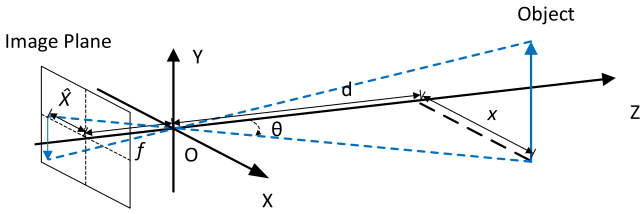


FIGURE 6. Estimating the aspect of an object.

C. NAVIGATION USING COGNITIVE MAPS

Recently, we studied SLAM (simultaneous localization and mapping) [43], [44] and AR (augmented reality) markers [45] for computer-assisted navigation. SLAM attempts to estimate the motion of the user and to map an indoor environment by analyzing camera footage. SLAM estimates the movement of the camera by comparing the changes between two frames. These changes include the movement of structure lines and feature points.

Another method that can assist indoor navigation is the determination of AR markers [46]. An AR maker is a square-based fiducial marker, which can easily be recognized by simple image processing techniques. The coding of an AR marker guarantees that any AR symbol will maintain a nonzero Hamming distance with itself after rotation [47]. The precision of the estimation depends on the size of the AR marker and the resolution of the camera. With a set of markers of 85 by 85 mm and a camera recording at 1920 by 1080 pixels, the error in the distance measurement can be kept below 50 mm within an angle of 0.2 rad [45].

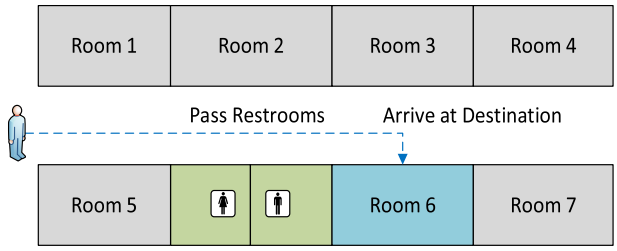


FIGURE 7. Navigation based on a cognitive map.

SLAM and AR markers do not provide information for constructing a cognitive map and, consequently, may not deliver a complete solution for navigating visually impaired people. People rely on cognitive maps using visual cues based on certain objects, landmarks, and signs. If visually impaired individuals can perceive visual cues as sighted people do, they will be able to navigate efficiently and intuitively. It is counterintuitive for people to keep track of their orientation and location step by step. Fig. 7 shows a sighted person searching for Room 6. Instead of counting the number of steps toward the intended location, this person will arrive in Room 6 right after passing the Restrooms. Therefore, to guide a visually impaired individual, a guiding system needs to detect the location of the restrooms and then convey this information to the user to construct a cognitive map. Therefore, the proposed STS-HMI system is designed to convert visual cues into acoustic cues and to guide visually impaired users by making them perceive a cognitive map.

III. ACOUSTIC-BASED HUMAN-MACHINE INTERFACE

A graphical user interface (GUI) is a practical and intuitive approach for human-machine interactions. Since it is impossible to assist visually impaired people using a GUI, the proposed STS-HMI system interacts with the user through predefined sound notes. These sound notes are binaural acoustic cues generated according to the type and the position of an object.

Mobile devices such as smartphones can assist visually impaired people by capturing and analyzing a scene for object detection, identification, and localization, as shown in Fig. 8. A smartphone empowered by neural networks can detect and classify objects in a scene [48]. The classified objects can

be translated into subband binaural acoustic cues. Through photogeometry, a machine vision system can calculate the location and the aspect of detected and classified objects. Based on an object's location and aspect, binaural acoustic cues can be generated for the user's perception.

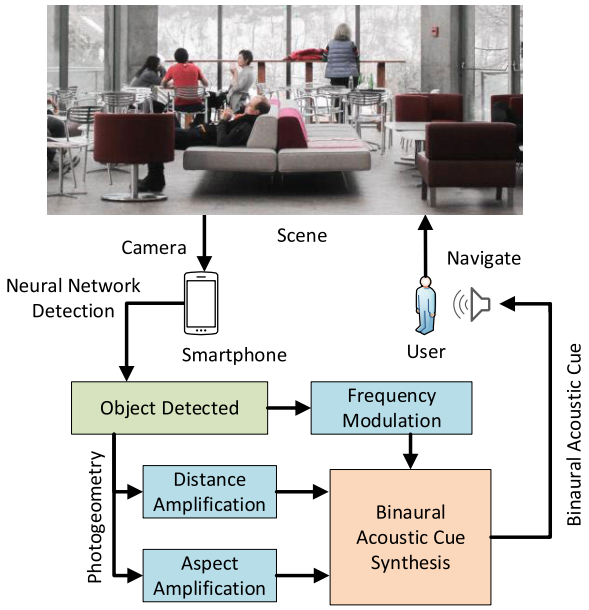


FIGURE 8. Design of the proposed acoustic-based user interface.

Typical navigation systems assist visually impaired people through voice prompts. For example, Microsoft's Seeing AI application [48] analyzes video footage captured by a smartphone and sends verbal commands to the user for guidance. In practice, it is difficult for a voice prompt to keep up with changes in a scene in a continuous manner. It takes a whole sentence for a voice prompt to describe a static scene. Hence, when a person is moving, verbal cues will not be able to pass sufficient information in real time without hampering the mobility of the person. Besides, the machine-generated voice prompt interferes with the natural communication activities of the user. Therefore, we developed efficient binaural acoustic cues in the STS-HMI system to counter such challenges.

A. ACOUSTIC CUES: A LANGUAGE FOR NAVIGATION

Human hearing perception can locate and unravel multiple sound sources. One can achieve such cognitive ability by analyzing interaural time differences (ITDs) and interaural level differences (ILDs) of sounds [49]. ITDs represent the difference between the arrival time of the same sound in both ears. ILDs represent the difference in the loudness of the sounds. By manipulating these two acoustic cues, a system can guide human perception with the location of a sound source. For example, surrounding sound technologies such as DTS (dedicated to sound) can create an immersive movie experience by mimicking the spatial arrangements of sound sources in a movie scene. By manipulating the quality of

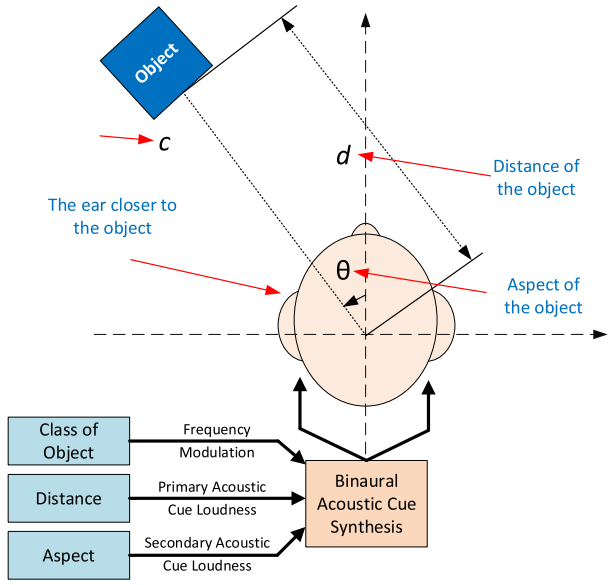


FIGURE 9. Top-down view of objects relative to the user.

sounds generated by the speakers, these technologies enable audiences to identify a sound source's location intuitively. The proposed STS-HMI system relies on human perception for sound source localization. Moreover, through training, visually impaired individuals can develop a more sensitive sense of hearing and sound source localization [50].

To create intuitive acoustic cues for navigation, the human-machine interface needs to encode the location and the aspect of each object by manipulating the loudness and the frequency of the acoustic cue. Based on this principle, an acoustic-based human-machine interface can be created. Such a system enables visually impaired users to visualize the context of their surroundings conveniently in real time. Once the user's surrounding has changed, the binaural acoustic cues are triggered to inform the user. Thus, users can be updated with their surrounding environment as if they can see.

Two attributes are defined for encoding the information describing an object in a scene: (i) the frequency associated with the class of the object and (ii) the loudness associated with the distance and the aspect of the object. As shown in Fig. 9, the human-machine interface chooses the frequency of binaural cues based on the class of the object c. Then, the ear closer to the object receives the primary acoustic cue corresponding to the distance d, and the other ear will receive the secondary acoustic cue corresponding to the aspect θ. The following sections explain how navigation instructions are encoded in the binaural acoustic cues.

B. OBJECT REPRESENTATION USING ACOUSTIC CUES

With a YOLO neural network, most common objects can be detected in real time. During navigation, it is important to identify those objects in the environment that might guide or block visually impaired people. Hence, for this study, seven classes of objects (see Fig. 10) have been defined. Six classes represent common objects, and one represents a

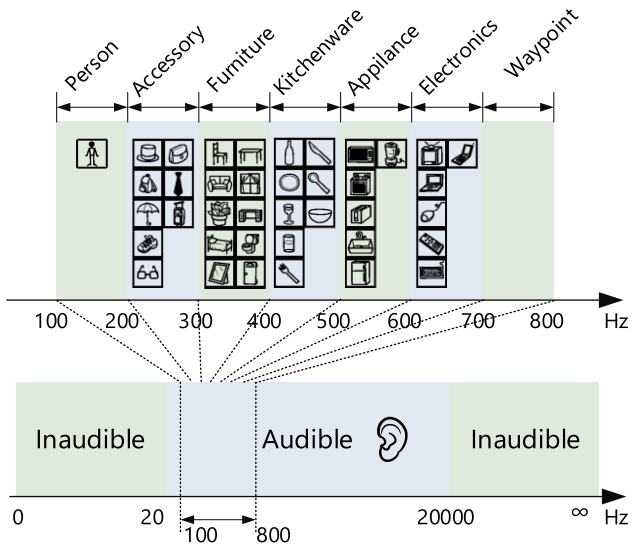


FIGURE 10. Audible subbands for object representation.

virtual object for the waypoint to direct the user. To represent these objects using acoustic cues, we split a band of the audible spectrum into seven subbands. Each of these subbands is associated with a group of similar objects in one of the seven classes. As shown in Fig. 10, the human audible spectrum ranges from 20 to 20000 Hz. For this study, the proposed STS-HMI system generates acoustic cues in a fraction of the audible spectrum (the frequency band from 100 to 800 Hz). This frequency band offers acceptable sensitivity and comfort [51]. Because of the limitation of using seven object groups to represent many objects, a voice command to the user may be the preferred method for alerting the user with the object type and distance.

C. ACOUSTIC CUES FOR OBJECT LOCATING

Locating an object necessitates the determination of the distance and the aspect of the object (see Fig. 9). The STS-HMI system encodes the distance and aspect information in acoustic cues. Depending on the position of the object, the ear closer to the object receives the primary acoustic cue, which encodes the distance information, while the other ear receives the secondary acoustic cue indicating aspect information. In the example shown in Fig. 9, the left ear receives the louder primary acoustic cue, and the right ear receives the weaker secondary acoustic cue. The user can differentiate and decode these cues by the loudness sensed in both ears.

1) DISTANCE ENCODING

The loudness perception is measured in phons [51]. Quantifying the loudness of generated sounds in phons allows the system to model the amplitude and loudness uniformly as a function of frequency. Equation (3) determines the amplitude of an acoustic cue A as a function of the distance d between the user and the object:

$$A = A_{\max} - \frac{A_{\min}}{1 + e^{-\frac{d-5}{2}}}, \quad \text{where } d \in [0, +\infty) \tag{3}$$

where the maximum and minimum amplitudes of the primary acoustic cue are $A_{\max}$ and $A_{\min}$, respectively. In the proposed STS-HMI system, the maximum loudness $A_{\max}$ is set at 60 phons, and the minimum loudness $A_{\min}$ is set at 30 phons. The upper limit is set at the loudness level of the human voice, while the lower limit is set at the intensity of ambient noise [52]. By limiting the minimum loudness of acoustic cues to 30 phons, users can maintain awareness of their surroundings in a typical environment. Moreover, by limiting the maximum loudness of acoustic signals to 60 phons, the user's verbal communication remains feasible. Equation (3) models the loudness of the acoustic cue by a logistic function. One of the benefits of using a logistic function is that the loudness of the generated cue increases quickly when objects nearly collide with the user (see Fig. 12). Such a quick increase in loudness can draw users' attention when objects are closer than 5 m. On the other hand, objects that are more than 10 m away can be ignored according to the acoustic cue loudness model shown in Fig. 12.

2) ASPECT ENCODING

The primary acoustic cue reflects distance information, and this information is insufficient for determining the location of an object relative to the user (see Fig. 9). A secondary acoustic cue is necessary to specify the object's aspect. If an object is on the left of the user, a louder acoustic cue is generated in the left ear. When an object is centered in front of the user, both ears receive acoustic cues with equal loudness. This approach for loudness in the left or right ear is intuitive, and (4) regulates the amplitude (loudness) of the secondary acoustic cue A' relative to the primary acoustic cue A:

$$A' = (A - 30)k + 30 \tag{4}$$

where

$$k = -\frac{2}{\pi}|\theta| + 1, \quad \text{where } \theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \tag{5}$$

The factor $k \in [0, 1]$ is determined by θ, the aspect of the object to the user. As a result, when the object is in front of the user ($\theta = 0$), both ears receive acoustic cues with the same loudness. Once the object moves to either side of the user, k decreases; the loudness, too, decreases as the aspect $|\theta|$ increases. The loudness of the secondary acoustic cue will decrease to 30 phons (i.e., background noise) when the aspect reaches $-\frac{\pi}{2}$ or $\frac{\pi}{2}$, a condition in which the object is no longer in the user's field of view.

During navigation, the STS-HMI system informs the user of surrounding objects. In the vicinity of an object, the user scans the environment by turning his or her head. The binaural cues have equal loudness when the user's head faces the object. As the user walks toward the object, the loudness of binaural cues will increase equally. Fig. 11 shows the characteristics of acoustic cues according to the positions of objects relative to the position of the user of the STS-HMI system. Fig. 11a shows the user facing toward kitchenware on the left and an appliance positioned on the right side.

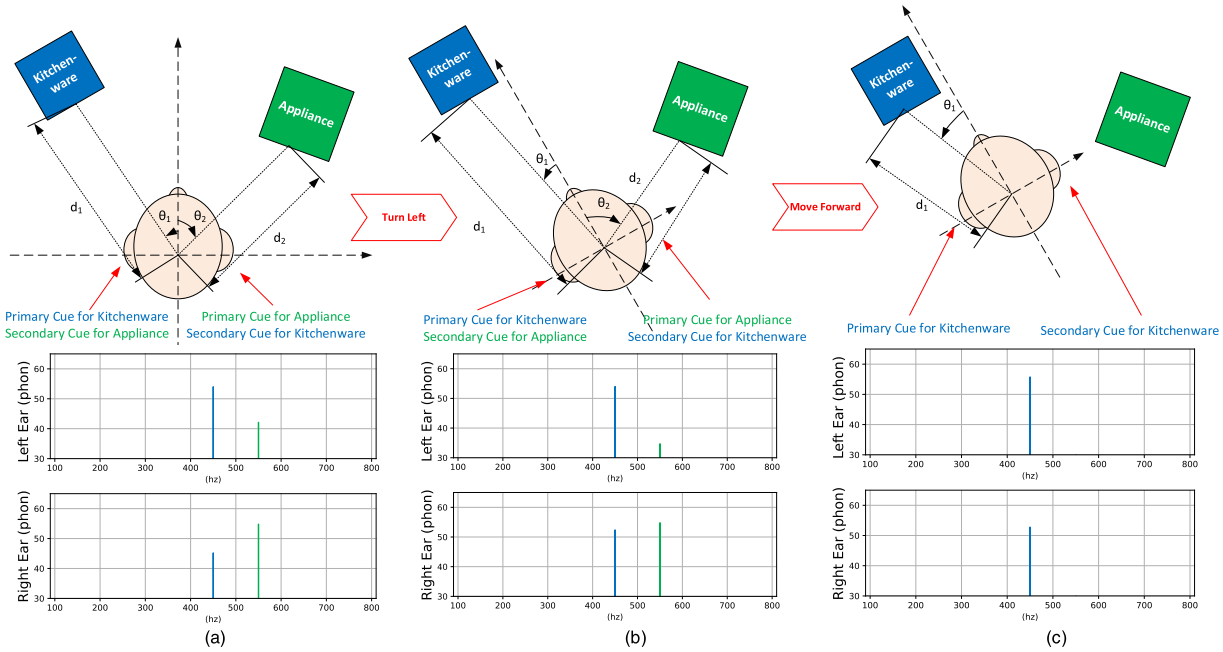


FIGURE 11. Signals changing in both ears while the user is moving.

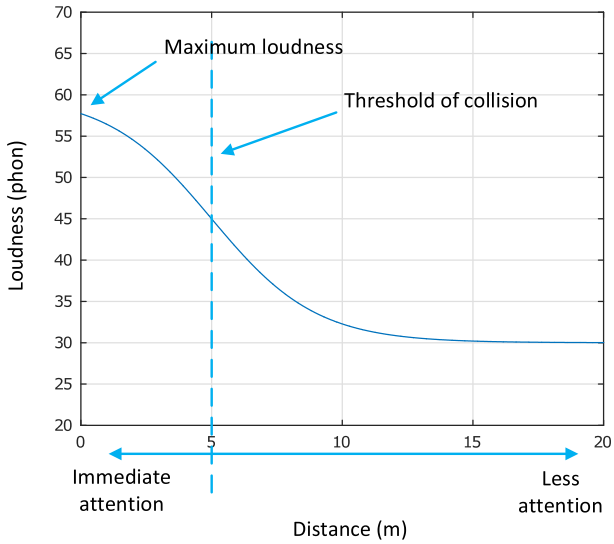


FIGURE 12. The amplitude (loudness) of acoustic cues versus distance of an object in the field of view for navigation.

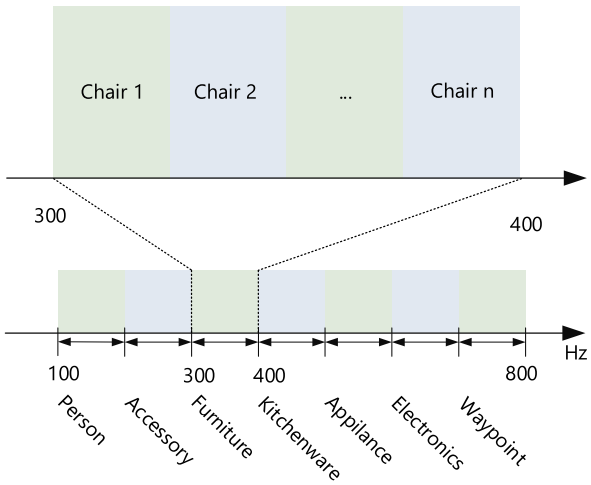


FIGURE 13. Subband splitting for multiple objects.

Fig. 11b shows that when the user turns left, the aspects of both objects change while the distances remain the same. Consequently, the distance acoustic cue (blue line) in the left ear and the aspect acoustic cue (green line) in the right ear remain the same. When the user moves forward, the appliance disappears from the user's field of view. This can be observed by inspecting Fig. 11c, in which the green acoustic cues vanish in both ears.

D. REPRESENTATION OF MULTIPLE OBJECTS USING ACOUSTIC CUES

As explained in Fig. 10, the STS-HMI system assigns seven subbands from the audible spectrum to seven classes of objects. However, in practice, it is common that multiple objects from the same class appear in the field of view. To inform the user of such a situation, the frequency of the acoustic cue within a class of objects needs to reflect the number of objects. Let the frequency ω_m be the cue representing object m from class c, where $i \in [1, n]$. Hence, ω_m can be calculated using the following equation:

$$\omega_m = \omega_c + \frac{m}{n+1} \times S \tag{6}$$

where S represents the reserved bandwidth for a given subband and ω_c represents the lower-end frequency of the class c subband. As a result, when multiple objects appear in the field of view, the acoustic cue becomes a summation of multiple discrete frequencies representing all objects:

$$\Omega = \Sigma_{i=1}^{n} \omega_i \tag{7}$$

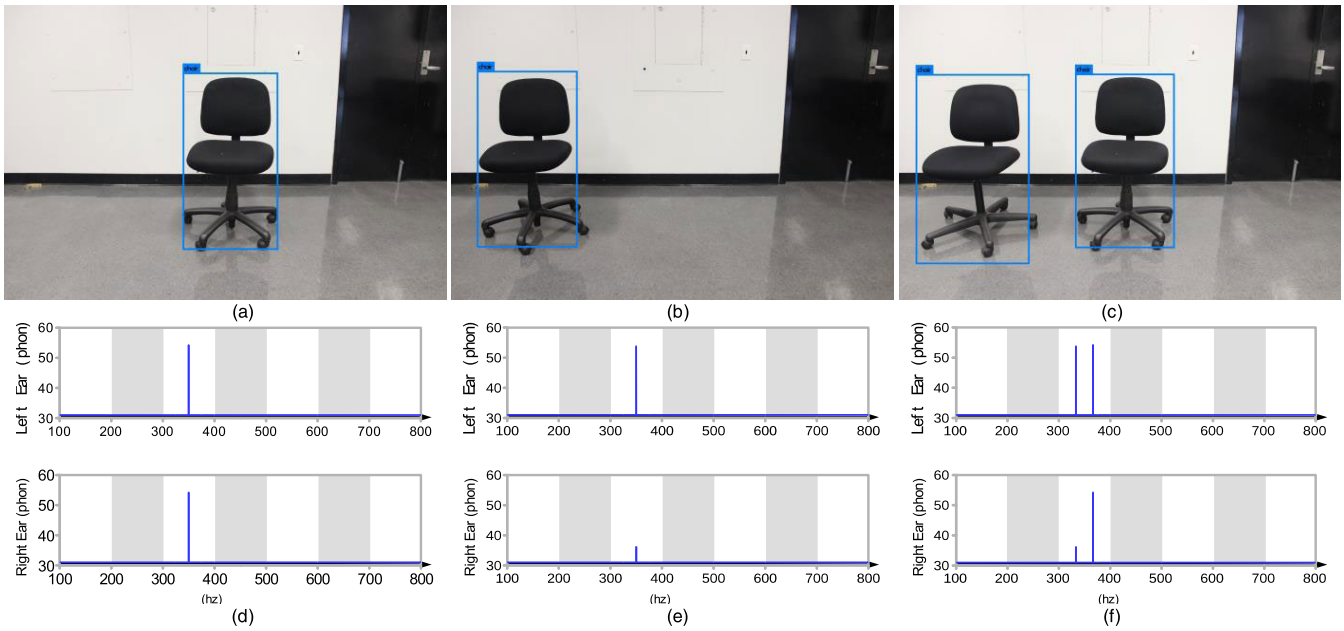


FIGURE 14. **Top: chair(s) detected by STS-HMI. Bottom: generated binaural acoustic cues.**

Equation (7) gives the acoustic cue's broadband spectrum for a complex scene with n objects from the same class. Subband splitting will be carried out on binaural cues; hence, primary and secondary acoustic cues representing the same object will occupy the same frequency band. Then, by comparing the loudness of the binaural cues, users can locate the objects. Fig. 13 shows an example of subband splitting. When n chairs appear in a scene, the furniture subband will be further split into n subbands. The frequency resolution of human hearing (pitch) can be as small as 3.6 Hz [52]. However, in our experiments, we observed that an untrained ear can resolve frequency increments of 20 Hz. Given a bandwidth of 100 Hz for each object class, an untrained person can differentiate five objects within that class.

IV. STS-HMI EXPERIMENTAL RESULTS AND DISCUSSION

Several experiments were conducted to assess the performance of the STS-HMI system. In particular, this section examines the performance of the acoustic-cue-guided navigation.

A. ACOUSTIC CUES FOR OBJECT IDENTIFICATION AND LOCALIZATION

For the first experiment, a chair was placed 1.8 m in front of the camera (see Fig. 14a). The chair was detected by YOLO, and the image of the chair was framed by a bounding box. The aspect angle is 0 since the chair is positioned in the center of the camera's field of view. Consequently, as shown in Fig. 14d, the loudness (representing distance) of the binaural acoustic cues is the same in both ears. In the second experiment (see Fig. 14b), the chair was placed in the front-left (1.8 m to the front and 0.9 m to the left) of the camera. Fig. 14e shows that the aspect of the chair was encoded in the acoustic cue applied to the right ear. In this experiment, the chair was closer to the left ear; hence, the primary cue carrying distance information was applied to the left ear while the secondary acoustic cue carrying the aspect information was applied to the right ear (for clarification of this action, see the description of Fig. 9). In the third experiment, two chairs were placed in front of the camera, with one in the center-front and the other in the front-left of the camera (see Fig. 14c). Fig. 14f shows that two discrete frequencies are required to represent both chairs, as mentioned in Section III-D. Both chairs are objects within the furniture class. Therefore, the subband associated with the furniture class will be further split into two subbands, each of them occupied by one chair. As shown in Fig. 14f, the primary and secondary acoustic cues associated with the chair on the left are sent to the left and the right ear, respectively. The primary and secondary acoustic cues associated with the chair directly in front of the camera are the same.

B. ACOUSTIC CUES FOR NAVIGATION

A series of experiments were created using Panda3D (i.e., a lightweight video game engine developed by Disney and Carnegie Mellon University) to test the effectiveness of the STS-HMI system for navigating visually impaired people. Before the experiment, all test subjects were given an oral instruction on how to use the system. Then, the test subjects learned the binaural acoustic cues by completing simple tasks such as identifying an object and its location for navigation. With a few trials, the test subjects were able to utilize the system in a complex scene. First, a virtual navigation course was designed for testing. The test subjects were asked to

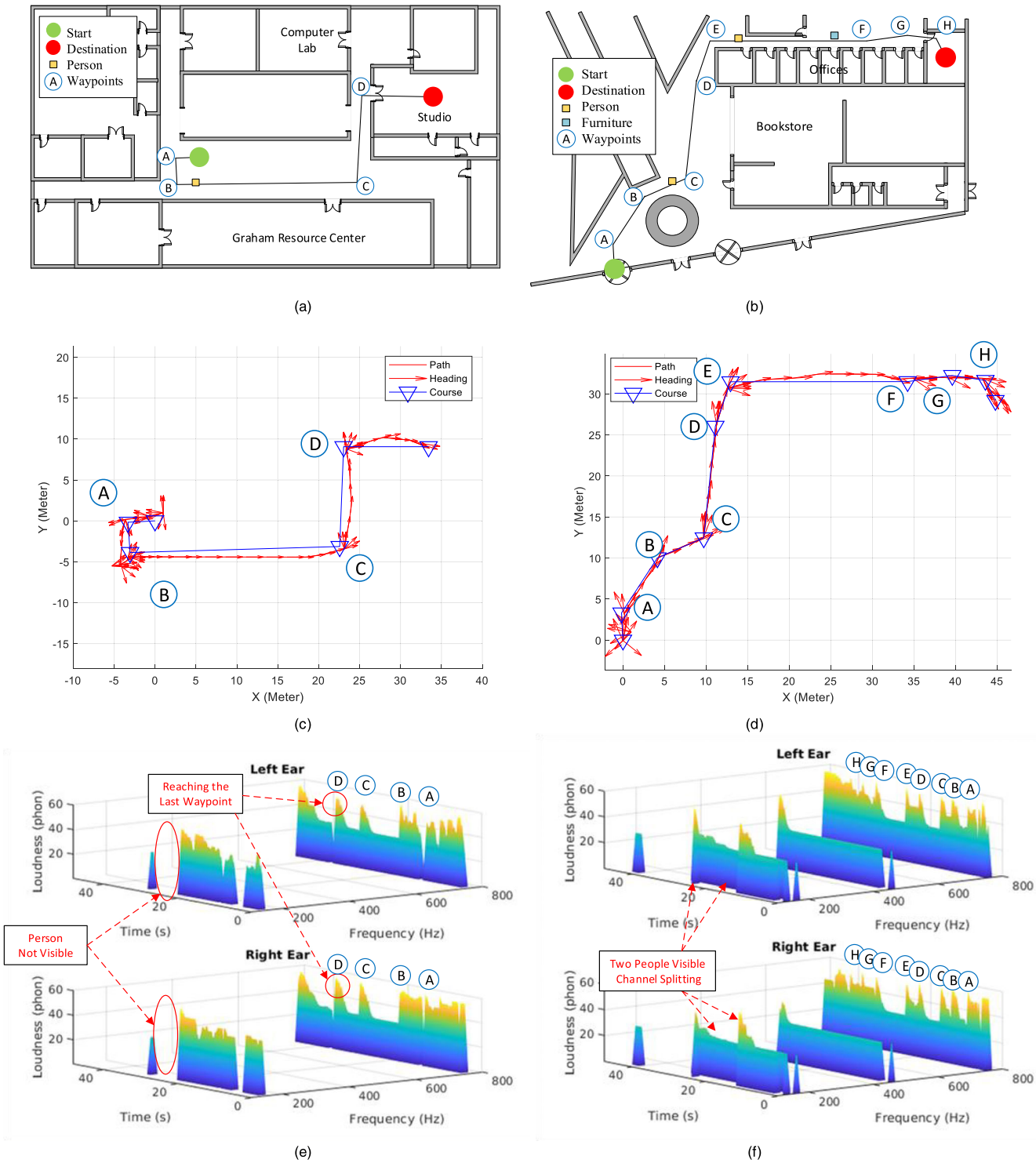


FIGURE 15. Top: chair(s) detected by STS-HMI. Bottom: generated acoustic cues.

follow the course using binaural acoustic cues. Like players of the first-person shooter video game, they traversed a virtual navigation course by moving an avatar via keyboards and mice. When a test subject reached a waypoint, a feedback sound was played, and the acoustic cues for the next waypoint were generated. By comparing the test subject's path and the designated navigation course, we can quantitively assess the effectiveness of the STS-HMI system.

The virtual navigation paths were modeled based on the floor plan of the Crown Hall (Fig. 15.a) and the McCormick Tribune Campus Center (MTCC) (Fig. 15.b) on the campus of the Illinois Institute of Technology. By simulating the path

in the virtual environment, the game engine was programmed to generate acoustic cues for navigation. The test subjects were blindfolded and guided by the binaural acoustic cues only. As the test subjects were walking toward the destination, the position and heading were recorded. Figs. 15c and 15d show the path and the heading of a test subject as red lines and red arrows, respectively. In these figures, the simulated paths and waypoints are shown as blue lines and blue triangles, respectively. As can be observed from Figs. 15c and 15d, the test subjects were able to follow the acoustic cue and complete the course with reasonable accuracy. In Figs. 15e and 15f, the binaural acoustic cues are shown in time-frequency distributions. As a subject walks closer to a waypoint, the system will increase the loudness of the acoustic cues (see (4)) at a predefined frequency of 750 Hz (see Fig. 10). Similarly, other detected objects will trigger their corresponding cues.

The performance of the STS-HMI system can be analyzed quantitively by measuring the deviation of the test subject's track from the actual navigation path. As shown in Fig. 16 (navigation path in the Crown Hall), the test subject was able to stay on the track within 2 m from the actual navigation path. Considering the tight corner in the Crown Hall, maintaining an error of less than 2 m (1.9% of the total distance traveled) is an acceptable accuracy for indoor navigation. Moreover, the divergence from the actual path is finite and nonaccumulating. The error surges temporarily each time the subject reaches a waypoint (see circled markers in Fig. 16).

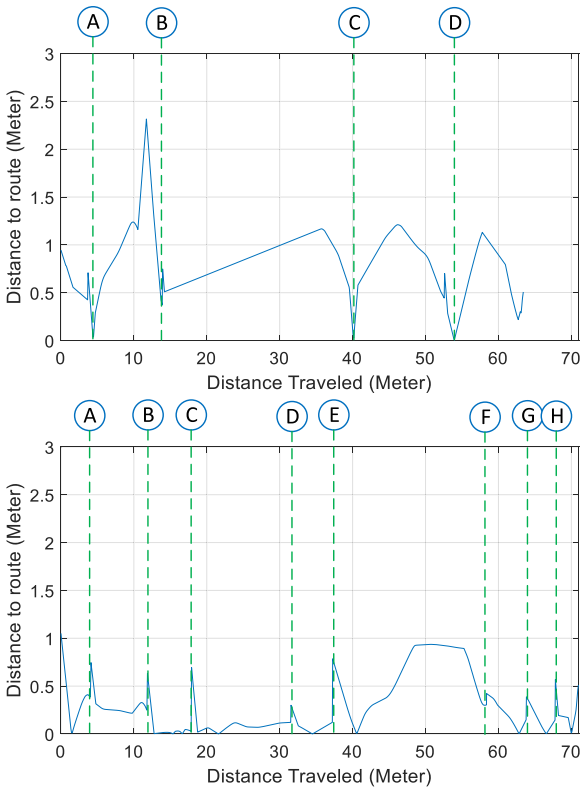


FIGURE 16. Deviation from the course while traveling the path in the Crown Hall (top) and the path in the MTCC (bottom).

Fig. 17 shows that the test subject's distance to the final destination decreases consistently as the subject advances through the course. This further confirms the intuitiveness and effectiveness of the binaural acoustic cues generated by the STS-HMI system. The STS-HMI system can be realized using common computational systems such as smartphones. These systems are highly efficient, and their power consumption may not exceed 10 W. Therefore, a typical 10-Wh battery can operate for at least one hour before recharging.

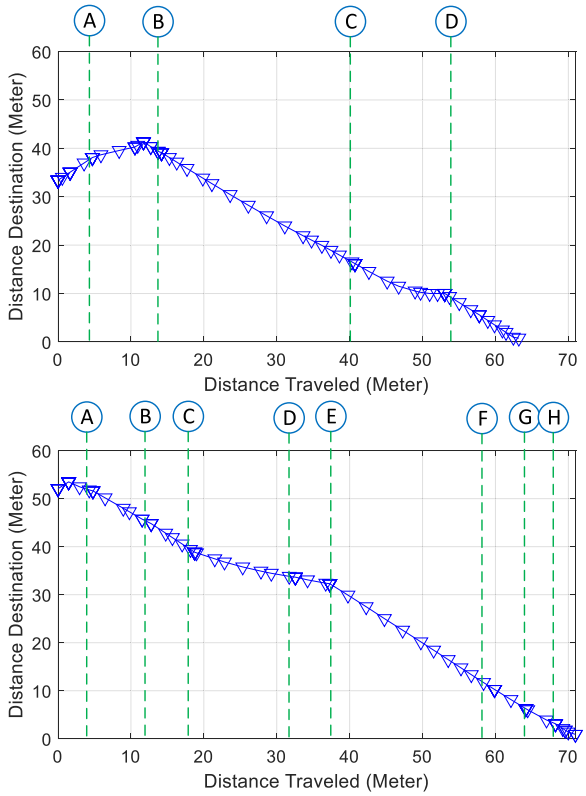


FIGURE 17. Distance to destination versus distance traveled in the Crown Hall (top) and the MTCC (bottom).

C. DISCUSSION

In the current design of the STS-HMI system, a regular RGB camera and a neural network are used to infer the depth information in the field of view. This approach is realizable using common devices such as smartphones, which are highly practical in terms of cost, size, and robustness. There are relatively costly application-specific devices available for depth measurements [53]. For example, LiDAR sensors [54] provide high-precision object localization. Other devices such as HoloLense AR goggles and stereovision cameras also provide depth measurements [55].

The scene analysis for object detection and navigation can be implemented on a portable device such as a smartphone or embedded devices (e.g jetson nano [56]). The YOLO network is trained using COCO, a data set with 1.5 million object images acquired in a variety of settings. In our experiments, the YOLO network exhibited

promising performance in terms of speed, accuracy, and robustness. Furthermore, YOLO can perform adequately in object recognition regardless of the size of the object, obstructions, brightness, contrast, hue, saturation, and/or noise [57]. The YOLO network can also detect objects by processing videos frame by frame. Video processing, although computationally intensive, can be realized in real time (6 – 207 frames per second depending on system specification, platform, and detection criteria) [57]–[59]. Moreover, the network can be retrained with a dataset eliminating uncommon classes for a higher frame rate. Furthermore, the 80 categories of objects in COCO are not unique for a specific domain, so the STS-HMI system can adapt to different environments such as hospitals, schools, and office buildings. Also, object recognition by the YOLO network can be extended by retraining it with additional classes of objects.

There is a trade-off between the cognitive bandwidth for the encoded information describing a scene and the intuitiveness for navigation. Short voice prompts offer limited context and, therefore, are inefficient in describing a changing scene while the user is navigating. It takes a few sentences for a verbal cue to describe a static scene. If the user is moving continuously, the voice prompt is unable to pass sufficient information in real time without hampering the user's mobility [48], [55], [60]. Moreover, it is hard to provide intuitive spatial information via descriptive sentences. To combat these limitations, the proposed STS-HMI system encodes information regarding objects' location and aspect by manipulating the frequency and the amplitude of the binaural acoustic cues. Future studies of binaural acoustic cues for scene analysis and navigation can be extended to include information encoded in pitches since human ears are also sensitive to pitch changes [52] and time differences [61].

In summary, the motivation for designing the STS-HMI system was to develop a machine vision system for rendering a cognitive map using binaural acoustic cues. Experimentally, we demonstrated the feasibility of realizing such a concept for indoor navigation since indoor spaces are often so complex that only machine vision can render cognitive information. On the other hand, for outdoor navigation, the STS-HMI system can be supplemented by GPS to further increase its versatility and accuracy.

V. CONCLUSION

This paper presents the STS-HMI system, a novel acoustic-cue-based navigation system for assisting visually impaired people. Existing navigation tools are task specific and limited. Hence, they are not suitable for object detection and navigation in a complex environment. However, it is desirable to develop a portable integrated system that is capable of object detection, localization, and navigation for visually impaired people. These design objectives are achievable with advancements in computer vision and machine learning.

The STS-HMI system creates a language that encodes information for intuitive navigation using binaural acoustic cues. By manipulating the frequency and amplitude of the cues, the system can convey the distance and aspect of objects to visually impaired people so that they can build a cognitive map. This approach is algorithm specific and does not require specific hardware and resources for scene analysis and human-machine interfaces. The STS-HMI system leverages neural networks and computer vision algorithms to detect and locate common objects. Then, this information is translated into binaural acoustic cues that enable visually impaired users to build cognitive maps for navigation. The experimental results support that the STS-HMI system is an intuitive and cost-effective mobile solution for navigating visually impaired people in difficult environments. The system is versatile and can readily be integrated with common indoor positioning systems such as AR markers and SLAM to achieve high-precision autonomous indoor navigation.

REFERENCES

[1] S. A. Marchette, A. Bakker, and A. L. Shelton, "Cognitive mappers to creatures of habit: Differential engagement of place and response learning mechanisms predicts human navigational behavior," *J. Neurosci.*, vol. 31, no. 43, pp. 15264–15268, Oct. 2011.

[2] R. Kitchin and S. Freundschuh, "Cognitive mapping and spatial decision-making," in *Cognitive Mapping: Past, Present, and Future*. Evanston, IL, USA: Routledge, 2000.

[3] A. A. Kalia, G. E. Legge, and N. A. Giudice, "Learning building layouts with non-geometric visual information: The effects of visual impairment and age," *Perception*, vol. 37, no. 11, pp. 1677–1699, Nov. 2008.

[4] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using wiFi," in *Proc. ACM Conf. Special Interest Group Data Commun.*, New York, NY, USA, 2015, pp. 269–282.

[5] B. Ilias, S. A. A. Shukor, A. H. Adom, M. F. Ibrahim, and S. Yaacob, "A novel indoor mobile robot mapping with USB-16 ultrasonic sensor bank and NWA optimization algorithm," in *Proc. IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Batu Feringghi, Malaysia, May 2016, pp. 189–194.

[6] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proc. 3rd Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, New York, NY, USA, 2005, pp. 205–218.

[7] D. Croce, P. Gallo, D. Garlisi, L. Giarre, S. Mangione, and I. Tinnirello, "ARIANNA: A smartphone-based navigation system with human in the loop," in *Proc. 22nd Medit. Conf. Control Autom.*, Palermo, Italy, Jun. 2014, pp. 8–13.

[8] J. Biswas and M. Veloso, "WiFi localization and navigation for autonomous indoor mobile robots," in *Proc. IEEE Int. Conf. Robot. Autom.*, Anchorage, Alaska, May 2010, pp. 4379–4384.

[9] S. Willis and S. Helal, "RFID information grid for blind navigation and wayfinding," in *Proc. 9th IEEE Int. Symp. Wearable Comput. (ISWC)*, Osaka, Japan, Oct. 2005, pp. 34–37.

[10] S. Chumkamon, P. Tuvaphanthaphiphat, and P. Keeratiwintakorn, "A blind navigation system using RFID for indoor environments," in *Proc. 5th Int. Conf. Electr. Eng./Electron., Comput., Telecommun. Inf. Technol.*, Krabi, Thailand, May 2008, pp. 765–768.

[11] M. Bessho, S. Kobayashi, N. Koshizuka, and K. Sakamura, "A space-identifying ubiquitous infrastructure and its application for tour-guiding service," in *Proc. ACM Symp. Appl. Comput. (SAC)*, Fortaleza, Brazil, 2008, pp. 1616–1621.

[12] B. Ding, H. Yuan, L. Jiang, and X. Zang, "The research on blind navigation system based on RFID," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Shanghai, China, Sep. 2007, pp. 2058–2061.

[13] S. Koide and M. Kato, "3-D human navigation system considering various transition preferences," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Waikoloa, HI, USA, vol. 1, Oct. 2005, pp. 859–864.

[14] T. Hollerer, D. Hallaway, N. Tinna, and S. Feiner, "Steps toward accommodating variable position tracking accuracy in a mobile augmented reality system," in *Proc. 2nd Int. Workshop Artif. Intell. Mobile Syst. (AIMS)*, Seattle, WA, USA, 2001, pp. 31–37.

[15] K. Nakamura, Y. Aono, and Y. Tadokoro, "A walking navigation system for the blind," *Syst. Comput. Jpn.*, vol. 28, no. 13, pp. 36–45, Nov. 1997.

[16] G. Retscher, "Pedestrian navigation systems and location-based services," in *Proc. 5th IEE Int. Conf. 3G Mobile Commun. Technol. (3G) Premier Tech. Conf. 3G Beyond*, London, U.K., 2004, pp. 359–363.

[17] A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," in *Proc. ACM SIGACCESS Conf. Comput. Accessibility (ASSETS)*, New York, NY, USA, 2004, pp. 147–152.

[18] L. Ran, S. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," in *Proc. 2nd IEEE Annu. Conf. Pervasive Comput. Commun.*, Orlando, FL, USA, Mar. 2004, pp. 23–30.

[19] G. Retscher and M. Thienelt, "NAVIO–A navigation and guidance service for pedestrians," *Positioning*, vol. 1, nos. 1–2, p. 8, 2004.

[20] S. Ertan, C. Lee, A. Willets, H. Tan, and A. Pentland, "A wearable haptic navigation guidance system," in *Dig. Papers. 2nd Int. Symp. Wearable Comput.*, Pittsburgh, PA, USA, Oct. 1998, pp. 164–165.

[21] J. Baus, A. Krüger, and W. Wahlster, "A resource-adaptive mobile navigation system," in *Proc. 7th Int. Conf. Intell. User Interfaces*, San Francisco, CA, USA, 2002, pp. 15–22.

[22] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proc. 6th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, New York, NY, USA, 2000, pp. 32–43.

[23] H. Huang, G. Gartner, M. Schmidt, and Y. Li, "Smart environment for ubiquitous indoor navigation," in *Proc. Int. Conf. New Trends Inf. Service Sci.*, Beijing, China, Jun. 2009, pp. 176–180.

[24] Y.-J. Chang, S.-K. Tsai, and T.-Y. Wang, "A context aware handheld wayfinding system for individuals with cognitive impairments," in *Proc. 10th Int. ACM SIGACCESS Conf. Comput. Accessibility (Assets)*, New York, NY, USA, 2008, pp. 27–34.

[25] A. Smailagic, R. Martin, B. Rychlik, J. Rowlands, and B. Ozceri, "Metronaut: A wearable computer with sensing and global communication capabilities," *Pers. Technol.*, vol. 1, no. 4, pp. 260–267, Dec. 1997.

[26] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[27] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet Things*, vol. 11, Sep. 2020, Art. no. 100188.

[28] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, Mar. 2017.

[29] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognit. Lett.*, vol. 137, pp. 37–52, Sep. 2020.

[30] T. Gonnot and J. Saniie, "Integrated machine vision and communication system for blind navigation and guidance," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, Grand Forks, ND, USA, May 2016, pp. 187–191.

[31] T. Gonnot and J. Saniie, "Image sensing system for navigating visually impaired people," in *Proc. IEEE Sensors*, Baltimore, MD, USA, Nov. 2013, pp. 1–4, doi: 10.1109/ICSENS.2013.6688594.

[32] A.-S. S. A. Malik, E.-Z. Ali, A.-S. Yousef, A.-H. Khaled, and G. Abdu, "Designing braille copier based on image processing techniques," *Int. J. Soft Comput. Eng. (IJSCE)*, vol. 4, no. 5, pp. 62–69, 2014.

[33] M. Arikawa, S. Konomi, and K. Ohnishi, "Navitime: Supporting pedestrian navigation in the real world," *IEEE Pervasive Comput.*, vol. 6, no. 3, pp. 21–29, Jul. 2007.

[34] P. Lei, M. Chen, and J. Wang, "Speech enhancement for in-vehicle voice control systems using wavelet analysis and blind source separation," *IET Intell. Transp. Syst.*, vol. 13, no. 4, pp. 693–702, Apr. 2019.

[35] N. Bourbakis and D. Kavraki, "A 2D vibration array for sensing dynamic changes and 3D space for blinds' navigation," in *Proc. 5th IEEE Symp. Bioinf. Bioeng. (BIBE)*, Minneapolis, MN, USA, Oct. 2005, pp. 222–226.

[36] V. Kulyukin, C. Gharpure, J. Nicholson, and G. Osborne, "Robot-assisted wayfinding for the visually impaired in structured indoor environments," *Auto. Robots*, vol. 21, no. 1, pp. 29–41, Aug. 2006.

[37] S. Hashino and S. Yamada, "An ultrasonic blind guidance system for street crossings," in *Computers Helping People With Special Needs* (Lecture Notes in Computer Science), vol. 6180, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Berlin, Germany: Springer, 2010, doi: 10.1007/978-3-642-14100-3_35.

[38] K. Patil, Q. Jawadwala, and F. C. Shu, "Design and construction of electronic aid for visually impaired people," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 2, pp. 172–182, Apr. 2018.

[39] L. Shangguan, Z. Yang, Z. Zhou, X. Zheng, C. Wu, and Y. Liu, "CrossNavi: Enabling real-time crossroad navigation for the blind with commodity phones," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.-UbiComp Adjunct*, Seattle, WA, USA, 2014, pp. 787–798.

[40] J. Redmon. (2016). *Darknet: Open Source Neural Networks in C*. [Online]. Available: https://pjreddie.com/darknet/

[41] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 2016, *arXiv:1612.08242*. [Online]. Available: http://arxiv.org/abs/1612.08242

[42] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and L. C. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[43] G. Yang, Z. Zhou, T. Gonnot, and J. Saniie, "Design flow of motion based single camera 3D mapping," in *Proc. IEEE Int. Conf. Electro/Inf. Technol. (EIT)*, Dekalb, IL, USA, May 2015, pp. 341–345.

[44] H. Lim, J. Lim, and H. J. Kim, "Real-time 6-DOF monocular visual SLAM in a large-scale environment," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Hong Kong, May 2014, pp. 1532–1539.

[45] G. Yang and J. Saniie, "Indoor navigation for visually impaired using AR markers," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, Lincoln, NE, USA, May 2017, doi: 10.1109/EIT.2017.8053383.

[46] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognit.*, vol. 51, pp. 481–491, Mar. 2016.

[47] B. Kim, M. Kwak, J. Lee, and T. T. Kwon, "A multi-pronged approach for indoor positioning with WiFi, magnetic and cellular signals," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Busan, South Korea, Oct. 2014, pp. 723–726.

[48] Microsoft. *Seeing AI*. Accessed: Feb. 12, 2020. [Online]. Available: https://www.microsoft.com/en-us/ai/seeing-ai

[49] M. E. Nilsson and B. N. Schenkman, "Blind people are more sensitive than sighted people to binaural," *Hearing Res.*, vol. 332, pp. 223–232, Feb. 2016.

[50] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS ONE*, vol. 6, no. 5, May 2011, Art. no. e20162, doi: 10.1371/journal.pone.0020162.

[51] *Normal Equal-Loudness-Level Contours*, Standard ISO 226:2003, 2003.

[52] H. F. Olson, "Theater, studio, and room acoustics," in *Music, Physics and Engineering*. New York, NY, USA: Dover, 1967, pp. 266–325.

[53] M. Bujacz and P. Strumillo, "Sonification: Review of auditory display solutions in electronic travel aids for the blind," *Arch. Acoust.*, vol. 41, no. 3, pp. 401–414, Sep. 2016.

[54] Y.-S. Shin, Y. S. Park, and A. Kim, "DVL-SLAM: Sparse depth enhanced direct visual-LiDAR SLAM," *Auto. Robots*, vol. 44, no. 2, pp. 115–130, Jan. 2020.

[55] Y. Liu, N. R. Stiles, and M. Meister, "Augmented reality powers a cognitive assistant for the blind," *ELife*, vol. 7, Nov. 2018, Art. no. e37841, doi: 10.7554/eLife.37841.001.

[56] G. Oltean, C. Florea, R. Orghidan, and V. Oltean, "Towards real time vehicle counting using YOLO-tiny and fast motion estimation," in *Proc. IEEE 25th Int. Symp. Design Technol. Electron. Packag. (SIITME)*, Cluj-Napoca, Romania, Oct. 2019, pp. 240–243.

[57] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*. [Online]. Available: http://arxiv.org/abs/2004.10934

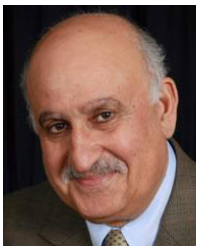
[58] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Feb. 2020.

[59] M. Hollemans. (May 20, 2017). *Real-Time Object Detection With YOLO*. Accessed: Sep. 19, 2020. [Online]. Available: https://machinethink.net/blog/object-detection-with-yolo/

[60] V. Shafiro and B. Gygi, "Perceiving the speech of multiple concurrent talkers in a combined divided and selective attention task," *J. Acoust. Soc. Amer.*, vol. 122, no. 6, pp. EL229–EL235, Dec. 2007, doi: 10.1121/1.2806174.

[61] A. D. Brown, "Temporal weighting of binaural cues for sound localization," Ph.D. dissertation, Dept. Speech Hearing Sci., Univ. Washington, Seattle, WA, USA, 2012.

GUOJUN YANG (Member, IEEE) received the B.E. degree in automation from the Beijing Information Science and Technology University, in 2013, and the M.S. degree in electrical engineering from the Illinois Institute of Technology, in 2015, where he is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department. He is an Research Assistant with the Electrical and Computer Engineering Department, Illinois Institute of Technology. His research interests include artificial intelligence, machine vision, image processing, embedded computing, and human–machine interface. His research was supported in part by Walter and Harriet Filmer Endowment from 2016 to 2020, and the Bhakta and Sushama Rath Endowed Research Award in 2019.



JAFAR SANIIE (Life Fellow, IEEE) received the B.S. degree (high honors) in electrical engineering from the University of Maryland, in 1974, the M.S. degree in biomedical engineering from Case Western Reserve University, Cleveland, Ohio, in 1977, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, Indiana, in 1981. In 1981, he joined the Department of Applied Physics, University of Helsinki, Finland, to conduct research in photothermal and photoacoustic imaging. Since 1983, he has been with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, where he is the Department Chair, the Filmer Endowed Chair Professor, and the Director of Embedded Computing and Signal Processing (ECASP) Research Laboratory. His research interests and activities are in ultrasonic signal and image processing, ultrasonic software-defined communication, artificial intelligence and machine learning, statistical pattern recognition, estimation and detection, data compression, time-frequency analysis, embedded digital systems, system-on-chip hardware/software co-design, the Internet of Things, computer vision, deep learning, and ultrasonic nondestructive testing and imaging. He has been a Technical Program Committee member of the IEEE Ultrasonics Symposium since 1987 (the Chair of the Sensors, NDE, and Industrial Applications Group, 2004–2013); Associate Editor of the IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control since 1994; Lead Guest Editor of the IEEE Ultrasonics, Ferroelectrics, and Frequency Control (UFFC) Special Issue on Ultrasonics and Ferroelectrics (August 2014); Lead Guest Editor of the IEEE UFFC Special Issue on Novel Embedded Systems for Ultrasonic Imaging and Signal Processing (July 2012); and Lead Guest Editor of the Special Issue on Advances in Acoustic Sensing, Imaging, and Signal Processing published in the *Journal of Advances in Acoustics and Vibration*, 2013. He was the General Chair of the 2014 IEEE Ultrasonics Symposium in Chicago. He has served as the IEEE UFFC Ultrasonics Awards Chair since 2018. He served as the Ultrasonics Vice President for the IEEE UFFC Society (2014–2017). He has over 340 publications and has supervised 35 Ph.D. dissertations and 22 M.S. theses to completion. He received the 2007 University (Illinois Institute of Technology) Excellence in Teaching Award. He is an IEEE Life Fellow for contributions to ultrasonic signal processing for detection, estimation, and imaging.