# "Listen2dRoom": Helping Blind Individuals Understand Room Layouts

**Myounghoon Jeon**
School of Psychology,
Georgia Institute of Technology
Atlanta, GA, 30332 USA
mh.jeon@gatech.edu

**Nazneen Nazneen**
School of Interactive Computing,
Georgia Institute of Technology
Atlanta, GA, 30332 USA
nazneen@gatech.edu

**Ozum Akanser**
School of Psychology,
Georgia Institute of Technology
Atlanta, GA, 30332 USA
ozum@gatech.edu

**Abner Ayala-Acevedo**
School of Electrical and Computer Engineering,
Georgia Institute of Technology
Atlanta, GA, 30332 USA
abner.ayala@gatech.edu

**Bruce N. Walker**
Sonification Lab, School of Psychology & Interactive Computing,
Georgia Institute of Technology
Atlanta, GA, 30332 USA
bruce.walker@psych.gatech.edu


*CHI'12*, May 5–10, 2012, Austin, Texas, USA.
ACM 978-1-4503-1016-1/12/05.

## Abstract
Over half a million Americans are legally blind. Despite much effort in assistive technology, blindness remains a major challenge to accessibility. For individuals who are blind, there has been considerable research on indoor/outdoor way finding, but there has been little research on room layout information. The purpose of the current research is to support blind individuals to understand the layout of an unfamiliar room. We found some important applications for this type of assistive technology such as safety, easy-to-use furniture and home appliances. To this end, we identified user needs and variables with blind participants, designed and evaluated prototype systems, and iteratively improved the system. The overall process, findings, and on-going future works are discussed. This effort is expected to enhance independence for persons who are blind.

## Keywords
Room Layout Information; Computer Vision; Auditory Display; Assistive Technology; Blind Individuals

## ACM Classification Keywords
H.5.2.1 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – interaction styles, user-centered design, multimedia information systems

K.4.2 [Computers and Society]: Social Issues – assistive technologies for persons with disabilities

## General Terms

Design, Experimentation, Human Factors

## Introduction

Over half a million Americans are legally blind [1]. Even though researchers have made great strides in developing assistive technology for persons who are blind, blindness still remains a priority for accessibility researchers [2], especially as many persons who are blind seek to go to school or to work. There have been many research projects about wayfinding for blind pedestrians [e.g., for outdoor navigation 3, for indoor navigation 4]. And once a person who is blind is sitting at a desk or computer workstation, there are accessibility tools to help them access their data, such as screen readers [5] like JAWS, Window-Eyes, or Apple's VoiceOver. However, ***between*** the task of navigating to a particular place or room, and getting down to work, it remains a challenge for that person to understand the layout of the space, determine what is in the room, and where, and then navigate through the space to their desk, for example. A few limited object-recognition systems have made use of computer vision [e.g., 6], but there have been few studies on providing room layout information. When entering both familiar and unfamiliar rooms, blind individuals have to either depend on others' explanation or explore the room themselves. Thus, providing room layout information is important for safety, use of furniture, and also use of home appliances.

In this project, we have been investigating various ways of supporting individuals who are blind to understand the layout of an unfamiliar room. To this end, we identified user needs and critical variables with blind participants, designed application concepts, evaluated our prototype systems, and created improved designs. Two major research questions were addressed: (1) What is the best way to scan the room? (2) How should the system translate the spatial information into non-visual modalities? Across these two questions, *modality* and *form factor* issues were considered. This participatory design effort is expected to facilitate *acclimation* and *adjustment* to an unfamiliar room and ultimately increase user independence.

## User Needs Analysis

*Procedure*

Following brainstorming sessions, we conducted user needs analysis with two totally blind participants in separate sessions (2 males; age = both above 40 years; both above undergraduate degree). We conducted semi-structured interviews to let them talk more comfortably with the questions.

*Results*

Through the discussion with users, we obtained insights for our system design. Both participants mentioned that a system to help find the necessary objects/items in an (1) *unfamiliar environment* would be more useful for them. They expressed frustration that comes with traveling, such as hotel or conference rooms. Participants also suggested that using (2) *speech sounds* would be easier and faster to learn than using non-speech sounds. Also, for people who are not familiar with the use of a smart phone "camera", participants suggested using a (3) *wearable camera* as an alternative input option. One more issue was that they have to carry too many devices for different purposes. One of the participants listed his assistive

Figure 1. The spatial auditory display. Objects on the left are heard from the left side and objects on the right are heard from the left side.

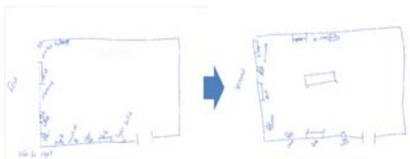


Figure 2. Drawing exercise. The drawings made with toe the directional audio output were more precise than those with the linear version.

technology items that he always carries as follows: a laptop, Braille note taker, talking PDA, book-reader, and iPhone. Therefore, they would prefer our system to be an (4) *"app"*. Both of them used their iPhone fairly well. Based on all of those suggestions, we decided to create an app so that it would reach more users and not require them to purchase and carry another device.

*Prioritization of "Objects of Interest"*

The first approach to create our initial prototype was to use Q/R codes to represent the items in the room to be detected with a smart phone camera. To figure out most important and frequently used objects that people want to know when they are in an unfamiliar room, we asked 12 sighted participants (all graduate and undergraduate students) to list such items in an office room. As a result, we identified the most common items from their lists: computer, chair, desk, door, light switch, microwave, cabinet, power outlet, printer, table, trash can, and window. Using this list, we made several speech outputs to demonstrate our system.

## Initial Prototype Design

*Input Modes*

The system has two input modes: the overview and details-on-demand. We named them after the visual information-seeking mantra [7]. The overview mode would be used to give a user an overview of the objects in the room. The user scans the room from left to right and the audio output lists all the key recognized items in the room sequentially. In the details-on-demand mode, the user can learn specific items in a specific direction. The user would direct the smart-phone camera or a wearable camera in the direction that he or she wants to get information and the system would provide an audio output of the specific items. After doing a scan, the user has the option to save the scan by giving it a specific name using voice commands, or the system would save it automatically with date and time.

*Audio Outputs*

The speech sounds were generated in a female voice via text-to-speech (TTS) using the AT&T Labs TTS Demo program with the 'Crystal-US-English' female voice (http://www.research.att.com). The sounds were edited in Cool Edit Pro 2.0. As an audio output, we came up with three different design alternatives: (1) *Linear*: Objects are listed by the order they were scanned from left to right (e.g., "table, trash can, chair"). (2) *Directional*: Objects are grouped and listed by their location with respect to the room walls and room center. We also envisioned the system to recognize landscape and portrait tables, if any (e.g., left wall: table, trash can, chair; room center: portrait table, etc). (3) *Spatial*: Objects on the left would be heard from the left side, and objects on the right would be heard from the right using a headset (see Figure 1).

To test our audio concepts, we conducted an assessment of drawing a room based on the alternative audio outputs. Three graduate students (2 females; 1 male) listened to the linear and directional audio outputs played via laptop speakers and drew the room layout. The drawings made with the directional audio output were much more precise compared to those with the linear version (see Figure 2).

## Concept Validation and Exploration

*Procedure*

We validated our prototype systems and explored further directions with key stakeholders including three totally blind individuals (2 females and 1 male; mean age = over 40; all employee) and an assistive technology specialist at the Center for the Visually Impaired (CVI) in Atlanta, GA. We conducted one-on-one sessions with each participant. Blind participants gave their consent in the presence of an impartial sighted witness. The session consisted of these six steps: (1) *Concept introduction;* (2) *User needs identification* (preferences, usage context, pros and cons, etc); (3) *Objects of interest;* (4) *Current practices of their assistive technologies;* (5) *Technology probes* (①*Input hardware*: smart phone camera vs. wearable camera, see Figure 3 (a), ②*Input software*: VoiceOver, mode selection, and gesture-based interaction, etc, ③*Scan experiment*: to detect various objects using attached Q/R tags with a smart phone camera, ④*Output hardware*: bone-conduction phones vs. headphones, ⑤*Output software*: e.g., use of proposition “Microwave on the table” vs. interval/pause “Microwave-Table”, ⑥*Audio experiments:* linear, directional, and *spatial* sounds); and (6) *Open discussion*.

*Results*

Notes taken during the study and transcribed video data were analyzed by creating thematic connections using a data-driven approach. Here, after describing users’ current practices, we briefly present *design considerations* for the system design.

*Current practices for room navigation and layout finding.* We found that blind individuals employ various strategies to find room layout based on their context. One participant mentioned “*…we listen to the room, such as refrigerator sound, running computer hardware.*”. Another approach is to find the room layout by touching various objects or with a cane while navigating across the room, especially when there is no one in the room. Other strategies include asking other sighted people. Participants agreed that there is a need for such a system that can help them understand room layout and for tracking objects in a room. Participants favored our ideas, mentioning that “*Yes, it could be useful, just depends upon the person’s level of vision and their mobility,*” “*I would use it when I travel …to unfamiliar places, new places…”*

*Current use of camera.* Participants mentioned that they use their phone cameras to take pictures and record videos. They do it to share their experiences and get feedback from their sighted friends and family members (e.g., color of clothes in the picture). However, it seemed difficult for one participant to detect Q/R tags using a smart phone camera.

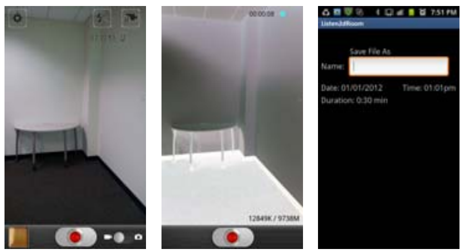
*Objects of interest to be identified.* Participants noted various objects they want to identify when they are in a room. The important objects were tables, chairs, cabinets, closets, counters, chairs, office equipments, trash cans, and phones.

*Design Considerations*

**Wearablility (physical and social comfort).** We found that it is important to take care of physical and social comfort when considering a wearable device. Not only size and weight of the device, but also look and feel of the device matter. One of the female participants highlighted that, *“It’s not heavy but it is bulky …, it should fit to my sun glasses or earrings….*” Moreover, one of the participants raised a concern that a wearable device can cause a social barrier.

(a) (b)

Figure 3. (a) Listen2dRoom main view and Looxcie 2 wearable camera. (b) Main view with menu selected.



(a) (b) (c)

Figure 4. (a) Scanning mode before starting. (b) Scanning mode on progress. (c) Save recording.

**Support Inference.** Participants identified various types of information for presenting details about the room. For instance, it would be useful to mention left/right/center, even far left, or immediately left, propositions, distance, and size of room in the audio description. An alternative way could direct the user using *clockwise* directions (e.g., table at 10 o'clock).

**Customization.** All participants emphasized the importance of a fully customizable system (e.g., level of detail, speed of audio, etc). One participant stated that "*…it needs flexibility…because every blind person is not the same. I think I should be able to modify it.*"

**Simplicity.** A very important consideration is that the system should be easy to use and simple.

**Autonomy.** The system should be designed such that it encourages user independence. The user should be able to use the system without asking anyone for help.

**Cost.** It is critical to keep the cost minimal. It can be achieved by allowing users to use an app on the existing device. Users can download the app and it is optional to use the wearable camera for input and bone-conduction headphones for output, which our participants preferred.

**Ecological integration.** The most important design consideration is related to promoting ecological integration. Because our system is designed for blind users, it is imperative that it should not only be easy to use, but it should be designed in a way to fit as seamlessly as possible into the users' existing routine. Important issues include minimum maintenance and graceful recovery.

## Enhanced System Design

Based on the user needs analysis and concept validation sessions, we concluded that the system could be a smart phone app. Even though not all blind users have a smart phone, the smart phone market is rapidly increasing. Considering those who are not familiar with the use of camera of the phone, our system allows users to choose either the camera on the phone or a over the ear wearable camera (Figure 3 (a)). Because all participants desired a customizable system, our app provides different output modes on the settings. Also, they could replay a saved recording of each scan. Figure 3 shows the main view of our app, "Listen2dRoom". To interact and navigate through the app, we use a gesture based approach that is compatible with that of Apple's VoiceOver, to ensure a short learning curve for new users.

### *Main Menu*

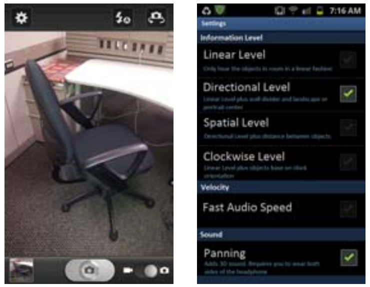
The main menu provides three options: scanning mode (overview), object recognition mode (details-on-demand), and recordings. To navigate the main menu, users use a one-finger flick right or left to move through the options with spoken outputs. One-finger tap speaks the current item and two-finger tap selects the desired item. To go to the menu seen in Figure 3 (b), users make a four-finger flick down.

### *Scanning Mode*

Once users select the scanning mode, they will be presented with the video camera view of the smart phone. Then, users can double-tap whenever they are ready to begin the room scanning from left to right, and they will double-tap again to let the system know they are done with the current scanning. After users finish scanning the room, they can then listen to the


(a) (b)

Figure 5. (a) Recordings list. (b) Recording audio.




(a) (b)

Figure 6. (a) Object recognition mode. (b) App settings.

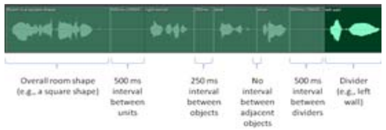



Figure 7. There are several predefined intervals (e.g., 0ms for adjacent objects, 250 ms for objects, 500ms for different units or dividers, etc).

room structure, which is the *directional* mode by default. Additionally, in a new version, there are several predefined intervals between different speech outputs in a more systematic way (e.g., 0ms for adjacent objects, 250 ms for objects [8], 500ms for different units or dividers. See Figure 7). Then, it will take users to Figure 4 (c) to save the recording.

*Recording Mode*

In the recording mode, Figure 5 (a), users can look through all saved files and re-listen to a specific scan. The recording view shows the recordings by date or by name. Users can also scroll to the window by using three-finger flick up or down, and go to the menu by using four-finger flick down. Once users select a specific recording by one-finger double tap, they are then presented with a graphical view of that recording, Figure 5 (b) (though not intended to be useful for totally blind users). Users can one-finger double tap to start listening to the recording and use two-finger flick left or right to forward the audio playback to the portion of the audio they wants to listen to.

*Details-on-Demand (Object Recognition) Mode*

Lastly, there is the object recognition mode, in which users can point the camera toward a specific object and the app will speak out the object type (see Figure 6 (a)). In contrast to the scanning mode, once this mode is selected, users will be presented with the camera mode already engaging.

*Settings*

For the fully customizable system, we provided users with a setting menu they can access from any page. Users can manipulate the amount of information and speed of the audio, with spearcons [8] or sped-up speech mode (see Figure 6 (b)). Participants can select a spatial mode as an option (not default) because they might sometimes wear headphones backwards.

## Conclusion and Future Works

We introduced the overall procedure of the participatory design of a new smart phone app for blind individuals to understand room layout. This iterative process allows us to design a system based on users' needs addressing their capabilities and limitations more accurately, which leads to a fully accessible system for blind individuals. Our next step is to implement the enhanced version app on iOS with a more sophisticated computer vision technology (e.g., OpenCV) and evaluate the accuracy and usability in a real context.

## References

[1] Glinert, E. P. and York, B. W. Computers and people with disabilities. *ACM Transactions on Accessible Computing*, 1, 2, 2008, 1-7.
[2] Newell, A. F. Accessible computing: Past trends and future suggestions. *ACM Transactions on Accessible Computing*, 1, 2, 2008, 9:1-7.
[3] Walker, B. N. and Lindsay, J. Development and evaluation of a system for wearable audio navigation. *HFES,* Orlando,2005.
[4] Ali, A. M. and Nordin, M. J. Indoor navigation to support the blind person using weighted topological map. *the International Conference on Electrical Engineering and Informatics*, Selangor, Malaysia, 2009.
[5] Asakawa, C. and Itoh, T. User interface of a home page reader. *ASSETS'98*, Marina del Rey, CA, USA, 1998.
[6] Hub, A., Diepstraten, J., and Ertl, T. Design and development of an indoor navigation and object identification system for the blind. *ASSETS'04*, Atlanta, GA, 2004.
[7] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualization. *IEEE Symposium on Visual Language*, Washington, USA. pp. 336-343. 1996.
[8] Walker, B. N., Nance, A., & Lindsay, J. Spearcons: Speech-based earcons improve navigation performance in auditory menus. *Proceedings of the International Conference on Auditory Display,* London, England. pp. 63-68. 2006.