การใช้เสียงเพื่ออธิบายฉากที่มีสิ่งกีดขวางอยู่นิ่งและเคลื่อนที่

นายกวิน เมศร์ศิริตระกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2560

USING SOUND TO DESCRIBE SCENES WITH STILL AND MOVING OBSTACLES

Mr. Kawin Metsiritrakul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2017

Thesis Title	USING SOUND TO DESCRIBE SCENES WITH STILL AND MOVING OBSTACLES
By	Mr. Kawin Metsiritrakul
Field of Study	Computer Engineering
Thesis Advisor	Assistant Professor Proadpran Punyabukkana, Ph.D.
Thesis Co-Advisor	Associate Professor Atiwong Suchato, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

....Dean of the Faculty of Engineering
(Associate Professor Supot Teachavorasinskun, D.Eng.)

THESIS COMMITTEE

....Chairman
(Assistant Professor Thanarat Chalidabhongse, Ph.D.)

....Thesis Advisor
(Assistant Professor Proadpran Punyabukkana, Ph.D.)

....Thesis Co-Advisor
(Associate Professor Atiwong Suchato, Ph.D.)

....External Examiner
(Sirinart Tangruamsub)

กวิน เมศร์ศิริตระกูล : การใช้เสียงเพื่ออธิบายฉากที่มีสิ่งกีดขวางอยู่นิ่งและเคลื่อนที่ (USING SOUND TO DESCRIBE SCENES WITH STILL AND MOVING OBSTACLES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.โปรดปราน บุณยพุกกณะ, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร.อติวงศ์ สุชาโต, หน้า.

ข้อมูลที่มนุษย์รับรู้จากภาพและเสียงมีความคล้ายคลึงกัน ด้วยความสัมพันธ์ของข้อมูลเหล่านี้ จึงมีงานวิจัยมากมายนำความคิดนี้ไปพัฒนางานเกี่ยวกับการอธิบายข้อมูลเชิงพื้นที่ด้วยเสียง งานวิจัยก่อนหน้านั้นมีการอธิบายข้อมูลภาพเชิงพื้นที่ด้วยเสียงโดยใช้ข้อมูลทิศทางและระยะทางของสิ่งกีดขวางที่ใกล้ที่สุดเท่านั้น เพื่อให้ผู้ใช้รับรู้และเข้าใจข้อมูลเชิงพื้นที่ในฉากมากขึ้น เราจึงเสนอการวิธีการอธิบายข้อมูลภาพเชิงพื้นที่เป็นเสียงโดยใช้ข้อมูลทิศทาง ระยะทาง ขอบเขต และความเร็ว พร้อมกับอธิบายข้อมูลเชิงพื้นที่ของสิ่งกีดขวางหลายชิ้นพร้อมกัน ผู้ทดสอบเข้ารับการทดสองโดยการควบคุมตัวละครและหลบหลีกสิ่งกีดขวางไปยังจุดหมายให้ได้ผ่านเกมจำลอง พวกเขาต้องใส่ผ้าปิดตาและฟังเสียงสามมิติของสิ่งกีดขวางรอบ ๆ ตัวแทนการมอง ผลการทดลองพบว่าผู้ทดสอบสามารถหลบหลีกสิ่งกีดขวางได้ดีกว่าการฟังเสียงแบบพื้นฐาน 2 จาก 3 ฉาก โดยมีอัตราการหลบหลีกอยู่ที่ร้อยละ 85.33 และ 95.33 ตามลำดับ

ภาควิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต	....
สาขาวิชา	วิศวกรรมคอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....
ปีการศึกษา	2560	ลายมือชื่อ อ.ที่ปรึกษาร่วม	....

5970105021 : MAJOR COMPUTER ENGINEERING

KEYWORDS: AVOIDANCE / NAVIGATION / SPATIAL INFORMATION / AUDITORY INFORMATION / STEREO SOUND / MOVING OBSTACLE

KAWIN METSIRITRAKUL: USING SOUND TO DESCRIBE SCENES WITH STILL AND MOVING OBSTACLES. ADVISOR: ASST. PROF. PROADPRAN PUNYABUKKANA, Ph.D., CO-ADVISOR: ASSOC. PROF. ATIWONG SUCHATO, Ph.D., pp.

The information human get from sight and hearing sense are similar. With these correlations, there are many works that take this idea to describe the spatial information with sound. The previous works only described the nearest obstacle with the direction and distance information to sound. To let the user clearly understand the spatial information in the scene, we aim to develop the describing solution from spatial information to auditory information including direction, distance, boundary and speed info with the multiple obstacle descriptions. The participants took the experiment by controlling the character and avoiding the obstacles in the different scenarios to the destination via simulator. They were also asked to wear blindfold and listen to the auditory information of the obstacles around them instead of using sight. The results show that the participants got 2 out of 3 scenarios higher performance than the baseline auditory information with avoidance rate 85.33% and 95.33%.

Department:	Computer Engineering	Student's Signature ....
Field of Study:	Computer Engineering	Advisor's Signature ....
Academic Year:	2017	Co-Advisor's Signature ....

ACKNOWLEDGEMENTS

I would especially like to thank Asst. Prof. Proadpran Punyabukkana, Thesis Advisor, and Assoc. Prof. Atiwong Suchato, Thesis Co-Advisor. They have taught me more than I could ever give them credit for here. They have shown me, by their example, what a good scientist should be. They also have been supportive of my goals and worked actively to provide me with the protected academic time to pursue those goals.

I am grateful to Surapol Vorapatratorn, Chonlathorn Kwankajornkiet and Kailerk Treetipsounthorn, the members in Spoken Language System (SLS) Lab, who provided me professional guidance and taught me a great deal about both scientific research and life in general.

This work would not have been possible without the financial support of the Junior Science Talent Project (JSTP) by National Science and Technology Development Agency (NSTDA).

Nobody has been more important to me in the pursuit of this project than the members of my family. I would like to thank my parents, whose love and guidance are with me in whatever I pursue. They are the ultimate role models.

CONTENTS

Page

Chapter 1

Introduction

1.1 Motivation

From the five senses of human including sight, hearing, taste, smell and touch. The visual perception is the most sense human uses in daily life. In additional, there is more information can be extracted by the vision than the other senses. With the two human eyes, the brain can provide each object information around the environment containing distance, direction, shape and color in one shot of the glance. Human also knows how quick the object moves when the brain processes multiple shots of what the human saw previously.

Besides the visual perception, the second sense that human most uses is hearing sense. This sense serves the auditory information with the source direction and distance. Moreover, it can also tell human which direction the source is going and how fast it is. The bats showed they can generate the image from the sound using echolocation [1]. The sound wave is sent from the bats and echoes the obstacles back to their ears. Then, they process the echo sound and decide which way they should fly through. Another example, the blind people show that they can know the environment around them by clicking their tongue and listening to the clicking echoed sound [2].

Obviously, the information provided by sight and hearing perception are similar. With these information, there are many video game companies develop the blindfolded games. The aim of these games is to listen the sound, imagine the situation and take the action. For example, Blindfold Runner game [3], the users must listen to the obstacle sound and decide which direction they should go or dodge while the character in the game is running. The Sixthsense 3D Sound Horror Shooting mobile game [4] was also developed by using sound to describe the scene. The users take the part of the police in the dark forest. They must listen which way the monsters are coming, turn the mobile direction to those monsters and tap the screen to shoot them. One of the solutions to create these games, firstly, is designing visual scenario. Then, the system transforms these visual things into sound format.

The works previously mentioned only transform direction and distance information from image to sound. To the best of our knowledge, most of the works play only the nearest obstacle sound. Some of them spawn the obstacle one by one. That means there is only one obstacle in the scene all the time.

The image not only give information about direction and distance, but it also gives speed and boundary information [5]. To let the user clearly understand the scene, we propose to develop the describing spatial information using sound with spatial information which is more than just distance and direction information.

1.2 Objective

To use sound to describe spatial information including direction, distance, boundary and speed info for avoidance and navigation task.

1.3 Scope and limitation

1.3.1 Scenario

- Scene with still obstacles
- Scene with moving obstacles
- Scene with still and moving obstacles

1.3.2 Obstacle

- Still obstacle
- Moving obstacle

1.3.3 Input

- Scenario dataset including position, boundary and speed of the all objects in front space of users
- The obstacle distance should be no more than 4 meters.

1.3.4 Output

- Stereo sound

1.3.5 Output Device

- Bone Conduction Headphone

1.4 Contribution

Develops and designs multiple obstacle sound representation more clearly from spatial information for navigation and avoidance. Thus, multiple obstacle sound with various information describing was developed to improve the existing works.

1.5 Research Plan

1.5.1 Preparation

We first studied previous works about transformation from spatial information to audio and characteristic information for navigation and avoidance. Then, we considered the advantage and limitation of each work.

1.5.2 Investigation

We carried out the sandbox testing by designing different kinds of audio to represent the spatial information and tested with the users. Then, we investigated the best performance of each audio that represents each spatial information.

1.5.3 Implementation

We followed this step by the methodology in chapter 3. The processes started with the input simulator development. Then, we developed sound reproduction parts. Lastly, sound rendering process (Merge and play) was proceeded.

1.5.4 Analysis and summary

We did the experiments with the participants and analyzed the results. The experiments were divided into 2 parts including baseline testing and proposed method testing.

1.6 Conference

Publication Title:

Obstacle Avoidance Feedback System for the Blind using Stereo Sound

Conference Proceedings:

The 11th International Convention on Rehabilitation Engineering and Assistive Technology (i-Create 2017)

Chapter 2

Literature Review

There are many researchers tried to investigate the different ways to convert visual information to audio. Politis, D., et al. [6] presented the transforming mechanism between music and colours. The 2D image was pixelized and segmented and then the music was composed by transforming each colour pixel element into note.

In 2015, Python released a tool called Pyc2Sound [7]. This tool can convert the drawing line into sound. The drawing line was mapped onto x-y correlation where x-axis represents time and y-axis represents frequency. Then, the drawing line was tracked and segmented into pieces of the edge which were sorted by time domain. Finally, the sound was synthesised from each edge and connected to each other sequentially.

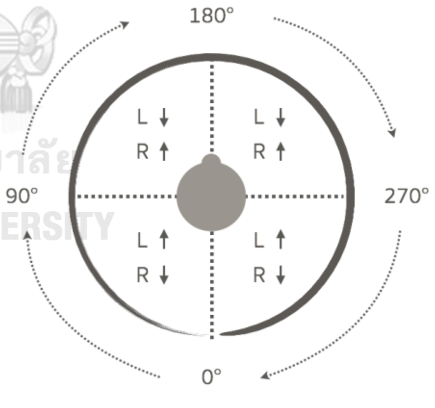

Figure 1. Headphone sound

The two researches as above-mentioned only give monotone feedback with the two-dimensional input. Zwinderman, M., et al. introduced the navigation system by transforming the spatial information of the target location to 3D-audio [8]. Balancing the signal intensity between the left and the right headphone give an indication about direction. The further sound position is, the lower the sound volume. The maximum volume in the left channel of headphone occurs when the angular distance compared with the head position is 90 degree and the minimum volume in the left channel of headphone exists when the angular distance is 270 degrees (Figure 1). Luca, Bruno and Thomas also support about the decreasing volume [9]. The research advised that sound source's intensity will fall by 6 dB every double distance. However, it is not clear what is the best model for distance-dependent intensity scaling.

The next study, the researchers investigated the way to represent echo sound from ultrasonic [10]. The echo signal was processed, and the system produced the wave form that described the object characteristics including distance, type (solid / not solid), and location. The result shows that the object and position were the only significant factors while the interaction between object-position and distance-position were also significant.

There is the different way to describe spatial information. Bower, C., et al. designed auditory feedback of the environment colours using orchestra [11]. The colours were represented by various instruments, different pitch, and chorus roles. Stereo sound was used to define the object location by convolving the colour sound with head-related impulse response (HRIR) which varies in azimuth, elevation, and distance [12]. HRIR is an approach to simulate the sound like what human ear hear. The stereo sound lets the user realise where the object source location is. The researchers produced spatial sound with the use of the HRIR measurements belonging to the CIPIC database (Figure 2) [13]. For the first experiment, the participants were blindfolded, and they were asked to pair sock pairs by listening to the instrument sound. The result showed that the participants could match coloured socks with high accuracy. The researchers had done another experiment by asking participants to listen the instrument sound and follow the red serpentine painted on the ground. They compared their approach with previous works which represented feedback sound using only depth data. Their feature map with spatial information allows for a better object detection.

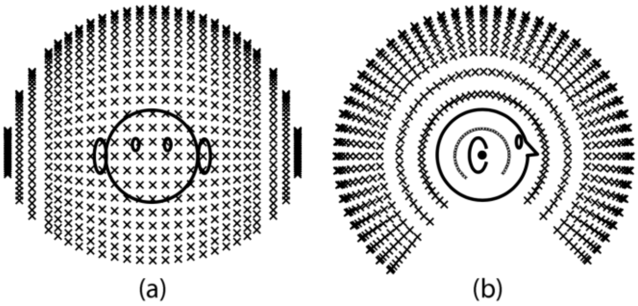

Figure 2. Location of CIPIC data points (a) front (b) side.

The converting obstacle information to audio methods are widely used for blind navigation. Dunai, L., et al. presented a prototype for being used as a travel aid for the blind [14]. The system turned obstacle information from camera into audio information. The obstacle distance was inversely proportional to frequency. The lowest frequency represented the furthest location. The obstacle direction was transfer to stereo sound using HRIR. Thus, the blind users could know the obstacle position around them by listening to the sound synthesis. Moreover, the system also provided the pitch changing process that blind users could notice how quick the obstacle is. The last review is still in the blind navigation field. The researchers developed the sound feedback using a polar grid representation [15]. This approach is the most appropriate representation for obstacle avoidance in robotics (Figure 3) [16].

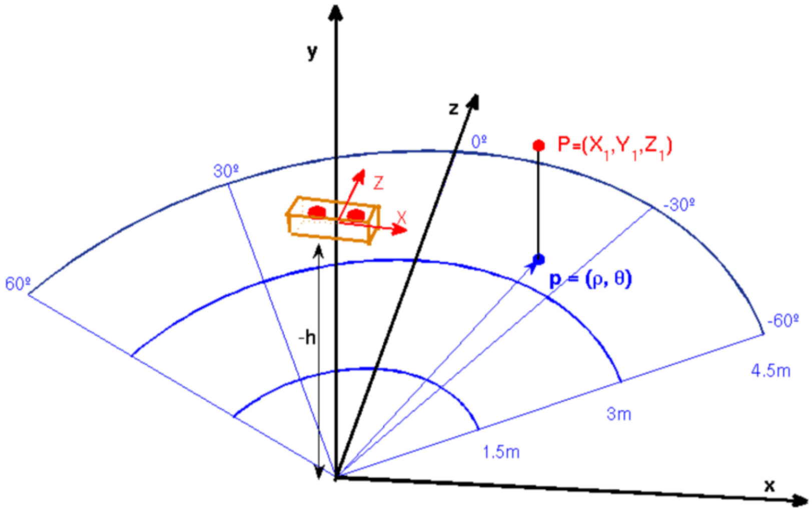

Figure 3. Polar grid.

The system segmented the area in front of the user into the polar grid (Figure 4). Each small piece of grid describes the different sound (Table 1). The distance and direction information were used to process the sound. In the same way as other studies, the direction was represented by stereo sound using HRIR. If the obstacle detected by the system drops into which piece of grid, the sound represented by that grid channel will be noticed to the user. For the experiment, the participants were asked to hold the camera and walk along the path. Then, they had to detect the obstacle location by listening to synthetic sound. The result showed that the participants got the high accuracy when the system warned them the obstacle location.

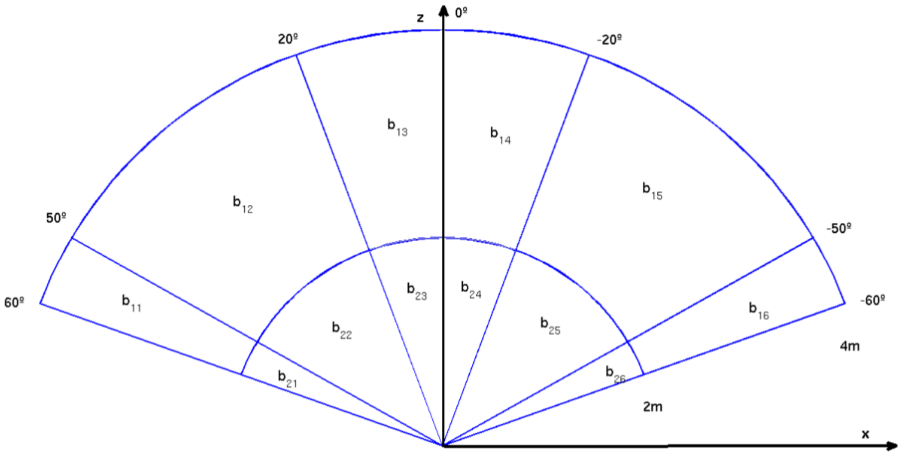

Figure 4. Redesigned polar grid.

	4 – 2 m	**2 – 0 m**
[−60°, −20°]	f = 261, l = 50, r = 1, c = right	f = 261, l = 50, r = 1, c = right
[−20°, 0°]	f = 2,637, l = 50, r = 1, c = right	f = 2,637, l = 50, r = 4, c = both
[0°, 20°]	f = 2,637, l = 50, r = 1, c = left	f = 2,637, l = 50, r = 4, c = both
[20°, 60°]	f = 261, l = 50, r = 1, c = left	f = 261, l = 50, r = 1, c = left

Table 1. Configuration of acoustic warnings corresponding to the deployment in Figure 4.

The most popular information which were used for audio synthesis are direction and distance information. However, the researches as above-mentioned provided only some spatial information. There are other compositions we should concern including obstacle speed and obstacle boundary. These compositions should be integrated to audio information to provide more comprehensible information.

Chapter 3

Methodology

3.1 System Overview

We first created a simulator which provide the spatial information of the obstacle using Unity3D. The system was created using MATLAB to synthesis the spatial sound. Unity3D is responsible for passing each obstacle information including position (direction and distance), boundary and speed to the system. Then, the system starts with sound selection process, boundary generation process and speed synthesis process in sequence. Finally, the spatial sound is generated and return the sound through user’s headphone.

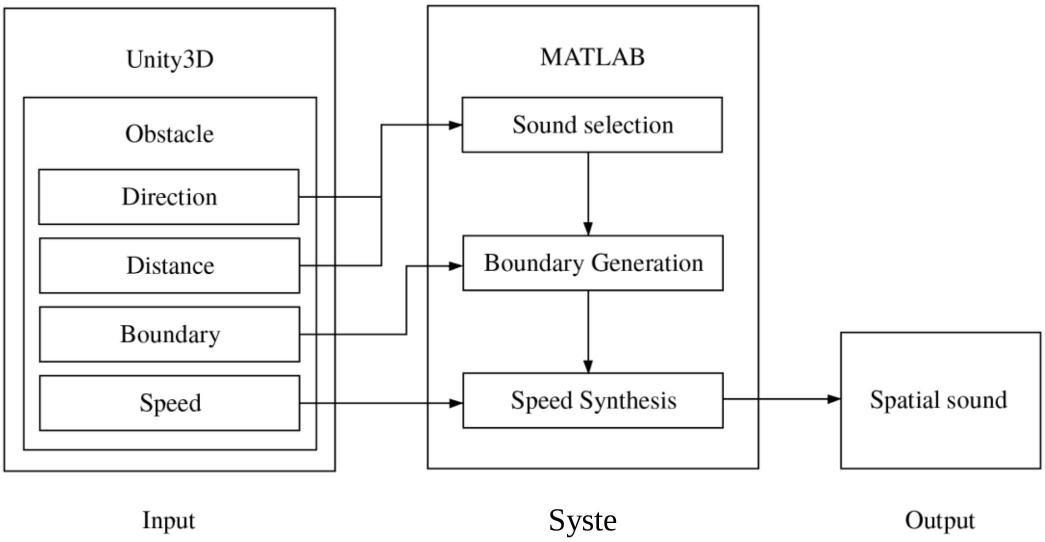

Figure 5. System overview flowchart

3.2 Procedure

3.2.1 Input simulator

We first created a simulator which provide the spatial information of the obstacle using Unity3D (Figure 6). The direction, distance, boundary and speed data of each obstacle seen by the character in the simulator are calculated and passed to the system in real-time. Figure 6 (left) shows the overall map simulation and Figure 6 (right) shows in the view of character.

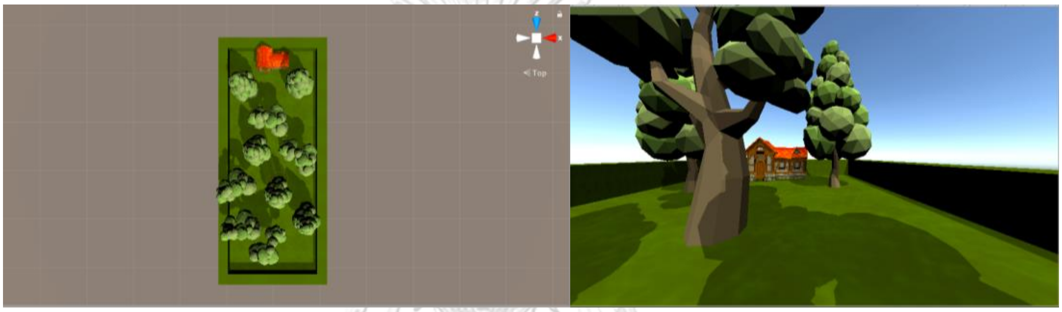

Figure 6. (left) Overall map, (right) view of character.

3.2.2 Audio preparation

Before starting the system, the system would prepare the basic set of the spatial audio combinations including 5 different directional sounds and 2 different distance sounds.

The Middle C (C_4) note (261.6 Hz) was used to be the original sound for preparation. The directional sound was divide into 5 directions including left (-75° to -45°), middle-left (-45° to -15°), front (-15° to 15°), middle-right (15° to 45°) and right (45° to 75°). Each directional sound also had different frequency. The furthest directional sound (left and right) was original sound. The system shifted the original sound frequency up 1 key (Middle D (D_4) 293.6 Hz) for the middle-left and middle-right sound. And the original sound frequency was shifted up 2 keys (Middle E (E_4) 329.6 Hz) to be the front sound. Then, HRIR belonging to the CIPIC database was implemented to the original sound to generate each directional sound.

The system fixed the set of directional sound as above-mentioned for the obstacles which the distance was less than 1.5 meters. The system also prepared the second set of directional sound for those obstacles which has the distance between 1.5 meters to 3 meters. Every sound frequency in the second set was shifted down 1 octave. In addition, there is the third set of directional sound for the obstacle which has the distance more than 3 meters. Every sound frequency in the third set was shifted down from the original sound 2 octaves. The spatial audio polar grid overview chart is showed in figure 7.

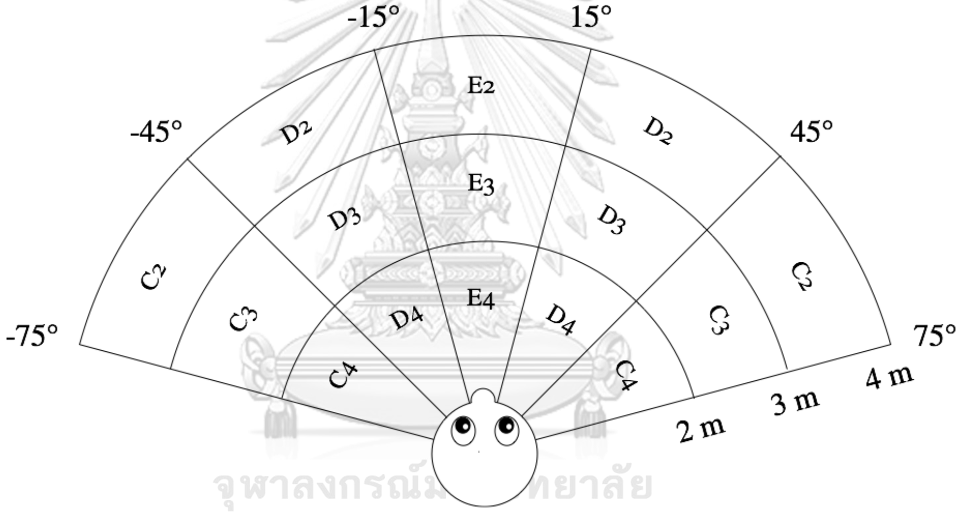

Figure 7. The spatial audio polar grid overview.

3.2.3 Sound selection

When the simulator started, the spatial information data of each obstacle would be sent to the system. Then, the system had to select which was the proper sound to play. With the distance and direction data, the system selected the sound by checking which spatial audio in polar grid match to the obstacle location. The selected sound would be transferred to the next process.

3.2.4 Adding boundary feature

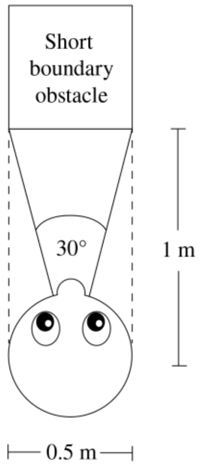

Figure 8. Short boundary obstacle.

The boundary is an angular distance between obstacle left and right edge in the view of character. We defined the obstacle which its boundary was less than the width of an average adult (0.5 m) [17] as a short boundary obstacle. These short boundary obstacles had the boundary angle less than 30 degrees when they were located 0.76 meters away from the character (Figure 8). The boundary is described by sound duration. Normally, the sound which the system selected from the previous process was about 0.1 seconds. For those obstacles which had boundary angle less than or equal to 30 degrees, the sound duration still was the same as the previous process. If the obstacle boundary angle was more than 30 degrees, this process would extend the sound duration by 0.1 seconds every 30 degrees.

3.2.5 Speed synthesis

There are 3 types of the obstacle speed including idle, slow and fast. The idle speed obstacle was defined as the obstacle which stays still. The slow speed obstacle was defined as the non-hazardous obstacle such as walking people (average 1.4m/s) [18]. The fast speed obstacle was defined as the obstacle which has risk to damage the character such as bike (average 4.3 m/s) [19] or car (16 m/s) [20]. The sound processed by system from boundary sound generation process would be made the repetition followed by the obstacle speed. Table 2 shows the relationship between sound and speed of the obstacle. The idle speed sound repeats nothing. The slow speed sound repeats once. The fast speed sound repeats twice.

Obstacle speed	Repetition
Idle speed (0 m/s)	1 times every 0.3s
Slow speed (0-4 m/s)	3 times every 0.2s
Fast speed (> 4 m/s)	5 times every 0.1s

Table 2. The relationship between sound and speed of the obstacle.

3.2.6 Merge and play

After the system had generated each obstacle sound in the scene, the final process was to play all sounds to the user. When the system received obstacle data from the simulator, the system played the sound immediately after finishing the sound synthesis process. The system played each sound continuously because sound produced continuously is better than using discontinuous sound [21]. Moreover, the system would take shorter time to return the sound to the user when using the continuous sound.

Chapter 4

Evaluation

4.1 Simulation maps and tasks

The main elements in each simulation map consisted of character, obstacles and destination. The participants were asked to listen to the spatial sound of the obstacles and control the character from starting point to the destination by avoiding all obstacles on the way.

Each simulation map provided 2 types of spatial sound. For the first round, the simulator described the obstacles by baseline sound [14] which was provided only the nearest obstacle sound with single audio feedback (1 beep per period). For the second round, the simulator described the obstacles by our proposed method. They could rest before testing in the next round.

The system also provided footstep sound effect to let the participants know how far they move the character. The footstep sound effect had different styles depending on what the character stands on. We classified the maps into 3 types including standing obstacle scenario, moving obstacle scenario and mixed obstacle scenario.

4.1.1 Scene with still obstacles

The scene with still obstacles was set as the forest scene (Figure 9). There are 10 obstacles located in this map. These obstacles consisted of 3 small obstacles (less than 30 degrees of angular distance), 4 medium obstacles (30 to 60 degrees of angular distance) and 3 large obstacles (more than 90 degrees of angular distance). Every obstacle speed was fixed as idle speed.

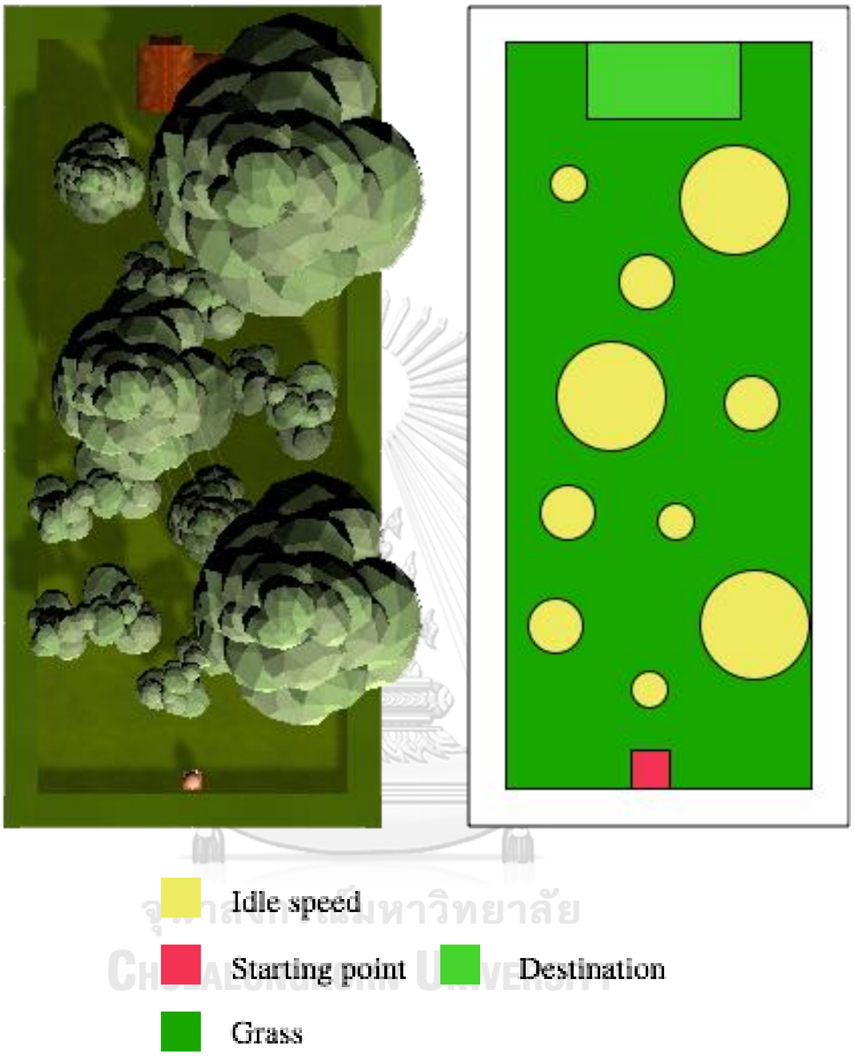

Figure 9. Standing obstacle scenario

4.1.2 Scene with moving obstacles

The scene with moving obstacles was set as the city scene (Figure 10). There were 10 obstacles located in this map. These obstacles consisted of 3 small obstacles (less than 30 degrees of angular distance), 4 medium obstacles (30 to 60 degrees of angular distance) and 3 large obstacles (more than 90 degrees of angular distance). The obstacles had different speed. The small obstacles were represented by children walking with slow speed. The medium obstacles were represented by adult people walking with slow speed. The large obstacles were represented by cars moving with fast speed.

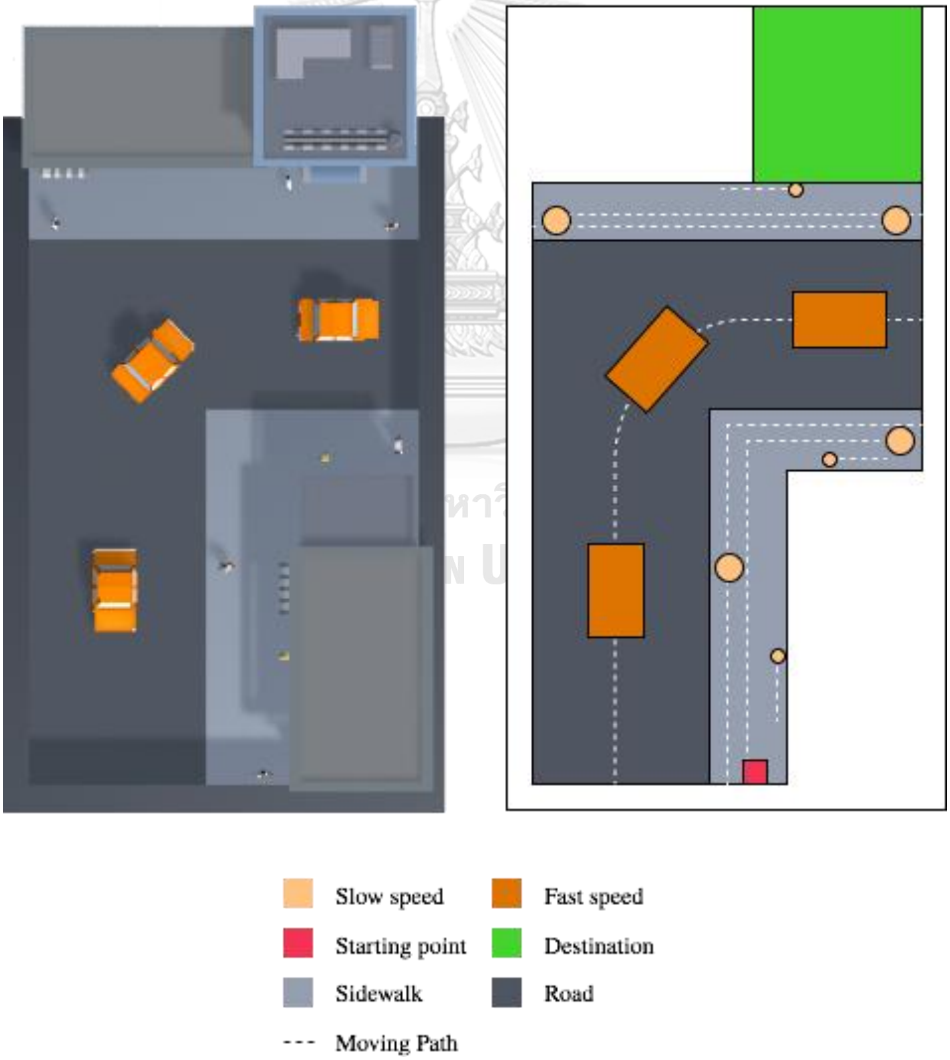

Figure 10. Moving obstacle scenario

4.1.3 Scene with still and moving obstacles

The scene with still and moving obstacles was set as the city scene (Figure 11). There are 15 obstacles located in this map. These obstacles consisted of 4 small obstacles with idle speed represented by trees, 6 medium obstacles with idle speed represented by trees and benches, 3 small obstacles walking with slow speed represented by children and 2 obstacles moving with fast speed represented by cars.

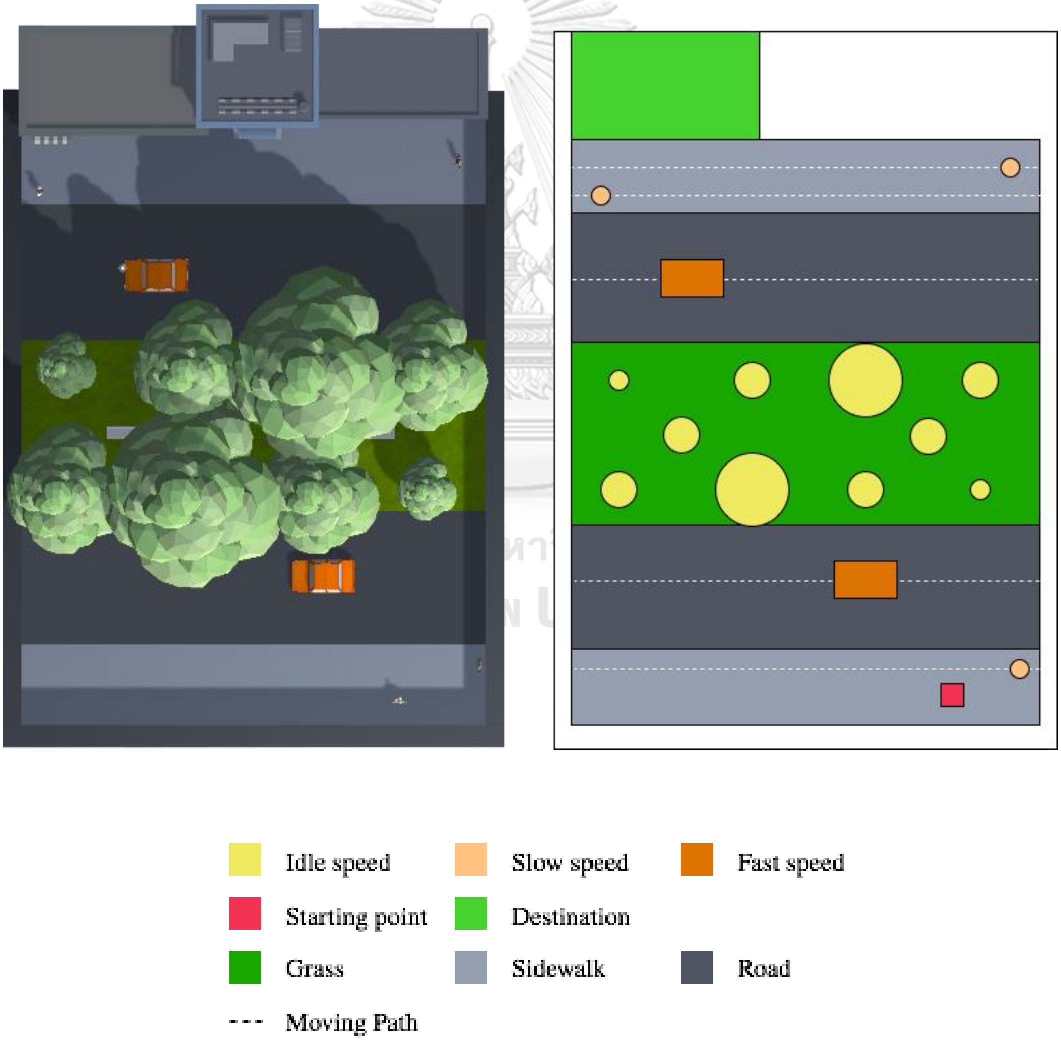

Figure 11. Mixed obstacle scenario

4.2 Tools

4.2.1 Bone conduction headphone

A bone conduction headphone is used to be the output device for listening the spatial sound. It allows the users to hear sound through the vibration on the bones of their face (jaw bones and cheek bone). The sound waves are passed through the outer and middle ear, where the eardrum is located, and directly stimulates the inner ear (hearing organ) [22]. The gain from using a bone conduction headphone is it does not plug user's auditory canal. The users can listen to the sound by a bone conduction headphone and environment sound at the same time without feeling stifle in their ear (Figure 12).

Figure 12. Bone conduction headphone

4.2.2 Gamepad

The users can control the character in simulator via wireless gamepad (Figure 13). The left stick is used to control the character direction and the right stick is used to move forward or backward.

Figure 13. Wireless gamepad F710 – Logitech

4.3 Experimental procedure

Before the experiments, we explained the meaning of the relative pitch difference, sound direction, duration and repetition to 30 participants (23 males and 7 females). Then, each participant was allowed time for practice with practice map (Figure 14). Each participant had enough practice ran with eyes open for perceiving each spatial audio information.

Then, the participants were asked to conduct the experiments following 3 above-mentioned simulation maps. They were also asked to wear blindfold while doing the tasks.

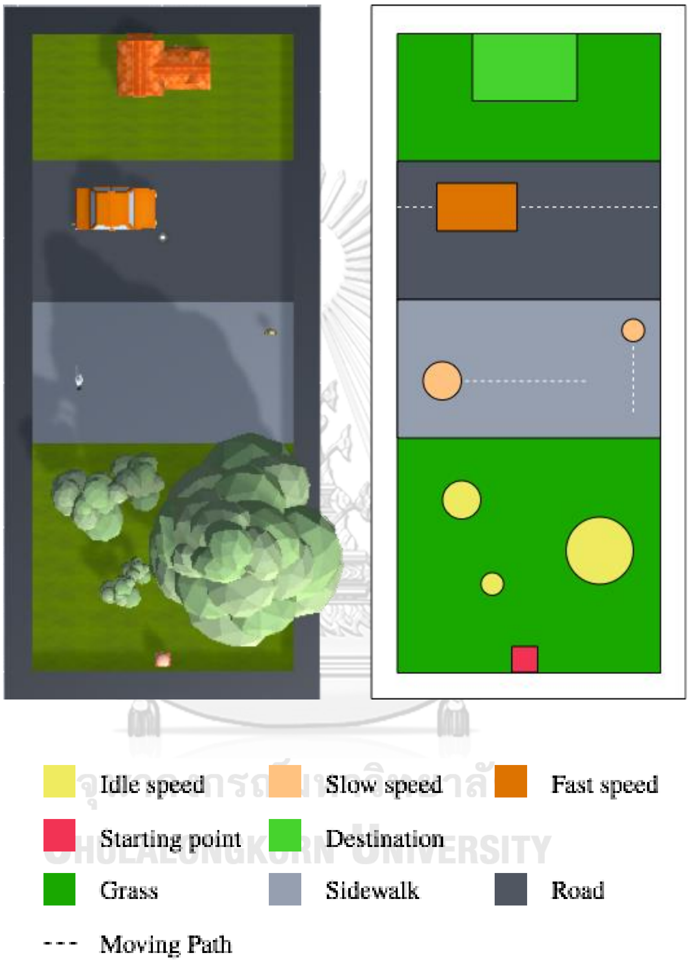

Figure 14. Practice map

4.4 Experimental condition

The number of collisions were counted when the character collided the obstacles which can be seen by him. In case of the obstacles collided the character side or back, the number of collisions weren't counted. Figure 15 shows the experimental condition when the character collides the obstacles.

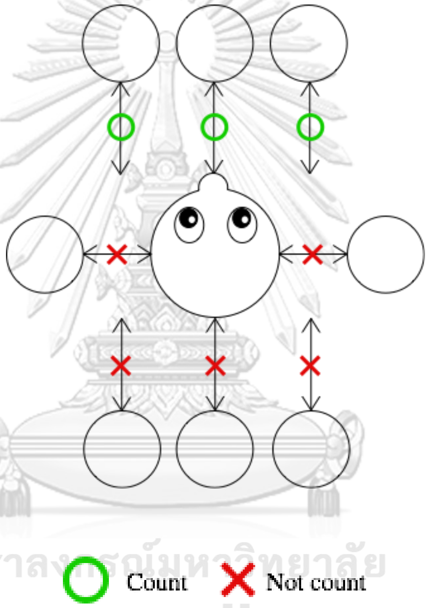

Figure 15. Experimental condition

4.5 Results

The evaluations were classified into 3 parts including size, speed and overall. The distance evaluation and direction evaluation were included in the overall part.

4.4.2 Boundary

We found that the avoidance rate classified by boundary size using proposed method was almost similar to the avoidance rate classified by size using baseline sound.

Baseline							
Size	**Scene with still obstacles**		**Scene with moving obstacles**		**Scene with still and moving obstacles**		**Average (%)**
	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	
S	3	94.44	3	88.89	5	92.67	92.12
M	4	93.33	4	79.17	6	97.78	91.19
L	3	91.11	3	84.44	4	92.50	89.35

Table 3. Avoidance rate classified by obstacle boundary size using baseline sound

Proposed method							
Size	**Scene with still obstacles**		**Scene with moving obstacles**		**Scene with still and moving obstacles**		**Average (%)**
	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	
S	3	77.78	3	93.33	5	94.00	89.39
M	4	91.67	4	82.17	6	93.89	90.48
L	3	93.33	3	78.89	4	99.17	90.46

Table 4. Avoidance rate classified by obstacle boundary size using proposed method sound

4.4.3 Speed

We found that the avoidance rate classified by speed using proposed method sound was higher than the avoidance rate classified by speed using baseline sound especially the slow speed. In the other hand, for the idle speed, the avoidance rate using proposed method sound was lower than the avoidance rate using baseline sound.

Baseline							
Speed	**Scene with still obstacles**		**Scene with moving obstacles**		**Scene with still and moving obstacles**		**Average (%)**
	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	
Idle	10	93.00	-	-	10	97.00	95.00
Slow	-	-	7	83.33	3	90.00	85.33
Fast	-	-	3	84.44	2	90.00	86.67

Table 5. Avoidance rate classified by obstacle speed using baseline sound

Proposed method							
Speed	**Scene with still obstacles**		**Scene with moving obstacles**		**Scene with still and moving obstacles**		**Average (%)**
	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	Obstacle number	Avoidance rate (%)	
Idle	10	88.00	-	-	10	95.67	91.83
Slow	-	-	7	88.10	3	92.22	89.33
Fast	-	-	3	78.89	2	98.33	86.67

Table 6. Avoidance rate classified by obstacle speed using proposed method sound

4.4.4 Overall

24 of 30 participants took time to complete the scene with still obstacles using proposed method sound shorter than using baseline sound as shown in figure 15. 18 of 30 participants took time to complete the scene with moving obstacles using proposed method sound shorter than using baseline sound as shown in figure 16. 23 of 30 participants took time to complete the scene with still and moving obstacles using proposed method sound shorter than using baseline sound as shown in figure 17.

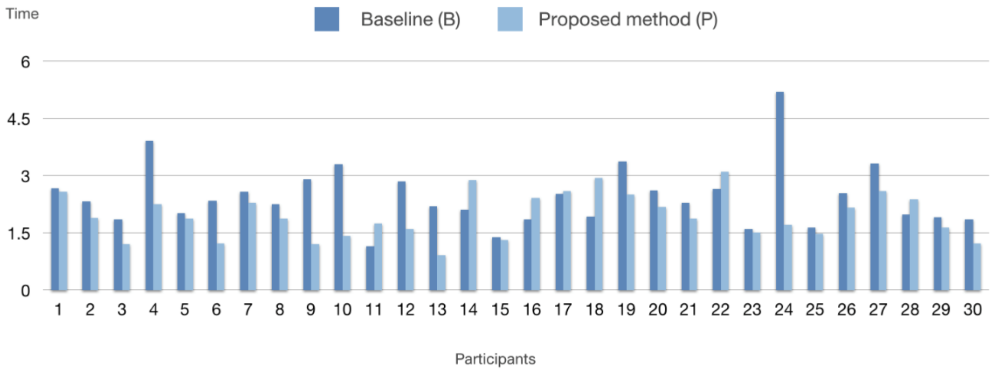

Figure 15. Average time of 30 participants in scene with still obstacles

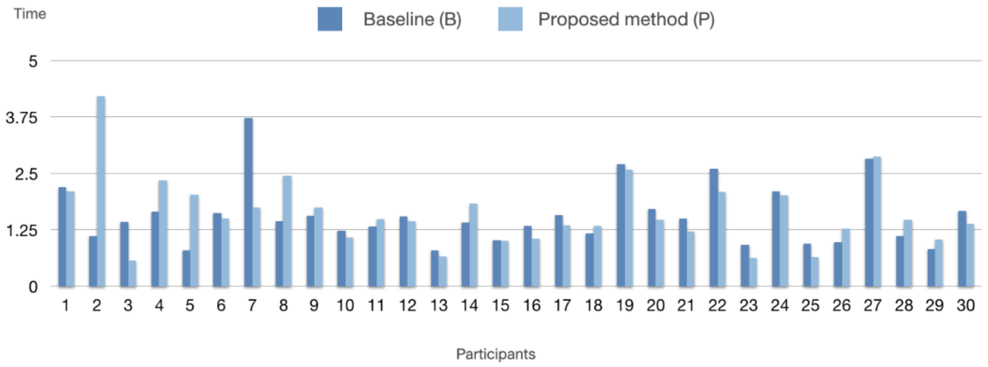

Figure 16. Average time of 30 participants in scene with moving obstacles

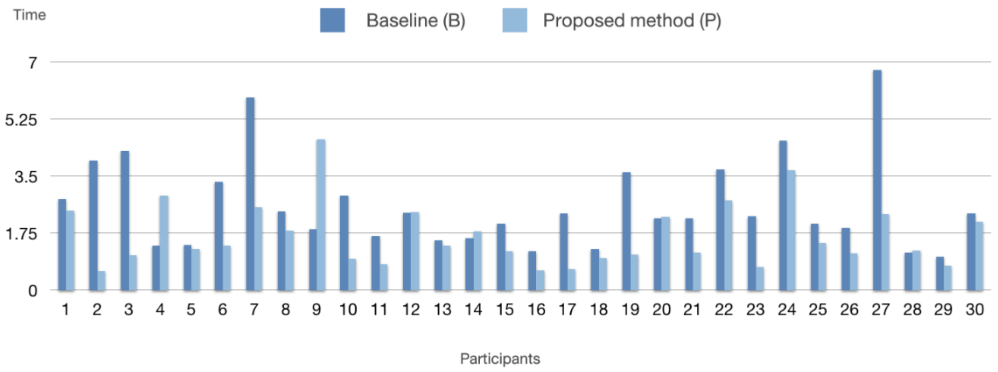

Figure 17. Average time of 30 participants in scene with still and moving obstacles

The participants got a better avoidance rate using proposed method sound in moving scenario and mixed scenario. In the other hand, the avoidance rate dropped, when the participants tested the experiment in standing scenario. The participants reached the goal faster when they avoided the obstacles using the proposed method sound.

Baseline			
Scene	**Obstacle number**	**Avoidance rate (%)**	**Average time**
Still obstacles	10	93.00	02:36
Moving obstacles	10	83.67	01:33
Still & moving obstacles	15	94.67	02:26

Table 7. Overall avoidance rate using baseline sound

Proposed method			
Scene	**Obstacle number**	**Avoidance rate (%)**	**Average time**
Standing scene	10	88.00	01:40
Moving scene	10	85.33	01:37
Mixed scene	15	95.33	01:57

Table 8. Overall avoidance rate using proposed method sound

Chapter 5

Conclusion

The results show that the avoidance rate using proposed method sound is slightly lower than the avoidance rate using baseline sound (no boundary classification). We asked the participants after finishing the experiment did they pay attention to the boundary sound while testing the experiment. Everyone came up with “No” answer. They said they only focused on how to walk away from all alert sounds. That means the boundary sound doesn’t matter for the participants to avoid the obstacles.

For speed representation sound, the results show that the avoidance rate is increased when the participants used the proposed method sound. There are only results from the standing scenario that the avoidance rate using proposed method sound is obviously lower than the baseline sound. Some participants said they used to listening the baseline sound. Some of them said they used to matching the sound with the car collision warning sound. Thus, with their experiences, they assumed that the closer to the obstacle, the more ofter the sound repeats. Their assumptions made them misunderstood about the proposed method warning sound.

In addition, with the same reason as the speed representation results, the overall results show the avoidance rate using proposed method sound is lower than the baseline sound in standing scenario. Figure 17a shows the 30 participants’ moving path of the standing scenario using proposed method sound. The nearest obstacle from the starting point was collided frequently. But after the participants more understood how the proposed method sound represents, they conducted the 2 scenarios left with the higher performance than baseline sound.

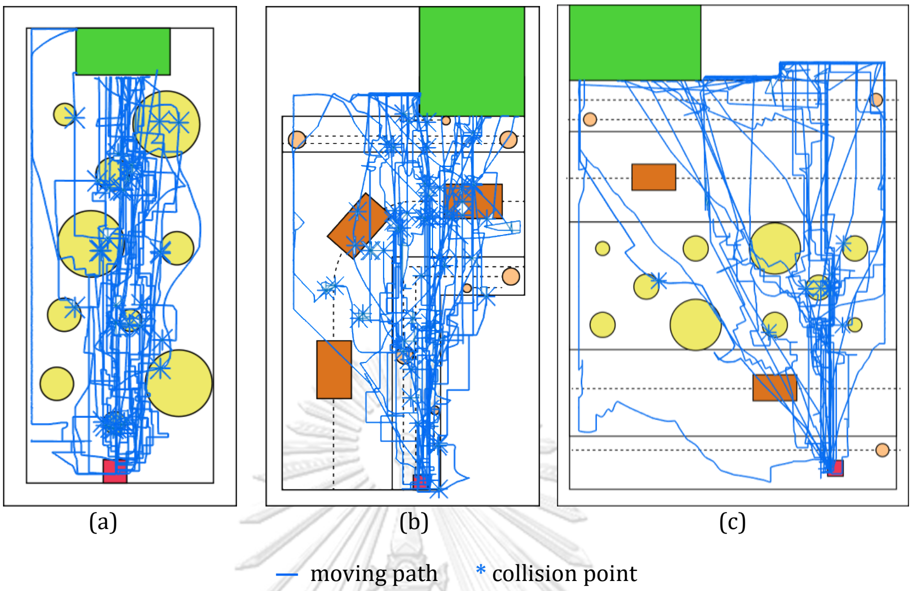

Figure 18. Participants’ moving path using baseline sound

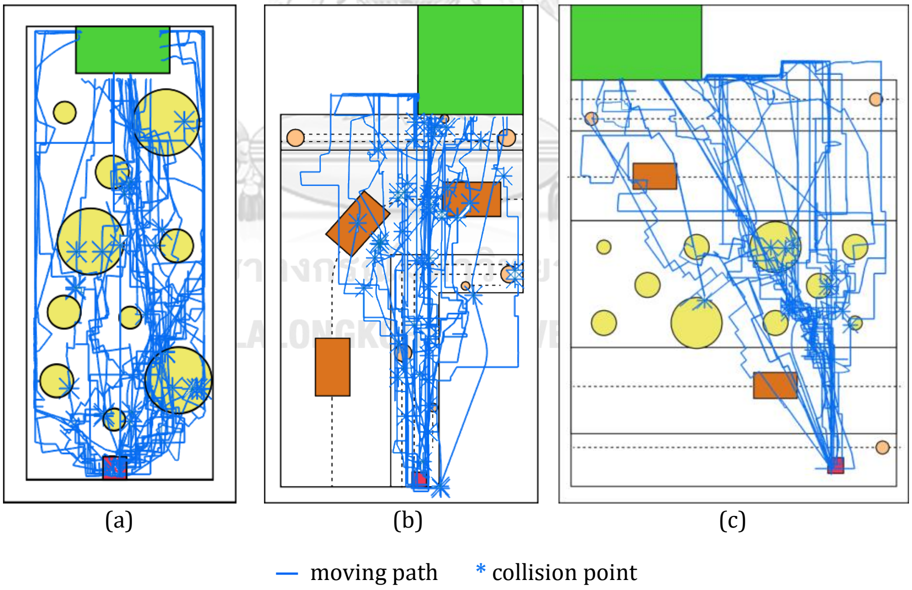

Figure 19. Participants’ moving path using proposed method sound

All participants control the character to the destination carefully. No one try to force collide the obstacles. The figure 16 and 17 show the moving path of the 30 participants in every scenario.

Lastly, we asked the participants which sound they prefer and why. 11 of 30 participants preferred the baseline sound because it is easy to understand. They just only walked away from the warning sound and headed to the destination. The 19 participants left preferred our proposed method sound because they could more understand the scene. Our proposed method sound lets them know which obstacles are moving or are there any obstacles located around them.

REFERENCES

[1] Obrist, M.K., 1995. Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. Behavioral ecology and sociobiology, 36(3), pp.207-219.

[2] Thaler, L., Reich, G.M., Zhang, X., Wang, D., Smith, G.E., Tao, Z., Abdullah, R.S.A.B.R., Cherniakov, M., Baker, C.J., Kish, D. and Antoniou, M., 2017. Mouth-clicks used by blind expert human echolocators–signal description and model based signal synthesis. *PLoS computational biology*, *13*(8), p.e1005670.

[3] Kid Friendly Software, Inc. (2018). Blindfold Runner (1.6.8) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/blindfold-runner/id1377675326?mt=8.

[4] Hyun Soo KIM (2013). SixthSense : All new 3D sound horror shooting game (1.2) [Mobile application software]. Retrieved from https://itunes.apple.com/us/app/sixthsense-all-new-3d-sound-horror-shooting-game/id580417319?mt=8.

[5] Ekkel, M.R., van Lier, R. and Steenbergen, B., 2017. Learning to echolocate in sighted people: a correlational study on attention, working memory and spatial abilities. *Experimental brain research*, *235*(3), pp.809-818.

[6] Politis, D., Margounakis, D., Tsaligopoulos, M. and Kyriafinis, G., 2015, October. Chromatic Reconstruction and Musical Aberration: a survey into the Brain Synapses of Musicality. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound* (p. 24). ACM.

[7] Bragard, V., Pellegrini, T. and Pinquier, J., 2015, October. Pyc2Sound: a Python tool to convert images into sound. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound* (p. 7). ACM.

[8] Zwinderman, M., Zavialova, T., Tetteroo, D. and Lehouck, P., 2011, August. Oh music, where art thou?. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (pp. 533-538). ACM.

[9] Mion, Luca, et al. "Real-time auditory-visual distance rendering for a virtual reaching task." Proceedings of the 2007 ACM symposium on Virtual reality software and technology. ACM, 2007.

[10] Bower, C., Duarte, C., Garrison, J., Reid, L., Seegmiller, J., Wilson, A., Bauer, D.H., Anderson, M. and Wixom, A., 2011, September. Seeing with Sound: Creating a Training Module to Use Echolocation for Object Detection. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 1230-1234). Sage CA: Los Angeles, CA: SAGE Publications.

[11] Deville, B., Bologna, G., Vinckenbosch, M. and Pun, T., 2009. See color: Seeing colours with an orchestra. In *Human Machine Interaction* (pp. 251-279). Springer, Berlin, Heidelberg.

[12] Brown, C. Phillip, and Richard O. Duda. "A structural model for binaural sound synthesis." IEEE transactions on speech and audio processing 6.5 (1998): 476-488.

[13] Algazi, V.R., Duda, R.O., Thompson, D.M. and Avendano, C., 2001. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the* (pp. 99-102). IEEE.

[14] Dunai, L., Fajarnes, G.P., Praderas, V.S., Garcia, B.D. and Lengua, I.L., 2010, November. Real-time assistance prototype—a new navigation aid for blind people. In *IECON 2010-36th Annual Conference on IEEE Industrial Electronics Society* (pp. 1173-1178). IEEE.

[15] Rodríguez, A., Yebes, J.J., Alcantarilla, P.F., Bergasa, L.M., Almazán, J. and Cela, A., 2012. Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. *Sensors*, *12*(12), pp.17476-17496.

[16] Borenstein, J.; Koren, Y.; Member, S. The vector field histogram-fast obstacle avoidance for mobile robots. *IEEE J. Robot. Autom.* 1991, *7*, 278–288.

[17] D. Littlefeild, Metric Handbook. Abingdon, U.L.:Routeledge, 2008.

[18] Mohler, B.J., Thompson, W.B., Creem-Regehr, S.H., Pick, H.L. and Warren, W.H., 2007. Visual flow influences gait transition speed and preferred walking speed. *Experimental brain research*, *181*(2), pp.221-228.

[19] Popular Mechanics. 2016. *New Human-Powered Speed Record Set at 89.6 Mph in Egg-Shaped Bike*. [ONLINE] Available at: https://www.popularmechanics.com/technology/a22946/human-powered-speed-record-aerovelo/. [Accessed 11 June 2018].

[20] Study.com. 2018. *Calculating Average Speed: Formula & Practice Problems*. [ONLINE] Available at: https://study.com/academy/lesson/calculating-average-speed-formula-practice-problems.html. [Accessed 11 June 2018].

[21] Watabe, T. and Niitsuma, M., 2013, December. Mental map generation assistance tool using relative pitch difference and angular information for visually impaired people. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on* (pp. 255-260). IEEE.

[22] "History of Bone Conduction, And 4 Industries Thriving With This Technology", Lindsey Banks, Au.D., 2015. [Online]. Available: http://www.everydayhearing.com/hearing-technology/articles/bone-conduction-headphones/. [Accessed: 24-Nov-2016].

APPENDIX

Before we started the methodology session, we had carried out the sandbox testing by designing different kinds of audio to represent the spatial information and tested with the participants. The sandbox testing was divided into 3 parts including distance testing, direction testing and speed testing. The aim of sandbox testing is to find the appropriate sound representation for navigation and avoidance. There are 7 participants (6 male and 1 female) participated the testing.

Distance testing

We designed 3 styles of the distance sound representation. For the first style, we used the representation style of Rodríguez, A., et al. [15]. This representation contains 2 audio sounds including near sound with Middle C Note (261.6 Hz) and far sound with C_2 Note (65.4 Hz). Figure 17 shows the distance sound representation for the first style.

"Middle C" is designated C_4 in scientific pitch notation with a frequency of 261.7 Hz, because of the note's position as the fourth C key on a standard 88 key piano keyboard.

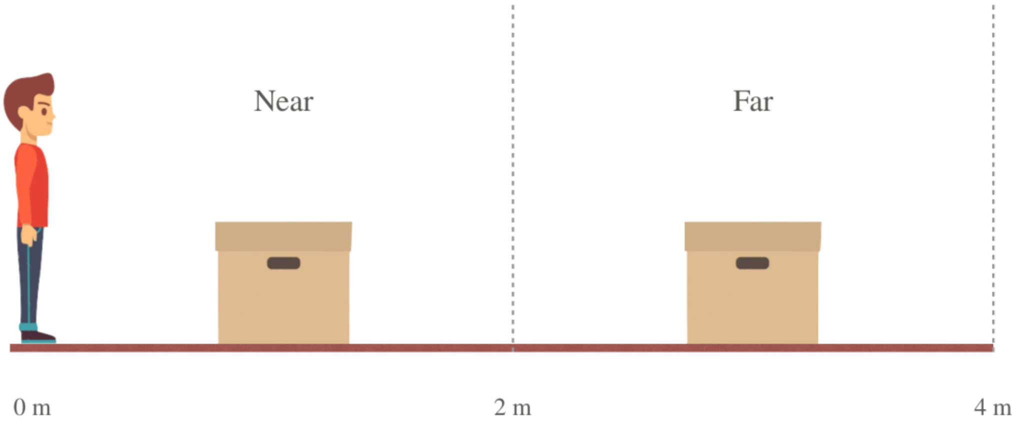

Figure 20. The first style of distance sound representation.

For the second style, we separated the distance into 3 parts. The medium distance was added to this style as shown in figure 18. The range was rearranged to near sound with Middle C Note (261.6 Hz), medium sound with C_3 Note (130.8 Hz) and far sound with C_2 Note (65.4 Hz).

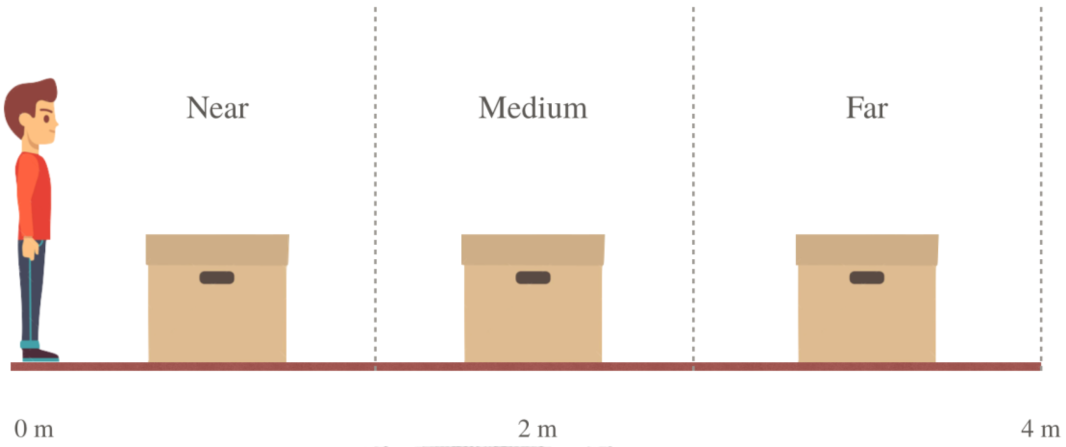

Figure 21. The second style of distance sound representation.

For the third style, we separated the distance into 4 parts (Figure 19). This representation contains very close sound with C_5 (523.3 Hz), near sound with Middle C Note (261.6 Hz), medium sound with C_3 Note (130.8 Hz) and far sound with C_2 Note (65.4 Hz).

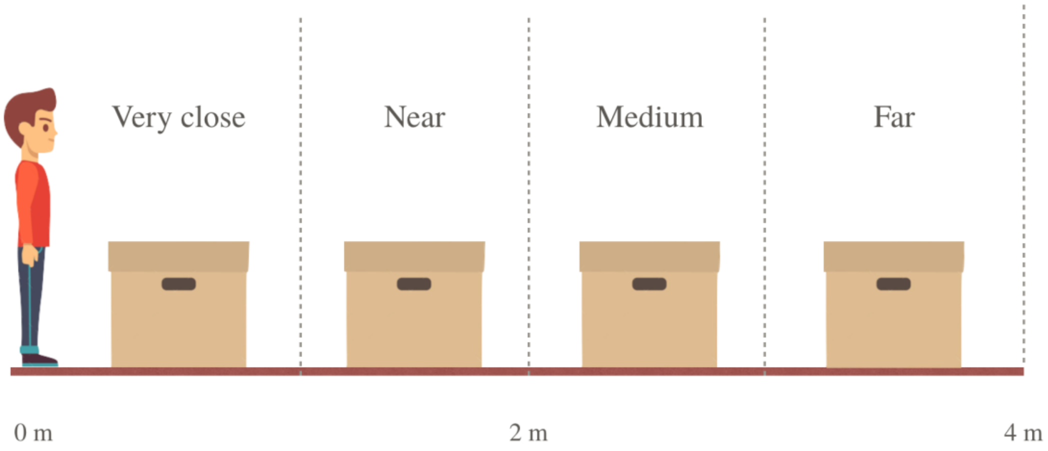

Figure 22. The third style of distance sound representation.

Before testing, we explained the meaning of each style of the distance sound to the participants. Then, for the testing session of each style, they were asked to listen to the sound randomly and answered what kind of sound they heard. The participants could take the rest for 3 minutes before starting the next testing of other styles.

The result shows that the accuracy of the second style testing is the highest (Table 10). We also asked the participants which style they prefer to use for navigation and avoidance. 4 of 7 participants selected the second style.

Style	**Accuracy (%)**	**Favourite number**
1st	85.71	3
2nd	88.89	4
3rd	76.19	0

Table 10. The distance sound selection testing.

Direction testing

We designed 3 styles of the direction sound representation. Each style sound is a stereo sound generated by using head-related impulse response (HRIR). For the first style, we used the representation style of Rodríguez, A., et al. [15]. This representation contains 4 audio sounds including left sound with C_4 Note (-60° to -30°), front-left sound with D_4 Note (-30° to 0°), front-right sound with D_4 Note (0° to 30°) and right sound with C_4 Note (30° to 60°). Figure 20 shows the direction sound representation for the first style.

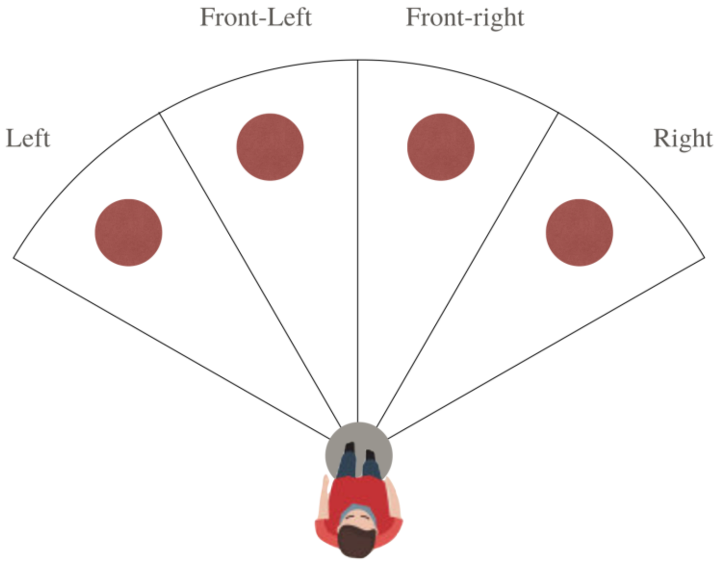

Figure 23. The first style of direction sound representation.

For the second style, we separated the direction into 5 parts. The front direction was added to this style as shown in figure 21. The range was rearranged to left with C_4

Note (-75° to -45°), font-left with D_4 Note (-45° to -15°), front with E_4 Note (-15° to 15°), front-right with D_4 Note (15° to 45°) and right with C_4 Note (45° to 75°).

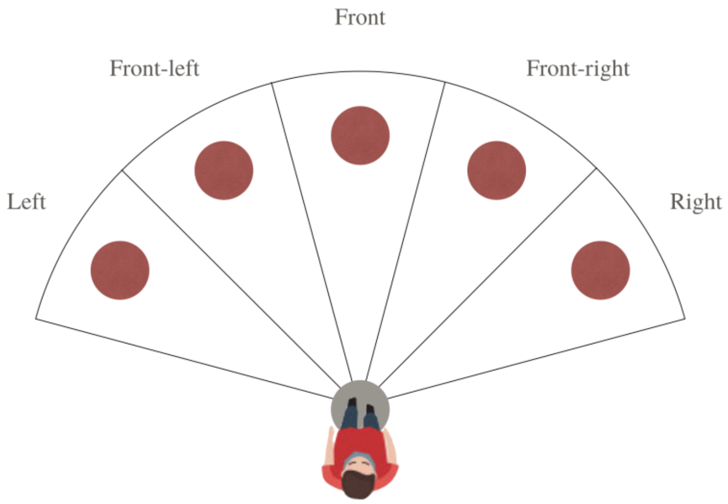

Figure 24. The second style of direction sound representation.

For the third style, we separated the direction into 6 parts (Figure 22). This representation contains left with C_4 Note (-60° to -40°), mid-left with D_4 Note (-40° to -20°), front-left with E_4 Note (-20° to 0°), front-right with E_4 Note (0° to 20°), mid-right with D_4 Note (20° to 40°) and right with C_4 Note (40° to 60°).

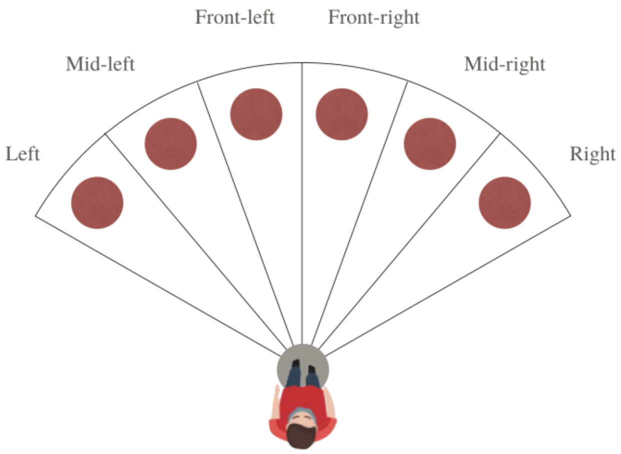

Figure 25. The third style of direction sound representation.

Before testing, with the same process as distance testing, we explained the meaning of each style of the direction sound to the participants. Then, for the testing session of each style, they were asked to listen to the sound randomly and answered what kind of sound they heard. The participants could take the rest for 3 minutes before starting the next testing of other styles.

The result shows that the accuracy of the second style testing is the highest (Table 11). We also asked the participants which style they prefer to use for navigation and avoidance. 5 of 7 participants selected the second style.

Style	Accuracy (%)	Favourite number
1st	87.50	1
2nd	88.57	5
3rd	86.90	1

Table 11. The distance sound selection testing.

Speed testing

We designed 3 styles of the speed sound representation. The sound repetition with the same duration was used to represent the speed sound. For the first style, the representation contains 2 audio sounds including standing idle speed sound (0 m/s) and moving sound (>0 m/s). The number of sound repetitions were 1 and 3 respectively.

For the second style, we divided the speed sound representation in to 3 types including idle speed sound (0 m/s), slow speed sound (0 - 4 m/s) and fast speed sound (>4 m/s). The number of sound repetitions were 1, 3, and 5 respectively. The slow speed sound was represented to people walking speed. The fast speed sound was represented to the speed that is faster than people walking speed which has a risk of harm for navigation.

For the third style, we separated the speed into 4 types including idle speed sound (0 m/s), slow speed sound (0 - 4 m/s), medium speed sound (4 - 12 m/s) and fast speed sound (> 12 m/s). The number of sound repetitions were 1, 3, 5 and 7 respectively. The medium speed sound was represented to people running speed or cycling speed. The fast speed sound was represented to the speed that is faster than people running or cycling such as car.

Before testing, we explained the meaning of each style of the speed sound to the participants. Then, for the testing session of each style, they were asked to listen to the sound randomly and answered what kind of sound they heard. The participants could take the rest for 3 minutes before starting the next testing of other styles.

The result shows that both accuracy of the second style testing and the third style testing are the highest (Table 12). We also asked the participants which style they prefer to use for navigation and avoidance. All of participants selected the second style.

Style	**Accuracy (%)**	**Favourite number**
1st	100.00	0
2nd	100.00	7
3rd	87.50	0

Table 12. The speed sound selection testing.

VITA

Kawin Metsiritrakul is a Master’s Degree student of Computer Engineering Departure, Faculty of Engineering, Chulalonkorn University.

He was born in Betong ,the southernmost city in Thailand. He studied at Mahidol Wittayanusorn school when he was high school student. In 10th grade, he had learnt programming class and he was interested in programming. So, he decided to attend to Computer Engineering Departure, Faculty of Engineering, Chulalonkorn University as a Bachelor's Degree student.

Not only was he interested in programming, he was also interested in computer graphic and animation. He spent his free time to create an graphic things or short film animation. In 2017, he attended an internship as an UX/UI designer for 4 months at National Institute of Informatics (NII), Tokyo, Japan.