Studies on the fluctuation of human consciousness, behavior and spatial comfort by auditory display according to individual and spatial contexts

- Summary -

March 2024

Noko Kuratomo

Studies on the fluctuation of human consciousness, behavior and spatial comfort by auditory display according to individual and spatial contexts

\- Summary -

Graduate School of Science and Technology

Degree Programs in Systems and Information Engineering

University of Tsukuba

March 2024

Noko Kuratomo

Abstract

Spatial comfort is an important element for spatial design to ensure that individuals are comfortable. One parameter is acoustics, and measures such as lowering the level of sounds that are not relevant to one's needs and masking by natural sounds are important. On the other hand, the use of auditory displays is intended to convey messages to people, so it is common for sounds to be intentionally and prominently transmitted. However, in some cases, the pursuit of spatial comfort through sound is hindered by the use of auditory displays. Therefore, there is a trade-off between the effectiveness of auditory display utilization and spatial comfort. This issue must be resolved.

In this dissertation, the author proposes an auditory display that senses whether an individual needs informational sound or not, whether the space requires the sound to be produced, and presents the sound only when needed. However, before this system will be implemented in public spaces, it is necessary to clarify the extent to which the trade-offs can be resolved. Regarding the effectiveness of the use of auditory displays, it will be seen whether people's awareness and behavior change from conventional auditory displays depending on the presence or absence of sound presentation and the sound genre. Regarding spatial comfort, it will also investigate changes in sound level and in acoustic comfort as a parameter. Through these research, the objective of this dissertation is to investigate the utilization effectiveness of auditory display and changes in spatial comfort.

Chapter 1 describes the issues to be addressed in this dissertation with background and previous research. The purpose of this thesis and the research questions are also stated.

Chapter 2 investigates the audible thresholds for words emitted by the parametric loudspeakers used in the auditory displays in this paper. Because parametric loudspeakers emit narrowly directional sound, it is possible to create a system that delivers sound only to those who are interested and not to others who might otherwise become noisy. However, the reduction in noise level is only qualitatively known, and the specific degree of reduction is not yet clear. Therefore, I aimed to quantitatively evaluate how much the noise level is reduced for people who do not need the sound by measuring the speech recognition threshold (SRT). As a result, the lower the background sound, the more noise is reduced. If the background sound level was 20 dBA, the SRT for non-target listeners was 27.25 dBA for the conventional loudspeakers, but 14.05 dBA for the parametric loudspeakers. When combined with background sound, the difference in loudness was 7.02 dBA. These results indicate that the parametric loudspeaker can enhance the spatial comfort level for non-target individuals.

In Chapter 3, I used advertising signage as an example to examine whether auditory displays based on individual conditions can resolve the trade-off between usage effectiveness and spatial comfort. Specifically, I proposed a two-step system: an on-demand pinpoint audio system in the first step and a pinpoint auditory glimpse in the second step. These systems change the sound level according to the individual's state of interest. It is verified whether this system can resolve the trade-off relationship between advertisement recognition as a utilization effect and spatial comfort. Experiments were conducted in which participants experienced shopping and viewing signage, and were verified through behavioral observation and questionnaires. The results showed that not only changing the sound

level according to the level of attention, but also changing the sound level as a "trigger" for passersby to look at the signage increased advertisement recognition. Also, even if an auditory display varies in sound level, it does not negatively impact the acoustic comfort. From the above, it was found that an auditory display that is tailored to the individual's state of interest can balance the effectiveness of use and spatial comfort.

In Chapter 4, the study examined whether an auditory display according to the spatial context could resolve the trade-off between guiding effects and spatial comfort, with a focus on pedestrian flow management and assessment of comfort in relation to streetscape. This study evaluated sound designs for guiding pedestrians while maintaining the streetscape under normal conditions. The results showed that under normal conditions, using comforting sounds, such as nature sounds, can effectively guide human flow as these sounds increase the motivation to follow the guidance.

From these three exploratory experiments, design guidelines for auditory displays that balance utilization effectiveness and spatial comfort were proposed, focusing on sound level, exposure time, and sound genre. This dissertation contributes as a foundation for implementing auditory displays in public spaces in the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Soundscape and Acoustic Comfort

We are constantly sensing various things through our sensory organs without even being aware of it. Through our ears, one such sensory organ, we perceive sounds. When we go out into the city, we involuntarily hear many sounds, i.e., the noise of our own footsteps while walking, the chatter of people around us, the sound of cars moving, and the rustling of tree leaves. These sounds inevitably enter our ears. Research on the perception of these sounds has a long history. In 1969, Southworth was the first to name this research 'soundscape,' defining it as "the quality and type of sounds and their arrangements in space and time" [1]. Subsequently, Schafer categorized the current sound environment into two types [2]. One is the Hi-Fi (High Fidelity) sound environment, defined as an ideal state where every sound can be heard clearly. The other is named the Lo-Fi (Low Fidelity) sound environment, characterized by a state where certain sounds or noises dominate, causing other sounds to be drowned out. Schafer noted that after the industrial revolution, the soundscape of the west rapidly transitioned to a Lo-Fi state, underscoring the importance of soundscape design.

In the design of urban soundscapes, acoustic comfort is an important aspect [3, 4]. The spatial comfort we experience in our daily lives involves elements such as temperature [5, 6], light [6], and humidity [7], with acoustics being one of them. Acoustic comfort is defined as "sound from all sound sources as modified by the environment"[8]. Previous studies have assessed the acoustic comfort of various spaces, ranging from noisy areas like stations[9] and shopping malls [3] to quiet environments like parks [10] and libraries [11]. These studies have reported that the sound level in the space is a major factor in determining acoustic comfort [12, 13]. In particular, it has been found that the acoustic comfort decreases when the sound level of unrelated noises, such as traffic noise in parks or conversations in stores, is high [14, 15]. As a method to prevent this decrease in acoustic comfort, techniques involving the masking of noise with natural sounds like bird chirps or the sound of rivers have been employed [16, 17].

Thus, in the design of spaces where people spend their time, it is important to enhance the acoustic comfort of the location. This can lead to an overall improvement in the spatial comfort.

1.2 Auditory Display in Public Space

In public spaces, auditory displays are used to convey messages to the people present. The content of the messages varies depending on the context of the space. For instance, in places like stations or bus stops, which are spaces frequented by people on the move, messages regarding transportation guidance are broadcasted. In commercial facilities, where people shop or receive services, advertising and promotional messages are played. In addition, in spaces like events that are crowded, announcements guiding people to detours are made to alleviate congestion. These messages are conveyed either through loudspeakers by people or through automated voices via loudspeakers.

These messages directly influence the consciousness and behavior of people. Transportation guidance messages can smooth the movement of passengers. Advertising messages can encourage customer purchasing behaviors, influencing product purchases or service utilization. Announcements for detour routes can lead to a reduction in congestion by directing some pedestrians away. Therefore, these sounds are emitted to be noticeable and audible.

Furthermore, sound plays an important role in shaping the atmosphere of a space. For example, in restaurants, playing fast tempo music can enhance the taste of food and the desire to purchase [18]. Customers also associate the music in stores with the brand and image of the products sold there [19]. In city streets, street music contributes to the vibrancy and character of the area [20, 21].

Thus, by using auditory displays in public spaces, people's consciousness and behaviors are being changed.

1.3 Relationships of Auditory Display and Acoustic Comfort

As mentioned in Section 1.1, acoustic comfort is an important factor for spatial design to ensure individual comfort. Therefore, measures such as reducing the sound level of irrelevant noises and masking with natural sounds are important. On the other hand, as stated in Section 1.2, the use of auditory displays is typically intended to convey messages to people, thus the sounds are emitted to be intentionally noticeable.

However, balancing these two points can be challenging. As previously mentioned, station announcements may be an important source of information for some passengers, but for others, they may be unnecessary and decrease acoustic comfort [9]. Similarly, advertising promotions may be effective for interested individuals or potential customers, but they can be noise for those who are not concerned. Announcements for detour routes can also become superfluous auditory information for those walking in the opposite direction or when such guidance is unnecessary. That is, in the same space, there are cases where some people need auditory information and others do not. Pursuing acoustic comfort for those who do not need it can be hindered by the use of auditory displays. This causes spatial comfort level decrease. Therefore, there is a trade-off issue between the effectiveness of auditory displays and spatial comfort, and resolving this trade-off issue is necessary.

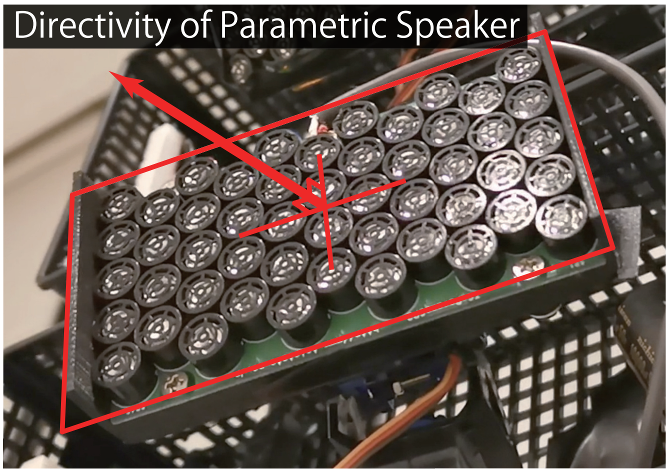

Figure 1.1: Sound Directivity of Parametric Loudspeaker

Thus, this dissertation proposes a system that delivers messages to those who need them and not to those who don't, aiming to ensure spatial comfort.

1.4 Parametric Loudspeaker

Currently, a loudspeaker used as a device for auditory display (hereafter referred to as a conventional loudspeaker) emit sound in a non-directional manner. This allows the sound to reach a wide area. However, this characteristic leads to the trade-off between the effectiveness of auditory displays and spatial comfort.

Therefore, this dissertation utilizes a parametric loudspeaker. A parametric loudspeaker is a device that arrays ultrasonic waves while transmitting a signal that demodulates them into audible sound by using the nonlinearity of their propagation through the air, resulting in audible sound with a narrow directionality [22, 23]. Whereas conventional loudspeakers diffuse sound, parametric loudspeakers emit sound in a straight line, which has the advantage of limiting the target area and causing less noise, as shown in Figure 1.1. Another feature of this system is that it is easier to localize than conventional loudspeaker stereophonic sound due to reduced crosstalk [24]. This allows the sound to be delivered only to people in a specific area. Thus, by using parametric loudspeakers, the aim is to deliver auditory information only to those who need it.

1.5 Dissertation Objective

As discussed in Section 1.4, parametric loudspeakers can deliver sound to specific individuals, but it's unclear in which situations and for whom the sound is necessary. Therefore, this dissertation proposes a system that not only utilizes parametric loudspeakers but also incorporates sensing to determine whether an individual needs informational sound and whether it's necessary to emit sound in a space. This will ensure that the message is delivered only to individuals who need it.

However, before implementing this system in public spaces, it is essential to clarify to what extent it contributes to resolving the trade-off. Regarding the effectiveness of auditory displays, I will examine whether the presence or absence of sound and the type of sound change people's behavior and consciousness compared to conventional auditory displays. For spatial comfort, I will investigate the fluctuations in sound level and the change in spatial comfort with sound as a parameter. This dissertation poses the following research questions and verifies them through three exploratory experiments.

- RQ.1 How much is the noise level reduced by the use of parametric loudspeakers?
- RQ.2-1 Does varying the sound level of an ad's auditory display improve recognition?
- RQ.2-2 Does varying the sound level of an ad's auditory display make the acoustic environment less comfortable?
- RQ.3-1 What kind of sound design effectively guides pedestrian flow?
- RQ.3-2 What is the impact of sound-guided intervention on streetscape perception?

Through these research questions, it will be verified to what extent the trade-off issue has been resolved. Specifically, the RQ.1 will examine how much controlling sound level improves spatial comfort. Through RQ.2-1 and 2-2, the effects will be tested by manipulating exposure time according to the degree of interest according to the individual context. As an example of this case, I will take up advertising digital signage. Through RQ.3-1 and 3-2, I will verify the effect of changing the type of sound according to the condition of the space. An auditory display that guides human flow will be employed as an example of this case study. This dissertation will examine the degree to which the trade-off issue is resolved and discuss how auditory displays should be designed, using sound level, exposure time, and sound genre as the three key parameters.

This dissertation is organized as follows: the remainder of Chapter 1 presents related research on auditory displays and the use of parametric loudspeakers, outlining the position of this study in relation to prior research. In the Chapter 2, I measure the Speech Recognition Threshold (SRT) of parametric loudspeakers and compare with that of conventional loudspeakers at various noise levels through experiments. It will also reveal how much the parametric loudspeaker can reduce the noise heard by non-target users, which is unavoidable with conventional loudspeakers. Chapter 3 proposes two-steps systems called "on-demand pinpoint audio system," which delivers sound only to a person by using directional sound, and "pinpoint auditory glimpse," which uses sound to notify passersby of the presence of signage. By conducting a participant experiment in a pseudo-store, research questions 2-1 and 2-2 will be answered. Chapter 4 proposes an auditory display for guide pedestrians and explores the sound design balancing the sound effectiveness and maintaining streetscape. Through the laboratory experiments, research questions 3-1 and 3-2 will be answered. In Chapter 5, the results of the above three chapters are summarized, and design guidelines for auditory displays, as well as their limitations and future directions, are discussed.

The contributions of this dissertation are following:

- The use of parametric loudspeakers has quantitatively shown that, even in the same space, sound is delivered only to those who need the information sound, and the noise level of those who do not need it is reduced. This also leads to a reduction in information masking. This increases the degree of flexibility in spatial design from the perspective of sound. (corresponds to Chapter 2)
- The system, which is audible only to those who are watching and delivers sound to passersby in a short period of time, is expected to have an advertising effect while maintaining the spatial comfort of the space. This will enable commercial facilities, which have traditionally had difficulty using advertising sound from a noise perspective, to make use of sound, thereby expanding the potential for business deployment. (corresponds to Chapter 3)
- Regarding pedestrian guidance using sound, it is shown that modifying the sound design according to the intensity of guidance, which changes with the spatial context, is effective in balancing effective use and streetscape perception. When considering the use of urban space, it leads to the easing of sound restrictions in cases where such restrictions have been made. (corresponds to Chapter 4)

1.6 Related Work

1.6.1 Auditory Display

This dissertation is associated not only with auditory display, but also with commercial facilities, digital signage, current pedestrian guidance, and spatial audio. This section describes two related studies, auditory display and parametric loudspeakers, that are relevant to all chapters. In addition, in each chapter, the related studies of the individually related topics will be described.

An *auditory display* is a method of conveying information using sound and has been applied in various fields. The information transmitted can range from complementing visual information as an assistive technology to sonifying data or even intentions [25].

One of the benefits of converting visual information into sound is that it becomes more readily accessible for individuals with certain disabilities. Several systems have been proposed to provide navigation assistance to individuals with visual impairments by conveying geographical information through sound or touch. These systems can be utilized outdoors using wireless communication or smartphones [26, 27, 28, 29], as well as indoors by using conventional loudspeakers in GPS-unavailable environments [30, 31, 32, 33]. Such systems can help to reduce the difficulty for visually impaired individuals to navigate various environments. Furthermore, by using sound to inform the visually impaired of the location of an object, it becomes easier to find lost items and avoid obstacles [34, 35, 36].

Another advantage of converting data into sound is the ease of perceiving data. For example, it has been reported that auditory data representation requires less learning time compared to visual data representation [37]. Additionally, auditory perception is said to be effective in detecting subtle differences between data [38]. Converting time-series data into audible formats allows the data to be perceived as a melody, enabling easier assimilation of the data [39].

Moreover, sound can be used to convey intentions through sound. For example, during emergencies, alarming sounds are played to convey a sense of crisis. In more everyday scenarios, sirens on emergency vehicles serve to communicate urgency, even when the vehicle is not visible, alerting drivers and pedestrians to make way. Research has been conducted to enhance the effectiveness of such signals [40] and reduce noise levels [41]. In the case of fire emergencies, various types of sound designs have been studied to quickly evacuate buildings [42]. Additionally, the development of systems that use abstract sounds alongside language to guide individuals to exits has been reported, proving to be effective even under conditions of reduced visibility due to smoke [43]. In a mobile context, notification sounds are used to convey information about new messages or updates. It has been reported that such auditory cues are intuitive and preferred by users [44].

In summary, auditory display serves various roles that cannot easily be achieved through visual means. This dissertation proposes an auditory display that fulfills two roles. The first, discussed in Chapter 3, is the presentation of advertising sounds. Advertising sounds use both verbal and non-verbal information, aiming to make them as cognitively recognizable to people as possible. The second, discussed in Chapter 4, involves guiding sounds. These sounds communicate to pedestrians at intersections which paths to follow. This role can be categorized as auditory displays intended to convey specific intentions to users.

1.6.2 Application Using Parametric Loudspeaker

Various techniques have been developed to localize sound sources using conventional loudspeakers. It has been reported that the precedence effect allows the sound to be localized approximately correctly even in the presence of reverberation [45, 46], and it is particularly important to be in the sweet spot to receive this effect [47]. However, crosstalk affects the localization [24]. A conventional loudspeaker also generally diffuses the sound and makes the surrounding environment noisy. In comparison, parametric loudspeakers have superdirective sound and suppress crosstalk. This makes it possible to transmit binaural information to the listener more stably than conventional loudspeakers [24]. An additional advantage is that their directivity does not diffuse the sound into the surrounding environment [48, 49].

In a previous study in which directional sounds were employed, an approach was proposed to realize a system that provides information in quiet environments such as museums [50, 51], utilizing a parametric loudspeaker and lenticular lens to make sound and display an image that was visible only to a specific person [52, 53]. Researchers have also developed systems that can transmit sounds that can track passersby wearing hats with markers [54] and those emitted sounds by phased arrays to produce point sources [55]. In addition, directional sound generated by sound lenses using optical lenses lowers the technical hurdle [56].

In the above mentioned studies, different techniques were developed for implementing directional sound. However, the psychological fluctuations of users when subjected to it were not fully discussed. This dissertation elucidates this unexplored aspect, especially human behavior and consciousness, and spatial comfort caused by acoustics, through applications and experiments.

Chapter 2

Speech Recognition Threshold

2.1 Chapter Objective

As stated in Chapter 1, by using parametric loudspeakers, sound can be delivered to a limited audience. This enables a system where only individuals who require auditory information can hear it. On the flip side, for those who do not need the sound and are not in the “sweet spot ” - those are now referred to as “non-target listeners ” - the sound is almost inaudible. This likely results in a lower noise level than what is experienced with conventional loudspeakers. However, this reduction in noise level is currently understood only qualitatively, and the exact degree of reduction is not clearly defined. Therefore, this chapter aims to quantitatively assess how much the noise level decreases for non-target listeners, addressing RQ.1: “How much is the noise level reduced by the use of parametric loudspeakers?” This analysis will help determine whether parametric loudspeakers can contribute to maintaining or enhancing acoustic comfort.

As the assessment, the Speech Recognition Threshold (SRT) of parametric loudspeakers is measured and compares with that of conventional loudspeakers at various noise levels through experiments. The threshold of understanding is calculated not only for the people around them who are not meant to receive it but also for the person receiving the spoken information. This information can reveal how much the parametric loudspeaker can reduce the noise heard by non-target users, which is unavoidable with conventional loudspeakers. It can be also used to quantitate how much louder a parametric loudspeaker must be compared with a conventional loudspeaker when it is actually used in a public place. This enables soundscape designers to evaluate the degree of contribution to spatial comfort. By examining environments ranging from quiet to noisy, such as museums and shopping malls, a variety of different conditions can be covered.

2.2 Experimental Environment

The experiment was conducted in an anechoic chamber to measure the SRTs of people who need information and those who do not. The experimental environment is described in Figure 2.1. In the experiment, the sound was emitted by a parametric loudspeaker and

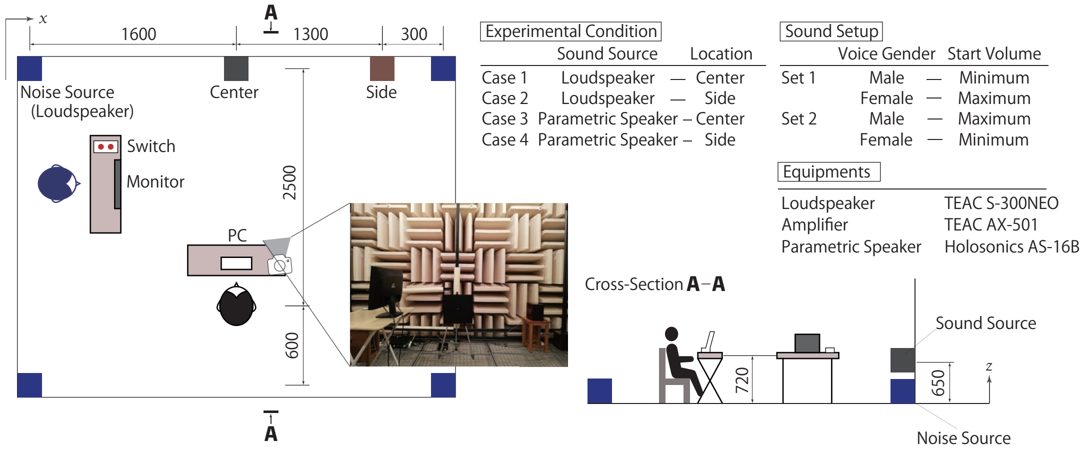

Figure 2.1: The experimental environment and conditions.

conventional loudspeaker at various noise levels. For background noise, conventional loudspeakers were placed in the four corners of the chamber to produce sound. The participants sat in chairs and judged whether they could hear the target sound. The conventional loudspeaker and parametric loudspeaker that emitted the target sound were to be installed in the center or at the sides. The loudspeakers were placed in front of the participants for the center placement and out of their line of sight for the side condition, as shown in Figure 2.1. A list of the equipment used is given in the table at the top right of Figure 2.1.

The background noise was white noise with an FIR digital filter (according to ITU-T Rec. G.227). The spectrum of this noise is shown in Figure 2.2 (a). The noise level was set at 20–80 dBA, and was changed by increments of 10 dBA. The sound level is defined as the average sound level measured by a sound level meter set up where the participants were sitting. For sound level calibration, a 1 kHz sine wave was used. The sound was calibrated to a sound pressure equivalent to that of the background noise or the target sound.

SRT measurement experiments need to use native languages with uniform syllable and parenthetical densities. If the target sound is not in the native language [57, 58, 59], or if the number of syllables or the degree of familiarity varies [60, 61], the recognition threshold will decrease and accurate data will not be obtained. Since this experiment was conducted entirely with Japanese participants, the FW07 dataset [62] in Japanese was used for the transmitted sound. This dataset contains words with four syllables and has been previously used in SRT experiments by Japanese researchers [63, 64]. Since familiarity density [65] can also be selected, this experiment used a high familiarity word set to create a language environment close to that of daily life. Loudness normalization (ITU BS.1770-3 standard) was performed for all spoken words speech, and the sound level was set to be the same. The time plot of the spoken words is shown in Figure 2.3, and the spectrum is shown in Figure 2.2.

The participants of this experiment were 15 undergraduate or graduate students (ten males and five females) with an average age of 22.64 years. The experiment was conducted

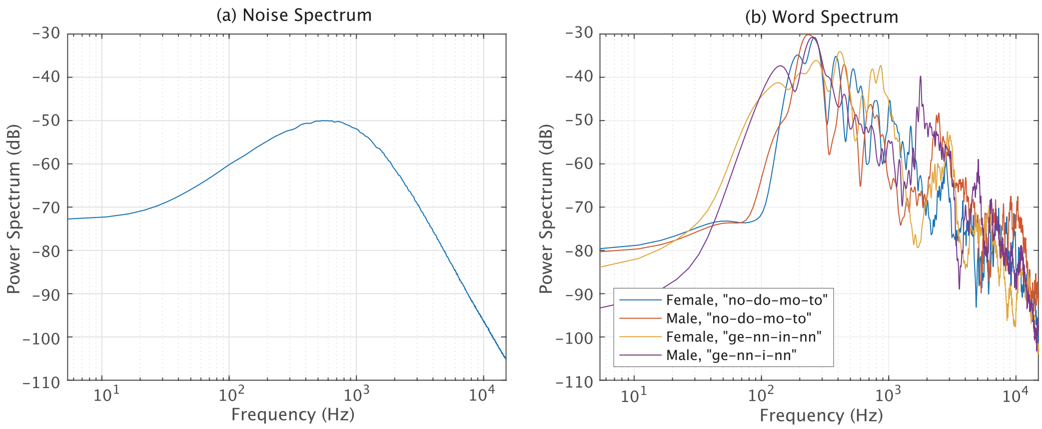

Figure 2.2: Spectra – (a) Background noise, (b)Examples of the spoken word

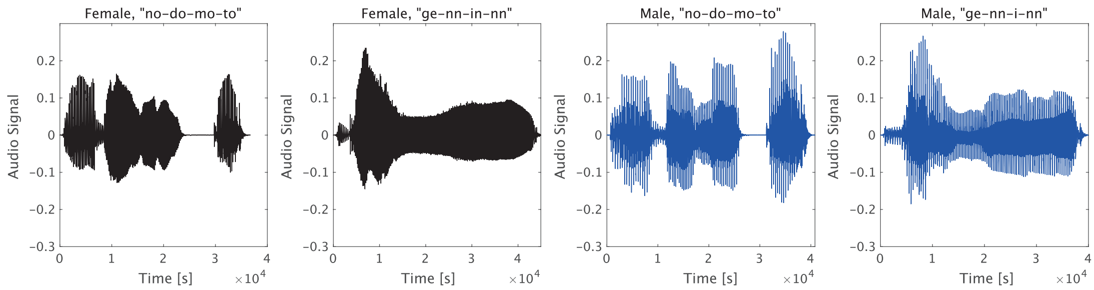

Figure 2.3: Examples of the time plot of the spoken words : "nodomoto" means throat, and "genninn" means reason.

after explaining the details of the experiment and obtaining the participants' consent in advance. Before the experiment, I conducted a sound field threshold measurement to confirm that the participants had normal hearing.

This experiment was reviewed and approved by the ethics committee of the University of Tsukuba (Faculty of Engineering, Information and Systems, Permission number: 2020R434-1).

2.3 Procedure

In the experiment, the words were played in an environment where background noise was always present. Participants typed the words that they heard on a PC. The researcher watched the input text by mirroring it to determine if it matched the actual words played and then used a switch to increase or decrease the sound level. If the test started at the minimum sound level, the sound was increased by 10 dBA every time a wrong answer was given. After that, the sound level was decreased by 2 dBA when a correct answer was given

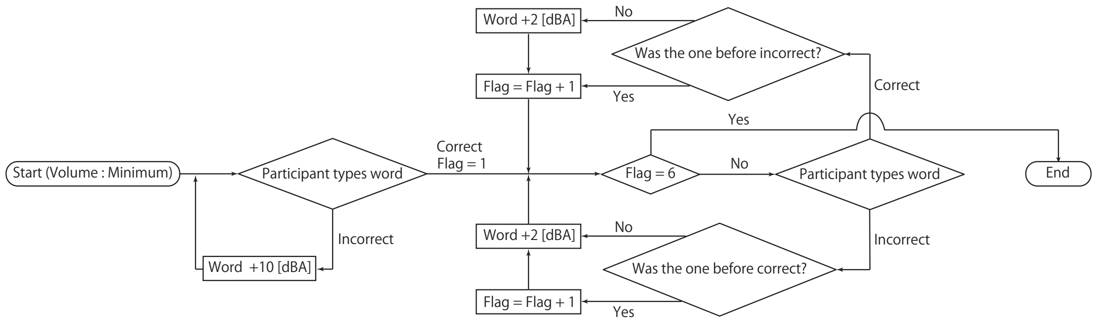

Figure 2.4: Experiment procedure of the test started at the minimum sound level.

and increased by 2 dBA after another wrong answer. Repeat these steps, and end the audio presentation if it turns into a wrong answer for the third time. These steps were repeated, and the audio presentation ended after a third wrong answer. This procedure was repeated but starting with the maximum sound level, and the sound presentation was ended after the third correct answer. The sound level at which the percentage of correct answers was 50% was set as the SRT, which was taken from the midpoint of the sound level when the answer reversed from incorrect to correct or vice versa. The above experimental procedure is shown in Figure 2.4.

There were four experimental conditions: a conventional loudspeaker or parametric loudspeaker was used; these loudspeakers were placed in a central or side location; the test started at the minimum or maximum sound level; and the words were read out loud in a male or female voice. The experimental conditions and sound settings are shown in the right corner of Figure 2.1. The level of the target sound is defined as the average level measured by a sound level meter placed where the participants were sitting with the loudspeaker placed in the center of the room. For example, a target level of 40 dBA means that this is the level that the participant heard when the loudspeaker was in the center, and the loudspeaker output the same level when it was at the side. The sound level range was 20–80 dBA for the conventional loudspeaker and 20–70 dBA for the parametric loudspeaker on the basis of equipment and ear protection.

2.4 Results

The SRT results for each loudspeaker and location are shown in Figure 2.5 (a). These results are pertain to the conditions of sound level variation and voice gender. For the conventional loudspeaker, the SRTs are almost the same for the cases in which the device is placed at the center and side at all background noise levels. The signal-to-noise ratios (SNRs), which represent the difference levels between the background noise and the SRT, and their average values were calculated. The calculated SNRs were approximately -1.41 dBA and -1.34 dBA for the center and side placements, respectively. These values show that non-target individuals can also hear the target sound when the target person can

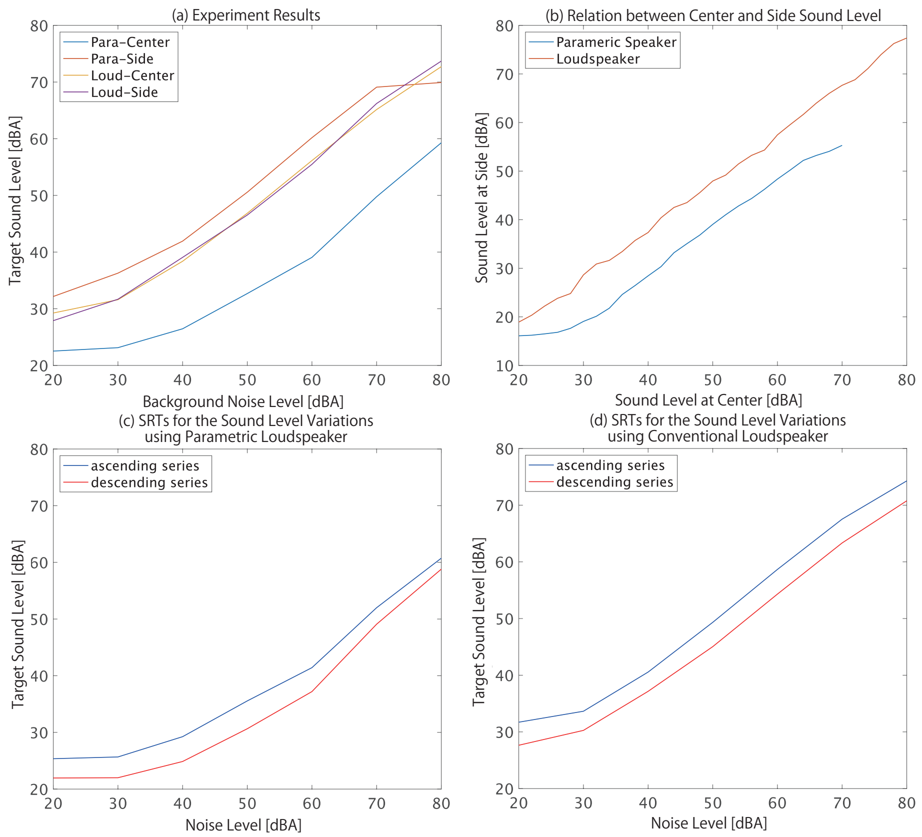

Figure 2.5: Experimental results of the Speech Recognition Threshold - (a) Experimental results, (b) Relation between Center and Side sound level, (c) SRTs for the sound level variations using parametric loudspeaker, (d) SRTs for the sound level variations using conventional loudspeaker

understand the sound. This feature of the conventional loudspeaker is appropriate for widespread announcement applications. However, if the sound is intended for only a specific individual or the surrounding environment must be maintained at a quiet level, this device is not suitable.

For the parametric loudspeaker, the SRTs were different between the center and side placements. The average SNRs of the SRTs are approximately -13.87 dBA and 1.45 dBA for the center and side placements, respectively. The sound level required to present an informative sound to a target person was approximately 12.46 dBA lower than that necessary in the case of the conventional loudspeaker. The results show that the person around the target listener cannot hear a considerable amount of the information contained in the sound. Moreover, the SRT of the non-target person was approximately 2.79 dBA in the case of the conventional loudspeaker. Furthermore, the sound level encountered by the non-

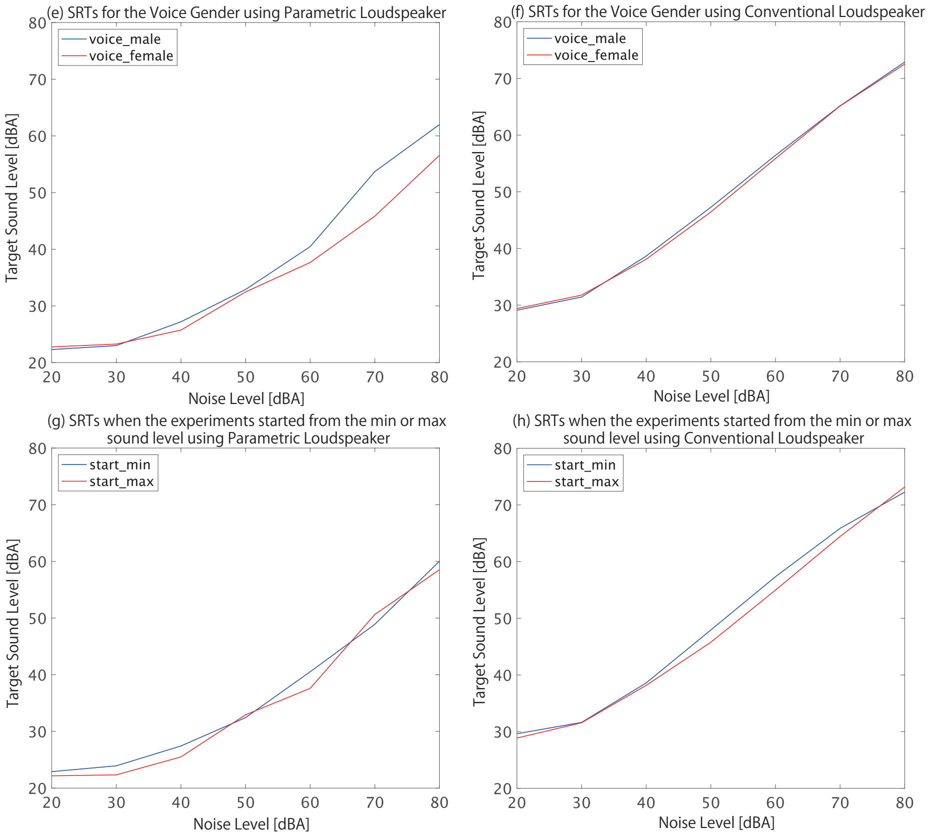

Figure 2.6: Experimental results of the Speech Recognition Threshold - (e) SRTs for the voice gender using parametric loudspeaker, (f) SRTs for the voice gender using conventional loudspeaker, (g) SRTs when the experiments started from the minimum or maximum sound level using parametric loudspeaker, (h) SRTs when the experiments started from the minimum or maximum sound level using conventional loudspeaker

target person when the target person heard the sound is shown in Figure 2.5 (b). These levels were varied in increments of 2 dBA and measured through the sound level meter. The average level heard by the non-target person was -2.26 dBA and -10.92 dBA lower than that heard by the target person for the conventional loudspeaker and parametric loudspeaker, respectively, corresponding to an average difference of 8.66 dBA.

To perform an extensive analysis for the target person, the SRTs for the sound level variations in descending and ascending series were calculated. The SRT of the descending series referred to the sound level at which a word was heard once and then not heard again. The SRT of the ascending series referred to the sound level at which the word was no longer heard and then heard again. The results for the conventional loudspeaker and parametric loudspeaker are shown in Figures 2.5 (c) and (d), respectively. In the case of

Table 2.1: The calculated sound level received by people in the center and surrounding area, including environmental noise, and the differences between conventional loudspeaker and parametric loudspeaker.

	Conventional Loudspeaker			Parametric Loudspeaker			
Noise	SRT	Side	Total	SRT	Side	Total	Diff
20	29.25	27.25	28.00	22.53	14.05	20.98	7.02
30	31.60	29.57	32.80	23.13	14.58	30.12	2.68
40	38.38	36.27	41.53	26.47	17.55	40.02	1.51
50	46.85	44.63	51.10	32.68	23.10	50.01	1.10
60	56.15	53.82	60.94	39.07	28.79	60.00	0.93
70	65.15	62.70	70.74	49.78	38.33	70.00	0.74
80	72.72	70.18	80.43	59.28	46.81	80.00	0.43

unit:[dBA]

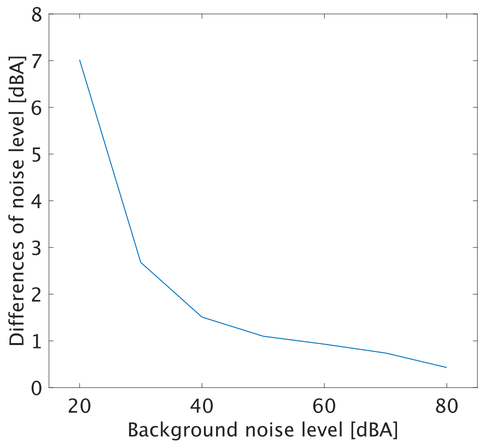

Figure 2.7: The graph of differences calculated in Table 2.1

the conventional loudspeaker, the average SNRs of the SRT were approximately -3.07 dBA and 0.82 dBA for the descending and ascending series, respectively. In the case of the parametric loudspeaker, the average SNRs of the SRT were approximately -15.06 dBA and -11.43 dBA for the descending and ascending series, respectively. These results showed that once a person understands the sound, it is possible to continue understanding it even if the sound level is decreased henceforth.

Furthermore, the SRTs associated with different voice genders were calculated. The cases involving the conventional loudspeaker and parametric loudspeaker are illustrated in Figures 2.6 (e) and (f), respectively. In the case of the conventional loudspeaker, the SRTs under various background noises were similar. However, in the case of the parametric loudspeaker, the SRTs were different for different voice genders under noise levels of 60 - 80 dBA. The SRTs for the male and female voice were 40.47 dBA and 37.67 dBA at a noise level of 60 dBA, 53.70 dBA and 45.83 dBA at a noise level of 70 dBA, and 62.00 dBA and 56.57 dBA at a noise level of 80 dBA, respectively. These results show that target individuals find it more difficult to hear the target word spoken by a male voice than that spoken by a female voice in a noisy environment.

The SRTs when the experiments were started from the minimum or maximum sound level are shown in Figures 2.6 (g) and (h), respectively. The SRTs were noted to be similar under various noise levels.

2.5 Discussion

Considering the results, the sound level received by the non-target individuals in the vicinity of the target individual, including the background noise, was calculated. For example, in a 60 dBA noise environment,the SRT of the target person was 56.15 dBA for the conventional loudspeaker. Through a linear approximation of the sound level relationship of the data shown in Figure 2.5 (b), the level received by the non-target person was calculated to be approximately 53.82 dBA. In other words, the non-target person was exposed to a combined noise level of 60 dBA + 53.82 dBA, and the result was approximately 60.94 dBA. In contrast, the sound level at which the target sound could be heard in a 60 dBA noise environment with the parametric conventional loudspeaker was approximately 39.07 dBA. In this case, the non-target person received a sound level of approximately 28.79 dBA. The overall noise level for the non-target person was 60 dBA + 28.79 dBA, and the result was approximately 60.00 dBA. Compared to that when the sound emitter was the conventional loudspeaker, when the sound emitter was the parametric loudspeaker, the people around the target person received a sound level that was approximately 0.94 dBA smaller. The same calculations for different noise levels are summarized in Table 2.1 and Figure 2.7. The values indicate that a smaller background noise level corresponds to a larger difference in the overall noise level. Compared with the conventional loudspeaker, the parametric loudspeaker is expected to enhance the spatial comfort level of non-target individuals in terms of the acoustic comfort, as depicted in Figure 2.8. Moreover, the overall sound level is decreased, which allows for more flexibility in the design of the sound space, for instance, through the addition of other announcements or environmental sounds. Furthermore, when considering information masking, when the level of one sound is reduced, it is easier to hear other sounds. This is effective even when the overall sound level has not changed much. In summary, it is quantitatively confirmed that the use of the parametric loudspeaker allows for more flexibility in the design of the sound space. This aspect can facilitate the space design in a quiet environment such as a museum, allowing people to listen to the explanations without additional noise.

From the above, the answer to RQ.1: "How much is the noise level reduced by the use of parametric loudspeakers?" is that noise reduction is more pronounced with lower background noise levels. For instance, if the background noise is 20 dBA, the SRT for non-target listeners was 27.25 dBA with conventional loudspeakers, but it was 14.05 dBA with parametric loudspeakers. When combined with the background noise, the difference in sound level was 7.02 dBA.

Additionally, the SRTs for different voice genders differed, likely because of the fre-

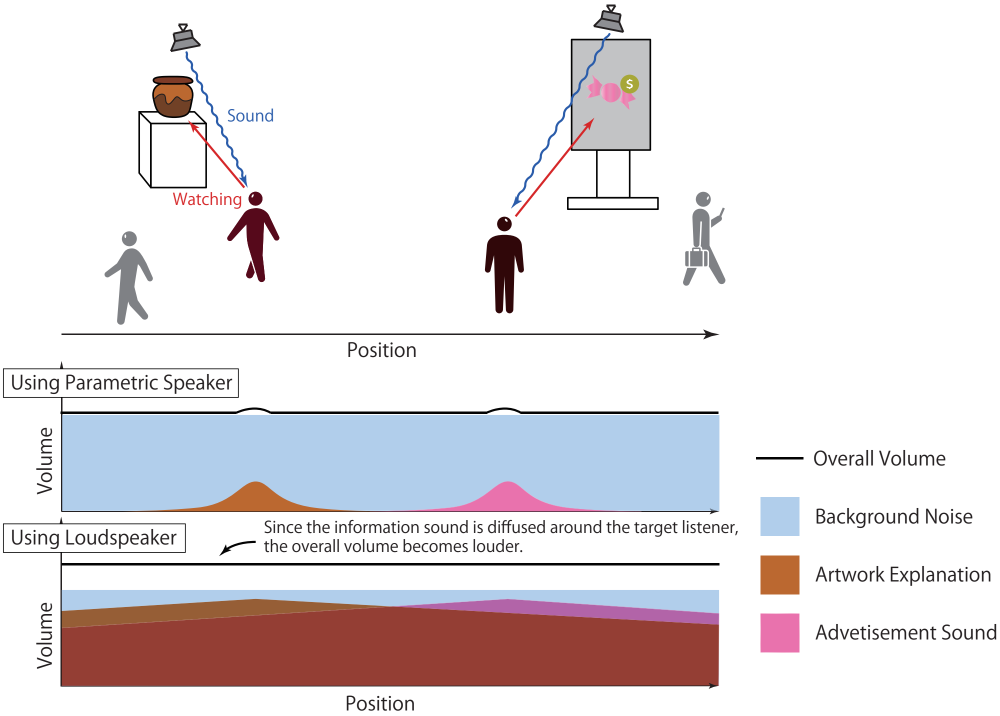

Figure 2.8: Discussion of overall sound level using parametric loudspeakers and conventional loudspeakers

quency of the parametric loudspeaker and fundamental frequency of the voice gender. The frequency characteristics of the conventional loudspeaker and the parametric loudspeaker are shown in Figures 2.9 (c) and (d), respectively. The measurement setups of (c) and (d) are below; Sampling frequency: 10kHz, Signal: 0 – 2.5kHz upchirp signal in 5 seconds, Microphone: Bruël & Kjær 4939. Above 100 Hz, the frequency characteristics of the conventional loudspeaker were nearly flat. In general, for a parametric loudspeaker, it is easier to output sounds with frequencies higher than 300 Hz. The fundamental frequency f_0 of the male voice is lower than that of the female voice. Consider the example of the word "*o-ya-su-mi*". The fundamental frequencies of the first syllable, 'o', are shown in Figures 2.9 (a) and (b). The four voices in the datasets corresponded to different persons [62]. The fundamental frequencies were (a-1) 120.30 Hz and (a-2) 113.74 Hz for the male voice, and (b-1) 213.33 Hz and (b-2) 235.29 Hz for the female voice, individually. Notably, the pitch affects the fundamental frequency and its overtone. Therefore, the sound with low f_0 is more difficult to output from the parametric loudspeaker than the sound with high f_0. Moreover, the soud with low f_0 may be masked by the loud background noise. These factors associated with the f_0 lead to differences in the audibility. Therefore, when emitted by the parametric loudspeaker under a noisy environment, information spoken by the voice of high f_0 or low f_0 equalized to a higher pitch is easier to understand. These hypotheses must be examined in further work.

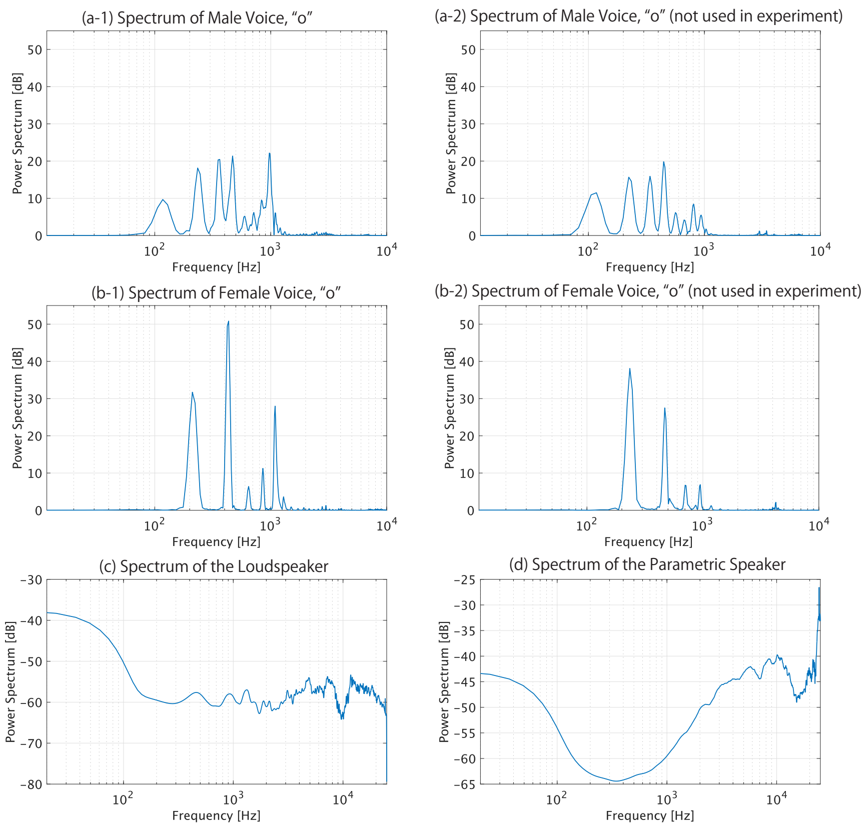

Figure 2.9: Spectra: (a) Male voice, (b) Female voice, (c) Conventional Loudspeaker, (d) Parametric Loudspeaker

2.6 Summary

The objective of this section is to evaluate the degree of noise reduction by using the parametric loudspeaker and answer to the research question 1 "How much is the noise level reduced by the use of parametric loudspeakers?". Besides, to clarify the differences pertaining to sound level variation and voice gender, the SRTs of a conventional loudspeaker and a parametric loudspeaker were measured at various noise levels. According to the results of experiments involving 15 participants, the target listener's SRT for a parametric loudspeaker is smaller than that for a conventional loudspeaker. The non-target person's SRT for a parametric loudspeaker is higher than that for a conventional loudspeaker. Moreover, the SRTs of the descending series are lower than those of the ascending series. In addition, the SRTs of the target sound with low f_0 is higher than that associated with that with high f_0.

The findings highlight that (1) the parametric loudspeaker can enhance the spatial comfort level for non-target individuals; (2) once a person understands the sound, it is possible to continue understanding it even if the sound level is lowered; (3) it is more difficult for target individuals to hear the target word spoken in a low f_0 than that spoken by a high f_0 in a noisy environment.

The findings can help facilitate loudness control and the design of spaces with different sound perceptions for different individuals in a crowd.

Chapter 3

Auditory Display on Digital Signage

3.1 Chapter Objective

From Chapter 2, it was demonstrated that the use of parametric loudspeakers can reduce noise levels and potentially contribute to spatial comfort. Therefore, in this chapter and the following Chapter 4, the study examines whether auditory displays according to individual and spatial context can resolve the trade-off between utilization effectiveness and spatial comfort. This is tested using specific scenarios. In this chapter, I will examine the subject of advertising signage, which is often installed in commercial facilities.

In public places, various advertisements are displayed using digital signage to make those who pass by aware of products and services. Consequently, people demonstrate increased recognition, interest, memory, and desire for the advertised products, eventually purchasing them [66]. Advertisement impression will lead to the recall of products and services in someone's memory. Therefore, it is important to make advertising recallable to influence purchasing behavior [67].

The implementation of digital signage in advertising makes it easier to update information as compared with an analog board, such as a leaflet. It is possible to display multiple pieces of information sequentially [68]. Digital signage has made it easier to obtain information in public spaces and has increased the ubiquity of information [69, 70, 71, 72]. However, digital signage attracts less attention from passersby, which is called "Display Blindness"[71, 73]. This makes advertising less recallable. Advertisements are also perceived as irrelevant to the passersby and are displayed passively rather than being picked up by them [71]; they are also generally presented without any sound. It has been found that the presence of sound is effective in making people notice the signage[74]. It has also been reported that the presence of background music brings changes in behavior, such as people staying longer and walking closer to the sound source[75, 76]. In other words, emitting sound may have a positive effect on advertising. This is because the advertising information can be perceived both visually and audibly, thus making it easier to make an impression. I consider that this effect can be extended to digital signage as well as online, and the presence of sound makes passersby recallable.

However, the emission of sound produces noise, which can be uncomfortable for passersby in the surrounding space. This leads to a decrease in the acoustic comfort level. In this study, I consider temperature and light as general conditions, and look specifically into the spatial comfort derived from sound [3, 4]. Sounds that match the ambience of the place, such as natural sounds and live music, are less likely to cause discomfort, whereas artificial and mechanical sounds are particularly likely to cause irritation [4, 77, 78]. Furthermore, even the music that is played in a store or during an announcement tends to be disliked if its sound level is high [79]. While the sounds have not been investigated, as illustrated in Figure 3.1 (b), they can be perceived as noisy when they are emitted towards people who are unwilling to pay attention to them. A high level of noise has negative effects not only on the mind, but also on the body; people subjected to it exhibit reduced co-operation, and may even suffer from health problems [80, 81]. In some cities, guidelines have been established to suppress noise in the surrounding space [82]. Therefore, digital signage should be prevented from constantly emitting sound.

In this chapter, I implemented two-steps proposed systems that delivers sound only to a person by using directional sound [52, 56, 83]. The first step is an on-demand pinpoint audio system, with a new on-demand feature that changes the sound level according to the degree of attention. This aims to control the amount of information by presenting the sound of the advertisement only during the time when the user is attentive to the advertisement, which is decreasing the overall levels and keeping the exposure only to the attentive individuals. The second step is a pinpoint auditory glimpse, which to passersby approaching the signage for a short time and changes the sound level depending on the degree of their attention in response. This system will be integrated into the first step system - the on-demand pinpoint audio system.

Participant experiments using the proposed system will examine how an auditory display according to individual context can change purchasing behavior and spatial comfort. Two specific research questions are as follows:

2-1 Does varying the sound level of an ad's auditory display improve recognition?

2-2 Does varying the sound level of an ad's auditory display make the acoustic environment less comfortable?

3.2 Related Work

3.2.1 Relationships between Advertisement Recognition and Signage

In this section,I will first explain how signage is related to advertisement recognition. Advertising is a tool used to promote products and brands, which can lead to consumer purchasing behavior. Recognition of the advertisement is a crucial and fundamental step in the purchasing process. There are three main purchasing models: the AIDMA model [84] corresponding to off-line, the AISAS model [66] corresponding to online, and Fogg's behavior model [85] corresponding to offline and online commonly.

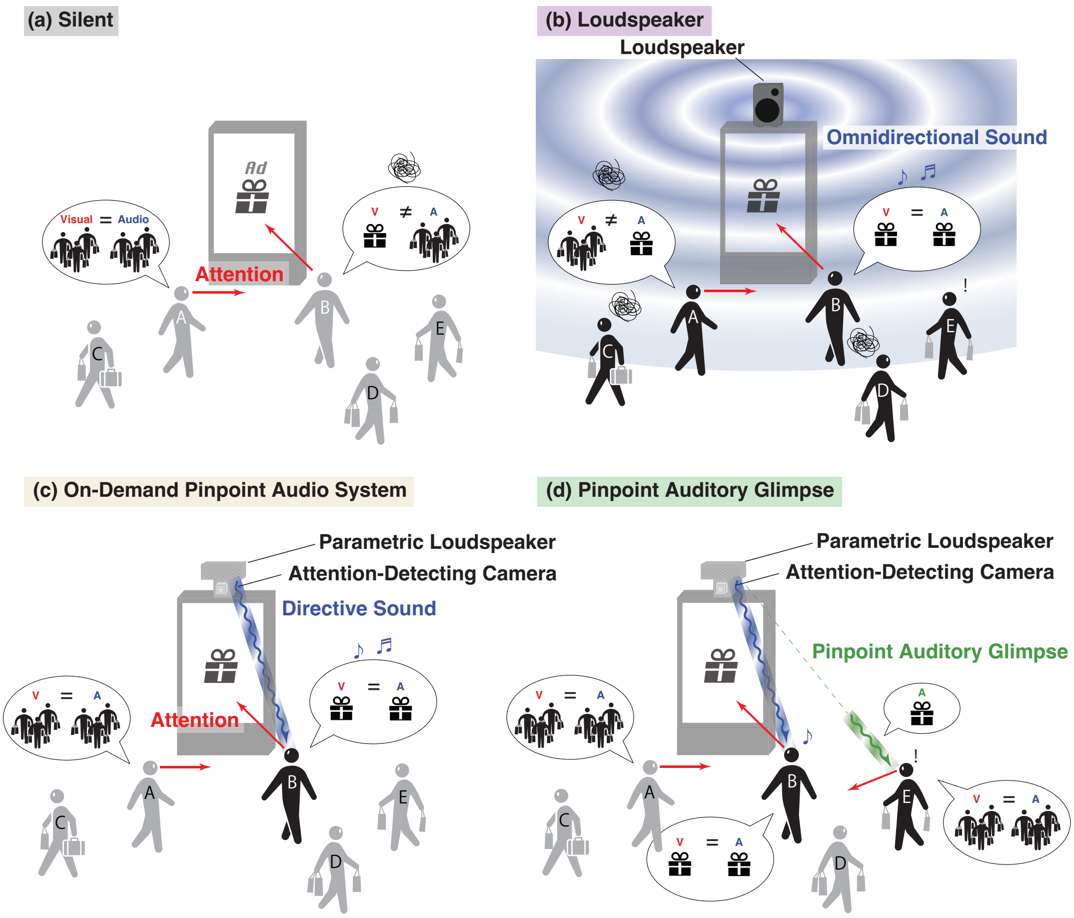

Figure 3.1: Conventional and Proposed Signage System - (a) Silent, (b) Conventional Loudspeaker, (c) On-Demand Pinpoint Audio System, (d) Pinpoint Auditory Glimpse

The flow of purchasing behavior can differ between offline and online environments. Offline purchasing behavior typically follows the stages of Attention (pay attention to an advertisement), Interest (interested in a product or service), Desire (desire to buy), Memory (memorize), and Action (buy in a physical store) [84]. The online flow is referred to as Attention, Interest, Search (search on the internet), Action (buy it online), and Share (spread the product or service on social networking sites or other sites) [66].

Attention, in particular, is defined as the behavior of "a consumer who notices a product, service, or advertisement" that leads to action, which is defined in this instance as "becomes a firm decision to make a purchase." There are cases where a purchase is made online after seeing an ad in digital signage. Attention, which is the scope of this chapter is to increase, and interest are common to both offline and online flows, as noted in Figure 3.2. However, not everyone who experiences attention moves on to action, but rather the percentage of those who do is decreasing rapidly. In other words, it is important to greatly increase

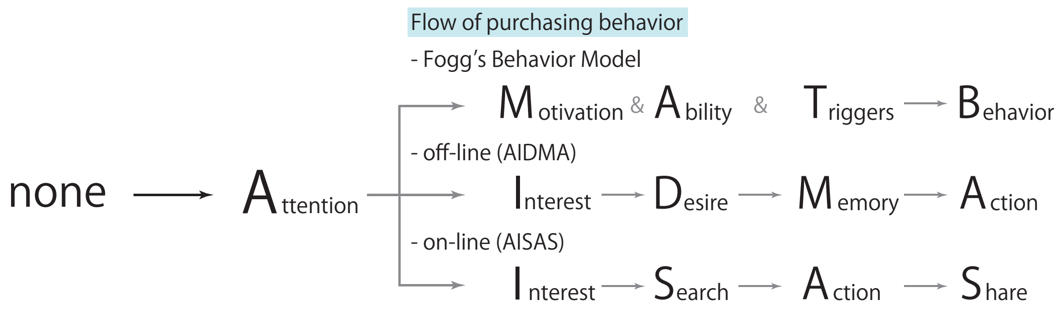

Figure 3.2: Flow of purchasing behavior - Fogg's behavior model, AIDMA, and AISAS

attention to increase action.

The Fogg's Behavior Model consists of three elements necessary to elicit human behavior: "motivation," "barriers to action," and "triggers." Once these elements exceed a certain threshold, they lead to "behavior." In this case, the "behavior" refers to a purchasing behavior. The advertising aims to appeal to these three elements and to increase purchasing behavior beyond the threshold. In this sense, it is also important to increase attention [85, 86].

Furthermore, the presence of in-store signage tends to make products seem like a bargain and increase customers' unplanned purchases [87, 88, 89]. This means that the signage can enlarge the actions in the purchase funnel. Consequently, it is both important and effective to stimulate people to actually see advertising signage. Hence, the objective of this chapter is to address display blindness to improve the recognition of advertising signage.

3.2.2 Public Display and Display Blindness

The previous section established that it is important for advertising signage to be seen. This includes not only advertising signage but also other information presentation displays and interactive displays and getting people to notice the presence of the display is the first step toward achieving effectiveness. Therefore, previous studies have contributed to the resolution of display blindness by devising various visual stimuli.

Huang et al. reported that the use of video content and signage at the eye level have made it easier to attract the attention of passersby [71]. Using animation in the content also improves the attention level [90, 91]. In addition, the vivid colors of the images make the display stand out and be seen easily [73, 92, 93]. Furthermore, visual stimuli such as changing luminance and the appearance of new objects also increase attention and contribute to increased interaction [72]. Studies which included a footprint positioned to guide the interaction [94] or physically waved a hand to call passersby [95] successfully notified people of the display's interactive elements and encouraged the passersby to initiate interactions. These studies addressed the first click problem [92], which refers to the difficulty of initiating interaction with signage. It has also been reported that placing mysterious objects in front of the display, completely unrelated to the display itself, can also attract attention to signage [96].

However, many passersby to whom the ads are presented have a negative attitude that the ads are of no interest to them. Field observations have reported that this has led many people to not look at ad signage in public spaces, even in prominent locations such as escalator signs, or to look at the signage only briefly as part of the scenery [97]. This means that there is a limit to the effectiveness of location devising, and it is obvious that people are even less likely to look at small signage installed among product shelves or similar locations. Therefore, in this study, I aim to overcome display blindness by using sound as an alternative to the traditional visual approach.

Practical experience, in Japan, the author's home country, digital signage has been used everywhere in the city. In Shibuya, a typical example, several digital signage panels are installed at a single scramble intersection, and most of them play sound. The annoyance of this image and sound has been raised as a problem in various ways. Jörg Müller, who has conducted research on display blindness, has described video advertisements in Shibuya and Times Square in New York City as "you face an overkill of advertisements, letting the vision presented in the movies seem ever more realistic" [98]. Yasushi Akimoto, who is a famous ryricist in Japan and abroad, wrote the lyrics about the sound environment surrounding signage as following: "Ditigal signs are everywhere in town / It's so annoying / All these messages poured at once make me wonder / Am I a multi-task listener or what?" [99]. What these suggest is that the bother of digital signage has been described not only as a phenomenon but also as a cultural form. In this way, the attempt to maximize the utilization effect of digital signage has become extremely complicated. This chapter uses digital signage as a specific example to eliminate the adverse effects of over-pursuing the effectiveness of use and the trade-off relationship with spatial comfort.

3.2.3 Audience Funnel

This chapter uses audio as an alternate solution to display blindness than the visual modality. However, when the sound is heard by people who are not interested or who are not shopping, it will become noise and reduce the acoustic comfort level. Thus, it is necessary to present sound with the appropriate timing and at the appropriate sound level to notify a passerby of the signage's presence. Accordingly, it is necessary to consider the behavioral characteristics of the passersby who are exposed to the sound as a result of the display.

Prior studies have made various attempts to model the display and its effect on the audience's behavior. In the studies of interactive displays, there are systems that divide the distance between the user and the display into zones and change the presented content accordingly. This improves information accessibility and personalization. Thus, it is easier to draw the user in and keep them interacting [100, 101]. There is a model that categorizes the concept and includes an interaction phase called the audience funnel [102]. This is meant to model the user's typical interaction-related behavior, which has allowed for a calculation of effectiveness based on the percentage of action phases, such as people passing by the display, initiating interaction, and continuing to the end. There is also research that creates a follow-up action at the end of the interaction based on the audience funnel to make people re-engage [103]. This is indeed a system that is based on the behavioral model of the audience, and its effectiveness has been verified.

Regardless of which model is used, they are all built on the idea that users have different degrees of interest in the display, and that it is important to present different information to each user. Based on this idea of the audience funnel, the second-step system creates a model of passerby behavior when confronted with signage using a pinpoint auditory glimpse, which is called "User-Glimpse Funnel". I will also discuss the behavior of participants for each User-Glimpse Funnel through experimental observations.

3.3 On-Demand Pinpoint Audio System

3.3.1 Section Objective

In this section, I propose an on-demand pinpoint audio system. This aims to control the amount of information by presenting the sound of the advertisement only during the time when the user is attentive to the advertisement, which is decreasing the overall levels and keeping the exposure only to the attentive individuals. The visual and audible senses of passersby can be constantly matched by emitting sound only toward a specific person, as shown in Figure 3.1 (c). Therefore, even in public spaces, the desired behavior from individuals, such as watching advertisements or enjoying shopping, can be realized. However, the psychological changes that occur when using this system have not yet been discussed. In particular, in the conventional silent advertisements displayed in public spaces, the vividness of videos is an important impression factor [73, 92, 93]. To the best of my knowledge, in the case of videos with sound, the recallability factors among features such as video and sound have not been examined.

The objective of this section is to employ an on-demand pinpoint audio system for displaying advertisements to examine the influence of advertisement recallability and acoustic comfort. The categories of advertisements that affect the advertisement recallability and acoustic comfort when digital signages emits sound was also investigated. Advertisement recallability is self-reported, and acoustic comfort is a subjective evaluation of the degree to which the people in the room are comfortable with the sound in the environment. This provides new insights into the psychological changes that advertising on digital signage can bring to people.

3.3.2 System Overview

The on-demand pinpoint audio system was developed to provide an appropriate amount of advertising information based on the conditions of the passersby. The system overview is depicted in Figure 3.3.

A camera is used to track the person's spatial coordinate position, and servo motors rotate the parametric loudspeaker in azimuth and elevation to keep it pointed at the person's area, thereby preventing the sound from reaching other people. This enables pinpoint sound presentation. In the case of multiple passersby, the master device manages the sensing data and passes the data to the slave device via socket communication. Based on this data, the system tracks the passersby individually to present them with the advertisement sound.

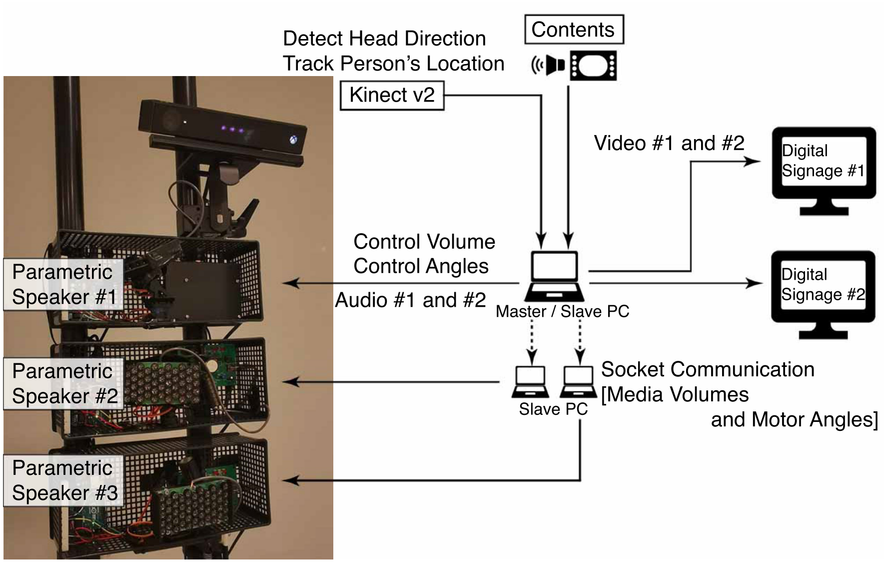

Figure 3.3: System overview of on-demand pinpoint audio system

The orientation of a person's face is also detected by the camera. Based on the orientation of the head, camera position, and the digital signage, it is calculated whether the direction of attention is toward the digital signage, and if not, how far away it is. According to the position of the attention, the sound sound level of the advertisement on the digital signage is changed. If the user is completely focusing on the display, the maximum sound level is output, and if the user is not looking at the display, the level decreases as the position of the attention direction is further away from the signage. A flowchart of passerby's attention and sound level is shown in Figure 3.4. If there is more than one digital signage, only the ads displayed in the direction of attention can be heard by the passersby, thus, preventing the noise caused by other advertisements from reaching them. The ad sound is output even at a minimum sound level to trigger the initial attention and is not completely muted.

The servo motor employed in this section was SG-90, which moves in 0.1° increments. The camera was Kinect v2, and the sensing during system runtime was approximately 30 fps. The parametric loudspeaker was a Tri-State K-02617 with as many as 50 transducers (AT40-10PB3) connected to it. Three microcomputer boards were used, which were Arduino UNO R3. Digital signage was SHARP PN-E703. The control application was developed in C++ and executed on Windows 10.

3.3.3 Advertisement Categories and Selection

Advertisement Categories

Besides the research questions 2-1 and 2-2, this section clarifies the type of features of advertising in digital signage will sound affect recallability. Here details the selected advertisements that were used in the experiment. The advertisements were selected from those

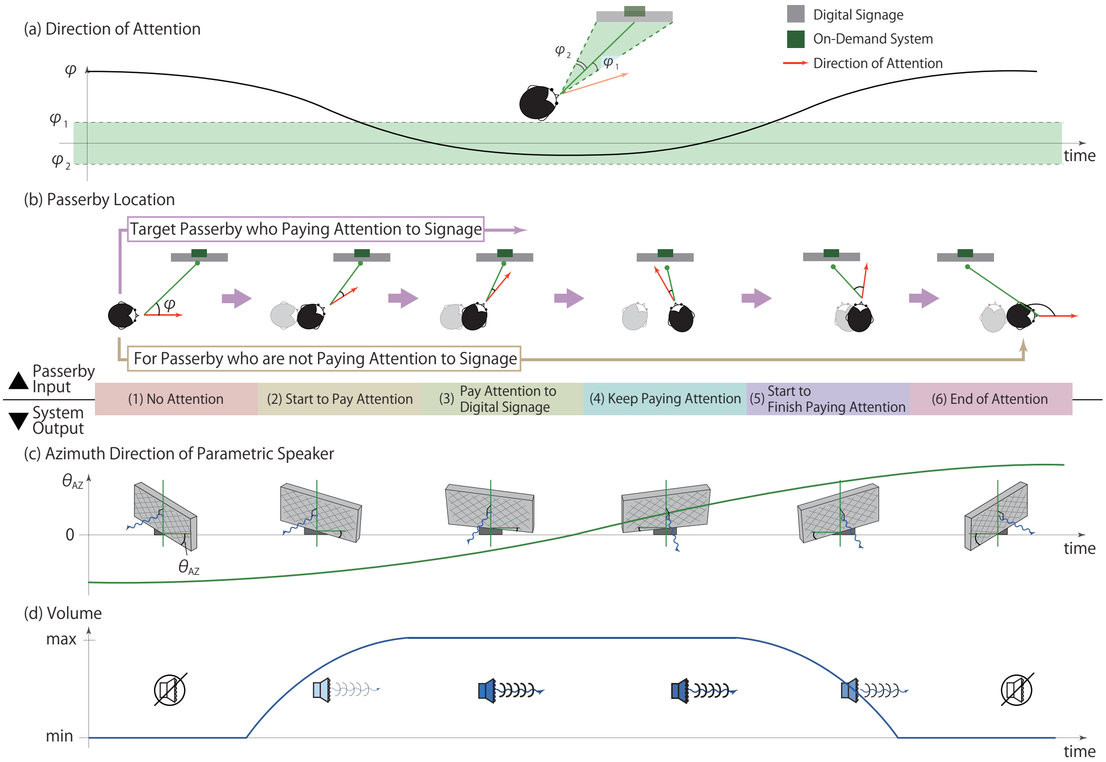

Figure 3.4: Flowchart of the passerby and on-demand pinpoint audio system

actually aired in the TV channel and were 15 s in duration. In this study, the video and audio of the advertisements were each classified into two categories as shown in the Figure 3.5: "subdued color" and "vivid color," and "narration dominating" and "music dominating", respectively [73, 104, 105]. In particular, advertisements in the "narration dominating" category predominantly contained an explanation in human voice. By contrast, "music dominating" referred to advertisements in which music was played almost entirely without any explanation by a human. Quantitative categorization of the advertisements was into two categories for video and two categories for audio, thus, resulting in four categories when multiplied together.

Quantitative Evaluation

First, I performed a quantitative classification. For the video categories, based on the hue, saturation, and value space, I calculated the hue and saturation variances of all pixels, which were measured per second of an advertisement image, for 15 images in total. The software was self-coded in Python. A high value of hue variance indicated that the advertisement used various colors. In contrast, a higher value of saturation variance indicated that the advertisement displayed clearer colors. In the sound categories, I selected advertisements with a narration duration of at least 10 and 0 s for "narration dominating" and "music

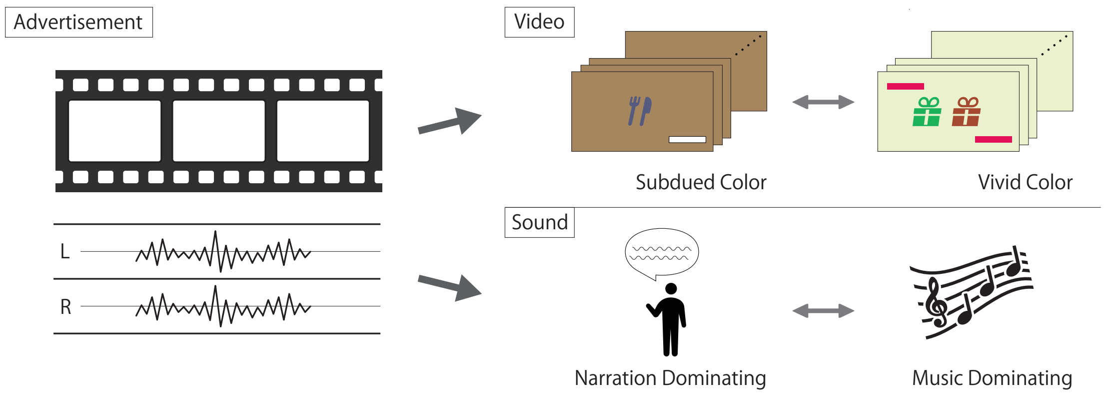

Figure 3.5: Video and Sound categories

Table 3.1: Quantitative evaluations of the selected advertisements

	Video Category	Sound Category	Time of Narration [s]	Hue Var.	Saturation Var.
Ad 1	Subdued Color	Narration Dominating	12:13	940.326	2990.611
Ad 2	Subdued Color	Music Dominating	0:00	2486.089	3148.739
Ad 3	Vivid Color	Narration Dominating	11:11	5277.715	5269.798
Ad 4	Vivid Color	Music Dominating	0:00	5927.326	2120.090

dominating", respectively. Finally, by authors, a dozen advertisements were selected that met each of these evaluation criteria, and so further videos were not considered.

Questionnaire Evaluation

I verified whether the quantitative classification was subjectively correct based on the questionnaire. After watching two average advertisements, seven-point Likert scale questionnaires were provided for each of a dozen advertisements on the scales of "subdued–vivid" and "narration–music", and was there a neutral value set at four. In addition, a seven-point Likert scale questionnaire was created on the preferences to ensure that the preferences of the audience and the uniqueness of the content of the advertisement did not affect the recallability evaluation in the experiment. Those advertisements whose preferences were close to the neutral were selected. Forty participants were collected via email and snowball sampling, and each ad was viewed on a Google form. A dozen advertisements were shown in random order.

From these questionnaires, I selected four advertisements, labeled as Ads 1-4, to be used in the experiment. The quantitative evaluations of the selected advertisements are listed in Table 3.1 and the questionnaire evaluations are shown in Figure 3.6. The relative amplitude of Ads 1-4 for 15 seconds is shown in Figure 3.7.

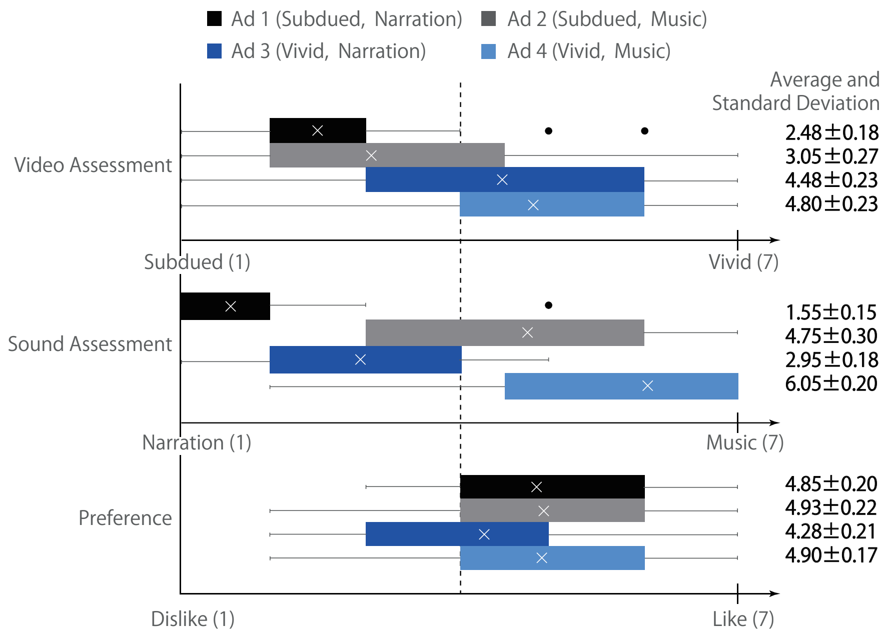

Figure 3.6: User evaluation for the video, sound classification, and preference of four advertisements

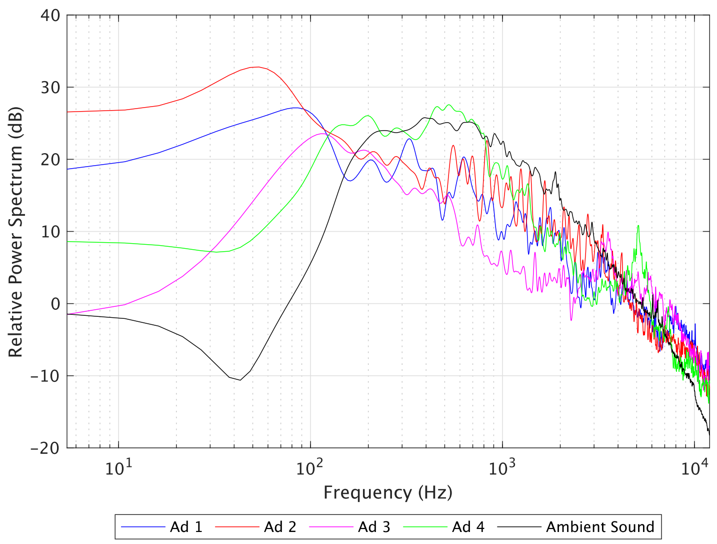

Figure 3.7: The relative amplitude for 15 seconds: Ads 1-4 and shopping mall sound used in participants experiment

3.3.4 Experiment

Measurements Method of On-Demand Pinpoint Audio System

This section confirms that the on-demand pinpoint audio system has the performance required in this study. The system must meet the requirement that the sound level changes according to the direction of attention while tracking a person. Accordingly, sine waved was emitted from the conventional loudspeaker and the on-demand pinpoint audio system in the anechoic chamber, recorded using a microphone (BEHRINGER ECM8000 [106, 107]), and then analyzed. The detail of the microphone is in Appendix 1. The experimental environment comprised two digital signages, as shown in Figure 3.8. The data of reverberation times was measured and its procedure is written in Appendix 2. The reverberation times were 116 ms in the anechoic chamber. The temperature of the experimental environment was approximately 26 °C. Accounting for overtone distortion, digital signages on the left and right presented sine wave of 600 and 1000 Hz, respectively.

Here, I describe the analysis procedure. The emitted sound at Persons A and B was recorded 10 times using a microphone, and the average data were analyzed for multiple time frequencies and power spectra. Subsequently, bandpass filters were applied to the 600 Hz signal for 1X, 2X, and 3X tones in ±10 Hz. The resulting power was then combined and multiplied by a 5-point moving average to produce a graph of the time-to-power spectrum of the 600 Hz component. This same procedure was also followed for the 1000 Hz component.

Person A walked from left to right, always facing forward. In the case of the on-demand pinpoint audio system, the closer the direction of attention was to the digital signage, the louder the sound presented was supposed to be. In contrast, Person B did not move and faced backwards, and hence, the direction of attention was completely off the digital signage. In this case, the sound was heard only when the current elevation position was the same as in the case of Person A. When using a conventional loudspeaker, it was assumed that Persons A and B could hear the sound from both the digital signages, regardless of their direction of attention. To measure the sound level at the listening position, Person A was recorded while holding a microphone in front of his head, and Person B was recorded with a microphone stand at the position of Person B, because he was facing backward.

Participants Experiment

The effects of advertising on the recallability and the acoustic comfort level were compared for different digital signage systems. Three digital signage systems were used: silent, conventional loudspeakers, and the on-demand pinpoint audio system. The experimental setup was prepared as shown in Figure 3.9. To recreate a public space in the best possible manner, an environmental sound of a shopping mall was played continuously in the experimental location. The relative amplitude of the sound for 15 seconds is shown in Figure3.7. The conventional loudspeaker used for the shopping sound reproduction was TEAC S-300NEO, which was installed at the center rear of the experiment environment. The noise levels at the experimental place when the environmental sound was being played were measured

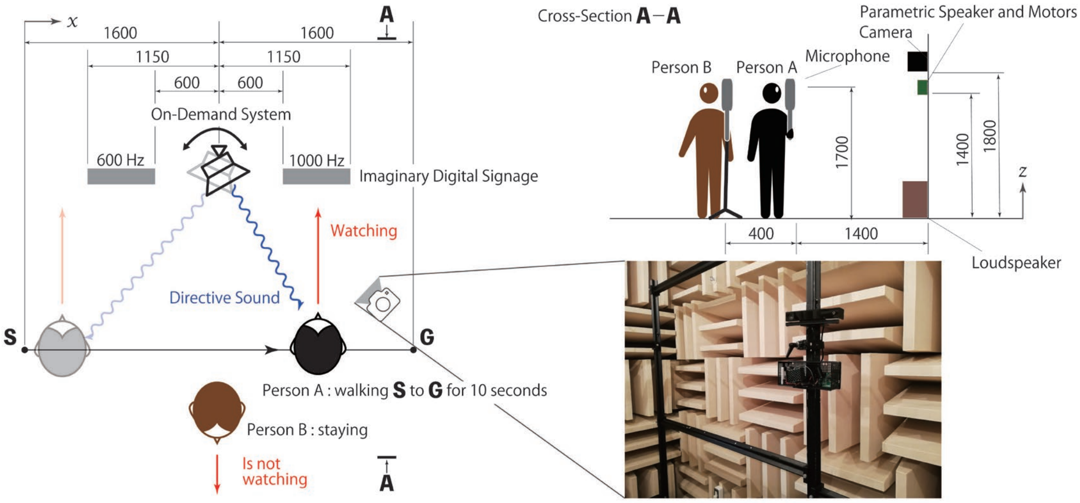

Figure 3.8: Overview of system performance measurement

using a noise level meter (RION NL-31, which is Class 1 compliant (IEC 61672-1)). The values were 36.48 and 62.14 dBA for the silent state and when the shopping mall sound was played as baseline stimulus, respectively. The data of reverberation times of the participants' experimental environment was also measured as following the procedure written in Appendix 2. The reverberation times were 652 ms. The temperature of the experimental site was approximately 18 °C.

A total of 30 participants, 24 males and 6 females with a mean age of 23.1, were considered as the test subjects in this study. The participants filled in their passing times on a paper at the beginning and goal desks, and walked in a single path between the tables, as shown in Figure 3.9. The participants were told that they were free to decide whether they wanted to watch the digital signage and that the quality of sound was not included in the evaluation. This was because this section was focused on the solution using sound and pinpoint presentation and ignored the parameter of sound quality. When using the on-demand pinpoint audio system, the parametric loudspeaker changed direction to track the user. To remove the dependence of user attention on the rotation of the device, even in the case of silent and conventional loudspeakers, the parametric loudspeaker tracked the user and rotated soundlessly.

To simulate how people would pass through a shopping mall, the experiment was simultaneously conducted in groups of three participants, and their order of walking direction was ensured to be random. Ads 1–4 of Table 3.1 were played for Cases 1–3 of Table 3.2 with two digital signages. The Ads and Cases were randomized to ensure maximum randomness. The conventional loudspeakers in the Case 2 used for advertisement sound were TEAC S-300NEO. The participants were evaluated at the end of each case based on the following aspects:

Q 1 and 2: Level of recallability of the advertisement in left (right) display.

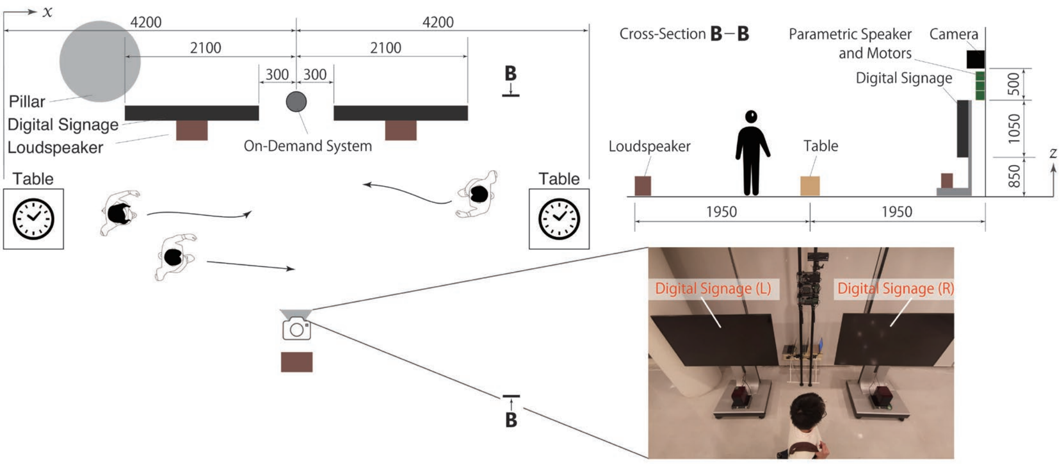

Figure 3.9: The schematic diagram and photograph of the experimental place

Table 3.2: List of sound output methods.

	Left Signage	Right Signage
Case 1	Silent	Silent
Case 2	Conventional Loudspeaker	Conventional Loudspeaker
Case 3	On-Demand Pinpoint Audio System	On-Demand Pinpoint Audio System

Q 3 and 4: Level of discomfort of the advertisement in left (right) display.

Q 5: Level of acoustic comfort.

Q 1–5 were assessed using a seven-point scale, where 1 was low and 7 was high. For Q 1 and 2, they were asked to subjectively rate how much the advertisement could be recalled. For Q 3 and 4, they were asked to subjectively rate how uncomfortable they felt about the advertisement. Advertisements are often present regardless of the intent of viewers; hence, users are likely to feel discomfort. Therefore, I also analyzed whether there was a change in the level of comfort depending on the system and advertisement category. For Q 5, they were asked to subjectively evaluate how comfortable they felt in the space related to the sound during the experiment. An additional comment box was provided as Q 6 for aspects

Table 3.3: The original text of the questionnaire in the participant experiment

	Question Text	
Q 1	How recallable was the advertisement on the left signage?	1 2 3 4 5 6 7
Q 2	How recallable was the advertisement on the right signage?	1 2 3 4 5 6 7
Q 3	How uncomfortable was the advertisement on the left signage?	1 2 3 4 5 6 7
Q 4	How uncomfortable was the advertisement on the right signage?	1 2 3 4 5 6 7
Q 5	How was the environmental comfort level regarding sound?	1 2 3 4 5 6 7
Q 6	(an additional comment box)	

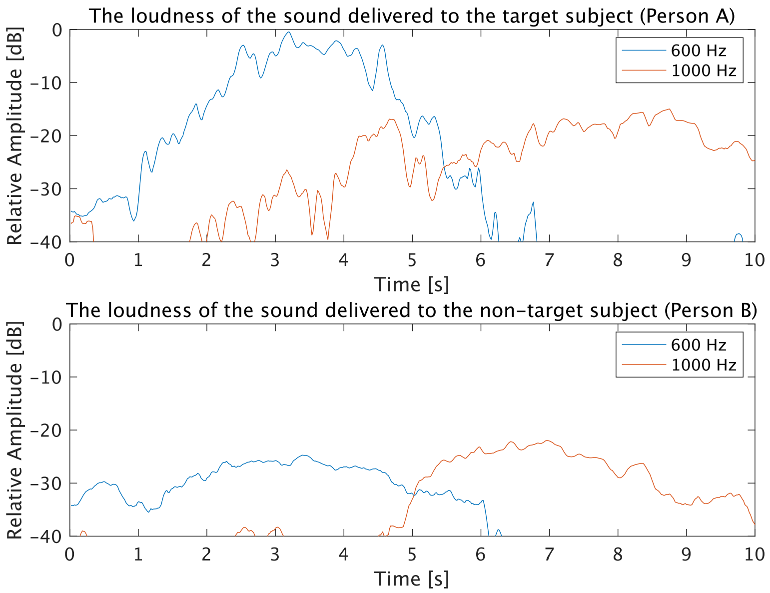

Figure 3.10: Behavior of the sound level on using on-demand pinpoint audio system

the participants noticed. The original text of the questionnaire is written in Table 3.3.

When conducting each of Cases 1-3, the left and right signage each play one of Ads 1-4 in a loop. For example, if three participants are engaging in Case1, the left signage plays Ad1 on a loop, the right signage plays Ad2 on a loop, and Ad3 and 4 are not played. This means that in the scenario, the participant experiences all of Cases 1-3, but within each Case, any two of Ads 1-4 will be played on a loop. The looped ads are randomly selected for each Case, and are designed to be exposed the same number of times through 30 participants. Therefore, the system evaluation for each case is 30 participants $\times$ 2 signages on each side = 60 evaluations. For the advertisement analysis, 60 evaluations / 4 ads = 15 evaluations per ad. The role of the three participants is just to walk randomly and change from left to right or vice versa for each Case1-3. This randomizes the order of viewing the signage.

This experiment was reviewed and approved by the ethics committee of the University of Tsukuba (Faculty of Engineering, Information and Systems, Permission number: 2019R310).

3.3.5 Results

Measurement Results of the On-Demand Pinpoint Audio System

The results of the on-demand pinpoint audio system are graphed in Figure 3.10. With the on-demand pinpoint audio system, Person A heard 600 Hz in the first half and 1000 Hz in the second half. As the parametric loudspeakers used were less powerful at 1000 Hz than at 600 Hz, there was a difference in their maximum sound pressure; nevertheless,

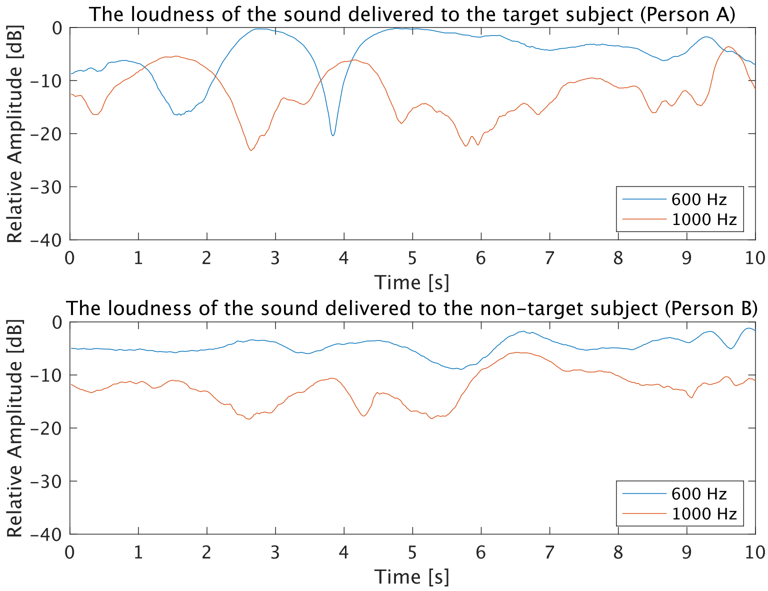

Figure 3.11: Behavior of the sound level on using conventional loudspeakers

the sound pressure variation was observable. Even while walking, a change in sound level occurred depending on the direction of attention. In the case of Person B, only 1000 Hz could be heard in addition to other negligible sounds after Person A passed near the center. Therefore, it can be confirmed that the on-demand pinpoint audio system only tracked Person A.

The results of the conventional loudspeakers are plotted in Figure 3.11. Both Persons A and B always heard both 600 and 1000 Hz of sine wave, regardless of their direction of attention.

Based on these results, I concluded that the on-demand pinpoint audio system did not diffuse the sound more than the conventional loudspeakers, and it was, thus, confirmed that the on-demand pinpoint audio system met the requirements for this study.

Effectiveness of the Systems

The number of evaluations for each system was 30 people $\times$ 2 left and right = 60 evaluations. The Tukey Honestly Significant Difference (HSD) Test was used to compare within systems. The results are shown in Figure 3.12.

The most recallable system was the advertisement using a conventional loudspeaker. A significant difference was found with the silent and on-demand pinpoint audio system. In addition, the on-demand pinpoint audio system was more recallable than the silent, and significant differences were also found. These indicates that although not as memorable as the conventional loudspeaker, the advertisement using the on-demand pinpoint audio system was found to be more recallable compared with the silent system.

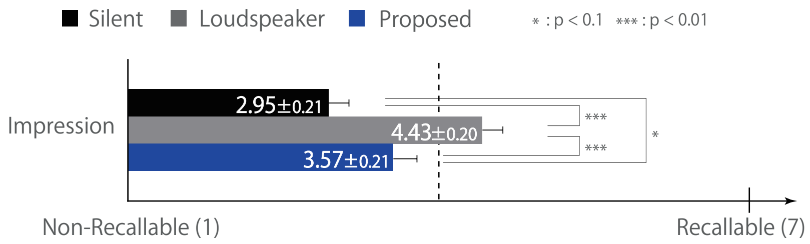

Figure 3.12: Results of self-reported recallabilities for each advertisement systems

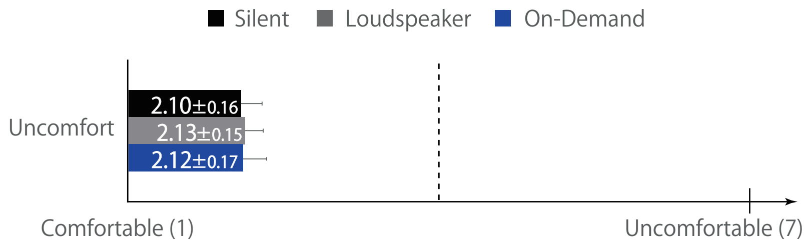

Figure 3.13: Results of the effects on discomfort for each advertisement systems

The results of advertisement discomfort are shown in Figure 3.13. Neither of these results is significantly different.

Thirty evaluations were obtained regarding evaluation of the acoustic comfort in terms of noise. Three cases were compared using Tukey HSD Test, and the results obtained are shown in Figure 3.14. As seen in the figure, no significant difference was found among these systems.

Effectiveness of Ads Features

An analysis was conducted for each category of advertisement for each of the self-reported recallabilities and discomfort levels of the advertisement. Fifteen evaluations were obtained for each system and each advertisement category. The Tukey HSD Test was used to compare and analyze the results between systems and advertisement categories.

First, the results of recallability for each video color category are shown in Figure 3.15. Comparing the systems, I observed no significant differences for the silent system and on-demand pinpoint audio system. In the conventional loudspeaker environment, advertisements with subdued colors were more recallable than vivid ones. Advertisements through conventional loudspeakers were more recallable than those in the silent and on-demand pinpoint audio systems in any video category. Additionally, conventional loudspeakers were more recallable than on-demand pinpoint audio systems for subdued-color advertisements.

The results of the effects on discomfort are displayed in Figure 3.16. None of these were significantly different.

The results of the recallability for each dominant-sound category are shown in Figure 3.17. Comparison within a system shows that music-dominated advertisements were more re-

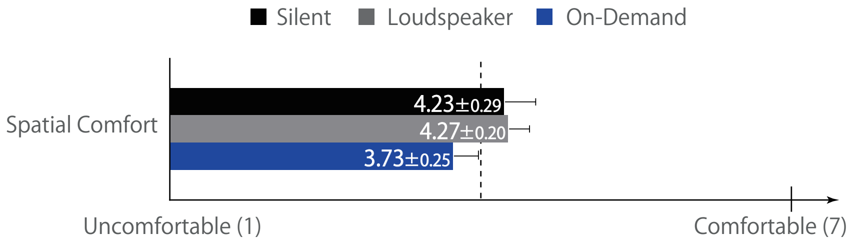

Figure 3.14: Results of acoustic comfort

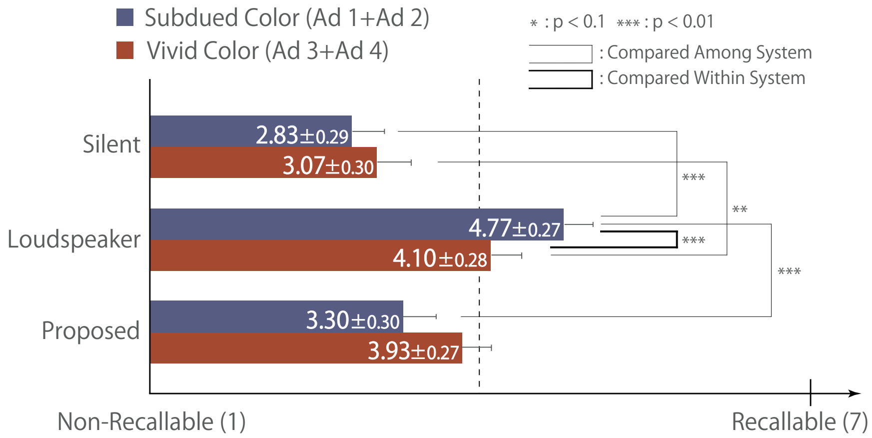

Figure 3.15: Results of self-reported recallabilities for each video color category

callable than narration-dominated ones in the case of conventional loudspeakers and on-demand pinpoint audio systems. Comparison within systems demonstrated that the advertisements with conventional loudspeakers were more recallable than those with the silent system in any sound category. In addition, in the on-demand pinpoint audio system, music-dominated advertisements were more recallable than those in the silent system.

The results of the effects on discomfort are shown in Figure 3.18, which indicate that none of these were significantly different.

3.3.6 Discussion

Advertisement Recallability and Discomfort

In summary, the experimental results showed that the on-demand pinpoint audio system was a more recallable advertising system than the silent system. In addition, although the discomfort level of the advertisement and the acoustic comfort were not shown to be different from the other systems, it can be seen that the noise level was quantitatively reduced compared to the conventional loudspeaker.

The reason for which conventional loudspeakers can leave a greater recallability is due

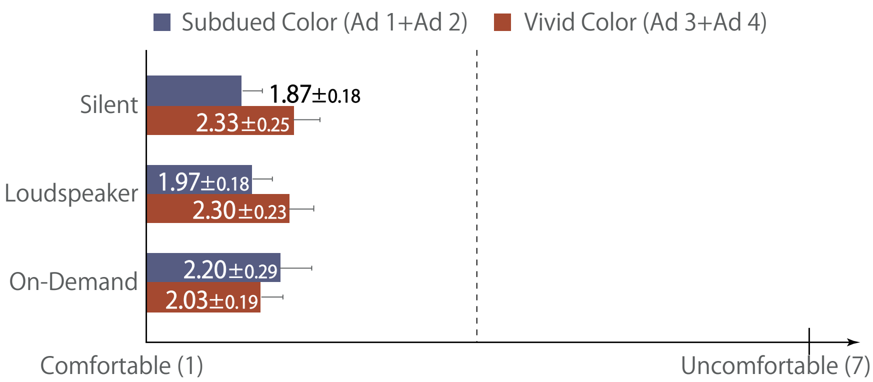

Figure 3.16: Results of the effects on discomfort for each video color category

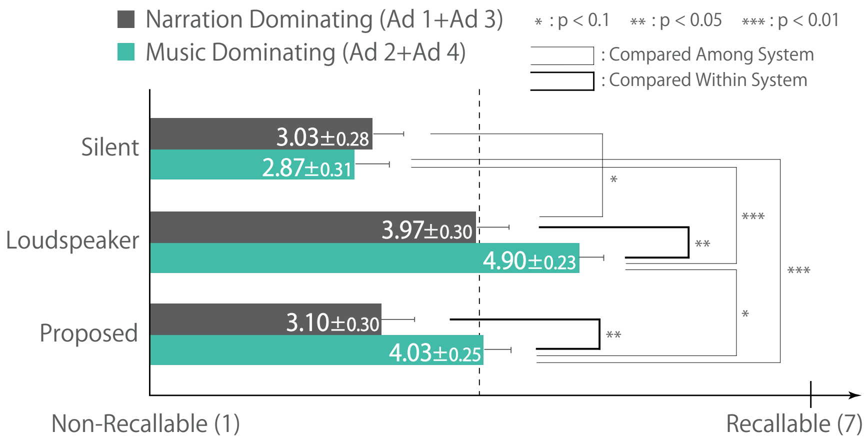

Figure 3.17: Results of self-reported recallabilities for each dominant-sound category

to their repetition effect. With conventional loudspeakers, the sound is repeated regardless of whether the passerby pays attention to the signage or not. I thought that this would make it noisy and uncomfortable, but I did not observe a difference in the participants' acoustic comfort. This may have been due to the short duration of the participants' stay. However, people who stay for a long period of time, such as mall clerks or game center employees, are always exposed to loud noise. Loud noise can not only produce discomfort but also pose a health risk [80, 81]. In addition, music played in stores and public address systems tends not to be preferred if it is too loud [79], and the same is considered for music-dominating advertisements. Therefore, it is still recommended to avoid placing conventional loudspeakers on any digital signage.

On-demand pinpoint audio systems can lead to a decrease in the noise levels perceived by people in the surrounding area, potentially lowering the risk of health problems. In

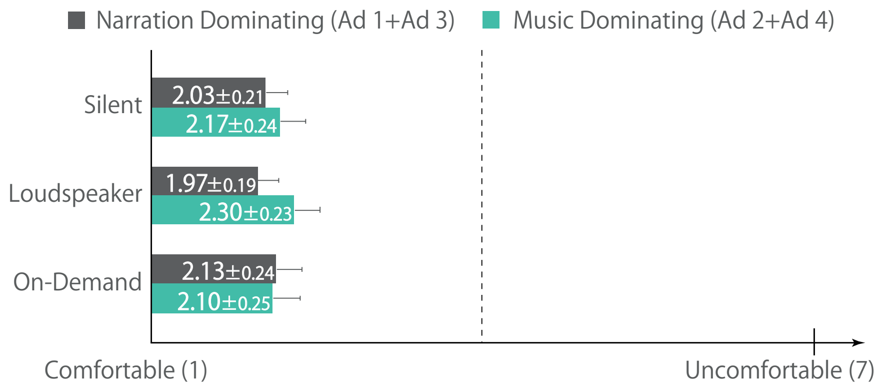

Figure 3.18: Results of the effects on discomfort for each dominant-sound category

summary, within the limitations of the public space, where noise must not be caused to maintain the acoustic comfort, the on-demand pinpoint audio system would be useful to improve the recallabilities of advertising.

Advertisement Category and Marketing Applications

There is a model of how to persuade people to buy, called the Elaboration Likelihood Model [108]. The elaboration likelihood model is divided into rational and long-term persuasion, called the central route, and emotional and short-term persuasion, called the peripheral route. Advertisements that present products for long-term persuasion are designed to be viewed entirely or repeatedly. As people pass by the digital signage in a short time, they may not see the entire advertisement from its beginning to end. Therefore, there is a fear that the viewer may be presented with only a partial explanation.

In contrast, music, which has the ability to move people's emotions [109], is more suited for short-term persuasion. Additionally, research found that playing music draws the flow of passersby to the sound source, slows down walking [110], and improves individual pleasure and dominance emotions [111]. From the results of the experiments in this study, it was found that advertisements in the music category made a better recallability on passersby within a short period of time.

However, musical advertising requires sound output. In this study, a possible methodology to promote purchase behavior is to first implement music-dominated advertisements on digital signage and then use the on-demand pinpoint audio system for sound output, as illustrated in Figure 3.19.

This can be seen as a new methodology for promoting purchasing behavior. Accordingly, it will be necessary to verify the concrete effects of the pilot program in public spaces.

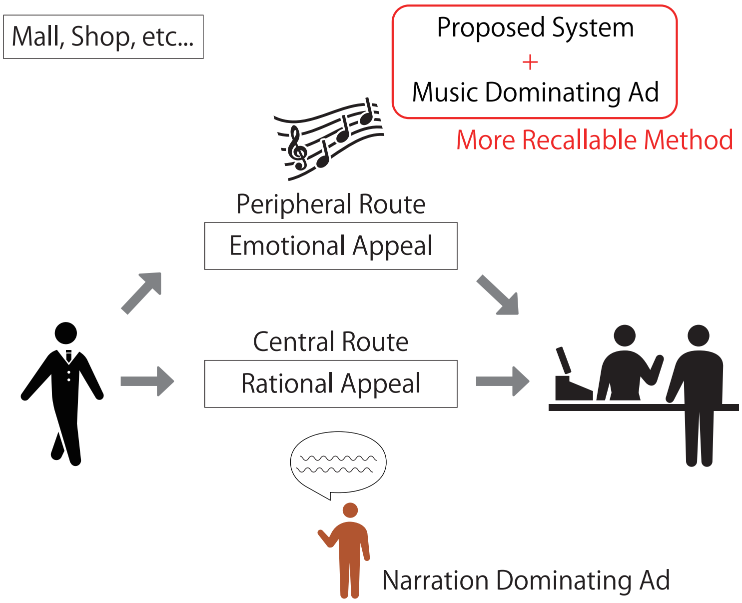

Figure 3.19: Elaboration Likelihood Model with the findings

Limitation and Future work

The limitations of the on-demand pinpoint audio system are described here. Although the Kinect v2, which was used as a camera, can support up to six people, the number of people who can be supported will increase if multiple Kinect or regular cameras are used. Parametric loudspeakers were used to provide one-on-one support for people; hence, as the number of units is increased, the number of people supported will grow. However, in a crowd, there is a risk that more than one person could be in the audible range of a parametric loudspeaker. Therefore, to cover a large number of people, it is necessary to equip the system with techniques to narrow the audible range, as seen in related studies [55].

The limitations of the experiment come mainly from the difference between laboratory and public space. To prevent bias by the participants in this experiment, a task unrelated to looking at digital signage was set. This ensured that the signage was present during the task in the laboratory experiment as well as on the way to the store in the shopping mall. However, the behavior of passersby in public spaces is more complicated. For example, it will be interesting to see whether some of the effects in human interaction with digital signage can be seen; such as the honeypot effect [112], where, if one person pays attention to the signage, others are drawn to pay attention to it, or the landing effect [113], where a person walks past the display and notices the signage gimmick and turns back. Moreover, this experiment was conducted indoors. Noise is also a problem outdoors, and the on-demand pinpoint audio system can be applied outdoors as well. Acoustic comfort is expected to see the same effects indoors and outdoors, but it could change because of outdoor noises such as road traffic or hydraulic breaker noises. New insights may be gained by installing the on-demand pinpoint audio system in actual stores and measuring its effectiveness.

This experiment was evaluated from the viewpoint of recallability, which is important in terms of purchase behavior [66]. However, I were not able to fully evaluate the effects of the ads. Therefore, the evaluation of the degree to which it leads to advertising effects needs to be examined in depth in future research.

Another limitation could be the bias introduced by the question about advertisement comfort where the exact wording used is negative instead of positive - I were asking how uncomfortable the advertisement is, not how comfortable, which is Q. 3 and 4 in Table 3.3. In addition, the measurements of on-demand pinpoint audio system was performed with an omnidirectional microphone. Binaural measurements will be expected in the future for more strict measurements.

Despite the fact that conventional loudspeakers were expected to be noisy, the results did not indicate any significant difference in the acoustic comfort when compared with other systems. This phenomenon could be attributed to the fact that advertisements have become a part of the daily lives of people. In the experiment, environmental sounds of a shopping mall were played to recreate a realistic setting. I believed that using conventional loudspeakers to emit two different types of advertising sounds would increase the level of discomfort because the sounds would be mixed up and difficult to distinguish, creating a considerable amount of noise. Against this backdrop, there was a participant comment that the advertising sounds were similar to the noise of the crowd. In other words, it was revealed that they were not perceived as annoying or even unpleasant.

This opinion is likely to vary depending on the culture. The participants in this experiment were all Japanese, and Japanese urban areas are routinely overflowing with crowds of people and advertisements. This appears to have led to the results of acoustic comfort. Therefore, I expect to see more diverse results in different cities and cultural environments, where advertisements are displayed less frequently. Research on the environment of advertising sound in public spaces is still inadequate, and this should be further investigated in future work.

3.4 Pinpoint Auditory Glimpse

3.4.1 Section Objective

In the previous section, I proposed an on-demand pinpoint audio system audible only to those who noticed it. Participant experiments indicated that the system could maintain spatial comfort for those who did not notice it, while effectively delivering advertising content to those who chose to engage with it. However, this system faces a challenge: it may not sufficiently signal the presence of advertising signage. Particularly in spaces crowded with visual information, like small signs inside a store, there is a significant risk that the signage might go unnoticed.

This problem has been noted in previous studies. It has been found that this signage fades from view in public spaces [71, 73]. Müller et al. call this phenomenon "Display Blindness" [73]. In particular, it has been reported that when an advertisement is publicly displayed, it is often irrelevant to passersby, which can lead to a passive attitude towards the advertisement. Observational studies conducted in shopping malls have also shown that many digital signs are only seen from a distance and briefly [97]. In other words, the problem is that even when advertisements are presented on signage, they are rarely seen by passersby and do not spread awareness. Therefore, the issue of display blindness must be solved to increase advertising effectiveness.

In previous Human-Computer Interaction (HCI) research, various solutions to display blindness have been attempted that rely on visual perception. These solutions have been proposed by studies that devise visual content meant to attract people's attention [71, 73, 90, 91, 92, 93] and by studies that physically place signs to lead people's visual focus to the signage [94, 95]. All of these have been reported to be effective in increasing the attention of passersby to signage. However, in urban areas, signage is often buried by many other visual things that compete for attention [97]. In addition, there are cases in which the display is small due to installation location constraints or in which the display is behind passersby. Thus, even if the display attempts to attract attention visually, passersby may not notice the presence of the display and not see it at all. Therefore, it is necessary to develop a method to attract attention to the display as an alternative to visual methods.

One potential solution to display blindness is to utilize sound in addition to vision. Even if a display is out of sight, its presence can be made known through sound. However, to my knowledge, there are few papers in HCI research that have considered sound-based solutions. In a limited study, Nordfält reported that sound can make passersby aware of the presence of a display even when it is behind them [74]. However, constantly emitting sound from a display can be considered noisy and may not be effective in attracting people's attention, especially if the sound is from an advertisement. I also proposed in the previous section a signage system that delivers sound only to those who are paying attention, using a camera to determine the level of attention of passersby and delivering pinpoint sound only to those people judged as attentive. This can reduce noise, but if passersby do not look at the signage, the sound will not be played and the goal of informing the user of the display's presence will not be achieved. Therefore, further research is needed to develop sound-based solutions to display blindness that minimize noise and effectively inform users of the display's presence.

This section proposes a system called "pinpoint auditory glimpse," which uses sound to notify passersby of the presence of signage. The system presents pinpoint sounds to a passerby walking near the signage as a kind of "pump-priming," as shown in the Figure 3.20. In other words, there is no sound until a passerby approaches the signage, but the sound is initiated the moment they approach the signage, with the intention of making them aware of the presence of signage. Additionally, the pinpoint auditory glimpse is designed to be presented only for a short period of time to prevent noise pollution. However, it is necessary to verify whether the brief presentation of pinpoint auditory glimpses serves to increase recognition of the advertisement. Furthermore, the suddenness of the sound presentation and the addition of the pinpoint auditory glimpse in a noisy environment, such as a shopping mall, may increase discomfort among passersby. Therefore, this section conducts an experiment in a pseudo-store from two perspectives: advertisement recognition and the comfort level of passersby.

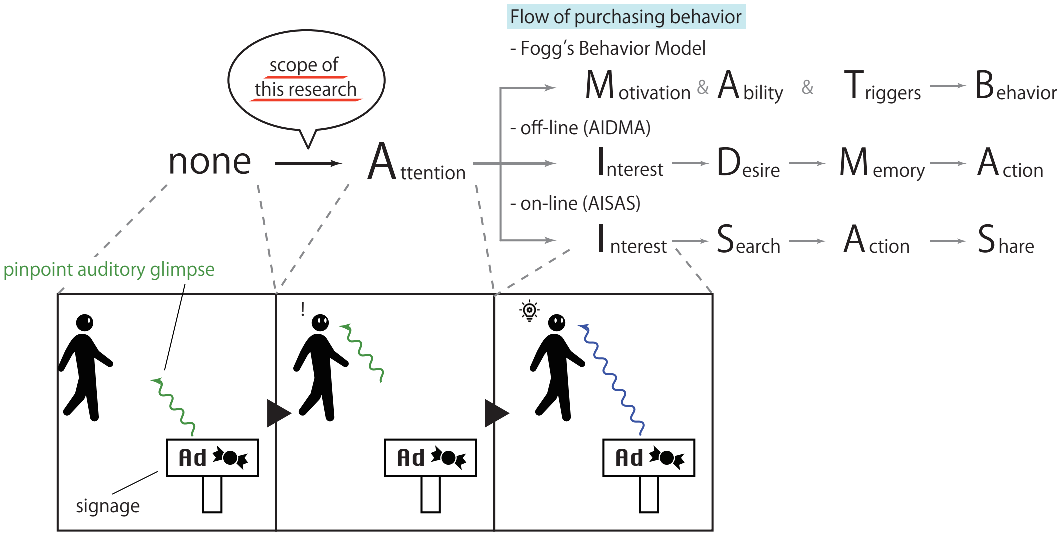

Figure 3.20: Scope of the Pinpoint Auditory Glimpse

3.4.2 Design

System Overview

In this section, I propose a system that presents a pinpoint auditory glimpse to passersby approaching the signage for a short time and changes the sound level depending on the degree of their attention in response. The system schematic is shown in Figure 3.21 a).

The parametric loudspeakers are mounted on two motors for emitting sound to the left and right ears, as illustrated in Figure 3.21 c), the two motors move in the azimuth and elevation directions. A camera is used to track a passerby and keep track of the spatial coordinates of their heads and ears. The displacement angle of the motor is calculated based on the spatial coordinates of the ears, and when a passerby is within the range of the camera, the parametric loudspeaker continues to move so that the parametric loudspeaker is facing the ear. This ensures that only specific people can hear the sound and that it does not reach nontarget users.

The pinpoint auditory glimpse is presented when the user approaches the signage. The content of a pinpoint auditory glimpse is the corresponded audio as the video content playing on the signage. Then, if the user remains interested in the signage, the sound level changes as described above. If the user is not interested in the signage, the sound level is reduced so that it does not become noise. This is important because, while pinpoint sound serves to attract users to the signage, it should not be noisy, so it is only played for short time intervals, such as one second.

A pinpoint auditory glimpse is included in the on-demand pinpoint audio system in section 3.3, i.e., after the pinpoint auditory glimpse is emitted, the sound level is changed according to the passerby's attention level. The sound level is changed according to the

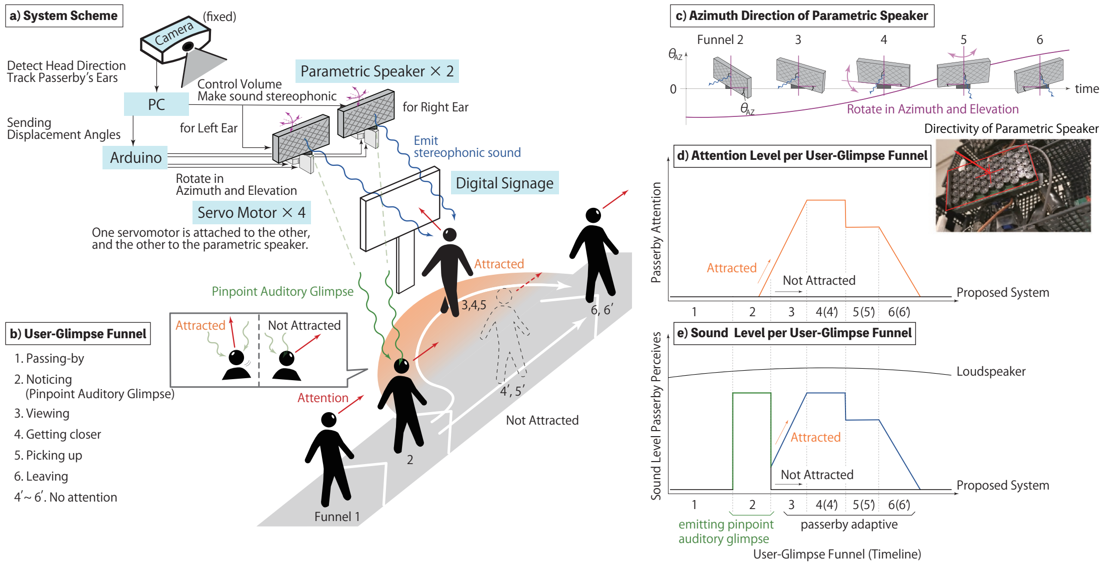

Figure 3.21: a) Architecture of the proposed system, b) User-Glimpse Funnel – Behavior flow of passerby when using digital signage with a pinpoint auditory glimpse, c) Orientation of the parametric loudspeaker in the proposed system, d) Attention level per User-Glimpse Funnel, e) Sound level fluctuation of the proposed system and conventional loudspeaker

direction of the face as observed by the camera, which calculates whether the face is facing the direction of the digital signage and, if not, how far away it is. If the face is looking at the signage, then the sound level is set to a maximum, and if it is not, then the sound level decreases as the face moves away, as shown in Figure 3.21 (d) and (e). If the passerby is looking away and ignores visual content, then the system will automatically fade the sound so that the passerby can ignore audio content. Therefore, the only noise that the user hears occurs during the pinpoint auditory glimpse, which is active for only a very short period of time. This scheme works with only one loudspeaker. For the sake of providing a sense of localization, however, the sound is presented stereophonically so as to specify its source [24].

By controlling the exposure time and sound level as described above, the objective is to make it sound as if it has suddenly played while not creating noise in the surrounding area, thereby drawing attention to it. Additionally, the sound is not a special sound like an alert, but is the same as the content of the video, so that the passerby does not feel alerted.

The servo motor used in this section was an SG-90, which moves in 0.1-degree angle increments. The camera was an Azure Kinect, and the sensing during system runtime was approximately 30 fps. The parametric loudspeaker was a Tri-State K-02617 with as many as 50 transducers (AT40-10PB3) connected to it. Three Arduino UNO R3 microcomputer boards were used. The control application was developed in C++ and executed on Windows 10. The system latency, from the first detection of a person to the presentation of a pinpoint auditory glimpse, was up to 41 ms (average 40.07 ±0.68 ms). This is below perceivable time, and the system operates in real time with almost no subjective delay.

Pinpoint Auditory Glimpse and User-Glimpse Funnel

I applied the audience funnel [102] to the new behavior flow, which is specialized for using the proposed system. The created funnel is called a User-Glimpse Funnel, which is shown in Figure 3.21 b). The level of attention and corresponding fluctuation in sound level compared to those of the conventional loudspeaker signage are depicted in Figure 3.21 d) and e). The user-glimpse funnel is divided into six stages, which are explained in conjunction with the operation of this system as follows:

1 . Passing-by: The passerby is walking toward the signage. The signage does not emit any sound.

2 . Noticing (Pinpoint Auditory Glimpse): The passerby approaches the signage. The signage emits a pinpoint auditory glimpse for a certain period of time.

3 . Viewing: The passerby reacts to the pinpoint auditory glimpse and pays attention to the signage. The signage stops the pinpoint auditory glimpse after a certain time.

4 . Getting closer: The passerby pays attention to the signage, and in some cases, approaches it. The signage increases the sound level depending on the direction of passerby attention.

5 . Picking up: The passerby is attracted to the product displayed by the signage and puts it in a shopping cart. The signage output is turned down slightly.

6 . Leaving: The passerby walks away. The signage decreases in sound level and eventually reaches a minimum.

4' . No attention: The passerby is not attracted to the signage and does not pay attention to it. The signage stops the pinpoint auditory glimpse after a certain time.

5', 6'. No attention: The passerby walks away without paying attention to the signage. The signage keeps the sound level to a minimum.

User-Glimpse Funnel has two features. The first one is the transition between stages 2 and 3 of the Funnel. The pinpoint auditory glimpse is presented at the time of approaching the signage to draw the attention of the passerby to the signage.

The second feature, Funnel 4 and beyond, adjusts the sound level based on the passerby's continued attention to the signage. If the passerby continues to pay attention, the sound level increases. However, if they pass by without paying attention, the sound level decreases. This allows for control of the amount of information received and the creation of an appropriate sound environment.

Traditional silent signages do not provide an opportunity for passersby to pay attention, resulting in display blindness. Additionally, systems that use conventional loudspeakers do not take into account the passerby's position or attention, leading to an uncomfortable sound environment. The proposed system aims to address these issues.

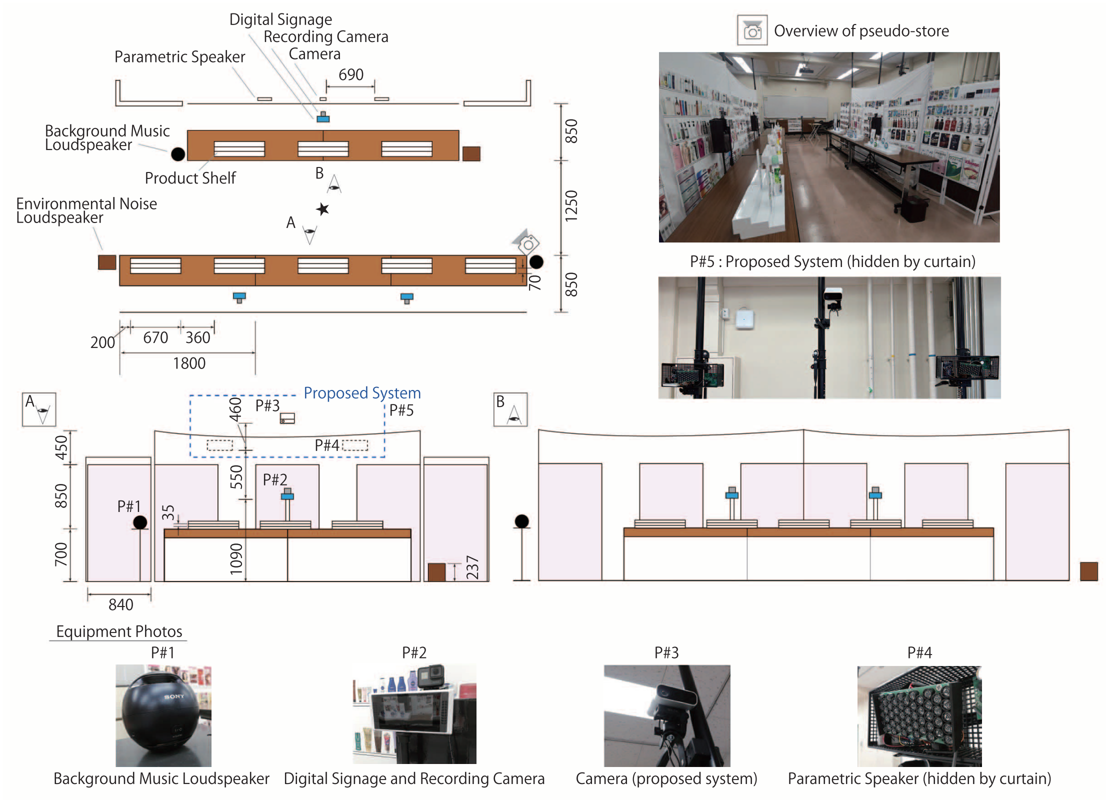

Figure 3.22: Setup of the user experimental environment

In terms of sound, the sound level required by listeners differs between conventional loudspeakers and parametric loudspeakers. Chapter 2 reported that the speech recognition threshold is lower when parametric loudspeakers are used than when conventional loudspeakers are used. Therefore, comparing the maximum sound level, this system is operated at a slightly lower level than the conventional loudspeaker, as shown in Figure 3.21 e).

3.4.3 Experiment

Experimental Environment

To see fluctuations in behavior and spatial comfortness with a pinpoint auditory display, a pseudo-store was setup in a laboratory room as the experimental environment. The setup is shown in Figure 3.22. The use of a pseudo-store is an experimental method often used in the field of service engineering, and its effectiveness has been measured [114].

In the pseudo-store, posters of the display shelves were placed on the wall. The sound level was set after conducting pilot study, as shown in the table in Table 3.4. The average sound levels were measured by a sound level monitor (RION NL-31) set up where the star mark appears in Figure 3.22. The calibration of the system sound level was performed using

Table 3.4: Sound level setup in the experiment, unit:[dBA]

Background Music	49.72
Proposed System (Funnel 4)	53.44
Environmental Noise	53.46
Proposed System (Funnel 5)	51.84
Conventional Loudspeaker System	67.26

Table 3.5: Equipments List used in the experiment except for proposed system

Digital Signage	GREEN HOUSE GH-EP4RW
Recording Camera	GoPro HERO5
Background Music Loudspeaker	SONY SRS-X1
Product Shelf	HPC-TOMOYA-30854XXX
Environmental Noise Loudspeaker	TEAC S-300NEO

the signage located in front of the star mark. To replicate a real store, the environmental sounds of a shopping mall and some instrumental background music were played. The equipment used is shown in the Table 3.5.

There were eight kinds of products in total, arranged in two rows across the aisle. Among the products in the store, daily necessities were displayed on the shelves without signage ads, while alcohol-based disinfection products were displayed on the shelves with signage ads. This composition was adopted because daily necessities and alcohol-based disinfection products, which are essential items during pandemics, do not fluctuate in demand based on preferences. Each product was displayed on a three-tiered platform, with the actual product in the middle of the front row and paper product pieces arranged around it. There were three digital signage units installed behind the products. Each signage unit was equipped with no conventional loudspeaker, a conventional loudspeaker, or the proposed sound system. The pinpoint auditory glimpse was emitted when the user walked past a neighboring product shelf. A camera was installed on top of the signage to record and measure the viewing time. These experimental environments are shown in Figure 3.22.

Advertisement Selection

The advertisements for the products on the shelves were played on the signage in a loop, with each advertisement lasting 15 seconds, the typical length of a TV commercial. The advertisements focused on the benefits of using the particular product, which is suitable for unplanned purchases [115].

However, advertising preferences can affect product preferences and evaluations [116]. To avoid this, I conducted a preliminary questionnaire to ensure user preferences for the advertisements used in the pseudo-store experiment. The questions in the preliminary questionnaire are as follows:

Q. (a) How recallable is the advertising for this product? (1: Not recallable at all. 7: Very recallable.)

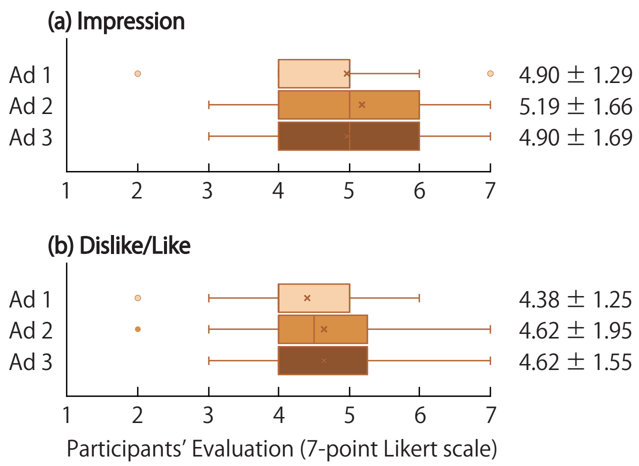

Figure 3.23: Results of the preliminary advertisement questionnaire

Q. (b) Please rate your preference for this ad. (1: Very dislikeable, 7: Very likeable)

The number of participants in the preliminary advertisement questionnaire was 21 (13 males and 8 females aged 22 to 34 with an average of 26.19) and consisted of people who did not participate in the main experiment (i.e., the pseudo-store experiment). They were recruited by snowball sampling. Not all participants were university staff or students. The results showed no significant differences among advertisement preferences and are illustrated in a Figure 3.23. To ensure randomness in the pseudo-store experiment, three signage systems and three advertisements were used at random.

System Performance Test

With sound level measurement, I confirmed if there is a fluctuation in acoustic comfort when using the proposed system compared to using a conventional loudspeaker. The sound level was measured in the same location as the user experiment. The author wore binaural microphones in both ears and recorded while walking in front of the signage. The conventional loudspeaker and proposed system were used to compare the sound level changes between when the participant was paying attention to the signage and when they were not. The ambient noise was set to only include ground noise, and the sound source was white noise. Subsequently, bandpass filters were applied to the range of 1 k to 3 k Hz and a 5-point moving average was used to remove other noise.

The measurement environment for sound level was the same as the user experimental environment, as shown in the top of Figure 3.24. The results of the measurement are shown in the bottom of Figure 3.24. It can be observed that the sound level changes in each stage, even though the time on the horizontal axis is not perfectly aligned in both systems. The proposed system played the pinpoint auditory glimpse in Funnel 2, and the sound level fluctuated depending on the degree of attention to the signage. On the other hand, the

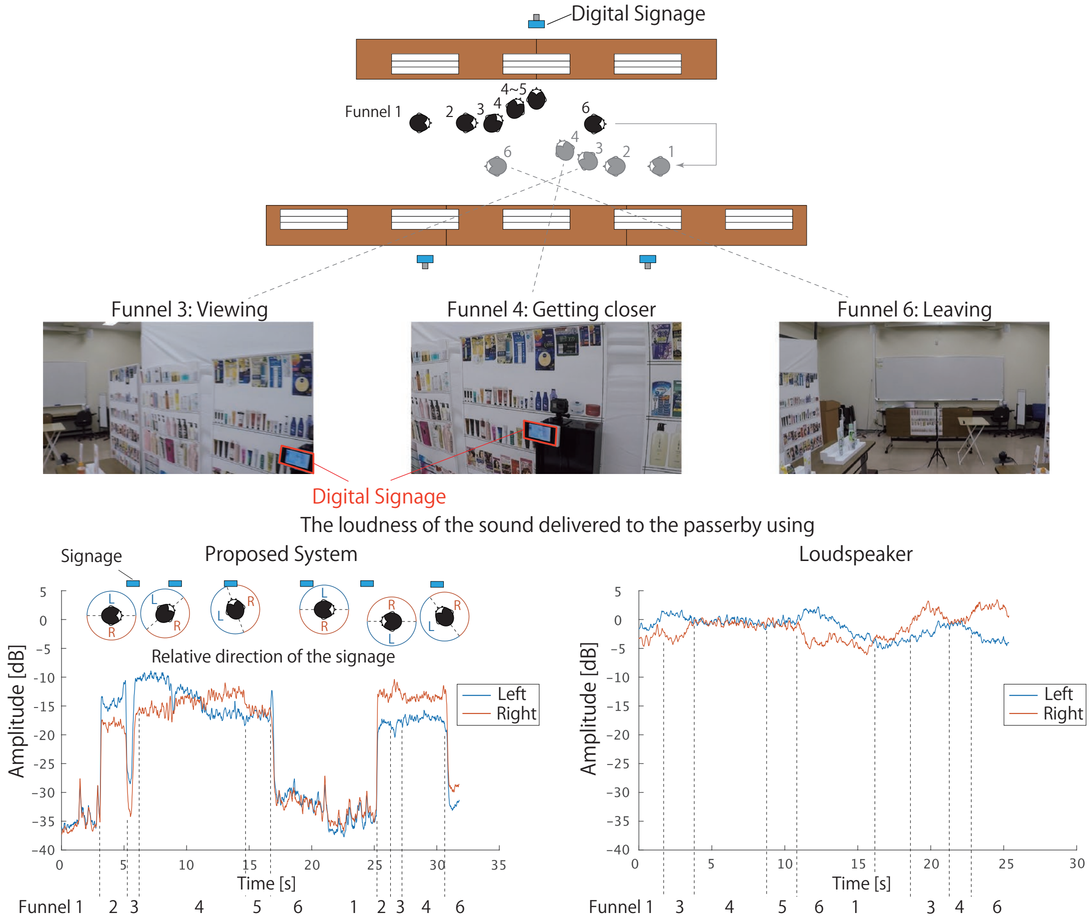

Figure 3.24: First person view of the User-Glimpse Funnel, the relative direction of the signage position as seen by a passerby, and the results of the system performance test of proposed and the system with a conventional loudspeaker

conventional loudspeaker system had a constant sound level, regardless of the degree of attention to the signage. As a result, it was confirmed that the proposed system created a more comfortable acoustic environment as the sound level was lowered when the user was not paying attention to the signage.

Methods

Participants were asked to shop at a pseudo-store. There were 33 participants recruited by snowball sampling in the pseudo-store experiment (27 males and 6 females aged 16 to 36 with an average age of 23.36), who were students and working adults. Not all of them were university staff or students. They had no visual or auditory problems by self-report. Besides, they didn't practice listening to the stereophonic sound before the study. The scenario for the experiment was as follows: Participants were given a shopping list and a basket of products in advance. Two products on the list were randomly chosen from five options, which are no signage ads, and participants were required to purchase them. They were free to choose from the remaining six products (three with and three without signage ads) as unplanned purchases. The items outside of the list were chosen based on the participants' preferences and did not affect the results of the experiment. It is important to note that only one of each item could be placed in the basket and participants were allowed to purchase products without consideration of the price. This setup aimed to simulate unplanned purchases and increase awareness of the products through watching the signage.

Participants can roam the aisles of the pseudo-store during the experiment. They completed their shopping, including the items on the provided shopping list, and then filled out questionnaires afterwards. There was no set time limit for shopping and participants were able to signal when they were finished. The questionnaire content was based on the characteristics of the advertisements [117], section 3.3, and the research questions being addressed.

The questionnaires included the following two questions:

(a) How recallable is the advertising for this product? (1: Not recallable at all. 7: Very recallable.)

(b) During shopping, how loud was the sound of this advertisement? (1: Not annoying at all. 7: Very annoying.)

These questions were answered for each of the three advertisements. Each questionnaire was based on a seven-point Likert scale, with 4 considered to be the midpoint.

Additionally, viewing time was used as the objective evaluation of recognition. Cameras (GoPro) were placed directly above each sign and recorded constantly during the experiment. The recorded video was reviewed by the author frame by frame to calculate the participant's signage viewing time. The recordings were made at 60 fps and converted to seconds. By analyzing the viewing times thus obtained, I measured the level of attention to each sign. By analyzing the participants' self-reports and viewing time, I evaluated the transition into the first phase of the AIDMA purchasing behavior, attention, as indicated in Figure 3.20.

3.4.4 Results

Analysis Overview

Through a participant experiment, viewing time was analyzed on the basis of three factors: the system type, the response to the pinpoint auditory glimpse and the purchase stage. Subsequently, the results regarding recallability and perceived loudness were analyzed on the basis of participant questionnaires. Analyzing the level of attention, viewing time and advertisement recallability from questionnaire responses reveals the attracting effect of the pinpoint auditory glimpse, which is a factor in the purchasing cycle (see Figure 3.20). Furthermore, since previous studies have already shown that the degree of attention to

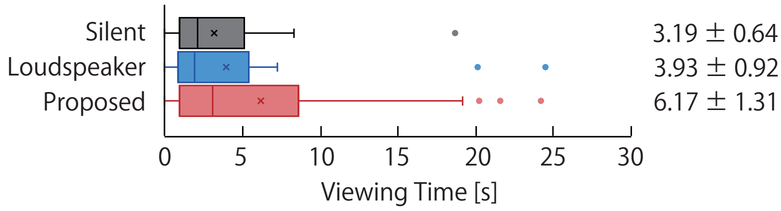

Figure 3.25: Results of the average participant viewing time for each system

signage changes depending on the purchasing stage [115], the purchasing stage was included in the analysis. The following is a summary of each section.

- Effect on Funnels 2–5: **Viewing Time for Each System**; The analysis of viewing time using the system type as a factor.
- Effect on Funnel 2: **Viewing Time with a Response to the Pinpoint Auditory Glimpse**: The analysis of viewing time with the response to the pinpoint auditory glimpse as a factor.
- Effect on Funnels 3–5: **Viewing Time Based on the Stage of the Purchasing Process**: The analysis of viewing time using the system type and the purchasing process as factors.
- **Participant Questionnaire**: An evaluation of the recallability and perceived loudness as obtained from participant questionnaires.

Viewing Time for Each System

Figure 3.25 shows the results of the average participant viewing time for each system. This provides an evaluation of the change in recognition level of the signage based on the systems used. The silent signage, signage with a conventional loudspeaker, and that using the proposed systems had average participant viewing times of 3.186 s, 3.929 s, and 6.179 s, respectively. A one-way independent-measure ANOVA (system type: three levels, silent signage, signage with a conventional loudspeaker and the proposed system) was performed on these data. There was no significant difference ($F(2{,}98)=2.451$, $p=0.092$).

The lack of a significant difference in the results is thought to be due to the large within-group variance. The silent signage, signage with a conventional loudspeaker, and the proposed system have variances of 13.390, 28.062, and 56.489, respectively. This suggests that the variance is particularly high when using the proposed method. Therefore, a comparison was made between those who responded to the pinpoint auditory glimpse and those who did not respond to it among the proposed system group.

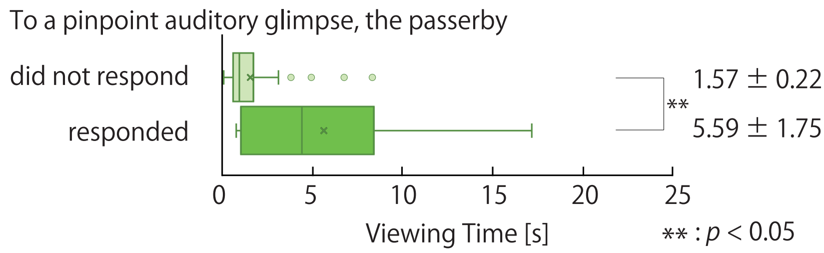

Figure 3.26: The results of viewing time with and without response to the pinpoint auditory glimpse when the number of items in the basket was 1 or 2

Viewing Time with a Response to the Pinpoint Auditory Glimpse

The proposed signage system presented a pinpoint auditory glimpse, and the experiment monitored the participants' responses to it. Participants were considered to have responded to the pinpoint auditory glimpse if they paid attention to the signage when the pinpoint auditory glimpse was played, but not if they were already paying attention to the signage before the pinpoint auditory glimpse was played. When the number of required products in the basket was 0, no one responded to the pinpoint auditory glimpse. However, when the number of required products in the basket was 1 or 2, five participants responded to the pinpoint auditory glimpse each time. These responses were made by different participants, and a total of 10 out of the 33 participants responded. To evaluate the effect of the pinpoint auditory glimpse on viewing time, the amount of time that the participants paid attention to the signage during and after the sound was played was measured.

The results of viewing times with and without a response to the pinpoint auditory glimpse are shown in Figure 3.26. The average viewing time for participants who responded to the pinpoint auditory glimpse was 5.588 seconds, while the average viewing time for those who did not respond was 1.571 seconds. I conducted an independent-means t-test, in which the factor was the mean viewing time in the presence or absence of a response to the pinpoint auditory glimpse. This showed a significant difference at the 5% level ($t(9)=2.277$, $p=0.049$). This suggests that the use of a pinpoint auditory glimpse in digital signage can increase the amount of time that passersby spend viewing the signage.

Viewing Time Based on the Stage of the Purchasing Process

The viewing time for the signage was calculated based on the number of items on the shopping list that had already been added to the basket. This method allows for the evaluation of not only the proposed system, but also the change in signage recognition due to the stage of purchase. In a real store, having 0 items in the basket indicates that the customer has just entered the store and is in the process of making a planned purchase. When the number of items in the basket is 2, the person has completed their planned purchases and may be in the process of making an unplanned purchase or about to finish shopping. The number of items in the basket can also be 1, and in this experiment, it was

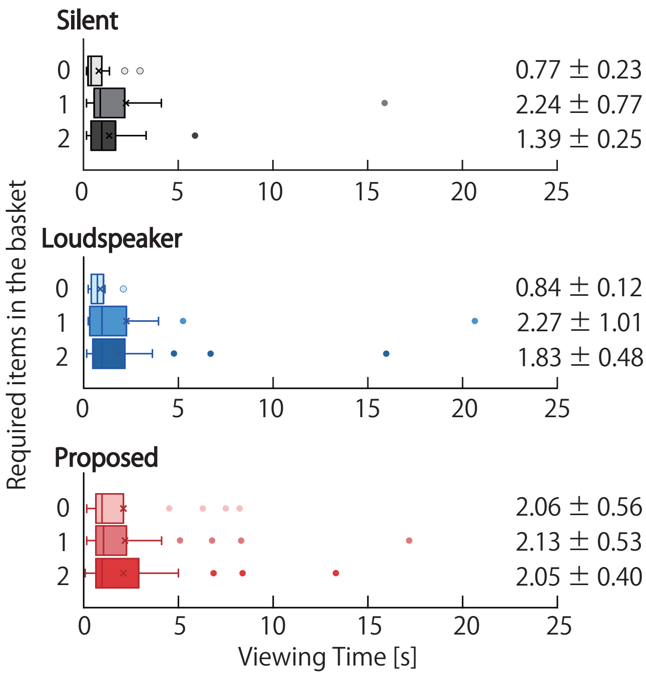

Figure 3.27: Results of the viewing time by the number of required products in the basket and the type of system

observed that some participants intermingled their unplanned purchases.

Figure 3.27 displays the results for the viewing time by the number of required products in the basket and the type of system. A two-way independent-measures ANOVA (system type: three levels, silent signage, signage with a conventional loudspeaker and the proposed system; purchasing stage: 0, 1 and 2) was performed on these data. The results showed that there were no significant differences in the system type factor ($F(2{,}237)=1.010$, $p=0.366$), in the type of stage factor ($F(2{,}237)=1.857$, $p=0.158$) or in both types ($F(4{,}237)=0.559$, $p=0.692$).

Moreover, when the average viewing time was independently calculated for the cases where the number of required items placed in the basket was [0,1] and 2 independent of the system, the results were 1.787 s for [0,1] and 1.779 s for 2 items. According to the independent-means t-test, when the purchasing stage was used as a factor, the two-tailed test showed no significant difference ($t(244)=0.023$, $p=0.981$).

The average times per viewing when the number of required items in the basket was 0 and [1,2] were 1.244 s and 1.949 s, respectively. According to the independent-means t-test when the purchasing stage was used as a factor, the two-tailed test showed that a significant difference was confirmed at the 5% level ($t(173)=2.312$, $p=0.022$). These results are shown in Figure 3.28.

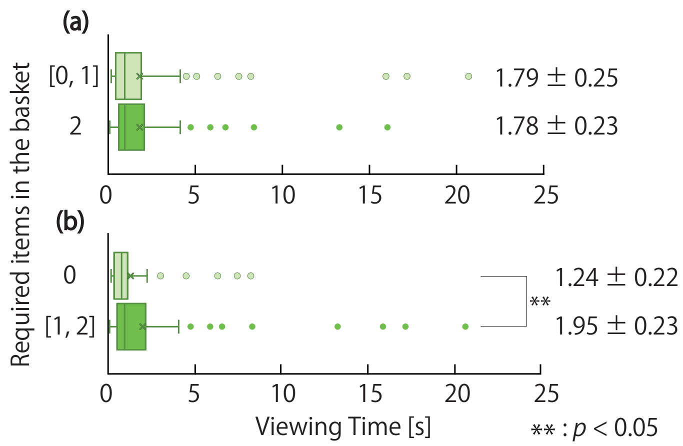

Figure 3.28: (a) Independent results for the average viewing time where the number of required items placed in the basket was [0,1] and 2 independent of the system,(b) Independent results for the average viewing time where the number of required items placed in the basket was 0 and [1,2] independent of the system

Participant Questionnaire

After finishing the purchasing experience, the participants answered the questionnaire. All answers were self-reported by the participants, and the results are shown in Figure 3.29.

The recallability ratings of the advertisement were 2.15, 3.33, and 2.84 in condition silent signage, signage with a conventional loudspeaker, and utilizing the proposed system, respectively. A one-way independent-measure ANOVA of whether the advertisement was strongly recallable was conducted, and the results showed no significant difference ($F(2,96)=3.091$, $p=0.073$).

The recallability levels of those who noticed the pinpoint auditory glimpse and those who did not in the proposed system were 2.80 ($n=10$) and 2.87 ($n=23$), respectively. A two-tailed t-test was performed and resulted in no significant difference ($t(23)=0.091$, $p=0.93$).

Subsequently, the loudness of the conventional loudspeaker and proposed systems were rated as 4.70 and 3.84, respectively. A two-tailed independent-means t-test using loudness as a factor showed no significant difference ($t(23)=1.839$, $p=0.073$)

3.4.5 Passerby Observations

Here, observations related to the proxemics of purchasing behavior other than that of viewing time and questionnaires are described. The following description is arranged on the basis of the user-glimpse funnel proposed in Section 3.4.2

1. **Passing-by, Noticing (Funnels 1 and 2)**: Many participants who received the shopping list tried to put the two products specified on the list in their baskets rather than looking carefully at the other products. Many participants' behavior consisted

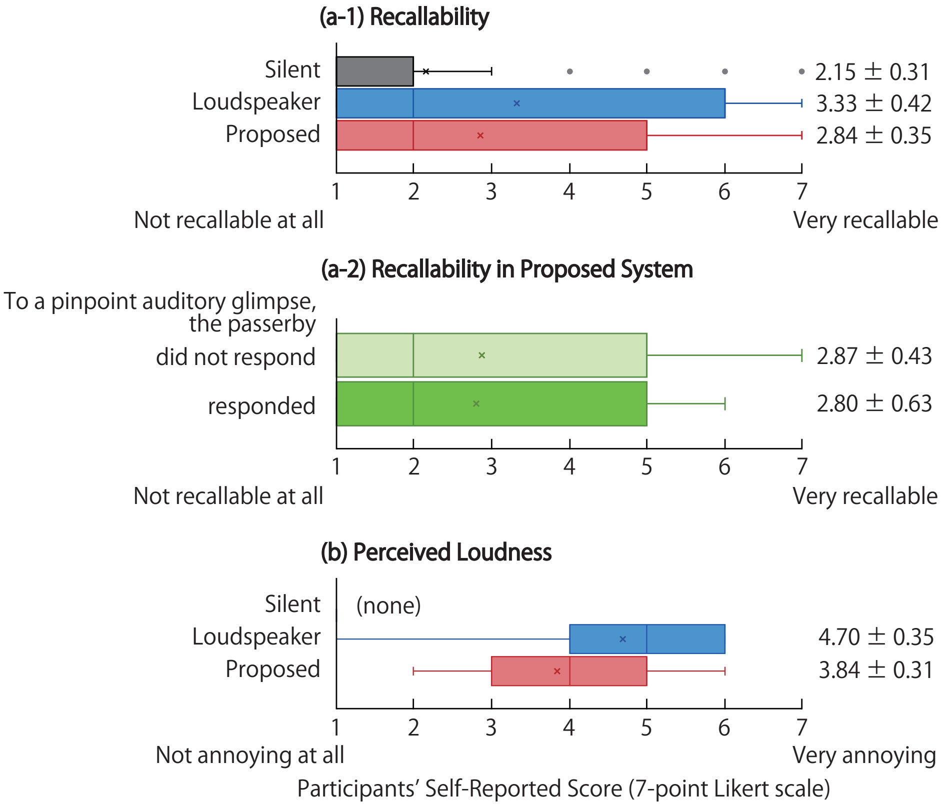

Figure 3.29: Results of the self-report questionnaire completed by participants.

of walking around comparing the list with the products displayed on the shelves. During this time, the signage was treated as reference information, and received little attention from the participants (refer to Figure 3.27, 3.28). When the basket contained one or two items from the list, more participants were observed viewing the signage for unplanned purchases. In some cases, as they approached to view a shelf across the aisle, the pinpoint auditory glimpse caused them to suddenly look back at the signage.

2. **Viewing, Getting Closer (Funnels 3 and 4)**: Some participants who were exposed to the pinpoint auditory glimpse paid attention to the signage, glancing around to determine the source of the sudden sound. Their walking pace slowed down, and some individuals briefly looked at the signage before continuing on their path, while others approached and stopped in front of the signage to observe it. Participants who did not respond to the pinpoint auditory glimpse were also observed to pass by the signage without paying attention to it, skipping the Funnel 3 of reacting to the pinpoint auditory glimpse.

3. **Picking up, Leaving (Funnels 5 and 6)**: Many participants who were observed

to pass close by the signage or spend a significant amount of time viewing the advertisements, considered whether to include the advertised product in their shopping basket while they were viewing the advertisements and products. Some participants also examined the back label of the product, which lists its active ingredients, and compared it with those of other products to make a purchase decision. After making a decision, they walked away from the shelf with their back turned. When they passed by the signage again, they almost completely avoided looking at it.

4. **No Attention (Funnels 4' – 6')**: There were some cases in which participants passed by without paying attention to the signage even when a the pinpoint audio glimpse was presented. These included those who focused on buying the products on the shopping lists and those who had already decided whether or not to buy the suggested products. This also included some participants who did not make unplanned purchases but bought only the listed products. In this case, their walking speeds did not decrease. Among them, one participant made purchases without paying attention to any digital signage.

As previously mentioned, the behaviors of participants were varied. However, most behaviors were consistent with the User-Glimpse Funnel, and I was able to observe interesting connections between the pinpoint auditory glimpse and purchasing behavior. These will be discussed in later sections.

3.4.6 Discussion and Limitations

Summary of Experiments

Attracting Effect: Considering the results of viewing time and Figure 3.26, it can be said that the time spent viewing the signage of the proposed system in response to the pinpoint auditory glimpse was significantly longer than that in no response to the pinpoint auditory glimpse. This implies that once a person responds to the sound, they are more likely to continue paying attention to the source of the sound, i.e., the signage. This is also observed in passerby observations. In several cases, passersby who were looking in another direction (Funnel 2) turned toward the signage (Funnel 3) when the pinpoint sound was presented. After this, some passersby continued to observe the signage and approached the product (Funnel 4).

I then calculated the percentage of passersby who responded to the pinpoint auditory glimpse. Pinpoint auditory glimpses were triggered in all 33 trials for a cumulative total of 71 instances. In ten of those instances, different participants responded to each pinpoint auditory glimpse. Thus, the percentage was 30.3 % when calculated based on the number of individuals and 14.1 % when calculated based on the number of instances. Here, the effect is discussed based on a comparison of the ratio with that of a previous study. The existing research that prompts passersby to interact with public displays aims to measure participant interaction with and viewing of advertisements [118]. Among all passersby, the ratios of persons who interacted with and watched traditional advertisements were approximately 10% and 0.1%, respectively. However, the proposed system could trigger 14.1% of participants to watch advertisements without interacting with displays. Since this experiment was not conducted in a public space and the advertisements were different from those used in existing research, the two results cannot be directly compared. Nonetheless, my results suggest that using a pinpoint auditory glimpse can be an effective way to encourage people to watch display advertisements without the need for interaction.

From these results, I obtain the findings as follows: even a very short sound presentation increases the ratio of passersby who look at the advertisement. Additionally, those who respond to the pinpoint auditory glimpse continue to look at it for a longer period than those who do not. These factors lead to improved advertisement recognition.

Acoustic Comfort: In terms of the acoustic comfort, the questionnaire results, as shown in Figure 3.29 (b), indicate that there was no significant difference in loudness between the proposed system and the system using a conventional loudspeaker. One participant commented "*The sudden sound was slightly distracting, but then the sound was lowered, so I didn't pay attention to it.*" This means that a sudden presentation of sound used a pinpoint auditory glimpse that momentarily disturbed the participant, but it was quickly muted in accordance with their choice not to see it, so it did not interfere with their activities. This comment suggests that the proposed system was able to provide acoustic comfort.

However, since there was no significant difference between the proposed system and the system using a conventional loudspeaker, and since only one proposed system was used in the experiment, it is necessary to further consider the possibility of the noise problems when multiple proposed systems are being used. Therefore, the finding is that when a single pinpoint auditory glimpse is used, it does not negatively impact the acoustic comfort of those nearby.

Sound Level

Traditionally, emitting a loud sound was an effective way to draw users' attention to digital signage [74]. However, the results of using the proposed system indicated that users can turn towards signage that emits sound for a brief period without necessarily requiring a loud sound. I assume this to be because the person who was presented with the pinpoint auditory glimpse became interested in the sound, which drew attention to the signage. In general, sound is heard gradually as one gets closer to the sound source, and the phenomenon of sudden hearing does not exist in nature. Therefore, the phenomenon of suddenly hearing the sound of an advertisement is considered to be mysterious and curious. This phenomenon is similar to a previous study in which a mysterious object was placed in front of the signage [96]. This previous study showed that passersby were interested in the mysterious object, which made them more likely to notice the presence of the signage nearby. The results of this section can be described as a sound version of this previous study. This can be explained by the fact that people are more interested in the pinpoint auditory glimpse and are more likely to pay attention to the source of the sound, the signage.

Passersby might be offended by the output of a sudden loud sound. Sound output is often restricted in public spaces [82], and there may be concerns that the output sound may be drowned out in places where other sounds are emitted loudly. However, prior section 2 has shown that parametric loudspeakers require less sound level than conventional loudspeakers, even in noisy environments. Additionally, the narrow directive sound of parametric loudspeakers means that they are barely audible to non-target listeners nearby. This allows for the use of the pinpoint auditory glimpse in a wide range of environments, from places where noise should be minimized to louder environments such as shopping malls. Overall, the proposed system is a useful solution to the limitations of sound level that hinder the improvement of public awareness, and it increases the flexibility of sound design in public spaces.

Stage of Purchasing Process – Design Implication for In-Store Signage

Further analysis in the viewing time and Figure 3.27 revealed that there was no significant relationship between the viewing time and the number of products using the proposed system. Additionally, it was found that the average viewing time for the signage was significantly longer when the number of required items in the basket was 1 or 2, regardless of the system being used.

Previous research, such as a study conducted in a real store [115], has shown that advertising placed near the cash register is more likely to lead to purchases. This phenomenon is linked to customers' behavior towards the cash register indicate the end of their purchasing behavior. The results of this current experiment reinforce these findings by showing that the use of the pinpoint auditory glimpse increases recognition even during the middle of the purchasing process.

Future studies should take into account various parameters within the store. As customers go through different products during the purchasing process, it is important to investigate the optimal timing for presenting pinpoint auditory glimpse to draw attention to specific products. Additionally, the duration and sound level of these cues should be adjusted based on factors such as the size of the product shelves and the layout of the store. For example, in smaller stores with limited space, auditory cues may need to be presented earlier or for a longer duration in order to be effective. By examining the relationship between store layout and these parameters, a system can be designed that improves product recognition while maintaining acoustic comfort.

Application for the Pinpoint Auditory Glimpse – Design Implication for Public Display

This section explores the use of pinpoint auditory glimpses in advertising signage. The use of this technique does not interfere visual information, haptic design, or context. Additionally, audio can be applied to a variety of displays beyond just advertising signage, such as interactive signage with playful elements. This solution has possibility for addessing the "first click problem" [92, 113, 119] where passersby may not understand that a display is interactive and thus do not engage with it. The use of a pinpoint auditory glimpse, such as a sudden callout (e.g., "Come talk with me!"), can increase the likelihood that they will turn to look.

The section can be also applying for strategies for maintaining audience engagement after initial interaction, such as predicting when people will leave and using new forms of interaction to retain their interest [103], or using auditory cues to surprise and deter them from leaving. The section suggests that future research should investigate the degree of attraction that different types of sound used in a pinpoint auditory glimpse may have. Overall, the section suggests that the use of a pinpoint auditory glimpse has the potential to change the business model of displays without disrupting conventional systems.

Limitations and Future work

Sound Reflection: When using a parametric loudspeaker, the reflection can be considered as the parametric loudspeaker emitting super directive sound. For this point, whereas the sound of a conventional loudspeaker may reflect from many different directions, the super directive sound of parametric loudspeakers limits reflections. In addition, by producing sound from a high point in a slightly downward direction, I adjusted the sound to not reach that person's ears after reflection on a wall. If the reflected sound were audible, the sound would be heard from a different direction than the intended direction, and the sound pressure difference between the left and right sides would be incorrect in the sound measurement. In contrast, when I look at the Proposed Systems Funnels 3 and 4 in Figure 3.24, the sound is delivered in such a way that the sound pressure in the intended direction becomes higher. For example, in Funnels 2–4 on the return route in Figure 3.24, since the signage is on the right side, the sound pressure presentation should be louder to the right ear. The measurement results show that it is indeed louder to the right ear. Therefore, I think that the participants in the experiment did not misunderstand the direction of the sound source due to reflections and that the reception of reflected sound was a rare case. From the above, I consider that the priming effect of the pinpoint auditory glimpse was maintained. In fact, there were no comments from participants that they heard sounds coming from different directions or that they felt uncomfortable. However, when parametric loudspeakers are used in a smaller space or when there are several people present at the same time and the number of parametric loudspeakers increases, the effect of reflected sound may be unavoidable. Countermeasures against these include the use of phased arrays and the use of sound-absorbing materials in reflective walls.

Differences between environments: In this experiment, I used a pseudo-store to evaluate the proposed system on the basis of viewing time. Although the pseudo-store was constructed to be as close to an actual store as possible, it has not fully reproduced an actual store. To further validate the effectiveness of the system, future researchers should conduct experiments in real-world public settings to gather a larger sample size and measure the attention of participants towards the signage over a specific period of time.

When using a pinpoint auditory glimpse, it is important to consider the transportation context. In outdoor settings, it is important to ensure that the auditory cues are only presented to pedestrians to avoid distracting bicyclists or drivers while they are operating vehicles. This can be achieved by limiting the use of the system to environments where only pedestrians are present, such as indoors, or by using cameras to detect the speed of movement and only playing the cues for pedestrians. When implementing the system in public places, it is essential to prioritize safety and take measures to minimize any potential risks to the public.

A possible approach for providing simultaneous coverage for multiple people is to increase the number of parametric loudspeakers or cameras in proportion to the number of people that the camera can detect. However, using many parametric loudspeakers at the same time may result in an excessive amount of directional sound in a given location, which could be as disruptive as using a conventional loudspeaker. Therefore, future research should investigate whether the "Honey Pot Effect" [112] can be observed. The honey pot effect is a phenomenon where one person or a small group's interaction with signage can stimulate others in the vicinity to also interact with the signage. This section did not involve direct interaction with the system, so it is unlikely that the honey pot effect would be observed in this experiment. However, when multiple passerby are present, it is possible that the behavior of passerby may change, and similar phenomena such as the honey pot effect could be observed. For example, a scenario could be created where one person in a group notices a pinpoint auditory glimpse and then tells the others, "Did you just hear something," which causes the whole group to pay attention to the signage. Although this is a hypothetical situation, such multi-person-specific phenomena are conceivable. Future research should investigate if other similar phenomena exist by conducting further experiments.

3.5 Answers to Research Questions

In this chapter, I proposed a digital signage system with a two-step system of On-demand Pinpoint Audio System and Pinpoint Auditory Glimpse, and verified it regarding two research questions through experiments.

For the RQ.2-1: "Does varying the sound level of an ad's auditory display improve recognition?," according to Figure 3.12, the on-demand pinpoint audio system by itself was inferior to the conventional loudspeaker in terms of recallability. However, when pinpoint auditory glimpse was integrated to it, no significant difference in viewing time was found between the conventional loudspeaker and the pinpoint auditory glimpse, as shown in Figure 3.25. Besides, it was also found that the viewing time increased when there was a response to the pinpoint auditory glimpse, as shown in Figure 3.26. In other words, the answer to the first research question is "Ad recognition increases when the sound level is not only changed according to attention level, but also when the sound level is changed as a 'trigger' for passersby to look at the signage."

For the RQ.2-2: "Does varying the sound level of an ad's auditory display make the acoustic environment less comfortable?," as shown in Figure 3.18 and Figure 3.29, the results of the proposed system were similar to those of the conventional loudspeaker. In other words, the acoustic comfort was found to be maintained. However, as stated in Section 3.4.6, the system was only validated with a limited number of units, so it is necessary to validate the case where more signage changes its sound level at the same time. Therefore, the answer to the second research question is "if only some auditory displays vary in sound level, it does not negatively impact the acoustic environment."

3.6 Summary

In this chapter, an examination using digital signage was conducted as an example to clarify whether an auditory display according to an individual context can resolve the trade-off between utilization effectiveness and spatial comfort.

In the section 3.3, I proposed an on-demand pinpoint audio system, and clarified the influence of advertisement recallability and acoustic comfort. The categories of advertisements that affected the advertisement recallability and the subjective acoustic comfort when digital signages emitted sound were also investigated. Through experiments with 30 participants, the results showed that advertisements using the on-demand pinpoint audio system were more recallable than silent advertisements. Although acoustic comfort was maintained, the on-demand pinpoint audio system reduced the noise level compared to a conventional loudspeaker. In addition, when analyzed by advertising category, music-dominating advertisements presented by the on-demand pinpoint audio system were more recallable than silent and other advertisement categories.

In the section 3.4, I proposed the use of a pinpoint auditory glimpse as a sound-based solution to address display blindness. The findings indicate that a pinpoint auditory glimpse can effectively draw people's attention to signage, and that once a person responds to the sound, they are more likely to continue paying attention to the signage. Additionally, the results suggest that the presence of a pinpoint auditory glimpse does not negatively impact participants' purchasing behavior. Furthermore, this system offers benefits not only for advertisers, but also for the general public by reducing the overall sound level and minimizing any potential harm to passersby. These results demonstrate the potential of the pinpoint auditory glimpse as a novel, sound-based method for solving display blindness, and highlight the potential for it to lead to changes in advertising display business models and increased flexibility in sound design in public spaces.

Chapter 4

Auditory Display for Guiding Pedestrians

4.1 Chapter Objective

In the previous chapter, it was found that an auditory display responsive to an individual's level of interest can achieve both improved utilization effectiveness and maintenance of acoustic comfort. In this chapter, I examine the auditory display that responds not to the individual but to the state of the space, using human flow control as the subject matter.

Managing pedestrians flow effectively is crucial for successful urban designs and large-scale events. Tourist congestion, for example, can affect sightseeing areas, making them difficult to access. There is now a need to alleviate these congestions [120, 121]. During massive events such as parades and concerts, unguided pedestrian flow can cause traffic congestion, reducing comfort and potentially leading to accidents [122, 123, 124]. In addition, to prevent the spread of infectious disease during a pandemic such as COVID-19, it is necessary to redirect passersby to other streets, which could reduce congestion rates, maintain physical distance and minimize contact [125]. Thus, managing the flow of pedestrian movement is essential.

Historically, visual methods have been employed to manage the flow of people. Stationary or temporary elements such as traffic signs (e.g., one-way streets) [126, 127] and digital signage [128, 129] were used along with pedestrian guidance through police officers [130]. While these methods have been useful in organizing pedestrian flow, they need to be visually prominent, which often requires bright colors. These colors can be disruptive to the aesthetics of the landscape. Furthermore, pedestrian guidance methods demand significant resources, as police or security personnel must be permanently stationed on the streets. Consequently, I need to consider alternatives to visual methods for managing pedestrian flow.

In this section, I propose an auditory display using non-linguistic sounds to guide pedestrians as shown in Figure 4.1. The corresponded research questions are following:

RQ.3-1: What kind of sound design effectively guides pedestrian flow?

RQ.3-2: What is the impact of sound-guided intervention on streetscape perception?

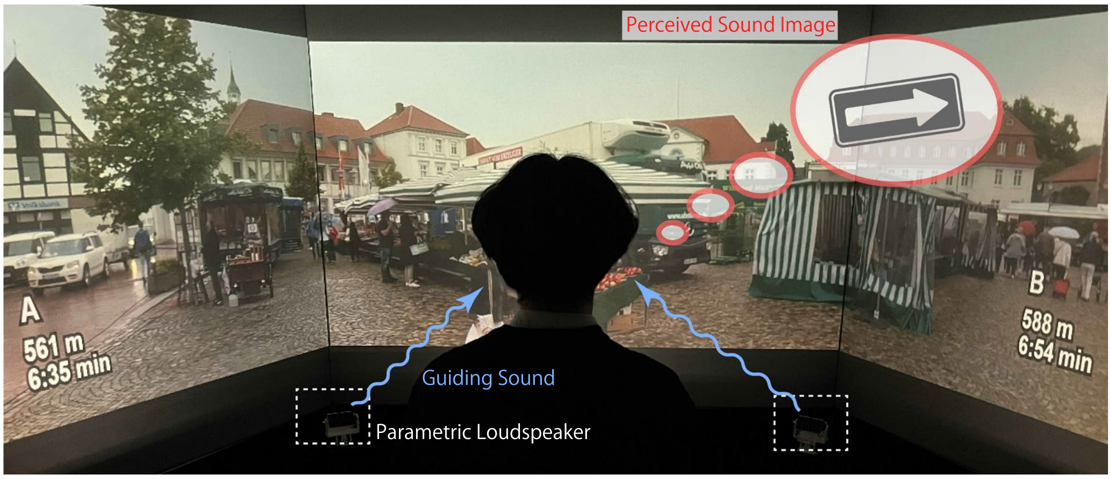

Figure 4.1: Overview of an auditory display for guiding pedestrians

These research questions are being investigated through participant experimentation. The experiment used a laboratory setup with parametric loudspeakers for the auditory display. This resulted in almost no sound reflections or noise, and only few head movement by the participants. Due to these limitations, the results are not entirely generalizable, but this chapter offers spatial audio of pedestrian guidance as an exploratory study.

4.2 Related work

4.2.1 Pedestrian Flow Guidance

Existing studies on pedestrian flow guidance have primarily focused on congestion alleviation and emergency evacuation, with visual information being the principal means of direction. For instance, traffic signs are ubiquitously used to regulate and direct traffic, ensuring safety via signs indicating temporary stops, speed limits, and so on [126, 127].

As a congestion mitigation strategy, some transit platforms visually display the expected congestion of the next arriving train [128]. The intention is to encourage passengers to adjust their boarding position to evenly distribute congestion on the train. Similarly, airports have proposed signs to depict the level of congestion in departure lobbies, nudging individuals towards less crowded areas [129]. These visual displays have been well-received and successfully guide the public.

Mobile navigation that adapts routes to control crowd flow has also been proposed [131]. This research indicates that this system helps organize pedestrian flow and direct individuals to specific destinations.

In emergency evacuations, safe and effective guidance is crucial. Past strategies have suggested the use of digital signage, studying factors such as size, distance, height, and graphic design to optimize visibility even in smoke-filled conditions [132, 133].

Auditory guidance has also been proposed for scenarios where visibility is hindered, such as fires in tunnels. Many participants reported basing their evacuation decisions on audio information [134]. Evacuation beacons using chime sounds and announcements have been effective in directing individuals towards optimal routes [43]. Further research has compared non-verbal auditory cues based on subjective evaluations and evacuation success rates [135].

While auditory methods have been utilized in emergencies, visual cues typically guide pedestrian flow. Few papers use audio guidance in non-emergency situations. However, it can have a wide range of applications, such as guiding visitors to tourist attractions and events, providing directions in zoos and museums, and easing congestion.

Research has also been conducted to guide walking through tactile presentations in non-emergency situations. Devices in which haptic transducers surrounding the user's waist vibrate in response to the destination's direction and distance have shown their design's effectiveness [136]. Stimulus presentation by galvanic vestibular stimulation (GVS) provided gait guidance and posture support [137], which provides direct navigation. Another method has been proposed to change the trajectory of gait by applying weak electrical signals with EMS (Electrical Muscle Stimulation) on the legs and neck to naturally rotate the muscles [138, 139]. Furthermore, a method has been suggested for attaching vibrotactile actuators to the entire head to assist with navigation and search tasks when using head-mounted displays [140]. These haptic navigation methods are undoubtedly effective. However, the haptic presentation uses invasive devices. Especially in the real world, not everyone wears such a device. This research uses an auditory display as a technique that does not need wearables or additional user equipment. Its objective is to investigate desirable sound design for guiding pedestrian flow in non-emergency situations.

4.2.2 Spatial Audio

Spatial audio is generally heard using headphones or other wearable devices. This is to present different sounds to the left and right ears, and because it is important to localize the sound accurately [141]. Various studies have been conducted on the perceptual localization of spatial sound; experiments on acoustic localization, including visual cues in AR environments, have reported high perceptual accuracy in the horizontal direction [142]. It has also been reported that stereoscopic sound image generation when using conventional loudspeakers is not sparse and can be perceived with good accuracy with proper spacing [143]. Concerning the advantages of stereophonic sound presentation, improved visual search performance results were observed, for instance, when tasks were performed with non-informative stereophonic sounds [144]. However, previous research has demonstrated that haptic sensation is more effective than auditory sensation in conveying warnings to car drivers [145]. Therefore, it is desirable to choose the most appropriate modality for the situation.

Currently, non-wearable systems that can present sound directly to the left and right ears by using directional audio have also been constructed. One such system [146] allows people to hear stereophonic sound even while walking. By sensing a person's position and calculating in real time, it is possible to perceive where the sound is located, no matter where the person is facing. Although this system has not yet been developed for public spaces, it can be used for visual applications such as mid-air AR. It has also been applied to stereophonic acoustic presentation while walking, such as hazardous sound presentation [147] as an accident prevention. Thus, the technology of spatial acoustic interfaces is developing, allowing greater freedom to the listener.

One advantage of spatial audio is that it is possible to perceive which direction the sound is coming from. One navigation system that makes use of this is a technology for assisting the visually impaired. Research on stereophonic presentation of sound icons [148] can perceive where something is by sound even if you cannot read a map. Kim et al. have made the system more ubiquitous, proposing a demodulated acoustic information presentation technique that can only be heard by the visually impaired [149]. This allows them to locate anywhere without having to touch the Braille. In addition, a system that can recognize the location and state of surrounding objects [150] enhances the safety of mobility for the visually impaired. As an object search, Komoda et al. have proposed a system that navigates by AR-like voice superimposition of object location [35]. Spatial audio presentation lightens the burden on the assistive person and facilitates object search for the visually impaired. As shown above, sound presentation with spatial acoustic direction is often used as an assistive technology for people with disabilities.

By using this direction-perceivable spatial audio, this chapter builds an interface that intuitively communicates which way one wants to be guided. In this way, I propose a new application of spatial audio.

4.3 Design of Auditory Display under Normal Condition

4.3.1 Section Objective

Initially, this section will conduct a study on the sound design, effectiveness, and the streetscape perception when guiding pedestrians in a situation where no specific event is occurring. To induce pedestrians to proceed to a particular path, it is necessary to be able to perceive which path it is and to motivate them to go that way. In other words, it is important to provide spatial sound and motivational design as the guiding sound. Thus, it is crucial to investigate the appropriate spatial sound design for inducing pedestrian flow. Further, I must consider whether this auditory intervention might adversely affect streetscape perception. Therefore, the purpose of this section is to identify what constitutes a good design for guiding sound that encourage pedestrians to follow a specific path, and to measure the effectiveness of it.

4.3.2 Design and Materials

Principles

This section outlines the design of an auditory display featuring guiding sound. The first step in presenting sounds to pedestrians involves determining the type of earcon to be used. To encourage pedestrians to follow the guidance, it is important to provide them with motivation.

Among these components, ability is inherently simple, as the only task required is path selection. Prompts can be provided at decision points such as intersections. Therefore, the sound must primarily convey motivation. The earcon must be designed such that pedestrians are motivated to follow the intent of the guiding sound and choose a path accordingly.

Moreover, pedestrians often have a brief window from noticing a sign to deciding their subsequent path. This means the design must swiftly convey the message to pedestrians. Additionally, the guiding sound must be distinctive, allowing people to understand which road the information pertains to and its location. Conventional loudspeakers diffuse sound widely, making it difficult for pedestrians to discern the specific road the guidance sound refers to. Consequently, the guiding sound should also incorporate angle information.

Sound Category

To act as an effective guiding sound providing motivation, the sound must incentivize pedestrians to listen. One such example is music. Street musicians sometimes perform impromptu sessions, infusing the environment with their music. Such street music is not only positively received by passersby but also tends to attract crowds [4, 77]. By playing music from a specific street, I aim to incentivize and encourage pedestrians to take that route. In this study, I used accordion music [151], a common choice for street music.

The second sound employed to create the illusion of a reward is applause [152]. By emanating the sound of clapping and whistles from a particular path, pedestrians might feel praised or welcomed, prompting them to choose that route. The sound of applause in this context includes multiple people clapping, supplemented by encouraging whistles [151].

Thirdly, I used nature sounds to promote a sense of comfort. According to various studies on soundscapes, nature sounds can induce relaxation [78, 153, 154] and enhance comfort by masking disruptive noises, such as traffic sounds [17, 155, 156, 157]. By directing this nature sound from a specific path, I aim to persuade pedestrians to follow that path in pursuit of this soothing ambiance. The nature sounds employed here comprise a blend of birdsong [158] and the babbling of a river [159].

All of these sound types are presented as brief excerpts, ensuring their content and intended meaning can be comprehended even if heard for only a short period. Furthermore, recognizing that pedestrians might need more time to decide their path after hearing the sound, each sound clip is designed to last for over 30 seconds. Sound clips were edited using adobe audition.

Visual Material

The previous study [160] examined the effect of visual signage in an immersive video environment (IVE). The IVE showed a video simulation of the environment (panoramic footage of a decision point) with overlays of the proposed pedestrian one-way signage. In addition, they added textual information on the panoramic video to indicate travel time and distance to a destination. Since participants can feel a high level of immersion in an IVE, it is suitable for simulating actual roads. In my study, I use the same visual environment except for the overlay of the visual one-way signs.

The videos consist of panoramic views of multiple decision points along a routes. There

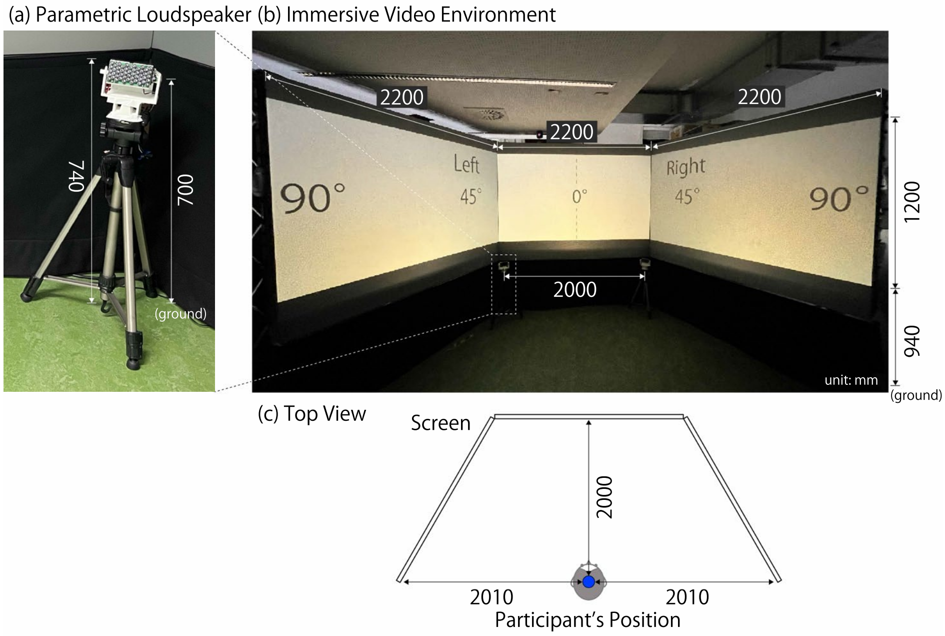

Figure 4.2: Experimental setup: this experiment is conducted using (a) two parametric loudspeakers and (b) an immersive video environment (IVE). The image on the screen in this figure is the reference for the angle perception experiment.

are three routes in total. Regardless of which route is chosen, each decision point of that single route leads to the same destination. To balance the realism of walking while avoiding the video-induced motion sickness [161], the footage between branch points is implemented as a transition of still images. These show the path from the current decision point to the next. Each route is composed of five to eight decision points.

The video footage utilized in this chapter was identical to that used in a previous study [160]. The method employed to create the video involved recording a 3-minute panoramic video at 37 intersections in the city of Quakenbrück, Lower Saxony, Germany. In order to capture different perspectives upon entering the same intersection from different directions, 180° videos were created from the recorded 360° videos. In addition, panoramic shots were taken every 20 meters between intersections to create transitions. All footage was recorded using a 360° camera (Kandao QooCam 8K Enterprise).

As shown in Figure 4.2 of the experimental environment diagram, the IVE is set up to cover the participant's field of view. The IVE operates as a single 5760*1080 display by connecting three 1920*1080 resolution displays, creating a highly immersive environment even without a headset. Additionally, at each decision point, the available paths lead to the left, center, or right direction, with no paths that are immediately adjacent to each other.

System Configuration

As outlined in the design principles, the guiding sound must allow pedestrians to determine which path the sound refers to. For this purpose, I utilized stereophonic sound to localize the sound image in the direction of the path. This enables pedestrians to identify the origin and angle of the guiding sound.

Stereophonic sound presentation comes in two formats: wearable and non-wearable. While headphones are a common wearable option [162], not all pedestrians wear headphones, and unauthorized Bluetooth connection is inappropriate. In addition, pedestrians have no clue as to where the sound source is, making it difficult to connect headphones. Consequently, I adopted a non-wearable approach to ensure universal applicability.

In stereophonic presentation, the direction of the sound image is perceived based on subtle discrepancies in arrival time and sound pressure between the left and right ears [141]. It's crucial that the sound source intended for the left ear reaches the left ear and the sound source intended for the right ear reaches the right ear.

Using conventional loudspeakers to present the sound might result in the sound intended for the left ear being heard by the right ear. This crosstalk issue is known to diminish presentation accuracy. To enhance accuracy, this chapter utilizes a parametric loudspeaker [22, 23]. This loudspeaker exhibits a narrow directivity feature, meaning it can be clearly heard only when aimed directly at the listener, thus eliminating the crosstalk problem. I deployed two such loudspeakers, one for each ear. The loudspeaker's model number is Tri-State K-02617, which has 50 transducers (AT40-10PB3).

The software needs to convert the sound source to stereophonic sound. However, stereophonic sound perception varies among individuals and depends on personal attributes like head and ear shapes. This attribute is known as the head-related transfer function (HRTF) [163, 164]. But capturing and utilizing the HRTF of every passerby is not feasible. Instead, a HRTF measured on a dummy head was used for generalization. The HRTF is part of an online publicly available dataset by [165]. The HRTF with subject number 046 was used for the study. For the specific measurement method, the reader is referred to the document [165] by the creators of the dataset.

The stereoacoustic procedure involves first normalizing the sound sources selected in the sound category to the same sound level (ITU-R BS.1770-3 standard) [166]. Then, among the selected HRTFs, I use the one where the vertical direction is 0°, indicating the sound source position is at the same height as the listener. HRTFs with 5° increments from 0° to 355° in the horizontal direction are prepared, and a convolution operation is performed on the sound source. The resulting left and right 2-channel signals are transmitted to the parametric loudspeakers corresponding to the left and right ears, respectively. This configuration is depicted in Figure 4.3.

Experimental Environment

An experiment involving participants was conducted to assess the effects of different types of stereophonic guiding sound. The chapter took place in an environment simulating an actual streetscape. The experimental protocol (equipment, procedure, and footage) used in

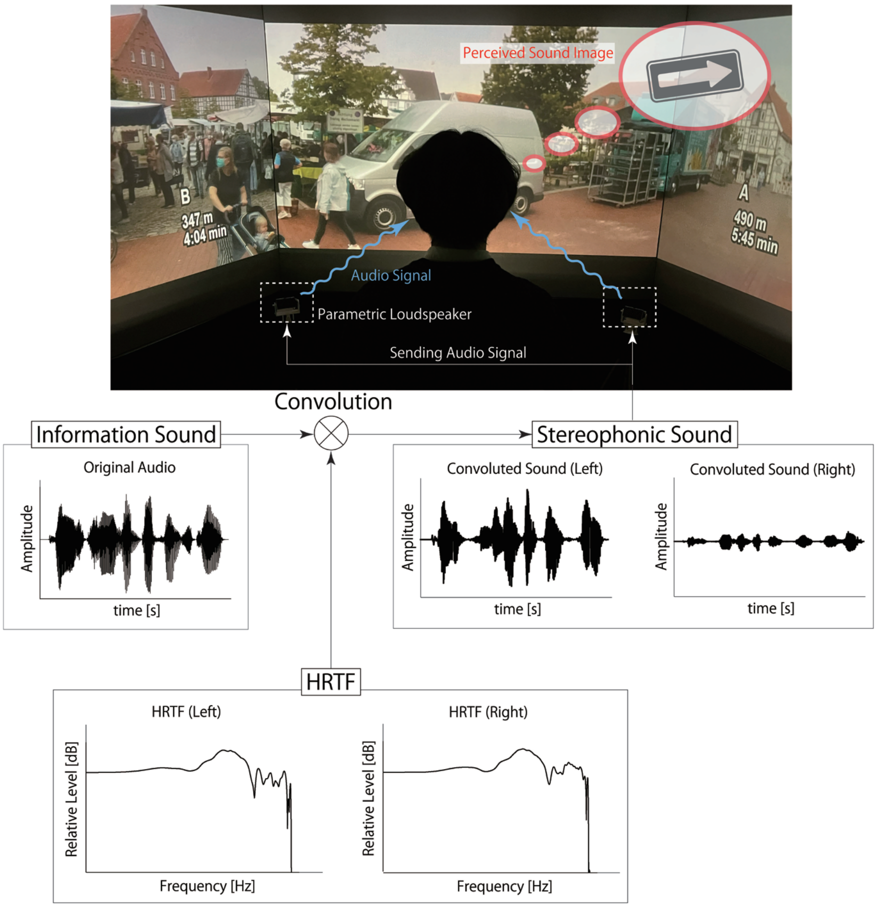

Figure 4.3: The system configuration of guiding sound

a previous study [160] was employed for this simulation. The immersive video environment (IVE) is equipped with a 180 degree screen setup to show panoramic video footage. It offers a VR-like real-world experience, making it ideal for my experiment, as shown in Figure 4.1.

Three routes with 5-8 junctions from the starting point to the destination were prepared, with each junction offering 2-3 paths. Participants walked each route during the routing scenario without a time constraint. As a scenario, the participants were given the following instructions: "You are out and about. You are on foot in a city on your way to a certain shop to buy something. You do not have an appointment so there is no time pressure to make a decision or to reach the destination." Static image transitions were used to simulate the movement between junctions, as continuous movement could potentially induce VR sickness. A video and audio recorded at the junction point are played at each junction. The audio is played from a conventional loudspeaker apart from the auditory display. Besides, each branching path's remaining walking distance and time to the destination were displayed as an overlay. As an action at a junction point, the participants were instructed that "There

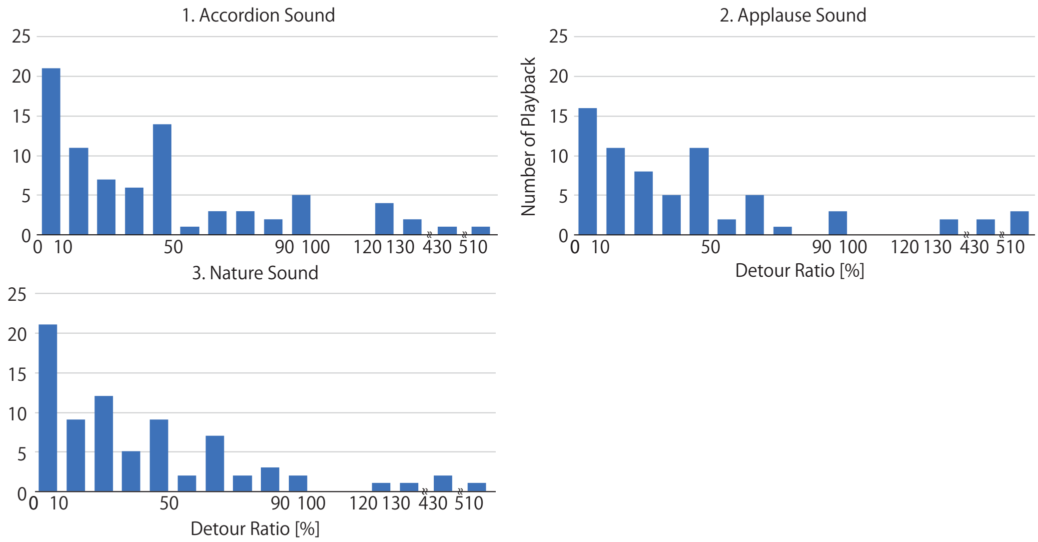

Figure 4.4: Random placement of the three guiding sounds and the distribution of the detour ratio

will be several branches before you reach your destination. There will be two choices, A and B, and sometimes three choices, including C. The choices are labeled with the distance and time to proceed to the next point. Please use these information, landscape, and soundscape as a reference to select a branch and tell me the letter of branch."

A previous study has shown that travel time and distance are decisive factors for pedestrians' route decisions [167]. Assuming that participants typically choose the shortest or quickest route, as per rational judgment, I played guiding sounds from routes that take a longer distance or time to reach the destination. This sound intervention tests if participants are willing to deviate from their rational judgment and take a detour.

The experimenter manually controled video and sound playback using the Oz system to make the experiment feel as automated as possible. After reaching a junction, the guidance sound was played. The branching paths were labeled A, B(, C), and participants indicated which path they wanted to take. After their selection, the experimenter turned off the guiding sound and played the video transition to the next junction. This process repeated until the destination location was reached.

Participants practiced on a short course before the experiment to familiarize themselves with the environment. The course featured a location that did not appear in the main experiment. It was designed to introduce participants to the experimental protocols, such as going to the destination, choosing a path at a junction, and the video transitioning to the next junction. I ended the exercise by confirming with the participants that they were ready for the main experiment. Each participant randomly followed one of the three routes in the experiment, with a random choice of sound source played at each junction. Due to the randomness, three participants each did not receive one type of sound. The omitted sounds were different for each of these three participants. The task took approximately 2 minutes for the practice and 5 minutes for the main, respectively.

In order to better compare the detour impact of the sound interventions across intersections with different remaining distances to the destination, I defined the "detour ratio," which is the ratio between detour taken and shortest available path:

$$\frac{D(p,i)}{S(i)} = R(p,i) \tag{4.1}$$

with *D(p,i)* being the length of the detour taken by participant p at intersection i; *SPL(i)* being the shortest available path length at intersection i; and *DR(p,i)* being the detour ratio of the detour taken by participant p at intersection i.

The distribution of presented sound types among intersections with varying detour ratios is shown in Figure 4.4. The y-axis shows the number of times each sound was played when the detour ratios at a given junction are aggregated into increments of size 0.1. This figure indicates that the randomly placed sound types are approximately evenly distributed.

Following the walking experience, participants completed a questionnaire about their awareness of the sound sources during the experiment and their streetscape assessment on the computer. First, they were asked if they noticed each sound - "Have you noticed this sound? (Yes/ Maybe/ No/ Other)" and "Did you want to head toward the road in the direction this sound was playing? (Yes/ Maybe/ No/ Other)," followed by "Why did you feel as you answered above?" To evaluate the streetscape, including sound, participants then watched a video from the starting point with the guiding sound playing in the direction of 0° and rated the overall comfort on a 7-point Likert scale - "How comfortable was the entire streetscape with this sound? (1: not comfortable at all, 7: very comfortable)". It should be mentioned that playing the video and sound again on the PC was intended to recall the experience in the IVE. The evaluation of those who did not notice the sound during the experiment was based solely on their experience during the replay on the PC. Hence, for evaluation of the streetscape, I only considered participants who noticed the sound in the IVE. That is, those who answered Yes to the question "Have you noticed this sound?".

Procedure

The participants' experiment consisted of four stages in total. The operation of parametric loudspeakers produces high frequency sounds that may cause discomfort or pain to some people. Therefore, initially, pink noise was presented as a sample sound source from the parametric loudspeaker to ensure that the experiment could proceed without any issues. The sample sound source consisted of the same stimuli as the subsequent perception experiment.

Next, I confirmed whether participants were able to correctly perceive the spatial sound of the auditory display. As discussed in Chapter 4.2, individual differences exist in the perception of stereophonic sound, but I used only one type of HRTF for this experiment. I used a pink noise as the sample sound source, which is commonly used in previous studies to measure spatial sound localization accuracy [24]. I convolved the pink noise with seven HRTFs, including 0 degrees in the center and 30, 60, and 90 degrees to the left and right. The order of the presentation was randomized.

As a reference for the participants, I projected an image of angles shown in Figure 4.2 on the screen. Participants were informed that they could give answers other than the angles shown in the image. Participants were asked to verbally tell the experimenter from which direction(s) (left/right) they heard the sound source. They were not required to specify left or right for 0 degrees. Since head position affected the accuracy of spatial sound perception, the experimenter instructed participants to face forward as much as possible.

After conducting the spatial sound perception test, I proceeded to verify the effects using the simulated routes. A practice route was used to explain the operation method. As shown in Figure 4.3, the decision points had choices with the information of the distance and time required to reach the destination. Participants were instructed to choose the path based on this information and consider the overall landscape and soundscape.

At the decision points, the control of the auditory display was manually triggered following the Wizard-of-Oz approach. When the participant reached the decision point or started to move toward the next decision point, the auditory display was manually played or stopped by the experimenter, synchronized as much as possible with the video.

Finally, the participants were asked to complete a questionnaire.

Participants

The participants were 16 individuals (8 males and 8 females, with an average age of 24.5 years, ranging from 18 to 34 years old) without any audiovisual impairments. Participants were recruited through snowball sampling, and they all were either university faculty members or acquaintances thereof. There were no restrictions regarding their backgrounds in acoustics or urban design. One participant completed the questionnaire in German. The responses were translated into English and included in the analysis. This experiment was approved by the ethical review committee of the ifgi at the University of Münster, and all participants provided informed consent before participation.

4.3.3 Results

Perceived Angle

First, the results are presented on whether participants were able to correctly perceive the spatial sound of the auditory display. The results for each angle were written in Figure 4.5. The results show that the angles are divided between those that are accurately localized and those that are not. In this experiment, there were no roads in the same direction at a single decision point, so it was important that the left-right direction cues were not confused. That is, it was necessary to be able to at least distinguish between left and right, and this condition was satisfied. Note that one participant answered partially reversed left and right. Thus, I exclude all of this participants' results, and report the perceived angle and following results of the experiment on 15 participants.

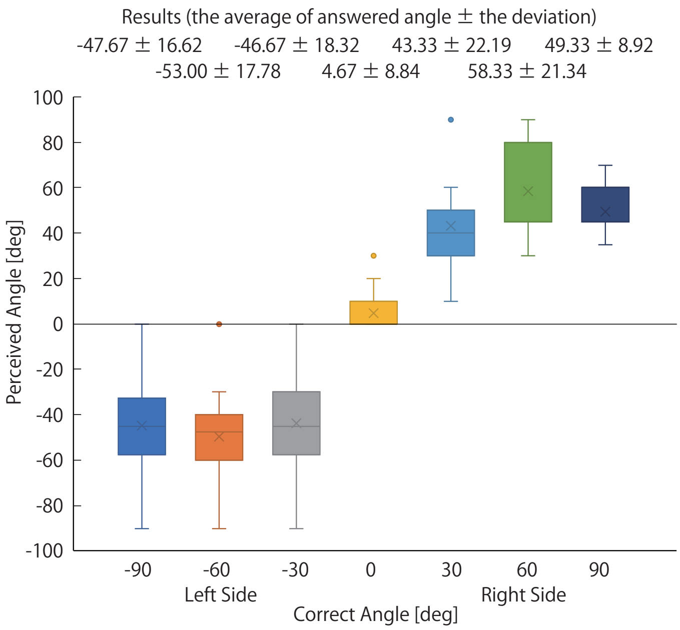

Figure 4.5: The results of the perceived angle of the stereophonic sound. Negative values on the x,y axis correspond to the left side and positive values to the right side.

Detour Ratio

First, the results of how far the participants were encouraged to go by being presented with guiding sounds are shown by the sound type. The detour ratio will be compared to data with no sound intervention as a baseline. The baseline data will use the results from the previous study [160] when there was no intervention and not in a pandemic situation ($n = 26$). As a note, the prior paper calculated per-route-detour ratio, whereas here I recalculated the data by per-intersection-detour ratio.

All results were shown in Figure 4.6. For the accordion-based guiding sound, the average detour ratio was 5.14 ± 1.42 %. The applause sound resulted in an average detour ratio of 8.72 ± 3.32 %. The nature sound led to an average detour ratio of 30.44 ± 8.90 %. For the baseline (= no intervention), the detour ratio was 2.77 ± 1.00 %.

These results were analyzed with a one-way independent-measures ANOVA (type: four levels - accordion, applause, nature sound, and baseline). There was a significant difference at the $< 1\%$ level ($F(3{,}243) = 5.505$, $p = 0.001$). A Tukey-HSD post-hoc test revealed that $p = 0.98$ for accordion and applause, $p = 0.021$ for accordion and nature, $p = 0.99$ for accordion and baseline, $p = 0.083$ for applause and nature, $p = 0.84$ for applause and baseline, $p < 0.001$ for nature and baseline. Hence, the significant differences were found between accordion and nature sound at 5 % level, applause and nature sound at 10 % level, and nature sound and baseline at 1 % level.

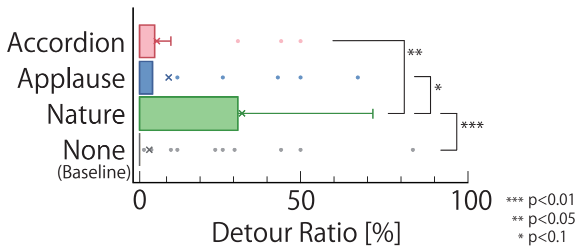

Figure 4.6: Results of the participants' detour ratio in the case of guiding sound

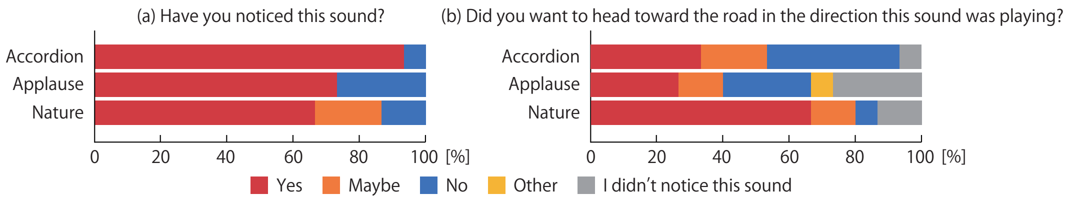

Figure 4.7: Participants' questionnaire results: (a) Have you noticed this sound? (b) Did you want to head toward the road in the direction this sound was playing?

Streetscape Perception

The responses to the questionnaire are presented below in summary form. Firstly, results regarding whether participants noticed the sound source and whether they wanted to head in the direction of the sound are displayed in Figure 4.7. Results show that almost all of the participants who noticed the nature sound wanted to walk its direction. In contrast, only about half of the participants expressed a desire to head towards the direction of the accordion or applause sounds.

The streetscape perception considered only those participants who indicated that they had heard the sound source during the experiment. Thus, evaluations from participants who did not notice the sound source (those who answered "No" to the question "Have you noticed this sound?") were excluded.

The results (see Figure 4.8) revealed an average Likert rating of 5.21 ± 2.03 for accordion, 4.36 ± 1.65 for applause, and 5.92 ± 1.91 for nature sounds. These results were subjected to a one-way independent-measures ANOVA (type: three levels - accordion, applause, and nature sound). A Tukey-HSD post-hoc test revealed $p = 0.29$ between accordion and applause, $p = 0.38$ between accordion and nature sound, and $p = 0.023$ between applause and nature sound. Therefore, there was a significant difference at the 5 % significance level between applause and nature sounds.

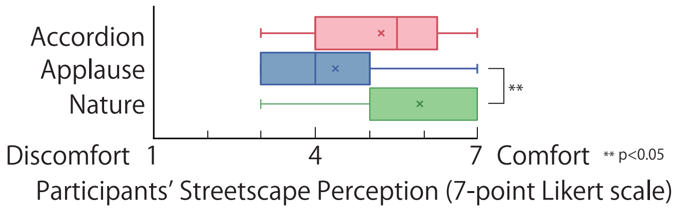

Figure 4.8: Results of the participants' streetscape perception in the case of guiding sound

4.3.4 Discussion

Participants' Feedback

The detour ratio and the results of the questionnaire showed that the nature sound was easier to guide while improving the streetscape perception compared to the accordion, applause sound, or no intervention. Regarding the participants' attitudes, the questionnaire also showed that they were more inclined to go in that direction when nature sound was presented.

Here are some feedback from participants regarding their decision (or lack thereof) to follow the path associated with the sound guidance:

Firstly, regarding the accordion sound, I had initially assumed that all participants would respond positively to it during the design stage, given that it is often played by street musicians. In fact, some participants said, "I want to know where the music is coming from" (P1), "I thought there might be a street musician" (P7), and "Sounds fun and there might be people there" (P12). Conversely, a few participants gave negative feedback about the type of music, stating, "I don't really like this kind of music" (P5) and "I don't like such music and the related bustle" (P13). Thus, it was found that individual music preferences can impact the effectiveness of sound guidance.

Secondly, for the applause sound, I considered to serve as an incentive for participants and easily boost their motivation to follow instructions. Some comments included, "Because it makes one curious and sounds as if something nice is going on" (P2), "I wanted to see what is going on" (P3), and "I was curious what was happening there" (P15). Therefore, it was discovered that the motivation was not directed towards the participants themselves but rather towards the expectation that there was some event taking place beyond the path indicated by the guiding sound. Some participants even avoided the direction of the applause sound due to the presumed presence of a crowd, for example, "The sound suggested there could be people there" (P4) and "It sounds like there would be a lot of people. So, I don't think it's comfortable to walk through" (P10). These reactions to the applause sound differed from my initial hypothesis.

Lastly, while nature sounds were initially thought to merely provide a comforting ambiance rather than serving as a guide, surprisingly, these sounds were reported to do both. Participants shared comments like, "Because it sounds as if a park would be nearby with birds singing and water flowing... nice to relax, especially in the city" (P2), "The sound suggested a quiet place in nature" (P3), and "Sounds like a very quiet and peaceful path... I love nature so I would enjoy the walk" (P8). The only negative feedback was, "reaching my destination was my goal, not the streetscape or anything else" (P6), suggesting that the nature sound was not relevant to the reason why this participant was not guided.

Design Implication of Guiding Sound

In this experiment, guiding sounds were employed solely to guide participants towards a particular path at intersections. Upon analyzing the participants' comments, it was found that individuals tend to imagine, what lies ahead of a junction when presented with guiding sounds. In the case of the applause sound, it evoked an 'expectation' of what could be happening ahead, and in the case of the natural sounds, it created an expectation of comfort, peace, and tranquility. These expectations consequently influenced their willingness to be guided in a particular direction. Conversely, the applause sound may also provoke a negative expectation of crowding, discouraging some individuals. Therefore, when designing a guiding sound, it is crucial to consider what the individual being guided may 'expect' to find ahead. This expectation will significantly influence their decision to follow the suggested path. However, there are also sounds, like music, which can divide individuals based on personal preference. In these instances, the individual might refuse to follow the guided path, not because of what they expect to find but because they do not like the sound itself. This suggests that these personal internal elements should also be considered as a parameter when designing a guiding sound. These expectations, however, were not met in my experimental setup. Depending on the interplay between the situation and the type of guiding sound used, there could be a risk of creating false expectations among the participants.

Each sound type proved effective under certain circumstances: when participants desired to be led towards an actual event, or when the intent was to foster relaxation or emphasize natural elements. Nevertheless, in other situations, the use of these sounds could unintentionally inflate passerby's expectations. Therefore, when designing guiding sounds, it's crucial to carefully select sounds that align well with the nature of the situation, thereby ensuring they serve their intended purpose without creating misleading cues.

4.4 Summary and Answers to Research Questions

In the section 4.3, I explored the sound designs for guiding pedestrians while maintaining the streetscape under normal conditions and evaluated their effectiveness. The findings suggest that sounds providing a sense of comfort, such as nature sounds, can enhance streetscape perceptions and motivate individuals to follow a certain path. However, I also discovered that certain sound types may give rise to undue expectations in pedestrians, highlighting the importance of careful sound selection according to specific situations.

The answer to research question 3-2 "What is the impact of sound-guided intervention on streetscape perception?" is that nature sounds can enhance streetscape perception.

Chapter 5

Conclusions: Guidelines and Limitations

5.1 Summary of the Dissertations

This doctoral dissertation proposed a system that includes sensing whether an individual needs information sound and whether it is necessary to emit sound based on the space, in order to resolve the trade-off between the effectiveness of auditory displays and spatial comfort. Furthermore, the objective was to investigate how the use of this system changed human consciousness, behavior, and spatial comfort with acoustics as a factor. From these investigations, the following research questions and answers were obtained:

RQ.1 How much is the noise level reduced by the use of parametric loudspeakers? (corresponds to Chapter 2)

A.1 Noise reduction is more pronounced with lower background noise levels. For instance, if the background noise is 20 dBA, the SRT for non-target listeners was 27.25 dBA with conventional loudspeakers, but it was 14.05 dBA with parametric loudspeakers. When combined with the background noise, the difference in sound level was 7.02 dBA.

RQ.2-1 Does varying the sound level of an ad's auditory display improve recognition? (corresponds to Chapter 3)

A.2-1 Ad recognition increases when the sound level is not only changed according to attention level, but also when the sound level is changed as a 'trigger' for passersby to look at the signage.

RQ.2-2 Does varying the sound level of an ad's auditory display make the acoustic environment less comfortable? (corresponds to Chapter 3)

A.2-2 If only some auditory displays vary in sound level, it does not negatively impact the acoustic environment.

RQ.3-1 What kind of sound design effectively guides pedestrian flow? (corresponds to Chapter 4)

A.3-1 During normal times, using comforting sounds such as nature sounds can effectively guide human flow because these sounds increase the motivation to follow the guidance.

RQ.3-2 What is the impact of sound-guided intervention on streetscape perception? (corresponds to Chapter 4)

A.3-2 Nature sounds can enhance streetscape perception.

5.2 Design Guidelines

When designing auditory displays, considering their effectiveness first brings forth a range of options. Extreme examples include making the sound level excessively loud or using conspicuous warning sounds. Extreme examples include making the sound level excessively loud or using conspicuous warning sounds. To resolve the trade-off issue, this dissertation poses the research question of sound level, exposure time, and sound genre as key parameters, and examines their effectiveness through examples of auditory displays that are according to individual and spatial contexts. While these parameters and examples are not comprehensive, it is concluded that manipulating these three parameters can achieve effectiveness while ensuring the spatial comfort The following are design guidelines for auditory displays based on these three parameters. These guidelines should be referenced based on what the purpose of the space is and what kind of space it is.

- **Sound Level** Previous research has indicated that sound level is an important parameter affecting acoustic comfort [12, 13]. Reducing the sound level being presented not only decreases the overall sound level in the space but also potentially avoids information masking [168]. Avoiding information masking leads to other sounds being more easily heard. One method for this is the parametric loudspeaker, which was examined in Chapter 2. It reduces the sound level to prevent it from becoming noise for the surroundings and also lowers the sound level for the target listener, which can be expected to improve acoustic comfort.

  However, beyond just altering the sound level, the use of noise cancellation [169] can also have a similar effect. While the technology for masking sound in open spaces is still developing [170], it effectively changes the level of the sound that is heard. Applying these technologies can ensure spatial comfort influenced by acoustics.

- **Exposure Time** Individuals vary in their context with presented content, such as level of interest, concern, and engagement. In previous research [171], cameras were used to identify an individual's gender to change the content displayed, which is a system responsive to the context of engagement. In contrast, this study adjusts exposure time to create a system that responds to the context of individual interest and concern. Chapter 3 demonstrated that by adapting this system to digital signage commonly found in stores, it is possible to improve comfort level without disrupting purchasing behavior. Adjusting exposure time in this way, by constantly changing the sound level, can avoid information masking, similar to the earlier discussed point.

Furthermore, it was found that exposing everyone to the signage for a short period only, rather than continuously, tends to increase the viewing duration. This can be understood as the audio version of the previous research on sudden object presentation [96], suggesting that the gimmick of sudden sound presentation is effective.

Thus, by controlling exposure time according to individual context, it is conceivable not only to maintain comfort level but also to improve utilization effectiveness.

- **Sound Genre** Traditionally, when conveying something only by sound, specific language messages or warning tones were often used [134, 43, 135]. This is largely due to the fact that spaces where only sound is presented are often in emergency situations. In contrast, this study examined sound design in situations where only sound is presented in non-emergency situations. Depending on the spatial context, the degree to which people want to hear the sound and the degree to which they want to follow the content of the sound varies. By changing the sound genre used depending on the context of the space, it was shown that it is possible to increase the effectiveness of sound presentation while ensuring comfort.

  For example, in chapter 4, a situation in which people are guided in a space where there is no emergency but congestion is to be alleviated was assumed. Through experiments, I found that pedestrians expect what is ahead when a sound is presented from the front. In terms of sound design, using natural sounds like streams and birds, which are positively evaluated in traditional soundscape research [155, 17, 156, 157], effectively guided people while maintaining the landscape. In cases where 100% guidance is not necessary, considering the genre of sound can lead to a comfortable acoustic environment.

5.3 Limitations and Future Directions

There are several limitations to this study.

First, the parametric loudspeakers used throughout the paper are highly directional and have the advantage of not scattering noise as described in chapter 2. However, as described in chapter 1.4, since ultrasonic waves are used to propagate sound, the restored audible sound is distorted. In particular, the high-frequency range is heard well and the low-frequency range is difficult to hear. This can be seen in the chapter 2.5 results, which show that sounds with low fundamental frequencies are difficult to hear. In cases where sound quality is important, the auditory display using parametric loudspeakers proposed in this thesis is not suitable.

Additionally, the auditory displays proposed in Chapters 3 and 4 are currently limited to specific spaces. The system overview in Chapter3 is designed for presenting to only one person at a time. In Chapter4, there is no sensing to differentiate which individuals are approaching a junction and which are not. Thus, as it stands, the system is only practical in spaces like store aisles or less busy paths. To expand the usable spaces in the future, it would be necessary to increase the number of parametric loudspeakers or manipulate sound phases within a large parametric loudspeaker to create multiple listening points simultaneously. Additionally, it's essential to investigate how auditory displays for groups impact each individual's behavior and consciousness.

The current experiments are conducted in a laboratory setting with a small sample size of approximately 15 - 30 participants. Generally, when implementing in society and observing people's behavior or examining the effectiveness of use, it is preferable to observe a larger and more diverse group of people. Therefore, it is necessary to conduct larger-scale experiments building upon the insights from this dissertation.

Another limitation is the lack of effectiveness verification and comfort surveys in spaces where a variety of sounds are intermingled. For example, in this thesis, auditory displays for advertisement sounds and natural sounds serve different purposes, as they are according to individual and spatial contexts, respectively. In future applications within the same space, the pinpoint audio presentation system proposed in this thesis, with camera sensing and parametric loudspeakers, could enable each person to hear only the sounds they need. However, to realize this, it is crucial to clarify what kind of sound each person needs, i.e., how the individual's behavior reflects their contexts and needs. It is also important for the design of auditory displays to observe the behavioral changes of individuals and the group as a whole when presented with sounds. Therefore, it is desirable to examine this issue in a space where multiple people with different behavioral intentions coexist.

In this dissertation, the evaluations are done by investigating each parameter separately to determine whether the effects of auditory display use and spatial comfort are compatible. However, it is desirable to develop a quantitative evaluation that encompasses both items to establish to what extent the trade-off issue has been resolved. For this purpose, it is better to establish an evaluation function for the trade-off relationship. In order to avoid the incorrect case where both indicators are compatible at low levels, the formula should also include whether the utilization effectiveness and spatial comfort are good values in themselves. It is a future task to propose and make available a formula to evaluate the degree to which the trade-off issue is resolved, which can be compared across different use scenarios.

As for potential social acceptance limitations, ethical issues with sensing and the possibility of people ignoring the system with repeated use are considered. In Chapter 3, the state of customers is sensed, such as whether they are approaching or paying attention to the signage. However, some individuals might reject this due to privacy concerns. With the increasing prevalence of smartphones, there is a growing focus on privacy and the concept of digital sovereignty [172] is becoming more relevant. This research is intrinsically linked to digital sovereignty. When spreading the system, it is desirable to consider ethical perspectives and provide the choice for people to either use or not use the system.

Furthermore, as mentioned, there is a phenomenon known as "Display Blindness" where people in public spaces fail to notice or ignore existing displays [71, 73]. Particularly, the pinpoint auditory glimpse in Section 3.4 was intended to solve this issue. However, constant use of auditory displays may lead to pedestrians ignoring these sounds, a phenomenon that could be termed "Auditory Blindness." This is anticipated as a social acceptance limitation. One possible cause could be habituation. Potential solutions include presenting the pinpoint auditory glimpse only when new content is delivered, or designing sounds that benefit pedestrians, such as the natural sounds discussed in Section 4.3. Addressing this Auditory Blindness is a problem to be solved in the future.

The current urban culture also dictates cases where the use should be avoided. In Japan, where the author grew up, advertising billboards are ubiquitous, and in urban areas like Shinjuku and Shibuya, advertising sounds are heard everywhere. Therefore, it is anticipated that the introduction of auditory displays will not be a significant barrier. However, this may not be the case in other countries and cities. In regions where sound use is restricted, the systems described in this dissertation could potentially create noise and detract from spatial comfort.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Associate Professor Keiichi Zempo (University of Tsukuba). Prof. Zempo has been incredibly supportive, offering numerous insightful consultations and accurate advice. He has also given me various opportunities to study, not only in research, but also in presentations at international conferences, study abroad, and exhibitions. Prof. Zempo knew my personality and tailored his guidance to it, making him an exceptional supervisor. Thanks to him, I had wonderful six years.

I would like to express my deep appreciation to Professor Naoto Wakatsuki (University of Tsukuba). I had many discussions with Prof. Wakatsuki about my research. Whenever he accurately pointed out the holes in my research, I realized how immature my research was, and he helped me to grow. I am also very grateful for the maintenance and management of the laboratory, including the internet environment. Thanks to you, I have been able to spend my laboratory life without any inconvenience.

I would like to express my sincere gratitude to Associate Professor Tadashi Ebihara (University of Tsukuba). Prof. Ebihara gave me unique and important perspectives in discussing my research. He also used humorous phrases and treated us in a close distance for students. Thanks to you, I was able to spend my time in the laboratory in a cheerful way.

I would like to express my great appreciation to Assistant Professor Yuka Maeda (University of Tsukuba). Prof. Maeda gave me sharp suggestions in my research. Also, when we met, she talked to me casually and I enjoyed spending time with her. When I was a teaching fellow in the experiment class that she was in charge of, I was able to absorb new knowledge about electronic circuits not only for students but also for myself.

I would like to express special thanks to Professor Emeritus Koichi Mizutani (University of Tsukuba). Prof. Mizutani gave me advice on various aspects of my research ideas. He also always gave me advice on how to design figures so that readers would be easy to read. Outside of research, he always made us laugh with jokes and took pictures with us, always keeping our hearts close. Thank you very much for letting me have a pleasant laboratory life.

I would also like to extend my gratitude to Professor Christian Kray (University of Münster), who has been of immense help since my study abroad experience. Prof. Kray graciously accepted my request to study abroad and has shown great kindness, from assisting with preparations to offering research advice. I have learned much from his approach of respecting personal life and fostering constant communication among lab members. I am deeply thankful for the valuable experience he provided.

I would like to thank Specially-Appointed Professor Toshimasa Yamanaka (University of Tsukuba) and Professor Soh Masuko (Shibaura Institute of Technology) for their great help in the joint research, and I am deeply grateful for their support. I have received various inspirations every time I talked with them. It was a fulfilling time for me to learn about a world I did not know even outside of my research. Thank you very much.

I would like to thank the Japan Society for the Promotion of Science (JSPS) for providing me with research funds and accommodation expenses. Thanks to them, my three years of research and six months of study abroad were fulfilling.

I would also like to thank all the participants who took part in the experiments. Thanks to your time and cooperation, I am able to write my doctoral dissertation.

Thank you very much to all of you with whom I have shared my laboratory life. I have many vivid memories. Especially, I would like to thank Mr. Yuichi Mashiba, Mr. Koki Shimoda, Mr. Ryotaro Chinone, Mr. Kazuki Yamada (currently at OPTiM Corporation), Ms. Haruna Miyakawa , Mr. Hiroki Uchida (currently at CRI Middleware Co., Ltd.), Dr. Misaki Mito (currently at Hitachi America, Ltd.), and Mr. Hirotaka Obo. Not only did they support me in my research, but they also made the time I spent inside and outside the lab enjoyable. I am deeply grateful to all of them.

I would like to thank the many people who supported me during my study abroad in University of Münster. In particular, Dr. Simge Özdal Oktay, Mr. Sven Heitmann, Dr. Tessio Novack, Ms. Eftychia Koukouraki, and Mr. Benjamin Karic have enriched my research life. They also provided me with various kinds of dedicated support after I suffered a serious injury early in my study abroad. They definitely made my inconvenient life less inconvenient. Thank you very much. I am also very grateful to the institute secretary, Mr. Karsten Höwelhans. He was very reliable and helpful, immediately resolving various questions I had regarding study abroad, treatment for injuries, and rewards for experiments. Without his support, I would not have been able to return to Japan as planned. Thank you very much.

My research life so far has required time to forget about research. Many thanks to all the people who have enriched this time for me. In particular, Mr. Ryoichi Ishijima, Ms. Fuyuri Koyama, Mr. Kenji Hashimoto, Mr. Kyohei Degura, Mr. Masahiro Yoshida, Ms. Mika Ohba, Ms. Ririko Nakata, Ms. Rumi Yamada, Mr. Yoichi Kanezaki, Ms. Yuka Funaki, Mr. Yukihiro Masaki, Ms. Yuri Hachisuka, and many others too numerous to mention. Thank you all very much.

To Mr. Dan Yoshikawa - *Thanks to my peers I was always able to smile* [173]. You've made my days.

Lastly, I would like to thank my family - Dr. Zoen So, Ms. Rie So, and Ms. Kiu Matsui - for watching over me during my student life so far. They are the ones who have supported me during my difficult times. I have always respected the humanity of my family. I will do my best to catch up with you all one day.

Bibliography

[1] Michael Frank Southworth. *The sonic environment of cities.* PhD thesis, Massachusetts Institute of Technology, 1967.

[2] R Murray Schafer. *The soundscape: Our sonic environment and the tuning of the world*. Simon and Schuster, 1993.

[3] Bing Chen and Jian Kang. Acoustic comfort in shopping mall atrium spaces—a case study in sheffield meadowhall. *Architectural Science Review*, 47(2):107–114, 2004.

[4] Wei Yang and Jian Kang. Acoustic comfort evaluation in urban open public spaces. *Applied acoustics*, 66(2):211–229, 2005.

[5] Aimee Desert, Emanuele Naboni, and David Garcia. The spatial comfort and thermal delight of outdoor misting installations in hot and humid extreme environments. *Energy and Buildings*, 224:110202, 2020.

[6] Wonyoung Yang and Hyeun Jun Moon. Combined effects of acoustic, thermal, and illumination conditions on the comfort of discrete senses and overall indoor environment. *Building and Environment*, 148:623–633, 2019.

[7] Deepak Amaripadath, Ramin Rahif, Mirjana Velickovic, and Shady Attia. A systematic review on role of humidity as an indoor thermal comfort parameter in humid climates. *Journal of Building Engineering*, page 106039, 2023.

[8] British Standards Institution. Bs iso 12913-1:201, acoustics - soundscape; part 1: Definition and conceptual framework. *ISO*, 2014.

[9] Yue Wu, Jian Kang, Wenzhong Zheng, and Yongxiang Wu. Acoustic comfort in large railway stations. *Applied Acoustics*, 160:107137, 2020.

[10] Man Sze Tse, Chi Kwan Chau, Yat Sze Choy, Wai Keung Tsui, Chak Ngai Chan, and Shiu Keung Tang. Perception of urban park soundscape. *The Journal of the Acoustical Society of America*, 131(4):2762–2771, 2012.

[11] Jieling Xiao and Francesco Aletta. A soundscape approach to exploring design strategies for acoustic comfort in modern public libraries: A case study of the library of birmingham. *Noise mapping*, 3(1), 2016.

[12] Prabal BANDYOPADHYAY, Sunil Kumar BHATTACHARYA, and Shri Kant KASHYAP. Assessment of noise environment in a major railway station in india. *Industrial health*, 32(3):187–192, 1994.

[13] Jikke Reinten, P Ella Braat-Eggen, Maarten Hornikx, Helianthe SM Kort, and Armin Kohlrausch. The indoor sound environment and human task performance: A literature review on the role of room acoustics. *Building and Environment*, 123:315–332, 2017.

[14] Paulo Henrique Trombetta Zannin, Andressa Maria Coelho Ferreira, and Bani Szeremetta. Evaluation of noise pollution in urban parks. *Environmental monitoring and assessment*, 118:423–433, 2006.

[15] Jian Ge and Kazunori Hokao. Research on the sound environment of urban open space from the viewpoint of soundscape–a case study of saga forest park, japan. *ACTA Acustica united with Acustica*, 90(3):555–563, 2004.

[16] Joo Young Hong, Zhen-Ting Ong, Bhan Lam, Kenneth Ooi, Woon-Seng Gan, Jian Kang, Jing Feng, and Sze-Tiong Tan. Effects of adding natural sounds to urban noises on the perceived loudness of noise and soundscape quality. *Science of the Total Environment*, 711:134571, 2020.

[17] Sabrina Skoda, J Steffens, and J Becker-Schweitzer. Road traffic noise annoyance in domestic environments can be reduced by water sounds. In *Forum Acusticum*, 2014.

[18] Felipe Pantoja and Adilson Borges. Background music tempo effects on food evaluations and purchase intentions. *Journal of Retailing and Consumer Services*, 63:102730, 2021.

[19] Michael Beverland, Elison Ai Ching Lim, Michael Morrison, and Milé Terziovski. Instore music and consumer–brand relationships: Relational transformation following experiences of (mis) fit. *Journal of Business Research*, 59(9):982–989, 2006.

[20] Andy Bennett and Ian Rogers. Street music, technology and the urban soundscape. *Continuum*, 28(4):454–464, 2014.

[21] Jonathan Adam. Street music, city rhythms: The urban soundscape as heard by street musicians, 2018.

[22] Kenichi Aoki, Tomoo Kamakura, and Yoshiro Kumamoto. Parametric loudspeaker—characteristics of acoustic field and suitable modulation of carrier ultrasound. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 74(9):76–82, 1991.

[23] Masahide Yoneyama, Jun-ichiroh Fujimoto, Yu Kawamo, and Shoichi Sasabe. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America*, 73(5):1532–1536, 1983.

[24] Shigeaki Aoki, Masayoshi Toba, and Norihisa Tsujita. Sound localization of stereo reproduction with parametric loudspeakers. *Applied Acoustics*, 73(12):1289–1295, 2012.

[25] Stephen Roddy. *Embodied sonification.* PhD thesis, Trinity College Dublin, 2016.

[26] Abdelsalam Helal, Steven Edwin Moore, and Balaji Ramachandran. Drishti: An integrated navigation system for visually impaired and disabled. In *Proceedings fifth international symposium on wearable computers*, pages 149–156. IEEE, 2001.

[27] Masaki Matsuo, Takahiro Miura, Ryosuke Ichikari, Karimu Kato, and Takeshi Kurata. Otasce map: a mobile map tool with customizable audio-tactile cues for the visually impaired. *Journal on Technology and Persons with Disabilities*, pages 82–103, 2020.

[28] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. Navigation system for the blind: Auditory display modes and guidance. *Presence*, 7(2):193–203, 1998.

[29] Chetan Bulla, Sourabh Zutti, Sneha Potadar, Swati Kulkarni, and Akshay Chavan. Low-cost automated navigation system for visually impaired people. In *Soft Computing: Theories and Applications: Proceedings of SoCTA 2020, Volume 1*, pages 125–135. Springer, 2022.

[30] Masaki Kuribayashi, Seita Kayukawa, Jayakorn Vongkulbhisal, Chieko Asakawa, Daisuke Sato, Hironobu Takagi, and Shigeo Morishima. Corridor-walker: Mobile indoor walking assistance for blind people to avoid obstacles and recognize intersections. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI):1–22, 2022.

[31] Masaki Kuribayashi, Tatsuya Ishihara, Daisuke Sato, Jayakorn Vongkulbhisal, Karnik Ram, Seita Kayukawa, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. Pathfinder: Designing a map-less navigation system for blind people in unfamiliar buildings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2023.

[32] Mohsen Bakouri, Naif Alyami, Ahmad Alassaf, Mohamed Waly, Tariq Alqahtani, Ibrahim AlMohimeed, Abdulrahman Alqahtani, Md Samsuzzaman, Husham Farouk Ismail, and Yousef Alharbi. Sound-based localization using lstm networks for visually impaired navigation. *Sensors*, 23(8):4033, 2023.

[33] Keiichi Zempo, Yuichi Mashiba, Takayuki Kawamura, Noko Kuratomo, and Hisham Elser Bilal Salih. Phonoscape: Auralization of photographs using stereophonic auditory icons. In *Adjunct Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 117–119, 2018.

[34] Samuel Paré, Maxime Bleau, Ismaël Djerourou, Vincent Malotaux, Ron Kupers, and Maurice Ptito. Spatial navigation with horizontally spatialized sounds in early and late blind individuals. *PLoS One*, 16(2):e0247448, 2021.

[35] Takayuki KOMODA, Fumina UTSUMI, Takayoshi YAMADA, and Keiichi ZEMPO. Auditory indication system for object-finding in remote collaborative assistance. 2022.

[36] Oliver Beren Kaul, Kersten Behrens, and Michael Rohs. Mobile recognition and tracking of objects in the environment through augmented reality and 3d audio cues for people with visual impairments. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2021.

[37] Jens C Hegg, Jonathan Middleton, Ben Luca Robertson, and Brian P Kennedy. The sound of migration: exploring data sonification as a means of interpreting multivariate salmon movement datasets. *Heliyon*, 4(2), 2018.

[38] Kongmeng Liew and PerMagnus Lindborg. A sonification of cross-cultural differences in happiness-related tweets. *Journal of the Audio Engineering Society*, 68(1/2):25–33, 2020.

[39] Stephen Roddy. Using conceptual metaphors to represent temporal context in time series data sonification. *Interacting with Computers*, 31(6):555–576, 2019.

[40] Terry Caelli and David Porter. On difficulties in localizing ambulance sirens. *Human Factors*, 22(6):719–724, 1980.

[41] K Catchpole and D McKeown. A framework for the design of ambulance sirens. *Ergonomics*, 50(8):1287–1301, 2007.

[42] Jesseca RI Taylor and Michael S Wogalter. Specific evacuation instructions enhance spoken fire warnings. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 59, pages 1486–1490. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[43] Sander J Van Wijngaarden, Adelbert W Bronkhorst, and Louis C Boer. Auditory evacuation beacons. *Journal of the Audio Engineering Society*, 53(1/2):44–53, 2005.

[44] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. Auditory icon and earcon mobile service notifications: intuitiveness, learnability, memorability and preference. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1513–1522, 2009.

[45] Sasha Devore, Antje Ihlefeld, Kenneth Hancock, Barbara Shinn-Cunningham, and Bertrand Delgutte. Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. *Neuron*, 62(1):123–134, 2009.

[46] Andrew D Brown, G Christopher Stecker, and Daniel J Tollin. The precedence effect in sound localization. *Journal of the Association for Research in Otolaryngology*, 16(1):1–28, 2015.

[47] William Evans, Jakob Dyreby, Søren Bech, Slawomir Zielinski, and Francis Rumsey. Effects of loudspeaker directivity on perceived sound quality-a review of existing studies. In *Audio Engineering Society Convention 126*. Audio Engineering Society, 2009.

[48] Woon-Seng Gan, Jun Yang, and Tomoo Kamakura. A review of parametric acoustic array in air. *Applied Acoustics*, 73(12):1211–1219, 2012.

[49] Yusuke Nakashima, Takeshi Yoshimura, Nobuhiko Naka, and Tomoyuki Ohya. Prototype of mobile super directional loudspeaker. *NTT DoCoMo Tech J*, 8(1):25–32, 2006.

[50] Chuang Shi and Woon-Seng Gan. Development of parametric loudspeaker. *IEEE potentials*, 29(6):20–24, 2010.

[51] Marco Alunno and Andres Yarce Botero. Directional landscapes: using parametric loudspeakers for sound reproduction in art. *Journal of new music research*, 46(2):201–211, 2017.

[52] Youngmin Kim, Joonku Hahn, Young-Hoon Kim, Jonghyun Kim, Gilbae Park, Sung-Wook Min, and Byoungho Lee. A directional-view and sound system using a tracking method. *IEEE transactions on broadcasting*, 57(2):454–461, 2011.

[53] Akihiko Shirai, Hisataka Suzuki, Yuta Yamaguchi, and Kazuhisa Yanaka. Glassless augmented display for public signage. In *Proceedings of the 2016 Virtual Reality International Conference*, pages 1–4, 2016.

[54] Naoya Muramatsu, Kazuki Ohshima, Ryota Kawamura, Ooi Chun Wei, Yuta Sato, and Yoichi Ochiai. Sonoliards: Rendering audible sound spots by reflecting the ultrasound beams. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology*, pages 57–59, 2017.

[55] Yoichi Ochiai, Takayuki Hoshi, and Ippei Suzuki. Holographic whisper: Rendering audible sound spots in three-dimensional space by focusing ultrasonic waves. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4314–4325, 2017.

[56] Gianluca Memoli, Letizia Chisari, Jonathan P Eccles, Mihai Caleap, Bruce W Drinkwater, and Sriram Subramanian. Vari-sound: A varifocal lens for sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[57] Takashi Shimizu, Kazumi Makishima, Masafumi Yoshida, and Hidetoshi Yamagishi. Effect of background noise on perception of english speech for japanese listeners. *Auris Nasus Larynx*, 29(2):121–125, 2002.

[58] Odette Scharenborg, Juul MJ Coumans, and Roeland van Hout. The effect of background noise on the word activation process in nonnative spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(2):233, 2018.

[59] Odette Scharenborg and Marjolein van Os. Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Communication*, 108:53–64, 2019.

[60] Shigeaki Amano. Effects of lexicon and coarticulation on phoneme perception. *Journal of the Acoustical Society of Japan (E)*, 14(2):91–97, 1993.

[61] Shigeaki Amano and Tadahisa Kondo. Estimation of mental lexicon size with word familiarity database. In *Fifth International Conference on Spoken Language Processing*, 1998.

[62] Shuichi Sakamoto, Tadahiro Yoshikawa, Shigeaki Amano, Yôiti Suzuki, and Tadahisa Kondo. New 20-word lists for word intelligibility test in japanese. In *Ninth International Conference on Spoken Language Processing*, 2006.

[63] Yosuke Kobayashi and Kazuhiro Kondo. Japanese speech intelligibility estimation and prediction using objective intelligibility indices under noisy and reverberant conditions. *Applied Acoustics*, 156:327–335, 2019.

[64] Katsuhiko Yamamoto, Toshio Irino, Shoko Araki, Keisuke Kinoshita, and Tomohiro Nakatani. Gedi: Gammachirp envelope distortion index for predicting intelligibility of enhanced speech. *Speech Communication*, 123:43–58, 2020.

[65] Shigeaki Amano, Tadahisa Kondo, and Kazumi Kato. Familiarity effect on spoken word recognition in japanese. In *Proceedings of the 14th ICPhS*, volume 2, pages 873–876, 1999.

[66] Kotaro Sugiyama and Tim Andree. *The Dentsu Way: 9 Lessons for Innovation in Marketing from the World's Leading Advertising Agency.* McGraw-Hill, 2011.

[67] Chu Chih-Chung, Chris Chang, Lee Wei-Chun Lin, et al. The effect of advertisement frequency on the advertisement attitude-the controlled effects of brand image and spokesperson's credibility. *Procedia-Social and behavioral sciences*, 57:352–359, 2012.

[68] John V Harrison and Anna Andrusiewicz. A virtual marketplace for advertising narrowcast over digital signage networks. *Electronic Commerce Research and Applications*, 3(2):163–175, 2004.

[69] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. How to evaluate public displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, pages 1–6, 2012.

[70] Charles Dennis, J Joško Brakus, Suraksha Gupta, and Eleftherios Alamanos. The effect of digital signage on shoppers' behavior: The role of the evoked experience. *Journal of Business research*, 67(11):2250–2257, 2014.

[71] Elaine M Huang, Anna Koster, and Jan Borchers. Overcoming assumptions and uncovering practices: When does the public really look at public displays? In *International Conference on Pervasive Computing*, pages 228–243. Springer, 2008.

[72] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and design space for interactive public displays. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1285–1294, 2010.

[73] Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzeck, Albrecht Schmidt, Tim Jay, and Antonio Krüger. Display blindness: The effect of expectations on attention towards digital signage. In *International conference on pervasive computing*, pages 1–8. Springer, 2009.

[74] Jens Nordfält. Improving the attention-capturing ability of special displays with the combination effect and the design effect. *Journal of Retailing and Consumer Services*, 18(3):169–173, 2011.

[75] Francesco Aletta, Federica Lepore, Eirini Kostara-Konstantinou, Jian Kang, and Arianna Astolfi. An experimental study on the influence of soundscapes on people's behaviour in an open public space. *Applied sciences*, 6(10):276, 2016.

[76] Qi Meng, Tingting Zhao, and Jian Kang. Influence of music on the behaviors of crowd in urban open public spaces. *Frontiers in psychology*, 9:596, 2018.

[77] Jian Kang and Mei Zhang. Semantic differential analysis of the soundscape in urban open public spaces. *Building and environment*, 45(1):150–157, 2010.

[78] Lei Yu and Jian Kang. Factors influencing the sound preference in urban open spaces. *Applied Acoustics*, 71(7):622–633, 2010.

[79] Wei Yang and Jian Kang. Soundscape and sound preferences in urban squares: a case study in sheffield. *Journal of urban design*, 10(1):61–80, 2005.

[80] RF Soames Job et al. Noise sensitivity as a factor influencing human reaction to noise. *Noise and Health*, 1(3):57, 1999.

[81] Richard A Page. Noise and helping behavior. *Environment and Behavior*, 9(3):311–334, 1977.

[82] World Health Organization et al. Environmental noise guidelines for the european region. 2018.

[83] Jussi Kuutti, Juhana Leiwo, and Raimo E Sepponen. Local control of audio environment: a review of methods and applications. *Technologies*, 2(1):31–53, 2014.

[84] S Roland Hall. *Retail Advertising and Selling: Advertising, Merchandise Display, Sales-planning, Salesmanship, Turnover and Profit-figuring in Modern Retailing, Including" Principles of Typography as Applied to Retail Advertising"*. McGraw-Hill book Company, Incorporated, 1924.

[85] Brian J Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, pages 1–7, 2009.

[86] Elad Yom-Tov, Jinia Shembekar, Sarah Barclay, and Peter Muennig. The effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. *NPJ digital medicine*, 1(1):24, 2018.

[87] Michel Chevalier. Increase in sales due to in-store display. *Journal of Marketing Research*, 12(4):426–431, 1975.

[88] Gary F McKinnon, J Patrick Kelly, and E Doyle Robison. Sales effects of point-of-purchase in-store signing. *Journal of Retailing*, 57(2):49–63, 1981.

[89] Arch G Woodside and Gerald L Waddle. Sales effects of in-store advertising. *Journal of Advertising Research*, 15(3):29–33, 1975.

[90] Guiying Du, Lukas Lohoff, Jakub Krukar, and Sergey Mukhametov. Comparing two methods to overcome interaction blindness on public displays. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, pages 243–244, 2016.

[91] Mojgan Ghare, Marvin Pafla, Caroline Wong, James R Wallace, and Stacey D Scott. Increasing passersby engagement with public large interactive displays: A study of proxemics and conation. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*, pages 19–32, 2018.

[92] Hannu Kukka, Heidi Oja, Vassilis Kostakos, Jorge Gonçalves, and Timo Ojala. What makes you click: exploring visual signals to entice interaction on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1699–1708, 2013.

[93] Jörg Müller and Antonio Krüger. Mobidic: context adaptive digital signage with coupons. In *European Conference on Ambient Intelligence*, pages 24–33. Springer, 2009.

[94] Hannah Limerick. Call to interact: communicating interactivity and affordances for contactless gesture controlled public displays. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*, pages 63–70, 2020.

[95] Wendy Ju and David Sirkin. Animate objects: How physical motion encourages public interaction. In *International Conference on Persuasive Technology*, pages 40–51. Springer, 2010.

[96] Steven Houben and Christian Weichel. Overcoming interaction blindness through curiosity objects. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1539–1544. 2013.

[97] Nicholas S Dalton, Emily Collins, and Paul Marshall. Display blindness? looking again at the visibility of situated displays using eye-tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3889–3898, 2015.

[98] Jörg Müller and Antonio Krüger. Competing for your attention: negative externalities in digital signage advertising. In *Workshop at Pervasive 2007 Designing and Evaluating Ambient Information Systems*, page 9. Citeseer, 2007.

[99] Sakurazaka 46 (Performance) Yasushi Akimoto (Lyrics). Shonin-yokkyu. https://youtu.be/x_BjvhMW9TE (Lyrics in English are available in the subtitles of the music video on YouTube), 2023.

[100] Thorsten Prante, Carsten Röcker, Norbert Streitz, Richard Stenzel, Carsten Magerkurth, Daniel Van Alphen, and Daniela Plewe. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, volume 2003, pages 277–278, 2003.

[101] Daniel Vogel and Ravin Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 137–146, 2004.

[102] Daniel Michelis and Jörg Müller. The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *Intl. Journal of Human–Computer Interaction*, 27(6):562–579, 2011.

[103] Florian Alt, Daniel Buschek, David Heuss, and Jörg Müller. Orbuculum-predicting when users intend to leave large public displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(1):1–16, 2021.

[104] David Allan. A content analysis of music placement in prime-time television advertising. *Journal of Advertising Research*, 48(3):404–417, 2008.

[105] Ann-Kristin Herget, Holger Schramm, and Priska Breves. Development and testing of an instrument to determine musical fit in audio–visual advertising. *Musicae Scientiae*, 22(3):362–376, 2018.

[106] Behringer. Measurement condenser microphone ecm8000 technical specifications. https://mediadl.musictribe.com/media/sys_master/h47/hde/8849927929886.pdf, 2013.

[107] Pasquale Bottalico, Juliana Codino, Lady Catherine Cantor-Cutiva, Katherine Marks, Charles J Nudelman, Jean Skeffington, Rahul Shrivastav, Maria Cristina Jackson-Menaldi, Eric J Hunter, and Adam D Rubin. Reproducibility of voice parameters: the effect of room acoustics and microphones. *Journal of Voice*, 2018.

[108] Richard E Petty and John T Cacioppo. The elaboration likelihood model of persuasion. In *Communication and persuasion*, pages 1–24. Springer, 1986.

[109] Carol L Krumhansl. Music: A link between cognition and emotion. *Current directions in psychological science*, 11(2):45–50, 2002.

[110] Meihui Ba, Jian Kang, and Zhongzhe Li. The effects of sounds and food odour on crowd behaviours in urban public open spaces. *Building and Environment*, 182:107104, 2020.

[111] Fashu Yi and Jian Kang. Effect of background and foreground music on satisfaction, behavior, and emotional responses in public spaces of shopping malls. *Applied Acoustics*, 145:408–419, 2019.

[112] Harry Brignull and Yvonne Rogers. Enticing people to interact with large public displays in public spaces. In *Interact*, volume 3, pages 17–24, 2003.

[113] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. Looking glass: a field study on noticing interactivity of a shop window. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 297–306, 2012.

[114] Young Eun Lee and Izak Benbasat. Interaction design for mobile product recommendation agents: Supporting users' decisions in retail stores. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(4):1–32, 2010.

[115] Stephanie van de Sanden, Kim Willems, and Malaika Brengman. How do consumers process digital display ads in-store? the effect of location, content, and goal relevance. *Journal of Retailing and Consumer Services*, 56:102177, 2020.

[116] Sung-Joon Yoon and Yong-Gil Choi. Determinants of successful sports advertisements: The effects of advertisement type, product type and sports model. *Journal of Brand Management*, 12(3):191–205, 2005.

[117] Shih-Chieh Chuang, Chia-Ching Tsai, Yin-Hui Cheng, and Ya-Chung Sun. The effect of terminologies on attitudes toward advertisements and brands: Consumer product knowledge as a moderator. *Journal of Business and Psychology*, 24(4):485–491, 2009.

[118] Gonzalo Parra, Joris Klerkx, and Erik Duval. Understanding engagement with interactive public displays: an awareness campaign in the wild. In *Proceedings of The International Symposium on Pervasive Displays*, pages 180–185, 2014.

[119] Robert Walter, Gilles Bailly, and Jörg Müller. Strikeapose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 841–850, 2013.

[120] Christine Landorf. Managing for sustainable tourism: a review of six cultural world heritage sites. *Journal of Sustainable tourism*, 17(1):53–70, 2009.

[121] Hilary Du Cros and Weng Hang Kong. Congestion, popular world heritage tourist attractions and tourism stakeholder responses in macao. *International Journal of Tourism Cities*, 2020.

[122] Todd Gabe. Measurement and analysis of neighborhood congestion: Evidence from sidewalk pedestrian traffic and walking speeds. *Growth and Change*, 52(3):1633–1651, 2021.

[123] Bruce W Landis, Venkat R Vattikuti, Russell M Ottenberg, Douglas S McLeod, and Martin Guttenplan. Modeling the roadside walking environment: pedestrian level of service. *Transportation research record*, 1773(1):82–88, 2001.

[124] Nowar Raad and Matthew I Burke. What are the most important factors for pedestrian level-of-service estimation? a systematic review of the literature. *Transportation research record*, 2672(35):101–117, 2018.

[125] Derek K Chu, Elie A Akl, Stephanie Duda, Karla Solo, Sally Yaacoub, Holger J Schünemann, Amena El-Harakeh, Antonio Bognanni, Tamara Lotfi, Mark Loeb, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis. *The lancet*, 395(10242):1973–1987, 2020.

[126] Mehmet Koyuncu and Sonia Amado. Effects of stimulus type, duration and location on priming of road signs: Implications for driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(2):108–125, 2008.

[127] Miguel S Prieto and Alastair R Allen. Using self-organising maps in the detection and recognition of road signs. *Image and Vision Computing*, 27(6):673–683, 2009.

[128] Leah Zhang-Kennedy, Saira Aziz, Oluwafunminitemi Oluwadare, Lyndon Pan, Zeyu Wu, Sydney EC Lamorea, Soda Li, Michael Sun, and Ville Mäkelä. Passenger perceptions, information preferences, and usability of crowding visualizations on public displays in transit stations and vehicles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2023.

[129] Haruno Kataoka, Kyoko Hashiguchi, Kae Wago, Yusuke Ichikawa, Hirohisa Tezuka, Shinichiro Yamashita, Yusaku Kuhara, and Tetsuo Akiyama. Dynamic guide signs system to control pedestrian flow. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1572–1577, 2016.

[130] Ulf Blanke, Gerhard Tröster, Tobias Franke, and Paul Lukowicz. Capturing crowd dynamics at large scale events using participatory gps-localization. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–7. IEEE, 2014.

[131] Ruiwei Shen, Tsutomu Terada, and Masahiko Tsukamoto. A navigation system for crowd flow control by controlling information presentation. In *Proceedings of the 13th International Conference on Advances in Mobile Computing and Multimedia*, pages 6–13, 2015.

[132] BL Collins, MS Dahir, and D Madrzykowski. Visibility of exit signs in clear and smoky conditions. *Fire technology*, 29(2):154–182, 1993.

[133] Ling Tim Wong and KC Lo. Experimental study on visibility of exit signs in buildings. *Building and environment*, 42(4):1836–1842, 2007.

[134] Penelope Burns, Garry Stevens, Kate Sandy, Arnold Dix, Beverley Raphael, and Bob Allen. Human behaviour during an evacuation scenario in the sydney harbour tunnel. *Australian Journal of Emergency Management, The*, 28(1):20–27, 2013.

[135] Tron Vedul Tronstad, Gunnar D Jenssen, Claudia Moscoso, Carl Södersten, and Eugene Zaikonnikov. Sound signals to improve evacuation in road tunnels. *Fire safety journal*, 125:103431, 2021.

[136] Jan BF Van Erp, Hendrik AHC Van Veen, Chris Jansen, and Trevor Dobbins. Waypoint navigation with a vibrotactile waist belt. *ACM Transactions on Applied Perception (TAP)*, 2(2):106–117, 2005.

[137] Taro Maeda, Hideyuki Ando, Tomohiro Amemiya, Naohisa Nagaya, Maki Sugimoto, and Masahiko Inami. Shaking the world: galvanic vestibular stimulation as a novel sensation interface. In *ACM SIGGRAPH 2005 Emerging technologies*, pages 17–es. 2005.

[138] Max Pfeiffer, Tim Dünte, Stefan Schneegass, Florian Alt, and Michael Rohs. Cruise control for pedestrians: Controlling walking direction using electrical muscle stimulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2505–2514, 2015.

[139] Yudai Tanaka, Jun Nishida, and Pedro Lopes. Electrical head actuation: Enabling interactive systems to directly manipulate head orientation. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2022.

[140] Oliver Beren Kaul and Michael Rohs. Haptichead: A spherical vibrotactile grid around the head for 3d guidance in virtual and augmented reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3729–3740, 2017.

[141] Volker Willert, Julian Eggert, Jürgen Adamy, Raphael Stahl, and Edgar Korner. A probabilistic model for binaural sound localization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(5):982–994, 2006.

[142] Jaka Sodnik, Saso Tomazic, Raphael Grasset, Andreas Duenser, and Mark Billinghurst. Spatial sound localization in an augmented reality environment. In *Proceedings of the 18th Australia conference on computer-human interaction: design: activities, artefacts and environments*, pages 111–118, 2006.

[143] Sascha Spors, Hagen Wierstorf, Alexander Raake, Frank Melchior, Matthias Frank, and Franz Zotter. Spatial sound with loudspeakers and its perception: A review of the current state. *Proceedings of the IEEE*, 101(9):1920–1938, 2013.

[144] Heng Zou, Hermann J Müller, and Zhuanghua Shi. Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision*, 12(5):2–2, 2012.

[145] JJ Scott and Robert Gray. A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. *Human factors*, 50(2):264–275, 2008.

[146] Hiroki Uchida, Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, and Keiichi Zempo. Spatialphonic360: Acoustic space for arbitrary sound image presentation

based on both ears tracking. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pages 123–125, 2022.

[147] Alexander Burka, Alaric Qin, and Daniel D Lee. An application of parametric speaker technology to bus-pedestrian collision warning. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 948–953. IEEE, 2014.

[148] Yuichi Mashiba, Ryunosuke Iwaoka, Hisham E Bilal Salih, Masayuki Kawamoto, Naoto Wakatsuki, Koichi Mizutani, and Keiichi Zempo. Spot-presentation of stereophonic earcons to assist navigation for the visually impaired. *Multimodal Technologies and Interaction*, 4(3):42, 2020.

[149] Kevin Kim, Saea Kim, and Anthony Choi. Ultrasonic sound guide system with eyeglass device for the visually impaired. *Sensors*, 22(8):3077, 2022.

[150] Larisa Dunai, Guillermo Peris-Fajarnés, E Lluna, and B Defez. Sensory navigation device for blind people. *The Journal of Navigation*, 66(3):349–362, 2013.

[151] Killy. Sound effect lab. `https://soundeffect-lab.info`, onlyavailable in Japanese.

[152] Winfried Menninghaus, Valentin Wagner, Eugen Wassiliwizky, Ines Schindler, Julian Hanich, Thomas Jacobsen, and Stefan Koelsch. What are aesthetic emotions? *Psychological review*, 126(2):171, 2019.

[153] Xu Zhang, Meihui Ba, Jian Kang, and Qi Meng. Effect of soundscape dimensions on acoustic comfort in urban open public spaces. *Applied acoustics*, 133:73–81, 2018.

[154] Mats E Nilsson and Birgitta Berglund. Soundscape quality in suburban green areas and city parks. *Acta Acustica united with Acustica*, 92(6):903–911, 2006.

[155] Bert De Coensel, Sofie Vanwetswinkel, and Dick Botteldooren. Effects of natural sounds on the perception of road traffic noise. *The journal of the acoustical society of America*, 129(4):EL148–EL153, 2011.

[156] Yiying Hao, Jian Kang, and Heinrich Wörtche. Assessment of the masking effects of birdsong on the road traffic noise environment. *The journal of the Acoustical Society of America*, 140(2):978–987, 2016.

[157] Wonyoung Yang and Hyeun Jun Moon. Effects of indoor water sounds on intrusive noise perception and speech recognition in rooms. *Building Services Engineering Research and Technology*, 39(6):637–651, 2018.

[158] Killy. Sound-dictionary. `https://sounddictionary.info`, only available in Japanese.

[159] Sound jay. `https://www.soundjay.com,`.

[160] Jan Stenkamp, Benjamin Karic, Paula Scharf, Judith A Verstegen, and Christian Kray. Using an immersive video environment to assess pedestrians' compliance with covid distance keeping interventions. *Interacting with Computers*, page iwad021, 2023.

[161] Robert S Kennedy, Julie Drexler, and Robert C Kennedy. Research in visually induced motion sickness. *Applied ergonomics*, 41(4):494–503, 2010.

[162] Durand R Begault and Leonard J Trejo. 3-d sound for virtual reality and multimedia. Technical report, 2000.

[163] Frederic L Wightman and Doris J Kistler. Headphone simulation of free-field listening. ii: Psychophysical validation. *The Journal of the Acoustical Society of America*, 85(2):868–878, 1989.

[164] Masayuki Morimoto and Katsuhiko Nomachi. Binaural disparity cues in median-plane localization. *Journal of the Acoustical Society of Japan (E)*, 3(2):99–103, 1982.

[165] Kanji Watanabe, Yukio Iwaya, Yôiti Suzuki, Shouichi Takane, and Sojun Sato. Dataset of head-related transfer functions measured with a circular loudspeaker array. *Acoustical science and technology*, 35(3):159–165, 2014.

[166] BS Series. Algorithms to measure audio programme loudness and true-peak audio level. 2011.

[167] Michael Ray Hill. *Spatial structure and decision-making aspects of pedestrian route selection through an urban environment.* The University of Nebraska-Lincoln, 1982.

[168] Charles S Watson. Some comments on informational masking. *Acta Acustica united with Acustica*, 91(3):502–512, 2005.

[169] Bernard Widrow, John R Glover, John M McCool, John Kaunitz, Charles S Williams, Robert H Hearn, James R Zeidler, JR Eugene Dong, and Robert C Goodlin. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE*, 63(12):1692–1716, 1975.

[170] Xun Wang, Yousuke Koba, Satoshi Ishikawa, and Shinya Kijimoto. Hybrid actice noise barrier with sound masking in open-plan offices. *Noise Control Engineering Journal*, 64(3):403–415, 2016.

[171] Riya Elizabeth Abraham and M Robert Kennedy. Intelligent digital signage system based on gender identification. In *Intelligent Embedded Systems: Select Proceedings of ICNETS2, Volume II*, pages 251–262. Springer, 2018.

[172] Simge Özdal Oktay, Sergio Trilles Oliver, Albert Acedo, Fernando Benitez-Paez, Shivam Gupta, and Christian Kray. Openness: A key factor for smart cities. In *Handbook of smart cities*, pages 1611–1642. Springer, 2021.

[173] Sakurazaka 46 (Performance) Yasushi Akimoto (Lyrics). Mamoribito. https://youtu.be/I8JZbNTl5rM (Lyrics in English are available in the subtitles of the music video on YouTube), 2023.

[174] EN ISO. 3382-2, 2008,"acoustics—measurement of room acoustic parameters—part 2: Reverberation time in ordinary rooms,". *International Organization for Standardization, Brussels, Belgium*, 2008.

Publication List

Publications with numbers in bold are relevant to this dissertation.

Journal with peer review

[J-3] Noko Kuratomo, Haruna Miyakawa, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, and Keiichi Zempo : "Attracting Effect of Pinpoint Auditory Glimpse on Digital Signage", IEEE Access,
https://doi.org/10.1109/ACCESS.2023.3270909 (open access)

[J-2] Noko Kuratomo, Haruna Miyakawa, Soh Masuko, Toshimasa Yamanaka, and Keiichi Zempo : "Effects of acoustic comfort and advertisement recallability on digital signage with on-demand pinpoint audio system", Applied Acoustics, Volume 184, 15 December 2021, 108359,
https://doi.org/10.1016/j.apacoust.2021.108359 (open access)

[J-1] Haruna Miyakawa, Noko Kuratomo, Hisham E. Bilal Salih, and Keiichi Zempo. "Auditory Uta-Karuta: Development and Evaluation of an Accessible Card Game System Using Audible Cards for the Visually Impaired." Electronics 10, no. 6 (2021): 750.
https://doi.org/10.3390/electronics10060750

International Conference with peer review

[I-11] Noko Kuratomo, Benjamin Karic, and Christian Kray: Design and Effect of Guiding Sound for Pedestrians While Maintaining the Streetscape Perception. In The 2023 ACM Symposium on Spatial User Interaction (SUI '23), October 13–15, 2023, Sydney, NSW, Australia. 10 pages.
https://doi.org/10.1145/3607822.3614514

[I-10] Moeki Horii, Azusa Yamazaki, Noko Kuratomo, and Keiichi Zempo: Effect of Spatial and Temporal Dilation of a Brand Logo Printed on a VR Shopping Bag, International Conference on Artificial Reality and Telexistence Eurographics Symposium on Virtual Environments (2022).
https://diglib.eg.org/xmlui/bitstream/handle/10.2312/egve20221297/021-022.pdf

[I-9] Noko Kuratomo, Hiroki Uchida, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, and Keiichi Zempo: Spatialphonic360: Accuracy of the Arbitrary Sound Image Presentation Using Surrounding Parametric Speakers, ISS 2022, Nobember 2022.
https://doi.org/10.1145/3532104.3571462

[I-8] Hiroki Uchida, Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, and Keiichi Zempo: Spatialphonic360: Acoustic Space for Arbitrary Sound Image Presentation Based on Both Ears Tracking, Ubicomp/ISWC 2022 Demos, September 2022.
https://doi.org/10.1145/3544793.3560319

[I-7] Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani, and Keiichi Zempo: Differences in Parametric-Speaker Emitted Speech Recognition Thresholds Caused by Sound Level Variations and Voice Gender. Proceedings of the 27th International Congress on Sound and Vibration, ICSV 28, Singapore, July 2022

[I-6] Ryo Kashiwabara, Ryota Sakuma, Keisuke Mizuno, Noko Kuratomo, and Keiichi Zempo: Directional Resolution of Haptic Display Consisted of a Pair of Vibrators Embedded Earrings, 2021 IEEE 10th GCCE, Kyoto Japan, October 2021.
https://doi.org/10.1109/GCCE53005.2021.9621878

[I-5] Noko Kuratomo, Tadashi Ebihara, Naoto Wakatsuki, Koichi Mizutani and Keiichi Zempo: How much is the noise level be reduced? -Speech Recognition Threshold in Noise Environments Using a Parametric Speaker-, INTERACT 2021, IFIP, Bari Italy, September, 2021.
https://doi.org/10.1007/978-3-030-85616-8_31

[I-4] Noko Kuratomo, Kazuki Yamada, Soh Masuko, Toshimasa Yamanaka and Keiichi Zempo: Effect of Attention Adaptive Personal Audio Deliverable System on Digital Signage, SIGGRAPH Asia 2019, ACM SIGGRAPH, Brisbane Australia, November 19, 2019.
https://doi.org/10.1145/3355088.3365171

[I-3] Noko Kuratomo, Yuichi Mashiba, Keiichi Zempo, Koichi Mizutani and Naoto Wakatsuki: Integrating a Binaural Beat into the Soundscape for the Alleviation of Feelings, INTERACT 2019, IFIP, Paphos Cyprus, September 5, 2019.
https://doi.org/10.1007/978-3-030-29390-1_13

[I-2] Haruna Miyakawa, Noko Kuratomo, Hisham Elser Bilal Salih and Keiichi Zempo: Auditory Uta-KARUTA: Sonificated Card Game towards Inclusive Design, UIST 2019 Posters, p90-92, New Orleans LA USA, October 22,2019.
https://doi.org/10.1145/3332167.3357125

[I-1] Keiichi Zempo, Yuichi Mashiba, Takayuki Kawamura, Noko Kuratomo, and Hisham Elser Bilal Salih. "Phonoscape: Auralization of Photographs using Stereophonic Auditory Icons." In The 31st Annual ACM Symposium on User Interface Software

and Technology Adjunct Proceedings, pp. 117-119. 2018.
https://doi.org/10.1145/3266037.3266120

Domestic Conference (in Japanese)

[D-4] Noko Kuratomo, Haruna Miyakawa, Ryoma Suizu, Soh Masuko, Toshimasa Yamanaka and Keiichi Zempo: Development of Ubiquitous Environment Consist of On-Demand Pinpoint Audio System for Persons with Visually Impaired, Human Communication Group Symposium 2020, Online, December 2020

[D-3] Haruna Miyakawa, Noko Kuratomo, Hisham Elser Bilal Salih and Keiichi Zempo: Design and evaluation of sound icons to realize an audible card game for the people with visually impairment, Human Communication Group Symposium 2019, Hiroshima, December 2019

[D-2] Noko Kuratomo and SeungHee Lee: Basic Study on Monitor Elderly People on Third Place, Human Communication Group Symposium 2019, Hiroshima, December 2019

[D-1] Noko Kuratomo, Keiichi Zempo, Koichi Mizutaniand Naoto Wakatsuki: Stereophonic Auditory Display of Several Time Series Data, Human Communication Group Symposium 2018, Mie, December 2018

Awards

[A-3] Outstanding Student Paper Award, IEEE GCCE 2021, October 2021

[A-2] Excellent Master's Thesis Award, Master's Program in Intelligent Interaction Technologies, University of Tsukuba, March 2021

[A-1] Dean's Prize, School of Science and Engineering, University of Tsukuba, March 2019

Grant

[G-3] Noko Kuratomo: "JSPS Research Fellowships (DC1)", Japan Society for the Promotion of Science, Japan (04.2021 - 03.2024, JPY 2,200,000)

[G-2] Noko Kuratomo: "Overseas Challenge Program for Young Researchers," Japan Society for the Promotion of Science, Japan (09.2022 - 03.2023, JPY 1,400,000)

[G-1] Noko Kuratomo, Kazuki Yamada, Haruna Miyawaka, Ryohei Nishida and Keiichi Zempo : "International Unicorn Development Program," University of Tsukuba, (JPY 1,500,000)

Others - Patent

[P-1] Keiichi Zempo (50%), Noko Kuratomo (50%). Information presentation device, information presentation system and information presentation method. Patent Application No. 2022-126439, filed Aug. 8, 2022. Patent pending.

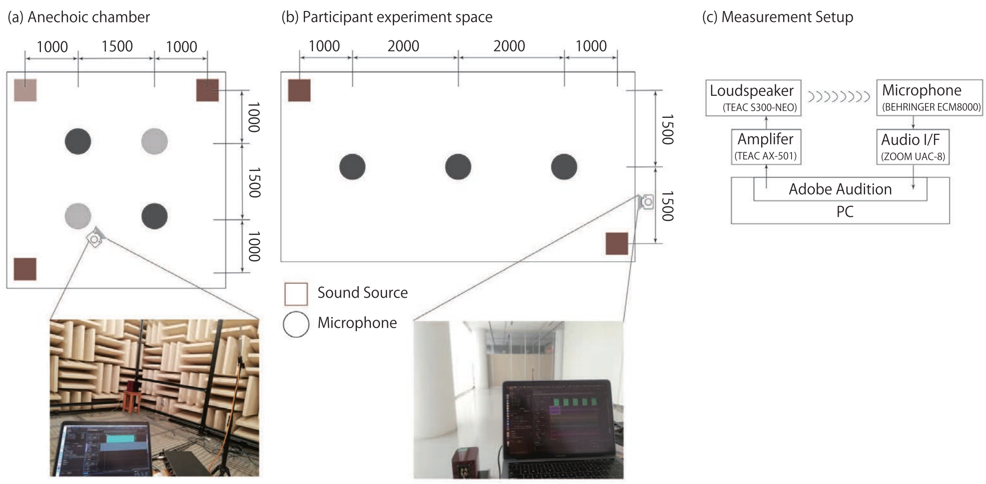

Figure 1: Procedure of Reverberation Time Measurement: (a) The anechoic chamber, (b) The participant experimental space, (c) Measurement setup of reverberation time

Appendix 1

The microphone used in chapter 3.3.4 is BEHRINGER ECM8000. A detailed data sheet of this microphone [106] shows that the directional characteristics differed within ±2.5 dB from 125 to 4k Hz. Also, the relative response differed within ±1 dB from 125 to 4k Hz. Thus, this microphone can record sound in all directions with a flat frequency response in this research. Before the experiment, I confirmed there were no lacks in the frequency band obtained from this microphone. Moreover, see [107] for a performance comparison with Class 1 microphones.

Appendix 2

This appendix describes the methods of measuring the reverberation time used in chapter 3.3.4. The reverberation time was measured according to the engineering method of ISO3382-2 [174]. Specifically, as shown in Figure 1 (a), in the anechoic chamber, the number of sound source positions was three and there were two microphone positions. Two setups were used: in brown, there was 1 source position and 2 microphones; in black, there were 2 source positions and 2 microphones, for a total of 6 combinations. In the room for the participants experiment, there is two source positions and three microphone locations, for a total of six combinations, as shown in Figure 1 (b). The setup is shown in Figure 1 (c). The number of repeated measurements per measurement point was five, and the one time when the measurement was clearly made was used for the calculation. Chirp signal was used as the sound source. The reverberation time was calculated from the impulse response using the integrated impulse method and averaged over six locations.