Title	A Study on Design Methodologies for Intuitive In-Vehicle Auditory Signals
Author(s)	神薗, 善規
Citation	大阪大学, 2025, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/101757
rights	
Note	

A Study on Design Methodologies for Intuitive In-Vehicle Auditory Signals

Submitted to
Graduate School of Information Science and Technology
Osaka University

January 2025

Yoshinori KAMIZONO

Publications

Journal Article (Refereed)

[J1] Yoshinori Kamizono, Wataru Kobayashi, Ittetsu Taniguchi, Hiroki Nishikawa, and Takao Onoye, “A Platform for Evaluation of Synthetic Reflected Sounds on 3D Sound Localization,” *IEEE Access*, vol. 12, pp. 86906-86916, 2024.

[J2] Fuma Sawa, Yoshinori Kamizono, Wataru Kobayashi, Ittetsu Taniguchi, Hiroki Nishikawa, and Takao Onoye, “An In-Vehicle Auditory Signal Evaluation Platform based on A Driving Simulator,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 106, no. 11, pp. 1368-1375, 2023.

International Conference Papers (Refereed)

[I1] Yoshinori Kamizono, Wataru Kobayashi, Hiroki Nishikawa, Ittetsu Taniguchi, and Takao Onoye, “In-Vehicle Auditory Signals for Driver Awareness through Danger Perception”, in *Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA)*, pp. 1-6, 2024.

[I2] Yoshinori Kamizono, Wataru Kobayashi, Hiroki Nishikawa, Ittetsu Taniguchi, and Takao Onoye, “Frequency and Timing Shifting of 3D-localized Advisory Signals for Awareness of Pedestrians”, in *Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA)*, pp. 27-32, 2024. (SISA Excellent Student Paper Award)

[I3] Fuma Sawa, Yoshinori Kamizono, Wataru Kobayashi, Ittetsu Taniguchi, Hiroki Nishikawa, and Takao Onoye, “Live Demonstration: In-Vehicle Auditory Signal Evaluation Platform in A Driving Simulator,” in *Proceedings of International Symposium on Circuits and Systems (ISCAS)*, pp. 1-1, 2023.

[I4] Yoshinori Kamizono, Wataru Kobayashi, Ittetsu Taniguchi, Hiroki Nishikawa, and Takao Onoye, “The 3D Sound Localization Platform with Reflected Sounds,” in *Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA)*, pp. 11-14, 2022. (SISA Best Student Paper Award)

[I5] Fuma Sawa, Yoshinori Kamizono, Wataru Kobayashi, Ittetsu Taniguchi, Hiroki Nishikawa, and Takao Onoye, "Development of In-Vehicle Auditory Signal Evaluation Platform in A Driving Simulator," in *Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA)*, pp. 37-40, 2022.

[I6] Mitsuhiro Watanabe, Yoshinori Kamizono, Fuma Sawa, Wataru Kobayashi, Ittetsu Taniguchi, and Takao Onoye, "Development of 3D Sound Localization Platform for Online Communications," in *Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA)*, pp. 41-44, 2022.

[I7] Yoshinori Kamizono, Takao Onoye, Wataru Kobayashi, Hidetaka Shibata, and Daisuke Okamoto, "Evaluation of Auditory Signals in an Automobile for Safe Driving," in *Proceedings of International Symposium on Multimedia and Communication Technology (ISMAC)*, pp. 1-5, 2019.

Summary

Driving safety is an extremely crucial aspect and the pressing business to reduce traffic fatalities has been increasingly highlighted in recent years, as indicated by the World Health Organization (WHO). This global urgency emphasizes the necessity of solution technologies to realize safer and more comfortable worldwide driving environments. Several advanced technologies, such as Advanced Driver-Assistance Systems (ADAS) and intelligent transportation systems (ITS), assist in safe driving by detecting potential risks and informing the drivers in advance coordinating between people, vehicles, and infrastructure. An autonomous driving system has become a reality in recent years through these technologies by which various functions have been devised to realize safe and comfortable driving. These latest technologies have been diversifying the information conveyed to drivers. Therefore, it is necessary to consider optimal information presentation methods, which takes into account the driver's driving situation.

Auditory signals have been taken the advantage of informing users of changes in a variety of events by hearing cues. The auditory signals have been also utilized in the automobile classically called in-vehicle auditory signals. However, the use of simple beep sounds has remained prevalent for a long term without much innovation. Especially, the in-vehicle auditory signals need to be designed diversely to promote an intuitive understanding of the content because the increment of electric systems in automobiles ever-increase the information to convey to drivers. Fundamental innovations based on various sound elements and human perceptions are required to achieve an in-vehicle auditory signal that communicates information in intuitive and natural manners.

This dissertation aims to realize a universal design of ideal in-vehicle auditory signals to present information according to danger levels intuitively with various sound elements and proposes methodologies that will serve as a foundation for such a universal design. The research subjects focus on methods of realizing directional perceptions, tone design based on the danger level of surroundings, and supporting multiple events in the special acoustic space of driving. Therefore, this dissertation addresses the following points to maximize the potential of intuitive in-vehicle auditory signals.

- How to communicate the direction of dangers by in-vehicle auditory signals
- How to communicate information appropriately according to the level of danger by in-vehicle auditory signals
- How to communicate multiple simultaneous events by in-vehicle auditory signals

The in-vehicle auditory signals ought to install directional perception to be intuitively aware of the information outside the conscious mind of drivers. Sound reflections exist inside the car due to several special shapes and materials, which seem to influence the 3D sound perception of drivers. However, there were no reports of platforms easily evaluating the effects of sound reflection on 3D sound perception in detail. This dissertation presents a platform designed to evaluate the impact of synthetic reflection on 3D sound perception. This proposed platform utilizes Head-Related Transfer Functions (HRTFs) to implement sound reflections by imaginary sound sources as well as direct arriving sounds from sound sources. This implementation method allows users to assess the impact of sound reflections on the directional perceptions of sound sources. The experiment is conducted to verify the usefulness of the proposed platform with an evaluation of differences in the sense of forward and backward sound sources with and without sound reflections. The results show the influences of sound reflections varied according to the direction of sound reflections and the types of sound sources, which verifies that the proposed platform is expected to reveal how different reflected sounds including reflections in vehicles affect the perception of sound direction in the future.

The intuitive in-vehicle auditory signals need to provide quick and correct communication with the hazardous information to drivers since the information to be communicated has diversified from low to high risks. On the other hand, the traditional design of in-vehicle auditory signals with a simple structure as monotonous beeps could cause misperception and discomfort for drivers. This dissertation explores designs for in-vehicle auditory signals to enhance driver intuitive awareness via danger perception. This research classifies in-vehicle auditory signals into three danger levels: warning, caution, and advisory, and then redesigns several samples of in-vehicle auditory signals based on musical and acoustic elements for intuitive communication with a driver, especially caution signals for dangerous situations and advisory signals for potential dangers with comfortable sounds. The redesigned in-vehicle auditory signals are evaluated by questionnaire and principal component analysis to demonstrate that the redesigned in-vehicle auditory signals improve intuitive awareness and to summarize the outline settings of sound elements as a guideline example for global design in designing in-vehicle auditory signals according to hazard levels.

In an ever-changing driving environment, different levels of hazard occur simultaneously. In-vehicle auditory signals are required to convey multiple risks to drivers according to the level of hazard in the driving environment. This dissertation proposes an in-vehicle auditory signal to distinguish among multiple potential collision risks due to focusing on pedestrians. Conventionally, the methods including 3D sound localization and tone designs for various hazards allow drivers to distinguish among multiple sounds easily and recognize risks from various angles simultaneously. However, if several sounds reach drivers to convey multiple risks from almost the same direction, drivers cannot judge multiple risks, such as pedestrians. The proposed methods tackle this problem by shifting the frequency and onset timing of the in-vehicle auditory signals enabling drivers to differentiate among multiple pedestrians in the blind spot of drivers based on Time-To-Collision (TTC) — the time remaining before a potential collision such as a pedestrian. The experiment evaluates the ability to decompose multiple signals with and without proposed methods. The results of experiments demonstrate a notable improvement in the proposed methods compared to conventional in-vehicle auditory signals, primarily due to the onset timing adjustments rather than frequency changes.

Acknowledgments

First of all, I would like to express my great appreciation and special thanks to my supervisor Associate Professor Ittetsu Taniguchi, Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University. He was a model of excellence in supervision, guiding and supporting his subordinates in a deeply formative way. In my case in particular, he became my supervisor after I entered the Ph.D. courses despite the difficulties of supervising my research from the middle of the program. Furthermore, he never gave up on me and supported me to the end even though my research skills were very limited and my mental condition was quite poor. Without his generous support, I would not have become the person I am today, nor would I have lived the life I now cherish. His insightful guidance and unwavering support were instrumental in helping me achieve advanced results in my research.

I would like to extend a special thanks to my supervisor Professor Takao Onoye from the Department of Information Systems Engineering at the Graduate School of Information Science and Technology, Osaka University. He gave me an opportunity to research sound perception which I was eager to study whereas it was not studied in our laboratory at the beginning, and he guided me to obtain my master's degree, which was the last time he could continue as my supervisor for moving to his next position. I am greatly indebted to him for insightful feedback and valuable advice throughout this research.

My sincere thanks are extended to Professor Yuki Uranishi from the Cybermedia Education Division, D3 Center, Osaka University. His in-depth reviews and recommendations greatly influenced the direction of my dissertation and led me on an improved and innovative path.

I would like to appreciate Fumio Kishino, Visiting Professor at Kwansei Gakuin Institute of Kansei Value Creation at Kwansei Gakuin University and Emeritus Professor at Osaka University, for detailed reviews and insightful suggestions that have driven my dissertation in a new and better direction.

My profound thanks go to Dr. Wataru Kobayashi who helped me with his invaluable ideas and pieces of advice from the beginning of my research until the time he passed away this year. He has always supported me with his rich ideas on sound research that no one else could come up with. Clearly, this dissertation could not have existed without him. In addition, he really assisted my life based on his rich experiences as a person who has lived a much longer life than me. I sincerely hope he rests in peace in the afterlife as compensation for his great work on this ground.

I express my sincere gratitude to Assistant Professor Hiroki Nishikawa, Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University, for his precious suggestions and his enormous support. I have always learned a lot from his insightful views on doing research and his constructive comments. I would not have been able to obtain my research achievements without his support, especially in paper writing and presentation of my research.

I extend my gratitude to Specially Appointed Assistant Professor Dafang Zhao, Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University. The support was invaluable I received from him with regular research to managing the lab operations.

I would like to thank Professor Yuichi Itoh, Department of Integrated Information Technology, Aoyama Gakuin University, for his helpful advice and valuable comments on my research.

I am grateful to Daisuke Okamoto, Representative of Dai1 Studio, for supporting the composition of several samples of in-vehicle auditory signals used in experiments for this dissertation.

I would like to acknowledge members of the Shosisha Foundation for their assistance with a scholarship and organizing social events. Special thanks are due to Mr. Yasuhiko Yamanaka, a board chairman in the Shosisha Foundation, for his invaluable input. Appreciation is also due to the chief of the secretariat of the Shosisha Foundation, Mr. Kohei Notoya, Ms. Satomi Takeuchi, and Ms. Yoko Inoue for their extensive support of my research life.

I would like to thank all members of the Onoye Laboratory for sharing a rich and vivid time. My appreciation goes to Mrs. Yuki Yoshida, the secretary of Onoye Laboratory, whose support was crucial in my lab life, allowing me to concentrate profoundly on my research. I am very grateful to Dr. Daichi Watari and Dr. Kenshiro Kato, alumnus of the Onoye Laboratory as well as former excellent Ph.D. colleagues. I would like to show my appreciation to them for their support and inspiration, both as friends and comrades, since we pushed forward together on our paths as Ph.D. students. Moreover, their exemplum as researchers have been an important influence in shaping who I am today. I am also very grateful to Mr. Sawa Fuma, a Ph.D. student of the Onoye Laboratory and a direct junior student of mine. I would like to thank his support for my research by creating a warm atmosphere with his wonderful personality in the life of the laboratory even though he was my junior. Additionally, I am grateful to the members of Group IoT and EA, Mr. Luyang Liu, Mr. Naoya Oda, Ms. Yuri Yano, Mr. Ryuhei Michizoe, Mr. Keita Sasaki, Ms. Kyoko Takii, and Mr. Shota Kajiya. Their feedback at meetings and assistance as experimental subjects are precious, and the pleasant times I have shared with them are a constant source of motivation to continue my research. I would also like to express my appreciation to my alumnus colleagues from the laboratory, Mr. Yoshitaka Ishihara, Mr. Masahiro Itoh, Mr. Shori Ueda, and Mr. Kento Sakurauchi. Their support and banter during our time together have been unforgettable and valuable.

To all my excellent friends, thank you for giving me exceedingly time of fun through warm friendship. I am firstly grateful to my dear childhood friend, Mr. Toshiki Hamasaki, who has been such a tremendous blessing in each other's lives. He has been such a dear friend that he is like a part of my family, and I am especially grateful that we have always been able to support each other through difficult times, crying and laughing together. My friendship with Mr. Masaki Yokoyama, the only male member in the same grades as myself in the Japanese Traditional Instrument Club at Osaka University, is one of the most precious things I have gained at the university. I want also to thank all the other members of the club for supporting and entertaining me, even though I was an imperfect head of the club.

I am deeply and sincerely grateful to all the brothers in the church I have been attending, who have been sustaining me with their Godly mercy and abounding grace. In particular, I would like to express my deepest gratitude to Pastor Joshua Jeong and Pastor Kenji Miyauchi for leading me to make my life full of blessings and love with the Lord, no matter what pain I am going through. I would also like to give special thanks to my brothers, Mr. Kenya Kumagai and Mr. Keijiro Takeda, who are friends and brothers of mine as well. They are irreplaceable to me, able to talk about the blessings they had received from heaven, challenging each other to live a life of treating themselves and their neighbors with dignity, without forgetting God's love and mercy no matter how hard we struggled.

My profound appreciation is directed towards my family: Mr. Fumiro Kamizono, Mrs. Atsuko Kamizono, Ms. Kaoru Kamizono, and Mrs. Emi Yoshinaga. My father Mr. Fumiro, whom I respect and admire, supported me financially and was a role model for me as a guy to live a strong life. My mother Mrs. Atsuko, who gave birth to me with an aching stomach, gave me strength even after her death, reminding me of the love I had in life. My two sisters have been like a surrogate for my mother and have been very supportive. I cannot thank them enough for how much they mean to me. The unwavering love of my family enabled me to complete this excruciatingly long way.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The advent of automobiles has profoundly revolutionized human society by enriching lives in countless ways. The ability to travel to previously unreachable places within a short period all while being safely enclosed in a sturdy and attractive exterior has transformed mobility domains. Driving is no longer merely a means of transportation and has become a source of leisure racing through natural landscapes or participating in motorsports. However, alongside these advancements lies the grim reality of countless lives lost in automobile accidents. Humanity has continually strived to combat this major challenge and produced numerous safety technologies. To hold driving pleasure while mitigating the risks of accidents, innovative driver assistance systems have been brought to life that engage various human senses. Among these techniques, auditory cues play a particularly important role in improving driving safety and awareness, known as *in-vehicle auditory signals*. Various research including this dissertation aims to alert drivers to surrounding dangers without disrupting their visual focus on driving operation by conveying critical information through auditory cues while simultaneously enhancing the sense of immersion that is feasible for driving pleasure. Consequently, this chapter introduces the background of a wide range of automotive safety technologies and then discusses in-vehicle auditory signals with their challenges and ideal versions. Above this description, the research scope of this dissertation is stated as research questions.

1.1 Background

Safety has always been one of the top priorities in the development of automotive technology. In fact, Fig. 1.1, in which horizontal and vertical axes indicate years and the number of fatalities from vehicle accidents respectively, shows more than a million people have been dead in traffic accidents each year [1, 6]. The World Health Organization (WHO) also reported that the global goal to halve road traffic deaths by the Decade of Action for Road Safety 2011–2020 was unfortunately not archived at the end of 2021,

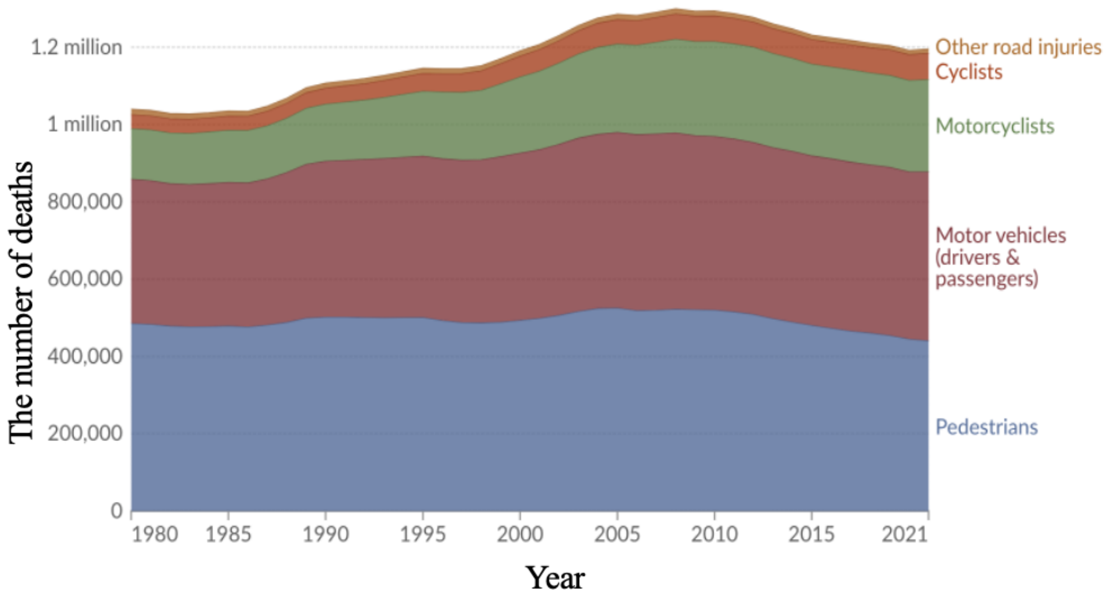


Figure 1.1: Annual deaths from road incidents in the world [1].

whereas some few countries accomplished the goal with results of great efforts [7]. On the contrary, death rates from traffic accidents increased in many low-income countries over the last decade such as African regions [8]. Moreover, developments in automated driving technology underscore the importance of safety these days [9]. For these reasons, further attempts are required across various disciplines to decrease death rates from traffic accidents and to realize safer and more comfortable worldwide driving environments.

To overcome these challenges, several technologies have been developed to generalize safe driving in recent days, called Advanced Driver-Assistance Systems (ADAS) [10–12]. These advanced technologies are prevailing systems that support safe driving by detecting potential risks and informing the drivers beforehand. The field of ADAS has developed significantly into more complex assistive functions that are applied to a wider range of applications. It has matured with a significant increase in its user base due to market penetration. Even personalization methods of ADAS have been proposed to achieve optimal driving experiments [13]. Such systems can work in conjunction with the surrounding infrastructure and feedback of driver's operations to achieve even more safety features to reduce the burden on the driver, as illustrated in Fig. 1.2 which represents an example of vehicle systems, infrastructure, and feedback of driver's operations in coordination with each other to present information to the driver by a display. For instance, the information detected by sensors or other means in the surrounding infrastructure can provide positions of vehicles and pedestrians that are difficult for drivers to see. This coordination between people, vehicles, and infrastructure are collectively known as intelligent transportation systems (ITS) [14], and it constantly utilizes state-of-the-art communicating information and control technologies to optimize road traffic, while at the same time eliminating accidents and traffic congestion, saving energy, and coexisting with the environment [15]. The destination of such a story with these driving assistant systems is an autonomous driving system [16, 17].

Autonomous transport of people by vehicles was a dream in the past, but it has become a reality in recent years. There are six levels of automated driving ranging from 0 (no automation) to 5 (full automation) according to the Society of Automotive Engineers (SAE) [5], and various studies are underway to realize each level depending on situations and technologies. Table 1.1 categorizes automation levels with a main operator based on the description of the driver and system's role. These roles can be divided into three categories: (1) primary tasks involve direct vehicle control, such as steering and braking; (2) secondary tasks are related to maintaining essential safety features, like using turn signals and windshield wipers; and (3) tertiary tasks encompass other functions that contribute to comfort [18]. The primary responsibility for vehicle operation remains with the human driver up to level 2, the vehicle itself takes on the main operational tasks from level 3 onward. The driving assistant systems including ADAS apply to levels 1 to 2 in which the main operation remains with the human driver. In level 3, passengers need to intervene in driving operations when any troubles occur suddenly, which is called a Take-Over Request (TOR). As inadequate TOR indication can cause accidents due to delayed transitions, many researchers have been exploring appropriate methods for TOR [19, 20]. At levels 4 and 5, occupants are expected to have a basic understanding of the autonomous driving system rather than the skills required for manual driving operation. As a result, it is anticipated that traditional driver's licenses may no longer be necessary. On the other hand, fully automated driving has various challenges and it is also said that not all activities can be replaced by fully automated driving because of factors such as sensor and camera performance, regulatory and safety restrictions, the long-standing use of level 2 and lower technologies in the market, and driving itself becoming a hobby. Manual driving will not disappear anytime soon, and autonomous and manual driving will coexist on public roads or be restricted to designated areas. In these circumstances, the continued development of driver-assistance technologies at levels 1–2, such as ADAS, which primarily supports the human driver, is expected to remain crucial.

When considering driver assistance, it is essential to establish information presentation methods for drivers to improve their safety and comfort in driving. Currently, research is being conducted on various presentation methods, such as using visual, aural, tactile, and olfaction senses, with balance and preconception as well as combinations thereof [21]. Table 1.2 summarizes the types of human perceptions and examples of methods to present information to drivers. Visual information, for instance, is presented through displays mounted at the steering wheel [22] and dashboard [23]. Head-Up Dis-

Table 1.1: Levels of autonomous driving [5].

Automation level	Name	Description	Operator
0	No automation	The driver is responsible for all vehicle control aspects.	Driver
1	Driver assistance	The driver has full control of the vehicle, and the automated driving system provides functions such as lane-keeping assist.	Driver
2	Partial automation	The automated system can take control of vehicle functions including accelerating, braking, and steering. The driver is required to constantly monitor the possibility of intervention by the automated driving system.	Driver
3	Conditional automation	The automated system takes over the vehicle's functions. The driver's presence is required, however, no need to constantly monitor the automated driving system.	System
4	High automation	Automated driving systems can perform all driving functions under certain conditions. The driver's control of the vehicle will be optional.	System
5	Full automation	Automated driving systems are fully automated under all circumstances. The driver intervention is not required.	System

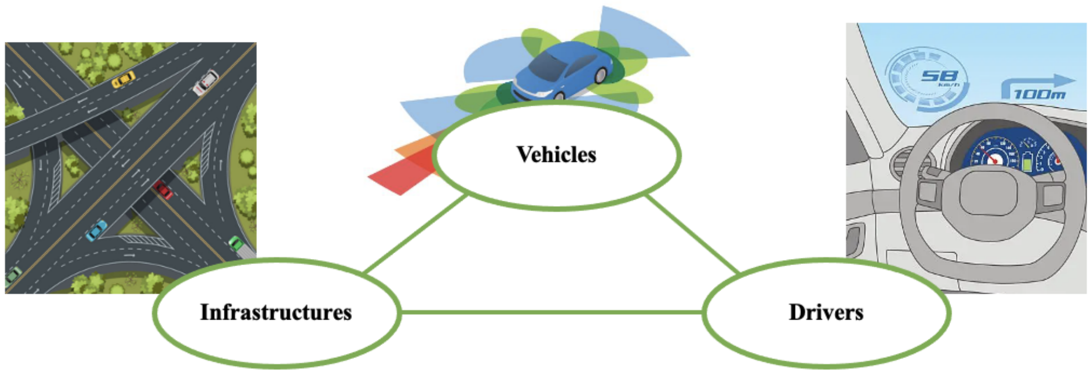


Figure 1.2: A schematic picture of driving assistant technologies.

Table 1.2: Presentation methods with human perceptions.

Types of perceptions	Examples of presentation methods
Visual	Head-Up Display (HUD), Head-Mounted Display (HMD)
Aural	Speech, Auditory icons, Earcon
Tactile	Vibration of steering wheels and driver's seats
Olfaction	Smell associated with presentation information

play (HUD) [24] and Head-Mounted Display (HMD) with Augmented Reality (AR) techniques [25] are used to communicate information efficiently to drivers. In the case of aural cues, short sounds are used with the composition of parts of existing sounds representing objects and arbitrary synthetic sounds, which are called *Auditory icons* [26] and *Earcon* [27]. The tactile methods reach out to drivers with such vibrations of steering wheels and driver's seats [28,29]. Some studies attempt to utilize olfaction to convey information [30].

Among various information presentation methods, visual or aural cues are predominant because other methods are not sufficiently clear about how much information can be transmitted reliably and properly perceived by the stimuli. For this aspect, visual and aural cues are often compared to which method is proper according to the driving situation [31, 32]. These studies mentioned that visual cues assist drivers more accurately, whereas aural cues support quicker notifications and responses. On the other hand, excessive trust in visual-based systems can overwhelm drivers by adding too many visual tasks as the drivers concentrate on the driving operation with their eyesight [33,34]. This can lead to driver fatigue, especially if they are required to monitor their surroundings continuously and follow visual cues simultaneously. In fact the work in [35] pointed out that visual information could lead to accidents by guiding the driver's eyesight to the display mounted in the cars. Moreover, novice drivers may be more susceptible to distractions from visual guidance, as they might focus more on the system's suggestions rather than the road itself. As a result, visual-based ADAS supports may not be ideal for drivers during driving. It is important to prioritize only essential monitoring tasks and propose alternative methods of assistance that reduce the reliance on visual input, particularly for less experienced drivers, to alleviate their cognitive load. From these points, aural cues are featured in driving situations as a safe and intuitive means of presenting information in driving. Additionally, there is an aural cue using brief speech messages. However, these methods are not critical for in-vehicle safety systems because the speech messages take a long time to concentrate on the contents until the end of the messages, which causes strong stress and no time to take action to avoid danger [36]. Therefore, this dissertation focuses on the aural cues of brief synthetic sounds such as *Auditory icons* and *Earcon* to communicate information in driving, which is called *in-vehicle auditory signals*.

1.2 In-Vehicle Auditory Signals

This section describes the background of in-vehicle auditory signals, which are aural cues featured as a safe and intuitive means of presenting information in driving situations, and proposes ideal in-vehicle auditory signals called *intuitive in-vehicle auditory signals* in this dissertation. First, challenges of in-vehicle auditory signals are introduced. Next, a flow of information transmission by auditory signals is introduced to consider ideal in-vehicle auditory signals. Then, the functions needed to achieve intuitive in-vehicle auditory signals are discussed.

1.2.1 Challenges for In-Vehicle Auditory Signals

Auditory signals have been taken advantage of informing users of changes in events by hearing cues in a variety of situations. For familiar instance, home appliances emit audible cues [37] when a control button is pressed, or provide an audible warning of an abnormality in the equipment. Then the auditory signals have been utilized in the automobile classically [38], called as in-vehicle auditory signals. However, the use of simple beep sounds has remained prevalent in consumer electronics and user interfaces for at least a decade without much innovation by 2016 according to the surveys of [39]. Their work introduced the use of melodies, voice prompts, and synthetic tones to provide users with more diverse auditory cues, intending to improve user awareness and create a richer auditory experience.

The in-vehicle auditory signals are recommended to be complex tones and less monotonous sounds [40]. Especially, the information to convey to drivers has been increasing gradually due to the increment of electric systems in automobiles, which is an ever-increasing trend according to the work in [2] as illustrated in Fig. 1.3 showing a pie chart of the increasing of electronic systems in the cost of a vehicle. For this

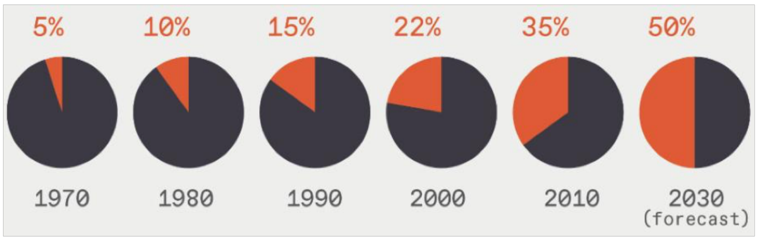


Figure 1.3: The trends of the percentage of electronic systems in the cost of automobiles [2].

factor, the in-vehicle auditory signals ought to be designed diversely from the simple beep sounds for an intuitive understanding of the content of information accurately and immediately. Moreover, in-vehicle auditory signals consisting of mainly beeps that are strongly focused on attracting the driver's attention can cause stress by the impression of great danger, even though the danger is of low level. This also creates the problem of turning off the function of the in-vehicle auditory signals itself by drivers. Thus, in-vehicle auditory signals require a fundamental re-innovation and a next-generation framework based on various sound elements, such as acoustic characteristics and music theory to cope with the context of driving scenarios in which the level of danger in the surroundings varies frequently [41]. However, composing sounds based on the perception of sound requires a great deal of experience and knowledge, as well as knowledge of automobiles. It is not easy to train such personnel on a site-by-site basis. Therefore, the above problems could be solved if there were some kind of guidelines for in-vehicle auditory signals that anyone could easily use to present information according to the degree of danger intuitively. In fact, several studies such as [42] have even emphasized the requirement for universal designs of in-vehicle auditory signals, which focused on the effectiveness of various sound elements.

1.2.2 A Flow of Information Transmission by Auditory Signals

Aural cues have been used since ancient times to understand information about human surroundings. To consider ideal in-vehicle auditory signals, a basic flow of information transmission by aural cues is introduced here. Fig. 1.4 shows a summary of the flow of information transmission by auditory signals, divided into three parts: sound environments, human perception, and behavior induced by sound. Based on each part, the factors that affect sound transmission are discussed.

Sounds generally travel from sound sources to human ears through various mediators such as air. During this transmission, these mediators influence the sound features by the time they reach human ears. If obstacles exist between the way of sound sources and human ears, diffraction [43] will occur by the obstacles, which may result in the

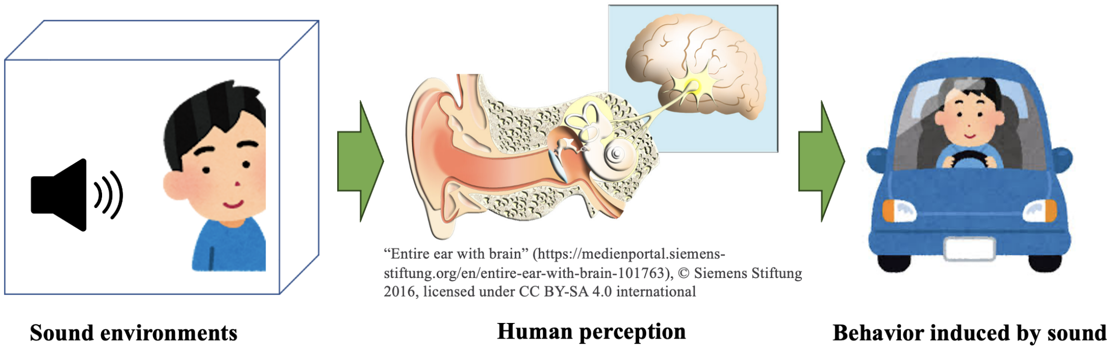


Figure 1.4: A flow of information transmission by auditory signals.

volume level of sound decreased and the sound waves changed by the phenomenon. In addition, as sounds do not only reach human ears directly, it is necessary to consider sound reflections and absorption caused by surrounding substances such as walls and ceilings [44]. As if the number of sound sources would increase by sound reflections, it is also needed to analyze the influence between directly reached sound and reflected sounds. Reflective mediums also could vary volume attenuation and change in sound shape with each reflection and absorption. Additionally, no sound exists alone in this real world, which can express that there are many sounds all around in other words. In this plenty-of-sound space, humans try to focus on the target sound source. However, sound identification might be difficult if the volume of background sounds is too large than the target sound, which is known as the masking phenomenon [45]. Thus, it is necessary to design auditory signals to take into account the surrounding sound, or noise, to deliver the desired sound to listeners. The ratio of the target sound and noise is also called the signal-to-noise (S/N) ratio, many researchers and sound designers have been working on improving the S/N ratio to situations [46, 47].

After the auditory signals reach human ears to overcome the factors in sound environments mentioned above, the sound is collected by the auricle as air vibrations and transmitted through the ear canal to the eardrum. The air vibrations cause the eardrum to vibrate, which is amplified by the ossicles of the ear, and the cochlea in the inner ear converts the vibrations into electrical signals that are then transmitted to the brain [48]. At this time, humans do not always accurately perceive the physical sounds occurring in the external world. Humans constantly create shortcuts to understanding information by adding interpretations to what they hear and aligning things based on past experiences and memories in addition to information from other sense organs [49]. This is a necessary cognitive mechanism for the brain to perceive the external world as quickly as possible. In ohter words, even if using a sound that is easy to hear as a physical phenomenon, it only means that it is easy to hear in the notice stage, and it does not necessarily promote human response or cognition to the sound. Therefore, auditory sig-

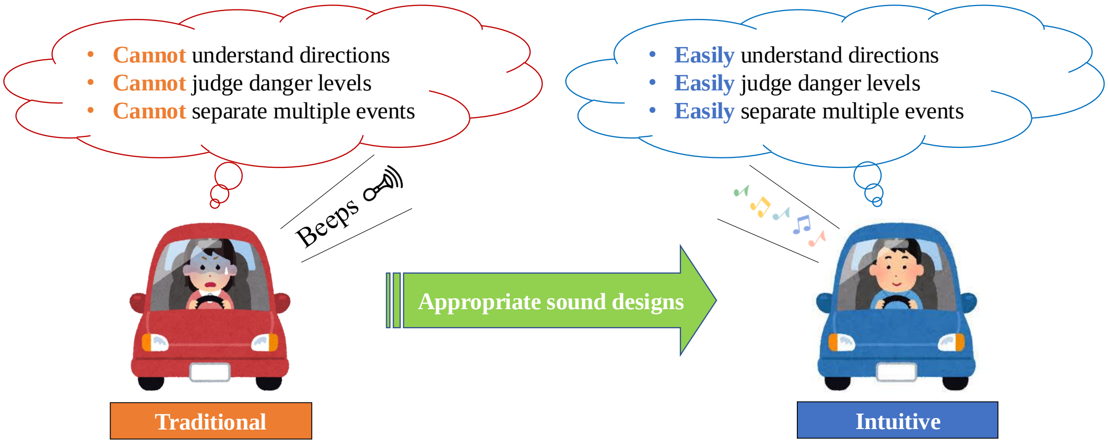


Figure 1.5: Redesigns from traditional in-vehicle auditory signals (e.g., beeps) to new-generational in-vehicle auditory signals with easy understanding of directions, danger levels of risk factors, and multiple events in driving surroundings.

nals should be carefully designed considering the differences between physical sound and human perception of sound, such as this illusion.

After the auditory signals reach human brains in which the signals are interpreted, humans act on that interpretation. It is necessary to design auditory signals with an awareness of how people behave and evaluate the ability of the function properly to achieve effective auditory signals. Therefore, it is important to design the auditory signal from retroactively the sound transmission discussed so far, starting from the actual behavior caused by sound and considering the interpretation of sound and the influence of the sound environment.

1.2.3 Intuitive In-Vehicle Auditory Signals

This dissertation aims to realize a universal design of ideal in-vehicle auditory signals to present information according to the degree of danger intuitively with various sound elements and proposes methodologies that will serve as a foundation for such a universal design. This ideal in-vehicle auditory signals called *intuitive in-vehicle auditory signals* (right figure) here are illustrated as an overall goal in Fig. 1.5 compared with traditional in-vehicle auditory signals (left figure). Traditional in-vehicle auditory signals (left figure) stereotypically consist of simple beeps, which are not effective in communicating necessary information to the driver and can even cause stress and fatigue through miscommunication while driving. These signals can result in misunderstandings, confusion, slower reaction times, and increasing the risk of accidents. Thus, intuitive in-vehicle auditory signals ought to be designed for easy understanding of surroundings, which is tailored to various driving scenarios [50, 51]. Also, these intuitive in-vehicle auditory signals ought to be evaluated repeatedly using a virtual environment in which safety is ensured [52, 53], and to realize the intended functions set according to driving scenarios. Then, intuitive in-vehicle auditory signals fulfill the natural presentation of information from vehicles to drivers through intuitive sounds immediately and accurately, not interfering with other activities while driving to be free of impressions appropriate to situations.

The main features required to achieve an intuitive in-vehicle auditory signal are as follows. If the direction of information can be determined from auditory signals, it is useful for guiding the driver's line of sight, which is an advantage over other sensory organs [36]. In addition, proper sound tonal designs for in-vehicle auditory signals based on acoustic elements and music theory may enhance driver awareness by facilitating the determination of which risk factors to priority focus on and may realize comfortable environments in driving [41]. Furthermore, it would be more helpful if the in-vehicle auditory signal could distinguish between multiple events such as those present in actual driving [54]. Considerations for achieving each of these functions are described below.

Directional Perception

Directional perception is effective for in-vehicle auditory signals because the aural sense immediately provides an approximate direction of sound sources. This is the advantage that other senses can not easily present. This is why some studies [55, 56] suggest that the in-vehicle auditory signals are designed to realize the directional perception as 3D sound depending on the purpose of promoting an understanding of where the incidents come from. On the other hand, it should be noted that this directional perception is an example of the difference between physical and human sound perception. As humans acquire directional perception as they grow from birth, various empirical rules and environmental information influence this directional perception [57]. Human ears are also dull concerning distances of sound sources in addition to sound directions, listeners tend to overestimate distances that are less than 1.5 m and underestimate distances that are greater than 1.5 m [58].

Background noises and reflected sounds also even fluctuate this positional perception. In driving situations, many sounds exist inside and outside of cars, including the noise of other vehicles, wind sounds, engine noises, speaking voices, and so on. Therefore, in-vehicle auditory signals are required to reach the ears of the driver to overcome these noises such as the work in [59]. Moreover, there are several special shapes and materials including window glass and cushions, which create unusual sound fields with attenuation and absorption when reflected. Designs of in-vehicle auditory signals are demanded to achieve the intended effect with consideration of acoustic features specialized inside a car [60,61]. The in-vehicle auditory signals are designed considering these perceptual differences to be mounted 3D sound effects.

At the time of implementation, acoustic devices could be equipped to realize directional perception for in-vehicle auditory signals while paying attention to acoustic features and background noises mentioned above. The points for choices of acoustic devices are position and performance, especially in automobiles. Available spaces are limited for acoustic devices in automobiles which is not so wide because other devices for driving systems occupy a lot of space. For this reason, it is necessary to carefully choose the equipped position of acoustic devices with the balance of distances from human ears and the range of occupation in automobiles, which greatly influences the positional perception of a sound source. The performance of acoustic devices is important needless to point out. Sound qualities ought to be guaranteed to convey the information with auditory signals. On the other hand, an acoustic device producing high-quality sounds tends to be bigger in size, and high-quality devices also might be expensive. The balance between performance, device size, and cost might be raised as a problem in the case of the implementation. The number of acoustic devices is also an important factor in the implementation. If more realistic sound quality is desired such as feeling of solidity, the number of acoustic devices needs to be incremented. In this case, the balance between performance, device numbers, and cost might also be raised as a problem.

Tone Design

Impressions of sound tone should be noted in designs of intuitive in-vehicle auditory signals to enhance the driver's danger awareness based on human physiology and psychology. Humans perceive some sounds commonly in terms of impressions, while others vary from person to person depending on personality such as memory and culture. After ensuring certain functions by realizing generally common sound impressions with physiologically based sound elements, various sound decorations are possible by customizing the tones to suit individual backgrounds and emotions [62]. Several studies such as [42] have even emphasized the requirement for universal designs of in-vehicle auditory signals, which focused on the effectiveness of various sound elements.

Impressions of sound tone fluctuate depending on the context of the surroundings in some cases. Especially, the driving situation keeps changing from time to time. According to ISO 15006 [63], the specifications for in-vehicle auditory presentation establish several factors of success for auditory signals for driving. The success factors include comprehensibility which defines the characteristics of an auditory signal allowing the driver to understand its meaning in the context of the situation. Because of this, appropriate designs of sound impressions are needed to adapt to changing ambient conditions in driving to promote understanding of the information contents.

Supporting Multiple Events

In the driving environment, there is always more than one piece of information to convey. It is necessary to effectively convey information by in-vehicle auditory signals in a driving situation where multiple events are occurring. The human ear can selectively extract and discriminate a sound from multiple sounds, which is known as the cocktail party effect [64]. By using such an ability, they are able to distinguish and hear multiple sounds.

There is a limit to the resolution at which humans can distinguish sounds. These limits of sound resolution depend on various elements of sound. For instance, humans can differentiate sounds with temporal modulations occurring up to about $20-50$ Hz under optimal conditions and spatial separation to distinguish concurrent sounds. It will be necessary to analyze the resolution and set the optimal function according to the acoustic characteristics of the actual in-vehicle auditory signals and the desired information to be conveyed.

1.3 Research Questions and Approaches

This section describes the remaining research questions to maximize the potential of intuitive in-vehicle auditory signals and proposes methodologies that provides a foundation for the universal design of intuitive in-vehicle auditory signals in response to these challenges. Fig. 1.6 shows the scope of this dissertation, which considered design methodologies of intuitive in-vehicle auditory signals based on three functions that are mentioned in the previous section. This dissertation addresses three questions regarding three key points for intuitive in-vehicle auditory signals.

Question 1: How could the direction of danger be communicated by in-vehicle auditory signals?

The directional perceptions should be installed on intuitive in-vehicle auditory signals because this is an advantage of the aural sense over other sensory organs. Multiple loudspeakers must be used to provide directional perception by in-vehicle auditory signals, however, the space and budget in an automobile are limited. Directional perception must be realized with a certain degree of accuracy in limited computational resources. Sound reflections also exist in automobiles because of several special shapes and materials, including window glass and cushions, which can influence directional perceptions of artificial sound sources such as auditory signals in three dimensions [65, 66]. Thus it is important to evaluate the effect of sound reflections on the three-dimensionality of auditory signals in an automobile where various sound reflections exist. However, the detailed effects of sound reflections have not been clarified on the perception of sound directions, and few studies report on sound reflections with limited evaluations [67].

Safety-assistive technologies for driving need to be evaluated in a virtual environment, such as a driving simulator, to ensure safety before implementation. For this reason, a platform is needed to allow for a deeper assessment of sound directional perception with sound reflections in virtual spaces for the goal of fulfilling intuitive in-vehicle

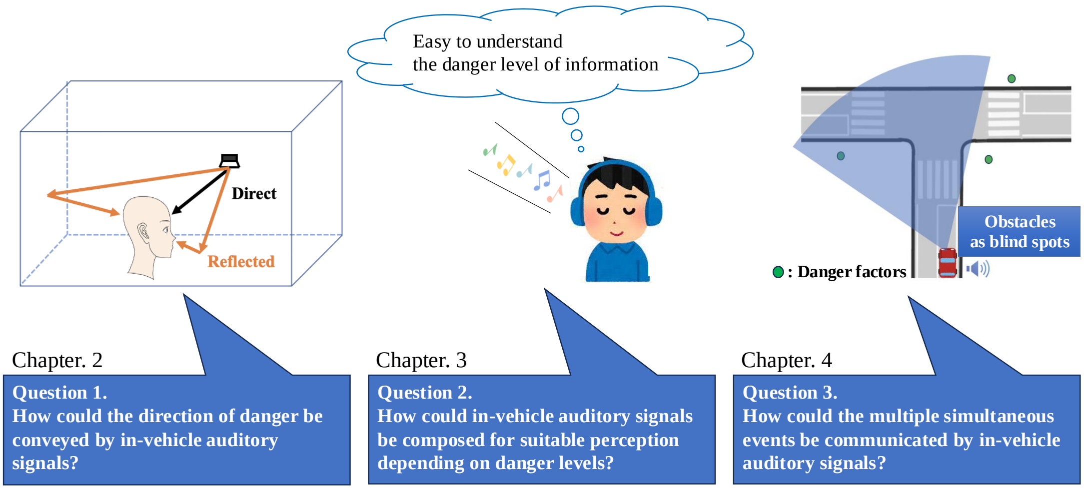


Figure 1.6: Study scope of this dissertation.

auditory signals where directional perceptions are installed. However, there are still no reports on a virtual environment in which the effect of reflections on 3D sound perceptions can be easily evaluated. Therefore, this research proposes a platform designed to evaluate the effect of reflections on 3D sound perceptions in the virtual environment to aim to realize auditory signals with 3D sound in automobiles, which consider situations with limited computational resources.

Question 2: How could in-vehicle auditory signals be composed for suitable perception depending on danger levels?

Intuitive in-vehicle auditory signals require to be designed with various sound elements to quickly identify and respond to surrounding hazards. However, the traditional in-vehicle auditory signals tend to be monotonous beep sounds to inform simply the driver of occurring danger factors, which is difficult to understand various information. Humans continually add interpretations to the aural information based on their memories and cultural backgrounds, which makes it difficult to implement a wide variety of sounds based on assessing what sounds effectively communicate for humans. Thus, this problem can be solved by initially focusing on the senses that humans generally have in common based on music theory and psychoacoustic. Then, the senses of individual experience and memory can be used to improve the comfort of auditory signals and optimize their function.

It seems to be effective to design in-vehicle auditory signals by classifying them according to the danger levels of the information to be conveyed when using the senses that humans generally have based on music theory and psychoacoustics in the presentation of information. This is because music theory has an aspect of a stepwise structure to systematize as the theory whereas sound has a continuous aspect. Therefore, this dissertation classifies the feature of in-vehicle auditory signals into stepwise danger levels and then explores suitable sound elements for each danger level based on music theory and psychoacoustic. Through summaries of explored sound elements, this dissertation contributes to realizing a foundation for a universal design of intuitive in-vehicle auditory signals.

Question 3: How could the multiple simultaneous events be communicated by in-vehicle auditory signals?

In the driving environment, there is always more than one event to convey. In an ever-changing driving environment, information on various levels of danger exists simultaneously. It is necessary to consider a method and a framework on how to effectively convey information with in-vehicle auditory signals according to the danger levels based on the variable segregation abilities of humans of each sound element in a driving situation where multiple events are occurring.

As a means of distinguishing between multiple sounds, it may be effective to add 3D sound localization and different sound impressions based on levels of danger as described so far. However, there are driving scenarios that cannot be resolved by these methods. For instance, if information on the close danger level exists in almost the same direction, it is difficult to communicate the information using solely 3D sound localization and sound impressions based on levels of danger. In particular, information to be communicated in advance of an accident occurs more often than critical information that could lead to an accident. It is required to consider effective means to replace the 3D sound localization and tone designs based on levels of danger. Therefore, this dissertation tackles this problem by focusing on the differences in frequency and onset timing, and then evaluates the segregation abilities of auditory signals using these methods.

1.4 Organization

The rest of this dissertation is organized as follows. Chapter 2 presents a design platform for 3D sound localization specialized in evaluating the effect of synthetic reflected sounds. The platform provides to evaluate the effect of reflected sounds on perceptions of sound direction with Head-Related Transfer Functions (HRTFs) in a rectangular virtual room composed of six reflective materials. The experimental results show the influences of reflected sounds varied according to the direction of reflected sounds, and the proposed platform is expected in the future to demystify the various effects of reflected sounds on perceptions of the direction of sound sources, which will contribute to realizing in-vehicle auditory signals with 3D sound perception inside automobiles where special sound reflections exist.

Chapter 3 explores the design of intuitive in-vehicle auditory signals to enhance driver awareness via danger perception. The proposed in-vehicle auditory signals, which are designed with musical and acoustic elements, can intuitively inform a driver, especially of caution signals for dangerous situations and advisory signals with comfortable signals. The designed in-vehicle auditory signals are evaluated by questionnaire and principal component analysis. The experiments demonstrate that designed in-vehicle auditory signals improve intuitive awareness through the perception of danger.

Chapter 4 proposes in-vehicle auditory signals to help drivers distinguish multiple potential collision risks posed by approaching pedestrians. Traditionally, by utilizing 3D sound localization and tonal designs based on various hazard levels, drivers can easily differentiate multiple sounds, which enhances their awareness of pedestrians from various angles simultaneously. However, if the auditory signals originate from almost the same direction, drivers may not be able to detect multiple pedestrians. The proposed method addresses this by altering the frequency and onset timing of auditory signals, allowing drivers to distinguish multiple potential collision risks, taking advantage of Time-To-Collision (TTC) which is the time until the driver might collide with a risk factor. Experimental results demonstrate a significant improvement with the proposed methods compared to the traditional one due to the timing shift rather than the frequency shift.

Chapter 5 serves as the concluding section of this dissertation providing a summary of the main points discussed and suggesting potential areas for further research.

Chapter 2

A Design Platform for 3D Sound Localization

The auditory system can play a role in receiving intuitively directional perception from all directions against other senses. This feature is especially effective in driving circumstances where drivers may concentrate on handling their automobiles and the road conditions in front. This is why directional information ought to be installed on intuitive in-vehicle auditory signals to awaken outside the conscious mind of drivers. On the other hand, sound reflections also subsist in the inside space of cars, which is caused by several special shapes and materials including window glasses and cushions. These sound reflections have the potential to influence the 3D sound perception of drivers. This chapter[1] presents a platform designed to evaluate the impact of synthetic reflection on 3D sound perception. The proposed platform implements sound reflections by imaginary sound sources using the method of Head-Related Transfer Functions (HRTFs) same as direct arriving sounds from sound sources, which enables users to assess the effect of sound reflections on the directional perceptions of sound sources. The experimentation is executed to verify the usefulness of the proposed platform and resulted in the influences of reflected sounds varied according to the direction of sound reflections which shows that the proposed platform is anticipated to clarify how different reflected sounds influence the perception of sound source direction in the future.

2.1 Introduction

In recent years, technologies that replicate the physical world in virtual spaces have been rapidly advancing, such as the Metaverse [68] and virtual reality (VR) [69]. With these concepts and cutting-edge technologies, human beings can now lead virtual lives almost

[1]This chapter is based on "A Platform for Evaluation of Synthetic Reflected Sounds on 3D Sound Localization" [J1] listed in Journal Article (Refereed) of Publications in this dissertation by the same author who owns the copyright of [J1], which appeared in IEEE Access.

indistinguishable from the real world, which enables activities including the beginnings of businesses and the creation of digital art with verifiable ownership. These advancements open the door to reimagining human society in virtual spaces, potentially creating an idealized version of reality. To fulfill the ideal, various sensory-based mechanisms have been developed to enhance the realism of virtual worlds. Moreover, state-of-the-art also enriches the real-world lives of human society. Augmented reality [69] called AR as an abbreviation overlays digital information such as images and sound data onto the real world in real-time. AR expands the real-world experience by adding virtual elements that facilitate users to interact while still being aware of their actual surroundings, whereas VR creates a completely immersive virtual environment.

In the aural realm, numerous research has focused on the virtual experiences of 3D sound perceptions [70, 71]. For instance, 3D sound localization is employed to identify the position of a sound source as if feeling the 3D sound fields [72, 73]. These techniques can pinpoint a specific sound source providing information on its direction and distance with methods focusing on the shape features of human bodies, including Head-Related Transfer Functions (HRTFs) which are frequency characteristics incorporating positional information into arbitrary monaural sound sources to enable 3D sound perceptions [74]. Furthermore, sound reflections from the surrounding environment are crucial in creating more natural 3D sound perceptions. Since humans naturally perceive sounds along with their reflections in the real world, these reflections cannot be ignored when aiming to replicate an environment more faithfully in virtual settings. As a result, various studies have introduced methods to implement sound reflections in virtual spaces tailored to different contexts and evaluated their impact on 3D sound perceptions [75–77].

The perspectives of 3D sound localization are also vital for in-vehicle auditory signals which demand perceiving intuitively the direction of events existing in the surrounding areas while driving. In this case, the driver cannot easily pay attention to visual assistance because of concentration on the road or dashboard for safe driving. For this reason, intuitive in-vehicle auditory signals mounted with directional information can induce effective awareness of hazardous factors without interfering with visibility behavior for driving operations. On the other hand, sound reflections caused by special materials, including window glasses and cushions, also exist inside automobiles. These reflections will alter the directional perception of sounds until reaching the ears of drivers from audio equipment. Consequently, these acoustic properties of sound reflections inside automobiles must be taken into consideration for that intuitive in-vehicle auditory signals should be designed to convey directional information.

Sound reflections serve primarily two functions in the virtual reproduction of 3D sound perceptions. First, reflections expand the sound field space and enhance the spatial feelings of the sound field. The features of these reflections vary based on the surrounding environment, promoting the realism of sound sources concerning their context. In this scenario, Binaural Room Impulse Responses (BRIRs) are commonly employed to describe the acoustic transmission from a sound source to the listener's outer ears [78,79]. The second function of sound reflections pertains to 3D sound localization. Recent research indicates that sound reflections can significantly impact the accuracy of 3D sound localization [67,80]. These studies have identified sound reflections as important cues for 3D sound localization. However, the specific effects of sound reflections on human perception of sound direction remain unclear. Ongoing research continues to explore how these reflections influence human perception of sound direction, as noted in [67].

A detailed analysis of sound reflections is crucial for the accurate simulation of 3D sound localization virtually. Various platforms have been developed to assess the effects of these reflections [67, 81]. However, evaluating the detailed impact of each sound reflection on human perception of sound direction poses challenges in these platforms. Typically, these platforms utilize BRIRs to enhance the realism of sound fields concerning their surrounding environments. The primary purpose of BRIRs fundamentally differs from assessing reflection's influence on sound directional perception; they express the overall sound fields concerning surroundings as a frequency characteristic that combines early and late reflections, which cannot discern the unique attributes of each reflected sound, such as differences in arrival time and precise positions. Consequently, a detailed comparative analysis of the impacts of sound reflections on directional perception proves difficult within BRIR evaluation environments. Although some studies have indicated that early reflections can assist in locating sound sources within a space [65, 66], the mechanisms by which early reflections affect sound directional perception remain poorly understood [67]. Understanding these primary interactions between sound waves and surfaces is critical for grasping sound localization. Early reflections are essential to be focused on at first as their influence is more pronounced and relatively simple compared to the total spectrum of acoustic phenomena. Thus, there is a need for a new method or platform that allows for a more detailed assessment of sound directional perception with a particular focus on initial reflections and allows for variations in sound reflection settings, but this has not yet been realized. This could significantly contribute to resolving the problems related to human perception of sound direction.

This chapter introduces a 3D sound localization platform enabling a detailed assessment of the impacts of sound reflections on the human perception of sound direction. An overview of the objective platform is illustrated in Fig. 2.1 explained below in detail. First of all, an arbitrary monaural sound source is inputted with its data of position and type, and the reflectors' data of position and reflection coefficient. Then, the positions of the sound reflections are calculated using conventional methods based on the positional date of the sound source and each reflector. Each sound reflection is localized using HRTFs, which are usually realized by methods such as BRIRs. The audio data identical to the sound source is duplicated, and the playback time of these duplicated audio data is delayed and mixed based on the arrival time differences of each reflected sound.

Figure 2.1: An overview of this research subject to design platform for 3D sound localization with sound reflections.

These processes of localizing, delaying, and mixing each reflected sound are new methods proposed in this research highlighted in Fig. 2.1 with the orange color. Through this process, the proposed platform provides free settings of the factors of early sound reflections from the surrounding environments. The demonstration is practiced to verify the usefulness of the proposed platform by evaluating the effective setting up of sound reflections related to the human perception of sound direction.

The contributions of this chapter are as follows:

1. This research introduces a platform with a new method designed to evaluate the detailed effects of sound reflections on 3D sound localization. This proposed platform provides free settings of the parameters of early reflections from six reflectors of a rectangular room.
2. The experiment is practiced to show the usefulness of the proposed platform and resulted in the influences of reflected sounds varied according to the direction of sound reflections and the types of sound sources. This result shows that the proposed platform contributes to clarifying how different sound reflections influence the directional perception of sound sources.

The subsequent section of this chapter is structured as follows. Section 2.2 describes related research to 3D sound localization. Section 2.3 proposes a platform for 3D sound localization with synthetic sound reflections. The results of the experiment are presented in Section 2.4, and this chapter is concluded in Section 2.5.

2.2 Related Research

2.2.1 A Head-Related Transfer Function (HRTF)

The fundamental factors for human perceptions of sound direction are related to differences in arrival times and volume levels between the left and right ears. These factors are called Inter-aural Time Difference (ITD) and Inter-aural Level Difference (ILD) [82], which allow listeners to perceive the left and right directions of the sound source. On the other hand, ITD and ILD cannot clearly present the perception regarding the front-

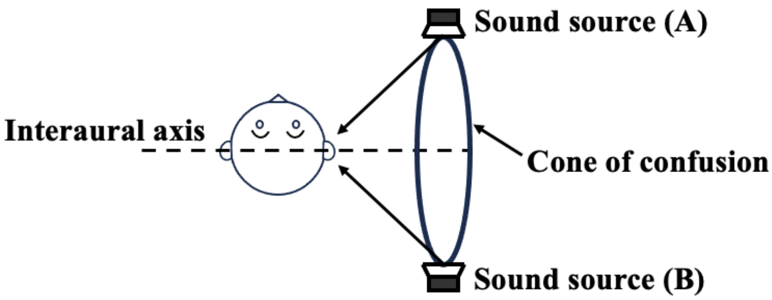


Figure 2.2: An example of the cone phenomenon [3].

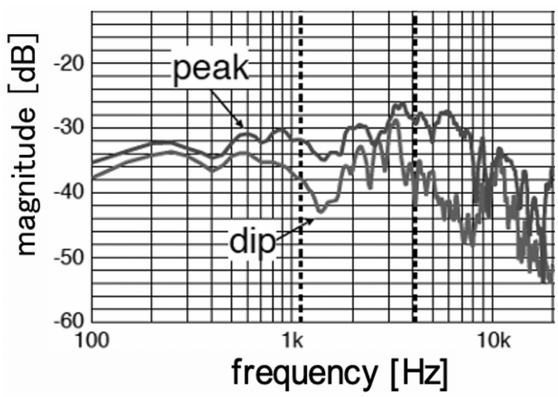


Figure 2.3: An example of a frequency-magnitude spectrum of HRTFs [4].

back and the top-bottom directions. This is known as the phenomenon called the cone of confusion and front-back confusion expressed in Fig. 2.2. In this figure, the listener can confuse the front and back positions of the sound source (A) and the sound source (B) because ITD and ILD are diluted between these positions.

A solution to the cone of confusion and front-back confusion is head motion by which the position of sound sources moves to a position where it is easier to determine the direction of the sound sources for human perceptions. However, the implementation of head motion on the artificial sounds binauralally is impractical compared with the real sound sources under the condition of using binaural devices such as headphones which are generally used to listen to artificial sounds binauralally by human ears.

A sound reaches human ears through reflections and diffraction by human bodies such as the head, auricle, and torso. By using these acoustic features influenced by the human body in addition to ITD and ILD, 3D sound localization can be refined. These acoustic features are represented in a frequency characteristic which is called Head-Related Transfer Functions (HRTFs) [74] illustrated in Fig. 2.3. As HRTFs indicate the influence of human body shape on the directional perception of sound sources, the

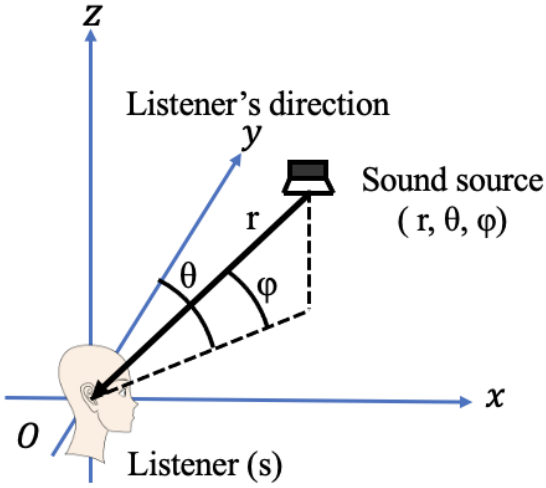


Figure 2.4: An example of the coordinate relationship between a sound source and a listener.

frequency characteristics vary individually as manifested in the shape of the peaks and dips. The HRTFs denoted as $H_{\mathrm{l,r}}$ for the left and right ears are defined by Eq. 2.1.

$$H_{\mathrm{l,r}}(s,r,\theta,\phi,\omega) = \frac{G_{\mathrm{l,r}}(s,r,\theta,\phi,\omega)}{F(r,\theta,\phi,\omega)}, \tag{2.1}$$

where $G_{\mathrm{l,r}}$ represents the transfer function from a sound source to the left and right ear canals, or eardrums, of a listener in a free field. On the other hand, F denotes the transfer function from a sound source to the center of a listener's head in a free field without the listener present. The subscripts $\mathrm{l,r}$ indicate the left and right ears, respectively. Furthermore, the variables used in the HRTF include the following parameters in Fig. 2.4.

- s: The listener's position (origin)
- r: The distance from the sound source to the listener
- θ: The lateral angle of the sound source
- ϕ: The elevation angle of the sound source
- ω: The angular frequency of the sound source

These HRTFs are essential for incorporating positional information into arbitrary monaural sound sources and enabling 3D sound localization that creates the perception of sound coming from specific directions relative to the listener.

Several approaches have attempted to recreate the frequency characteristics of HRTFs individually and accurately [83]. For instance, some methods customize HRTFs based on anthropometric measurements of the listener pinnae of the ears [84, 85] by which front-back confusion rates can be reduced [86]. In other approaches, advanced technologies including deep learning [87] and a GPU accelerator [88] have been integrated in recent years to recreate HRTFs with high speed and high precision. However, the methods described above demand enormous computational costs to prepare individual HRTFs and cause obstacles to integration on mobile devices. Especially, as the usage of artificial sounds in virtual environments has become more familiar to the common public in recent years, it is significant to realize 3D sound perception on terminal devices where computing resources are limited.

Various trials have been undertaken to produce generic HRTFs for mobile devices in [81, 89]. For instance, some methods attempt to reduce computation costs to produce generic HRTFs focusing on the shape features of HRTFs including the peaks and dips [4]. The work on [90] synthesizes and interpolates HRTFs efficiently to reduce computation costs substantially. Nevertheless, generic HRTFs still cause errors due to inaccuracies in front-back and vertical directional perception.

2.2.2 Sound Reflections

Ordinarily, humans listen to a sound reflected off the walls in addition to sounds reaching ears directly from sound sources. Some research has noted that humans utilize sound reflections to acquire hints about surrounding environments. For instance, echolocation [91] is a method to understand profoundly what exists around one by how one's voice reflects. In the study realm of robot engineering, using sound reflections greatly improves the accuracy of sound source identification [92]. Sound reflections also apply an effect of natural feelings to music sounds which is a well-known called reverberation effect [93, 94]. For these factors, evaluating the impacts of sound reflections is vital for producing a virtual sound environment to simulate an acoustic field realistically and comfortably. Furthermore, a comprehensive analysis of sound reflection is required to communicate crucial information such as severe danger via 3D sound localization within a virtual sound environment that remains true to reality.

Some recent studies reported that sound reflections support humans in localizing sound sources in three-dimensional spaces. The work on [80] even stated that the usage of sound reflections has the potential to solve the front-back confusion of HRTFs mentioned in Section 2.2.1. The other work on [95] also stated that sound reflections can effectively reduce the rate of front-back confusion in artificial sounds compared to head motion, which has been described as a typical solution for front-back confusion. On the other hand, the in-depth impact of sound reflections on the directional perception of sound sources has not been thoroughly studied, and the available research offers a limited analysis of these impacts [67].

Some research has reported methodologies and platforms for recreating sound reflections with artificial sound sources due to the required circumstances, by which the impact of sound reflections on the 3D sound perceptions has been evaluated [96, 97]. Binaural Room Impulse Responses (BRIRs) are commonly used in these researches to recreate sound reflections by capturing the acoustic transmission from a sound source to the listener's ears. BRIRs are generated using multiple microphone arrays such as ambisonics allowing the addition of spatial acoustic information to artificial sound sources. This method enhances the 3D auditory experience and contributes to a more realistic representation of the surrounding environment. However, BRIRs present challenges regarding the detailed analysis of sound reflections' effects on the directional perception of sound sources. The BRIRs complicate the process of separating and examining individual reflected sounds in depth. Since BRIRs blend early and late reflections into a combined frequency characteristic, the method cannot allow precise analysis of factors such as the arrival times and exact locations of each reflected sound. This limitation has caused research to focus on only a few parameters of sound reflections, which leaves many aspects of their impacts on 3D sound localization unexplored. For instance, whereas early reflections are known to influence 3D sound localization and even can help in resolving front-back confusions [65, 66], the exact mechanisms are still not fully understood [67]. It is crucial to prioritize the study of early reflections, as its effects are pronounced and relatively simple compared with the full spectrum of acoustic phenomena.

Early reflections carry more weight on 3D sound localization than secondary or higher-order reflections, which are weakened by attenuation [76, 98] and masking effects [99]. By concentrating on early reflections, researchers can better understand the fundamental mechanisms of 3D sound localization without the complications introduced by more complex acoustic phenomena. Several pieces of research have shown that these early reflections play a role in resolving sound directional ambiguities in the directional perception of sound sources, including front-back confusions [67, 80]. Despite this, there has been no established method to easily evaluate the effects of early reflections on 3D sound localization.

This research presents a novel platform aimed at assessing the effects of early reflections on 3D sound localization. Beyond contributing to the development of precise auditory perception models, the platform is anticipated to serve as a foundation for incorporating these insights into more advanced simulations encompassing various forms of acoustic reflections.

2.3 A 3D Sound Localization Platform with Synthetic Sound Reflections

2.3.1 Overviews

A platform designed to evaluate the impact of synthetic reflection on 3D sound localization is intended for implementation in a handy device such as a smartphone with headphones illustrated as Fig. 2.5. The input data involved in a sound source and sound

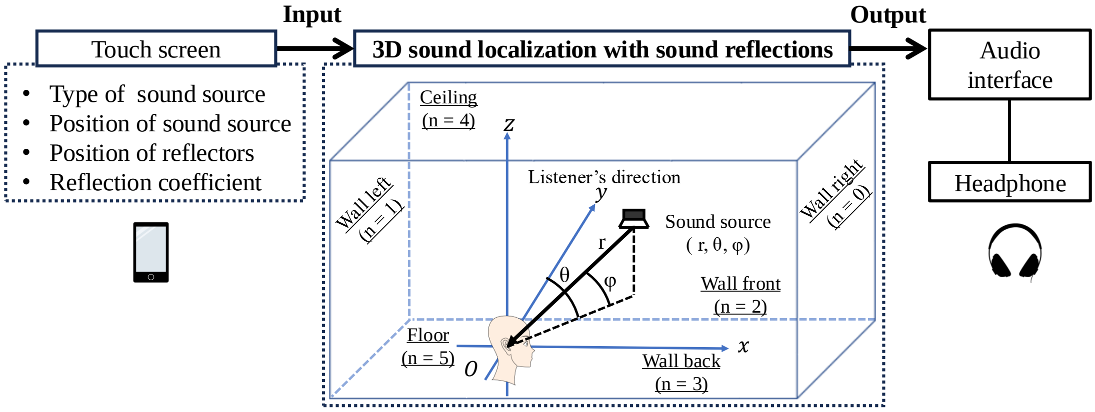


Figure 2.5: An overview of the proposed platform.

reflections are set manually by users through an input device, such as the smartphone's touch screen. This platform presents a rectangular space with six reflectors causing sound reflections: ceiling, floor, front, back, right, and left wall in Fig. 2.5. A sound source is placed in the rectangular space, and one early reflection is generated on each reflector based on input data. The directly arriving sound from the sound source and the reflections from all of the reflectors are subjected to a process for 3D sound localization using HRTFs with positional information added to each sound. The process of 3D sound localization uses spherical coordinates, in which the distance between the listener and the sound source (r), horizontal angle (θ), and vertical angle (ϕ) are defined as parameters for coordinates of a sound source and sound reflections. The 3D localized sound data synthesized with sound reflections is outputted through an audio interface and a headphone.

The proposed platform can evaluate the sound reflections from each frequency characteristic by adding positional information to the sound reflections using HRTF. The platform enables users to freely set parameters for sound sources, including type, gain, distance, horizontal angle, and vertical angle. This platform also allows for flexible adjustment of both the distance and the reflection coefficient (h) of the reflective material for each sound reflection, which enables precise control over the gain and delay of the sound reflections.

2.3.2 Preparation Settings of the Position of Reflected Sounds

This proposed platform employs the mirror image method to calculate the position of each reflected sound [81]. As illustrated in Fig. 2.6, when a sound reflects off a reflector and arrives at the listener, the sound reflection can be conceptualized as arriving from a virtual sound source denoted as *Imaginary sound source*, positioned symmetrically opposite the original sound source with the reflector acting as the axis of symmetry.

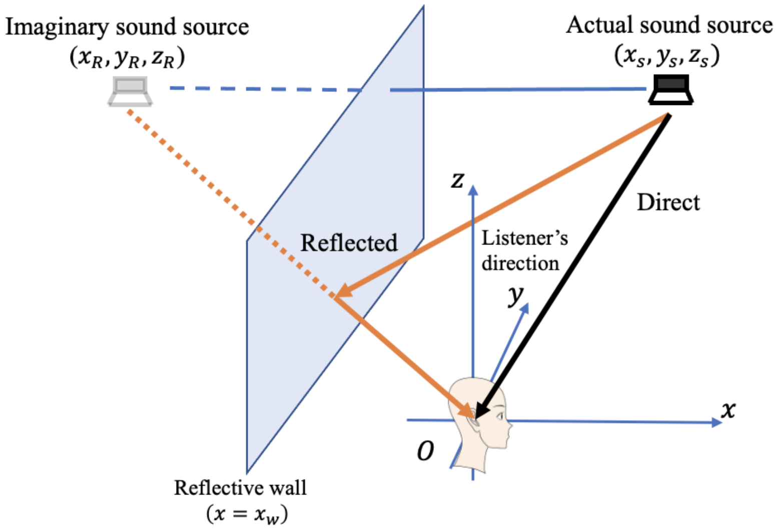


Figure 2.6: The imaginary sound source for a reflected sound.

Thus, if the listener is positioned at the origin of the Cartesian coordinates, the reflector perpendicular to x-axis is located at a distance of x_w, the location of the sound source is at coordinates x_s, y_s, z_s, the location of the virtual sound source for the reflected sound can be expressed by Eq. (2.2). The n in (2.2) indicates the type of reflectors in Fig. 2.5, which corresponds to the negative or positive symbols depending on the positional relation with the listener and reflectors. Likewise, if the distances from the listener to the reflectors perpendicular to y and z axes are defined as y_w and z_w, the positions of the virtual sound sources for those reflected sounds are given by Eq. (2.3) and Eq. (2.4). The difference in distances from the origin to the positions of the sound sources represented as d is determined using Eq. (2.5). Consequently, the time delay of the sound reflections, denoted as t is calculated using Eq. (2.6) where c represents the speed of sound.

$$\begin{cases} x_R = 2(-1)^n x_w - x_s \\ y_R = y_s \\ z_R = z_s \end{cases} \tag{2.2}$$

$$\begin{cases} x_R = x_s \\ y_R = 2(-1)^n y_w - y_s \\ z_R = z_s \end{cases} \tag{2.3}$$

$$\begin{cases} x_R = x_s \\ y_R = y_s \\ z_R = 2(-1)^n z_w - z_s \end{cases} \tag{2.4}$$

$$d = \sqrt{x_R^2 + y_R^2 + z_R^2} - \sqrt{x_s^2 + y_s^2 + z_s^2} \tag{2.5}$$

$$t = d/c \tag{2.6}$$

2.3.3 HRTFs Employed in the Proposed Platform

The 3D sound localization process in the proposed platform employs a method of embedded 3D sound localization processing [4]. This method reduces computational complexity and generates generic HRTFs, which assumes the future development of various applications.

As illustrated in Fig. 2.7, the low-computation approach from [4] minimizes computational costs by applying approximate arithmetic for each frequency band based on its characteristics. A monaural sound input is divided into sub-bands using a high-pass filter (HPF), band-pass filter (BPF), and low-pass filter (LPF). Then, delay times are added to each sub-band signal via a delay buffer (Delay). Following this, the sub-band signals undergo appropriate signal processing with peaking filters (Peak) and a comb filter (Comb) used to replicate the distinctive curve shape of the HRTFs for each frequency band. Finally, the three sub-bands are mixed, resulting in reduced computational load and enabling implementation on embedded systems. The proposed platform can simulate both directly arrived sound and multiple reflected sounds from any monaural sound source by utilizing the method from [4] within the 3D sound localization process.

2.3.4 A Design Method of Synthetic Sound Reflections

The process for localizing 3D sound with synthetic sound reflections is outlined in Fig. 2.8. This process begins with the input of an arbitrary monaural sound source and its positional data. Then, the positions of an imaginary sound source for the sound reflections are calculated using the mirror image method described in Section 2.3.2 based on the positional date of the sound source and each reflector. In Fig. 2.8, sound reflections from the right and left reflectors are used as examples of how these positions of imaginary sound sources are determined. A maximum total of sound sources is assumed in seven sources: *Direct* sound refers to the sound coming directly from a sound source, whereas *Reflected* sounds are those resulting from reflections off reflectors. Each sound source is localized using HRTFs generated following the method in Section 2.3.3. *Direct* source defined as (x_s, y_s, z_s) for the positions of the sound sources, and imaginary sound sources for the sound reflections defined as (x_{R0}, y_{R0}, z_{R0}) and (x_{R1}, y_{R1}, z_{R1}) for two reflected sounds denoted as *Reflected 0* and *Reflected 1*, are then calculated here as an example. A delay is applied to these reflected sounds according to their positions to determine their arrival times. Let t represent the delay time for each reflected sound, where $t = 0$ corresponds to the arrival time of *Direct* sound at the listener. Audio data identical to the sound source is duplicated, and playback timing of this duplicated audio

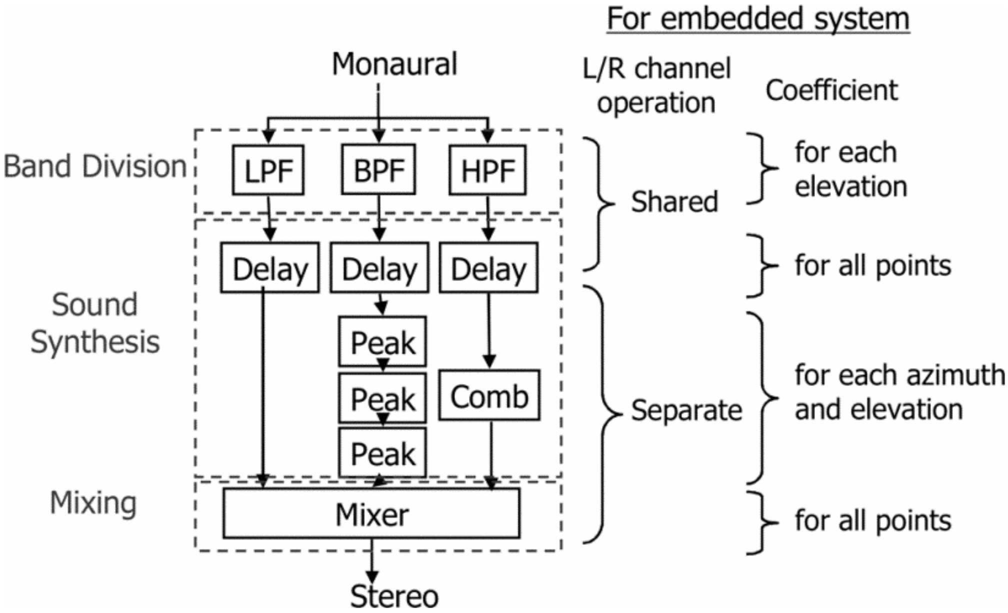


Figure 2.7: The low-computation method of 3D sound localization for embedded systems [4].

data is delayed based on t which reflects the arrival time differences of each reflected sound. This delayed audio is defined as *Reflected* sounds and mixed with the original audio data of the sound source defined as *Direct* sound to prevent any further delays caused by parallel processing.

Typically, sound reflections are generated differently than *Direct* sound by using methods BRIRs in many cases. However, most frameworks and platforms that use BRIRs aim to enhance the atmosphere of 3D sound rather than focusing on precise directional cues including arrival time and position, which is not easy to assess the effect of sound sources on directional perception in detail. To address this, the proposed platform synthesizes each reflected sound by calculating its arrival time difference using HRTFs to provide more accurate localization than BRIR-based methods and to allow a detailed evaluation of how sound reflections influence the directional perception of sound sources. Additionally, the platform employs the HRTF generation method for embedded systems discussed in Section 2.3.3, which allows for the implementation of numerous reflected sounds with HRTFs on mobile devices and facilitates easy setup for 3D sound localization with sound reflections for any monaural sound input. The proposed platform has been successfully implemented on an Apple iPad using the AudioToolbox framework with Objective-C++ in Xcode.

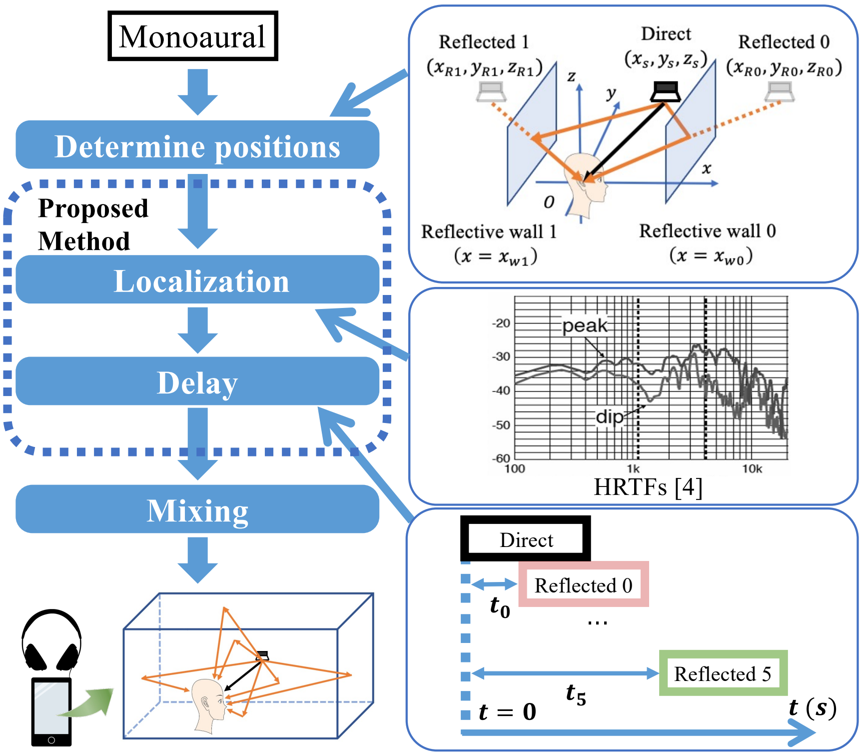


Figure 2.8: Processes to localize 3D sound with sound reflections.

2.4 Experiments

2.4.1 Setup

Overviews

The experiment is carried out to demonstrate the usefulness of the proposed platform for assessing the impact of sound reflections on the directional perception of sound sources. The research aims to develop a platform rather than solely analyzing the effects of sound reflections. Thus, this experiment investigates whether the proposed platform can evaluate without problems that the presence or absence of sound reflections affects the directional perception of the sound source. This experiment validates the effect of sound reflections with the proposed platform by focusing on the precedence effect as an example of how sound reflections affect the directional perception of sound sources.

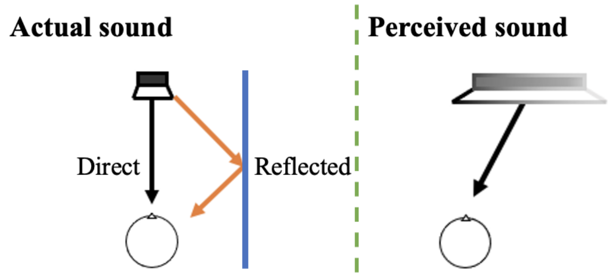


Figure 2.9: An example of the precedent effect.

Evaluation Methods

This experiment employs the precedence effect to verify the effect of sound reflection by the proposed platform. In this context, sound reflections do not significantly alter the perceived location of a sound source but instead enhance the sound image. The precedence effect is a well-documented phenomenon related to sound reflections [98]. This phenomenon occurs when a series of sounds reach the ears within a brief interval, resulting in as if these sounds are a single auditory event and primarily localized to the position of the first sound that arrives. The precedence effect is particularly relevant when sounds reach a listener alongside sound reflections [100]. When a subsequent sound including sound reflection arrives further from the initial sound, the sound image is perceived as broadened, which is illustrated in Fig. 2.9. This expansion occurs because of differences in arrival time and sound intensity, which can obscure the exact location of the sound source.

On a different note, humans tend to perceive the front-back location of a sound source more effectively when the sound source is positioned obliquely, as opposed to directly in front or on the side [101]. Consequently, it appears that humans can enhance their perception of front-back position by leveraging the precedence effect of sound reflections and magnifying the sound image of a sound source near the front or side. Thus, sound reflections play a crucial role in 3D sound localization, which highlights the necessity to evaluate the various effects of these sound reflections. Building on this premise, the experiment illustrates the potential benefits of the proposed platform by validating how sound reflections influence the directional perception of sound sources with a particular emphasis on the precedence effect as a key example in assessing the impact of sound reflections.

This experiment is conducted using two approaches involving a questionnaire[2]. In

[2]This experiment is carried out with the permission of the Osaka University Research Ethics Review Committee. Before the experiment, experiment conductors explain the details of the experiment, its precautions, and how the experimental data will be used. Participants are also informed how to respond in the event of unforeseen circumstances and that the experiment can be terminated under any circumstances. After the participants agree to the above, the experiment begins with them.

Table 2.1: Response items and corresponding scores.

Items	Scores (pt.)
Felt a difference	2
Relatively felt a difference	1
Not sure	0
Relatively not felt a difference	-1
Not felt any difference	-2

the first approach, participants answer whether they perceive a difference in sound localization between the front and back positions when the sound sources play the same type. The order is randomized, and the sound sources are played on the front and the back. The participants answer questions with response items for each sound source, which is played for approximately five seconds at a time that can be repeated. These response items are summarized in Table 2.1 along with their corresponding scores. The average scores are then calculated to assess the differences in perceptions of front and back localization, particularly regarding the presence or absence of sound reflections. The second approach for the evaluation involves participants choosing between two options: whether they hear the sound from the front or the back. This approach allows us to compare the percentage of correct responses based on the presence or absence of sound reflections. By analyzing the results from this second approach alongside the first, it is possible to assess how the varying perceptions of sound direction arising from different positions and arrival times of sound reflections affect the participant's ability to determine the exact locations of the sound sources. Throughout these evaluations, the differences in the effects of sound reflections are assessed due to variations in the type and position of the sound sources.

Environments

The experiment involved five male participants in their twenties without hearing impairments who uses headphones plugged into the proposed platform to assess differences in front-back sound perception depending on the presence or absence of sound reflections. The primary focus of this research is the varying effects of the position and types of sound reflections. For this reason, this research decided that it is permissible to disregard differences in the abilities of the participants' hearing for this particular analysis. The analysis of the variability in hearing abilities among participants is planned for a future extension of this work. For this experiment, the terms of *Front* and *Back* are defined as shown in Fig. 2.10. If the listener's straight-ahead direction is defined as 0° on the horizontal plane, *Front* is described as the range $0^\circ \leq \theta < 90^\circ$ and $270^\circ \leq \theta < 360^\circ$, whereas *Back* refers to the range $90^\circ \leq \theta < 270^\circ$. Simply put, *Front* includes the areas in front and on either side from the ears of listeners, whereas *Back* covers the area

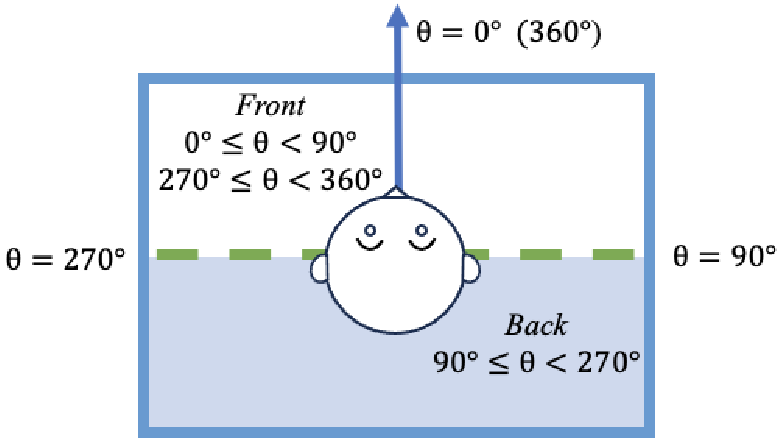


Figure 2.10: A definition of *Front* and *Back*.

behind them.

This experiment is practiced under three specific conditions, as illustrated in Fig. 2.11. First in *Case 1*, the sound source is positioned at *Front* (0° in horizontal angle) and *Back* (180° in horizontal angle). Sound reflections from the left and right walls are utilized in this case, which is anticipated to enhance the judgment of the front-back location of the sound source. This is predicted to be achieved by stretching the sound images towards the right and left, where perceiving front-back orientation is more sensitive. Second, *Case 2* involves evaluating sound reflections for determining the front-back position of a sound source that is located closer to the sides of the listeners. In this condition, the sound source is located at horizontal angles of 85° as *Front* and angles of 95° as *Back*. Sound reflections from the front and back reflectors are employed, which is expected to improve the front-back judgment of the location of sound sources by stretching the sound image into a more perceptible front-back orientation. Third in *Case 3*, the sound source remains in the same position as in *Case 2*; however, the position of reflected sounds differ. Only the sounds reflected from the right wall are considered. In this condition, it is predicted that the sound reflections will not have a stretching effect on the sound image and lead to less pronounced effects of the sound reflections compared to the other cases. Since it is assumed that sound reflections from the left wall would have a similar effect due to the symmetry with the right wall's reflections, this experiment does not evaluate the effect of left wall reflections on the directional perception of sound sources in this case.

This experiment uses six types of sound sources: 1/3-octave band noise at 100 Hz (low-frequency narrow-band noise), 1/3-octave band noise at 1 kHz (mid-frequency narrow-band noise), 1/3-octave band noise at 10 kHz (high-frequency narrow-band noise), pink noise (wide-band noise with a focus on low-frequency components), blue noise (wide-band noise with a focus on high-frequency components), and white noise

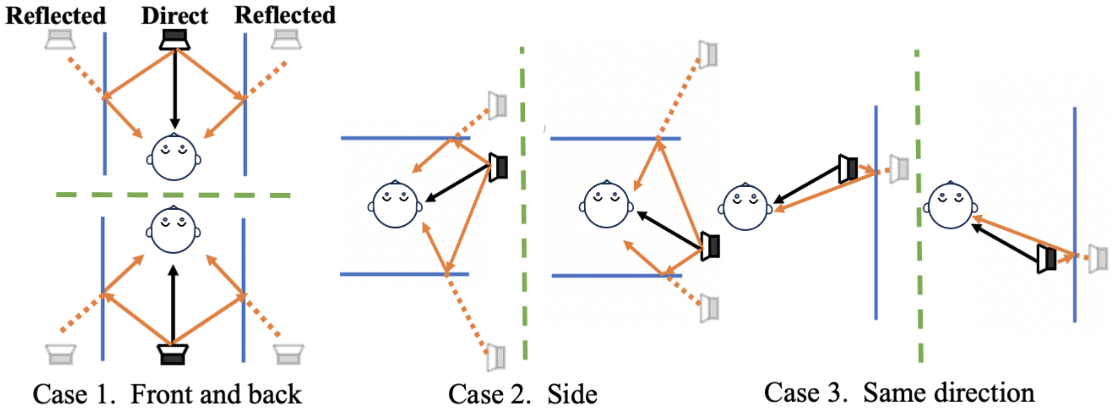


Figure 2.11: Experimental environments.

(wide-band noise covering the full audible range). It evaluates the effect of each frequency band on the localization by sound reflections. For other parameters, the distance between a sound source and a listener is fixed at 1.0 m. The reflection walls are placed 2.0 m from a listener, and the reflection coefficient is set at $h = 0.9$ assuming a natural sound level that is slightly lower in volume than the directly arriving sound. In this study, the distance and reflection coefficient of sound reflections, which are important evaluation elements, are excluded from this evaluation to focus on the directions of sound sources and sound reflections as the first step of the evaluation of sound reflections on directional perception.

2.4.2 Results

The results are analyzed using a two-tailed paired-sample t-test to reveal statistically significant differences for certain sound sources. The results of the first method are presented in Table 2.2 and Table 2.3, which are categorized by the frequency band ranges. For each type of sound source, the results are represented as averaged scores reflecting the perceived level of front-back differentiation in environments both with and without sound reflections. The tables display results for *Case 1* through *Case 3*, which correspond to the experimental environments described in Section 2.4.1. *No Reflection* shows the averaged perception scores for front-back differentiation in each *Case* without sound reflections, whereas *Reflection* presents the scores with sound reflections according to Table 2.1. *P Value* indicates the results of the two-tailed paired-sample t-test to compare the scores between the absence and presence of sound reflections in each *Case*. Bold numbers highlight the sound sources that showed statistically significant differences in perception when the same type of sound source is played from *Front* and *Back* positions.

For sound sources of narrow-band noise, a significant difference of $p < 0.05$ was found for the 1/3-octave band noise at the sound source of 10 kHz (high-frequency

Table 2.2: Results of the averaged scores representing the perceived level of front-back differentiation with and without sound reflections (narrow-band noises).

	100 Hz		
	No Reflection	Reflection	P Value
Case 1	-1.000	0.000	0.394
Case 2	-0.400	-0.400	1.000
Case 3	-0.600	0.109	0.600

	1 kHz		
	No Reflection	Reflection	P Value
Case 1	-0.200	-0.200	1.000
Case 2	0.000	-1.600	0.140
Case 3	-0.600	0.400	0.142

	10 kHz		
	No Reflection	Reflection	P Value
Case 1	0.000	1.000	**0.034**
Case 2	0.400	0.800	0.587
Case 3	1.200	-0.800	**0.047**

Table 2.3: Results of the averaged scores representing the perceived level of front-back differentiation with and without sound reflections (wide-band noises).

	Pink Noise		
	No Reflection	Reflection	P Value
Case 1	1.600	0.800	0.456
Case 2	0.000	1.200	0.070
Case 3	0.400	0.000	0.772

	Blue Noise		
	No Reflection	Reflection	P Value
Case 1	0.600	-0.400	0.351
Case 2	0.000	0.600	0.529
Case 3	1.600	-0.200	0.070

	White Noise		
	No Reflection	Reflection	P Value
Case 1	1.200	0.000	0.305
Case 2	-0.200	1.800	**0.011**
Case 3	0.400	-0.200	0.553

narrow-band noise) in *Case 1* and *Case 3*. In the presence of sound reflections from the left and right walls, listeners experienced a different perception of *Front* and *Back* sound source positions compared to when no reflected sounds were present. However, when the sound source was positioned at 85° and 95° horizontal angles and sound reflections from the right reflector, which was nearly aligned with the sound source, were used, the difference in the directional perception was less noticeable with sound reflections. It appears that when the sound reflections were positioned in the same direction as the sound source and almost in a straight line, these reflections did not improve the perception of *Front* and *Back* positions as effectively. For sound sources of narrow-band noise, reflections from the left and right walls, which were farther from the sound source, seemed to have a stronger effect. Additionally, high-frequency sound sources appeared to be more influenced by sound reflections. This aligns with the findings of [102], which reported that high-frequency noises play a role in distinguishing *Front* and *Back* direction of sounds due to the shape of HRTFs. The study also demonstrated that adjusting the level of high-frequency components to modify the shape of HRTFs can reduce front-back confusion in sound localization. Therefore, it seems that sound reflections particularly affect directional perception for high-frequency sources as supported by these findings.

For sound sources of wide-band noises, when the sound source was placed at 85° and 95° in horizontal angles and sound reflections from *Front* and *Back* walls were used in *Case 2*, a significant difference of $p < 0.05$ was observed for white noise (a wide-band noise covering the full audible range). Related research has indicated that white noise sound sources influence the perception of *Front* and *Back* sound direction more effectively than other sound sources [103, 104]. In contrast, when reflections from the left and right walls were applied, no effect of the sound reflections was found for any of the three sound sources at horizontal angles of 0° and 180°. For wide-band noise sources that encompass both low and high frequencies, front and back reflections seem to effectively enhance the perception of *Front* and *Back* sound positional differences. The reason for the variation in the effectiveness of sound reflections between sound sources of narrow-band and wide-band noises is not fully understood. However, this experiment confirms that the effects of sound reflections depend on both the position and type of sound sources. Furthermore, the variation in the position of effective sound reflections among different sound sources may be influenced by the frequency composition of the sound sources, which affects the shape of HRTFs and alters the perception of sound direction. Further studies could potentially clarify why the position of effective sound reflections differs between sound sources of narrow-band and wide-band noises.

The second method comparing the percentage of correct responses based on the presence or absence of sound reflections did not reveal any significant differences in the percentage of correct responses for *Front* and *Back* across any sound sources. Detailed results from this second method for evaluation are presented in Tables 2.4 and 2.5. These results are categorized according to the frequency bands of sound sources in the same manner as the first method's results. Each type of sound source presents the

Table 2.4: Comparison of the percentage of correct responses with and without sound reflections (narrow-band noises).

	100 Hz		
	No Reflection	Reflection	P Value
Case 1	60%	50%	0.591
Case 2	50%	60%	0.193
Case 3	50%	50%	1.000

	1 kHz		
	No Reflection	Reflection	P Value
Case 1	50%	60%	0.726
Case 2	50%	30%	0.443
Case 3	50%	50%	1.000

	10 kHz		
	No Reflection	Reflection	P Value
Case 1	40%	60%	0.343
Case 2	50%	50%	1.000
Case 3	50%	50%	1.000

Table 2.5: Comparison of the percentage of correct responses with and without sound reflections (wide-band noises).

	Pink Noise		
	No Reflection	Reflection	P Value
Case 1	60%	30%	0.279
Case 2	40%	30%	0.726
Case 3	40%	40%	1.000

	Blue Noise		
	No Reflection	Reflection	P Value
Case 1	30%	50%	0.168
Case 2	60%	40%	0.343
Case 3	40%	60%	0.343

	White Noise		
	No Reflection	Reflection	P Value
Case 1	80%	60%	0.343
Case 2	60%	50%	0.726
Case 3	60%	50%	0.678

average percentage of correct responses for front and back distinctions under conditions where sound reflections are both absent and present. *Case 1* to *Case 3* in the result tables correspond to the experimental environments described in Section 2.4.1. *No Reflection* column indicates the average percentage of correct responses for front and back distinctions in each *Case* where sound reflections were absent. *Reflection* column indicates the average percentage of correct responses for front and back distinctions in each *Case* where sound reflections were present. The *P Value* represents the results of a two-tailed paired-sample t-test to compare the average percentages between the absence and presence of sound reflections in each *Case*.

The table did not show any significant differences at $p < 0.05$ in the percentage of correct responses for front-back distinctions across any sound sources. The effects arising from the differences in sound reflection positions on the horizontal plane alone did not seem sufficient to significantly reduce front-back confusion, even though some differences in front-back direction were perceived between the presence and absence of sound reflections. It also appeared that other effects of sound reflections beyond the precedence effect associated with arrival time differences and sound reflection positions might reinforce directional perception. It is anticipated that combining additional parameters of sound reflections along with the effects evaluated in the first method of this experiment could help resolve front-back confusion. For instance, future research could explore the influence of sound reflections from other positions, such as those on the vertical plane excluded from this evaluation. Furthermore, the Interaural Time Difference (ITD) and Interaural Level Difference (ILD) of each sound reflection could play a key role in influencing the directional perception of sound sources. The proposed platform has the potential to analyze these parameters of sound reflection individually. Thus, further exploration of the optimal settings for sound reflection is necessary to achieve more accurate 3D sound localization using the proposed platform and various evaluation methods.

2.4.3 Discussion

There are perceptual differences in distinguishing front and back sound sources based on the positions of sound reflections. Additionally, the sound reflections influencing front-back perception vary according to the positions and types of sound sources. It is crucial to evaluate the effects of each sound reflection individually on directional perception taking into account the listener's environment and using appropriate parameter settings for sound reflections. The proposed platform offers the capability to analyze various sound reflections and their impact on directional perception by allowing flexible settings for each sound reflection. The experiments demonstrate the usefulness of the proposed platform in enabling the customization of sound reflections to influence the perception of sound direction. Whereas some perceptual differences between the presence and absence of sound reflections, significant improvement in front-back confusion was not achieved solely by manipulating the sound reflection positions on the horizontal plane based on the precedence effect. It appears that additional effects of sound reflections beyond the parameters evaluated in this experiment may enhance the directional perception of sound sources. Therefore, further exploration of optimal sound reflection settings is necessary for more accurate 3D sound localization using the proposed platform and various evaluation methods. For instance, other parameters such as upward and downward sound reflection directions, as well as the ITD and ILD of sound reflections, can also be analyzed using this platform. Ultimately, the platform can be applied in a wide range of 3D sound localization evaluations with sound reflections tailored to user preferences.

This study proposed a platform that can evaluate the effect of sound reflections on human perception of the three-dimensionality of sound, which aims to realize natural information presentation using intuitive in-vehicle auditory signals with 3D sound localization minimizing eye movement without looking aside. In a real driving environment, the driver's seat and window glass are located close to the driver, from which the initial reflected sound may affect the three-dimensional impression of the in-vehicle auditory signals. It is believed that this evaluation experiment has demonstrated the possibility that the proposed platform can evaluate the effects of such sound reflections on the three-dimensional sense of in-vehicle auditory signals, including the sense of forward and backward sound. It will also be possible to evaluate the effect of the initial sound reflections from the driver's seat and window glass on the three-dimensional sense of in-vehicle auditory signals by developing this evaluation experiment and setting the reflectance of the back and side wall of the proposed platform to that of cushions and glasses. In reality, transformations in the shape of the walls and increments in the number of reflections would be necessary to strictly reproduce the real driving environment. On the other hand, since initial reflections have a considerable influence on sound reflections, the current platform might be sufficient to evaluate the effect of sound reflections on the three-dimensional impression of in-vehicle auditory signals based on the real driving environment. The distance of dangers should also be understood intuitively by in-vehicle auditory signals, which are excluded from this evaluation to focus on directions of sound sources and sound reflections as the first step of the evaluation of sound reflections on directional perception. Since the current platform can evaluate the difference in the perception of different sound distances, additional evaluation is required to mount a feature realizing distance perception on intuitive in-vehicle auditory signals. The proposed platform should be implemented in a system including driving simulators so that the effects of stereoscopic perception and sound reflections on driving behavior based on various driving scenarios can also be considered. For instance, an event of a window opening is achieved by setting the reflection coefficient of the reflective wall corresponding to the window to 0 in the proposed platform.

2.5 Summary

This chapter introduced a platform designed to evaluate the impact of synthetic early reflection on 3D sound perception. The proposed platform allows users to freely adjust the parameters of sound reflections, which appear to influence the directional perception of sound sources and to evaluate the effects of these reflections in detail. The key contribution of this research is the development of an evaluation environment that enables the exploration of optimal sound reflection settings required for more precise 3D sound localization. Two experiments were conducted to demonstrate the usefulness of the platform. In the first experiment, the difference in front-back sound perception with and without sound reflections was assessed. This experiment revealed that the position of sound reflections impacted the front-back perception of sound source location, which demonstrated the need to analyze individual sound reflections based on the listener's environment. The results showed that the proposed platform contributes to this need by allowing the evaluation of various parameters, such as the positions of sound reflections. Thus, the first experiment successfully demonstrated the platform's ability to customize sound reflections to influence sound directional perception. However, the second experiment found that modifying sound reflection positions on the horizontal plane alone based on the precedence effect did not lead to a significant improvement in the front-back confusion rate. This suggests that additional effects of sound reflections beyond those evaluation in this experiment may be needed to enhance the directional perception of sound sources.

Further exploration of optimal settings for sound reflections is necessary for more accurate 3D sound localization in future works. The experiments were practiced focusing on the precedence effect, conversely, more exploration with the other methods for evaluation of sound reflection effects including detailed analysis of HRTF shapes. Also, other parameters, including the distance between a sound source and a listener, the positions of reflection walls, and the reflection coefficient were fixed in this experiment. An analysis of these parameters is needed for further findings on the impacts of sound reflection on 3D sound perception, which is allowed by the proposed platform. Moreover, the experiments used only sound reflections at horizontal angles toward sound sources. Sound reflections at a vertical angle also need to be evaluated for further verification of the impacts of sound reflections. The platform proposed in this chapter presents a rectangular room with six reflector walls causing sound reflections, where a sound source is located and one early reflected sound is generated on each reflector wall. The further this platform is augmented, the more the effects of sound reflections can be evaluated. As for sound reflection, late reverberations should be implemented with the combination of early reflections to analyze human directional perceptions. Concerning the environment in the platform, the variable shape of reflector walls is feasible to assess various material properties with the settings of the reflection coefficient. This development is crucial to evaluating the effects of sound reflection assumed in the driving environment. Additionally, augmentation of the number of sound sources and sound reflections can be helpful for the analysis of various sound reflections and acoustic environments. Combining advanced technologies such as deep neural networks (DNN) can improve the platform by reproducing HRTFs and reflected sounds at high speed and accuracy.

Chapter 3

Designs of In-Vehicle Auditory Signals for Drivers Awareness Control

For safe driving, it is important for drivers to identify and respond to hazards around quickly. Intuitive in-vehicle auditory signals are required to communicate this hazardous information to drivers quickly and correctly. Moreover, when the information to be communicated has diversified from low-hazard to high-hazard in recent years, the traditional design with a monotonous structure of in-vehicle auditory signal may cause misperception of non-dangerous sounds as dangerous and discomfort for drivers in the stressful environment of driving. This chapter[1] explores designs for in-vehicle auditory signals to control driver awareness via danger. The proposed in-vehicle auditory signals are designed with musical and acoustic elements to intuitively inform a driver, especially of caution signals for dangerous situations and advisory signals to communicate potential dangers with comfortable sounds. The designed in-vehicle auditory signals are evaluated by questionnaire and principal component analysis to demonstrate that the designed auditory signals improve intuitive awareness and to explore the best settings of sound elements for intuitive in-vehicle auditory signals.

3.1 Introduction

Various efforts have recently been implemented in the real world to create a safer and more comfortable driving environment. In recent years, advances in electronic computer systems, such as Advanced Driver Assistance Systems (ADAS) [105, 106] and

[1]This chapter is based on "In-Vehicle Auditory Signals for Driver Awareness through Danger Perception" [I1] listed in International Conference Papers (Refereed) of Publications in this dissertation by the same author, which appeared in the Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA), Copyright(C)2020 IEICE. The material in this paper was presented in part at the Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA) [I1], and all the figures reused from [I1] in this paper are under the permission of the IEICE.

autonomous driving systems [16], have led to an increasing variety of information to be communicated to drivers. The emphasis is on improving safety by conveying information according to the level of risk and providing a comfortable sense in stressful situations including driving. On the other hand, ADAS mainly introduces communication methods based on visual information by displays with more suitable fixation of eye locations [107, 108], which may lead to over-reliance on visual details and increase the cognitive load on the driver. Against this background, in-vehicle auditory signals that provide auditory information about the surrounding environment have been attracting attention for several decades. In driving environments where visual attention is focused on the driving task, in-vehicle auditory signals that can communicate information to drivers without obstructing the field of view are especially important.

In the mobilities domain, in-vehicle auditory signals have been increasingly important for the creation of a safe and comfortable driving environment. However, these signals face challenges in that nearly identical auditory cues are used in different situations, even at varying levels of danger [41, 42]. The key problems and the goal of in-vehicle auditory signals are illustrated in Fig. 3.1. The left figure indicates traditional warning signals which usually emit a simple beep. These simple beep sounds often cannot effectively communicate the intended information to drivers. Since beep sounds are designed to feel quite dangerous to attract the attention of drivers, there may be discrepancies in information transmission if information that is not highly dangerous is conveyed by such overstimulated sounds. Moreover, these auditory signals can mislead and delay decision-making, potentially leading to traffic accidents. Therefore, designing in-vehicle auditory signals that are intuitively tailored to different driving scenarios due to danger levels is essential to improving safe and comfortable perceptions. This chapter explores designs for in-vehicle auditory signals that intuitively control driver awareness via danger perception. As shown in the right figure of Fig. 3.1, redesigned in-vehicle auditory signals composed of music theory and acoustic properties, called caution and advisory signals, can improve overall responses and safety of drivers with an intuitive perception of current danger levels.

The contributions of this chapter are as follows:

1. It is discovered of the appropriate components of sound elements for designing in-vehicle auditory signals to control intuitive driver awareness based on danger perception.
2. Designed in-vehicle auditory signals can be easily distinguished for danger perception through tonal diversification as a result of principle component analysis.

The structure of the remaining chapter is outlined as follows: Section 3.2 delves into relevant literature and studies. Section 3.3 introduces in-vehicle auditory signals classified into several categories. It then explores sound elements suitable for in-vehicle auditory signals based on human perceptions of sounds. Section 3.4 addresses experimental scenarios to show the effectiveness of the work of this chapter and the results are discussed from the perspective of analyzing the first and second principal components.

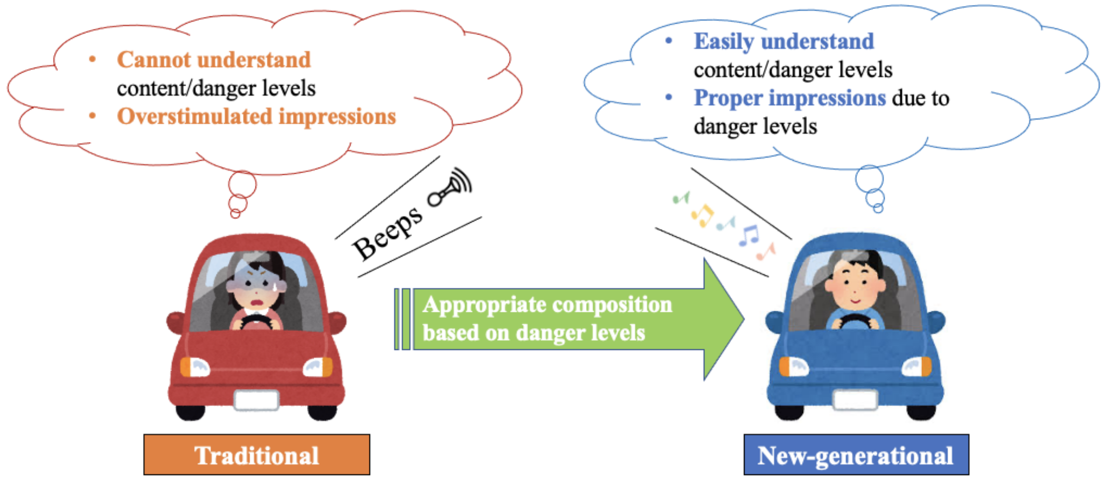


Figure 3.1: An overview of this research subject to redesign in-vehicle auditory signals with a composition based on music theory and acoustic properties.

Section 3.5 finally concludes this chapter with future remarks.

3.2 Related Research

Auditory signals carry a variety of information from objects to people, including characteristics and impressions of the object [109]. The authors in [39] have surveyed various auditory signals in consumer electronics and user interfaces, and showed that these auditory signals mostly used beep sounds which had not changed for at least 10 years as of the year 2016. The work in [39] recommended using melody, voice, and synthetic tones to receive various impressions from sounds to improve awareness.

For vehicular systems, collision avoidance systems have attracted public attention in auditory signals [110]. Many works have highlighted warnings for accident hotspots, including intersections, merges, etc. The authors in [111] attempted to propose a combination system with visual presentation and warning sounds designed according to the National Highway Traffic Safety Administration (NHTSA). However, even though the warning tone was newly designed based on NHTSA, it was not much different from the conventional monotone beep. In addition, it is not easy for drivers to pay attention simultaneously to visual information and monotonous warning sounds, resulting in cognitive burden and stress for drivers. Rather, several studies have reported that in-vehicle auditory signals are more effective and safer than visual supports since drivers do not use their eyes to look at anything other than the driving controls [112, 113]. Many studies have searched proper sounds for in-vehicle auditory signals from current in-vehicle auditory signals in [40, 114, 115].

Several studies, such as [42], have even emphasized the need for universal designs of in-vehicle auditory signals, which focused on the effectiveness of various sound elements. Additionally, some researchers have proposed platforms for assessing in-vehicle auditory signals, as seen in works by [116]. However, limited research has been conducted on the auditory aspect, particularly regarding how specific auditory signals can communicate clear impressions to drivers who can intuitively understand the driving situation. In-vehicle auditory signals, especially, should be designed in alignment with the danger levels since vehicle safety systems must be able to define specific hazardous situations and suggest appropriate subsequent actions [117]. The work in [41] also highlighted the need for a radical overhaul and the development of a next-generation framework for in-vehicle auditory signals, incorporating diverse sound elements such as music theory and acoustic characteristics to adapt to dynamic driving scenarios where the level of danger fluctuates continuously.

3.3 In-Vehicle Auditory Signals for Danger Perception

First of all, this section classifies various in-vehicle auditory signals used in current driving according to the degree of danger for the exploration of the sound elements that could express the impression of each danger level. Then, this research selected sound elements that have been proven at some general level to enhance the impression including brightness, anxiety, etc. [62, 118, 119], related to the information to be conveyed according to each danger level, and redesigned new in-vehicle auditory signals.

3.3.1 Classification of In-Vehicle Auditory Signals

This research categorizes various in-vehicle auditory signals into three groups—*Warning*, *Caution*, and *Advisory*—based on the perceived level of danger with the intended information. Table 3.1 provides a summary of the representative in-vehicle auditory signals organized into these three groups according to the corresponding danger levels.

Warning categorizes in-vehicle auditory signals intended to be used immediately before an accident. These auditory signals have a high risk of prompting the driver to take action preparing for an accident. For instance, *Auto-brake* feature activates automatically when a vehicle is on the verge of a collision. In addition, some in-vehicle auditory signals referred to as *Near collision* are triggered when the vehicle approaches a potential crash.

In-vehicle auditory signals categorized in *Caution* indicate a moderate level of danger and are employed in situations that could lead to an accident, which prompts the driver to take evasive action. For instance, when a vehicle loses or strays out of its lane, in-vehicle auditory signals labeled *Lane lost* and *Lane deviation* will activate to warn the driver and help prevent these hazardous maneuvers. Auto-cruise provides a function that automatically manages the gas pedal and brakes to assist with driving. When a

Table 3.1: Classification of commonly used in-vehicle auditory signals.

Levels of danger	Types of in-vehicle auditory signals
Warning (High)	Auto-brake
	Near collision
Caution (Middle)	Auto-cruise sudden lost
	Lane lost
	Lane deviation
Advisory-reminder (Low)	Reminder of seat belt
	Reminder of turning off light
	Reminder of forgotten key
	Reminder of half-opened door
	Reminder of inspection period
	Reminder of oil pressure drop
	Reminder of decrease in cooling water
	Reminder of decrease in gasoline
	Reminder of decrease in electric charge
	Reminder of driving more than 2 hours

vehicle is detected ahead, this function maintains a following distance safely. If a potential collision is imminent, *Auto-cruise sudden lost* activates to alert drivers temporarily disabling the auto-cruise function.

Advisory signals are characterized by a low level of danger and serve to avert potential accidents. Many *Advisory* signals share similar characteristics functioning primarily as reminders for drivers, such as urging them to use seat belts or notifying them of a decline in gasoline. Because *Advisory* signals are activated more frequently than *Warning* or *Caution* signals, these in-vehicle auditory signals are designed to be pleasant to listen to repeatedly.

3.3.2 Tone Designs for Intuitive In-Vehicle Auditory Signals

Firstly, this research excludes *Warning* signals from the tone design for in-vehicle auditory signals. These signals are meant to be as prominent as possible to instantly alert drivers to danger, which means leaving little room for improvement regarding the impression of danger. Consequently, the research focus is on designing in-vehicle auditory signals for *Caution* and *Advisory* from Table 3.1, including signals for *Auto-cruise sudden lost*, *Lane lost*, *Lane deviate*, and *Reminder of seat belt*. It is important to note that this research chose *Reminder of seat belt* from *Advisory* signals because, of all *Advisory* signals, it is the most frequent during driving and shares similar characteristics with the others. Table 3.2 presents a summary of the designed in-vehicle auditory signals in this research and their requirements. Each proposed in-vehicle auditory sig-

Table 3.2: Requirements for designed in-vehicle auditory signals: *NC*, *NLL*, *ND*, and *NB* are the abbreviations of newly designed auditory signals, *Auto-cruise sudden lost*, *Lane lost*, *Lane deviation*, and *Reminder of seat belt*, respectively.

Types	Intended expressions
NC	Unusually shifted from normal to abnormal conditions
NLL	A feeling that a driver gets lost
ND	A driver faces the danger of being out of the lane
NB	Remind a driver to wear a seat belt with more comfortably

Table 3.3: Feature summaries of designed in-vehicle auditory signals.

Types	BPM / Tempo name	Melody features
NC	104 / Allegretto	Transition from stable *Maj* to unstable *aug*
NLL	144 / Allegro	Quite dark *dim*, Repeated ascent/descent
ND	99 / Allegretto	*Maj* to dark *min*, *Maj* is a 6-note tension chord
NB	87 / Moderato	*Maj* descent, Less 1 kHz fundamental frequency

nal is crafted using musical elements and acoustic features summarized in Table 3.3 as described below[2], which have been proven at some general level to enhance the impression [62, 118, 119] including brightness, anxiety, etc.

The newly designed *Auto-cruise sudden lost* referred to as *NC* is intended to convey the sudden transition of the auto-cruise function from a normal operation to an abnormal state. The sound transitions from a C major (*CMaj*) chord, which is stable and bright, to a C augmented (*Caug*) chord, which is more unstable. This shift between stability and instability reflects the change from normal to abnormal conditions. Unlike the tone design of traditional *Auto-cruise sudden lost* including all harmonics heightening the perception of danger, the newly designed signal omits a part of harmonics to moderate the perceived level of danger. The tempo is set at 104 beats per minute (BPM) categorized in *Allegretto* meaning slightly faster.

The new *Lane lost* signal labeled *NLL* is designed to evoke the feeling of being lost. The tone design alternates between ascending and descending tones using the C diminished (*Cdim*) chord, which has a more unstable and darker structure than a major and minor chord. This movement between darker tones mirrors the sensation of losing one's way or lane. The tempo is set at 144 BPM classified as *Allegro* meaning a fast speed.

The newly created *Lane deviation* signal labeled *ND* represents the danger of veering out of a lane. The tone design alternates from the G major (*GMaj*) chord with its bright and stable structure to the F minor seventh-five (*Fm7-5*) chord which is darker and more tense. The irregular melody and rhythm represent the driver's anxiety. The use of a six-

[2]The newly designed in-vehicle auditory signals were composed by Dai1 Studio.

note tension chord in *GMaj* adds a sense of tense to amplify the perception of danger. A normal chord consists of three or four notes, but a tension chord includes more notes to form the chord. The tempo is set at 99 BPM, falling under the *Allegretto* category and suggesting a slightly faster speed.

Reminder of seat belt labeled *NB* is designed to remind drivers to wear their seat belts with a sound that is more comfortable compared to other in-vehicle auditory signals, which is classified as *Advisory* signals with the lowest danger level. This signal builds on the actual sound of wearing a seat belt and is easily recognizable due to added musical elements and acoustic characteristics. The base chord is C major (*CMaj*) which is bright and stable to provide a calm and non-threatening auditory cue. The descending melody reinforces a sense of calmness, as opposed to the tension evoked by ascending notes. The tempo is set at 88 BPM classified as *Moderato*. This tempo is slowest among the designed in-vehicle auditory signals to further enhance the perception of calmness. Additionally, whereas most traditional in-vehicle auditory signals have a fundamental frequency of 2 kHz or higher, the newly designed *Reminder of seat belt* uses a frequency below 1 kHz as higher frequencies tend to be perceived as unpleasant for low-danger signals. The volume peak is also reduced from over -7.1 integrated LUFS in traditional in-vehicle auditory signals to -21.2 integrated LUFS in the new design to enhance the perception of calmness. The Loudness Unit Full Scale (LUFS) is a unit that measures the signal level corrected for human hearing sensations and *integrated* expresses how loud it is on average in the LUFS. Furthermore, a longer envelope has been added to the proposed in-vehicle auditory signal for a comfortable and smoother auditory experience.

3.4 Experiments

3.4.1 Setup

To demonstrate the significant difference between traditional and newly designed in-vehicle auditory signals in terms of danger awareness, this research conducts an experiment using a questionnaire with impression rating scales. Twenty-four participants (16 males and 8 females in their 20s) listened to both traditional and newly designed in-vehicle auditory signals through headphones in an anechoic room. Although differences in hearing ability and age of the participants need to be analyzed, the research subject can be more complex and obscure in the early stages of exploring the elements of effective auditory signals, and this is a topic for future work. The in-vehicle auditory signals are played as many times as requested by the participants who then rate their impressions using a 7-point Likert scale. The questionnaire items shown in Table 3.4 are selected through a trial study and consisted of 16 items. Participants assign a score from one to seven for each item. This research summarizes and analyzes the questionnaire results using principal component analysis (PCA) to reduce the multivariate data and present the key findings in a visually accessible form.

Table 3.4: Rating scales used in the questionnaire.

Scale names	Likert scale	Counter scale names
Powerful	1-2-3-4-5-6-7	Not enough
With alarm sense		Without alarm sense
Strong		Weak
Safe		Dangerous
Luxury		Cheap
Colorful		Uncharacteristic
Rich		Skinny
Tense		No Tense
Thick		Thin
Quiet		Noisy
Calm		Shrill
Restful		Irritable
Comfort		Discomfort
Spacious		Cramped
Sharp		Dull
Large		Small

A total of nine in-vehicle auditory signals are used in this experiment. Table 3.5 represents the examined auditory signals for comparison, which consists of *Warning* signal used as a baseline, four traditional in-vehicle auditory signals, and designed auditory signals *Caution* and *Advisory* respectively. The newly designed auditory signals do for more than five seconds, whereas the traditional in-vehicle auditory signals sustain for less than three seconds. This is because if the intent is to convey fine nuances to the driver, the signal duration should be a bit longer.

3.4.2 Results

The experimental analysis by PCA resulted in Table 3.6, which presents the eigenvectors derived from the analysis of in-vehicle auditory signals. The first principal component accounts for 42.6% of the total explained variance of traditional in-vehicle auditory signals. The newly designed in-vehicle auditory signals exhibited similar first principal component to the traditional one. In this research, eigenvalues with absolute values greater than 0.3 are regarded as significant values. The eigenvalues for the dimensions *Tense - No tense*, *With alarm sense - Without alarm sense*, and *Safe - Dangerous* were notably higher compared to other eigenvectors. These components seem to capture information related to danger, which leads to the definition of the first principal component as the *Danger* evaluation principal component with values along the horizontal axis representing the degree of danger associated with each in-vehicle auditory signal.

Table 3.5: Types of in-vehicle auditory signals used in the experiment and their classification according to the danger levels.

Warning (as baseline)	Warning(W)
Caution of traditional auditory signals	Old auto-cruise sudden lost (OC)
	Old lane lost (OLL)
	Old lane deviation (OD)
Caution of proposed auditory signals	New auto-cruise sudden lost (NC)
	New lane lost (NLL)
	New lane deviation (ND)
Advisory	Old reminder of seat belt (OB)
	New reminder of seat belt (NB)

The second principal component for traditional in-vehicle auditory signals is mainly associated with dimensions such as *Powerful - Not enough*, *Strong - Weak*, *Luxury - Cheap*, *Large - Small*, and *Thick - Thin*. In contrast, it is more related to *Colorful - Uncharacteristic*, *Rich - Skinny*, and *Thick - Thin* in the newly designed in-vehicle auditory signals. This suggests that traditional in-vehicle auditory signals tend to be more monotonous and distinguished primarily by the strength, whereas the newly designed ones represent relatively greater diversity incorporating various melodies and tones. As a result, the second principal component can be interpreted as representing timbre appearing monotonous in traditional in-vehicle auditory signals and various in the newly designed in-vehicle auditory signals.

Fig. 3.2 and Fig. 3.3 visualize the PCA results where the x-axis corresponds to danger level and the y-axis indicates timbre based on the previous analysis. The red circle represents a cluster of *Warning* (high danger), the green circle represents *Caution* (medium danger), and the blue circle represents *Advisory* (low danger) respectably in these figures. In Fig. 3.3, participants can distinguish between the three types of in-vehicle auditory signals (*Warning*, *Caution*, and *Advisory*) in the newly designed in-vehicle auditory signals, whereas no such distinction is evident in the traditional ones shown in Fig. 3.2. Due to the variety of tonal features in the designed in-vehicle auditory signals, participants could intuitively identify the auditory impressions of *Warning*, *Caution*, and *Advisory*. The *Warning* and *Advisory* signals correspond to higher and lower danger levels with each horizontal axis value. The *Caution* signals overlap with both *Warning* and *Advisory* signals but are distinguishable by their timbre, which helped participants differentiate between the types of in-vehicle auditory signals.

3.4.3 Discussion

From the principal component analysis results, this research evaluated the improvement in the perception of in-vehicle auditory signals. Fig. 3.2 illustrates that the three types of

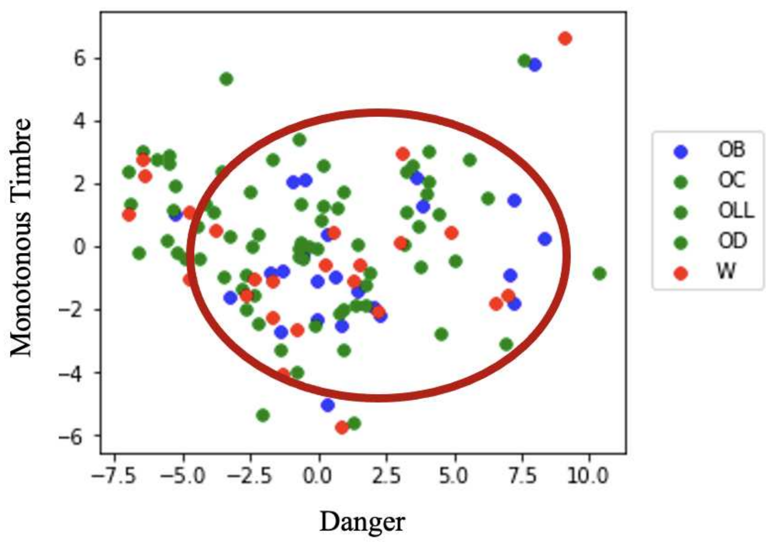


Figure 3.2: Results of traditional in-vehicle auditory signals.

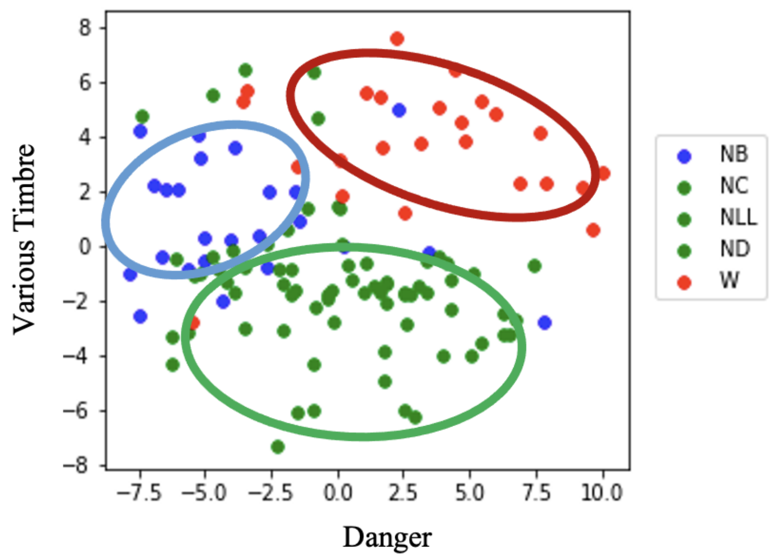


Figure 3.3: Results of proposed in-vehicle auditory signals.

Table 3.6: Eigenvector of the principal component analysis.

	Old		New	
Principal component	Danger	Monotonous	Danger	Various
Tense - No Tense	**0.365**	-0.119	**-0.398**	0.034
With alarm sense - Without alarm sense	**0.362**	-0.025	**-0.386**	-0.027
Safe - Dangerous	**-0.342**	0.063	**0.321**	-0.119
Spacious - Cramped	-0.292	-0.121	**0.320**	0.178
Calm - Shrill	-0.265	-0.189	**0.319**	0.193
Powerful - Not enough	0.205	**-0.488**	-0.164	0.271
Strong - Weak	0.180	**-0.427**	-0.194	0.297
Luxury - Cheap	-0.090	**-0.372**	0.081	0.175
Large - Small	0.133	**-0.354**	-0.121	0.289
Thick - Thin	-0.162	**-0.326**	0.083	**0.386**
Colorful - Uncharacteristic	-0.116	-0.252	0.077	**0.469**
Rich - Skinny	-0.154	-0.254	0.095	**0.407**
Quiet - Noisy	-0.291	-0.033	0.276	-0.170
Restful - Irritable	-0.287	-0.048	0.277	-0.110
Comfort - Discomfort	-0.279	-0.079	0.233	-0.165
Sharp - Dull	0.234	0.054	-0.268	-0.206
Explained variance	**0.426**	**0.138**	**0.394**	**0.211**

traditional in-vehicle auditory signals are indistinguishable with no clear separation in terms of danger level. This means that the traditional in-vehicle auditory signals do not effectively convey the degree of danger of information. In contrast, Fig. 3.3 illustrates a clear separation of the newly designed in-vehicle auditory signals into three categories: *Warning*, *Caution*, and *Advisory*, corresponding to their intended danger levels. This separation suggests that the newly designed in-vehicle auditory signals improve the perception of danger by offering a more intuitive differentiation with the various tones between degrees of danger, unlike the traditional in-vehicle auditory signals dominated by monotonous beep tones. This improvement can be attributed to the careful designs of musical elements and acoustic features aligned with the type of information being conveyed.

For a deeper discussion, the result of *Safe-Dangerous* scale is extracted from the questionnaire with the average values of the Likert scale, shown in Fig. 3.4. The horizontal scale indicates the sort of in-vehicle auditory signals, and the vertical scale indicates the scale 1 as *Safe* to 7 as *Dangerous*. The bar colors correspond to the results of the PCA: The red bar represents *Warning* (high danger), the green bars represent *Caution* (medium danger), and the blue bar represents *Advisory* (low danger) respectably. The results are analyzed using a two-tailed paired-sample t-test which showed the sta-

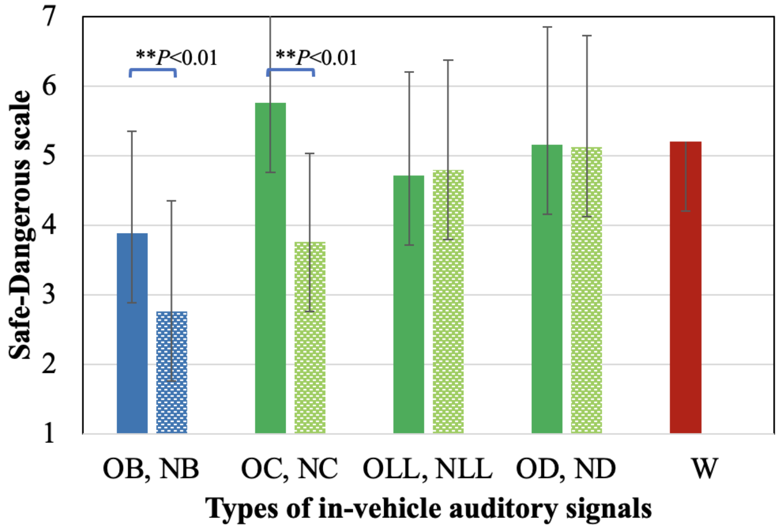


Figure 3.4: Results of averaged values of *Safe-Dangerous* scale.

tistically significant differences between traditional in-vehicle auditory signals and proposed ones, especially between *OB* and *NB*, *OC* and *NC* with $p < 0.01$. Proposed in-vehicle auditory signals produce a stepwise sense of danger based on danger levels.

The following discusses each musical element and acoustic feature focusing on *Reminder of seat belt* and *Auto-cruise sudden lost* signals. In terms of musical elements such as pitch scale and tempo, the traditional *Reminder of seat belt* had an upward progression with an unstable chord structure, whereas the new version featured a descending major chord progression. This change likely contributed to a clearer perception of danger in the *Advisory* signal. The new *Auto-cruise sudden lost* signal transitioned from a stable *CMaj* chord to an unstable *Caug* chord to create a sense of unease. This chord structure likely led to a stronger perception of danger compared to the *Advisory* signal for the *Reminder of seat belt*. Additionally, the slower tempo of the *Reminder of seat belt* advisory signal contributed to a reduced sense of urgency and danger, as slower tempos are generally perceived as less threatening. Regarding acoustic features, the newly designed in-vehicle auditory signals have a fundamental frequency below 1 kHz, whereas traditional in-vehicle auditory signals typically exceed 2 kHz. The volume of the new *Reminder of seat belt* is also lower at -21.2 integrated LUFS compared to -7.1 integrated LUFS for the traditional ones. Furthermore, the traditional *Auto-cruise sudden lost* contains louder volume and all harmonics contributing to a higher perception of danger due to its higher frequency. These findings suggest that the newly designed in-vehicle auditory signals are perceived as less dangerous. The improvement in danger perception between the traditional and newly designed *Auto-cruise sudden lost* signals may also be due to the absence of all harmonics and its longer envelope duration in the new version.

Table 3.7: Relationship between sound elements and perception of danger.

	Advisory	Caution	Warning
Spectral	Less 1 kHz	1 to 2 kHz	More 2 kHz
Volume	The lower the sense of risk, the smaller the volume		
Tempo (BPM)	The lower the sense of risk, the slower the tempo		
Envelope	The lower the sense of risk, the longer the envelope		
Chord structure	Maj	min, dim, aug	Dissonance, Tension
Melody	Descending	Ascending	Intermittent sound

The musical elements and acoustic features that may have contributed to better danger perception are summarized in Table 3.7 according to danger levels discussed so far. This table is a guideline example for a global design in designing in-vehicle auditory signals for different hazard levels. This research focused on the six sound elements that are often used to evoke different senses of sound based on the physiological effects that humans seem to have in common, such as a sense of chord structure and musical progression, and compiled them into a design guideline. Based on this current design guideline example, it is possible to create in-vehicle auditory signals that realize the function of judgment according to the degree of danger by using sound elements for the desired degree of danger. On the other hand, since this research is limited in terms of sound elements and scenarios, it is necessary to conduct evaluations based on other sound elements and actual driving scenarios. Furthermore, it is important to evaluate the interaction of various sound elements and the differences in the perception of danger in different contexts due to changes in the surrounding conditions. In particular, the more the sound element increases, the greater the danger is expected to be. By performing the above evaluation and sorting out the sound elements that are related or unrelated to danger, it will be possible to freely create in-vehicle auditory signals that match the atmosphere of the car while performing the function of judgment according to the degree of danger. Ultimately, it is believed that by organizing sound elements related or unrelated to hazards as a global standard guideline, it will be possible for humans and AI to create in-vehicle auditory signals that are appropriate for each situation and provide intuitive and natural information presentations.

In this research, designed in-vehicle auditory signals according to the degree of danger did not evaluate the relation between danger degrees of information and the length of the auditory signals. In fact, the length and composition of some designed in-vehicle auditory signals might be inappropriate due to feedback from the questionnaire. It is important to consider the length of sound according to the danger level of information. The signals should be short when the level of urgency is high and immediate driving action is required, whereas the signals can be long when the level of urgency is low and there is time to listen to the information. On the other hand, rich synthetic sounds with a lot of information are likely to sound unintelligible until drivers listen to them for a while. Especially in the initial stages when the user is not accustomed to the signals, it is highly likely that the user will have to concentrate on the signals. Thus, it is necessary to balance the familiarity of the signals and the length of the signals with the amount of information. If tones with different characteristics are assigned to each piece of information to be presented, the user will be able to respond to the signals in the moment drivers hear it once the drivers become accustomed to it. In particular, sounds with distinctive characteristics that are easily associated with the information to be presented are more likely to be retained in memory, and thus it should be possible to promote instantaneous judgment.

3.5 Summary

This chapter categorized the in-vehicle auditory signals based on the level of risk due to the content of information used in current driving. According to this category of in-vehicle auditory signals, samples of in-vehicle auditory signals were designed for hazardous perceptions. The experiments evaluated the dangerous impression of the designed in-vehicle auditory signals compared to the traditional in-vehicle auditory signals. The results showed that it is possible to add dangerous impressions according to the danger level of the in-vehicle auditory signals by designing acoustic characteristics and musical elements according to the content of the information and its danger level. The musical elements and acoustic features that may have contributed to better danger perception are summarized according to danger levels. These summaries are a guideline example for a global design in designing in-vehicle auditory signals for different hazard levels.

In this research, designed in-vehicle auditory signals according to the degree of danger did not evaluate the relation between danger degrees of information and the length of the auditory signals. The length and composition of some designed in-vehicle auditory signals might be inappropriate due to feedback from the questionnaire. It is important to consider the length of sound according to the danger level of information. Moreover, rich synthetic sounds with a lot of information are likely to sound unintelligible until drivers listen to them for a while, especially in the initial stages when the user is not accustomed to the signals. On the other hand, sounds with distinctive characteristics that are easily associated and accustomed with the information to be presented are more likely to be retained in memory and prompt immediate response. Thus, it is necessary to balance the familiarity of the signals and the length of the signals with the amount of information. In addition, this research focused on the sound elements based on the physiological effects that humans seem to have in common, such as a sense of chord structure and musical progression. Since sound perceptions involved in the context of cultures and memory can be effective in enhancing danger perception, these elements should be analyzed for intuitive in-vehicle auditory signals as the future perspective. It is also important to evaluate the interaction of various sound elements and the differences in the perception of danger in different contexts due to changes in the surrounding conditions. In particular, it is expected that the more sound elements are added, the greater the increase in danger level. By performing the above evaluation and sorting out the sound elements that are related or unrelated to danger, it will be possible to freely create in-vehicle auditory signals that match the atmosphere of the car while performing functions according to the degree of danger. Furthermore, the danger level was divided into three categories—*Warning*, *Caution*, and *Advisory*—based on the perceived level of danger according to the design purposes of each in-vehicle auditory signal to confirm the possibility of sound-based representation of the gradual increase in danger level in this research. On the other hand, the number of categories may be incremented depending on the design purpose of safety systems including in-vehicle auditory signals. In some cases, it would also be beneficial to seamlessly and continuously increase the sense of urgency. As a future direction of this research, the best definition of the danger levels should be explored and organized according to driving scenarios. More evaluation experiments also need to be conducted with greater diverse individuals of various ages and hearing abilities for more general results. Ultimately, by organizing sound elements related or unrelated to hazardous perceptions as a global standard guideline, it will be possible for humans and AI to create in-vehicle auditory signals that are appropriate for each situation and provide intuitive and natural information presentations.

Chapter 4

Generation Methods of In-Vehicle Auditory Signals for Multiple Events

In a driving environment where the surroundings are constantly changing, various events happen simultaneously at various hazard levels. In-vehicle auditory signals need to communicate multiple events with drivers according to the level of danger in the driving situations. This chapter[1] proposes an in-vehicle auditory signal to distinguish between multiple potential collision risks due to focusing on pedestrians. Conventionally, 3D sound localization and tone designs for various hazards allow drivers to distinguish between multiple sounds easily and simultaneously recognize pedestrians from various angles. However, if several sounds emanate from almost the same direction, drivers cannot detect multiple pedestrians. Proposed methods address this problem by varying the frequency and timing of the in-vehicle auditory signal so that drivers can distinguish multiple risks potentially colliding by Time-To-Collision (TTC) which is the time before a driver may collide with a pedestrian. The Experiment results in a significant improvement of the proposed methods compared to the conventional in-vehicle auditory signal due to timing shift rather than frequency shift.

4.1 Introduction

Driving operations take place in a constantly changing environment. Various events occur simultaneously at various levels of danger, which causes great stress on the drivers and adverse effects on driving operations. Safety support systems are imperative to

[1]This chapter is based on "Frequency and Timing Shifting of 3D-localized Advisory Signals for Awareness of Pedestrians" [I2] listed in International Conference Papers (Refereed) of Publications in this dissertation by the same author, which appeared in the Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA), Copyright(C)2020 IEICE. The material in this paper was presented in part at the Proceedings of International Workshop on Smart Info-Media Systems in Asia (SISA) [I2], and all the figures reused from [I2] in this paper are under the permission of the IEICE.

reduce these burdens on drivers while communicating the simultaneous risks occurring in the surroundings [120]. In particular, it is necessary to communicate multiple events with drivers through in-vehicle auditory signals without burdening the driver's vision focused on driving operations.

Recent studies have attempted various ways to facilitate more intuitively distinguishing information by in-vehicle auditory signals. For instance, focusing on tonal design subdivided into various sound components is studied to distinguish in-vehicle auditory signals according to the danger levels in driving surroundings [41, 42], as introduced in Chapter 3. In addition, several works in [54, 55] implemented in-vehicle auditory signals by using 3D sound localization, which is the method introduced in Chapter 2, to enable the determination of the direction and distance of surrounding hazards. However, according to these studies, multiple pieces of information can be difficult to distinguish based solely on the effects of 3D sound localization or slight differences in tonal design such as pitch, repetition rate, or volume.

An in-vehicle auditory signal that foreshadows danger within a low level of risk has also been proposed from [54]. In this case, simultaneous incidents to be communicated may happen frequently since low-risk events are far more numerous than high-risk ones. Moreover, the distinction of multiple risks by in-vehicle auditory signals, especially existing in almost the same direction that can be difficult for in-vehicle auditory signals to identify through 3D localized sound, has not been discussed in related research.

Fig. 4.1 illustrates an overview of this research. A car indicated in red attempts to turn right at a T-junction. On the other hand, an obstruction exists as a blind spot where several pedestrians attempt to cross or are crossing pedestrian crossings indicated by a green circle, which is not visible to the red car. This figure also shows the driver's original view as an orange line and the actual view blocked by the obstacle as a blue line and a fan shape. When events categorized as the same level of low danger exist in almost the same direction as this figure shows, it is difficult to convey multiple dangerous incidents to drivers using only 3D localized sound or tonal design techniques based on the level of danger, as has been done recently for in-vehicle auditory signals. Such in-vehicle auditory signals may cause confusion and delay in judgment, potentially leading to traffic accidents. This is the situation that needs to be resolved to achieve an intuitive in-vehicle auditory signal.

This chapter proposes generation methods of in-vehicle auditory signals for multiple events. The proposed in-vehicle auditory signals utilize methods of shifting the frequency and timing of signals to distinguish various events. The experiment evaluates proposed in-vehicle auditory signals in case pedestrians exist in the location of the blind spot of the driver.

The contributions of this chapter are as follows:

1. This research presents generation methods to communicate to drivers with multiple low-level risks such as the presence of pedestrians in almost the same direction.

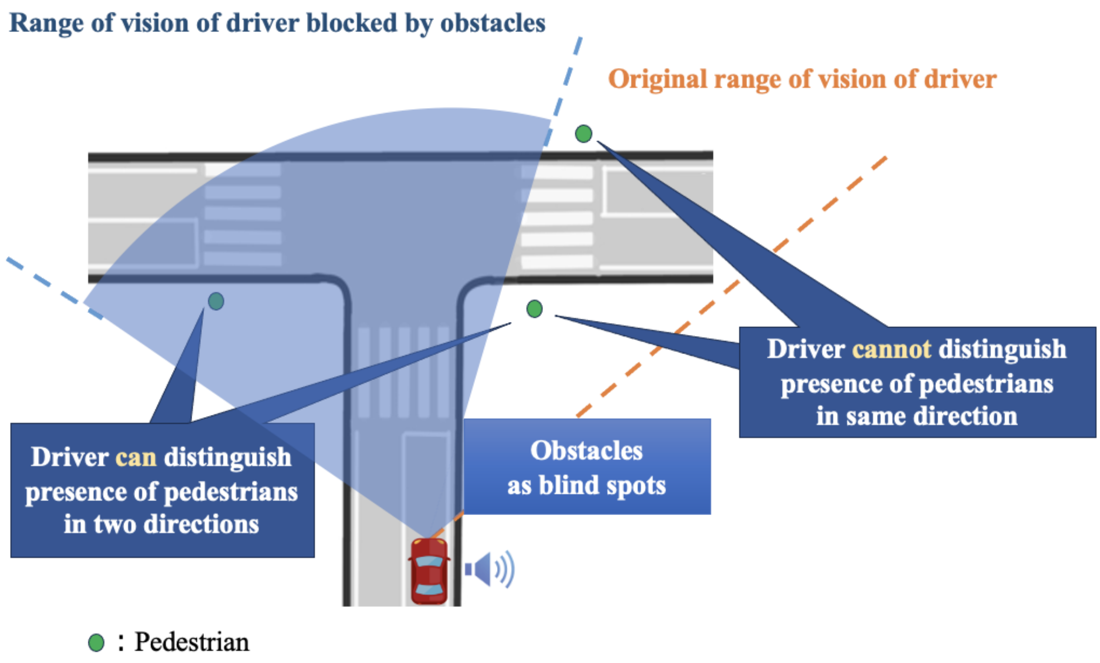


Figure 4.1: An overview of this research subject with a concrete driving scenario.

2. Proposed in-vehicle auditory signals are designed with different frequency spectrums and sound onset timing according to the number of pedestrians. Experimental results using questionnaires demonstrate that the proposed methods show a significant improvement over the conventional signals.

The subsequent section of this chapter is structured as follows. Section 4.2 discusses related work according to the context of in-vehicle auditory signals in general use-case and mobility domains. Sections 4.3 introduces a system of in-vehicle auditory signals, it then presents the proposed methods to determine the frequency spectrum and the sound onset timing of in-vehicle auditory signals. Section 4.4 addresses experimental scenarios to show the effectiveness of proposed methods and the results are discussed from the perspective. Section 4.5 finally concludes this chapter with future remarks.

4.2 Related Research

Differences in the arrival direction and tone designs of sound sources are often used as cues for humans to distinguish between multiple sounds. The methods to realize these cues have already been discussed with detailed mechanisms and related works in Chapter 2 and Chapter 3. This section introduces related research using both methods involved in the arrival direction and tone designs of sound sources for in-vehicle auditory signals.

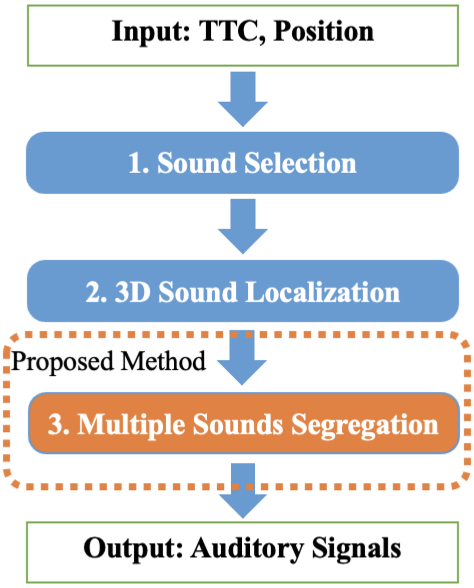
Tan and Lerner [55] conducted an extensive study on how subjects perceived warning sounds in a driving environment by incorporating directional information and sound impressions using multiple speakers. Their work as one of the earliest studies on sound in cars introduced the concept of in-vehicle auditory signals with directional and tonal cues. Similarly, the work in [54] introduced the safety assistant systems of in-vehicle auditory signals equipped with 3D sound and tone designs based on the risk level of surrounding incidents, which shows that 3D sound and tonal design support drivers judge the priority to address surrounding incidents in driving with the direction and danger level. However, this study stated that it is difficult to convey multiple events simultaneously by solely 3D sound and tonal design with slight volume and frequency differences. These studies also highlight several challenges related to using 3D sound localization in vehicles.

The factors, including ITD, ILD and HRTFs, play a key role in human perception of sound direction as mentioned in Chapter 2. Whereas these cues allow listeners to distinguish left and right directions, they struggle with front-back and top-bottom localization. This phenomenon called the cone of confusion and front-back confusion discussed in Chapter 2 as well has been a long-standing problem and remains a major challenge in embedded systems, especially in limited computational resources such as automotive systems. The work in [54] discussed that front-back confusion is one of the factors causing difficulties in conveying multiple events simultaneously by solely 3D sound. The research in Chapter 2 also addressed solving this long-standing problem by incorporating sound reflections in the real world and introduced a platform for analyzing the impact of sound reflections on 3D sound localization. A study including this approach would solve front-back confusion one day.

Distinguishing multiple risks in almost the same direction using 3D-localized sound also can be challenging. In addition, there may frequently be multiple risks to be communicated, such as pedestrians crossing a crosswalk or attempting to cross, especially if in-vehicle auditory signals foreshadow dangerous incidents at a low-risk level as proposed in [54]. In such cases, it is also difficult to distinguish between multiple risks by different tonal designs according to the degree of danger because these risks are the same or close to the degree of danger. However, no methods have been reported to allow drivers to distinguish these multiple risks in almost the same direction by in-vehicle auditory signals.

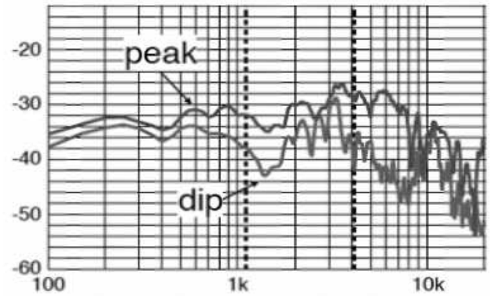
4.3 In-Vehicle Auditory Signal System

An in-vehicle auditory signal system is outlined in Fig. 4.2. The process flow of this system is in three steps as shown in Fig. 4.2a from the input of TTC and position of danger to the output of generalized in-vehicle auditory signals. This research assumes that one's own vehicle called as *ego-vehicle* can acquire relative information regarding its

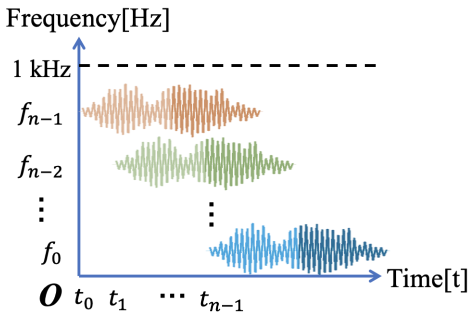


(a) The process flow.

(b) Select from each hazard level.



(c) Localize with HRTFs [4].



(d) Shift frequency and timing.

Figure 4.2: Proposed in-vehicle auditory signal system model.

position and the position of Vulnerable Road Users (VRUs). VRUs include pedestrians, cyclists, and others who are more prone to accidents or injuries compared to typical road users. Additionally, this research assumes that the ego-vehicle can gather this information through its sensing mechanisms and from shared data via infrastructure systems. Based on this information, the proposed in-vehicle auditory signal system triggers the auditory alert with the following three key features.

- Sound Selection: The proposed system selects one of three types of auditory signals according to the hazard level shown in Fig. 4.2b.
- 3D Sound Localization: The signals are spatially oriented to help the driver identify the direction of the hazard by HRTFs shown in Fig. 4.2c.
- Multiple Sounds Segregation: The proposed system handles multiple in-vehicle auditory signals simultaneously by shifting the frequency and onset timing of each

signal shown in Fig. 4.2d to ensure that each signal is distinct and identifiable even in complex scenarios involving several VRUs.

This framework is designed to enhance effectively the driver's ability to assess and respond to varying levels of danger.

4.3.1 Sound Selection

Firstly, this system selects an appropriate in-vehicle auditory signal from three types—warning, caution, and advisory—as shown in Fig. 4.2b. These in-vehicle auditory signals are specifically designed based on the Time-To-Collision (TTC) which refers to the time remaining between the ego-vehicle and a Vulnerable Road User (VRU) before a potential collision if assuming the current speed is maintained.

The implications of the length of TTC are defined along with the necessary actions according to the hazard levels. The previous work in Chapter 3 explored designs for in-vehicle auditory signals with musical and acoustic elements according to the hazard levels. Building on these findings, the design outlines for in-vehicle auditory signals based on musical and acoustic elements have been summarized in Table 3.7. This table provides a framework for designing appropriate in-vehicle auditory signals that reflect the hazard levels, which allows users to tailor the sound characteristics of in-vehicle auditory signals to match the severity of the situation effectively.

4.3.2 3D Sound Localization

After setting the in-vehicle auditory signals based on the danger levels, these signals are spatially localized as 3D sounds to incorporate the direction and distance of VRUs relative to the ego-vehicle. The system assumes the usage of binaural audio processed with Head-Related Transfer Functions (HRTFs). Usually, sounds reach a listener with delay and pressure differences between the two ears through reflections and absorption by the head, auricles, and torso. These factors allow the listener to locate the position of a sound source, and these detailed effects have been explained in Chapter 2. In 3D sound localization, positional information is granted to artificial sound sources by virtually reproducing these spatial effects by human bodies, whose characteristics are modeled as HRTFs [74] shown as the frequency in Fig. 4.2c.

Low-computation techniques for 3D sound localization have been developed, such as those in [4] explained in Section 2.3.3, which provide generic HRTFs suitable for embedded systems. These methods are particularly beneficial for implementation in environments with spatial and computational limitations, such as vehicles. Furthermore, there are also sound reflections in the interior space of vehicles due to several special shapes and materials, including window glasses and cushions. These sound reflections can affect the perception of the three-dimensional sound of the driver. The previous study in Chapter 2 focuses on the use of a generic HRTF low-computation generation method suitable for embedded systems to deal with sound reflection effects including the vehicle cabin and to achieve more realistic 3D sound. This approach aims to establish a 3D sound environment tailored to automotive applications to enhance spatial awareness in real-time driving scenarios.

4.3.3 Multiple Sounds Segregation

Generally, multiple risk incidents occur on the road while driving. For this reason, in-vehicle auditory signals ought to effectively convey multiple risk incidents to drivers simultaneously. This research focuses on two methods for achieving this subject: Frequency shifting and timing shifting, which are believed to enhance sound source segregation and feasible to easily distinguish between multiple risks using in-vehicle auditory signals, as supported by [121]. This research presents methods to shift the frequency and timing of in-vehicle advisory signals for different risks, as shown in Fig. 4.2d. It is thought that both frequency shifting and timing shifting methods contribute to effective sound source segregation and are feasible to easily identify multiple auditory warnings for drivers in a complex driving environment. Combined, these methods result in a more sophisticated and perceivable system for simultaneously alerting drivers to multiple hazards. Firstly, the method of frequency shifting is explained, followed by the method of timing shifting in each section.

Frequency Shifting Method

First, the method of frequency shifting for in-vehicle advisory signals is defined in Eq. 4.1, Eq. 4.2, and Table 4.1.

$$f = a \times b \times 2^{\frac{N}{12}} \quad (N = 4, 5, 7, 9, 12), \tag{4.1}$$

$$ID = y \times \frac{n - x}{n}, \tag{4.2}$$

which employs the equal temperament dividing the octave, referred to as the interval between the fundamental frequency and its double frequency, into 12 equal parts. For advisory signals that are non-urgent, the goal is to use harmonic and bright structures (e.g., major chords) rather than dissonant or minor structures. In light of these considerations, harmonic frequencies are selected using Eq. 4.1. For instance, suppose the fundamental frequency a of the basic advisory signal is 131 Hz represented by *C2* in musical notation, harmonic frequencies such as *E2, F2, G2*, and *A2* are then determined without dissonance and the octave difference. The proposed formula for harmonic frequency shifting is based on equal temperament, where the multiplication of 2 to the $N/12$ power provides the harmonic frequency. The parameter N represents the harmonic order, and an octave difference is calculated by doubling the frequency at each

Table 4.1: IDs of candidate frequencies and musical notes are determined for shifting of advisory signals of 131 Hz represented by *C2* of musical notes.

ID	Frequency (Hz)	Musical notes
0	131	C2
1	165	E2
2	175	F2
3	196	G2
4	220	A2
5	262	C3
6	330	E3
7	349	F3
8	392	G3
9	440	A3
10	523	C4
11	659	E4
12	698	F4
13	784	G4
14	880	A4

octave, represented by the integer b. The resulting candidate frequencies with an upper limit of 1 kHz for advisory signals are listed in Table 4.1 per the advisory signal design from Table 3.7.

The ID assigned for each candidate from the number 0 to the maximum number minus one is set for frequency-shifted advisory signals by Eq. 4.2. In this equation, y is defined as the total number of candidate frequency, n denotes the total number of risk incidents around the ego-vehicle, and x indicates the order of risk incidents around the ego-vehicle. Because of the need to simultaneously manage the design of non-urgent advisory signals and the segregation of multiple sounds, the frequency should be determined between the lowest possible frequency and as far away from the fundamental frequency of the basic advisory signal as possible. Higher frequencies are more likely to be perceived as dangerous and are assigned to risk factors that result in shorter TTC, which is represented on the vertical axis of Fig. 4.3. Each signal shown as from *signal 1* to *signal n* is allocated the frequency of f_0 to f_{n-1} in Fig. 4.3. Each color of sound wave in Fig. 4.3 also shows the frequency differences.

Taking the situation in which four pedestrians exist as shown in Fig. 4.1 as an example, the nearest pedestrian is determined by using Eq. 4.2 and Table 4.1 as ID 11 by rounding down the value. In this case, y defined as the total number of frequency candidates is determined in 15, n defined as the total number of risk incidents around the ego-vehicle is determined in 4, and x defined as the order of risk incidents around the ego-vehicle is determined in 1 in Eq. 4.2.

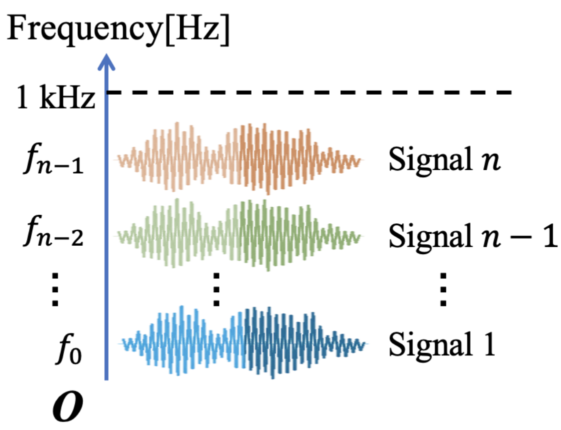


Figure 4.3: An illustration of frequency shifting method.

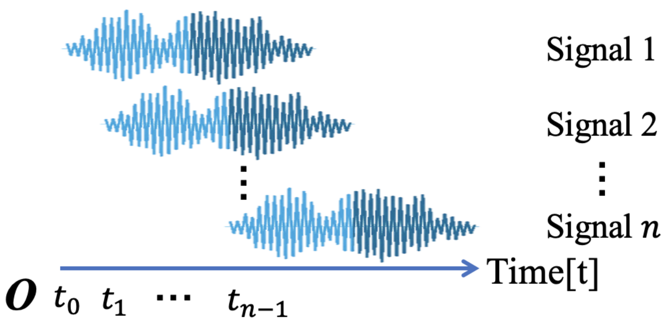


Figure 4.4: An illustration of timing shifting method.

Timing Shifting Method

In this method, onset times of advisory signals are shifted based on beat intervals. If there is an onset of sound at the same time as the beat, it becomes difficult to isolate multiple sounds. Therefore, this method staggers the onset of in-vehicle auditory signals between beat intervals to create separation.

The method for timing shifting of advisory signals is explained in Eq. 4.3.

$$t_x = \frac{60}{BPM} \times \frac{x-1}{n}, \tag{4.3}$$

where *BPM* denotes the signal's tempo in beats per minute, x indicates the order of risk incidents around ego-vehicle, and n represents the total number of risk incidents around ego-vehicle. The shorter the TTC, the earlier the onset is allocated for the in-vehicle auditory signal, which is reflected on the horizontal axis of Fig. 4.4. Each signal shown as from *signal 1* to *signal n* is allocated the frequency of t_0 to t_{n-1} in Fig. 4.4.

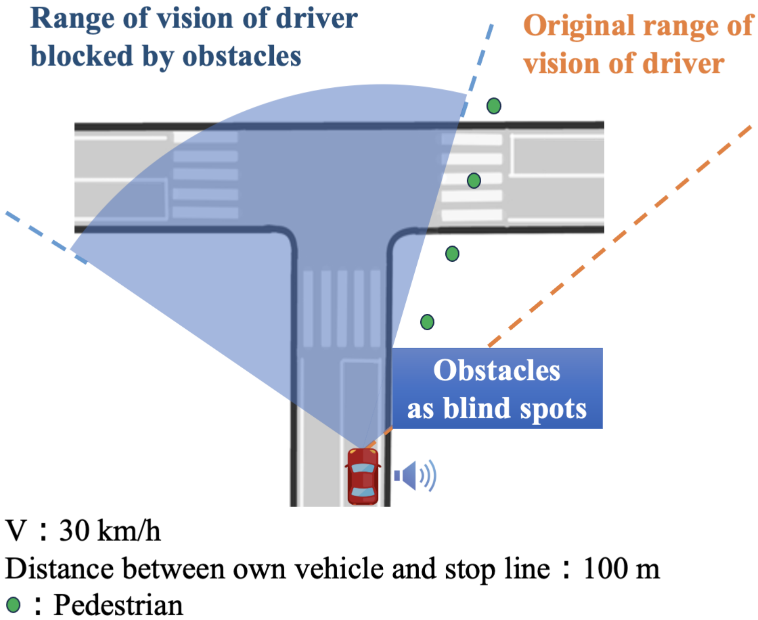


Figure 4.5: A driving scenario in the experiment.

4.4 Experiments

4.4.1 Setup

Experiments are conducted to demonstrate the effectiveness of the proposed methods. The experiment evaluates that the proposed methods are feasible recognizing the number of pedestrians in almost the same direction by achieving a high percentage of correct responses. This experiment demonstrates that multiple pedestrians in almost the same direction can be distinguishable by the proposed methods. Only pedestrians among VRUs are treated as relatively stationary objects for simplicity in this experiment.

This experiment simulates a scenario where a driver approaches a T-junction with poor visibility and faces the risk of colliding with multiple pedestrians as depicted in Fig. 4.5. This scenario assumes that the ego-vehicle is equipped with sensing capabilities and can share information about potential risks via infrastructure. The ego-vehicle is driving at the legal speed limit of 30 km/h and 100 m away from pedestrians crossing near the T-junction, where it begins to decelerate gradually.

In this scenario, Time-to-Collision (TTC) is used, defined by the following risk levels.

- $TTC_{warning}$ is within 5 seconds indicating close proximity to pedestrians.
- $TTC_{caution}$ is within 10 seconds marking the beginning of deceleration as the vehicle approaches pedestrians.
- $TTC_{advisory}$ is within 20 seconds covering the early stages of deceleration and

Table 4.2: Sound elements for advisory level excerpted from Table 3.7.

	Advisory
Spectral	Less 1 kHz
Volume	The smaller the volume
Tempo (BPM)	The slower the tempo
Envelope	The longer the envelope
Chord	Maj
Melody	Descending

increases of awareness of pedestrians.

Due to obstacles obstructing the view of the driver in the ego-vehicle, the system tries to convey risks via in-vehicle auditory signals based on TTC to alert the driver to pedestrians in the blind spot. The driver unaware of pedestrians receives auditory cues indicating various levels of danger based on TTC. Fig. 4.5 depicts pedestrians as perfectly aligned in a particular direction, but there is some misalignment in reality, which only emphasizes the fact that drivers perceive that the risk factors are present to some extent in the same direction.

This experiment compares four types of signals in advisory levels used in Chapter 3 as well. The tonal features of advisory signals are excerpted with sound elements from Table 3.7 to Table 4.2, and the four types of advisory levels using the proposed methods are summarized below.

1. **Baseline signal**: A fundamental frequency of *C2* with a descending progression and 88 BPM, which are same characteristics as newly designed *Reminder of seat belt* of *Advisory* signals in Table 3.3, and 3D-localized to append the direction of pedestrians on the signal. The signal volume should be 3 dB louder as the pedestrian's TTC is shorter. These features are also common across all other signal conditions.
2. **Frequency shifted signal**: This signal shifts frequency according to the pedestrians' TTC.
3. **Timing shifted signal**: This signal shifts timing according to the pedestrians' TTC.
4. **Frequency and timing shifted signal**: This signal combines frequency and timing shifting according to the pedestrians' TTC.

The experiment evaluates participants' ability to segregate auditory signals based on the number of pedestrians with the number ranging from two to four in their blind spot. Each participant listens to twelve signals, each lasting approximately three seconds, and answers a questionnaire by selecting from four figures showing one to four pedestrians as in Fig. 4.5. The one pedestrian option is included to add confusion, even though two to four pedestrian signals are presented. In the experiment, the usefulness of the proposed methods is measured by the percentage of correct answers given by the par-

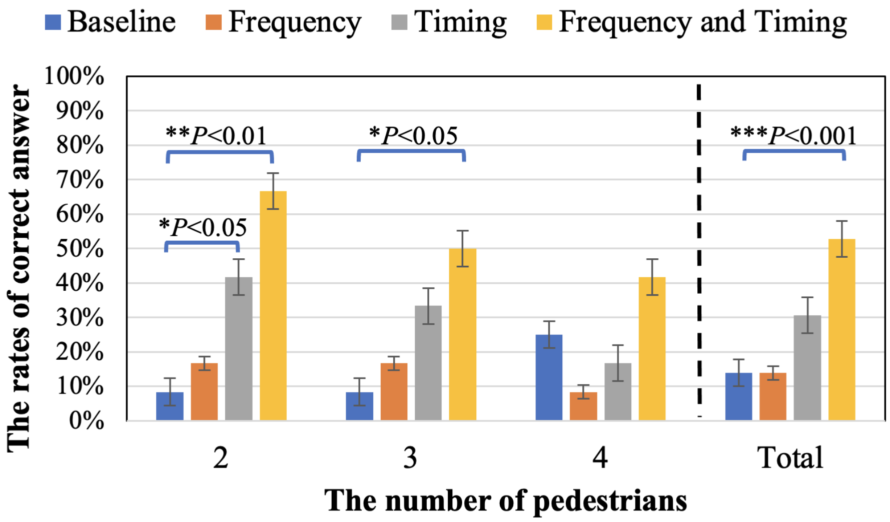


Figure 4.6: Results of the evaluation experiment.

ticipants. Participants can listen to each signal repeatedly during answering played for approximately 3 seconds at once. To familiarize participants with the structure of the sound, a baseline signal, i.e., a sound without multiple sound sources, can be heard before the experiment begins. The order of sound types and pedestrian numbers is randomized, and only the signal type is explained to participants before listening. The experiment involves twelve male subjects in their twenties using headphones, all with normal hearing. Therefore, differences in hearing ability are not considered in the analysis for simplifying, which is a topic for future work same as Chapter 3.

4.4.2 Results

The results of the experiment are evaluated using a two-tailed paired-sample t-test as shown in Fig. 4.6. In this figure, the x-axis represents the number of pedestrians, whereas the y-axis represents the rate of correct answers. The total score on the x-axis separated by dotted lines aggregates all the scores regardless of the number of pedestrians. The results are color-coded to indicate the different in-vehicle auditory signal types: Blue represents the baseline signal, orange represents the frequency-shifting signal, gray represents the timing-shifting signal, and yellow represents the signal with both frequency and timing shifting. The results indicate statistically significant differences in the correct answer rates between the baseline and the other signals. The significance is shown by the p value, where one asterisk represents $p < 0.05$, two asterisks represent $p < 0.01$, and three asterisks represent $p < 0.001$ respectively.

Firstly, there are significant improvements in correct response rates when using frequency and timing shifting compared to the baseline. The combination of both frequency and timing shifting indicates the most significant improvement ($p < 0.001$) in the total scores across two to four pedestrians, which suggests that combining both effects results in higher accuracy. Additionally, significant differences are observed for two pedestrians ($p < 0.01$) and three pedestrians ($p < 0.05$), although there is no significant difference for four pedestrians. This implies that as the number of signals increases, signal segregation becomes more challenging, leading to a decrease in the correct response rate due to the complexity of the tones.

Secondly, there is a significant difference in correct response rate ($p < 0.05$) for two pedestrians between the timing-shifted signals and the baseline, but no significant difference between the frequency-shifted signals and the baseline. These suggest that timing shifting is more effective than frequency shifting when used separately, though frequency shifting complements timing shifting when combined as shown in the results of in-vehicle auditory signals using both methods. However, the correct response rate does not show significant differences for three and four pedestrians, which further supports the idea that the complex tones hinder signal segregation as the number of signals increases. The signals using both frequency and timing shifting show better performance compared to those using only timing shifting, particularly for three pedestrians and in the total score. This implies that a combination of frequency and timing shifting is more effective than using either method alone.

4.4.3 Discussion

The results showed that the timing-shifting method is more effective than the frequency-shifting method. Additionally, combined methods of frequency and timing shifting resulted in the most effectively distinguishing signals. These results imply that the frequency-shifting method is complementary to the timing-shifting method. These results was expected and demonstrated the usefulness of the proposed methods to some extent. This research limited the baseline signal to *Reminder of seat belt* proposed in Chapter 3, and evaluated the auditory perception of the signals based on frequency and onset timing differences. However, the degree to which sounds are discernible may vary depending on other sound sources and driving conditions. For instance, the effect of these methods may be varied for signals composed of various tones whose fundamental frequencies are difficult to determine, or for signals with relatively fast rhythms and various sound lengths. It may be necessary to switch the method to distinguish signals according to the sound elements of the sound sources.

The other ambient sounds such as talking voices and ambulance sirens or other car audio systems such as music and radio may also have an effect on sound discrimination, which is necessary to consider the ease of distinguishing signals from the surrounding environment, for instance, by adjusting the volume setting according to the ambient sound or reducing the volume of other sounds being played by the car audio system. By organizing sound elements that be related or unrelated to hazard levels in conjunction with the research of Chapter 3, it may be possible to create a in-vehicle auditory signal that has a natural impression that matches the atmosphere inside the car while maintaining the necessary information conveying function. Moreover, these methods may be limited to conveying information about two to three pedestrians. For scenarios with more pedestrians or additional complexity, further refinement of in-vehicle auditory signals would be necessary to convey more information simultaneously. Furthermore, this experiment was limited to evaluating pedestrians as VRUs. Further research is needed to assess performance in scenarios where various types of VRUs, such as cyclists on a sidewalk or other vehicles in a roadway, are mixed on real roads.

4.5 Summary

This chapter introduces methods for communicating the multiple risks of pedestrians in almost the same direction by in-vehicle auditory signals, which are difficult to distinguish in driving situations. These proposed methods shift the frequency and timing of in-vehicle auditory signals according to TTC. This research evaluated the proposed methods to compare the separation rates with the baseline which is not shifting the frequency and timing of the signals. The experimental results showed that the timing-shifting method is more effective than the frequency-shifting method. Additionally, combined methods of frequency and timing shifting resulted in the most effectively distinguishing signals. These results imply that the frequency-shifting method is complementary to the timing-shifting method. Meanwhile, these proposed methods are limited to communicating to drivers with two to three pedestrians.

Distinguishable differences should be evaluated according to sound elements of sound sources. Environmental sounds such as ambulance sirens and talking voices may also affect sound discrimination, and the ease of distinguishing signals from the surrounding environment should be considered. Other methods should be explored or evaluated in combination with the method proposed in this research to communicate more information simultaneously by in-vehicle auditory signals. Moreover, only pedestrians were evaluated as VRUs in this experiment. Since various VRUs are mixed on actual roads and constantly on the move, the effective method for multiple hazards with in-vehicle auditory signals can be varied depending on the situations. It is necessary to explore further and evaluate the best method for in-vehicle auditory signals in various driving scenarios. In addition, experiments with many more diverse individuals of various ages and hearing abilities are needed to validate with more general results.

Chapter 5

Conclusion and Future Work

The various design methodologies for intuitive in-vehicle auditory signals have been discussed so far. This chapter concludes with an overall summary of each chapter of this dissertation and the future work for the direction of this research discussed. In the conclusion, summaries of each chapter are provided in response to the research questions stated in the introduction. Future work describes the research directions beyond each chapter to achieve intuitive in-vehicle auditory signals.

5.1 Conclusion

The development of automobiles is required in various fields to decrease the fatality rate from traffic accidents with the creation of a safer and more comfortable driving environment. One of the critical approaches to safe and comfortable driving is in-vehicle auditory signals. Whereas visual-based support systems can lead to fatigue since drivers are required to continuously monitor their surroundings and follow visual cues simultaneously, the auditory signals are more effective in driving. The in-vehicle auditory signals are desired to have diverse designs to enhance driver awareness and facilitate an intuitive understanding of information accurately and immediately. Therefore, this dissertation presents design methodologies for intuitive in-vehicle auditory signals based on three functions that need to be mounted: directional perception, tone design, and supporting multiple events, which can contribute to serving as a foundation for such a universal design of ideal in-vehicle auditory signals to present information according to the degree of danger intuitively with various sound elements. To realize these intuitive in-vehicle auditory signals, this dissertation has addressed three research questions defined in Chapter 1.

5.1.1 Answer to Question 1: How could the direction of danger be communicated by in-vehicle auditory signals?

Sound reflections exist in automobiles because of several special shapes and materials including window glass and cushions. These reflections can influence localizing artificial sound sources in three dimensions, including in-vehicle auditory signals. Chapter 2 presented a design platform to evaluate the impact of synthetic early reflections on 3D sound localization and explored the best settings of sound reflections to improve the human directional perception of artificial sound sources, especially in the front and back positions, which aims to realize natural information presentation using intuitive in-vehicle signals with 3D sound localization minimizing eye movement without looking aside. The demonstration with two experiments evaluated the usefulness of the proposed platform focusing on the precedence effect involved in the impacts of sound reflections on human directional perception of sound sources. The first experimental results indicate that the impacts of sound reflections on the human directional perception of sound sources vary depending on the types of sound sources and positions of sound reflections. On the other hand, the second experimental results discovered that modifying sound reflection positions on the horizontal plane alone based on the precedence effect did not significantly improve the exact judgment of the front and back locations of the sound sources. These results suggest that additional effects of sound reflections beyond those evaluated in these experiments may be required to enhance the directional perception of sound sources. Throughout these experiments, the usefulness of the proposed platform was successfully demonstrated as the contribution of this chapter to assess the differences in the effects of sound reflections due to variations in the type and position of the sound sources with the proposed platform. Further exploration of optimal settings for sound reflections will create more accurate 3D sound localization with this platform, which will contribute to realizing in-vehicle auditory signals with 3D sound perception inside automobiles where special sound reflections exist.

In a real driving environment, the driver's seat and window glass are located close to the driver, from which the initial reflected sound may affect the three-dimensional impression of the in-vehicle auditory signals. It is believed that the evaluation experiment in this research has demonstrated the possibility that the proposed platform can evaluate the effects of such sound reflections on the three-dimensional sense of in-vehicle auditory signals, including the sense of forward and backward sound. It will also be possible to evaluate the effect of the initial sound reflections from the driver's seat and window glass on the three-dimensional sense of in-vehicle auditory signals by developing this evaluation experiment and setting the reflectance of the back and side wall of the proposed platform to that of cushions and glasses. In reality, transformations in the shape of the walls and increments in the number of reflections would be necessary to strictly reproduce the real driving environment. On the other hand, since initial reflections have a considerable influence on sound reflections, the current platform might be sufficient to evaluate the effect of sound reflections on the three-dimensional impression of in-vehicle auditory signals based on the real driving environment.

5.1.2 Answer to Question 2: How could in-vehicle auditory signals be composed for suitable perception depending on danger levels?

To ensure driving safety, quick identification and correct response to surrounding hazards are significant. Intuitive in-vehicle auditory signals are required to communicate quickly and correctly this hazardous information to drivers according to danger levels of surrounding incidents and comfort perception for drivers in the stressful driving environment. Chapter 3 explored designs of in-vehicle auditory signals to control driver awareness via danger perception. This research presented the categorization of current in-vehicle auditory signals based on the level of risk due to the content of information and samples of auditory signals were designed for the impression of each danger level. The experiments evaluated the impression of danger comparing the designed in-vehicle auditory signals to the traditional in-vehicle auditory signals. The results showed that the method of designing acoustic characteristics and musical elements according to the content of the information and its danger level was feasible to add dangerous impressions according to the danger level of the in-vehicle auditory signals. On the other hand, the length and composition of some designed in-vehicle auditory signals might be inappropriate due to feedback from this questionnaire. It is important to consider the length of sound according to the danger level of information. Future work is needed to balance the information content, frequency of occurrence, and stimulus of the in-vehicle auditory signals with acoustic characteristics and musical elements.

The musical elements and acoustic features that may have contributed to better danger perception are summarized according to danger levels. These summaries are a guideline example for a global design in designing in-vehicle auditory signals for different hazard levels. This research focused on the six sound elements that are often used to evoke different senses of sound based on the physiological effects that humans seem to have in common, such as a sense of chord structure and musical progression, and compiled them into a design guideline. Based on this current design guideline example, it is possible to create in-vehicle auditory signals that realize the function of judgment according to the degree of danger by using the sound elements for the desired degree of danger. On the other hand, since this research is limited in terms of sound elements and scenarios, it is necessary to conduct evaluations based on other sound elements and actual driving scenarios. Furthermore, it is important to evaluate the interaction of various sound elements and the differences in the perception of danger in different contexts due to changes in the surrounding conditions. In particular, the more the sound element increases, the greater the danger is expected to be. By performing the above evaluation and sorting out the sound elements that are related or unrelated to danger, it will be possible to freely create in-vehicle auditory signals that match the atmosphere of the car while performing functions of discrimination according to the degree of danger, beyond which it could lead to a global standard guideline for ideal in-vehicle auditory signals.

5.1.3 Answer to Question 3: How could the multiple simultaneous events be communicated by in-vehicle auditory signals?

More than one event is continuously occurring in driving environments. In this ever-changing situation where multiple events are occurring, a method and a framework are needed to effectively convey information with in-vehicle auditory signals. Chapter 4 introduced methods for communicating the multiple risks by in-vehicle auditory signals, which focused on the situation where pedestrians in almost the same direction are difficult to distinguish in driving. These proposed methods shifted the frequency and timing of in-vehicle auditory signals. The experiment evaluated the proposed methods to compare the separation rates with the baseline not shifting the frequency and timing of in-vehicle auditory signals, and resulted in the timing-shifting method being more effective than the frequency-shifting method. Additionally, the method combining frequency and timing shifting resulted in the most effective distinguishing signals, which implies the frequency-shifting method is complementary to the timing-shifting method. These results was expected and demonstrated the usefulness of the proposed method to some extent.

In the future, these proposed methods need to expand with other methods to communicate more information simultaneously by in-vehicle auditory signals because of the limitations of communicating to drivers with two to three pedestrians in this research. Further research is also needed to assess performance in scenarios where various types of VRUs are mixed on real roads. This research also limited the baseline signal to *Reminder of seat belt* proposed in Chapter 3. It may be necessary to switch the method to distinguish signals according to the sound elements of the sound sources. The other ambient sounds such as talking voices and ambulance sirens or other car audio systems such as music and a radio may also have an effect on the sound discrimination, which is necessary to consider the ease of distinguishing signals from the surrounding environment. By organizing sound elements that are related or unrelated to hazard levels in conjunction with the research of Chapter 3, it may be possible to create an in-vehicle auditory signal that has a natural impression that matches the atmosphere inside the car while maintaining the necessary information conveying function.

5.2 Future Work

The works presented in this dissertation will be able to extend in several directions.

Sound Design and Evaluation for Various Scenarios

This research focused on a limited number of sound elements in a limited number of driving scenarios. In order to realize a universal design of intuitive in-vehicle auditory signals, it is necessary to evaluate various sound elements and driving scenarios. It is important to balance the information content, frequency of occurrence, and stimulus of the in-vehicle auditory signals with sound elements, and balance the familiarity of the signals and the length of the signals with the amount of information. Ultimately, by organizing sound elements related or unrelated to sound hazards as a global standard guideline, it will be possible for humans and AI to create in-vehicle auditory signals that are appropriate for each situation and provide intuitive and natural information presentations.

Implementation and Evaluation in a Driving Simulator

A design platform to evaluate the impact of synthetic early reflection on 3D sound localization has been proposed in Chapter 2. This platform can be implemented in a driving simulator where it is possible to evaluate the effect of 3D sound localization with sound reflection on driving behaviors. Moreover, designed in-vehicle auditory signals by the sound elements explored in Chapter 3 and the proposed methods for sound segregation in Chapter 4 can also be evaluated in the driving simulator with the effect of 3D sound localization and sound reflection on driver's impressions and driving behaviors. In fact, the sound elements explored in Chapter 3 were evaluated with a driving simulator in a driving scenario by the work in [122]. As in this case, the evaluation needs to be based on various driving scenarios to verify the effective sound elements and 3D sound localization settings for intuitive in-vehicle auditory signals.

Implementation and Evaluation in a Real Driving

After the verification of the safety and usefulness of design methodologies for intuitive in-vehicle auditory signals in virtual environments such as a driving simulator, these methodologies ought to be implemented and evaluated in a real automobile. In the implementation of the proposed methodologies, the selection of appropriate acoustic devices ought to be considered, which produce the in-vehicle auditory signals. The points for choices of acoustic devices are position and performance, especially in automobiles where available spaces are limited for acoustic devices because other devices for driving systems occupy a lot of space. The performance of acoustic devices is also important needless to point out. Sound qualities ought to be guaranteed to convey the information with in-vehicle auditory signals. On the other hand, an acoustic device producing high-quality sounds tends to be bigger, and high-quality devices also might be expensive. The number of acoustic devices is also an important factor in the implementation. If more realistic sound quality is desired such as feeling of solidity, the number of acoustic devices needs to be incremented. For these reasons, it is necessary to carefully choose the equipped position of acoustic devices with the balance between distances from human ears, the range of occupation in automobiles, performance, device size, device numbers, and cost.

Experiments with a Large Number of Diverse Subjects

In every chapter of this dissertation, it is determined to disregard differences in the ages and hearing abilities of the participants because the research subjects can be more complex and obscure in the early stages of each research. This is a topic for future work to evaluate the differences in the ages and hearing abilities of the participants for universal designs of intuitive in-vehicle auditory signals as some works in [123]. In addition, the number of subjects in this dissertation is limited. A large number of diverse subjects are needed to verify more reasonable results for the design methodologies of intuitive in-vehicle auditory signals. Designing sounds for the main target, such as driving ability and cultural background, is also important in creating intuitive auditory signals. Throughout these extensive verifications, a universal design for intuitive in-vehicle auditory signals is expected to be realized.

Bibliography

[1] H. Ritchie, "More than a million people die from road injuries every year," *Our World in Data*, 2024.

[2] R. N. Charette, *How software is eating the car*. IEEE Spectrum, 2021.

[3] T. Fischer, M. Caversaccio, and W. Wimmer, "A front-back confusion metric in horizontal sound localization: The FBC score," in *Proceedings of ACM Symposium on Applied Perception (SAP)*, 2020, pp. 1–5.

[4] T. Matsumura, N. Iwanaga, W. Kobayashi, T. Onoye, and I. Shirakawa, "Embedded 3D sound movement system based on feature extraction of head-related transfer function," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 262–267, 2005.

[5] SAE, "Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles," [Accessed on 11/01/2024]. [Online]. Available: https://www.sae.org/standards/content/j3016_202104/

[6] IHME, "Global burden of disease (2024)," *Institute for Health Metrics and Evaluation (IHME)*, 2024.

[7] WHO, "Global status report on road safety 2023," *Technical Report of World Health Organization (WHO)*, 2023.

[8] H. E. Rosen, I. Bari, N. Paichadze, M. Peden, M. Khayesi, J. Monclús, and A. A. Hyder, "Global road safety 2010–18: An analysis of global status reports," *Injury*, 2022.

[9] T. Myklebust, T. Stålhane, G. D. Jenssen, and I. Wærø, "Autonomous cars, trust and safety case for the public," in *Proceedings of 2020 Annual Reliability and Maintainability Symposium (RAMS)*, 2020, pp. 1–6.

[10] V. K. Kukkala, J. Tunnell, S. Pasricha, and T. Bradley, "Advanced driver-assistance systems: A path toward autonomous vehicles," *IEEE Consumer Electronics Magazine*, vol. 7, no. 5, pp. 18–25, 2018.

[11] P. M. Greenwood, J. K. Lenneman, and C. L. Baldwin, "Advanced driver assistance systems (ADAS): Demographics, preferred sources of information, and accuracy of ADAS knowledge," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 86, pp. 131–150, 2022.

[12] B. A. Jumaa, A. M. Abdulhassan, and A. M. Abdulhassan, "Advanced driver assistance system (ADAS): A review of systems and technologies," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 8, no. 6, pp. 231–234, 2019.

[13] M. Hasenjäger, M. Heckmann, and H. Wersing, "A survey of personalization for advanced driver assistance systems," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 2, pp. 335–344, 2020.

[14] R. Ghosh, R. Pragathi, S. Ullas, and S. Borra, "Intelligent transportation systems: A survey," in *Proceedings of 2017 International Conference on Circuits, Controls, and Communications (CCUBE)*, 2017, pp. 160–165.

[15] P. Arthurs, L. Gillam, P. Krause, N. Wang, K. Halder, and A. Mouzakitis, "A taxonomy and survey of edge cloud computing for intelligent transportation systems and connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6206–6221, 2022.

[16] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58 443–58 469, 2020.

[17] A.-K. Frison, Y. Forster, P. Wintersberger, V. Geisel, and A. Riener, "Where we come from and where we are going: A systematic review of human factors research in driving automation," *Applied Sciences*, vol. 10, no. 24, pp. 1–36, 2020.

[18] P. K. Murali, M. Kaboli, and R. Dahiya, "Intelligent in-vehicle interaction technologies," *Advanced Intelligent Systems*, vol. 4, no. 2, pp. 1–27, 2022.

[19] W. Morales-Alvarez, O. Sipele, R. Léberon, H. H. Tadjine, and C. Olaverri-Monreal, "Automated driving: A literature review of the take over request in conditional automation," *Electronics*, vol. 9, no. 12, pp. 1–34, 2020.

[20] M. Colley, L. Gruler, M. Woide, and E. Rukzio, "Investigating the design of information presentation in take-over requests in automated vehicles," in *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI)*, 2021, pp. 1–15.

[21] P. Jansen, M. Colley, and E. Rukzio, "A design space for human sensor and actuator focused in-vehicle interaction based on a systematic literature review," in

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, 2022, pp. 1–51.

[22] F. Diederichs, A. Muthumani, A. Feierle, M. Galle, L.-A. Mathis, V. Bopp-Bertenbreiter, H. Widlroither, and K. Bengler, "Improving driver performance and experience in assisted and automated driving with visual cues in the steering wheel," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4843–4852, 2022.

[23] S. Kim, A. K. Dey, J. Lee, and J. Forlizzi, "Usability of car dashboard displays for elder drivers," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. ACM, 2011, pp. 493–502.

[24] M. Tonnis, C. Lange, and G. Klinker, "Visual longitudinal and lateral driving assistance in the head-up display of cars," in *Proceedings of 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007, pp. 91–94.

[25] K.-M. Chen, L.-L. Chen, and S.-T. Shen, "Development and comparison of a full-scale car display and communication system by applying augmented reality," *Displays*, vol. 29, no. 1, pp. 33–40, 2008.

[26] W. W. Gaver, "The SonicFinder: An interface that uses auditory icons," *Human–Computer Interaction*, vol. 4, no. 1, pp. 67–94, 1989.

[27] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human–Computer Interaction*, vol. 4, no. 1, pp. 11–44, 1989.

[28] S. S. Borojeni, T. Wallbaum, W. Heuten, and S. Boll, "Comparing shape-changing and vibrot-actile steering wheels for take-over requests in highly automated driving," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '17, 2017, pp. 221–225.

[29] M. Schwalk, N. Kalogerakis, and T. Maier, "Driver support by a vibrotactile seat matrix–recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving," *Procedia Manufacturing*, vol. 3, pp. 2466–2473, 2015.

[30] D. Dmitrenko, C. T. Vi, and M. Obrist, "A comparison of scent-delivery devices and their meaningful use for in-car olfactory interaction," in *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. Automotive'UI 16, 2016, pp. 23–26.

[31] G. Labiale, "In-car road information: Comparisons of auditory and visual presentations," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 34, no. 9, 1990, pp. 623–627.

[32] M. Wang, Y. Liao, S. L. Lyckvi, and F. Chen, "How drivers respond to visual vs. auditory information in advisory traffic information systems," *Behaviour & Information Technology*, vol. 39, no. 12, pp. 1308–1319, 2020.

[33] T. C. Ojsteršek and D. Topolšek, "Influence of drivers' visual and cognitive attention on their perception of changes in the traffic environment," *European Transport Research Review*, vol. 11, pp. 1–9, 2019.

[34] V. Kondyli, M. Bhatt, D. Levin, and J. Suchan, "How do drivers mitigate the effects of naturalistic visual complexity? On attentional strategies and their implications under a change blindness protocol," *Cognitive Research: Principles and Implications*, vol. 8, pp. 1–30, 2023.

[35] M. Bassani, L. Catani, A. Hazoor, A. Hoxha, A. Lioi, A. Portera, and L. Tefa, "Do driver monitoring technologies improve the driving behaviour of distracted drivers? A simulation study to assess the impact of an auditory driver distraction warning device on driving performance," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 95, pp. 239–250, 2023.

[36] M. Jeon, P. T. Lautala, C. Nadri, and D. N. Nelson, "In-vehicle auditory alerts literature review," *Technical Report of U.S. Department of Transportation, Federal Railroad Administration*, 2022.

[37] K. Groß-Vogt, M. Weger, R. Höldrich, T. Hermann, T. Bovermann, and S. Reichmann, "Augmentation of an institute's kitchen: An ambient auditory display of electric power consumption," in *Proceedings of the 24th International Conference on Auditory Display*, 2018, pp. 105–112.

[38] F. Bella and M. Silvestri, "Vehicle–pedestrian interactions into and outside of crosswalks: Effects of driver assistance systems," *Transport*, vol. 36, no. 2, pp. 98–109, 2021.

[39] S. H. Wake, M. Inoue, E. Kokubu, and Y. Kobayashi, "A survey of UI sounds used in consumer electronics and user awareness," in *Proceedings of 2016 IEEE 5th Global Conference on Consumer Electronics*. IEEE, 2016, pp. 1–3.

[40] C.-F. Chi, R. S. Dewi, and M.-H. Huang, "Psychophysical evaluation of auditory signals in passenger vehicles," *Applied Ergonomics*, vol. 59, pp. 153–164, 2017.

[41] A. Sigman and N. Misdariis, "Alarm/will/sound: Sound design, modelling, perception and composition cross-currents," *Organised Sound*, vol. 24, no. 1, pp. 54–70, 2019.

[42] K. Yamauchi, J.-D. Choi, R. Maiguma, M. Takada, and S. Iwamiya, "A basic study on universal design of auditory signals in automobiles," *Journal of Physiological Anthropology and Applied Human Science*, vol. 23, no. 6, pp. 295–298, 2004.

[43] A. D. Pierce, "Diffraction of sound around corners and over wide barriers," *The Journal of the Acoustical Society of America*, vol. 55, no. 5, pp. 941–955, 1974.

[44] K. F. Herzfeld, "Reflection of sound," *Physical Review*, vol. 53, no. 11, pp. 899–906, 1938.

[45] R. C. Chanaud, "Progress in sound masking," *Acoustics Today*, vol. 3, no. 4, pp. 21–26, 2007.

[46] G. Box, "Signal-to-noise ratios, performance criteria, and transformations," *Technometrics*, vol. 30, no. 1, pp. 1–17, 1988.

[47] J. Bérubé and C. F. J. Wu, "Signal-to-noise ratio and related measures in parameter design optimization: an overview," *Sankhyā: The Indian Journal of Statistics, Series B*, vol. 62, no. 3, pp. 417–432, 2000.

[48] D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," *The Journal of the Acoustical Society of America*, vol. 100, no. 1, pp. 408–427, 1996.

[49] N. J. Bullot and P. Égré, "Objects and sound perception," *Review of Philosophy and Psychology*, vol. 1, pp. 5–17, 2010.

[50] J. Wang, A. Pun, J. Tu, S. Manivasagam, A. Sadat, S. Casas, M. Ren, and R. Urtasun, "Advsim: Generating safety-critical scenarios for self-driving vehicles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 9909–9918.

[51] W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao, "A survey on safety-critical driving scenario generation—a methodological perspective," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 6971–6988, 2023.

[52] L. Bruck, B. Haycock, and A. Emadi, "A review of driving simulation technology and applications," *IEEE Open Journal of Vehicular Technology*, vol. 2, pp. 1–16, 2020.

[53] R. A. Wynne, V. Beanland, and P. M. Salmon, "Systematic review of driving simulator validation studies," *Safety Science*, vol. 117, pp. 138–151, 2019.

[54] M. Wang, S. L. Lyckvi, C. Chen, P. Dahlstedt, and F. Chen, "Using advisory 3D sound cues to improve drivers' performance and situation awareness," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 2814–2825.

[55] A. K. Tan and N. D. Lerner, "Acoustic localization of in-vehicle crash avoidance warnings as a cue to hazard direction," *Technical Report of U.S. Department of Transportation, National Highway Traffic Safety Administration*, 1996.

[56] L. Xun-xiang, Z. Duo, C. Dongyan, and C. Dingfang, "Implementation of 3D sound effect modeling technology based on vehicle driving simulator," in *Proceedings of 2008 9th International Conference on Computer-Aided Industrial Design and Conceptual Design*, 2008, pp. 63–68.

[57] R. Y. Litovsky and S. P. Godar, "Difference in precedence effect between children and adults signifies development of sound localization abilities in complex listening tasks," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 1979–1991, 2010.

[58] J. C. Middlebrooks, "Sound localization," *Handbook of Clinical Neurology*, vol. 129, pp. 99–116, 2015.

[59] E. Šabić, J. Chen, and J. A. MacDonald, "Toward a better understanding of in-vehicle auditory warnings and background noise," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 63, no. 2, pp. 312–335, 2021.

[60] M. Binelli, A. Venturi, A. Amendola, and A. Farina, "Experimental analysis of spatial properties of the sound field inside a car employing a spherical microphone array," in *Proceedings of 130th Audio Engineering Society Convention*, 2011.

[61] S.-K. Lee, G.-H. Lee, and J. Back, "Development of sound-quality indexes in a car cabin owing to the acoustic characteristics of absorption materials," *Applied Acoustics*, vol. 143, pp. 125–140, 2019.

[62] R. Panda, R. Malheiro, and R. P. Paiva, "Audio features for music emotion recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 68–88, 2023.

[63] ISO, "ISO 15006: Road vehicles — Ergonomics aspects of transport information and control systems — Specifications for in-vehicle auditory presentation," *International Organization for Standardization (ISO)*, 2011.

[64] B. Arons, "A review of the cocktail party effect," *Journal of the American Voice I/O society*, vol. 12, no. 7, pp. 35–50, 1992.

[65] N. A. Gumerov and R. Duraiswami, “Modeling the effect of a nearby boundary on the HRTF,” in *Proceedings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 2001, pp. 3337–3340.

[66] B. Rakerd and W. M. Hartmann, “Localization of sound in rooms, II: The effects of a single reflecting surface,” *The Journal of the Acoustical Society of America*, vol. 78, no. 2, pp. 524–533, 1985.

[67] D. Poirier-Quinot and B. F. G. Katz, “On the improvement of accommodation to non-individual HRTFs via VR active learning and inclusion of a 3D room response,” *Acta Acustica*, vol. 5, no. 25, pp. 1–17, 2021.

[68] L.-H. Lee, T. Braud, P. Y. Zhou, L. Wang, D. Xu, Z. Lin, A. Kumar, C. Bermejo, and P. Hui, “All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda,” *Foundations and Trends® in Human-Computer Interaction*, vol. 18, no. 2–3, p. 100–337, 2024.

[69] P. A. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt, “What is XR? Towards a framework for augmented and virtual reality,” *Computers in Human Behavior*, vol. 133, pp. 1–18, 2022.

[70] C. Wang and C.-W. Cheng, “Perceiving the world with hearing in virtual reality,” in *Proceedings of 2019 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW)*, 2019, pp. 1–2.

[71] J. Broderick, J. Duggan, and S. Redfern, “The importance of spatial audio in modern games and virtual environments,” in *Proceedings of 2018 IEEE Games, Entertainment, Media Conference (GEM)*, 2018, pp. 1–9.

[72] R. B. Masterton and T. J. Imig, “Neural mechanisms for sound localization,” *Annual Review of Physiology*, vol. 46, pp. 275–287, 1984.

[73] F. Zotter and M. Frank, *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature, 2019.

[74] J. R. Clarke and H. Lee, “The effects of decreasing the magnitude of elevation-dependent notches in HRTFs on median plane localization,” in *Proceedings of 142nd Audio Engineering Society Convention*, 2017.

[75] B. I. Bacila and H. Lee, “360° binaural room impulse response (BRIR) database for 6DOF spatial perception research,” *Proceedings of 146th Audio Engineering Society Convention*, 2019.

[76] N. Kolotzek, P. G. Aublin, and B. U. Seeber, "The effect of early and late reflections on binaural unmasking," in *Proceedings of Fortschritte der Akustik-DAGA'21*, 2021.

[77] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," in *Proceedings of the 2001 International Workshop on Spatial Media*, 2001.

[78] V. R. Algazi and R. O. Duda, "Headphone-based spatial sound," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 33–42, 2011.

[79] S. V. A. Garí, W. O. Brimijoin, H. G. Hassager, and P. W. Robinson, "Flexible binaural resynthesis of room impulse responses for augmented reality research," in *Proceedings of the EAA Spatial Audio Signal Processing Symposium*, 2019, pp. 161–166.

[80] D. K. Reed and R. C. Maher, "An investigation of early reflection's effect on front-back localization in spatial audio," in *Proceedings of 127th Audio Engineering Society Convention*, 2009.

[81] S.-N. Yao, "Headphone-based immersive audio for virtual reality headsets," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 3, pp. 300–308, 2017.

[82] L. Rayleigh, "XII. On our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.

[83] C. Guezenoc and R. Seguier, "HRTF individualization: A survey," *arXiv:2003.06183*, 2020.

[84] K. Iida, Y. Ishii, and S. Nishioka, "Personalization of head-related transfer functions in the median plane based on the anthropometry of the listener's pinnae," *The Journal of the Acoustical Society of America*, vol. 136, no. 1, pp. 317–333, 2014.

[85] S. Spagnol, "HRTF selection by anthropometric regression for improving horizontal localization accuracy," *IEEE Signal Processing Letters*, vol. 27, pp. 590–594, 2020.

[86] R. Bomhardt and J. Fels, "The influence of symmetrical human ears on the front-back confusion," in *Proceedings of 142nd Audio Engineering Society Convention*, 2017.

[87] R. Miccini and S. Spagnol, "HRTF individualization using deep learning," in *Proceedings of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 390–395.

[88] J. A. Belloch, M. Ferrer, A. Gonzalez, F. J. Martinez-Zaldivar, and A. M. Vidal, “Headphone-based virtual spatialization of sound with a GPU accelerator,” *Journal of the Audio Engineering Society*, vol. 61, no. 7/8, pp. 546–561, 2013.

[89] G. Ramos, M. Cobos, B. Bank, and J. A. Belloch, “A parallel approach to HRTF approximation and interpolation based on a parametric filter model,” *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1507–1511, 2017.

[90] J. A. Belloch, G. Ramos, J. M. Badia, and M. Cobos, “An efficient implementation of parallel parametric HRTF models for binaural sound synthesis in mobile multimedia,” *IEEE Access*, vol. 8, pp. 49 562–49 573, 2020.

[91] L. Thaler, G. M. Reich, X. Zhang, D. Wang, G. E. Smith, Z. Tao, R. S. A. B. R. Abdullah, M. Cherniakov, C. J. Baker, D. Kish, and M. Antoniou, “Mouth-clicks used by blind expert human echolocators - signal description and model based signal synthesis,” *PLOS Computational Biology*, vol. 13, no. 8, pp. 1–17, 2017.

[92] I. An, M. Son, D. Manocha, and S.-E. Yoon, “Reflection-aware sound source localization,” in *Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 66–73.

[93] N. W. Adelman-Larsen, E. R. Thompson, and A. C. Gade, “Suitable reverberation times for halls for rock and pop music,” *The Journal of the Acoustical Society of America*, vol. 127, no. 1, pp. 247–255, 2010.

[94] J.-M. Jot, “Efficient models for reverberation and distance rendering in computer music and virtual audio reality,” in *Proceedings of International Computer Music Conference (ICMC)*, 1997.

[95] A. Nykänen, A. Zedigh, and P. Mohlin, “Effects on localization performance from moving the sources in binaural reproductions,” in *Proceedings of International Congress and Exposition on Noise Control Engineering*, vol. 4, 2013, pp. 3193–3201.

[96] C. Pörschmann, P. Stade, and J. M. Arend, “Binauralization of omnidirectional room impulse responses-algorithm and technical evaluation,” in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, 2017, pp. 345–352.

[97] V. Gunnarsson and M. Sternad, “Binaural auralization of microphone array room impulse responses using causal wiener filtering,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2899–2914, 2021.

[98] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, “The precedence effect,” *The Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 1633–1654, 1999.

[99] J. P. A. Lochner and J. F. Burger, "The influence of reflections on auditorium acoustics," *Journal of Sound and Vibration (JSV)*, vol. 1, no. 4, pp. 426–448, 1964.

[100] J. Braasch, "Sound localization in the presence of multiple reflections using a binaurally integrated cross-correlation/auto-correlation mechanism," *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. EL143–EL148, 2016.

[101] T. Usagawa, A. Saho, K. Imamura, and Y. Chisaki, "A solution of front-back confusion within binaural processing by an estimation method of sound source direction on sagittal coordinate," in *Proceedings of TENCON 2011-2011 IEEE Region 10 Conference*, 2011, pp. 1–4.

[102] M. Frank and F. Zotter, "Simple reduction of front-back confusion in static binaural rendering," *Fortschritte der Akustik, DAGA*, 2018.

[103] O. Balan, A. Moldoveanu, and F. Moldoveanu, "A systematic review of the methods and experiments aimed to reduce front-back confusions in the free-field and virtual auditory environments," in *Proceedings of 15th International Conference on Human Computer Interaction, RoCHI 2018*, 2018, pp. 24–29.

[104] G. Parseihian and B. F. G. Katz, "Rapid head-related transfer function adaptation using a virtual auditory environment," *The ournal of the Acoustical Society of America*, vol. 131, no. 4, pp. 2948–2957, 2012.

[105] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[106] L. Li, D. Wen, N.-N. Zheng, and L.-C. Shen, "Cognitive cars: A new frontier for ADAS research," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 395–407, 2012.

[107] J. Xu, K. Guo, F. Menchinelli, and S. H. Park, "Eye fixation location recommendation in advanced driver assistance system," *Journal of Electrical Engineering & Technology*, vol. 14, pp. 965–978, 2019.

[108] M. Araki, H. Taniguch, and K. Tanaka, "The vehicular display," in *Proceedings of 2018 25th International Workshop on Active-Matrix Flatpanel Displays and Devices (AM-FPD)*. IEEE, 2018, pp. 1–3.

[109] S. Merchel and M. E. Altinsoy, "Psychophysical comparison of the auditory and tactile perception: A survey," *Journal on Multimodal User Interfaces*, vol. 14, pp. 271–283, 2020.

[110] F. Meng and C. Spence, "Tactile warning signals for in-vehicle systems," *Accident Analysis & Prevention*, vol. 75, pp. 333–346, 2015.

[111] B. Ryder, B. Gahr, P. Egolf, A. Dahlinger, and F. Wortmann, "Preventing traffic accidents with in-vehicle decision support systems-The impact of accident hotspot warnings on driver behaviour," *Decision Support Systems*, vol. 99, pp. 64–74, 2017.

[112] X. Wu, L. N. Boyle, D. Marshall, and W. O'Brien, "The effectiveness of auditory forward collision warning alerts," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 59, pp. 164–178, 2018.

[113] D. Bouis, M. Voss, G. Geiser, and R. Haller, "Visual vs auditory displays for different tasks of a car driver," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 23, no. 1, 1979, pp. 35–39.

[114] C.-F. Chi, R. S. Dewi, Y. Y. Surbakti, and D.-Y. Hsieh, "The perceived quality of in-vehicle auditory signals: a structural equation modelling approach," *Ergonomics*, vol. 60, no. 11, pp. 1471–1484, 2017.

[115] N. I. M. Zaki, S. M. C. Husin, M. K. A. Husain, N. A. Husain, A. Ma'aram, S. N. A. Marmin, A. F. Adanan, Y. Ahmad, and K. A. A. Kassim, "Auditory alert for in-vehicle safety technologies: a review," *Journal of the Society of Automotive Engineers Malaysia*, vol. 5, no. 1, pp. 88–102, 2021.

[116] L. M. Sparrer, S. Winkelmann, A. von Hoffmann, and D. S. Morillo, "A novel framework for the assessment of the impact of affective UI sounds in a driving situation through virtual reality and physiological signals," *Working Paper of Technische Hochschule Nürnberg Georg Simon Ohm*, pp. 1–7, 2023.

[117] ANSI, "Z535: American national standard for safety colors," *American National Standards Institute (ANSI)*, 1991.

[118] S. Bannister, "A survey into the experience of musically induced chills: Emotions, situations and music," *Psychology of Music*, vol. 48, no. 2, pp. 297–314, 2020.

[119] M. de Witte, A. Spruit, S. van Hooren, X. Moonen, and G.-J. Stams, "Effects of music interventions on stress-related outcomes: a systematic review and two meta-analyses," *Health Psychology Review*, vol. 14, no. 2, pp. 294–324, 2020.

[120] M. J. Lochner and L. M. Trick, "Multiple-object tracking while driving: the multiple-vehicle tracking task," *Attention, Perception, & Psychophysics*, vol. 76, pp. 2326–2345, 2014.

[121] R. Lipp, P. Kitterick, Q. Summerfield, P. J. Bailey, and I. Paul-Jordanov, "Concurrent sound segregation based on inharmonicity and onset asynchrony," *Neuropsychologia*, vol. 48, no. 5, pp. 1417–1425, 2010.

[122] K. Takii, F. Sawa, W. Kobayashi, S. Park, M. Citir, H. Nishikawa, I. Taniguchi, and T. Onoye, "Enhancing driver awareness of vulnerable road users through in-vehicle auditory signals," in *Proceedings of 2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, 2024, pp. 1–6.

[123] J. Y. An, Y. J. Kim, and H. S. Yoo, "User preference for vehicle warning sounds to develop AUI guideline focusing on differences between sex and among age groups," in *Proceedings of HCI International 2020 – Late Breaking Papers: Digital Human Modeling and Ergonomics, Mobility and Intelligent Environments*, 2020, pp. 3–13.