Journal of Data Intelligence, Vol. 3, No.2 (2022) 232-251

A DIGITAL PLATFORM FOR SHARING COLLECTIVE HUMAN HEARING

RISA KIMURA

Waseda University, Tokyo, JAPAN
r.kimura@dcl.cs.waseda.ac.jp

TATSUO NAKAJIMA

Waseda University, Tokyo, JAPAN
tatsuo@dcl.cs.waseda.ac.jp

People are hearing various natural and artificial sounds, but it is hard to imagine how hearing the sounds from people collectively can be effectively used in our everyday lives if those sounds become sharable. The sharing economy, which uses digital technologies to share a variety of physical resources in a peer-to-peer manner, has been attracting attention in recent years. Investigating the feasibility of sharing human hearing offers promising opportunities with which to expand the current scope of the sharing economy. In this study, we have developed a digital platform named CollectiveEars to share collective human hearing and explore the opportunities and pitfalls of sharing human physical senses on a digital platform. The first contribution of the study is to present an overview of CollectiveEars. The second contribution is that we reveal opportunities and pitfalls of CollectiveEars by extracting insights from two experiments. The third contribution is to show three examples to extend CollectiveEars.

Key words: Collective Human Hearing, Head Gesture, Sharing Economy, 3D Sound

1 Introduction

In the near future, we will see the development of wearable smart devices such as smart glasses [10] and smart earphones [12]. Such wearable devices should be able to gather information around people more easily. The sounds collected from these devices can be used by others to refine their hearing. Having access to the hearing of others will allow people to be flexible in their thinking and think more deeply. Our world is filled with a variety of sounds, both natural and artificial, and we consciously or unconsciously hear the sounds in our everyday lives. These sounds may strongly influence our daily lives, but it is hard to imagine the effects of the sounds that other people are hearing if those sounds become available for us anytime and anywhere. The term soundscape refers to the totality of the sounds that can be heard at any moment in any given place [31]. The soundscape offers only what an individual can listen to and not what a collection of people hear. However, using sounds in the world as sharing collective resources has not been well investigated in future smart urban environments.

In this paper, we propose a digital platform named CollectiveEars that allows collective human hearing to be shared more easily. The platform is based on smart earbuds and provides an experience where a user can make others' hearing his/her own. The digital platform enabled us to listen to what people are currently hearing, allowing us to enhance our imagination about our world. Modern urban life has become very busy and stressful, and we tend to forget a variety of important issues that contribute to our well-being, such as living in nature or diverse chances to do so in the future. The various sounds in the world can offer new possibilities that make us aware of our world's diversity and expand our chances to experience it. The contributions of the study are that we proposes a novel digital platform named CollectiveEars and present its preliminary user studies to extract some current characteristics of CollectiveEars and we show an enhancement of its key abstraction to expand the use cases to expand its possibilities. This paper expands our brief report [20] and enhances our previous conference paper [23] through extended design, analysis, implementations and evaluations.

In the remainder of this paper, after we present related work and compare our approach with the related work, we present key components and functions of CollectiveEars, especially, several interactive methods to present and navigate the sounds that are heard by other people in a user's personal listening space, which occurs after presenting the related work. Then, we show two experiments that were conducted to explore some opportunities and pitfalls to share people's physical hearing. Then, we show three examples to extend CollectiveEars by enhancing the theme channel abstraction. Finally, we conclude the paper and show some future directions.

2 Background and Related Researches

Multisided digital platforms, termed the sharing economy — referring to the peer-to-peer-based sharing of access to various resources by sharing them with digital technologies — have recently attracted a great deal of attention [7]. Traditional sharing economies generally involve people's belongings, such as cars, or logical resources, such as people's spare time. Investigating the feasibility of sharing other types of physical resources, particularly human bodies, might offer promising opportunities to reconsider an interesting direction in which to expand the current scope of the sharing economy because navigating people's behavior is necessary for accessing expected situations through others' shared human bodies. The sharing economy includes a sprawling range of multisided digital platforms and offline activities such as Airbnb[a], a peer-to-peer lodging service, and Uber[b], a peer-to-peer transportation network to share goods and services. TaskRabbit is a digital marketplace that matches freelance labor with the demand, allowing consumers to find immediate help with everyday tasks such as cleaning[c]. In CollectiveEars, people allow users to share their ears with very low effort. Thus, most people would allow access to their ears without economic rewards if proper incentives were given and their privacy was not violated. Various types of incentives can be considered to encourage people to offer their hearing capabilities.

Building on definitions of technology through a large number of research perspectives and thinking has generated debates regarding technology-in-use and sociomateriality and its grounds [29, 30]. In developing agential realism, Karen Barad suggests that further attention must be given to how meaning and matter are held together [1]. Sociomateriality denotes the enactment of a specific set of

[a] https://www.airbnb.com/
[b] https://www.uber.com/
[c] http://www.taskrabbit.com/

practices that combine materiality with institutions, norms, discourses, cultures, and other social phenomena. The related research emphasizes nonhuman agencies, whereas in our approach, we view virtualized human hearing as a first-class object and consider its agency to analyze its effect on society.

Cities and Memory [2] is a collaborative worldwide project for field recordists, sound artists, musicians and sound enthusiasts who contribute sound recordings collected from various cities in the world. Over 500 contributors have currently contributed to record more than 2,000 audio sounds from more than 80 countries around the world on the Web. In contrast, CollectiveEar*s* presents multiple sounds from around the world, which are collected from people who use it simultaneously, and users navigate the sounds in their personal listening spaces to hear various sounds in the world.

Adaptive walk on a fitness soundscape [32] is a new kind of interactive evolutionary computation for musical works. It provides a virtual two-dimensional grid in which each grid point corresponds to a listening point that generates a sound environment. People's localization and selective listening abilities make them walk toward the grid points that generate more favorable sounds. Audio augmented reality is another direction to enhance the use of sounds. For example, NavigaTone integrates the needed navigational cues into the regular stream of music in an unobtrusive way [11]. Instead of moving the entire track around in stereo panorama, we only move a single voice, instrument, or instrument group. CollectiveEars is also considered to enhance our real world in terms of sounds. A user interface to music repositories called nepTune creates a virtual landscape for the arbitrary collection of digital music files, letting users freely navigate the collection [13]. Automatically extracting features from the audio signal and clustering the music pieces accomplishes this. The clustering helps generate a 3D island landscape. The interface projects multiple sounds only in a spatial way. On the other hand, CollectiveEars offers both spatial and temporal ways to project multiple sounds in a user's listening space.

360 Reality Audio [33] offers a new music experience that uses an object-based spatial audio technology. Individual sounds such as vocals, chorus, piano, guitar, bass and even sounds of the live audience can be placed in a 360 degree spherical sound field, giving sound artists and creators a new way to express their creativity. Listeners can be immersed in a field of sound exactly as intended by sound artists and creators. The approach will offer CollectiveEars new opportunities to locate sounds in a 3D space in a more immersive way.

The design of traditional digital platform has been based on the human centred design [26], where the assumption is that humans should naturally be given primacy over nonhuman actors in the design process. The approach is effective to extract the opportunities and pitfalls in terms of end-users. However, the real potentiality of CollectiveEars is to offer abilities beyond human's hearing capabilities. New technological capabilities are starting to give nonhuman actors decision-making abilities, thereby allowing for an active form of agency [3, 4]. This new trend will be meaningful to improve the potential power in future CollectiveEars for offering a deep impact on our social creativity.

A digital platform to share collective human eye views named CollectiveEyes has been proposed [14, 17, 22]. The platform offers spatial and temporal multiple view modes similar to CollectiveEars, and the views can be navigated through gaze-based gestures. Similar to CollectiveEars, CollectiveEyes broadens people's thinking abilities by being aware of the diversity in the world. It can also be used to make us happy by offering positive past visual memories captured from collective people, and the same effect can be expected by sharing people's ears [18, 19, 21].

Emerging digital technologies allow us to augment the meaning of our real world to influence our attitudes and behaviors [15, 16, 25]. For our daily lives to become more sustainable and flourish, influencing our attitudes and behaviors is essential. To influence our attitudes and behaviors, the meaning of the real world needs to be refined to make people believe that their augmentation has meaningful effects on our real daily lives. CollectiveEars offers a new possibility to enhance the meaning of our real world from the auditory angle not the visual angle that the previous work mainly focuses on.

3 Designing CollectiveEars

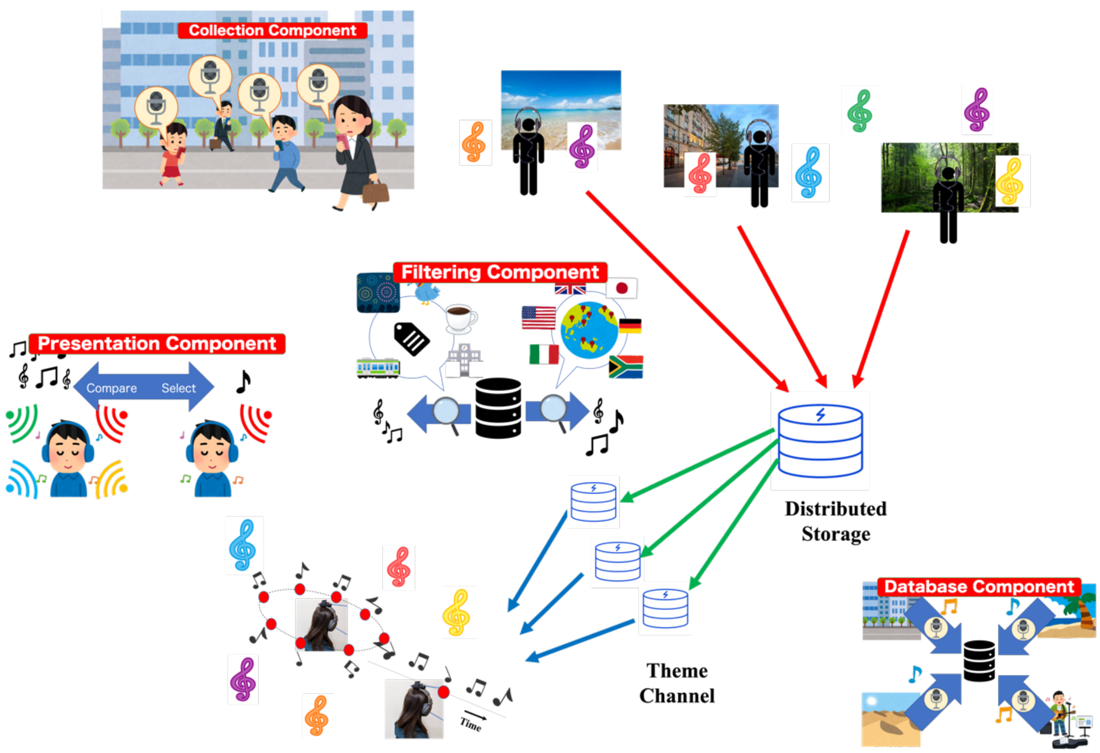

Figure 1 An overview of the CollectiveEars digital platform

We have developed CollectiveEars to demonstrate its feasibility. The current CollectiveEars consists of the following four components as shown in Figure 1. The first component is the *Collection Component* that collects the heard sounds captured through people's hearing capabilities around the world. The current component assumes that the people use wearable microphones such as the eSense device [12] to share their hearing. The second component is the *Database Component* that stores all hearing sounds gathered in the collection component. The third component is the *Filtering Component* that selects the stored sounds from the shared database based on the selected theme channel. The last component is the *Presentation Component* that shows multiple sounds selected by the theme channel.

We adopt Unity[d] to present multiple sounds in a user's 3D listening space. An end-user of the CollectiveEars wears a headphone equipped with the Nintendo JoyCon device [24], as shown in Figure 2 to simulate eSense in the current implementation. The device contains an acceleration sensor for detecting the user's head gesture and transmits the sensor data through Bluetooth.

The purpose of the current research is to investigate the feasibility of CollectiveEars in which we would like to present the opportunities and pitfalls in terms of an end-user's point of view. The current CollectiveEars focuses on presenting multiple sounds that directly offer a user experience to end-users. However, we do not focus on gathering and recording a large number of collective people's hearing in the current prototype platform because the current research's focus is to investigate the feasibility of CollectiveEars from the end-user's point of view.

Figure 2 Using CollectiveEars

The remainder of this section focuses on the prototype platform's main four characteristics that are essential for the end-user experience when using CollectiveEars.

3.1 Sound Presenting Methods

CollectiveEars offers the following two methods to present multiple sounds. The first is the spatial presentation method and the second is the temporal presentation method. In the spatial presentation method shown in the left part of Figure 3, sound sources are placed around a user, and multiple

---

[d] https://unity.com/

sounds are presented simultaneously. Here, we used the Unity's 3D sound function and prepared one Listener object as a user's listening position and multiple Audio Source objects as each sound source.

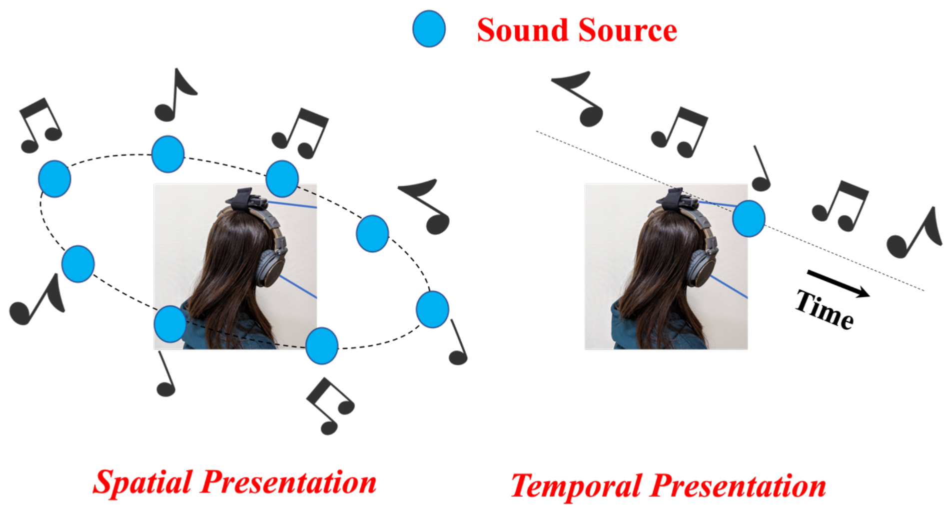

Figure 3 Spatial and temporal presentation

In the temporal presentation method shown in the right part of Figure 3, multiple sounds are presented one by one at regular intervals. The currently heard sound can be selected, and then the selected sound can be skipped if a user explicitly specifies to switch to the next sound.

3.2 Presenting Multiple Sounds

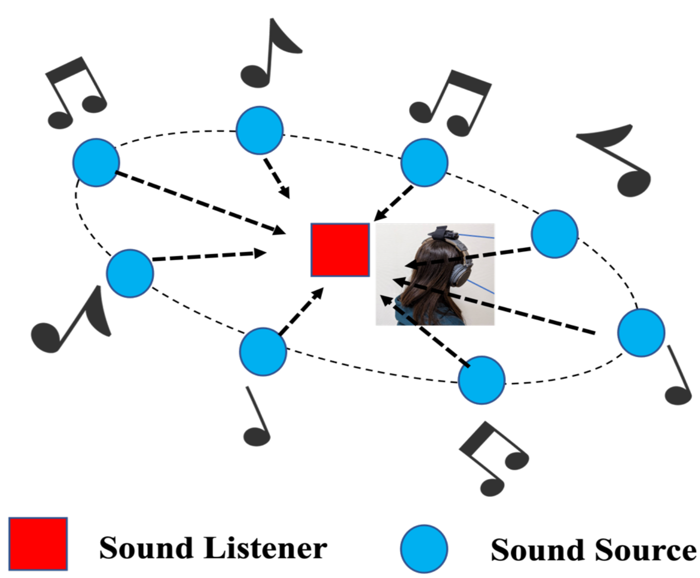

Figure 4 Spatially presenting multiple sounds

All Audio Source objects were placed equidistant from the Listener object so that all sound sources could be heard at about the same size by the Listener object. Figure 4 shows the arrangement for the eight sound sources in CollectiveEars. In the figure, the Listener object at the listening position is represented by a square, and the Audio Source object at the sound source is represented by a circle.

3.3 Choosing Sounds from a Sound Storage

A user specifies a theme channel to select multiple sounds that are presented around the user, such as selecting a TV channel from the stored sounds in the database. Currently, the following four theme channels are offered in CollectiveEars: the *similar sounds in different places* channel, the *different sounds in the same place* channel, the *speaking sounds on streets* channel and the *random sounds* channel. The *similar sounds in different places* channel offers multiple similar sounds such as birds' songs and rivers' sounds in various places. The channel allows a user to be aware of the diversity among similar sounds. The *different sounds in the same place* channel offers various different sounds in the space where a user specifies. The channel allows the user to be aware that there are diverse sounds in the target place. The *speaking sounds on streets* channel offers multiple human speaking sounds on various urban streets. The channel allows a user to hear several conversations that are currently active in various urban cities. The *random sounds* channel is the last channel that offers diverse multiple sounds randomly. The head gesture that is described below can replace the currently presented sounds within the specified channel so that a user can hear a large number of sounds from around the world by continuously replacing the presented sounds.

3.4 Navigating Sounds with Head Movement

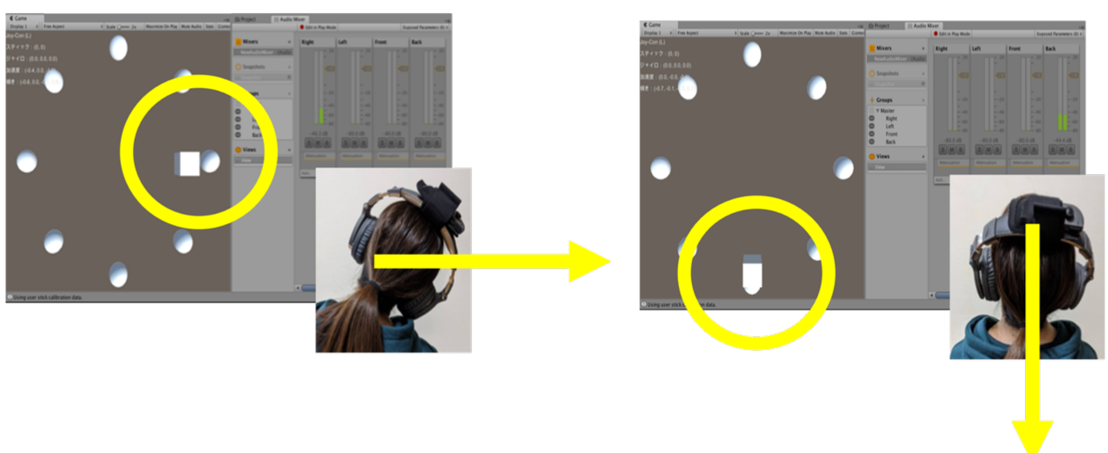

Figure 5 Head gesture based sound navigation

When using the spatial presentation method, the *sound focusing function* was implemented, where the function makes it possible to loudly hear sounds in the direction of a head tilt. For example, if a user tilts his/her head to the right, the sound of the Audio Source object located on the right side of the Listener object in the Unity's 3D space will be heard loudly; if he/she keeps listening for a few seconds, the sound will be selected. The left part of Figure 4 shows the arrangement of each object and the loudness when the head is tilted to the right. By tilting the head, the Listener object moves to the right, and the sound to the right becomes louder.

The right part of Figure 5 shows the arrangement of each object and the loudness when the head is lowered to the back. By lowering the head, the Listener object moves to the back, and the sound to the back becomes louder. When using the temporal presentation method, the currently presented sound is automatically switched to the next sound after the predefined time. The head gesture allows a user to switch to the next sound without waiting for the predefined time. It also allows him/her to select the current sound and stops to switch to the next sound or to return back to the previous sound.

4 Extracting Insights through Two Experiments

The purpose of these experiments shown in this section is to investigate the opportunities and pitfalls of the current CollectiveEars prototype platform in terms of the end-user experience perspective to share people's hearing and to navigate multiple sounds with head gestures. We conducted the following two experiments: the first experiment is mainly to extract insights into experience with CollectiveEars by end-users, and the second experiment focuses on the usability of the head gesture-based sound navigation to manage multiple sounds offered by current CollectiveEars.

4.1 Experiment 1: Investigating Opportunities and Pitfalls in CollectiveEars

In the first experiment, we wanted to explore the opportunities and pitfalls of CollectiveEars based on the research through design method [9]. We knew that the proposed approach might not be familiar to most of the participants in the experiment. We predicted that they might not understand how to use the current CollectiveEars prototype platform just by experimenting with it. Additionally, some essential aspects of CollectiveEars are missing in the current prototype platform, and we did not strongly focus on its usability. Because participants may focus on the weaknesses that we do not want to investigate in this experiment if we focus on the usability. Therefore, we decided to use four use case scenarios shown in Table 1. Each scenario uses a different theme channel explained in Section 3.3. By playing a role in the scenarios, the participants could better understand the potential merits of the design and explore the opportunities and pitfalls of the essential ideas in CollectiveEars. Additionally, they did not investigate the aspects outside of this experiment's main focuses. The experiment used the sounds captured by Cities and Memory [2] that gathered sounds in various cities and natural places to simulate the gathered sounds from collective people's hearing around the world.

In the current experiment, 14 males and 2 female participants (average age = 24.1) participated. After providing an overview of the proposed concepts in CollectiveEars and the purpose of the experiment, each participant played the role of the main character in each scenario based on user enactments [28] with CollectiveEars. In each trial, a participant used the CollectiveEars prototype platform with both the spatial presentation method and temporal presentation method.

The participants experienced multiple sounds with the spatial presentation method to compare and select them under six conditions: 1-6 shown in Table 2. Each condition differs the number of sounds presented and the presence or absence of the sound focusing function (SFF) using a participant's head gestures. They also experienced multiple sounds with the temporal presentation method under two conditions: 7-8, shown in Table 2, with and without head gestures.

Table 1 Four scenarios used in the experiment

Scenario 1
Takumi is a university student who lives in Japan. Takumi uses CollectiveEars and listens to the sounds provided by other people almost everyday. One day, he had to go to France to present his research in an academic conference. He has never been abroad before, so he wanted to get a sense of the French atmosphere. He searched for sounds recorded in France with CollectiveEars. He was able to feel the French atmosphere by selecting one sound from multiple sounds, such as in a church, at a restaurant, near the sea, in an old town, or on a tram by comparing multiple sounds at the same time.
Scenario 2
Asuka is a university student who lives in Japan. He uses CollectiveEars to listen to the sounds provided by other people almost everyday. One day, He learned about Japanese zoos in his university class. Therefore, he wanted to know the differences between foreign and Japanese zoos, so he used CollectiveEars to get an idea of the atmosphere of each. He searched for sounds associated with zoos with CollectiveEars. He was able to feel the atmosphere of each country's zoos by comparing the sounds of zoos in Japan, France, Australia, Egypt, and other countries, and by selecting a single sound from the multiple sounds.
Scenario 3
Makoto is a university student who lives in Japan. He uses CollectiveEars and listens to the sounds provided by other people almost everyday. One day, he had to work at home using his desktop computer. He wanted to hear some sounds as background music, so he used CollectiveEars. He compared several randomly selected sounds from all over the world, and picked up one sound while working.
Scenario 4
Daichi is a university student who lives in Japan. He uses CollectiveEars to listen to the sounds provided by other people almost everyday. He is planning to study English in England next year. He wanted to hear native British English, so he used CollectiveEars. He studied British English by comparing several "conversation" sounds from England to the sounds provided from around the world, or by selecting one sound from the multiple sounds.

Table 2 Eight conditions in the experiment

Condition Number	Presentation Method	The Number of Simultaneous Sounds	Head Gesture
1	Spatial	2	Unused
2	Spatial	4	Unused
3	Spatial	8	Unused
4	Spatial	2	Used
5	Spatial	4	Used
6	Spatial	8	Unused
7	Temporal	1	Unused
8	Temporal	1	Used

Figure 6 and Figure 7 show the number of participants who answered which scenario they would like to use in each condition in. In these figures, “don’t want to use” means that a participant do not want to use CollectiveEars in any scenarios. Figure 6 shows the results when using the spatial presentation method, and Figure 7 shows the results when using the temporal presentation method. After asking the questions, we also interview them to ask the reasons of their answers.

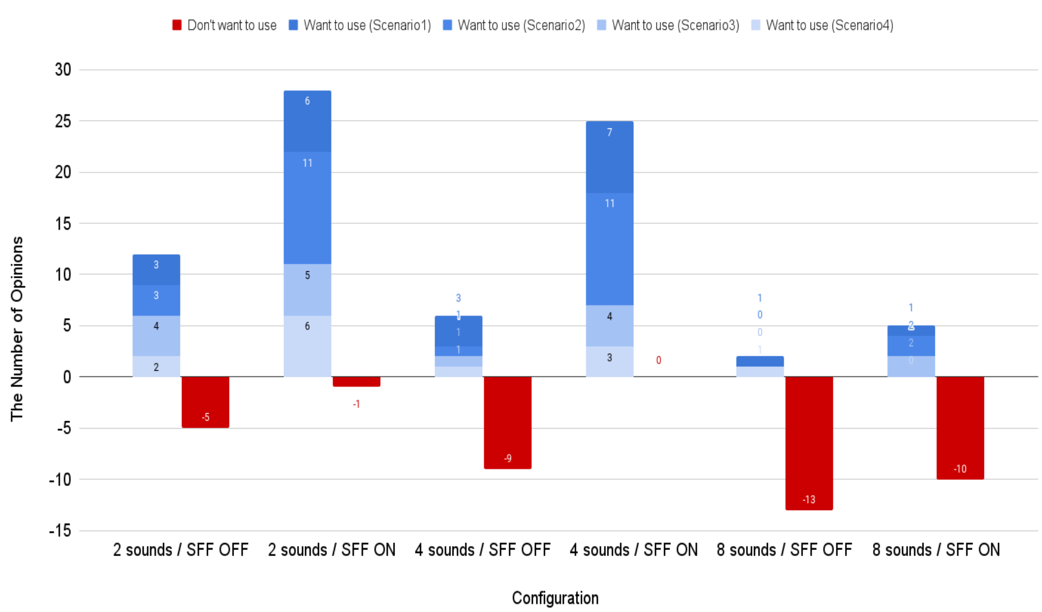

Figure 6 The answers on scenarios when using the spatial presentation method

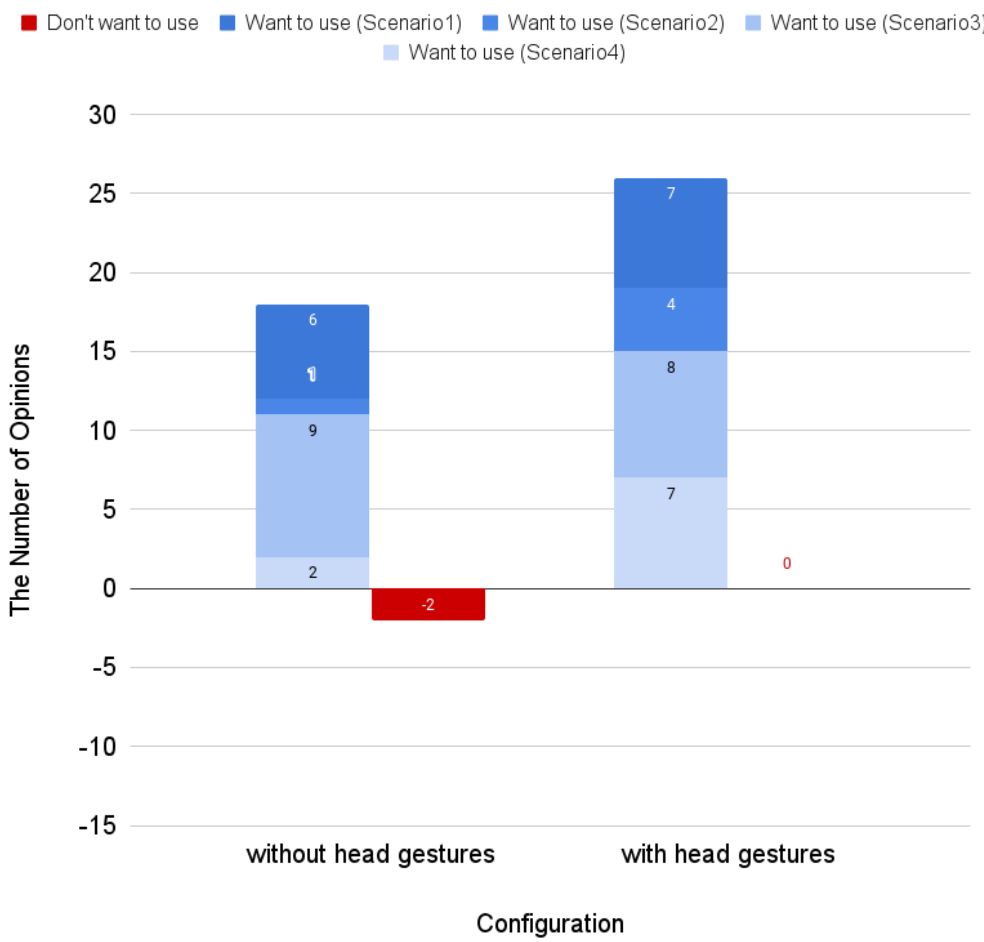

Figure 7 The answers on scenarios when using the temporal presentation method

As the results shown in Figure 6, when presenting multiple sounds using the spatial presentation method, in the condition where the sound focusing function by head gestures is not used, the number of opinions that they did not want to use CollectiveEars increases as the number of sounds presented simultaneously increases. This fact suggests that it would be desirable to minimize the number of sounds presented at the same time when the sound focusing function is not used. On the other hand, in the case where the sound focusing function is used, a large number of participant wanted to use Scenario 2 in both cases where the number of sounds played simultaneously is two or four. The opinions supporting this result were: "I thought I could enjoy the difference of multiple sounds by comparing them." and "I felt it was convenient to be able to compare sounds smoothly by moving my head." On the other hand, the majority did not want to use ColletiveEars when the number of sounds played simultaneously was eight. This result may be because they feel that it is difficult to distinguish the sounds when there are eight sounds playing at the same time, even though the sound focusing function can increase the sound in the direction they tilt their heads. A participant answered, "It was difficult to understand which direction the sound was coming from even though I listened to the sound in eight directions while moving my head." Another participant said "It was difficult to listen to only one sound because it felt like the sounds were mixed.

From the result shown in Figure 7, it seems that many participants wanted to use the temporal presentation method in Scenario 3 regardless of whether or not the sounds are navigated by head gestures. One of the reasons for the answer is an opinion that "When I'm listening to a sound in a stream, it is convenient to easily access the next sound." Compared to the spatial presentation method, the number of participants to claim "don't want to use" was much smaller when using the temporal presentation method. The reason for the claim was that it is convenient that the next sound will be heard when listening to the sound. Since the way to select sounds in terms of time is almost the same as the function of playing songs one by one that is widely used, it seems that the method is familiar to the participants

We also interviewed the participants to ask about their experiences performing the scenarios. From the results of the interviews, we classified the participants' opinions into the following two categories of insights in accordance with the thematic analysis method [27].

Use Cases of Collective People's Hearing: The most popular opinion from the participants is that they want to use CollectiveEars to hear or experience sounds that they had never heard before rather than obtaining information on the sounds from familiar places. For example, a participant said that they wanted to hear many sounds from places where he has never been before. Another participant said that the participant wants to feel the realism of the places by sharing the participant's hearing experience with people who attend the live music concerts of famous artists or the various festivals around the world. This result is probably because it is difficult to obtain detailed information in an identified place only from the sounds.

Other opinions indicate that CollectiveEars can be used to enhance our imagination due to the limited information. For example, one opinion is that it seems that only hearing offers the participant a good inspiration and another is that it is difficult to determine what is happening with only sound, but it stirs up his/her inspiration.

On Privacy Regarding What People Are Hearing: Some opinions about privacy are that there is no problem sharing their hearing with anyone since it is hard to identify the location and other characteristics from only sounds. In addition, they would definitely like to refrain from sharing their hearing that contains private contents, but if anyone can listen to their hearing when they are there, there is no problem sharing their hearing. On the other hand, some participants want to share their hearing only with their well-known acquaintances. For example, they said that they do not want to disclose what they are currently hearing to people who do not know and they are reluctant to provide strangers with their hearing containing their personal information. The results show that some people will never want to offer their hearing to CollectiveEars.

4.2 Experiment 2: Evaluating Multiple Sounds Presentation and Head Gestures

In the second experiment, we investigated several insights on the multiple sounds presentation and the head gesture based sound navigation in the current CollectiveEars prototype platform. In each trial, each participant used CollectiveEars with the following two modes. The first mode adopts the spatial presentation method with the sound focusing function, and the second mode adopts the spatial presentation method without the sound focusing function. We also changed the number of sound sources to 2, 4 and 8, when adopting the first and second modes.

We hired 16 participants (ages m=24.12, 14 males). The participants were asked to complete a listening test using the spatial and temporal presentation methods, and their correct response rates were investigated. In the experiment, 15 tests were conducted, including 5 for each configuration of the different number of sound sources presented. To simulate the sounds heard by people in various locations around the world, sounds were selected from mainly Cities and Memory [2], and they were classified into five categories: electronic sounds, animal sounds, bird sounds, environmental sounds, and forest sounds. We expected that the rate of correct answers would decrease as the sounds became more similar and difficult to distinguish, which is why we categorized them in this way. Figure 8 shows the results of this experiment. Each column shows the percentages of correct, similar and wrong answers. Similar answers mean that people chose the sound source located next to the "target sound" in the question where the number of sound sources was eight. For example, if the "target sound" is located on the "right", the person answered "right oblique front" or "right oblique back".

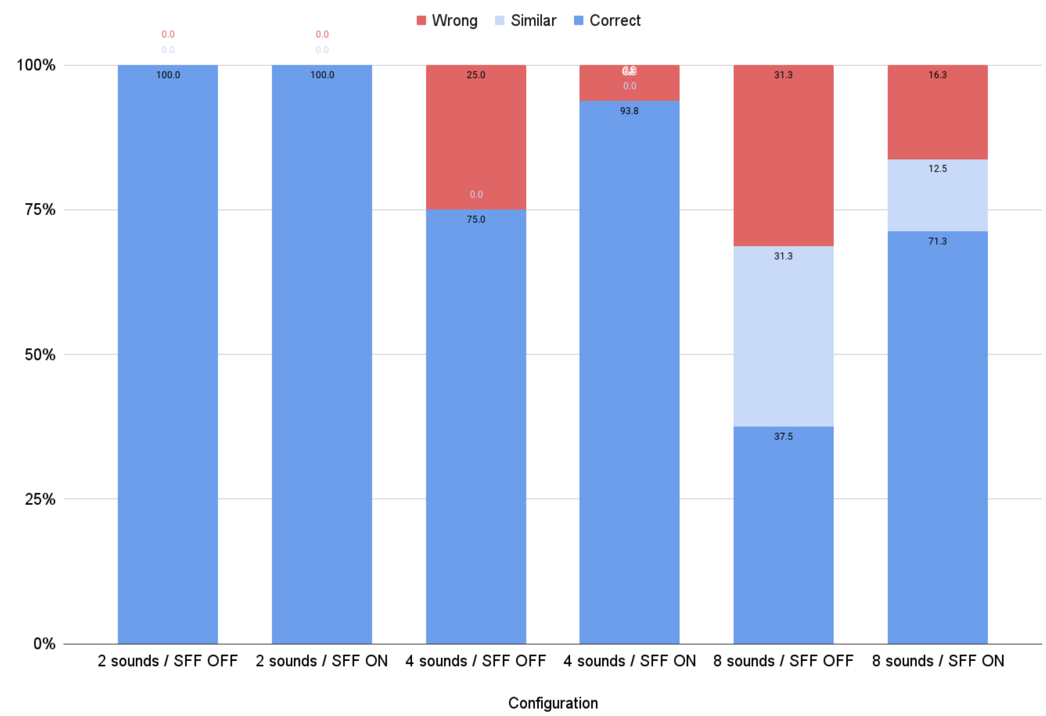

Figure 8 The results of experiment 2

When looking at Figure 8, the correct answer rate for the two sound sources configuration was 100%. Therefore, it is possible to distinguish between left and right sounds whether the sound focusing function is turned on or off. In the case of listening to the four sound sources configuration, the correct answer rate was approximately 75% when the sound focusing function is turned off, while it improved to approximately 90% when the sound focusing function is turned on. The result indicates that the sound focusing function is more appropriate than without the function in the case of the four sound sources configuration.

Regarding the case of the eight sound sources configuration, the rate of correct answers is extremely low in the case without the sound focusing function. When the sound focusing function is turned on, the correct answer rate increases significantly, but it remains low compared to the cases of the four and two sound sources configurations. If we sum up the similar answer rate and the correct answer rate in the case of the eight sound sources configuration without the sound focusing function, we can see that the correct answer rate for the four sound sources configuration without the sound focusing function is not much different. In addition, it can be seen that the rate of correct answers of the eight sound sources configuration with the sound focusing function increased at almost the same rate as the rate of similar answers of the eight sound sources configuration without the sound focusing function. Therefore, accurate listening in the eight sounds configuration is extremely difficult without the sound focus function, but it is possible to determine the approximate location of the sounds in the current prototype platform. Even with the sound focusing function, it is more difficult than in the two or four sounds configuration, but it is possible to improve the accuracy of listening.

5 Three Examples to Extend CollectiveEars

In this section, we examine to enhance CollectiveEars through exploiting the flexibility of the theme channel abstraction as an example to show the power of the abstraction based on the experiences with designing, developing and evaluating the original version of CollectiveEars. We show three enhancements: Mindful Sound Space, Ideological Talking Things and Onomatopoeic Music Space, where each enhancement explores different potential opportunities of CollectiveEars.

5.1 Mindful Sound Space

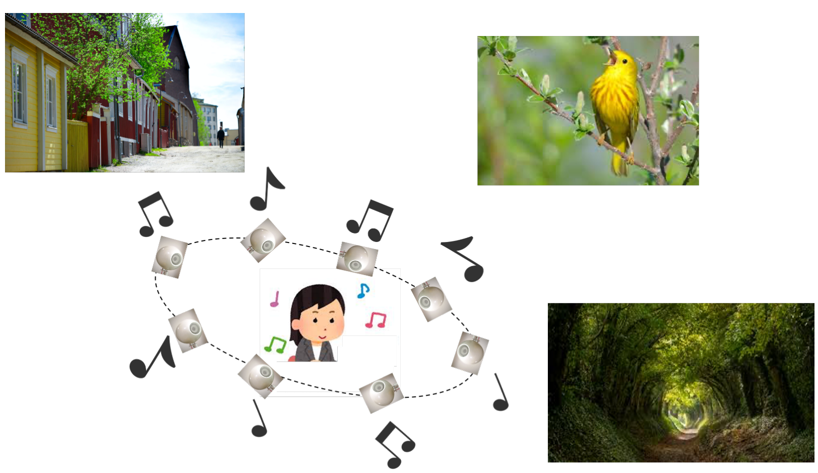

Figure 9 Mindful Sound Space

Mindful Sound Space offers mindful sounds to a user. A user is located within multiple loudspeakers that produce different sounds while using Mindful Sound Space, where original CollectiveEars to share human hearings to use headphones to produce sounds. Mindful Sound Space specializes the theme channel abstraction to select sounds that a user wants to listen to. The sounds are gathered by people who hear various sounds in the world through the CollectiveEars platform.

Mindful Sound Space offers three theme channels as shown in Figure 9. The first theme channel is "Calm", and the second theme channel is "Ambient" and the third theme channel is "Aesthetic". A user chooses the channel that he/she likes to hear through head-based gestures. After selecting the channel, several sounds are rendered in the 3D sound space and he/she can choose one of them if he/she likes. People need to tag the sounds that they are hearing before registered the sounds in Mindful Sound Space. One interesting issue of the approach is that the tagging is determined by human, so the decision of the tagging is based on human' ambiguous perception about the sounds that they hear. Therefore, a user can consider how people think that the current sound is calm, ambient or aesthetic.

5.2 Ideological Taking Things

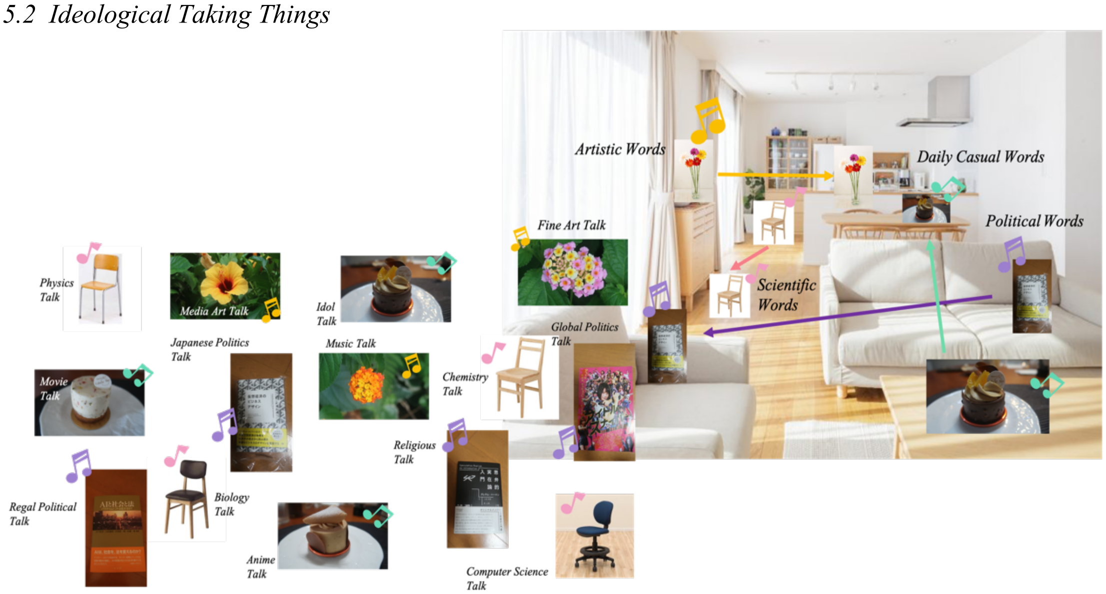

Figure 10 Ideological Talking Things

Ideological Talking Things gathers sounds and presents the sounds from physical objects commonly encountered in a living room. A user can move these objects to any location in the room as shown in Figure 10. The user can shuffle sounds by touching the objects, and the volume of the sounds they emit is automatically configured according to the brightness of the surroundings. Some users come to consider these objects really living things and feel empathy with the objects. In this manner, the user can become more aware of these objects and come to cherish them such that he/she may develop a desire to maintain the objects more carefully so as to use them longer.

In the current design, we focus on four object types found in a living room: chairs, books, flowers and cakes. Ideological Talking Things uses theme channels to group the sounds that each object speaks, and offers four theme channels. The chairs speak scientific words selected by the first theme channel, the books speak political words selected by the second theme channel, the flowers speak artistic words selected by the third theme channel and the cakes speak casual words typical of daily conversational words selected by the fourth theme channel. Each object is converted into an IoT device, whereby a small attached computer detects each object's location. A user wears headphones and hears 3D sounds emitted by these objects. The user can move an object from its current position to a new one, and the location of the sound produced by the object changes according to its current position.

Each object speaks different words extracted from sounds in the world. Thus, a user can easily distinguish the objects and their current positions in the room, so the user determines the location of each object in an easy way. Because the sound volume changes according to the brightness of the objects' surroundings, users may feel that the objects resemble living beings. Additionally, the user may experience new relationships with the objects because he/she can move them according to his/her feeling and mood.

A user can passively listen to a conversation among these objects, but he/she can also change the spoken content through interaction with them. Each object independently speaks its own words without considering the other objects. However, Ideological Talking Things offers an interesting experience in which a typical person speaks selfishly in a daily casual conversational mode, without displaying consideration for the feelings of others. This experience offers a chance for the user to become aware of his/her own selfish speaking in the course of daily life.

5.3 Onomatopoeic Music Space

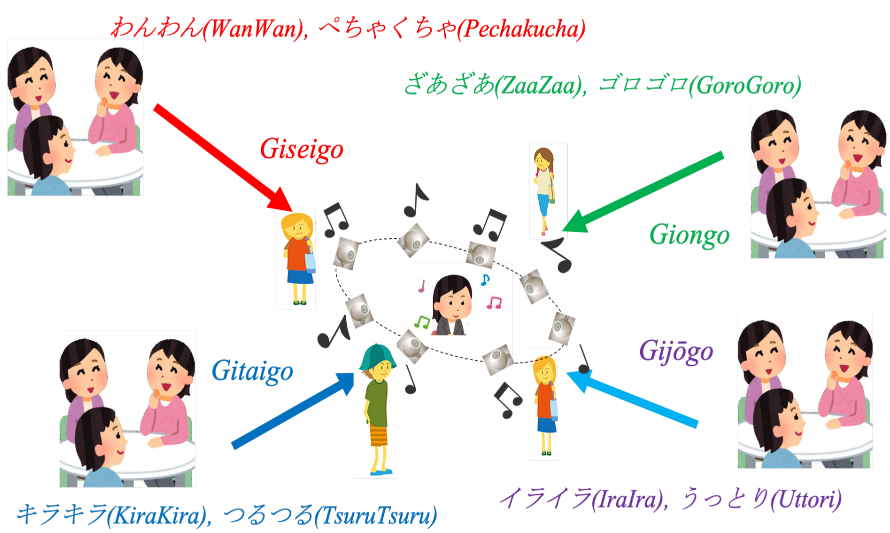

Figure 11 Onomatopoeic Music Space

Onomatopoeic Music Space offers mindful, music-like sounds constructed from onomatopoeic sounds filtered from the sounds in the world. The user of Onomatopoeic Music Space is surrounded by multiple loudspeakers that produce different sounds as shown in Figure 11. The current version of Onomatopoeic Music Space offers four theme channels: *Giseigo*, *Giongo*, *Gitaigo*, and *Gijōgo*. Each channel groups onomatopoeic sounds based on their Japanese classification [8], whereby *Giseigo* mimics humans and animals[e], *Giongo* mimics general noises in nature or made by inanimate objects[f], *Gitaigo* describes states of the external world[g], and *Gijōgo* describes psychological states or bodily feelings[h].

Several onomatopoeic sounds are rendered in the 3D sound space. A user can listen to the four types of onomatopoeia simultaneously. The experience with using Onomatopoeic Music Space offers a strange conversational atmosphere constructed only from onomatopoeic sounds. Onomatopoeic Music Space also creates an artful experience, similar to listening to a musical work like "ZawaZawa[i]" [6], created by Dai Fujikura who is Japanese-born composer of contemporary classical music and in which each person in a chorus speaks different onomatopoeic sounds.

5.4 Other Future Opportunities

Digital twin technologies have recently increased in popularity, particularly in manufacturing and smart cities [5]. Certain aspects of the real world are transformed into data with digital twin technologies, and the data are used to optimize physical resource use. Our approach advances one step further from the current digital twin approach by virtualizing these aspects as design materials for developing innovative digital services. Our approach virtualizes a variety of sounds in the real world, so it offers new ideas for investigating unexplored opportunities for future digital services. For example, certain abstract complex concepts can be represented as human-like agents, and the agents can explain the concepts through their agencies, a technique popular in fiction. Such virtualized materials also offer new opportunities for enhancing traditional sharing economy approaches [6]. Sharing economy platforms should treat the sharing entities as digital information, and virtualizing certain aspects of the real world enables us to use these aspects as sharable entities for developing innovative services. Enhanced digital twin technologies enable us to use more sharable entities by virtualizing diverse aspects of the real world.

6 Conclusion

In this paper, we presented a digital platform named ColectiveEars for developing services that uses collective humans' auditory capabilities. The study conducted several experiments and explorations about the proposed approach and showed the diverse opportunities and potential pitfalls of sharing a collection of what people have heard.

---

[e] e.g., "WanWan" (a dog's bark)
[f] e.g., "ZaaZaa" (the sound of water flowing down vigorously)
[g] e.g., "KiraKira" (glittering)
[h] e.g., "Uttori" (enchanted)
[i] murmuring sound

Currently, we used a simple 3D audio interface to locate multiple sounds in a 3D space. The technology has limitations in recognizing a large number of sounds, but a new technology such as that in [33] will offer new opportunities to locate sounds in a 3D space in a more immersive way. The capability of sharing a collection of what people are seeing was discussed in [14, 17, 22]. We would like to investigate whether digital platforms to “share” people’s eyes and ears will make it possible to refine how we see and hear the world and to enhance our thinking and creative abilities in the future.

A very few researches have investigated potential opportunities of collective sound hearing, where the current study focused on end-users’ individual sound hearing. One new promising research direction is to study how the sounds in worldwide influence collective people’s daily behaviors and thinking through collective hearing experiences. We like to investigate that our proposed concept offers new opportunities for people to reconsider a way of hearing and listening and to develop a new capability to experience our world.

References

1. Barad, K. Meeting the University Halfway: Quantum Physics and the Entanglement of Matter and Meaning, Duke University Press, 2007.
2. Cities and Memory, https://citiesandmemory.com/ (Accessed January 10 2022)
3. Coulton, P., and Lindley, J,G. More-Than Human Centred Design: Considering Other Things, The Design Journal: An International Journal for All Aspects of Design, 22(4), 2019, https://doi.org/10.1080/14606925.2019.1614320
4. Cruickshank, L. and Nina Trivedi, N. Beyond Human-Centred Design: Supporting a New Materiality in the Internet of Things, or How to Design When a Toaster is One of Your Users, The Design Journal: An International Journal for All Aspects of Design, 20(5), 2017, https://doi.org/10.1080/14606925.2017.1349381
5. Farsi, M., Daneshkhah, A., Hosseinian-Far, A., and Jahankhani, H. Digital Twin Technologies and Smart Cities, Springer, 2020.
6. Fujikura, D. Zawazawa, https://www.youtube.com/watch?v=jznd5JCy2As (Accessed January 10 2022)
7. Hamari, J. Sjöklint, M. and Ukkonen, A. The Sharing Economy: Why People Participate in Collaborative Consumption, Journal of the Association for Information Science and Technology, 67(9), 2047-2059, 2015, https://doi.org/10.1002/asi.23552
8. Hayasaka, M. Japanese Onomatopoeia, Rockwaller Books, 2019.
9. Giaccardi E. and Stappers, P.J. Research through Design., In Chapter 43: The Encyclopedia of Human-Computer Interaction. 2nd Ed. Interaction Design Foundation, 2017).
10. Google Glass, https://www.google.com/glass/ (Accessed January 10 2022)
11. Heller, F. and Schöning, J. NavigaTone: Seamlessly Embedding Navigation Cues in Mobile Music Listening, In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 637, 2018, https://doi.org/10.1145/3173574.3174211
12. Kawsar, F, Min, C. Mathur, A. and Montanari, A. Earables for Personal-Scale Behavior Analytics, IEEE Pervasive Computing, 17(3), 2018, https://doi.org/10.1109/MPRV.2018.03367740.
13. Knees, P. Schedl, M., Pohle, T., and Widmer, G. Exploring Music Collections in Virtual Landscapes. IEEE Multimedia, 14(3), 46–54, 2007, https://doi.org/10.1109/MMUL.2007.48

14. Kimura R. and Nakajima T. Sharing Collective Human's Eye Views for Stimulating Reflective Thinking. In Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia (MUM 2018), 341–349, 2018, https://doi.org/10.1145/3282894.3289724
15. Kimura, R. and Nakajima, T. Gamifying Human Behavior in Urban Crowdsourcing for a Sustainable Smart City, In Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, 2019, https://doi.org/10.1145/3366030.3366031
16. Kimura, R. and Nakajima, T. Digitally Enhancing Society through Structuralism: Virtualizing Collective Human Eyesight and Hearing Capabilities as a Case Study, In Proceedings of the 8th International Conference on Distributed, Ambient and Pervasive Interactions, 2020, https://doi.org/10.1007/978-3-030-50344-4_29
17. Kimura, R. and Nakajima, T. Collectively Sharing People's Visual and Auditory Capabilities: Exploring Opportunities and Pitfalls, SN Computer Science, 1(5), 298:1-24, Springer Nature, 2020, https://doi.org/10.1007/s42979-020-00313-w
18. Kimura, R. and Nakajima, T. Gathering People's Happy Moments from Collective Human Eyes and Ears for a Wellbeing and Mindful Society, In Proceedings of the 14th International Conference on Augmented Cognition, 2020, https://doi.org/10.1007/978-3-030-50439-7_14
19. Kimura, R., Jiang, K., Zhang, D. and Nakajima, T. Society of "Citizen Science through Dancing", In Proceedings of the 11th International Symposium on Ambient Intelligence, 2020, https://doi.org/10.1007/978-3-030-58356-9_2
20. Kimura, R. and Nakajima, T. Opportunities to Share Collective Human Hearing, In Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20), 4 pages, 2020, https://doi.org/10.1145/3419249.3421244
21. Kimura, R. and Nakajima, T. Collectively Sharing Human Hearing in Artful CollectiveEars, In Proceedings of the 9th International Conference on Distributed, Ambient and Pervasive Interactions (DAPI 2021), 2021, https://doi.org/10.1007/978-3-030-77015-0_16
22. Kimura, R. and Nakajima, T. Hunting Lovely and Serendipitous Eye Sights through Sharing Collective Human Eye Views, In Proceedings of the 23rd International Conference on Information Integration and Web-based Applications & Services (iiWAS 2021), 2021, https://doi.org/10.1145/3487664.3487675
23. Kimura, R. and Nakajima, T. CollectiveEars: Sharing Collective People's Hearing Capability, In Proceedings of the 23rd International Conference on Information Integration and Web-based Applications & Services (iiWAS 2021), 2021, https://doi.org/10.1145/3487664.3487801
24. McWhertor. M. Nintendo Switch Joy-Con controller does some amazing things, https://www.polygon.com/2017/1/12/14260790/nintendo-switch-joy-con-controller-features (Accessed January 10 2022)
25. Nakajima, T., Kimura, R., Ishizawa, F. and Sakamoto, M. Analyzing User Experiences in Incorporated Fictionality, In: Moura Oliveira P., Novais P., Reis L. (eds) Progress in Artificial Intelligence. EPIA 2019. Lecture Notes in Computer Science, vol 11805. Springer, Cham, 2019, https://doi.org/10.1007/978-3-030-30244-3_2
26. Norman, D. The Design of Everyday Things, Newprint, 2002.
27. Nowell, L.S., Norris, J. M., White, D. E. and Modules, N. J. Thematic Analysis: Striving to Meet the Trustworthiness Criteria, International Journal of Qualitative Methods, 16, 1-13, 2017, https://doi.org/10.1177/1609406917733847
28. Odom, W., Zimmerman, J., Davidoff, S., Forlizzi, J., Dey, A. K. and Lee, M. K. A fieldwork of the future with user enactments, In Proceedings of the Designing Interactive Systems Conference (DIS '12), 338-347, 2012, https://doi.org/10.1145/2317956.2318008

29. Orlikowski, W.J., Yates, J., Okamura, K. and Fujimoto, M. Shaping electronic communication: The metastructuring of technology in the context of use, Organisation Science, 6(4), 423-444, 1995, https://doi.org/10.1287/orsc.6.4.423
30. Orlikowski, W.J. and Scott, S.V. Exploring material-discursive practices, Journal of Management Studies, 52, 697-705, 2015, https://doi.org/10.1111/joms.12114
31. Rudi, J. Soundscape and Listening, In Soundscape in the Arts (ed. Jøran Rudi), NOTAM. 2011.
32. Suzuki, R. and Yamaguchi, S. Adaptive Walk on a Fitness Soundscape, Lecture Notes in Computer Science, vol.6625, 2011, https://doi.org/10.1007/978-3-642-20520-0_41
33. 360 Reality Audio, https://www.sony.com/electronics/360-reality-audio (Accessed January 10 2022)