





FootstepsMixer: A Tool to Express Multiple People's Footsteps in a Footstep Transmission System for Awareness Support

Hiroki Echigo^(✉) , Masakatsu Kanno, Kana Matsu, Taira Endo,
and Minoru Kobayashi 

Meiji University, 4-21-1 Nakano, Nakano-ku, Tokyo, Japan
{hiroki.echigo,masakatsu.kanno,matsu.kana,taira.endo}@koblab.org,
minoru@acm.org

Abstract. Recently, the number of people living in single-person households has increased. Although the numbers of sensor-based services that monitor personal safety are increasing, they do not provide the feeling of living in a multi-person household. In this study, we attempt to realize a system that shares the sounds of a distant family household, thereby creating the impression that the family household is nearby. Using this system, we devised a method to express people's activity in a remote location and evaluated the system experimentally. We considered multiple-footstep expression methods and based on experimental results, identified the most suitable method. Herein, we propose a "FootstepsMixer," a tool that expresses the actions of multiple people using the optimal footstep expression method. With this tool, users can easily express footsteps in an exaggerated manner and can set the number of people and the maximum volume of each footstep. We also consider what type of motion expression can be realized using the proposed tool.

Keywords: Awareness support · Footsteps · Mixer

1 Introduction

Living in a family household, we are constantly surrounded by ambient sounds, such as your mother cooking in the kitchen, your sister going up and down the stairs, or your father working on his computer. Alone in your room, you can imagine what your family members are doing without seeing the source of those sounds. These sounds often provide reassurance and happiness; however, when you live alone away from your family, you cannot hear such sounds because of work or college. We can sense the activities and liveliness of the family members because of these sounds. Herein, "liveliness" refers to the degree of vigor of the person; if you hear footsteps of your sister going upstairs quickly, you understand

that she is vigorous, and if you hear slow footsteps, you understand that she might be exhausted.

The number of one-person households is increasing globally [1]. According to the Organisation for Economic Co-operation and Development (OECD), the number of one-person households is expected to grow in all the OECD countries for which projections are available [2].

Some people who are unable to live with their families owing to various circumstances sometimes feel sad and nervous. To solve this problem, we attempted to realize a system that shares the real-life household sounds of a distant family, creating the impression that the family is nearby (see Fig. 1). Among the various possible sounds, we focused on footsteps because footsteps differ from person to person, and we considered that the sound of footsteps easily represented the liveliness of a family living away.

To achieve this goal, we are developing a system that satisfies the following requirements.

- It enables us to feel the liveliness and activity of a family living away.
- It enables us to share a sense of space with remote partners, while protecting the privacy. (This is in a trade-off relation with the first point.)

Initially, we implemented a microphone and loudspeaker-based system that recorded and reproduced footstep sounds. We found a way to express footstep sounds that could, to some extent, sound like a single person walking around on a floor above the listener. However, the initial system could not effectively express the activities and liveliness of multiple people.

In this paper, we explore various ways of expressing footstep sounds and find a method to reproduce footstep sounds that let remote family members feel as though they were in their family home. To facilitate this investigation, we propose a tool that we refer to as “FootstepsMixer.” Using this tool, we search for an optimal footstep expression method. In addition, we consider what type of motion can be expressed using the proposed tool.

2 Related Work

2.1 Awareness Communication

Considerable research has focused on awareness of remote areas [3–8]. For example, Dourish et al. [3] proposed systems called Portholes that arranged still images captured at regular intervals at various points on a single screen such that people at multiple remote points can see at any time. Robert et al. [4] proposed a system called the Video Window System that employed a large super wide screen video, and that keeps video and audio connected to remote rooms at all times. These systems [3, 4] can convey a sense of the entire space; however, there are significant privacy concerns.

Ishii et al. [5] proposed the ambientROOM, which was an interface to transmit information variedly in the background of awareness. It conveyed the motion



Fig. 1. Image of the proposed system. Our goal is to realize a system that shares the sound of footsteps from the household of a distant family to create the impression that the family is nearby.

of people and animals from remote areas in a surrounding environment using various display media, such as water, light, and shadows. Siio et al. [6] proposed a system called Peek-A-Drawer that automatically transfers photos in drawers to remote locations and shares the contents of the drawers. In addition, Meeting Pot, a system that conveys the scent of brewed coffee to a remote place, has been proposed. Rowan et al. [7] proposed a system called Digital Family Portrait that shares family life with elderly relatives who live in remote areas by connecting via portraits. Although these systems [5–7] can provide a sense of sharing aspects of life while preserving privacy and awareness, they share a limited space and do not fully share the living space. Tsujita et al. [8] proposed a system called SyncDecor that synchronizes furniture used in daily life, such as a trash can whose opening and closing is synchronized with a remote area and a lamp whose brightness is synchronized with a remote area. Although these are highly sensitive to privacy, they are shared only in limited spaces such as drawers, portraits, and pots, and it is difficult to say that they share the sense of life completely.

2.2 Stereophonic Sound

Many studies have attempted to reproduce stereophonic sound based on three clues, i.e., time gap, volume, and phase and change of frequency responses. In the groupware system developed by Cohen et al. [9], the positions of participants in a virtual meeting room are associated with the positions of sound sources. Here stereophonic sound was used to construct an acoustic environment in a virtual conference space in the conference. The system developed by Seligmann et al. [10]

shows participants' activity by arranging not only voice but also keyboard strikes and clicks on the space.

These previous studies indicate that the use of stereophonic sound in a virtual space can provide clues about human actions and a feeling of being near to others. Reproducing stereophonic sound is difficult and requires specialized equipment. In the current study, stereophonic sound does not need to be reproduced perfectly because the purpose is to convey the movement of people in remote places. The goal is to express rough movement and liveliness by the strength of the sound heard via a speaker.

3 Footstep Expression Methods

In this section, we describe the method of expressing footsteps. When footsteps are recorded using a stationary microphone, the volume decreases sharply as the distance between the microphone and the source of the footsteps increases. This sudden change hinders the smooth expression of footstep movement and adjusting the sound like smoothly moving footsteps is difficult. To overcome this difficulty, we developed a method to express the movement of footsteps by mixing and connecting prerecorded footstep sounds. We investigated several ways to express footsteps and conducted experiments to compare them.

3.1 Footstep Expression Models

The four footstep expression models investigated are listed in Table 1. Figure 2 shows how the volume changes over time for Model 1, Model 3, and Model 4. In Fig. 2, the horizontal axis represents time, and the vertical axis represents sound intensity. Model 1 is a raw recording of the sound of moving footsteps using a microphone. The method used in the initial system described in Sect. 1 is used to record this model. Models 2–4 are synthesized reproductions of footstep sounds generated by repeating a single footstep sound while gradually changing the volume. Model 2 (not shown in Fig. 2) is produced by imitating the loudness change of Model 1. Model 3 is designed to test a gradual change in the loudness of footsteps. Here, the loudness decreases linearly up to -18 dB in 2.91 s. The change of loudness over time is expressed as follows.

$$G = -\frac{600}{97}t (= -\frac{18}{2.91}t) \quad (1)$$

Here, G is volume in decibels (dB) and t is time in seconds. The method represented by Model 4 changes volume twice as fast as Model 3; thus, the volume reaches -18 dB in half the time taken by Model 3. The relation between G (dB) and t (s) in Model 4 is shown below.

$$G = -\frac{1200}{97}t (= -\frac{18}{1.455}t) \quad (2)$$

Decibels are represented using a logarithmic scale; thus, they must be represented by a power factor. The relation between G (dB) and Power is expressed as follows.

$$G = 20 \log_{10} Power \quad (3)$$

Table 1. Experimental conditions.

Model	Example
Model 1	Footsteps recorded with microphone
Model 2	Footsteps imitating Model 1
Model 3	Footsteps changed gradually
Model 4	Footsteps changed gradually in half the time of Model 3

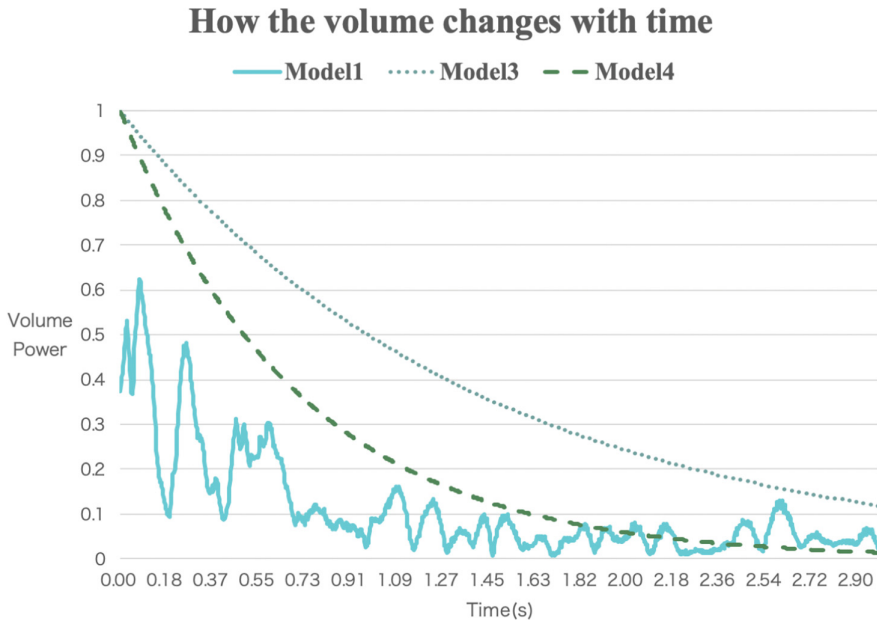


Fig. 2. Comparison of Model 1, Model 3, and Model 4.

3.2 Experiment

The experiment was performed to identify the best model from Models 1 to 4. The best model should not sound as if the people in remote areas are moving momentarily and should not feel like several people are walking. The meaning “moving momentarily” refers to the feeling when a person in a remote area teleports or jumps between footsteps playback devices. In addition, the footsteps

in the experiment are one person's footsteps; thus, the feeling that several people are walking is not suitable.

Participants. Twenty college and graduate students volunteered to participate in the experiment. The participants included fifteen males and five females, and ranged in age between 22 and 25 years. Prior to conducting the experiment, the participants were asked the question, "Do you have experience of hearing footsteps from the upper floor in an apartment, or a house?" 85% of the participants answered "Yes."

Experimentation Space. The experimental space was a square (2.61 m per side). This size is sufficiently large for participants to understand the movement of footsteps. We installed four footstep playback devices on poles at the corners of the square space. The footstep playback devices were placed facing downward at a height of 2 m from the floor.

Procedure. The participants listened to the playback of the four types of footstep sounds listed in Table 1 individually. Participants answered the questionnaire for each Model (see Table 1). Table 2 lists questions and options the range of potential responses. The participants were told in advance that they would hear footsteps; however, they were not told how many people's footsteps they would hear. In Table 2, Q1, Q2, and Q4 were evaluated on a seven-point Likert scale. The Latin square method was used in consideration of the influence of the order on the results. After the experiment, we told the participants in the experiment that footsteps sound you heard is of a single person and the participants were asked the questions shown in Table 3. The experimental environment is shown in Fig. 3.

3.3 Results

Results of Questionnaire. For "Q1. Did you feel that it was footsteps?," all models received scores of 5 or more. There was no significant difference between the four models. Figure 4 shows the results for "Q2. Did this footstep make you feel that people in remote areas were moving momentarily?" Here, two of the four models were selected, and the Wilcoxon signed-rank test was performed. A significant difference was confirmed at a 5% level between Model 2 and Model 3 and between Model 2 and Model 4. In Model 2, the participants who responded that they felt that people in remote areas were moving momentarily (answered by 5 or more) had the following opinions.

- "It seems that footsteps were generated in an unnatural place."
- "It sounds like footsteps moving from the right front to the left back in an instant."
- "It seems that footsteps were interrupted on the way."

Table 2. Questionnaire

No.	Question
Q1	Did you feel that it was footsteps? (1. Not at all. - 7. Very much.)
Q2	Did this footstep make you feel that people in remote areas were moving momentarily? (1. Not at all. - 7. Very much.)
Q3	Please enter the reason for your response to Q2
Q4	Did you feel like several people were walking? (1. Not at all. - 7. Very much.)
Q5	Please enter the reason for your response to Q4

Table 3. Questionnaire administered after the experiment.

No.	Question
After-Q1	Which one model felt most realistic liveliness? (first, second, third, fourth, about the same)
After-Q2	Please enter the reason for your response to After-Q1

**Fig. 3.** Experimental environment. Participants sat at the center of the experiment space and answered questions while listening to footsteps output sequentially from multiple devices counterclockwise.

Figure 5 shows the results for “Q4. Did you feel like several people were walking?” Two of the four models were selected, and the Wilcoxon signed-rank test was performed. A significant difference was confirmed at a 5% level between Model 1 and Model 4. In addition, a significant difference was confirmed at a 1% level between Model 2 and Model 4. In Model 4, the participants who responded that they did not feel that several people were walking (answered by 2 and fewer) had the following opinions.

- “I felt that the footsteps were moving continuously.”
- “I felt that one person was walking around.”
- “I felt only one person's footsteps.”

Q2. Did this footsteps make you feel that people in remote areas were moving momentarily?

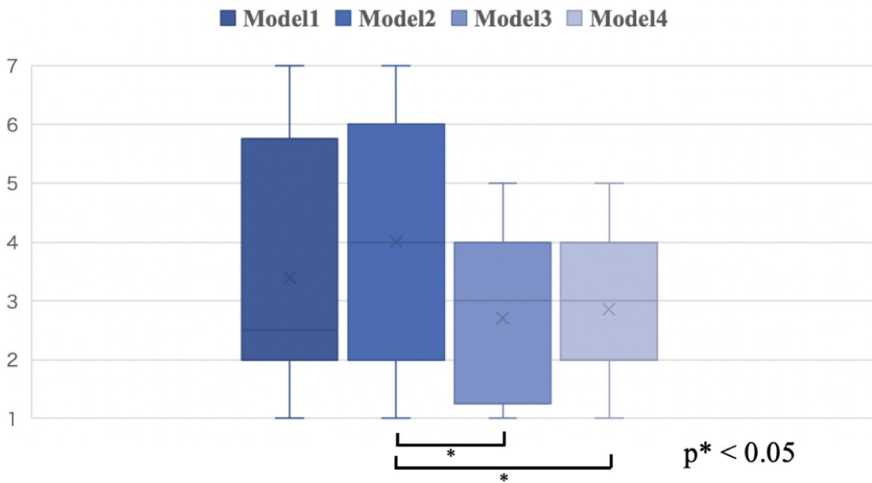


Fig. 4. Q2 results.

Results of After Questionnaire. Figure 6 shows the result for “After-Q1. Which one model felt most realistic liveliness?” The number of participants who answered Model 4 was the greatest (50% of the total number of participants). The reasons provided for selecting Model 4 were as follows.

- “Movement of footsteps was easy to understand and there were no elements that felt particularly unpleasant.”
- “I felt that one person was walking around.”
- “There were few overlapping footsteps sounds.”

Q4. Did you feel like several people were walking?

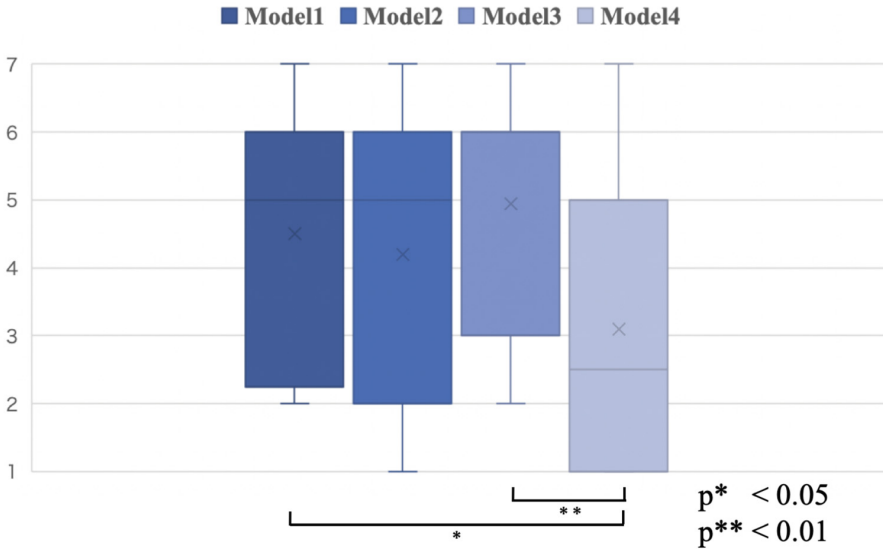


Fig. 5. Q4 results.

3.4 Best Model

The results provided in Sect. 3.3 are summarized as follows.

- Compared to Model 2, Models 3 and 4 did not sound like moving instantaneously.
- Compared to Models 1 and 3, Model 4 did not feel that some people were walking.
- Model 4 felt most realistic liveliness.

Notably, the reason for these results was the difference in volume of the first 1.455 s (half of 2.91 s). According to Fig. 2, the volume up to 1.455 s is in the following order: *Model3* > *Model4* > *Model1* (\approx *Model2*). In the case of Model 3, the volume power is greater than 0.3 in 1.455 s; thus, the footsteps sounds overlap with the footsteps sounds of the next speaker, thereby appearing like there are several people. Conversely, in the case of Model 1, the volume power is smaller than 0.2 for about the first 0.7 s; thus, it does not sound like movement between the speakers, but it sounds like a different person is walking. Therefore, Model 4 is considered to be the best among four models. In the follow, we describe FootstepsMixer, the tool we developed a tool to express footsteps of multiple people based on Model 4.

After-Q1. Which one model felt most realistic liveliness?

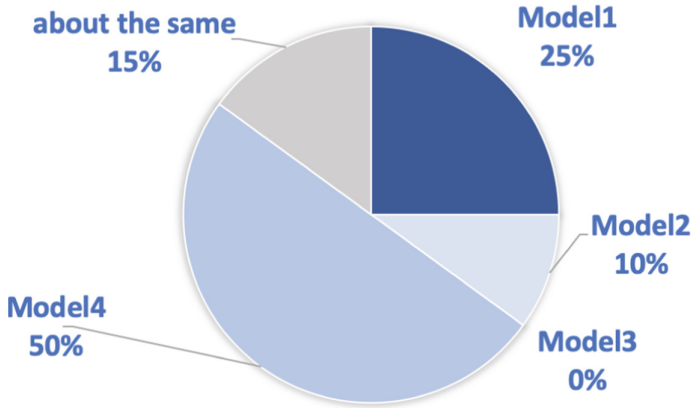


Fig. 6. Results of questionnaire administered after the experiment (see Table 3).

4 FootstepsMixer

4.1 Problems Expressing Footsteps of Multiple People

According to Sect. 3, when transmitting a single person’s footsteps to a remote space, it is possible that listener felt realistic liveliness because of the impression that he/she is nearby. However, when transmitting multiple people’s footsteps to a remote place, balancing the volume of each footstep is problematic. In this study, we needed to transmit a sense of liveliness as well as motion, but when multiple people’s footsteps are played simultaneously from a single playback device, there is a possibility that sound cannot be discerned as footsteps. In addition, it may be necessary to change the mixing balance depending on the number of people in the remote space. To address these problems, we developed the FootstepsMixer tool that expresses the movements of multiple people. In the following sections, we describe FootstepsMixer’s features and system configuration.

4.2 FootstepsMixer Features

Initially, the FootstepsMixer user sets the number of “walkers” in the space. To generate an audio stream to be played from one of the system speakers, FootstepsMixer calculates the positions of the walkers based on the walking speed defined in the system, calculates the intensity of each walker’s footsteps based on the distance between the walker and the speaker, generates the footstep sound by repeating a single prerecorded footstep, and mixes the footstep sounds of the walkers at an intensity that corresponds to each of the walkers. Note that the intensity is calculated based on Model 4. The implemented system has four

speakers, and this calculation and generation of sound based on a prerecorded footstep is repeated for each of the speakers.

In addition to the basic generation based on the Model 4, FootstepsMixer includes some adjustment functions that can be used to express multiple walkers' footsteps effectively. These functions include the following.

- The user can set footstep exaggeration (weighting).
- The user can set the number of people and the maximum volume of each footstep.
- The user can preview how each of the speaker sounds.

4.3 FootstepsMixer System Configuration

A screenshot of FootstepsMixer's control panel is shown in Fig. 7. FootstepsMixer can generate footsteps of multiple walkers by mixing footstep sounds according to their position in the space. As shown in Fig. 7, the FootstepsMixer user interface is divided into three areas, i.e., Floor Area, Sensor Select Area, and User Select Area.

Floor Area. The square frame in this area indicates the space in which the footstep transmission system can be used. This area plays a role like an execution screen; however, the user cannot set anything in this area. The four circles in the square frame represent the speakers. Here we refer to the speakers as “sensors” to indicate that they represent the audio sensors (microphones) at the remote site. As the user select a given sensor in the Sensor Select Area, the corresponding circle is displayed in dark blue and the corresponding circle for the sensors not selected are displayed in light blue. The user can set the number of walkers in the space in User Select Area; human-shaped icons representing the number of walkers are displayed in the square frame. In Fig. 7, the number of walkers is set to three. The human-shaped icons move in the square frame at random speeds.

Sensor Select Area. In this area, the user can select one of the sensors to playback the sound. Only one sensor at a time can be selected. In Fig. 7, Sensor1 is selected; therefore, in the Floor Area, the upper left circle is displayed dark blue, and the audio stream corresponding to Sensor1 is played back.

User Select Area. In this area, the user can set the intensity of exaggeration (i.e., weighting), the number of walkers, and the maximum volume of each walker's footsteps. To facilitate the differentiation of closely located multiple walkers, the system makes one of the footstep sounds louder than the others. Notably, the intensity of exaggeration refers to the extent to which the footsteps of the walker closest to the sensor are louder than other footsteps.

The algorithm used to calculate the percentage of exaggeration is explained in the next section. The number of walkers (one to five) can be selected. The user can hear footsteps for the selected number of walkers and can set the maximum volume of each footstep. Volume can be from -18 dB to 0 dB. The mixed footsteps change dynamically each time the user changes the setting, and the user can adjust the setting based on a preview playback.

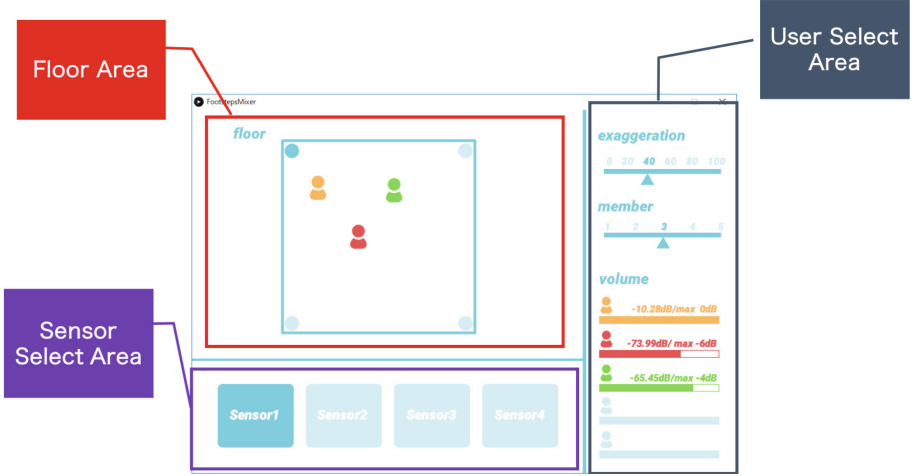


Fig. 7. FootstepsMixer. (Color figure online)

4.4 Algorithm

Here, we will explain the algorithm with and without exaggeration.

Based on the results given in Sect. 3, we use the footstep expression method of Model 4. The number of walkers who create footsteps is represented as n ($1 \leq n \leq 5$). When the distance between sensors is L (pixel), the distance between the sensor and the foot is d_n (pixel), and the maximum volume is M_n (dB). The volume $G(d_n, M_n)$ can be expressed as follows.

$$G(d_n, M_n) = -36 \frac{d_n}{L} + M_n \quad (4)$$

The equation representing d_n (pixel) with the coordinates ($sensor_x, sensor_y$) of any sensor in the Floor Area and the coordinates of the human-shaped icon ($human_x, human_y$) is as follows.

$$d_n = \sqrt{(sensor_x - human_x)^2 + (sensor_y - human_y)^2} \quad (5)$$

When the number of walkers is set to two or more, this algorithm compares the distance between the sensor and the walkers, and exaggerates the footsteps

of the walker closest to the sensor. When w is the exaggeration rate ($0 \leq w \leq 1$), the volume $G(d_n, M_n)$ is expressed as follows.

$$G(d_n, M_n) = \begin{cases} -36\frac{d_n}{L} + M_n & (d_n = \min D) \\ (-36\frac{d_n}{L} + M_n)\frac{1}{w} & (else) \end{cases} \quad (6)$$

Here, a set of $d_n (1 \leq n \leq 5)$ is represented as D .

The user can adjust the exaggeration rate in the User Select Area after pre-listening to the footstep sounds. However, in the current implementation, each time the distance between the sensor and the walker changes, it takes time for the system to determine the walker closest to a sensor. To compensation for this limitation, the exaggeration rate w can only be adjusted in six stages (see Fig. 7).

5 Discussion

5.1 Further Use of the Proposed FootstepsMixer

Currently, FootstepsMixer is designed to be used as a tool to explore ways of expressing footstep sounds effectively such that people can sense remote members of their family. Finding a way to control the sound would represent an improvement of the current implementation; however, even if we do not, the tool is can realize our initial objective. To support this contention, we provide the following two scenarios.

Scenario A: Father Wants to Know About the Life of His Two Daughters Who Live at a Distant. Due to an unavoidable circumstance related to his work, the father lives separately from his family (his wife and two daughters). To know that they are doing okay, the father wants to hear his daughters' footsteps rather than those of his wife. Thus, he sets the volume of footsteps of his daughters to be louder compared to that of his wife. Even if the daughters' footsteps are heard simultaneously, he does not perceive the sound as noise; therefore, he sets the system not to exaggerate.

Scenario B: Grandmother Wants to Know the Life of Her Family. A grandmother is living separately from five family members, i.e., her daughter, daughter's husband, and three grandchildren. She wants to know about her grandchildren's every day. Therefore, she sets their footsteps to be louder than those of her daughter or her daughter's husband. However, the sound is a bit noisy because her grandchildren run a lot. Consequently, to compensate for the noisiness, she set the footsteps to exaggerate, which allows her to perceive her grandchildren's liveliness.

5.2 Exaggeration Rate

In the current implementation, the degree of exaggeration was limited to six stages (0%, 20%, 40%, 60%, 80%, 100%). The purpose of this study is to enable people who are listening to the footsteps to understand the liveliness and activities of people in remote areas. Therefore, we assumed that a fine adjustment of the degree of exaggeration was not necessary. However, it is still necessary to consider whether six steps are appropriate. In addition, we recognize that we need to develop a better exaggeration algorithm.

5.3 Limitation

The newly developed FootstepsMixer allows the user to set parameters to control the simultaneous playbacks of multiple walkers' footsteps. The current system limited to a setting in which sensors are placed at the corners of a square. Based on feedback from actual users, it will be necessary to provide additional setting options, including a different arrangement of sensors.

6 Future Work

6.1 Application to Footstep Transmission System

We intend to apply the proposed method to a footstep transmission system that records the activity of people at a remote site and reproduces their footstep sound. This system will transmit data about people's position captured by the foot activity recording device, such as position sensors, and will play footstep sounds based on the position data.

6.2 Long-Term Experiment in an Actual Residence

We need to perform experiments using the footstep recording transmission system in an actual home. By conducting experiments over a long period, we would like to investigate how much awareness can be felt based on footstep activity in a home. Using the tool in an actual home may reveal previously unidentified problems and identify ways to improve the proposed tool.

7 Conclusion

In this study, we attempted to realize a system that shares the living sounds of a household of a distant family to create the impression that the household is nearby. To date, we have not been able to express the activity and liveliness of several people; therefore, we proposed the FootstepsMixer tool that expresses the activity of several people using the footstep expression method we devised. Using this tool, users can easily express footsteps exaggeratingly, and can set the number of people and the maximum volume of each footstep. In future, we would like to apply this tool to a footstep transmission system and conduct long-term experiments in an actual home.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Number JP18K11410. The authors would like to thank Enago (www.enago.jp) for the English language review.

References

1. Chamie, J.: The Rise of One-Person Households. IPS. <http://www.ipsnews.net/2017/02/the-rise-of-one-person-households/>. Accessed 20 Apr 2019
2. OECD: THE FUTURE OF FAMILIES TO 2030 A Synthesis Report. OECD (2011)
3. Dourish, P., Bly, S.: Portholes: supporting awareness in a distributed work group. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1992), pp. 541–547. ACM, New York (1992). <https://doi.org/10.1145/142750.142982>
4. Robert, E.K., Robert, S.F., Barbara, L.C.: The VideoWindow system in informal communications. In: Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work, CSCW 1990, pp. 1–11. ACM, New York (1990). <https://doi.org/10.1145/99332.99335>
5. Ishii, H., et al.: ambient- ROOM: integrating ambient media with architectural space. In: CHI 1998 Conference Summary on Human Factors in Computing Systems, CHI 1998, pp. 173–174. ACM, New York (1998). <https://doi.org/10.1145/286498.286652>
6. Siiio, I., Rowan, J., Mima, N., Mynatt, E.: Digital decor: augmented everyday things. In: Graphics Interface 2003, pp. 159–166 (2003). <https://doi.org/10.20380/GI2003.19>
7. Rowan, J., Mynatt, E.D.: Digital family portrait field trial: support for aging in place. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, CHI 2005, pp. 521–530. ACM, New York (2005). <https://doi.org/10.1145/1054972.1055044>
8. Tsujita, H., Tsukada, K., Siiio, I.: SyncDecor: appliances for sharing mutual awareness between lovers separated by distance. In: CHI 2007 Conference Proceedings and Extended Abstracts, Conference on Human Factors in Computing Systems, pp. 2699–2704. ACM, New York (2007). <https://doi.org/10.1145/1240866.1241065>
9. Cohen, M., Koizumi, N., Aoki, S.: Design and control of shared conferencing environments for audio telecommunication. In: Proceedings the International Symposium on Measurement and Control in Robotics, Society of Instrument & Control Engineers, pp. 405–412 (2005)
10. Seligmann, D.D., Mercuri, R.T., Edmark, J.T.: Providing assurances in a multimedia interactive environment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1995, pp. 250–256. ACM, Denver (1995). <https://doi.org/10.1145/223904.223936>