Cognitive Computation and Systems

Research Article

IET Journals
The Institution of Engineering and Technology

Retrieval and management system for layer sound effect library

eISSN 2517-7567
Received on 22nd June 2020
Revised 19th August 2020
Accepted on 28th September 2020
E-First on 16th November 2020
doi: 10.1049/ccs.2020.0027
www.ietdl.org

Jiale Yang[1], Ying Zhang[1] ✉, Yang Hai[1]

[1]Shanghai Film Academy, Shanghai University, Shanghai 200072, People's Republic of China

✉ *E-mail: luciezy@shu.edu.cn*

Abstract: Here, the authors present a novel interactive prototype system that enhances the effectiveness and ingenuity for sound designers to explore the sound effect library created by layering in multi-methods. They combine the explored methods of semantic keyword, acoustic feature, and layer relationship. In particular, the system visualises the layer relationship via circle pack, which facilitates the sound designers' understanding on the components of the mixed sound effect by the designed layer and sourced layer. In order to evaluate the proposed method, they conduct a timing experiment along with a five-point Likert scale survey to analyse the searching efficiency, the user experience, and the interactive user behaviours. The studies performed by the authors show that the proposed system is capable of enhancing the sound designers' ability for sound effects searching, thus creating new possible combination and design.

1 Introduction

The effective retrieval of a target sound effect in the vast sample library for the production of film, game, and music has received considerable scholarly attention. For instance, Font *et al.* [1] conducted a review of the related research on sound sharing and retrieval and focused on improving the reusability of sound effects. Most of the studies in the design of the sound effect retrieval system focus on improving the efficiency of search. However, sound designers also need some creative inspirations when they have no idea about their target sound effect. In this study, we investigate the sound effect retrieval in terms of efficiency and creativity. We aim to provide some design thinking when searching for a sound effect.

In the process of sound design, layering different sound effects to compose a new sound effect is a common way used by the sound designers [2]. In this study, layer designed (LD) refers to the LD by changing volume, inserting plugin, or using other design tools. Layer sourced (LS) refers to the layer with no modifications on the original sample. The final sound effects mixed by the LD sound effects are called mixed sound effects. The LD and the LS are saved into the sound effect library. Generally, sound designers create a folder to keep each kind of sound effect or save the project from the digital audio station (DAW) directly for reuse and redesign in the future. However, the folder hierarchy is destroyed after keyword searching in the traditional sound effect retrieval system. For reusable purposes, the project from DAW is not convenient for the management and retrieval of sound effects. Thus, we propose a scheme to save the layer information and visualise the layer relationship (LR) between sound effects to let sound designers reuse the ideas of layering and share these ideas with others easily. In this proposed scheme, sound designers can make utterly different sound effects or similar variance with the layer information shown in the system.

We classify the methods used to explore sound effects into two main categories: semantic description and acoustic feature (AF) similarity. The former is the most commonly utilised search method in sound effect retrieval systems based on the semantic list. Some of the examples under the semantic description category are Soundminer, BaseHead, and AudioFinder. Sound designers input or select keywords to describe their target and filter the unnecessary sound effects. However, poorly described keywords result in low search efficiency of the sound effect library. The latter is often applied to the visualisation systems based on the 2D scatter plot [3]. Some of the examples under the AF similarity are Freesound Explorer [4, 5], Audio Quilt [6], and Sound Torch [7, 8]. These systems represent the similarity of sound effects in the AF by using distance, such as Euclidean distance between points to help sound designers find similar sound effects quickly [9, 10]. Furthermore, the application of the AFs in sound effect retrieval benefits from the development of music information retrieval. In music production, using the 2D or 3D visualisation in retrieving instrumental samples can improve the overall exploration experience [11–13]. Dupont *et al.* [14] presented an interactive prototype tool for browsing sound libraries by similarity. Urbain *et al.* [15] combined the semantic keyword (SK) and AF to design a system that uses list and scatter plot to present the sound effects for browsing extensive collections of Foley sound effects. The function of the keyword filter consists of the tag cloud and facet fields. Okamoto *et al.* [16] developed a system that applies three types of similarities: context, AFs and the symbol of onomatopoeia. In accordance with the above research, we find that combining multiple retrieval methods can help sound designers find their target sound effects. Therefore, we combine these two methods in the system.

In terms of presenting the LR, we utilise the circle pack to show the hierarchy information between the mixed sound effect and the layers. Lafay *et al.* [17] used the circle pack to visualise the semantic relationship between the sound effects without keywords. Following the above research, we consider the circle pack a good way to present hierarchy information. They separately compared the scatter plot based on the AF similarities with the semantic circle pack. However, a single method limits the ability of sound designers to search for sound effects. We combine these two components in the proposed system. We find that the visualisation, including colour, size and structure, can facilitate people to memorise the spatial information [18–22]. Thus, we change each circle in the 2D scatter plot into a circle pack wherein the outside circle and the inside circles denote the mixed sound effect and the LD and the LS sound effect, respectively. Fig. 1 shows the visualisation from the digital audio workstation to the circle pack. The three-layered tracks below correspond to the inside circle in the circle pack. Specifically, one inside circle represents two kinds of sound effects that can be played via different interactions. The mixed track above corresponds to the outside circle in the circle pack.

We divide the innovation points of this work into three aspects. Firstly, we originally use the circle pack to visualise the LR. By

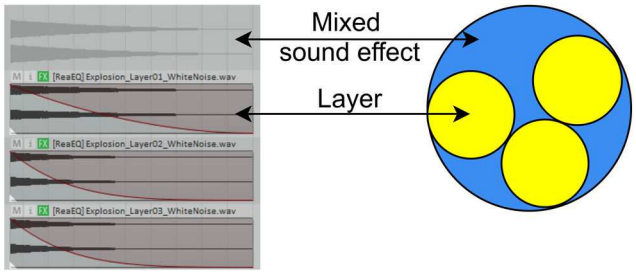

Fig. 1 *Layer visualisation from digital audio workstation to circle pack*

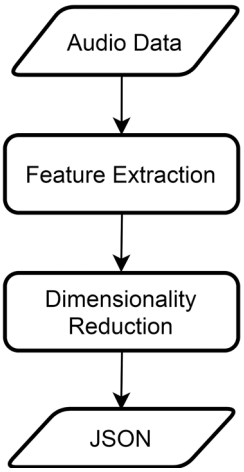

Fig. 2 *Preprocessing audio data before visualising into the user interface of the system*

doing so, sound designers can find the target sound effect and think of a layered design. We also focus on the ability of sound designers to search for sound effects by using single or multiple methods.

The remaining part of this study is organised as follows. We introduce the system architecture in Section 2. Then, we illustrate the main idea about how to use the proposed system in Section 3. Next, we evaluate the proposed system and present the analysis results in Sections 4 and 5, respectively. Subsequently, we discuss the outcome through user observation in Section 6. Finally, we provide our conclusion and recommendations for future work in Section 7.

2 System architecture

2.1 Data preprocessing

To utilise the information about sound effects, we need to preprocess the data and save the results into a JSON file, which consists of the LRs, the names, the design notes and the coordinates in the 2D plane after reducing the dimensionality. Fig. 2 shows the procedure of the audio data preprocessing before the visualisation of information into the user interface.

The dataset we used in the experiment consists of 164 sound effects, including 20 mixed sound effects, 72 LD sound effects, and 72 LS sound effects. To differentiate between these two types, the names of the LD and the LS sound effects are added with the prefix of 'LD' and 'LS', respectively. We also use red colour to make them stand out in the user interface. The length of all the sound effects varies from 0.1 to 3 s.

Firstly, in the system, the AFs we extracted via the Python library of Librosa are Mel Frequency Cepstral Coefficients (MFCCs), which have been utilised to achieve the tasks of speech recognition and timbre classification [23, 24]. In the proposed system, we only apply the similarity of MFCCs to the mixed sound effects.

Secondly, to visualise the sound effects on the 2D plane, we need to reduce the dimensionality by using the machine learning algorithm. For the application of dimensionality reduction algorithms in the system, we consider four types of algorithms, namely, principal component analysis (PCA) [25], multidimensional scaling (MDS) [26, 27], *t*-distributed stochastic neighbour embedding (*t*-SNE) [28, 29], and uniform manifold approximation and projection (UMAP) [30], which can preserve the structure between sound effects in the high-dimensional space whilst projecting the data on the 2D plane. These algorithms belong to the unsupervised category. Thus, we do not need to annotate the sound effects manually. Our analysis of these four algorithms is listed as follows:

(i) PCA is a linear dimensionality reduction algorithm that aims to map high-dimensional data to low-dimensional space through some linear projection. It also expects the variance of the data to be the largest in the projected dimension by using few data dimensions whilst retaining the characteristics of multiple original data points. However, after the projection, the distinction of the data is not apparent.
(ii) MDS is also a linear dimensionality reduction algorithm that preserves the interdistance between points. However, it is not still conducive to find the best structure when the data is from multiple categories.
(iii) *t*-SNE is a non-linear dimensionality reduction algorithm that can keep the local and the global structures concurrently. However, it consumes considerable time when dealing with numerous dataset.
(iv) UMAP is a non-linear dimensionality reduction algorithm that has a better performance in keeping the local and the global structures and consumes less time than other algorithms.

The visualisation of data structure and the operation time are the main components we value in the proposed system. Therefore, we select UMAP to reduce the dimensionality and obtain the coordinates of each point. The neighbours and distance are two components that would influence the final distribution to the 2D plot. The former represents the number of nearest neighbours. The latter refers to the distance from each *i*th sound effect to its first nearest neighbour. To utilise the parameters, we follow the rule that the points should have a certain distance between each other to avoid overlapping in the final user interface, as the overlap will influence users to play the sound effects individually.

Finally, we save the names, the coordinates, the design notes and the LRs to the JSON file.

2.2 User interface

The interactive retrieval system is mainly written in D3 and Webaudiox, which perform user interaction (UI) and audio operations. Fig. 3 shows the final user interface, which is a single page application with the following three modules:

Graphic visualisation module: This module shows the layer information and the similarity of the AF between mixed sound effects. Sound designers can hover the cursor on a circle and click the circle to play the mixed and the LD sound effects. By hovering on the circle, they can also obtain the corresponding name of the sound effects, which is linked with the fundamental information module. The LS sound effects are played by hovering the cursor on a circle and pressing the spacebar on the keyboard. If sound designers want to stop playing the sound effect, they need to press the 'ESC' button on the keyboard. By using the mouse's scroll wheel, sound designers can zoom in and out of the scatter plot. They can also pan by dragging and pressing the left mouse button when the target is out of the vision by zooming in too much.
Semantic retrieval module: Sound designers can type in the keywords to filter the library, including the function of 'AND' and 'OR' search. The list consists of interactive buttons with the name of the sound effects, which can be played by clicking. The blue buttons represent the mixed sound effects, whereas the yellow buttons denote the LD and the LS sound effects. Sound designers can use the 'DEL' button on the right side to delete the sound effects from the filter list without influencing the graphic visualisation module. If they delete the related sound effect wrongly, they can click the 'Search' button again to find the sound effect back. Each time, they filter the list by using the search bar

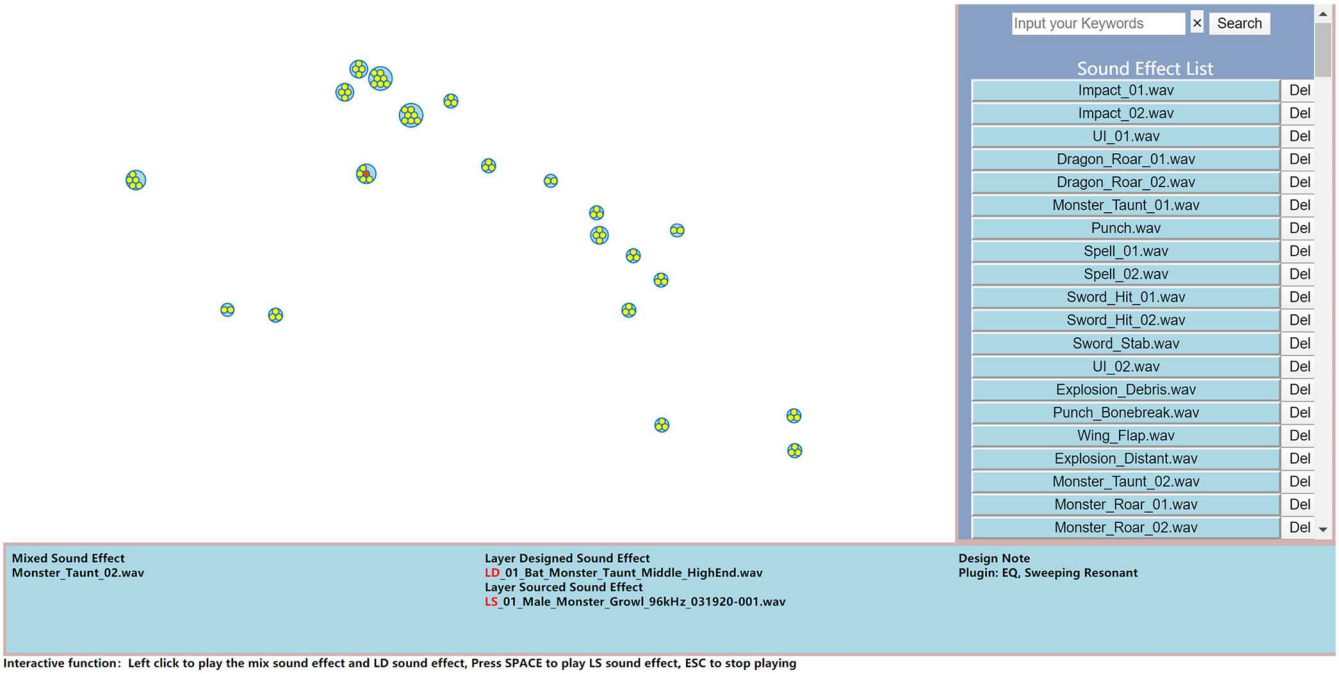

Fig. 3 *User interface of the retrieval and management system for layer sound effect library*

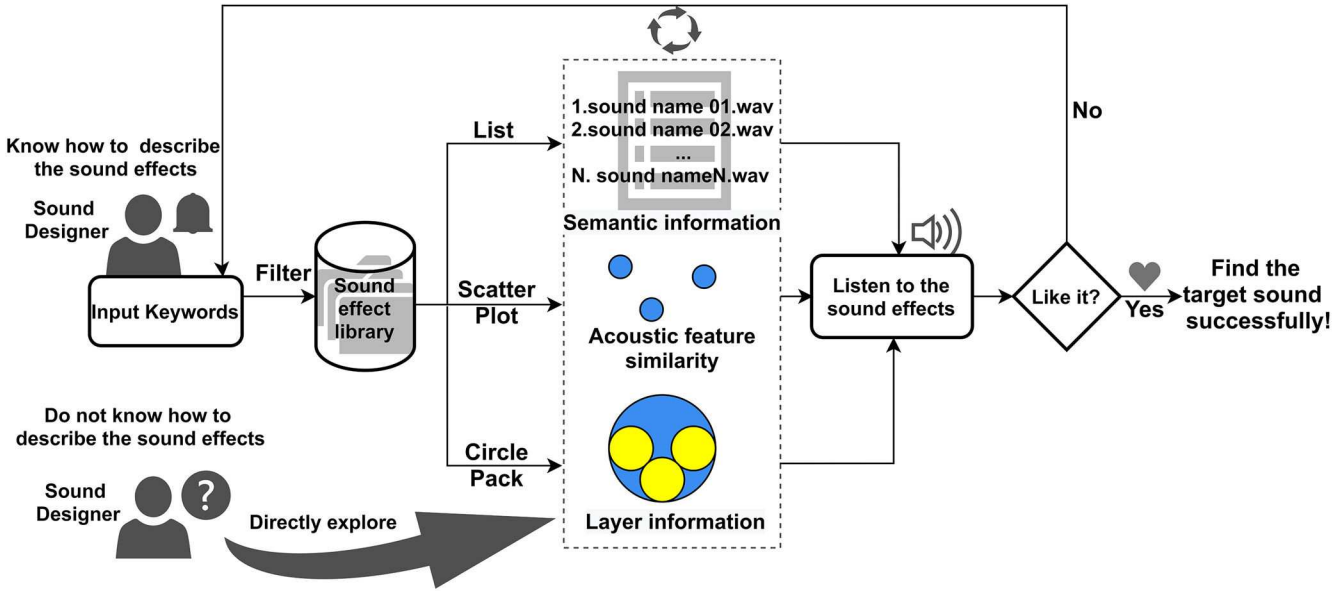

Fig. 4 *Sound effect search process proposed in this paper*

and change the opacity of each point in the visualisation module. The opacity of related sound effects will be maintained, but others will become lighter.

Fundamental information module: This module linked with the visualisation module presents the naming information and the design note when sound designers hover on the scatter plot and the circle pack.

3 Searching method of the system

Fig. 4 shows the process of exploring sound effects in two situations: the first situation implies that sound designers know how to describe the sound effect by using keywords, whereas the second situation implies that they have no idea how to describe the sound effect by using keywords. In the former case, sound designers search the sound effect from the visualisation after filtering the sound effect library by using SKs. In the latter case, they directly explore the sound effect from the viewing. Therefore, in the whole process, sound designers in both situations employ one of the following visualisations: list, scatter plot and circle pack.

In this study, we propose to preview the mixed sound effects preferentially. According to the cocktail effect, people focus on the target sound and neglect the unrelated sound. We employ this effect to help sound designers effectively identify the LD sound effects by listening to these sound effects together. Consequently, they find that the mixed sound effect has the target sound effect of the LD or the LS. As shown in Fig. 5, sound designers can check the location where the colour of the circle pack turns red for a second by clicking the button with the corresponding name on the list. then, they can preview the sound effects that have the LR with the mixed sound effect in the circle pack. If the sound effect is not the final one, they can click the 'DEL' button on the right of the sound effect button. Deleting unrelated items is helpful for them to avoid listening to the same sound effect repeatedly.

Overall, we classify the search methods into three types. Sound designers can apply one of these ways or combine them to realise the corresponding destination:

Semantic keyword (SK): Sound designers describe the sound effects by using keywords directly and utilise the semantic list to

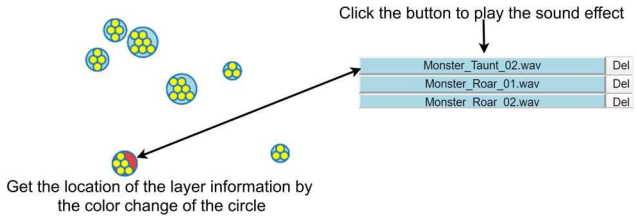

Fig. 5 *Connection between the graphic visualisation module and the semantic retrieval module*

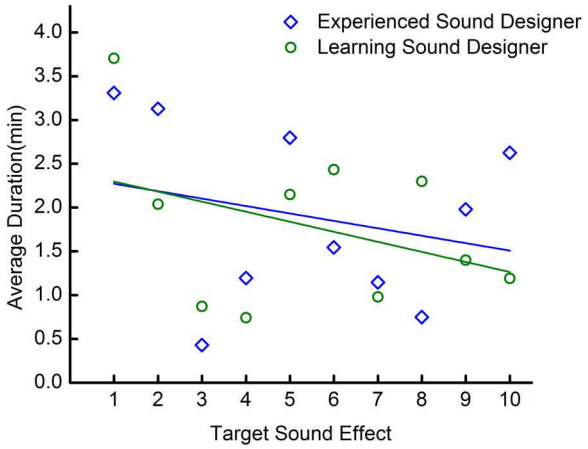

Fig. 6 *Progression of the average duration of each target sound effect retrieval observed for experienced sound designers and learning sound designers with the corresponding linear regression*

find the target efficiently, which is a traditional method in the industry.

Acoustic feature (AF): Relying on the similarity of the mixed sound effects is an effective method to filter unrelated sound effects in terms of the audio impression. In this study, the method focuses on searching the mixed sound effect in the system.

Layer relationships (LRs): The layer information includes the name and the design note of the sound effects in the mixed sound effect, which can be used to judge what kind of sound effects are designed in the mixed sound effect. Moreover, sound designers can use the number of the yellow points in the circle pack to know the complexity of the mixed sound effect.

4 System evaluation

4.1 Evaluation designed

We divide the evaluation into three parts: learning system, timing experiment and subjective assessment. In total, eight participants took part in the experiment, four sound designers have a year of industrial experience in sound design and four sound designers are still learning in the school. We observe whether the industrial experience is a threshold for utilising the system and making difference between the sound designers.

4.1.1 Learning system: In this part, we teach the participants to use the system. To avoid the participants to remember the dataset of timing experiment, we set a different dataset, which consists of 2 mixed sound effects, 13 LD sound effects and 13 LS sound effects. The participants need to find one LD sound effect and one LS sound effect to practice the system, and they can ask any questions whilst learning the system. This experiment lasts about 20 min.

4.1.2 Timing experiment: The participants need to search ten target sound effects, including eight LD and two LS sound effects. The first and the fifth are the LS sound effects. Each target sound effect is previewed before starting the exploration. The name and waveform are not provided to them. They have to judge the type of the target. Then, they can apply different retrieval methods to explore the target sound effects and ask questions about the system during the timing experiment. The time of finding each target sound effect is limited in 8 min. If the duration of the retrieval is beyond the limit, the participants must execute the next search directly. This part lasts about 40 min. In this experiment, two types of data that we observe and record are listed as follows:

- The duration of each exploration, which includes the time that the participants ask the question during searching.
- The frequency of applying and combining different methods to search target sound effects. There are seven different combinations that consist of semantic keyword (SK), acoustic feature (AF), and layer relationship (LR).

4.1.3 Subjective assessment: We design a five-point Likert scale questionnaire to survey the participants' user experience based on the learning experiment and timing experiment. The score of 1–5 denotes strongly disagree to strongly agree. There are ten questions out of order in the questionnaire. Five questions refer to evaluate the feature of the functional diversity. Five questions refer to evaluate the assist feature in the procedure of the sound design. The question with the same answer will be excluded from the results, which means the question is meaningless. The final mean of the dimensional score S_M is obtained by:

$$S_M = \frac{\sum_{n=1}^{N} S(Q_n)}{N} \quad (1)$$

where $S(Q_n)$ is the average score of the corresponding question, N is the total number of the dimensional questions. If the dimensional score's final mean is higher than 3, it shows the participants have a positive attitude to the system.

4.2 Hypotheses

We propose the following hypotheses before the evaluation:

(i) The overall average duration of exploring the target sound effect is <30% of all the sound effects (12 min): 3.6 min.

(ii) Whilst using the system, the participants are capable of applying multiple methods to search the target sound effect, and they rely more on the LRs.

(iii) The graphic presentation provides some new design schemes to the sound designers and improves the sound designers' ability to understand the LR during exploration.

5 Results and analysis

5.1 Results on timing experiment

Fig. 6 presents the progression of the average retrieval duration of each target for experienced sound designers and learning sound designers. The linear regression of the data shows that efficiency increases with the growth of the index. The average duration at the first target sound effect is the highest one as the participants are not familiar with the system and the dataset. Therefore, the familiarity with the system and the dataset is a critical component whilst using the system. The industrial experience has no significant influence on applying the system in terms of the search, which means the proposed system has no industrial experience threshold. As presented in Fig. 7, the overall average duration of retrieval of all the participants conforms to the first hypothesis, which shows the system has an acceptable efficiency in searching sound effects.

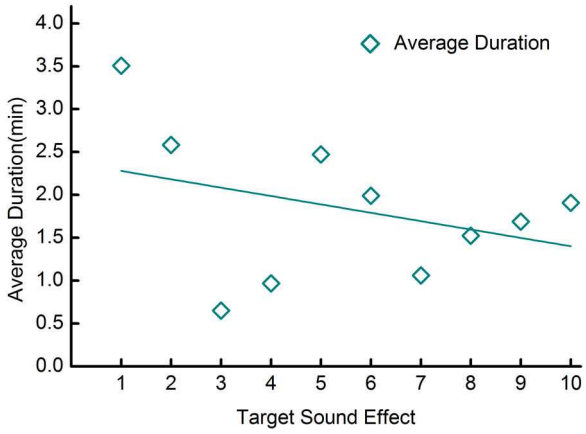

Fig. 7 *Progression of the overall average duration of each target sound effect retrieval with the linear regression*

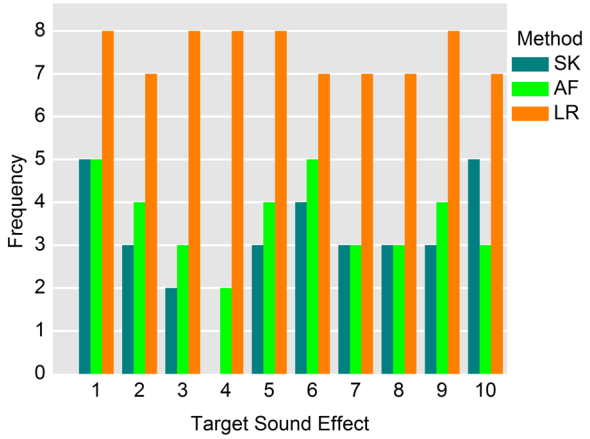

Fig. 8 *Frequency of each method used in the retrieval for the target sound effects*

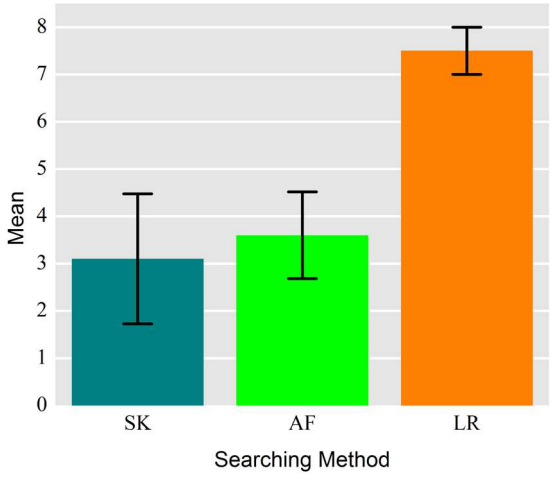

Fig. 9 *Mean of the frequency using different methods in the participants*

Fig. 8 shows the frequency of each method used in the retrieval for the target sound effects. In the fourth target search, the SK is not used by any participants, which shows the participants are not limited in the traditional method, and they can switch the searching style between different ways. As shown in Fig. 9, the mean of the frequency using LR is up to 7.5 (93.75%) out of 8, which is in line with the second hypothesis. Also, it makes sense that the system is designed for layer sound effect retrieval and management. The average frequency of using the semantic information and AF is 3.1 (38.75%) and 3.6 (45.00%), respectively. These results confirm that the LR is a prioritised choice and a significant method to search the target in the experiment in most cases.

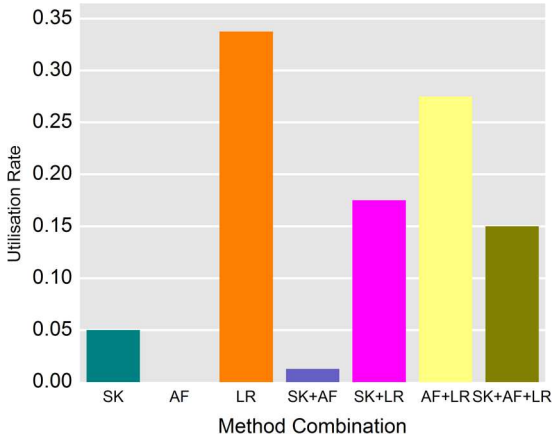

Fig. 10 *Utilisation rate of different methods combination*

We compared the utilisation rate that the participants utilised different combinations of the methods to explore the target sound effect, as shown in Fig. 10. Using only the LR to find the target sound effects occupies the highest utilisation rate, which is 0.3375 out of 1. Also there is zero in utilising only the AF as the experiment has no task of finding mixed sound effects. When the participants have an idea about describing the target, they can search it directly by filtering keywords and preview efficiently. However, this situation happened at a low utilisation rate (0.05) in the experiment. In most cases, the participants apply two methods to search the target sound effect more frequently, compared with one method or three methods ($M2 = 0.4625 > M1 = 0.3875 > M3 = 0.15$).

5.2 Results on subjective assessment

We excluded three questions with the same answer after the assessment, one question from the dimension of the functional diversity and two questions from the dimension of the auxiliary design feature. The final results are shown in Table 1, the first set of questions from Q1 to Q4 were designed to assess the functional diversity and the second set of questions from Q5 to Q7 were set to evaluate the assist feature in the designing procedure.

Table 1 illustrates that the participants have a positive attitude with the system. The average score of each dimension is higher than 3, which makes sense to the second and the third hypotheses. However, there are some specific questions with a low score, such as Q2–Q5.

In terms of Q2, we find that the function of the 'DEL' button designed beside the sound effect cannot memorise the delete history for the second keywords filter. For example, the participants filter the library one time and delete the sound effects from the list via 'DEL' button, then they filter the library for the second time, the sound effects that have been deleted at the last filter search shown again, which adds more work to the sound designer to remove the sound effect one more time.

In terms of Q3, we find that the participants focus more on the graphic present and the hearing than the naming information. The visual distribution is easy to memory to this kind of user.

In terms of Q4, the function of 'AND' and 'OR' search relies on whether the sound effect library is described sufficiently. If the keyword is not included in the semantic library, the user will fail in search.

In terms of Q5, although the structure of the circle pack supplies the information like the number of samples and the design note documented in the file, it lacks distribution of the waveform for each layer sound effect, which causes the participants cannot judge which is the layer corresponding to the hearing from the mixed sound effect.

6 Discussion

Our observation of UI behaviour is explained as follows.

When using the proposed system for the first time, the participants tend to apply keywords to filter the library to search

Table 1 Five-point Likert scale survey with results

Question	1	2	3	4	5	Mean
Q1 Is the system helpful to find the target sound effect by three methods?	0	0	3	2	3	3.58
Q2 Is the system convenient to delete the unrelated sound effect in the list?	1	2	1	4	0	
Q3 Is the naming information important to the exploration?	0	1	0	5	2	
Q4 Is the function of 'AND' and 'OR' helpful to the exploration?	0	2	0	2	4	
Q5 Is the structure of the circle pack conducive to understand the LR in the mixed sound effect?	0	1	1	1	5	4.29
Q6 Does the layer information apply unexpected design idea?	0	0	2	2	4	
Q7 Is the graphic present convenient to manage the layer sound effect library?	0	0	1	3	4	

for the sound effects. This method is typically utilised in traditional software. For example, they would consider the material ('wood', 'metal', or 'water') or the action ('hit', 'punch', or 'flag') of the target sound effects. However, the library's naming standard is different between traditional systems, and some sound effects' description is not sufficient. Consequently, sound designers fail to search for the target sound effects easily. If they are not familiar with the descriptions, they cannot reuse the resource in the library plenty. Therefore, the target sound effect is difficult to obtain when sound designers use only keywords. Moreover, they would try to combine different methods to explore, which confirm that the design of a sound effect retrieval and management system needs to consider multiple features. The proposed system does not limit the flexibility and creativity of sound designers by providing various methods. Specifically, sound designers can utilise semantic description, AF, and layering thought as main routes. The results show that the participants combined these methods successfully.

The similarity of the AF represented by the distance is also hard for the participants to understand when using the proposed system for the first time, although they can find which area includes different kinds of mixed sound effects. The x-axis and y-axis have no meaning, which makes the participants feel confused about the distribution of the mixed sound effects. These findings suggest that specific information about the similarity of the AF is an important component of the system design.

The LS sound effects are challenging to search, as the participants are not the original designer of the sound effect library. After inserting multiple plugins, the LS sound effects can be changed into a different sound effect in terms of the audio impression. Therefore, this phenomenon suggests that the sound designers who made the layer sound effect library can search the LS sound effects efficiently. However, for sharing design ideas, the participants show a positive attitude as they observe the design note frequently to the LD sound effect that they never made.

The function of deleting unrelated sound effects in the list by hitting the 'DEL' button is not utilised sufficiently during the experiment, which causes the participants to consume plenty of time to listen to the same sound effects repeatedly. After noticing this feature, they exhibit efficiency in the experiment. This finding suggests that the function of deleting unrelated sound effects is not common in traditional systems. As shown in Table 1, the participants present a strong desire to obtain this function in the system design.

7 Conclusion and future work

This study proposes and investigates a new system to search and manage the layer sound effect library. The proposed system highlights efficiency and creativity. In terms of creativity, we study how to educate sound designers to design a new sound effect in the search process. Learning new design methods can make them productive at work. We also evaluate the possibility of the application of the proposed system in actual situations. The results indicate that the proposed system has an acceptable efficiency, and sound designers can combine multiple methods to search for the target sound effect. Through user observations, we confirm that the proposed system helps sound designers explore sound effects in multiple methods and supplies some design thoughts that can be used in the exploration process.

In future research, we hope to investigate the overlapping problem whilst visualising a massive layer sound effect library, making the design note and naming information editable, adding the waveform and the spectrum to visualise additional details and utilising the colour of the circle to present various acoustic information.

8 References

[1] Font, F., Roma, G., Serra, X.: 'Sound sharing and retrieval', in Virtanen, T., Plumbley, M.D., Ellis, D. (Eds.): '*Computational analysis of sound scenes and events*' (Springer, Cham, Switzerland, 2018, 1st edn), pp. 279–301

[2] Viers, R.: 'Sound design', in Somerviller, B. (Ed.): '*The Sound Effects Bible*' (Michael Wiese Productions, Studio City, CA, 2008, 1st edn.), pp. 165–173

[3] Ahlberg, C., Shneiderman, B.: 'Visual information seeking: tight coupling of dynamic query filters with starfield displays', in Bederson, B.B., Shneiderman, B. (Eds.): '*The craft of information visualization*' (Morgan Kaufmann, San Francisco, CA, 2003, 1st edn), pp. 7–13

[4] Font, F., Bandiera, G.: 'Freesound explorer: make music while discovering freesound!'. Web Audio Conf. WAC, London, UK, August 2017, pp. 1–2

[5] Font, F., Roma, G., Serra, X.: 'Freesound technical demo'. Proc 21st ACM Int. Conf on Multimedia MM '13. ACM Press, Barcelona, 2013, pp. 411–412

[6] Fried, O., Jin, Z., Oda, R.: 'Audioquilt: 2D Arrangements of audio samples using metric learning and kernelized sorting'. In NIME, Goldsmiths, University of London, UK., 2014, pp. 281–286

[7] Heise, S., Hlatky, M., Loviscach, J.: 'Aurally and visually enhanced audio search with soundtorch'. CHI'09 Extended Abstracts on Human Factors in Computing Systems, Boston, MA, USA., 2009, pp. 3241–3246

[8] Heise, S., Hlatky, M., Loviscach, J.: 'Soundtorch: quick browsing in large audio collections'. Audio Engineering Society Convention 125. Audio Engineering Society, San Francisco, CA, 2008, pp. 1–8

[9] Favory, X., Font, F., Serra, X.: 'Search result clustering in collaborative sound collections'. Proc. of the 2020 Int. Conf. on Multimedia Retrieval, Dublin, Ireland, June 2020, pp. 207–214

[10] Hemgren, D.: 'Fuzzy Content-Based Audio Retrieval Using Visualization Tools'. Master thesis, School of Electrical Engineering and Computer Science (EECS), 2019

[11] Turquois, C., Hermant, M., Gómez-Marín, D., *et al.*: 'Exploring the benefits of 2D visualizations for drum samples retrieval'. Proc. of the 2016 ACM on Conf. on Human Information Interaction and Retrieval, Carrboro, North Carolina, USA., 2016, pp. 329–332

[12] Nuanáin, C.Ó., Herrera, P., Jordá, S.: 'Rhythmic concatenative synthesis for electronic music: techniques, implementation, and evaluation', *Comput. Music J.*, 2017, **41**, (2), pp. 21–37

[13] Berthaut, F., Desainte-Catherine, M., Hachet, M.: 'Combining audiovisual mappings for 3D musical interaction'. Int. Computer Music Conf., New York, USA., June 2010, pp. 100–108

[14] Dupont, S., Frisson, C., Siebert, X., *et al.*: 'Browsing sound and music libraries by similarity'. Audio Engineering Society Convention 128, Audio Engineering Society, Novel London West, London, UK., 2010, pp. 1–7

[15] Urbain, G., Frisson, C., Moinet, A., *et al.*: 'A semantic and content-based search user interface for browsing large collections of Foley sounds'. Proc. of the Audio Mostly, Norrkoping, Sweden, 2016, pp. 272–277

[16] Okamoto, K., Yamanishi, R., Matsushita, M.: 'Sound-effects exploratory retrieval system based on various aspects'. *IEE J. Trans. Electron., Inf. Syst.*, 2016, **136**, (12), pp. 1712–1720

[17] Lafay, G., Misdariis, N., Lagrange, M., *et al.*: 'Semantic browsing of sound databases without keywords', *J. Audio Eng. Soc.*, 2016, **64**, (9), pp. 628–635

[18] Yang, J., Hermann, T.: 'Interactive mode explorer sonification enhances exploratory cluster analysis', *J. Audio Eng. Soc.*, 2018, **66**, (9), pp. 703–711

[19] Robertson, G., Czerwinski, M., Larson, K., *et al.*: 'Data mountain: using spatial memory for document management'. In Proc. of the 11th Annual ACM Symp. on User Interface Software and Technology, 1998, pp. 153–162

[20] Richan, E., Rouat, J.: 'A proposal and evaluation of new timbre visualization methods for audio sample browsers', *Pers. Ubiquitous Comput.*, 2020, pp. 1–14, Available at: https://doi.org/10.1007/s00779-020-01388-1

[21] Evreinova, T.V., Evreinov, G., Raisamo, R.: 'An exploration of volumetric data in auditory space', *J. Audio Eng. Soc.*, 2014, **62**, (3), pp. 172–187

[22] Adeli, M., Rouat, J., Molotchnikoff, S.: 'Audiovisual correspondence between musical timbre and visual shapes', *Front. Hum. Neurosci.*, 2014, **352**, (8), pp. 1–12

[23] McFee, B., Raffel, C., Liang, D., *et al.*: 'Librosa: audio and music signal analysis in python'. Proc. of the 14th Python in Science Conf., Austin, Texas, 2015, (8), pp. 18–24

[24] Zheng, F., Zhang, G., Song, Z.: 'Comparison of different implementations of MFCC', *J. Comput. Sci. Technol.*, 2001, **16**, (6), pp. 582–589

[25] Wold, S., Esbensen, K., Geladi, P.: 'Principal component analysis', *Chemometr. Intell. Lab. Syst.*, 1987, **2**, (1–3), pp. 37–52
[26] Cox, M.A., Cox, T.F.: 'Multidimensional scaling', in Chen, C., Hardle, W.K., Unwin, A. (Eds.): '*Handbook of data visualization*' (Springer, Berlin, Heidelberg, 2008, 1st edn), pp. 315–347
[27] Schwarz, D., Schnell, N.: 'Sound search by content-based navigation in large databases'. Sound and Music Computing (SMC)', 2009, pp. 1–1
[28] Maaten, L.V.D., Hinton, G.: 'Visualizing data using t-SNE', *J. Mach. Learn. Res.*, 2008, **9**, (Nov), pp. 2579–2605
[29] Van Der Maaten, L.: 'Accelerating t-SNE using tree-based algorithms', *The J. of Mach. Learn. Res.*, 2014, **15**, (1), pp. 3221–3245
[30] McInnes, L., Healy, J., Melville, J.: 'UMAP: uniform manifold approximation and projection for dimension reduction', *J. Open Source Softw.*, 2018, pp. 1–51, arXiv:1802.03426v3