Exploratory Searches for sound effects: Verification of similarity based on the acoustic features of sound effects

Kahori Okamoto[†] **Ryosuke Yamanishi**[‡] **Mitsunori Matsushita**[†*]
[†]Graduate School of Informatics, Kansai University
[‡]College of Information Science and Engineering, Ritsumeikan University
[*]mat@res.kutc.kansai-u.ac.jp

Abstract

The goal of this study is to develop an exploratory search system for sound effects (SEs). SEs appreciably influence viewers' impressions of a movie; thus, SE editors must skillfully select the most appropriate SEs for a given scene. However, existing SE search methods have three main difficulties, involving the diversity of SE purposes, the representation of sound using text, and the conceptualization of SEs for a given scene. These difficulties lead to inefficient SE searches, because the SE editor must perform repeated searches. To solve the problems, this paper proposes a framework to define similarities among SEs with three types of features: context, acoustic features, and symbols of onomatopoeia. As the first step in this study, SEs were clustered based on their acoustic features. The relationships between the clustered SEs and their onomatopoeia characteristics are acquired through subjective evaluation experiments. As a result, it was confirmed that the classifications could help determine the content of the SE.

Keywords: Sound effects, Music informatics, Onomatopoeia, Exploratory Search

1 Introduction

Sound effects (SEs) are essential elements in many movies. Assigning SEs to a scene is one of the most important steps in the movie production process; viewer impressions of the scene greatly depend on the SEs that are used. Thus, SE editors must prepare appropriate SEs for each scene to make a movie more engaging.

A common SE may be used in multiple scenes of a movie, as well as scenes in other movies. For example, "Rainy Sound" might be appropriate and available for any scene where it is rainy. From this characteristic, it is probable that many similar SEs would be stored in an SE database. In most movie composition scenarios, the SE editors retrieve SEs from a database. The editors then listen to each SE and compare them, to determine the appropriate SE for the scene.

An SE editor conceptualizes their idea of the SE for the scene, and retrieves the appropriate SE from a database using the SE name and description. However, retrieving SEs using only a name and description may be difficult for the editor; common SE names are assigned to many different sounds, while entirely different descriptions may be given to similar sounds. Moreover, without listening to the sounds, it is difficult to fully understand the tone color and nuance based only on the descriptions. Therefore, a significant amount of time may be required to listen to each sound individually; if the editor has only a vague idea of the SE required for the scene, additional search time may be required. To address these issues, we strive for developing a system that supports a user's efficient exploration of SEs.

This paper, as the first step in this study, focuses on the relationships between the acoustic features of SEs and the onomatopoeia given to the SEs. Hierarchical clustering of SEs with acoustic features is conducted. We discuss the relationships between the clustered SEs and their representative onomatopoeia information.

2 Difficulties in SE search activities

SE searches consist of three processes: selection, comparison, and decision. To execute the processes, the searcher will face the following three difficulties.

Diversity of SE purposes

One SE can be used for different situations. For example, an SE for "Knocking on a wooden door" can also be applied as "Cutting vegetables on a wooden cutting board." It is difficult to determine the most appropriate SEs using a simple search based on names or descriptions, or both. This suggests that SE searchers will not find the

correct SEs before listening to them.

Using text to represent sound

It is difficult to express the detailed nuances of sound effects with names alone (e.g., "wind 1") or descriptions (e.g., "wind blowing softly"). SE nuances such as "reverberation" and "strength" cannot be obtained from the text information. Therefore, searchers must listen to all searched SEs to determine the most appropriate example; as a result, a significant amount of time is expended listening to irrelevant SEs.

Conceptualizing SEs for a given scene

Vague ideas for the SE might be clarified through the search cycle; the relationships among the searched SEs may help the user conceptualize the desired SE. Subsequently, effective exportation environments should be provided to allow the user to repeat the search while reformulating the query.

This study aims to provide a framework for intuitive exploratory SE searches, to reduce the effort required to locate the most appropriate SE for a scene. To fulfill this objective, this paper proposes several SE search design guidelines to solve the three difficulties described above.

3 Design concepts

This paper emphasizes the idea that SEs can be expressed with onomatopoeia [1]. Onomatopoeia is a term that emotionally and delicately describes sound and conditions [2]. Moreover, three important elements of SEs, i.e., acoustic features, onomatopoeia, and listening impression caused by sound, are compatible. The existing research has reported that onomatopoeia is effective for capturing the acoustic impressions of SEs [3]. From these facts, using onomatopoeia provides the searcher with an intuitive and specific SE search method. The SE searcher would be able to grasp rough nuances of SEs without listening.

SEs have been expressed by only "context" and "acoustics." This system proposed in this study utilizes "onomatopoeia" when searching for SEs as an additional means of identifying their characteristics. As a result, the similarities among SEs can be expressed from three different aspects: "context," "acoustic," and "symbol of onomatopoeia." Table 1 shows detailed descriptions and the purpose of each aspect used to represent SEs. This study aims to realize efficient and intuitive exploratory SE searches, using each aspect to meet the requirements of the SE searcher.

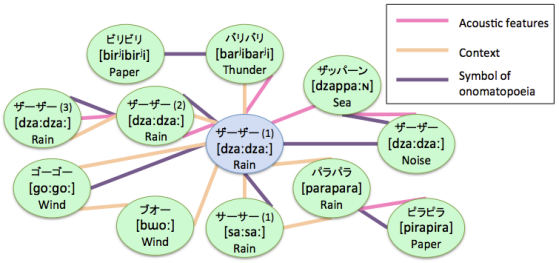

Figure 1. Schematic diagram of a relationship among SEs.

As shown in Figure 1, the three aspects can be used to form relationships among SEs: the blue node is related to other nodes by each of the three aspects. For example, "ザーザー (1), rain," which is the blue node, is similar to "サーサー (1), rain" based on its relation to "symbol of onomatopoeia," and is similar to "ザッパーン, sea" from its relation to "acoustic." The proposed system enables us to perform an exploratory search while visually surveying the SEs' similarities to each other. Therefore, the SE searcher sequentially performs the following three experiments while efficiently searching for SEs: 1) Determining which queries are effective for SE searches, 2) Gaining associative ideas from the similarities among SEs, and 3) Discovering new SEs. Moreover, a linked SE structure such as the example shown in Figure 1 may save many steps when regenerating queries; SE searchers only have to follow the link.

This paper hierarchically clusters SEs by using acoustic features. In addition, onomatopoeia sounds representing the SEs are obtained through subjective evaluation experiments. As the first step in developing the proposed SE search system, this paper discusses the relationships between acoustic features and onomatopoeia sounds.

4 Associations between SEs and onomatopoeia sounds

A previous study reported that onomatopoeia is effective for representing impressions of SEs [3]. In this previous study, onomatopoeia is a sym-

Table 1. Detailed descriptions and the purpose of each characteristic used to represent SEs.

Characteristic	Description	Purpose
Context	Title and descriptions of SEs. Sources and scenes of SEs are specified in the description.	SE searchers capture causality of source and sound, and roughly narrow down SEs.
Acoustic features	Features calculated from the acoustic waveform.	Searchers can capture detailed nuances. They are used to finely adjust the nuances of SEs.
Symbol of onomatopoeia	Onomatopoeia representing SEs. Text form features.	SE searchers capture characteristic sound of SE. They enable SE searchers to roughly narrow down SEs. The relationships among SEs can be spatially understood.

bol representing signal information that is difficult for humans to naturally understand. Using onomatopoeia, sound information with artificial temporal and acoustic characteristics becomes visual and spatial information. The similarities among SEs may be checked with text.

In order to associate onomatopoeia to SEs, we conducted an experiment, in which participants freely describe onomatopoeia corresponding to certain SEs. The experiments were conducted online, with 114 participants. Prior to the experiment, 100 SEs were obtained from a SE distribution site and the SEs were divided into five groups. To reduce the burden on the participants, each participant listened to and evaluated SEs in only one group. Each SE group contained 20 sounds, presented in random order. Each participant listened to the 20 SEs in arbitrary times, then asked to verbalize each SE as onomatopoeia. The Onomatopoeias collected through the experiments were subjected to pretreatment to control fluctuations. In addition, the modal onomatopoeia for an SE was assumed to be the onomatopoeia representing the SE.

5 Acoustic feature extraction and its clustering

In the field of music informatics, it has been reported that music can be clustered using acoustic similarity [4]. The relationships between SE clusters, based on their acoustic features and onomatopoeia, are discussed in Section 4. Subsequently, this paper focuses on the morphological features and sound symbolism of onomatopoeia.

MIRtoolbox [5], which is a music information processing tool, is used to extract acoustic features from SEs. In this paper, eight acoustic features (RMS energy, Low energy, Tempo, Zero cross, Roll off, Brightness, Inharmonicity, Mode) are extracted from each SE; f_i shows each feature in the table and i shows the index of the feature. The acoustic features were determined by referring to the previous study; acoustic features are effective in classifying music, and related sound tones [6]. It is impossible to extract tempo feature f_3 if an SE does not have periodicity; in these cases, f_3 is set to 0. The ranges of the acoustic feature values differ; thus, the ith feature value of an SE s is normalized using the following equation:

$$f'_i(s) = \frac{f_i(s) - f_i^{\min}}{f_i^{\max} - f_i^{\min}} \quad (1)$$

where $f_i^{\min}$ and $f_i^{\max}$ respectively show the minimum and maximum values for f_i in all SEs. Normalized ith feature value: $f'_i(s)$ is regarded as the ith feature of s.

Using the acoustic features, SEs are hierarchically clustered based on the Ward method [7]. Hierarchical clustering generates the classified results as a dendrogram. From the dendrogram, it is possible to visually confirm the relationships among SEs.

6 Analysis

The system proposed in this study provides three viewpoint types for SE searches: "context," "acoustic," and "symbol of onomatopoeia." As a result, searchers can progressively clarify vaguely defined SEs by traversing these viewpoints. To perform such a search, SE clusters should be optimized for each viewpoint. However, "acoustic" and "symbol of onomatopoeia" are expected to have a certain correlation because onomatopoeia expresses a sound with words. Therefore, in order to improve accessibility in SEs searches, clusters containing diverse onomatopoeia must be formed, in addition to those containing similar onomatopoeia.

To verify this point, the dendrogram from Section 5 was analyzed from two points of view.

Table 2. Onomatopoeia evaluation example

Onomatopoeia	Conversion	Categories
ザー [ʣaː]	A ー [1]	Macron
コッ [kop]	A ッ [2]	DC
シャキン [ɕakʲiɴ]	AB ン [3]	SN
プルルル [puɾuɾuɾuɯ]	ARRR	R
カチャカチャ [kaʨakaʨa]	ABAB	Repetitive
チャリン [ʨaɾʲiɴ]	AR ン	R, SN
チャリンチャリン [ʨaɾʲiɴʨaɾʲiɴ]	AR ン AR ン	R, SN, Repetitive

[1] "ー" is a Japanese macron.
[2] "ッ" is a Japanese double consonant.
[3] "ン" is a Japanese syllabic nasal.

Table 3. A presence or absence of more than 50% for items in each group

Group	Macron	DC	SN	R	Repetitive
Group 1	◯	–	–	–	–
Group 2	–	–	–	–	◯
Group 3	–	–	◯	–	◯
Group 4	–	–	◯	◯	–
Group 5	–	–	–	–	◯
Group 6	–	–	◯	–	–
Group 7	–	–	◯	–	–
Group 8	–	–	–	–	◯
Group 9	–	–	◯	–	–

To perform the analysis, the dendrogram was divided according to certain rules. One division rule was "a cluster cannot consist of only a single SE." As a result, 100 SEs were classified into nine groups; the respective features of each group are discussed below.

6.1 Morphological features of onomatopoeia

Morphological features of onomatopoeia were analyzed when determining the features of the group. The analysis considered whether onomatopoeia existed in any form for each group. Onomatopoeia were replaced with a common symbol to remove any spelling variants. Specifically, the first sound was replaced with "A", the second sound was replaced with "B". If it appeared alternately, it was represented as "repetitive." Liquid sound (the Sound of "R") was expressed as "R" because it is frequently used to represent fast amplitude fluctuations such as a beat sound. In addition, double consonant (DC), syllabic nasal (SN), and macron were used without modifications. Table 2 shows these conversion examples. Table 3 shows the items that occupy more than 50% of the group are characterized as converted. This table suggests that each group has a dominant morphological feature. In contrast, DC was less than 50% in each group.

Both Group 2 and Group 8 contained onomatopoeias of which morphological features were similar. Similarly, Group 6, Group 7,and Group 9 also contained onomatopoeias that morphological features were similar. For example, more than 50% of Group 6 was occupied by short onomatopoeia such as "A–DC" and "A–SN." Similarly, Group 7 was a group formed by short onomatopoeia. These two groups are located in close proximity to each other on the dendrogram. Therefore, it can be presumed that they were classified by morphological features.

Group 5 has a unique characteristics. It can be separated further into two subgroups according to the distance between the groups. In one subgroup, onomatopoeia with DCs accounted for 80%. In the other subgroup, onomatopoeia with macrons accounted for 50%. In addition, the ARAR types in Group 9 were in close proximity to each other.

Table 4. Sound symbolism of vowels

Type	Element	Sound symbolism
i	/i/	Straight line, High–pitched sound
ii	/a/	Flatter, Spread, Glamorous, Flashy
iii	/o/	Round, Small, Inconspicuous and modest things
iv	/u/	Things that are relevant to a small round hole, Soft and modest sound
v	/e/	Inadequacy or vulgarity of behavior

Table 5. Sound symbolism of consonant

Type	Element	Sound symbolism
I	/p, b/	Strike the object, Sudden explosive, Taut, Suddenly, Strength. /p/: Associated with dynamism and active behavior
II	/t, d/	Blow
III	/k, g/	Contact with hard surfaces such as metal
IV	/s, z/	1 Mora: Smoothness, Action that is not abrupt. 2 Mora: Light touch, Friction, Small movement, Liquid flowing, The spacious movement, Tranquility, Calm, Refreshing. Personality: Neatness, Smart, Coolness
V	/h/	Breath. 2 Mora: Uncertainty, Unreliability, Weakness, Delicate elegance
VI	/m/	Obscure state, Lack of calmness. 1Mora: Suppression, ambiguity
VII	/w/	1 Mora: sound that emitted by animal and human

6.2 Impressions when listening to sound

The human impression of a sound is analyzed according to the sound's symbolic onomatopoeia meaning. Sound symbolism is the direct linkage between sound and meaning [8]. Onomatopoeia sounds corresponding to the SEs were classified according to each sound metaphor, to be ana-

Table 6. Classification of each group

Group	Vowel	Consonant
Group 1 (5)	Type i (2), Type iii (3), Type iv (1)	Type I (4), Type II (1)
Group 2 (11)	Type i (2), Type ii (2), Type iii (5), Type iv (5)	Type I (4), Type II (5), Type III (4), Type IV (1)
Group 3 (9)	Type ii (2), Type iii (5), Type iv (3)	Type I (3), Type II (2), Type III (5)
Group 4 (4)	Type i (3), Type ii (2)	Type II (3), Type III (2)
Group 5 (15)	Type i (4), Type ii (11), Type iv (1)	Type II (2), Type III (2), Type IV (10)
Group 6 (10)	Type i (1), Type ii (3), Type iii (6), Type iv (2)	Type I (2), Type II (5), Type III (6)
Group 7 (9)	Type i (2), Type ii (5), Type iii (2)	Type I (2), Type II (6), Type III (4)
Group 8 (14)	Type i (5), Type ii (8), Type iii (1), Type iv (4), Type v (1)	Type I (4), Type II (6), Type III (6), Type IV (1)
Group 9 (23)	Type i (9), Type ii (12), Type iii (3), Type iv (3)	Type I (4), Type II (8), Type III (9), Type IV (7)

The ( ) after the group number contains the number of onomatopoeia classified into the group.

lyzed according to the sound's symbolism.

A summary of phonological form and sound symbolism is shown in Table 4 (vowel) and Table 5 (consonant). The summary refers to the work of Tamori et al. [9] and is supplemented with data from Hamano et al. [10]. In these tables, a type name was attached to each item to perform this analysis. Onomatopoeia that were classified into each type were counted, and the features of the group have been summarized. The summary is shown in Table 6.

First, vowels were classified into groups and their features were analyzed by using Table 6.

All groups were formed using a plurality of vowel types. Type v appeared only in Group 8, because no other onomatopoeia collected in Section 4 contained this vowel type. Group 2 was composed of 4 types — Type i, ii, iii, and iv. They were present at the same rate in the Group. As shown by Type ii in Group 5, there were cases in which certain vowels were far more prominent than others. However, most of the groups did not differ significantly in the types of vowels in the group. To some extent, this occurred because the heights of the classified SEs were varied. The vowels used to represent onomatopoeia change according to the frequency of the sound.

Next, consonants were classified into groups and their features were analyzed by using Table 6. All groups had multiple consonant types, similar to the case of vowels. Type V, VI, and VII did not exist in any group. This occurred because no onomatopoeia with these consonant types were collected in Section 4. In some types, percentage of consonants included in the type were almost the same (i.e. Type I and Type III in Group 2). On the other hand, consonant type percentages differed significantly in Group 5.

Finally, the symbolic sound meanings of each group were analyzed by combining the features of consonants and vowels. The analysis produced nine groups, as follows:

Group 1 was a sound a human can feel roundly and dynamically. Group 2 was a soft modest impact sound. Group 3 was a modest metallic sound. Group 4 was a high–pitched spacious sound. Group 5 was a light touch sound that spread. Group 6 was modest metallic sound. Group 7 was a flashy striking sound. Group 8 was a sound that hits a hard surface such as metal. Group 9 was a loud sound in contact with a hard surface such as metal.

Simple experiments were performed to validate whether SEs similar to these group were classified correctly. The experiment was conducted with three participants. They listened to the SEs, and SEs that belonged to the same cluster. As a result, 50% of the SEs in the group were confirmed to have a similar sound.

6.3 Summary

Cluster features were analyzed by combining the morphological features and symbolic sound perspectives of onomatopoeia. Each group was found to be composed of features that were different from the other groups. Group included not only similar onomatopoeia but also onomatopoeia that were not similar. The results were expected to draw a searcher's awareness in doing SE exploration.

The foregoing suggests that, among the features of the system proposed in this study, acoustic features are a key factor in identifying similar SEs. In the future, this functionality will be verified with experiments using human participants.

7 Related research

Wake et al. proposed an SE retrieval system using three query types [1]. The queries are onomatopoeia, the source of the sound, and subjec-

tive representation. The user enters at least one or more queries into the system from among the three types of queries. The input query calculates search scores according to the degree of similarity between the label assigned to the SE in the database. By calculating the search scores, the SEs and other similar SEs are presented to the searcher. Even if the queries do not match the label assigned to the SEs, the searcher can still obtain various SEs with some relation to the query.

Wake et al. have performed comparative experiments with this system and a system using a list. The system using the list classifies SEs by the source of the sound. The experimental results confirm that Wake's system provides the searcher with many more results than the list-based system.

Music searches are mentioned as other examples of sound searches. M. C. Schraefel et al. have proposed a system that can perform exploratory searches for online music by hierarchically presenting it [11]. The hierarchy is organized in the order of Period, Composer, Form (e.g. Concerto), Arrangement (e.g. Orchestra), and Piece. When a searcher selects a cue in the hierarchy, the system displays cues from the next hierarchy. In addition, these cues are associated with each of the related characteristics. For example, Agricola is associated with Agricola's representative music, and the Romantic period is associated with Romantic pieces. The searcher can repeat the search at any time using the path displayed in the upper left corner of the window.

M. C. Schraefel et al. have experimented using this system and a system that presents results in a single column. The results show that all participants using this system can easily discover new music that matches their preferences. In addition, the system reduces the number of playbacks required for discovering new music.

Unlike these studies, the current study focused on exploratory searches for SEs.

8 Conclusion

This paper proposed a system framework for efficient SE exploration. The characteristics of the framework is to utilize similarities among SEs with three types of features: context, acoustic features, and symbols of onomatopoeia. As the first step in this study, SEs were clustered based on their acoustic features. The relationships between the clustered SEs and their onomatopoeia characteristics were acquired through subjective evaluation experiments.

In the experimental results, SE groups were found to contain SEs that contained approximately similar characteristics and meaning, according to the group's classification features. When a participant listened to an SE that belong to the group, it was suggested to be classified according to the characteristics of that SE.

Verification of the proposed system primarily compared the onomatopoeia associated with the SEs. Thus, sources and sound quality of the SE were not considered when determining whether SEs were similar. In future studies, when a searcher is listening to an SE belonging to the group, its similarity to the other SEs in the group will be verified with a listening experiment.

Acknowledgement

This work was supported by JSPS KAKENHI Grand Number 15H02780 and 15K12151.

References

[1] Wake, S. and Asahi, T.: Sound retrieval with intuitive verbal expressions. *ICAD1998*, 1998.

[2] Ono, M.: *Giongo Gitaigo 4500 Nihongo Onomatope Jiten (Onomatopoeia Dictionary)*. Shogakukan, 2007. in Japanese.

[3] Iwamiya S.: *Science of sound color: evaluation and creation of timbre and sound quality* , vol, 1, *Sound science series*. Coronasha, 2010. in Japanese.

[4] Orio, N.: *Music Retrieval: A Tutorial and Review*. Now Pub., 2006.

[5] Toiviainen, P., Lartillot, O. and T Eerola, T.: A matlab toolbox for music information retrieval. *Data Analysis, Machine Learning and Applications*, pp. 261–268, 2008.

[6] Helmholtz, H.: *On the Sensations of Tone*. Dover Pub., 1954.

[7] Ward, J. H.: Hierarchical grouping to optimize an objective function. *J. of American Statistical Association*, 58(301), pp. 236–244, 1963.

[8] Hinton, L., Nichols, J. and Ohala, J. J.: *Sound Symbolism*. Cambridge Univercity Press, 2006.

[9] Tamori, I. and Schourup, L. C.: *Onomatope: Keitai-to imi (Onomatopoeia: Form and Meaning)*. Kuroshio Pub., 1999. in Japanese.

[10] Hamano, S. S.: *Sound-symbolic system of Japanese*. Cambridge University Press, 1986.

[11] Schraefel, M. C. *et al.*: Listen to the music: Audio preview cues for exploration of online music. *INTERACT'03*, pp. 192–195, 2003.