Vector-Based Sound Retrieval using Successive Relative Search

Kiyoaki Aikawa and Kanako Yajima

School of Media Science, Tokyo University of Technology
aik@media.teu.ac.jp

ABSTRACT

This paper proposes a new algorithm for retrieving sound based on successive relative search. In retrieving musical sound focusing on its sound features, emotional representations are more appropriate than conventional keywords of the genre or the composer. A vector-based sound retrieval system "Sound Advisor" was built up using emotional parameters. The problem is how to find a better candidate, if the first candidate is not satisfactory. The reason is that it is difficult to quantitatively image the emotional vector space of sounds. This paper proposes a method for successively retrieving sounds using relative search regarding the found candidate as the reference base. This method can contribute to sound-effect retrieval for producing movies, and emotional communication via sound.

KEYWORDS

Multimedia communication, Information retrieval, Search methods, Speech communication, Multimedia systems

INTRODUCTION

Human brains directly receive sensory information from the five senses without any translation. However, the human needs to translate the sensed information into some expression which can be understood by other human. The primary media is verbal representation in general. The problem is what expression should be used for communicating the sensed information from one person to another. Metaphor is sometimes effective for conveying such sensory information. Sounds, patterns and colors are also available.

A vector-based language generation method has been reported as the Weather Reporter system [1-3]. The Weather Reporter System generated verbal expressions evoking meteorological imagery from the measured weather data. The system selected meteorological expression depending on the similarity between an authorized weather expression and the measured weather vectors. The system could give more appropriate expression than just a literal translation. For example, let us take a day in May, when the temperature is 24 degrees centigrade, humidity is 50%, and the wind is a gentle breeze. People may refer to such weather as ‘bracing’, ‘refreshing’ or ‘invigorating’, rather than listing the meteorological measurements. The 62-dimensional vector was composed of seven categories such as ‘month’, ‘part of the month’, ‘part of the day’, ‘weather’, ‘temperature’, ‘humidity’ and ‘wind velocity’. The cosine similarity between the query vector and each weather-expression vector was computed and compared. The similarity between a sound and measured vectors indicated a confidence of the decided expression. Therefore, a confidence expression was added depending on the similarity value [3-4].

This paper proposes a new relative search method for retrieving sound based on the vector-based search. The sound search offers a problem of difficulty in image the distribution of the sounds in the multi-dimensional search space of emotional representations. This results in a problem how to set the parameters in searching more appropriate candidate when the first candidate is not satisfactory. One of the conventional methods selects the second closest candidate indicating the second large similarity to the given input emotional vector. Another conventional method repeats trial and error by changing parameters. However the emotional expression is difficult in quantitative representation. Therefore, it is usually difficult to obtain the ideal candidate by these methods. It is a good idea to start the next search under the condition that the first candidate is already found, because it is easy to specify a relative request based on the found sound. This paper proposes the mathematical derivation to obtain the next candidate by the relative search. A vector-based music retrieval method itself has been reported [5]. However, the relative search has not been reported, yet.

VECTOR-BASED SEARCH

Figure 1 schematically shows a vector-based sound search. At first, similarities are measured between the query request vector and all the reference sound vectors. Three parameters corresponds "Delight", "Sadness", etc.

The sound feature is decoded as the set of parameters in the N-dimensional vector space. The search request vector is also given in the N-dimensional feature space. This paper denotes the emotional vector for sounds as the sound vector, and the request vector as the search vector.

Let the search vector be x, and k-th sound vector be $a_{k.,}$ the similarity between the search vector and the k-th sound vector is given by the following cosine similarity measure.

$$s(x,a_k)=\frac{(x,a_k)}{|x||a_k|} =\frac{(x,a_k)}{\sqrt{(x,x)(a_k,a_k)}}, =\cos\theta_k \quad (1)$$

where, θ_k is the angle between the two vectors. (x, a_k) denotes the inner product between two vectors x and a_k given by

$$(x,a_k)=\sum_{j=1}^{N}x(j)a_k(j) \,. \quad (2)$$

The first candidate is the sound that shows the highest similarity.

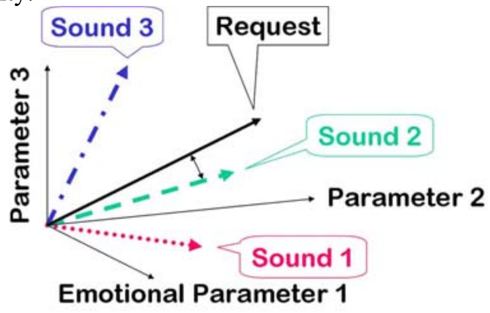


Fig. 1 Schematic illustration of vector-based sound search.

RELATIVE SEARCH

Figure 2 shows how the first sound candidate is selected using the proposing method. The angle between the request vector and the sound-2 vector is smallest. Thus the sound-2 is selected as the first choice. If the result is not satisfactory, the user needs to search another candidate.

This paper proposes a new sound retrieval method for obtaining better candidate. The new method searches the next choice regarding the first candidate vector as the new origin with giving the relative vector indicating the search direction.

Figure 2 schematically depicts how the sound-3 is obtained by the relative search. The method searches the vector toward the direction of the relative vector starting from the first candidate. The maximum similarity is obtained between the sound-3 and the sum of the first candidate and the relative vector multiplied by an appropriate multiplier. The maximum similarity is calculated for all the sound vectors. Given the first candidate, the relative vector, and a sound vector, the optimal multiplier is obtained so as to maximize the similarity between the sound vector and the sum of the first candidate vector and the relative vector multiplied by an appropriate multiplier. The multiplier should be a positive real number. The sound which achieves the maximum of the maximum similarity is selected as the next candidate.

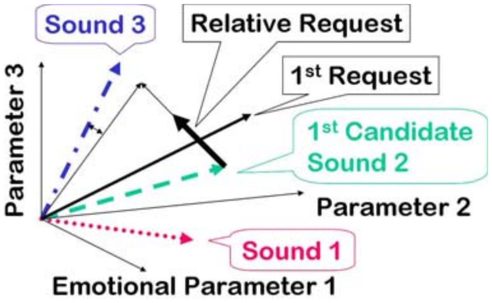


Fig. 2 Relative search of the next candidate by changing the multiplier of the relative vector. The origin of the relative vector is the first candidate.

FORMULATION

Suppose a sound vector a_k has been selected by the first search. Let the relative vector be r, multiplier be λ, the sum of the first candidate vector and the relative vector is given by,

$$a_k+\lambda r\,. \quad (3)$$

The problem is how to obtain the m-th sound vector a_m which maximize the similarity $s(a_k+\lambda r, a_m)$. Let the number of sound vector a_m be M, the maximum similarity is obtained by changing the multiplier λ for all the sound vectors $\{a_m, 1<m<M\}$. The a_m is calculated by the following equations.

$$\begin{aligned} a_m &= \arg\max_{m,\lambda} s(a_k+\lambda r, a_m) \\ &= \arg\max_{m} z_m \\ z_m &= \arg\max_{\lambda} s(a_k+\lambda r, a_m) \end{aligned} \quad (4)$$

$$\begin{aligned} &s(a_k+\lambda r, a_m) \\ &=\frac{(a_k+\lambda r, a_m)}{|a_k+\lambda r||a_m|} \\ &=\frac{(a_k+\lambda r, a_k)}{\sqrt{(a_k+\lambda r, a_k+\lambda r)(a_m,a_m)}} \end{aligned} \quad (5)$$

The similarity is maximized at a multiplier value at which the derivative is zero as the following equation.

$$\frac{\partial s(a_k + \lambda r, a_m)}{\partial \lambda} \to 0 \qquad (6)$$

The similarity $s(a_k+\lambda r,\ a_m)$ for the given m-th sound is given by

$$\begin{aligned} & s(a_k + \lambda r, a_m) \\ & = \frac{(a_k + \lambda r, a_m)}{|a_k + \lambda r||a_m|} \\ & = \frac{(a_k, a_m) + \lambda (r, a_m)}{\sqrt{(a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r)}|a_m|} \end{aligned} \qquad (7)$$

Therefore, the derivative is given by

$$\begin{aligned} & \frac{\partial s(a_k + \lambda r, a_m)}{\partial \lambda} \\ & = \frac{1}{|a_m|} \frac{(r, a_m)}{\sqrt{(a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r)}} \\ & - \frac{1}{2|a_m|} \frac{(a_k, a_m) + \lambda (r, a_m)}{((a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r))^{3/2}} \\ & \times (2(a_k, r) + 2\lambda (r, r)) \\ & = \frac{1}{|a_m|} \frac{(r, a_m)((a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r))}{((a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r))^{3/2}} \\ & - \frac{1}{|a_m|} \frac{((a_k, a_m) + \lambda (r, a_m))((a_k, r) + \lambda (r, r))}{((a_k, a_k) + 2\lambda (a_k, r) + \lambda^2 (r, r))^{3/2}} \end{aligned} \qquad (8)$$

The numerator of Eq. (8) is denoted as n and is given by

$$\begin{aligned} n & = (r, a_m)(a_k, a_k) + 2\lambda (r, a_m)(a_k, r) \\ & + \lambda^2 (r, a_m)(r, r) \\ & - (a_k, a_m)(a_k, r) - \lambda (r, a_m)(a_k, r) \\ & - \lambda (a_k, a_m)(r, r) - \lambda^2 (r, a_m)(r, r) \\ & = (r, a_m)(a_k, a_k) + \lambda (r, a_m)(a_k, r) \\ & - (a_k, a_m)(a_k, r) - \lambda (a_k, a_m)(r, r) \end{aligned} \qquad (9)$$

The multiplier λ which maximize the similarity is give by

$$\lambda = -\frac{(r, a_m)(a_k, a_k) - (a_k, a_m)(a_k, r)}{(r, a_m)(a_k, r) - (a_k, a_m)(r, r)} \qquad (10)$$

The maximum similarity value for a sound vector a_m is obtained by substituting the multiplier λ with the Eq. (10). Finally the m-th sound vector is selected that maximizes the maximum similarity.

SOUND ADVISOR SYSTEM

A Sound Advisor system has been developed using the vector-based relative search. The system was implemented using MATLAB.

FEATURE VECTOR

The system used eight emotional categories for representing sound impressions shown in Table 1. The vectors for 88 sounds were obtained based on subjective tests. The number of subjects was 14. Each answer was one of the five choices shown in Table 2.

Table 1 Emotional categories for sound search.

Category	No. of Levels
Delight	5
Sadness	5
Terror	5
Stability	5
Anger	5
Eerie	5
Brightness	5
Refreshing	5

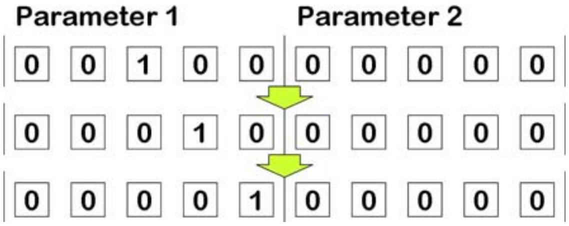
Table 2 Emotional levels.

Level	Meaning
1	very far
2	rather far
3	even
4	rather close
5	very close

This paper used the combination of the category and the level as a vector element to represent the distribution function of the user response. Therefore, total vector size was 40.

GUI

Figure 3 shows the GUI of the Sound Advisor system. The left column is the eight pop-up menus each corresponding to an emotional category. Each pop-up menu has five choices shown in Table 2. The first menu concerning "Delight" is set to level four and the seventh menu concerning "Brightness" is set to five. The base vector is obtained by pushing the button "SEARCH". The result is displayed in the text box "TITLE". The sound is played by pushing the button "PLAY". "LESS" and "MORE" check-boxes specify the relative vector toward negative and positive directions, respectively. The relative vector is set to (0, 0, 1, 0, 0) when "MORE" or "LESS" check-box are filled. The relative search is executed by pushing the button "RELATIVE". The result is displayed in the text box. Then the position of "1" is shifted toward right in case "MORE" checkbox is filled. It shifts the position of "1" toward left in case that "LESS" checkbox is filled.



Parameter 1					Parameter 2				
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0

Fig 6 Shift of the position of "1" in the relative vector after the relative search in case "MORE" check-box for the parameter-1 is filled.

EXPERIMENT

Sound retrieval tests were conducted for 10 subjects. Number of trials was three for each subject. When the subjects were requested to use categories "Delight", "Brightness", and "Refreshing" as the major search condition together with the emotional level of three and upper, the 92% of the retrieved sounds were satisfactory. However, when "Sadness", "Terror", and "Stability" were used for the major search condition, the proportion of the satisfaction was 50%. The important result was that the 60% of the unsatisfactory results were improved by the proposed relative search.

CONCLUSIONS

This paper proposed a new sound retrieval method using vector-based relative search. The method is characterized by the relative vector beginning at the vector of the firstly retrieved sound as the next origin. The mathematical algorithm was formulated on the basis of maximization of the cosine similarity measure. Experimental results demonstrated that the 60% of the unsatisfactory results were improved by the relative search.

ACKNOWLEDGMENT

The authors would like to thank Dr. Akemi Iida for providing the confidence expression generation program.

REFERENCES

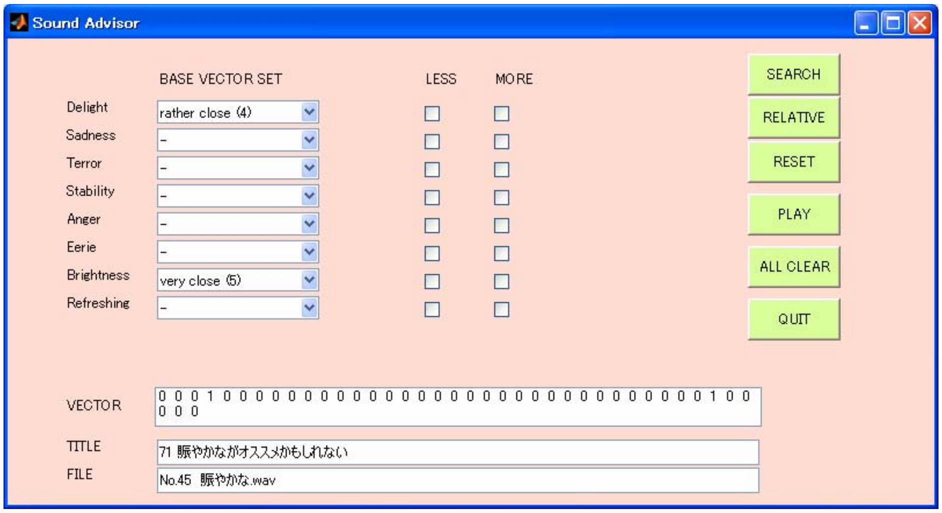
[1] Kuo, H. -K. J. and Lee, C.-H. “A portability study on natural language call steering”, In: Proceedings of the Eurospeech-01, Aalburg, Denmark, 2001.

[2] Zitouni, I., Kuo, H. -K. J. and Lee, C.-H. “Boosting and combination of classifiers for natural language call routing systems”, Speech Communication, 41, 647-661, 2003.

[3] Iida, A., Ueno, Y., Matsuura, R., Aikawa, K. “A Vector-based Method for Efficiently Representing Multivariate Environmental Information”, In proceedings of ICSLP 2004, pp.269-272, 2004.

[4] Aikawa, K. and Iida, A., “Vector-based language generation for associatively evoking environmental images”. J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, pp.3038, (2006-11).

[5] Ohta, K., Kumamoto, T., and Isahara, H., “Design of an impression-based music retrieval system”, J. Acoust. Soc. Am., Vol. 120, No. 5, Pt. 2, pp.3236, (2006-11).



	BASE VECTOR SET	LESS	MORE
Delight	rather close (4)	□	□
Sadness	-	□	□
Terror	-	□	□
Stability	-	□	□
Anger	-	□	□
Eerie	-	□	□
Brightness	very close (5)	□	□
Refreshing	-	□	□

VECTOR 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0

TITLE 71 賑やかながオススメかもしれない

FILE No.45 賑やかな.wav

Fig. 3 Relative sound search GUI