

Sound-Environment Monitoring Technique Based on Computational Auditory Scene Analysis

Mitsuru Kawamoto

Service Sensing, Assimilation, and Modeling Research Group, Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 1, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8560 Japan

Abstract— Monitoring techniques are a key technology for examining conditions in various scenarios (e.g., structural and weather conditions and disasters). The appropriate extraction of the features of these scenarios from observation data is important in understanding them. This study proposes a monitoring technique that allows sound environments to be expressed as a sound pattern. To this end, the concept of synesthesia is exploited. The keys, tones, and pitches of the monitored sound are expressed using the three elements of color, namely hue, saturation, and brightness, respectively. The hue, saturation, and brightness are assumed herein to be detected from the chromagram, sonogram, and sound spectrogram, respectively, based on a previous synesthesia experiment. The sound pattern can then be drawn using color, yielding a “painted sound map.” The usefulness of the proposed monitoring technique is verified using environmental sound data observed at a Galleria.

Keywords— Sound environment visualization; Environmental sounds; Monitoring; Painted sound patterns; Synesthesia

I. INTRODUCTION

The recent analysis of large data sets or the so-called “big data” has allowed a variety of information to be extracted. This information can help create certain services. Furthermore, monitoring techniques can be useful for determining the phenomena that initially generated the recorded data. Thus, monitoring techniques are regarded as those that allow the identification of monitored environment conditions through an analysis of the data observed within the area. For example, in the case of structural monitoring, which is known as building health monitoring, the deterioration and damage to buildings can be checked using findings obtained through sensor data (e.g., data acquired from acceleration sensors and cameras) analysis [1]. Sound environments are assumed in this study to be the target field of the monitoring problem, that is, sound environment monitoring is addressed.

Various methods for understanding sound environments have been proposed to date. However, almost all researchers focused on topics related to environmental sound recognition [2]. For example, a method of understanding environmental sounds by employing a matching pursuit algorithm was proposed [3]. To the best of the author’s knowledge, no studies focused on determining sound environments. Therefore, such a method is presented here.

This study proposes an unconventional method that allows the analysis of sound environments using color. Recently, the author proposed a monitoring method of sound environments using color [14]. However, differently from [14], in this study, an information formulated by a different method from the one used in [14] is utilized to analyze sound environments with color, and hence the visualization result obtained in this study has a different pattern from the one shown in [14]. The efficacy of the proposed monitoring method is evaluated using environmental sound data observed at a Galleria.

II. PROPOSED METHOD

A. Overview of the proposed method

For the application of the proposed method, environmental sounds are first collected using a microphone array (Fig. 1). Various sound environment conditions can be estimated using these sounds. These scenarios are then expressed using colors, where the color rules are based on the concept of synesthesia [4]. In other words, sound positions can be estimated, and a color based on three features extracted from the observed environmental sounds can be painted at the estimated position. Hence, painted sound patterns referred to as “painted sound maps” are obtained, from which sound environment scenarios can be recognized.



Fig. 1 Microphone array utilized to collect environmental sounds at a Galleria.

Synesthesia is a phenomenon in which one kind of sensory stimulation is expressed as another sensation [5]. In the case of synesthesia relating sound and color, Nagata et al. reported an experimental result in which keys, tones, and pitches were related to hue, saturation, and brightness, respectively [6]. The present study utilizes information on the keys, tones, and pitches of environmental sounds based on this result to draw painted sound maps.

The keys, tones, and pitches are assumed to be detected by chromagram, sonogram, and sound spectrogram, respectively. Hence, the hue score is calculated using the key histogram yielded by the chromagram. Similarly, the saturation and brightness scores are calculated using the frequency-element histograms produced by sonogram and sound spectrogram, respectively, where a clustering method is then applied to the environmental-sound spectrogram, and similar frequency components can be categorized so that the frequency component dispersion of the environmental sounds is clarified. This dispersion information is then used to calculate the histogram with respect to the spectrogram frequency elements.

The proposed method of the sound environment analysis is presented in detail in the section that follows. The sound data $y(t)$ can be obtained using the microphone array shown in Fig. 1, where a short-term Fourier transform $y(\omega, t)$ is applied to $y(t)$. The amplitude of $y(t)$ has $\max_{\omega} |y(\omega, t)|^2 > a$, where a is a constant value, and the maximum period of $y(t)$ is 2 s. The sound position estimation is conducted using multiple signal classification (MUSIC) [7] utilizing the microphone array outputs.

B. Key information extraction from the chromagram

The environmental sound chromagram is calculated using the MATLAB chroma toolbox [8]. First, the environmental-sound pitch features can be computed using the `audio_to_pitch_via_FB` function. Figs. 2 and 3 show an environmental sound and its pitch features, respectively, obtained using `audio_to_pitch_via_FB`.

A chromagram can be calculated (Fig. 4) based on pitch features, such as those shown in Fig. 3.

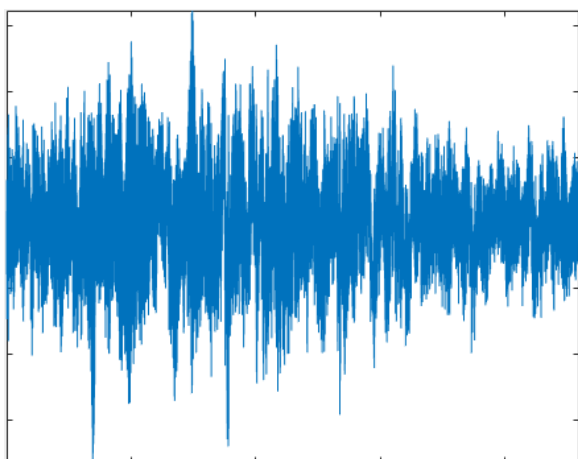


Fig. 2 Environmental sound.

Subsequently, a histogram showing the key information indicated in the chromagram is calculated. Fig. 5 shows the histogram calculated based on the chromagram shown in Fig. 4.

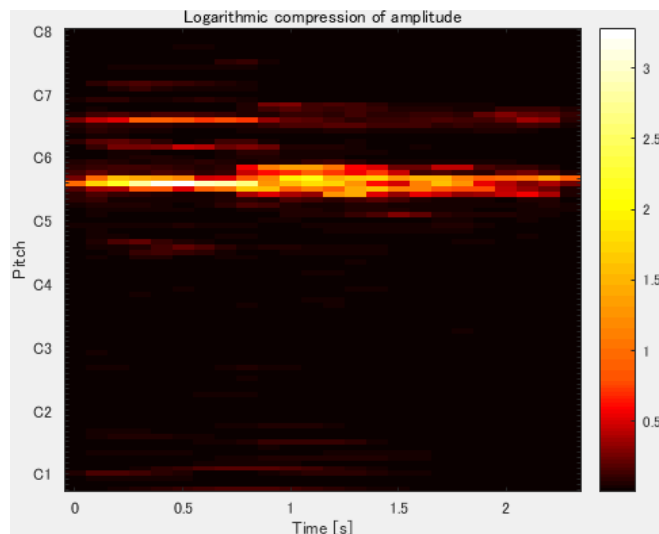


Fig. 3 Pitch features of the environmental sound shown in Fig. 2.

Note that differently from [14], the histogram herein is formulated using such a calculation that the dispersion of key information can be expressed, and hence the painted sound map obtained by the histogram herein would have a different pattern from the ones presented in [14].

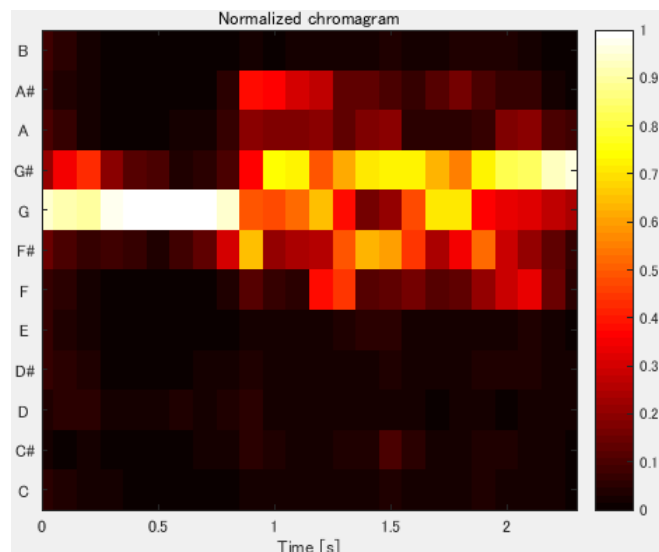


Fig. 4 Chromagram of the environmental sound shown in Fig. 2.

The histogram data can be transformed into an 8-bit binary code as follows depending on the histogram variability: 1) the histogram mean is calculated (dashed line in Fig. 5); 2) values greater or less than the mean are replaced with “1” or “0,”

respectively; and 3) an 8-bit binary code is obtained (the code corresponding to Fig. 5 is “00011000”). Hence, the hue score is determined by converting the binary code to decimal values. A low score apparently indicates that the dominant key of the environmental sound is low, whereas a higher score basically indicates that the dominant key of the environmental sound is high, and an environmental sound comprises various keys.

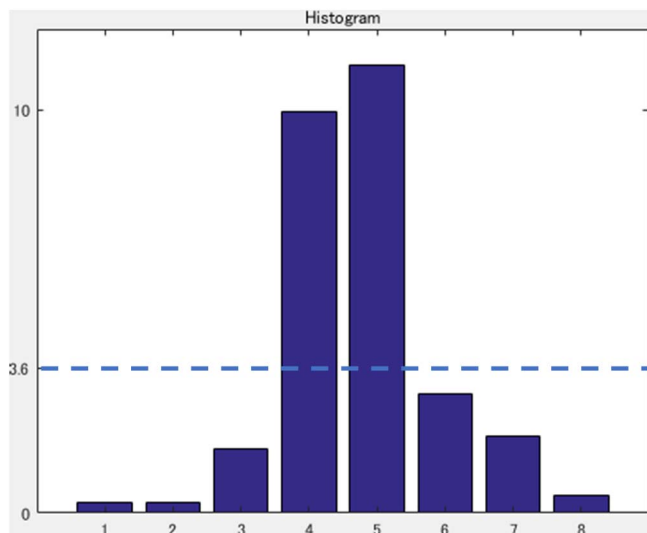


Fig. 5 Key histogram obtained from the chromagram in Fig. 4 with a mean.

C. Tonal information extraction from the sonogram

The sonogram can be calculated using the MATLAB Music Analysis (MA) toolbox [9], where the loudness sensation per frequency band is estimated using auditory models and the `ma_sone` function of the MA toolbox. Fig. 6 shows the sonogram of the environmental sound shown in Fig. 2.

The frequency band histogram of the sonogram is computed [9]. The saturation score is then determined using the same approach as that used for the hue score. Therefore, apparently, a higher score basically indicates an environmental sound with a large number of variations in the loudness sensation per frequency band.

D. Pitch information extraction from the spectrogram

A spectrogram can also be calculated. Fig. 7 shows the spectrogram of the environmental sound shown in Fig. 2. An edge-extraction image-processing technique is applied to the spectrogram. The number of pixels in its frequency characteristic areas and their centroid frequencies are then computed. The frequency characteristic areas of the spectrogram detected by the edge extraction technique are categorized using an improved affinity propagation (IAP) method [14]. The readers are referred to [10] for the details of the affinity propagation.

Each exemplar centroid frequency obtained by the IAP is classified into a low-, medium-, or high-frequency group. The frequency group histogram can then be acquired. The brightness score is obtained from the histogram in a similar manner to that

of the hue score. However, the threshold determining “1” or “0” values is set to zero when the 8-bit binary code is obtained.

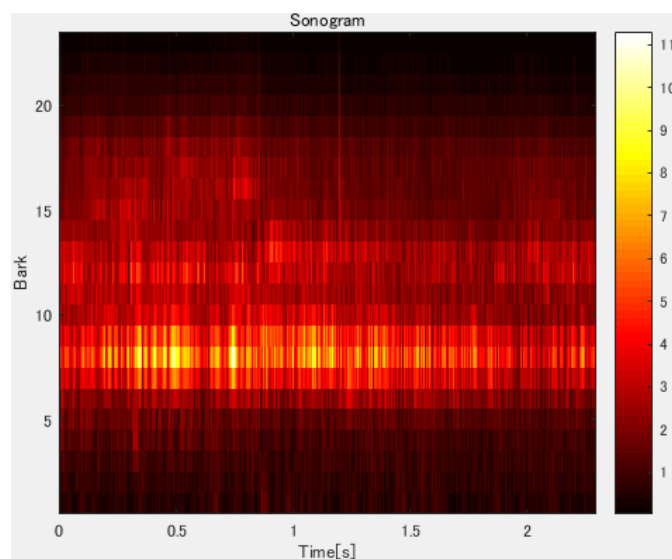


Fig. 6 Sonogram of the environmental sound shown in Fig. 2.

Therefore, a low score indicates that the dominant frequency of the environmental sound is low, whereas a higher score indicates that the environmental sound comprises various frequencies.

E. Painted sound map from three scores

The hue, saturation, and brightness scores are used to draw the painted sound map, where the hue–saturation–brightness color model obtained using these three scores is converted to a red–green–blue color model.

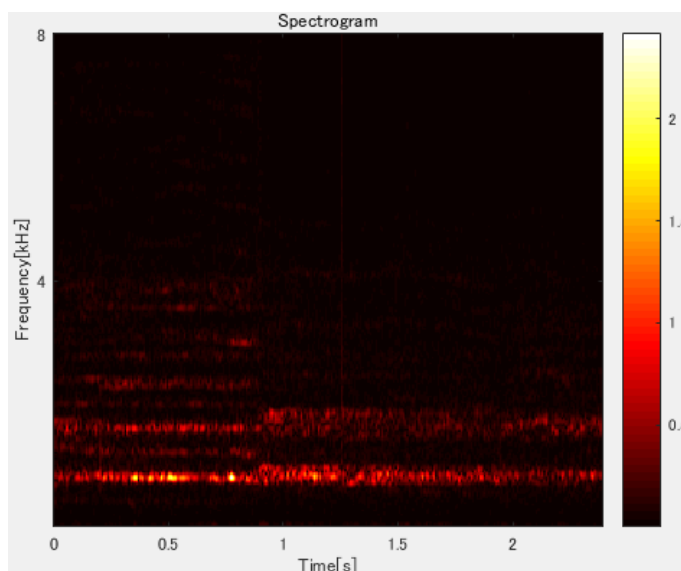


Fig. 7 Spectrogram of the environmental sound shown in Fig. 2.

III. EXPERIMENTAL RESULTS

The efficacy of the painted sound map method is demonstrated in this section using the environmental sounds observed in the sound environment shown in Fig. 1. The painted sound map in each demonstration has been drawn using the environmental sounds generated in one day. The position of the microphone array is denoted by a red circle in each figure that follows.

A. Painted sound map of the sound environment on a typical day



Fig. 8 Painted sound map of the sound environment in Fig. 1.

The sound environment shown in Fig. 1 is a shopping center galleria. Fig. 8 shows its painted sound map. A train station is located near the galleria, outside the left side of Fig. 8. Therefore, train sounds are intermittently generated. Furthermore, rattling sounds from chairs and desks are generated during the galleria preparation time, along with the voices of children and students visiting the place.

Fig. 9 presents the painted sound map for a different day. Notably, the generated sound patterns are similar to those in Fig. 8. Fig. 9 shows a painted sound map of the same year as that in Fig. 8.



Fig. 9 Painted sound map showing similar sound patterns to Fig. 8.



Fig. 10 Painted sound map at the galleria after one year and half a year from the year denoted in Figs. 8 and 9.

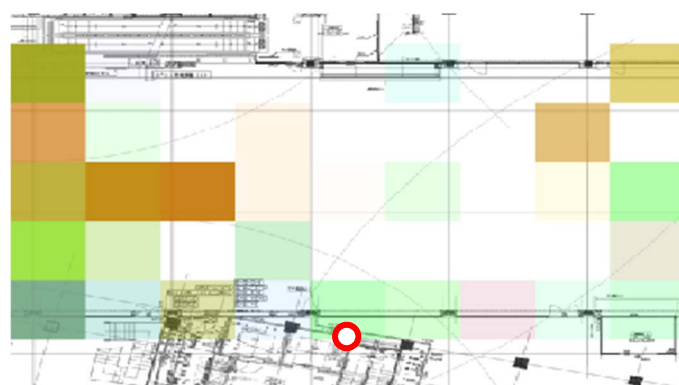


Fig. 11 Another painted sound map showing the sound patterns in the same year as Fig. 10.

These four maps show that the painted sound maps can be utilized to determine the similarities and the differences in the sound patterns of sound environments.

B. Painted sound map on a windy day

Fig. 12 shows a painted sound map obtained on a windy day. A comparison of Figs. 8–11 apparently shows that the painted

sound map varies with the state of the sound environment. Hence, painted sound maps can be utilized to detect the variations in a sound environment.

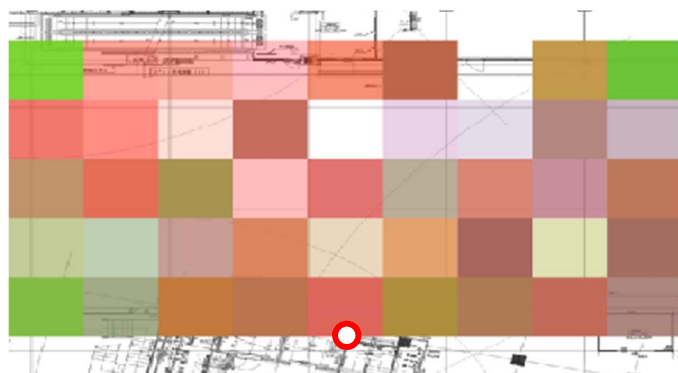


Fig. 12 Painted sound map obtained on a windy day.

C. Mini-concert event

Fig. 13 shows the painted sound map obtained for the same area during a mini-concert event. Blue tones are emphasized at the event location, which is on the left-hand side of the map.

All these results indicate that the proposed painted sound map drawn using the three scores discussed earlier is effective for the sound environment analysis. This approach is particularly useful for visually detecting and determining the sound environment conditions and their variations.



Fig. 13 Painted sound map obtained on an event day.

IV. CONCLUSIONS

This study proposes a method of monitoring sound environments based on a computational auditory scene analysis. The proposed visualization technique allows sound environment conditions to be determined and represented using colors. The proposed monitoring technique can be applied for monitoring superannuated building structural conditions as a future research work.

ACKNOWLEDGMENTS

The author thanks Dr. Sashima and Dr. Kurumatani for their helpful discussions. This work was partly supported by a JSPS KAKENHI Grant (grant number: 16H02911).

REFERENCES

- [1] T. Hamamoto, "Structural Health Monitoring of Buildings" (in Japanese), *Trans. of Foundation Engineering & Equipment*, Vol. 43, No. 11, pp. 17–20, 2015.
- [2] J.S. Chachada and C.-C. Jay Kuo, "Environmental Sound Recognition: A Survey," *SIP* (2014), Vol. 3, e14, pp. 1–15, 2014.
- [3] S. Chu, S. Narayanan, and C.-C. Jay Kuo, "Environmental Sound Recognition with Time-Frequency Audio Features," *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 17, No. 6, pp. 1142–1158, 2009.
- [4] The Color Science Association of Japan, *Handbook of Color Science* (in Japanese), 3rd ed., University of Tokyo Press, Japan, 2011.
- [5] R.E. Cytowic, *The Man Who Tasted Shapes*, MIT Press, USA, 2003.
- [6] N. Nagata, D. Iwai, M. Tsuda, S.H. Wake, and S. Inokuchi, "Non-verbal Mapping between Sound and Color-mapping Derived from Colored Hearing Possessors and Its Applications," *IEICE A*, Vol. J86-A, No. 11, pp. 1219–1230, 2003.
- [7] R.O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. on Antennas and Propagation*, Vol. 34, No. 3, pp. 276–280, 1986.
- [8] M. Muller and S. Ewert, "Chroma Toolbox: Matlab Implementations for Extracting Variants of Chroma-based Audio Features," *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pp. 215–220, 2011.
- [9] <http://www.pampalk.at/ma/documentation.html>
- [10] B.J. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *Science*, Vol. 315, No. 5814, pp. 972–976, 2007.
- [11] R. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, "Adaptive Affinity Propagation Clustering," *Acta Automatica Sinica*, Vol. 33, No. 12, pp. 1242–1246, 2007.
- [12] <https://jp.mathworks.com/matlabcentral/fileexchange/18244-adaptive-affinity-propagation-clustering>
- [13] T. Calinski and J. Harabaz, "A Dendrite Method for Cluster Analysis," *Commun. Stat.*, Vol. 3, No. 1, pp. 1–27, 1974.
- [14] M. Kawamoto, "Sound Environment Monitoring Method Based on Computational Auditory Scene Analysis," *Journal of Signal and Information Processing*, 8, pp. 65–77, 2017.