

## Building Health Monitoring using Visualization of Sound Features Based on Sound Localization<sup>†</sup>

Mitsuru Kawamoto<sup>\*1</sup>, Takuji Hamamoto<sup>2</sup>

<sup>1</sup>Human Augmentation Research Center, National Institute of Advanced industrial Science and Technology (AIST), Kashiwa, 277-0882, Japan

<sup>2</sup>Tokyo City University, Tokyo, 158-8557, Japan

### ARTICLE INFO

Article history:

Received: 21 July, 2021

Accepted: 11 October, 2021

Online: 06 November, 2021

Keywords:

Understanding sound environment

Sound visualization technique

Sound features

Sound map

Monitoring the collapse of buildings

### ABSTRACT

*This paper describes what can be accomplished by understanding sound environments. Understanding sound environments is achieved by extracting the features of the sound and visualizing the features. The visualization is realized by converting the three features, namely, loudness, continuity, and pitch, into RGB values and expressing the sound with color, where the color is painted in the estimated direction of the sound. The three features can distinguish falling objects within a building and roughly estimate the direction of the generated sounds. The effectiveness of the proposed sound visualization was confirmed using the sounds of cans and stones falling in a building; hence, it is shown that the proposed visualization method will be useful for monitoring the collapse of buildings by sound.*

## 1. Introduction

Techniques for understanding environments are used to obtain sensor-related information in environmental measurements. For example, in terms of monitoring buildings, the analysis of camera and accelerometer sensor data is used to understand the current state of the building and the differences from its past state (e.g., [1]). A camera can capture images of an entire building, and an accelerometer can detect building vibrations. However, the camera only detects events within its field of view and the accelerometer only detects vibrations in its surroundings. Further, these sensors can only be installed in the monitoring area, thereby providing limited coverage, since it is difficult and risky to install sensors under collapse risk. Alternatively, sounds can reach sensors such as microphones for detection. Moreover, if a microphone array is used as the sound sensor, the sound source direction can be estimated. Therefore, sound sensors can allow more flexible monitoring of buildings than cameras and accelerometers.

Considering the benefits of sound sensing, we propose a method for analyzing environmental sounds aiming to evaluate the

difference from the past state of buildings, which is called building health monitoring, by extracting their features and estimating the direction of sound sources. As precursory sounds often occur before building collapse, such sounds may be detected and characterized by extracting sound features.

Aiming to perform building health monitoring, a visualization method for sound features based on sound localization to facilitate analysis has been introduced. In most cases, sound monitoring is simply performed by recognizing measured environmental sounds for applications, such as elderly people (e.g., [2], and references therein). On the other hand, the proposed method can provide awareness of changes in buildings by providing a visual representation of environmental sounds. The sound features of loudness, continuity, and pitch are considered. These features are quantified using spectrograms and chromagrams. Then, a visual representation of the sound features along with their estimated source position are obtained. The proposed visualization technique was evaluated using experimental sound signal data generated at a building in Gunkanjima (Hashima), Japan. Our method is a novel paradigm for sound monitoring and a novel contribution to building health monitoring.

<sup>\*</sup>Corresponding Author: Mitsuru Kawamoto, Kashiwa II Campus, University of Tokyo, 6-2-3 Kashiwanoha, Kashiwa, Chiba 277-0882, +81 3 3599 8543 & [m.kawamoto@aist.go.jp](mailto:m.kawamoto@aist.go.jp).

<sup>†</sup> A preliminary version of this paper was presented at DCOSS2020.

## 2. Materials and Methods

### 2.1. Data Acquisition

We evaluate the proposed visualization technique which can be used for building health monitoring, using measurements obtained from a sound sensor (microphone array). Figure 1 shows Gunkanjima (Hashima), a World Heritage site in Nagasaki, Japan. The microphone array is installed as shown in Figure 2 on the second floor of building No. 30 (enclosed in black circle of Figure 1(a)) of Gunkanjima. The method is applied to verify the feasibility of determining the building status by analyzing environmental sounds. Building No. 30 is the oldest reinforced concrete building in Japan and is at risk of collapsing. The microphone array is comprised of 16 microphones (each arrow in the righthand side of Figure 2 shows a cluster of four microphones).



(a)



(b)

Figure 1: (a) Bird's eye view and (b) cityscape of Gunkanjima, Japan

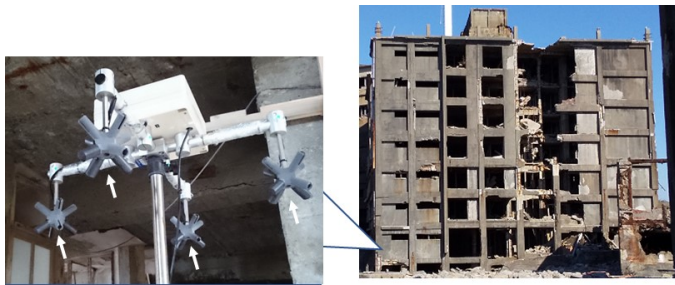


Figure 2: Microphone array installed in Building No. 30 of Gunkanjima (Hashima Island), Japan.

### 2.2. Sound Model of an Array with $M$ Microphones

The sounds measured with the microphone array (Figure 2) are analyzed in the frequency domain. Then, taking the short-time Fourier transform of each microphone input at time  $t$ , the following model is obtained [3]:

$$\mathbf{y}(t, \omega) = \mathbf{A}(t, \omega)\mathbf{s}(t, \omega) + \mathbf{n}(t, \omega), \quad (1)$$

where  $\mathbf{y}(t, \omega) = [Y_1(t, \omega), \dots, Y_M(t, \omega)]^T$  is the input vector of the microphone array, with  $m$ -th element  $Y_m(t, \omega)$  of the vector, at time  $t$  and frequency  $\omega$ ;  $M$  ( $= 16$ ) denotes the number of microphones in the array; and the superscript  $T$  denotes the transpose. Additionally,  $\mathbf{A}(t, \omega)$  is a matrix of transfer function vectors defined as

$$\mathbf{A}(\omega) = [\mathbf{a}_1(\omega), \dots, \mathbf{a}_L(\omega)], \quad (2)$$

where  $\mathbf{s}(t, \omega) = [S_1(t, \omega), \dots, S_L(t, \omega)]^T$  is the source spectrum vector of the  $L$  sound sources in the measurement environment and  $\mathbf{n}(t, \omega) = [N_1(t, \omega), \dots, N_M(t, \omega)]^T$ , a background noise spectrum vector, carries the assumption of following a zero-mean Gaussian distribution. We assume  $\mathbf{a}_l(\omega) = [A_{l1}(\omega), \dots, A_{lM}(\omega)]^T$  to be a transfer function that can be premeasured using time-stretched pulses [4]. We design time-stretched pulses to measure the impulse response, and the energy of the impulse signal is dispersed over time using a filter [4]. This filter is used to advance (or delay) the phase in proportion to the square of the frequency. Therefore,  $\mathbf{a}_l(\omega)$  can be obtained by applying the inverse of the filter to the time-stretched pulse response.

### 2.3. Sound Source Features

In building health monitoring, it is necessary to detect the location, magnitude, and duration of an eventual collapse. Hence, the loudness, continuity, and pitch features are extracted from the sounds observed by the microphone array. To visualize the sound features, a color map is drawn at the estimated sound locations to analyze the types of generated sounds. The proposed method for the building health monitoring displays information in the color map, which we call a sound map.

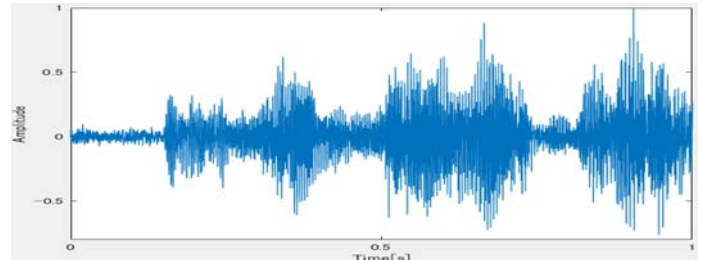


Figure 3: Sample environmental sound.

The observed environmental sounds are divided into 1-second segments, as illustrated in Figure 3. The short-time Fourier transform is applied to each segment to determine the loudness, continuity, and pitch, as detailed below.

#### 1) Loudness Feature

Loudness is important to characterize environmental sounds during building collapse. This information is represented by a sonogram, which we obtain via the MATLAB Music Analysis toolbox [5]. The estimation of the loudness per frequency band is performed using auditory models and the function `ma_sone` in the MATLAB toolbox, where the specific loudness sensation (in sones) per critical band (in Bark scale) is calculated in six steps. 1) The fast Fourier transform is used to calculate the power spectrum of the audio signal. 2) According to the Bark scale [6], the frequencies are bundled into 20 critical bands. 3) The spectral masking effects are calculated as in [7]. 4) The loudness is

calculated in decibels relative to the threshold of hearing (decibels with respect to sound pressure level—dB-SPL). 5) From the dB-SPL values, equal loudness levels in unit phones are calculated. 6) The loudness is calculated in sones based on [8]. (Regarding detail of the six steps, see [9]). Figure 4 shows the sonogram in the Bark scale of the sound depicted in Figure 3.

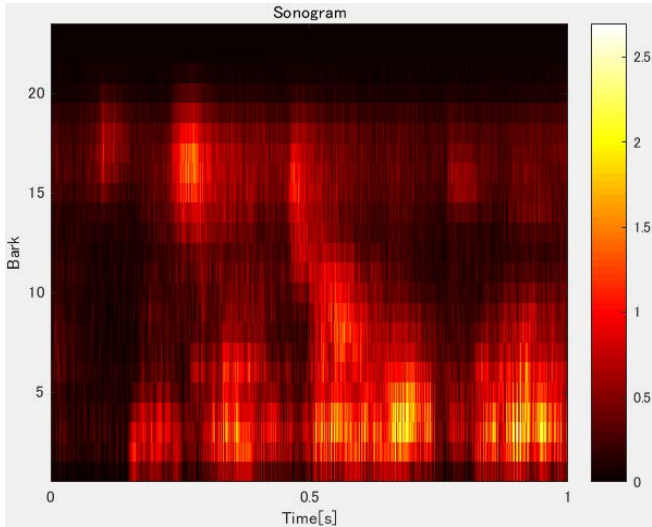


Figure 4: Sonogram in Bark scale of the sound in Figure 3.

The frequency histogram of the sonogram is also computed [5], and the histogram is resampled to 8 bits using the magnitude relation for the median of the histogram. The score expressing the information can then be obtained by converting each 8-bit value into the corresponding decimal value. An environmental sound with several variations in loudness sensation per frequency band is indicated by a higher score.

2) Continuity Feature

We use pitch variations of environmental sounds with respect to time  $t$  to measure the continuity of sounds. Therefore, continuity is expressed as a score obtained from a feature representing the pitch variation of the environmental sound.

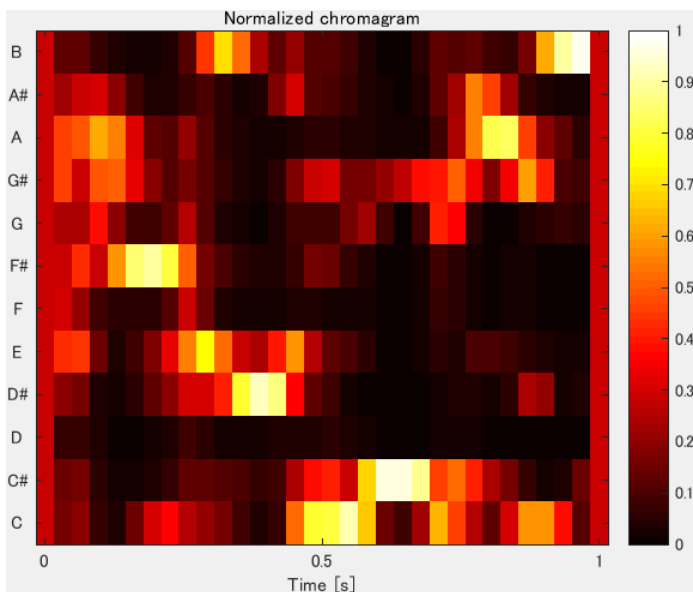


Figure 5: Chromagram of the sound in Figure 3

The corresponding chromagram, like the one shown in Figure 5, can be calculated using pitch features as those illustrated in Figure 3 by applying the method in [10]. Namely, we use the MATLAB Chroma toolbox to calculate the chromagram [11]. Subsequently, a differential chromagram is determined from the initial chromagram (Figure 5) using command diff in MATLAB, and the discrete cosine transform is applied to the differential chromagram (Figure 6). Based on this result, a time-domain histogram is then calculated. Continuity information is thus represented by a score obtained from the time-domain histogram in a method analogous to that used to obtain the loudness.

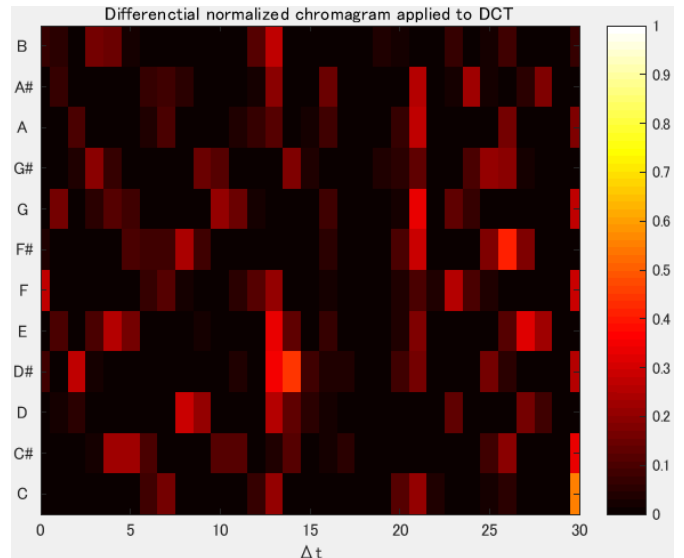


Figure 6: Differential chromagram after applying discrete cosine transform for the sound in Figure 3.

The histogram represents the variation of keys over time  $t$ . Therefore, a low score suggests that key variations of the environmental sound do not occur frequently, that is, the sound has low continuity. Conversely, a high score indicates frequent key variations of the environmental sound, indicating high continuity. As environmental sounds contain various keys, we use this score to represent continuity.

3) Pitch Feature

Pitch information is represented by the corresponding score, which is used to determine the type of environmental sound.

The spectrogram of the sound depicted in Figure 3 is shown in Figure 7. Edge extraction is applied to the spectrogram allowing the calculation of the number of pixels in its frequency feature areas and centroid frequencies. The improved affinity propagation method [12] is used to categorize the detected frequency characteristic areas of the spectrogram. More details on affinity propagation can be found in [13] and [14].

Each centroid frequency obtained by improved affinity propagation is classified into low-, medium-, or high-frequency groups. Then, a frequency group histogram can be established. The pitch is obtained from the histogram as a score, for which the calculation details can be found in [15]. A low score indicates a low dominant frequency of the environmental sound, whereas a high score indicates the presence of various frequency components.

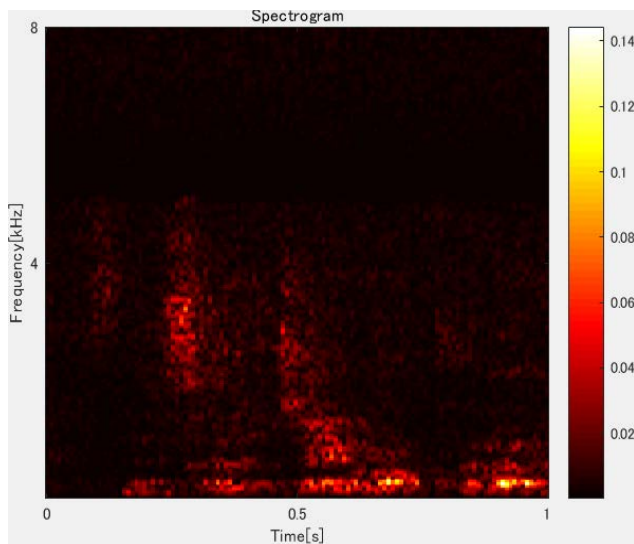


Figure 7: Spectrogram of the sound in Figure 3.

### 2.4. Sound Localization

Various methods for estimating sound source directions has been proposed (e.g., [16], [17], [18]). In this paper, using  $A(\omega)$  obtained from (2), the MUSIC [19] is used to estimate the locations of sound sources. Details about the estimation of virtual 3D sound source positions can be found in [10].

### 2.5. Sound Map from Features

The loudness, continuity, and pitch scores are used to construct a sound map by representing the three scores in a red–green–blue color model. The obtained colors are then overlaid on the estimated sound source positions. Therefore, unlike conventional sound visualization methods such as power spectrum, the proposed sound map reflects not only the pitch but also the loudness and continuity of the sound in a color model, thus establishing a novel representation.

## 3. Experimental Results

### 3.1. Experimental Setup

The effectiveness of the proposed method is evaluated by using the sounds of falling cans and stones measured by the microphone array installed as shown in Figure 2.



Figure 8: Experimental scenario using falling cans and stones that hit the second floor of the building as sound sources

The three hitting points for evaluation are marked with an X in Figure 8 and are located on the 2nd floor of the building. The cans

are dropped from the 4th and 6th floors of the building to investigate the difference in the sound features caused by the drop height. Figure 8 also shows the placement of the microphone array and the person in charge of dropping the cans and stones.

### 3.2. Sound Maps

Figure 9 shows the sound maps obtained from the sound sources generated by cans (Figure 9(a)) and stones (Figure 9(b)) hitting on the second floor. The color differences mainly correspond to loudness and continuity, as listed in Table 1, where each feature score is the average across nine trials. There is also a difference in pitch. Hence, the sounds of hitting stones have more frequency components than those of hitting cans. The three extracted features allow to distinguish differences in objects that cause sudden sounds, and Figure 9 also demonstrates the correct localization of the sound sources.



(a)



(b)

Figure 9: Sound maps of sound sources generated by (a) cans and (b) stones hitting the 2nd floor of the building. The red circle indicates the position of the microphone array.

Table 1: The averages of loudness, continuity, and pitch scores for the sounds generated by falling cans and stones hitting on second floor of the building.

	Loudness score	Continuity score	Pitch score
Can	158.2	107.3	19.2
Stone	115.3	179.8.8	32.2



(a)



(b)

Figure 10: Sound maps of the sound sources of dropping cans at the (a) 4th and (b) 6th floors of the building.

Therefore, if there are frequency differences in sounds generated before and after a building collapses, the three features, especially the pitch, can be used to detect them.

Figs. 10 shows the sound maps of the dropping cans from the fourth (Figure 10(a)) and sixth (Figure 10(b)) floors of the building, respectively, averaged across three trials. The X marks in Figure 11 show the corresponding dropping points for the maps in Figure 10. As expected, the colors on the estimated directions of the sound sources for both floors are similar.



(a)



(b)

Figure 11: Dropping points of cans (X marks) from the (a) 4th and (b) 6th floors of the building.

Table 2: Averages of loudness, continuity, and pitch scores for the generated sounds of falling cans at the 4th and 6th floors of the building.

Floor	Loudness score	Continuity score	Pitch score
2	240.0	113.7	25.0
4	240.0	84.3	16.3
6	245.3	28.3	26.0

The colors of the sound maps in Figs. 10(a) and (b) reflect the difference in floors and are mainly related to variations in the continuity score, as listed in Table 2.

The experimental results indicate the feasibility of using the proposed method for visualizing sound features in a building health monitoring system that can roughly determine the locations and types of environmental sounds in buildings.

#### 4. Discussion and Conclusions

Figure 12 shows snapshots of building 30 before (Figure 12(a)) and after (Figure 12(b)) a floor collapse (red area). These snapshots illustrate the difficulty in determining the collapse through the use of images from cameras. Thus, the use of sounds may allow effective monitoring of buildings such as those in Gunkanjima, which is an uninhabited island.



(a)



(b)

Figure 12: Building No. 30 (a) before and (b) after the collapse of the floor (red area).

We propose a method of visualizing sounds which can be used for implementing building health monitoring by calculating the direction and features of environmental sounds. The proposed visualization technique of sound features considering sound localization relies on sound maps that reflect the loudness, continuity, and pitch of multiple sounds.

Experiments considering falling cans and stones were considered to simulate sounds of a collapsing building. The experimental results suggest that, using the proposed method, collapsing floors and walls attributable to the damage and deterioration of buildings can be localized. Moreover, the mechanism of building collapse may be analyzed and clarified by using the proposed method.

The proposed method for building health monitoring using the sound measurement will be used to continuously monitor building No. 30 of Gunkanjima, Japan, and the building monitoring data will be collected. In addition, we will further improve the accuracy of sound localization.

### Conflict of Interest

The authors declare no conflict of interest.

[www.astesj.com](http://www.astesj.com)

### Acknowledgment

This work was partly funded by Japan Society for the Promotion of Science (grant number 16H02911 and 17HT0042) and JST CREST under grant JPMJCR18A4.

### References

- [1] H. Hamamoto, N. Kurata, S. Saruwaari, M. Kawamoto, A. Tomioka, T. Daigo, "Field test of change detection system of building group in preparation for unexpected events in Gunkanjima," *AIJ J. Technol. Des.*, **24**(57), 553-558, 2018, doi:10.3130/aijt.24.553 (article in Japanese with an abstract in English).
- [2] A. Sasou, N. Odontselgel, S. Matsuoka, "An acoustic-based tracking system for monitoring elderly people living alone," in 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health, **2018-March**, 89-95, 2018, doi:10.5220/0006664800890095.
- [3] A. Quinlan, M. Kawamoto, Y. Matsusaka, H. Asoh, F. Asano, "Tracking Intermittently Speaking Multiple Speakers Using a Particle Filter," *EURASIP Journal on Audio, Speech, and Music Processing*, **2009**(1), 1-11, 2009, doi:10.1155/2009/673202.
- [4] Y. Suzuki, F. Asano, H. Kim, T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, **97**(2), 1119-1123, 1995, doi: 10.1121/1.412224.
- [5] E. Pampalk, Music analysis toolbox [Online]. Available: <http://www.pampalk.at/ma/documentation.html>.
- [6] H. Fastl, E. Zwicker, *Psychoacoustics Facts and Models*, Springer-Verlag Berlin Heidelberg, 1999.
- [7] M. R. Schroeder, D.S. Atal, J.L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *The Journal of the Acoustical Society of America*, **66**(6), 1647-1652, 1979, doi:10.1121/1.383662.
- [8] R. A. W. Bladon, B. Lindblom, "Modeling the judgment of vowel quality differences," *The Journal of the Acoustical Society of America*, **69**(5), 1414-1422, 1981, doi: 10.1121/1.385824.
- [9] E. Pampalk, A. Rauber, D. Dueck, "Content-based organization and visualization of music archives," in the 10th ACM International Conference on Multimedia 2002, 570-579, 2002, doi:10.1145/641007.641121.
- [10] M. Kawamoto, T. Hamamoto, "Building Health Monitoring Using Computational Auditory Scene Analysis," in 16th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2020, 144-146, 2020, doi:10.1109/DCOSS49796.2020.00033.
- [11] M. Müller, S. Ewert, "Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features," in the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, 215-220, 2011.
- [12] M. Kawamoto, "Sound-Environment Monitoring Method Based on Computational Auditory Scene Analysis," *Journal of Signal and Information Processing*, **8**(2), 65-77, 2017, doi:10.4236/JSIP.2017.82005.
- [13] B. J. Frey, D. Dueck, "Clustering by passing messages between data points," *Science*, **315**(5814), 972-976, 2007, doi:10.1126/SCIENCE.1136800.
- [14] K. Wang, J. Zhang, D. Li, X. Zhang, T. Guo, "Adaptive Affinity Propagation Clustering," *Acta Automatica Sinica*, **33**(12), 1242-1246, 2007.
- [15] M. Kawamoto, "Sound-environment monitoring technique based on computational auditory scene analysis," in 25th European Signal Processing Conference, EUSIPCO 2017, 2017-January, 2516-2524, 2017, doi:10.23919/EUSIPCO.2017.8081664.
- [16] T. Nishiura, T. Yamada, S. Nakamura, K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," in 2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **2**, 1053-1056, 2000, doi:10.1109/ICASSP.2000.859144.
- [17] H. L. Van Trees, *Optimum Array Processing*, Wiley Interscience: New York, 2002, doi:10.1002/0471221104.
- [18] D. H. Johnson, D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, New Jersey: Englewood Cliffs: PTR, Prentice Hall, 1993.
- [19] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, **AP-34**(3), 276-280, 1986, doi:10.1109/TAP.1986.1143830.