Title	
Author(s)	,
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17135
Rights	
Description	Supervisor: , ,

Study on anomaly sound detection using objective indices related to timbral attribute

1910245 Seigo Kura

Maintenance and inspection work is essential for the safe operation of machine in factories. This work is usually carried out by machine maintenance technicians, but there are various problems such as ensuring safety and maintaining costs. In order to solve these problems, technology for anomaly detection is being developed.This technology generally consists of feature extraction and detectors. In particular, for sound-based anomaly detection, emphasis has been placed on the development of highly accurate detectors using machine learning. However, in order to establish an anomaly detection method with higher performance than existing techniques, it is necessary to examine what acoustic features and perceptual factors are important for anomaly detection. A sound consists of three attributes: loudness, pitch, and timbre, which represent the three aspects of sound as an auditory impression. Unlike loudness and pitch, the psychological properties of timbre are expressed in a multidimensional manner and are complex. The problem of difficulty in anomaly sound detection may be due to the fact that only skilled technicians can identify the acoustic features and perceptual factors that are important for anomalous sound detection.

The purpose of this study is to clarify whether the evaluation index related to timbre is important for anomaly sound detection. In this study, Timbral Attribute used to investigate this issue. A timbre attribute is an adjective that describes the characteristics of a timbre, and is an index that is associated with the psychological quantity of a person's perception of sound. In this study, skilled engineers is considered that they rely on the complex differences in timbre between normal and abnormal sounds to judge abnormalities, and investigate which of the Timbral Attributes are important in anomaly sound detection.

To date, many research results on abnormal sound detection have been reported. One of the problems in anomaly sound detection is that it is difficult to collect a large amount of abnormal sounds themselves. In order to solve this problem, research has been reported on anomaly sound detection by using a large amount of data of only normal sounds and learning the features of normal sounds by machine learning. In Uematsu et al.'s research on anomaly sound detection using deep learning, they were able to detect anomaly sounds for a water discharge pump, a 3D printer, and a water supply pump using spectrograms as features. In the previous study, Suefusa et al.'s reported on the problems faced by autoencoders, which are often used

in unsupervised anomaly sound detection. The autoencoder learns to resynthesize normal sound and detects outliers by producing different outputs when abnormal sound is input. However, when the target machine sound is non-stationary, the reconstruction error tends to be large regardless of the abnormality due to the problem that prediction by reconstruction is difficult. To solve this problem, the input mel-spectrogram is divided into frames and the central frame is excluded as the input, which is then input to an autoencoder designed using a deep neural network. The reconstructed features as the output are interpolated with the center frame that was excluded in the input. As a result, the reconstruction error was matched between normal and abnormal sounds, and the identification accuracy was improved especially for non-stationary machine sounds. As described above, representative acoustic features such as mel-spectrogram are often extracted from sound data and used, and how to use these representative features for anomaly sound detection has been reported.

There are several acoustic features that are considered to be related to timbre. Typical examples are the spectral centroid and spectral slope, which are related to "brightness". In addition, some indices have been reported to calculate the sensation of sound such as brightness and hardness from acoustic features. Timbral attributes are adjectives used to describe the characteristics of timbre. Some of them have been implemented as Timbral models in a project called AudioCommons at the University of Surrey. In particular, eight metrics, Hardness, Depth, Brightness, Roughness, Warmth, Sharpness, Boominess, and Reverb, have been implemented. It was implemented because the most used search term when people search for sound sources on freesound.org, where a variety of sound sources are available, is the timbral attribute to which they belong. Each model (except Reverb) is implemented to take a value from 0 to 100. For example, Hardness means that the higher the value, the harder the sound.

The dataset of ToyADMOS and MIMII is used in this paper. ToyADMOS is a dataset created by NTT Media Intelligence Laboratory and designed to detect abnormal machine operation sounds. It contains normal and abnormal sounds for three types of toys. In this study, the ToyTrain is used, which is data for model trains. The MIMII dataset was used in DCASE together with ToyADMOS, and is an industrial equipment sound dataset for anomaly sound detection.

The proposed method consists of a feature extraction part and a detector. In the feature extraction part, eight timbral attribute are calculated using Timbral models. For each of the eight calculated values, all combinations (255 possible combinations) are extracted and input to the ditector. In the detector, a support vector machine is used to detect between normal and

abnormal.

In the study on ToyADMOS, the purpose is the timbre attribute can be used for anomaly sound detection using a simple detector such as threshold judgment by Receiver Operatorating Characteristic curve. In addition, the effectiveness of the proposed method is clarified by comparing the proposed method with the aforementioned results. In addition, the results of the proposed method is compared with the results of Mizuno's research to clarify whether the timbre attribute can be used for anomaly sound detection as well as the sound quality metrics.

In the study on the MIMII dataset, the purpose is the proposed method is effective for industrial equipment, instead of the toy failure sound used in the ToyADMOS study. In order to evaluate the effectiveness of the proposed method, we compared it with Baseline, which uses log-Mel-spectrograms as features. All the sound data in the MIMII dataset was used for the evaluation.

The results for the study on the ToyADMOS, Brighness and Sharpness showed the difference between normal and abnormal sounds except for one abnormal sound. The proposed method was able to detect all the normal and abnormal sounds.

Thre results for the study on the MIMII dataset, the proposed method was found to be more effective than the log-mel-spectrogram for all the individuals in the Valve and some of the individuals in the Pump and Slide rail.

By using Timbral attribute as a feature, high identification accuracy was achieved. Through the two studies, it is found that Brightness was the most frequently used evaluation metric in the proposed method. This indicates that Brightness is the most important of the timbral attributes. In addition, the number of dimensions of the features used in the proposed method is eight at most. From this, it is concluded that the timbral attribute represents the difference between normal and abnormal in some models better than other features that use a huge number of dimensions.