


Data Augmentation By Finite Element Analysis for Enhanced Machine Anomalous Sound Detection

Zhixian Zhang¹ , Yucong Zhang¹, and Ming Li^{1*} 

¹Suzhou Municipal Key Laboratory of Multimodal Intelligent Systems,
Duke Kunshan University, Kunshan, China
`ming.li369@dukekunshan.edu.cn`

Abstract. Current data augmentation methods for machine anomalous sound detection (MASD) suffer from insufficient data generated by real world machines. Open datasets such as audioset are not tailored for machine sounds, and fake sounds created by generative models are not trustworthy. In this paper, we explore a novel data augmentation method in MASD using machine sounds simulated by finite element analysis (FEA). We use Ansys, a software capable for acoustic simulation based on FEA, to generate machine sounds for further training. The physical properties of the machine, such as geometry and material, and the material of the medium is modified to acquire data from multiple domains. The experimental results on DCASE 2023 Task 2 dataset indicates a better performance from models trained using augmented data.

Keywords: Data augmentation · Finite element analysis · Machine anomalous sound detection.

1 Introduction

Machine anomalous sound detection (MASD) is the task to identify whether a sound clip is generated by a machine working normally or anomalously. It is commonly used in automatic machine condition monitoring due to its capability to reduce labor work and detect subtle changes in machine working condition that could be otherwise missed by manually monitoring. However, these advantages have the premise that the detection system has seen enough anomalous data. Due to the low probability of occurrence and high variance of the anomalous sounds, the training process often involves only normal sounds, and the task is often considered as an unsupervised learning problem [4,5,15].

To deal with the problems mentioned above, both generative methods and self-supervised methods are proposed in recent years. Without the anomalous data, generative methods allow the researchers to directly model the distribution of the normal data, and any data that does not fit the distribution is regarded as anomaly. Autoencoders (AE) [7,9], interpolation deep neural networks (IDNN) [20] and generative adversarial networks (GAN) based approaches [8,13] are popular generative models for MASD. However, generated

* Corresponding Author: Ming Li

methods are hard to extract effective features [17]. The other way to model the normal data is to build a feature extractor by self-supervised methods [2,11,18,25]. By categorizing various machine sounds into different machine types and IDs, self-supervised methods can learn more effective and more compact representations than the generative methods. The anomalies can then be identified in the feature space using distance metrics, such as cosine distance.

However, the lack of data is still a limitation when training a feature extractor in the self-supervised method. One way to overcome this issue is to use large open-source dataset to train the model. In [19], the researchers explore the effectiveness of using various pre-trained models that trained from human speech [1,3,12,22]. Despite the good performance in the field of automatic speech recognition (ASR), those pre-trained models are not trained by machine sounds and might not be suitable for MASD. [21] uses a pre-trained model PANNs [16], which is trained on AudioSet [6]. Although AudioSet is not a dataset for speech, it still contains sounds from various kinds of categories and is not tailored for machine sounds. Another way to address the lack of data is to use simulated data. Recently, the AI-generated content (AIGC) is a hot topic, so researchers who study MASD try to apply the AIGC techniques to generate machine sounds [21,13] or embeddings [24] for data augmentation. However, the AI-generated sounds are not explainable and trustworthy, since it is generated using deep learning techniques.

To tackle the issues above, we propose a novel data augmentation method by finite element analysis in this paper. FEA is a numerical approach that represents geometries as a system of linear equations built up by meshed nodes and links. The solution is formula-based and has actual physical meaning. FEA is widely applied in engineering to approximate solutions for various problems. The acoustic simulation based on FEA is a reliable data augmentation tool. To the best of our knowledge, we are the first to explore the FEA acoustic simulation for data augmentation in MASD.

2 Methods

We utilize FEA to conduct modal analysis aiming to identify its natural vibrational modes. Additionally, we perform harmonic frequency response analysis to ascertain the localized stresses occurring within the speaker. We use Ansys 2023 R1, a software capable for performing acoustic analysis using FEA, to simulate real world sound. We build a speaker model as the geometry to generate sound. To make the speaker vibrates, we simulate a harmonic force exert on the diaphragm of the speaker. The vibration of the diaphragm will generate sound by various kinds of mediums in the defined enclosure. A far-field microphone is set in front of the speaker to record sound signals. The generated sound is trained together with the machine sounds from DCASE 2023 dataset. The acoustic simulation of the speaker consists of geometry design, model analysis, harmonic response analysis and acoustic simulation. The overall pipeline is shown in the figure 1.

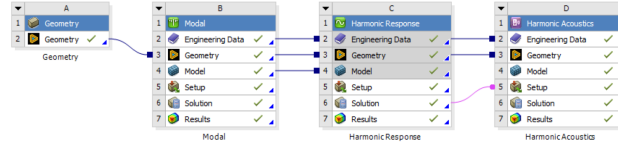


Fig. 1. Overall pipeline of the simulation

2.1 Geometric model

The geometry where we perform acoustic simulation is a simple speaker model. The design is done using SpaceClaim in Ansys. The speaker consists of a diaphragm and body holding the diaphragm. An enclosure is defined outside of the speaker to simulate the medium that the sound transmits. Figure 2 shows the geometry of the speaker and the enclosure.

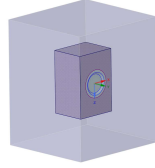


Fig. 2. Imported geometry in Ansys

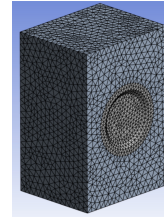


Fig. 3. Mesh of the speaker

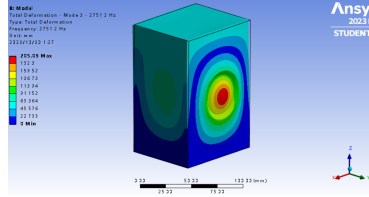


Fig. 4. Example of deformation of the speaker

2.2 Model analysis

The model analysis is performed to get the mesh of the model, creating the finite element model shown in Fig. 3. The natural frequencies of resonance which depends on the material assignment is also calculated to provide basic dynamic information of the model. The boundary condition of the speaker is fixed to the bottom of the body. The mesh of the model is created with the size of 5mm for the body and 3mm for the diaphragm. The speaker is therefore represented by the finite element model consists of nodes and links generated by the mesh.

2.3 Harmonic analysis

In harmonic analysis, we first import a constant force of magnitude 1N on the center of the diaphragm on the speaker. Harmonic response results in a sinusoidal excitation at each frequency. So this analysis can simulate vibration of

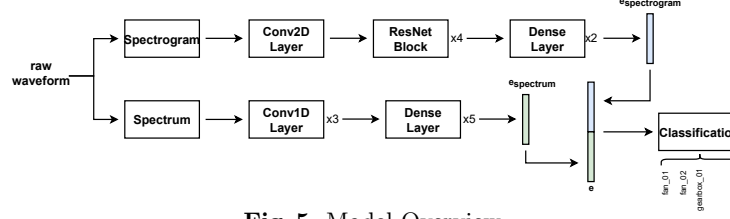


Fig. 5. Model Overview

the diaphragm propelled by the electromagnetic voice coil in real world speakers. An example of deformation response of the speaker is shown in Fig. 4. When calculating frequency response, the frequency range 0-20000Hz is divided into 50 parts with an attenuation ratio of 10% is assumed. The solution contains the deformation in every frequency bins.

2.4 Harmonic acoustics

After we get the vibration frequencies and velocity results in harmonic analysis, we can import the computed information to acoustic simulation. The frequency information in harmonic analysis will be synchronized to harmonic acoustics. The acoustic simulation will suppress the solid part of the model, and calculate the response in fluid domain. The material of the medium in the enclosure can be any fluids. In this paper, we use air and water as fluid options to further increase the sensitivity of the classifier. The mesh size is calculated based on the wavelength of the sound and the maximum frequency we want to achieve. Based on the sound speed c in the medium, the wavelength $\lambda = c/f$, where f is the maximum frequency we want to get. The size of the mesh is $\lambda/6$. After we have the mesh size for each fluid, we can create mesh of the fluid in the enclosure. We then set the boundary condition to radiation boundary, which means that when the energy hits the boundary of the enclosure it gets radiated away instead of being reflected back. That means our acoustic analysis is accurate and is focused solely on the vibration of the speaker. We set a far-field microphone in front of the speaker to collect sound information. After solving the acoustic scenario described above, we get information of sound pressure level. We then export the result in the microphone to wav file to get the augmented data.

2.5 Backbone Model

We adopt and re-implement the similar model described in [23] for MASD, where researchers show the effectiveness of this model on previous DCASE 2022 challenge. We use this model as our backbone model for two reasons. First, it is a popular and effective model. State-of-the-art performance is reported on both DCASE 2022 [23] and DCASE 2023 challenge [14]. Second, the model has two encoding pathways, which is investigated already by [18] and proved to be effective for MASD. Hence, in this paper, we use this dual-path model to test the effectiveness of our data augmentation technique.

As is shown in Fig. 5, the input waveform is encoded by two pathways. In the spectrum pathway, the machine sound is transformed into utterance-level spectrum, which aims at finding patterns over the whole utterance. The spectrum is processed by three 1D convolutional layers and five dense layers. For the spectrogram pathway, the machine sound is first converted to spectrogram

Table 1. Structure of the encoder. n indicates the number of layers or blocks, c is the number of output channels, k is the kernel size and s is the stride. h and w are the output height and width of the ResNet Blocks.

Operator	n	c	k	s
Conv2D 7x7	1	32	(7,7)	(2,2)
MaxPooling	-	-	(3,3)	(2,2)
ResNet block	4 (64, 128, 128, 128)	(3,3)	(2,2)	
MaxPooling	-	-	(h, w)	(h, w)

by short-term fourier transform (STFT), and then fed to a 2D convolutional layer and 4 ResNet [10] blocks shown in Table 1.

3 Experiments and results

3.1 Acoustic simulation

To simulate the condition of anomalous machines, we manually create holes of different sizes and locations on the diaphragm. Since the diaphragm is the key component of sound generation on the speaker, the changes on the diaphragm can be well represented by the sound the speaker generates. In the experiment, we have four speaker geometry designs and each of them contains different holes on the diaphragm to create different types of sounds. The comparison of spectrogram of sound generated by four geometry designs is shown in Fig. 6.

For material assignment, normally we have wood on the speaker body and polypropylene for the diaphragm of the speaker. To make the classifier detect more subtle differences between sounds generated by the speaker, we expanded the material set for body: {wood, steel, plastic} and diaphragm: {steel, polypropylene}. In acoustic analysis, the set for material of the medium in the enclosure is {air, water}. We created geometry model on every combination of the materials. The far-field microphone is located 30mm away from the center of the diaphragm of the speaker.

For every speaker geometry design, we assign all possible material combinations in the material sets mentioned above. After solving the result in the microphone and get the sound pressure level at each frequency, we use inverse fourier transform to generate the sounds with a fixed sample rate of 16kHz. As a result, we get 48 different classes of simulated sounds. Within each class, we get 100 audio clips of 10 seconds long each.

3.2 Dataset

The experiments are conducted on the development part of DCASE 2023 Task 2 dataset [4], containing audio clips from seven distinct machine types. Each machine type has about 1,000 audio clips. Each audio clip lasts 10 seconds with a sampling rate of 16 kHz. Only normal data in the training dataset is used for training, and the results are evaluated on the test dataset.

3.3 Result

We compare the AUC results of training the MASD model with and without the augmented data. The result comparing model performance using only DCASE

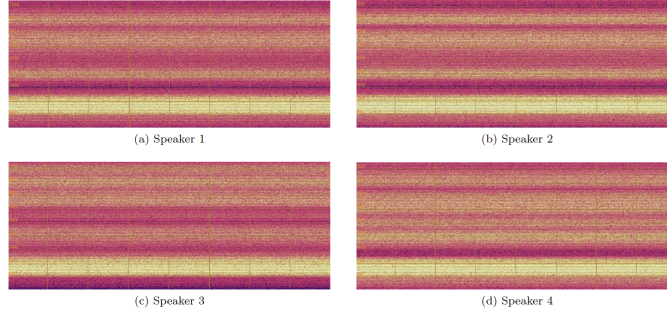


Fig. 6. Spectrogram of generated sounds on 4 speakers with different holes

2023 dataset for training or adding the simulated data together is shown in Table 2. The data augmentation method of adding simulated data can improve the model performance on most of the machine types in the development set. The average AUC of all machines is also enhanced by the introduction of augmented data. This means that the data augmentation method using acoustic simulation based on FEA is effective in increasing the performance of the MASD model. In addition, we show the results of the official baseline [7] and top-ranked ensemble model [19] using multiple pre-trained models for comparison. From Table 2, we can see that our system outperforms the baseline for a large margin and achieves comparable performance with the ensemble model that uses large pre-trained models.

4 Conclusion

This paper proposes a novel data augmentation method of simulating machine sounds based on finite element analysis (FEA). Subtle changes in geometry design and different combinations of materials enable the MASD model to be more sensitive to sound variation under different machine types and different domains. The performance of MASD model trained with simulated data outperforms the one that is trained without simulated data. It can be proved that the acoustic simulation based on FEA is effective in enhancing the performance of the MASD model. In the future, we might change the basic structure of the geometry to add more varieties to the machine types or even build geometry that is tailored for certain machine type, so that the simulated data can be more effective.

Acknowledgement This research is funded in part by Science and Technology Program of Suzhou City (SYC2022051) and National Natural Science Foundation of China (62171207). Many thanks for the computational resource provided by the Advanced Computing East China Sub-Center.

References

1. Baevski, A., Zhou, H., rahman Mohamed, A., Auli, M.: wav2vec 2.0: A framework for self-supervised learning of speech representations. ArXiv **abs/2006.11477** (2020)

Table 2. Average AUC% and pAUC % of every machine type on DCASE 2023 development dataset

Machine type		Simulated Data		Official	Z. Lv
		No	Yes	Baseline [7]	et al. [19]
Valve	AUC	87.73	75.00	53.74	73.66
	pAUC	62.15	59.47	51.28	53.68
Gearbox	AUC	78.81	73.79	71.58	82.28
	pAUC	63.05	55.32	54.84	62.47
Fan	AUC	68.63	70.52	61.89	65.97
	pAUC	58.63	58.10	58.42	56.32
Bearing	AUC	57.90	69.78	59.77	78.80
	pAUC	49.89	51.05	50.68	62.26
ToyTrain	AUC	51.29	57.66	48.73	64.82
	pAUC	48.05	50.36	48.05	49.32
ToyCar	AUC	52.40	58.10	59.20	65.47
	pAUC	50.74	52.79	49.18	49.47
Slide Rail	AUC	86.11	92.28	79.25	94.74
	pAUC	65.57	78.63	56.18	76.68
Average	AUC	68.98	71.02	62.02	75.11
	pAUC	56.86	57.96	52.66	58.60

2. Chen, H., Song, Y., Dai, L., Mcloughlin, I., Liu, L.: Self-supervised representation learning for unsupervised anomalous sound detection under domain shift. Proc. ICASSP 2022 pp. 471–475 (2022)
3. Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., Li, J., Kanda, N., Yoshioka, T., Xiao, X., Wu, J., Zhou, L., Ren, S., Qian, Y., Qian, Y., Zeng, M., Wei, F.: Wavlm: Large-scale self-supervised pre-training for full stack speech processing. IEEE J-STSP **16**, 1505–1518 (2021)
4. Dohi, K., Imoto, K., Harada, N., Niizumi, D., Koizumi, Y., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Kawaguchi, Y.: Description and discussion on dcase 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring. ArXiv [abs/2305.07828](#) (2023)
5. Dohi, K., Imoto, K., Harada, N., Niizumi, D., Koizumi, Y., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Yamamoto, M., Kawaguchi, Y.: Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques. In: Proc. of DCASE 2022 Workshop (2022)
6. Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M., Ritter, M.: Audio set: An ontology and human-labeled dataset for audio events. In: Proc. of ICASSP 2017. pp. 776–780 (2017)
7. Harada, N., Niizumi, D., Ohishi, Y., Takeuchi, D., Yasuda, M.: First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline. ArXiv [abs/2303.00455](#) (2023)
8. Hatanaka, S., Nishi, H.: Efficient gan-based unsupervised anomaly sound detection for refrigeration units. In: Proc. of ISIE. pp. 1–7. IEEE (2021)

9. Hayashi, T., Yoshimura, T., Adachi, Y.: Conformer-based id-aware autoencoder for unsupervised anomalous sound detection. Tech. rep., DCASE 2020 Challenge (2020)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. of CVPR 2016. pp. 770–778 (2015)
11. Hojjati, H., Armanfard, N.: Self-supervised acoustic anomaly detection via contrastive learning. Proc. of ICASSP 2022 pp. 3253–3257 (2021)
12. Hsu, W.N., Bolte, B., Tsai, Y.H.H., Lakhotia, K., Salakhutdinov, R., rahman Mohamed, A.: Hubert: Self-supervised speech representation learning by masked prediction of hidden units. IEEE/ACM TASLP **29**, 3451–3460 (2021)
13. Jiang, A., Zhang, W.Q., Deng, Y., Fan, P., Liu, J.: Unsupervised anomaly detection and localization of machine audio: A gan-based approach. In: Proc. of ICASSP. pp. 1–5. IEEE (2023)
14. Jie, J.: Anomalous sound detection based on self-supervised learning. Tech. rep., DCASE 2023 Challenge (June 2023)
15. Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., Purohit, H., Suefusa, K., Endo, T., Yasuda, M., Harada, N.: Description and discussion on dcase2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring. In: Proc. of DCASE 2020 Workshop (2020)
16. Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., Plumbley, M.D.: Panns: Large-scale pretrained audio neural networks for audio pattern recognition. IEEE/ACM TASLP **28**, 2880–2894 (2019)
17. Kuroyanagi, I., Hayashi, T., Takeda, K., Toda, T.: Improvement of serial approach to anomalous sound detection by incorporating two binary cross-entropies for outlier exposure. In: Proc. of EUSIPCO 2022. pp. 294–298 (2022)
18. Liu, Y., Guan, J., Zhu, Q., Wang, W.: Anomalous sound detection using spectral-temporal information fusion. In: Proc. of ICASSP 2022. pp. 816–820 (2022)
19. Lv, Z., Han, B., Chen, Z., Qian, Y., Ding, J., Liu, J.: Unsupervised anomalous detection based on unsupervised pretrained models. Tech. rep., DCASE 2023 Challenge (June 2023)
20. Suefusa, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., Kawaguchi, Y.: Anomalous sound detection based on interpolation deep neural network. In: Proc. of ICASSP 2020. pp. 271–275. IEEE (2020)
21. Tian, J., Zhang, H., Zhu, Q., Xiao, F., Liu, H., Mei, X., Liu, Y., Wang, W., Guan, J.: First-shot anomalous sound detection with gmm clustering and finetuned attribute classification using audio pretrained model. Tech. rep., DCASE 2023 Challenge (June 2023)
22. Wang, C., Wu, Y., Wu, Y., Qian, Y., Kumatani, K., Liu, S., Wei, F., Zeng, M., Huang, X.: Unispeech: Unified speech representation learning with labeled and unlabeled data. ArXiv **abs/2101.07597** (2021)
23. Wilkinghoff, K.: Design choices for learning embeddings from auxiliary tasks for domain generalization in anomalous sound detection. In: Proc. of ICASSP 2023. pp. 1–5 (2023)
24. Zeng, X., Song, Y., McLoughlin, I., Liu, L., Dai, L.: Robust prototype learning for anomalous sound detection. In: Proc. of INTERSPEECH 2023 (2023)
25. Zhang, Y., Hongbin, S., Wan, Y., Li, M.: Outlier-aware Inlier Modeling and Multi-scale Scoring for Anomalous Sound Detection via Multitask Learning. In: Proc. of INTERSPEECH 2023. pp. 5381–5385 (2023)