Hammering Inspection System Using HPSS and Gradient Boosting with a Wall-Climbing Robot

Nichika Koyama* and Nari Tanabe† and Masaya Fujisawa‡
* Tokyo University of Science, Tokyo, Japan
E-mail: 4624517@ed.tus.ac.jp
† Suwa University of Science, Nagano, Japan
E-mail: nari@rs.sus.ac.jp
‡ Tokyo University of Science, Tokyo, Japan
E-mail: fujisawa@rs.tus.ac.jp

***Abstract*—This paper proposes an automatic judgment system that suppresses noise in the observed signal, which contains both hammering sound and noise, using Harmonic/Percussive Sound Separation (HPSS) to extract hammering sound, and then uses machine learning to judge whether the inspection area is healthy or damaged. The proposed method calculates hammering sounds from the observed signal using HPSS, and extracts features considering the timing of the hammering sound occurrence from the hammering sound. Subsequently, Gradient Boosting is used to automatically judge whether the area is healthy or damaged. The feature of the proposed method is that it suppresses noise while retaining the characteristic hammering sound necessary for judging healthy and damaged areas, and by extracting features considering the timing of the hammering sound occurrence, it allows for highly accurate classification.**

I. Introduction

One of the nondestructive inspection methods to determine the condition of invisible internal damage in bridges, highways and tunnels is hammering inspection. Hammering inspection involves lightly tapping the concrete structure's wall surface with a hammer and detecting internal abnormalities from the differences in the sounds produced. Recently, in Japan, there was an accident in which a ceiling panel of a tunnel fell.[1]

The cause was the detachment, loosening and corrosion of the anchor bolts, along with cracks in the protective concrete. However, the damage was overlooked because hammering inspection had not been conducted for 12 years. This accident has led to renewed recognition of the importance of hammering inspection and the appearance of three issues.

The first issue is that the noise overlaps the hammering sound, making them difficult to hear. Since the road cannot be closed during inspection, the inspection is conducted in an environment where vehicles are constantly passing. Therefore, the noise of vehicle traffic and surrounding environmental sounds is constantly heard as noise.

The second issue is that the accuracy of determining whether the inspection area is healthy or damaged varies depending on the level of skill of the operator. Experienced operators can distinguish whether the inspection area is healthy or damaged by listening to the frequency, amplitude, and timbre differences of the hammering sound. However, if the operator's ability to distinguish these sounds is not high, it is difficult to accurately determine whether the area is healthy or damaged.

The third issue is the danger to life due to hazardous inspection environments. On highways and bridges, work at high altitudes is necessary, and in tunnels, vehicles pass very close to the operators, putting them at risk.

Recently, research has been conducted using wall-climbing robots to perform hammering inspection to eliminate the third issue of life-threatening danger. However, the drive noise of the wall-climbing robots and the vacuum noise used to adhere to the wall are observed as noise, leaving the first issue unresolved. Additionally, since operators still need to listen to the hammering sound to determine whether the area is healthy or damaged, the second issue remains unresolved. Therefore, this paper proposes a system that suppresses noise suitable for the characteristics of hammering sound and noise identified through spectrogram analysis and extracts features considering the timing of the hammering sound occurrence. Gradient Boosting is then used to automatically judge whether the area is healthy or damaged.

II. Previous Method

In recent years, research has been conducted on the digitization of hammering inspection. Previous research [2][3] have focused on the features that represent the differences between healthy and damaged areas by conducting hammering inspection on bridges where deterioration was confirmed in real environments. In previous research [2], hammering inspection were performed using inspection hammers and rotary hammers, and the characteristics of acceleration and sound pressure obtained from the data were compared, demonstrating the effectiveness of the inspection method that evaluates sound pressure using a rotary hammer. In previous research [3], hammering inspection were performed on bridges with recognized deterioration and damage, specifically in areas with noticeable delamination and separation. The research verified changes in sound pressure characteristics between healthy and damaged areas, and a clear difference was observed.

Additionally, artificial damage areas were created to confirm the effectiveness of the frequency characteristics of sound pressure. As a result, it was confirmed that the frequency spectrum varies depending on the scale and condition of the damaged areas, but using a self-organizing map, judgment is possible. Moreover, research on hammering inspection using machine learning has also been reported in order to automate the judgment of healthy and damaged areas [4]. Previous research [4] compared multiple machine learning methods and neural networks for judgement. However, considering the use in the field, it is difficult to extract accurate features because the noise overlaps with the hammering sound. In previous research, noise was recorded separately and used as a reference or white noise was assumed, but noise suppression was not the main focus, leading to errors in judgment accuracy in real environments. In response, previous research [5] proposed an automatic judgment system using noise suppression with a Butterworth filter and a colored source-driven Kalman filter and a nonlinear Support Vector Machine, a type of machine learning, considering the time transition of the hammering sound for feature extraction. However, these studies could not solve the problems of life-threatening work at high altitudes, which is one of the problems with hammering inspection. Therefore, in recent years, research has been conducted on hammering inspection using drones and wall-climbing robots [6]. Hammering inspection using these robots can be conducted at high altitudes without life-threatening risks, but drive noise and vacuum noise overlap with hammering sound as noise.

Therefore, this paper proposes a hammering inspection support system with noise suppression suitable for the characteristics of hammering sound and noise identified through spectrogram analysis, a feature extraction method that extracts only hammering sound features, and Gradient Boosting.

III. Proposed Method

First, we define the observed signals obtained during hammering sound inspection. The observed signals obtained during hammering inspection are stereo audio signals, and the observed signals at time n can be expressed as $x_{L/R}(n)$, represented by the following equation:

$$x_{L/R}(n) = s_{L/R}(n) + v_{L/R}(n) \tag{1}$$

where $s_{L/R}(n)$ denotes the hammering sound signal, and $v_{L/R}(n)$ denotes the noise signal.

A. Noise suppression using HPSS

The observed spectrogram $X_{L/R}(\lambda, k)$, obtained by performing a Short-Time Fourier Transform (STFT) on the observed signal $x_{L/R}(n)$, is given by

$$X_{L/R}(\lambda, k) = \sum_{\mu=0}^{W-1} x_{L/R}(\lambda N + \mu) \cdot h(\mu) \cdot \exp\left(-j\frac{2\pi\mu k}{W}\right) \tag{2}$$

where λ is the frame number, k is the frequency bin number, μ is the bin number within the frame, W is the Discrete Fourier Transform (DFT) frame length, and N is the frame interval. A Hanning window function $h(\mu)$ is used.

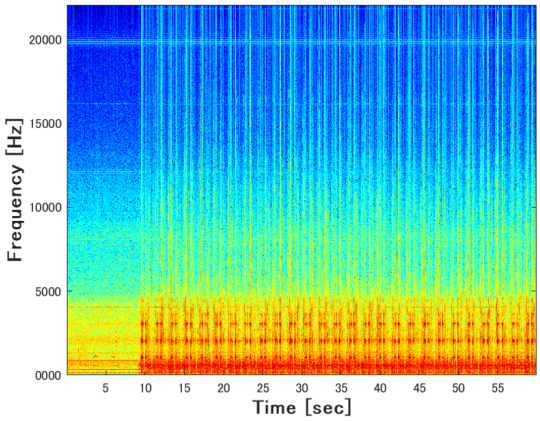


Fig. 1: Frequency analysis

From the spectrogram analysis (Fig. 1), it was found that hammering sound are continuous in the frequency direction, while noise is continuous in the time direction. Therefore, we applied Harmonic/Percussive Sound Separation (HPSS), an effective sound source separation method when such characteristics are present. However, while noise was effectively suppressed, the characteristic sounds (hereafter referred to as "characteristic hammering sound") disappeared, causing the feature distributions of sound and damaged areas to overlap, making judgment difficult.

Focusing on noise and characteristic hammering sound, it was found that noise is continuous in the time direction, while characteristic hammering sound occur intermittently in the time direction at the timing of hammering sound occurrences. This suggests that the disappearance of characteristic hammering sound is due to them being suppressed as noise since they occur intermittently in the time direction.

Additionally, the enhanced hammering spectrogram calculated in the next section uses a median filter applied to the frequency direction of the observed spectrogram. However, it was found that the relatively large characteristic hammering sound existing in a certain frequency band could converge to the median value within the range and disappear when the median filter is applied. Therefore, we determine the processing range in the frequency direction of the median filter for each frame based on the signal-to-noise (SN) ratio of the estimated hammering spectrum $\bar{S}_{L/R}(\lambda - 1, k)$ and the estimated noise spectrum $\bar{V}_{L/R}(\lambda - 1, k)$ from the previous frame. By narrowing the filter range when hammering sound are present and broadening it when they are not, we avoid applying the median filter to characteristic hammering sound.

The average previous frame SNR $\overline{\mathrm{SNR}}(\lambda, k)$ is given by

$$\overline{\mathrm{SNR}}(\lambda, k) = \frac{1}{K}\sum_{k=0}^{K-1} \mathrm{SNR}(\lambda, k) \tag{3}$$

where K is the number of samples in one frame, and

$\mathrm{SNR}(\lambda, k)$ is the length of the DFT frame:

$$\mathrm{SNR}(\lambda, k) = 20 \log_{10} \frac{\bar{S}_{L/R}(\lambda - 1, k)}{\bar{V}_{L/R}(\lambda - 1, k)} \tag{4}$$

Next, using the calculated average previous frame SNR, we compute the frequency direction processing range l_h for each frame. Since large noise often exists in the low-frequency range in the wall-climbing robot-based hammering sound inspection environment, we define the frequency direction processing range l_h as follows:

$$l_h\,[\mathrm{Hz}] = \begin{cases} f_n & (\overline{\mathrm{SNR}}(\lambda) < -5) \\ 1500 & (-5 \leq \overline{\mathrm{SNR}}(\lambda) < 5) \\ 1300 & (5 \leq \overline{\mathrm{SNR}}(\lambda) < 15) \\ 1000 & (15 \leq \overline{\mathrm{SNR}}(\lambda)) \end{cases} \tag{5}$$

where f_n is the Nyquist frequency.

By applying median filters in both the frequency and time directions, we calculate the enhanced hammering spectrogram and enhanced noise spectrogram. A median filter outputs the median of the input data, as expressed by the following equation:

$$y(n) = M\left\{x\left(n - \frac{d}{2} : n + \frac{d}{2}\right), d\right\} \tag{6}$$

where $x(n)$ is the input data and d is the filter length. the enhanced hammering spectrogram $\hat{S}_{L/R}(\lambda, k)$, which emphasizes the hammering components that are continuous in the frequency direction, is calculated by applying the median filter with the frequency direction processing range l_h obtained in Eq.(5). The enhanced hammering spectrogram $\hat{S}_{L/R}(\lambda, k)$ is given by

$$\hat{S}_{L/R}(\lambda, k) = M\left\{X_{L/R}\left(k - \frac{l_h}{2} : k + \frac{l_h}{2}\right), l_h\right\} \tag{7}$$

Next, the enhanced noise spectrogram $\hat{V}_{L/R}(\lambda, k)$, which emphasizes the noise components, is calculated by applying the median filter with an arbitrarily set filter length l_n in the time direction. The enhanced noise spectrogram $\hat{V}_{L/R}(\lambda, k)$ is given by

$$\hat{V}_{L/R}(\lambda, k) = M\left\{X_{L/R}\left(\lambda - \frac{l_n}{2} : \lambda + \frac{l_n}{2}\right), l_n\right\} \tag{8}$$

The values of the enhanced spectrograms in the current frame number λ_{now} for each enhanced spectrogram are denoted as the enhanced hammering spectrum $\hat{S}_{L/R}(\lambda_{now}, k)$ and the enhanced noise spectrum $\hat{V}_{L/R}(\lambda_{now}, k)$. Applying a Wiener filter to each enhanced spectrum, we calculate the estimated hammering spectrum $\bar{S}_{L/R}(\lambda_{now}, k)$. The estimated hammering spectrum $\bar{S}_{L/R}(\lambda_{now}, k)$ is given by

$$\bar{S}_{L/R}(\lambda_{now}, k) = X_{L/R}(\lambda_{now}, k) \cdot Q_{L/R}(\lambda_{now}, k) \tag{9}$$

where the observed spectrum is $X_{L/R}(\lambda_{now}, k)$, and the estimated hammering coefficient $Q_{L/R}(\lambda_{now}, k)$ is given by

$$Q_{L/R}(\lambda_{now}, k) = \frac{\hat{S}_{L/R}(\lambda_{now}, k)}{\hat{S}_{L/R}(\lambda_{now}, k) + \rho \cdot \hat{V}_{L/R}(\lambda_{now}, k)} \tag{10}$$

where ρ is the noise suppression parameter. Finally, by adding phase information to the estimated hammering spectrum $\bar{S}_{L/R}(\lambda_{now}, k)$ obtained in Eq.(9) and performing an inverse short-time Fourier transform (ISTFT), we calculate the estimated hammering signal $\overline{s}_{L/R}(n)$.

B. Feature extraction

Next, we extract features from the estimated hammering signal $\overline{s}_{L/R}(n)$. In this study, we used spectral peaks and their frequencies as features. However, since the rotary hammer used generates impacts in an arc, there is a pause when the hammer stops at both ends of the arc. Therefore, extracting one feature per frame (0.046 s) could lead to extracting features during periods when no impact is occurring. To address this, we consider selecting one feature from multiple frames.

First, we perform short-time fourier transform (STFT) on the estimated hammering signal $\overline{s}_{L/R}(n)$ to calculate the estimated hammering spectrum $\overline{S}_{L/R}(\lambda_{now}, k)$. Next, we search for peaks in the estimated hammering spectrum $\overline{S}L/R(\lambda now, k)$ and calculate the estimated hammering spectrum peak P_{λ_{now},k_p} and its frequency f_{λ_{now},k_p}. The estimated hammering spectrum peak is given by:

$$P_{\lambda_{now},k_p} = \max\left\{\hat{S}_{\lambda_{now},1}, \hat{S}_{\lambda_{now},2}, \cdots, \hat{S}_{\lambda_{now},K}\right\} \tag{11}$$

where k_p is the frequency bin number of the spectral peak.

Next, we prepare feature extraction frames of length l_T and select the maximum value $\overline{P}_\tau$ of the estimated hammering spectrum peaks and its frequency $\overline{f}_\tau$ from these frames to form the feature vector $\boldsymbol{\theta}_\tau$. Here, τ is an index that increments for each feature extraction frame l_T. The maximum value of the estimated hammering spectrum peaks for the feature extraction frame l_T is given by:

$$\overline{P}_\tau = \max\left\{P_{\lambda_i,k_{p_i}}, P_{\lambda_i,k_{p_{i+1}}}, \cdots, P_{\lambda_i,k_{p_{i+l_T}}}\right\} \tag{12}$$

C. Classification of healthy and damaged areas using Gradient Boosting

To classify healthy areas as 1 and damaged areas as 0, we define the class label t_τ and combine it with the extracted feature vector $\boldsymbol{\theta}_\tau$ to create the training data $\boldsymbol{r}_\tau$, as follows:

$$\boldsymbol{r}_\tau = \{t_\tau, \boldsymbol{\theta}_\tau\} = \{t_\tau, (f_\tau, \overline{P}_\tau)\} \tag{13}$$

We use Gradient Boosting to classify healthy and damaged areas using training data $\boldsymbol{r}_\tau$. First, we calculate the initial model $F_0(\boldsymbol{\theta})$ using:

$$F_0(\boldsymbol{\theta}) = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{N} L(t_i, c) \tag{14}$$

where N is the number of samples and $L(t_i, c)$ is the cross-entropy loss function, given by:

$$L(t_i, c) = -[t_i \log(c) + (1 - t_i) \log(1 - c)] \tag{15}$$

Next, we calculate the residuals $r_i^{(m)}$ between the predicted labels and the true labels of the $m-1$-th model $F_{m-1}(\boldsymbol{\theta})$

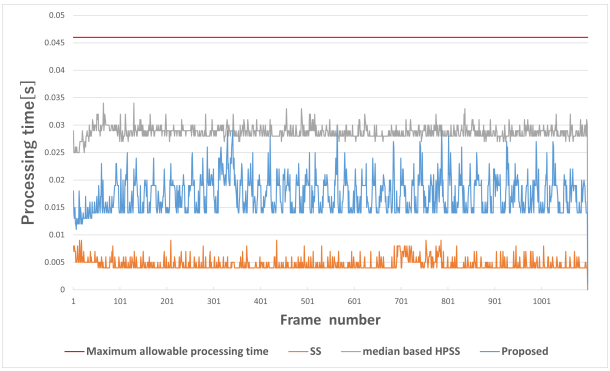


Fig. 2: Processing speeds

using:

$$r_i^{(m)} = -\left[\frac{\partial L(t_i, F(\boldsymbol{\theta}))}{\partial F(\boldsymbol{\theta})}\right]_{F(\boldsymbol{\theta})=F_{m-1}(\boldsymbol{\theta})} \tag{16}$$

Then, a decision tree h_m is fitted to the dataset $\{(\boldsymbol{\theta}_i, r_i^{(m)})_{i=1}^n\}$. We update the model by adding the new model to the $m-1$-th model:

$$F_m(\boldsymbol{\theta}) = F_{m-1}(\boldsymbol{\theta}) + \alpha h_m(\boldsymbol{\theta}) \tag{17}$$

where α represents the learning rate. By updating the model $M-1$ times in this manner, we obtain the final strong classifier $F_M(\boldsymbol{\theta})$. In this study, we use LightGBM[7], which offers fast processing speeds.

IV. Simulation

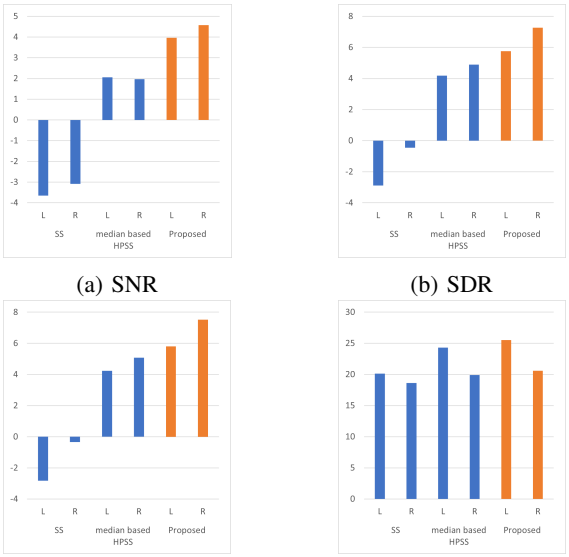
In the simulation, we compare the proposed method with previous methods in terms of noise suppression performance and classification performance. Noise suppression performance is evaluated by measuring processing speed and assessing noise suppression degree and hammering sound quality. Classification performance is evaluated by extracting features from each noise-suppressed hammering signal and comparing evaluation metrics.

A. Noise suppression performance evaluation

In the noise suppression performance evaluation experiment, we compare the proposed method with previous methods such as Spectral Subtraction (SS) [8][9] and Harmonic/Percussive Sound Separation (HPSS) [10][11], to demonstrate the effectiveness of the proposed method.

In the processing speed evaluation experiment, we measure the time required to estimate the hammering signal from the observed signal within one frame. If the proposed method and previous methods process within the maximum allowable processing time per frame (= 0.046 s), they are considered capable of real-time processing. Fig. 2 shows the processing speed evaluation. From Fig. 2, it can be confirmed that the processing time of the proposed method, indicated by the blue line, is below the maximum allowable processing time (= 0.046 s), demonstrating its capability for real-time processing.

Next, we evaluate noise suppression performance using SNR (Average Signal to Noise Ratio), SDR (Average Source to Distortion Ratio), SAR (Average Source to Artifacts Ratio), and SIR (Average Source to Interferences Ratio)[12]. The results of these evaluations are shown in Fig. 3. Here, SNR represents the ratio of the target sound to noise, SDR represents the quality of the target sound, SAR represents the minimal distortion caused by the separation process, and SIR represents the degree of separation. Higher values for each metric indicate better performance.



(a) SNR (b) SDR

(c) SAR (d) SIR

Fig. 3: Evaluation

TABLE I: Healthy area

	healthy area		
	Precision	Recall	F1-score
No processing	0.81	0.81	0.81
SS	0.79	0.73	0.76
Median based HPSS	0.76	0.87	0.81
proposed	0.86	0.93	0.89

TABLE II: Damaged area

	damaged area		
	Precision	Recall	F1-score
No processing	0.80	0.81	0.80
SS	0.74	0.80	0.77
Median based HPSS	0.85	0.72	0.78
proposed	0.92	0.85	0.89

From Fig. 3(a), 3(b), 3(c), and 3(d), it is evident that the proposed method outperforms previous methods in all evaluation metrics. This confirms that the proposed method provides superior noise suppression, hammering sound quality, minimal distortion, and degree of separation compared to previous methods.

B. Classification performance evaluation

For classification performance evaluation, we used cross-validation. The experimental data included 50-second WAV files collected using a wall-climbing robot, with three samples each for healthy and damaged areas. After noise suppression,

we compared precision, recall, F1-score, and accuracy for feature extraction and LightGBM application to verify the effectiveness of the proposed method. The results of the classification performance evaluation are shown in TABLE I and II. From TABLE I and II, the proposed method showed higher precision, recall, and F1-score compared to no processing, SS method, and median-based HPSS. Additionally, accuracy was 0.89 for the proposed method, 0.81 for no processing, 0.77 for the SS method, and 0.80 for median-based HPSS. Thus, the proposed method achieved the highest values for all classification evaluation metrics and was able to classify healthy and damaged areas with an average accuracy of 89%.

V. Conclusions

This paper proposed a Hammering Sound Inspection System Using HPSS and Gradient Boosting with a Wall-Climbing Robot and demonstrated its effectiveness through noise suppression and classification performance evaluations.

The proposed method has two main features. First, it determines the frequency direction median filter application range in the median-based HPSS from the SNR of previous frames, enabling noise suppression while preserving distinctive hammering sound present in the observed signal. Second, it performs feature extraction considering the timing of hammering sound generation, avoiding the extraction of features from periods when no hammering sound is occurring. The effectiveness of the proposed method was demonstrated by its superior performance in both noise suppression and classification metrics compared to previous methods.

References

[1] Research and R. Committee, "Report of the investigation and review committee on the falling tunnel ceiling plate accident," 2013.

[2] M. Omagari, Y. Sonoda, and O. Munemoto, "Fundamental study on rotational impact inspection of deteriorated concrete bridges," concrete engineering annual journal," *Concrete Engineering Annual Journal*, vol. 31, no. 1, 2009.

[3] Y. Sonoda and T. Watanabe, "Quantitative evaluation of defect detection capability in rotational impact inspection," *Journal of Structural Engineering A*, vol. 59, pp. 682–692, 2013.

[4] E. Hisao, B. Naoyuki, A. Hiromoto, and N. Yamato, "Evaluation of flaw detection in impact testing using ai techniques," *ScienAI and Dce Journal*, vol. 1, no. J1, pp. 514–521, 2020.

[5] K. Hori, N. Tanabe, and M. Fujisawa, "Nonlinear svm-type automatic dicision algorithm in noisy environment for hammering test system," *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, pp. 275–281, 2021.

[6] K. Yamashita, S. Ogata, and T. Shimamura, "Improved spectral subtraction utilizing iterative processing," *Electronics and Communications in Japan*, vol. 90, no. 4, pp. 39–51, 2006.

[7] G. K. et al, "Lightgbm a highly efficient gradient boosting decision tree," *in Advances in Neural Information Processing Systems*, vol. 30, 2017.

[8] R. Martin, "Spectral subtraction based on minimum statistics," *power*, vol. 6, no. 8, pp. 1182–1185, 1994.

[9] S. K. Roy and K. K. Paliwal, "A noise psd estimation algorithm using derivative-based high-pass filter in non stationary noise conditions," *EURASIP Journal on Audio, Speech, and Music Processing*, no. 1, p. 32, 2021.

[10] D. FitzGerald, "Harmonic/percussive separation using median filtering," *in Proc. 13th Int. Conf. Digital Audio Effects, Graz, Austria*, 2010.

[11] S. Oyabu, D. Kitamura, and K. Yatabe, "Linear multichannel blind source separation based on timefrequency mask obtained by harmonic/percussive sound separation," *in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021.

[12] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.