

Detection Of Micro Vibrations In Air Duct Structures Using Phase-Based Motion Magnification And Fast Fourier Transform

Muhammad Hideyosi Berry^{1,2*}, Andi Adriansyah¹

¹Department of Electrical Engineering, Universitas Mercu Buana, Jakarta

²Papaja Maju Mandiri, PT, Jakarta

*muhammadhideyosi@gmail.com

Abstract— Micro-vibrations in HVAC air-duct structures can compromise energy efficiency and structural integrity, yet conventional contact-based vibrometry is constrained by installation complexity and environmental sensitivity. This study develops a non-contact micro-vibration detection system by integrating Phase-Based Motion Magnification (PBMM) with Fast Fourier Transform (FFT) analysis, and validates its performance against a Micro-Electro-Mechanical Systems (MEMS)-based ADXL345 accelerometer. High-resolution 4K video (60 fps) of a galvanized, rigid square duct (0.8 m × 0.8 m) was processed through a steerable complex pyramid, ideal band-pass temporal filtering, and phase amplification ($\alpha = 20$). The magnified phase sequences were reconstructed and subjected to Hann-windowed, detrended FFT, with dominant frequency peaks refined via parabolic interpolation. Validation with the ADXL345 time-series yielded a mean absolute frequency error of 2.53 Hz ($\sigma = 1.06$ Hz) across five measurement points. Temporal band-pass settings effectively suppressed spectral artifacts, and phase gain influenced magnification clarity. Computational performance averaged around 10 s/frame on CPU, indicating real-time feasibility with GPU acceleration. The proposed PBMM+FFT approach offers a practical, non-invasive alternative for structural vibration monitoring in HVAC systems.

Key Words— ADXL345 accelerometer, Fast Fourier Transform, micro-vibration detection, non-contact vibration sensing, Phase-Based Motion Magnification.

DOI: 10.22441/jte.2026.v17i3.002

1. INTRODUCTION

It is said that the best method to prevent mechanical damage is to detect it early [1]. A 65% increase in energy used by cooling fans and a 10% increase in cooling coil load for a 20% air leak [2] and additional 0.17 scm/m²/year of natural gas consumption for pre-heating outdoor air [3] indicates that a small structural failure in HVAC air ducts, such as cracked joints or metal fatigue due to repeated micro-vibrations, can cause energy efficiency problems in those air ducts. The conventional accelerometer-based vibration monitoring method requires physical contact, necessitates stopping operations, is difficult to apply at remote locations, is easily affected by temperature changes, and adds weight to the workpiece [4, 5]. In general, accelerometers measure acceleration at the point of contact in real time with a bandwidth of up to hundreds of Hz, but only provide localized data at the location where the sensor is installed. Installation requires a stable sensor mount and recalibration of the sensor, as

well as cable routing that adds to the complexity of the field, so an alternative solution is needed.

PBMM offers a non-contact solution by magnifying sub-pixel shifts in video, allowing micro-vibrations to be visualized and measured. The latest study [6] highlights that PBMM is highly promising for monitoring thin structures, including air ducts, as it does not require markers.

The key objective of this research is to develop a prototype of a PBMM-based microstructure vibration detection system for air ducts that can estimate vibration frequency with a mean absolute error of < 2.5 Hz compared to the ADXL345 sensor in a non-contact manner.

2. LITERATURE REVIEW

The PBMM concept is based on the Python implementation [7] of a research project titled Phase-Based Video Motion Processing [8].

A. Fundamentals of PBMM

Each frame of a video is taken away into a steerable pyramid and stored with different levels of details. Steerable Pyramid is a variation of Image Pyramid with linear, multi-level, and multi-orientation image decomposition characteristics. Steerable Pyramid is designed to capture all spatial frequencies and orientations in an image, particularly for analyzing localized signals and movements such as micro-vibrations.

Each frame image from the video $I(x, y)$ is decomposed into several levels and orientations using a steerable pyramid, each level-orientation band yields a complex coefficient:

$$C_{\ell,o}(x, y) = A_{\ell,o}(x, y) \exp(j\phi_{\ell,o}(x, y)) \quad (1)$$

where ℓ is the spatial level in the image pyramid, o is the filter orientation on the steerable pyramid, while $A_{\ell,o}$ is the local amplitude, and $\phi_{\ell,o}$ is the local phase at that level and orientation.

For a video sequence, which is a sequence of frames sorted by time $\{I_t\}$, each frame is decomposed into the complex coefficient $\{I_t\} \rightarrow \{A_{\ell,o}(t), \phi_{\ell,o}(t)\}$, then apply a band-pass temporal filter which consisted of a low-cut and a high-cut filter.

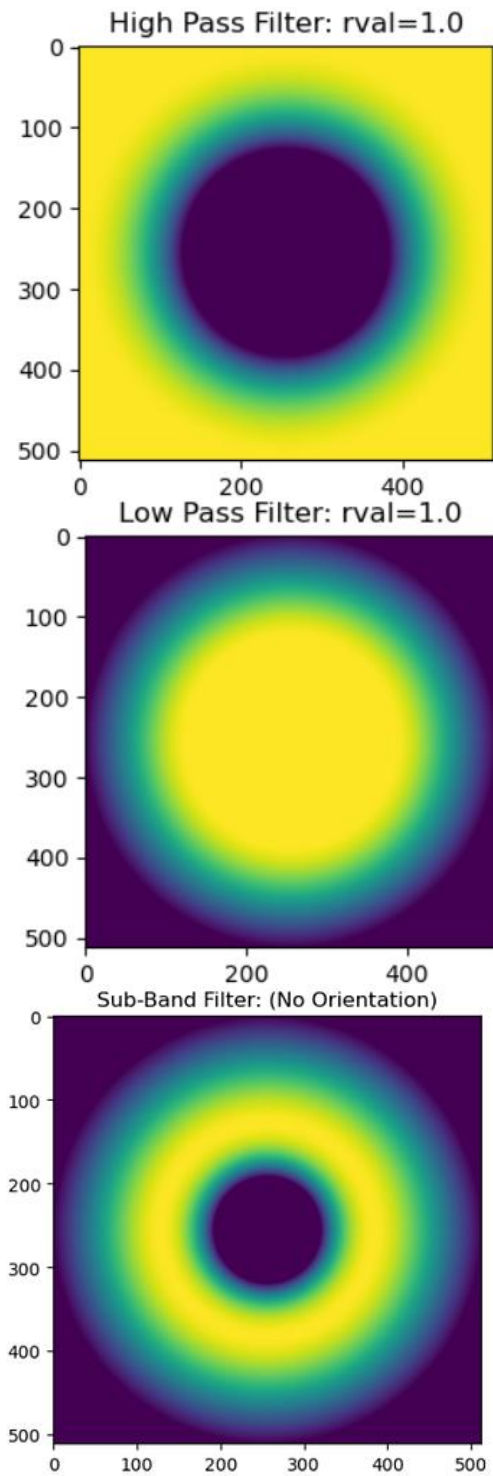


Figure 1. From the top to the bottom: (a) a high pass; and (b) a low pass filter; combined into (c) a sub-band filter without orientation.

After combining the high pass and low pass filters into a sub-band filter, then we determine several orientations as phase filters (steering masks).

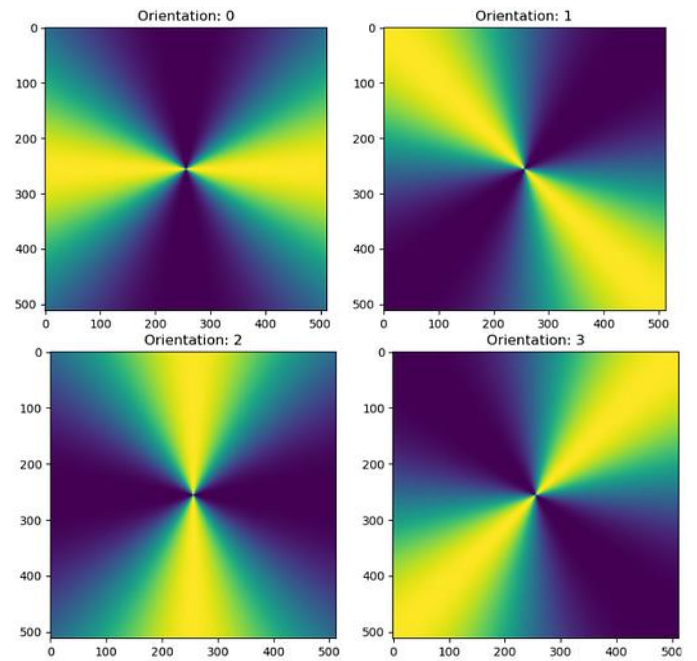


Figure 2. Steering masks with 4 orientations.

These filters serve to steer the direction of the sub-band filters, and combining them with the sub-band filter will create the phase filter for the steerable pyramid.

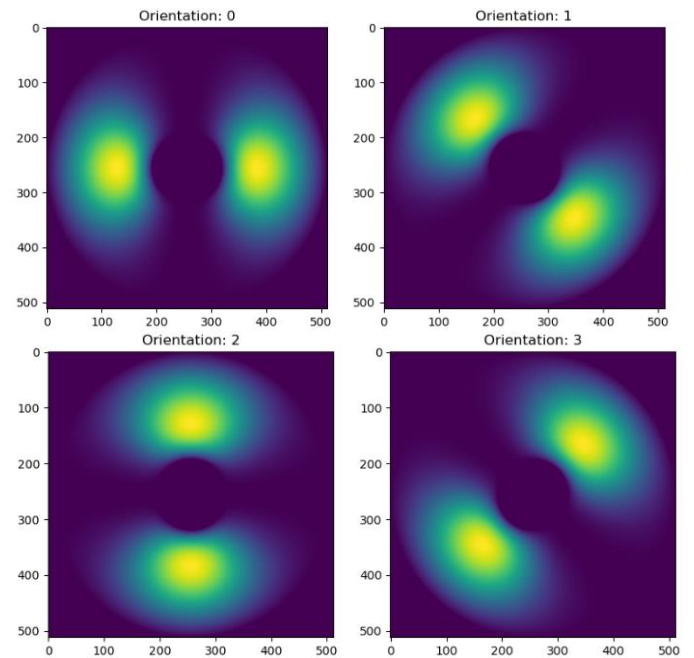


Figure 3. Combined sub-band filter and steering masks results in the phase filter for the steerable pyramid

The result is a filtered phase $\phi_{\ell,o}^f(t)$ where the difference with the initial phase is calculated:

$$\Delta\phi_{\ell,o} = \phi_{\ell,o} - \phi_{\ell,o}^f \quad (2)$$

which then multiplied by magnification factor α to form the magnified phase $\phi_{\ell,o}'$ after adding the initial phase.

$$\phi'_{\ell,o} = \phi_{\ell,o}^f + \alpha \Delta \phi_{\ell,o} \quad (3)$$

The resulting magnified phase from Equation (3) is merged back with the original amplitude to form a new complex coefficient for each image:

$$\hat{C}_{\ell,o}(x, y) = A_{\ell,o}(x, y) \exp(j\phi'_{\ell,o}(x, y)) \quad (4)$$

Perform this on the entire 2D grayscale image to obtain a video that has undergone PBMM processing $\{\hat{I}_t\}$, since the amplitude does not change, the brightness and contrast levels of the video remain the same as the original sample.

B. Processing PBMM's result with FFT

The resulting magnified video of PBMM is then passed through FFT to obtain the frequency of those micro-vibrations. The Fourier transform maps the time domain signal $x(t)$ to the frequency domain $X(f)$ allowing the identification of how strong each frequency component is [9], where:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi f t} dt \quad (5)$$

In this implementation, because the source signal itself is a discrete a derivative algorithm of the Discrete Fourier Transform (DFT), namely the Fast Fourier Transform (FFT), is used which produces complex values $X[k]$ for each frequency bin k , where:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi \frac{k}{N} n} \quad (6)$$

This process shifts the original signal domain from time based into frequency based, but signals recorded by cameras often have offsets (consistent deviations from actual values), drifts (offsets that change in value over time), and truncations (signals that are cut off) depending on the recording duration. Therefore, to reduce said effect can be done by performing detrending and windowing before FFT is performed.

The detrending process is a preprocessing step in which the constant (DC) component is removed from a signal [10] prior to frequency analysis which subtracts the mean, equivalent to fitting a zero-degree (constant) polynomial. In this study, detrending is implemented by subtracting the mean value of the signal from each sample, as expressed in Equation (7):

$$x_{dt}(n) = x(n) - \frac{1}{N} \sum_{m=0}^{N-1} x(m) \quad (7)$$

where $x_{dt}(n)$ is the detrended signal, $x(n)$ is the original signal, and N is the total number of the samples. This approach eliminates the DC offset, a non-zero DC offset manifests as a large peak near 0 Hz in the Fourier spectrum. By centering the signal around zero, the detrending process ensures that the FFT output more accurately represents the true frequency content of the measured data.

When a discrete signal such as from video is captured, there are cut-offs at the beginning and end of the video, namely at time $t = 0$ and at the end of the video. To make the transition of the signal cut-off smoother, Hann window is applied, as expressed by Equation (8):

$$\omega(n) = \frac{1}{2} \left[1 - \cos \left(\frac{2\pi n}{N-1} \right) \right] \quad (8)$$

where $\omega(n)$ is the value of Hann window at discrete point n , n is the sample index starting from 0 through $N - 1$, and N is the total number of samples. The visualization of these steps can be observed in Figure 4-5.

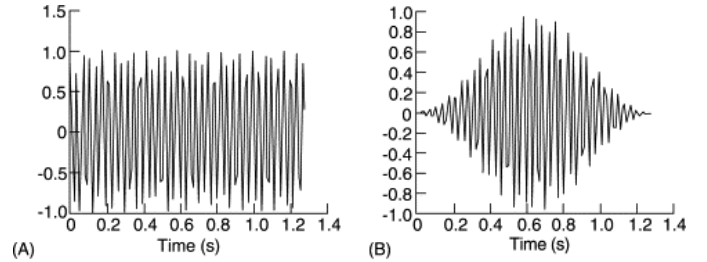


Figure 4. Effect of window on dynamic range: (a) composite signal, rectangular window; (b) composite signal, Hanning window [11].

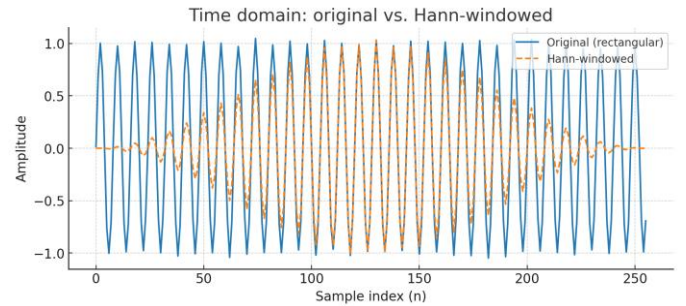


Figure 5. The effect of Hann window vs a rectangular window on the signal at their extremities.

The detrending and windowing method from Equation (7) and Equation (8) then combined with the original FFT of Equation (6):

$$X_{\omega}[k] = \sum_{n=0}^{N-1} (x_{dt}(n) \omega(n)) e^{-j2\pi \frac{k}{N} n} \quad (9)$$

the result of this equation is a frequency spectrum with minimal leakage, enabling accurate detection of frequency peaks.

C. Validation Method

To determine whether the results of PBMM + FFT are reliable, validation and calibration are necessary to ensure that the results truly represent the actual physical conditions. Validation is performed using a conventional accelerometer in the form of a 3-axis Micro Electro Mechanical Sensor (MEMS), ADXL345, controlled by Arduino Uno.



Figure 6. ADXL345 Sensor [12]

The ADXL345 sensor has the capability to communicate with two-wire, half-duplex I2C (Inter-Integrated Circuit) and

four-wire, full-duplex SPI (Serial Peripheral Interface), with an acceleration range of ± 2 g to ± 16 g [12].

The sensor is then attached to the surface of the air duct and offset calibration is performed in zero-g conditions by placing the sensor on a relatively flat surface. Then take a sample (~ 1000 samples), add them up, and find the average to determine the bias as in Equation (10):

$$b_x = \frac{1}{N} \sum_{i=1}^N a_{x,i}, b_y = \frac{1}{N} \sum_{i=1}^N a_{y,i}, b_z = \frac{1}{N} \sum_{i=1}^N a_{z,i} \quad (10)$$

The result of the measurement is a time series of acceleration on the X, Y, and Z axes. We estimated the one-sided power spectral density (PSD) using Welch's method [13], which averages modified periodograms of overlapped, Hann-windowed segments to reduce variance. The measurement signal, which is a long time series $\{x[n]\}_{n=0}^{N-1}$ is then divided into a segment L with length M so that it becomes:

$$x_l[n] = x[n + lD] \quad (11)$$

where l is the segment index, n is the length index of that segment, and $D = M/2$ is called hop-size which ensures an overlap of 50% between each segment.

The resulting segmented data is then subjected to windowing of Equation (8) and transformed into frequency domain using DFT:

$$y_l[n] = \omega(n)x_l[n] \quad (12)$$

$$Y_l[k] = \sum_{n=0}^{M-1} y_l[n] e^{-j2\pi \frac{k}{M} n} \quad (13)$$

Because the window scales do not sample equally, the raw $|Y_l[k]|$ does not have the right power. The solution is to divide by the window power per sample, the window power U is defined by Equation (14):

$$U := \frac{1}{M} \sum_{n=0}^{M-1} \omega^2[n] \quad (14)$$

With f_s is the sampling frequency and with k is the bin index of frequency bin $\Delta f = f_s/M$ we then proceed to convert power per bin to power spectral density (power per Hz), divide by f_s (or equivalently by $\Delta f M$) to get the two-sided periodogram of power spectral density (PSD) $I_l(f_k)$ at segment l and bin k can be expressed as Equation (15):

$$I_l(f_k) = \frac{1}{f_s M U} |Y_l[k]|^2 \quad (15)$$

Because this signal is a real ($\mathbb{R}$) number we can collapse negative frequencies by doubling the positive-frequency bins (except DC and, if present, Nyquist), this is defined as:

$$a_k \begin{cases} 1, & k = 0 \text{ (DC)} \\ 1, & k = \frac{M}{2}, \text{ if } M \text{ is even (Nyquist)} \\ 2, & \text{otherwise} \end{cases}$$

then the one-sided periodogram is defined as:

$$I_l^{(1)}(f_k) = a_k \frac{1}{f_s M U} |Y_l[k]|^2 \quad (16)$$

hence the Welch's PSD estimate is expressed in Equation (17) as:

$$\hat{S}_{xx}(f_k) = \frac{1}{K} \sum_{l=0}^{K-1} I_l^{(1)}(f_k) \quad (17)$$

The result is a PSD with the maximum frequency reading from the ADXL345 accelerometer sensor in g^2/Hz .

3. EXPERIMENTAL SETUP

We recorded, at 5 sampling point, frame-synchronized 4K/60 fps video of an HVAC duct and tri-axial acceleration with an ADXL345. The video was processed by phase-based motion magnification (steerable complex pyramid; temporal band-pass 5–25 Hz; phase gain factor $\alpha=20$). The trace was mean-removed, Hann-windowed, and transformed using a one-sided FFT to estimate the dominant frequency.

Accelerometer data were detrended, band-pass filtered (5–25 Hz), and analyzed with Welch's method (Hann window, segment length M , 50 % overlap, window-power normalization $U = \frac{1}{M} \sum_{n=0}^{M-1} \omega^2[n]$). One-sided PSD result is reported in g^2/Hz . Peaks were identified with a minimum separation of 0.5 Hz to avoid adjacent-bin duplicates. Frequencies from video and accelerometer were compared using mean absolute error (MAE).

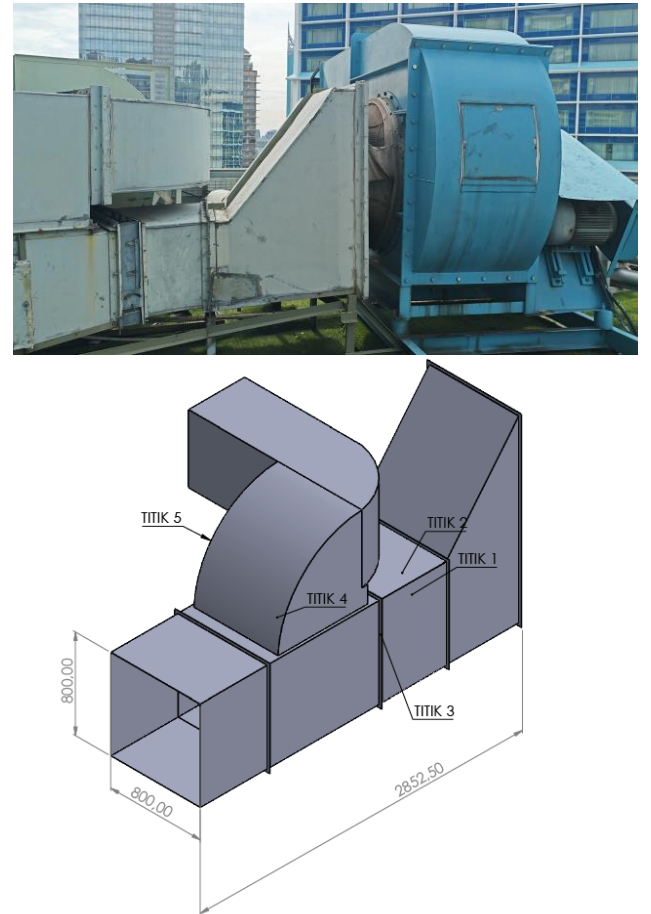


Figure 7. The target of the experiment, from the top is (a) the entirety of the duct; (b) the details and dimension of sampling point

Key parameters:

- FPS, $f_s = 60$, overlap, $D = M/2$, PBMM: levels=5, $\alpha = 20$, band=5–25 Hz.

- Software/versions: Python 2.7.18, NumPy 1.13.3, SciPy 1.2.1, OpenCV 3.2.0
- Hardware: Camera: POCO F6; Laptop for PBMM+FFT: Lenovo Ideapad 3-14ITL6; ADXL 345 range: 2-16g, Mounting Orientation: Point 1 = -X, Point 2 = -Z, Point 3 = -X, Point 4 = -X, Point 5 = -X.

The experiment is conducted in adequate daylight to ensure minimal artifact in the magnification process.

4. EXPERIMENT RESULTS

The experiment was performed across five different points to obtain measurement data that was representative of the actual conditions. At each point, the experiment was performed for 30 seconds.

Table 1. Experimental Data Results

<i>Point</i>	<i>Number of Frames</i>	<i>Number of ADXL345 Data</i>
1	1802	1323
2	1802	1323
3	1829	1323
4	1801	1323
5	1790	1323

After applying the PBMM+FFT pipeline to the video footage, the result is a frequency spectrum as shown in Figure 8.

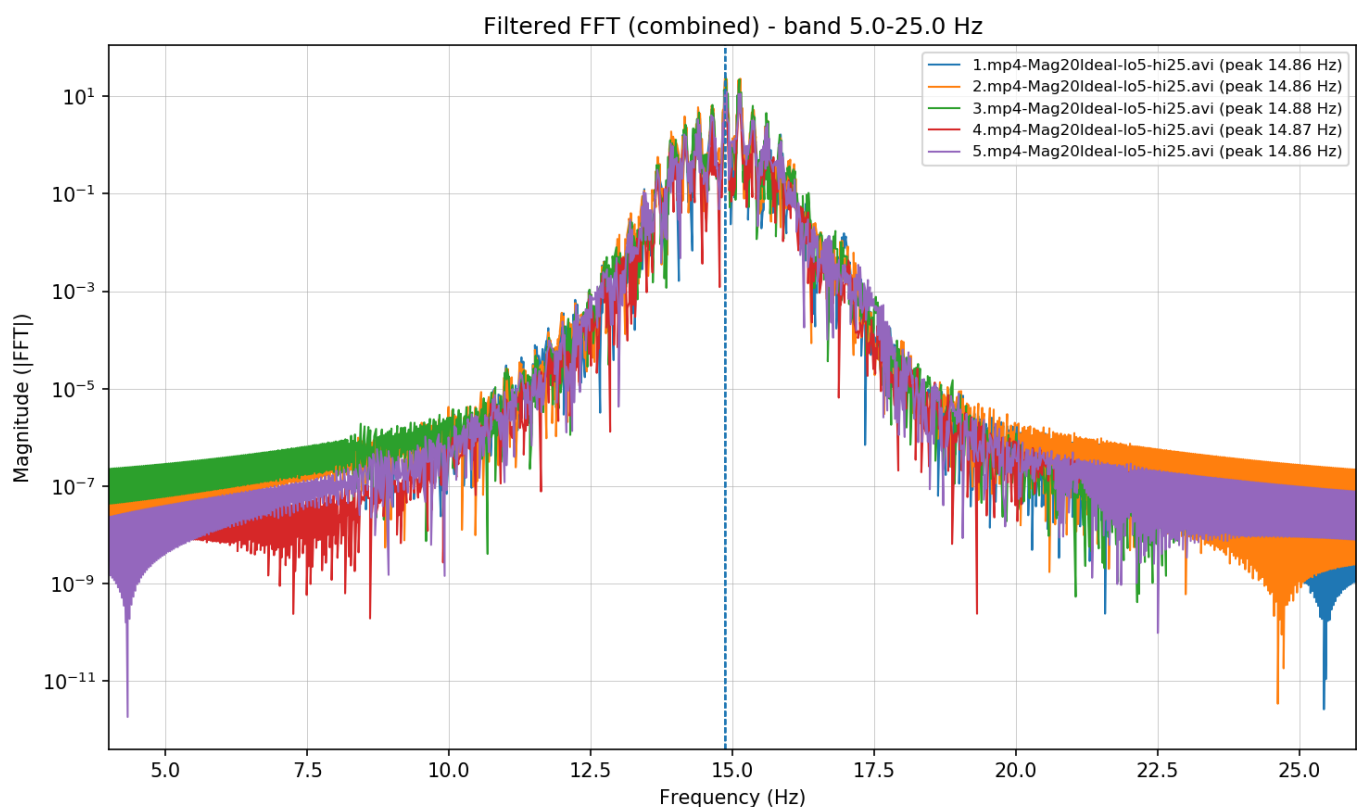


Figure 8. PBMM+FFT Results at Points 1-5, Filtered FFT shows that when diff-trace 5-25 Hz frequency has been done, the peak frequency is 14.86 Hz.

In parallel, data from ADXL345 is read using the time domain peak finding method, then processed using the Welsch method.

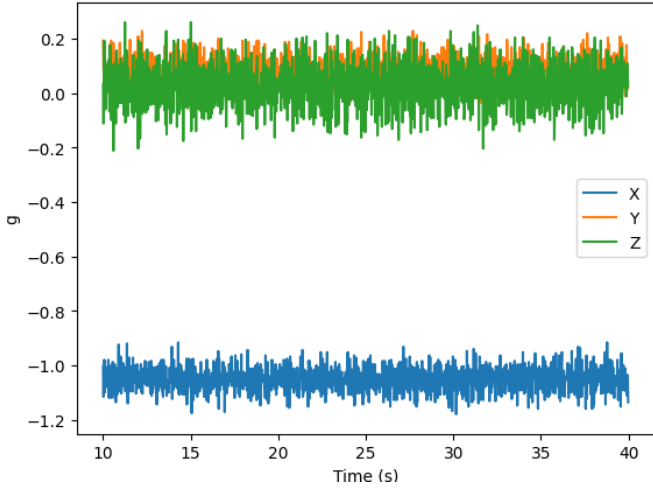


Figure 9. Raw data of ADXL345 at Point 1.

Then after processing the raw data at Point 1, we got the Root Mean Squared (RMS) acceleration in Figure 10:

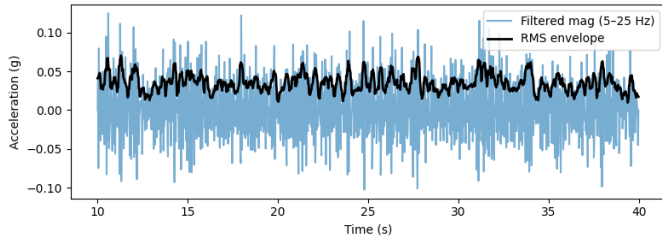


Figure 10. Acceleration spectrum with RMS envelope values at a band-pass filter of 5-25 Hz.

The result is a PSD of Point 1:

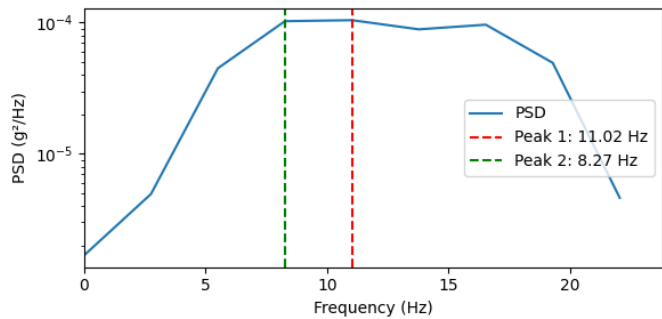


Figure 11. The final PSD results from the ADXL345 data log show the highest peaks at 11.02Hz and 8.27Hz.

Accuracy is measured by the Mean Absolute Error (MAE) between the PBMM+FFT result frequency and the ADXL345 reference frequency. For the i -th measurement, the absolute error is defined as:

$$\varepsilon_i = |f_{PBMM+FFT}^i - f_{ADXL345}^i| \quad (18)$$

MAE is then calculated as:

$$MAE_f = \frac{1}{N} \sum_{i=1}^N \varepsilon_i \quad (19)$$

where N is the number of data measurement pairs of PBMM+FFT and ADXL345.

Table 2. Frequency Estimation and Absolute Error

Point	$f_{PBMM+FFT}^i$	$f_{ADXL345}^i$	ε_i
1	14.86 Hz	11.02 Hz	3.84 Hz
2	14.86 Hz	16.54 Hz	1.68 Hz
3	14.88 Hz	16.54 Hz	1.66 Hz
4	14.87 Hz	16.53 Hz	1.66 Hz
5	14.86 Hz	11.02 Hz	3.84 Hz

The table shows the frequencies measured by PBMM+FFT compared to the frequencies measured by ADXL345 and the MAE_f is calculated at 2.53 Hz.

To determine the performance of the device used in this experiment, the time required to perform the PBMM+FFT process was calculated. The measurement was taken by subtracting the end time from the start time in HH:MM:SS format.

Table 3. Processing Time for Each Measurement Point.

Point	Time (HH:MM:SS)	Number of Frames	Average Time / Frame (s)
1	04:50:40	1802	9.678
2	04:58:31	1802	9.939
3	05:23:35	1829	10.615
4	04:47:06	1801	9.564
5	03:46:15	1790	7.583

The table shows that the average time/frame is around 10 seconds. For real-time detection, this can be accelerated by using devices with Graphic Processing Unit (GPU).

5. CONCLUSIONS

This study demonstrates that a non-contact approach based on Phase-Based Motion Magnification (PBMM) followed by FFT analysis is capable of reliably detecting micro-vibration frequencies in air ducts up to <2.5 Hz of MAE. The proposed pipeline detrending (mean-removal), Hann window, and Welch's method for ADXL345 data—produces a spectrum with low leakage and consistent frequency peaks in the 5–25 Hz range. The frequency estimates from the video align with the ADXL345 PSD peaks at five test points, thereby validating the physical validity of the PBMM+FFT results.

Compared to contact-based methods, this approach is more practical for field monitoring because it does not require the installation of sensors on objects and does not interfere with operations. Computationally, the current implementation runs well on CPUs but is not yet real-time; thus, applying GPU

acceleration and parameter optimization have the potential to enable real-time detection.

Some limitations of this system include:

1. The selection of frequency peaks at several points resulting from the ADXL345 shows that there is more than one peak of approximately the same magnitude, which may cause the PBMM value to tend closer to the average between these peaks.
2. This method is limited by the Nyquist frequency, which is half the sampling frequency. Therefore, with FPS=60, the maximum frequency that can be measured is 30 Hz. This study focuses on vibrations in the 5-25 Hz range.
3. This method does not yet have the ability to estimate the amplitude of the vibrations occurring, as this study is more focused on estimating the vibration frequency.

Further development could include the following:

1. Utilizing a Graphical Processing Unit (GPU) to perform real-time detection.
2. Adding an amplitude estimation method by calibrating displacement per pixel.

ACKNOWLEDGEMENT

The authors thank PT Papaja Maju Mandiri for providing the facilities, site access, and equipment that made this research possible. We are also grateful to R. S. F. for their supportive measures and guidance throughout the preparation of this manuscript. Any remaining errors are the authors' own.

DECLARATIONS

A. Data Availability

An anonymized, view-only link to the derived data (ADXL345 logs, ROI traces) has been provided to the editor for peer review. After acceptance, the dataset will be deposited in a public repository. Raw HVAC videos are available under restricted access as noted above.

B. Code Availability

Source code to reproduce the analysis is available at <https://github.com/deyoyow/Vreq>.

REFERENCES

- [1] M. Śmieja, J. Mamala, K. Prażnowski, T. Ciepliński, and Ł. Szumilas, "Motion Magnification of Vibration Image in Estimation of Technical Object Condition-Review," *Sensors*, vol. 21, no. 19, p. 6572, Sep. 2021, doi: 10.3390/s21196572.
- [2] W. Cai, L. Huang, and Z. Zou, "Actively-exploring thermography-enabled autonomous robotic system for detecting and registering HVAC thermal leaks," *Automation in Construction*, vol. 152, p. 104901, Aug. 2023, doi: 10.1016/j.autcon.2023.104901.
- [3] S. Krishnamoorthy and M. Modera, "Impacts of duct leakage on central outdoor-air conditioning for commercial-building VAV systems," *Energy and Buildings*, vol. 119, pp. 340–351, May 2016, doi: 10.1016/j.enbuild.2016.03.057.
- [4] I. Yildiz, "Alternative to laser vibrometers: a non-contact vibration measurement based on enhanced laser speckle pattern and stroboscopic sampling," *Metrology and Measurement Systems*, pp. 767–767, Dec. 2024, doi: 10.24425/mms.2024.152050.
- [5] J. Guo et al., "Non-contact vibration sensor using deep learning and image processing," *Measurement*, vol. 183, p. 109823, Oct. 2021, doi: 10.1016/j.measurement.2021.109823.
- [6] M. Civera and C. Surace, "Non-Destructive Techniques for the Condition and Structural Health Monitoring of Wind Turbines: A Literature Review of the Last 20 Years," *Sensors*, vol. 22, no. 4, p. 1627, Feb. 2022, doi: 10.3390/s22041627.
- [7] J. van Gemert and J. Kooij. "pbMoMa," (2015). Accessed: May 6, 2025. [Online]. Available: <https://github.com/jvngemert/pbMoMa>.
- [8] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, "Phase-based video motion processing," *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 1–10, Jul. 2013, doi: 10.1145/2461912.2461966.
- [9] M.-A. Delsuc and P. O'Connor, "The Fourier transform in analytical science," *Nature Reviews Methods Primers*, vol. 4, no. 1, p. 49, Jul. 2024, doi: 10.1038/s43586-024-00326-2.
- [10] S. Dermisi, "The Impact of Terrorism Fears on Downtown Real Estate Chicago Office Market Cycles," *Journal of Real Estate Portfolio Management*, vol. 13, no. 1, pp. 57–73, Jan. 2007, doi: 10.1080/10835547.2007.12089759.
- [11] S. G. Braun, D. J. Ewins, and S. S. Rao, *Encyclopedia of vibration*. Academic Press, 2002. Accessed: Aug. 13, 2025. [Online]. Available: <https://cir.nii.ac.jp/crid/1970586434828142628.bib?lang=en>
- [12] Analog Devices, MA, USA. *ADXL345 (Rev. G)*. (2022). Accessed: June 6, 2025. [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/adxl345.pdf>
- [13] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.