Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cutting Condition Monitoring in Milling within Enclosed Environments Using a Microphone Array and Random Forest

Joel Martins Crichigno Filho

crichigno@joinville.udesc.br

State University of Santa Catarina

Research Article

Keywords: Process monitoring, End milling, Microphone array, Machine learning

Posted Date: July 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10361315/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Cutting Condition Monitoring in Milling within Enclosed Environments Using a Microphone Array and Random Forest

First Joel Martins Crichigno Filho

Department of Mechanical Engineering, Santa Catarina State University, Rua Paulo Malschitzki, 200, Joinville, 89.219-710 , Santa Catarina, Brazil.

Contributing authors: joel.crichigno@udesc.br;

Abstract

In modern industry, monitoring systems have become essential for detecting process deviations and enabling preventive adjustments before defective parts are manufactured, thereby reducing production costs. Acoustic emission is an attractive source of information for monitoring machining processes. However, acoustic signals can be contaminated by background noise and, within the enclosed environment of a machine tool, by reverberation caused by the reflective metallic surfaces of the enclosure. This study evaluates the feasibility of monitoring cutting conditions in end milling using a microphone array positioned inside this highly reverberant environment, which exhibits acoustic behavior close to that of a diffuse field. Experiments were conducted on a machining center during end-milling operations using a four-microphone array mounted close to the spindle. The results showed that the enclosure produced a diffuse acoustic field without completely masking the acoustic signature of the cutting process. To mitigate the influence of the diffuse sound field, the proposed signal-processing strategy combined octave-band decomposition with Wiener filtering. Its effectiveness was assessed by comparing two feature extraction strategies. The first employed conventional broadband time- and frequency-domain features commonly reported in the literature, whereas the second extracted root-mean-square (RMS) descriptors from octave-band signals after Wiener filtering. The results demonstrated that the conventional broadband descriptors were highly susceptible to the homogenizing effects of reverberation. In contrast, the octave-band RMS descriptors preserved significantly greater sensitivity to variations in the machining parameters, demonstrating that frequency-selective analysis retains process-related information that is largely lost when global broadband metrics are

employed. The Random Forest algorithm was subsequently used to further validate the proposed feature extraction strategies by classifying cutting conditions across six independent experimental trials. Whereas the conventional broadband strategy achieved an average classification accuracy of 64.7% with a standard deviation of 26.5%, the proposed frequency-selective strategy achieved an average accuracy of 95.5% with a standard deviation of only 8.1%. These findings indicate that the proposed frequency-selective methodology provides substantially greater robustness against reverberation and background noise typically encountered in enclosed industrial machining environments.

Keywords: Process monitoring, End milling, Microphone array, Machine learning

1 Introduction

Process instabilities and deviations during machining can lead to out-of-specification parts, accelerate tool wear, and cause unplanned production downtimes, resulting in significant financial losses. In this context, real-time process monitoring has become a strategy for preserving the productivity, reliability, and competitiveness of modern manufacturing systems.

Machining monitoring methods are classified into direct and indirect approaches. While direct methods require process interruption to measure the variable of interest with high precision, indirect methods are based on establishing a correlation between the variable of interest and sensor signals.

Indirect methods have been widely adopted due to their lower cost, ease of implementation, and the possibility of continuous acquisition without interfering with the operation. Among the variables used, spindle motor current, structural vibration, acoustic emission, and audible sound signals are noteworthy [1].

Audible acoustic signals generated during machining are traditionally exploited by experienced operators to assess process conditions and diagnose anomalies such as tool wear, breakage, and *chatter*. Microphone-based monitoring offers important advantages: non-intrusive nature, positioning flexibility, wide frequency response, and low cost.

Microphones capture acoustic emissions, the spectral content of which is influenced by the dynamic behavior of the tool–workpiece–machine system during cutting. Consequently, acoustic sensing has been widely employed in turning, milling, and drilling operations. Recent literature describes a wide variety of experimental configurations (microphone types, positioning, and acquisition strategies) focused primarily on *Tool Condition Monitoring* (TCM) and *chatter* detection [2].

Milling is an intermittent cutting process in which successive tooth engagements produce periodic force excitations that induce structural vibrations in the tool–workpiece–machine system. These vibrations generate pressure fluctuations in the surrounding air, resulting in acoustic radiation whose spectral characteristics reflect both the dynamic response of the machining system and cutting force excitation.

Consequently, in-process audible sound measurements have recently been employed to identify the natural frequencies of a workpiece during material removal [3].

Despite these advantages, microphone-based monitoring in industrial environments is significantly affected by background noise. In addition to the sound generated by the cutting process, microphones capture acoustic contributions from neighboring machine tools, electric motors, ventilation systems, coolant pumps, and operator conversations [4]. Consequently, the sound pressure levels on the shop floor may exceed 90 dB over a broad frequency range, reducing the signal-to-noise ratio and hindering the extraction of information directly related to the machining process [5].

In addition to background noise, the interior of enclosed machine tools presents further acoustic complexity owing to their highly reflective metallic surfaces and compact geometry. Under these conditions, the direct sound radiated from the cutting zone is accompanied by multiple reflections from the enclosure walls, giving rise to a reverberant sound field. As the number of reflections increases, the acoustic energy becomes progressively more homogeneous and isotropic throughout the enclosure, allowing the reverberant field to be approximated as a diffuse field, which is widely adopted as a statistical model for enclosed acoustic environments [6, 7].

The coexistence of direct and reverberant sounds alters the spectral and spatial characteristics of the measured acoustic signal, making it more difficult to isolate information directly associated with the cutting process. To better understand these effects, recent studies have modeled sound propagation inside enclosed machine tools to optimize microphone placement for tool wear monitoring [8].

The placement of microphones in enclosed machine tools involves balancing the acoustic performance with sensor protection. Positioning the microphones close to the cutting zone generally improves the signal-to-noise ratio, but exposes the sensors to chips, coolant, and geometric constraints. Conversely, microphones mounted in protected regions or outside the enclosure are less susceptible to mechanical damage, although the recorded signals contain stronger contributions from reverberation and industrial background noise [9, 10]. In particular, external mounting increases the influence of neighboring machines, whereas internal mounting benefits from the enclosure walls, which partially shield external acoustic sources.

The superposition of direct and reverberant sounds alters the spectral content of the measured signal, making acoustic descriptors dependent not only on the cutting process but also on the propagation characteristics of the enclosure. Consequently, the statistical and spectral features commonly employed for machining monitoring may exhibit reduced sensitivity to variations in cutting conditions. Previous studies have shown that reverberation, structural resonances, and acoustic reflections degrade spectral analysis and reduce the reliability of microphone-based chatter detection [2, 11].

Several signal processing strategies have been proposed to improve the robustness of acoustic monitoring in industrial environments. For example, Sio-Sever et al. [12] employed a four-microphone array mounted outside a machine enclosure to monitor milling operations. To reduce the effects of industrial noise and reverberation,

the authors combined multichannel signal processing with frequency-domain feature extraction, demonstrating improved accuracy in machining parameter estimation under noisy conditions.

Sallese et al. [13] utilized a parametric equalization inverse filter to reconstruct the amplitude of cutting noise in *chatter* detection applications. However, a large portion of the literature still assumes near-free-field conditions, with few studies specifically investigating sound propagation in highly reverberant enclosed environments [14].

Most studies on machining process monitoring employ *Machine Learning* algorithms to predict process conditions from features extracted from the monitored signals. Among the available classifiers, Random Forest has consistently demonstrated superior performance compared with other methods widely used in TCM, such as Support Vector Machines (SVM) and artificial neural networks, owing to its ability to model nonlinear relationships and handle noisy datasets [15]. In a recent comparative study on tool wear monitoring in micromachining, Random Forest achieved the highest classification accuracy among the evaluated methods, outperforming XGBoost, SVM with an RBF kernel, and a multilayer perceptron (MLP) neural network [15]. Similar findings were reported by [16], where Random Forest outperformed classifiers such as FURIA and Hoeffding Tree in tool condition classification, further demonstrating its robustness to the variability and noise typically present in machining signals.

In parallel with advances in Machine Learning, the use of microphones for machining process monitoring has attracted increasing attention because of their low cost, non-contact operation, and ease of installation. However, most existing studies implicitly assume that the recorded acoustic signals accurately represent the cutting process, while little attention has been devoted to characterizing the audible acoustic field inside enclosed machine tools or assessing the influence of reverberation on the acoustic descriptors used for process monitoring. This gap is particularly important because reverberation may significantly alter the characteristics of the measured signals, potentially degrading the performance of feature extraction methods and, consequently, the accuracy of machine learning classifiers.

Therefore, this study evaluatd the feasibility of monitoring milling cutting conditions using a microphone array operating in a reverberant environment approaching a diffuse sound field. Specifically, it investigates (i) the effects of the machine enclosure on the captured acoustic signals, (ii) the influence of cutting parameters on features extracted in both the time and frequency domains, and (iii) the robustness of two feature extraction strategies combined with Random Forest classifiers against the noise and reverberation typically encountered in industrial machining environments.

2 Methodology

The proposed methodology consisted of acquiring audible acoustic signals during milling, extracting descriptors using two independent feature extraction strategies, and evaluating their ability to discriminate machining conditions using a Random Forest classifier.

2.1 Experimental Setup

Experiments were carried out on a Romi D600 vertical machining center. The machine tool features a maximum spindle speed of 10,000 rpm and a power output of 15 kW.

The workpieces were machined using a 32 mm diameter end mill (model R390–032B32–11, Sandvik Coromant) equipped with two carbide inserts (R390–11T308M–PM1010, Sandvik Coromant) featuring a TiAlN coating deposited via the PVD (*Physical Vapor Deposition*) process. The tool overhang length was set to 64.5 mm to maximize system stiffness and minimize the occurrence of regenerative vibrations. Machining was performed under dry conditions.

The workpiece consisted of a cylindrical specimen with a diameter of 150 mm, clamped to the machine table using a three-jaw chuck. The experiments were performed by face milling one of the flat faces of the cylinder using a linear tool path under a down-milling configuration. Signal acquisition was restricted to the central region of the cut surface; lateral sections of approximately 30 mm on both sides were excluded to avoid the influence of tool entry and exit transient phenomena. The material was SAE 8620, a low-carbon alloy steel widely used in machining studies, machined in its as-received condition.

To investigate the sensitivity of acoustics-derived features with respect to variations in the machining process parameters, a Design of Experiments (DOE) methodology was employed. A 2^{4-1} fractional factorial design was adopted, considering four machining factors at two levels each: cutting speed (v_c), feed (f_z), axial depth of cut (a_p), and radial depth of cut (a_e). The specific levels for each parameter are presented in Table 1. This experimental design resulted in 8 distinct parameter combinations, which define the operational conditions (classes) to be subsequently detected and classified by the Random Forest model.

Table 1: Machining conditions

Class	v_c (m/min)	f_z (mm/rot)	a_e (mm)	a_p (mm)
1	100	0.1	5	1
2	200	0.2	5	1
3	100	0.2	20	1
4	200	0.1	20	1
5	100	0.2	5	2
6	200	0.1	5	2
7	100	0.2	20	2
8	200	0.2	20	2

For acoustic signal acquisition, an array consisting of four MEMS microphones arranged in a square configuration, similar to the ReSpeaker 4-Mic Array *HAT* [17], was utilized. This array employs four digital MEMS microphones with an inter-microphone distance of approximately 57 mm. The system was positioned alongside the spindle, with its face oriented parallel to the axis of rotation. The array was

positioned that the system maintains a constant distance from the tool–even during its movement–while also protecting it from chip ejection and cutting fluid flow. As illustrated in Fig. 1, the distance between the center of the microphone array and the cutting tool was approximately 350 mm, in both parallel and perpendicular orientations relative to the spindle.

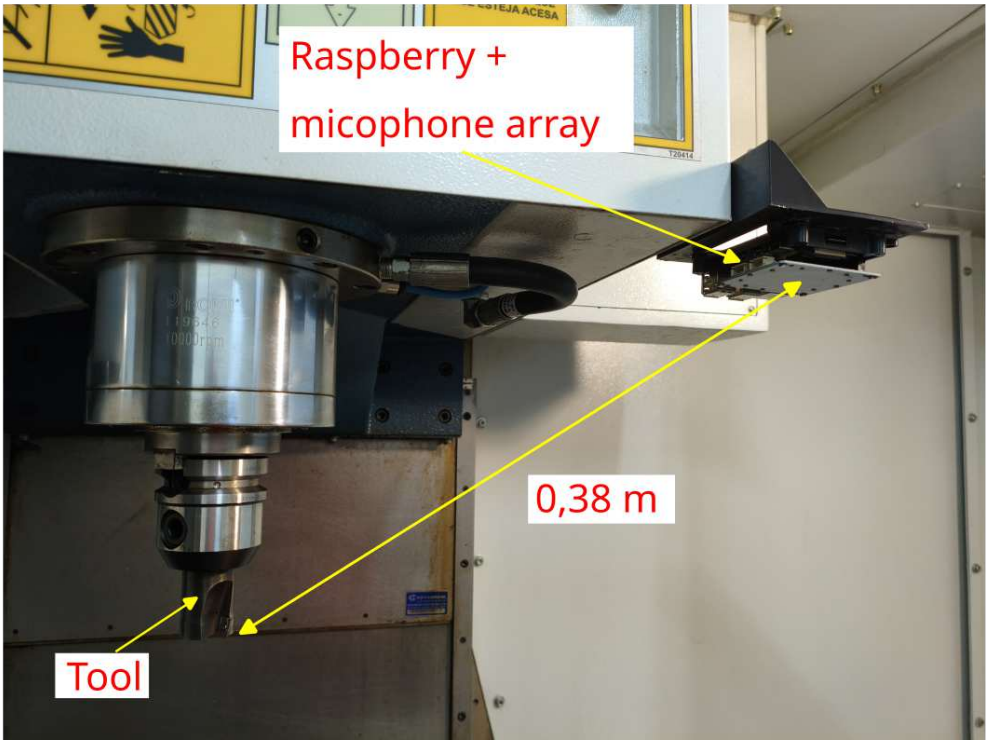


Fig. 1: Experimental setup.

Since the analysis was limited to octave bands up to 8 kHz, a sampling frequency of 16 kHz satisfied the Nyquist criterion. The signals were acquired and stored on a 32 GB microSD card. Remote control of the acquisition initiation and file transfers was managed by a PC via a Wi-Fi connection. Subsequent signal processing was implemented in Python, encompassing segmentation, filtering, feature extraction, and classification stages.

The acoustic signal corresponding to each experimental run was automatically extracted from the raw data files using a custom Python routine. To ensure objective and reproducible segmentation, the boundaries defining the start and end of the active cutting stage were determined using an energy-based method computed via the Root Mean Square (RMS) envelope of the entire signal.

2.2 Acoustic and Dynamic Characterization of the Environment

2.2.1 Acoustic Reverberation Tests

To characterize the acoustic response of the machine tool enclosure, the reverberation time (RT_{60}) was measured. This parameter is defined as the time interval required for the sound energy to decay by 60 dB after the source is interrupted, quantifying the persistence of sound reflections within the environment.

In this work, a firecracker was used as an impulsive excitation source due to its practicality in confined spaces. According to Papadakis and Stavroulakis [18], who evaluated fifteen alternative sound sources to the traditional dodecahedron loudspeaker for acoustic characterization, explosive sources stand out as a viable impulsive choice for impulse response measurements.

The RT_{60} values are estimated by calculating the slope of the Schroeder curve, which is a plot of the energy (squared values) of the impulse response that is backwards integrated (summed starting from the end and moving backwards). The vertical axis of the plot is in dB. It is not always possible to measure a full 60 dB of reverberant decay, so smaller ranges can be used instead. In this work, the evaluation range T_{30} is used. This range is measured between two points on the Schroeder curve: the first point lies where the signal has decreased by 5 dB relative to the initial level, and the endpoint is located where the signal has decreased by 35 dB (providing a 30 dB evaluation dynamic range).

2.2.2 Spatial Coherence and Diffuse Field Validation

To validate the diffuse field hypothesis and verify whether the enclosure environment corrupted the local wavefronts, the spatial coherence function (γ^2) between the microphones in the array was analyzed. In a perfectly diffuse reverberant field, the theoretical coherence between two points separated by a distance d follows the Rayleigh spatial model, defined by the *sinc* function according to Equation 1:

$$\gamma^2_{\text{Rayleigh}}(f, d) = \left| \text{sinc}\left(\frac{2\pi f d}{c}\right) \right|^2 \tag{1}$$

where c is the speed of sound in air and f is the frequency.

To quantify the agreement between the measured spatial coherence and the coherence predicted by the Rayleigh diffuse field model, a similarity index based on the mean squared error (MSE) between the curves was defined. Initially, the MSE between the measured and theoretical coherence values across the analyzed frequency bands was computed. Subsequently, the similarity index (S) was obtained by:

$$S = 1 - \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\gamma^2_{\text{med},i} - \gamma^2_{\text{teo},i}\right)^2} \tag{2}$$

where N is the number of frequency bands considered, γ^2_{med} is the measured coherence, and γ^2_{teo} is the coherence predicted by the Rayleigh model. The index was bounded

between 0 and 1, with values close to 1 indicating high agreement between the model and the experimental results.

2.2.3 Acoustic Response of the Machine Enclosure

The acoustic frequency spectrum of the machine enclosure was obtained via the Fast Fourier Transform (FFT) of the measured time-domain acoustic impulse response, utilizing a firecracker as the impulsive acoustic excitation source. The experimentally obtained spectrum was subsequently subjected to a spectral smoothing process using 1/3-octave filters.

The utilization of this filtering type applies a logarithmic-scale spectral resolution, characterized by narrow bands at low frequencies and progressively wider bands at higher frequencies. This approach preserves the primary acoustic resonance modes of the enclosure while smoothing out the rapid spectral oscillations associated with the high density of peaks and valleys in the mid- and high-frequency regions.

2.2.4 Cutting Tool FRF

Investigating the dynamic characteristics and dominant natural frequencies of the cutting tool via its Frequency Response Function (FRF) is essential to establish correlations with the sound measured during machining. To this end, the tool was subjected to impact tests using an instrumented hammer equipped with a force transducer (model 8206–003, Brüel & Kjær). An accelerometer (model 35C22, PCB Piezotronics) was attached to the free end of the cutting tool to capture the vibratory response. Simultaneous acquisition of the force, vibration, and acoustic signals was performed using a National Instruments USB–4432 module. Data processing and analysis were executed using CutPro software developed by MAL Inc. [19].

2.3 Feature Extraction Strategies

Once the acoustic data were acquired, the signals were pre-processed through amplitude normalization relative to the digital full-scale range of the acquisition format. The recorded samples were divided by the maximum representable digital value according to the signal bit depth; for the 16-bit signed acquisition system, the samples were normalized by the corresponding full-scale value of 2^{15}. This procedure scales the signals to a dimensionless range independent of the absolute microphone sensitivity and acquisition gain.

The RMS value of the acoustic signal was calculated to determine the boundaries corresponding to the beginning and end of the cutting process. The maximum average RMS value obtained among the eight DOE experiments was then adopted as a reference value for an additional normalization stage. This normalization procedure was applied consistently to both the training dataset (Trial 1) and the validation datasets (Trials 2–6). The purpose of this additional normalization was to improve the performance of the machine learning algorithms by reducing the influence of numerical features measured on different scales or acquired under varying sensor conditions, as discussed by Li and Lin [20].

Consequently, the acoustic signal was segmented, and feature extraction was performed according to the statistical and spectral descriptors described below. These features were employed during the training and validation stages of the classification models based on the Random Forest algorithm.

2.3.1 Time and Frequency Domain Analysis

Acoustic signals are commonly analyzed in the time, frequency, and time–frequency domains. Features extracted from these domains are widely employed for training classification models, as discussed in [21]. In the present study, analyses in the time and frequency domains were adopted. Time-domain descriptors were employed to characterize the statistical and energetic properties of the acoustic signals, while spectral descriptors were used to represent the distribution of acoustic energy over the frequency spectrum.

Temporal features such as root mean square, mean, and standard deviation are associated with signal energy and amplitude variability. Specifically, the mean of absolute values acts as a central tendency indicator, while the root mean square reflects the effective signal energy. Higher-order statistical moments, including skewness and kurtosis, provide information regarding signal asymmetry and impulsiveness. Additionally, amplitude-based indicators including peak amplitude, the crest factor, shape factor, impulse factor, and clearance factor are commonly employed in condition monitoring applications because they are sensitive to transient and impulsive events embedded in noisy environments.

Spectral features were included to characterize the frequency distribution of the acoustic energy. Parameters such as spectral centroid and spectral spread provide information about dominant frequency regions and effective bandwidth, whereas spectral entropy quantifies the degree of spectral randomness associated with machining stability conditions. Furthermore, the spectral root mean square was computed to evaluate the energy content within the frequency domain, while spectral skewness and spectral kurtosis were utilized to monitor asymmetry and the presence of dominant tonal or impulsive components in the spectrum. The temporal and spectral features extracted in this work are summarized in Table 2.

2.3.2 Octave-Band Decomposition and Wiener Filtering

The utilization of sub-band processing and Wiener filtering for noise reduction in reverberant environments and diffuse fields has been explored by previous works. In Hussain et al. [22] and Abutalebi et al. [23], adaptive sub-band filters were combined with Wiener filtering to minimize residual noise components. Subsequently, Li and Akagi [24] explicitly modeled the frequency dependence of spatial coherence in diffuse fields, utilizing different post-filtering strategies for low and high frequencies.

Since, in this work, the microphone array is positioned in a region that transitions between free-field and reverberant-field conditions, the influence of diffuse acoustic components depends strongly on frequency. Consequently, the proposed strategy combines octave-band decomposition and Wiener filtering. The acoustic signals were decomposed into octave bands with center frequencies ranging from 125 Hz to 8 kHz,

Table 2: Acoustic signal features employed in this study.

Feature	Description	Abbreviation
Time-Domain		
Root Mean Square	Effective signal energy	RMS_T
Mean	Mean of absolute values	ME_T
Standard Deviation	Amplitude variability	SD_T
Skewness	Distribution asymmetry	SK_T
Kurtosis	Signal impulsiveness	KU_T
Crest Factor	Peak-to-RMS ratio	CF_T
Shape Factor	RMS-to-mean absolute value ratio	SF_T
Impulse Factor	Impulsiveness indicator	IF_T
Clearance Factor	Sensitive to incipient faults	CLF_T
Peak	Peak signal amplitude	PK_T
Frequency-Domain		
Spectral Centroid	Mean dominant frequency of the spectrum	SC_F
Spectral Spread	Effective signal bandwidth	SS_F
Spectral Skewness	Energy distribution asymmetry in frequency	SK_F
Spectral Kurtosis	Presence of tonal or impulsive components	KU_F
Spectral RMS	Spectral energy content	RMS_F
Spectral Entropy	Degree of spectral randomness	ENT_F

where the lower and upper cutoff frequencies of each band were defined according to the IEC 61260 standard.

After the octave-band decomposition, a Wiener filter was independently applied to each band to reduce the influence of environmental noise and reverberation. Considering that the measured signal $y(t)$ is composed of the desired signal $s(t)$ and an additive noise component $n(t)$,

$$y(t) = s(t) + n(t), \tag{3}$$

the Wiener filter in the frequency domain can be expressed as:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_n(\omega)}, \tag{4}$$

where $P_s(\omega)$ and $P_n(\omega)$ correspond to the power spectral densities of the desired signal and noise, respectively. The filtered signal estimate is obtained through:

$$\hat{S}(\omega) = H(\omega)Y(\omega), \tag{5}$$

where $Y(\omega)$ denotes the Fourier transform of the measured signal and $\hat{S}(\omega)$ represents the filtered signal estimate.

To operationalize the Wiener filter under the dynamic, non-stationary conditions of the milling process, the power spectral densities $P_n(\omega)$ and $P_s(\omega)$ must be dynamically estimated and updated for each isolated octave band. In this work, a data-driven recursive update scheme was implemented.

First, the noise power spectral density $P_n(\omega)$ is estimated during the air-cutting or machine-idle periods preceding tool engagement, where the acoustic field inside the enclosure is composed solely of background structural noise and ventilation. To

accommodate potential slow variations in environmental noise, $P_n(\omega)$ is recursively tracked using a smoothing factor α associated with each specific frequency band:

$$P_n(\omega, k) = \alpha P_n(\omega, k-1) + (1-\alpha)|Y(\omega, k)|^2, \tag{6}$$

where k denotes the discrete time-frame index. The smoothing parameter α controls the adaptation rate and stationarity assumptions of the background noise baseline within that specific spectral region.

Subsequently, the clean signal power spectral density $P_s(\omega, k)$ is updated via a decision-directed approach to prevent musical noise artifacts and handle process non-stationarities. The filter gain $H(\omega, k)$ is bounded by a frequency-dependent minimum gain threshold ($G_{\min}$) to preserve a controlled floor of the residual acoustic field, which prevents excessive signal distortion in bands critical to tool-holder dynamics. The operational transfer function is thus formulated as:

$$H(\omega, k) = \max\left(\frac{P_s(\omega, k)}{P_s(\omega, k) + P_n(\omega, k)}, G_{\min}\right). \tag{7}$$

The Wiener filter parameters were empirically adjusted through an iterative trial-and-error procedure, using the Random Forest classification accuracy as the evaluation criterion. This calibration was performed exclusively on Trial 1 – specifically, on its internal training and validation subsets (80%/20% split, Section 2.4.1) – with no access to the external validation trials (Trials 2–6) at any stage of the tuning process.

2.4 Random Forest classifier

The Random Forest algorithm is an ensemble machine learning method based on the combination of multiple decision trees, applicable to both classification and regression tasks [25]. Each tree is independently trained using a bootstrap sample drawn with replacement from the original dataset (Bagging), while each split is determined from a randomly selected subset of predictor variables (random subspace method). The final prediction is obtained by aggregating the outputs of all trees through majority voting for classification or averaging for regression. This ensemble strategy reduces variance and mitigates overfitting compared to individual decision trees, making Random Forest particularly suitable for nonlinear and noisy datasets. A comprehensive description can be found in Refs. [25, 26].

2.4.1 Training and Validation

The Random Forest classifier was implemented in Python using the *scikit-learn* machine learning library. For each of 6 trials, the acquired acoustic signal was segmented into 50 consecutive temporal windows, each containing 5,000 samples. Each window was treated as an independent observation, resulting in a dataset composed of multiple samples for classifier training and evaluation.

To evaluate the generalization capability and robustness of the classifier under varying experimental setups, a strict validation strategy was adopted. The first experimental trial (Trial 1) was dedicated exclusively to model development, whereas the remaining trials (Trials 2–6) were held out entirely as independent validation datasets for external performance comparison.

To conduct an internal performance assessment during the development phase, the dataset originating from Trial 1 was partitioned into development training and validation subsets using the `train_test_split` function with a test size of 20%. Stratified sampling was adopted to strictly preserve the original class distribution across both subsets. Prior to the training procedure, the categorical classification labels were converted into numerical targets. The Random Forest architecture was configured with 100 decision trees and a maximum tree depth of 10. To mitigate potential class imbalance effects within the machining conditions, balanced class weighting was enabled. The random seed was fixed at 42 to guarantee absolute reproducibility of the results, and parallel processing was deployed utilizing all available CPU cores.

2.4.2 Metrics for Evaluation

The performance of the Random Forest classifier on both internal testing and external validation trials was quantitatively evaluated using a confusion matrix and several derived classification metrics. The confusion matrix represents the classifier predictions in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). From these elements, accuracy, precision, recall, and F_1-score were computed.

Accuracy quantifies the overall proportion of correct predictions along the main diagonal relative to the total number of samples:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Precision evaluates the proportion of correctly predicted positive instances among all samples classified as positive, serving as an indicator of the model's robustness against false alarms:

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

Recall (or sensitivity) measures the proportion of correctly identified positive samples out of all actual positive instances, representing the capability of the system to detect the target machining conditions:

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

The F_1-score corresponds to the harmonic mean of precision and recall, providing a balanced performance metric for classification, especially in the presence of uneven class distributions:

$$F_1 = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

3 Results

3.1 Acoustic and Dynamic Characterization of the System

The characterization of the system constitutes a fundamental step to support the interpretation of the experimental results, as it enables a comprehensive understanding of the intrinsic properties of both the measurement environment and the tool–machine assembly prior to the machining tests. In this section, two complementary aspects are investigated: (i) the acoustic characterization of the enclosure through reverberation measurements, including the estimation of parameters such as the reverberation time (RT_{60}), Schroeder frequency, and critical distance; and (ii) the dynamic characterization of the cutting tool through Frequency Response Functions (FRFs), allowing the identification of its dominant resonance frequencies.

3.1.1 Acoustic characterization of the enclosure

Figure 2 illustrates the experimental results obtained for the acoustic reverberation test using an impulsive excitation signal generated by a firecracker inside the enclosure of the machine tool.The corresponding Schroeder decay curve used for RT_{60} estimation is presented in Fig. 3.

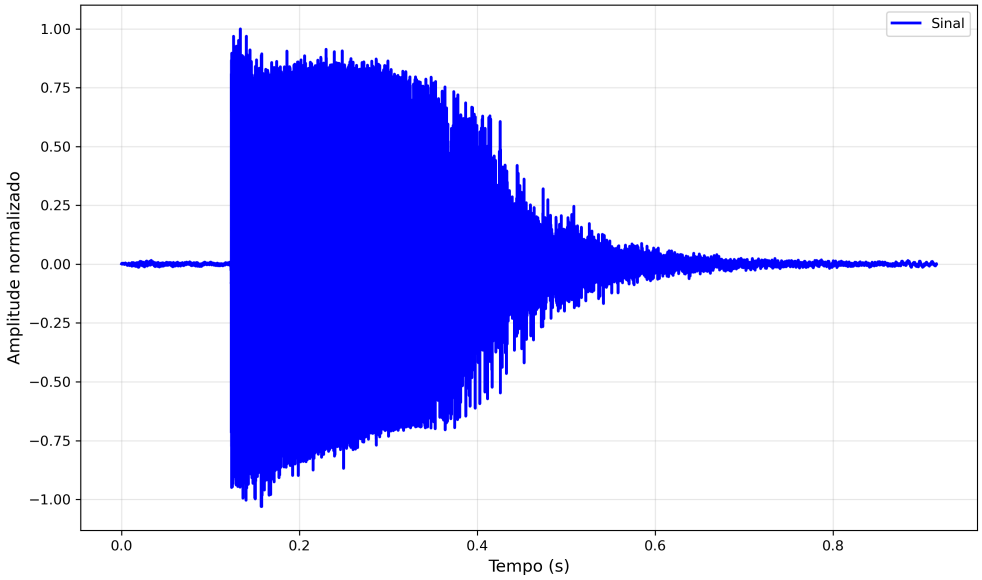


Fig. 2: Excitation signal in the time domain.

The experimental results indicate a moderate reverberation time, with an RT_{60} value of approximately 0.64 s, which is typical of small to medium-sized enclosures with partially reflective surfaces. This behavior is directly associated with the structural characteristics of the machine tool enclosure, whose internal walls are predominantly

composed of steel. Due to its high density and low acoustic absorption coefficient, steel strongly promotes sound reflections and energy persistence, particularly in the mid- and high-frequency ranges.

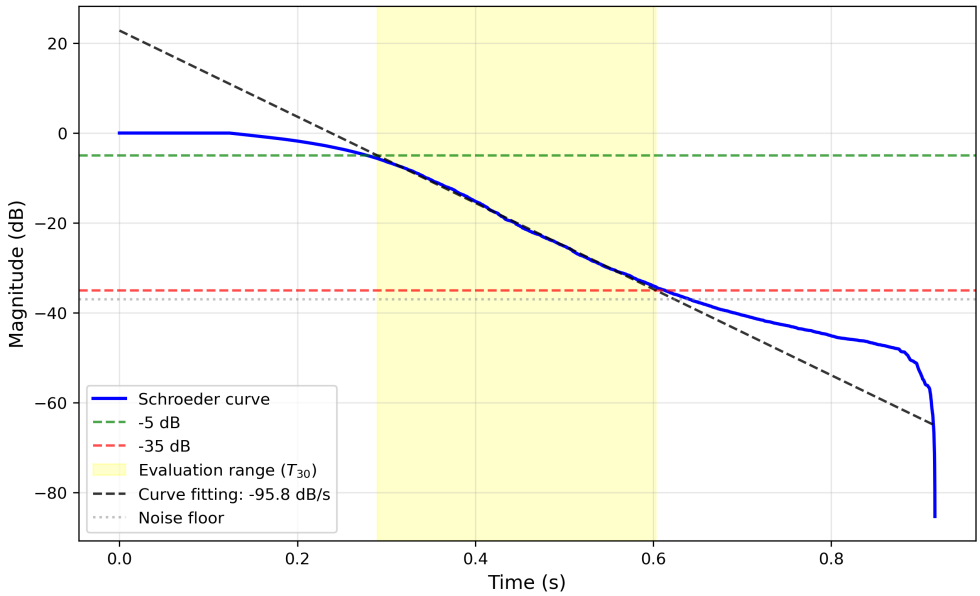


Fig. 3: Schroeder decay curve.

Based on the approximate internal dimensions of the machine enclosure ($2.5 \times 1.2 \times 1.5$ m) and accounting for the volume displaced by the machine table and spindle assembly, the net internal volume (V) of the cavity was estimated at 4.5 m^3.

These reverberation measurements allow for the characterization of the spatial behavior of the acoustic field inside the enclosure. The Schroeder frequency (f_s), which defines the transition between the discrete modal region and the high modal density region where the sound field approaches diffuse-field behavior, was calculated as follows:

$$f_s \approx 2000\sqrt{\frac{RT60}{V}} \approx 2000\sqrt{\frac{0.64}{4.5}} \approx 752 \text{ Hz} \tag{12}$$

For frequencies above approximately 752 Hz, the acoustic field can be treated under the assumptions of a statistically diffuse sound field.

The critical distance (r_c) was also calculated to evaluate the relationship between the direct acoustic energy radiated from the cutting process and the reverberant energy reflected by the enclosure walls. If the microphones are positioned at a distance shorter than r_c, the direct sound field dominates; conversely, at greater distances, the reverberant field becomes dominant. When the distance equals r_c, a balanced mixture of

both fields occurs. Assuming an omnidirectional sound source ($Q = 1$), the critical distance is approximated by:

$$r_c \approx 0.14\sqrt{\frac{V}{RT60}} \approx 0.14\sqrt{\frac{4.5}{0.64}} \approx 0.37 \text{ m} \tag{13}$$

The center of the microphone array was positioned approximately 0.38 m from the cutting zone, placing the acquisition system in a condition close to the critical distance ($r \approx r_c$). This configuration characterizes the measurement region as a hybrid acoustic field, where the energy densities of the direct field and the reverberant field exhibit comparable magnitudes.

The impulsive firecracker signal was also filtered into octave bands, as described in the methodology section. The resulting RT_{60} and critical distance values indicate that the enclosure operates predominantly under reverberant conditions over the entire analyzed frequency range, as shown in Fig. 4. A comparison between the microphone-to-source distance ($r = 0.38$ m) and the frequency-dependent critical distance (r_c) reveals that the ratio (r/r_c) remained greater than unity for all frequency bands, confirming the predominance of the reverberant field over the direct field.

It can be observed that RT_{60} decreases progressively with increasing frequency, varying from approximately 1.08 s at 125 Hz to 0.43 s at 8000 Hz. This behavior is characteristic of metallic enclosures, where low-frequency acoustic waves experience lower absorption and consequently exhibit longer energy persistence. At higher frequencies, acoustic losses associated with atmospheric absorption and wall scattering become more significant, thereby reducing the reverberation time [7].

As a direct consequence of this RT_{60} reduction, the critical distance increases gradually with frequency, ranging from approximately 0.17 m at 125 Hz to 0.28 m at 8000 Hz. This trend indicates that higher-frequency acoustic components propagate more efficiently before the reverberant field becomes dominant. Nevertheless, the microphone array remained positioned above the critical distance limit for all analyzed bands, with (r/r_c) ratios ranging from 2.2 to 1.4. Therefore, the measurement system operated entirely within the reverberant field, although the degree of reverberation was non-uniform throughout the spectrum. Low-frequency bands exhibited stronger reverberant behavior, whereas higher-frequency bands approached the transition region between the direct and reverberant fields.

3.1.2 Spatial coherence analysis

The similarity index values obtained for the six microphone pairs ranged from 0.88 to 0.95, as presented in Table 3. The average value was 0.913, indicating a good agreement between the experimentally measured coherence functions and those predicted by the Rayleigh diffuse-field model.

The relatively small variation among the six microphone pairs indicates that the agreement with the theoretical model is not restricted to a particular sensor spacing. Both adjacent and diagonal microphone pairs exhibited comparable similarity values,suggesting that the spatial coherence throughout the array is consistent with the behavior expected for a diffuse acoustic field.

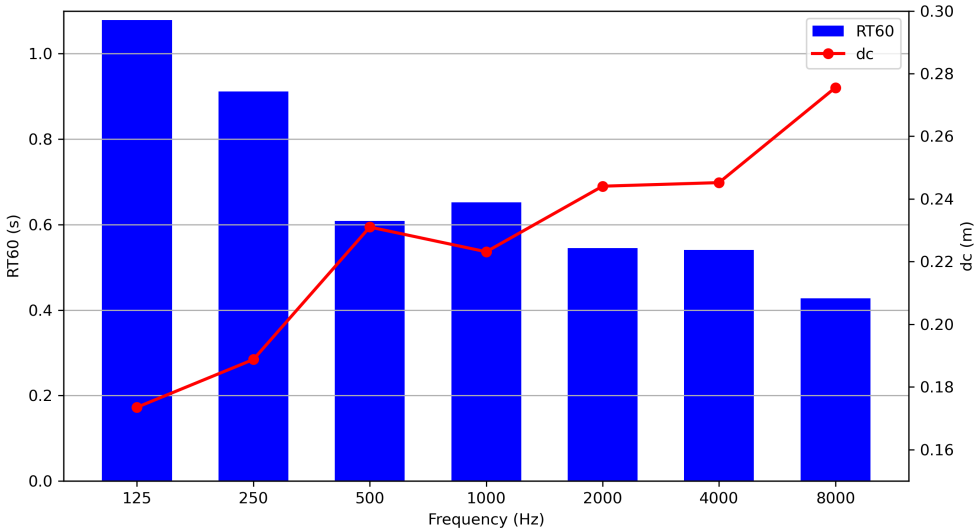


Fig. 4: **RT60 and critical distance as a function of frequency.**

Table 3: Similarity values obtained between measured and theoretical coherence functions for each microphone pair.

Microphone pair	**Similarity**
Adjacent 0–1	~ 0.92
Diagonal 0–2	~ 0.91
Adjacent 0–3	~ 0.92
Adjacent 1–2	~ 0.88
Diagonal 1–3	~ 0.89
Adjacent 2–3	~ 0.95

These results support the assumption that, over a substantial portion of the analyzed frequency range, the measured spatial coherence is consistent with that expected for an approximately diffuse sound field.

The coherence analysis complements the reverberation-time and frequency-response measurements presented previously. While the FRF analysis revealed the transition from discrete resonances to a dense modal distribution around the estimated Schroeder frequency, the high agreement between the measured and theoretical coherence functions further supports the interpretation that the spatial characteristics

of the sound field are consistent with diffuse-field behavior over most of the analyzed frequency range. Together, these independent observations provide a consistent description of the acoustic environment inside the machine enclosure.

3.1.3 Structural and acoustic FRFs

The FRF of the machine enclosure obtained from the impulsive excitation generated by the firecracker source is presented in Fig. 5. The measured FRF characterizes the global acoustic behavior of the enclosure and reveals the frequency regions where the cavity exhibits greater acoustic amplification.

The measured response shows a limited number of well-defined resonance peaks below approximately 750 Hz, which are attributed to the dominant low-frequency acoustic resonances of the enclosure. As discussed in the previous section, the estimated reverberation time was $RT_{60} = 0.64$ s, yielding a Schroeder frequency of approximately 752 Hz.

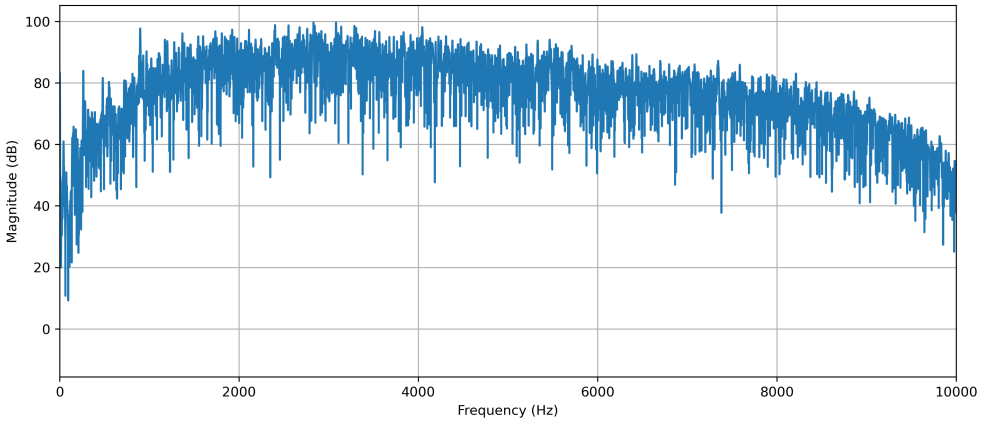


Fig. 5: Acoustic FRF of the enclosure.

The overall spectral magnitude gradually increases from low frequencies up to approximately 2 kHz, where the measured response reaches its highest levels. Between approximately 2 kHz and 5 kHz, the response remains relatively high despite the large number of closely spaced peaks and notches produced by modal overlap. Above approximately 5 kHz, the average response gradually decreases, becoming more pronounced beyond 8 kHz, which is consistent with increasing acoustic and structural damping at higher frequencies.

Firecrackers exhibit reduced acoustic energy at low frequencies compared with the mid-frequency range. Nevertheless, Papadakis and Stavroulakis [18] demonstrated that they provide sufficient bandwidth for room impulse-response measurements up to 8 kHz. Accordingly, the measured frequency response should be interpreted considering the spectral characteristics of the excitation source, particularly at low frequencies.

To better interpret the low-frequency response ($f < f_s$), the acoustic modes of the enclosure were estimated assuming an ideal rigid rectangular cavity. The first axial modes were predicted at 68.6 Hz, 114.3 Hz, and 142.9 Hz. Although the actual enclosure contains several internal components, including the machine table and spindle, which modify the ideal cavity geometry, these analytical estimates provide a physical reference for the dominant resonances observed experimentally in the low-frequency region.

Fig. 6 illustrates the Frequency Response Function (FRF) of the cutting tool measured up to 10 kHz. Several well-defined resonance peaks can be observed at approximately 316, 749, 936, 1038, 2101, 2509, 3482, 4199, 4528, 6074, 7263 and 8266 Hz, corresponding to the dominant vibration modes of the tool structure. Among these modes, the resonance located near 936 Hz exhibits the highest receptance magnitude, indicating the most compliant structural mode of the tool–holder assembly, followed by the mode around 2101 Hz and 316 Hz.

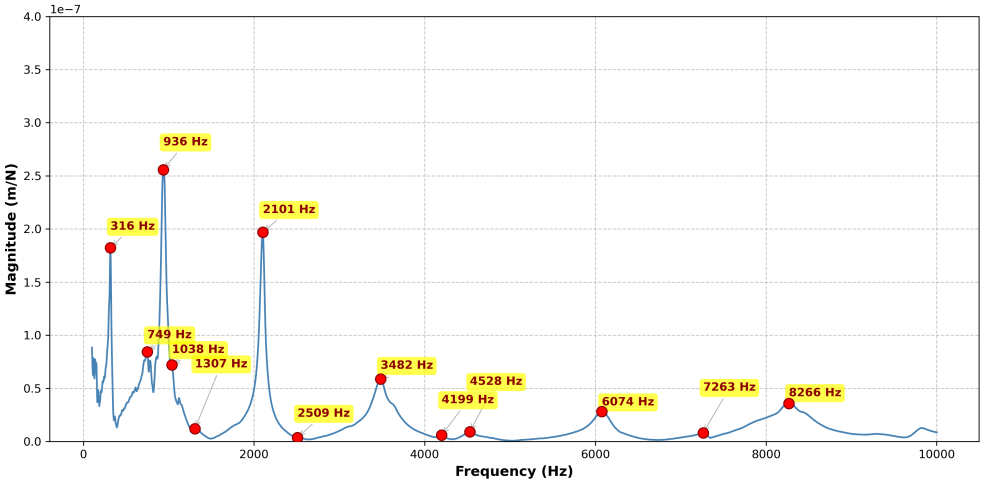


Fig. 6: FRF of the cutting tool.

When compared with the acoustic FRF of the machine enclosure (Fig. **??**), two important observations can be made. First, most structural resonance frequencies of the cutting tool are located above the estimated Schroeder frequency ($f_s \approx 752$ Hz), where the enclosure exhibits a highly reverberant sound field characterized by significant modal overlap. Second, the most compliant structural modes, particularly those near 936 Hz and 2101 Hz, coincide with the frequency range in which the measured acoustic response of the enclosure reaches its highest levels. Consequently, structural vibrations generated during milling are expected to radiate into an acoustic environment where numerous overlapping cavity modes efficiently sustain the acoustic energy, increasing the persistence and complexity of the measured sound field.

3.2 Analysis of milling acoustic signals

This section presents the experimental results obtained during the milling tests, focusing on the analysis of the acquired acoustic signals in both the time and frequency domains.

3.2.1 Signal in Time domain

As shown in Fig. 7, the temporal evolution of the acoustic signal acquired during the complete machining cycle for the eight experimental conditions investigated in this work is presented. These signals correspond to the eight runs defined by the fractional factorial design of experiments (DOE), where the cutting parameters were systematically varied.

The total machining time varies across the different experimental conditions due to the distinct operational kinematics dictated by the combinations of cutting speed and feed rate. Furthermore, since the operations consisted of face milling the top surface of a cylindrical workpiece, the actual physical tool path length changes across all tests.

As observed in the plot, the eight distinct periods of higher signal amplitude correspond to the effective engagement of the tool with the workpiece. Conversely, the lower amplitude intervals between these periods represent the "air cutting" stages, where the tool transitions between experiments without material removal.

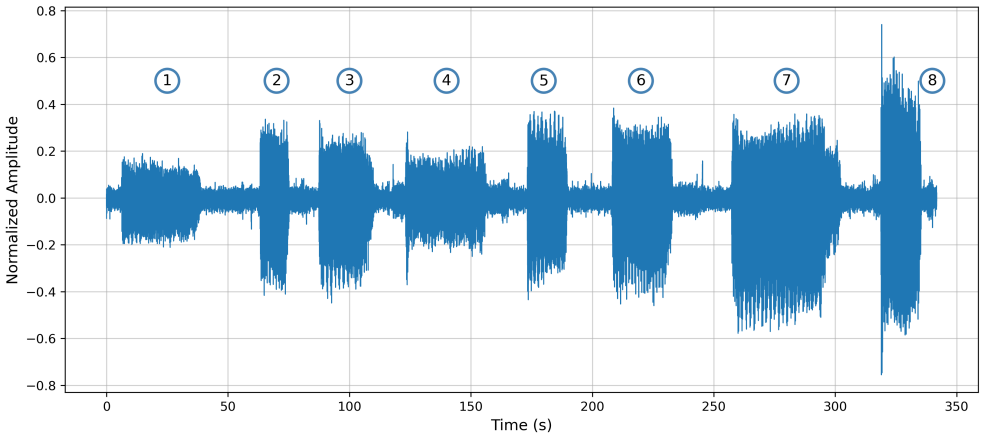


Fig. 7: Time-domain acoustic signal. Conditions 1–8 correspond to the machining parameters described in the Materials and Methods section.

By evaluating the intersection points where the localized RMS signal permanently crosses the predefined energy threshold, the active cutting zone was isolated. Only the data blocks within this stable cutting region were retained for subsequent feature extraction and machine learning classification, thereby preventing transient entry/exit acoustic bursts from corrupting the model's training features, as can be seen in Fig 8.

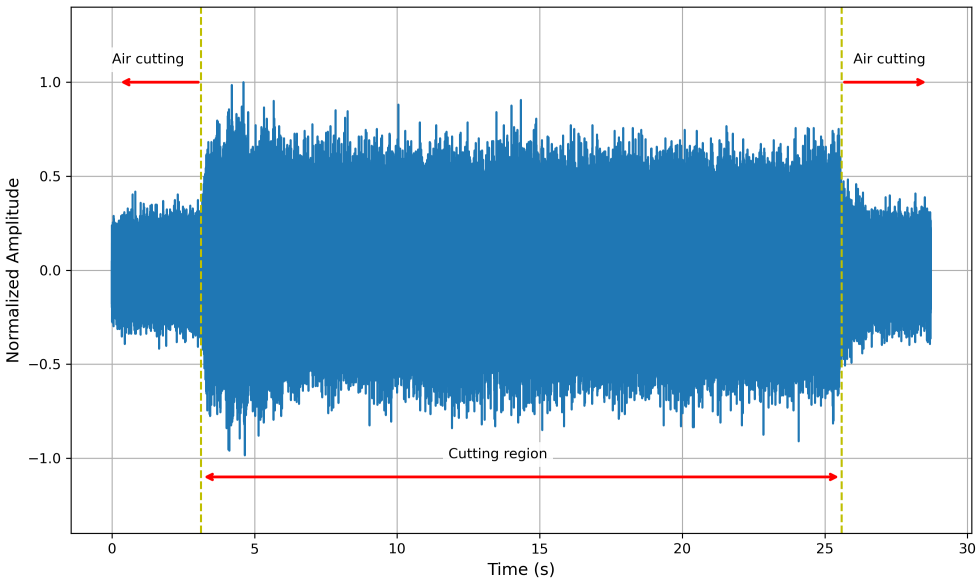


Fig. 8: Time-domain acoustic signal for the array positioned perpendicular to the spindle axis.

3.2.2 Frequency-Domain Analysis

Fig 9 presents the Fast Fourier Transform (FFT) spectrum obtained for the cutting condition corresponding to $v_c = 200$ m/min, $f_z = 0.2$ mm/rev, $a_e = 5$ mm, and $a_p = 1$ mm. The experimental spectrum is analyzed together with the spindle pass frequency (SPF), the tooth pass frequency (TPF), and their corresponding higher-order harmonics. These fundamental excitation frequencies are determined from the spindle rotational speed N (rpm), where $SPF = N/60$ and $TPF = (N \cdot z)/60$, with z representing the number of inserts on the milling cutter.

From the spectral distribution, several integer multiples of the SPF and TPF can be observed to coincide with the dominant peaks in the measured FFT spectrum. These harmonic components are further amplified when they overlap with the structural resonances of the tool-holder-spindle assembly. Their amplitudes are also affected by the acoustic frequency response of the machine enclosure, which selectively reinforces specific frequency components through reverberation and cavity resonances, as can be seen in Figs 9 (b) and (c), which present a zoomed-in view around 1000 Hz and 2000 Hz, respectively.

Furthermore, it can be explicitly stated that the highest spectral amplitudes are concentrated within the frequency range between 1000 Hz and 3000 Hz, with the absolute maximum amplitude peak occurring around 2100 Hz. Rather than exhibiting isolated, sharp resonance lines, the spectrum in this region displays a pronounced energy scattering (spectral smearing) around specific frequencies. This behavior is fundamentally associated with the cutting tool's most flexible structural modes,

specifically those identified at 936 Hz and 2101 Hz,interacting under operational conditions.

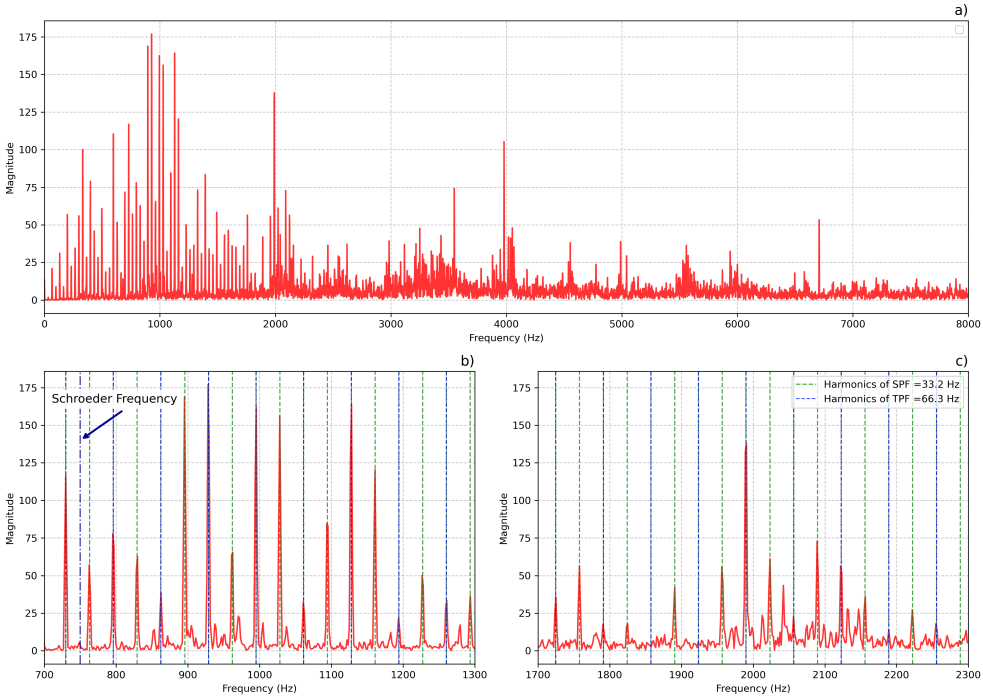


Fig. 9: FFT spectrum of the acoustic signal for the selected cutting condition: a) complete range up to 8000 Hz, b) interval between 0 and 1200 Hz, and c) interval between 1400 and 2600 Hz showing the harmonics of SPF and TPF

The physical mechanism driving this spectral broadening comprises the high structural damping introduced by the tool–workpiece engagement, coupled with the high modal density of the enclosure cavity above the Schroeder frequency ($f_s = 752$ Hz). Furthermore, the continuous variation of cutting forces inherent to intermittent tooth impacts induces frequency modulation effects. Consequently, when the SPF and TPF harmonics excite these structural-acoustic frequencies, the energy is distributed across a continuous bandwidth rather than discrete peaks.

This frequency-dependent sensitivity maps distinctly across the analyzed spectrum. While low-frequency components (below 500 Hz) are predominantly driven by the global structural dynamics of the machine–tool system and cavity modal amplification, the intermediate frequency range, extending up to 3000 Hz, is heavily governed by the coupled interaction between tool vibrations, cutting dynamics, and reverberant energy accumulation. This intense structural–acoustic coupling makes these intermediate bands highly sensitive to variations in machining conditions and particularly relevant for chatter detection.

3.2.3 Influence of cutting parameters

The Design of Experiments (DOE) approach was employed to investigate the influence of cutting speed (v_c), feed per tooth (f_z), radial depth of cut (a_e), and axial depth of cut (a_p) on the acoustic response generated during the milling process for both feature extraction strategies.

Fig. 10 illustrates the standardized main effects of the cutting parameters on the acoustic descriptors extracted using the conventional Time- and Frequency-Domain Analysis (Strategy 1). Overall, the results reveals that the axial depth of cut (a_p) is the dominant machining parameter, followed by the feed per tooth (f_z). The largest standardized effects were observed for energy-related descriptors, including RMS, standard deviation, and peak amplitude, with maximum values reaching approximately 1.53, 1.44, and 1.43, respectively. In contrast, the cutting speed (v_c) produced only moderate effects, whereas the radial depth of cut (a_e) generally exhibited the lowest sensitivity among the investigated machining parameters. These results suggest that descriptors associated with signal energy are more sensitive to changes in axial depth of cut than to the other investigated machining parameters.

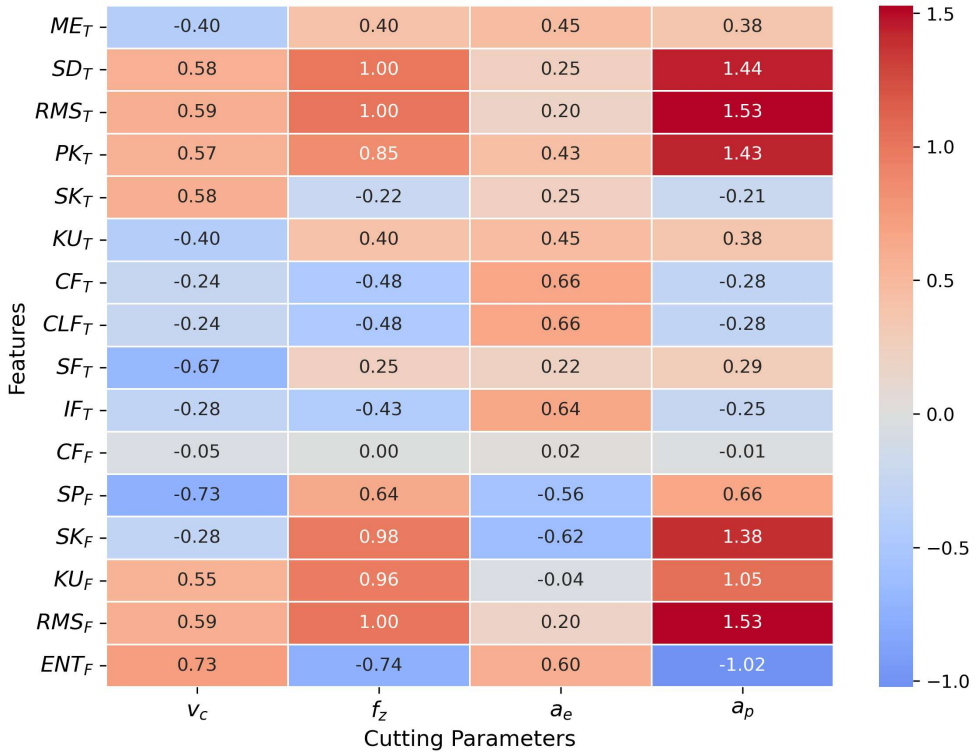


Fig. 10: Heatmap of the relative sensitivity to cutting parameters.

An interesting observation is that the RMS descriptors extracted in the time and frequency domains exhibit nearly identical standardized effects. This behavior is expected from Parseval's theorem, which states that the total signal energy is preserved between the time and frequency domains. Consequently, both descriptors contain essentially the same information regarding the influence of the machining parameters.

However, a different behavior is observed for the remaining frequency-domain descriptors. While spectral spread, spectral skewness, spectral kurtosis, and spectral entropy remained sensitive to changes in the cutting parameters, the spectral centroid exhibited almost no variation under any experimental condition, with standardized effects remaining close to zero for all factors. This finding indicates that although the overall acoustic energy changes considerably with machining conditions, the average spectral balance of the measured sound field remains nearly unchanged.

A plausible explanation for this behavior is the reverberant acoustic environment inside the machine enclosure. Multiple reflections from the enclosure walls redistribute acoustic energy throughout the cavity, producing a more diffuse sound field. Under these conditions, broadband spectral descriptors become less sensitive to localized changes in the sound generation mechanisms associated with the cutting process. Consequently, variations in the machining parameters can substantially modify the overall signal energy while producing only minor changes in descriptors that depend on the spectral energy distribution, such as the spectral centroid.

Fig. 11 presents the standardized main effects obtained for Strategy 2, in which the acoustic signals were first decomposed into octave bands and subsequently processed using Wiener filtering. Three RMS-based descriptors were extracted from each frequency band, namely the mean RMS, RMS standard deviation, and RMS range (ΔRMS), computed across the four microphones of the array.

From that results, the feed per tooth (f_z) is the dominant machining parameter over most of the analyzed frequency range. The largest standardized effects were observed for the mean RMS descriptor, reaching values of 1.31, 1.33, 1.41, and 1.42 in the 125, 250, 1000, and 2000 Hz octave bands, respectively. In contrast, the influence of cutting speed (v_c) increases progressively with frequency, becoming the dominant factor in the 4000 and 8000 Hz bands, where standardized effects reached 1.47 and 1.41, respectively.

Axial depth of cut (a_p) produced moderate positive effects, particularly at frequencies below 2000 Hz, whereas radial depth of cut (a_e) consistently exhibited the smallest standardized effects across nearly all frequency bands. These observations indicate that the proposed frequency-selective analysis is capable of separating the influence of different machining parameters according to their spectral content, revealing trends that are not apparent when broadband descriptors are employed.

Compared with Strategy 1, a noticeable difference becomes evident. The conventional broadband descriptors summarize the entire acoustic spectrum into a small number of global statistical features, inevitably averaging information generated by different physical mechanisms. As a result, several descriptors, particularly the spectral centroid, exhibited little sensitivity to the machining parameters. Conversely,

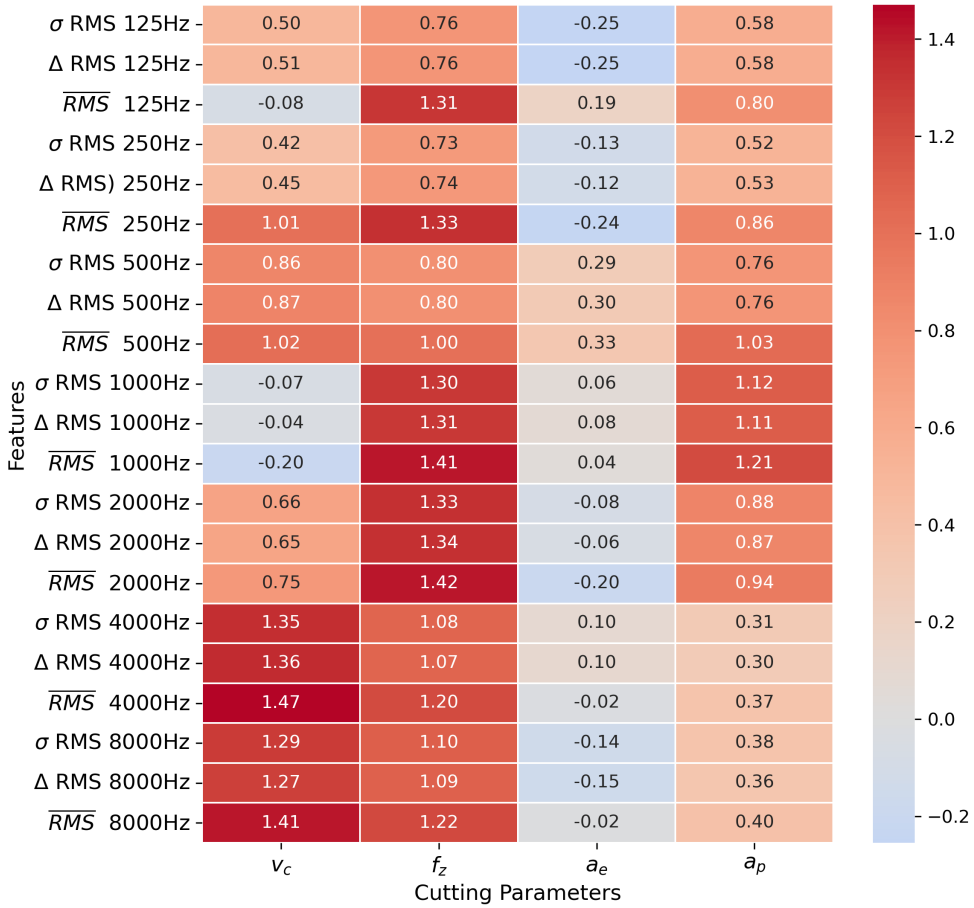


Fig. 11: Normalized RMS sensitivity heatmap for the investigated cutting parameters.

the octave-band decomposition preserved the parameter sensitivity over multiple frequency regions, demonstrating that relevant process information remains embedded within specific portions of the spectrum.

These observations are consistent with the acoustic characteristics of the machine enclosure. In a reverberant environment, multiple reflections redistribute acoustic energy throughout the enclosure, reducing the effectiveness of broadband spectral descriptors. However, by first separating the signal into octave bands and then applying Wiener filtering, the proposed approach improves the signal-to-noise ratio within each frequency region while preserving frequency-dependent process information. Consequently, variations associated with chip formation, cutting forces, and structural vibrations remain detectable despite the reverberant sound field. This explains the substantially higher sensitivities observed for the octave-band RMS descriptors and provides a physical justification for the superior classification performance obtained with Strategy 2.

3.3 Performance of the classification strategies

The Random Forest algorithm was employed to classify the acoustic signals acquired during the experimental milling tests. The models were initially trained using the data obtained from Trial 1 and subsequently evaluated using five independent experimental Trials not included in the training stage. This procedure allowed the investigation of the generalization capability of the different feature-extraction strategies under varying acquisition conditions.

Table 4: Random Forest classification accuracy for the investigated feature-extraction strategies.

Trial	**Strategy 1 (%)**	**Strategy 2 (%)**
1	99.74	100.00
2	34.38	99.74
3	36.46	93.88
4	64.84	95.92
5	67.97	99.74
6	85.68	79.34
Mean	64.85	94.77
Standard deviation	26.07	7.92

Table 4 presents the classification accuracies obtained for both feature-extraction strategies. The obtained results demonstrate that Strategy 2 significantly outperformed Strategy 1 during the validation stage. While Strategy 1 achieved an average accuracy of 64.69% with a standard deviation of 24.54%, Strategy 2 reached an average accuracy of 94.77% with a considerably lower standard deviation of 7.92%. In addition to the higher average accuracy, Strategy 2 exhibited substantially more consistent performance across the independent experimental trials, indicating superior robustness and generalization capability.

Although Strategy 1 achieved high accuracy during training (98.70%) and satisfactory performance in Trial 6 (85.68%), a severe performance degradation was observed in Trials 2, 3, 4, and 5. Particularly in Trials 2 and 3, the classification accuracy dropped to 34.38% and 36.46%, respectively.

Conversely, Strategy 2 maintained classification accuracies above 90% in most validation Trials, reaching values close to 100%, with the exception of Trial 6, which still presented a satisfactory performance of 79.34%.

Overall, the results show that the octave-band representation combined with Wiener filtering produced features that were far more representative of the actual machining conditions and less dependent on acquisition-specific variations. A particularly relevant aspect is that the preprocessing stages incorporated into Strategy 2, namely octave-band decomposition and Wiener filtering, produced a substantial improvement in classification performance when compared with the direct feature extraction approach employed in Strategy 1.

Fig 12 and Fig. 13 present the pooled confusion matrices obtained by summing the classification counts across the six experimental trials and subsequently normalizing each row by the total number of samples of the corresponding true class, for both feature extraction strategies.

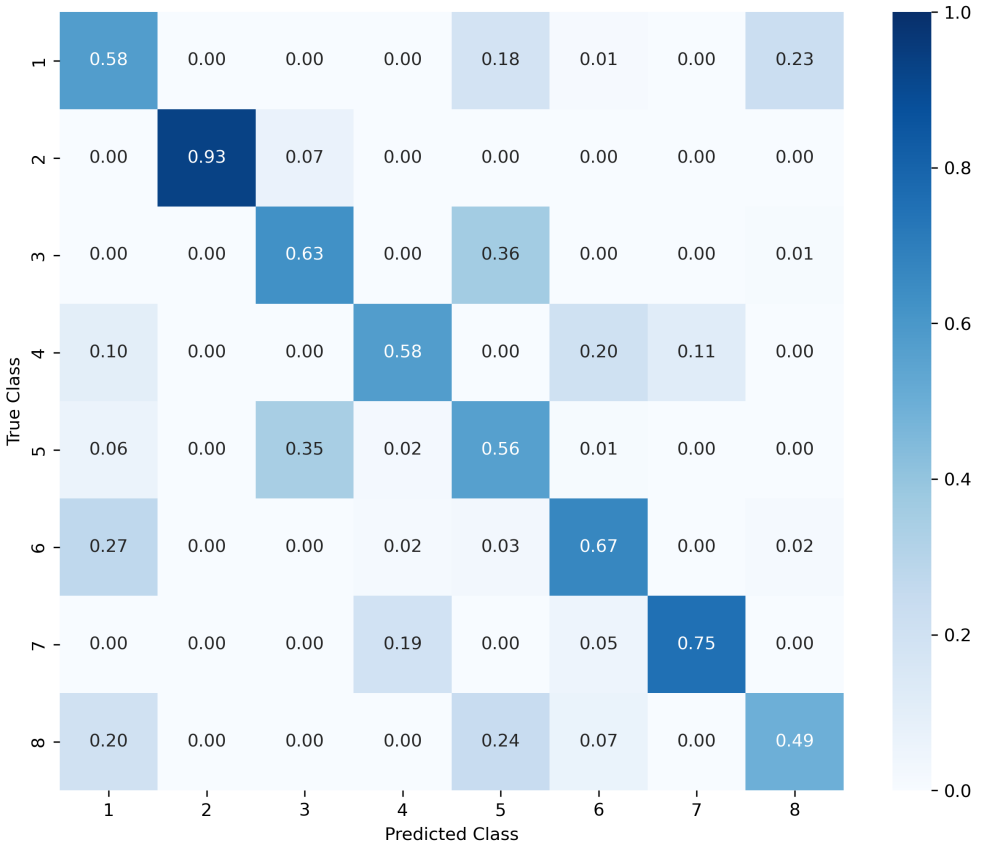


Fig. 12: Confusion matrix for the feature extraction in Strategy 1..

The analysis of these pooled confusion matrices allows for a comprehensive evaluation of the global classification behavior under each strategy. It is worth noting that the acoustic signal isolated within the active cutting region is partitioned into n consecutive time windows. Since the Random Forest model performs the classification on a per-segment basis, the confusion matrices depict the normalized prediction probabilities (ranging from 0 to 1) along the true and predicted classes.

For Strategy 1, a significant dispersion of misclassifications is observed outside the main diagonal, with errors distributed across multiple neighboring classes. Although some specific cutting conditions exhibit a high true positive rate, there is a prominent presence of systematic errors, particularly among classes characterized by adjacent

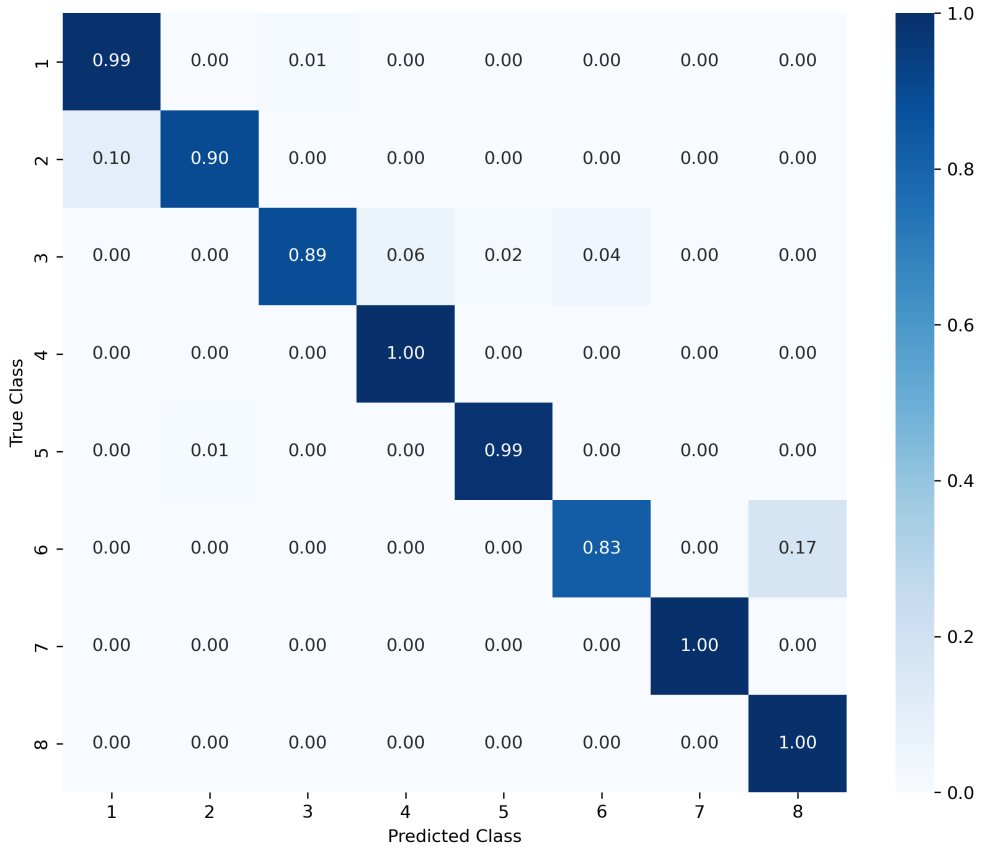


Fig. 13: Confusion matrix for the feature extraction in Strategy 1.

experimental parameters in the design of experiments. The wide error distribution reflects the features' high sensitivity to inter-trial variations, which aligns with the elevated accuracy standard deviation observed in the previous analysis.

In contrast, Strategy 2 displays a predominantly diagonal confusion matrix, demonstrating a high separability across all investigated machining states. The majority of the time segments are correctly classified, with minor residual errors tightly concentrated at specific boundary transitions between highly similar cutting configurations.

Table 5 depicts the classification metrics obtained for both strategies. The comparative analysis between the two feature extraction approaches demonstrates a substantial and systematic improvement in classification performance when octave-band decomposition and Wiener filtering are incorporated into the signal processing pipeline. Strategy 2 achieved an average accuracy of 94.77%, which is higher than the 64.85% obtained via Strategy 1. In addition to the higher mean accuracy, Strategy 2 also presented considerably lower standard deviation values across all evaluated metrics,

Table 5: Comparison of classification metrics obtained for the two feature extraction strategies.

Metric	Strategy 1		Strategy 2	
	Mean (%)	**Std (%)**	**Mean (%)**	**Std (%)**
Accuracy	64.85	26.07	94.77	7.92
Precision	65.00	26.14	96.17	5.02
Recall	64.83	26.37	94.17	8.02
F1-score	63.83	26.39	93.50	9.33

indicating superior stability and generalization robustness across the independent experimental validation trials.

The comparative analysis of precision, recall, and F1-score values further reinforces this trend. While Strategy 1 exhibited average metrics bounded around 64% and coupled with high standard deviations of approximately 26%, Strategy 2 consistently achieved precision, recall, and F1-score values well above 93%, while maintaining standard deviations below 10%. These macro-results confirm that the sub-band features extracted after octave-band decomposition and Wiener filtering provided a more discriminative representation of the actual machining states, successfully mitigating the classifier's sensitivity to trial-to-trial environmental fluctuations and reverberant acoustic interference.

3.4 Feature Importance Analysis

The feature importance analysis obtained from the Random Forest model provides insights into which descriptors contributed most to the classification process for each feature extraction strategy. Fig 14 and 15 present the ten most important attributes ranked according to their relative importance scores for both strategies.

For Strategy 1, the ranking is dominated by conventional time- and frequency-domain descriptors. The temporal standard deviation (SD_T) was identified as the most influential feature (17.5%), followed by the spectral RMS (RMS_F, 16.0%), temporal RMS (RMS_T, 16.0%), and spectral kurtosis (KU_F, 15.6%). Features such as spectral skewness (SK_F), temporal peak amplitude (PK_T), and spectral entropy (ENT_F) exhibited intermediate importance, whereas the temporal mean (ME_T), temporal kurtosis (KU_T), and spectral spread (SS_F) showed comparatively smaller contributions.

A relevant aspect is that the ten highest-ranked features account for approximately 94% of the total feature importance. This result indicates that the Random Forest classifier relies primarily on a relatively small subset of descriptors, while the remaining variables contribute only marginally to the classification decisions.

The prominence of RMS-, standard deviation-, and kurtosis-based descriptors suggests that global variations in signal energy and statistical distribution provide the most discriminative information for distinguishing the investigated cutting conditions.

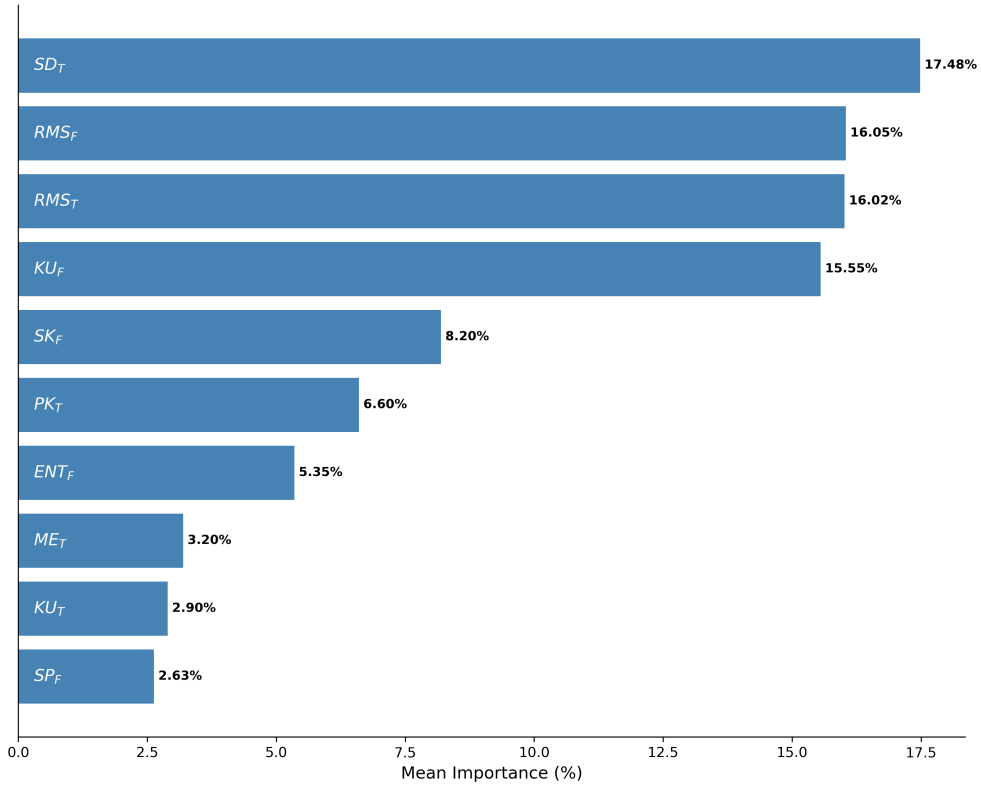


Fig. 14: Most 10 significant features in Strategy 1.

For Strategy 2, the feature ranking was dominated by the mean octave-band RMS descriptors ($\overline{RMS}$) computed across the four-microphone array. The five most important variables corresponded to the mean RMS values in the 4 kHz (14.9%), 8 kHz (14.5%), 2 kHz (12.5%), 250 Hz (11.7%), and 1 kHz (11.2%) octave bands. The mean RMS at 500 Hz also presented a noticeable contribution (6.4%). In contrast, the remaining top-ranked features consisted primarily of the RMS range (ΔRMS) extracted from the 1 kHz, 4 kHz, and 8 kHz bands, together with the mean RMS in the 125 Hz band.

The ten highest-ranked features account for approximately 84% of the total feature importance, indicating that the classification model relies mainly on a relatively small subset of frequency-selective descriptors.

Compared with Strategy 1, a relevant difference can be observed. Whereas the most relevant descriptors in Strategy 1 were global time- and frequency-domain statistics, Strategy 2 was dominated almost exclusively by octave-band mean RMS features. This result suggests that preserving frequency-selective information provides descriptors that proved to be more effective for distinguishing the investigated cutting conditions.

These feature ranking results complement the sensitivity analysis presented in the previous section. While the DOE heatmaps quantify the influence of the machining

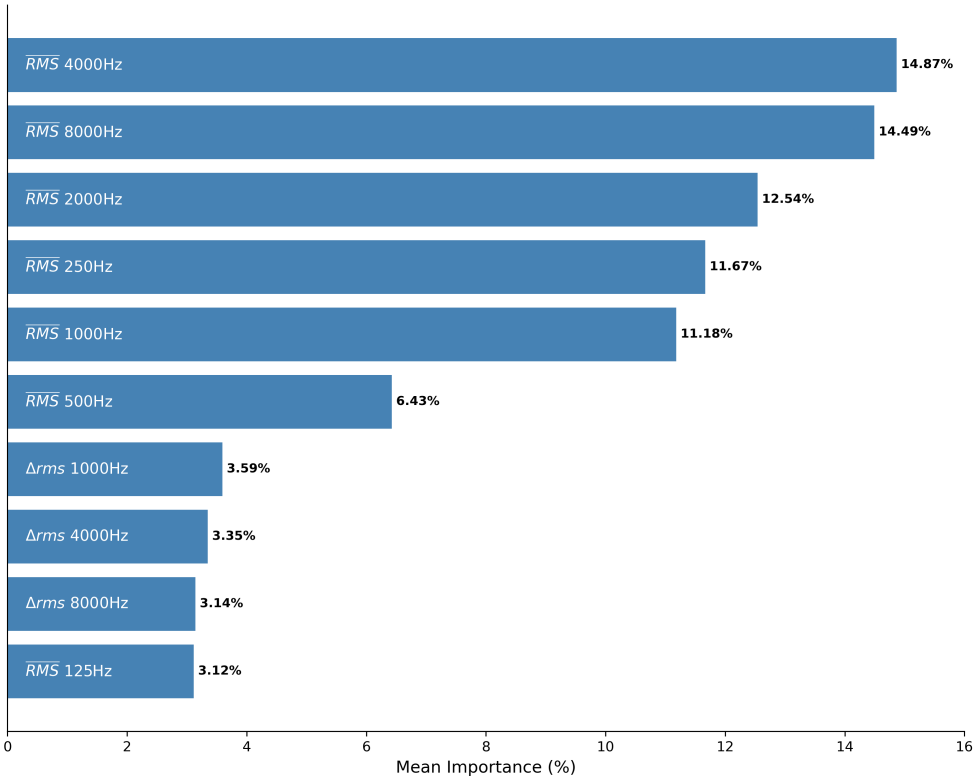


Fig. 15: Most 10 significant features in Strategy 2.

parameters on each descriptor individually, the Random Forest feature importance identifies which descriptors contributed most to the classification task when all variables were considered simultaneously. Although the two analyses address different aspects of the data, both consistently highlight the relevance of the octave-band RMS descriptors.

Several of the highest-ranked descriptors are located in frequency bands that overlap the frequency region where the tool-holder FRFs exhibit pronounced dynamic responses. Although the feature importance analysis does not establish a causal relationship, this observation is consistent with the hypothesis that these frequency regions contain useful information for discriminating the investigated cutting conditions.

The feature importance analysis reinforces the previous classification results by showing that the Random Forest model relies primarily on frequency-selective RMS descriptors rather than on broadband statistical features. Together with the classification metrics presented in the previous section, these findings support the use of octave-band analysis combined with Wiener filtering as an effective feature extraction strategy for acoustic monitoring of the investigated milling process.

4 Conclusions

This work investigated the deployment of a microphone array operating within the reverberant acoustic environment of an enclosed machine tool during milling operations. Particular attention was given to the influence of enclosure acoustics on the acquired signals, the sensitivity of the extracted features to variations in cutting parameters, and the robustness of two distinct feature extraction strategies when combined with a Random Forest classifier under realistic, noise-dense industrial conditions.

The acoustic characterization showed that the machine tool enclosure exhibited a predominantly reverberant behavior, closely approaching a diffuse sound field above the Schroeder frequency. However, despite the presence of multiple reflections and acoustic energy storage within the cavity, the distinct acoustic signatures generated by the milling process were not completely masked, demonstrating that critical process-related information remains accessible for monitoring purposes.

Feed per tooth exhibited the strongest influence over most of the analyzed frequency range, whereas cutting speed became dominant in the highest-frequency octave bands. This behavior is consistent with the interaction between the dynamic characteristics of the tool-holder system, the modal response of the enclosure, and the acoustic energy generated by the periodic tooth-passing impacts on the workpiece.

A comparison between the two feature extraction strategies demonstrated that conventional broadband descriptors based on global time- and frequency-domain statistics suffer a reduced sensitivity in highly reverberant environments, where multiple boundary reflections tend to homogenize the measured acoustic field. In contrast, the proposed strategy based on octave-band decomposition, frequency-selective Wiener filtering, and spatial descriptors derived from the microphone array successfully preserved localized process information, yielding superior classification performance. The stark improvement observed across all classification metrics confirms that frequency-selective representations provide a far more robust description of the machining process than broadband statistical measures.

Furthermore, the feature importance analysis showed that the Random Forest classifier relied primarily on octave-band RMS descriptors. Most of the highest-ranked features were concentrated in frequency bands that overlapped with structural resonance regions identified in the experimental FRFs. Although this observation does not establish a causal relationship, it is consistent with the experimental characterization of the machining system and reinforces the relevance of frequency-selective descriptors for acoustic monitoring.

Overall, the experimental results show that acoustic monitoring of milling operations inside enclosed machine tools is feasible using a low-cost MEMS microphone array combined with a Random Forest classifier. The proposed frequency-selective framework consistently achieved higher classification accuracy than the conventional broadband approach, indicating that octave-band analysis combined with Wiener filtering provides a more informative representation of the acoustic response under reverberant conditions.

Acknowledgements.

References

[1] Teti, R., Jemielniak, K., O'Donnell, G., Dornfeld, D.: Advanced monitoring of machining operations. CIRP Annals **59**(2), 717–739 (2010) https://doi.org/10.1016/j.cirp.2010.05.010

[2] Guochao, L., Shang, X., Li, S., Bofeng, F., Lei, Y., Honggen, Z.: Application of audible sound signals in tool wear monitoring: A review. Journal of Advanced Manufacturing Science and Technology **5**(1), 2025003 (2025)

[3] Talebloo, M., Tuysuz, O.: An in-process dominant vibration mode frequencies identification model for thin-walled parts using a microphone. The International Journal of Advanced Manufacturing Technology, 1–11 (2025) https://doi.org/10.1007/s00170-025-15369-2

[4] Delio, T., Tlusty, J., Smith, S.: Use of audio signals for chatter detection and control. Journal of Engineering for Industry **114**(2), 146–157 (1992) https://doi.org/10.1115/1.2899767

[5] Lu, L., Kurfess, T., Saldana, C.: Effects of extrinsic noise factors on machine learning–based chatter detection in machining. Smart and Sustainable Manufacturing Systems **5**(1), 167–180 (2021) https://doi.org/10.1520/SSMS20210007

[6] Brandão, E.: Acústica de Salas: Projeto e Modelagem. Editora Blucher, São Paulo (2018). (In Portuguese)

[7] Kuttruff, H.: Room Acoustics, 6th edn. CRC Press, Boca Raton (2016)

[8] Yang, L., Li, G., Xie, D., Kumar, A., Sun, L., Shang, X., Wu, J.: Modeling and analysis of the machining-generated audible sound field in the closed machine space. Arabian Journal for Science and Engineering, 1–12 (2025) https://doi.org/10.1007/s13369-025-10005-3

[9] Parsian, A., Magnevall, M., Beno, T., Eynian, M.: Sound analysis in drilling, frequency and time domains. Procedia CIRP **58**, 411–415 (2017) https://doi.org/10.1016/j.procir.2017.03.242

[10] He, M., Li, G., Si, M., Wang, Z., Sun, L., Gong, Y., Dong, Y.: Analysis of the machining-generated audible sound field outside the closed machine space based on and multiple microphone measurement test. The International Journal of Advanced Manufacturing Technology, 1–13 (2026) https://doi.org/10.1007/s00170-026-17679-5

[11] Wegener, K., Bleicher, F., Heisel, U., Hoffmeister, H.-W., Möhring, H.-C.: Noise and vibrations in machine tools. CIRP Annals **70**(2), 611–633 (2021) https://doi.org/10.1016/j.cirp.2021.05.010

[12] Sio-Sever, A., Lopez, J.M., Asensio-Rivera, C., Vizan-Idoipe, A., Arcas, G.: Improved estimation of end-milling parameters from acoustic emission signals using a microphone array assisted by ai modelling. Sensors **22**(10), 3807 (2022) https://doi.org/10.3390/s22103807

[13] Sallese, L., Grossi, N., Scippa, A., Campatelli, G.: Investigation and correction of actual microphone response for chatter detection in milling operations. Measurement and Control **50**(2), 45–52 (2017) https://doi.org/10.1177/0020294017702285

[14] Liu, Y., Gu, X., Liu, Y., Dai, S.: Acoustic bearing fault diagnosis based on lcmv beamformer and mgigram via an mems microphone array. Transactions of the Institute of Measurement and Control, 01423312251361576 (2025) https://doi.org/10.1177/01423312251361576

[15] Abu, L., Huo, D., Liu, Z.: Integration of acoustic emission and dynamometer systems for tool condition monitoring in micro-machining brittle materials. The International Journal of Advanced Manufacturing Technology **144**(1), 1067–1086 (2026) https://doi.org/10.1007/s00170-026-17960-7

[16] Dahe, S.V., Manikandan, G.S., Jegadeeshwaran, R., Sakthivel, G., Lakshmipathi, J.: Tool condition monitoring using random forest and furia through statistical learning. Materials Today: Proceedings **46**, 1161–1166 (2021) https://doi.org/10.1016/j.matpr.2021.02.059

[17] Seeed Studio: ReSpeaker 4-Mic Array for Raspberry Pi. https://wiki.seeedstudio.com/ReSpeaker_4_Mic_Array_for_Raspberry_Pi/. Accessed 10 June 2026 (2024)

[18] Papadakis, N.M., Stavroulakis, G.E.: Review of acoustic sources alternatives to a dodecahedron speaker. Applied Sciences **9**(18), 3705 (2019) https://doi.org/10.3390/app9183705

[19] Manufacturing Automation Laboratories Inc.: Manufacturing Automation Laboratories Inc. http://www.malinc.com/. Accessed 15 Feb 2026 (2026)

[20] Li, K.-M., Lin, Y.-Y.: Tool wear classification in milling for varied cutting conditions: with emphasis on data pre-processing. The International Journal of Advanced Manufacturing Technology **125**(1), 341–355 (2023) https://doi.org/10.1007/s00170-022-10701-6

[21] Jombo, G., Zhang, Y.: Acoustic-based machine condition monitoring—methods and challenges. Eng **4**(1), 47–79 (2023) https://doi.org/10.3390/eng4010004

[22] Hussain, A., Squartini, S., Piazza, F.: Novel sub-band adaptive systems incorporating Wiener filtering for binaural speech enhancement. In: Faundez-Zanuy, M., Janer, L., Esposito, A., Satue-Villar, A., Roure, J., Espinosa-Duro, V. (eds.) Nonlinear Analyses and Algorithms for Speech Processing, pp. 318–327. Springer,

Berlin, Heidelberg (2005). https://doi.org/10.1007/11613107_28

[23] Abutalebi, H.R., Sheikhzadeh, H., Brennan, R.L., Freeman, G.H.: A hybrid subband adaptive system for speech enhancement in diffuse noise fields. IEEE Signal Processing Letters **11**(1), 44–47 (2004) https://doi.org/10.1109/LSP.2003.819348

[24] Li, J., Akagi, M.: A hybrid microphone array post-filter in a diffuse noise field. Applied Acoustics **69**(6), 546–557 (2008) https://doi.org/10.1016/j.apacoust.2007.01.005

[25] Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001) https://doi.org/10.1023/A:1010933404324

[26] Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edn. Springer Series in Statistics. Springer, New York, NY (2009). https://doi.org/10.1007/978-0-387-84858-7