 https://doi.org/10.24191/jeesr.v29i1.002

Adaptive Spectral Subtraction for HF Audio Denoising in Maritime Intelligence Applications

Mohd Azman Hanif Sulaiman, Ihsan Mohd Yassin*, Khairul Khaizi Mohd Shariff, Khairuddin Md Tap, Muhamad Haslimin Mohd Nasushardin, Aiddil Fariz Saidin, and Nooritawati Md Tahir

***Abstract*—The Royal Malaysian Navy (Tentera Laut Diraja Malaysia, TLDM) conducts continuous wideband radio-frequency surveillance across maritime operational areas using an array of high-frequency (HF) wideband antennas. These intercepts are severely degraded by channel noise, co-channel interference, and atmospheric propagation effects, which reduce signal quality and clarity. This paper presents a systematic evaluation of an adaptive spectral subtraction method for denoising these HF recordings. A corpus of 15 real-world HF recordings totaling 4,723.65 seconds was obtained from surveillance operations and processed using three parameter configurations to identify the most suitable setting. The objective was to design and implement a signal-cleaning procedure that suppresses unwanted noise while preserving the information-bearing signal. Root-mean-square (RMS) energy reduction, Pearson correlation, and log-spectral distance (LSD) were computed for each file-parameter combination, yielding 45 evaluation points. The adaptive spectral subtraction method with a noise percentile of 5% and a Savitzky-Golay smoothing window of 51 achieved the best balance between noise suppression and signal fidelity, with a mean correlation of 0.989 and a mean LSD of 3.92 dB across all noise categories.**

***Index Terms*—Adaptive spectral subtraction, denoising, high-frequency analysis, wideband antennas.**

I. INTRODUCTION

The Royal Malaysian Navy (*Tentera Laut Diraja Malaysia*, TLDM) is Malaysia's principal maritime defence force, mandated to safeguard national waters, enforce maritime sovereignty, and support broader national security objectives across the strategically vital Strait of Malacca and surrounding littoral zones. As part of its intelligence collection mandate, TLDM operates a continuous signals intelligence (SIGINT) program in which trained intercept operators monitor a broad swath of the radio frequency spectrum from 1 kHz to 512 MHz using wideband antenna arrays to capture voice communications, Morse code transmissions, and electronic data signals of intelligence value.

Utilizing advanced wideband antenna arrays, the TLDM scans a massive spectrum from 1 kHz to 512 MHz a span that traverses seven distinct International Telecommunication Union (ITU) bands and covers nearly the entire lower half of the radio frequency spectrum. This comprehensive reach allows for the capture of a wide variety of signals, ranging from tactical voice communications on walkie-talkies and amateur radio channels to more specialized electronic data, such as Morse code bursts and various digital tones.

The intercepted recordings are, however, severely corrupted by multiple potential noise sources: atmospheric noise inherent to HF propagation [1], [2], co-channel interference from adjacent frequency allocations [3], [4], receiver thermal noise [4], and impulsive interference from nearby electronic equipment [3]. These degradations render the recordings difficult for military analysts to interpret, directly impacting the intelligence value of the collected data [5].

Classical Spectral Subtraction (CSS), first proposed by Boll [6], provides the foundational framework upon which the adaptive method evaluated in this paper is constructed, and remains among the most widely deployed single-channel speech enhancement techniques owing to its low computational cost and parameter interpretability. By estimating and subtracting a noise spectrum in the Short-Time Fourier Transform (STFT) domain, CSS offers a direct and computationally tractable approach to signal recovery. However, traditional formulations specifically magnitude spectral subtraction and power spectral subtraction rely on subtracting a static average noise floor from the noisy input and are well-documented to produce "musical noise" artifacts arising from half-wave rectification of residual spectral peaks [7]. These limitations are particularly consequential under the low signal-to-noise ratio (SNR) conditions characteristic of HF maritime intercepts, where a fixed noise estimate cannot track the inherently non-stationary interference environment.

The flexibility of the CSS framework makes it a natural candidate for adaptive refinement. Incorporating SNR-adaptive over-subtraction scaling, percentile-based noise estimation, and Savitzky-Golay frequency-domain smoothing [8] can substantially improve interference suppression while preserving the signal fidelity required for operational intelligence analysis. This augmented pipeline is the subject of the present work.

This paper evaluates an SNR-adaptive spectral subtraction method that incorporates per-frame over-subtraction scaling,

This manuscript is submitted on June 16, 2026, revised on August 10, 2026, accepted on August 13, 2026 and published on October 31, 2026.

M. A. H. Sulaiman, I. M. Yassin, N. M. Tahir, and K. K. M. Shariff are with the Faculty of Electrical Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

K. M. Tap, M. H. M. Nasushardin, and A. F. Saidin are with the Royal Malaysian Navy (TLDM), Malaysia.
(email: khairuddin4081@navy.mil.my, haslimin@navy.mil.my, and aiddil.fariz@navy.mil.my).

*Corresponding author
Email address: ihsan_yassin@uitm.edu.my

1985-5389/© 2026 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

percentile-based noise estimation, and Savitzky-Golay spectral smoothing for real-world HF recordings under varying noise conditions. It identifies parameter configurations that balance noise suppression and signal distortion for TLDM surveillance applications. The contributions are: (1) a reproducible, fully automated batch-analysis pipeline applied to 15 real-world HF recordings under three parameter configurations, producing 45 evaluation logs; (2) a systematic parameter-sensitivity analysis with validated recommendations; and (3) a comprehensive set of objective metrics for direct comparison.

The remainder of this paper is organized as follows: Section II presents recent related works, followed by this research's theoretical background (Section III), methodology (Section IV), and discussion of results (Section V). Finally, concluding remarks are presented in Section VI.

II. Literature Review

This section reviews denoising research relevant to the noise challenges of HF maritime radio, including classical and adaptive spectral subtraction, multiband spectral subtraction, learning-based speech enhancement, and Savitzky-Golay filtering for signal denoising.

A. HF Maritime Radio Noise Environment

The effectiveness of high-frequency (HF) maritime surveillance is inherently tied to the quality of the signal environment. It is frequently compromised by a diverse array of noise sources and complex interference patterns. In the context of naval intelligence, these degradations are not merely technical artifacts but significant barriers to accurate signal interpretation and situational awareness [9]. Understanding the specific mechanisms of these disturbances is critical for implementing effective restoration techniques that can isolate intelligence from background corruption. Consequently, the following discussion outlines the complexities of the maritime RF landscape, where co-channel interference and non-stationary atmospheric noise present the primary challenges to signal intelligibility.

HF maritime communications are subject to co-channel interference and non-stationary atmospheric noise that severely limit signal intelligibility. Afsari *et al.* [3] demonstrated that co-channel interference is exacerbated when multiple RF transceivers operate in proximity on the same platform, a scenario directly analogous to shipboard HF installations. Recent work by Uljasz [10] examined how environmental noise power affects the probability of receiving Global Maritime Distress and Safety System (GMDSS) alarm signals, further demonstrating the operational impact of noise on maritime communications reliability.

Atmospheric noise in the HF band is characterized by pronounced non-stationarity. Johansson and Lindgren [11] conducted extensive man-made noise measurements in Sweden, reporting diurnal and spatial variability that complicates receiver design. Giesbrecht *et al.* [12] established that HF noise amplitude distributions are often non-Gaussian, with heavy tails that conventional noise models fail to capture. Li *et al.* [13] highlighted that dynamic oceanic conditions, including atmospheric variability, vessel mobility, and wave disturbances contribute to severe signal attenuation and high transmission latency in maritime networks, while Wang and Wyatt [14] showed that radio frequency interference in shared HF bands can occupy critical spectral regions essential for signal analysis, reducing data quality. These findings collectively motivate the use of robust, nonparametric noise estimation strategies such as the percentile-based approach adopted in this work, which does not rely on Gaussian assumptions.

B. Classical and Adaptive Spectral Subtraction

Spectral subtraction, introduced by Boll [6], remains the most widely deployed single-channel speech enhancement technique due to its low computational cost and parameter interpretability. An eminent limitation is the generation of musical noise isolated spectral peaks resulting from random fluctuations in the noise estimate. Subsequent work has sought to address this through adaptive over-subtraction. Hasan *et al.* [15] proposed a modified *a priori* SNR estimator that adapts the averaging factor in conventional spectral subtraction, improving speech quality across varying noise types and SNR levels. Stahl and Mowlaee [16] introduced a pitch adaptive decision-directed method that smooths the *a priori* SNR estimate along harmonic trajectories, achieving greater noise attenuation than fixed-frequency-bin approaches. Lin *et al.* [17] developed an adaptive noise estimation algorithm that updates noise estimates by smoothing noisy signal power in sub-bands, eliminating the need for explicit speech/silence detection, a constraint that is particularly restrictive in HF intercept scenarios where speech boundaries are unknown.

C. Multi Band Spectral Subtraction

Multiband spectral subtraction (MBSS) divides the noisy spectrum into frequency bands and applies independent over-subtraction in each band, enabling frequency dependent noise control. Upadhyay [18] proposed an improved MBSS method that leverages the masking properties of the human auditory system and uses an adaptive noise estimation approach controlled by a linear function of the *a posteriori* SNR. Sunitha and Prasad [19] evaluated multiple frequency-spacing strategies for MBSS and their effects on objective quality measures, finding that nonuniform band spacing aligned with auditory critical bands yield superior results. Lezzoum *et al.* [20] proposed a nonlinear multiband adaptive gain control that operates entirely in the temporal domain without requiring SNR or noise parameter estimation, reporting improved Perceptual Evaluation of Speech Quality (PESQ) scores and reduced musical noise. These multiband approaches demonstrate that frequency-dependent processing improves upon broadband techniques, a principle reflected in the per-bin noise estimation and -bin gain computation of the present work.

D. Savitzky-Golay Filtering for Signal Denoising

The Savitzky-Golay (SG) filter [21] performs local polynomial regression to smooth data while preserving signal morphology. Its application to noise suppression has been validated across diverse signal processing domains. Agarwal et

al. [22] demonstrated cascaded SG filters for Electroencephalogram (EEG) denoising, achieving superior SNR, mean-square error, and correlation metrics compared to alternative filtering approaches. Huang et al. [23] developed a low-distortion adaptive SG filter for Electrocardiogram (ECG) denoising that adapts window parameters using discrete curvature estimation. Wei and Liu [24] applied multilevel singular value decomposition combined with SG filtering for communication radar signal noise reduction, showing effective noise suppression under varying SNR conditions. Despite this broad adoption, no prior work has applied SG smoothing in the frequency domain as a post-processing step for spectral subtraction gain masks, the specific application investigated in this paper.

Furthermore, [25] established that the combination of percentile-based noise estimation with SG frequency-domain smoothing for heterogeneous signal types (voice and Morse code) appears to be novel, as confirmed by a systematic search of the literature.

Although learning-based speech-enhancement methods have demonstrated strong performance and some compact architectures support real-time CPU execution [26], their operational effectiveness depends on representative training data and validation under target-domain conditions; performance on synthetic data may not transfer reliably to real recordings [27]. The present corpus contains heterogeneous HF voice, Morse-code, and electronic signals affected by non-stationary maritime interference, for which paired clean and noisy training examples were unavailable. Adaptive spectral subtraction was therefore selected because it requires no model training, operates entirely on a CPU, provides interpretable control over noise estimation and suppression strength, and can be adjusted without retraining when operating conditions change. This selection reflects the deployment and data constraints of the present study rather than a general claim of superiority over learning-based methods. A controlled comparison with a representative low-complexity learning-based method is included in the future work plan.

Table I provides a qualitative comparison of representative denoising studies and the proposed work. The comparison distinguishes application domain, implementation requirements, principal strengths, and remaining gaps; it is not presented as a direct performance ranking because the studies use different datasets and evaluation protocols.

III. Theoretical Background

This section presents the mathematical framework underlying the adaptive spectral subtraction method. All processing is performed in the short-time Fourier transform (STFT) domain.

A. Short Time Fourier Transform

To effectively isolate intelligence from the non-stationary noise floors characteristic of HF maritime intercepts, the intercepted audio must be analyzed in both the time and frequency domains simultaneously. The short-time Fourier transform (STFT) serves as the primary analytical framework for this decomposition, segmenting continuous recordings into overlapping frames for localized spectral estimation. This transformation is essential for the adaptive spectral subtraction pipeline, as it allows for the precise manipulation of signal magnitudes while preserving the temporal structure necessary for speech intelligibility.

TABLE I. Comparison of Representative Denoising Studies with the Proposed Work

Study and approach	Domain	Data / compute profile	Principal strength	Gap or distinction
Boll [6]: classical spectral subtraction	Speech	Training-free; low computation	Foundational, transparent frequency-domain suppression	Static noise estimation and musical-noise risk under non-stationary interference
Hasan et al. [15]: adaptive a priori SNR spectral subtraction	Speech	Training-free; low-to-moderate computation	Adapts suppression strength to estimated SNR	Conventional speech focus; not evaluated on heterogeneous operational HF signals
Lin et al. [17]: adaptive subband noise estimation	Speech	Training-free; adaptive noise tracking	Updates noise estimates without requiring explicit silence	Not evaluated on mixed HF voice, Morse-code, and electronic recordings
Upadhyay [18]; Lezzoum et al. [20]: multiband enhancement	Speech / hearing protection	Training-free; additional band controls	Frequency-dependent suppression can reflect nonuniform noise	Greater tuning complexity and primarily speech-oriented evaluation
Huang et al. [23]; Wei and Liu [24]: Savitzky-Golay-assisted denoising	ECG / communication radar	Training-free smoothing	Preserves useful morphology or signal structure during smoothing	Does not apply SG smoothing to an adaptive spectral gain mask for HF audio
Schroter et al. [29]: DeepFilterNet 2	Full-band speech	Trained model; real-time CPU execution demonstrated	Low-complexity neural speech enhancement	Requires representative training and target-domain validation; not tested here on mixed HF classes
Proposed work: percentile noise estimation, SNR-adaptive subtraction, and SG gain smoothing	Operational HF voice, Morse code, and electronic signals	Training-free; CPU-based; operator-adjustable	Combines adaptive estimation and interpretable gain smoothing on authentic maritime recordings	Single collection platform; no paired clean references or direct controlled benchmark

In STFT, the input audio signal $x[n]$, sampled at $f = 44{,}100\ Hz$, is transformed into the time-frequency domain via the STFT:

$$X(k,m) = \sum_{n=0}^{N-1} x\,[n + mH]w[n]e^{-j\frac{2\pi kn}{N}} \quad (1)$$

where N is the frame size (DFT length), H is the hop length

(frame advance), $w[n]$ is the analysis window, k is the frequency bin index, and m is the frame index. The STFT yields a complex-valued matrix whose magnitude $|X(k,m)|$ and phase $\angle X(k,m)$ are separated for subsequent processing. The frame size N and hop length H jointly govern the time-frequency resolution trade-off: at $N = 1024$ samples and $f_s = 44{,}100$ Hz, each frame spans 23.2 ms with a frequency resolution of $f_s/N = 43.1$ Hz.

B. Percentile Based Noise Spectrum Estimation

Unlike classical spectral subtraction, which estimates the noise spectrum from explicitly labeled silence segments, this method employs a nonparametric percentile-based estimator that requires no voice activity detection:

$$\widehat{N}(k) = \text{Percentile}_P(|X(k,m)|), \quad m = 1, \ldots, M \tag{2}$$

where $\widehat{N}(k)$ is the estimated noise magnitude at frequency bin k, computed as the P-th percentile of the magnitude values across all M frames. This estimator exploits the statistical observation that in noisy recordings with intermittent speech, the noise-only magnitude at each frequency bin will dominate the lower tail of the magnitude distribution. The percentile parameter P directly controls the aggressiveness of the noise estimate: lower values produce conservative estimates that preserve weak signal components, whereas higher values capture more noise energy at the risk of eroding low-amplitude signal content.

C. SNR Adaptive Over Subtraction

The standard spectral subtraction formula is augmented with a frame-dependent over-subtraction factor α(m) that adapts according to the estimated instantaneous SNR:

$$G(k,m) = 1 - \alpha(m) \cdot \frac{\widehat{N}(k)}{|X(k,m)|+\epsilon} \tag{3}$$

where $G(k, m)$ is the spectral gain (suppression mask) and $\epsilon = 10^{-12}$ is a regularization constant preventing division by zero. The per-frame SNR estimate is derived from the ratio of frame signal power to estimated noise power:

$$\widehat{\text{SNR}}(m) = 10\log_{10}\left(\frac{\frac{1}{K}\sum_{k=0}^{K-1}|X(k,m)|^2}{\frac{1}{K}\sum_{k=0}^{K-1}\widehat{N}(k)^2+\epsilon}\right) \tag{4}$$

The base factor α_{base} is then scaled by a piecewise rule:

$$\alpha(m) = \begin{cases} 0.8\,\alpha_{\text{base}} & \text{if } \widehat{\text{SNR}}(m) > 10 \text{ dB} \\ 1.5\,\alpha_{\text{base}} & \text{if } \widehat{\text{SNR}}(m) < 0 \text{ dB} \\ \alpha_{\text{base}} & \text{otherwise} \end{cases} \tag{5}$$

This adaptive rule implements the principle that high SNR frames require less aggressive subtraction (the signal dominates), while low or negative SNR frames benefit from stronger suppression to attenuate the noise floor.

D. Spectral Floor

To prevent the gain from producing negative magnitudes, which would invert the phase and create artifacts, a spectral floor parameter β enforces a minimum gain value:

$$G_{\text{floor}}(k,m) = \max(G(k,m),\, \beta) \tag{6}$$

The parameter β controls the maximum attenuation depth: $\beta = 0.005$ permits suppression up to −46 dB ($20\log_{10}(0.005)$), while $\beta = 0.05$ limits it to −40 dB.

E. Savitzky-Golay Frequency Domain Smoothing

Musical noise consists of spurious tonal artifacts that manifest as faint, rapidly fluctuating tones or chirps after spectral subtraction. It is caused by the random frame-to-frame survival of isolated spectral peaks after gain attenuation. In HF denoising, this effect is exacerbated by non-stationary wideband interference, which can make the noise estimate locally inaccurate across many frequency bins.

The Savitzky-Golay (SG) smoothing filter [33] suppresses the bin-to-bin fluctuations that produce musical noise. It performs a local polynomial fit that preserves spectral trends, including formant peaks and harmonic structure, while eliminating isolated gain spikes between adjacent bins. The filter is applied to the raw spectral gain along the frequency axis of each frame:

$$\tilde{G}(\cdot,m) = \text{SG}(G(\cdot,m);\, W,\, p) \tag{7}$$

where W is the smoothing window length (number of frequency bins) and p is the polynomial order. Unlike simple moving-average smoothing, the polynomial fit preserves spectral edges, which is critical for speech quality. The complete pipeline from noisy input to enhanced output is depicted in Fig. 1.

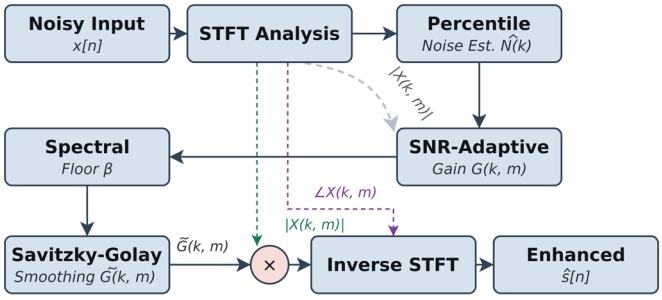

Fig. 1. Block diagram of the adaptive spectral subtraction pipeline. The noisy input undergoes STFT analysis, percentile-based noise estimation, SNR-adaptive gain computation, Savitzky-Golay spectral smoothing, and inverse STFT reconstruction.

F. Signal Reconstruction

The enhanced magnitude spectrum is combined with the original noisy phase to form the reconstructed STFT:

$$\hat{S}(k,m) = |X(k,m)| \cdot \tilde{G}(k,m) \cdot e^{j\angle X(k,m)} \tag{8}$$

where $\hat{S}(k,m)$ is the estimated clean-signal magnitude, $X(k,m)$ is the original noisy-signal magnitude, $\tilde{G}(k,m)$ is the gain used to suppress noise, and $e^{j\angle X(k,m)}$ provides temporal alignment and phase information. The time-domain signal is recovered via the inverse STFT with overlap-add synthesis. The use of the original phase is a standard assumption in magnitude-domain spectral subtraction.

G. Chunked Processing with Linear Cross Fading

For recordings whose length exceeds a duration threshold, the signal is segmented into fixed-length chunks with a short

overlap. Each chunk is processed independently, and the results are stitched using a linear cross-fade. For chunk outputs $y_{i+1}[n]$ with an overlap region of L samples, the stitched output is:

$$y[n] = \frac{w_i[n] \cdot y_i[n] + w_{i+1}[n] \cdot y_{i+1}[n]}{w_i[n] + w_{i+1}[n]} \quad (9)$$

where w_i and w_{i+1} are triangular (linear ramp) cross-fade weights.

H. Evaluation Metrics

Three complementary objective metrics were selected to evaluate the balance between noise suppression and signal preservation: root-mean-square (RMS) energy, Pearson correlation, and log-spectral distance (LSD). RMS energy reduction measures the overall attenuation achieved, while Pearson correlation indicates how well the temporal structure of the original signal is preserved after processing. LSD measures spectral distortion and, when interpreted alongside correlation, indicates whether the enhanced audio retains the fidelity required for intelligence analysis. The output-evaluation methods are described in Sections III-H-1 to III-H-3.

1) Root Mean-Square (RMS) Energy

Root-mean-square (RMS) energy serves as the most direct quantitative indicator of the overall attenuation achieved by the denoising pipeline, capturing the aggregate reduction in signal power across the entire processed recording. In the context of HF audio enhancement, where the noise floor contributes substantially to the total waveform energy, a measurable RMS reduction provides empirical confirmation that the spectral subtraction gain is actively suppressing noise-dominant frequency bins rather than leaving the signal energetically unchanged. The RMS of a discrete audio signal is defined as:

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} s\,[n]^2} \quad (10)$$

where N is the total number of samples and $s[n]$ is the audio signal amplitude at sample index n. The ratio of processed to original RMS quantifies the overall energy reduction achieved by the enhancement.

2) Pearson Correlation Coefficient

While RMS reduction confirms that noise energy has been attenuated, it provides no indication of whether the underlying signal structure has been preserved in the process. The Pearson correlation coefficient addresses this complementary concern by quantifying the linear correspondence between the original and processed waveforms at the sample level, serving as a direct measure of temporal fidelity critical in HF intercept scenarios where prosodic patterns, syllable boundaries, and keying transients must survive enhancement intact. In the absence of a clean reference, correlation against the noisy original provides a conservative estimate of waveform preservation:

$$r = \frac{\sum_n (x[n]-\bar{x})(y[n]-\bar{y})}{\sqrt{\sum_n (x[n]-\bar{x})^2 \sum_n (y[n]-\bar{y})^2}} \quad (11)$$

where $x[n]$ and $y[n]$ are the original and processed signal samples at index n, respectively, $\bar{x}$ and $\bar{y}$ are their respective sample means, and the summation is taken over all N samples. A correlation close to 1.0 indicates that the processing preserves the overall temporal structure of the signal, confirming selective noise attenuation rather than signal distortion. Conversely, values falling measurably below 1.0 indicate that the gain function has begun eroding signal-bearing spectral components, manifesting perceptually as smeared transients, attenuated consonants, or degraded Morse keying edges.

3) Log Spectral Distance (LSD)

While Pearson correlation operates at the sample level and captures waveform-domain fidelity, it is insensitive to the frequency-specific redistribution of spectral energy that spectral subtraction inevitably introduces. Log-spectral distance (LSD) addresses this gap by operating directly in the STFT domain, providing a frame-by-frame, bin-by-bin measure of how substantially the enhancement pipeline has reshaped the spectral magnitude of the signal relative to the noisy original:

$$\text{LSD} = \frac{1}{M}\sum_{m=1}^{M}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(20\log_{10}\frac{|X(k,m)|}{|Y(k,m)|}\right)^2} \quad (12)$$

where X and Y are the STFTs of the original and processed signals, respectively. LSD captures spectral distortion: lower values indicate less spectral modification, while higher values indicate more aggressive spectral reshaping. In the absence of a clean reference, LSD measures the magnitude of the spectral change rather than improvement and must be interpreted alongside correlation.

IV. Methodology

This section describes implementation decisions, parameter selections, and experimental design. The adaptive spectral subtraction pipeline incorporates four algorithmic components: percentile-based noise estimation, SNR-adaptive over-subtraction, spectral-floor enforcement, and Savitzky-Golay frequency-domain smoothing. Their mathematical formulations were presented in the preceding section.

A. Computational Environment

All experiments were conducted on a single workstation running Windows 11 with an AMD Ryzen Threadripper 3990X 64-core processor and 64 GB of DDR4 RAM. Processing was entirely CPU-based, with no GPU acceleration. The high CPU core count and memory bandwidth facilitate parallel input/output and vectorized NumPy computation. The core STFT loop uses NumPy broadcasting and eliminates per-frame Python loops, reducing processing time by approximately one order of magnitude compared with a frame-by-frame implementation.

B. Experimental Setup

Processing proceeds as follows for each of the 15 input files and 3 parameter configurations (45 total evaluations). The processing flow is illustrated in Fig. 2.

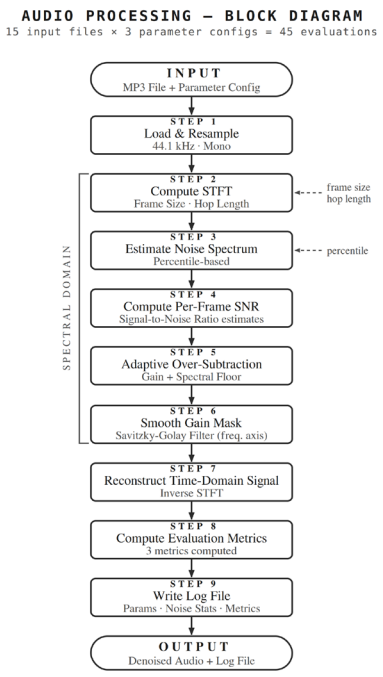

Fig. 2. Audio Processing Flow Chart

C. Processing Pipeline

The complete pipeline was implemented in Python 3.14 using the *librosa (v0.10)* library for STFT computation, *NumPy* for vectorized array operations, *SciPy* for Savitzky–Golay filtering, and *matplotlib* for visualization. All STFT computations use the default Hann window.

The signal-processing workflow begins by loading each MP3 file and resampling it to 44.1 kHz mono to provide consistent algorithm input. The Short-Time Fourier Transform (STFT) is then computed using the specified frame size and hop length. The noise spectrum is estimated using the selected percentile, after which per-frame SNR values are calculated.

An adaptive over-subtraction gain is applied to suppress unwanted noise, with a spectral floor used to limit musical-noise artifacts associated with aggressive filtering. The resulting gain mask is smoothed along the frequency axis using a Savitzky-Golay filter. The time-domain signal is then reconstructed by inverse STFT, and the three-evaluation metrics are calculated.

For recordings exceeding 120 seconds, steps 2 to 7 are applied independently to each 60-second chunk, and the results are stitched using the linear cross-fade procedure.

D. Dataset Composition

The dataset comprises 15 HF audio recordings provided by the Royal Malaysian Navy (TLDM) intelligence division. Table II presents the full dataset inventory. The corpus contains 4,723.65 seconds (approximately 78.7 minutes) of audio, with individual recordings ranging from 21.2 seconds to 25.1 minutes. The recordings were originally encoded in MP3 format at various bit rates and were resampled to 44.1 kHz during loading to ensure consistent STFT analysis. Domain experts manually classified the recordings into five categories based on noise level and signal type.

TABLE II. DATASET COMPOSITION BY NOISE CATEGORY

Category	Files	Duration (s)	Signal Type
High noise vocal	8	2 264.1	Voice
Low noise Morse	2	1 225.0	CW/Morse
Low noise vocal	2	1 040.9	Voice
Medium electronic Morse	1	84.4	CW/Electronic
Medium noise vocal	2	109.3	Voice
Total	**15**	**4 723.7**	—

The "high noise vocal" category contains eight recordings with the most severe degradation, dominated by broadband atmospheric noise and co-channel voice interference. The "low noise" categories exhibit comparatively cleaner signals with narrowband interference patterns. The Morse/electronic categories contain continuous-wave (CW) signals and data tones rather than speech, presenting a distinct signal structure.

E. STFT Configuration

A frame size of $N = 1{,}024$ samples at 44.1 kHz corresponds to a frame duration of 23.2 ms and a frequency resolution of 43.1 Hz. This balance is appropriate for both the temporal dynamics of speech signals (which require sub-50 ms resolution to capture consonant transitions) and the frequency resolution needed to distinguish narrowband interference from speech harmonics. A hop length of $H = 256$ samples (5.8 ms, 75% overlap) ensures adequate temporal oversampling, producing smooth gain transitions between frames and reducing blocking artifacts in the reconstructed signal. All STFT computations use a Hann analysis window.

F. Noise Percentile Selection

The percentile values were configured to span the conservative and moderate spectrum of noise estimation:

- $P = 5$(conservative): Estimates only the lowest 5% of magnitudes as noise. This preserves weak signal components that might otherwise be misidentified as noise, at the risk of underestimating the true noise floor.
- $P = 10$ (moderate): Captures more of the noise energy, yielding stronger suppression but increasing the risk of removing low-energy speech segments such as unvoiced consonants.

The rationale for these specific values derives from the observation that in intelligence intercept recordings, the duty cycle of the target signal varies widely. A percentile of 5 is conservative enough for recordings with near-continuous speech, while a higher percentile tier accommodates sparse intercept signals that are heavily interspersed with long, noise-only propagation periods.

G. Over-Subtraction Factor

The base over-subtraction factor $\alpha_{base} = 1.5$ was selected as the primary test value because it represents a mild degree of over subtraction. The piecewise SNR adaptive rule in (Eq. 5) scales this base value by 0.8 for high SNR frames (> 10 dB) and by 1.5 for low SNR frames (< 0 dB). The threshold values of 0

dB and 10 dB were chosen based on the observation that HF recordings span this range: segments where the target speaker is transmitting at close range may achieve SNRs above 10 dB, while segments with only background interference typically fall below 0 dB.

H. Spectral Floor Selection

The spectral floor values were included in the design grid to control the tradeoff between noise suppression depth and residual noise texture. Only $\beta = 0.005$ was evaluated in the final three configuration experiment. This value allows attenuation to −46 dB, providing deep suppression suitable when the noise is significantly louder than any residual artifacts, effectively silencing noise dominated frequency bins.

I. Savitzky-Golay Smoothing Window

To mitigate the "musical noise" artifacts inherent in spectral subtraction, an SG filter was employed to smooth the gain mask and ensure spectral continuity. The smoothing process is defined by a fixed polynomial order of $p = 3$, with two specific window lengths evaluated to identify the most effective configuration for maritime signal recovery. These window settings allow for a systematic comparison between localized artifact suppression and broader spectral smoothing suited for the high-volatility noise environments of HF intercepts. Two window lengths were tested:

- $W = 31$: Spans 31 frequency bins ($31 \times 43.1 = 1{,}336\ Hz$ at $N = 1{,}024$), smoothing over roughly one octave of spectral content. This is sufficient to suppress isolated musical noise peaks while preserving broader formant structure.
- $W = 51$: Spans 51 frequency bins ($51 \times 43.1 = 2{,}198\ Hz$), providing stronger smoothing that more effectively suppresses musical noise at the cost of slightly blurring closely spaced spectral features. The wider window was hypothesized to be advantageous for high noise recordings where the spectral gain mask exhibits more erratic fluctuations.

J. Chunked Processing for Long Recordings

Several recordings in the dataset exceed two minutes in duration, with one recording reaching 25.1 minutes (1,507.3 seconds, approximately 66.5 million samples at 44.1 kHz). Direct STFT computation of such signals would require storing magnitude matrices of dimensions $513 \times 259{,}000$, which is both memory-intensive and computationally prohibitive. Recordings exceeding a threshold of 120 seconds are therefore segmented into 60-second chunks with 0.5-second overlap and stitched using the linear cross-fade of Eq. (9). The 60-second chunk duration was selected to be long enough to capture sufficient statistical context for noise estimation (each chunk yields approximately 10 to 300 STFT frames).

K. Configuration Selection

For computational tractability, the grid was reduced to three representative configurations based on the most influential parameters: frame size N, hop length H, base over-subtraction factor αbase, noise percentile P, and smoothing window W. The configurations span conservative to aggressive settings: C1 (P = 5, W = 31), C2 (P = 5, W = 51), and C3 (P = 10, W = 31). All use the same STFT parameters (N = 1024, H = 256), base over-subtraction factor (αbase = 1.5), and spectral floor (β = 0.005), as shown in Table III.

TABLE III. EVALUATED PARAMETER CONFIGURATIONS

Config	N	H	α_{base}	P	W
C1	1024	256	1.5	5	31
C2	1024	256	1.5	5	51
C3	1024	256	1.5	10	31

V. RESULTS AND DISCUSSION

This section presents and analyzes the adaptive spectral subtraction results across 45 evaluation points (15 files × 3 configurations). The settings are conservative (C1: P = 5, W = 31), smooth (C2: P = 5, W = 51), and moderate (C3: P = 10, W = 31). Fig. 3 shows representative spectrograms for a high-noise vocal recording under each configuration.

A. Aggregate Performance and the Noise Fidelity Trade

Table IV summarizes the mean evaluation metrics across all 15 files. The noise-fidelity trade-off is markedly nonlinear. Doubling the noise percentile from 5 to 10 increases mean LSD by 59%, from 3.98 to 6.32 dB, and reduces mean correlation from 0.990 to 0.977.

Although the absolute correlation change appears small, the original phase is preserved within the spectral subtraction framework. A decline of 0.013 therefore corresponds to a perceptually meaningful increase in processing artifacts that could affect signal clarity.

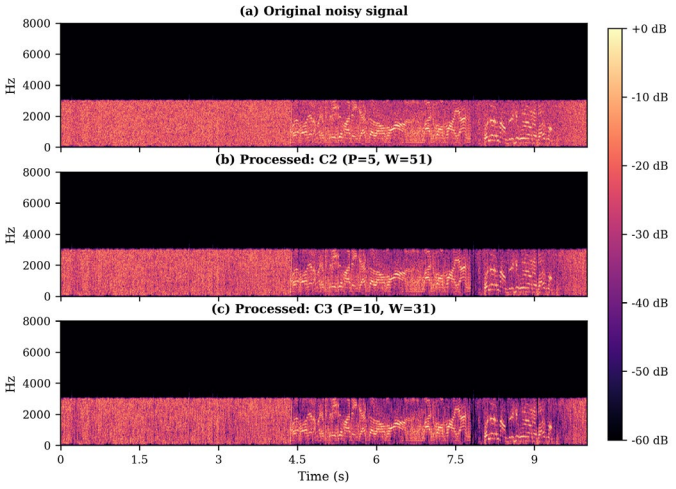

Fig. 3. Spectrogram comparison for a representative high-noise vocal recording (ori (13).mp3). Top: original noisy signal. Middle: processed with C2 (P =5, W =51). Bottom: processed with C3 (P =10, W =31). The conservative configuration preserves speech formant structure while reducing the broadband noise floor; the aggressive configuration achieves deeper suppression but visibly attenuates weak speech segments.

This nonlinear sensitivity suggests that the noise magnitude distribution in HF recordings does not exhibit a smooth transition between noise-dominated and signal-dominated percentiles. Rather, the data imply that the 5th–10th percentile range sits precisely at the boundary where noise floor estimates begin to encroach on the tails of the signal magnitude distribution, causing the subtraction to progressively erode weak but information-bearing spectral components such as unvoiced consonants and low-amplitude Morse keying transients, as illustrated in Fig. 3.

TABLE IV. MEAN METRICS ACROSS ALL 15 FILES BY CONFIGURATION

Config	P	W	$\bar{r}$	$\overline{\text{LSD}}$ (dB)	RMS Reduction
C1	5	31	0.9899	4.031	27.4%
C2	5	51	0.9893	3.922	29.7%
C3	10	31	0.9771	6.319	37.2%

The best-performing configuration overall is C2 (P = 5, W = 51), which achieves the lowest mean LSD of 3.922 dB while maintaining a mean correlation of 0.9893. It also produces a 29.7% mean RMS reduction, indicating effective noise-energy removal without the spectral-distortion penalty associated with the higher percentile setting.

Increasing the smoothing window from 31 to 51 bins improves LSD by approximately 0.1 dB across the corpus without degrading waveform-level correlation. This result confirms that broader Savitzky-Golay smoothing reduces the bin-to-bin fluctuations that contribute to musical noise and inflate LSD.

Practically, an appropriately wide frequency-domain polynomial smoothing window can mitigate a well-known limitation of spectral subtraction [1] without requiring more computationally expensive spectral-floor reshaping or Wiener post-filtering.

B. Performance Across Noise Regimes

Table V presents performance by the five manually categorized noise categories, revealing how the algorithm adapts to fundamentally different signal and noise characteristics.

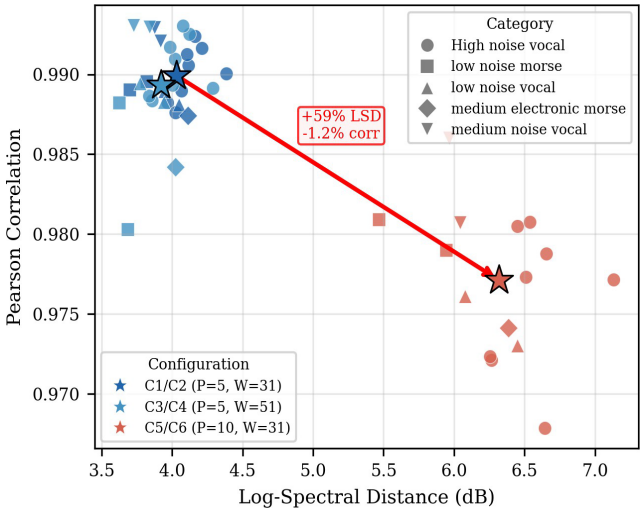

Fig. 4. Noise-fidelity trade-offs across the three evaluated configurations. Each marker represents one of the 15 files, and marker shape indicates the noise category. Doubling the noise percentile from 5 to 10 increases mean LSD by 59% while reducing mean correlation by only 1.3%, indicating a disproportionate distortion cost.

The category-level analysis shows that the method remains robust across heterogeneous signal conditions, with correlation above 0.974 for every category-configuration combination. This is notable because standard speech-enhancement benchmarks generally use controlled signal types at known SNRs, whereas the present corpus includes vocal speech, continuous-wave Morse code, and electronic data tones under varying noise intensities.

The strong performance of a single configuration (P = 5, W = 51) across all five categories suggests that the SNR-adaptive over-subtraction mechanism in Equation (5) functions as an implicit signal classifier. It scales suppression according to the instantaneous characteristics of each frame rather than relying on an externally imposed signal model, as illustrated in Fig. 4.

The medium-noise vocal category achieves the highest correlation in the study (r = 0.9930 with C2) and the lowest LSD (3.785 dB). This result reflects its SNR operating point. When the initial noise level is moderate, most frames have SNR estimates between 0 and 10 dB, where $\alpha(m) = \alpha_{\text{base}}$ and no scaling is applied.

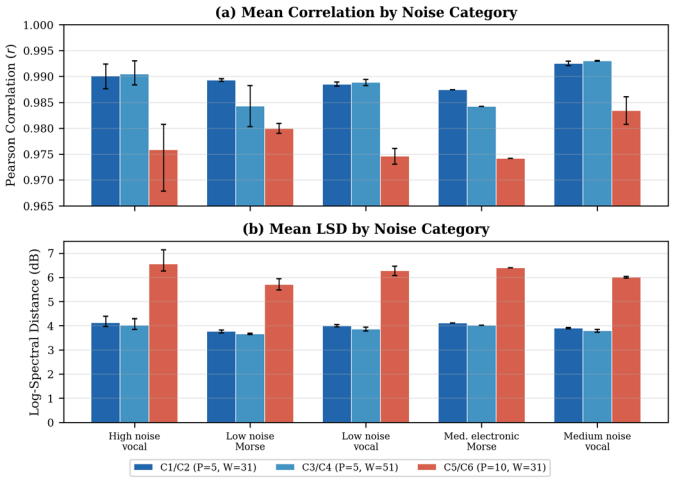

Fig. 5. Mean correlation (top) and log spectral distance (bottom) by noise category for each effective configuration. Error bars indicate the range across files within each category. The medium noise vocal category achieves the highest correlation under all settings, while the high noise vocal category incurs the greatest LSD penalty when moving from P =5 to P =10.

By contrast, in the high-noise vocal category, a greater proportion of frames have estimated SNR values below 0 dB, triggering the 1.5αbase amplification rule. This more aggressive subtraction removes noise energy, with RMS reductions of up to 37.2%, but increases LSD to 6.556 dB under P = 10, the highest value observed. Although the piecewise SNR adaptation prevents over-subtraction in high-SNR frames, a gradual interpolation in the 0-to-10 dB transition region may

produce a smoother noise-fidelity curve than the current step function, as illustrated in Fig. 5.

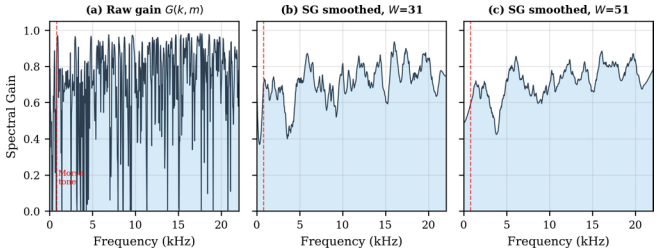

Fig. 6. Spectral gain mask for a single frame of a low-noise Morse recording. Left: raw gain G(k, m) before smoothing, exhibiting erratic bin to bin fluctuations (musical noise). Centre: smoothed gain G˜(k, m) with W =31. Right: smoothed gain with W =51. The wider window produces a smoother mask at the cost of slightly broadening the suppression around the Morse tone frequency (marked by the dashed vertical line).

The low noise Morse category is particularly instructive for understanding how broadband noise estimation interacts with narrowband signal content. Morse-code signals concentrate their energy in a narrow frequency band (typically a single tone near 600–1000 Hz), whereas the percentile-based noise estimator operates independently on each frequency bin across the full spectrum. In the signal-bearing bins, the 5th percentile magnitude is elevated by the presence of the Morse tone, producing an inflated noise estimate for those specific bins. This explains the slightly lower correlations observed for Morse recordings under the $W = 51$ smoothing configuration ($r = 0.9843$ versus $r = 0.9893$ $r = 0.9893$ with $W = 31$): the wider smoothing window, which spans approximately 2.2 kHz, smooths the gain mask across the signal-bearing bins and adjacent noise-only bins, partially attenuating the Morse tone energy, as shown in Fig. 6. This observation has practical significance for intelligence signal processing: when the signal of interest is known to be narrowband (e.g., Morse, single-sideband voice, or data tones), a narrower smoothing window should be preferred to avoid spectral leakage of the suppression gain into the signal passband.

TABLE V. MEAN METRICS BY NOISE CATEGORY AND CONFIGURATION

Category	Config	Correlation (r)	LSD (dB)	RMS Original	RMS Processed	RMS Reduction
High noise vocal	C1 (P =5, W =31)	0.9901	4.130	0.1381	0.1006	27.2%
	C2 (P =5, W =51)	0.9905	4.025	0.1381	0.0986	28.6%
	C3 (P =10, W =31)	0.9759	6.556	0.1381	0.0867	37.2%
Low noise Morse	C1 (P =5, W =31)	0.9893	3.760	0.2630	0.1881	28.5%
	C2 (P =5, W =51)	0.9843	3.656	0.2630	0.1771	32.7%
	C3 (P =10, W =31)	0.9800	5.706	0.2630	0.1622	38.3%
Low noise vocal	C1 (P =5, W =31)	0.9885	3.998	0.1600	0.1203	24.8%
	C2 (P =5, W =51)	0.9889	3.859	0.1600	0.1168	27.0%
	C3 (P =10, W =31)	0.9746	6.264	0.1600	0.1055	34.1%
Med. electronic Morse	C1 (P =5, W =31)	0.9874	4.114	0.1762	0.1205	31.6%
	C2 (P =5, W =51)	0.9842	4.024	0.1762	0.1146	35.0%
	C3 (P =10, W =31)	0.9741	6.387	0.1762	0.1020	42.1%
Medium noise vocal	C1 (P =5, W =31)	0.9925	3.898	0.1304	0.0964	26.1%
	C2 (P =5, W =51)	0.9930	3.785	0.1304	0.0941	27.8%
	C3 (P =10, W =31)	0.9834	6.004	0.1304	0.0839	35.7%

The medium electronic Morse category, represented by one 84.4-second recording, exhibits the largest RMS reduction under the aggressive configuration (42.1%). This suggests that the recording contains substantial broadband noise alongside the narrowband electronic signal. Correlation nevertheless remains at 0.9741 under C3, indicating preservation of the core signal structure. The deterministic electronic Morse waveform, with its sharp on-off keying, is more resilient to spectral subtraction than stochastic speech because its energy is concentrated in time-frequency regions with high local SNR that the adaptive algorithm identifies and protects.

C. Energy Removal and Its Relationship to Noise Regime

Across all 45 evaluations, the processed signals uniformly exhibit lower RMS energy than the originals, with category level reductions ranging from 24.8% (low noise vocal, conservative) to 42.1% (medium electronic Morse, aggressive). This monotonic energy reduction is expected, because spectral subtraction is a gain attenuation method: the spectral gain $G(k, m)$ is always ≤ 1, and the spectral floor β ensures that it never reaches zero. The degree of reduction is strongly correlated with two factors: the noise percentile P and the original noise level of the recording, as illustrated in Fig. 7.

The noise-percentile effect is quantifiable: increasing P from 5 to 10 raises the mean RMS reduction by approximately 10 percentage points across all categories. This relationship is nearly linear, suggesting that each percentile increment increases the estimated noise floor and the magnitude of the subtracted component.

The relationship between energy reduction and signal quality is not linear. The additional 10 percentage points of RMS reduction are accompanied by a 2.3 dB increase in LSD and a 0.013 decrease in correlation, representing a disproportionate quality penalty. This asymmetry suggests that the 5th percentile captures noise that is readily separable from the signal, whereas the 5th-to-10th percentile range contains both noise and low-amplitude signal energy that cannot be cleanly separated by magnitude-only processing.

An additional category-level observation is that the low-noise Morse recordings have the highest original RMS value in the dataset (0.2630), despite their low-noise classification.

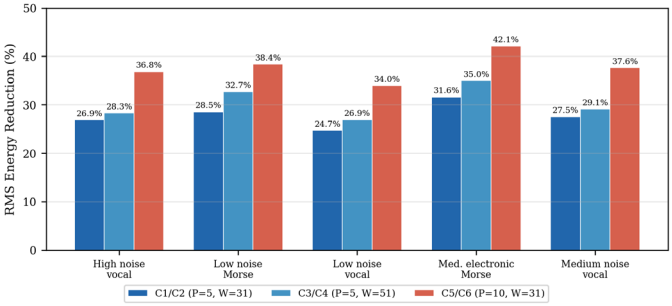

Fig. 7. Mean RMS energy reduction (%) by noise category and configuration. The aggressive configuration (P =10) consistently removes 8 to 12 percentage points more energy than the conservative setting (P =5), with the largest absolute reduction observed in the medium electronic Morse category (42.1%).

The Morse signal itself contributes substantial energy because the continuous-wave tone is a high-amplitude, near-constant signal that dominates the overall RMS measurement. The low-noise classification refers to perceptual recording quality as assessed by domain experts, not to total energy. This distinction highlights a limitation of RMS as a noise metric: it conflates signal and noise energy and is therefore an unreliable proxy for perceptual noise in signals with high duty cycles.

The Morse signal itself contributes substantial energy: the continuous wave tone is a high amplitude, near constant signal that dominates the overall RMS measurement. The “low noise” classification refers to the perceptual quality of the recording as assessed by domain experts, not to the total energy content. This highlights a limitation of RMS as a noise assessment metric: it conflates signal energy with noise energy and is therefore an unreliable proxy for perceptual noise levels in signals with high duty cycles.

D. Per File Variability and Outlier Analysis

Examination of the individual file level results reveals important variability concealed by category-level averages, as shown in Fig. 8. The longest recording in the corpus, ori (13).mp3 (1 507 seconds, high noise vocal), achieves a correlation of 0.9893 under C2 but drops to 0.9679 under C3, the lowest file level correlation observed in the entire experiment.

This recording was processed in 26 chunks of 60 seconds. Its lower correlation under the aggressive setting likely reflects accumulated noise-estimation errors because each chunk independently estimates its noise floor. When a chunk contains transient noise spikes, the 10th percentile may exceed the true floor and cause localized over-subtraction. For very long recordings, a global noise estimate computed once for the entire file may therefore produce more stable results than per-chunk estimation, although it would be less adaptable to non-stationary noise.

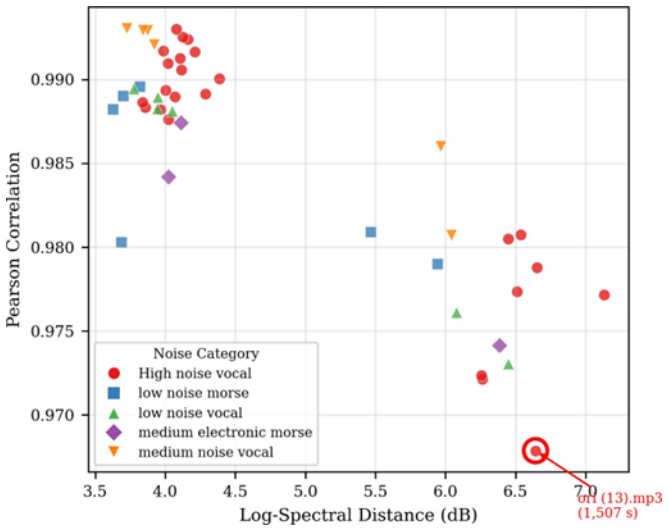

Fig. 8. Per-file correlation versus LSD for all 45 evaluation points (15 files × 3 configurations). Each point is labelled by file index; color indicates noise category. The longest recording (ori (13).mp3, 1 507 s) is highlighted as an outlier under the P =10 setting, exhibiting the lowest correlation in the dataset.

At the opposite extreme, the short recording ori (99).mp3 (30.7 seconds, high noise vocal) achieves the highest correlation of any high noise file ($r = 0.9930$ under C2). Its brevity means it is processed as a single chunk, providing the noise estimator with a statistically coherent block of frames. This contrast between long and short recordings illustrates an estimation trade off longer recordings provide more frames for statistical estimation but are more likely to contain non-stationary noise profiles that violate the estimator’s implicit stationarity assumption.

E. The Dominance of Noise Percentile as a Control Parameter

As shown in Fig. 9, the parameter sensitivity analysis yields a clear hierarchy: the noise percentile P is the dominant parameter, the smoothing window W has a secondary but consistent effect, and all other parameters (frame size, hop length, spectral floor, adaptation rate) have no measurable impact within the tested configurations because they were held constant across the three evaluated settings.

The dominance of P is quantifiable. Between C1 and C3, which differ only in P (5 versus 10), mean LSD increases by a factor of 1.57 and mean correlation decreases by 1.3%. Between C1 and C2, which differ only in W (31 versus 51), mean LSD decreases by a modest 2.7% while correlation remains statistically unchanged (0.9899 versus 0.9893). The ratio of these effect sizes is approximately 21:1 in favor of P over W demonstrates that noise percentile selection is by far the most critical design decision in deploying adaptive spectral subtraction to HF recordings.

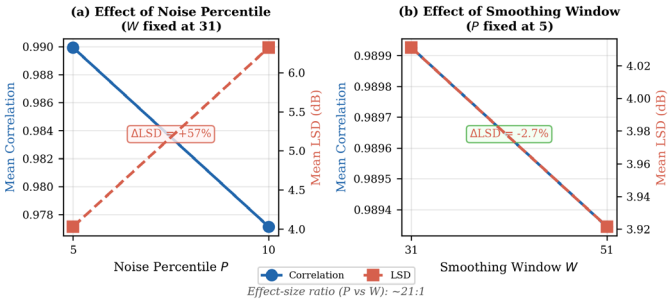

Fig. 9. Parameter sensitivity analysis. Left: effect of noise percentile P on mean LSD and correlation (with W fixed at 31). Right: effect of smoothing window W on mean LSD and correlation (with P fixed at 5). The noise percentile exhibits approximately 21× the effect size of the smoothing window, confirming its dominance as the primary control parameter.

This finding carries an important practical recommendation: when deploying spectral subtraction in the field with limited opportunity for parameter tuning, the noise percentile should be the single parameter exposed to the operator, with all other parameters fixed at the default values identified in this study ($N = 1{,}024, H = 256, \alpha_{base} = 1.5, \beta = 0.005, W = 51$). The operator can then adjust P upward from the conservative default of 5 only when noise is overwhelming, accepting the associated increase in spectral distortion as a conscious trade off.

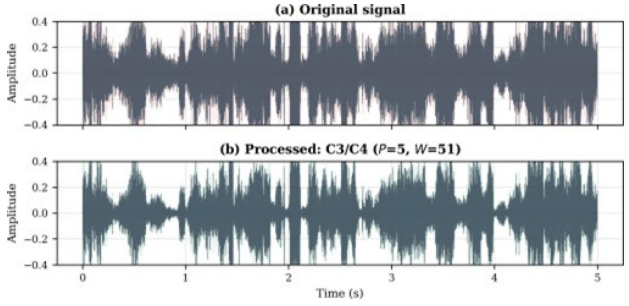

Fig. 10. Time-domain waveform comparison for a 5-second excerpt of a medium-noise vocal recording. Top: original signal with visible noise floor. Bottom: processed signal (C2, P =5, W =51) showing reduced noise floor with preserved speech envelope structure. The temporal alignment confirms the high correlation (r = 0.993) reported in Table IV.

The time-domain effect of the best-performing configuration is shown in Fig. 10. Although secondary, increasing W from 31 to 51 consistently reduces LSD by 0.1 dB without a detectable quality penalty. Further increases may provide additional benefit until formant resolution begins to degrade, which warrants investigation in future work.

F. Implications for Intelligence Signal Processing

Several aspects of these findings are directly relevant to the operational context of maritime intelligence. First, the consistently high correlations ($r > 0.97$ across all configurations) indicate that adaptive spectral subtraction preserves the waveform envelope of the original recording with high fidelity, as demonstrated in Fig. 10. For intelligence analysts, this means that temporal cues critical to speaker identification and speech segmentation, such as prosodic patterns, pause structure, and syllable timing are retained in the processed output, even when aggressive noise percentiles are used.

Second, the method's ability to handle heterogeneous signal types without modification is operationally valuable. In real-world intelligence collection, the content of a recording is often unknown prior to processing. A method that performs robustly on both speech and electronic signals (Morse, data tones) without requiring signal-type specific configuration eliminates the need for a classification stage before enhancement, simplifying the processing workflow.

Third, the RMS reduction metric, while imperfect as a quality indicator, provides a useful proxy for the "visual" improvement that analysts observe when comparing waveform displays and spectrograms of original versus processed recordings. A 25% to 40% reduction in RMS translates to a noticeably cleaner spectrogram with reduced noise floor, facilitating both auditory analysis and automated transcription by downstream speech recognition systems.

VI. Conclusion

This paper presented a comprehensive evaluation of adaptive spectral subtraction for enhancing HF intelligence recordings from TLDM intercepts. The batch-analysis pipeline processed 15 recordings under three parameter configurations and produced 45 evaluation logs covering high-noise voice, low-noise Morse, and mixed electronic signals.

The principal findings are: (1) adaptive spectral subtraction with conservative parameters ($P = 5, W = 51, \alpha_{base} = 1.5$) achieves a mean correlation of 0.989 and a mean LSD of 3.92 dB, representing effective noise reduction with minimal signal distortion; (2) the noise percentile parameter is the dominant factor governing the noise-distortion trade-off, with a change from $P = 5$ to $P = 10$ increasing mean LSD by 59%; and (3) Savitzky-Golay spectral smoothing consistently reduces LSD by approximately 0.1 dB without degrading correlation, confirming its value as a musical noise suppression tool.

These results provide actionable guidance for the deployment of spectral subtraction methods in maritime intelligence signal processing and establish a reproducible benchmark for future comparison with alternative enhancement techniques.

VII. Limitations and Future Work

Several limitations should be acknowledged. First, the evaluation relies on proxy metrics, namely correlation and LSD measured against the noisy original, because a clean reference is unavailable. High correlation with the noisy signal indicates waveform preservation but does not directly confirm noise removal; a method could obtain perfect correlation by passing the noisy signal through unchanged. RMS reduction partially addresses this concern by confirming that energy is removed, but optimal evaluation would require clean references or perceptual listening tests.

Second, the corpus comprises 15 operational recordings

collected using a single platform. The sample provides proof-of-concept evidence across 4,723.65 seconds (approximately 78.7 minutes) of authentic material and five manually identified signal and noise categories, but it should not be interpreted as a population-representative HF benchmark.

The 45 processing evaluations arise from three parameter configurations applied to the same 15 recordings and do not constitute 45 independent recordings. Differences in receiver design, antenna configuration, geographic region, propagation conditions, observation period, and interference environment may affect the recorded noise statistics and the relative effectiveness of individual parameter settings.

Consequently, the results establish feasibility under the evaluated operational conditions but not cross-platform generalizability. Future validation should include independently collected recordings from multiple platforms, receivers, locations, propagation conditions, and time periods, with predefined signal categorization and, where feasible, clean or expert-validated references.

Third, only three parameter combinations were tested for computational tractability, and they represented relatively conservative settings. Interactions among frame size, hop length, and spectral floor therefore warrant further investigation. Future work should conduct a full factorial experiment or use Bayesian optimization to explore the high-dimensional parameter space more efficiently.

Direct experimental benchmarking against classical spectral subtraction, Wiener filtering, multiband spectral subtraction, and learning-based enhancement methods was not undertaken. The operational recordings lack time-aligned clean references; therefore, RMS reduction, Pearson correlation, and LSD relative to the noisy input describe attenuation, waveform preservation, and spectral modification rather than absolute enhancement quality.

A ranking based only on these proxy measures could be misleading because the corpus includes voice, Morse code, and electronic signals with different perceptual and structural characteristics. The absence of a controlled baseline comparison is therefore acknowledged as a limitation.

Future work will establish a controlled comparative benchmark by combining clean or expert-validated HF voice, Morse-code, and electronic signals with measured maritime noise at known SNRs. Classical spectral subtraction, Wiener filtering, multiband methods, the proposed adaptive approach, and a representative low-complexity learning-based method will be evaluated under identical sampling and processing conditions.

Voice evaluation will include appropriate full-reference intelligibility and speech-quality measures together with structured listening tests. Morse and electronic recordings will use signal-specific measures such as tone-detection accuracy, keying-envelope preservation, and character-decoding error.

Processing time, real-time factors, CPU utilization, and peak memory use will also be reported. Recordings from additional collection platforms and propagation conditions will be included to assess generalizability.

ACKNOWLEDGMENT

The authors would like to thank Universiti Teknologi MARA (UiTM) for infrastructure and equipment support, and TLDM for the HF recording dataset for this research.

REFERENCES

[1] International Telecommunication Union (ITU-R), "Radio noise," Recommendation ITU-R P.372-17, Geneva, Switzerland, Aug. 2024. [Online]. Available: https://www.itu.int/rec/R-REC-P.372

[2] B. A. Witvliet, R. M. Alsina-Pagès, D. Altadill, E. van Maanen, and G. J. Laanstra, "Separation of ambient radio noise and radio signals received via ionospheric propagation," *Atmosphere*, vol. 14, no. 3, Art. no. 529, Mar. 2023, doi: 10.3390/atmos14030529.

[3] A. Afsari, B. Van Veen, and N. Behdad, "Co-site interference mitigation in full-duplex HF antenna arrays employing electronically reconfigurable impedance matching and decoupling networks," in *Proc. USNC-URSI Nat. Radio Sci. Meeting*, 2025.

[4] M. Yang, X. Zhang, W. Zhang, and C. Wang, "Research and simulation experiment on HF communication noise reduction method based on wavelet transform," in *Proc. 4th Int. Conf. Comput. Commun. Artif. Intell. (CCAI)*, 2024.

[5] M. D. Keller, J. M. Ziriax, W. Barns, B. Sheffield, and D. Brungart, "Performance in noise: Impact of reduced speech intelligibility on Sailor performance in a Navy command and control environment," *Hear. Res.*, vol. 349, pp. 58–67, 2017, doi: 10.1016/j.heares.2016.10.007.

[6] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, Apr. 1979, doi: 10.1109/TASSP.1979.1163209.

[7] C. Yang, "The application of spectral subtraction and Kalman filtering for audio noise reduction," in *Proc. 6th Int. Conf. Intell. Control, Meas. Signal Process. (ICMSP)*, 2024.

[8] M. J. Niedźwiecki, M. Ciołek, A. Gańcza, and P. Kaczmarek, "Application of regularized Savitzky–Golay filters to identification of time-varying systems," *Automatica*, vol. 133, Art. no. 109865, 2021, doi: 10.1016/j.automatica.2021.109865.

[9] X. Li, Y. Wang, W. Sun, and Y. Wu, "A novel vessel detection method for SHFSWR using ocean dynamic environmental parameters," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2025.

[10] B. Uljasz, "The influence of environmental noise power on the probability of receiving GMDSS alarm signals," *TransNav, Int. J. Mar. Navig. Saf. Sea Transp.*, vol. 19, no. 3, 2025, doi: 10.12716/1001.19.03.33.

[11] B. Johansson and T. Lindgren, "Man-made noise measurement in Sweden: Results, lessons learned and concept for future measurement," in *Proc. IEEE Int. Symp. Electromagn. Compat. (EMC)*, 2016, pp. 868–873.

[12] J. Giesbrecht, R. Clarke, and D. Abbott, "An empirical study of the probability density function of HF noise," *Fluct. Noise Lett.*, vol. 6, no. 2, pp. L117–L125, 2006.

[13] S. Li, X. Wang, D. Li, and L. Li, "MDJSCC: Multiscale feature learning-based semantic communications for maritime vessel networks," *IEEE Sensors J.*, 2025.

[14] W. Wang and L. R. Wyatt, "Radio frequency interference cancellation for sea-state remote sensing by high-frequency radar," *IET Radar, Sonar Navig.*, vol. 5, no. 4, pp. 405–415, 2011.

[15] Md. K. Hasan, S. Salahuddin, and M. R. Khan, "A modified a priori SNR for speech enhancement using spectral subtraction rules," *IEEE Signal Process. Lett.*, vol. 11, no. 4, pp. 450–453, Apr. 2004.

[16] J. Stahl and P. Mowlaee, "A simple and effective framework for a priori SNR estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2018, pp. 5474–5478.

[17] L. Lin, W. H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," *Electron. Lett.*, vol. 39, no. 9, pp. 754–755, 2003, doi: 10.1049/el:20030480.

[18] N. Upadhyay, "An improved multi-band speech enhancement utilizing masking properties of human hearing system," in *Proc. 5th Int. Symp. Electron. Syst. Design (ISED)*, 2014, pp. 42–46.

[19] P. Sunitha and K. S. Prasad, "Multi band spectral subtraction for speech enhancement with different frequency spacing methods and their effect on objective quality measures," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 6, pp. 18–27, 2019.

[20] N. Lezzoum, G. Gagnon, and J. Voix, "Noise reduction of speech signals using time-varying and multi-band adaptive gain control for smart digital hearing protectors," *Appl. Acoust.*, vol. 109, pp. 37–43, 2016.

[21] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964, doi: 10.1021/ac60214a047.

[22] S. Agarwal, A. Rani, V. Singh, and A. P. Mittal, "EEG signal enhancement using cascaded S-Golay filter," *Biomed. Signal Process. Control*, vol. 36, pp. 194–204, 2017.

[23] H. Huang, S. Hu, and Y. Sun, "A discrete curvature estimation based low-distortion adaptive Savitzky–Golay filter for ECG denoising," *Sensors*, vol. 19, no. 7, Art. no. 1617, 2019, doi: 10.3390/s19071617.

[24] X. Wei and S. Liu, "Noise reduction method of communication radar signal based on multilevel singular value decomposition and Savitzky–Golay," *J. Wuhan Univ. Technol. (Transp. Sci. Eng.)*, vol. 44, no. 3, pp. 537–541, 2020.

[25] M. A. Selver, M. Secmen, and E. Y. Zoral, "Savitzky–Golay filtering for scattered signal de-noising," *J. Phys., Conf. Ser.*, vol. 1065, Art. no. 212018, 2018.

[26] C. K. A. Reddy *et al.*, "The INTERSPEECH 2020 Deep Noise Suppression Challenge: Datasets, subjective testing framework, and challenge results," in *Proc. Interspeech 2020*, 2020, pp. 2492–2496, doi: 10.21437/Interspeech.2020-3038.

[27] H. Schröter, A. N. Escalante-B., T. Rosenkranz, and A. Maier, "DeepFilterNet2: Towards real-time speech enhancement on embedded devices for full-band audio," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*, 2022, pp. 1–5, doi: 10.1109/IWAENC53105.2022.9914782.