

Audio Noise Reduction Based on Spectral Attenuation

Abril Donahi Castro-Sánchez, Gengis Cruz-Coronel

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación,
Mexico

{cs226570036,cc226570037}@alm.buap.mx

Abstract. This paper presents an experimental characterization of an audio denoising method based on the Short-Time Fourier Transform (STFT) and spectral attenuation. The study analyzes how processing parameters, noise-profile estimation, input signal-to-noise ratio (SNR), and noise characteristics influence the performance of the method. A controlled synthetic signal was first used to study the influence of the attenuation factor, spectral floor, STFT window length, and frame overlap. The analysis identified a favorable attenuation region around $\alpha = 1.7$, whereas low values of the spectral-floor parameter showed comparatively small influence on SNR improvement. The selected configuration was subsequently evaluated for input SNR levels from 0 to 15 dB using both known and estimated noise profiles. A complementary experiment with a speech signal contaminated by white, cafeteria, and traffic noise was also performed. With an estimated noise profile and an input SNR of 10 dB, the mean SNR improvements were 8.15 ± 0.18 dB for white noise, 4.12 ± 1.27 dB for cafeteria noise, and 8.66 ± 3.04 dB for traffic noise. The results show that the performance of spectral attenuation is influenced by both parameter selection and noise characteristics.

Keywords: Audio denoising, short-time Fourier transform, spectral attenuation, noise estimation, signal-to-noise ratio.

Reducción de ruido de audio basada en atenuación espectral

Resumen. Este artículo presenta una caracterización experimental de un método de reducción de ruido en audio basado en la Transformada de Fourier de Tiempo Reducido (STFT) y la atenuación espectral. El estudio analiza cómo los parámetros de procesamiento, la estimación del perfil de ruido, la relación señal-ruido (SNR) de entrada y las características del ruido influyen en el rendimiento del método. Inicialmente se utilizó una señal sintética controlada para estudiar la influencia del factor de atenuación, el nivel de suelo espectral (*spectral floor*), la longitud de la ventana de la STFT y el solapamiento de tramas. El análisis

identificó una región de atenuación favorable en torno a $\alpha = 1,7$, mientras que los valores bajos del parámetro de suelo espectral mostraron una influencia comparativamente menor en la mejora de la SNR. Posteriormente, se evaluó la configuración seleccionada para niveles de SNR de entrada de 0 a 15 dB, utilizando tanto perfiles de ruido conocidos como estimados. También se llevó a cabo un experimento complementario con una señal de voz contaminada por ruido blanco, ruido de cafetería y ruido de tráfico. Con un perfil de ruido estimado y una SNR de entrada de 10 dB, las mejoras medias de la SNR fueron de $8,15 \pm 0,18$ dB para ruido blanco, $4,12 \pm 1,27$ dB para ruido de cafetería y $8,66 \pm 3,04$ dB para ruido de tráfico. Los resultados demuestran que el rendimiento de la atenuación espectral está influido tanto por la selección de parámetros como por las características del ruido.

Palabras clave: Reducción de ruido en audio, transformada de Fourier de tiempo reducido, atenuación espectral, estimación de ruido, relación señal-ruido.

1 Introduction

Audio signals acquired in practical environments are frequently affected by unwanted acoustic components originating from recording conditions, transmission systems, surrounding sources, or sensor noise. Such contamination can degrade perceived quality and may also affect subsequent signal analysis when relevant information must be extracted from temporal or spectral representations. Consequently, noise reduction is an important preprocessing task in audio signal processing [5, 8].

Audio signals commonly exhibit time-varying spectral characteristics, making a purely temporal or global frequency-domain representation insufficient for describing their local behavior. The Short-Time Fourier Transform (STFT) addresses this limitation by analyzing short signal segments and providing a time-frequency representation of their spectral content [1, 7]. The resulting representation can be used not only for visualization through spectrograms, but also for modifying spectral components associated with unwanted noise.

Among classical approaches to noise reduction, spectral subtraction and related attenuation methods are particularly attractive because of their computational simplicity and interpretability. These methods estimate the spectral characteristics of the noise and reduce their contribution to the magnitude spectrum of the contaminated signal, while the phase of the observed signal can be preserved for reconstruction [3]. Despite this relatively simple formulation, their practical behavior depends on several design choices, including the quality of the noise estimate, the amount of spectral attenuation, the spectral floor, and the parameters used to construct the STFT.

This dependence creates an important experimental problem. Reporting a single set of processing parameters and a single noise condition does not indicate whether the observed improvement is robust to parameter changes, different

contamination levels, or changes in the temporal and spectral characteristics of the noise. In particular, a configuration that performs well for stationary white noise may behave differently for environmental noise whose spectrum changes over time or overlaps with components of the signal of interest.

Therefore, the objective of this work is to experimentally characterize an STFT-based spectral attenuation method under controlled and progressively more realistic conditions. The study investigates two main questions: how sensitive is the achieved noise reduction to the principal processing parameters, and how does its performance change with the contamination level, noise-profile acquisition strategy, and type of acoustic noise.

The experimental study is organized in three stages. First, a controlled synthetic signal is used to analyze the influence of the attenuation factor α , the spectral-floor parameter β , STFT window length, frame overlap, and the interaction between the attenuation parameters. Second, the selected configuration is evaluated at input SNR levels of 0, 5, 10, and 15 dB using multiple generated noise signals and comparing known and estimated noise profiles. Finally, a recorded speech signal is contaminated at a controlled input SNR with white, cafeteria, and traffic noise to evaluate the method under different acoustic characteristics. The main contribution of this work is a reproducible experimental analysis aimed at identifying the sensitivity of the main processing parameters, evaluating the effect of noise-profile estimation, and characterizing the performance of the method under different contamination levels and noise conditions.

The remainder of this paper is organized as follows. Section 2 reviews related work on time–frequency audio analysis and spectral noise reduction. Section 3 describes the implemented denoising methodology. Section 4 presents the experimental setup and evaluation protocol. Section 5 reports and discusses the experimental results and their validity. Finally, Section 6 presents the conclusions and future research directions.

2 Related Work

Noise reduction based on spectral processing has been studied since the earliest work on speech enhancement in the frequency domain. One of the classical methods is spectral subtraction, proposed by Boll [3], in which the noise spectrum is estimated and its contribution to the spectrum of the noisy signal is reduced.

The signal is then reconstructed using the modified magnitude and the phase information from the observed signal. Although this procedure can reduce broadband noise with a relatively simple implementation, it can also produce residual components and perceptible distortions when the noise estimate or the degree of attenuation is inadequate.

To address this limitation, Berouti et al. [2] studied a variant of spectral subtraction in which the noise is overestimated and a minimum level is imposed on the spectral components after subtraction. These two elements correspond to the parameters denoted by α and β in this work. The parameter α controls the

amount of noise subtracted from the observed spectrum, whereas β establishes a spectral floor that prevents the magnitude from decreasing without bound. This formulation makes it possible to adjust the trade-off between residual noise reduction and alteration of the signal of interest.

In addition to the attenuation parameters, estimation of the noise profile is an important part of spectral processing. Martin [6] proposed a method for estimating the power spectral density of nonstationary noise by tracking spectral minima, without relying directly on a voice activity detector. Cohen [4] subsequently introduced an estimation procedure based on minima-controlled recursive averaging, designed to track changes in the noise under nonstationary conditions and at low signal-to-noise ratios. These studies show that the quality of the noise estimate directly affects the performance of the subsequent noise reduction stages.

The time–frequency representation also requires several analysis choices. The STFT requires the selection of a window function, its length, and the hop size between frames [1, 7]. These parameters determine the trade-off between temporal and frequency resolution and may therefore affect the representation used during noise estimation and attenuation.

In view of these previous studies, this work focuses on the experimental characterization of an STFT-based spectral attenuation scheme. The analysis examines the influence of the main processing parameters, the input SNR, the method used to obtain the noise profile, and the characteristics of the additive noise. Controlled conditions with a synthetic signal are considered first, followed by a speech signal corrupted by white and environmental noise.

3 Methodology

The evaluated method follows a sequence of processing steps in the time–frequency domain. The input signal is preprocessed to remove its DC component and maintain a common scale among the clean reference, the noisy signal, and the noise. The Short-Time Fourier Transform (STFT) is then computed, a spectral noise profile is obtained, and the magnitude of the noisy signal is modified through spectral attenuation. The phase of the observed signal is retained, and the processed signal is recovered using the inverse STFT. Finally, the result is compared with the clean signal using the signal-to-noise ratio (SNR). Figure 1 summarizes the main processing stages.

3.1 Preprocessing

Before the spectral analysis, two preprocessing operations were applied: DC component removal and normalization using a common scale. The DC component is removed by centering each signal around zero:

$$x_{\text{dc}}[n] = x[n] - \mu_x, \quad (1)$$

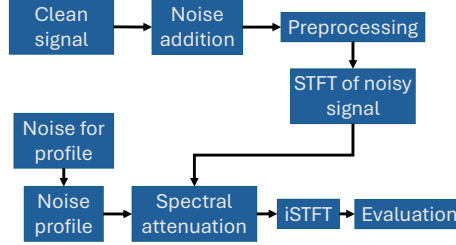


Fig. 1. General workflow of the evaluated audio denoising procedure.

where μ_x is the mean value of the signal:

$$\mu_x = \frac{1}{N} \sum_{n=0}^{N-1} x[n]. \quad (2)$$

The clean signal, the noisy signal, and the noise were then divided by the same scaling factor, computed as the maximum absolute value of the noisy signal after removing its DC component:

$$x_{\text{proc}}[n] = \frac{x_{\text{dc}}[n]}{\max(|x_{\text{mix,dc}}[n]|)}. \quad (3)$$

Using a common scaling factor preserves the relative amplitudes of the signals within the same experimental condition [5].

3.2 Time–Frequency Representation Using the STFT

The preprocessed signal was represented in the time–frequency domain using the short-time Fourier transform. For a discrete-time signal $x[n]$, the STFT can be expressed as:

$$X(k, m) = \sum_{n=0}^{N_w-1} x[n + mH] w[n] e^{-j \frac{2\pi kn}{N_w}}. \quad (4)$$

where $w[n]$ is the window function, N_w is its length, H is the hop size between frames, k is the frequency index, and m is the time index [1, 7, 8].

A Hann window was used in the implementation. The window length and frame overlap were later examined as part of the sensitivity analysis because both parameters affect the temporal and frequency resolution of the representation.

3.3 Noise Profile Estimation

Two approaches were considered for obtaining the spectral noise profile. The first, referred to as the known profile, directly uses the noise employed to corrupt the

signal. The profile is calculated by averaging the magnitude of its STFT across the time frames:

$$N_p(k) = \frac{1}{M} \sum_{m=0}^{M-1} |N(k, m)|. \quad (5)$$

where M is the number of frames available from the noise used in the evaluation.

The second condition uses a profile estimated from a 0.30 s noise-only segment that is not included in the 2 s interval used to evaluate the noisy signal. The profile is calculated in the same manner:

$$\hat{N} * p(k) = \frac{1}{M_0} \sum_{*m=0}^{M_0-1} |N_{\text{est}}(k, m)|. \quad (6)$$

where M_0 is the number of frames obtained from the segment used for estimation.

This separation allows a condition in which the noise used during evaluation is known to be compared with another in which only an independent sample from the same acoustic environment is available.

3.4 Spectral Attenuation

Noise reduction was performed by modifying the magnitude of the STFT of the noisy signal. Let $|X(k, m)|$ denote its magnitude and $N_p(k)$ the noise profile used for the corresponding condition. The processed magnitude is obtained as:

$$|\hat{S}(k, m)| = \max(|X(k, m)| - \alpha N_p(k), \beta N_p(k)). \quad (7)$$

The parameter α controls the amount of the noise profile subtracted from the observed magnitude. Larger values produce stronger attenuation, but they can also remove components of the signal of interest. The parameter β sets a spectral floor that limits how far each component can be reduced [3, 2].

After modifying the magnitude, the phase of the noisy signal is retained to form a complex representation:

$$\hat{S}(k, m) = |\hat{S}(k, m)| e^{j \angle X(k, m)}. \quad (8)$$

3.5 Signal Reconstruction

The processed signal was recovered in the time domain using the inverse short-time Fourier transform (iSTFT). To allow a direct comparison among the signals, the reconstructed output was adjusted to match the length of the clean reference signal.

This procedure retains the frame-based analysis structure used in the STFT and produces a time-domain signal from the modified spectrum [1].

3.6 Evaluation Metric

Quantitative performance was evaluated using the signal-to-noise ratio (SNR). For a clean reference signal $x_{\text{ref}}[n]$ and a signal under evaluation $x_{\text{test}}[n]$, the SNR was calculated as:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{n=0}^{N-1} x_{\text{ref}}^2[n]}{\sum_{n=0}^{N-1} (x_{\text{ref}}[n] - x_{\text{test}}[n])^2} \right). \quad (9)$$

In addition to the output SNR, the improvement produced by the processing was calculated as:

$$\Delta \text{SNR} = \text{SNR} * \text{out} - \text{SNR} * \text{in}. \quad (10)$$

Thus, the comparison considers both the final SNR and the level of contamination present before processing.

The implementation was developed in Python within a Jupyter Notebook environment. General-purpose libraries were used for numerical operations, STFT and inverse-STFT computation, audio loading and resampling, and visualization. The specific noise-reduction logic, including preprocessing, noise-profile construction, spectral attenuation, signal reconstruction, and SNR-based evaluation, was explicitly implemented for this study. No noise-reduction function or package was used to directly perform the complete denoising procedure.

4 Experimental Setup

The evaluation consisted of three experiments. The first examined the sensitivity of the method to changes in the main processing parameters. The second analyzed the performance of the selected configuration at different contamination levels and compared the use of a known noise profile with an estimated profile. Finally, the third experiment used a speech signal corrupted by different types of noise to evaluate the performance of the method under a wider range of acoustic conditions.

4.1 Synthetic Signal and Controlled Noise Generation

A synthetic signal composed of three sinusoidal components was used in the first two experiments:

$$x_{\text{clean}}(t) = 0.6 \sin(2\pi 440t) + 0.3 \sin(2\pi 880t) + 0.2 \sin(2\pi 1760t). \quad (11)$$

The sampling rate was $f_s = 16000$ Hz, and each evaluation signal had a duration of 2 s, corresponding to 32000 samples.

To prevent the noise level from depending on an arbitrary amplitude value, the generated white noise was scaled to a predefined input SNR. For a clean signal with power P_x and a noise signal with power P_n , the scaling factor was calculated as:

$$g = \sqrt{\frac{P_x}{P_n \cdot 10^{\text{SNR}_{\text{obj}}/10}}}. \quad (12)$$

The noisy signal was then obtained as:

$$x_{\text{mix}}[n] = x_{\text{clean}}[n] + g, n[n]. \quad (13)$$

This procedure allowed the experimental conditions to be compared using known contamination levels.

4.2 Parameter Sensitivity Analysis

The first experiment examined the influence of the main parameters of the method before selecting a configuration for the subsequent tests. All evaluations in this stage used an input SNR of 10 dB and the known noise profile so that the comparison focused only on the processing parameters.

The attenuation factor α was first evaluated over a broad range from 0.2 to 3.0, in increments of 0.2. After identifying the best-performing region, a finer evaluation was conducted between 1.5 and 2.1, in increments of 0.05.

The spectral-floor parameter β was then studied using values between 0 and 0.2. The interaction between values of α near the best-performing region and low values of β was also evaluated to determine whether the effects of the two parameters were independent or whether performance was particularly sensitive to a specific combination.

Finally, STFT window lengths of 256, 512, 1024, 2048, and 4096 samples were analyzed, together with overlap values of 25%, 50%, and 75%. A Hann window was used in all cases.

4.3 Evaluation at Different SNR Levels

After the sensitivity analysis, a common configuration was selected for the remaining tests: $\alpha = 1.7$, $\beta = 0.02$, a Hann window of 1024 samples, and an overlap of 512 samples, equivalent to 50%.

Using this configuration, input SNR levels of 0, 5, 10, and 15 dB were evaluated. Ten trials were conducted at each level, each using a white-noise signal generated with a random seed.

Each trial was evaluated under two conditions. In the first, a known profile was obtained directly from the same noise used to corrupt the signal. In the second, an independent 0.30 s segment was generated with the same statistical characteristics and scaling level, and this segment was used to estimate the noise profile.

The output SNR and the improvement ΔSNR were calculated for each condition. The results of the ten trials were summarized using the mean and sample standard deviation.

4.4 Evaluation Using a Speech Signal and Different Types of Noise

To complement the evaluation with a real signal, a speech recording originally sampled at 48 kHz was used. The signal was resampled to 16 kHz, and a 2 s segment containing continuous speech was selected.

The same signal was corrupted by three types of noise: artificially generated white noise, cafeteria noise, and traffic noise. The two environmental noises were obtained from the DEMAND database (Diverse Environments Multichannel Acoustic Noise Database), which contains environmental noise recordings from different indoor and outdoor settings [9].

An input SNR of 10 dB was used in all three cases, and the previously selected parameters were retained. For white noise, ten trials were conducted using signals generated with different random seeds. For cafeteria and traffic noise, ten nonoverlapping time segments from each recording were evaluated.

For each environmental noise segment, 0.30 s was used to estimate the profile, and the following 2 s was used to corrupt the speech signal. Thus, the segment used to construct the profile did not overlap with the interval used to calculate the SNR.

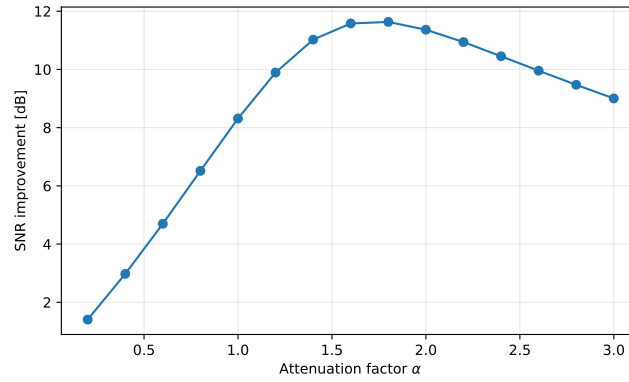


Fig. 2. SNR improvement as a function of the attenuation factor α .

5 Results and Discussion

The results are presented in the same order as the three experiments described in Section 4. The sensitivity of the attenuation and STFT parameters is examined first. The performance of the selected configuration is then studied at different noise levels and using different approaches to obtain the noise profile. Finally, the same configuration is evaluated using a speech signal and different types of noise.

5.1 Parameter Sensitivity

The first experiment showed that the parameters examined do not affect the performance of the method equally. The most pronounced variation occurred when the attenuation factor α was changed. At an input SNR of 10 dB, the improvement

increased steadily from low values of α and reached a best-performing region around 1.6–1.8. Beyond this region, further increases led to a gradual decrease in the output SNR.

In the initial evaluation, $\alpha = 1.0$ produced an improvement of 8.31 dB, whereas values between 1.6 and 1.8 yielded improvements above 11.5 dB. The finer evaluation conducted between 1.5 and 2.1 produced the highest improvement at $\alpha = 1.70$, with a value of 11.66 dB. The trend obtained over the complete evaluation of this parameter is shown in Figure 2.

After identifying the best-performing region for α , the value was fixed at $\alpha = 1.70$ to examine the influence of the spectral floor β . Unlike α , this parameter showed lower sensitivity within the range of small values. For $\alpha = 1.70$, the improvement was 11.69 dB with $\beta = 0$, 11.68 dB with $\beta = 0.01$, and 11.66 dB with $\beta = 0.02$. The difference between $\beta = 0$ and $\beta = 0.02$ was only 0.034 dB. As β increased beyond this range, the improvement gradually decreased.

Although the two parameters were initially analyzed separately, their combined behavior was also evaluated to verify that the selection of α did not depend on a single value of β . For all the low values of β tested, the best value of α within the analyzed range remained 1.70. Therefore, the selection of α was not tied to an isolated combination with a particular value of β .

After examining the spectral attenuation parameters, the influence of the STFT representation was studied. Window length also produced marked changes. For the synthetic signal used in the experiment, the improvement increased from 7.13 dB with a 256-sample window to 14.33 dB with a 4096-sample window. This behavior is consistent with the characteristics of the test signal, which consists of stationary sinusoidal components for which a longer window provides a more detailed frequency representation.

However, the 4096-sample window was not automatically selected as the best option for the subsequent tests. At a sampling rate of 16 kHz, it corresponds to a duration of 256 ms, which considerably reduces temporal resolution and may be unsuitable for time-varying signals such as speech. For this reason, an intermediate window length of 1024 samples, equivalent to 64 ms, was retained for the subsequent evaluations.

After selecting an intermediate window length of 1024 samples, this value was held constant while the effect of frame overlap was examined. Overlap values of 25%, 50%, and 75% produced improvements of 9.30, 11.66, and 11.97 dB, respectively. Although 75% yielded the highest value, the difference relative to 50% was approximately 0.31 dB. A 50% overlap was therefore selected, reducing the number of frames without a substantial loss in performance.

Based on these results, the configuration used in the subsequent experiments was $\alpha = 1.70$, $\beta = 0.02$, a 1024-sample Hann window, and an overlap of 512 samples. This selection is not regarded as a generally optimal configuration, but rather as a working configuration obtained under the controlled conditions of the first experiment.

Table 1. SNR improvement at different noise levels.

Input SNR [dB]	Known profile [dB]	Estimated profile [dB]
0	12.51 ± 0.24	11.65 ± 0.23
5	12.04 ± 0.24	11.28 ± 0.23
10	11.51 ± 0.23	10.87 ± 0.22
15	11.00 ± 0.22	10.46 ± 0.21

Table 2. SNR improvement for speech with different types of noise.

Noise type	Known profile [dB]	Estimated profile [dB]
White	8.28 ± 0.10	8.15 ± 0.18
Cafeteria	4.95 ± 0.83	4.12 ± 1.27
Traffic	8.93 ± 2.70	8.66 ± 3.04

5.2 Effect of Noise Level and Noise Profile

Once the processing configuration had been fixed, the effect of the input SNR was examined. Table 1 shows the mean improvement obtained over the ten trials conducted at each noise level.

Both types of profile produced a positive improvement at every SNR level tested. The improvement was greater when the initial noise level was higher and gradually decreased as the input SNR increased. For the known profile, the improvement decreased from 12.51 dB at an initial SNR of 0 dB to 11.00 dB at an initial SNR of 15 dB.

The estimated profile produced slightly lower values than the known profile under all conditions. The mean difference between them was 0.86 dB at an input SNR of 0 dB and decreased to 0.54 dB at an input SNR of 15 dB. The standard deviations remained close to 0.2 dB, indicating consistent behavior across the different white-noise signals generated.

5.3 Evaluation Using Speech and Different Types of Noise

The results show that the type of noise affects both the mean improvement and the variability of the method. White noise produced a high and stable improvement, with a standard deviation below 0.2 dB for the estimated profile. Cafeteria noise produced the lowest mean improvement, at 4.12 ± 1.27 dB, whereas traffic noise produced the highest mean improvement, at 8.66 ± 3.04 dB, but also showed the greatest variation across segments. Table 2 summarizes the SNR improvement obtained for the different types of noise.

Therefore, the higher mean observed for traffic noise does not imply that this condition is always more favorable. The standard deviation shows that the result depends strongly on the time segment analyzed. In contrast, white noise produced much more uniform behavior.

The lower performance obtained with cafeteria noise may be related to its more complex acoustic content, which includes speech components and

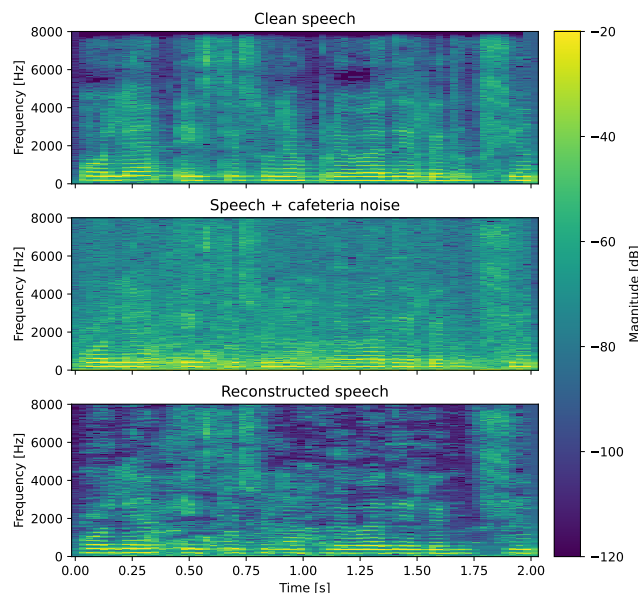


Fig. 3. Spectrogram comparison for clean speech, speech contaminated with cafeteria noise, and reconstructed speech using the estimated noise profile.

other sources that may overlap spectrally with the signal of interest. During listening, a clear reduction in the overall environmental noise was observed, although some background speech components remained after processing. This observation was used only as qualitative support and does not constitute a formal perceptual evaluation. Figure 3 shows an example for cafeteria noise. The comparison illustrates the spectral distribution of the clean speech, the increase in background content after contamination, and the reduction obtained after applying attenuation using the estimated profile.

5.4 Experimental Validity and Scope of the Results

The experimental design sought to maintain comparable conditions across the trials. In the synthetic-signal experiments, the same clean signal was used, and the noise power was explicitly adjusted to obtain the input SNR defined for each case. Therefore, the changes observed during the sensitivity analysis can be attributed mainly to changes in the processing parameters rather than to differences in the initial noise level.

After the sensitivity analysis, the same processing configuration was retained for the remaining tests. This decision allowed the method to be evaluated at different noise levels and with different types of noise without tuning the parameters to favor any particular condition. In addition, ten signals generated with different random seeds were used in the white-noise tests, whereas ten

nonoverlapping time segments from each recording were analyzed for the environmental noises. The results were reported using the mean and standard deviation to account for variation across the trials rather than a single case.

For the environmental noises, the different segments provide information about temporal variation within the same recording. However, these segments were taken from only one cafeteria recording and one traffic recording. The results therefore represent the acoustic conditions examined rather than all possible environments within these categories. Similarly, the speech evaluation was conducted using a single speaker and a two-second segment.

The configuration used in the speech and environmental-noise tests was initially selected using a synthetic signal and white noise. Retaining this configuration without further adjustment made it possible to examine its behavior outside the condition used for its selection, although this does not imply that the selected values are optimal for every signal or acoustic environment. In particular, the favorable behavior observed with longer windows during the sensitivity analysis is related to the stationary nature of the synthetic signal.

6 Conclusions

This work presented an experimental evaluation of an STFT-based spectral attenuation scheme for noise reduction. The study examined the effects of the main processing parameters, the method used to obtain the noise profile, the initial noise level, and the characteristics of the noise. Controlled experiments using a synthetic signal were combined with a subsequent evaluation using a speech signal.

The sensitivity analysis showed that the parameters examined did not affect the results in the same way. The attenuation factor α had a clear effect on SNR improvement, with a favorable range around 1.6–1.8 and the highest tested improvement obtained at $\alpha = 1.70$. In contrast, β produced only small variations within the range of low values. Window length also had a marked effect on performance, although longer windows benefited from the stationary nature of the synthetic signal. An intermediate configuration was therefore selected for the subsequent tests instead of simply using the value that produced the highest SNR under the initial condition.

The evaluation at different noise levels showed positive improvements at all the SNR levels examined. A profile estimated from an independent noise segment produced slightly lower results than the known profile but remained consistent across the trials. This indicates that the method retains its noise-reduction capability even when the noise used to corrupt the signal is not directly available for constructing the spectral profile.

The speech test showed that performance also depends on the characteristics of the noise. White noise produced a high improvement with little variation across trials, whereas traffic noise achieved a similar or higher mean improvement but with considerably greater dispersion. Cafeteria noise produced the lowest

improvement, illustrating the difficulty faced by the method when the noise contains acoustic components that overlap with the speech signal.

Overall, these results characterize the behavior of the method under different processing and noise conditions. The main contribution of the study is the systematic evaluation of these factors through controlled conditions, repeated trials, and subsequent validation using speech and environmental noises.

Future work could extend the evaluation to a larger number of speakers and independent environmental recordings and include complementary metrics related to perceptual quality and speech intelligibility. Automatic adaptation of the attenuation parameters to the characteristics of the observed noise could also be investigated.

References

1. Allen, J.B., Rabiner, L.R.: A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE* 65(11), 1558–1564 (1977)
2. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of speech corrupted by acoustic noise. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. pp. 208–211 (1979)
3. Boll, S.F.: Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27(2), 113–120 (1979)
4. Cohen, I.: Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing* 11(5), 466–475 (2003)
5. Lerch, A.: *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*. Wiley-IEEE Press (2012)
6. Martin, R.: Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing* 9(5), 504–512 (2001)
7. Müller, M.: *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer (2015)
8. Oppenheim, A.V., Schaffer, R.W.: *Discrete-Time Signal Processing*. Pearson, 3rd edn. (2010)
9. Thiemann, J., Ito, N., Vincent, E.: The diverse environments multi-channel acoustic noise database (DEMAND): A database of multichannel environmental noise recordings. *Proceedings of Meetings on Acoustics* 19, 035081 (2013)