ELEKTROTEHNIŠKI VESTNIK 93(3): 125-131, 2026
ORIGINAL SCIENTIFIC PAPER

An Improved Multi-Band Spectral Subtraction Technique for Cochlear Implant Users

Zouhir Amokrane

Laboratory of Renewable Energy and Materials (LREM), Department of Electrical Engineering, Faculty of Technology, University Yahia Fares of Medea, Medea, Algeria.

E-mail: amokrane_zouhir@yahoo.fr

Abstract. A novel approach is proposed to reduce the noise in cochlear implants by utilizing an enhanced multi-band spectral subtraction method. The main objective is to optimize the subtraction factors used in the spectral subtraction process to improve the overall performance of the stimulation strategy in cochlear implants designed to decompose the incoming audio signals into multiple sub-bands for a further processing. However, one of the persistent challenges in the cochlear implant technology is maintaining the speech intelligibility in noisy environments. To solve the issue, the subtraction factors are dynamically adapted based on common noise patterns observed in real-world scenarios.
The method effectiveness is assessed by conducting extensive simulations using noisy sentences corrupted by eight distinct types of the environmental noise, including the street noise, machinery, and crowd chatter. The objective evaluations are performed by using the established metrics for the speech intelligibility and audio quality. The results demonstrate that the proposed method significantly enhances the signal-to-noise ratio,thus ensuring are improved speech clarity for users of cochlear implants.
The results obtained in our simulations show that the approach is highly promising. The more precise signal processing enabled improves the speech intelligibility and assures a user satisfaction.

Keywords: multi-band spectral subtraction, speech enhancement, speech intelligibility, cochlear implant

Izboljšana tehnika večpasovnega spektralnega odštevanja za uporabnike polževih vsadkov

V tem članku predlagamo nov pristop k zmanjšanju šuma pri polževih vsadkih z uporabo izboljšane večpasovne metode spektralnega odštevanja. Glavni cilj je optimizirati faktorje odštevanja, ki se uporabljajo v procesu spektralnega odštevanja, in s tem izboljšati splošno učinkovitost strategije stimulacije v polževih vsadkih. Ti vsadki so zasnovani tako, da razgradijo dohodne zvočne signale v več podpasov za nadaljnjo obdelavo. Eden stalnih izzivov pri tej tehnologiji je ohranjanje razumljivosti govora v hrupnem okolju. Za reševanje te težave naš pristop vključuje dinamično prilagajanje faktorjev odštevanja na podlagi tipičnih vzorcev šuma, ki jih opažamo v realnih situacijah.
Za oceno učinkovitosti metode smo izvedli obsežne simulacije z uporabo hrupnih stavkov, popačenih z osmimi različnimi vrstami okoljskega hrupa, vključno s hrupom ulic, strojev in klepeta množice. Rezultati kažejo, da predlagana metoda znatno izboljša razmerje signal/šum in s tem zagotavlja večjo jasnost govora za uporabnike polževih vsadkov.

Ključne besede: Večpasovno spektralno odštevanje, Izboljšanje govora, Razumljivost govora, Polžev vsadek.

Received: 30 September 2025
Accepted:12 Maj 2026

1 INTRODUCTION

People with a hearing loss often rely on hearing aids, which are critical devices for improving their ability to perceive sounds. Among these, the cochlear implants (CIs) are designed for individuals with a profound hearing loss, providing a more advanced solution when conventional hearing aids are insufficient. CI consists of an external component which captures and processes the sound and an internal implant that directly stimulates the cochlea. Electrodes inserted into the cochlea replace the function of the damaged ear cells by stimulating the auditory nerve, enabling the user to perceive the sound [1], [2].

Despite their transformative impact, CIs face persistent challenges in noisy environments. The noise interference significantly reduces the speech intelligibility, affecting the CI users more than the individuals with a normal hearing (NH). Even the low levels of the background noise, tolerable for the NH individuals, can disrupt the speech comprehension for the CI users. Moreover, the limited number of the stimulation channels (typically 16 to 22, depending on the manufacturer) restricts the auditory resolution, making the sound perception even more challenging. Addressing these issues is crucial, as enhancing the speech intelligibility in noisy environments can significantly improve the quality of life for the CI users [2].

From the theoretical perspective, research into the noise reduction techniques for cochlear implants advances our understanding of signal processing in dynamic auditory environments. The practical contributions include the potential for improved CI technologies that better meet the user needs, particularly in real-world noisy conditions. Among the noise reduction methods, the spectral subtraction has garnered attention for its simplicity and computational efficiency, which makes it suitable for real-time applications. The multiband spectral subtraction (MBSS) is a notable advancement that aligns well with the signal decomposition strategies used in the CI processors. By applying the noise reduction in individual spectral sub-bands, MBSS effectively the enhances speech intelligibility while maintaining a low computational complexity [3], [4].

However, the conventional spectral subtraction methods, including MBSS, encounter limitations, particularly with the non-stationary noise. Addressing these limitations is essential to ensure a better performance in diverse real-world auditory scenarios [5], [6]. The previous research proposed various enhancements to MBSS, such as over-subtraction techniques and psychoacoustic models that ensure the inaudible noise without affecting the speech intelligibility [7].

Our objective is to propose and evaluate an enhanced multiband spectral subtraction (IMBSS) to strengthen the MBSS performance by addressing its limitations and improving its effectiveness for speech processing in cochlear implants. By focusing on both the theoretical advancements and practical applications, seeks our aim is to contribute to the development of more robust noise reduction techniques, ultimately improving the speech intelligibility for the CI users in real-world environments.

2 MULTI-BAND SUBSTRACTION ALGORITHM

Figure 1 shows a block diagram describing the multiband spectral subtraction method [8], [9].

Suppose that y(n), the corrupted speech, is composed of clean speech s(n) and additive noise d(n), that is:

$$y(n) = s(n) + d(n) \tag{1}$$

The power spectrum of the noise-corrupted input signal can be approximately expressed as:

$$|Y(k)|^2 \approx |S(k)|^2 + |D(k)|^2 \tag{2}$$

S(k) are the magnitude spectra of the clean speech and D(k) is the noise. A magnitude spectra are unknown; an estimate of D(k) is calculated by its average value during speech pauses. Several variations of the spectral subtraction method have been proposed. The majority of the implementations of the spectral subtraction approach are variants of the approach proposed by M. Berouti. This approximation of the clean speech spectrum $|S(k)|$ is obtained as [10]:

$$\left|\hat{S}(k)\right|^2 = |Y(k)|^2 - \alpha\left|\hat{D}(k)\right|^2 \tag{3}$$

α is an over-subtraction factor (α ≥ 1), which subtracts the over-estimate of the noise.

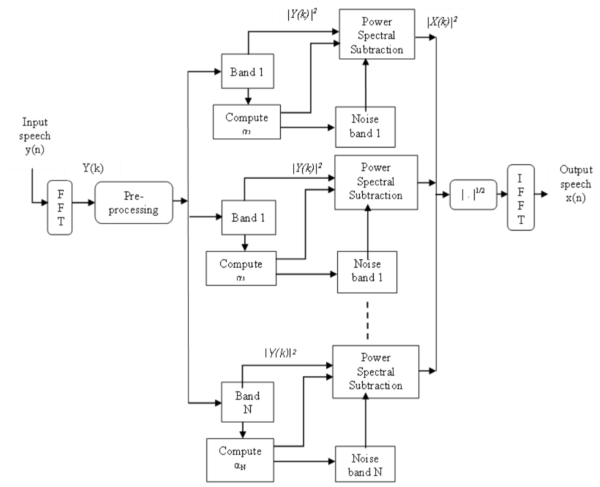

Figure 1. Block diagram of the multiband spectral subtraction method.

The approach supposes that the noise affects the speech spectrum equally. In reality, the noise affects the speech spectrum differently at various frequencies. Some frequencies will be affected more adversely than others. The MBSS approach estimates one subtraction factor for each band. The speech spectrum is divided into N non-overlapping bands, and the spectral subtraction is performed independently in each band. The clean speech spectrum in the ith band is estimated as:

$$\left|\hat{S}_i(k)\right|^2 = |Y(k)|^2 - \alpha_i\delta_i\left|\hat{D}(k)\right|^2 \ ; \ b_i \le k \le e_i \tag{4}$$

bi and ei are the beginning and ending frequency bins of the ith frequency band, αi is the over-subtraction factor of the ith band and δi is an additional band-subtraction factor that can be independently set for each frequency band to modify the noise removal process. Band - specific over-subtraction factor αi is a function of segmental signal to noise ration of the ith frequency band and is computed as follows:

$$\alpha_i = \begin{cases} 5 & SNR_i \le 5 \\ 4 - \frac{3}{20}(SNR_i) & -5 \le SNR_i \le 20 \\ 1 & SNR_i > 20 \end{cases} \tag{5}$$

Using the SNRi value calculated in Equation (5), α_i can be determined as:

$$SNR_i(dB) = 10\,log\,1\,0\left(\frac{\sum_{k=b_i}^{e_i}|Y_i(k)|^2}{\sum_{k=b_i}^{e_i}|\hat{D}_i(k)|^2}\right) \quad (6)$$

While the use of over-subtraction factor αi provides a degree of the control over the noise subtraction level in each band, the use of multiple frequency bands and the use of the δi weights provide an additional degree of the control within each band. The negative values in the enhanced spectrum in equation (4) are floored to the noisy spectrum as:.

$$\left|\hat{S}_i(k)\right|^2 = \begin{cases} \left|\hat{S}_i(k)\right|^2 & if \quad \left|\hat{S}_i(k)\right|^2 > 0 \\ \beta\left|\hat{Y}_i(k)\right|^2 & else \end{cases} \quad (7)$$

where b_i and $\beta = 0.002$ is the spectral floor parameter.

Although the use of over subtraction factor α_i provides a degree of the control over the noise subtraction level in each band, the use of multiple frequency bands and the use of the δi weights provides an additional degree of the control within each band. The values for δi are empirically determined and set to the values determined and set to:

$$\delta_i = \begin{cases} 1 & f_i \leq 1KHz \\ 2.5 & 1KHz \leq f_i \leq \frac{Fs}{2} - 2KHz \\ 1.5 & f_i \geq \frac{Fs}{2} - 2KHz \end{cases} \quad (8)$$

where fi is the upper frequency of the i^{th} band and Fs is the sampling frequency in Hz.

Figure 2 shows a graph of the over-subtraction factor as a function of SNR.

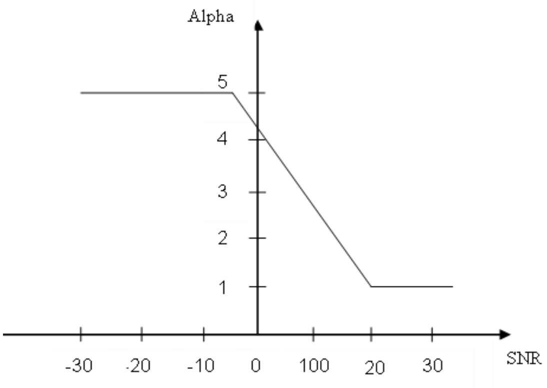

Figure 2. Plot of the over-subtraction factor as a function of SNR.

Figure 3 shows a plot of weights δi as a function of the frequency.

where fi is the upper frequency of the i^{th} band and Fs is the sampling frequency in Hz.

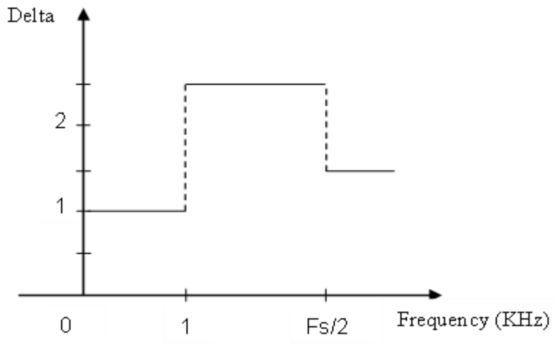

Figure 3. Plot of the δi weights as a function of the frequency.

3 Speech enhancement approach

In the MBSS method, factor α in equation (6) is defined as a function of segmental SNR of each frame and is empirically independent of factor δ. In each sub-band, δi is defined in equation (8) according to the noise present in the sub-band.

To improve factor δ, our factor δ does not depend only on the frequency band but also on the noise level in each band, this is ensured by introducing a more dynamic and noise-adaptive approach than the traditional spectral subtraction that introduces a more noise-adaptive and dynamic approach compared to the MBSS method.

We therefore propose a new framework in which factor δi is also a function of SNR calculated in each sub-frame which makes the spectral subtraction more adapted to the noise conditions in each frequency band.

3.1 Noise estimation using the Normalized Spectral Distance (NSD)

As it is generally difficult to estimate the noise in a subframe our method solves the problem by calculating the normalized spectral distance (NSD). NCM is a measure of the amount of the noise in each sub-band, calculated it is based on the difference between the spectral content of the noisy signal and a reference segment. Subtraction parameters δi are modified according to NCM which allows a more precise reduction. NSD in each frequency band represents the degree to which the noise spectrum deviates from the considered spectral shape of the speech, which makes this approach advantageous.

3.2 Noise signal decomposition into frequency bands

The sound is first decomposed into N frequency bands. We choose the decomposition according to the critical band distribution because it corresponds to the human hearing system. Which processes the sound similarly according to the different frequency ranges. This expresses an approach more adapted to CIs.

For a signal sampled at 8 kHz and windowed into frames of the size of 256 samples, each bin in the Fast Fourier Transform (FFT) frame represents a frequency increment of 31.25 Hz (as shown in Table 1).

Customizing the subtraction parameters according to the noise ensures independent processing. In our approach, the risk of distortion of the speech spectrum is minimal because the reduction is applied only where it is most needed in each sub-band.

Thus, our contribution is based on adapting the spectral subtraction parameters according to the noise in each frequency band, which improves the speech intelligibility for the CI users in noisy environments.

NSD between the noise and the noisy speech spectrum in each band. (Figure (4)):

$$NSD_i = \frac{\sqrt{\sum_{k=b_i}^{k=e_i}\left(F_i^s(k)-F_i^n(k)\right)^2}}{\sqrt{\left(F_i^s(k)\right)^2}} \tag{9}$$

where:

F_i^s is the i^th^ band energy of the noisy speech.

F_i^n is the i^th^ band energy of the noise.

Table 1: Mapping from critical bands at sampling frequency of 8 khz and frame size n=256

Critical Band Number	Intervals	Number of FFT bins	Frequencies (Hz)
01	01-03	03	0000-0094
02	03-06	03	0094-0187
03	07-10	04	0187-0312
04	10-13	03	0312-0406
05	13-16	03	0406-0500
06	16-20	04	0500-0625
07	20-25	05	0625-0781
08	25-29	04	0781-0906
09	29-35	06	0906-1094
10	35-41	06	1094-1281
11	41-47	06	1281-1469
12	47-55	08	1469-1719
13	55-64	09	1719-2000
14	64-74	10	2000-2312
15	74-86	12	2312-2687
16	86-100	14	2687-3125
17	100-118	18	3125-3687
18	118-128	10	3687-4000

The NSD value is between 0 and 1. When NSD between the estimated noise and the noisy speech is above 0.5, there is a small amount of the noise in this band, so δ takes the value 1 for all bans without affecting equation (4). Else, if with the NSD value below 0.5, which means there is much noise, δ takes the values according to equation (8). The proposed enhancement scheme is shown in Figure 4.

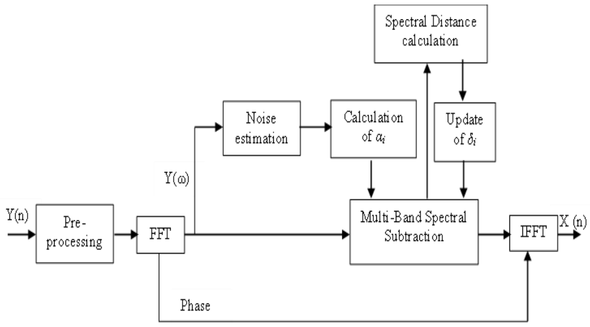

Figure 4. Block diagram of the IMBSS spectral subtraction method.

The initial algorithm is given as:

$BEGIN$
IF NSD $\geq$ 0.5

$$\delta_i = \begin{cases} 1 & f_i \leq 1KHz \\ 2.5 & 1KHz \leq f_i \leq \frac{Fs}{2} - 2KHz \\ 1.5 & f_i \geq \frac{Fs}{2} - 2KHz \end{cases}$$

$ELSE$
$\delta_i = 1 \quad \forall \quad f_i$
END

The performance of our method is evaluated using 30 sentences from the NOIZEUS noisy speech corpus, an IEEE database that contains audio recordings from both female and male speakers. The original audio files are sampled at 25 kHz and subsequently down-sampled to 8 kHz to match the typical sampling rate used in cochlear implants. The speech signals are corrupted by many forms of a real-world noise, recorded in diverse environments such as restaurants, streets, and train stations. This variation in the noise types provides a realistic and challenging test scenario for assessing the robustness of the noise reduction algorithm [11].

We use the Itakura-Saito (IS) distortion measure to evaluate the IMBSS performance [12]. This objective measure calculates the difference between the clean (noise-free) speech spectrum and the noisy speech spectrum. IS compares the spectral characteristics of the two signals in the spectral domain. Lower values indicate a better speech quality. SI varies between 0 and 10. SI close to zero means that a minimum distortion has been achieved.

4 SPEECH ENHANCEMENT FOR THE CI USERS

Vocoders that simulate the speech processing in CIs well are an effective approximation of the signal processing experience in the CI processors. There are strong correlations between the intelligibility scores of the CI users and normal-hearing (NH) listeners when compared to the vocoded speech [13]. To adapt our parameters to the CI users, we use a model based on vocoder principles.

The FFT-based N-of-M channel selection vocoder is one of the most widely used well-known vocoders among the cochlear stimulation strategies. The N-of-M channel selection is a peak selection method, where only a subset of the most dominant channels (usually "N") is selected from the total number of the channels (usually "M") to select the most critical spectral components of the sound.

4.1 Signal processing for the CI simulation

The block diagram of our signal processing approach is shown in Figure 4. The input speech signal is decomposed into N frequency bands by the Fast Fourier Transform (FFT). This decomposition mimics the frequency analysis performed by the IC processors, where the sound is decomposed into multiple frequency bands corresponding to the number of the stimulation electrodes available in the implant. Spectral subtraction is then applied with the optimized δi parameters. The acoustic waveform is then formed by summing the modulated filtered noises. This process simulates how an IC modulates the signal to stimulate the auditory nerve, creating an auditory representation of the speech signal for the user [14].

4.2 Perceptual evaluation of the speech quality

The perceptual speech quality can be measured using measures such as the Normalized Covariance Measure (NCM). NCM has the ability to predict the states subjective speech quality under different acoustic conditions with a high accuracy. This objective measure determines the covariance between the input and output envelope signals of the vocoder [15].

NCM is highly correlated with the human perceptual ratings of the speech intelligibility and quality, making it a reliable tool to evaluate the IMBSS effectiveness for different noise types. This is done by combining the vocoder-based CI simulation with the IMBSS measure enabling an objective evaluation of the impact of our improved spectral subtraction approach on the speech quality and intelligibility for different noise types.

5 Simulation Results and Discussions

Figures 5.a, 5.h i.e describe the results of the two methods, the classical and the proposed IMBSS method.

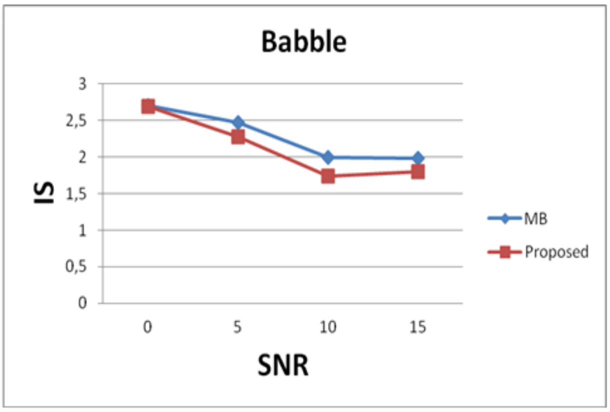

Figure 5.a. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Babble).

For the babble noise, our IMBSS approach consistently outperforms MBSS in terms of the IS distortion, demonstrating a superior reduction capability in this type of the noise.

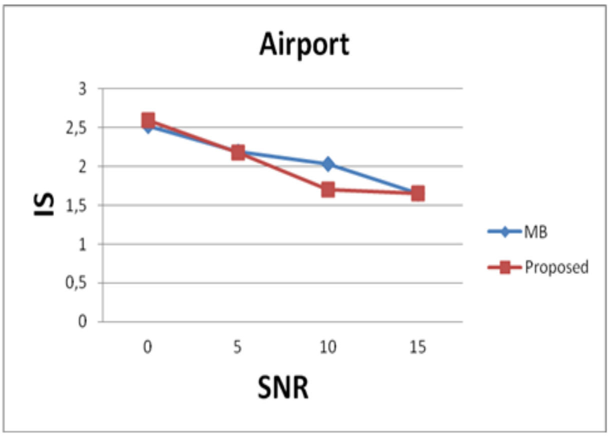

Figure 5.b. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Airport)).

In a case of the airport noise, the IMBSS and MBSS methods a show similar performance at lower SNR values, ranging from 0 dB to 5 dB. Beyond 5 dB, a significant improvement is seen when IMBSS is applied. This shows that IMBSS is more effective in reducing the airport noise as the SNR increases.

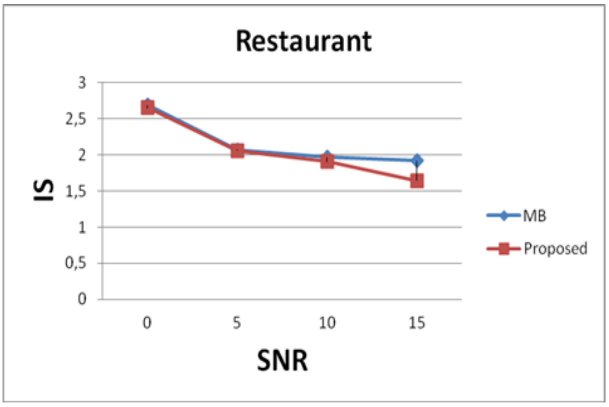

Figure 5.c. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Restaurant).

For the restaurant noise, both methods perform well at all SNR levels except at 10 dB SNR where IMBSS has a lower IS value. This suggests that IMBSS may show a marginal improvement over MBSS in some environments.

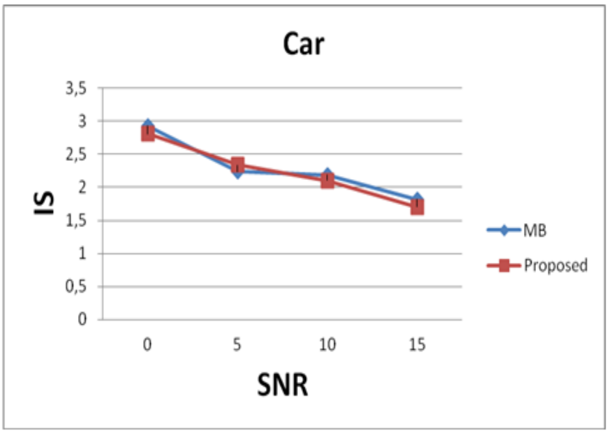

Figure 5.d. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Car).

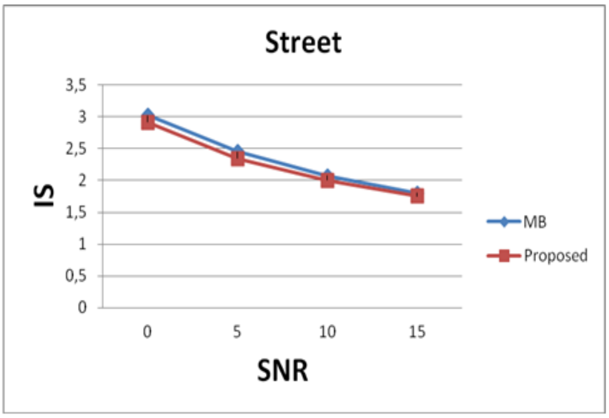

Figure 5.e. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Street).

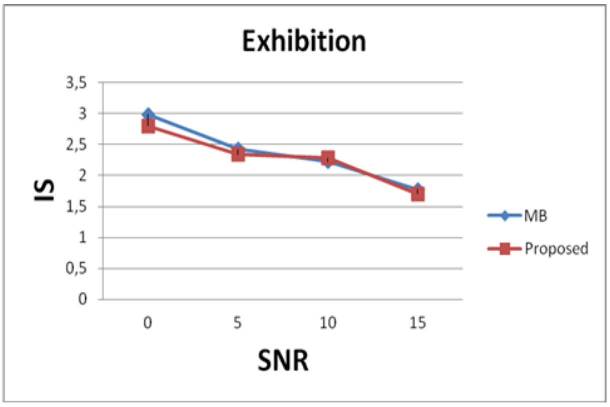

Figure 5.f. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Exhibition).

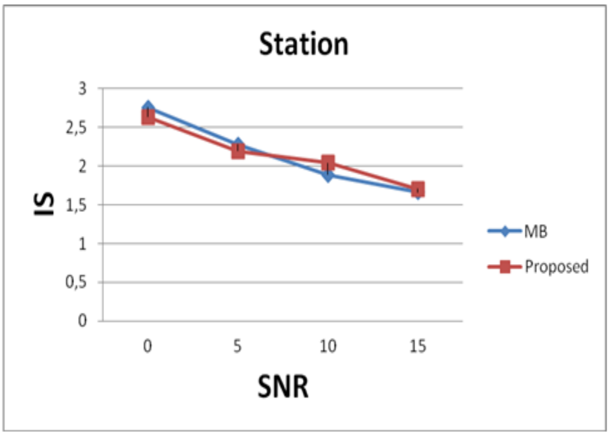

Figure 5.g. Comparison of the performance, in terms of the IS measure, of the MBSS with the IMBSS method over 30 sentences at different SNRs (Station).

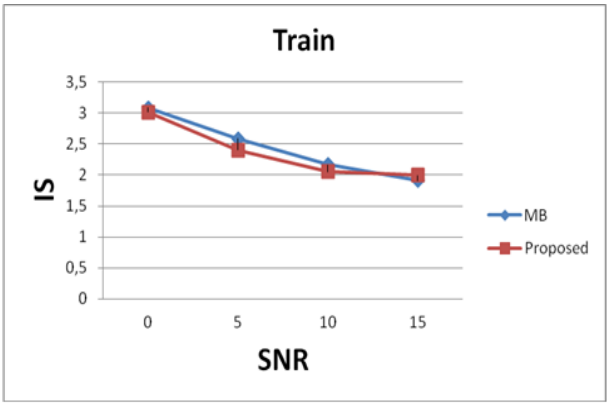

Figure 5.h. NCM mean scores for the MBSS and IMBSS methods at different SNRs (Train).

For other the noise types, such as the street noise and the train station noise, the performance of the two methods is similar, with no significant advantage at either approach under the conditions tested.

Figure 6 shows the variation of the Normalized Covariance Metric (NCM) based on average SNR across all the noise types in the database.

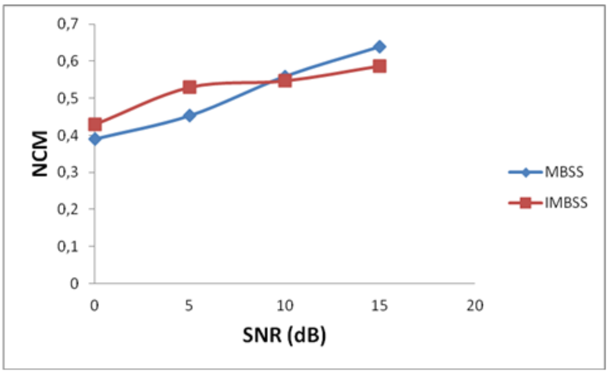

Figure 6. NCM mean scores for the MBSS and IMBSS methods at different SNRs.

When estimating the IMBSS performance using NCM, there are some weaknesses noted in improving the speech intelligibility at lower SNR levels (under 10 dB) compared to the MBSS method. Starting from 10 dB and higher SNR values, IMBSS shows a notable improvement in NCM, showing a better perceptual quality and the speech intelligibility in higher SNRs.

Overall, while the IMBSS method performs well under certain noise types and SNR conditions, its advantages are most noticeable at higher SNRs. In the environments with a more challenging noise (e.g., babble or airport noise), IMBSS clearly outperforms MBSS, for other noise types. Both methods deliver comparable results.

6 Conclusion and Perspectives

An enhanced multiband spectral subtraction (IMBSS) method is proposed to minimise the noise for the cochlear implant (CI) users, particularly in environments dominated by a colored noise, such as the babble and airport noise. The results demonstrate that the IMBSS method significantly outperforms the conventional MBSS method, especially in handling the non-stationary noise, making it a robust solution for improving the speech intelligibility in challenging acoustic environments.

Our work has practical implications for the CI users. The proposed method enhances communication in noisy settings such as transportation hubs and public spaces, contributing to a better quality of life. It also provides valuable insights into advancing the noise-reduction techniques, addressing the limitations of the existing methods, and ensuring compatibility with the real-world applications.

Our future work will focus on implementing IMBSS in real-time CI systems and on validating its performance through clinical trials, thus paving the way for its integration into modern CI processors.

References

[1] D.M. Zeitler, S.M. Prentiss, S.A. Sydlowski, C.C. Dunn, “Recommendations for Determining Cochlear Implant Candidacy in Adults,” American Academy of Audiology, The Laryngoscope, 133(10), 15314995, 2023.

[2] T. Arras, L. Rachman, A.V. Wieringen, D. Basçkent, “Perception of voice cues and speech-in-speech by children with prelingual single-sided deafness and a cochlear implant,” Hearing Research, 109133, 2024.

[3] Y. Meng, S. Yang, F. Haihong, L.Tan, “A Speech Enhancement Method for Cochlear Implant Listeners,” 35th Annual International Conference of the IEEE EMBS Osaka, Japan, 3 - 7 July, 2013.

[4] G. Kim, Y. Lu, P.C. Loizou, “An algorithm that improves speech intelligibility in noise for normal-hearing listeners,” J. Acoust. Soc. Am, 126(3), pp. 1486-1494, 2009.

[5] Y. Hu, P.C. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms,” Elsevier, Speech Communication, 49, pp. 588–601, 2007.

[6] S.F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. Acoustic, Speech and Signal Processing, 27, pp. 113-120, 1979.

[7] Y. Hu, P.C. Loizou, “A geometric approach to spectral subtraction,” Elsevier, Speech Communication, 50, pp. 453–466, 2008.

[8] G. J. Jang, J. S. Park, J. H. Kim, Y. H. Seo, “Line Spectral Frequency-based Noise Suppression for Speech-Centric Interface of Smart Devices,” Advances in Electrical and Computer Engineering, 11, pp. 3–8, 2011.

[9] S. Kamath, P. C. Loizou, “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise,” In Proc. Of ICASSP2002, Orlando, 2002.

[10] M. Berouti, R. Schwartz, J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” Proc IEEE Int Conf Acoust, Speech, Signal Process, pp. 208–211, 1979.

[11] R. Badri, J.H. Siegel, B. A. Wright, “Auditory filter shapes and high-frequency hearing in adults who have impaired speech in noise performance despite clinically normal audiograms,” J. Acoust. Soc. Am.,129(2), pp. 852–863, 2011.

[12] J. Jr. Deller, J. Hansen, J. Proakis, “Discrete-Time Processing of Speech Signals,” NY: IEEE Press, 2000.

[13] P.C. Loizou, “Speech processing in vocoder-centric cochlear implants. Cochlear and Brainstem Implants.,” Adv. Otorhinolaryngol, 64, pp. 109–143, 2006.

[14] Y. Hu, P.C. Loizou, “Subjective comparison and evaluation of speech enhancement algorithms,” Elsevier, Speech Communication, 49, pp. 588–61, 2007.

[15] Y. Hu, P.C. Loizou, “Objective measures for predicting speech intelligibility in noisy conditions based on new band importance functions,” J. Acoust. Soc. Am, 125(5), pp. 3387–3405, 2009.

Zouhir Amokrane received his engineering degree in electrical engineering from the University of Medea, Algeria, in 1997, and his Ph.D. degree in electrical engineering from the Ecole Polytechnique Algiers, Algeria, in 2018. Since 2019 he has been an Associate Professor with the Electrical Engineering Department of the same university. His research interests are in power conversion control, renewable energy, system signal processing, real-time implementation of systems and embedded systems.