EEMD-IF Based Method for Underwater Noisy Acoustic Signals Enhancement in Time-Domain

A. Caldeira, *Student Member, IEEE*, and R. Coelho, *Senior Member, IEEE*

***Abstract*—This letter introduces a novel signal enhancement method to mitigate the effects of underwater acoustic ambient noises. The combination of ensemble empirical mode decomposition (EEMD) and iterative filtering (IF) is adopted to analyze the underwater speech and chirp noisy signals. The main novelty relies on the estimation and selection of the noise components using the index of non-stationarity of each decomposition mode. The idea is to better suppress the noise components, especially for highly non-stationary environments. The proposed method is compared to four baseline approaches to enhance speech and chirp signals. For this purpose, signals of interest are corrupted by three underwater ambient noises with different non-stationarity degrees. Experiments results demonstrate that the proposal achieves the best results, especially in the presence of severe non-stationary noisy conditions. In this situation, the reduction in the root mean square error achieves 11.2%.**

***Index Terms*—Underwater acoustics, signal enhancement, empirical mode decomposition, index of non-stationarity.**

I. Introduction

The investigation of the underwater ambient noise with its distinct dynamics and statistics has become an active area of research [1], [2], [3]. These masking interferences severely impact the acoustic systems such as sonar and underwater speech communications [4], [5], and a vast range of applications: oceanographic studies [6], source detection [7], localization [8]. Generally, ambient noise varies according to geographic location, proximity of shipping lanes, season of the year, and depth of the sea (shallow or deep water) [1], [6]. The reduction of the background noise is an essential task for the quality and intelligibility improvement of underwater communication. The major challenge is to estimate the noise components, particularly in non-stationary real environments [9], [10].

Since 1970, signal enhancement solutions have been proposed to attenuate interferences caused by acoustic noises. Techniques such as the unbiased minimum mean-square error (UMMSE) [11] and the optimally-modified log-spectral amplitude (OMLSA) [9] apply the short-time Fourier transform (STFT) to cope with non-stationary noises in the spectral domain. Alternatively, methods based on time-frequency analysis, such as the empirical mode decomposition (EMD) [12] and its variations, have also been adopted for signal enhancement. Although EMD lacks of mathematical formulations, it has proven to be a powerful tool for non-stationary signal analysis and has been successfully applied in several research areas [13]. For instance, the EMDH [10], [14] solution decomposes the noisy speech signal in time domain and employs the Hurst exponent [15] to detect and attenuate the most corrupted components. The non-stationary noise estimation for speech enhancement (NNESE) [16] is another method defined to enhance non-stationary noisy signals in time domain. In this approach, a robust estimator [17] is adopted to obtain the noise standard deviation on a frame-by-frame basis directly from the corrupted speech samples.

This Letter introduces a novel method, hereinafter named EIF-NS, to enhance signals corrupted by non-stationary underwater acoustic noise. Ensemble EMD (EEMD) [18] and iterative filtering (IF) [19] are combined to analyze the noisy signals in the time domain. The EEMD-IF decomposition is able to deal with the mode mixing problem, while its convergence is formally established under certain conditions [19]. The index of non-stationarity (INS) [20] is here defined to detect the most corrupted intrinsic mode functions (IMF). The frame-by-frame criterion satisfactorily identifies those segments where the noise is dominant. Thus, the least corrupted components are selected to reconstruct the enhanced version of the signal of interest.

Several experiments are conducted to examine the proposed EIF-NS for speech and chirp signals. Speech signals are commonly used for communication with and in between submarines and divers [21]. Chirp is a low Doppler sensitive signal that presents good ability for interference rejection [22]. It is therefore used in wireless underwater acoustic communications. The target signals are corrupted by three underwater ambient noises: Bubbles, Killer Whale, and Ocean Liner. The NNESE, OMLSA, UMMSE, and EMDH approaches are adopted for comparative purposes. The EIF-NS technique is evaluated in terms of speech quality and intelligibility improvement with three objective measures. Chirp signals are examined using the signal-to-noise ratio (SNR) and the root mean square error (RMSE). Results show that the EIF-NS solution outperforms the baseline approaches for both speech and chirp target signals. In particular, results are substantially better than those attained by the EMDH, which reinforces the importance of the INS-based noise detection criterion.

The main contributions of this work are as follows:

- Introduction of the EIF-NS signal enhancement method to mitigate the effects of underwater ambient noise in the time domain;

Manuscript received 17 December 2022; revised 2 March 2023; accepted 2 March 2023. Date of publication 16 March 2023; date of current version 31 March 2023. This work was supported in part by the National Council for Scientific and Technological Development under Grant (CNPq) 305488/2022-8 and in part by Fundação de Amparo á Pesquisa do Estado do Rio de Janeiro (FAPERJ) under Grant 200518/2023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Daniele Giacobello. *(Corresponding author: R. Coelho.)*

The authors are with the Laboratory of Acoustic Signal Processing, Military Institute of Engineering (IME), Rio de Janeiro, RJ 22290-270, Brazil (e-mail: coelho@ime.eb.br).

Digital Object Identifier 10.1109/LSP.2023.3258219

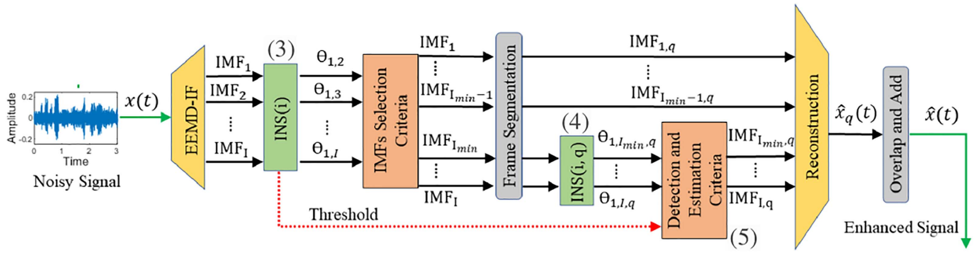

Fig. 1. Block diagram of the proposed EIF-NS method for signal enhancement based on EEMD-IF for non-stationary underwater noisy acoustic environment.

- Definition of a criterion for the detection of the non-stationary masking components based on the INS index;
- Adoption of the EEMD-IF to avoid the use of the original noisy phase for the enhanced signal reconstruction;
- Description of a scheme for underwater noise detection to yield quality and intelligibility improvement of speech and chirp noisy signals.

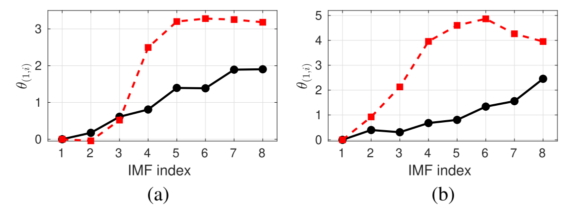

Fig. 2. The continuous black lines indicate the $\theta_{(1,i)}$ values of IMFs obtained for clean (a) speech and (b) chirp signal, while the dashed red lines represent the corresponding values from the noisy signals with SNR of 0 dB.

II. Proposed Enhancement Method

Fig. 1 illustrates the block diagram of the EIF-NS signal enhancement method, which is implemented in four phases: decomposition of the noisy signal with EEMD-IF; selection of IMFs that are most affected by noise; detection and estimation of noise components; and target signal reconstruction.

A. EEMD-IF

The combination of EEMD and IF is here adopted for the noisy signal decomposition. The EEMD was introduced in [18] to overcome the mode mixing problem, i.e., when widely distinct oscillations are concentrated in the same IMF. To this end, EEMD modes are computed by averaging IMFs obtained from the EMD of noisy versions of $x(t)$ considering different realizations of white Gaussian noise. The IF algorithm replaces the cubic splines interpolation required during the sifting process of the original EMD [12]. Instead, a data dependent moving average $\mathcal{L}(x)$ is applied to achieve lower processing time. The high-frequency component of $x(t)$ is given by

$$\mathcal{T}(x) = x - \mathcal{L}(x), \tag{1}$$

where $\mathcal{L}(x)$ refers to the local trend component. This separation is iteratively repeated over $\mathcal{L}(x)$ to define a series of IMFs and a residual $r(t)$, such that $x(t) = \sum_{i=1}^{I} \mathrm{IMF}_i(t) + r(t)$.

In [19], the function $\mathcal{L}(x)$ is represented by

$$\mathcal{L}(x)(t) = \sum_{j=-m}^{m} \left[\frac{m+1-j}{(m+1)^2}\right] x(t+j), \tag{2}$$

where the window size m is defined as $m = {}^{\alpha T}/_{\lambda}$, T is the number of samples of $x(t)$, λ is the number of local maxima and minima, and α is an adjustment factor. The mathematical support for the algorithm convergence and the lower computational cost when compared to the EMD method are the two main reasons to adopt the IF method. Furthermore, strategies such as the fast IF (FIF) [23] and the direct FIF (dFIF) [24] have also been derived to improve even further the IF speed. However, it is important to highlight that the following phases may also be applied to IMFs obtained from other methods, e.g., the variational mode decomposition (VMD) [25], [26].

B. IMF Selection

After the corrupted signal decomposition, the index of non-stationarity (INS) is applied to detect the IMFs that are most corrupted by noise. The INS was described in [20] to objectively examine the non-stationarity of a target signal. For this purpose, the signal is compared to a set of stationary references called surrogates. For each time scale T_h/T, where T_h is a short-time analysis window and T is the total duration of the signal, a threshold $\gamma \approx 1$ is defined to keep the stationarity assumption, considering a confidence degree of 95%. Thus, a signal is considered as stationary if INS $\leq \gamma$, and non-stationary, otherwise.

Chirp and speech signals are highly non-stationary. The acoustic noise interference generally attenuates the non-stationary behavior of such signals. Since acoustic noises are concentrated at low frequencies, the first IMFs have higher INS compared to the last ones. Furthermore, the first IMF is composed of faster oscillations than the others, resulting in higher SNR. Based on these facts, let $\theta_{(1,i)}$ be defined as the ratio between the logarithmic of the maximum INS value ($\mathrm{INS}_{\max}$) computed from the first and i-th IMFs,

$$\theta_{(1,i)} = \frac{\log(\mathrm{INS}_{\max,1})}{\log(\mathrm{INS}_{\max,i})}, \; i = 1, \ldots, I. \tag{3}$$

The ratio $\theta_{(1,i)}$ represents the attenuation of the non-stationary behavior observed in the i-th IMF when compared to the first mode. It also reflects the difference between the noise energy content in these IMFs. Fig. 2 exhibits the $\theta_{(1,i)}$ values obtained from clean and noisy versions of speech and chirp signals corrupted by a Ocean Liner noise with SNR of 0 dB. It can be seen that the the distance between $\theta_{(1,i)}$ values of clean and noisy signals significantly increases from third and fourth IMF for chirp and speech, respectively. It means that $\mathrm{IMF}_i(t), i \geq I_{\min}$ are the most corrupted modes, where $I_{\min} = 4$ for the speech signal and $I_{\min} = 3$ for the chirp signal. Thus, the frame-by-frame detection criterion is restricted to these selected IMFs.

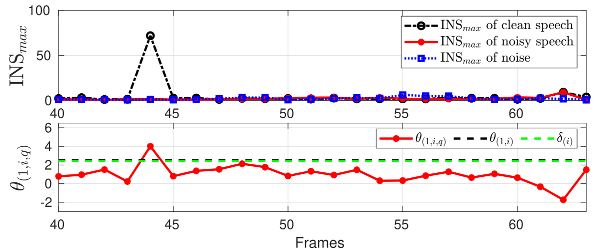

Fig. 3. INS_{max} of clean, noisy speech signal with 0 dB and noise for IMF 4 (top) and the corresponding $\theta_{(1,i,q)}$ values and threshold (bottom).

C. Detection and Estimation Criterion

Analogously to the definition of $\theta_{(1,i)}$, the detection of the most corrupted frames of each IMF is based on the frame-by-frame ratio $\theta_{(1,i,q)}$ computed as

$$\theta_{(1,i,q)} = \frac{\log(\text{INS}_{\text{max},(1,q)})}{\log(\text{INS}_{\text{max},(i,q)})},\ i = I_{\text{min}}, \ldots, I, \tag{4}$$

where $\text{INS}_{\text{max},(i,q)}$ refer to the maximum INS value obtained from frame q of $\text{IMF}_i(t)$. Let $\delta_{(i)} = \log(\text{INS}_{\text{max},i})$, in this work the following decision rule is proposed to determine the most corrupted frames:

$$\theta_{(1,i,q)} \begin{cases} \leq \min\{\theta_{(1,i)}, \delta_{(i)}\}, \text{frame is preserved;} \\ > \min\{\theta_{(1,i)}, \delta_{(i)}\}, \text{frame is removed.} \end{cases} \tag{5}$$

IMFs with greater indices may achieve lower $\delta_{(i)}$ and higher $\theta_{(1,i)}$ values than the first mode. Fig. 3 shows the INS_{max} of noise, clean and noisy speech signal (top), and the corresponding $\theta_{(1,i,q)}$ values and threshold for noise components detection for $i = 4$ (bottom). Note that in frame 44 (highlighted with black arrow), the INS_{max} values of noisy speech signal are close to INS_{max} values of noise and distant from the INS_{max} of clean speech signal. This indicates a low SNR in this frame, which is related to the fact that $\theta_{(1,i,q)}$ is above the threshold. On the other hand, in frame 62 (highlighted with blue arrow) the INS_{max} of noisy signal is close to INS_{max} of clean signal and distant from INS_{max} of noise. In this case, most of the energy content belongs to the speech and the $\theta_{(1,i,q)}$ is below the threshold.

D. Signal Reconstruction

After the detection of the most corrupted frames from the selected modes $\text{IMF}_i(t), i \geq I_{\text{min}}$, the preserved frames are added to the remaining modes, i.e., $\text{IMF}_i(t), i < I_{\text{min}}$. Finally, the reconstructed frames are concatenated considering Hanning window and 50% overlapping.

III. Experiments and Discussion

Several experiments are conducted to evaluate the proposed method in terms of quality and intelligibility improvement. Experiments are divided according to the signal of interest, i.e., speech and linear chirp. The first scenario considers 240 speech signals spoken by 16 male and 8 female speakers, sampled at 16 KHz and with 3 s average duration, selected from the TIMIT [27] database. For the second scenario, a set of 10 linear chirp signals are generated as in [28] with an exponent function to account for transmission loss.

Three underwater ambient noises are considered for both scenarios. Bubbles is collected from Freesound.org[1], Killer Whale from San Francisco Maritime National Park Association database[2], and Ocean Liner from ShipsEar database [29]. Signals of interest are corrupted considering three SNR values, -5 dB, 0 dB, and 5 dB, leading to 2250 tests. The underwater acoustic noises are captured by sensor networks, that is, by hydrophones present in the acoustic systems. Fig. 4 depicts spectrograms and INS values of the acoustic noises. Note that the noises energies are mainly concentrated at low frequencies, while the target signals contain relevant components at high frequencies. Moreover, acoustic noises present different non-stationarity, with $\text{INS}_{\text{max}} = 92$ for Bubbles and $\text{INS}_{\text{max}} = 47$ for Killer Whale. The Ocean Liner noise is classified as stationary since most of the INS values are below the threshold.

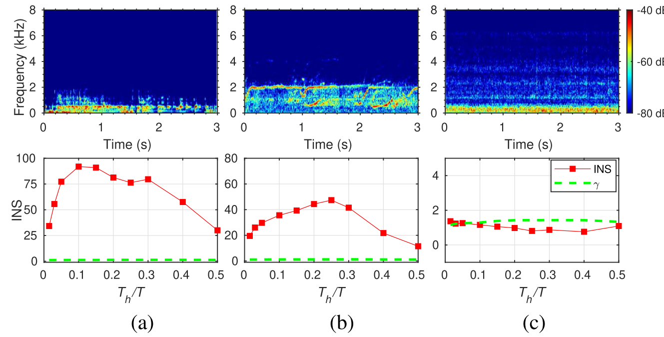

Fig. 4. Spectrograms (top) and INS (bottom) of (a) Bubbles, (b) Killer Whale, and (c) Ocean Liner. In lower part, the continuous red lines and green dashed lines indicate the value for INS and stationarity test threshold, respectively.

For the implementation of the EIF-NS method, 8 IMFs are attained with EEMD-IF using 50 ensembles and $\alpha = 2.5$. Five surrogates sequences are considered for the INS computation. Four speech enhancement solutions are adopted as baseline: EMDH [10] and NNESE [16] in the time domain, OMLSA [9] and UMMSE [11] in the frequency domain. The oversubtraction factor of the NNESE source code provided by the authors is set to 0.35 and 0.1 for quality and intelligibility improvement, respectively. For the second scenario, it is adopted the Hurst exponent threshold value of 0.5 instead of 0.9 for the EMDH [10], in order to achieve better quality improvement for chirp signals.

A. First Scenario: Speech Signal

The speech quality prediction is here objectively examined with the perceptual evaluation of speech quality (PESQ) [30] and the perceptual evaluation of audio quality (PEAQ) [31] measures. Table I presents the PESQ values achieved with unprocessed (UNP) and processed speech signals with the EIF-NS and competing enhancement methods. *Highlighted values correspond to the best results*. Note that EIF-NS achieves the best PESQ improvement for the non-stationary Bubbles and Killer Whale noises. It outperforms the competing time-domain solutions also for the stationary Ocean Liner noise. EIF-NS attains an overall PESQ value of 2.90, which is 0.14 and 0.19 greater than NNESE and UMMSE, respectively.

Fig. 5 depicts the objective quality results in terms of PEAQ scores. Once again, EIF-NS outperforms the competing time- and frequency-domain methods for the non-stationary noises. The spectral approaches achieve the best results for the stationary noise. EIF-NS obtains significant improvement when compared to the EMDH, which is also based on the time-frequency

[1] Available at http://www.freesound.org

[2] Available at https://maritime.org/sound

TABLE I
PESQ RESULTS WITH THE EIF-NS AND BASELINE METHODS

NOISES	SNR	UNP	EIF-NS	NNESE	EMDH	UMMSE	OMLSA
Bubbles $INS_{max} = 92$	-5 dB	2.60	**2.88**	2.79	2.62	2.63	2.55
	0 dB	2.86	**3.09**	3.00	2.87	2.87	2.84
	5 dB	3.13	**3.34**	3.25	3.14	3.14	3.12
Killer Whale $INS_{max} = 47$	-5 dB	2.02	**2.47**	2.34	2.07	1.96	1.81
	0 dB	2.57	**2.84**	2.69	2.57	2.44	2.34
	5 dB	2.85	**3.15**	2.98	2.85	2.77	2.70
Ocean Liner Stationary	-5 dB	2.12	2.46	2.21	2.12	2.44	**2.56**
	0 dB	2.51	2.85	2.61	2.51	2.87	**2.98**
	5 dB	2.92	3.07	3.00	2.92	3.24	**3.33**
Overall		2.62	**2.90**	2.76	2.63	2.71	2.69

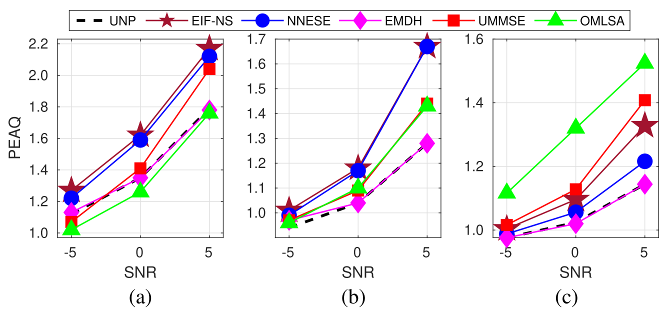

Fig. 5. Average PEAQ for (a) Bubbles, (b) Killer Whale and (c) Ocean Liner.

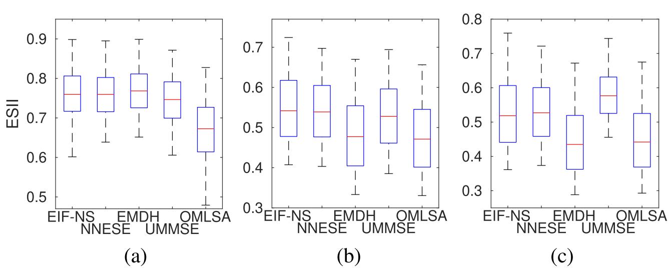

Fig. 6. ESII box-plots for (a) Bubbles, (b) Killer Whale and (c) Ocean Liner.

EMD analysis. This reinforces the contribution of the INS-based criterion in (5) to detect noise components.

The extended speech intelligibility index (ESII) [32] is here applied to evaluate speech intelligibility. Fig. 6 illustrates the ESII intelligibility prediction results considering SNR values of -5 dB, 0 dB, and 5 dB. It can be noted that, for the Bubbles noise, the time-domain EIF-NS, NNESE, and EMDH techniques attain better results than the spectral solutions. For Killer Whale, EIF-NS presents the highest average ESII values, followed by the NNESE. The UMMSE achieves the best average value for the stationary Ocean Liner noise. For this same noise source, the EIF-NS outperforms the other time-domain methods.

B. Second Scenario: Chirp Signal

The SNR and RMSE measures are considered for the evaluation of the proposed solution for chirp as target signal. Fig. 7 depicts the SNR improvement, i.e., the difference between output and input SNR. Table II shows the average RMSE results obtained from frames of 32 ms. Note that both measures show

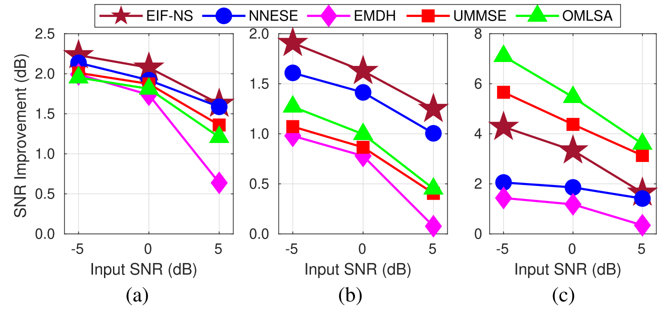

Fig. 7. The SNR improvement obtained with the EIF-NS and baseline techniques for (a) Bubbles, (b) Killer Whale and (c) Ocean Liner.

TABLE II
RMSE RESULTS [$\times 10^{-2}$] WITH THE EIF-NS AND BASELINE METHODS

NOISES	SNR	UNP	EIF-NS	NNESE	EMDH	UMMSE	OMLSA
Bubbles $INS_{max} = 92$	-5 dB	3.58	**3.18**	3.21	3.27	3.26	3.29
	0 dB	2.01	**1.82**	1.85	1.89	1.86	1.88
	5 dB	1.35	**1.08**	**1.08**	1.21	1.11	1.13
Killer Whale $INS_{max} = 47$	-5 dB	3.92	**3.30**	3.41	3.67	3.63	3.55
	0 dB	2.20	**1.92**	1.96	2.11	2.09	2.06
	5 dB	1.33	**1.13**	1.15	1.29	1.24	1.24
Ocean Liner Stationary	-5 dB	4.05	2.51	3.25	3.49	2.14	**1.82**
	0 dB	2.28	1.57	1.87	2.02	1.40	**1.23**
	5 dB	1.33	1.08	1.10	1.25	0.91	**0.86**

consistent results: the higher the SNR improvement, the lower the RMSE between the clean and enhanced signal. EIF-NS outperforms the baseline techniques for the non-stationary Bubbles and Killer Whale noises, followed by NNESE. For the Bubbles noise, EIF-NS and NNESE achieve SNR improvement of 2.2 dB and 2.1 dB for input SNR of -5 dB, respectively. For the Killer Whale, EIF-NS attains an SNR gain of 1.9 dB, against 1.6 dB of NNESE for the most severe noisy condition. The spectral OMLSA and UMMSE techniques outperform the other methods for Ocean Liner. For this noise, EIF-NS achieves the highest SNR improvement among the time-domain solutions, with an improvement of 4.3 dB for input SNR of -5 dB.

IV. CONCLUSION

This Letter introduces the EIF-NS signal enhancement approach based on EEMD-IF for non-stationary underwater acoustic environment. The estimation and selection of the most corrupted decomposition modes were applied in a frame-by-frame basis using the index of non-stationarity. Several experiments were conducted with speech and chirp signals corrupted by three underwater acoustic noises with different non-stationarity behavior. The results with speech signals scenario demonstrated that the proposed method outperforms four baseline solutions in terms of quality, particularly in the presence of severe non-stationary noisy conditions. The EIF-NS method achieved an average gain of 10.7% and 21.4% in terms of the PESQ and PEAQ measures, respectively. It also led to interesting speech intelligibility improvement for the non-stationary noises. For the chirp signals scenario, the novel approach also obtained the highest output SNR and lowest RMSE values in non-stationary noisy environments.

REFERENCES

[1] G. Wenz, "Acoustic ambient noise in the ocean: Spectra and sources," *J. Acoust. Soc. Amer.*, vol. 34, no. 12, pp. 1936–1956, 1962.

[2] S. Siddagangaiah, Y. Li, X. Guo, and K. Yang, "On the dynamics of ocean ambient noise: Two decades later," *Chaos: An Interdiscipl. J. Nonlinear Sci.*, vol. 25, no. 10, 2015, Art. no. 103117.

[3] W. Wilcock, K. Stafford, R. Andrew, and R. Odom, "Sounds in the ocean at 1–100 hz," *Ann. Rev. Mar. Sci.*, vol. 6, pp. 117–140, Jan. 2014.

[4] Y. Liu, H. Wang, L. Cai, X. Shen, and R. Zhao, "Fundamentals and advancements of topology discovery in underwater acoustic sensor networks: A review," *IEEE Sensors J.*, vol. 21, no. 19, pp. 21159–21174, Oct. 2021.

[5] Z. Huang, S. Li, X. Yang, and J. Wang, "OAE-EEKNN: An accurate and efficient automatic modulation recognition method for underwater acoustic signals," *IEEE Signal Process. Lett.*, vol. 29, no. 2, pp. 518–522, Feb. 2022.

[6] P. Dahl, J. Miller, D. Cato, and R. Andrew, "Underwater ambient noise," *Acoust. Today*, vol. 3, pp. 23–33, Jan. 2007.

[7] A. Trucco, R. Bozzano, E. Fava, S. Pensieri, A. Verri, and A. Barla, "A supervised learning approach for rainfall detection from underwater noise analysis," *IEEE J. Ocean. Eng.*, vol. 47, no. 1, pp. 213–225, Jan. 2022.

[8] M. Alves, R. Coelho, and E. Dranka, "Effective acoustic energy sensing exploitation for target sources localization in urban acoustic scenes," *IEEE Sensors Lett.*, vol. 4, no. 2, pp. 1–4, Feb. 2020.

[9] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, 2001.

[10] L. Zão, R. Coelho, and P. Flandrin, "Speech enhancement with EMD and hurst-based mode selection," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 5, pp. 899–911, May 2014.

[11] T. Gerkmann and R. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[12] N. Huang et al., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A: Math., Phys., Eng. Sci.*, vol. 454, pp. 903–995, 1998.

[13] A. Stallone, A. Cicone, and M. Materassi, "New insights and best practices for the successful use of empirical mode decomposition, iterative filtering and derived algorithms," *Nature Sci. Rep.*, vol. 10, no. 15161, pp. 1–15, 2020.

[14] R. Coelho and L. Zão, "Empirical Mode Decomposition Theory Applied to Speech Enhancement," in *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering and Processing*, R. V. Coelho, R. N. Queiroz, J. Romano, and C. Cavalcante, eds. Boca Raton, Florida, USA: CRC, 2015, ch. 5, pp. 123–153.

[15] E. Hurst, "Long-term storage capacity of reservoirs," *Trans. Amer. Soc. Civil Eng.*, vol. 116, no. 1, pp. 770–799, 1951.

[16] R. Tavares and R. Coelho, "Speech enhancement with nonstationary acoustic noise detection in time domain," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 6–10, Jan. 2016.

[17] D. Pastor and F.-X. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1545–1555, Apr. 2012.

[18] Z. Wu and N. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adaptive Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.

[19] L. Lin, Y. Wang, and H. Zhou, "Iterative filtering as an alternative algorithm for empirical mode decomposition," *Adv. Data Sci. Adaptive Anal.*, vol. 1, no. 4, pp. 543–560, 2009.

[20] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3459–3470, Jul. 2010.

[21] B. Woodward and H. Sari, "Digital underwater acoustic voice communications," *IEEE J. Ocean. Eng.*, vol. 21, no. 2, pp. 181–192, Apr. 1996.

[22] W. Lei, D. Wang, Y. Xie, B. Chen, X. Hu, and H. Chen, "Implementation of a high reliable chirp underwater acoustic modem," in *Proc. Oceans - Yeosu*, 2012, pp. 1–5.

[23] A. Cicone and H. Zhou, "Numerical analysis for iterative filtering with new efficient implementations based on FFT," *Numerische Mathematik*, vol. 147, no. 1, pp. 1–28, 2021.

[24] A. Cicone, "Iterative filtering as a direct method for the decomposition of nonstationary signals," *Numer. Algorithms*, vol. 85, no. 3, pp. 811–827, 2020.

[25] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[26] S. Chen, X. Dong, Z. Peng, W. Zhang, and G. Meng, "Nonlinear chirp mode decomposition: A variational method," *IEEE Trans. Signal Process.*, vol. 65, no. 22, pp. 6024–6037, Nov. 2017.

[27] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "TIMIT acoustic-phonetic continuous speech corpus," *Linguist. Data Consortium*, Philadelphia, PA, USA, 1993.

[28] H. Ou, J. S. Allen, and V. L. Syrmos, "Frame-based time-scale filters for underwater acoustic noise reduction," *IEEE J. Ocean. Eng.*, vol. 36, no. 2, pp. 285–297, Apr. 2011.

[29] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Appl. Acoust.*, vol. 113, pp. 64–69, 2016.

[30] "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Rec. P.862 Int. Telecommu. Uni*., 2001.

[31] M. Torcoli, T. Kastner, and J. Herre, "Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence," *IEEE/ACM Trans. Audio Speech. Lang. Process.*, vol. 29, no. 3, pp. 1530–1541, Mar. 2021.

[32] K. Rhebergen and N. Versfeld, "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *J. Acoust. Soc. Amer.*, vol. 117, pp. 2181–2192, 2005.