Realce de Sinais em Ambiente com Variações Acústicas Subaquáticas

A. Caldeira e R. Coelho

***Resumo*— Este artigo apresenta um estudo comparativo de métodos de realce de sinais acústicos na presença de ruídos submarinos. Para esta avaliação, serão utilizados os métodos OMLSA, UMMSE e NNESE, que são capazes de lidar com a não-estacionariedade dos ruídos, para realce de sinais de voz e de *chirp*. Experimentos são conduzidos considerando três ruídos acústicos não-estacionários, utilizando duas medidas de qualidade (PESQ e OQCM) e de inteligibilidade (STOI) para os sinais de voz e uma medida de qualidade (SegSNR) para o sinal *chirp*. Os resultados apontam que o NNESE apresenta os maiores aprimoramentos em ambos os cenários experimentais.**

***Palavras-Chave*— Realce de Sinais, acústica submarina, índice de não-estacionariedade**

***Abstract*— This paper presents a comparative study of acoustic signal enhancement methods in the presence of underwater noises. For this evaluation, the OMLSA, UMMSE and NNESE methods will be used, which are able to deal with nonstationary noises, for enhancing speech and chirp signals. Experiments are conducted considering three nonstationary acoustic noises, using two quality measures (PESQ and OQCM) and intelligibility measure (STOI) for speech signals and one quality measure (SegSNR) for chirp signal. The results show that NNESE presents higher improvements in both experimental scenarios.**

***Keywords*— Signal Enhancement, underwater acoustics, index of nonstationarity**

I. INTRODUÇÃO

Desde a década de 1970 [1], as soluções de realce de sinais têm sido propostas para atenuar os efeitos causados por ruídos acústicos. No contexto da acústica submarina, o realce de sinais pode abranger diversas aplicações, como estudos oceanográficos, exploração de petróleo *offshore*, comunicações e operações de defesa [2][3]. O realce de sinais de voz também é de vital importância na comunicação entre mergulhadores [4]. O principal desafio consiste em estimar as estatísticas do ruído em ambientes acústicos naturais, principalmente quando suas características variam ao longo do tempo, ou seja, são não-estacionários.

As técnicas de realce de sinais propostas na literatura podem ser classificadas como espectrais e temporais. Técnicas espectrais convencionais, como a subtração espectral [1] e mínimos erros quadráticos médios [5] utilizam a STFT (*short-time fourier transform*) e o VAD (*voice activity detector*) para estimar os componentes espectrais do ruído na ausência de voz. Estes métodos têm um bom desempenho quando os ruídos são estacionários, porém outros algoritmos são necessários para lidar com a não-estacionariedade dos ruídos. Alguns métodos, como MS (*minimum statistics*) [6] e IMCRA (*improved minima controlled recursive averaging*) [7] foram propostos para estimar o ruído com estas características, atualizando o espectro do ruído quadro a quadro, até mesmo na presença de voz. A solução UMMSE (*unbiased minimum mean-square error*) [8] visa extrair as variações espectrais de ruídos não-estacionários com um atraso menor que outros estimadores espectrais. Entretanto, estes métodos ainda apresentam baixo desempenho para a estimação de ruídos altamente não-estacionários.

As técnicas temporais, por outro lado, são geralmente baseadas na decomposição *wavelets* ou EMD (*empirical mode decomposition*). Diferentemente destas soluções, em [9], foi proposto o método NNESE (*non-stationary noise estimation for speech enhancement*) baseado em um estimador robusto [10] para estimar o desvio padrão do ruído quadro a quadro. Neste, as componentes mais afetadas pelo ruído são eliminadas e os valores restantes são atenuados com o desvio padrão do respectivo quadro. O NNESE apresentou interessantes resultados no aprimoramento da qualidade e inteligibilidade dos sinais de voz na presença de ruídos não-estacionários quando comparado a outros métodos competitivos [9].

Este trabalho apresenta um estudo comparativo de técnicas espectrais e temporais de realce de sinais acústicos na presença de ruídos subaquáticos naturais e não-estacionários. No ambiente acústico submarino, os ruídos podem ser classificados como naturais, como os sons gerados pelos animais marinhos, ondas, chuvas, terremotos submarinos, e os antrópicos, como embarcações, pistolas de ar (*air guns*) para obtenção de dados sísmicos submarinos e sonares ativos militares [11].

Experimentos foram realizados neste trabalho, sendo o primeiro visando realçar um sinal de voz, e outro para realçar um sinal *chirp*. Sinais *chirp* têm baixa sensibilidade ao efeito Doppler e boa capacidade de rejeição de interferências, sendo por isso comumente usados em aplicações sonar, em especial nas comunicações acústicas submarinas [12].

Para ambos os cenários, são considerados três ruídos acústicos naturais (*Bubbles*, *Killer Whale* e *Underwater Earthquake*) com diferentes graus de não-estacionariedade. Estes ruídos são utilizados para corromper o sinal acústico com valores de SNR variando entre -5 dB e 5 dB. Os métodos de realce OMLSA (*optimally-modified log-spectral amplitude*), UMMSE e NNESE são avaliados considerando quatro medidas objetivas, sendo duas medidas de qualidade (PESQ, *perceptual evaluation of speech quality* e OQCM, *overall quality composite measure*) [13][14] e uma de inteligibilidade (STOI, *extended short-time objective intelligibility*) [15] para o realce dos sinais de voz, e uma medida de qualidade (SegSNR,

A. Caldeira é mestrando do Programa de Pós-Graduação em Engenharia de Defesa do Instituto Militar de Engenharia (IME). O trabalho dos autores A. Caldeira e R. Coelho é desenvolvido no Laboratório de Processamento de Sinais Acústicos (LASP/IME) e parcialmente financiado pelo CNPq (308155/2019-0) e pela FAPERJ (203075/2016). E-mails: {awscaldeira,coelho}@ime.eb.br.

segmental signal-to-noise ratio) [16] para o realce do sinal *chirp*.

O restante deste trabalho está organizado da seguinte forma. Na Seção II, são introduzidos os métodos de realce de sinais acústicos. A Seção III aborda as medidas de qualidade e inteligibilidade utilizadas para uma avaliação objetiva dos métodos aplicados na Seção anterior. A descrição dos cenários experimentais e as discussões dos resultados são apresentadas na Seção IV. Por fim, na Seção V são apresentadas as principais conclusões deste trabalho.

II. MÉTODOS DE REALCE DE SINAIS ACÚSTICOS

Esta Seção descreve, de forma sucinta, os três métodos de realce de sinais acústicos adotados neste trabalho. Tanto o OMLSA quanto o UMMSE utilizam estimadores distintos para obter as componentes espectrais do ruído, enquanto o NNESE emprega o algoritmo DATE (*d-dimensional trimmed estimator*) [10] para estimação robusta do desvio padrão do ruído.

A. OMLSA

Este método emprega o IMCRA para atualização das estimativas do espectro de potência dos ruídos. Para este fim, o IMCRA realiza duas iterações de suavização do espectro de potência do sinal ruidoso e localização por estatísticas mínimas.

A primeira iteração fornece uma detecção de atividade de voz aproximada em cada banda de frequência. A suavização na segunda exclui componentes de voz mais significantes, resultando em uma maior robustez das estatísticas mínimas durante atividade de voz. Os valores mínimos da estimação suavizada do espectro de potência do sinal ruidoso são multiplicados por um fator de compensação de viés, que pode ser obtido de maneira empírica.

Após a implementação do IMCRA, o algoritmo OMLSA [17] é utilizado para obter o espectro do sinal de voz através da minimização do erro quadrático médio entre os logaritmos das magnitudes espectrais dos sinais de voz limpo e realçado. A função ganho que resulta na amplitude espectral da reconstrução ótima da voz é definida em [17] como:

$$G_{OMLSA}(\kappa,\tau) = G_{LSA}(\kappa,\tau)^{p(\kappa,\tau)} G_{min}^{1-p(\kappa,\tau)}, \qquad (1)$$

sendo τ e κ os índices de quadro e frequência, respectivamente, $G_{LSA}(\kappa,\tau)$ um ganho calculado em função do SNR *a priori* e deduzido em [18], e G_{min} um limiar mínimo para o ganho e correspondente à -25 dB [17].

B. UMMSE

O método UMMSE é derivado da técnica proposta em [19] para estimação das componentes espectrais do ruído a partir da minimização dos erros médios quadráticos. Diferentemente do IMCRA, no UMMSE não é necessário captar informações de vários quadros anteriores para estimação espectral do ruído, resultando em um menor atraso na captação das variações espectrais de ruídos não-estacionários. Outra vantagem do UMMSE sobre o IMCRA é que não é necessário calcular um fator de compensação de viés.

A estimação do espectro de potência do ruído pode ser atualizado de um quadro para outro através da seguinte equação:

$$|\hat{N}(\kappa,\tau)|^2 = \alpha_p |\hat{N}(\kappa,\tau-1)|^2 + (1-\alpha_p) E[|N(\kappa,\tau)|^2 \mid Y(\kappa,\tau)], \quad (2)$$

sendo α_p uma constante de suavização. A estimação do periodograma do ruído $E[|N(\kappa,\tau)|^2 \mid Y(\kappa,\tau)]$ depende das probabilidades de presença e ausência da voz e do espectro de potência do ruído calculado no quadro anterior.

Após a estimação das componentes espectrais do ruído, o espectro do sinal de voz é obtido pela supressão destes componentes através de uma técnica baseada no filtro de Wiener apresentada em [20]. O ganho de Wiener depende do SNR *a priori*, que pode ser estimado pelo método da direção direta [5].

C. NNESE

O método NNESE é composto de três etapas. A primeira etapa consiste na estimação do desvio padrão do ruído $\hat{\sigma_q}$ através do algoritmo DATE [10], que é aplicado em segmentos de 32 ms do sinal. Na segunda etapa, as componentes do ruído são selecionadas a partir de um limiar $y(b_q)$ na amplitude do sinal derivado da etapa anterior. Assim, componentes cujas amplitudes são inferiores a este limiar são tratadas como ruído. Por fim, na última etapa é feita a reconstrução do sinal de voz. Dado o quadro q com $k = 1, ..., K$ amostras, os valores de amplitude do sinal estimado de voz são dados pela equação

$$\tilde{y}_q = \begin{cases} y_q(k) - \alpha\hat{\sigma_q}, & \text{se } y_q(k) \geq y(b_q). \\ \beta y_q(k), & \text{caso contrário.} \end{cases} \qquad (3)$$

sendo α o fator de subtração para a reconstrução do sinal acústico e $\beta = 1-\alpha$ o fator de piso para valores negativos de amplitude. Para atuação do NNESE com o sinal de voz, foram utilizados valores de α = 0,35 e α = 0,1 para o aprimoramento da qualidade e inteligibilidade, respectivamente. Para o sinal *chirp* foi adotado α = 0,65.

III. MEDIDAS OBJETIVAS DE QUALIDADE E INTELIGIBILIDADE

Esta Seção apresenta as quatro medidas objetivas para avaliação do desempenho dos métodos de realce. As medidas PESQ, OQCM e SegSNR são utilizadas para avaliar a melhora na qualidade do realce dos sinais acústicos. A medida STOI é empregada para examinar a inteligibilidade da voz.

A. PESQ

A medida PESQ, inicialmente desenvolvida para avaliar a qualidade em codificadores de voz e canais telefônicos de banda estreita, é também largamente empregada para técnicas de realce. Os sinais de voz limpo e degradado são mapeados para o domínio de tempo-frequência, e a diferença entre eles fornece uma medida de erro audível. A medida PESQ é convertida em uma medida subjetiva MOS (*mean opinion score*), cujos valores variam de 1,0 (ruim) a 4,5 (sem distorção).

Neste trabalho, com o propósito de prover um resultado mais consistente entre a melhora observada no sinal processado e a

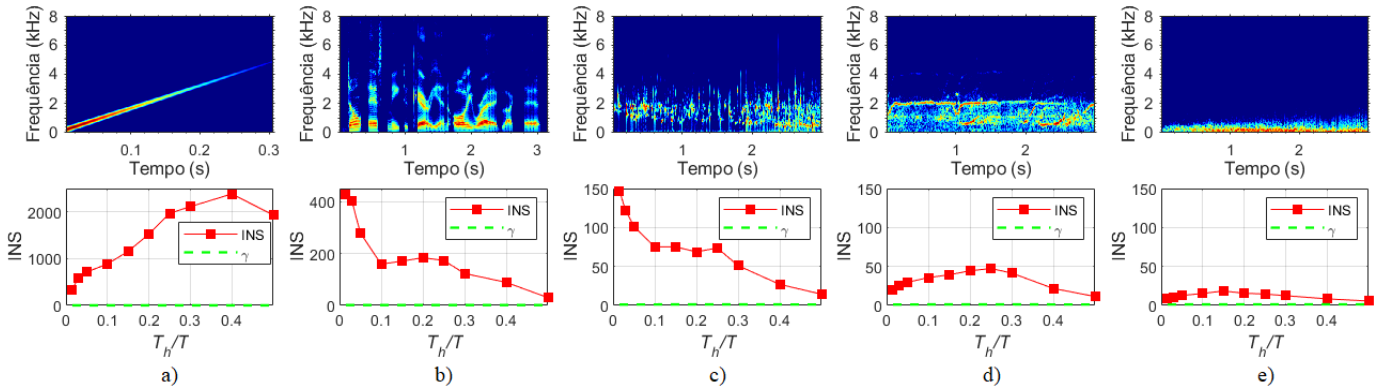

Fig. 1: Espectrograma e INS do (a) sinal *chirp*, (b) sinal de voz, (c) *Bubbles*, (d) *Killer Whale* e (e) *Underwater Earthquake*.

medida de qualidade, foram selecionados 30% dos quadros mais relevantes à melhora na qualidade para o cálculo da medida.

B. OQCM

A medida OQCM, proposta em [14], estabelece uma combinação linear de medidas de qualidade existentes na literatura com o intuito de obter maior correlação com os testes subjetivos. Neste trabalho, a correlação entre as medidas objetivas e testes subjetivos foi estudada com sinais de voz realçados por treze algoritmos distintos de realce de voz. As medidas subjetivas avaliadas nos experimentos foram a distorção do sinal de voz, distorção do ruído e qualidade total do sinal. Os autores demonstraram que as medidas PESQ, LLR (*log-likelihood ratio*) [21] e WSS (*weighted spectral slope*) [22] foram as que apresentaram o maior coeficiente de correlação com os testes subjetivos. A medida OQCM, então, é calculada conforme a equação a seguir.

$$\text{OQCM} = 1{,}594 + 0{,}805\text{PESQ} - 0{,}512\text{LLR} - 0{,}007\text{WSS}. \quad (4)$$

C. SegSNR

A razão sinal-ruído segmental consiste em obter a média entre os valores de SNR, em dB, calculados em quadros de curta duração do sinal de voz. Seja $x(t)$ um sinal de voz limpo, e $\hat{x}(t)$ uma versão corrompida ou distorcida deste mesmo sinal, a SegSNR de $\hat{x}(t)$ é estimada conforme equação a seguir :

$$\text{SegSNR} = \frac{10}{Q}\sum_{\tau=0}^{Q-1}\log\frac{\sum_{t=\tau T_{sh}}^{\tau T_{sh}+T_d-1} x^2(t)}{\sum_{t=\tau T_{sh}}^{\tau T_{sh}+T_d-1}[x(t)-\hat{x}(t)]^2} \quad (5)$$

Sendo T_d a quantidade de amostras para cada quadro, T_{sh} o deslocamento, em amostras, entre quadros consecutivos e Q o total de quadros. Para cômputo da SegSNR, os valores obtidos de SNR em cada quadro são limitados entre -10 dB e 35 dB.

D. STOI

A medida STOI [15] estima a degradação na inteligibilidade da voz causada por algoritmos de supressão de ruídos através do cálculo do coeficiente de correlação entre os espectros dos sinais limpo e realçado. O sinal de voz limpo e corrompido são reamostrados a 10 kHz e divididos em janelas de Hamming de 256 amostras e com 50% de sobreposição.
Na sequência, aplica-se DFT com 512 pontos em cada quadro. Estes pontos resultantes da DFT são agrupados em 15 bandas cujas frequências centrais variam de 150 Hz a 4300 Hz, com três bandas por oitava. Para cada quadro τ e banda j, uma medida de inteligibilidade intermediária $\text{STOI}_{(j,\tau)}$ é definida como o coeficiente de correlação entre os vetores de envoltória temporal obtidos para a voz limpa e corrompida.
Por fim, a medida STOI é calculada a partir da média de todos os valores de $\text{STOI}_{(j,\tau)}$:

$$\text{STOI} = \frac{1}{15Q}\sum_{j=1}^{15}\sum_{\tau=1}^{Q}\text{STOI}_{(j,\tau)}. \quad (6)$$

IV. EXPERIMENTOS: RESULTADOS E DISCUSSÃO

Nesta Seção, são apresentados os resultados dos dois experimentos realizados neste trabalho. O primeiro cenário consiste em um subconjunto de 10 áudios de voz de 24 locutores, sendo 16 homens e 8 mulheres, extraídos da base TIMIT [23], totalizando 240 áudios. Cada áudio tem uma frequência de amostragem de 16 kHz e uma duração média de 3 segundos. O segundo cenário experimental consiste em um sinal contendo 10 *chirps* lineares com um decaimento exponencial considerando perdas no sinal [3] e com duração total de 3,125 segundos. A expressão para cada *chirp* é dada por:

$$s[x] = \sin(2\pi f_i x + \frac{\pi(f_f - f_i)x^2}{T}).e^{-x/\beta}, \quad (7)$$

sendo f_i, f_f e T, respectivamente, as frequências inicial e final e a duração do sinal *chirp*. Neste trabalho, foi proposto f_i = 100 Hz, f_f = 5 kHz, T = 312,5 ms e β = 0,067.
Três ruídos acústicos submarinos, *Bubbles*, *Killer Whale* e *Underwater Earthquake*, são adicionados a estes sinais acústicos, considerando cinco valores diferentes de SNR: -5 dB, -3 dB, 0 dB, 3 dB e 5 dB.
Os espectrogramas de um sinal *chirp*, sinal de voz e dos ruídos subaquáticos adotados neste estudo podem ser observados na Figura 1. O ruído *Bubbles* foi obtido na base de dados da Freesound.org[1], e os ruídos *Killer Whale* e *Underwater Earthquake* foram extraídos da base de dados da *San Francisco Maritime National Park Association*[2].

[1]Disponível em http://www.freesound.org
[2]Disponível em https://maritime.org/sound

A. *Índices de não-estacionariedade*

O índice de não-estacionariedade (INS) [24] foi adotado para a análise da não-estacionariedade dos sinais e ruídos. Esta medida é obtida a partir da comparação do sinal de voz com seus referenciais estacionários (*surrogates*). O INS é então obtido de acordo com uma escala de observação T_h/T, que estabelece uma razão entre o tamanho da janela de tempo adotada na análise espectral (T_h) e a duração total do sinal (T). O valor de γ indica o limiar do teste de estacionariedade, calculado para cada valor de janela T_h, considerando uma precisão de 95%. Os ruídos são não-estacionários quando o INS é maior do que o limiar.

Para este trabalho, os seguintes critérios foram adotados para a classificação dos ruídos de acordo com o INS máximo (INS_{max}):

- $INS_{max} > 100\gamma$: altamente não-estacionário;
- $20\gamma < INS_{max} \leq 100\gamma$: não-estacionário; e
- $\gamma < INS_{max} \leq 20\gamma$: moderadamente não-estacionário.

Na Figura 1, segundo o INS máximo, o sinal *chirp*, sinal de voz e o ruído *Bubbles* são classificados como altamente não-estacionários, o ruído *Killer Whale* não-estacionário e o ruído *Underwater Earthquake* moderadamente não-estacionário, com o INS_{max} atingindo 2384, 428, 147, 47 e 19, respectivamente.

B. *Experimento 1: Realce para sinais de voz*

1) Resultados da Qualidade: A Tabela I apresenta os resultados obtidos com a PESQ, tanto para os métodos de realce quanto para os sinais UNP (*unprocessed*). O NNESE apresentou os maiores incrementos na qualidade para os ruídos *Bubbles* e *Killer Whale*, obtendo a média PESQ de 2,78 e 2,67, respectivamente, sendo o maior aprimoramento observado no ruído não-estacionário, de 0,17 com relação ao UNP. Por outro lado, o UMMSE teve melhores resultados com o ruído *Underwater Earthquake*, moderadamente não-estacionário, obtendo a média PESQ de 3,11, o que significa um aprimoramento de 0,18 com relação ao UNP. Nota-se que, para os ruídos *Bubbles* e *Killer Whale*, somente o NNESE foi capaz de aprimorar a qualidade da voz. A maior média geral de qualidade, segundo a medida PESQ, também foi obtida pelo NNESE, de 2,83, seguido por UMMSE e OMLSA.

A Figura 2 mostra as curvas *box-plot* para a OQCM, que confirmam os resultados da Tabela I. Novamente, o NNESE obteve o melhor aprimoramento para os ruídos *Bubbles* e *Killer Whale*, enquanto o UMMSE conseguiu o melhor resultado para o ruído *Underwater Earthquake*.

2) Resultados da Inteligibilidade: A Tabela II apresenta os resultados da medida de inteligibilidade STOI para os sinais UNP, enquanto a Figura 3 mostra os aprimoramentos do STOI com os métodos de realce em relação ao UNP (ΔSTOI). O NNESE teve o melhor aprimoramento, seguido pelo UMMSE. Os maiores incrementos foram observados no ruído *Bubbles*, em que o NNESE obteve 1% de incremento com a SNR de -5 dB, seguido pelo *Killer Whale*, com um máximo de 0,5% obtido pelo NNESE com a SNR de 3 dB. Apesar da alta inteligibilidade dos sinais corrompidos pelo ruído *Underwater Earthquake*, observa-se que os métodos de realce aprimoraram a inteligibilidade para valores positivos de SNR.

TABELA I: Resultados da PESQ para diferentes ruídos e SNR

RUÍDOS	SNR	UNP	NNESE	UMMSE	OMLSA
Bubbles $INS_{max} = 147$	-5 dB	2,25	**2,39**	2,22	2,10
	-3 dB	2,44	**2,57**	2,40	2,29
	0 dB	2,69	**2,80**	2,65	2,54
	3 dB	2,90	**3,01**	2,87	2,79
	5 dB	3,02	**3,12**	3,00	2,92
MÉDIA		2,66	**2,78**	2,63	2,53
Killer Whale $INS_{max} = 47$	-5 dB	2,02	**2,34**	1,96	1,81
	-3 dB	2,33	**2,48**	2,20	2,07
	0 dB	2,57	**2,69**	2,44	2,34
	3 dB	2,73	**2,86**	2,64	2,56
	5 dB	2,85	**2,98**	2,77	2,70
MÉDIA		2,50	**2,67**	2,40	2,30
Underwater Earthquake $INS_{max} = 19$	-5 dB	2,60	2,72	**2,76**	2,49
	-3 dB	2,72	2,84	**2,90**	2,65
	0 dB	2,93	3,04	**3,11**	2,90
	3 dB	3,12	3,23	**3,33**	3,13
	5 dB	3,27	3,38	**3,47**	3,29
MÉDIA		2,93	3,04	**3,11**	2,89
MÉDIA GERAL		2,70	**2,83**	2,71	2,57

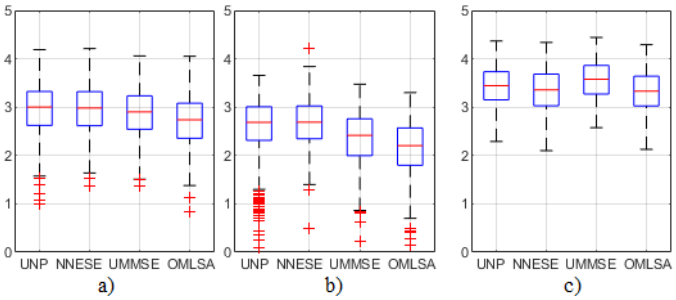

Fig. 2: Curvas *box-plot* com os resultados OQCM dos ruídos: (a) *Bubbles*, (b) *Killer Whale* e (c) *Underwater Earthquake*.

TABELA II: Resultados STOI para sinais UNP

	STOI				
SNR (dB)	-5	-3	0	3	5
Bubbles	0,609	0,655	0,721	0,782	0,818
Killer Whale	0,625	0,674	0,743	0,805	0,841
Underwater Earthquake	0,744	0,761	0,787	0,812	0,828

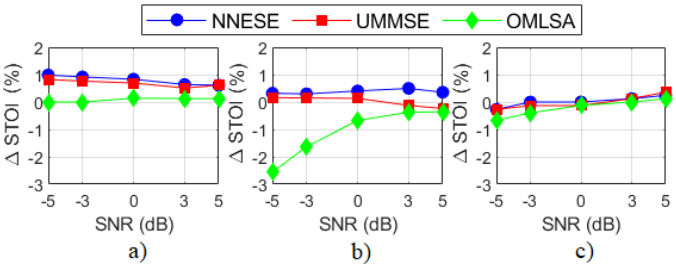

Fig. 3: Incremento da medida STOI com relação ao sinal corrompido pelos ruídos: (a) *Bubbles*, (b) *Killer Whale* e (c) *Underwater Earthquake*.

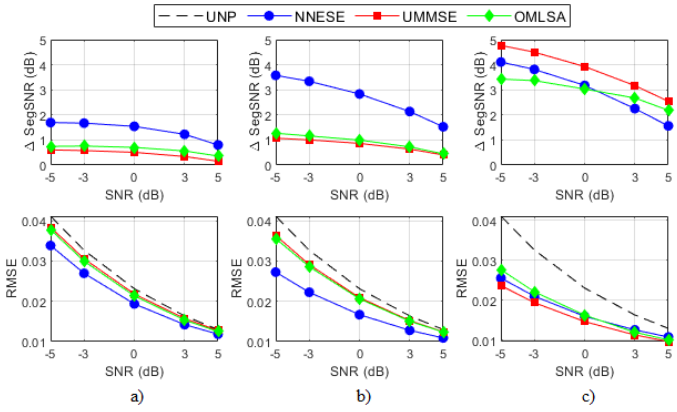

Fig. 4: Resultados da SegSNR e RMSE para os ruídos: (a) *Bubbles*, (b) *Killer Whale* e (c) *Underwater Earthquake*.

C. Experimento 2: Realce para sinal chirp

1) Resultados da Qualidade: A Figura 4 ilustra os resultados da medida SegSNR para os métodos de realce aplicados ao sinal *chirp*. Além disso, o RMSE (*root mean square error*) também foi empregado como em [2] para calcular o erro entre o sinal realçado e o de referência. Como esperado, foi constatada uma relação direta entre as medidas SegSNR e RMSE, uma vez que, quanto maior o aprimoramento da SegSNR, menor o RMSE do *chirp* após o realce. Assim como para os sinais de voz, o NNESE obteve os maiores ganhos na SegSNR para os ruídos com maior INS, enquanto o UMMSE obteve melhores resultados com o ruído *Underwater Earthquake*, moderadamente não-estacionário.
Nota-se, também, que os menores incrementos foram obtidos para o ruído *Bubbles*. Neste caso, o maior incremento foi de 1,7 dB com o NNESE com SNR de -5 dB. Os maiores aprimoramentos foram observados no ruído *Underwater Earthquake*, em que o UMMSE obteve um ΔSegSNR de 5 dB com SNR de -5 dB. Por fim, o NNESE obteve a maior média geral no aprimoramento do SegSNR, com 2,35 dB, seguido por UMMSE, com 1,67 dB, e OMLSA, com 1,49 dB.

V. CONCLUSÃO

Este artigo apresentou um estudo comparativo de três métodos de realce de sinais acústicos para aplicações em ambiente acústico submarino. Experimentos distintos foram conduzidos utilizando três ruídos subaquáticos não-estacionários considerando um cenário de avaliação que incluiu sinais de voz e *chirp*. Os resultados mostraram que o método NNESE apresentou o melhor aprimoramento da qualidade e inteligibilidade dos sinais acústicos, especialmente para os ruídos com maiores índices de não-estacionariedade. Além disso, os resultados de qualidade obtidos pelo NNESE também demonstraram interessante aprimoramento para o sinal *chirp* em relação aos demais métodos competitivos.

REFERÊNCIAS

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[2] Y. Y. Al-Aboosi and A. Z. Sha'ameri, "Improved signal de-noising in underwater acoustic noise using s-transform: A performance evaluation and comparison with the wavelet transform," *Journal of Ocean Engineering and Science*, vol. 2, no. 3, pp. 172–185, 2017.

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

[4] B. Woodward and H. Sari, "Digital underwater acoustic voice communications," *IEEE Journal of Oceanic Engineering*, vol. 21, no. 2, pp. 181–192, 1996.

[5] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[6] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 5, pp. 504–512, 2001.

[7] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, 2003.

[8] T. Gerkmann and R. C. Hendriks, "Unbiased mmse-based noise power estimation with low complexity and low tracking delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383–1393, 2012.

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

[10] D. Pastor and F. Socheleau, "Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1545–1555, 2012.

[11] J. A. Hildebrand, "Anthropogenic and natural sources of ambient noise in the ocean," *Marine Ecology Progress Series*, vol. 395, pp. 5–20, 2009.

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

[13] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, pp. 749–752 vol.2, 2001.

[14] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in *In: Proc. of INTERSPEECH*, 2006.

[15] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[16] J. H. L. Hansen and B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Proceedings of the International Conference on Speech and Language Processing*, pp. 2819–2822, 1998.

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

[18] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.

[19] R. C. Hendriks, R. Heusdens, and J. Jensen, "Mmse based noise psd tracking with low complexity," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4266–4269, 2010.

[20] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 2, pp. 629–632 vol. 2, 1996.

[21] S. R. Quackenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality*. Ellis Horwood Series in Artificial Intelligence, Prentice Hall, 1988.

[22] D. Klatt, "Prediction of perceived phonetic distance from critical-band spectra: A first step," in *ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 7, pp. 1278–1281, 1982.

[23] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "Darpa timit acoustic phonetic continuous speech corpus cdrom," 1993.

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