APLICAÇÃO DE TÉCNICAS DE REALCE DE VOZ E MÁSCARAS ACÚSTICAS PARA APRIMORAMENTO DAS COMUNICAÇÕES POR VOZ NO TELEFONE SUBMARINO

Application of speech enhancement techniques and acoustic masks for improving speech communications on the underwater telephone

Antônio Walkir Sibanto Caldeira[1], Rafael Marinati de Barros Martiny[2], Jefferson Osowsky[3], Renato Peres Vio[4]

Resumo: Este artigo apresenta um estudo sobre a aplicação de métodos de realce e máscaras acústicas, usualmente empregadas na acústica aérea, no aprimoramento dos aspectos perceptuais da voz em um telefone acústico submarino. Para esta análise, foi desenvolvida uma base própria de dados de voz composta por palavras empregadas na comunicação entre os navios da Marinha do Brasil, utilizando locutores nativos de língua portuguesa. Foram propostos dois cenários experimentais, ambos distorcidos por dois tipos de ruídos acústicos distintos. No primeiro cenário, os sinais de voz foram corrompidos somente com ruído ambiente, e os métodos de realce OMLSA e UMMSE foram avaliados por meio de medidas de qualidade (SegSNR) e inteligibilidade (ESII). No segundo cenário, foi incluído o efeito da reverberação, considerando seis cenários reverberantes, e os métodos anteriores, juntamente com uma máscara acústica ideal (IRM) e cega (BRM), foram avaliados pela medida SRMR. Em ambos os cenários experimentais, o OMLSA apresentou melhores resultados quando comparado às soluções cegas.

Palavras-chave: Acústica submarina. Telefone submarino. Realce de voz. Máscara acústica.

Abstract: This paper presents a study focused on the application of enhancement methods and acoustic masks, generally used in airborne acoustics, to improve the perceptual aspects of speech in an underwater acoustic telephone. For this analysis, an own speech database was developed with words used among the ships of the Brazilian Navy, with native portuguese speakers. Two experimental scenarios were proposed, both corrupted by two distinct acoustic noises. In the first, speech signals were corrupted only with ambient noise, and the OMLSA and UMMSE enhancement methods were evaluated by a quality (SegSNR) and intelligibility (ESII) measure. In the second, the effect of reverberation was included, considering six reverberant scenarios, and the previous methods, in addition to ideal (IRM) and blind (BRM) acoustic mask, were evaluated by the SRMR measure. In both experimental scenarios, OMLSA showed better results when compared to blind solutions.

Keywords: Underwater acoustics. Underwater telephone. Speech enhancement. Acoustic mask.

1. Capitão-Tenente (EN). Mestre em Engenharia de Defesa pelo Instituto Militar de Engenharia. Ajudante da Divisão de Comunicações Submarinas do Instituto de Estudos do Mar Almirante Paulo Moreira, Niterói, RJ – Brasil. E-mail: antonio.caldeira@marinha.mil.br

2. Primeiro-Tenente (RM2-EN). Mestre em Engenharia Elétrica pelo Instituto Militar de Engenharia. Assessor para Sistemas de Comando e Controle/Relação com ICTs do Programa SisGAAz da Diretoria de Gestão de Programas da Marinha, Rio de Janeiro, RJ – Brasil. E-mail: rafael.martiny@marinha.mil.br

3. Analista de Pesquisa e Desenvolvimento do Departamento de Inovação na OceanPact Serviços Marítimos, Rio de Janeiro, RJ – Brasil. Email: jefferson.osowsky@oceanpact.com

4. Capitão-de-Fragata (EN). Doutor em Engenharia Acústica pela Naval Postgraduate School da US Navy. Professor do Programa de Pós-graduação em Acústica Submarina e Encarregado da Divisão de Engenharia Acústica do Instituto de Estudos do Mar Almirante Paulo Moreira, Niterói, RJ – Brasil. E-mail: peres.vio@marinha.mil.br

1. INTRODUÇÃO

A comunicação acústica submarina desempenha um papel fundamental em uma ampla gama de atividades humanas e naturais, atraindo a atenção de diversas instituições mundo afora devido às suas aplicações de interesse dual, voltadas para a exploração dos recursos marinhos, monitoramento ambiental, pesquisa científica e operações militares. Dentre os diversos sistemas de comunicação, destacam-se aqueles que empregam o sinal de voz, frequentemente adotados em situações de elevada criticidade, nas quais uma mensagem mal compreendida ou não recebida pode levar a sérios riscos operacionais. Mergulhadores utilizam transdutores acoplados às suas máscaras para manter contato contínuo com sua equipe submersa ou na superfície (WOODWARD; SARI, 1996). No âmbito militar, os telefones submarinos são cruciais para salvamento e resgate de submarinos em emergência e são regulados pela norma internacional *Material Interoperability Requirements for Submarine Escape and Rescue* (OTAN STANAG 1475).

A recepção de um sinal de voz que trafega no ambiente acústico submarino com boa qualidade e inteligibilidade[1] é um grande desafio em razão das particularidades do canal submarino. Durante a propagação, as ondas sonoras são atenuadas em virtude do espalhamento geométrico do pulso acústico, à medida que se afastam da fonte, dos processos de absorção e das interações dessas ondas com o fundo e a superfície do oceano. Essas interações resultam em múltiplas reflexões do sinal, causando o indesejado efeito de reverberação, que pode comprometer a compreensão da mensagem recebida. Soma-se a isso a dificuldade em estimar a resposta ao impulso do canal, que varia principalmente em função dos parâmetros que afetam o perfil de velocidade do som no mar (temperatura, salinidade e pressão da água) e da geometria do canal (distância, profundidade e batimetria).

Além disso, o sinal de interesse também sofre distorções do ruído ambiente, cuja multiplicidade de fontes e a natureza não estacionária podem corromper severamente a qualidade e inteligibilidade da mensagem. No ambiente submarino, os ruídos podem ser classificados como naturais ou antropogênicos. Os ruídos naturais dividem-se em bióticos, gerados pela vida marinha, como cardume de peixes, cetáceos e camarões, e abióticos, causados por atividades sísmicas, ondas e chuvas. Dentre os ruídos antropogênicos, destacam-se os provenientes de embarcações, canhões de ar (*airguns*) e sonares ativos.

Esses desafios exigem a implementação de técnicas capazes de aprimorar a qualidade e a inteligibilidade dos sinais acústicos de interesse, mitigando os efeitos da reverberação e dos ruídos acústicos, garantindo a robustez e a confiabilidade das comunicações. Nesse contexto, métodos de realce de voz têm sido estudados com o objetivo de atenuar as distorções causadas pelos ruídos nos sinais de interesse, visando ao aumento da qualidade do sinal (CALDEIRA; COELHO, 2021). Máscaras acústicas também são mencionadas na literatura como soluções para simular o sistema auditivo humano e melhorar a inteligibilidade de sinais em cenários acústicos reverberantes e ruidosos (MARTINY; ALCÂNTARA; COELHO, 2022). Como o emprego dessas soluções tem sido majoritariamente voltado para a acústica aérea, este trabalho realizará um estudo sobre a aplicação dessas técnicas no cenário acústico submarino, considerando as distorções e variações causadas pelo ruído ambiente e pela reverberação, além das especificidades do processamento de sinais da voz dos telefones submarinos.

2. OBJETIVOS

O objetivo principal deste trabalho foi analisar e comparar os resultados obtidos por métodos de realce e máscaras acústicas, originalmente desenvolvidos para a acústica aérea, no aprimoramento da qualidade e inteligibilidade dos sinais de voz no contexto da comunicação acústica submarina utilizando o Telefone Submarino. Para este estudo, foram considerados diferentes graus de reverberação da voz e ruídos acústicos submarinos com características temporais e espectrais variadas. O objetivo específico deste trabalho foi desenvolver um banco de dados próprio contendo palavras empregadas na comunicação entre os meios navais da Marinha do Brasil, incluindo numerais, com locutores nativos da língua portuguesa, para a realização dos experimentos.

3. METODOLOGIA

3.1. COMUNICAÇÃO ACÚSTICA POR TELEFONE SUBMARINO

A Figura 1 apresenta um esquema da comunicação acústica por voz no Telefone Submarino adotada neste artigo. Os sinais

[1] A inteligibilidade reflete o quanto uma mensagem acústica é compreensível, podendo ser avaliada objetivamente pelo número de palavras ou fonemas identificados corretamente por um locutor.

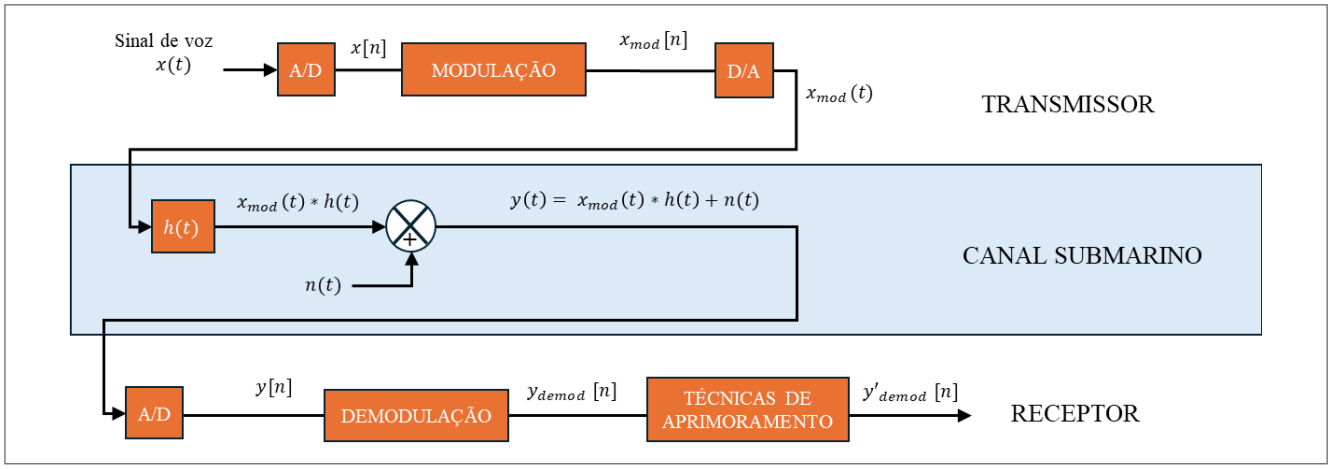

Figura 1. Esquemático da comunicação por voz no Telefone Submarino.

de áudio captados pelo microfone, $x(t)$, são digitalizados e modulados analogicamente em amplitude com portadora suprimida de banda única (*Amplitude Modulation Single Sideband - Suppressed Carrier* – AM SSB-SC), na frequência de 8,0875 kHz. Essa modulação é a mais comumente empregada nas comunicações submarinas por voz (OTAN STANAG 1475). O sinal modulado é então convertido para sinal analógico e transmitido por uma fonte acústica no canal submarino, onde é distorcido pelas múltiplas reflexões no fundo e na superfície, além de sofrer a adição de ruído ambiente. Sendo assim, o sinal $y(t)$, captado pelo hidrofone no sistema de recepção, pode ser modelado matematicamente como (Equação 1):

$$y(t) = x_{mod}(t) * h(t) + n(t), \quad (1)$$

sendo $x_{mod}(t)$ o sinal de voz modulado a partir de $x(t)$, $h(t)$ a resposta ao impulso do canal e $n(t)$ o ruído ambiente. No sistema de recepção, o sinal é demodulado e passa por um filtro passa-faixa de 300Hz a 3kHz, que corresponde à faixa de frequência adotada na telefonia analógica convencional para comunicação submarina. Nos experimentos propostos neste trabalho, as distorções causadas pela reverberação e pelo ruído ambiente no canal submarino foram simuladas com recursos computacionais, e os métodos para mitigar os efeitos da reverberação e do ruído ambiente foram aplicados após a etapa de filtragem.

3.2. BANCO DE DADOS

Para este estudo, foi desenvolvido um banco de dados de voz próprio, composto por 32 locutores voluntários, sendo 12 mulheres e 20 homens, que pronunciaram 36 palavras em áudios com duração de 2 segundos. O conjunto de palavras inclui as 26 letras do alfabeto fonético da OTAN, de Alfa a Zulu, e 10 algarismos em português, de Zero a Nove, utilizados nos meios navais. As gravações foram realizadas em uma sala fechada, com interferência mínima de reverberação ou ruído ambiente. Para a captura dos áudios, foi utilizada uma mesa de som Behringer Xenyx qx1002usb e um microfone Shure SM58, com frequência de amostragem de 48 kHz. Com o objetivo de analisar as distorções provocadas pelos ecos nas palavras subsequentes em cenários com reverberação, todas as palavras gravadas foram concatenadas em um único sinal de áudio para cada locutor.

3.3. CENÁRIOS EXPERIMENTAIS

Dois ruídos acústicos submarinos foram selecionados para corromper o sinal de voz: Biológico e Chuva. O Ruído Biológico foi gravado em uma estação submarina pertencente ao Instituto de Estudos do Mar Almirante Paulo Moreira, localizada próxima ao costão rochoso em Arraial do Cabo-RJ, sendo predominante os sons gerados por animais invertebrados marinhos presentes no costão, como o camarão-estalo. O Ruído Chuva foi obtido da base de dados do *Discovery of Sound in the Sea*. Esses ruídos foram utilizados para corromper os sinais de interesse em níveis de *signal-to-noise ratio* (SNR) de -5, 0 e 5 dB. A SNR foi calculada considerando a energia do ruído na mesma faixa de frequência dos sinais de voz modulados. A Figura 2 ilustra os espectrogramas e os índices de não-estacionariedade (INS) dos ruídos, calculados na faixa de frequência que corrompe a voz modulada durante 3 segundos do ruído. O INS é uma medida objetiva para quantificar a não-estacionariedade de um sinal (BORGNAT *et al.*, 2010). A curva em vermelho representa

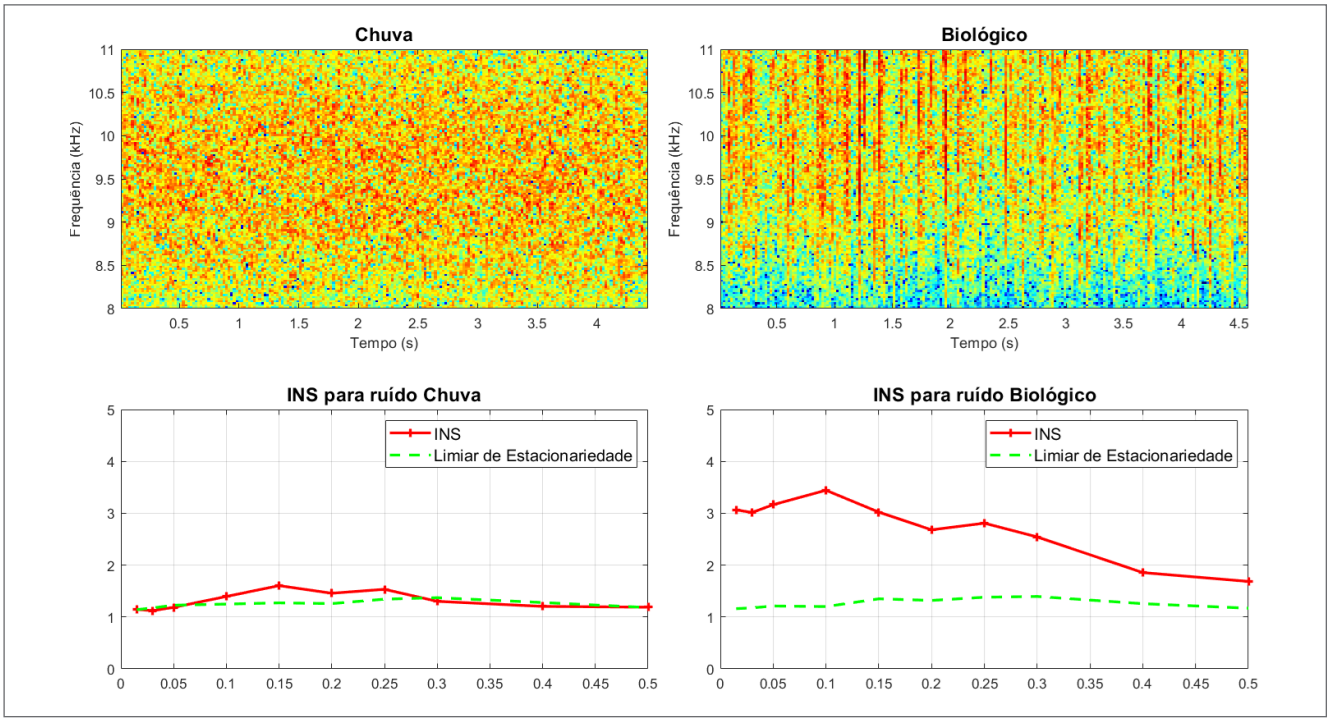

Figura 2. Espectrograma dos ruídos ambientais (acima) e seus respectivos INS (abaixo).

os valores de INS medidos para cada escala de observação $\frac{T_h}{T}$, sendo T_h é a janela de tempo adotada na análise e T é a duração total do sinal. A linha tracejada em verde representa o limiar de estacionariedade. Com base nessa análise, os ruídos Chuva e Biológico podem ser classificados como estacionário e não-estacionário, respectivamente.

O primeiro cenário consistiu em analisar e comparar os resultados obtidos pelos métodos de realce em um canal submarino sem o efeito da reverberação, considerando somente a distorção causada pelo ruído ambiente. As soluções de realce de voz *optimally-modified log-spectral amplitude* (OMLSA) (COHEN; BERDUGO, 2001) e *unbiased minimum mean-square error* (UMMSE) (GERKMANN; HENDRIKS, 2011) foram avaliadas com base em duas medidas objetivas que analisam os aspectos perceptuais da voz: uma de qualidade (*segmental signal-to-noise ratio* – segSNR) (HANSEN; PELLOM, 1998) e outra de inteligibilidade (*extended speech intelligibility index* – ESII) (RHEBERGEN; VERSFELD, 2005).

O segundo cenário considerou o canal reverberante e ruidoso, utilizando-se o modelo computacional de propagação acústica submarina BELLHOP (PORTER, 2016), baseado na teoria de traçados de raios, para estimar a resposta ao impulso do canal. O ambiente simulado consistiu em uma fonte acústica posicionada a uma profundidade de 80 metros, transmitindo um sinal até um hidrofone situado a 15 quilômetros de distância da fonte, na mesma profundidade, em um canal com profundidade constante de 100 metros. Foram obtidas seis respostas ao impulso do canal, resultantes de combinações de dois perfis de velocidade do som (*sound speed profile* – SSP) registrados em experimentos reais realizados em Arraial do Cabo, com três fundos distintos, *Sand* (Areia), *Gravel* (Cascalho) e *Chalk* (Pedra Calcária), nesta ordem de menor para maior reverberação. As propriedades geoacústicas dos fundos, adotadas para o modelo, foram extraídas da Tabela 1.3 do livro de Jensen *et al.* (2011). Os perfis de SSP e exemplos de espectrogramas obtidos para os três tipos de fundos podem ser observados na Figura 3.

Neste segundo cenário, os sinais foram processados pelos mesmos métodos de realce utilizados no primeiro cenário, além de duas máscaras acústicas descritas na literatura para aprimorar a inteligibilidade de sinais reverberantes e ruidosos. A *Ideal Reverberant Mask* (IRM) é uma máscara ideal que utiliza informações do sinal de voz limpo, proporcionando os melhores resultados no aprimoramento da inteligibilidade. Já a *Binary Reverberant Mask* (BRM) (HAZRATI; LEE; LOIZOU, 2012) é uma máscara não ideal (cega), proposta para supressão de reverberação em situações nas quais as informações do sinal de interesse não

estão disponíveis *a priori*. Para este cenário, foi adotada uma medida não intrusiva da qualidade e inteligibilidade empregada em cenários reverberantes (*speech to reverberation modulation energy ratio* – SRMR) (FALK; ZHENG; CHAN, 2010).

4. RESULTADOS

4.1. CENÁRIO EXPERIMENTAL SEM REVERBERAÇÃO

A Figura 4 exibe os resultados da média dos incrementos obtidos no SegSNR (*ΔSegSNR)* em relação aos resultados não processados (NP, ou seja, sem os métodos de realce). O método OMLSA apresentou os melhores aprimoramentos para ambos os ruídos, destacando-se um incremento de 2,45 dB contra 1,06 dB do UMMSE no Ruído Chuva com SNR de 5 dB. Além disso, observou-se que, para o ruído Biológico, a diferença entre os resultados obtidos pelo OMLSA e pelo UMMSE foi ainda mais significativa, especialmente nos casos de SNR de -5 e 0 dB. De modo geral, verificou-se que, quanto maior o SNR, maior o *ΔSegSNR* obtido para ambos os métodos. Por fim, os métodos de realce demonstraram os maiores *ΔSegSNR* para o ruído Chuva, indicando um desempenho superior dessas soluções na mitigação de ruídos estacionários.

A Figura 5 apresenta um gráfico *boxplot* para os resultados da medida ESII, cujos valores variam entre 0 e 1, indicando que maiores valores refletem maior inteligibilidade. Observa-se, inicialmente, que os valores de ESII para o ruído Chuva são ligeiramente menores do que para o ruído Biológico, principalmente em SNRs mais baixas, sugerindo que o ruído estacionário distorce mais as componentes da voz no tempo e na frequência. Por outro lado, os maiores aprimoramentos na inteligibilidade também foram associados ao ruído Chuva, evidenciando o bom desempenho dos métodos de realce em lidar com ruídos estacionários. Por exemplo, para os sinais corrompidos com SNR de 0 dB, os aumentos percentuais nas médias do ESII obtidos pelo OMLSA (ESII de 0,506) e pelo UMMSE (ESII de 0,428) foram de 32,8 e 12,3% em relação ao NP (ESII de 0,381) para o ruído Chuva, e 20,2% (ESII de 0,506) e 5,5% (ESII de 0,444), em relação ao NP (ESII de 0,421) para o ruído Biológico, respectivamente. Assim como no aprimoramento da qualidade, o OMLSA obteve os melhores resultados de inteligibilidade pelo ESII em todos os cenários analisados.

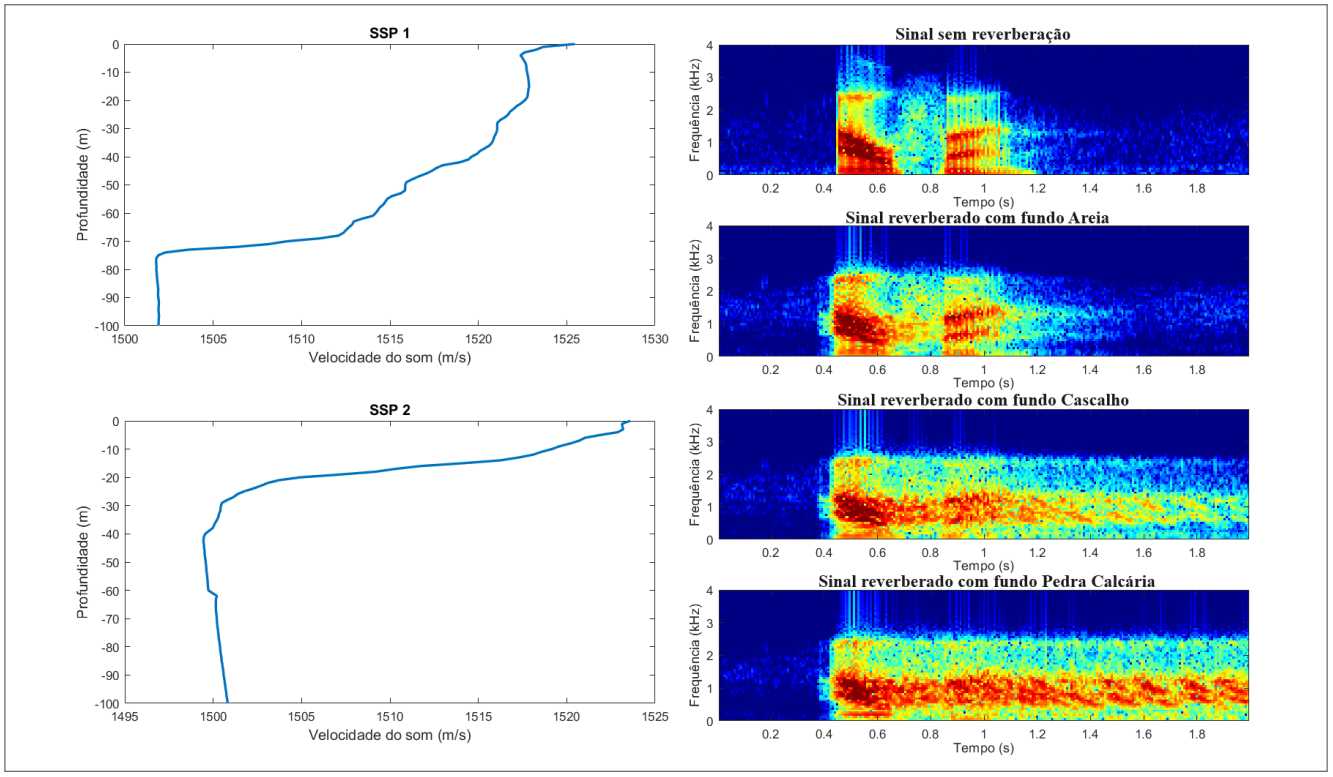

Figura 3. Os SSP (à esquerda) e os espectrogramas de um sinal de voz com reverberação para os 3 fundos adotados neste trabalho comparado ao mesmo sinal de voz sem reverberação (à direita).

4.2. CENÁRIO EXPERIMENTAL COM REVERBERAÇÃO

A Figura 6 ilustra os resultados da medida SRMR para o cenário com reverberação, calculados como a média dos 32 sinais concatenados de cada locutor, considerando as três SNR e 2 SSP, totalizando 192 sinais de áudio para cada coluna. Como esperado, a máscara ideal IRM obteve os melhores resultados em todos os cenários avaliados. Contudo, o método de realce OMLSA, ainda que não tenha sido originalmente proposto para cenários reverberantes, apresentou os maiores aprimoramentos

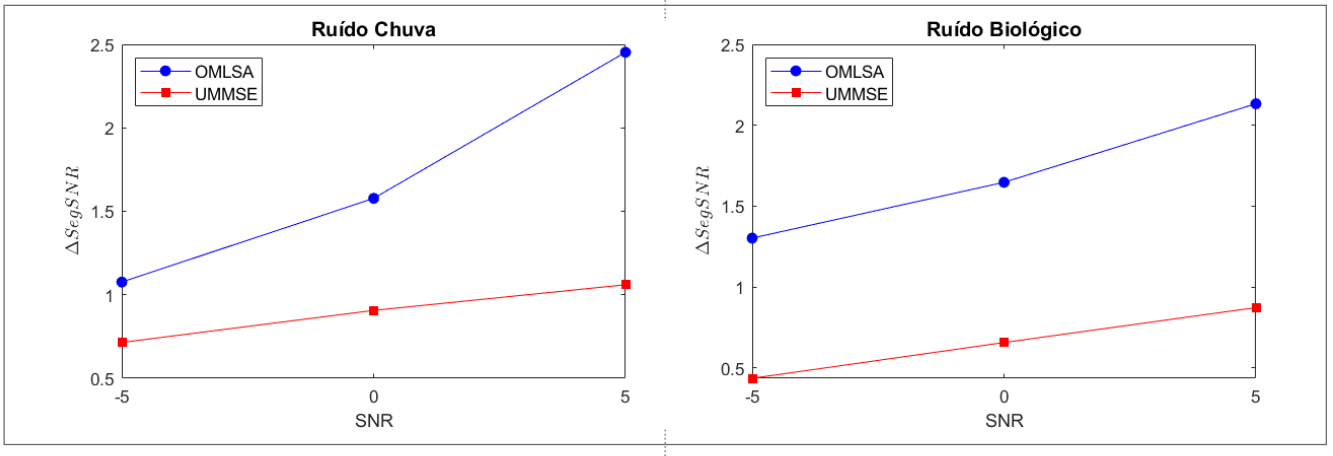

Figura 4. Resultados do incremento do SegSNR.

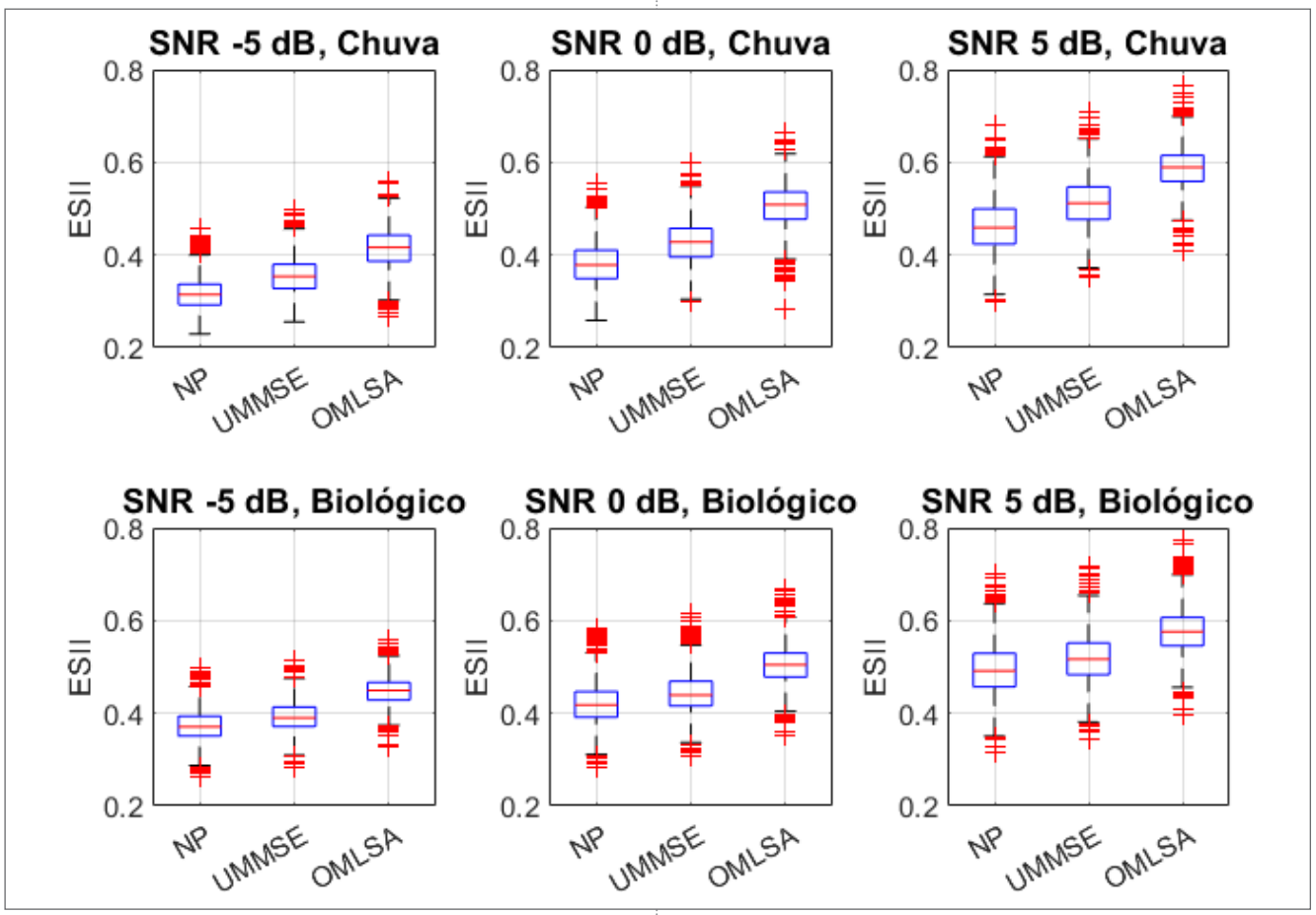

Figura 5. Resultados do ESII.

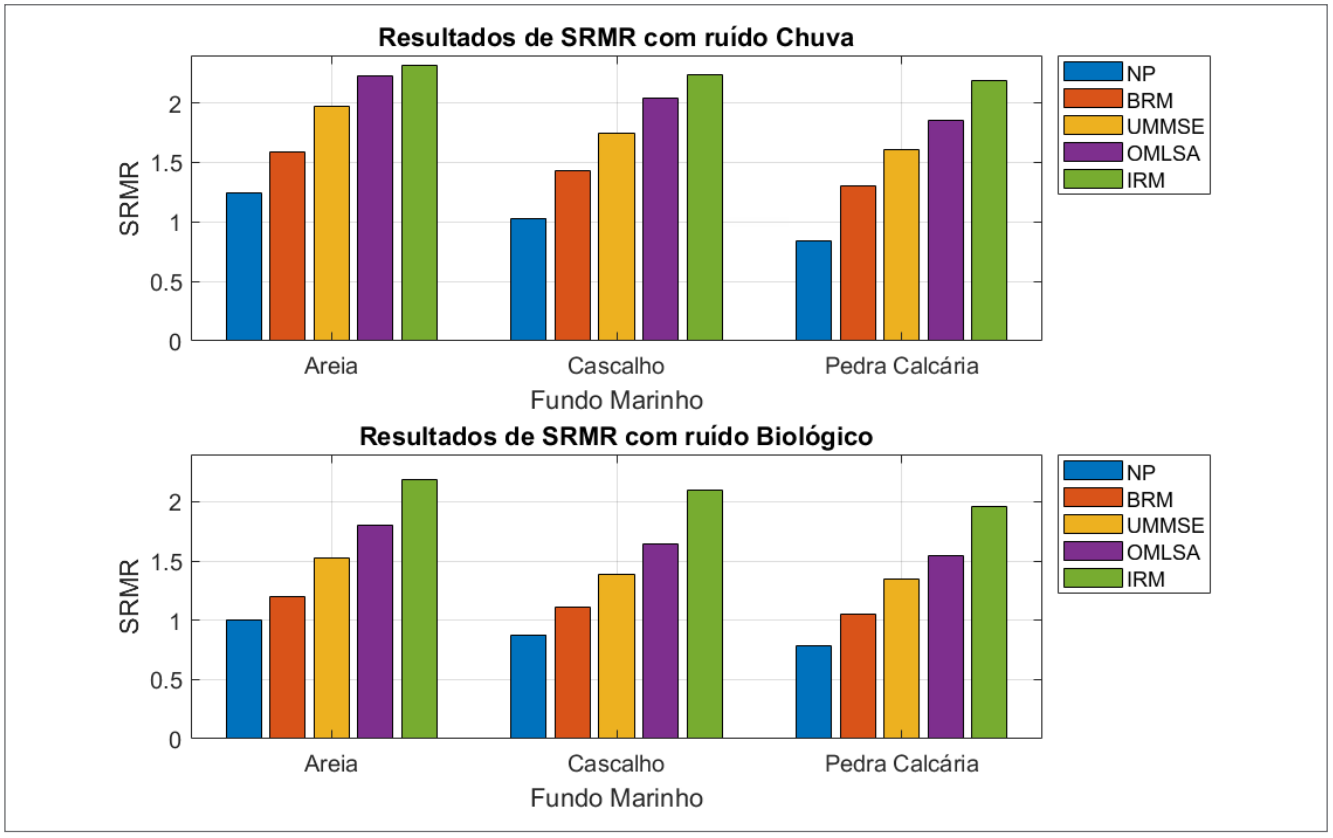

Figura 6. Resultados da medida SRMR.

entre as demais soluções, seguido pelo UMMSE e pela BRM. Esse desempenho manteve-se mesmo no cenário com maior reverberação (fundo Pedra Calcária). De modo geral, observou-se que, quanto maior a reverberação, maior o distanciamento entre a média do SRMR obtida pela IRM e as demais soluções. Ao comparar os resultados entre os ruídos, observou-se que esse distanciamento é ainda maior com o ruído não-estacionário. No ruído Chuva, a média de SRMR do OMLSA (2,22) chegou a 96% do resultado obtido pela IRM (2,31) no cenário com menor reverberação (fundo Areia), contrastando com os 79% obtido com o fundo Pedra Calcária corrompido com o ruído Biológico (1,55 de 1,96).

5. CONCLUSÃO

Este trabalho apresentou um estudo sobre o emprego de métodos de realce de sinais e máscaras acústicas, usualmente empregados na acústica aérea, em um sistema de telefonia acústica submarina, adotando uma base de dados de voz específica para essa finalidade. Foram propostos dois cenários experimentais, sendo um corrompendo o sinal de voz somente com ruído ambiente e outro incluindo o efeito da reverberação, considerando dois ruídos e seis cenários reverberantes distintos. Os resultados indicaram que o método OMLSA apresentou os melhores aprimoramentos na qualidade e inteligibilidade da voz no cenário sem reverberação, e uma melhora na SRMR superior às soluções cegas no cenário com reverberação. Para trabalhos futuros, sugere-se a implementação de métodos capazes de estimar a resposta ao impulso do canal submarino para a mitigação da reverberação, além da implementação de soluções de realce baseadas em aprendizado de máquina e redes neurais, com o objetivo de comparar seus resultados com os obtidos neste trabalho.

AGRADECIMENTOS

Os autores agradecem a todos os locutores que, voluntariamente, participaram da montagem do banco de dados de voz próprio para este trabalho, cujas contribuições foram essenciais para a conclusão desta pesquisa.

REFERÊNCIAS

BORGNAT, P.; FLANDRIN, P.; HONEINE, P.; RICHARD, C.; XIAO, J. Testing stationarity with surrogates: A time-frequency approach. *IEEE Transactions on Signal Processing*, v. 58, n. 7, p. 3459-3470, 2010. https://doi.org/10.1109/TSP.2010.2043971

CALDEIRA, A.; COELHO, R. Realce de sinais em ambiente com variações acústicas subaquáticas. *In:* SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS, 39., 2021. *Anais* [...]. 2021.

COHEN, I.; BERDUGO, B. Speech enhancement for non-stationary noise environments. *Signal Processing*, v. 81, n. 11, p. 2403-2418, 2001. https://doi.org/10.1016/S0165-1684(01)00128-1

DISCOVERY OF SOUND IN THE SEA. *Audio gallery*. Disponível em: https://dosits.org/galleries/audio-gallery/. Acesso em: 5 fev. 2024.

FALK, T. H.; ZHENG, C.; CHAN, W.-Y. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 18, n. 7, p. 1766-1774, 2010. https://doi.org/10.1109/TASL.2010.2052247

GERKMANN, T.; HENDRIKS, R. C. Unbiased MMSE-based noise power estimation with low complexity and low tracking delay. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 20, n. 4, p. 1383-1393, 2011. https://doi.org/10.1109/TASL.2011.2180896

HANSEN, J. H. L.; PELLOM, B. L. An effective quality evaluation protocol for speech enhancement algorithms. *In:* INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 5., 1998. *Anais* [...]. 1998. p. 2819-2822.

HAZRATI, O.; LEE, J.; LOIZOU, P. C. Binary mask estimation for improved speech intelligibility in reverberant environments. *In:* INTERSPEECH, 2012. *Anais* [...]. 2012. p. 162-165.

JENSEN, F. B.; KUPERMAN, W. A.; PORTER, M. B.; SCHMIDT, H. *Computational ocean acoustics*. New York: Springer, 2011.

MARTINY, R.; ALCÂNTARA, R.; COELHO, R. Avaliação da predição objetiva da inteligibilidade de sinais reverberantes e ruidosos com uso de máscaras acústicas. *In:* SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES E PROCESSAMENTO DE SINAIS, 40., 2022. *Anais* [...]. 2022.

PORTER, M. B. *Bellhop3d user guide*. Technical report, Heat, Light, and Sound Research Inc., 2016.

RHEBERGEN, K. S.; VERSFELD, N. J. A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *Journal of the Acoustical Society of America*, v. 117, n. 4, p. 2181-2192, 2005. https://doi.org/10.1121/1.1861713

WOODWARD, B.; SARI, H. Digital underwater acoustic voice communications. *IEEE Journal of Oceanic Engineering*, v. 21, n. 2, p. 181-192, 1996. https://doi.org/10.1109/48.486793