

Realce de Sinais em Ambiente com Variações Acústicas Subaquáticas

Antônio Walkir Sibanto Caldeira
Rosângela Fernandes Coelho

Laboratório de Processamento de Sinais Acústicos
Instituto Militar de Engenharia

XXXIX Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - SBrT

26-29 de Setembro de 2021

Plano de Apresentação

- Motivação e Desafios
- Objetivo
- Índice de Não-Estacionariedade
- Métodos de Realce de Sinais
- Medidas Objetivas de Predição: Qualidade e Inteligibilidade
- Resultados Experimentais
- Conclusão

Motivação e Desafios

- As interferências acústicas causadas pelo ruído ambiente submarino afetam a qualidade de sistemas sonar e comunicações.

Motivação e Desafios

- As interferências acústicas causadas pelo ruído ambiente submarino afetam a qualidade de sistemas sonar e comunicações.
- As soluções de realce de sinais visam atenuar os efeitos causados por estes ruídos.

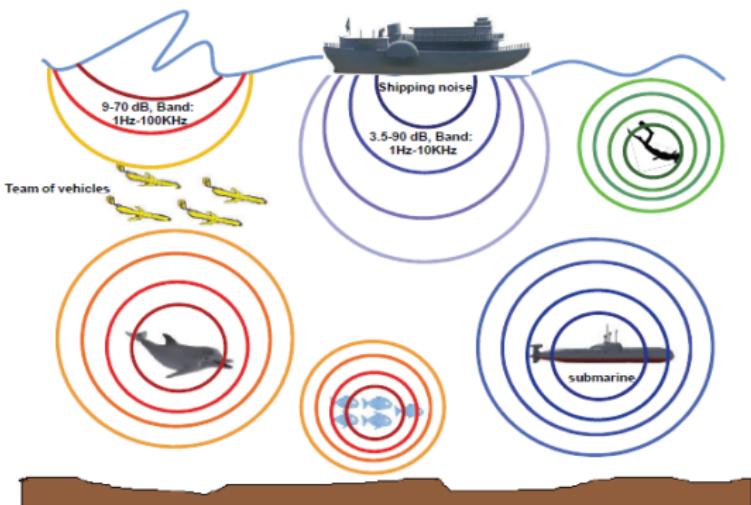
Motivação e Desafios

- As interferências acústicas causadas pelo ruído ambiente submarino afetam a qualidade de sistemas sonar e comunicações.
- As soluções de realce de sinais visam atenuar os efeitos causados por estes ruídos.
- Aprimoramento da qualidade e inteligibilidade dos sinais acústicos.

Motivação e Desafios

- As interferências acústicas causadas pelo ruído ambiente submarino afetam a qualidade de sistemas sonar e comunicações.
- As soluções de realce de sinais visam atenuar os efeitos causados por estes ruídos.
- Aprimoramento da qualidade e inteligibilidade dos sinais acústicos.
- Exemplos de Aplicações:
 - Exploração de petróleo *offshore*
 - Comunicação por voz entre mergulhadores
 - Operações de defesa

Motivação e Desafios



Fonte: M. Rahmati, P. Pandey and D. Pompili, "Separation and classification of underwater acoustic sources" in Underwater Communications and Networking (UComms)2014, IEEE, pp. 1-5, 2014.

Motivação e Desafios

- Classificação dos Ruídos no Ambiente Submarino:
 - Ruídos Antropogênicos
 - embarcações
 - pistolas de ar (*air guns*)
 - sonares ativos
 - Ruídos Naturais
 - sísmicos (vulcões, terremotos, tsunamis)
 - biológicos (baleias, cardume de peixes, golfinhos)
 - associados às condições climáticas (ondas, vento e chuvas)

Motivação e Desafios

- Classificação dos Ruídos no Ambiente Submarino:

- Ruídos Antropogênicos

- embarcações
 - pistolas de ar (*air guns*)
 - sonares ativos

- Ruídos Naturais

- sísmicos (vulcões, terremotos, tsunamis)
 - biológicos (baleias, cardume de peixes, golfinhos)
 - associados às condições climáticas (ondas, vento e chuvas)

- Desafios

- Ruídos provocados por fontes acústicas de diferentes características temporais e espectrais; e
 - Ruídos não-estacionários.

Objetivo

Estudo comparativo de soluções de **realce de sinais** para aplicações na acústica submarina, considerando ruídos de diferentes **fontes acústicas** e graus de **não-estacionariedade**, para realce de **sinais de voz** e **sinal chirp**.

Índice de Não-Estacionariedade

- Medida proposta para determinar, objetivamente, o grau de não-estacionariedade dos sinais acústicos.*
- Comparação do sinal com seus referenciais estacionários (*surrogates*).
- INS obtido de acordo com uma escala de observação $\frac{T_h}{T}$.
- Limiar do teste de estacionariedade γ calculado para cada T_h .
- Sinal não-estacionário se $\text{INS} > \gamma$.

* [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.]

Métodos de Realce de Sinais

- Métodos Espectrais

- OMLSA (*optimally-modified log-spectral amplitude*)*
- UMMSE (*unbiased minimum mean-square error*)**

- Método Temporal

- NNESE (*non-stationary noise estimation for speech enhancement*)***

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

** [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.]

*** [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.]

Métodos de Realce de Sinais

• OMLSA

- Estimativas do espectro de potência dos ruídos pelo IMCRA (*improved minima controlled recursive averaging*).^{*}
- Reconstrução do sinal pelo algoritmo OMLSA, que minimiza o erro quadrático médio do logaritmo da magnitude espectral.
- $G_{OMLSA}(\kappa, \tau) = G_{LSA}(\kappa, \tau)^{p(\kappa, \tau)} G_{min}^{1-p(\kappa, \tau)}$
 - τ e κ índices de quadro e frequência, respectivamente;
 - $G_{LSA}(\kappa, \tau)$ ganho calculado em função do SNR *a priori*;
 - G_{min} limiar mínimo para o ganho, igual à -25 dB.

* [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.]

Métodos de Realce de Sinais

• UMMSE

- Estimação das componentes espectrais do ruído baseada na minimização dos erros quadráticos médios.
- $|\hat{N}(\kappa, \tau)|^2 = \alpha_p |\hat{N}(\kappa, \tau - 1)|^2 + (1 - \alpha_p) E[|N(\kappa, \tau)|^2 | Y(\kappa, \tau)]$
 - τ e κ índices de quadro e frequência, respectivamente;
 - α_p constante de suavização;
 - $E[|N(\kappa, \tau)|^2 | Y(\kappa, \tau)]$ periodograma do ruído.
- Supressão das componentes ruidosas feita por uma técnica baseada no filtro de Wiener.

Métodos de Realce de Sinais

• NNESE

- Estimação do desvio padrão do ruído obtida a partir do algoritmo DATE(*d-dimensional trimmed estimator*)^{*} em cada segmento de 32 ms do sinal.
- Componentes do ruído selecionadas a partir de um limiar $y(b_q)$.
- Reconstrução:

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

- α fator de subtração para a reconstrução do sinal acústico; e
- $\beta = 1 - \alpha$ fator de piso para valores negativos de amplitude.

* [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.]

Medidas Objetivas de Predição

- Medidas de Qualidade
 - PESQ (*perceptual evaluation of speech quality*)^{*}
 - OQCM (*overall quality composite measure*)^{**}
 - SegSNR (*segmental signal-to-noise ratio*)^{***}
- Medida de Intelligibilidade
 - STOI (*extended short-time objective intelligibility*)^{****}

* [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.]

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

*** [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.]

**** [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.]

Medidas Objetivas de Predição

- Medidas de Qualidade
 - PESQ
 - Sinais de voz limpo e degradado são mapeados para o domínio de tempo-frequência.
 - Medida convertida em uma medida subjetiva MOS (*mean opinion score*), de 1,0 (ruim) a 4,5 (sem distorção).
 - Análise em 30% dos quadros.

Medidas Objetivas de Predição

- Medidas de Qualidade

- PESQ

- Sinais de voz limpo e degradado são mapeados para o domínio de tempo-frequência.
 - Medida convertida em uma medida subjetiva MOS (*mean opinion score*), de 1,0 (ruim) a 4,5 (sem distorção).
 - Análise em 30% dos quadros.

- OQCM

- Combinação linear de medidas de qualidade existentes na literatura.
 - $OQCM = 1,594 + 0,805PESQ - 0,512LLR^* - 0,007WSS^{**}$.

LLR - *log-likelihood ratio*

WSS - *weighted spectral slope*

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

** [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.]

Medidas Objetivas de Predição

- Medidas de Qualidade
 - SegSNR
 - Média entre valores de SNR calculados em quadros de curta duração.
 - Valores em cada quadro limitados entre -10 dB e 35 dB.

Medidas Objetivas de Predição

- Medidas de Qualidade
 - SegSNR
 - Média entre valores de SNR calculados em quadros de curta duração.
 - Valores em cada quadro limitados entre -10 dB e 35 dB.
- Medida de Inteligibilidade
 - STOI
 - Cálculo do coeficiente de correlação entre os espectros dos sinais limpo e realçado.
 - Pontos resultantes da DFT agrupados em 15 bandas, com frequências centrais variando de 150 Hz a 4300 Hz.
 - Média de todos os valores da medida calculada para cada segmento de tempo e banda.

Resultados Experimentais

- Cenário Experimental I

- 10 sinais de voz de 24 locutores;
- 16 homens e 8 mulheres;
- Base TIMIT*;
- Frequência de Amostragem: 16 kHz;
- Duração média de 3 segundos;
- PESQ, OQCM, STOI;
- Ruídos *Bubbles**, *Killer Whale***, *Underwater Earthquake***; e
- SNR: -5 dB, -3 dB, 0 dB, 3dB e 5 dB.

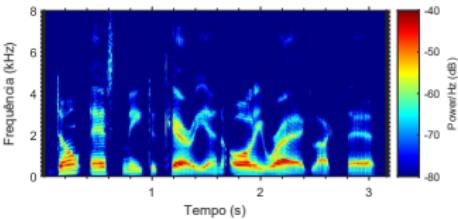
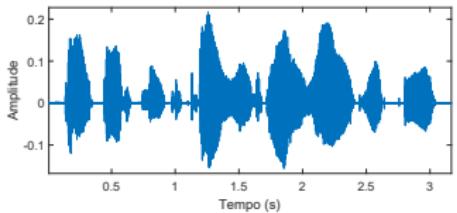
* [J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "Darpa timit acoustic phonetic continuous speech corpus cdrom," 1993.]

** [Disponível em <http://www.freesound.org>]

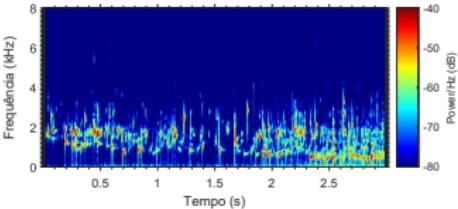
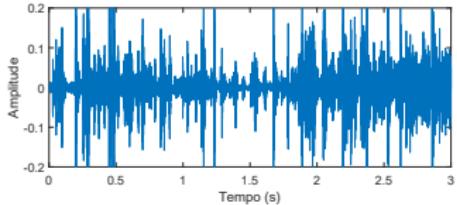
*** [Disponível em <https://www.maritime.org/sound>]

Resultados Experimentais

- *Speech*

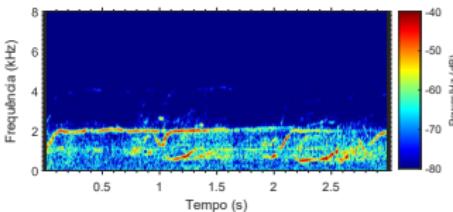
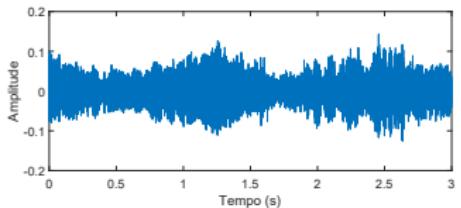


- *Bubbles:*

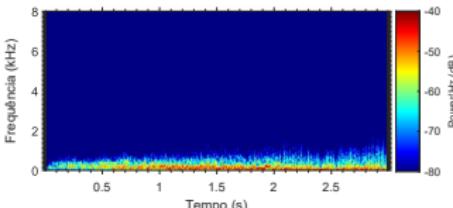
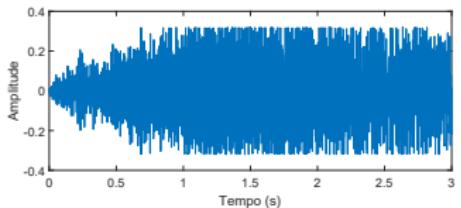


Resultados Experimentais

- *Killer Whale:*



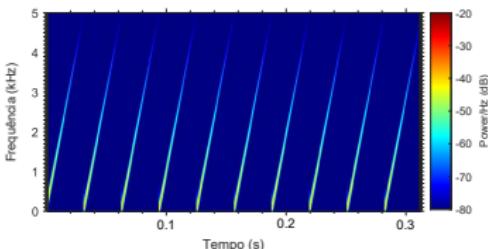
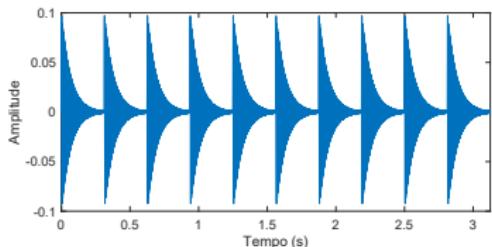
- *Underwater Earthquake:*



Resultados Experimentais

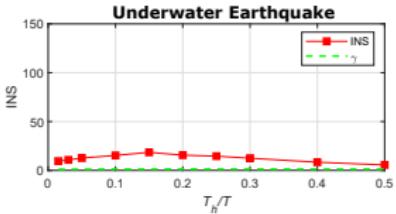
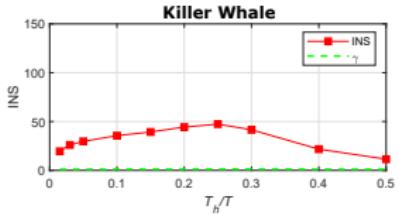
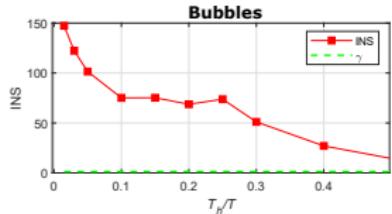
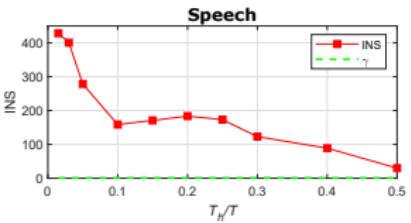
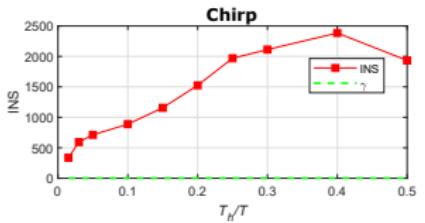
- Cenário Experimental II

- 10 sinais *chirp* lineares com decaimento exponencial;
- $f_i = 100$ Hz, $f_f = 5$ kHz, $T = 312,5$ ms, $\beta = 0.067$;
- Duração total de 3,125 segundos;
- Frequência de Amostragem: 16 kHz;
- SegSNR; e
- Mesmos ruídos e SNR do Cenário Experimental I.



Resultados de INS

- Índices de Não-Estacionariedade



Resultados de INS

Classificação dos sinais acústicos de acordo com o INS_{max} :

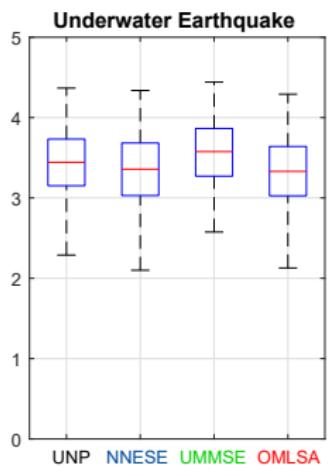
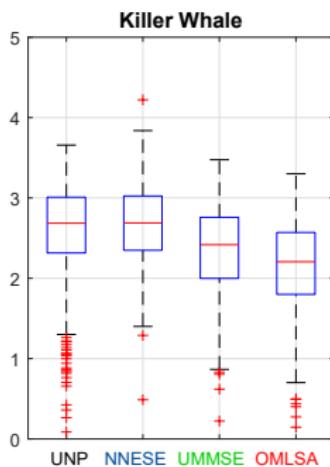
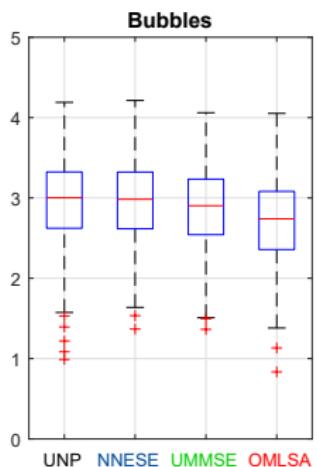
- $INS_{max} > 100\gamma$: Altamente Não-Estacionários
 - Chirp
 - Speech
 - Bubbles
- $20\gamma < INS_{max} \leq 100\gamma$: Não-Estacionários
 - Killer Whale
- $\gamma < INS_{max} \leq 20\gamma$: Moderadamente Não-Estacionários
 - Underwater Earthquake

Resultados de PESQ - Cenário Experimental I

RUÍDOS	SNR	UNP*	NNESE	UMMSE	OMLSA
<i>Bubbles</i> $INS_{max} = 147$	-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
<i>Killer Whale</i> $INS_{max} = 47$	-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
<i>Underwater Earthquake</i> $INS_{max} = 19$	-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

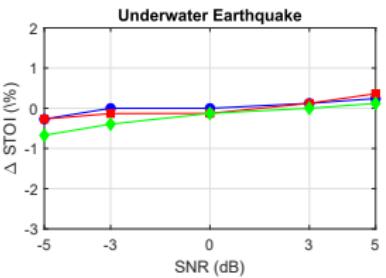
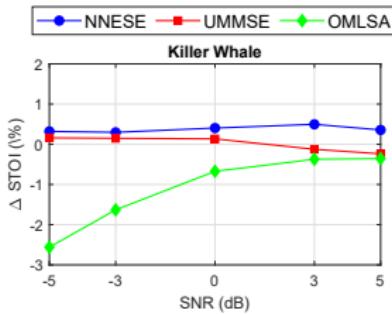
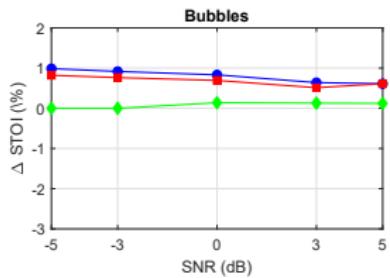
* [UNP - unprocessed]

Resultados de OQCM - Cenário Experimental I

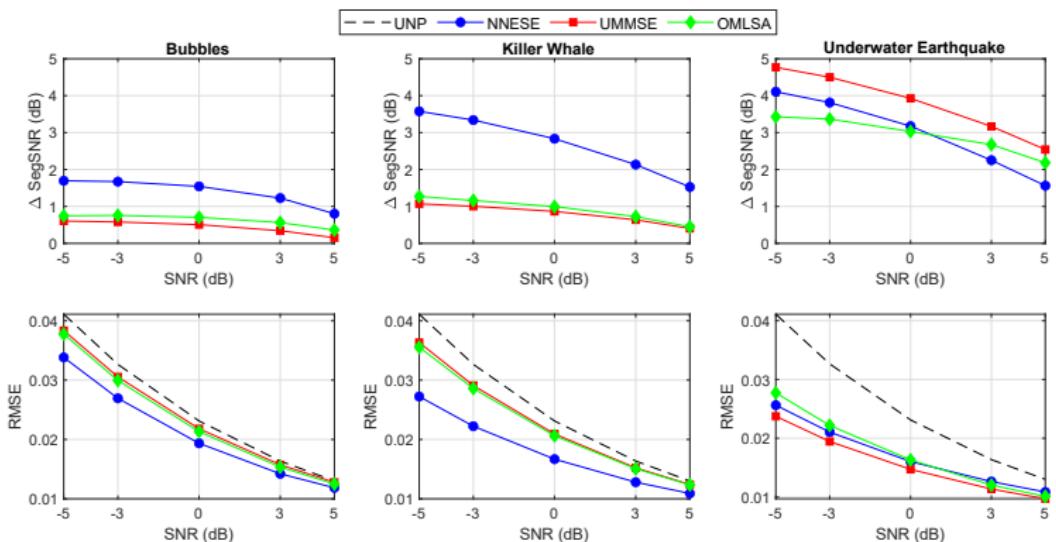


Resultados de STOI - Cenário Experimental I

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



Resultados de SegSNR - Cenário Experimental II



Resultados de SegSNR - Cenário Experimental II

Média Geral do Incremento do SegSNR:

- **OMLSA** : 1,49 dB
- **UMMSE** : 1,67 dB
- **NNESE** : 2,35 dB

Conclusão

- Comparação de desempenho dos métodos de realce de sinais aplicados em ambiente acústico submarino.
- Realce de sinais de voz e sinal *Chirp*.
- Ruídos de diferentes fontes acústicas e INS.
- **NNESE** apresentou melhores aprimoramentos na qualidade e inteligibilidade dos sinais acústicos, principalmente para ruídos de maior INS.

Agradecimentos



Conselho Nacional de Desenvolvimento Científico e Tecnológico



Fundação Carlos Chagas Filho de Amparo Pesquisa do Estado do Rio de Janeiro



Marinha do Brasil



OBRIGADO!