MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE DEFESA

ANTÔNIO WALKIR SIBANTO CALDEIRA

REALCE DE SINAIS EM AMBIENTE COM VARIAÇÕES ACÚSTICAS SUBAQUÁTICAS

RIO DE JANEIRO
2021

ANTÔNIO WALKIR SIBANTO CALDEIRA

REALCE DE SINAIS EM AMBIENTE COM VARIAÇÕES ACÚSTICAS SUBAQUÁTICAS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Defesa.

Orientador(es): Rosângela Fernandes Coelho, Docteur ENST

Rio de Janeiro

2021

INSTITUTO MILITAR DE ENGENHARIA
Praça General Tibúrcio, 80 – Praia Vermelha
Rio de Janeiro – RJ CEP: 22290-270

Sibanto Caldeira, Antônio Walkir.
Realce de Sinais em Ambiente com Variações Acústicas Subaquáticas / Antônio Walkir Sibanto Caldeira. – Rio de Janeiro, 2021.
68 f.

Orientador(es): Rosângela Fernandes Coelho.

Dissertação (mestrado) – Instituto Militar de Engenharia, Engenharia de Defesa, 2021.

1. Realce de Sinais de Voz. 2. Realce de Sinais Acústicos. 3. Acústica Subaquática. 4. Medidas de Qualidade. 5. Medidas de Inteligibilidade. i. Fernandes Coelho, Rosângela (orient.) ii. Título

ANTÔNIO WALKIR SIBANTO CALDEIRA

Realce de Sinais em Ambiente com Variações Acústicas Subaquáticas

Dissertação apresentada ao Programa de Pós-graduação em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Ciências em Engenharia de Defesa.

Orientador(es): Rosângela Fernandes Coelho.

Aprovada em 14 de janeiro de 2022, pela seguinte banca examinadora:

Prof. **Rosângela Fernandes Coelho** - Docteur ENST do IME - Presidente

---

Prof. **Paulo Fernando Ferreira Rosa** - Ph.D. do IME

Prof. **Marco Antonio Grivet Mattoso Maia** - Ph.D. da PUC-Rio

---

Dr. **Eduardo Esteves Vale** - D.Sc. da PUC-Rio

Rio de Janeiro

2021

AGRADECIMENTOS

À minha orientadora, Professora Rosângela Fernandes Coelho, pela dedicação, empenho, disponibilidade e pelos valiosos ensinamentos transmitidos nestes dois anos e que foram cruciais para a conclusão desta Dissertação.

À minha esposa Camilla e meu filho Oliver, pelo companheirismo e por sempre me dar razões para seguir em frente.

Aos colegas do Laboratório de Processamento de Sinais Acústicos, Rafael, Anderson, Zucatelli e Zão, pela amizade e suporte ao longo desta intensa jornada.

À Marinha do Brasil, por conceder a oportunidade de estudar em uma instituição de ensino de excelência por dedicação exclusiva.

Ao Instituto Militar de Engenharia, que me concedeu a oportunidade de realizar este curso de Mestrado.

"As dificuldades fortalecem a mente,
assim como o trabalho o faz com o corpo."
(Sêneca)

RESUMO

Nesta Dissertação de Mestrado, são abordados os desafios associados às interferências acústicas causadas por ruído do ambiente subaquático. Neste contexto, um método para realce de sinais acústicos subaquáticos no domínio do tempo é proposto, baseado na decomposição EEMD-IF (*ensemble empirical mode decomposition - iterative filtering*) e no índice de não-estacionariedade. No estudo, quatro soluções de realce de sinais propostas na literatura foram empregadas para comparação dos resultados com o método proposto. Os métodos foram examinados para realce de sinais de voz e um conjunto de sinais *chirp*. Os sinais de interesse, voz e *chirp*, foram corrompidos por ruídos acústicos subaquáticos com diferentes graus de não-estacionariedade e valores distintos de razão sinal-ruído. Quatro medidas de predição objetiva de qualidade e três medidas de inteligibilidade sonora foram adotadas para avaliar os métodos de realce. De modo geral, o método proposto alcançou os melhores resultados no aprimoramento das medidas de qualidade e inteligibilidade para realce dos sinais de voz, além de obter resultados interessantes para a qualidade dos sinais *chirp*, principalmente na presença de ruídos com maior grau de não-estacionariedade.

Palavras-chave: Realce de Sinais de Voz. Realce de Sinais Acústicos. Acústica Subaquática. Medidas de Qualidade. Medidas de Inteligibilidade.

ABSTRACT

This work adresses the challenges associated with acoustic interference caused by underwater ambient noise. A signal enhancement method in the time domain is proposed, based on the EEMD-IF decomposition (ensemble empirical mode decomposition - iterative filtering) and index of non-stationarity. In this work, four signal enhancement solutions proposed in the literature were considered for comparison in the experiments. The methods were evaluated for enhancing speech and chirp signal. The signals of interest, speech and chirp, were corrupted by underwater acoustic noises with different degrees of non-stationarity and signal-to-noise ratios. Four quality measures and three intelligibility measures were adopted to evaluate the enhancement methods. In general, the proposed method achieved the best results for the improvement of quality and intelligibility measures for speech signals, in addition to obtaining interesting results for the quality of chirp signals, mainly in the presence of noises with higher degree of non-stationarity.

Keywords: Speech Enhancement. Acoustic Signal Enhancement. Underwater Acoustics. Quality Measures. Intelligibility Measures.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ASII_{ST}	*Short-Time Approximated Speech Intelligibility Index*
BIF	*Band-Importance Function*
DATE	*D-Dimensional Trimmed Estimator*
DFT	*Discrete Fourier Transform*
DI	*Distortion Index*
EEMD	*Ensemble Empirical Mode Decomposition*
EMD	*Empirical Mode Decomposition*
EMD	*Empirical Mode Decomposition - Hurst*
ESII	*Extended Speech Intelligibility Index*
FFT	*Fast Fourier Transform*
IF	*Iterative Filtering*
INS	*Index of Non-Stationarity*
IMF	*Intrinsic Mode Function*
MMSE	*Minimum Mean Square Error*
MOS	*Mean Opinion Score*
MOV	*Model Output Variable*
NNESE	*Non-Stationary Noise Estimation For Speech Enhancement*
NP	Não Processados
ODG	*Overall Difference Grade*
OMLSA	*Optimally-Modified Log-Spectral Amplitude*
PEAQ	*Perceptual Evaluation of Audio Quality*
PESQ	*Perceptual Evaluation of Speech Quality*
PRO	Proposto
RMSE	*Root Mean Square Error*

SDR	*Signal-to-Distortion Ratio*
SegSNR	*Segmental Signal-to-Noise Ratio*
SII	*Speech Intelligibility Index*
SNR	*Signal-to-Noise Ratio*
STFT	*Short-Time Fourier Transform*
STOI	*Short-Time Objective Intelligibility*
UMMSE	*Unbiased Minimum Mean Square Error*
VAD	*Voice Activity Detector*

SUMÁRIO

1 INTRODUÇÃO

O oceano é composto por um vasto e diverso ambiente de ruídos acústicos. Há décadas, o estudo da multiplicidade da origem destas fontes ambientais, com suas distintas dinâmicas e características estatísticas, é amplamente reportado na literatura (WENZ, 1962), (SIDDAGANGAIAH et al., 2015), (URICK; KUPERMAN, 1989), (WILCOCK et al., 2014). Estas interferências acústicas são um grande desafio para diversos sistemas e aplicações do meio subaquático. Particularmente, por provocarem severas degradações na qualidade de seus principais sistemas acústicos: sonar e de comunicações. Outras aplicações de interesse deste meio como estudos oceanográficos, exploração de petróleo *offshore* e operações de defesa (AL-ABOOSI; SHA'AMERI, 2017) também podem ser impactadas pela presença destes ruídos ambientais.

Os ruídos acústicos ambientais presentes no cenário subaquático são comumente originados por dois tipos de fontes: as antropogênicas, como as embarcações, sonares ativos, pistolas de ar (*air guns*), e as naturais, como chuvas, vento, vida marinha, sísmicos (AL-ABOOSI; SHA'AMERI, 2017), (WENZ, 1972). A redução dos efeitos causados por estes ruídos, que corrompem os sinais acústicos de interesse, torna-se uma tarefa essencial para evitar a perda de qualidade e inteligibilidade destes sinais. O principal desafio para mitigar esses impactos consiste em obter estimativas das componentes de mascaramento provocadas pelas diversas e distintas fontes dos ruídos. Principalmente, quando suas características variam no tempo, ou seja, são não-estacionários.

Desde a década de 1970 (BOLL, 1979), as soluções de realce de sinais de voz têm sido propostas para atenuar estas interferências causados por ruídos acústicos. Os métodos de realce de sinais podem ser classificados, segundo seu domínio de atuação, como espectrais e temporais. Técnicas espectrais convencionais utilizam a transformada de Fourier de tempo curto (STFT *short-time Fourier transform*) para estimar o espectro do ruído. Para isso, é necessário identificar os trechos do sinal em que não há atividade de voz com uso de detectores de atividade de voz (VAD - *voice activity detector*). Todavia, os métodos clássicos de estimação geralmente assumem a hipótese de que os ruídos são estacionários (BOLL, 1979).

Para lidar com as interferências causadas por ruídos não-estacionários, o método OMLSA (*optimally-modified log-espectral amplitude*) (COHEN; BERDUGO, 2001) adota o estimador IMCRA (*improved minima controlled recursive averaging*) (COHEN, 2003) para realizar as estimações do espectro do ruído baseado em quadros anteriores do sinal. Contudo, este estimador torna-se lento para acompanhar as variações espectrais de ruídos não-estacionários (MANOHAR; RAO, 2006). Uma solução desenvolvida para capturar, com menor tempo de resposta, estas variações, é o UMMSE (*unbiased minimum mean-square*

error) (GERKMANN; HENDRIKS, 2012), que propõe obter a estimação do espectro do ruído a partir da minimização de erro médio quadrático.

Dentre as técnicas temporais descritas na literatura, algumas são baseadas em análise tempo-frequência, como a decomposição *wavelets* (DONOHO; JOHNSTONE, 1994) ou a decomposição empírica de modos (EMD - *empirical mode decomposition*) (HUANG et al., 1998), (FLANDRIN; RILLING; GONCALVES, 2004). Uma limitação das *wavelets* é que as funções base utilizadas na decomposição do sinal são fixas (OMITAOMU; PROTOPOPESCU; GANGULY, 2011).

A decomposição EMD decompõe o sinal em um conjunto de funções intrínsecas de modo (IMF - *intrinsic mode functions*), que são totalmente dependentes do próprio sinal, ou seja, as bases são adaptativas. Os modos de menor índice correspondem às oscilações de alta frequência, enquanto os de maior índice são de baixa frequência. Nos métodos de realce baseados em EMD, um critério de decisão precisa ser definido para identificar e suprimir os modos mais afetadas pelo ruído. O EMDF (EMD-*based filtering*) detecta estas IMFs baseado em um estudo dos seus valores de variância. O EMDH (EMD-*Hurst*) (ZÃO; COELHO; FLANDRIN, 2014) utiliza o expoente de Hurst para detecção e estimação dos quadros das IMFs mais corrompidos pelos ruídos não-estacionários que apresentam altas concentrações de energia nas baixas frequências.

Para realçar o sinal de voz na presença de ruídos não-estacionários, o método NNESE (*non-stationary noise estimation for speech enhancement*) (TAVARES; COELHO, 2016) utiliza um estimador robusto (PASTOR; SOCHELEAU, 2012) para estimação do desvio padrão do ruído acústico em segmentos curtos de tempo, a partir do sinal degradado no domínio do tempo. Para a reconstrução do sinal realçado, as amplitudes constituídas predominantemente por ruídos são extraídas e as demais são atenuadas baseadas na estimação do desvio padrão do ruído.

Geralmente, para avaliação dos métodos de realce, adota-se medidas objetivas de predição de qualidade, que está associada à atenuação das distorções causadas pelos ruídos acústicos. Apesar dos testes subjetivos perceptuais serem a forma mais precisa para avaliação do aprimoramento do sinal, estes são frequentemente substituídos por medidas objetivas em virtude do tempo despendido e do alto custo. Todavia, uma medida objetiva somente é considerada satisfatória quando demonstra alta correlação com os resultados dos testes perceptuais obtidos de forma subjetiva. Além da qualidade, a predição da inteligibilidade também é examinada como análise complementar dos métodos de realce. A inteligibilidade é definida como a medida que reflete o quanto uma mensagem acústica é compreensível.

Nesta Dissertação, é proposto um método de realce para atuação em dois sinais de interesse (voz e *chirp*) no ambiente acústico subaquático, com foco na supressão de ruídos ambientais com características não-estacionárias. Este método emprega o EEMD-IF

(*ensemble* EMD - *iterative filtering*) (WU; HUANG, 2009; LIN; WANG; ZHOU, 2009) para a decomposição do sinal corrompido e um critério baseado no índice de não-estacionariedade (INS - *index of nonstationarity*) (BORGNAT et al., 2010) para detecção e estimação das componentes mais afetadas pelos ruídos.

O método proposto, denominado PRO, é avaliado em termos de qualidade e inteligibilidade, em dois cenários experimentais distintos, considerando sete medidas objetivas de predição. Os ruídos acústicos considerados em ambos os experimentos são originados de fontes reais do ambiente subaquático, com diferentes características temporais e espectrais. Para comparação da solução proposta neste trabalho, foram utilizados quatro métodos de realce competitivos como referência. Dentre estes métodos, foram adotados duas soluções espectrais (UMMSE e OMLSA) e duas temporais (NNESE e EMDH).

Os cenários experimentais se distinguem pelos sinais de interesse, voz e *chirp*, e suas composições ruidosas. O primeiro experimento utiliza um subconjunto de 24 locutores, sendo 8 mulheres e 16 homens, provenientes da base de voz TIMIT (GAROFOLO et al., 1993). Cada locutor possui 10 gravações com frequência de amostragem de 16 kHz e uma duração média de 3s, totalizando 240 sinais de voz. Neste cenário, as medidas de predição objetiva de qualidade PESQ (*perceptual evaluation of speech quality*) (RIX et al., 2001) e PEAQ (*perceptual evaluation of speech quality*) (COLOMES et al., 1999) são adotadas para examinar os métodos de realce. Para avaliação da inteligibilidade, são empregadas as medidas STOI (*short-time objective intelligibility*) (TAAL et al., 2011), ESII (*extended speech intelligibility index*) (RHEBERGEN; VERSFELD, 2005) e ASII_{ST} (*short-time approximated speech intelligibility index*) (HENDRIKS et al., 2015).

No segundo experimento, o sinal de interesse consiste em um conjunto de sinais *chirp*, que são comumente empregados nas comunicações subaquáticas em virtude de sua baixa sensibilidade ao efeito *Doppler* e boa capacidade de rejeição de interferências (HE et al., 2009). As medidas de qualidade SegSNR (*segmental signal-to-noise ratio*) (HANSEN; PELLOM, 1998) e RMSE (*root mean square error*) são adotadas para avaliar os métodos de realce neste experimento.

Em ambos os experimentos, os sinais são corrompidos por quatro ruídos ambientais com diferentes graus de não-estacionariedade: Bolhas (*Bubbles*), Orca (*Killer Whale*, Terremoto Submarino (*Underwater Earthquake*) e Transatlântico (*Ocean Liner*). Estes ruídos foram adicionados aos sinais de interesse para três valores de SNR: -5 dB, 0 dB e 5 dB.

1.1 Objetivos

O objetivo principal da Dissertação é propor um método para realce, no domínio do tempo, de sinais acústicos de interesse (voz e *chirp*) corrompidos por ruídos subaquáticos

de diferentes índices de não-estacionariedade e originados por distintas fontes acústicas reais, ou seja, de diferentes características temporais e espectrais. Os objetivos específicos são:

- investigar o uso de um critério baseado no índice de não-estacionariedade para detecção e estimação das componentes de ruído nas decomposições do sinal.
- avaliar e comparar o método de realce proposto com outros métodos competitivos propostos na literatura pelos resultados das medidas preditivas, observando a influência dos índices de não-estacionariedade dos ruídos acústicos nestes resultados. Para esta avaliação, são adotadas quatro medidas objetivas de qualidade e três medidas objetivas de inteligibilidade, considerando os dois cenários experimentais distintos.

1.2 Resultados obtidos

Os principais resultados advindos do desenvolvimento desta Dissertação são:

- Proposta de um método de realce que atua no domínio do tempo para sinais acústicos de interesse corrompidos por ruídos acústicos subaquáticos.
- Os resultados observados no experimento de realce de sinais de voz demonstrou que o método proposto obteve aprimoramentos interessantes na qualidade. Destaca-se, aqui, o aprimoramento obtido na medida PESQ com a maior média geral dentre as soluções de realce, e as maiores médias do PEAQ para os ruídos com maior grau de não-estacionariedade.
- Para estes ruídos, o método proposto também apresentou ganhos interessantes e superiores aos algoritmos espectrais na inteligibilidade.
- No realce dos sinais *chirp*, os resultados no incremento do SegSNR e redução do RMSE também foram relevantes, principalmente para os ruídos não-estacionários.
- Definição de um critério baseado na não-estacionariedade para detecção e estimação das componentes pertencentes aos ruídos acústicos. A medida INS mostrou ser eficiente para estimar as componentes mais corrompidas do sinal, em virtude dos efeitos de atenuação da não-estacionariedade de sinais de voz ou sinais *chirp* causados pelas distorções provocadas pelos ruídos. Esta característica pode ser explorada em futuros trabalhos para desenvolvimento de novos métodos de realce ou máscaras acústicas.

1.3 Organização da Dissertação

O restante da Dissertação está organizado da seguinte forma:

- **Capítulo 2:** Neste Capítulo, inicialmente são introduzidos os métodos de realce de sinais competitivos abordados neste trabalho, sendo duas soluções espectrais, OMLSA (COHEN; BERDUGO, 2001) e UMMSE (GERKMANN; HENDRIKS, 2012), e duas temporais, EMDH (ZÃO; COELHO; FLANDRIN, 2014) e NNESE (TAVARES; COELHO, 2016). Ainda neste Capítulo, são apresentadas as medidas objetivas de predição, tanto de qualidade quanto de inteligibilidade, empregadas nesta Dissertação. As medidas de qualidade adotadas aqui são a PESQ (RIX et al., 2001), PEAQ (COLOMES et al., 1999), SegSNR (HANSEN; PELLOM, 1998) e RMSE. Para a inteligibilidade, foram empregadas as seguintes medidas: STOI (TAAL et al., 2011), ESII (RHEBERGEN; VERSFELD, 2005) e ASII_{ST} (HENDRIKS et al., 2015).

- **Capítulo 3:** Neste Capítulo, são descritas as três etapas do método de realce de sinais acústicos proposto neste trabalho. Na primeira etapa, é feita a decomposição do sinal corrompido em IMFs pelo EEMD-IF (WU; HUANG, 2009),(LIN; WANG; ZHOU, 2009). Na segunda etapa, é feita a seleção das IMF para atuação do realce. Para este fim, um critério baseado no INS (BORGNAT et al., 2010) é proposto, considerando as diferenças no grau de não-estacionariedade entre os sinais de interesse e os ruídos acústicos subaquáticos. Finalmente, na terceira e última etapa, um critério de detecção e estimação das componentes do ruído é adotado para cada IMF selecionada em segmentos curtos de tempo. A implementação quadro a quadro permite acompanhar as variações nas características temporais e espectrais ao longo do tempo, especialmente para ruídos não-estacionários. Ainda nesta etapa, é descrita a reconstrução do sinal realçado a partir dos quadros remanescentes de cada modo de oscilação.

- **Capítulo 4:** Os dois experimentos para avaliação de desempenho dos métodos de realce de sinais competitivos e o proposto neste trabalho são descritos neste Capítulo. Inicialmente, apresenta-se os resultados de INS dos sinais de interesse e dos quatro ruídos adotados neste trabalho. Estes ruídos foram coletados em diferentes fontes acústicas subaquáticas reais, com características temporais e espectrais díspares. O primeiro experimento é voltado para o realce de sinais de voz da base TIMIT corrompidos pelos ruídos acústicos selecionados e duas medidas objetivas de qualidade (PESQ e PEAQ) e três de inteligibilidade (STOI, ESII e ASII_{ST}) são empregadas para avaliação do realce. No segundo experimento, as soluções de realce são implementadas para aprimorar um conjunto de sinais *chirp*, sendo estes resultados avaliados por duas medidas de qualidade (SegSNR e RMSE).

- **Capítulo 5:** Por fim, neste Capítulo, são expostas as principais conclusões e contribuições desta Dissertação. Além disso, também são apresentadas sugestões para trabalhos futuros.

2 MÉTODOS DE REALCE DE SINAIS ACÚSTICOS E MEDIDAS OBJETIVAS DE PREDIÇÃO

O realce de sinais acústicos subaquáticos desperta interesse científico, tendo em vista que a redução das distorções causadas pelo ruído ambiente representa um grande desafio nas aplicações associadas à acústica subaquática. Principalmente, quando o ruído é não-estacionário. Os ruídos afetam o desempenho de equipamentos sonar na detecção e classificação de alvos e das comunicações acústicas subaquáticas. Além disso, uma boa comunicação por voz no ambiente subaquático é um requisito essencial para preservar a integridade física de mergulhadores em diversas situações, como no monitoramento e controle de paradas de descompressão, avisos de perigo ou no diagnóstico de problemas médicos (WOODWARD; SARI, 1996).

As fontes de ruído no meio subaquático incluem, principalmente, o ruído ambiente, o ruído próprio e a reverberação do som (OU; ALLEN; SYRMOS, 2011). A reverberação é um efeito gerado pelas múltiplas reflexões de uma onda sonora no fundo do mar ou em objetos submersos, antes desta onda ser captada pelos transdutores, sendo um grande desafio para sonares ativos (WENZ, 1972).

O ruído próprio é aquele gerado na plataforma onde o sistema acústico está instalado, como por exemplo ruídos causados nos circuitos elétricos dos transdutores ou por vibrações mecânicas na estrutura. Este ruído pode ser reduzido ou até mesmo eliminado, tomando-se medidas adequadas durante a instalação e montagem do sistema acústico.

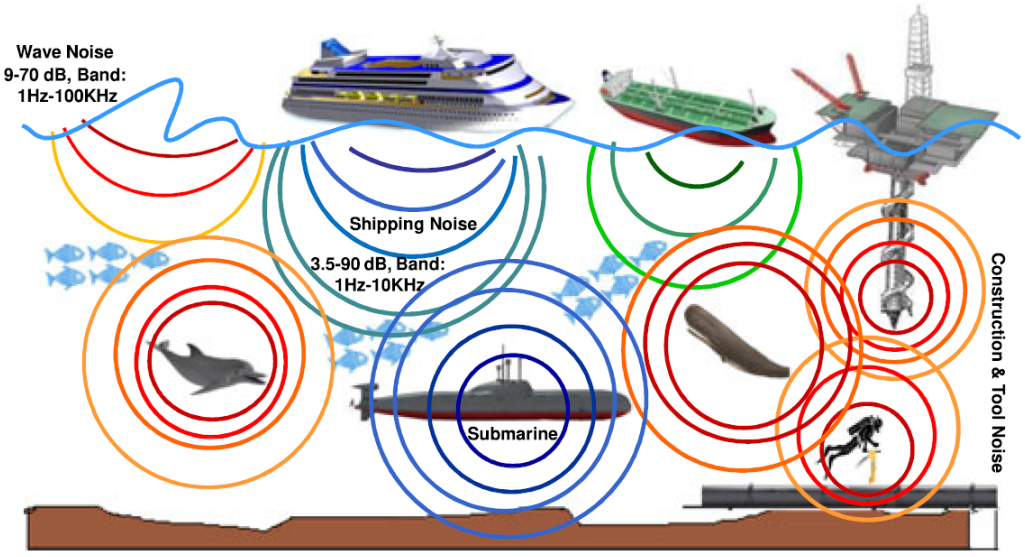

Figura 1 – Exemplos de fontes acústicas de ruídos ambientais subaquáticos. Retirado de (RAHMATI; POMPILI, 2018).

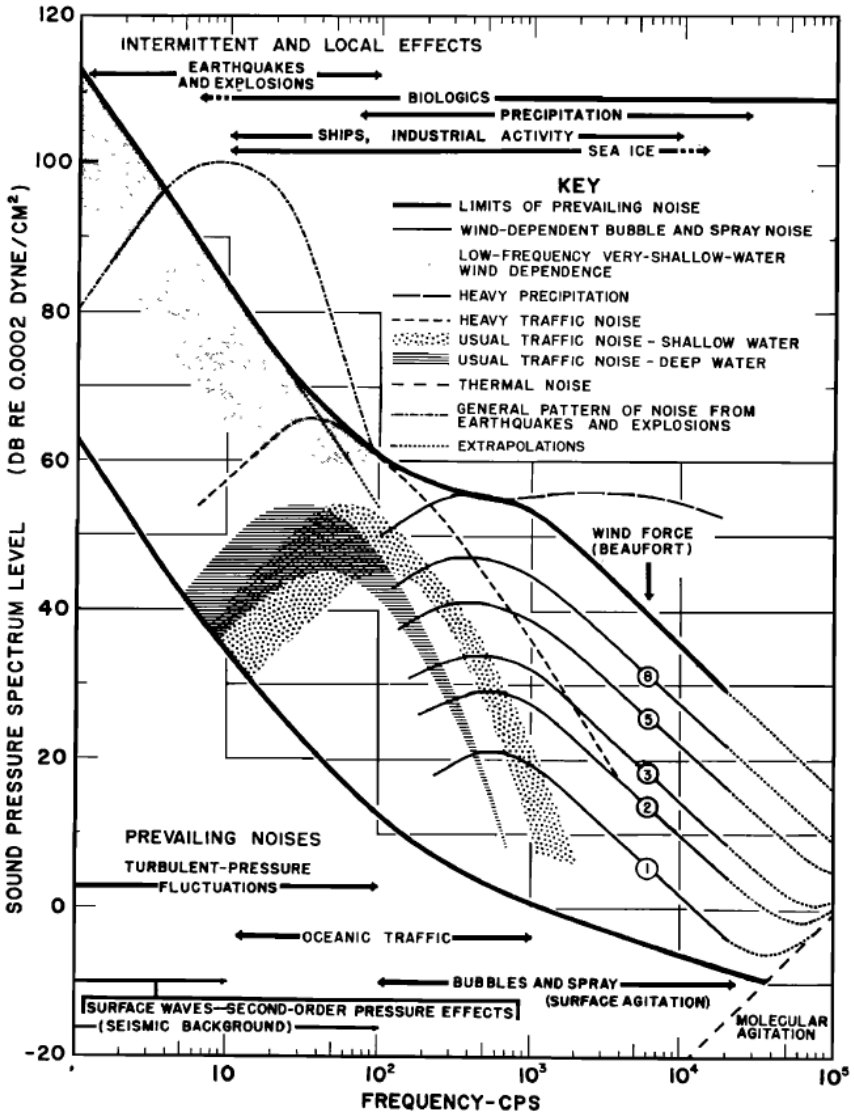

Figura 2 – Curva que relaciona o nível de pressão sonora, em dB, e a faixa de frequência, em Hz, dos ruídos ambientais provocados por fontes acústicas subaquáticas. Retirado de (WENZ, 1962).

Por outro lado, o ruído ambiente subaquático é independente do sistema acústico e pode variar de acordo com a localização geográfica, proximidade das rotas de navegação, estação do ano, profundidade do mar (águas rasas ou profundas) e é resultado de contribuições de diferentes fontes acústicas, como as antropogênicas e as naturais (sísmicas, biológicas e aquelas associadas às condições climáticas). A figura 1 apresenta alguns exemplos de fontes acústicas de ruídos ambientais presentes no meio subaquático.

A figura 2 apresenta um gráfico que relaciona o nível de pressão sonora, em dB, e a faixa de frequência, em Hz, dos ruídos ambientais provocados pelas fontes acústicas subaquáticas, mostrando as distintas características espectrais dos ruídos que compõem o ambiente subaquático.

Neste contexto, as soluções de realce de sinais têm como propósito atenuar os

efeitos causados por estes ruídos acústicos. Os métodos de realce propostos na literatura podem ser classificados, segundo seus domínios de atuação, como espectrais e temporais. Técnicas espectrais convencionais utilizam a STFT (*short-time Fourier transform*) e um mecanismo de detecção de atividade de voz (VAD - *voice activity detector*) para estimar os componentes espectrais do ruído na ausência de voz (BOLL, 1979). Estes métodos têm um bom desempenho quando os ruídos são estacionários, porém outros algoritmos são necessários para lidar com a não-estacionariedade dos ruídos.

O método OMLSA (*optimally-modified log-spectral amplitude*) (COHEN; BERDUGO, 2001) foi proposto para estimar o ruído com estas características, atualizando o espectro do ruído quadro a quadro. A solução UMMSE (*unbiased minimum mean-square error*) (GERKMANN; HENDRIKS, 2012) 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.

Outras soluções de realce são baseadas na análise tempo-frequência utilizando a decomposição empírica de modos (EMD - *empirical mode decomposition*)(HUANG et al., 1998). No EMD, a decomposição resulta em um conjunto de funções intrínsecas de modo (IMF - *intrinsic mode functions*), que são totalmente dependentes do próprio sinal, ou seja, as bases são adaptativas (*data driven*). Dentre estas soluções, pode-se destacar o EMD-DT (*EMD-based detrending*) (FLANDRIN; GONÇALVÈS; RILLING, 2004), EMDF (*EMD-based filtering*) e EMDH (*EMD-Hurst*)(ZÃO; COELHO; FLANDRIN, 2014), (COELHO et al., 2015).

Outra solução temporal presente na literatura é o NNESE (*non-stationary noise estimation for speech enhancement*) (TAVARES; COELHO, 2016), baseado no estimador robusto DATE (*d-dimensional trimmed estimator*) (PASTOR; SOCHELEAU, 2012) para estimar o desvio padrão do ruído a partir do sinal corrompido no domínio do tempo. 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 (TAVARES; COELHO, 2016).

Os métodos de realce de sinais acústicos são geralmente avaliados por medidas de qualidade. A qualidade expressa a relação entre um sinal de voz limpo e a sua versão atenuada. A inteligibilidade é uma medida que reflete o quanto uma mensagem acústica é compreensível e é aferida perceptualmente pelo reconhecimento de palavras.

O aprimoramento dos sinais acústicos obtido pelos métodos de realce pode ser examinado de duas formas:

- avaliação perceptual subjetiva; e

- avaliação objetiva de predição.

Na avaliação perceptual subjetiva, ouvintes são utilizados a fim de julgar a qualidade ou a inteligibilidade do sinal de voz. Embora os testes subjetivos perceptuais sejam a forma mais precisa para julgamento da qualidade de um sinal acústico, o alto custo operacional faz com que estes sejam frequentemente substituídos por medidas objetivas (QUACKENBUSH; BARNWELL; CLEMENTS, 1988), (HU; LOIZOU, 2008).

Uma medida objetiva de predição, seja de qualidade ou de inteligibilidade, é considerada satisfatória quando ela possui alta correlação com os resultados obtidos por testes subjetivos. Contudo, a melhora na qualidade não necessariamente resulta em aprimoramento da inteligibilidade (HU; LOIZOU, 2007). Por conta disso, julga-se necessário que os métodos de realce de sinais acústicos sejam avaliados por medidas apropriadas para analisar estes diferentes aspectos perceptuais.

Inicialmente, são descritos os algoritmos espectrais OMLSA (COHEN; BERDUGO, 2001) e UMMSE (GERKMANN; HENDRIKS, 2012). Depois, são introduzidas as técnicas temporais EMDH (ZÃO; COELHO; FLANDRIN, 2014) e NNESE (TAVARES; COELHO, 2016). Em seguida, são abordadas as medidas objetivas de predição de qualidade e inteligibilidade. Primeiramente, são apresentadas as quatro medidas de qualidade PESQ (*perceived evaluation of speech quality*) (RIX et al., 2001), PEAQ (*perceptual evaluation of audio quality*) (COLOMES et al., 1999) e SegSNR (*Segmental SNR*) (HANSEN; PELLOM, 1998) e RMSE (*root mean square error*). Finalmente, expõe-se as medidas de inteligibilidade STOI (*short-time objective intelligibility*) (TAAL et al., 2011), ESII (*extended speech intelligibility index*) (RHEBERGEN; VERSFELD, 2005) e ASII_{ST} (*short-time approximated SII*).

Neste trabalho, os métodos de realce são empregados para tratar sinais de voz e um conjunto de sinais *chirp*. Estes sinais são degradados por quatro ruídos acústicos ambientais do meio subaquático, com diferentes características espectrais e graus de não-estacionariedade. Para os sinais de voz, as medidas PESQ e PEAQ são adotadas para a predição de qualidade obtida pelos métodos de realce, enquanto as medidas ESII e ASII_{ST} avaliam a inteligibilidade. As medidas objetivas SegSNR e RMSE são consideradas para a predição de qualidade para o realce do conjunto de sinais *chirp*.

2.1 MÉTODOS DE REALCE DE SINAIS ACÚSTICOS

Esta seção descreve os métodos de realce de sinais acústicos competitivos adotados neste trabalho. Primeiramente, são apresentadas as técnicas espectrais OMLSA e UMMSE, que utilizam a estimação do espectro de potência do ruído. Em seguida, são abordadas as propostas de realce que estimam as componentes do ruído no domínio do tempo, EMDH e

NNESE.

2.1.1 OMLSA

O método OMLSA, definido em (COHEN; BERDUGO, 2001), emprega o estimador IMCRA (COHEN, 2003) para aferir o espectro de potência dos ruídos acústicos. Este estimador realiza duas iterações, sendo que cada iteração é composta por uma etapa de suavização do espectro de potência do sinal ruidoso e outra etapa de localização por estatísticas mínimas (MARTIN, 2001).

Na primeira iteração, o sinal degradado é mapeado para o domínio de tempo-frequência por meio do STFT (*short-time Fourier transform*). Em seguida, uma versão suavizada de $|T(\kappa,\tau)|^2$ na frequência ($S_f(\kappa,\tau)$) e no tempo ($S(\kappa,\tau)$) é obtida por

$$\begin{cases} S_f(\kappa,\tau) = \sum_{i=-w}^{w} W(i)|Y(\kappa-i,\tau)|^2, \\ S(\kappa,\tau) = \delta_s S(\kappa,\tau-1) + (1-\delta_s)S_f(\kappa,\tau), \end{cases} \tag{2.1}$$

onde $W(i)$ é uma janela normalizada para calcular a média entre valores adjacentes em frequência de $|Y(\kappa,\tau)|^2$, e $\delta_s \in [0,1]$ é o parâmetro de suavização no tempo que atualiza os valores de $S(\kappa,\tau)$ de forma recursiva. O espectro de potência do ruído é, então, estimado a partir dos valores mínimos de $S(\kappa,\tau)$ em um conjunto de Q quadros passados, obtendo-se $S_{min}(\kappa,\tau)$. Dessa forma, considera-se que, no mínimo, em um destes Q quadros anteriores, a voz estará ausente, e

$$E\{S_{min}(\kappa,\tau)\} = B_{min}^{-1} E\{|\mathcal{N}(\kappa,\tau)|^2\} \tag{2.2}$$

em que B_{min} é um fator de correção de tendência (*bias*) que pode ser determinado de forma empírica. Em (COHEN, 2003), o valor atribuído a este fator foi de $B_{min} = 1,66$.

Ao término da primeira iteração, um VAD é definido para cada quadro e cada índice de frequência, por meio dos seguintes parâmetros:

$$\begin{aligned} \gamma_{min}(\kappa,\tau) &\triangleq \frac{|Y(\kappa,\tau)|^2}{B_{min}S_{min}(\kappa,\tau)} \\ \zeta(\kappa,\tau) &\triangleq \frac{S(\kappa,\tau)}{B_{min}S_{min}(\kappa,\tau)} \end{aligned} \tag{2.3}$$

A ausência ou presença de voz, em cada quadro e índice de frequência, é determinada de acordo com o seguinte critério de decisão:

$$I(\kappa,\tau) = \begin{cases} 1, \text{se } \gamma_{min}(\kappa,\tau) < \gamma_0 \text{ e } \zeta(\kappa,\tau) < \zeta_0 \text{ (voz ausente)} \\ 0, \text{ caso contrário (voz presente)} \end{cases} \tag{2.4}$$

Na segunda iteração, um novo espectro suavizado $\tilde{S}_f(\kappa,\tau)$ é determinado a partir dos segmentos em que o algoritmo não detectou atividade da voz, ou seja, $I(\kappa,\tau)=1$. Analogamente, a partir de $\tilde{S}_f(\kappa,\tau)$ são executados os mesmos passos observados na primeira iteração. Considerando-se, respectivamente, as hipóteses de ausência e presença de voz $\mathcal{H}_0(\kappa,\tau)$ e $\mathcal{H}_1(\kappa,\tau)$ no quadro τ e índice de frequência κ, a probabilidade condicional de presença de voz $p(\kappa,\tau) \triangleq P(\mathcal{H}_0(\kappa,\tau)|\gamma(\kappa,\tau))$ pode ser calculada pela seguinte equação:

$$p(\kappa,\tau) = (1+\frac{q(\kappa,\tau)}{1-q(\kappa,\tau)}(1+\xi(\kappa,\tau))\exp\{\upsilon(\kappa,\tau)\})^{-1}, \tag{2.5}$$

em que $\upsilon \triangleq \dfrac{\gamma\xi}{(\xi+1)}$ e a probabilidade *a priori* de ausência de voz, $q(\kappa,\tau)=P(\mathcal{H}_0(\kappa,\tau))$ pode ser estimada por

$$\hat{q}(\kappa,\tau) = \begin{cases} 1, \text{ se } \hat{\gamma}_{min}(\kappa,\tau) \leq 1 \text{ e } \hat{\zeta}(\kappa,\tau) < \zeta_0; \\ \dfrac{\gamma_1 - \tilde{\gamma}_{min}(\kappa,\tau)}{\gamma_1 - 1}, \text{ se } 1 < \hat{\gamma}_{min}(\kappa,\tau) \leq \gamma_1 \text{ e } \hat{\zeta}(\kappa,\tau) < \zeta_0; \\ 0, \text{ em outros casos.} \end{cases} \tag{2.6}$$

Em seguida, o espectro de potência do ruído do quadro subsequente $|\bar{\mathcal{N}}(\kappa,\tau+1)|^2)$ pode ser recursivamente estimado por

$$|\bar{\mathcal{N}}(\kappa,\tau+1)|^2 = \tilde{\delta}_\eta(\kappa,\tau)|\bar{\mathcal{N}}(\kappa,\tau)|^2 + [1-\tilde{\delta}_\eta(\kappa,\tau)]|Y(\kappa,\tau)|^2, \tag{2.7}$$

em que $\tilde{\delta}_\eta(\kappa,\tau)$ é um parâmetro de suavização que depende de $p(\kappa,\tau)$ e uma constante $\delta_\eta \in [0,1]$,

$$\tilde{\delta}_\eta(\kappa,\tau) \triangleq \delta_\eta + (1-\delta_\eta)p(\kappa,\tau). \tag{2.8}$$

Por fim, a versão final do espectro do ruído é obtida multiplicando-se o espectro do ruído estimado por um novo fator de correção de tendência B, em virtude deste valor ser subestimado pelo estimador IMCRA, uma vez que este é derivado do método de estatísticas mínimas. Assim:

$$|\hat{\mathcal{N}}(\kappa,\tau)|^2 = B|\bar{\mathcal{N}}(\kappa,\tau)|^2 \tag{2.9}$$

Após a implementação do IMCRA, o algoritmo OMLSA (COHEN; BERDUGO, 2001) 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, ou seja,

$$E_{min}\{(\log|X(\kappa,\tau)| - \log|\hat{X}(\kappa,\tau)|)^2\}. \tag{2.10}$$

A função ganho que multiplica o espectro do sinal de entrada e resulta na amplitude espectral da reconstrução ótima da voz é definida em (COHEN; BERDUGO, 2001) como:

$$G_{OMLSA}(\kappa,\tau) = G_{LSA}(\kappa,\tau)^{p(\kappa,\tau)}G_{min}^{1-p(\kappa,\tau)}, \tag{2.11}$$

sendo τ e κ os índices de quadro e frequência, respectivamente, $G_{LSA}(\kappa,\tau)$ um ganho do estimador LSA calculado em função do SNR *a priori* e deduzido em (EPHRAIM; MALAH, 1984), e G_{min} um limiar mínimo para o ganho e correspondente à -25 dB (COHEN; BERDUGO, 2001).

2.1.2 UMMSE

O estimador UMMSE (GERKMANN; HENDRIKS, 2012) é derivado do estimador MMSE (*minimum mean-square error*) proposto em (HENDRIKS; HEUSDENS; JENSEN, 2010) que tem como objetivo estimar as componentes espectrais do ruído a partir da minimização dos erros quadráticos médios. Diferentemente do IMCRA, nesta solução não é necessário processar informações de vários quadros anteriores para a estimação do espectro do ruído. Esta característica faz com que o UMMSE apresente um menor atraso para captar as variações no espectro dos ruídos não-estacionários, reduzindo a sua complexidade computacional. Também vale ressaltar que o UMMSE não requer um fator de compensação de tendência. No estimador MMSE, considera-se a hipótese de que os componentes de espectro do ruído e do sinal de voz apresentam distribuição Gaussiana (HENDRIKS; HEUSDENS; JENSEN, 2010). Assim, o valor do periodograma do ruído $|\mathcal{N}(\kappa,\tau)|^2$ é definido por

$$E[|\mathcal{N}(\kappa,\tau)|^2 \mid Y(\kappa,\tau)] = (\frac{1}{1+\hat{\xi}(\kappa,\tau)})^2|Y(\kappa,\tau)|^2 + \frac{\hat{\xi}(\kappa,\tau)}{1+\hat{\xi}(\kappa,\tau)}|\hat{\mathcal{N}}(\kappa,\tau-1)|^2. \quad (2.12)$$

O SNR *a posteriori* $\hat{\gamma}(\kappa,\tau)$ é determinado considerando o espectro de potência do ruído obtido no quadro anterior, uma vez que a variação das características espectrais do ruído entre quadros subsequentes tende a ser menor que a da voz. Assim:

$$\hat{\gamma}(\kappa,\tau) = \frac{|Y(\kappa,\tau)|^2}{|\hat{\mathcal{N}}(\kappa,\tau-1)|^2}, \quad (2.13)$$

enquanto o SNR *a priori* é estimado por

$$\hat{\xi}(\kappa,\tau) = \max\{\hat{\gamma}(\kappa,\tau) - 1, 0\}. \quad (2.14)$$

Finalmente, a estimação do espectro de potência do ruído é atualizada quadro a quadro de forma suavizada e recursiva,

$$|\hat{\mathcal{N}}(\kappa,\tau)|^2 = \alpha_p|\hat{\mathcal{N}}(\kappa,\tau-1)|^2 + (1-\alpha_p)E[|\mathcal{N}(\kappa,\tau)|^2 \mid Y(\kappa,\tau)], \quad (2.15)$$

sendo $\alpha_p = 0,8$ uma constante de suavização definida em (HENDRIKS; HEUSDENS; JENSEN, 2010).

No UMMSE, foi proposta uma alteração deste estimador, reformulando a equação (2.14) utilizando as probabilidades condicionais de ausência ($P(\mathcal{H}_0|Y)$) e presença de voz ($P(\mathcal{H}_1|Y)$) para a estimação do periodograma do ruído $E[|\mathcal{N}(\kappa,\tau)|^2|Y(\kappa,\tau)]$,

$$E(|\mathcal{N}|^2|Y) = P(\mathcal{H}_0|Y)|Y|^2 + P(\mathcal{H}_1|Y)|\hat{\mathcal{N}}^2. \quad (2.16)$$

Para resolver a equação 2.16, a probabilidades condicionais são calculadas como

$$P(\mathcal{H}_1|Y(\kappa,\tau)) = (1 + (1+\xi_{opt})e^{-\hat{\gamma}(\kappa,\tau)\frac{\xi_{opt}}{1+\xi_{opt}}})^{-1}, \tag{2.17}$$

e $P(\mathcal{H}_0|Y(\kappa,\tau)) = 1 - P(\mathcal{H}_1|Y(\kappa,\tau))$. O valor ótimo do SNR *a priori* ξ_{opt} foi definido como 15 dB (GERKMANN; HENDRIKS, 2012).

Após a estimação das componentes espectrais do ruído, o ganho de Wiener G_W (SCALART; FILHO, 1996) é aplicado no espectro do sinal corrompido para reconstrução do sinal de interesse. O filtro de Wiener é um estimador ótimo para a minimização do erro quadrático médio dos coeficientes espectrais obtidos para o sinal de voz limpo. Este filtro depende do SNR *a priori* $\xi(\kappa,\tau)$,

$$G_W(\kappa,\tau) = \frac{\xi(\kappa,\tau)}{1+\xi(\kappa,\tau)}, \tag{2.18}$$

sendo $\xi(\kappa,\tau)$ obtido pelo método da decisão direta desenvolvida em (EPHRAIM; MALAH, 1984) como

$$\hat{\xi}(\kappa,\tau) = \alpha_W G_W^2(\kappa,\tau-1)\gamma(\kappa,\tau-1) + (1-\alpha_W \max\{\gamma(\kappa,\tau)-1,0\}, \tag{2.19}$$

sendo o valor adotado em (GERKMANN; HENDRIKS, 2012) para a constante de suavização de α_W igual a 0,98 (SCALART; FILHO, 1996).

2.1.3 NNESE

O método NNESE (TAVARES; COELHO, 2016) faz a detecção das componentes ruidosas presentes no sinal de voz por meio da estimação do desvio padrão do ruído a partir do sinal corrompido, pela adoção do algoritmo DATE (PASTOR; SOCHELEAU, 2012). Neste estimador, não é necessário conhecimento *a priori* da distribuição das amostras do sinal para obter a estimativa do desvio padrão do ruído.

Além disso, o algoritmo DATE considera duas hipóteses: a norma das amplitudes do sinal deve estar acima de um limiar inferior conhecido e a probabilidade de ocorrência do sinal de voz deve ser menor que 0,5. Nesta solução de realce, o DATE foi adaptado para estimar o desvio padrão de ruídos acústicos não-estacionários.

A atuação do NNESE é composta de três etapas. A primeira etapa consiste na identificação e estimação das componentes do ruído. Inicialmente, o limiar de estimação $\xi(\rho)$ e o grau de confiança Q são calculados segundo as fórmulas abaixo:

$$\begin{cases} \xi(\rho) = \dfrac{1}{2}\rho + \dfrac{1}{\rho} log(1+\sqrt{1-exp(-\rho^2)}) \\ Q \leq 1 - \dfrac{K}{4(\dfrac{K}{2}-1)^2}. \end{cases} \tag{2.20}$$

em que K é o tamanho total de $y_{(k)}$, e ρ é a razão entre a média de todos os valores de amplitude do sinal corrompido e o desvio padrão dos seus valores mínimos.

Em (PASTOR; SOCHELEAU, 2012), para ruídos Gaussianos adotou-se os valores de $\rho = 4$, $\xi(\rho) = 3,4742$ e $Q = 95\%$. Em seguida, a sequência amostral do sinal corrompido $\{y(1), y(2), \ldots, y(k)\}$ é rearranjada na ordem crescente do valor de amplitude $Y_1, \leq Y_2, \ldots, \leq Y_k$.

Para estimação do desvio padrão do ruído acústico, inicialmente calcula-se o k_{min}, que representa a quantidade de amostras na qual os k primeiros valores do sinal corrompido são constituídos apenas por ruídos, dado um grau de confiança. O valor de k_{min} é determinado conforme a desigualdade de Bienaymé-Chebyshev (ROUSSEEUW; RONCHETTI, 1981):

$$k_{min} = \frac{K}{2} - hK \tag{2.21}$$

em que $h = \dfrac{1}{\sqrt{4K(1-Q)}}$. O próximo passo consiste em verificar se existe um valor inteiro mínimo b em $\{Y_{(kmin)}, \ldots, Y_{(k)}\}$ tal que

$$||Y_{(k-1)}|| \leq \frac{[\sum_{i=1}^{k} ||Y||\xi(\rho)]}{\lambda k} < ||Y_{(k+1)}||, \tag{2.22}$$

sendo $||\cdot||$ a norma euclidiana e λ um fator de ajuste do limiar de estimação em função da dimensão da sequência de amostras. Caso exista este valor mínimo, $b = k$, caso contrário $b = k_{min}$.

Com este valor de b_q, o desvio padrão estimado do ruído quadro a quadro é dado por

$$\sigma_q = \frac{[\sum_{i=1}^{b_q} ||Y||]\xi(\rho)}{\lambda b}. \tag{2.23}$$

Na segunda etapa, as componentes ruidosas são selecionadas a partir de um critério de decisão baseada no valor de amplitude $y(b_q)$. Os valores de amplitude do sinal abaixo deste limiar são correspondentes às estas componentes, enquanto os valores acima do limiar são considerados componentes do sinal de voz limpo.

Por fim, na última etapa é feita a reconstrução do sinal de voz, conforme pode ser observada na equação a seguir:

$$\tilde{y}_q = \begin{cases} y_q(k) - \alpha\hat{\sigma}_q, & \text{se } y_q(k) \geq y(b_q). \\ \beta y_q(k), & \text{caso contrário.} \end{cases} \tag{2.24}$$

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. O sinal realçado, então, é obtido pela concatenação

de todos os quadros obtidos em (2.24), ou seja,

$$\tilde{y}(k) = \sum_{q=0}^{Q-1} \tilde{y}_q(k - qK). \tag{2.25}$$

Neste trabalho, foram adotados os valores de $\alpha = 0{,}35$ e $\alpha = 0{,}1$ para o aprimoramento da qualidade e inteligibilidade do sinal de voz. Para o aprimoramento da qualidade do sinal *chirp* foi adotado $\alpha = 0{,}65$. Em (TAVARES; COELHO, 2016), o NNESE apresentou resultados promissores no aprimoramento das medidas de predição para realce de sinais de voz, principalmente na presença de ruídos altamente não-estacionários.

2.1.4 EMDH

No método de realce EMDH (ZÃO; COELHO; FLANDRIN, 2014), inicialmente o sinal é decomposto via EMD. O EMD foi proposto em (HUANG et al., 1998) para análises de sinais não-estacionários oriundos de sistemas não-lineares. O método é baseado em um conjunto de funções intrínsecas de modo (IMF), resultado do EMD, que são inteiramente dependentes do sinal analisado, e um resíduo. As primeiras IMF possuem oscilações mais rápidas (altas frequências), ao passo que as IMFs de maior índice possuem oscilações mais lentas (baixas frequências). Esta característica é explorada por diversas técnicas de realce de sinais acústicos (CHATLANI; SORAGHAN, 2012),(ZÃO; COELHO; FLANDRIN, 2014), que adotam um critério de seleção visando detectar as IMF mais corrompidas pelos ruídos, que geralmente estão concentrados nas baixas frequências.

Na proposta EMDH, expoente de Hurst (HURST, 1951) é adotado como critério de detecção das IMF mais corrompidas pelo ruído quadro a quadro. Desta forma, é possível identificar as variações das características do ruído no tempo, sendo um método adequado para realçar sinais acústicos na presença de ruídos não-estacionários. O expoente de Hurst de uma série temporal $y(t)$ é determinado pela taxa de decaimento de sua função de autocorrelação normalizada $\rho(k)$. Sendo assim, o valor de H também está relacionado com as características espectrais de $y(t)$, uma vez que a densidade espectral de potência ($S_y(f)$) é a Transformada de Fourier de $\rho(k)$. A DEP é proporcional à f^{1-2H}, ou seja,

$$S_y(f) = \mathcal{F}\{\rho(k)\} \propto f^{1-2H}, f \to 0 \tag{2.26}$$

em que $\mathcal{F}\{\cdot\}$ representa a Transformada de Fourier.

A partir desta relação, é possível concluir que:

- $H < 1/2$: $S_y(f)$ é composta, predominantemente, por altas frequências;
- $H = 1/2$: $S_y(f)$ é aproximadamente constante ao longo de todo o espectro de potências, sendo correspondente ao ruído branco; e

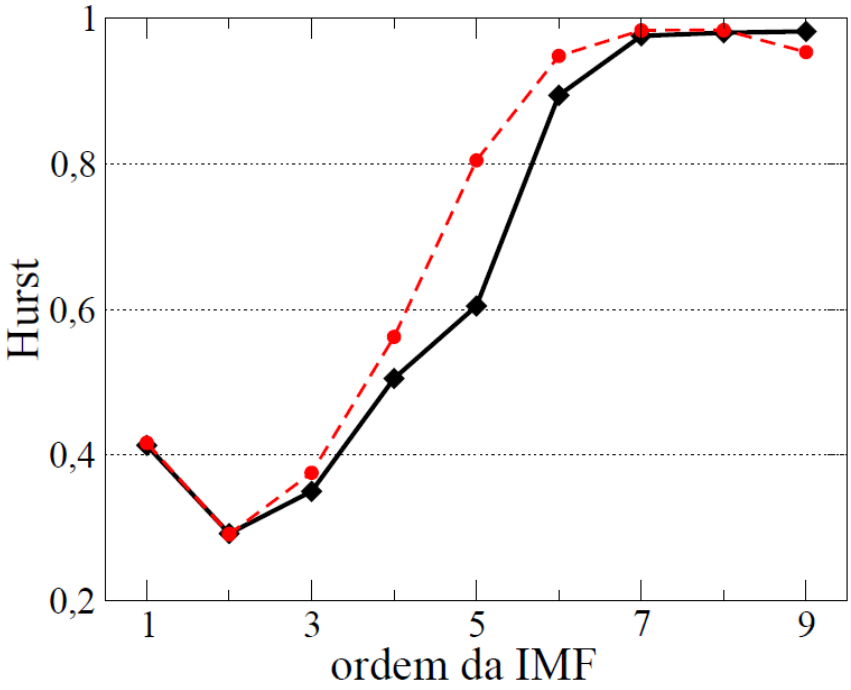

Figura 3 – Valores de Expoente de Hurst para cada IMF de um sinal de voz limpo e ruidoso. A linha preta contínua indica os valores de H estimados das IMF de um sinal de voz limpo. A linha vermelha tracejada apresenta os valores de H deste mesmo sinal de voz corrompido por ruído fábrica com SNR de 0 dB (ZÃO; COELHO; FLANDRIN, 2014).

- $H > 1/2$: $S_y(f)$ é composta, predominantemente, por baixas frequências.

Dessa forma, o Expoente de Hurst possibilita a detecção das IMF com maior presença das componentes dos ruídos acústicos de baixas frequências. Na figura 3, é possível observar que as primeiras IMFs, que concentram as componentes de alta frequência, possuem valores de $\hat{H}$ no intervalo $(0, 1/2)$. Por outro lado, os modos de maior índice, como as IMF de 7 a 9, possuem $H \approx 1$, sendo correspondentes às componentes em que os ruídos acústicos geralmente estão concentrados.

No EMDH, o sinal de voz ruidoso $y(t)$ é decomposto em M modos, e em seguida cada IMF é dividida em quadros não sobrepostos de curta duração, ou seja,

$$\text{w-IMF}_{m,q}(t) = \begin{cases} \text{IMF}_m(t + qT_d), t \in [0, T_d], \\ 0, \text{ caso contrário}, \end{cases} \tag{2.27}$$

em que $q \in \{0, \ldots, Q-1\}$ representa o índice de quadros e T_d a duração de cada quadro. Neste técnica, cada quadro possui 512 amostras, ou 32 ms com taxa de amostragem de 16 kHz. Para cada quadro q, estima-se o valor do expoente de Hurst H_m referente à IMF janelada $\text{w-IMF}_{m,q}(t)$. Nesta proposta, a estimação do H foi obtida pelo método baseado em *wavelets* (VEITCH; ABRY, 1999). Estes valores compõem o vetor H_q com M coeficientes $(m = 1, \ldots, M)$.

A etapa seguinte consiste em determinar a última IMF janelada cujo valor estimado do Expoente de Hurst para aquele quadro seja inferior a um limiar H_{lim}. A escolha do limiar H_{lim} é muito importante, uma vez que representa uma relação de compromisso entre a parcela do ruído de baixa frequências que será removida e a distorção causada pela supressão das componentes do sinal de interesse. Em (ZÃO; COELHO; FLANDRIN, 2014), o limiar determinado empiricamente e adotado nos experimentos foi de $H_{lim} = 0,9$, valor considerado satisfatório para reduzir as interferências causadas pelo ruído, sem distorcer drasticamente o sinal de interesse.

Cada quadro do sinal de voz realçado $\hat{x}_q(t)$ é reconstruído pela seguinte equação:

$$\hat{x}_q(t) = \sum_{m=1}^{N_q} \text{w-IMF}_{m,q}(t), q = 0, \ldots, Q-1, \tag{2.28}$$

e o sinal de voz completo $\hat{x}(t)$ é,finalmente, obtido pela somatório desses quadros,

$$\hat{x}(t) = \sum_{q=0}^{Q-1} \hat{x}_q(t - qT_d). \tag{2.29}$$

2.2 MEDIDAS OBJETIVAS DE QUALIDADE

As medidas objetivas de qualidade aplicadas neste estudo são, brevemente, apresentadas para avaliar o desempenho dos métodos de realce no aprimoramento da qualidade dos sinais acústicos de interesse. As medidas de qualidade têm o propósito de medir a atenuação das interferências ruidosas obtida pelos métodos de realce.

2.2.1 PESQ

A medida PESQ (RIX et al., 2001), inicialmente recomendada pela ITU (*International Telecommunications Union*) para avaliação da qualidade em codificadores de voz e canais telefônicos de banda estreita, também é amplamente empregada como medida de qualidade para técnicas de realce de sinais.

Inicialmente, os sinais limpo e degradado são nivelados para um nível de audição padrão, levando-se em consideração os ganhos e perdas do sistema de comunicação em teste. Em seguida, os sinais são filtrados via FFT (*fast Fourier transform*) e alinhados no tempo. Os sinais são separados em elocuções e a estimação do atraso para cada elocução é repassada ao modelo perceptual. Este modelo mapeia os sinais em uma representação da percepção de volume em tempo-frequência, levando em consideração as percepções dos sinais sonoros pelo sistema auditivo humano.

Dois parâmetros de distúrbio, simétrico (d_{SYM}) e assimétrico (d_{ASYM}), são extraídos através da diferença entre as representações do sinal limpo e o degradado. Uma combinação

linear destes parâmetros fornece uma medida MOS (*mean opinion score*), definida como na equação a seguir,

$$\text{PESQMOS} = 4.5 - 0.1d_{SYM} - 0.0309d_{ASYM}. \tag{2.30}$$

Os valores dessa medida geralmente ficam limitados entre 1,0 (ruim) e 4,5 (sem distorção), porém em situações de grandes distorções os valores do MOS podem atingir valores inferiores à 1,0.

2.2.2 PEAQ

O algoritmo PEAQ (COLOMES et al., 1999) mede a qualidade de sinais de áudio, definida na Recomendação ITU-R BS.1387, e possui as versões básica e avançada. Na versão básica, o modelo auditivo é baseado apenas na FFT e é designada para aplicações que necessitem de alta velocidade de processamento. Na versão avançada, além da FFT também é adotado um banco de filtros para modelar o ouvido humano, e é voltada para aplicações que requerem maior acurácia na avaliação, ao custo de maior complexidade computacional (TORCOLI; KASTNER; HERRE, 2021).
Em ambas as versões, são adotadas variáveis de saída do modelo (MOV - *model output variables*), que são atributos perceptuais relevantes que medem as distorções causadas por ruídos no tempo e na frequência entre o sinal de referência e o corrompido. Estas MOVs são ponderadas e combinadas por uma rede neural treinada, gerando duas variáveis de saída: Índice de Distorção (DI - *distortion index*) e Grau de Diferença Objetiva (ODG - *overall difference grade*).
Neste trabalho, o resultado final do PEAQ foi obtido de acordo com o critério a seguir:

$$\begin{cases} \text{PEAQ} = \text{DI, se ODG } \leq -3,6\text{; e} \\ \text{PEAQ} = \dfrac{\text{ODG} + \text{DI}}{2}\text{, caso contrário.} \end{cases} \tag{2.31}$$

O resultado final varia de -4, que seria uma distorção muito desagradável, a 0, que seria uma distorção imperceptível ao sistema auditivo humano. Para converter para a escala MOS, soma-se 5 neste resultado.

2.2.3 SegSNR

A medida SegSNR (HANSEN; PELLOM, 1998), ou razão sinal-ruído segmental, é uma medida de qualidade cujo valor representa 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:

$$\text{SegSNR} = \frac{10}{Q}\sum_{\tau=0}^{Q-1}\log\frac{\sum_{t=\tau T_{sh}}^{\tau T_{sh}+T_d-1}x^2(t)}{\sum_{t=\tau T_{sh}}^{\tau T_{sh}+T_d-1}[x(t)-\hat{x}(t)]^2} \tag{2.32}$$

onde T_d é a quantidade de amostras para cada quadro, T_{sh} o deslocamento, em amostras, entre quadros consecutivos e Q o total de quadros. Uma limitação decorrente desta definição é que, nos quadros onde a diferença entre as energias do sinal de interesse e do ruído é muito significativa, o logaritmo calculado dentro do somatório pode resultar em valores muito pequenos ou muito grandes, comprometendo o cálculo final do somatório. Para solucionar este problema, no cômputo da SegSNR os valores obtidos de SNR em cada quadro são limitados entre -10 dB e 35 dB. Desta forma, não é necessário implementar um detector de atividade da voz.

2.2.4 RMSE

A raiz do erro quadrático médio RMSE é amplamente empregada na avaliação de métodos que visam reduzir as distorções causadas por ruídos no ambiente subaquático (AL-ABOOSI; SHA'AMERI, 2017). Seja $x(n)$ o sinal de referência e $\hat{x}(n)$ o sinal realçado, o RMSE é calculado da seguinte maneira:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[\hat{x}(n) - x(n)]^2}. \tag{2.33}$$

O RMSE fornece uma medida de erro entre o sinal de referência e o realçado e possui uma relação inversamente proporcional com o SNR, ou seja, quanto maior o SNR, menor o RMSE.

2.3 MEDIDAS OBJETIVAS DE INTELIGIBILIDADE

As medidas de inteligibilidade empregadas neste estudo são descritas para avaliar a melhora na inteligibilidade dos sinais de voz. As medidas de inteligibilidade visam avaliar o número de acertos de sentenças obtidas a partir de um sinal de voz realçado.

2.3.1 STOI

A medida STOI foi proposta em (TAAL et al., 2011) para estimar a degradação na inteligibilidade de sinais de voz causada por algoritmos de supressão de ruídos através do cálculo do coeficiente de correlação entre os espectros dos sinais limpo e realçado. Na obtenção desta medida de inteligibilidade, o sinal de voz limpo $x(t)$ é, inicialmente, reamostrado a taxa de 10 kHz e divididos em janelas de Hamming de 256 amostras e com 50% de sobreposição. Na sequência, cada segmento é transformado para o domínio da frequência utilizando-se a DFT com 512 pontos. 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. A norma da j-ésima banda, $j = 1, 2, \ldots, 15$, é computada a partir de

$$\overline{X}_j(\tau) = \sqrt{\sum_{\kappa=\kappa_l(j)}^{\kappa_u(j)-1} |X(\kappa,\tau)|^2}, \tag{2.34}$$

sendo $\kappa_u(j)$ e $\kappa_l(j)$ são, respectivamente, os limites superior e inferior. A partir desta norma, o envelope temporal de cada banda do sinal limpo é definido em cada região de tempo e frequência como:

$$x_{(j,\tau)} = [\overline{X}_j(\tau - N + 1), \overline{X}_j(\tau - N + 2), \ldots, \overline{X}_j(\tau)]^T. \tag{2.35}$$

sendo $N = 30$ coeficientes para o vetor $x_{(j,\tau)}$, ou 30 quadros consecutivos na análise temporal, correspondente a 384 ms, ou seja, um quadro a cada 12,8 ms. Analogamente ao $x_{(j,\tau)}$ para o sinal limpo, obtém-se o vetor $y_{(j,\tau)}$ referente ao sinal corrompido. O vetor $y_{(j,\tau)}$ é, então, normalizado para compensar eventuais diferenças de energia em relação ao vetor $x_{(j,\tau)}$. A versão normalizada do n-ésimo coeficiente do vetor $y_{(j,\tau)}$ é dado por:

$$\bar{y}_{j,\tau}(n) = \min(\frac{||x_{(j,\tau)}||}{||y_{(j,\tau)}||} y_{(j,\tau)}, (1 + 10^{-\beta_{SDR}/20}) x_{(j,\tau)}(n)), \tag{2.36}$$

em que $||\cdot||$ representa norma do vetor e β_{SDR} o valor máximo para a grandeza SDR (*signal-to-distortion ratio*) que é definida em (TAAL et al., 2011).

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, ou seja,

$$\text{STOI}_{(j,\tau)} = \frac{(x_{(j,\tau)} - \mu_{x(j,\tau)})^T (\bar{y}_{(j,\tau)} - \mu_{\bar{y}(j,\tau)})}{||(x_{(j,\tau)} - \mu_{x(j,\tau)})||\, ||\bar{y}_{(j,\tau)} - \mu_{\bar{y}(j,\tau)}||}, \tag{2.37}$$

em que $\mu_{(\cdot)}$ representa a média do vetor correspondente. Por fim, a medida STOI é calculada a partir da média de todos os valores de $\text{STOI}_{(j,\tau)}$ calculados em (2.37):

$$\text{STOI} = \frac{1}{15Q} \sum_{j=1}^{15} \sum_{\tau=1}^{Q} \text{STOI}_{(j,\tau)}, \tag{2.38}$$

sendo Q o número total de quadros.

2.3.2 ESII

A medida ESII foi definida em (RHEBERGEN; VERSFELD, 2005) como uma adaptação da medida SII definida em ANSI S3.5-1997 (PAVLOVIC, 2018). Na medida SII, um grau de importância γ_k é atribuído a cada banda de frequência, de acordo com a contribuição desta banda na inteligibilidade do sinal de voz, resultando na função de importância de banda (BIF - *band-importance function*), de modo que $\sum_k \gamma_k = 1$. Esta função multiplica a SNR obtida em cada banda de frequência, resultando em um valor

que representa a inteligibilidade para cada banda. A medida SII, então, é determinada pela média destes valores nas diferentes faixas de frequência.

Uma desvantagem na adoção do SII é o fato de que as variações nas características temporais dos sinais acústicos são perdidas, principalmente na presença dos ruídos não-estacionários, uma vez que o SII calcula a média da inteligibilidade. Para corrigir este problema, na ESII os sinais são divididos em pequenos quadros de tempo, e o SII é calculado para cada quadro. O SNR calculado para cada quadro τ e banda de frequência κ pode ser definido por

$$\xi(\kappa,\tau) = \frac{\sigma_S^2(\kappa,\tau)}{\sigma_N^2(\kappa,\tau)}. \tag{2.39}$$

Na medida ESII, os SNRs são limitados aos valores de 15 dB e -15 dB e normalizados para que os valores finais $d(\kappa,\tau)$ sejam limitados a um intervalo de 0 a 1:

$$d(\kappa,\tau) = \frac{\max(\min(10\log_{10}\xi(\kappa,\tau),15),-15)}{30} + \frac{1}{2}. \tag{2.40}$$

Finalmente, a medida ESII é, então, computada como a média ponderada de todos os valores de $d(\kappa,\tau)$:

$$\text{ESII} = \frac{1}{Q}\sum_{\tau=1}^{Q}\sum_{\kappa=1}^{K}\gamma_k d(\kappa,\tau), \tag{2.41}$$

onde Q e K o número total de quadros e bandas críticas de frequência, respectivamente.

2.3.3 ASII_{ST}

Assim como o ESII, o ASII_{ST} (HENDRIKS et al., 2015) também foi uma medida proposta para adaptação de tempo curto da métrica clássica SII e estende os trabalhos desenvolvidos no ASII (TAAL; JENSEN; LEIJON, 2013) podendo levar também em consideração o efeito da reverberação. Alternativamente ao ESII, a medida ASII_{ST} propõe obter as funções $d(\kappa,\tau)$ da seguinte forma:

$$d(\kappa,\tau) = \frac{\xi(\kappa,\tau)}{\xi(\kappa,\tau)+1}, \tag{2.42}$$

sendo a média das funções $d(\kappa,\tau)$ calculadas analogamente à Equação 2.41. Os valores computados para o ASII_{ST} são baseados nas mesmas funções γ_k adotadas no ESII.

2.4 Resumo

Neste Capítulo, foram apresentados os métodos presentes na literatura que visam realçar sinais de voz e *chirp* na presença de ruídos acústicos subaquáticos. Estes métodos

podem ser classificados como espectrais (OMLSA e UMMSE) ou temporais (EMDH e NNESE). As soluções espectrais empregam métodos de estimação visando atualizar o espectro de potência do ruído. Por sua vez, o NNESE implementa um estimador robusto de desvio padrão para estimar o desvio padrão do ruído. O método EMDH é baseado na análise tempo-frequência e utiliza a decomposição empírica de modos para detecção e estimação das componentes ruidosas em cada IMF. Por fim, foram apresentadas as medidas objetivas de predição, tanto de qualidade quanto de inteligibilidade, com o intuito de avaliar o desempenho dos métodos de realce. Para a aferição da qualidade, foram adotadas as medidas PESQ, PEAQ, SegSNR e RMSE. Para a avaliação da inteligibilidade, foram implementadas as medidas STOI, ESII e ASII_{ST}.

3 MÉTODO DE REALCE DE SINAIS: PROPOSTA

As soluções de realce de sinais são essenciais para atenuar os efeitos das interferências acústicas causadas pelos ruídos no meio submarino. Algumas das principais técnicas de realce de sinais competitivas que atuam tanto no domínio da frequência, quanto no domínio do tempo, bem como as medidas objetivas de predição adotadas para avaliar o desempenho destas técnicas foram descritas no Capítulo 2.

Neste presente Capítulo, uma proposta de realce de sinais acústicos no domínio do tempo é introduzida, com o propósito de aprimorar os sinais de voz e *chirp* corrompidos por ruídos acústicos subaquáticos.

3.1 Método de Realce Proposto

Esta seção descreve o método de realce de sinais acústicos proposto nesta Dissertação. Inicialmente, o sinal degradado é decomposto em IMF via uma técnica chamada EEMD-IF (*Ensemble EMD - Iterative Filtering*) (WU; HUANG, 2009),(LIN; WANG; ZHOU, 2009). A estimação das componentes do ruído é obtida quadro a quadro, em cada IMF, e é baseada no índice de não-estacionariedade (INS) (BORGNAT et al., 2010) e, que é uma medida objetiva que mensura o grau de não-estacionariedade dos sinais. As IMF mais corrompidas pelos ruídos podem ser identificadas uma vez que há uma relação direta entre a razão sinal-ruído e o INS do sinal corrompido. De forma geral, a solução proposta é dividida nas seguintes etapas:

- decomposição tempo-frequência do sinal corrompido em modos de oscilação (IMF);
- seleção das IMF para atuação do algoritmo; e
- detecção e estimação das componentes do ruído e reconstrução do sinal.

3.1.1 EEMD-IF

A técnica adotada neste trabalho para decomposição do sinal corrompido em modos de oscilação foi o EEMD-IF, que implementa o algoritmo EEMD proposto em (WU; HUANG, 2009) para a decomposição do sinal, e a filtragem iterativa proposta em (LIN; WANG; ZHOU, 2009) para a interpolação das envoltórias máxima e mínima durante os processos de *sifting*. O EEMD é baseado na decomposição empírica de modos (EMD), proposta em (HUANG et al., 1998) para análises de sinais não-estacionários oriundos de sistemas não-lineares. O resultado desta decomposição é um conjunto de funções intrínsecas de modo (IMF), que são inteiramente dependentes do sinal analisado, e um resíduo.

Dado um sinal $y(t)$, a decomposição empírica de modos segue os seguintes passos (HUANG et al., 1998), (FLANDRIN; RILLING; GONCALVES, 2004):

a Identificar todos os extremos de $y(t)$, sejam os pontos de máximo $y_{max}(t)$ ou mínimo $y_{min}(t)$ locais;

b Obter as envoltórias $e_{max}(t)$ e $e_{min}(t)$ pela interpolação dos pontos máximo e mínimo, respectivamente. Neste caso, adota-se a interpolação polinomial com *spline* cúbico;

c Calcular o resíduo, definido como a média entre as envoltórias, ou seja, $r(t) = \frac{e_{max}(t)+e_{min}(t)}{2}$;

d Extrair os componentes de detalhes $d(t) = y(t) - r(t)$;

e Repetir a iteração sobre o resíduo $r(t)$.

Em (HUANG et al., 1998) as seguintes propriedades foram definidas para a IMF:

- o número de extremos $y_{max}(t)$ ou mínimo $y_{min}(t)$ e de cruzamentos em zero devem ser iguais ou se diferenciar em uma unidade; e
- o valor médio definido pelas envoltórias $\frac{e_{max}(t)+e_{min}(t)}{2}$ deve ser nulo.

Os procedimentos citados em (a-d) são repetidos toda vez que o componente de detalhes $d(t)$ não apresenta as propriedades mencionadas acima, de acordo com algum critério de parada a ser definido no algoritmo. Este processo iterativo é denominado *sifting* e visa garantir que $d(t)$ atenda aos requisitos de uma IMF. O número de máximos e mínimos locais diminuem com o aumento do índice da IMF, o que significa que as primeiras IMF possuem oscilações mais rápidas (altas frequências), ao passo que as últimas frequências possuem oscilações mais lentas (baixas frequências).

Após um número finito de iterações, o sinal pode ser definido como

$$y(t) = \sum_{m=1}^{M} d_m(t) + r(t), \tag{3.1}$$

em que $d_m(t)$, $1 \leq m \leq M$, são os componentes de detalhes obtidos no passo (d) de cada iteração, r(t) o sinal residual obtido na última iteração.

O EEMD foi descrito na literatura para solucionar a mistura de modos (*mode mixing*) observada no EMD. A mistura de modos ocorre quando uma mesma IMF contém oscilações que se diferem significativamente uma das outras, ou quando oscilações de mesmo grau estão decompostas em IMF distintas.

Para resolver esta limitação, o EEMD adota diferentes sequências amostrais de ruído gaussiano branco aditivo para corromper o sinal de entrada. Esta solução reside

no fato de que o ruído branco adicionado é distribuído uniformemente em todo o espaço tempo-frequência do sinal, e uma vez que as amostras deste ruído são descorrelatadas, as IMF de mesmo índice obtidas pela média das diferentes sequências corrompidas por ruído branco também serão descorrelatadas. Dessa forma, no resultado final desse processo as amostras referentes ao ruído branco acabam se cancelando, enquanto o sinal de entrada original permanece preservado.

Primeiramente, o EEMD obtém múltiplas versões do sinal corrompido (*ensembles*) pela adição de amostras distintas de ruído gaussiano branco, ou seja,

$$x^i(t) = x(t) + w^i(t); i = 1, \ldots, I, \tag{3.2}$$

sendo $x^i(t)$ a i-ésima versão do sinal $x(t)$ e $w^i(t)$ o ruído branco gaussiano para a versão i do sinal.

Na etapa seguinte, cada versão do sinal corrompido é decomposta via EMD em K modos de oscilação $IMF_k^i(t)$, sendo $k = 1, \ldots, K$ o índice de cada modo. Por fim, a média destas decomposições é calculada para cada IMF correspondente,

$$\overline{IMF}_k(t) = \frac{1}{I}\sum_{i=1}^{I} IMF_k^i(t); k = 1, \ldots, K, \tag{3.3}$$

resultando na decomposição definitiva do algoritmo EEMD.

No EEMD-IF, algoritmo de filtragem iterativa é proposto para substituir a interpolação por *spline* cúbica nas envoltórias máxima e mínima realizada recursivamente no processo de *sifting*, em virtude de algumas limitações deste processo. Uma limitação está associada à natureza altamente adaptativa do *sifting* que faz com que o processo seja instável, no sentido de que uma pequena mudança no sinal pode acarretar em diferentes decomposições. A outra limitação é a ausência de fundamentos matemáticos que permitam garantir, por exemplo, a convergência deste processo recursivo e a ortogonalidade das IMF. Além disso, destaca-se o menor custo computacional do EEMD-IF comparado ao EEMD.

Alternativamente, a filtragem iterativa propõe substituir a média dos envelopes obtidos nas interpolações por *splines* por uma média móvel dependente do sinal, ou seja,

$$S(X) = X - \frac{1}{2}(e_{max} + e_{min}) \tag{3.4}$$

é substituído por

$$\mathcal{T}(X) = X - \zeta(X), \tag{3.5}$$

sendo X o sinal a ser decomposto, e_{max} e e_{min} os envelopes máximo e mínimo, respectivamente, e $\zeta(X)$ uma função que representa a média móvel de X. Em (LIN; WANG; ZHOU, 2009), o $\zeta(X)$ adotado é uma média ponderada adaptativa local, por ser de fácil implementação e garantir a convergência na iteração.

Assim, sendo $X(n)$ o sinal em tempo discreto, tem-se a média móvel

$$\zeta(X) = \sum_{j=-m}^{m} a_j(n)X(n+j) \tag{3.6}$$

sendo $(a_j)_{j=-m}^{m}$ a máscara para $\zeta(X)$ em n. Para garantir a convergência no *sifting*, foi adotada uma máscara uniforme cujos coeficientes do filtro obedecem à relação $a_m = (a_j)_{j=-m}^{m}$, sendo $a_j = \dfrac{m+1-j}{(m+1)^2}$.

O tamanho da janela m pode ser calculado conforme a equação a seguir:

$$m = \frac{\alpha N}{k}, \tag{3.7}$$

onde N é o número de amostras do sinal, k o número de máximos e mínimos locais no sinal e α é um fator de ajuste.

A figura 4 ilustra seis IMFs obtidas pelas decomposições EEMD (à esquerda) e EEMD-IF (à direita) para um mesmo sinal de voz corrompido por ruído Terremoto Submarino à uma SNR de 0 dB.

A Tabela 1 apresenta um comparativo do custo computacional considerando os métodos EMD, EEMD, EMD-IF e EEMD-IF, com os resultados normalizados em relação ao algoritmo EMD-IF, que apresenta o menor tempo de processamento. Para esta comparação, foram adotados 8 IMF, 600 *siftings*, 50 *ensembles* e $\alpha = 2,5$. Observa-se que tanto o EMD-IF, quanto o EEMD-IF apresentam um custo computacional menor quando comparados aos métodos EMD e EEMD, respectivamente.

Tabela 1 – Custo computacional (tempo médio de processamento normalizado) entre os métodos EMD-IF, EMD, EEMD-IF e EEMD

Método	**Custo Computacional**
EMD-IF	1,0
EMD	3,3
EEMD-IF	39,2
EEMD	127,6

3.1.2 Seleção das IMF

Na solução de realce de sinais acústicos subaquáticos apresentada nesta Dissertação, um critério baseado no conceito do índice de não-estacionariedade é implementado para selecionar as IMF mais corrompidas pelas componentes do ruído. Após esta seleção, as componentes ruidosas nos quadros de cada umas destas IMF são detectadas quadro a quadro com o objetivo de identificar as variações nas características temporais e espectrais ao longo do tempo. Dessa forma, a solução permite aprimorar sinais que são distorcidos por efeitos de ruídos não-estacionários.

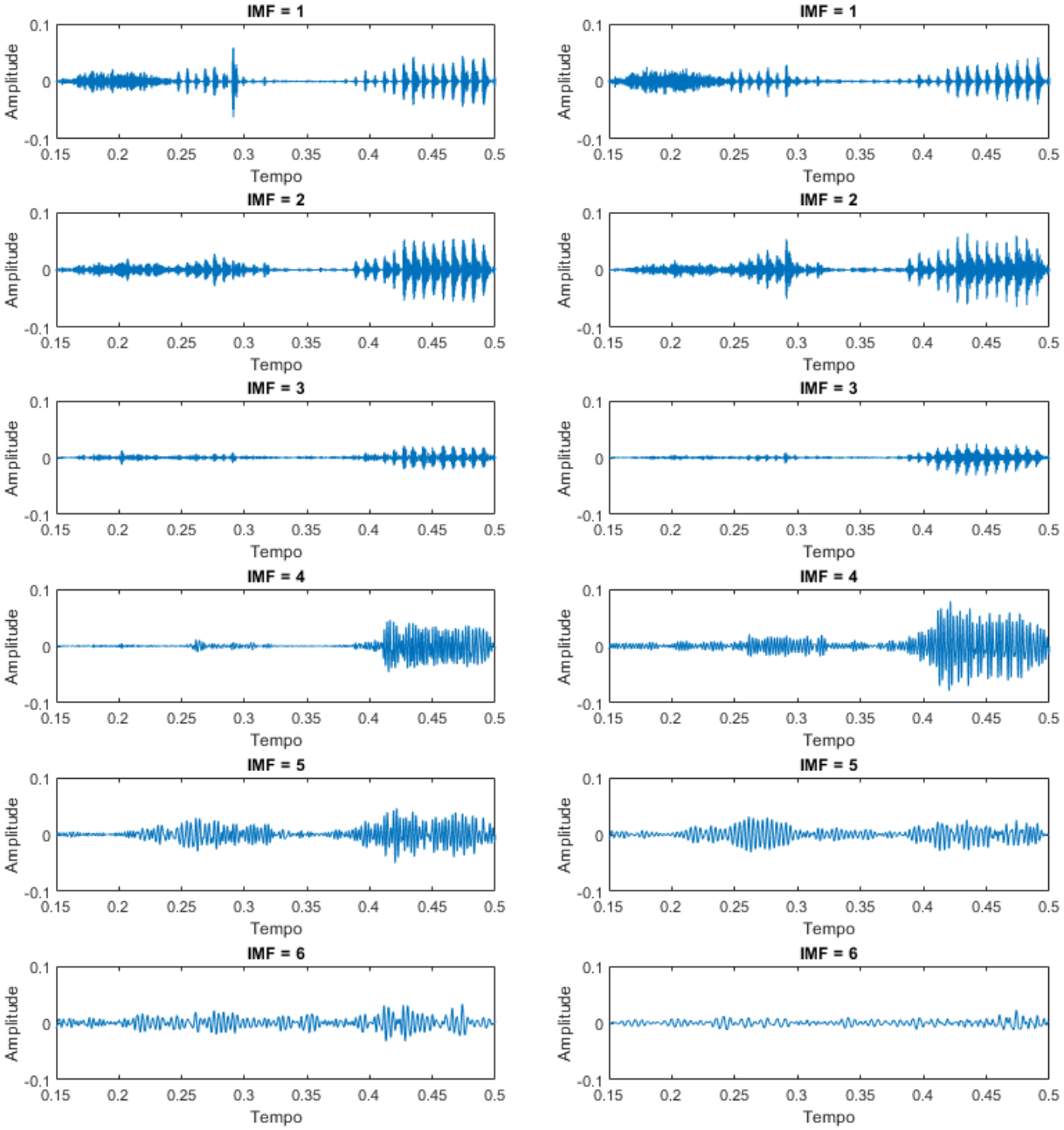

Figura 4 – Comparação das 6 primeiras IMF de um sinal de voz corrompido por ruído Terremoto Submarino a 0 dB decomposto por EEMD (à esquerda) e EEMD-IF (à direita).

Nesta subseção, são apresentados o conceito do INS e a definição do critério de seleção das IMF para atuação nos curtos quadros de tempo nas decomposições.

- **Índice de não-estacionariedade**

O índice de não-estacionariedade (INS) foi o proposto em (BORGNAT et al., 2010) com objetivo de determinar, objetivamente, o grau de não-estacionariedade dos sinais e ruídos. Esta medida é calculada a partir da comparação entre as componentes espectrais

do sinal original e os seus referenciais estacionários (*surrogates*). Considerando $X[k]$ a DFT (*discrete Fourier transform*) do sinal a ser analisado $x(t)$, no domínio da frequência o sinal $X[k]$ pode ser representado em termos de sua magnitude $A[k]$ e fase $\phi[k]$, ou seja,

$$X[k] = A[k]\exp(i\phi[k]), \tag{3.8}$$

em que i é a unidade imaginária. Para a obtenção dos referenciais estacionários de $x(t)$, a fase do sinal sob análise $\phi[k]$ é substituída por uma sequência aleatória com amostras independentes e uniformemente distribuídas no intervalo $[-\pi, \pi]$, $\psi[k]$, ou seja,

$$\tilde{X}[k] = A[k]\exp(i\psi[k]), \tag{3.9}$$

Este procedimento é repetido com o intuito de gerar um conjunto de referenciais estacionários $\tilde{x}_j(t), j = 1, 2, \ldots, J$, sendo J a quantidade de referenciais estacionários a serem utilizados no cômputo do INS.

Na etapa seguinte, o sinal sob análise é comparado com os seus *surrogates*. Para este fim, adota-se a distância de Kullback-Leibler (D_{KL}) (BASSEVILLE, 1989) simétrica que compara a média dos espectrogramas com os próprios espectrogramas obtidos em cada um dos pontos. Esta comparação é feita através de $c_n^{(x)}$ e $c_n^{(\tilde{x}_j)}$, que são, respectivamente, as distâncias obtidas para o sinal analisado e o conjunto de distâncias obtidas de todos os referenciais estacionários:

$$\begin{cases} c_n^{(x)} := D_{KL}(S_{x,K}(t_n, .), \langle S_{x,K}(t_n, .)\rangle), n = 1, \ldots, N \\ c_n^{(\tilde{x}_j)} := D_{KL}(S_{x,K}(t_n, .), \langle S_{x,K}(t_n, .)\rangle), n = 1, \ldots, N, \end{cases} \tag{3.10}$$

sendo $S_{x,K}(t, f)$ e $S_{\tilde{x}_j,K}(t, f)$ espectrogramas obtidos com um janelamento multi-ortogonal (*multitaper*) do sinal analisado e os referenciais estacionários, respectivamente.

Por fim, o INS é definido como a razão entre a média das variâncias obtidas dos sinais referenciais ($\Theta_0(j)$) e a variância das distâncias observadas do sinal analisado (Θ_1), ou seja:

$$\text{INS} := \sqrt{\frac{\Theta_1}{\langle \Theta_0(j)\rangle_j}} \tag{3.11}$$

em que $\Theta_0(j)$ e Θ_1 são definidos por

$$\begin{cases} \Theta_0(j) = \text{Var}(c_n^{(\tilde{x}_j)})_{n=1,\ldots,N}, j = 1, \ldots, J. \\ \Theta_1 = \text{Var}(c_n^{(x)})_{n=1,\ldots,N} \end{cases} \tag{3.12}$$

O INS é 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 analisado (T). Um limiar do teste de estacionariedade γ é definido para cada

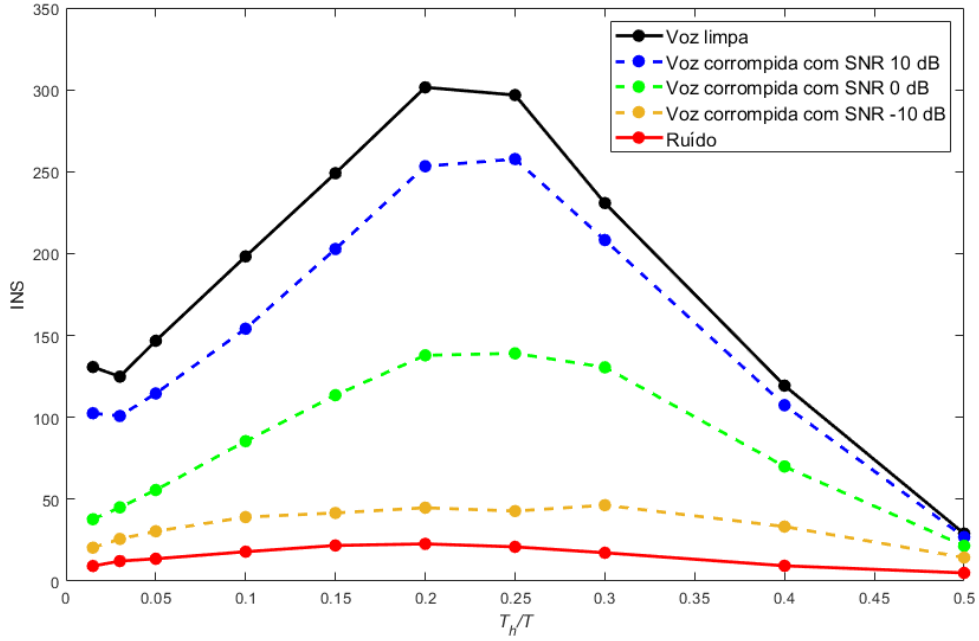

Figura 5 – Comparação do INS do sinal de voz limpo, INS do sinal de voz corrompido por ruído Terremoto Submarino com diferentes SNR e o INS do ruído Terremoto Submarino. O Terremoto Submarino é um ruído não-estacionário, comINS_{max} igual à 19.

valor de janela T_h, considerando uma precisão de 95%. Os ruídos são não-estacionários quando o INS é maior do que o limiar, ou seja,

$$INS \begin{cases} \leq \gamma, \text{ o sinal é estacionário;} \\ > \gamma, \text{ o sinal é não-estacionário;} \end{cases} \tag{3.13}$$

Os sinais de voz e *chirp* são sinais com alto índice de não-estacionariedade. As interferências causadas pelos ruídos acústicos presentes no ambiente, que apresentam um menor INS, alteram, significativamente, as características temporais e espectrais destes sinais, atenuando severamente o comportamento não-estacionário dos sinais de interesse. Nas Figuras 5 e 6, o sinal de voz, de INS máximo (INS_{max}) igual a 265, é corrompido por Terremoto Submarino (INS_{max} igual à 19) e Transatlântico (estacionário), respectivamente.

Pode-se observar que há uma relação entre a não-estacionariedade da voz corrompida e a energia do ruído presente neste sinal. Assim, quanto menor o SNR, menor o INS do sinal de voz ruidoso. Em segundo, também nota-se que o ruído Transatlântico, estacionário, provoca uma maior atenuação no comportamento não-estacionário do sinal de voz quando comparado ao Terremoto Submarino, que é não-estacionário.

Além disso, uma vez que os ruídos acústicos estão concentrados nas baixas frequências, e como as primeiras IMF representam as oscilações de maior frequência, também é esperado que o INS das primeiras IMF sejam superiores ao INS das últimas IMF. Isso

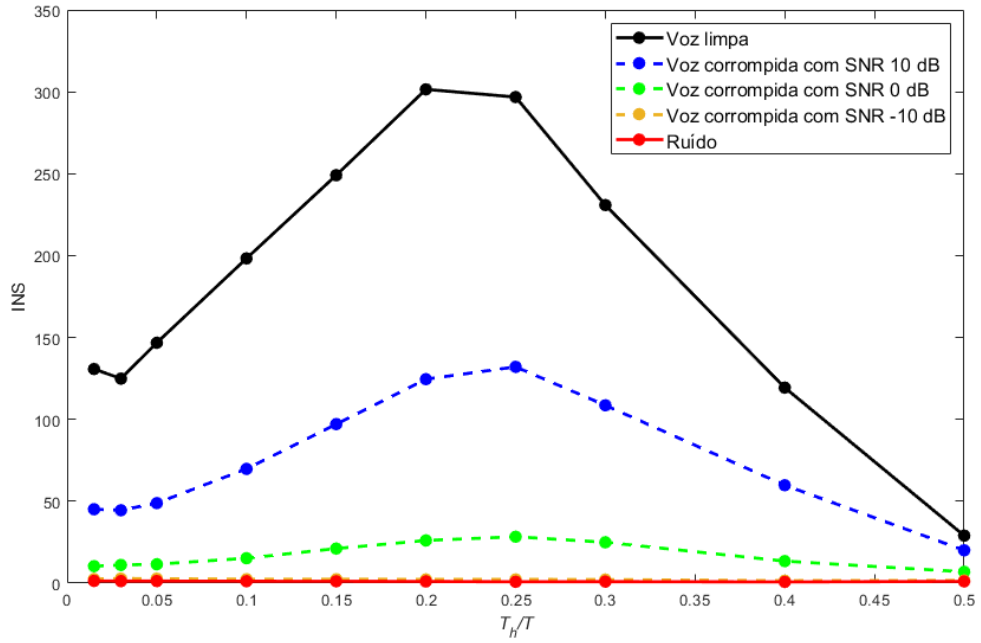

Figura 6 – Comparação do INS do sinal de voz limpo, INS do sinal de voz corrompido por ruído Transatlântico com diferentes SNR e o INS do ruído Transatlântico. O Transatlântico é um ruído estacionário.

ocorre pois a maior parte das componentes das primeiras IMF são pertencentes ao sinal de voz ou *chirp*. Portanto, estas IMF apresentam um comportamento mais não-estacionário. A Figura 7 ilustra os valores de INS para 8 IMF obtidas pelo algoritmo EEMD-IF aplicado em um sinal de voz ruidoso, sendo possível observar que a redução do INS do sinal corrompido acompanha o aumento do índice da IMF.

- **Critério de seleção das IMF**

Para a implementação deste critério, inicialmente é necessário diferenciar as IMF que concentram a maior parte da energia do sinal de interesse das IMF que são consideravelmente compostas por ruídos. Esta etapa é necessária para identificar o menor índice da IMF em que a solução de realce vai atuar no sinal decomposto. Esta diferenciação garante que o algoritmo não agrave a distorção nas primeiras IMF, compostas predominante por componentes de alta frequência. Para isso, foi definida uma variável $\theta_{1,i}$ que representa o logaritmo da razão entre o INS_{max} da IMF 1 e o INS_{max} da IMF i, sendo $i = 2, \ldots, N$ e N a última IMF, ou seja,

$$\theta_{1,i} = \frac{\log(\text{INS}max_1)}{\log(\text{INS}max_i)} \tag{3.14}$$

Como a IMF 1 representa oscilação de mais alta frequência do sinal, considera-se que estes modos apresentam o maior SNR. Sendo assim, $\theta_{1,i}$ representa a diferença na atenuação do comportamento não-estacionário observada entre a IMF 1 e a IMF i

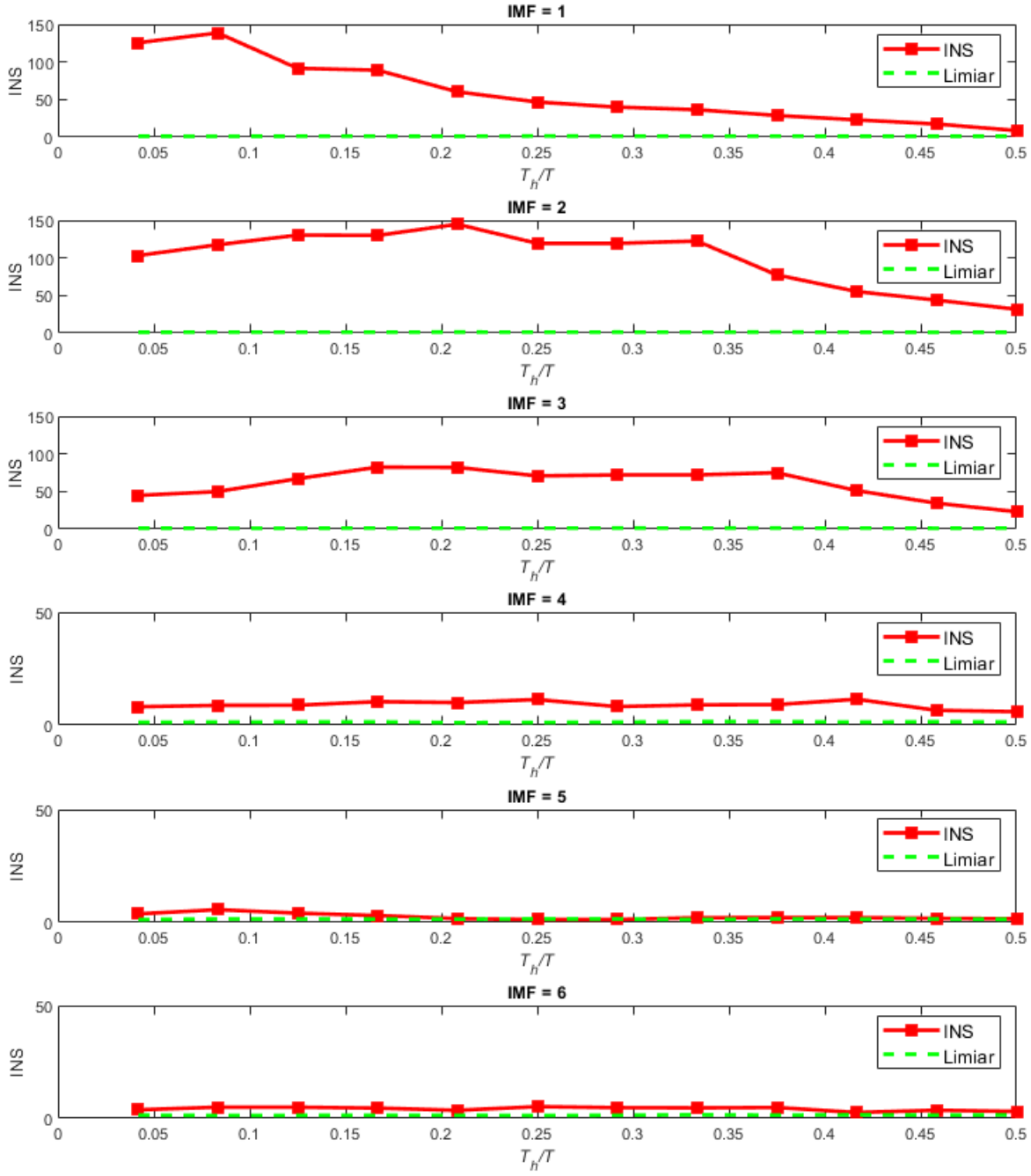

Figura 7 – Valores de INS para as IMF de um sinal de voz corrompido por ruído Terremoto Submarino com 0 dB e decomposto por 8 IMFs por EEMD-IF.

ou, em outras palavras, a diferença na quantidade de energia do ruído presente nestas decomposições.

As Figuras 8 e 9 ilustram os valores de $\theta_{1,i}$ do sinal limpo (linha contínua preta) e do sinal ruidoso (linha tracejada vermelha), para o sinal de voz e sinal *chirp* corrompidos por ruído Terremoto Submarino à 0 dB, respectivamente. Na figuras 8, observa-se que a distância entre o $\theta_{1,i}$ do sinal limpo e do sinal ruidoso aumenta significativamente a partir da IMF 4. Por conta disso, a solução proposta preservou, para realce de sinais de voz, pelo

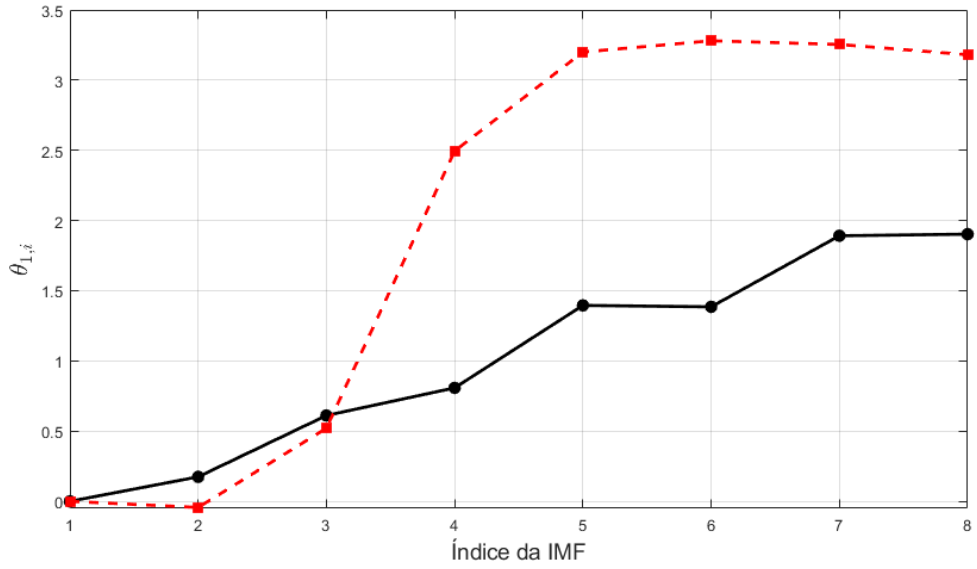

Figura 8 – Cálculo do $\theta_{1,i}$ para cada IMF de um sinal de voz corrompido por ruído Terremoto Submarino com 0 dB e decomposto por 8 IMF por EEMD-IF.

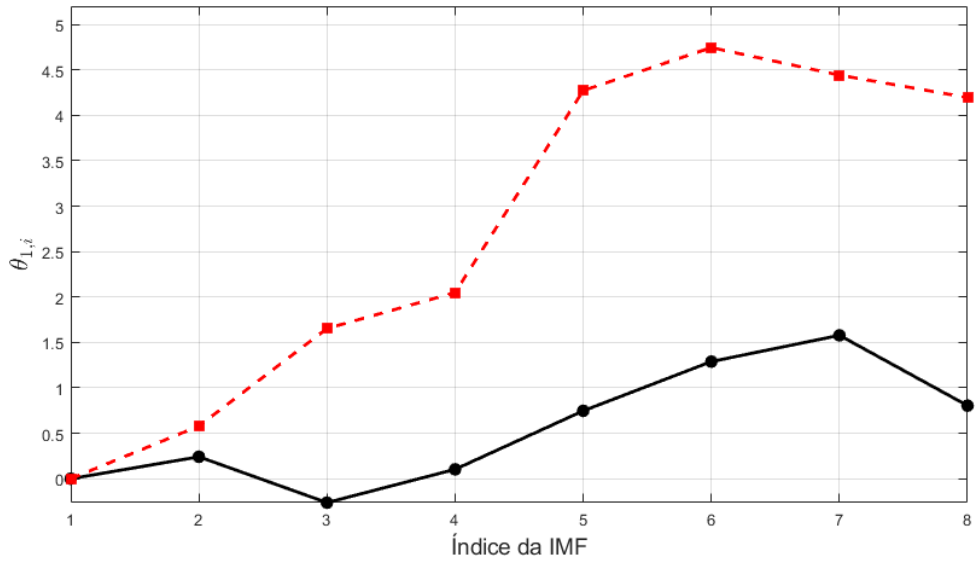

Figura 9 – Cálculo do $\theta_{1,i}$ para cada IMF de um sinal *chirp* corrompido por ruído Terremoto Submarino com 0 dB e decomposto por 8 IMF por EEMD-IF.

menos as três primeiras IMF na etapa de reconstrução do sinal, a fim de se evitar eventuais distorções no sinal realçado por conta da remoção dos componentes de alta frequência do sinal de interesse. Portanto, os valores de i na atuação do algoritmo são limitados a $i \geq 4$.

Por outro lado, na figura 9, observa-se que o $\theta_{1,i}$ do sinal limpo e ruidoso começam a se afastar a partir da IMF 3. No caso do realce dos sinais *chirp*, os valores de i na atuação do algoritmo estão restritos a $i \geq 3$.

3.1.3 Detecção e estimação das componentes ruidosas e reconstrução do sinal

Após a seleção das IMF, um critério é adotado para detectar e estimar as componentes mais afetadas pelos ruídos acústicos em segmentos curtos de tempo em cada IMF selecionada. A reconstrução do sinal realçado é feita somando-se os quadros remanescentes em cada modo de oscilação.

- **Critério de detecção e estimação**

Analogamente ao cálculo de $\theta_{1,i}$, para a atuação no quadro q calcula-se $\theta_{(1,i,q)}$, que é resultado da diferença do INS_{max} entre a IMF 1 e a IMF i em cada quadro,

$$\theta_{(1,i,q)} = \frac{\log(\text{INS}max_{(1,q)})}{\log(\text{INS}max_{(i,q)})}. \tag{3.15}$$

Dessa forma, os maiores $\theta_{(1,i,q)}$ representam aqueles que devem ser extraídos do sinal. Neste trabalho, foram adotados quadros de 20 ms, com 50% de sobreposição.

Seja $\delta_{(i)} = \log(\text{INS}max_i)$, para a determinação do limiar foram definidas as seguintes regras de decisão, para cada IMF i e quadro q:

$$\begin{cases} \text{se } \theta_{(1,i,q)} \leq \min\{\theta_{(1,i)}, \delta_{(i)}\}, \text{ a IMF } i \text{ do quadro } q \text{ é preservada; } e \\ \text{se } \theta_{(1,i,q)} > \min\{\theta_{(1,i)}, \delta_{(i)}\}, \text{ a IMF } i \text{ do quadro } q \text{ é removida.} \end{cases} \tag{3.16}$$

Nas Figuras 8 e 9, é possível notar que $\theta_{(1,i)}$ tende a aumentar com o aumento do índice da IMF. Por outro lado, como $\delta_{(i)}$ está associado diretamente ao INS_{max}, seu valor diminui para as IMF de maior índice. O limiar adotado na regra de decisão para a detecção das componentes ruidosas é definido como o valor mínimo entre esses dois parâmetros. Valores acima deste limiar são detectados como componentes do ruído, enquanto valores menores ou igual ao limiar são considerados como componentes do sinal de interesse, seja voz ou *chirp*.

A Figura 10 mostra uma comparação entre os INS_{max} dos sinal de voz limpa, ruidoso e ruído (Terremoto Submarino, à 0 dB) correspondentes à IMF 4 enquanto a figura 11 exibe $\theta_{(1,i,q)}$ do sinal de voz ruidosa e os limiares $\theta_{(1,i)}$ e $\delta_{(i)}$ referentes à esta mesma IMF, sendo possível verificar o funcionamento deste critério na detecção dos quadros mais corrompidos em uma IMF.

Na figura 11, pode-se observar que o quadro 44 (identificado pela seta preta) encontra-se acima de ambos os limiares, enquanto na figura 10, neste mesmo quadro, nota-se que o INS_{max} do sinal de voz ruidosa e do ruído estão praticamente iguais e distantes do INS_{max} do sinal de voz limpa. Isto significa que, neste quadro, há uma maior energia do ruído presente no sinal.

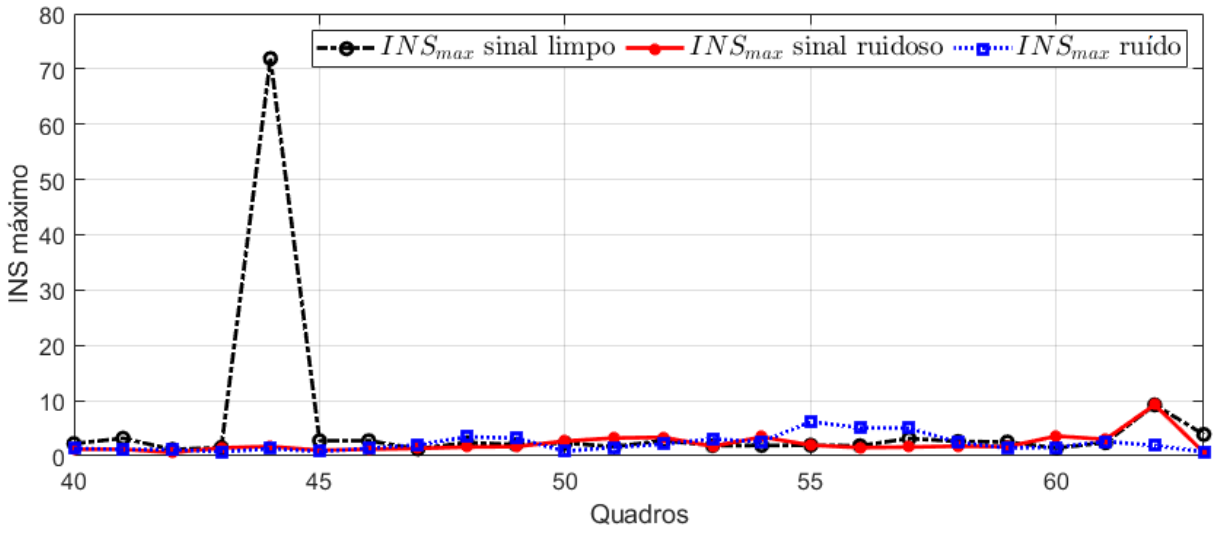

Figura 10 – INS_{max} de sinal de voz limpo, sinal de voz corrompido por ruído Terremoto Submarino à 0 dB e do ruído Terremoto Submarino, em cada quadro de 20 ms para a IMF 4.

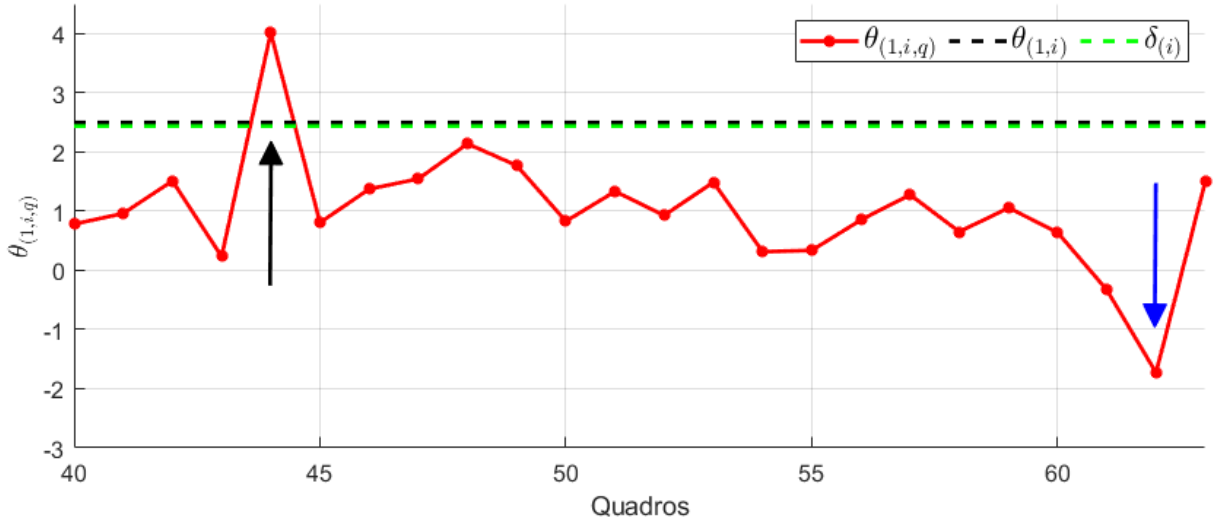

Figura 11 – Valores de $\theta_{1,i,q}$ do sinal de voz ruidoso e os limiares $\theta_{1,i}$ e δ_i referentes à IMF 4, de um sinal de voz corrompido por ruído Terremoto Submarino à 0 dB.

Por outro lado, nota-se que o quadro 62 (identificado pela seta azul na figura 11), o INS_{max} do sinal de voz ruidoso está próximo do INS_{max} do sinal limpo, e ao mesmo tempo distante do INS_{max} do ruído. Neste caso, a maior parte da energia presente no sinal pertence à voz e, por conta disso, o $\theta_{1,i,q}$ encontra-se bem abaixo do limiar.

- **Reconstrução do sinal**

Para a reconstrução do sinal realçado, cada quadro do sinal aprimorado $\hat{x}_q(t)$ é reconstruído somando-se as IMF janeladas (janelamento de *Hamming*) que não foram

excluídas na etapa anterior. Dessa forma, tem-se para $\hat{x}_q(t)$:

$$\hat{x}_q(t) = \sum_{i=1}^{N_q} \text{w-IMF}_{i,q}(t) \tag{3.17}$$

onde N_q é o número de IMF a serem somadas para cada quadro q e w-IMF$_{i,q}(t)$ é a i-ésima IMF janelada no quadro q.

Por fim, o sinal é obtido da mesma forma que em (ZÃO; COELHO; FLANDRIN, 2014), ou seja, somando-se todos os Q quadros do sinal realçado $\hat{x}_q$:

$$\hat{x}(t) = \sum_{q=0}^{Q-1} \hat{x}_q(t - qT_d). \tag{3.18}$$

3.2 Resumo

Neste Capítulo, apresentou-se uma proposta de realce de sinais acústicos de voz e *chirp* corrompidos por ruídos acústicos ambientais não-estacionários. Este método atua no domínio do tempo e utiliza, inicialmente, a decomposição tempo-frequência EEMD-IF para decompor o sinal em distintos modos de oscilação.

Para a detecção das componentes do sinal que são compostas predominantes por ruído, adotou-se o índice de não-estacionariedade (INS), que é uma medida utilizada para avaliar o grau de não-estacionariedade dos sinais de forma objetiva. Como os sinais de interesse apresentam um alto índice de não-estacionariedade comparado aos ruídos, as distorções causadas pelos ruídos podem ser identificadas a partir da atenuação do comportamento não-estacionário do sinal. A partir desta característica, um parâmetro foi definido para mensurar a diferença do INS máximo entre a primeira e as demais IMF, uma vez que a primeira IMF representa as oscilações de mais alta frequência e os ruídos estão concentrados nas baixas frequências.

O critério foi examinado em quadros de 20 ms, com o propósito de aprimorar os sinais corrompidos por ruídos acústicos ambientais não-estacionários. Um limiar baseado no INS máximo referente à IMF inteira é proposto com o intuito de detectar e estimar os quadros das IMF mais corrompidos por ruído. Finalmente, a reconstrução do sinal é feita excluindo-se estes quadros das IMF estimados, aplicando-se um janelamento nas IMF restantes e somando-se todos os quadros.

4 RESULTADOS DAS MEDIDAS OBJETIVAS DE PREDIÇÃO

Neste Capítulo, são apresentados os resultados de predição objetiva de qualidade e inteligibilidade, obtidos para os métodos de realce de sinais investigados neste estudo, sendo duas soluções espectrais (UMMSE e OMLSA) e três temporais (PRO, NNESE e EMDH). Para examinar o desempenho obtido pelas soluções de realce no aprimoramento de sinais acústicos de interesse, são realizados neste trabalho dois cenários experimentais distintos.

No primeiro cenário experimental, os sinais de interesse são sinais de voz, provenientes de diferentes locutores, homens e mulheres. As soluções de realce são examinadas por medidas de qualidade e de inteligibilidade. As medidas de qualidade adotadas são a PESQ (RIX et al., 2001) e PEAQ (COLOMES et al., 1999), enquanto as medidas de inteligibilidade utilizadas são a STOI (TAAL et al., 2011), ESII (RHEBERGEN; VERSFELD, 2005) e ASII_{ST} (TAAL; JENSEN; LEIJON, 2013).

No segundo cenário experimental, o sinal de interesse é um conjunto de sinais *chirp*. Estes sinais são bastante explorados nas comunicações acústicas subaquáticas em virtude da baixa sensibilidade ao efeito Doppler e boa capacidade de rejeição de interferências (HE et al., 2009). Neste cenário, os métodos são avaliados pelas medidas de qualidade SegSNR (HANSEN; PELLOM, 1998) e RMSE.

Em ambos os cenários, os sinais de interesse são corrompidos por ruídos acústicos ambientais subaquáticos com distintos valores de SNR. Os quatro ruídos adotados neste trabalho (Bolhas, Orca, Terremoto Submarino e Transatlântico) são provenientes de diferentes fontes acústicas subaquáticas, exibem distintos valores de INS, além de diferentes características espectrais. Os resultados de INS são apresentados, e tanto os sinais de interesse quanto os ruídos acústicos são classificados conforme seus graus de não-estacionariedade, de acordo com um critério adotado para o valor máximo de INS. O intuito desta análise é compreender o impacto na melhora da inteligibilidade e qualidade obtida pelos métodos de realce em cada cenário.

4.1 Descrição dos Experimentos de Realce de Sinais Acústicos

Nesta Seção, são descritos os cenários experimentais realizados neste trabalho para avaliação de desempenho dos métodos de realce. O primeiro cenário consiste em um subconjunto de 10 sinais de voz provenientes de 24 locutores, sendo 16 homens e 8 mulheres. Estes sinais são extraídos da base TIMIT (GAROFOLO et al., 1993), totalizando 240 locuções.

Cada sinal tem uma frequência de amostragem de 16 kHz e uma duração média de 3 segundos. Quatro ruídos acústicos ambientais e subaquáticos, Bolhas, Orca, Terremoto Submarino e Transatlântico, são adicionados aos sinais de voz, considerando-se os seguintes valores de SNR: -5 dB, 0 dB e 5 dB. A escolha destes ruídos se deu em função dos diferentes graus de não-estacionariedade e características espectrais. Além disso, estes ruídos são provenientes de diferentes fontes acústicas, sendo um ruído hidrodinâmico (Bolhas), biológico (Orca), sísmico (Terremoto Submarino) e antropogênico (Transatlântico).

O segundo cenário experimental consiste em um sinal contendo 10 sinais *chirp* lineares com um decaimento exponencial considerando perdas no sinal (OU; ALLEN; SYRMOS, 2011) e com duração total de 3,125 segundos, também com frequência de amostragem de 16 kHz. Para corromper o sinal *chirp*, foram utilizados os mesmos ruídos e SNR do primeiro cenário experimental.

A expressão para cada *chirp* é dada por:

$$s[x] = \sin(2\pi f_i x + \frac{\pi(f_f - f_i)x^2}{T}).e^{-x/\beta}, \tag{4.1}$$

sendo f_i, f_f, T e β, respectivamente, a frequência inicial, frequência final, duração do sinal *chirp* e a constante de decaimento exponencial. Neste trabalho, foi adotado $f_i = 100$ Hz, $f_f = 5$ kHz, $T = 312{,}5$ ms e $\beta = 0{,}067$.

A Figura 12 ilustra os espectrogramas de segmentos de 3 segundos de um sinal *chirp*, sinal de voz e dos ruídos subaquáticos adotados neste estudo. Nesta Figura, é possível notar que os ruídos acústicos estão concentrados nas baixas frequências, ao passo que os sinais de interesse contém componentes relevantes nas altas frequências. O ruído Bolhas foi obtido na base de dados da Freesound.org[1], os ruídos Orca e Terremoto Submarino foram extraídos da base de dados da *San Francisco Maritime National Park Association*[2], e o ruído Transatlântico foi obtido da base *ShipsEar* (SANTOS-DOMÍNGUEZ et al., 2016).

4.2 Resultados do Índice de Não-Estacionariedade

Nesta seção, são apresentados os resultados dos índices de não-estacionariedade tanto para os sinais de interesse (sinais de voz e *chirp*), quanto para os ruídos acústicos ambientais. Para classificar os sinais de interesse e os ruídos neste trabalho, os seguintes critérios foram adotados, baseados na comparação do INS máximo (INS_{max}) destes sinais com o limiar de estacionariedade (γ):

- $\text{INS}_{max} > 80\gamma$: altamente não-estacionário;
- $20\gamma < \text{INS}_{max} \leq 80\gamma$: não-estacionário;

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

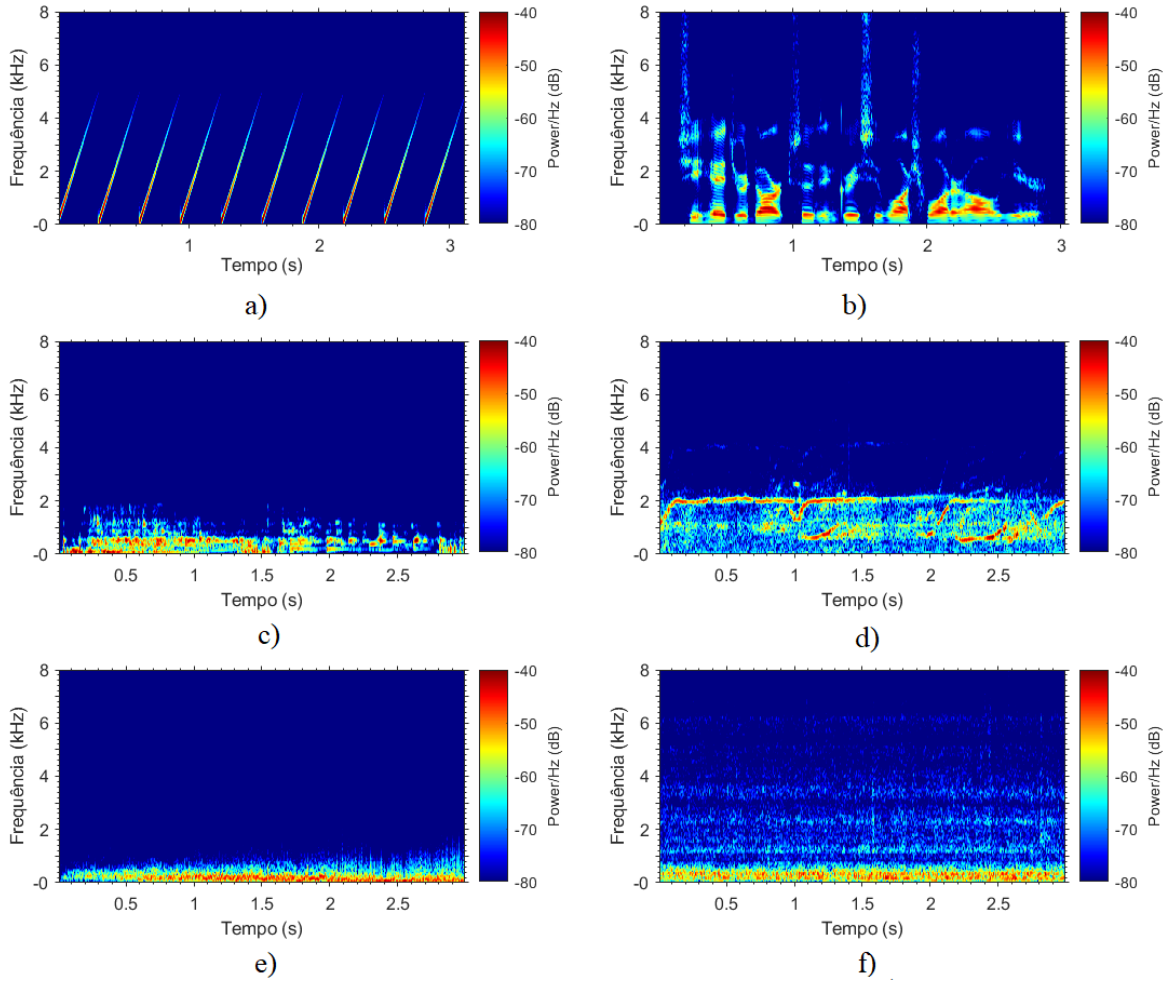

Figura 12 – Espectrogramas de segmentos de 3 segundos de duração dos sinais de interesse (a) *chirp* e (b) voz, e dos ruídos (c) Bolhas, (d) Orca, (e) Terremoto Submarino e (f) Transatlântico.

- $\gamma < \text{INS}_{max} \leq 20\gamma$: moderadamente não-estacionário; e
- $\text{INS}_{max} \leq \gamma$: estacionário.

No cômputo do INS, foram adotados 50 *surrogates* para trechos de 3 s do sinal de voz e dos ruídos sob análise, e 312 ms de um sinal *chirp*. Na Figura 13, são apresentados os valores do índice de não-estacionariedade para cada escala temporal T_h/T (linha vermelha contínua), e o limiar de estacionariedade para cada um dos sinais sob análise (linha verde tracejada).

O sinal de voz, o sinal *chirp* e os ruídos Bolhas, Orca e Terremoto Submarino apresentam valores de INS acima do limiar de estacionariedade e com distintos valores de INS_{max}, ou seja, estes sinais contém diferentes graus de não-estacionariedade. Para o ruído Transatlântico, observa-se que os valores de INS encontram-se predominantemente abaixo deste limiar, sendo assim classificado como estacionário.

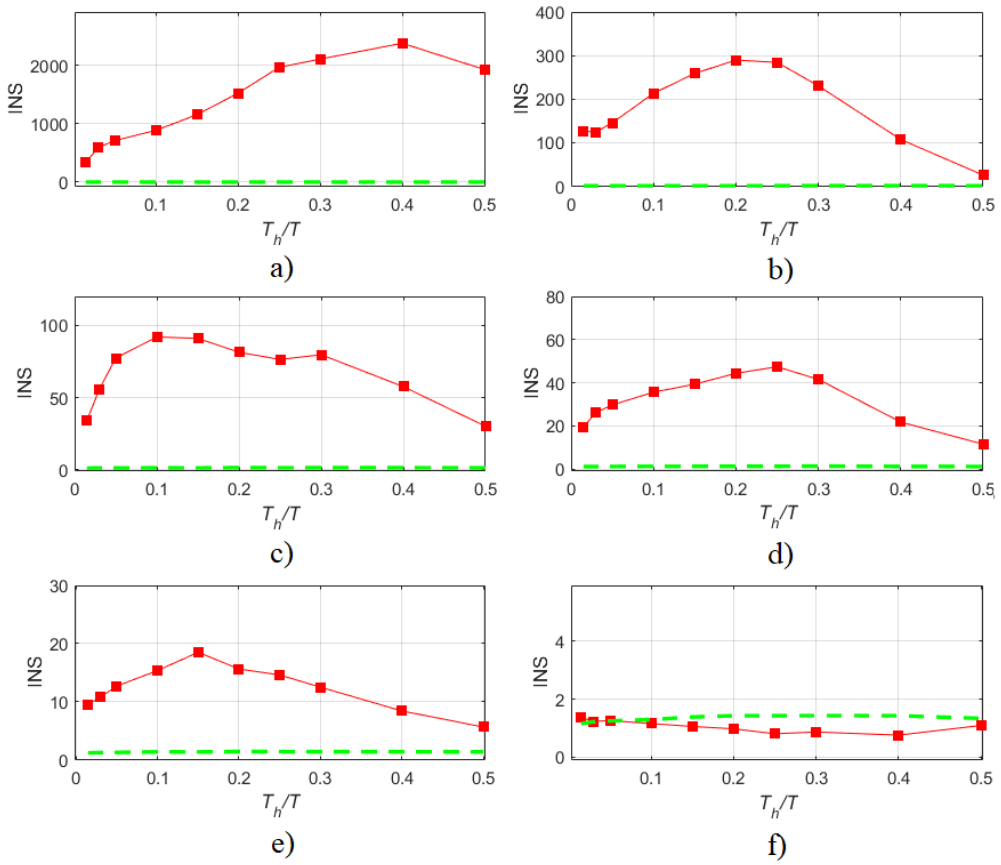

Figura 13 – Valores de INS obtidos de um sinal (a) *chirp* com 312,5 ms, e de segmentos de 3 s de um sinal de (b) voz, e dos ruídos (c) Bolhas, (d) Orca, (e) Terremoto Submarino e (f) Transatlântico. As linhas verdes tracejadas indicam os valores correspondentes do limiar γ para os testes de estacionariedade, enquanto as linhas vermelhas contínuas expõe os valores de INS calculados para cada escala de tempo T_h/T.

A Tabela 2 exibe o INS_{max} dos sinais de interesse e dos ruídos e as respectivas classificações destes sinais de acordo com este critério adotado.

Tabela 2 – Resultados de INS_{max} e classificação dos sinais acústicos quanto aos seus graus de não-estacionariedade

Sinal de Interesse	**INS_{max}**	**Classificação**
Voz	290	altamente não-estacionário
Chirp	2384	altamente não-estacionário
Ruído	**INS_{max}**	**Classificação**
Bolhas	92	altamente não-estacionário
Orca	47	não-estacionário
Terremoto Submarino	19	moderadamente não-estacionário
Transatlântico	abaixo do limiar de estacionariedade	estacionário

Tabela 3 – Resultados de PESQ para sinais de voz corrompidos por diferentes ruídos e SNR

RUÍDOS	SNR	NP	PRO	NNESE	EMDH	UMMSE	OMLSA
BOLHAS $\text{INS}_{max} = 95$	-5 dB	2,60	**2,89**	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	**3,25**	**3,25**	3,14	3,14	3,12
MÉDIA		2,86	**3,08**	3,01	2,88	2,88	2,84
ORCA $\text{INS}_{max} = 47$	-5 dB	2,02	**2,50**	2,34	2,07	1,96	1,81
	0 dB	2,57	**2,85**	2,69	2,57	2,44	2,34
	5 dB	2,85	**3,06**	2,98	2,85	2,77	2,70
MÉDIA		2,48	**2,80**	2,67	2,50	2,39	2,28
TERREMOTO SUBMARINO $\text{INS}_{max} = 19$	-5 dB	2,60	**2,81**	2,72	2,61	2,76	2,49
	0 dB	2,93	**3,12**	3,04	2,94	3,11	2,90
	5 dB	3,27	3,30	3,38	3,28	**3,47**	3,29
MÉDIA		2,93	3,08	3,05	2,94	**3,11**	2,89
TRANSATLÂNTICO Estacionário	-5 dB	2,12	2,47	2,21	2,12	2,44	**2,56**
	0 dB	2,51	2,86	2,61	2,51	2,87	**2,98**
	5 dB	2,92	3,04	3,00	2,92	3,24	**3,33**
MÉDIA		2,52	2,79	2,61	2,52	2,85	**2,96**
MÉDIA GERAL		2,70	**2,94**	2,83	2,71	2,81	2,74

4.3 Cenário Experimental 1

Os resultados das medidas objetivas de qualidade (PESQ e PEAQ) e inteligibilidade (STOI, ESII e ASII_{ST}) são apresentados e discutidos para o Cenário Experimental 1, cujo propósito é avaliar o desempenho das soluções de realce para sinais de voz. Para o método proposto, a decomposição EEMD-IF adotou 50 *ensembles*, um limite máximo de 600 *siftings*, 8 IMFs e um fator de ajuste da filtragem iterativa α de 2,5. Para o cálculo do INS, tanto na IMF inteira quanto em cada quadro, foram adotados 5 *surrogates*.

4.3.1 Resultados de PESQ

Os resultados da predição de qualidade obtidos pela medida objetiva PESQ para os métodos de realce podem ser observados na Tabela 3. Os ruídos acústicos subaquáticos estão organizados de forma decrescente quanto aos máximos valores de INS. A sigla NP corresponde aos sinais não processados, ou seja, aqueles em que não foram aplicadas as soluções de realce. Os valores em destaque correspondem àqueles de maior valor dentre todas as soluções de realce.

O método PRO teve os maiores aprimoramentos de PESQ para os ruídos Bolhas e

Orca, com média de 3,08 e 2,80, representando ganhos de 7,7% e 12,9%, respectivamente. Além disso, os maiores ganhos são observados nos menores valores de SNR, com destaque para a variação de 2,02 para 2,50 obtida para o ruído Orca a -5 dB, o que corresponde a um ganho de 23,8%. Em seguida, as soluções de realce NNESE e EMDH, que também foram desenvolvidas para lidar com os ruídos de maior não-estacionariedade, apresentaram interessantes aprimoramentos de PESQ para estes ruídos.

Para o ruído Terremoto Submarino, moderadamente não-estacionário, o UMMSE apresentou a maior média de PESQ, com 3,11, seguido da PRO com 2,92 e o NNESE com 2,89. A técnica PRO obteve os incrementos mais significativos para as SNR de -5 dB e 0 dB, com PESQ de 2,81 e 3,12, o equivalente a um ganho de 8,1% e 6,5%, respectivamente.

Quanto ao ruído Transatlântico, estacionário, os algoritmos espectrais apresentaram os maiores aprimoramentos, uma vez que os estimadores IMCRA e UMMSE conseguem bons resultados na estimação do espectro de potência destes ruídos. O OMLSA obteve a média de 2,96, seguido do UMMSE com 2,85, um acréscimo de 0,44 e 0,33 respectivamente, no PESQ em comparação a média de 2.52 do NP. Por sua vez, a solução PRO exibiu a maior média entre os algoritmos temporais, com média de 2,61.

Por fim, considerando todos os ruídos, pode-se perceber que o método proposto apresenta a melhor predição de qualidade com valor médio geral de PESQ de 2,94, comparado à média de 2,70 obtida para o NP, o que representa um ganho de 8,9%. Em seguida, os métodos NNESE e UMMSE apresentam ganhos de 4,8% e 4,1%, respectivamente.

4.3.2 Resultados de PEAQ

A Tabela 4 exibe os resultados da medida de qualidade PEAQ para os métodos de realce. O PRO obteve o maior aprimoramento para o ruído Bolhas, de maior INS, com uma média de 1,66, uma variação de 16,9% com relação ao NP, de média 1,42. Para o ruído Orca, o NNESE apresentou o maior valor de PEAQ, com 1,28, seguido do PRO com 1,26, o que representa um ganho de 18,5% e 16,7%, em relação ao PEAQ de 1,08 do NP, respectivamente.

Por outro lado, o UMMSE obteve o aprimoramento mais interessante para o ruído Terremoto Submarino, com média de 1,94, seguido do NNESE, com 1,89, e PRO, com 1,82, em comparação à média de 1,62 do NP. Para o ruído Transatlântico, os algoritmos espectrais apresentaram os melhores resultados, com o OMLSA e UMMSE obtendo uma média de 1,80 e 1,46 respectivamente, considerando a média de 1,12 do NP. A solução PRO apresentou os maiores resultados dentre os algoritmos temporais, com média de 1,30, o que representa uma melhora na qualidade de 16,1% segundo o PEAQ.

Por fim, os resultados médios globais obtidos com a medida PEAQ reforçam que os maiores aprimoramentos na qualidade foram obtidos pelos métodos PRO e NNESE,

Tabela 4 – Resultados de PEAQ para diferentes ruídos e valores de SNR

RUÍDOS	SNR	NP	PRO	NNESE	EMDH	UMMSE	OMLSA
BOLHAS $\text{INS}_{max} = 95$	-5 dB	1,12	**1,28**	1,22	1,13	1,07	1,02
	0 dB	1,35	**1,63**	1,59	1,35	1,41	1,26
	5 dB	1,79	2,07	**2,12**	1,78	2,04	1,76
MÉDIA		1,42	**1,66**	1,64	1,41	1,37	1,24
ORCA $\text{INS}_{max} = 47$	-5 dB	0,94	**1,02**	0,99	0,97	0,97	0,96
	0 dB	1,04	**1,18**	1,17	1,04	1,09	1,10
	5 dB	1,28	1,59	**1,67**	1,28	1,44	1,43
MÉDIA		1,08	1,26	**1,28**	1,08	1,17	1,16
TERREMOTO SUBMARINO $\text{INS}_{max} = 19$	-5 dB	1,19	1,28	**1,36**	1,20	**1,36**	1,17
	0 dB	1,54	1,88	1,90	1,55	**1,99**	1,51
	5 dB	2,12	2,30	2,41	2,12	**2,48**	2,08
MÉDIA		1,62	1,82	1,89	1,60	**1,94**	1,59
TRANSATLÂNTICO Estacionário	-5 dB	0,94	1,01	0,97	0,94	1,04	**1,29**
	0 dB	1,06	1,25	1,14	1,05	1,32	**1,80**
	5 dB	1,36	1,64	1,54	1,36	2,02	**2,31**
MÉDIA		1,12	1,30	1,22	1,12	1,46	**1,80**
MÉDIA GERAL		1,31	1,51	1,51	1,30	**1,52**	1,47

ambos com média 1,51, e UMMSE, com 1,52, e comparado à média de 1,12 apresentada pelo NP.

4.3.3 Resultados STOI

Os resultados da predição de inteligibilidade obtidos pela medida objetiva STOI para as quatro condições de ruídos são expostos na figura 14. Para o ruído Bolhas, verifica-se que os métodos temporais, PRO, NNESE e EMDH, obtiveram valores de STOI superiores aos espectrais. No caso dos ruídos Orca e Transatlântico, os métodos temporais, juntamente com o UMMSE, apresentam valores de STOI superiores aos obtidos pelo OMLSA, sendo esta diferença mais nítida para SNR de -5 dB.

Para o ruído Terremoto Submarino, observa-se que os resultados de todas as soluções encontram-se muito próximos, fato que pode ser justificado pelos altos valores de STOI. Nota-se que, até mesmo para SNR de -5 dB, os valores de STOI estão próximos de 0,74, o que é um valor considerado elevado para esta medida.

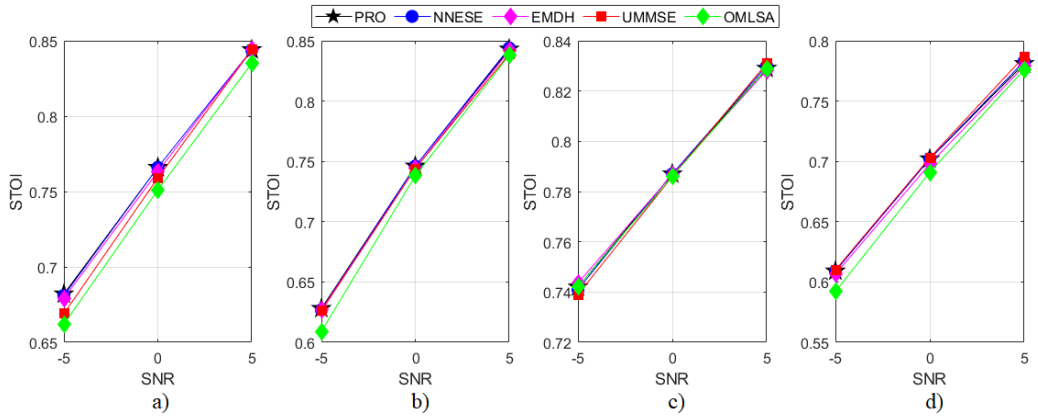

Figura 14 – Resultados de STOI das soluções de realce OMLSA, UMMSE, EMDH, NNESE e PRO para sinais de voz corrompidos pelos ruídos (a) Bolhas, (b) Orca, (c) Terremoto Submarino e (d) Transatlântico, considerando SNR de -5, 0 e 5 dB.

4.3.4 Resultados de ESII e ASII_{ST}

As figuras 15 e 16 ilustram os resultados das medidas de inteligibilidade ESII e ASII_{ST}, respectivamente. De modo geral, os algoritmos temporais apresentaram melhores resultados para os ruídos Bolhas e Orca, quando comparado aos espectrais.

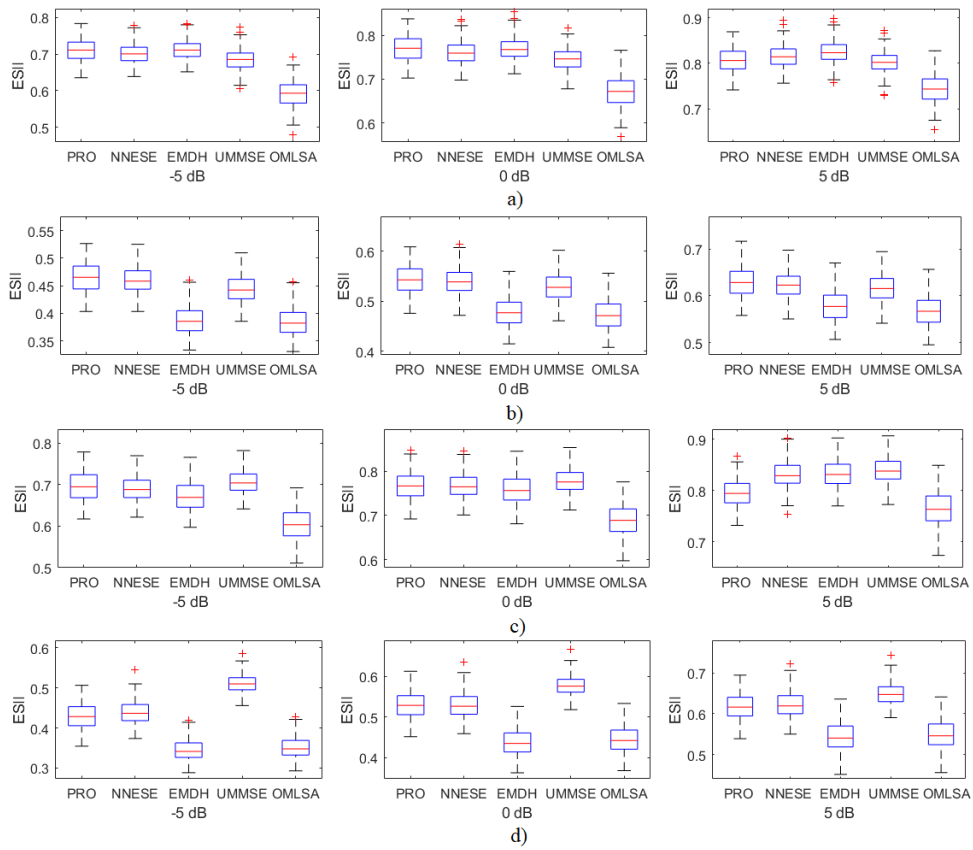

Figura 15 – Resultados de ESII das soluções de realce OMLSA, UMMSE, EMDH, NNESE e PRO para sinais de voz corrompidos pelos ruídos (a) Bolhas, (b) Orca, (c) Terremoto Submarino e (d) Transatlântico, considerando SNR de -5, 0 e 5 dB.

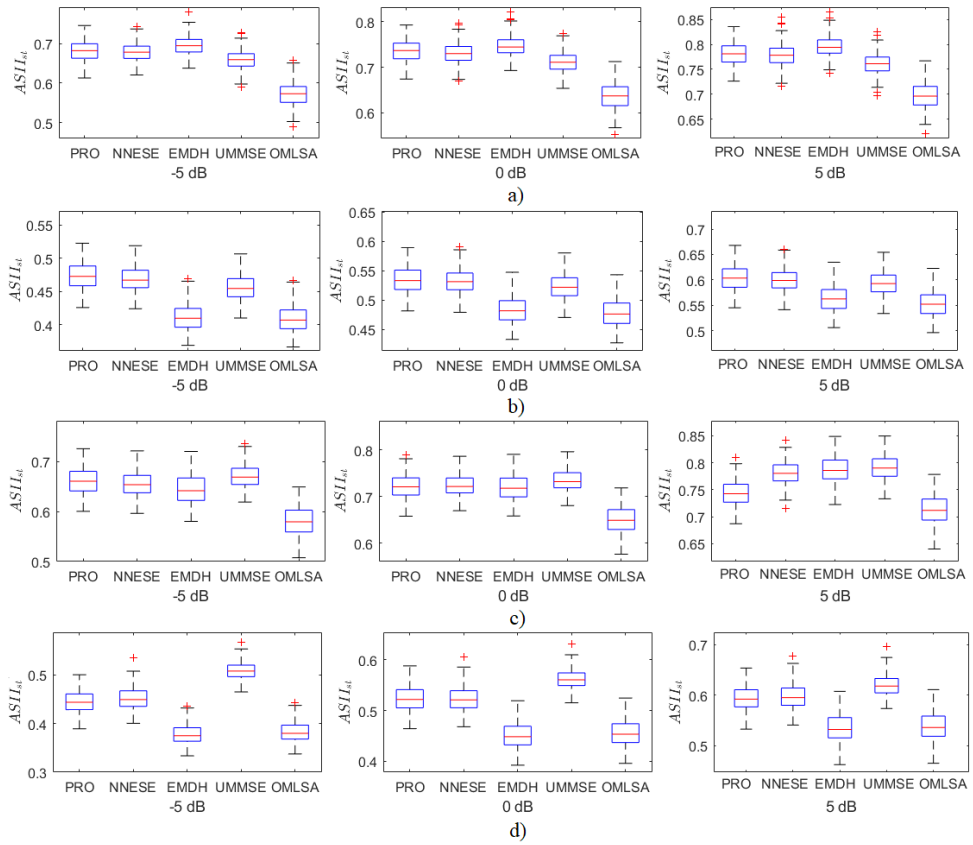

Figura 16 – Resultados de $ASII_{ST}$ das soluções de realce OMLSA, UMMSE, EMDH, NNESE e PRO para sinais de voz corrompidos pelos ruídos (a) Bolhas, (b) Orca, (c) Terremoto Submarino e (d) Transatlântico, considerando SNR de -5, 0 e 5 dB.

No ruído Bolhas, o EMDH apresentou um resultado levemente superior aos demais, seguido do PRO e do NNESE. No ruído Orca, o PRO obteve o melhor resultado em ambas as medidas, seguido do NNESE. No ruído Terremoto Submarino, os maiores ganhos de inteligibilidade observados para o ESII e $ASII_{ST}$ foram obtidos pelo UMMSE, seguido pelos demais algoritmos temporais. O PRO também apresentou aprimoramentos relevantes para os SNR de -5 dB e 0 dB deste ruído.

O UMMSE também obteve melhores resultados para o ruído Transatlântico, especialmente para os SNR de -5 dB e 0 dB, seguido do NNESE e PRO, que também obtiveram valores interessantes em ambas as medidas para o ruído estacionário.

4.4 Cenário Experimental 2

Nesta seção, são apresentados e discutidos os resultados das medidas objetivas de qualidade (SegSNR e RMSE) para o Cenário Experimental 2, cujo propósito é avaliar o desempenho das soluções de realce para sinais *chirp*. Para a SegSNR, os valores calculados representam os incrementos em dB alcançados para esta medida (ΔSegSNR) com as

técnicas de realce, para cada SNR.

Para o método PRO, foram adotados os mesmos parâmetros do Cenário Experimental 1, porém a atuação do critério de detecção das componentes ruidosas inicia-se a partir da IMF 3, ao invés da IMF 4. Para o EMDH, esta mesma alteração foi implementada em relação ao algoritmo original implementado em (ZÃO; COELHO; FLANDRIN, 2014). Além disso, o limiar do Expoente de Hurst foi alterado de 0,9 para 0,5, visando obter maiores incrementos do SegSNR para o sinal *chirp*.

4.4.1 Resultados de SegSNR e RMSE

Os resultados das medidas de qualidade SegSNR e RMSE podem ser notados nas Figuras 17 e 18, respectivamente. Primeiramente, observa-se que ambas as medidas apresentam resultados similares, ou seja, quanto maior o incremento obtido pela SNR segmental, menor o erro entre o sinal corrompido e o realçado.

O método PRO obteve ganhos interessantes de SegSNR, principalmente para os ruídos mais não-estacionários como Bolhas e Orca, obtendo resultados superiores aos apresentados pelas técnicas espectrais e do EMDH. No ruído Bolhas, para SNR de 5 dB, o PRO auferiu o maior aprimoramento com 1,58 dB. Nos demais casos, o NNESE apresentou maiores incrementos para estes ruídos, seguido do PRO. Para o ruído Orca, o NNESE também obteve os maiores aprimoramentos, seguido do PRO, em todos os SNR. Os ganhos mais relevantes foram observados no SNR de -5 dB, com o NNESE e PRO atingindo um ΔSegSNR de 4,0 dB e 1,7 dB, respectivamente.

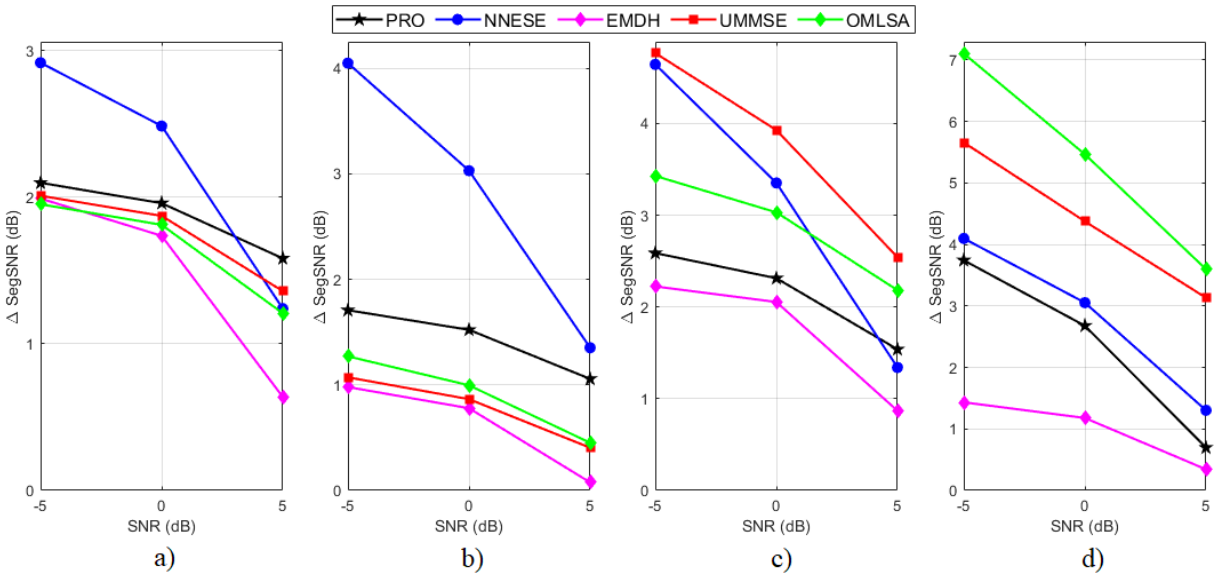

Figura 17 – Resultados de SegSNR das soluções de realce OMLSA, UMMSE, EMDH, NNESE e PRO para sinais chirp corrompidos pelos ruídos (a) Bolhas, (b) Orca, (c) Terremoto Submarino e (d) Transatlântico, considerando SNR de -5, 0 e 5 dB.

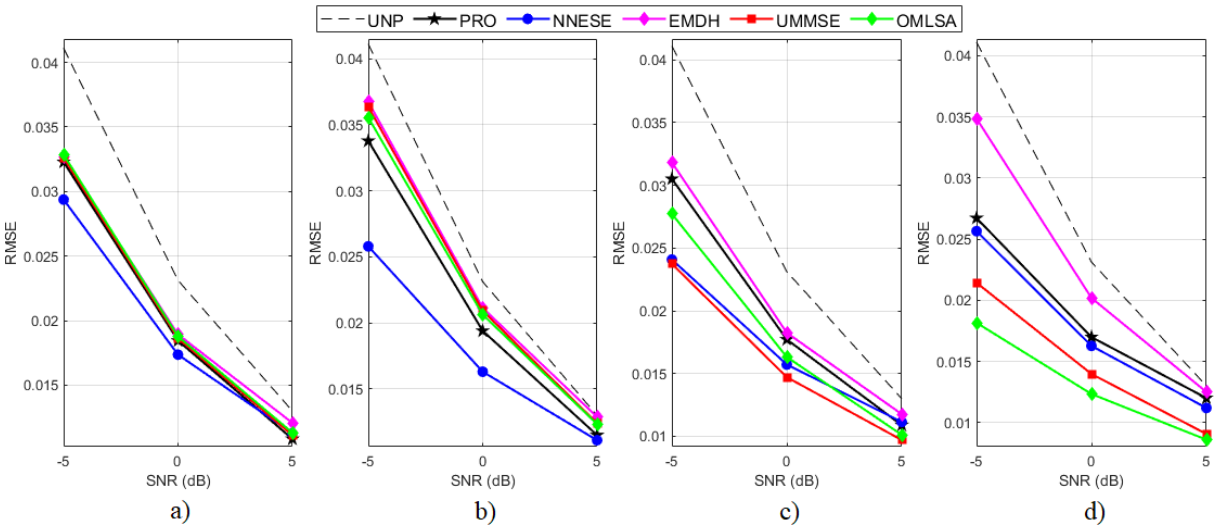

Figura 18 – Resultados de RMSE das soluções de realce OMLSA, UMMSE, EMDH, NNESE e PRO para sinais *chirp* corrompidos pelos ruídos (a) Bolhas, (b) Orca, (c) Terremoto Submarino e (d) Transatlântico, considerando SNR de -5, 0 e 5 dB.

Para o ruído Terremoto Submarino, os ganhos mais expressivos de SegSNR foram observados para o UMMSE. Para o SNR de 5 dB neste ruído, o método PRO obteve um ΔSegSNR de 1,54 dB, superior aos obtidos pelo NNESE e EMDH, de 1,34 dB e 0,87 dB, respectivamente. Finalmente, para o ruído Transatlântico, o OMLSA obteve os maiores aprimoramentos para todos os SNR, com destaque para o ΔSegSNR de 7,1 dB para SNR de -5 dB.

4.5 Discussão

No Cenário Experimental 1, a solução PRO apresentou o maior aprimoramento na medida de qualidade PESQ, com a maior média para os ruídos de maior INS, Bolhas e Terremoto, com ganhos na qualidade de 7,7% e 12,9%. O PRO também apresentou a melhor predição de qualidade no geral, com um ganho de 8,9%. Para o ruído Bolhas, altamente não-estacionário, o método PRO obteve o melhor aprimoramento na qualidade segundo a medida PEAQ, em comparação aos demais métodos competitivos, com um incremento de 16,9% na qualidade. O método PRO ainda apresentou uma das maiores médias gerais, juntamente com o NNESE e UMMSE.

Ainda para o realce de sinais de voz, considerando as medidas objetivas STOI, ESII e ASII_{ST}, a medida proposta apresentou resultados interessantes para o incremento da inteligibilidade, principalmente para os ruídos não-estacionários Bolhas e Orca. Para as medidas ESII e ASII_{ST}, o aprimoramento na inteligibilidade também foi interessante para o ruído Terremoto Submarino considerando o SNR de -5 e 0 dB. Neste ruído, o UMMSE apresentou os melhores resultados na predição de inteligibilidade.

No Cenário Experimental 2, o NNESE obteve os maiores valores de ΔSegSNR para os ruídos Orca e Bolhas, o que também equivale a um baixo RMSE quando comparado ao sinal limpo. Para estes mesmos ruídos, a solução PRO apresentou o segundo ganho mais significativo de ΔSegSNR. O UMMSE e OMLSA apresentaram incrementos mais relevantes para os ruídos Terremoto Submarino e Transatlântico, respectivamente.

4.6 Resumo

Neste Capítulo, foram apresentados dois experimentos realizados para avaliação do desempenho dos métodos de realce de sinais competitivos e do método proposto neste trabalho. O primeiro experimento consistiu em realçar sinais de voz e avaliar o aprimoramento obtido por estes métodos sob o aspecto da qualidade e da inteligibilidade. Para examinar a qualidade da voz, duas medidas objetivas foram adotadas, PESQ e PEAQ. Para julgar a inteligibilidade, foram adotadas três medidas objetivas, STOI, ESII e ASII_{ST}. No segundo experimento, o sinal acústico de interesse a ser aprimorado foi um conjunto de sinais *chirp*, sendo aqui adotadas as medidas de qualidade SegSNR e RMSE para avaliar o desempenho das técnicas de realce.

Em ambos os cenários experimentais, foram utilizadas cinco soluções de realce de sinais, sendo duas espectrais (UMMSE e OMLSA) e três temporais (PRO, NNESE e EMDH). Os métodos foram aplicados nos sinais acústicos de interesse corrompidos, com diferentes SNR, por quatro ruídos acústicos coletados de diferentes fontes reais pertencentes ao ambiente subaquático, com características temporais e espectrais distintas. Segundo os seus valores de INS máximo, um critério foi adotado para discriminar os ruídos da seguinte maneira: um ruídos altamente não-estacionário (Bolhas), um não-estacionário (Orca), um moderadamente não-estacionário (Terremoto Submarino), e o último estacionário (Transatlântico).

O método de realce PRO apresentou resultados relevantes no incremento da qualidade e inteligibilidade dos sinais de voz, além do aprimoramento da qualidade para sinais *chirp*, principalmente para os ruídos com maiores índices de não-estacionariedade.

5 CONCLUSÃO

Esta Dissertação de Mestrado expõe os desafios associados à redução das interferências acústicas causadas pelos ruídos ambientais do meio subaquático. Estas distorções afetam os sistemas acústicos de sonar e comunicações, principalmente quando estes ruídos variam suas características ao longo do tempo. Neste contexto, este estudo propõe uma solução de realce de sinais acústicos baseada na decomposição empírica de modos e no índice de não-estacionariedade, com o propósito de mitigar este efeitos e, assim, aprimorar a qualidade e inteligibilidade dos sinais de interesse mesmo na presença de ruídos não-estacionários.

Para avaliação deste método de realce, foram realizados dois experimentos, sendo o primeiro direcionado para sinais de voz, e o segundo voltado para sinais *chirp*. Em ambos os experimentos, os sinais de interesse foram corrompidos por quatro ruídos acústicos com diferentes índices de não-estacionariedade e provenientes de distintas fontes acústicas. Além disso, quatro soluções de realce propostas na literatura, sendo duas técnicas espectrais e duas temporais, foram implementadas para comparação com o método proposto.

Os resultados para o aprimoramento da qualidade dos sinais de voz confirmam o bom desempenho da solução proposta neste trabalho. O método PRO apresentou a maior média geral do valor de PESQ dentre todos os métodos de realce competitivos, com um ganho de 8,9% em relação aos sinais não processados. Para a medida PEAQ, esta solução apresentou uma das maiores médias gerais, com uma melhora de 15,6%, além do maior ganho na qualidade para os ruídos com maior INS. Para estes mesmos ruídos, esta solução apresentou ganhos na inteligibilidade superiores aos obtidos pelos algoritmos espectrais, de acordo com as medidas STOI, ESII e ASII_{ST}. Para o realce de sinais *chirp*, o método proposto obteve relevantes aprimoramentos no SegSNR, especialmente para os ruídos com elevados graus de não-estacionariedade.

Os principais resultados deste trabalho podem ser resumidos da seguinte maneira:

- Proposta de um método de realce de sinais acústicos corrompidos por ruídos subaquáticos com diferentes características temporais e espectrais, que utiliza a decomposição empírica de modos do sinal juntamente com um critério baseado em INS para detecção e estimação das componentes ruidosas.
- O método proposto aprimorou a qualidade e inteligibilidade dos sinais de voz, de acordo com as medidas objetivas de predição. Os resultados apresentados por esta solução, em comparação às demais técnicas de realce adotadas como referência, foram particularmente relevantes para os ruídos com maior grau de não-estacionariedade.

- Além disso, a solução apresentada também apresentou incrementos interessantes na qualidade do sinal *chirp* para os ruídos com maiores valores de INS.

5.1 Sugestões para Trabalhos Futuros

Nesta Seção são destacadas algumas sugestões para trabalhos futuros:

- Investigar o uso de novas variações da decomposição empírica de modos que possibilitem um menor custo computacional, juntamente com o critério de detecção e estimação adotado neste trabalho.
- Examinar a aplicação do critério baseado no INS em máscaras acústicas para aprimoramento da inteligibilidade dos sinais de voz.
- Investigar a adaptação do método de realce proposto para treinamento de soluções de aprendizado por máquinas (*machine learning*) (CHOI; CHOO; LEE, 2019) e estocásticos (SIDDAGANGAIAH et al., 2015).
- Analisar a inclusão do efeito da reverberação e examinar as predições objetivas de qualidade do método proposto neste trabalho.
- Examinar a classificação de fontes acústicas de ruídos ambientais subaquáticos para aprimoramento do método proposto.
- Verificar a aplicação dos métodos propostos e avaliados para localização de fontes acústicas em ambiente subaquático.

5.2 Comentários Finais

Nesta Dissertação, é apresentada uma nova proposta de realce de sinais acústicos na presença de ruídos ambientais subaquáticos. Esta solução emprega o EEMD-IF para a decomposição do sinal em modos de oscilação e um critério baseado no INS para detecção e estimação dos modos mais corrompidos pelos ruídos em segmentos curtos de tempo. Os cenários experimentais abrangendo o realce de sinais de voz e *chirp* mostraram que o método proposto apresentou resultados promissores no aprimoramento da qualidade e inteligibilidade, principalmente para ruídos com maior grau de não-estacionariedade.

REFERÊNCIAS

AL-ABOOSI, Y. Y.; SHA'AMERI, A. Z. 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*, v. 2, n. 3, p. 172–185, 2017.

BASSEVILLE, M. Distance measures for signal processing and pattern recognition. *Signal Processing*, v. 18, n. 4, p. 349–369, 1989.

BOLL, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 27, n. 2, p. 113–120, 1979.

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.

CHATLANI, N.; SORAGHAN, J. J. EMD-based filtering (EMDF) of low-frequency noise for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 20, p. 1158–1166, 2012.

CHOI, J.; CHOO, Y.; LEE, K. Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning. *Sensors (Basel, Switzerland)*, v. 19, 2019.

COELHO, R. F.; NASCIMENTO, V. H.; QUEIROZ, R. L. D.; ROMANO, J. M. T.; CAVALCANTE, C. C. *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*. [S.l.]: CRC Press, 2015.

COHEN, I. Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. *IEEE Transactions on Speech and Audio Processing*, v. 11, n. 5, p. 466–475, 2003.

COHEN, I.; BERDUGO, B. Speech enhancement for non-stationary noise environments. *Signal Processing*, v. 81, n. 11, p. 2403–2418, 2001.

COLOMES, C.; SCHMIDMER, C.; THIEDE, T.; TREURNIET, W. C. Perceptual quality assessment for digital audio: PEAQ-the new ITU standard for objective measurement of the perceived audio quality. *Journal of the Audio Engineering Society*, September 1999.

DONOHO, D.; JOHNSTONE, I. Threshold selection for wavelet shrinkage of noisy data. *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, v. 1, p. A24–A25, 1994.

EPHRAIM, Y.; MALAH, D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 32, n. 6, p. 1109–1121, 1984.

FLANDRIN, P.; GONÇALVÈS, P.; RILLING, G. Detrending and denoising with empirical mode decompositions. *12th European Signal Processing Conference*, p. 1581–1584, 2004.

FLANDRIN, P.; RILLING, G.; GONCALVES, P. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, v. 11, n. 2, p. 112–114, 2004.

GAROFOLO, J. S.; LAMEL, L.; FISHER, W. M.; FISCUS, J. G.; PALLETT, D. S.; DAHLGREN, N. L. TIMIT Acoustic-Phonetic Continuous Speech Corpus. 1993.

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, 2012.

HANSEN, J. H. L.; PELLOM, B. L. An effective quality evaluation protocol for speech enhancement algorithms. *Proceedings ICSLP*, 1998.

HE, C.; HUANG, J.; ZHANG, Q.; LEI, K. Reliable mobile underwater wireless communication using wideband chirp signal. *WRI International Conference on Communications and Mobile Computing*, v. 1, p. 146–150, 2009.

HENDRIKS, R. C.; CRESPO, J. B.; JENSEN, J.; TAAL, C. H. Optimal near-end speech intelligibility improvement incorporating additive noise and late reverberation under an approximation of the short-time sii. v. 23, n. 5, p. 851–862, 2015.

HENDRIKS, R. C.; HEUSDENS, R.; JENSEN, J. MMSE based noise PSD tracking with low complexity. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 4266–4269, 2010.

HU, Y.; LOIZOU, P. A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America*, v. 122, p. 1777, 10 2007.

HU, Y.; LOIZOU, P. C. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 16, n. 1, p. 229–238, 2008.

HUANG, N.; SHEN, Z.; LONG, S.; WU, M.; SHIH, H.; ZHENG, Q.; YEN, N.-C.; TUNG, C.-C.; LIU, H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, v. 454, p. 903–995, 1998.

HURST, H. E. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, v. 116, n. 1, p. 770–799, 1951.

LIN, L.; WANG, Y.; ZHOU, H. Iterative filtering as an alternative algorithm for empirical mode decomposition. *Advances in Adaptive Data Analysis*, v. 01, n. 04, p. 543–560, 2009.

MANOHAR, K.; RAO, P. Speech enhancement in nonstationary noise environments using noise properties. *Speech Commun.*, v. 48, p. 96–109, 2006.

MARTIN, R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, v. 9, n. 5, p. 504–512, 2001.

OMITAOMU, O. A.; PROTOPOPESCU, V. A.; GANGULY, A. R. Empirical mode decomposition technique with conditional mutual information for denoising operational sensor data. v. 11, n. 10, 2011.

OU, H.; ALLEN, J. S.; SYRMOS, V. L. Frame-based time-scale filters for underwater acoustic noise reduction. *IEEE Journal of Oceanic Engineering*, v. 36, n. 2, p. 285–297, 2011.

PASTOR, D.; SOCHELEAU, F.-X. Robust estimation of noise standard deviation in presence of signals with unknown distributions and occurrences. *IEEE Transactions on Signal Processing*, v. 60, n. 4, p. 1545–1555, 2012.

PAVLOVIC, C. SII—speech intelligibility index standard: ANSI S3.5 1997. *The Journal of the Acoustical Society of America*, v. 143, n. 3, p. 1906–1906, 2018.

QUACKENBUSH, S. R.; BARNWELL, T.; CLEMENTS, M. *Objective Measures of Speech Quality*. [S.l.]: Prentice Hall, 1988.

RAHMATI, M.; POMPILI, D. Unisec: Inspection, separation, and classification of underwater acoustic noise point sources. *IEEE Journal of Oceanic Engineering*, v. 43, n. 3, p. 777–791, 2018.

RHEBERGEN, K.; VERSFELD, N. A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, v. 117, p. 2181–92, 05 2005.

RIX, A.; BEERENDS, J.; HOLLIER, M.; HEKSTRA, A. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. *IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, v. 2, p. 749–752, 2001.

ROUSSEEUW, P. J.; RONCHETTI, E. Influence curves of general statistics. *Journal of Computational and Applied Mathematics*, v. 7, n. 3, p. 161–166, 1981.

SANTOS-DOMÍNGUEZ, D.; TORRES-GUIJARRO, S.; CARDENAL-LÓPEZ, A.; PENA-GIMENEZ, A. ShipsEar: An underwater vessel noise database. *Applied Acoustics*, v. 113, p. 64–69, 2016.

SCALART, P.; FILHO, J. Speech enhancement based on a priori signal to noise estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, v. 2, p. 629–632, 1996.

SIDDAGANGAIAH, S.; LI, Y.; GUO, X.; YANG, K. On the dynamics of ocean ambient noise: Two decades later. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, v. 25, n. 10, p. 103117, 2015.

TAAL, C. H.; HENDRIKS, R. C.; HEUSDENS, R.; JENSEN, J. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 7, p. 2125–2136, 2011.

TAAL, C. H.; JENSEN, J.; LEIJON, A. On optimal linear filtering of speech for near-end listening enhancement. v. 20, n. 3, p. 225–228, 2013.

TAVARES, R.; COELHO, R. Speech enhancement with nonstationary acoustic noise detection in time domain. *IEEE Signal Processing Letters*, v. 23, n. 1, p. 6–10, 2016.

TORCOLI, M.; KASTNER, T.; HERRE, J. Objective measures of perceptual audio quality reviewed: An evaluation of their application domain dependence. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 29, p. 1530–1541, 2021.

URICK, R.; KUPERMAN, W. A. Ambient noise in the sea. *The Journal of the Acoustical Society of America*, v. 86, n. 4, p. 1626–1626, 1989.

VEITCH, D.; ABRY, P. A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory*, v. 45, n. 3, p. 878–897, 1999.

WENZ, G. M. Acoustic ambient noise in the ocean: Spectra and sources. *The Journal of the Acoustical Society of America*, v. 34, n. 12, p. 1936–1956, 1962.

WENZ, G. M. Review of underwater acoustics research: Noise. *The Journal of the Acoustical Society of America*, v. 51, n. 3B, p. 1010–1024, 1972.

WILCOCK, W. S.; STAFFORD, K. M.; ANDREW, R. K.; ODOM, R. I. Sounds in the ocean at 1–100 hz. *Annual Review of Marine Science*, v. 6, n. 1, p. 117–140, 2014.

WOODWARD, B.; SARI, H. Digital underwater acoustic voice communications. *IEEE Journal of Oceanic Engineering*, v. 21, n. 2, p. 181–192, 1996.

WU, Z.; HUANG, N. E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, v. 01, n. 01, p. 1–41, 2009.

ZÃO, L.; COELHO, R.; FLANDRIN, P. Speech enhancement with EMD and Hurst-based mode selection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, v. 22, n. 5, p. 899–911, 2014.