

Predição Objetiva de Inteligibilidade em Ambientes Acústicos Industriais

F. Borges e R. Coelho

Resumo— Este artigo apresenta um estudo sobre a predição objetiva da inteligibilidade em ambientes acústicos industriais. No experimento, foram consideradas as medidas objetivas ESTOI, ESII e ASII_{ST} para a avaliação da inteligibilidade em diversos cenários acústicos. Adicionalmente, foi investigado o efeito do ruído e da reverberação, típicos desses ambientes, sobre os sinais de voz, levando em conta aspectos de não-estacionariedade e comportamento tempo-frequência. Os resultados obtidos indicaram que a adição de reverberação contribui para a atenuação dos ruídos acústicos, resultando em ganhos de inteligibilidade nos diferentes cenários ruidosos analisados.

Palavras-Chave— Predição objetiva da inteligibilidade; reverberação; ruído acústico.

Abstract— This article presents a study on the objective prediction of intelligibility in industrial acoustic environments. In the experiment, the objective measures ESTOI, ESII, and ASII_{ST} were employed to evaluate intelligibility across various acoustic scenarios. Additionally, the effects of noise and reverberation, which are characteristic of these environments, were investigated on speech signals, considering aspects of non-stationarity and time-frequency behavior. The results indicate that the addition of reverberation contributes to the attenuation of acoustic noise, resulting in intelligibility improvements across the different analyzed noisy scenarios.

Keywords— Objective intelligibility prediction; reverberation; acoustic noise.

I. INTRODUÇÃO

A reverberação é um efeito comumente presente em cenários urbanos. Em ambientes industriais fechados, o impacto acústico é mais intenso. Ruídos com variados graus de não-estacionariedade estão igualmente presentes, podendo corromper os sinais de voz. Os ruídos alteram as características temporais e espectrais, acarretando a redução da qualidade e da inteligibilidade [1], [2]. Diversos métodos de aprimoramento foram propostos na literatura para mitigar esses efeitos [3], [4]. A reverberação, por outro lado, é “absorvida” pelo sistema nervoso e fornece informações relevantes a respeito do ambiente [5]–[7]. Em situações onde a trajetória de propagação do som é prejudicada [8], a reverberação é fundamental para a comunicação.

Pesquisas evidenciam a relevância dos efeitos associados à reverberação na inteligibilidade de sinais acústicos. A incorporação das reflexões primárias [9]–[11] e o incremento da difusividade espacial [12] são exemplos relevantes. Não obstante, a influência da reverberação na inteligibilidade dos

sinais de voz ainda carece de compreensão abrangente [6], [12].

O presente artigo propõe analisar o efeito da adição da reverberação na inteligibilidade de sinais em ambientes industriais.

Os testes são realizados em cenários acústicos compostos por ruídos típicos da indústria. Esses ruídos incluem Gerador, Motor, Soprador e Máquina, provenientes da base NIGENS [13]. Os sinais de voz utilizados pertencem a um subconjunto da base TIMIT [14]. A esses sinais, foram adicionados os ruídos com níveis de Razão Sinal-Ruído (*Signal-to-Noise Ratio*, SNR) de -10 dB, -5 dB, 0 dB e 5 dB. Os sinais são reverberados com Respostas ao Impulso (*Room Impulse Response*, RIR). Neste estudo, são utilizadas RIR de quatro salas reais: *Meeting* e *Stairway*, da base AIR [15], além de LASP 1 e LASP 2, da base LASP¹. Para a avaliação da inteligibilidade, são empregadas as medidas objetivas de predição ESTOI (*Extended Short-Time Objective Intelligibility*) [16], ESII (*Extended Speech Intelligibility Index*) [17] e ASII_{ST} (*Approximated Short-Time Speech Intelligibility Index*) [18]. Adicionalmente, o efeito do ruído e da reverberação nos sinais de voz é analisado com base na não-estacionariedade, em espectrogramas e nos espectros de Hilbert.

O restante deste artigo está estruturado da seguinte forma: A Seção II demonstra os efeitos do ruído e da reverberação nos sinais acústicos. A Seção III aborda as medidas objetivas de predição utilizadas para a avaliação da inteligibilidade. A descrição dos cenários experimentais e a discussão dos resultados são realizadas na Seção IV. Por fim, na Seção V é apresentada a conclusão do trabalho.

II. EFEITOS DO RUÍDO E DA REVERBERAÇÃO NOS SINAIS ACÚSTICOS

O sinal acústico reverberado resulta da combinação do som propagado pelo caminho direto com suas múltiplas reflexões [6]. A reverberação é tida como um efeito convolutivo no sinal de voz. Os ruídos acústicos, em contrapartida, são interpretados como um efeito aditivo [19]. Deste modo, um sinal reverberante e ruidoso, $r(t)$, é definido por:

$$r(t) = s(t) * h(t) + w(t), \quad (1)$$

onde $s(t)$ é o sinal de voz e $w(t)$ o ruído acústico. A resposta ao impulso do ambiente, representada por $h(t)$, é caracterizada pelo Tempo de Reverberação (RT_{60}^2) e pela Razão entre os Sinais Direto e Reverberado (*Direct-to-Reverberant Ratio*, DRR) [7].

¹Disponível em lasp.ime.eb.br

²Tempo necessário para que a energia da RIR decaia em 60 dB [7].

F. Borges é doutorando do Programa de Pós-Graduação em Engenharia Elétrica do Instituto Militar de Engenharia (IME). O trabalho dos autores F. Borges e R. Coelho é desenvolvido no Laboratório de Processamento de Sinais Acústicos (LASP/IME) e parcialmente financiado pelo CNPq (308155/2019-0) e pela FAPERJ (200518/2023). E-mails: filipe.diniz@ime.eb.br, coelho@ime.eb.br.

Os ruídos acústicos provocam o mascaramento do sinal de voz, modificando suas características naturais [20]. Neste trabalho, o Índice de Não-Estacionariedade (*Index of Non-Stationarity*, INS) [21] é adotado para esta análise. O INS é obtido a partir da comparação de um determinado sinal acústico com referenciais estacionários, denominados *surrogates*. A escala de observação T_h/T corresponde à razão entre as durações da janela utilizada na análise espectral (T_h) e do sinal (T). Para cada valor de T_h é estabelecido um limiar, γ , associado ao nível de confiança de 95%. A avaliação da hipótese de estacionariedade é realizada comparando este limiar com o valor de INS em cada uma das escalas de observação, da seguinte forma:

$$\text{INS} : \begin{cases} \leq \gamma, & \text{sinal é estacionário;} \\ > \gamma, & \text{sinal é não-estacionário.} \end{cases} \quad (2)$$

A Fig. 1 ilustra as curvas de INS para os ruídos selecionados para os experimentos. Note que os sinais dos quatro ruídos apresentam não-estacionariedade na maior parte das escalas de observação. O ruído Gerador apresenta $\text{INS}_{\max}$ de 2,51, Motor de 2,35 e Máquina de 2,82. O ruído com maior valor de INS é Soprador, cujo $\text{INS}_{\max}$ é de 8,20.

Em cenários ruidosos, a presença da reverberação pode promover a atenuação da interferência, favorecendo a inteligibilidade [5], [6], [8], [12]. Este efeito é demonstrado na Fig. 2. Ela apresenta, na parte superior, os espectrogramas; ao centro, os espectros de Hilbert; e, na parte inferior, as respectivas curvas de INS. Quatro condições distintas são consideradas: sinal de voz limpo (a), ruído Soprador (b), sinal de voz ruidoso (c) e sinal de voz ruidoso com adição de reverberação (RIR *Meeting*, $\text{RT}_{60} = 0,31$ s, DRR = 4,70 dB) (d). Nos casos (c) e (d), os sinais são contaminados pelo ruído Soprador com SNR de 0 dB.

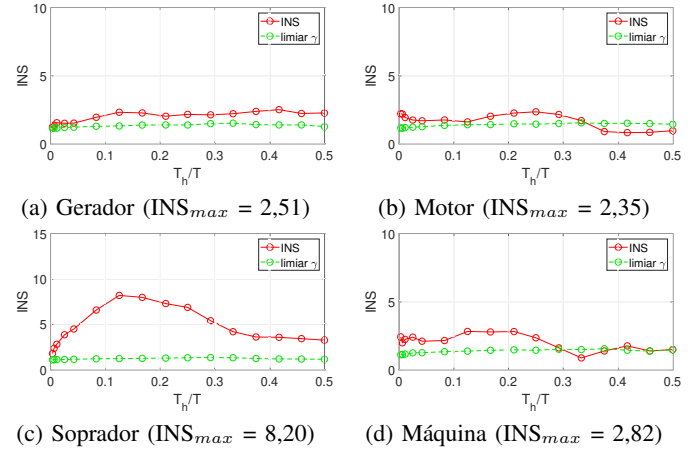


Fig. 1. INS dos ruídos acústicos industriais analisados: Gerador (a), Motor (b), Soprador (c) e Máquina (d).

Os espectrogramas e espectros de Hilbert evidenciam que o sinal de voz (a) apresenta componentes temporais e espectrais bem definidas. O ruído Soprador (b), por outro lado, exibe distorções ao longo do espectro, especialmente nas frequências abaixo de 2 kHz. Tanto o sinal de voz quanto o ruído apresentam características não-estacionárias em todas as escalas de observação T_h/T . Os sinais ruidoso (c) e ruidoso com reverberação (d) preservam parcialmente a não-estacionariedade do sinal de voz. Esta característica, no entanto, é mais fortemente atenuada no sinal ruidoso. No sinal ruidoso com reverberação, o conteúdo espectral associado ao sinal de voz limpo também é mais evidente. Esses resultados sugerem que a adição de reverberação favorece a inteligibilidade. Esse efeito ocorre tanto pela atenuação da interferência ruidosa quanto pela preservação da não-estacionariedade do sinal de voz.

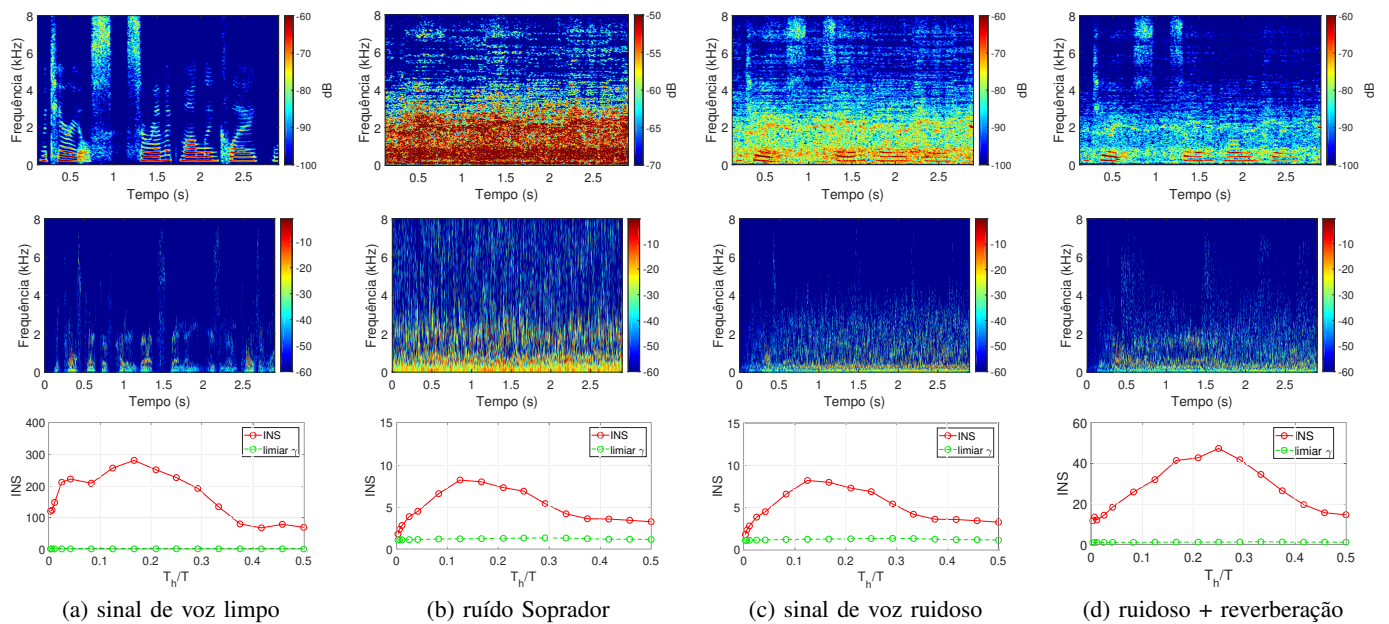


Fig. 2. Espectrogramas (parte superior), espectros de Hilbert (ao centro) e os respectivos valores de INS (parte inferior) para: (a) sinal de voz limpo, (b) ruído Soprador, (c) sinal de voz ruidoso (ruído Soprador com SNR=0 dB), (d) sinal de voz ruidoso com adição de reverberação (ruído Soprador com SNR=0 dB e ambiente *Meeting*, $\text{RT}_{60} = 0,31$ s).

III. MEDIDAS OBJETIVAS DE PREDIÇÃO DE INTELIGIBILIDADE

Esta Seção descreve brevemente as três medidas objetivas de predição de inteligibilidade (ESTOI, ESII e $ASII_{ST}$) empregadas neste trabalho.

A. ESTOI

A ESTOI foi desenvolvida para avaliar a inteligibilidade de sinais corrompidos por ruídos não-estacionários [16]. A medida representa a correlação entre as envoltórias espectrais dos sinais de referência e distorcido. Com isto, captura melhor a relação de dependência entre bandas e leva em consideração a mudança nas características do sinal e do ruído com o tempo.

A determinação da ESTOI envolve os seguintes processos [22]:

- 1) *Segmentação*: os sinais de referência e distorcido são segmentados em janelas de curta duração e decompostos a nível espectral por FFT (*Fast Fourier Transform*);
- 2) *Extração das regiões de silêncio*: quadros contendo apenas silêncio são descartados com base em sua energia;
- 3) *Agrupamento em bandas*: os espectros são agrupados em 15 bandas de 1/3 de oitava, contendo 30 quadros cada, e;
- 4) *Normalização*: o sinal distorcido é normalizado em relação ao sinal de referência.

A ESTOI corresponde à média dos coeficientes de correlação entre os segmentos espectrais processados dos sinais de referência ($\mathbf{x}_{m,n}$) e distorcido ($\mathbf{y}_{m,n}$), sendo expressa por

$$ESTOI = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \rho(\mathbf{x}_{m,n}, \mathbf{y}_{m,n}), \quad (3)$$

em que $\rho(\cdot)$ representa o coeficiente de correlação linear, M corresponde ao número de segmentos temporais e N ao número de bandas analisadas.

A medida de inteligibilidade ESTOI é adequada para diversos tipos de degradações e possui correlação ($\rho = 0,95$) com testes subjetivos de inteligibilidade [23].

B. ESII

O ESII foi desenvolvido a partir do Índice de Inteligibilidade da Fala (*Speech Intelligibility Index*, SII) [23] para ser capaz de lidar com o comportamento não-estacionário das distorções. A SII baseia-se na premissa de que a inteligibilidade depende da proporção de informação espectral audível. Esta proporção é calculada a partir da divisão do espectro em bandas com igual contribuição, ponderadas pelo SNR. A medida possibilita a predição da inteligibilidade em cenários com ruídos estacionários. Entretanto, para ruídos não-estacionários, a estimativa de inteligibilidade não tem a mesma precisão [24]. A versão estendida do SII (ESII) foi proposta para possibilitar a predição de inteligibilidade mesmo com ruídos não-estacionários [17]. A ESII é determinada em duas etapas [25]:

- 1) *Obtenção dos espectros*: obtenção dos espectros de tempo curto do sinal de fala e do ruído em bandas críticas, e;

- 2) *Cálculo do SNR*: cálculo do SNR em cada banda e quadro temporal, limitado ao intervalo entre -15 dB e 15 dB.

A ESII corresponde à média temporal dos valores de SNR normalizados em cada região de tempo, ponderados de acordo com a faixa de frequência, ou seja

$$ESII = \frac{1}{T} \sum_{\tau=1}^T \sum_{j=1}^J \gamma_j \xi(\tau, j), \quad (4)$$

onde $\xi(\tau, j)$ representa o SNR normalizado de cada quadro temporal τ e banda de frequência j , T corresponde ao número total de quadros considerados e γ_j é uma função de importância da banda.

C. $ASII_{ST}$

A aproximação realizada pela medida ESII é imprecisa para bandas com valores de SNR inferiores a -15 dB [22]. Como consequência, a energia da voz pode ser distribuída incorretamente entre as bandas de frequência, prejudicando a predição da inteligibilidade [16]. Para mitigar este problema, foi originada a medida $ASII_{ST}$ [26]. Diferentemente do ESII, esta medida considera adequadamente bandas com SNR inferiores a -15 dB e superiores a 15 dB, empregando uma normalização contínua do SNR.

A medida $ASII_{ST}$ é então obtida pela combinação das contribuições normalizadas de todas as bandas e quadros temporais, isto é

$$ASII_{ST} = \frac{1}{T} \sum_{\tau=1}^T \sum_{j=1}^J \gamma_j \frac{\xi(\tau, j)}{1 + \xi(\tau, j)}, \quad (5)$$

em que T representa o número de quadros temporais analisados, J o número de bandas críticas, γ_j é uma função de importância da banda j e $\xi(\tau, j)$ o SNR de um determinado quadro τ com banda de frequência j .

IV. EXPERIMENTOS, RESULTADOS E DISCUSSÃO

Nesta Seção, o efeito da adição da reverberação nos cenários acústicos ruidosos é avaliado empregando as medidas objetivas de predição de inteligibilidade ESTOI, ESII e $ASII_{ST}$.

Os experimentos foram realizados em oito cenários distintos, sendo quatro ruidosos e quatro ruidosos com reverberação. Nos cenários ruidosos sem adição de reverberação (UNP), o ruído acústico é adicionado ao sinal de voz nos níveis de SNR de -10 dB, -5 dB, 0 dB e 5 dB. Já nos cenários com reverberação, o ruído é adicionado apenas ao componente direto [8], [12], mantendo os mesmos níveis de SNR dos cenários UNP. Nesse caso, o SNR é definido em relação ao sinal direto [27]. O componente reverberado, obtido pela convolução do sinal de voz com a RIR, é então somado ao sinal direto ruidoso e o resultado é equalizado ao componente direto [8], [12]. Um subconjunto da base de sinais de voz TIMIT [14] foi selecionado para os testes. Esta base é composta por 192 locuções amostradas a uma taxa de 16 kHz e padronizadas para uma duração de 3 segundos. Quatro ruídos acústicos típicos de ambientes industriais (Gerador, Motor, Soprador e Máquina)

foram extraídos da base de sinais NIGENS [13], cuja taxa de amostragem é de 48 kHz. Antes de serem adicionados, os ruídos foram padronizados para a mesma duração e a mesma taxa de amostragem dos sinais de voz. As RIR das salas reais *Meeting* e *Stairway*, da base AIR [15], e LASP 1 e LASP 2, da base LASP, foram selecionadas para os experimentos. As RIR da base AIR, com taxa de amostragem de 48 kHz, e da base LASP, com 44,1 kHz, também foram reamostradas para a mesma taxa dos sinais de voz. Ao todo, foram realizados 12288 testes, sendo 1536 por cenário acústico.

Os valores de RT_{60} , de DRR, as dimensões das salas e as distâncias de captação das RIR, D_{fm} , são apresentados na Tabela I. Os dados demonstram que as RIR *Meeting* (M) e LASP 1 (L1) possuem os menores valores de RT_{60} e, em contrapartida, os maiores valores de DRR.

TABELA I
PARÂMETROS DAS REVERBERAÇÕES UTILIZADAS.

RIR	RT_{60} (s)	DRR (dB)	Dimensões (m ³)	D_{fm} (m)
<i>Meeting</i> (M)	0,31	4,70	[8,0 x 5,0 x 3,1]	1,45
<i>Stairway</i> (S)	0,82	0,92	–	1,70
LASP 1 (L1)	0,73	6,31	[5,7 x 4,5 x 3,1]	1,20
LASP 2 (L2)	1,59	4,34	[5,7 x 4,5 x 3,1]	1,50

A Tabela II exhibe os valores médios de ESTOI para os quatro cenários ruidosos UNP estudados. Os resultados apontam que a medida ESTOI foi mais impactada pelo cenário ruidoso definido pelo ruído Soprador, de maior índice de não-estacionariedade, independentemente do SNR considerado. O cenário ruidoso menos desafiador é dado pelo ruído Motor, que possui o menor INS dentre os ruídos selecionados.

TABELA II

RESULTADOS DE ESTOI PARA OS QUATRO CENÁRIOS RUIDOSOS (UNP).

Cenário Ruidoso (UNP)	SNR (dB)				Média
	-10	-5	0	5	
Gerador	0,16	0,26	0,38	0,52	0,33
Motor	0,20	0,31	0,44	0,58	0,38
Soprador	0,12	0,23	0,36	0,51	0,31
Máquina	0,19	0,31	0,42	0,54	0,37

Na Fig. 3 são mostrados os aprimoramentos médios de ESTOI (Δ ESTOI) obtidos com a adição da reverberação nos diferentes cenários ruidosos, para cada SNR testado. Os valores de Δ ESTOI de cada cenário nos diversos níveis de SNR testados, bem como as médias dos valores das curvas de Δ ESTOI, são exibidas na Tabela III. Os resultados denotam que a reverberação decorrente da RIR LASP 1, de maior DRR, alcançou o maior valor de Δ ESTOI ($12,07 \times 10^{-2}$) para o ruído Soprador e SNR de -5 dB.

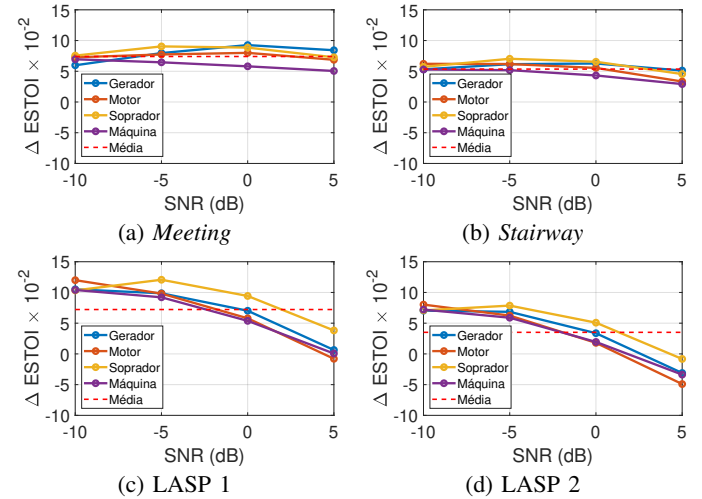


Fig. 3. Resultados de Δ ESTOI dos cenários ruidosos com reverberação.

TABELA III
RESULTADOS DE Δ ESTOI ($\times 10^{-2}$) PARA CENÁRIOS RUIDOSOS COM REVERBERAÇÃO.

Ruído	Reverberação Adicionada				Média
	M	S	L1	L2	
Gerador	7,97	5,68	7,17	3,60	6,11
Motor	7,47	5,17	6,48	2,81	5,48
Soprador	8,16	5,97	9,15	5,11	7,10
Máquina	6,02	4,75	6,72	3,07	5,14
Média	7,41	5,39	7,38	3,65	–

Os resultados das medidas objetivas de inteligibilidade ESII e $ASII_{ST}$ são apresentados na Fig. 4. Ela ilustra a dispersão

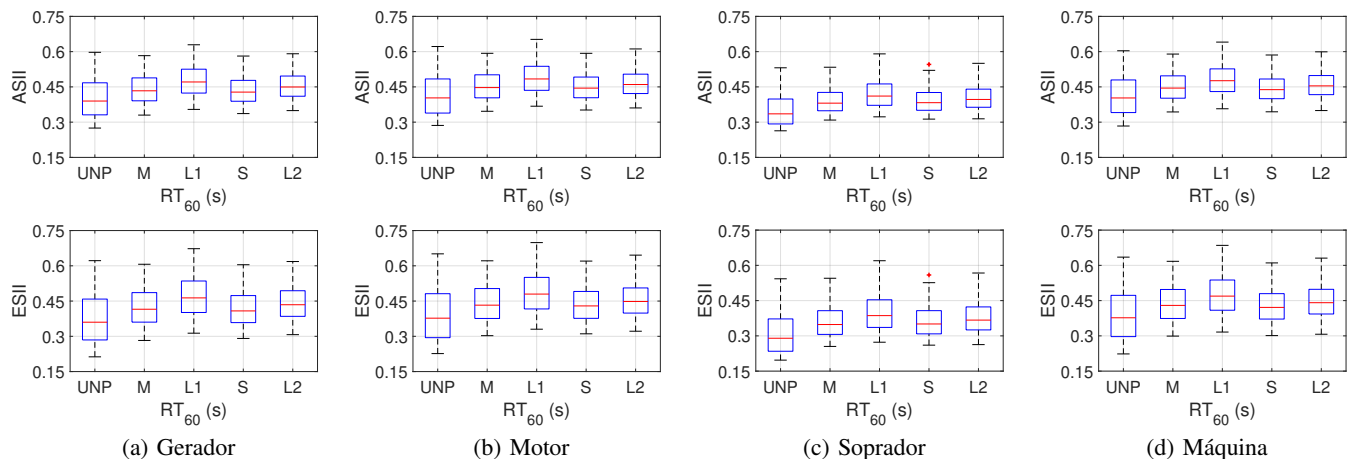


Fig. 4. Resultados de ESII e $ASII_{ST}$ dos cenários ruidosos com e sem adição de reverberação. Ruídos Gerador (a), Motor (b), Soprador (c) e Máquina (d).

das medidas por meio de diagramas de caixa (*boxplots*). Cada diagrama de caixa representa a distribuição dos valores da medida para um determinado ruído, com e sem a adição de reverberação, considerando todos os níveis de SNR avaliados. A Tabela IV apresenta os valores médios destas medidas. Os resultados obtidos permitem inferir que a adição da RIR LASP 1 (L1) resulta nos maiores valores médios de inteligibilidade ESII e $ASII_{ST}$, independentemente do ruído analisado. Os valores médios das medidas de inteligibilidade alcançados no cenário ruidoso formado pelo ruído Soprador foram os menores, independentemente da reverberação adicionada. Como esperado, os valores médios de ESII e $ASII_{ST}$ foram os maiores no cenário ruidoso menos desafiador (Motor) após a adição da reverberação.

TABELA IV

RESULTADOS MÉDIOS DE ESII E $ASII_{ST}$ DOS CENÁRIOS RUIDOSOS SEM (UNP) E COM A ADIÇÃO DA REVERBERAÇÃO.

Ruído	UNP	Reverberação Adicionada				Média
		M	L1	S	L2	
ASII _{ST}						
Gerador	0,40	0,44	0,47	0,44	0,45	0,44
Motor	0,41	0,45	0,49	0,45	0,46	0,45
Soprador	0,35	0,39	0,42	0,39	0,40	0,39
Máquina	0,41	0,45	0,48	0,44	0,46	0,45
Média	0,39	0,43	0,46	0,43	0,44	–
ESII						
Gerador	0,37	0,42	0,47	0,42	0,44	0,42
Motor	0,39	0,44	0,48	0,44	0,45	0,44
Soprador	0,31	0,36	0,40	0,36	0,38	0,36
Máquina	0,39	0,44	0,47	0,43	0,45	0,43
Média	0,37	0,42	0,46	0,41	0,43	–

V. CONCLUSÃO

Este artigo apresentou um estudo sobre a predição objetiva da inteligibilidade em ambientes acústicos industriais. Os experimentos envolveram testes em oito cenários, resultantes da combinação de quatro condições de reverberação e quatro tipos de ruído típicos da indústria. Os resultados indicam que a adição de reverberação aos sinais de fala ruidosos pode contribuir para o aumento da inteligibilidade. Entre as respostas ao impulso analisadas, a RIR LASP 1 apresentou os maiores índices de inteligibilidade, seguida pela RIR Meeting, que, em média, demonstrou os maiores aprimoramentos médios. Ambas estão associadas aos valores máximos de DRR. O cenário ruidoso mais desafiador, correspondente ao ruído Soprador, apresentou os maiores valores médios de $\Delta ESTOI$. Além disso, a análise da não-estacionariedade dos sinais, complementada por espectrogramas e espectros de Hilbert, sugere que a adição de reverberação pode contribuir para a atenuação da interferência ruidosa, principal fator responsável pela redução da inteligibilidade.

REFERÊNCIAS

- [1] P. Assman and Q. Summerfield, "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*. Springer, 2004.
- [2] A. Queiroz and R. Coelho, "Frequência fundamental de sinais reverberantes e ruidosos com classificação de atributos harmônicos," in *Proceedings of the XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt)*, 2023.
- [3] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, 1979.
- [4] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, 2001.
- [5] A. K. Nábělek and P. K. Robinson, "Monaural and binaural speech perception in reverberation for listeners of various ages," *The Journal of the Acoustical Society of America*, vol. 71, 1982.
- [6] J. Traer and J. H. McDermott, "Statistics of natural reverberation enable perceptual separation of sound and space," *Proceedings of the National Academy of Sciences of the United States of America*, 2016.
- [7] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, 2012.
- [8] J. S. Bradley and H. Sato, "On the importance of early reflections for speech in rooms," *The Journal of the Acoustical Society of America*, vol. 113, 2003.
- [9] H. Haas, "The influence of a single echo on the audibility of speech," *Journal of the Audio Engineering Society*, 1949.
- [10] J. P. A. Lochner and J. F. Burges, "The influence of reflections on auditorium acoustics," *Journal of Sound and Vibration*, 1964.
- [11] A. K. Nábělek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," *Journal of Speech and Hearing Research*, vol. 17, 1974.
- [12] J. Grosse and S. van de Par, "Improved binaural speech intelligibility by adding reverberation to the target speaker," in *Proceedings of the International Congress on Acoustics*, 2019.
- [13] I. Trowitzsch, J. Taghia, Y. Kashef, and K. Obermayer, "The NIGENS general sound events database," arXiv:1902.08314, 2019.
- [14] J. S. Garofolo *et al.*, "DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM," NIST Speech Disc 1-1.1, 1993.
- [15] M. Jeub, M. Schäfer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of the IEEE Digital Signal Processing Workshop*, 2009.
- [16] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2009–2022, 2016.
- [17] K. S. Rhebergen and N. J. Versfeld, "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, 2005.
- [18] J. B. Crespo, J. Jensen, and C. H. Taal, "Optimal near-end speech intelligibility improvement incorporating additive noise and late reverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 5, 2015.
- [19] A. Queiroz and R. Coelho, "Métodos de mitigação de efeitos reverberantes e ruidosos com ganho de inteligibilidade e qualidade," in *Proceedings of the XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt)*, 2021.
- [20] G. Zucattelli, F. Farias, and R. Coelho, "A two-stage approach for noisy-reverberant speech intelligibility improvement," in *Proceedings of the XXXVIII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt)*, 2020.
- [21] S. P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Transactions on Signal Processing*, vol. 58, 2010.
- [22] F. S. Farias and R. Coelho, "Adaptive gain methods to improve intelligibility under reverberation," in *Proceedings of the XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt)*, 2018.
- [23] R. Martiny, R. Alcântara, and R. Coelho, "Avaliação da predição objetiva da inteligibilidade de sinais reverberantes e ruidosos com uso de máscaras acústicas," in *Proceedings of the XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt)*, 2022.
- [24] H. J. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *The Journal of the Acoustical Society of America*, vol. 67, 1980.
- [25] R. M. Cox and G. C. A. I. M. Rivera, "Comparison of objective and subjective measures of speech intelligibility in elderly hearing-impaired listeners," *Journal of Speech and Hearing Research*, vol. 34, 1991.
- [26] C. H. Taal, J. Jensen, and A. Leijon, "On optimal linear filtering of speech for near-end listening enhancement," *IEEE Signal Processing Letters*, vol. 20, no. 3, 2013.
- [27] G. Zucattelli and R. Coelho, "Adaptive reverberation absorption using non-stationary masking components detection for intelligibility improvement," *IEEE Signal Processing Letters*, vol. 27, 2020.