Inteligibilidade Objetiva de Sinais com Reverberação e com Uso de Diferentes Máscaras Acústicas

R. Alcântara, R. Coelho e B. S. Masiero

***Resumo*— Este artigo apresenta um estudo do efeito da reverberação acústica na inteligibilidade de sinais de voz. A avaliação inclui o uso das máscaras ideais clássicas IdBM e reverberante e a máscara não-ideal cega BRM. Três medidas objetivas fwSegSNR, CSII e STOI, além da medida de qualidade SegSNR são consideradas nos experimentos. Os resultados mostram que a reverberação impacta a inteligibilidade do sinal de voz e que as máscaras acústicas são capazes de melhorar a inteligibilidade degradada. Além disso, é demonstrado que uso da máscara BRM incrementou a inteligibilidade dos sinais reverberados nos diferentes cenários.**

***Palavras-Chave*— Máscara acústica, inteligibilidade, desreverberação.**

***Abstract*— This article presents a study of the effect of acoustic reverberation on speech intelligibility. This evaluation includes the use of the classics ideal binary masks IdBM and IRM and the blind non-ideal mask BRM. Three objective measures fwSegSNR, CSII, and STOI in addition to the quality measure SegSNR are considered in the experiments. The results show that reverberation impacts speech intelligibility and that binary masks are capable of improving the degraded intelligibility. Moreover, it is demonstrated that BRM increased the intelligibility of reverberated signals in different scenarios.**

***Keywords*— Binary mask, intelligibility, dereverberation.**

I. Introdução

O efeito da reverberação é causado pelas múltiplas reflexões que ocorrem com uma onda sonora em superfícies e objetos antes desta ser captada por um microfone ou um ouvinte. No dia-a-dia, este efeito é mais facilmente notado em locais fechados como salas de aula, auditórios, igrejas ou teatros. Em sinais de voz, a reverberação tem impacto negativo na sua qualidade e inteligibilidade [1], afetando principalmente idosos e usuários de implantes cocleares. Esta degradação tem diversas consequências indesejáveis, como o agravamento do desempenho escolar [2], além de fazer cair as taxas de acerto de sistemas de reconhecimento de palavras e de locutor [3].

A literatura apresenta diferentes técnicas para desreverberação de sinais de voz. Entre elas, estão algoritmos que utilizam filtragem inversa [4] e arranjos de microfones para estimar a RIR (*Room Impulse Response*) da sala [5]. Geralmente, os métodos propostos são avaliados segundo o critério da qualidade de áudio do sinal resultante do processamento.

As máscaras acústicas [6] são soluções baseadas em seleção de canal e foram inicialmente propostas para aprimorar a inteligibilidade de sinais de voz corrompidos por interferências ou ruídos acústicos. Isto é realizado através de uma divisão do sinal corrompido em quadros tempo-frequência e na exclusão dos quadros que forem considerados dominantes pela interferência. A IBM (*Ideal Binary Mask*) é considerada pela literatura como um limite superior do desempenho das máscaras acústicas. Nela, são utilizadas informações a priori para se preservar os quadros em que a SRR (*Signal-to-Reverberation Ratio*) está acima de um limiar predeterminado e excluir os demais. O uso de máscaras acústicas em situações de reverberação se demonstrou eficiente em melhorar a inteligibilidade dos sinais de voz. A IRM (*Ideal Reverberant Mask*) apresentou ganhos de até 72% em testes subjetivos de inteligibilidade realizados com usuários de implantes cocleares [7]. A máscara cega (não-ideal) BRM (*Binary Reverberant Mask*), com foco na reverberação [8], mostrou melhorar a inteligibilidade em testes subjetivos. As máscaras não-ideais têm a vantagem de não serem limitadas ao conhecimento prévio do sinal e apresentam bons resultados. Por estes aspectos, estas máscaras são mais adaptadas a situações reais.

Este artigo apresenta um estudo com medidas objetivas de inteligibilidade e de qualidade para avaliar o efeito causado pela reverberação e o desempenho das máscaras acústicas nestes casos. A qualidade é medida através da SegSNR (*Segmental Signal-to-noise Ratio* [9]). Para a avaliação da inteligibilidade acústica, são adotadas três medidas: fwSegSNR (*Frequency-Weighted* SegSNR [10]), CSII (*Coherence and Speech Intelligibility Index* [11]) e STOI (*Short-Time Objective Intelligibility* [12]). Na literatura, estas medidas foram aplicadas com sucesso para investigar situações de distorção por ruídos [12] [13] [14]. Os resultados indicam que a reverberação degradou a qualidade e a inteligibilidade da voz. Em uma mesma sala, esta degradação ocorreu em maior magnitude com o aumento da d_{fm} (distância fonte-microfone) e de RT_{60} (*Reverberation Time*). O uso das máscaras acústicas nos sinais de voz com reverberação aprimorou a sua qualidade e inteligibilidade.

O restante deste artigo está organizado da seguinte maneira: Na Seção II são descritas as implementações das máscaras utilizadas neste trabalho. A Seção III descreve brevemente as medidas SegSNR, fwSegSNR, CSII e STOI. Na Seção IV são apresentados os resultados das medidas aplicadas aos sinais

R. Alcântara, mestrando no Programa de Pós-graduação da Faculdade de Engenharia Elétrica e de Computação (FEEC), UNICAMP; R. Coelho*, Laboratório de Processamento de Sinais Acústicos (lasp.ime.eb.br), Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil; B. S. Masiero, Departamento de Comunicações, UNICAMP. E-mails: {raoni@decom.fee.unicamp.br, coelho@ime.eb.br, masiero@unicamp.br}. *Este trabalho foi parcialmente financiado pelo CNPq/307866/2015-7.

de voz com reverberação e após o uso das máscaras. Por fim, a Seção V conclui este trabalho.

II. Máscara Acústica para Sinais com Reverberação

Nesta Seção é apresentada uma breve descrição das três máscaras acústicas IdBM, IRM e BRM. O objetivo principal do emprego das máscaras acústicas é a redução dos efeitos da reverberação no sinal alvo, i.e., sinal de voz, e consequentemente, o aprimoramento da qualidade e inteligibilidade do sinal.

A. Máscara Acústica: Ideal

No problema do "*cocktail party*" [15], um ouvinte é capaz de selecionar e compreender uma única fonte sonora em meio a diversas interferências. As máscaras ideais foram propostas para simular esta capacidade perceptual humana. Geralmente, elas estão definidas pelos seguintes passos [16]:

1) *Decomposição em tempo-frequência*: O sinal reverberado é janelado e, em seguida, é aplicada a FFT (*Fast Fourrier Transform*) em cada um dos quadros. O sinal $Y(k,t)$ representa o espectro do sinal reverberado na sub-banda k e tempo t.
2) *Critério de seleção*: Define-se um critério $C(k,t)$ que determinará se o quadro $Y(k,t)$ será considerado dominante pela voz ou pela reverberação. No caso da máscara ideal, além da representação tempo-frequência do sinal reverberado, também é necessário o conhecimento do sinal sem reverberação para a obtenção de $C(k,t)$.
3) *Mascaramento*: Os quadros que comporão o sinal "mascarado" $\hat{X}(k,t)$ são definidas por:

$$\hat{X}(k,t) = \begin{cases} Y(k,t), & \text{se} \quad C(k,t) \geq \gamma, \\ 0, & \text{caso contrário,} \end{cases} \tag{1}$$

onde γ é o limiar de seleção.
4) *Reconstrução do sinal*: A FFT inversa é aplicada em $\hat{X}(k,t)$ para reconstruir os quadros no domínio do tempo. Em seguida, os quadros reconstruídos são usadas para concatenar e obter o sinal mantendo as sobreposições utilizadas inicialmente.

As máscaras IdBM [17] e IRM [18] utilizadas neste estudo estão detalhadas abaixo:

1) IdBM: Em [17] é empregada a FFT como forma de decomposição em frequência dos quadros do sinal. O janelamento foi realizado com duração de quadro de 20 ms e 50% de sobreposição. O critério de seleção escolhido é a razão sinal-reverberação $\text{SRR}(k,t) \geq -5$ dB.

2) IRM: Os filtros *gammatone* [19] [20] [21] foram propostos para descrever o comportamento da função de resposta ao impulso do sistema auditivo humano no domínio do tempo. Sendo assim, este banco de filtros é amplamente aplicado para modelar ou simular o sistema auditivo. Por esta interessante característica, estes filtros foram adotados nas propostas da máscaras acústicas IRM e BRM. Nela, é utilizado um banco de 128 filtros *gammatone* de quarta ordem para realizar a decomposição tempo-frequência. As frequências centrais estão espaçadas entre si de acordo com a escala ERB (*Equivalent rectangular bandwidth*) distribuída entre 50 Hz e 8 kHz. Em seguida, os sinais filtrados de cada sub-banda são divididos em quadros de 20 ms com 50% de sobreposição. Este processo é realizado com o sinal reverberado e com o sinal sem reverberação para a obtenção da SRR de cada quadro tempo-frequência. O critério de seleção utilizado é $\text{SRR}(k,t) \geq -5$ dB.

Para reconstruir o sinal, as 128 sub-bandas são obtidas a partir de $\hat{X}(k,t)$ e invertidas no tempo. Em seguida, é aplicado um filtro *gammatone* em cada uma e estas são invertidas no tempo novamente. Ao final, as sub-bandas são somados e o sinal de voz com redução do efeito de reverberação é obtido.

B. Máscara Acústica para Reverberação

As máscaras acústicas ideais têm a limitação de necessitarem de informações do sinal de voz limpo (sem reverberação) para o cálculo de $\text{SRR}(k,t)$. A BRM [8] é uma máscara cega não-ideal que não necessita das informações do sinal sem reverberação. Para isto, é necessário utilizar um critério de seleção diferente da SRR.

Para a obtenção da representação tempo-frequência, os autores propõem um banco de 64 filtros *gammatone* de quarta ordem espaçados logaritmicamente entre 50 Hz e 8 kHz. Em seguida, para cada quadro tempo-frequência $r(k,t)$ é calculado um coeficiente dado por:

$$f_M(k,t) = 10 \cdot \log_{10}\left(\frac{\sigma^2_{r'}(k,t)}{\sigma^2_{|r|}(k,t)}\right), \tag{2}$$

onde $r'(t,j) = |r(k,t)|^\alpha$ e $|r(t,j)|$ é o valor absoluto do quadro no tempo t e sub-banda j. Depois, os valores de f_M são suavizados no tempo através de um filtro mediana de ordem 3. Para determinar o critério de seleção da máscara é utilizado o histograma $f_{hist}(k,t)$, computado a partir dos valores de f_M dos Q_p quadros anteriores a t até os seus Q_f quadros seguintes.

Cada histograma $f_{hist}(k,t)$ normalizado possui L classes com pesos p_i $(i = 1, ..., L)$. A partir destes valores, são calculadas a média global $m_G = \sum_{i=1}^{L} i.p_i$, a média cumulativa $m(l) = \sum_{i=1}^{l} i.p_i$ e a soma cumulativa $P_s(l) = \sum_{i=1}^{l} p_i$. O limiar ótimo l^* é definido como o valor de l que maximiza a variância entre classes $\sigma^2_B(l)$, dada por:

$$\sigma^2_B(l) = \frac{(m_G P_s(l) - m(l))^2}{P_s(l)(1 - P_s(l))}. \tag{3}$$

O valor l^* é empregado como critério de seleção para definir se o conteúdo do quadro $r(k,t)$ é predominante pela voz e será mantido após o mascaramento. Isto ocorre de acordo com,

$$\hat{X}(k,t) = \begin{cases} Y(k,t), & \text{se} \quad f_M(k,t) > \max(l^*(k,t), l_0), \\ 0, & \text{caso contrário,} \end{cases} \tag{4}$$

onde l_0 é o limiar de silêncio.

A reconstrução do sinal mascarado é realizada primeiramente em cada sub-banda de frequência. Os quadros são concatenados de acordo com as suas sobreposições iniciais e invertidas no tempo. Um filtro *gammatone* é aplicado em cada sub-banda e, em seguida, o sinal é invertido no tempo novamente. Por fim, os sinais são somados para a obtenção do sinal reconstruído.

III. MEDIDAS DE INTELIGIBILIDADE ACÚSTICA

Esta Seção descreve as três medidas objetivas de inteligibilidade fwSegSNR, CSII e STOI aplicadas neste estudo. Estas medidas permitem avaliar o efeito causado pela reverberação nos sinais de voz e a eficiência das máscaras acústicas em recuperar a inteligibilidade desses sinais.

A. fwSegSNR

Esta medida é calculada a partir da soma ponderada da SNR de cada região tempo-frequência e é definida por,

$$\text{fwSegSNR} = \frac{1}{Q}\sum_{\tau=0}^{Q-1}\frac{\sum_{j=1}^{K} W_f(j,\tau)\text{SNR}(j,\tau)}{\sum_{j=1}^{K} W_f(j,\tau)}, \quad (5)$$

onde t e τ são os índices do quadro e da sub-banda. O valor de $\text{SNR}(j,\tau)$ é obtido a partir de $10\cdot\log\frac{|X(j,\tau)|^2}{(|X(j,\tau)|-|\hat{X}(j,\tau)|)^2}$. $|X(j,\tau)|$ e $|\hat{X}(j,\tau)|$ representam os espectros dos sinais sem reverberação e após a utilização das máscaras, respectivamente, e são obtidos a partir do janelamento com quadros de 32 ms de duração e 75% de sobreposição, seguido da divisão dos quadros em K sub-bandas de frequência com filtros Gaussianos. A ponderação de frequência é feita por $W_f(j,\tau) = |X(j,\tau)|^\gamma$, onde $\gamma = 0,2$. O valor é identificado em [10] por refletir maior correlação com resultados perceptuais de inteligibilidade. Os valores de SNR de cada quadro são limitados entre -10 dB e 35 dB.

B. CSII

Para a CSII [11], o sinal de referência sem reverberação $x(t)$ e o sinal resultante do uso das máscaras $y(t)$ são janelados com tamanho de quadro de 16 ms com 50% de sobreposição. A partir da aplicação de uma DFT (*Discrete Fourier Transform*), são obtidos os respectivos espectros $X_j(f)$ e $Y_j(f)$, com $f = 0,...,F$, referentes ao quadro j. A medida MSC (*magnitude-squared coherence*) é dada por,

$$\text{MSC}(f) = \frac{|\sum_{j=0}^{Q-1} X_j(f)Y_j^*(f)|^2}{(\sum_{j=0}^{Q-1}|X_j(f)|^2)(\sum_{j=0}^{Q-1}|Y_j(f)|^2)}, \quad (6)$$

onde Q é o número total de quadros. Em seguida, a SRR é calculada por,

$$\text{SRR}(j) = \frac{\sum_{f=0}^{F} I_j(f)MSC(f)S_y(f)}{\sum_{f=0}^{F} I_j(f)[1-MSC(f)]S_y(f)}, \quad (7)$$

onde $S_y(f)$ é a amostra f da densidade espectral de potência de $y(t)$ e $I_b(f)$ é um filtro que atribui um peso à frequência f relativo à inteligibilidade.

A obtenção de $\text{SDR}(j)$ é realizada em três níveis de amplitudes diferentes do sinal de entrada. Assim, o $CSII_{\text{alto}}$ é obtido a partir das regiões com amplitude acima do valor RMS (*root mean square*). O $CSII_{\text{médio}}$ é calculado com as regiões entre 0 e 10 dB abaixo de RMS. A partir das regiões restantes, é obtido $CSII_{\text{baixo}}$. O resultado desta composição é dado por $c = -3,47 + 1,84CSII_{\text{baixo}} + 9,99CSII_{\text{médio}} + 0.00CSII_{\text{alto}}$. A função de mapeamento deste índice e a predição de inteligibilidade é descrita por,

$$I_3 = \frac{100}{1+exp(ac+b)}. \quad (8)$$

onde $a = -10,9$ e $b = 4,65$.

C. STOI

Na STOI [12], o coeficiente de correlação entre os espectros dos sinais limpo e realçado é utilizado para avaliar a degradação da inteligibilidade de algoritmos de redução de ruídos. Primeiramente, o sinal de voz limpo $x(t)$ é reamostrado a 10 kHz e dividido em janelas de Hamming de 256 amostras com 50% de sobreposição. Em seguida, aplica-se uma DFT de 512 pontos em cada quadro, formando a matriz X, onde $X(\kappa,\tau)$ representa o κ-ésimo ponto da DFT do quadro τ. Os pontos $X(\kappa,\tau)$ são então agrupados em 15 sub-bandas de frequência cujo centro variam entre 150 Hz e 4300 Hz. A norma para cada sub-banda é definida por,

$$\bar{X}_j(\tau) = \sqrt{\sum_{\kappa=\kappa_l(j)}^{\kappa_u(j)-1}|X(\kappa,\tau)|}, \quad (9)$$

onde $\kappa_l(j)$ e $\kappa_u(j)$ são, respectivamente, os limites inferior e superior da sub-banda j ($j = 1,2,...,15$). Com os valores das normas, define-se a envoltória temporal de cada sub-banda pelo seguinte vetor:

$$\boldsymbol{x}_{(j,\tau)} = [\bar{X}_j(\tau-29), \bar{X}_j(\tau-28), ..., \bar{X}_j(\tau)]^T. \quad (10)$$

A partir do mesmo processo com o sinal de voz corrompido $y(t)$ obtém-se $\boldsymbol{y}_{(j,\tau)}$. Este é normalizado segundo,

$$\bar{\boldsymbol{y}}_{(j,\tau)} = \min\left(\frac{\|\boldsymbol{x}_{(j,\tau)}\|}{\|\boldsymbol{y}_{(j,\tau)}\|}\boldsymbol{y}_{(j,\tau)}, (1+10^{-\frac{\beta}{20}})\boldsymbol{x}_{(j,\tau)}(n)\right), \quad (11)$$

com $\beta = -15$ dB representando o valor mínimo de SRR.

O valor de $\text{STOI}_{(j,\tau)}$ é dado por:

$$\text{STOI}_{(j,\tau)} = \frac{(\boldsymbol{x}_{(j,\tau)} - \mu_{\boldsymbol{x}_{(j,\tau)}})^T(\bar{\boldsymbol{y}}_{(j,\tau)} - \mu_{\bar{\boldsymbol{y}}_{(j,\tau)}})}{\|\boldsymbol{x}_{(j,\tau)} - \mu_{\boldsymbol{x}_{(j,\tau)}}\|\|(\bar{\boldsymbol{y}}_{(j,\tau)} - \mu_{\bar{\boldsymbol{y}}_{(j,\tau)}}\|}, \quad (12)$$

sendo μ a média do vetor correspondente. Por fim, a medida STOI é calculada a partir da média de todos os valores de $\text{STOI}_{(j,\tau)}$, dados por:

$$\text{STOI} = \frac{1}{15Q}\sum_{j=1}^{15}\sum_{\tau=1}^{Q}\text{STOI}_{(j,\tau)}, \quad (13)$$

onde Q é o número total de quadros.

O mapeamento dos valores da medida STOI com os resultados de inteligibilidade obtidos pelos testes subjetivos é definido pela seguinte função,

$$f(\text{STOI}) = \frac{100}{1+\exp(a\text{STOI}+b)}, \quad (14)$$

onde $a = -13,45$ e $b = 9,36$.

IV. RESULTADOS EXPERIMENTAIS E DISCUSSÃO

Diversos experimentos foram realizados para a avaliação objetiva da inteligibilidade resultante do emprego das máscaras IdBM, IRM e BRM. As medidas foram aplicadas em sinais de voz em diferentes situações de reverberação sem aplicação das máscaras (SM) e após o uso das máscaras IdBM, IRM e BRM. Um subconjunto de 168 locutores da base de voz TIMIT [22] foi selecionado para os experimentos. Cada um dos 128

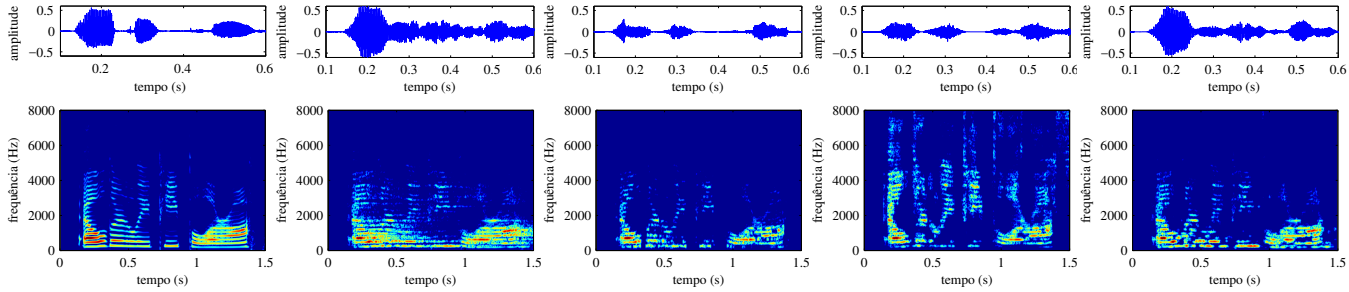

Fig. 1. Sinais de voz e seus respectivos espectrogramas após serem reverberados e com utilização das máscaras acústicas: (a) sem reverberação, (b) voz reverberada, (c) IdBM, (d) IRM, e (e) BRM.

TABELA I
REVERBERAÇÕES SELECIONADAS DA BASE DE DADOS AIR.

Reverberação	RT_{60} (s)	d_{fm} (m)	SRR (dB)
Escritório 1	0,51	1,00	17,58
Escritório 2	0,56	2,00	17,88
Escritório 3	0,59	3,00	17,96
Sala de aula 1	0,79	2,25	16,90
Sala de aula 2	0,82	7,10	15,64
Sala de aula 3	0,83	10,20	21,83

sinais de voz têm duração de 3 segundos e taxa de amostragem de 16 kHz. Estes sinais foram reverberados através de uma convolução com as respostas ao impulso de um subconjunto da base de dados AIR [23]. A Tabela I descreve as condições dos sinais adotadas neste trabalho para reverberar os sinais de voz. As reverberações foram extraídas de duas salas com três valores de distância fonte-microfone (d_{fm}) distintas e foram escolhidas com base nos seus valores de RT_{60} de 0,51 a 0,83 s, uma faixa considerada de média a alta intesidade sonora. Em uma mesma sala, o aumento da d_{fm} faz com que o valor de RT_{60}[1] seja incrementado, provocando um maior efeito da reverberação na inteligibilidade da voz.

A Figura 1 ilustra um sinal de voz em 5 condições: limpo (sem reverberação), após ser reverberado e depois de aplicadas as máscaras IdBM, IRM e BRM. Os testes com a medida SegSNR foram realizados com quadros de 32 ms de duração com sobreposição de 75%. Os valores de SNR de cada quadro foram limitados entre -10 e 35 dB. Os resultados apresentados na Figura 2 indicam que as máscaras aumentaram o valor de SegSNR do sinal de voz com reverberação. A BRM incrementou a inteligibilidade em todas as reverberações, com ganho médio de 1,55 dB. O maior ganho ocorre com a máscada IdBM, com aumento de 2,74 dB. A IRM incrementou o resultado médio em 0,31 dB.

A. Resultados de inteligibilidade: fwSegSNR, CSII, STOI

1) fwSegSNR: A Tabela II mostra os resultados obtidos com a medida fwSegSNR. Pode-se perceber que a BRM obteve o melhor aprimoramento de inteligibilidade, de 0,76 dB, para Sala de aula 3, condição de maior SRR (vide Tabela I).

[1]Tempo necessário para que a RIR decaia em 60 dB.

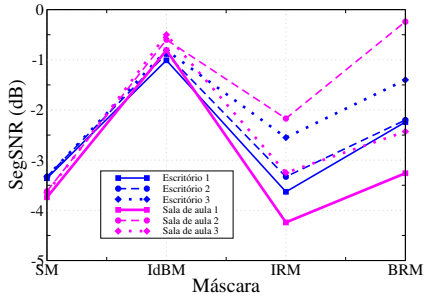

Fig. 2. Resultados de SegSNR para os sinais de voz com reverberação e após a aplicação das máscaras acústicas.

Para esta mesma condição, a melhora é de 0,70 dB para a IRM. Para a IdBM, a medida fwSegSNR apresenta o melhor aprimoramento, de 3,09 dB, também para a Sala de aula 3. Estes resultados confirmam que a fwSegSNR depende dos valores de SRR introduzidos pela reverberação.

TABELA II
RESULTADOS DE FWSEGSNR (DB) PARA OS SINAIS DE VOZ COM REVERBERAÇÃO E APÓS A APLICAÇÃO DAS MÁSCARAS ACÚSTICAS.

Reverberação	SM	IdBM	IRM	BRM
Escritório 1	7,97	8,41	5,79	5,57
Escritório 2	7,20	9,00	5,87	4,86
Escritório 3	6,43	8,39	5,96	4,58
Sala de aula 1	8,30	9,10	4,15	6,07
Sala de aula 2	4,58	7,48	**5,83**	4,86
Sala de aula 3	3,89	**6,98**	4,60	**4,65**

2) CSII: A Figura 3 apresenta os resultados de inteligibilidade obtidos com a medida CSII. Note que a máscara BRM obteve um aprimoramento médio na inteligibilidade em 27,13 p.p. (pontos percentuais), com o maior incremento para a reverberação Escritório 3, de 40,63 p.p.. As máscaras IdBM e IRM melhoraram os resultados em 6,16 p.p. e 34,17 p.p., para as mesmas condições, respectivamente.

Os resultados obtidos sem máscara mostram que o impacto da reverberação na inteligibilidade aumenta com o valor de RT_{60} e a distância d_{fm} em um mesmo ambiente. O aumento da distância d_{fm} em 1 m em Escritório reduz a inteligibilidade em até 30,41 p.p.. Em Sala de aula, os resultados diminuem de 30,47 p.p. para 1,13 p.p. com o crescimento de d_{fm} em

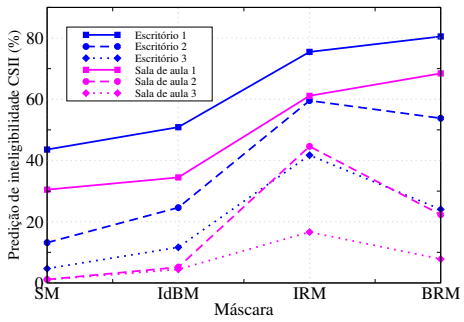

Fig. 3. Predição de inteligibilidade (%) da CSII para as condições SM, IdBM, IRM e BRM.

4,85 m.

3) STOI: A Tabela III ressalta os resultados de inteligibilidade obtidos pela medida STOI. A máscara BRM tem ganho médio de 21,45 p.p. em relação aos testes SM. Seu melhor resultado ocorre em Sala de aula 2, com incremento de 50,49 p.p.. O maior aumento acontece com a máscara IRM, de 35,51 p.p.. A máscara IdBM aumenta o resultado médio em 30,48 p.p..

TABELA III
PREDIÇÃO DE INTELIGIBILIDADE (%) DA STOI PARA AS CONDIÇÕES SM, IDBM, IRM E BRM.

Reverberação	SM	IdBM	IRM	BRM
Escritório 1	**81,30**	81,34	82,60	**89,16**
Escritório 2	46,82	79,51	80,68	75,63
Escritório 3	27,06	74,70	79,99	48,62
Sala de aula 1	84,38	76,55	66,46	75,25
Sala de aula 2	**2,21**	60,83	81,30	**52,70**
Sala de aula 3	0,59	52,34	70,38	29,68

Os resultados de predição para os sinais de voz SM indicam que a inteligibilidade diminuiu quando aumentou-se a d_{fm} em uma mesma sala. Em Escritório, a diminuição da predição de taxa de acerto de palavras foi de até 34,48 p.p. em um distanciamento de 1 m de d_{fm}. Em Sala de aula, a diminuição chegou a 82,17 p.p. com um afastamento de 4,85 m de d_{fm}.

V. CONCLUSÃO

Este artigo apresentou um estudo da inteligibilidade de sinais de voz reverberados e da eficiência de máscaras acústicas ideais e não-ideais em recuperar esta característica. Neste trabalho, foram utilizadas reverberações de duas salas com diferentes distâncias entre fonte e microfone. A influência da reverberação e das máscaras foi analisada a partir de três medidas objetivas de inteligibilidade e uma de qualidade. Os resultados mostraram que, em uma mesma sala, a reverberação diminui a inteligibilidade de acordo com o aumento da distância entre a fonte e o receptor. Além disso, foi mostrado que o uso de máscaras acústicas incrementa a inteligibilidade e a qualidade degradada pelo efeito da reverberação. Os resultados confirmaram que a BRM (não-ideal e cega) é bastante promissora. Vale ressaltar que os resultados de inteligibilidade obtidos para as máscaras ideais IdBM e IRM demonstraram o potencial dos filtros *gammatone* para a detecção do efeito de reverberação obtidos pela máscara não-ideal BRM.

REFERÊNCIAS

[1] R. H. Bolt and A. D. MacDonald, "Theory of speech masking by reverberation," *The Journal of the Acoustical Society of America*, vol. 21, no. 6, pp. 577–580, 1949.

[2] A. T. V. Rabelo, J. N. Santos, R. C. Oliveira, and M. d. C. Magalhaes, "Effect of classroom acoustics on the speech intelligibility of students," *CoDAS*, vol. 26, pp. 360–366, october 2014.

[3] B. Gold and N. Morgan, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. New York, NY, USA: John Wiley & Sons, Inc., 1st ed., 1999.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 145–152, Feb 1988.

[5] K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 1579–1591, July 2007.

[6] P. C. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL, USA: CRC Press, Inc., 2nd ed., 2013.

[7] K. Kokkinakis, O. Hazrati, and P. Loizou, "A channel-selection criterion for suppressing reverberation in cochlear implants," *Journal of the Acoustic Society of America*, vol. 129, pp. 3221–3232, may 2011.

[8] O. Hazrati, J. Lee, and P. C. Loizou, "Binary mask estimation for improved speech intelligibility in reverberant environments," in *INTERSPEECH*, pp. 162–165, ISCA, 2012.

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

[10] J. Ma, Y. Hu, and P. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *Journal of the Acoustic Society of America*, vol. 125, pp. 3387–3405, may 2009.

[11] J. Kates and K. Arehart, "Coherence and the speech intelligibility intex," *Journal of the Acoustic Society of America*, vol. 117, pp. 381–384, april 2005.

[12] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 2125–2136, september 2011.

[13] R. Tavares and R. Coelho, "Speech enhancement with nonstationary acoustic noise detection in time domain," *IEEE Signal Processing Letters*, vol. 23, pp. 6–10, Jan 2016.

[14] L. Zao, R. Coelho, and P. Flandrin, "Speech enhancement with emd and hurst-based mode selection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, pp. 899–911, May 2014.

[15] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica united with Acustica*, vol. 86, pp. 117–128, January 2000.

[16] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, 2006.

[17] N. Li and P. Loizou, "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," *Journal of the Acoustic Society of America*, vol. 123, pp. 1673–1682, march 2007.

[18] R. Patterson, I. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," pp. 357–366, december 1987.

[19] P. I. M. Johannesma, "The pre-response stimulus ensemble of neurons in the cochlear nucleus," pp. 58–69, 1972.

[20] R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," *Frequency selectivity in hearing*, pp. 123–177, 1986.

[21] M. Cooke, *Modelling Auditory Processing and Organisation*. New York, NY, USA: Cambridge University Press, 1993.

[22] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Timit acoustic phonetic continuous speech corpus," 1993.

[23] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*, pp. 1–5, July 2009.