

# Análise do impacto de métodos de redução de ruído na inteligibilidade da fala em implantes cocleares

Gustavo L. Mourão, Maria M. C. Pinheiro, Márcio H. Costa, Stephan Paul e Bernardo H. P. Murta

**Resumo**—Este trabalho apresenta uma análise comparativa de desempenho entre três máscaras tempo-frequência de redução de ruído. O objetivo é a avaliação do impacto desses métodos na inteligibilidade da fala em usuários de implantes cocleares. Para tanto, métricas objetivas e experimentos psicoacústicos foram aplicados em sinais de fala contaminados por ruído aditivo e sujeitos a processamento por vocoder. Testes estatísticos indicam que a técnica de Mínimo Erro Quadrático Médio (MMSE) produz aumento significativo na inteligibilidade da fala em comparação às máscaras binária e de Wiener. Esse fato decorre da característica de preservação de envoltória da máscara MMSE.

**Palavras-Chave**—Implante coclear, Inteligibilidade, Redução de ruído, Vocoder.

**Abstract**—This work presents a comparative performance analysis between three noise reduction time-frequency masks. The aim is the evaluation of their impact on speech intelligibility in cochlear implant users. Therefore, objective metrics and psychoacoustic experiments were applied to noise contaminated speech signals processed by a vocoder. Statistical tests indicate that the Minimum Mean-Square Error (MMSE) technique results in significant speech intelligibility improvements as compared to the Binary and Wiener masks. This fact results from the envelope preservation characteristic of the MMSE mask.

**Keywords**—Cochlear implant, Intelligibility, Noise reduction, Vocoder.

## I. INTRODUÇÃO

A perda auditiva é uma limitação física com grande impacto na vida social de um indivíduo. Embora graus leves e moderados possam ser compensados através de aparelhos de amplificação sonora individual (AASI), perdas auditivas bilaterais de grau severo e profundo requerem a utilização de implantes cocleares.

Implantes cocleares são dispositivos cirurgicamente implantados que transformam a informação acústica em estimulação elétrica aplicada diretamente no nervo auditivo através de um conjunto de eletrodos [1].

Os implantes cocleares atuais são capazes de produzir elevados níveis de inteligibilidade [2]. Entretanto, apesar dos inúmeros avanços, o desempenho do usuário em ambientes ruidosos ainda é um desafio [3] [4]. Estudos têm demonstrado que em ambientes ruidosos o limiar de recepção da fala (*speech reception threshold* - SRT), definido pela relação sinal ruído (*signal to noise ratio* - SNR) necessária para o entendimento correto de 50 % da fala, é tipicamente de 7 a 20 dB mais

elevado em usuários de implantes que em pessoas de audição normal [5] [6]. Dessa forma, a aplicação de métodos de redução de ruído e ênfase da fala é essencial para a melhoria da qualidade de vida de usuários de implantes cocleares.

Métodos de redução de ruído utilizando múltiplos microfones têm apresentado excelentes resultados quando as fontes acústicas de interesse e ruído se encontram espacialmente separadas. Por outro lado, métodos que utilizam um único microfone são esteticamente mais adequados e de fácil incorporação ao módulo externo do implante [4]. Adicionalmente, as restrições na complexidade computacional e no consumo do dispositivo fazem com que sistemas de um único microfone sejam mais interessantes comercialmente.

Diversos trabalhos têm sido publicados na área de redução de ruído para implantes cocleares. A classe das máscaras tempo-frequência tem atraído grande interesse por suas características de implementação e por seu desempenho. Entre elas, a máscara binária e a máscara de Wiener são as mais populares [7] [8]. Apesar dos resultados interessantes obtidos, esses métodos não têm como motivação original a preservação da envoltória dos sinais de interesse (fala). Esse fato pode ser limitante no desempenho global dos implantes cocleares, visto que nesses sistemas a estrutura fina dos sinais não é transmitida ao usuário. Dessa forma, métodos de redução de ruído baseados na preservação da envoltória do sinal de interesse podem apresentar resultados mais relevantes.

O presente trabalho apresenta um estudo comparativo sobre o desempenho de máscaras tempo-frequência, utilizadas para redução de ruído, na inteligibilidade da fala em usuários de implantes cocleares. Para tanto, são utilizados resultados de critérios objetivos associados à inteligibilidade, como também de experimentos psicoacústicos realizados com normo-ouvintes e um sistema vocoder.

O artigo é estruturado da seguinte forma: Na Seção II é apresentada a modelagem do problema estudado e na Seção III as máscaras tempo-frequência analisadas. A Seção IV apresenta os procedimentos metodológicos associados à avaliação de desempenho, como o processo de simulação dos sinais utilizados, a descrição do Vocoder, os critérios objetivos e o método utilizado para a avaliação subjetiva. Em sequência, as Seções V a IV apresentam, respectivamente, os resultados obtidos, a discussão e a conclusão do trabalho.

Ao longo desse documento, letras maiúsculas e minúsculas em negrito representam, respectivamente, matrizes e vetores, enquanto letras em itálico referem-se a variáveis escalares.

## II. PROCESSAMENTO DE SINAIS EM IMPLANTES COCLEARES

Em um implante coclear, o fluxo de processamento do sinal consiste basicamente na captação do sinal acústico seguida por uma decomposição espectral realizada através de um banco de filtros ou da Transformada de Fourier de Curto Termo (*Short Time Fourier Transform - STFT*). No caso da STFT, assumindo-se quase-estacionariedade [9], o sinal é descrito como:

$$x(\lambda, k) = s(\lambda, k) + v(\lambda, k) \quad (1)$$

em que  $s(\lambda, k)$  é a STFT do sinal de fala relativo ao *frame*  $\lambda$  e bin  $k$ ,  $v(\lambda, k)$  é a STFT do ruído e  $x(\lambda, k)$  é a STFT do sinal de fala contaminado. Cada bin  $k$  é então submetido individualmente a um processo de redução de ruído que se dá pela aplicação de um fator de atenuação. Esse fator é definido em função da SNR estimada (em cada *bin*), de forma que:

$$y(\lambda, k) = G(\lambda, k)x(\lambda, k), \quad (2)$$

em que  $y(\lambda, k)$  é uma estimativa de  $s(\lambda, k)$ .

## III. MÉTODOS DE REDUÇÃO DE RUÍDO

Os métodos de redução de ruído baseados em máscaras têm sido amplamente estudados na literatura de aparelhos auditivos, sendo as máscaras binária e de Wiener as mais utilizadas. A máscara *Minimum Mean Square Error* (MMSE), pouco explorada em implantes cocleares, também é apresentada neste trabalho.

### A. Máscara Binária

A máscara binária [10] apresenta um fator de atenuação:

$$G_{MB}(\lambda, k) = \begin{cases} 1, & \xi(\lambda, k) \geq \xi_o(k) \\ 0, & \xi(\lambda, k) < \xi_o(k) \end{cases}, \quad (3)$$

em que  $\xi(\lambda, k)$  é a estimativa da SNR *a priori* [11] e  $\xi_o(k)$  é um limiar previamente definido, geralmente sendo um valor fixo de 0 dB.

### B. Máscara de Wiener

O fator de atenuação de Wiener é obtido através da minimização do erro quadrático médio entre os espectros estimado e real do sinal de interesse [9], sendo definido como:

$$G_{FW}(\lambda, k) = \frac{\xi(\lambda, k)}{1 + \xi(\lambda, k)}. \quad (4)$$

### C. Máscara MMSE

Diferentemente do filtro de Wiener, o fator de atenuação associado à máscara MMSE é obtido através da minimização do valor quadrático da diferença entre as magnitudes, estimada e real, do sinal de fala. Esse método requer não apenas a informação da SNR *a priori*, como também da SNR *a posteriori*, sendo calculado como:

$$G_{MMSE}(\lambda, k) = \frac{\sqrt{\pi} \sqrt{v(\lambda, k)}}{2 \gamma(\lambda, k)} e^{-v(\lambda, k)/2} \times \left\{ [1 + v(\lambda, k)] I_0\left(\frac{v(\lambda, k)}{2}\right) + v(\lambda, k) I_1\left(\frac{v(\lambda, k)}{2}\right) \right\} \quad (5)$$

em que  $I_0(\cdot)$  e  $I_1(\cdot)$  são as funções modificadas de Bessel, de ordem zero e um, respectivamente, enquanto que  $v(\lambda, k)$  é definido como:

$$v(\lambda, k) = \frac{\xi(\lambda, k)}{1 + \xi(\lambda, k)} \gamma(\lambda, k) \quad (6)$$

em que  $\gamma(\lambda, k)$  é a SNR *a posteriori* [11].

### D. Estimação da SNR

Neste trabalho, a estimação das SNRs é realizada através do método *Decision-Directed Approach* modificado [9], sendo que a SNR *a priori* é definida como:

$$\xi(\lambda, k) = \max \left[ a \frac{|y(\lambda-1, k)|^2}{\hat{\sigma}_v^2(\lambda-1, k)} + (1-a) \max[\gamma(\lambda, k) - 1, 0], \xi_{\min} \right] \quad (7)$$

em que  $\hat{\sigma}_v^2(\lambda, k)$  é a estimativa da variância do ruído  $v(\lambda, k)$  no *frame*  $\lambda$  e bin  $k$ , obtida conforme [12],  $a = 0,98$  é um fator de ponderação e  $\xi_{\min} = 0,0316$  (-15 dB) é um patamar mínimo permitido para a estimação de  $\xi(\lambda, k)$ . A SNR *a posteriori* é estimada como

$$\gamma(\lambda, k) = \frac{|x(\lambda, k)|^2}{\hat{\sigma}_v^2(\lambda, k)}. \quad (8)$$

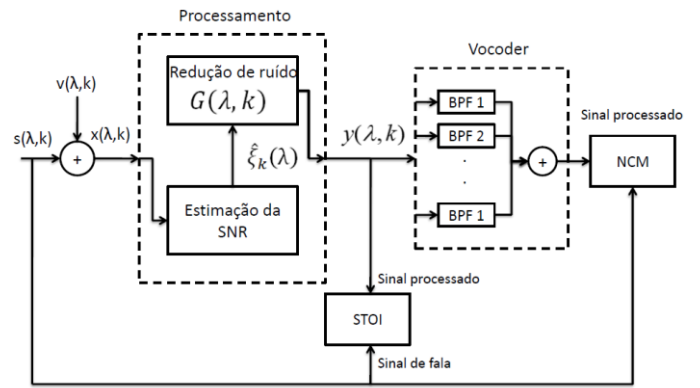


Fig. 1. Diagrama em blocos do processo de avaliação objetiva das máscaras de redução de ruído.

## IV. AVALIAÇÃO

A Fig. 1 apresenta o diagrama em blocos do processo de avaliação objetiva das máscaras de redução de ruído. A metodologia apresentada teve como objetivo a análise comparativa da inteligibilidade da fala em usuários de implantes cocleares. Para tanto, foram escolhidas as três máscaras descritas na Seção III.

Os sinais de fala foram obtidos a partir do banco de dados apresentado em [13], no qual as sentenças foram gravadas em português e organizadas em 20 listas com 10 áudios cada. O ruído do tipo *babble* [14] foi obtido de [9]. Todos os sinais foram reamostrados para uma frequência de amostragem de 16 kHz.

Foi definido um cenário acústico simulando a origem dos sinais de fala e ruído a uma distância de 3 m do usuário de implante, estando a fala diretamente a frente do usuário (azimute de 0°) e o ruído no lado direito (azimute de +90°), ambos no plano horizontal. A composição do cenário acústico foi obtida artificialmente através da convolução dos sinais com respostas impulsivas relativas à cabeça gravadas em câmara anecoica [15]. Definiu-se o uso de um implante monocanal com o microfone de recepção situado em um molde BTE sobre

o pavilhão auditivo direito.

Após contaminação aditiva, para uma faixa SNRs de  $-12$  dB a  $12$  dB, em passos de  $2$  dB, o sinal de fala contaminado foi transformado para o domínio frequência através de uma Transformada Discreta de Fourier com  $128$  bins. Foi utilizado janelamento do tipo Hanning e sobreposição de  $50\%$  [7]. As SNRs *a priori* e *a posteriori* foram estimadas de acordo com os métodos apresentados na Seção III.D. Na sequência foi aplicado o fator de atenuação de cada uma das máscaras descritas na Seção III. A reconstrução dos sinais filtrados no domínio tempo foi realizada pelo método *Weighted Overlap and Add* [16].

#### A. Vocoder

O vocoder é um dos métodos amplamente utilizado para análise da inteligibilidade em implantes cocleares [9]. Sua grande vantagem é a possibilidade de realizar inferências através de experimentos psicoacústicos com voluntários normo-ouvintes. Por outro lado, as limitações desse método são amplamente conhecidas, visto que a sensação auditiva evocada é estimada através de conceitos de processamento de sinais em detrimento das características fisiológicas auditivas reais [17].

Neste trabalho foi implementado o vocoder descrito em [18]. O sinal acústico é decomposto em  $22$  canais por um banco de filtros do tipo gammatone [19]. As frequências centrais desses filtros variam de  $250$  Hz a  $7438$  Hz, de acordo com o modelo de funcionamento da cóclea [20]. Em sequência a envoltória de cada canal é estimada e submetida à estratégia *Advanced Combination Encoder* (ACE) [21]. Essa estratégia seleciona, a cada *frame*  $\lambda$ , um subconjunto de  $8$  canais que apresentam as maiores magnitudes de envoltória para estimulação dos eletrodos, sendo então aplicado um trem de pulsos com uma taxa média de  $1000$  pulsos por segundo.

Com a finalidade de emular o espalhamento espacial do campo elétrico relacionado à distância entre os eletrodos na perilínfa da escala timpânica [22], cada pulso é multiplicado por uma função exponencial decrescente dada por  $\exp(-|d|/0,009)$ , em que  $|d|$  representa a distância absoluta entre o eletrodo de estimulação e os demais. A distância intereletrodo assumida foi de  $d = 0,7$  mm. Finalizando, cada canal é filtrado por um banco de filtros do tipo gammatone de terceira ordem com frequências centrais entre  $664$  Hz e  $11950$  Hz. A finalidade desse segundo processo de filtragem é simular o mapeamento de cada canal à respectiva localização do eletrodo [22].

#### B. Medidas Objetivas de Inteligibilidade

Duas medidas objetivas de inteligibilidade foram utilizadas para comparar o desempenho dos métodos de redução de ruído sob análise. A métrica *Short-Time Objective Intelligibility*, também conhecida como STOI [23], apresenta alta correlação com predições de inteligibilidade em usuários de implantes cocleares [24]. A segunda métrica utilizada foi a *Normalized Covariance Measure* (NCM) [25] que apresenta alta correlação com a predição de inteligibilidade em experimentos com vocoder em normo-ouvintes [26]. Ambas são métricas intrusivas, de forma que o sinal de fala limpo foi utilizado como sinal de referência.

As simulações utilizando STOI foram realizadas com os sinais reconstruídos no domínio tempo através do método *Weighted Overlap and Add*, conforme visualizado na Fig. 1, enquanto que as simulações da inteligibilidade usando NCM foram realizadas com os sinais reconstruídos pelo vocoder.

#### C. Experimentos Psicoacústicos com Voluntários

De forma a corroborar os resultados obtidos através das métricas objetivas, foram realizados experimentos psicoacústicos com voluntários. Foram selecionados  $4$  [27] indivíduos normo-ouvintes, caracterizados pela inexistência de histórico ou reclamações em relação a limitações de audição. Todos os voluntários eram homens e falantes nativos da língua portuguesa, apresentando média de  $27$  anos de idade e desvio padrão de  $2,6$ .

Foram apresentadas a cada voluntário sequências de áudios processados relativos às técnicas analisadas (MMSE, Wiener, máscara binária e não-processado). Para tanto, foram avaliados  $4$  níveis de SNR:  $0$  dB,  $2,5$  dB,  $5$  dB e  $7,5$  dB. Para cada SNR foram apresentados  $6$  áudios. Os sinais processados foram normalizados na mesma potência do sinal contaminado para compensação da diminuição de potência decorrente da aplicação das máscaras tempo-frequência. Todas as sentenças foram aleatoriamente misturadas. Foi utilizado um fone de ouvido marca Sennheiser, modelo HD 20, em um nível confortável ao usuário.

Durante a sessão, os voluntários foram instruídos a repetir verbalmente a sentença apresentada. Após a finalização dos experimentos a taxa de palavras corretas foi calculada dividindo-se o número de palavras corretamente identificadas pelo total de palavras nas sentenças.

Os experimentos com voluntários foram aprovados pelo Comitê de Ética em Pesquisa com Seres Humanos através do certificado CEP-UFSC número 49741615.2.0000.0121.

## V. RESULTADOS

Nessa seção são apresentados os resultados da aplicação das métricas objetivas STOI e NCM, como também os resultados de experimentos subjetivos.

Nas Figs. 2 e 3 são apresentados os resultados relativos às métricas NCM e STOI, respectivamente. Na abscissa é apresentada a SNR e na ordenada o valor da métrica, no qual  $1$  corresponde ao índice de inteligibilidade de  $100\%$ .

As Figs. 4-6 apresentam os espectrogramas (fala, fala contaminada processada pelo MMSE e fala contaminada processada pela máscara binária) da frase “Hoje dormirei bem” para SNR =  $0$  dB. Cada trecho corresponde a um *frame* de  $8$  ms e os canais são definidos de acordo com o filtro gammatone utilizado. As altas frequências estão relacionadas aos canais posicionados na região basal da cóclea (índices maiores), enquanto que as baixas frequências associam-se à região apical (índices menores).

A Fig. 7 apresenta diagramas de caixa para os dados obtidos nos experimentos psicoacústicos. Foram avaliados  $4$  voluntários e utilizado um total de  $36$  palavras, em média, para cada máscara e SNR. O conjunto dos valores da amostra compreendidos entre o  $1^\circ$  e o  $3^\circ$  quartis (denominados  $q_1$  e  $q_3$ , respectivamente) é representado por um retângulo cuja mediana é indicada por uma barra em seu interior. Se os valores das amostras forem maiores do que  $q_3 + \varpi(q_3 - q_1)$  ou menores do que  $q_3 - \varpi(q_3 - q_1)$ , considerando  $q_1$  e  $q_3$  como sendo os valores percentuais de  $25\%$  e  $75\%$ , respectivamente, eles serão considerados artefatos. Para a variável  $\varpi$  foi definida o valor padrão de  $1,5$  [28].

O teste de hipótese t de Student foi utilizado para verificação da diferença estatística entre os resultados obtidos com os métodos STOI e NCM para as três máscaras avaliadas.

Foi verificada diferença significativa para as SNRs negativas.

## VI. DISCUSSÃO

Duas métricas objetivas, NCM e STOI, foram utilizadas para a previsão do percentual de inteligibilidade em aplicações de implantes cocleares. Note-se, que os resultados numéricos obtidos não correspondem diretamente aos índices percentuais de inteligibilidade obtidos em testes psicoacústicos, podendo apenas ser interpretados de forma relativa [4].

A aplicação do teste de hipótese t de Student nos resultados das métricas STOI e NCM resultou em diferença significativa ( $p < 0,05$ ) de inteligibilidade entre as três máscaras e o sinal não processado para SNRs negativas. O método MMSE resultou, consistentemente, em maior inteligibilidade que as máscaras binária e de Wiener. O motivo desse resultado pode ser explicado pela característica intrínseca de preservação da envoltória da fala limpa, pela máscara MMSE, como pode ser verificado através das Figs. 4-6. Comparando-se os espectrogramas da fala limpa e da fala contaminada processada pela máscara binária verifica-se que a máscara binária resulta em mudanças consideráveis na distribuição de magnitudes do sinal. O espectrograma do sinal contaminado processado pela máscara de Wiener é muito semelhante ao do MMSE e, portanto, não é apresentado.

Os resultados dos experimentos psicoacústicos corroboram os obtidos através das métricas objetivas. Consistentemente, a máscara MMSE produziu taxas de reconhecimento de palavras maiores que as demais máscaras para todos os SNRs analisados.

Diversos trabalhos têm sido publicados na literatura analisando o desempenho de técnicas para redução de ruído em implantes cocleares [7], [8]. Todavia, poucos trabalhos têm abordado o desempenho inerente à máscara MMSE [29], sobretudo em relação ao contexto proposto neste trabalho. A partir dos resultados obtidos nesse trabalho é possível inferir que uma maior atenção deve ser dada à máscara MMSE, em detrimento do filtro de Wiener e da máscara binária.

## VII. CONCLUSÕES

Esse trabalho apresentou uma análise comparativa de desempenho entre métodos de redução de ruído, baseados em máscaras tempo-frequência, para aplicação em implantes cocleares. Foram utilizadas métricas objetivas e experimentos psicoacústicos com voluntários normo-ouvintes em sinais processados por um vocoder. Testes estatísticos indicam que o método MMSE resulta em maior inteligibilidade da fala em comparação às máscaras binária e de Wiener. Esse resultado é sustentado pela característica intrínseca de preservação da envoltória do MMSE, sendo essa a característica principal utilizada na codificação do sinal sonoro pelo implante coclear.

## REFERÊNCIAS

- [1] F.-G. Zeng, et al., Cochlear implants: system design, integration, and evaluation, *IEEE Reviews in Biomedical Engineering*, v. 1, pp. 115-142, 2008.
- [2] B. S. Wilson, M. F. Dorman, The surprising performance of present-day cochlear implants, *IEEE Transactions on Biomedical Engineering*, v. 54, no. 6, pp. 969-972, 2007.
- [3] V. Gopalakrishna, et al., Real-time automatic tuning of noise suppression algorithms for cochlear implant applications, *IEEE Transactions on Biomedical Engineering*, v. 59, no. 6, pp. 1691-1700, 2012.
- [4] Y.-H. Lai, et al., A deep denoising autoencoder approach to improving the intelligibility of vocoded speech in cochlear implant simulation, *IEEE Transactions on Biomedical Engineering*, v. 64, no. 7, pp. 1568-1578, 2017.
- [5] D. Wang, J. H. L. Hansen, Speech enhancement based on harmonic estimation combined with MMSE to improve speech intelligibility for cochlear implant recipients, in *Interspeech*, pp. 186-190, 2017.
- [6] R. Koning, et al., Perceptual and model-based evaluation of ideal time-frequency noise reduction in hearing-impaired listeners, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, v. 26, no. 3, pp. 687-697, 2018.
- [7] R. Koning, N. Madhu, J. Wouters, Ideal time-frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners, *IEEE Transactions on Biomedical Engineering*, v. 62, no. 1, pp. 331-341, 2015.
- [8] N. Madhu, et al., The potential for speech intelligibility improvement using the ideal binary mask and the ideal wiener filter in single channel noise reduction systems: Application to auditory prostheses, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 21, no. 1, pp. 63-72, 2013.
- [9] P.C. Loizou. *Speech enhancement: theory and practice*. CRC press, 2013.
- [10] D. Wang, *On ideal binary mask as the computational goal of auditory scene analysis, Speech separation by humans and machines*. Springer, Boston, MA, pp.181-197, 2005.
- [11] Y. Ephraim, D. Malah, Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator, *IEEE Transactions on acoustics, speech, and signal processing*, v. 32, no.6, pp. 1109-1121, 1984.
- [12] X. Li, et al., Non-stationary noise power spectral density estimation based on regional statistics, in *ICASSP*, pp.181-185, 2016.
- [13] A. Alcain, J. A. Solewicz, J. A. Moraes, Frequência de ocorrência dos fonos e listas de frases foneticamente balanceadas no português falado no Rio de Janeiro, *Revista da Sociedade Brasileira de Telecomunicações* v. 7, no. 1, pp. 23-41, 1992.
- [14] N. Krishnamurthy, J. H. L. Hansen, Babble noise: modeling, analysis, and applications, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, no. 7, pp. 1394-1407, 2009
- [15] H. Kayser, et al., Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses, *EURASIP Journal on Advances in Signal Processing*, v. 2009, pp. 1-10, 2009.
- [16] R. Crochiere, A weighted overlap-add method of short-time Fourier analysis/synthesis, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 28, no. 1, pp. 99-102, 1980.
- [17] N. El Boghdady, et al., A neural-based vocoder implementation for evaluating cochlear implant coding strategies, *Hearing research*, v. 333, pp. 136-149, 2016.
- [18] F. Chen, P. C. Loizou. Predicting the intelligibility of vocoded speech, *Ear and Hearing*, v. 32, no. 3, pp. 331-338, 2011.
- [19] V. Hohmann, Frequency analysis and synthesis using a Gammatone filterbank, *Acta Acustica united with Acustica*, v. 88, no. 3, pp. 433-442, 2002.
- [20] W. Nogueira, et al., A psychoacoustic N of M-type speech coding strategy for cochlear implants, *EURASIP Journal on Applied Signal Processing*, v. 2005, pp. 3033-3059, 2005.
- [21] J. Wouters, H. J. McDermott, T. Francart, Sound coding in cochlear implants: from electric pulses to hearing, *IEEE Signal Processing Magazine*, v. 32, no.2 pp. 67-80, 2015.
- [22] F. Langner, T. Jürgens, Forward-masked frequency selectivity improvements in simulated and actual cochlear implant users using a preprocessing algorithm, *Trends in hearing*, v. 20, pp. 1-14, 2016.
- [23] C. H. Taal, et al, An algorithm for intelligibility prediction of time-frequency weighted noisy speech, *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, no. 7, pp. 2125-2136, 2011.
- [24] T. H. Falk, et al, Objective quality and intelligibility prediction for users of assistive listening devices: advantages and limitations of existing tools, *IEEE signal processing magazine*, v. 32, no. 2, pp. 114-124, 2015.
- [25] J. Ma, Y. Hu, P. C. Loizou, Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions, *The Journal of the Acoustical Society of America*, v. 125, no. 5, pp. 3387-3405, 2009.
- [26] F. Chen, P. C. Loizou, Predicting the intelligibility of vocoded and wideband Mandarin Chinese, *The Journal of the Acoustical Society of America*, v. 129, no. 5, pp. 3281-3290, 2011.
- [27] H. Jiang, Q.-J. Fu, Statistical noise compensation for cochlear implant processing, in *Interspeech*, pp. 2085-2088, 2005.

- [28] D. Shannon, Box-and-whisker plots with the SAS, Pharmaceutical Statistics, v. 2, pp. 291-295, 2003.
- [29] H. Jiang, Q. J. Fu, Statistical noise compensation for cochlear implant processing, Ninth European Conference on Speech Communication and Technology, 2005.

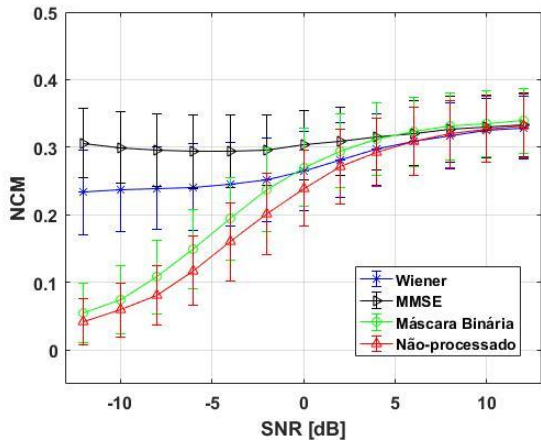


Fig. 2. Inteligibilidade segundo a métrica objetiva NCM.

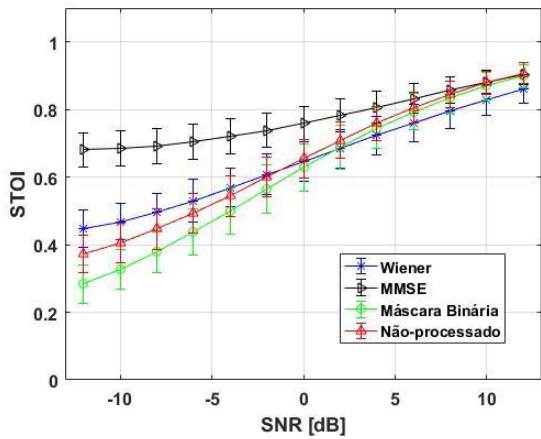


Fig. 3. Inteligibilidade segundo a métrica objetiva STOI.

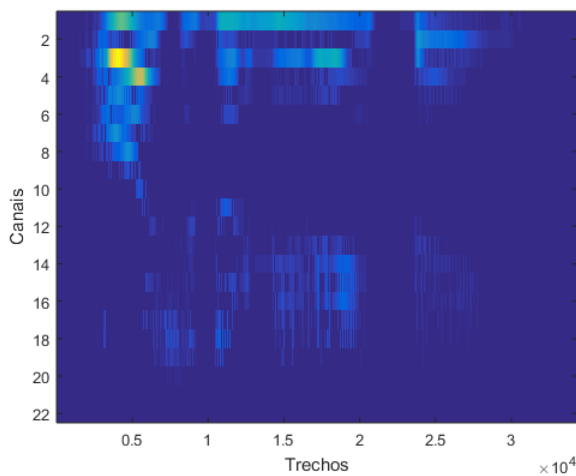


Fig. 4. Espectrograma do sinal de fala limpo. SNR = 0dB.

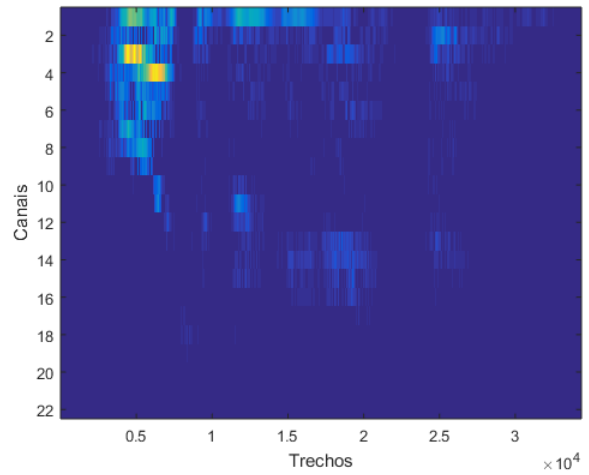


Fig. 5. Espectrograma do sinal contaminado processado pela máscara MMSE. SNR = 0dB.

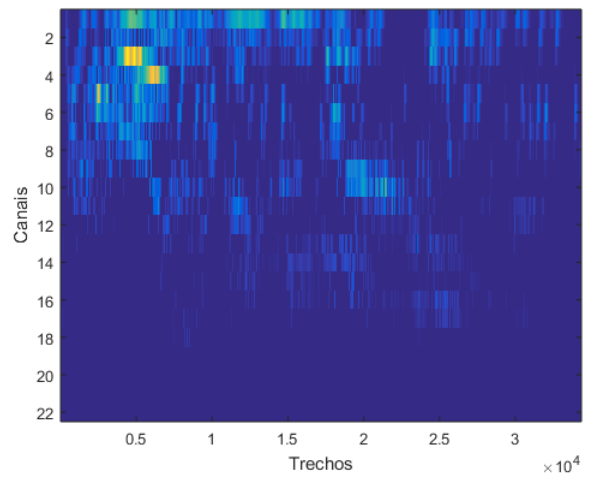


Fig. 6. Espectrograma do sinal contaminado processado pela máscara binária. SNR = 0dB.

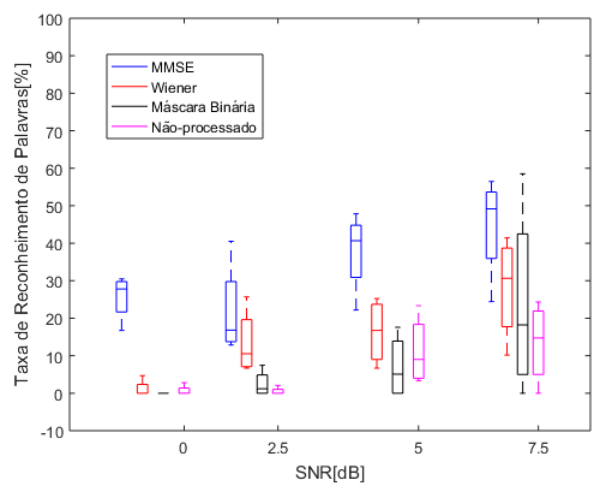


Fig. 7. Taxa de reconhecimento de palavras resultante do experimento psicoacústico com voluntários normo-ouvintes e sinais processados por vocoder.