

# Modelagem do Efeito da Redução da Taxa de Comunicação em Aparelhos Auditivos Biauriculares

Vitor P. Curtarelli<sup>1</sup>, Márcio H. Costa<sup>2</sup>

Programa de Pós-Graduação em Engenharia Elétrica, UFSC, Florianópolis, SC

**Resumo.** Aparelhos auditivos biauriculares têm como objetivo a compensação de limitações na audição humana. Consistem em dois dispositivos de amplificação sonora posicionados nas orelhas, que trocam informações entre si, de forma a propiciar maior inteligibilidade e qualidade da fala. Este trabalho tem como objetivo apresentar um novo modelo matemático para a predição de desempenho do processo de redução de ruído quando a taxa de comunicação entre os dois aparelhos é diminuída para economizar energia. O modelo convencional para a quantização é generalizado para incluir também a saturação. Resultados de simulação indicam maior acurácia na capacidade de predição de desempenho.

**Palavras-chave.** Aparelho auditivo, Redução de ruído, Taxa de comunicação, Quantização, Saturação.

## 1 Introdução

Aparelhos auditivos são dispositivos cujo propósito é a compensação de limitações na audição humana. Apesar de sua elevada eficiência em condições controladas, a existência de ruído ambiente pode acarretar grandes prejuízos à inteligibilidade e ao conforto acústico do usuário. Dessa forma, a inclusão de um sistema de redução de ruído é de grande importância para possibilitar a comunicação em situações cotidianas.

Aparelhos auditivos biauriculares (BHAs, do inglês “Binaural Hearing Aids”) consistem em um par de aparelhos auditivos (um em cada orelha) que se comunicam, possibilitando a troca de sinais e informações [1, 2]. Essa característica permite a utilização de métodos de redução de ruído mais eficientes. Entretanto, como principal desvantagem, apresentam maior consumo de energia [3] em função da necessidade de transmissão sem fio dos dados entre os aparelhos. Dessa forma, estabelece-se um compromisso entre o desempenho do processo de redução de ruído e a duração da bateria.

O filtro de Wiener multicanal (MWF, do inglês “Multichannel Wiener Filter”) tem sido amplamente estudado na literatura como método de redução de ruído para BHAs. Ele consiste em um par de filtros lineares que processam os sinais captados pelos microfones de ambos os aparelhos. O cálculo dos coeficientes desses filtros depende da estimação dos momentos estatísticos de segunda ordem dos sinais envolvidos.

O efeito da redução da taxa de comunicação em BHAs (realizado através da redução do número de bits das palavras de dados, ou seja, do processo de quantização) sobre os coeficientes ótimos do MWF tem sido estudado por um modelo baseado em ruído aditivo com distribuição uniforme [4, 5, 6, 7]. A grande limitação dessa abordagem é a inexistência da previsão de ocorrência de

---

<sup>1</sup>vitor.curtarelli@gmail.com

<sup>2</sup>costa@eel.ufsc.br

saturação, o que pode ocasionar erros de predição. Neste trabalho, é apresentado um novo modelo matemático de predição para os coeficientes do MWF sujeito à limitações na taxa de comunicação. A principal contribuição é a previsão do efeito da saturação dos dados quando o número de bits é reduzido. Essa situação é de grande interesse pelo projetista para aumentar o tempo entre recargas das baterias, mantendo a eficiência do processo de redução de ruído.

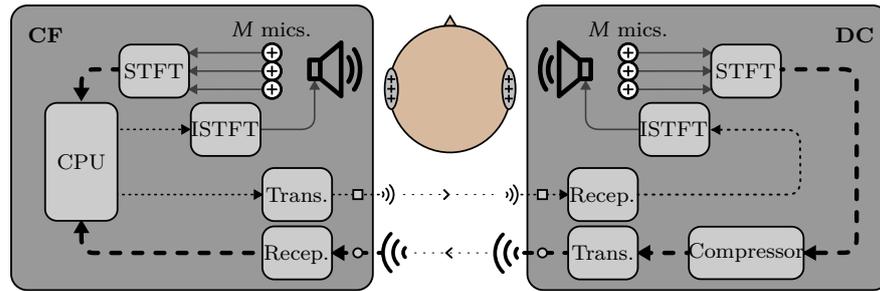
## 2 Redução de Ruído em Aparelhos Auditivos Biauriculares

Nesta seção é apresentada a descrição do problema abordado e o modelo clássico para a descrição do efeito da quantização.

### 2.1 Aparelhos auditivos biauriculares

A Figura 1 apresenta o diagrama de blocos de um aparelho auditivo biauricular, composto por dois dispositivos: o centro de fusão (CF) na orelha esquerda, responsável por processar os sinais captados; e o dispositivo contralateral (DC) na orelha direita, cuja função é transmitir informações ao CF e receber e reproduzir o sinal enviado pelo CF. Cada dispositivo conta com  $M$  microfones, para aquisição de dados.

**Figura 1:** Esquemático do sistema biauricular para o problema proposto.



Os sinais acústicos captados pelos microfones são discretizados e processados pela transformada de Fourier de tempo curto (STFT, do inglês “Short-Time Fourier Transform”), ou outra técnica de análise tempo-frequência, tal que:

$$y_m(\lambda, k) = x_m(\lambda, k) + r_m(\lambda, k), \quad (1)$$

onde:  $(\lambda, k)$  são os índices de tempo e frequência;  $y_m$  é a fala contaminada no microfone  $m$ ;  $r_m$  é o ruído global; e  $x_m$  é o sinal desejado (fala). Assume-se que  $x_m(\lambda, k)$  e  $r_m(\lambda, k)$  sejam decorrelacionados [8] e de média nula [9]. Os sinais adquiridos podem ser organizados vetorialmente, tal que  $\mathbf{y}(\lambda, k) = [y_1(\lambda, k) \cdots y_M(\lambda, k) \ y_{M+1}(\lambda, k) \cdots y_{2M}(\lambda, k)]^T$ , o mesmo para  $\mathbf{x}(\lambda, k)$  e  $\mathbf{r}(\lambda, k)$ ; em que  $(\cdot)^T$  é o operador transposto. Os primeiros  $M$  microfones correspondem ao CF e os últimos  $M$  ao DC.

A taxa de comunicação entre os dispositivos é definida como o número de bits transmitidos por segundo. Para uma determinada frequência de amostragem, a diminuição da taxa de comunicação é obtida pela redução do número de bits da palavra de dados transmitida.

### 2.2 Filtro de Wiener multicanal

O filtro de Wiener multicanal (MWF) é uma das estratégias mais estudadas para redução de ruído em aparelhos auditivos [10]. O processo de filtragem é dado por  $\hat{x}_F(\lambda, k) = \mathbf{w}_F^H(\lambda, k)\mathbf{y}_q(\lambda, k)$ , onde:  $\hat{x}_F(\lambda, k)$  é a representação tempo-frequência do sinal filtrado do CF (posteriormente sendo

reconstruído no tempo e enviado ao alto-falante) e  $\hat{x}_C(\lambda, k) = \mathbf{w}_C^H(\lambda, k)\mathbf{y}_q(\lambda, k)$  é o sinal equivalente no DC.  $\mathbf{w}_F$  e  $\mathbf{w}_C$  são os vetores de coeficientes do CF e DC, respectivamente, e  $\mathbf{y}_q(\lambda, k)$  é o vetor de sinais contaminados disponível no CF (após a transmissão pelo DC).  $(\cdot)^H$  é o operador conjugado transposto. Os coeficientes ótimos do MWF são obtidos pela minimização de  $J(\lambda, k) = E[|x_F(\lambda, k) - \hat{x}_F(\lambda, k)|^2] + E[|x_C(\lambda, k) - \hat{x}_C(\lambda, k)|^2]$ , em que  $E[\cdot]$  é o operador valor esperado, com o qual se obtém

$$\mathbf{w}_F^*(\lambda, k) = \mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}^{-1}(\lambda, k)\mathbf{R}_{\mathbf{x}\mathbf{y}_q}(\lambda, k)\mathbf{q}_F, \quad (2)$$

e o mesmo para  $\mathbf{w}_C^*(\lambda, k) = \mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}^{-1}(\lambda, k)\mathbf{R}_{\mathbf{x}\mathbf{y}_q}(\lambda, k)\mathbf{q}_C$ ; no qual  $\mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}(\lambda, k) = E[\mathbf{y}_q(\lambda, k)\mathbf{y}_q^H(\lambda, k)]$  é a matriz de coerência de  $\mathbf{y}_q(\lambda, k)$ ;  $\mathbf{R}_{\mathbf{x}\mathbf{y}_q}(\lambda, k)$  é a matriz de coerência cruzada entre  $\mathbf{x}(\lambda, k)$  e  $\mathbf{y}_q(\lambda, k)$ ; e  $\mathbf{q}_F$  é o vetor de seleção do sinal de referência do CF (idem para  $\mathbf{q}_C$ ).

### 2.3 Modelo clássico para quantização

O modelo clássico de quantização apresentado em [11] assume a inexistência de saturação. Caso o passo de quantização  $\Delta$  seja suficientemente pequeno ou seja usado *dithering* aditivo [6, 7] é possível assumir que  $\mathbf{y}_q(\lambda, k) \approx \mathbf{y}_\Delta(\lambda, k) = \mathbf{y}(\lambda, k) + \mathbf{e}(\lambda, k)$ , em que  $\mathbf{e}(\lambda, k) = [0 \ \cdots \ 0 \ e_{M+1}(\lambda, k) \ \cdots \ e_{2M}(\lambda, k)]^T$  é o vetor erro de quantização, cujos elementos referentes aos  $M$  sinais do DC podem ser considerados variáveis aleatórias brancas, com distribuição uniforme, média zero, descorrelacionadas de  $\mathbf{y}(\lambda, k)$  e com variância  $\sigma_e^2(\lambda, k) = \Delta^2/6$  (caso complexo). Dessa forma, estimativas  $\bar{\mathbf{R}}_{\mathbf{x}\mathbf{y}_q}(\lambda, k)$  e  $\bar{\mathbf{R}}_{\mathbf{y}_q\mathbf{y}_q}(\lambda, k)$  podem ser obtidas através de

$$\bar{\mathbf{R}}_{\mathbf{x}\mathbf{y}_q}(\lambda, k) \triangleq \mathbf{R}_{\mathbf{x}\mathbf{y}_\Delta}(\lambda, k) = \mathbf{R}_{\mathbf{x}\mathbf{x}}(\lambda, k), \quad (3a)$$

$$\bar{\mathbf{R}}_{\mathbf{y}_q\mathbf{y}_q}(\lambda, k) \triangleq \mathbf{R}_{\mathbf{y}_\Delta\mathbf{y}_\Delta}(\lambda, k) = \mathbf{R}_{\mathbf{y}\mathbf{y}}(\lambda, k) + \mathbf{R}_{\mathbf{e}\mathbf{e}}(\lambda, k), \quad (3b)$$

Substituindo a Equação (3) na Equação (2) obtém-se os coeficientes do MWF, a partir dos quais pode-se avaliar critérios de desempenho relativos à inteligibilidade, qualidade e conforto acústico.

## 3 Modelo Proposto

Nesta seção é apresentado um novo modelo para a estimação das matrizes de coerência  $\mathbf{R}_{\mathbf{y}_q\mathbf{y}_q}(\lambda, k)$  e  $\mathbf{R}_{\mathbf{x}\mathbf{y}_q}(\lambda, k)$  que generaliza o modelo clássico com a inclusão do efeito da saturação.

### 3.1 Modelagem para quantização e saturação

Considera-se um processo de quantização uniforme, na forma *midtread quantization* [6]. Para  $N$  bits obtém-se 1 nível zero e  $\mathcal{N} = 2^{N-1} - 1$  níveis positivos (idem negativos). O limiar de saturação é dado por  $V = \Delta\mathcal{N}$ , em que  $\Delta$  é o passo de quantização. A partir dessa descrição pode-se definir o processo de quantização (com saturação) por:

$$\mathcal{L}_{\Delta, \mathcal{N}}\{z\} = \Delta \sum_{i=1}^{\mathcal{N}} [u(z - [i - 1/2]\Delta) - u(-z - [i - 1/2]\Delta)], \quad (4)$$

onde:  $u(\cdot)$  é a função degrau unitário. A Equação (4) pode ser descrita como:

$$\mathcal{L}_{\Delta, \mathcal{N}}\{z\} = \mathcal{S}_V\{\mathcal{Q}_\Delta\{z\}\}, \quad (5)$$

onde:  $\mathcal{S}_V\{z\} = \text{sgn}(z) \cdot \min(|z|, V)$ ,  $\text{sgn}(\cdot)$  é a função sinal,  $\min(\cdot)$  é o mínimo entre os argumentos e  $\mathcal{Q}_\Delta\{z\}$  é a quantização sem limitação ( $\mathcal{L}_{\Delta, \mathcal{N}}\{z\}$  para  $\mathcal{N} \rightarrow \infty$ ). Considerando que os dados transmitidos do DC ao CF são valores complexos, é possível representá-los como:

$$\mathbf{y}_{qm}(\lambda, k) \triangleq \mathcal{L}_{\Delta, \mathcal{N}}^C\{y_m(\lambda, k)\} \approx \mathcal{S}_V^C\{y_{\Delta m}(\lambda, k)\}, \quad (6)$$

onde:  $\mathcal{L}_{\Delta, N}^C\{z\} = \mathcal{L}_{\Delta, N}\{z^R\} + j\mathcal{L}_{\Delta, N}\{z^I\}$  (de forma semelhante para  $\mathcal{S}_V^C\{z\}$  e  $\mathcal{Q}_\Delta^C\{z\}$ ), com  $z^R \triangleq \text{Re}(z)$  e  $z^I \triangleq \text{Im}(z)$ ; e o modelo clássico é utilizado para caracterizar  $y_{\Delta m}(\lambda, k)$ , conforme a [Seção 2.3](#).

### 3.2 Matrizes de coerência

Cada elemento das matrizes complexas  $\mathbf{R}_{xy_q}(\lambda, k)$  e  $\mathbf{R}_{y_q y_q}(\lambda, k)$  pode ser descrito como:

$$\left[\mathbf{R}_{xy_q}\right]_{ij} = \mathbb{E}\left[x_i^R y_{qj}^R\right] + j\mathbb{E}\left[x_i^I y_{qj}^R\right] - j\mathbb{E}\left[x_i^R y_{qj}^I\right] + \mathbb{E}\left[x_i^I y_{qj}^I\right], \quad (7a)$$

$$\left[\mathbf{R}_{y_q y_q}\right]_{ij} = \mathbb{E}\left[y_{qi}^R y_{qj}^R\right] + j\mathbb{E}\left[y_{qi}^I y_{qj}^R\right] - j\mathbb{E}\left[y_{qi}^R y_{qj}^I\right] + \mathbb{E}\left[y_{qi}^I y_{qj}^I\right], \quad (7b)$$

no qual  $(i, j)$  são os índices de linha e coluna das matrizes, e  $(\lambda, k)$  foi omitido por clareza.

Assume-se que as partes real e imaginária de  $x_m$  sejam variáveis aleatórias de média nula [9] e  $\sigma_{x_m^R}^2 = \sigma_{x_m^I}^2 = 1/2 \sigma_{x_m}^2$  (idem para  $\sigma_{y_{qm}^R}^2$  e  $y_{qm}$ ) [9]. Utilizando os resultados do [Apêndice A](#) e assumindo que as variáveis aleatórias citadas sejam Gaussianas [9], é possível demonstrar que estimativas para os elementos das matrizes de coerência  $\tilde{\mathbf{R}}_{xy_q}(\lambda, k)$  e  $\mathbf{R}_{y_q y_q}(\lambda, k)$  podem ser obtidas através de

$$\left[\tilde{\mathbf{R}}_{xy_q}\right]_{ij} \triangleq \kappa_{ij} \cdot \left[\mathbf{R}_{xy_\Delta}\right]_{ij}, \quad (8a)$$

$$\left[\tilde{\mathbf{R}}_{y_q y_q}\right]_{ij} \triangleq \frac{1}{2} \left[ \left(\tau_{ij}^R + \tau_{ij}^I\right) \cdot \left[\mathbf{R}_{y_\Delta y_\Delta}\right]_{ij} + \left(\tau_{ij}^R - \tau_{ij}^I\right) \cdot \left[\mathbf{R}_{y_\Delta y_\Delta}\right]_{ij}^* \right], \quad (8b)$$

com  $\kappa_{ij} = \text{erf}\left\{\frac{1}{\sqrt{2}} \xi_{y_j}\right\}$  e

$$\tau_{ij}^F = \frac{2}{\pi} \int_{-1}^1 \int_0^{\xi_{y_j}} \int_0^{\xi_{y_i}} \frac{1}{\sqrt{1 - \left(\rho_{ij}^F \alpha\right)^2}} \exp\left\{-\frac{1}{2} \cdot \frac{u^2 - 2uv\rho_{ij}^F \alpha + v^2}{1 - \left(\rho_{ij}^F \alpha\right)^2}\right\} du dv d\alpha. \quad (9)$$

onde:  $F \in [\mathbb{R}, \mathbb{I}]$  representa genericamente real ou imaginário;  $\rho(\lambda, k)$  é o coeficiente de correlação (complexo) entre  $y_{qi}(\lambda, k)$  e  $y_{qj}(\lambda, k)$ ;  $\xi_{y_m}(\lambda, k) = \sqrt{2}v/\sigma_{y_m}(\lambda, k)$ ; e as matrizes de coerência  $\mathbf{R}_{xy_\Delta}(\lambda, k)$  e  $\mathbf{R}_{y_\Delta y_\Delta}(\lambda, k)$  são obtidas através do modelo clássico (vide [Seção 2.3](#)).

## 4 Simulações Computacionais

O cenário acústico utilizado foi construído por simulação. Sinais reais de fala e ruído interferente artificial, amostrados com  $f = 16$  kHz, foram convoluídos com respostas ao impulso obtidas a partir de um ambiente anecoico, conforme descrito em [12]. Cada dispositivo do BHA contém 3 microfones. Assumindo que o centro da cabeça do usuário está posicionado na origem do sistema de coordenadas, a fonte de fala foi posicionada artificialmente em (80 cm, 0°, 0°) (em coordenadas esféricas); ou seja, no plano horizontal médio da cabeça, à frente e próxima do usuário.

As falas utilizadas são provenientes de um indivíduo do gênero feminino [13]. O sinal interferente utilizado foi um ruído ICRA [14], cujas características de frequência são semelhantes às da voz. A fonte interferente encontra-se em (300 cm, 0°, -60°) (à frente e à esquerda do usuário). Esse azimute foi escolhido propositalmente para que os sinais captados pelo CF apresentassem maior nível de contaminação que os do DC. De forma complementar, foi adicionado ruído branco Gaussiano a todos os microfones, para emular ruído ambiente difuso e ruído eletrônico dos microfones. As

médias da razão sinal-interferência e razão sinal-ruído para os microfones de referência (frontais) de cada dispositivo são, respectivamente, 0 dB e 27 dB.

Os sinais de fala contaminada, no domínio tempo, foram normalizados na faixa  $(-1,1)$ , sendo analisadas quantizações para um número de bits  $3 \leq N \leq 9$  e limiar de saturação  $V \in \{2, 5, 20\}$  (aplicado no domínio frequência). O limiar de  $V = 2$  implica em saturação de cerca de 3% dos valores, enquanto que  $V = 5$  resulta em apenas 0,5%.  $V = 20$  corresponde ao caso sem saturação.

A STFT foi implementada com janelas de 256 amostras e sobreposição de 50%. A reconstrução dos sinais foi realizada através de uma STFT inversa, seguida por uma rotina de *overlap and add* [15].

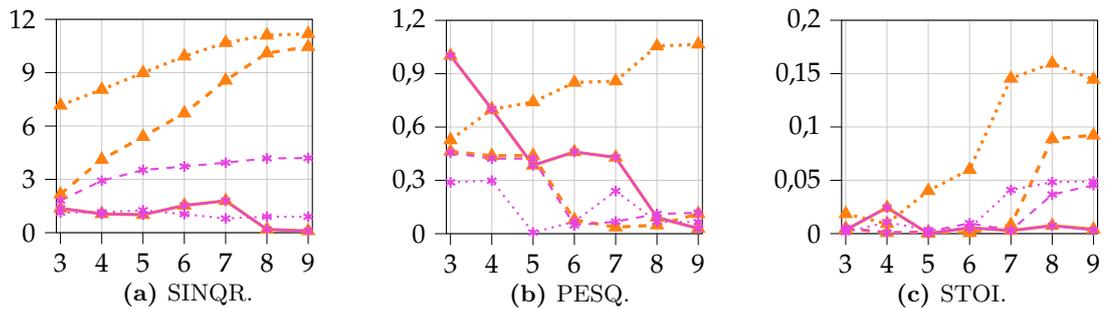
Assume-se que os sinais são estacionários [16], e portanto  $\mathbf{R}_{xy}(\lambda, k) \triangleq \mathbf{R}_{xy}(k)$  (idem para  $\mathbf{R}_{xx}(k)$ ,  $\mathbf{R}_{yy}(k)$  e  $\mathbf{R}_{y_q y_q}(k)$ ). De forma análoga, para o filtro de Wiener tem-se  $\mathbf{w}_F^*(\lambda, k) = \mathbf{w}_F^*(k)$  e  $\mathbf{w}_C^*(\lambda, k) = \mathbf{w}_C^*(k)$ .

O conforto acústico, qualidade da fala e inteligibilidade foram (respectivamente) estimados através dos seguintes critérios objetivos: razão sinal-ruído-interferência-e-quantização (SINQR, do inglês “Signal-to-Interference-and-Noise Ratio”) usando a razão sinal-ruído *a posteriori* [17]; PESQ [18]; e STOI [19]. Para cada critério, a avaliação de desempenho dos dois modelos (clássico e proposto) foi realizada através do erro absoluto médio entre o resultado obtido utilizando o MWF com matrizes de coerência estimadas a partir dos sinais sujeitos à restrição da taxa de comunicação (quantização e saturação) e o resultado obtido pelo modelo em questão.

## 5 Resultados e Discussão

Os resultados das simulações descritas na seção anterior são apresentados na [Figura 2](#). Em todas as figuras, o eixo das abscissas indica o número de bits e o das ordenadas o erro absoluto médio relacionado à métrica avaliada. Linhas laranjas com  $\Delta$  representam o erro do modelo clássico em relação à simulação e linhas magentas com  $*$  representam o erro obtido pelo modelo proposto em relação à simulação. Linhas contínuas são o resultado para  $V = 20$ , tracejadas para  $V = 5$  e pontilhadas para  $V = 2$ .

**Figura 2:** Erro absoluto médio para os modelos clássico (laranja) e proposto (magenta) em relação ao obtido com sinais simulados; com  $V = 20$  (linha contínua),  $V = 5$  (tracejado) e  $V = 2$  (pontilhado).



Considerando a ausência de saturação ( $V = 20$ ), verifica-se que ambos os modelos apresentam erros pequenos e de mesma ordem de grandeza para todos os três critérios objetivos. Nessa situação  $\kappa_{ij}$  e  $\tau_{ij}^F \approx 1$ . Para o caso em que o efeito da saturação é moderado ( $V = 5$ ), o modelo clássico apresenta erros significativos para SINQR. No caso de saturação severa ( $V = 2$ ), os erros do modelo

clássico são substancialmente maiores que os do modelo proposto para as três métricas.

Uma ampla variedade de simulações adicionais foi realizada. Foram alterados o tempo de reverberação do ambiente acústico, a localização da fonte interferente e a sua razão sinal-interferência. A análise dos resultados encontrados indica grande aderência ao exemplo apresentado nesta seção. A partir dessas observações, o modelo proposto pode ser considerado uma generalização do modelo clássico, apresentando maior acurácia de predição na existência de saturação.

## 6 Conclusões

Neste trabalho foi apresentado um novo modelo para estimação das matrizes de coerência necessárias para calcular os coeficientes do filtro de Wiener multicanal em aparelhos auditivos biauriculares com limitação da taxa de comunicação. O modelo proposto generaliza o modelo clássico que tem sido amplamente utilizado na literatura, permitindo também a modelagem de efeitos de saturação. Resultados de simulação utilizando cenários acústicos reais emulados e critérios objetivos de conforto acústico, qualidade e inteligibilidade indicam que o modelo proposto apresenta o mesmo desempenho que o modelo clássico em situações sem o efeito da saturação. Sob o efeito de saturação, o modelo proposto apresenta uma maior acurácia que o modelo clássico, principalmente em relação às previsões da razão sinal-ruído-interferência-e-quantização. Os resultados apresentados são de interesse para o projeto de aparelhos auditivos biauriculares, de forma a estabelecer o compromisso desejado entre o consumo de energia e a eficiência do processo de redução de ruído.

## Referências

- [1] DAY, G. A.; BROWNING, G. G.; GATEHOUSE, S. Benefit from binaural hearing aids in individuals with a severe hearing impairment. **British Journal of Audiology**, Taylor & Francis, v. 22, p. 273–277, 1988.
- [2] AVAN, P.; GIRAUDET, F.; BÜKI, B. Importance of binaural hearing. **Audiology & Neuro-otology**, Karger Publishers, v. 20, Suppl. 1, p. 3–6, 2015.
- [3] LOTTERMAN, S. H.; KASTEN, R. N.; MAJERUS, D. M. Battery life and nonlinear distortion in hearing aids. **Journal of Speech and Hearing Disorders**, ASHA, v. 3, p. 82–118, 1967.
- [4] ROY, O.; VETTERLI, M. Rate-Constrained Collaborative Noise Reduction for Wireless Hearing Aids. **IEEE Transactions on Signal Processing**, v. 57, n. 2, p. 645–657, fev. 2009.
- [5] ZHANG, J.; LI, C. Quantization-Aware Binaural MWF Based Noise Reduction Incorporating External Wireless Devices. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 29, p. 3118–3131, 2021.
- [6] LIPSHITZ, S.; WANNAMAKER, R.; VANDERKOOY, J. Quantization and dither: A theoretical survey. **Journal of the Audio Engineering Society**, v. 40, p. 355–374, mai. 1992.
- [7] WANNAMAKER, R.; LIPSHITZ, S.; VANDERKOOY, J.; WRIGHT, J. A theory of nonsubtractive dither. **Signal Processing, IEEE Transactions on**, v. 48, p. 499–516, mar. 2000.
- [8] HADAD, E.; MARQUARDT, D.; DOCCLO, S.; GANNOT, S. Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 23, n. 12, p. 2449–2464, 2015.
- [9] LOIZOU, P. C. **Speech Enhancement: Theory and Practice**. 2nd. USA: CRC Press, Inc., 2013.
- [10] CHEN, J.; BENESTY, J.; HUANG, Y.; DOCCLO, S. New insights into the noise reduction Wiener filter. **IEEE Transactions on Audio, Speech, and Language Processing**, v. 14, n. 4, p. 1218–1234, 2006.

- [11] SRIPAD, A.; SNYDER, D. A necessary and sufficient condition for quantization errors to be uniform and white. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 25, n. 5, p. 442–448, 1977.
- [12] KAYSER, H.; EWERT, S.; ANEMÜLLER, J.; ROHDENBURG, T.; HOHMANN, V.; KOLLMEIER, B. 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, p. 6, dez. 2009.
- [13] HENTER, G.; MERRITT, T.; SHANNON, M.; MAYO, C.; KING, S. Measuring the perceptual effects of modelling assumptions in speech synthesis using stimuli constructed from repeated natural speech. **Proceedings of the Annual Conference of the International Speech Communication Association**, p. 1504–1508, jan. 2014.
- [14] DRESCHLER, W.; HERSCHUURE, H.; LUDVIGSEN, C.; WESTERMANN, S. ICRA noises: Artificial noise signals with speech-Like spectral and temporal properties for hearing aid assessment. **Audiology**, v. 40, p. 148–157, jan. 2001.
- [15] SMITH, S. W. **FFT convolution and the overlap-add method**. [S.l.]: EE Times, 2007.
- [16] CHRISTIANSEN, C.; PEDERSEN, M. S.; DAU, T. Prediction of speech intelligibility based on an auditory preprocessing model. **Speech Communication**, v. 52, n. 7, p. 678–692, 2010.
- [17] BORGA, M. Canonical Correlation: a Tutorial, out. 2001.
- [18] 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. **2001 IEEE International Conference on Acoustics, Speech, and Signal Processing**, v. 2, p. 749–752, 2001.
- [19] TAAL, C.; HENDRIKS, R.; HEUSDENS, R.; JENSEN, J. A short-time objective intelligibility measure for time-frequency weighted noisy speech. **ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings**, p. 4214–4217, abr. 2010.
- [20] PAWULA, R. A modified version of Price’s theorem. **IEEE Transactions on Information Theory**, v. 13, n. 2, p. 285–288, 1967.

### Apêndice A: Cálculo de valores esperados com variáveis aleatórias sujeitas à saturação

Sejam duas variáveis aleatórias reais  $\beta$  e  $\gamma$  de média nula e coeficiente de correlação  $\rho$ , definindo  $\beta = \sigma_\beta \beta_1$ , em que  $\beta_1$  é uma variável aleatória com variância unitária e  $\sigma_\beta^2$  é a variância de  $\beta$  (e da mesma forma para  $\gamma = \sigma_\gamma \gamma_1$ ), através do Teorema Modificado de Price [20] é possível demonstrar que

$$E[\beta \mathcal{S}_V\{\gamma\}] = \kappa E[\beta\gamma], \tag{10a}$$

$$E[\mathcal{S}_V\{\beta\} \mathcal{S}_V\{\gamma\}] = \tau E[\beta\gamma], \tag{10b}$$

em que  $\mathcal{S}_V\{\cdot\}$  é a operação de saturação conforme a [Seção 3.1](#), e:

$$\kappa = \int_{-\xi_\gamma}^{\xi_\gamma} f_{G_1}(\gamma_1) d\gamma_1, \tag{11a}$$

$$\tau = \left( \int_{-1}^1 \left[ \int_{-\xi_\gamma}^{\xi_\gamma} \int_{-\xi_\beta}^{\xi_\beta} f_{B_1, G_1}(\beta_1, \gamma_1; \alpha\rho) d\beta_1 d\gamma_1 \right] da \right), \tag{11b}$$

onde:  $f_{G_1}(\gamma_1)$  é a função de densidade de probabilidade de  $\gamma_1$ ;  $f_{B_1, G_1}(\beta_1, \gamma_1; \alpha\rho)$  é a função de densidade de probabilidade conjunta de  $\beta_1$  e  $\gamma_1$ ,  $\rho$  o coeficiente de correlação entre  $\beta_1$  e  $\gamma_1$ , e  $\xi_\beta = \frac{\sqrt{2}V}{\sigma_\beta}$  (idem para  $\xi_\gamma$ ).