

Redução de Ruído Usando Subtração Espectral, Filtragem de Wiener e Log MMSE

Ênio dos Santos Silva*

Resumo— Este trabalho apresenta a investigação de técnicas de redução de ruídos. Especificamente, as técnicas de subtração espectral, filtragem de Wiener iterativa, filtragem de Wiener com estimativa de ruído e logMMSE são investigadas. Análises subjetivas e objetivas são utilizadas para avaliar o desempenho das técnicas supracitadas.

Palavras-Chave— Filtragem de Wiener, LogMMSE, Redução de Ruído, Subtração Espectral.

I. INTRODUÇÃO

Neste trabalho, com base em [1], diferentes técnicas de redução de ruído são investigadas. Para cada técnica utilizada (descrita na Seção II), medidas objetivas e subjetivas são discutidas. Para as discussões sobre qualidade dos sinais resultantes dessas técnicas, ressalta-se a importância da percepção sonora do ouvido humano, referenciada como “audibilidade” e apresentada através de curvas de “loudness” em [2].

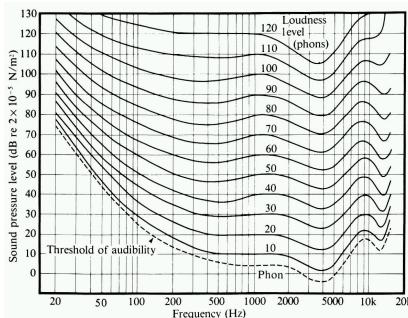


Fig. 1. Percepção do ouvido humano (“audibilidade”).

A descrição detalhada das técnicas de redução de ruído, adotadas neste trabalho, é apresentada em [1] e uma breve descrição é dada a seguir:

II. TÉCNICAS DE REDUÇÃO DE RUÍDO

A. Subtração Espectral

Historicamente, os algoritmos de subtração espectral foram os primeiros algoritmos destinados para a redução de ruído em sinais de fala [1]. Tais algoritmos foram amplamente baseados em princípios intuitivos e heurísticos. Mais especificamente, esses algoritmos exploraram o fato de que o ruído é aditivo, e de que é possível obter uma estimativa do espectro de sinal limpo simplesmente subtraindo o espectro de ruído do espectro de fala com ruído.

$$s_n(n) = s(n) + r(n) \quad (1)$$

Ênio dos Santos Silva*, *Universidade Federal de Santa Catarina - UFSC, Brasil. E-mails: enio@linse.ufsc.br.

onde $s_n(n)$ denota o sinal corrompido pelo ruído, $s(n)$ denota o sinal original e $r(n)$ representa o sinal de ruído. Analisando no domínio da frequência, temos que:

$$|\hat{S}(\omega)|^2 = H^2(\omega) \cdot |S_n(\omega)|^2 \quad (2)$$

onde

$$H(\omega) = \sqrt{1 - \frac{|\hat{R}(\omega)|^2}{|S_n(\omega)|^2}} \quad (3)$$

e $S_n(m)$, $R(m)$ denotam os espectros do sinal ruidoso $s_n(n)$ e o espectro do ruído $r(n)$, respectivamente.

B. Filtragem de Wiener

Estimadores de Wiener são considerados estimadores especiais (complexos: magnitude e fase) ótimos, no sentido de erro médio quadrático, quando ambos os coeficientes complexos da transformada discreta de Fourier (*discrete Fourier transform - DFT*), dos sinais de fala original $s(t)$ e do sinal estimado $\hat{s}(t)$, são considerados como variáveis aleatórias gaussianas independentes. Assim, tais estimadores minimizam o erro entre os modelos lineares dos espectros do sinal original (livre de ruído) e do sinal estimado (sinal tratado).

Assim, deseja-se projetar um filtro linear tal que o sinal de saída $\hat{s}(n)$ é o mais próximo possível, no sentido do erro médio quadrático, do sinal desejado $s(n)$. Para isso, filtros de resposta finita ao impulso são assumidos (*finite impulse response - FIR*) como apontados na Eq. (4)

$$\hat{s}(n) = \sum_{k=0}^{M-1} h_k \cdot s_n(n-k) \quad n = 0, 1, 2, \dots \quad (4)$$

Onde h_k são os coeficientes do filtro FIR e M é o número de coeficientes.

Analizando no domínio da frequência, temos para cada coeficiente k da DFT (*Bin*):

$$H(\omega_k) = \frac{\xi_k}{\xi_k + 1} \quad (5)$$

Onde ξ_k representa a SNR *a priori* na frequência ω_k .

A rigor, os filtros de Wiener são não causais e consequentemente não realizáveis, no entanto as próximas seções abordam soluções alternativas para a obtenção dos coeficientes $H(\omega_k)$.

1) *Wiener Iterativo* ($i = 3$): Aqui, a estimativa de Wiener é obtida de maneira iterativa (adotando $i = 3$ [1]), onde:

$$|\hat{S}_{i+1}(\omega_k)| = H_i(\omega_k) S_n(\omega_k) \quad (6)$$

em que $H_i(\omega_k)$ denota o filtro de Wiener obtido na i -ésima iteração, de acordo com o procedimento a seguir:

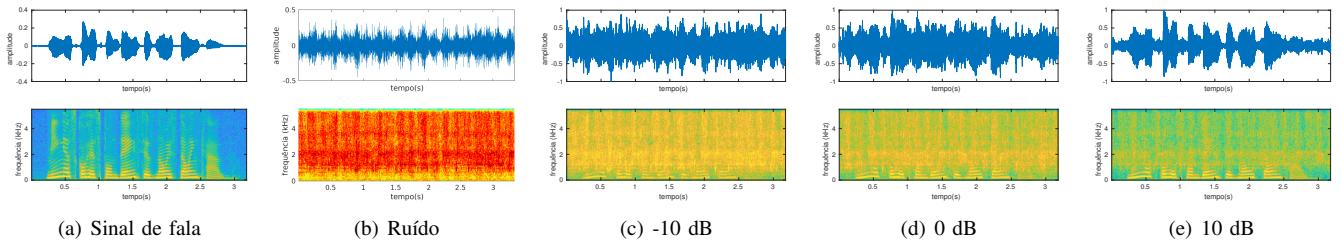


Fig. 2. Exemplo de sinais nos domínios do tempo e da frequência: (a) Sinal de fala original; (b) Sinal de ruído; (c), (d) e (e) Sinais de fala contaminados com $SNR = -10 \text{ dB}$, $SNR = 0 \text{ dB}$ e $SNR = 10 \text{ dB}$, respectivamente.

- 1) **Passo 1.** Obter uma estimativa de $H_i(\omega_k)$ baseado no sinal estimado $\hat{s}(n)$. Para a primeira iteração, inicializar $\hat{s}(n)$ com o sinal ruidoso $s_n(n)$. Neste trabalho adotou-se 120 ms de $s_n(n)$ para a estimativa inicial de $\hat{s}(n)$.
- 2) **Passo 2.** Filtrar o sinal ruidoso $s_n(n)$ com os novos coeficientes $H_i(\omega_k)$, de acordo com a Eq. (6), e obter o novo sinal estimado $\hat{s}_{(i+1)}(n)$. Repetir o passo 1 usando $\hat{s}_{(i+1)}(n)$ no lugar de $\hat{s}_i(n)$.
- 2) **Wiener com Estimação do Ruído:** Outra forma de estimar os coeficientes de $H_i(k)$ é através da estimação da SNR do ruído *a priori* ξ_k considerando uma estimativa de baixa variância de ξ_k . Segundo [1], tal consideração pode eliminar os artefatos geradores de ruídos musicais.

Portanto, ξ_k , para o quadro (frame) m é estimado como:

$$\hat{\xi}_k(m) = \alpha \frac{|\hat{S}_k(m-1)|^2}{|R_k(m-1)|^2} + (1-\alpha) \max \left(\frac{|S_{nk}(m)|^2}{|R_k(m)|^2} - 1, 0 \right) \quad (7)$$

onde $\alpha = 0.98$ é uma constante de suavização; $|\hat{S}_k|(m-1)$ representa o espectro do sinal estimado obtido no quadro $m-1$; e $S_{nk}(m)$, $R_k(m)$ denotam os espectros do sinal ruidoso ($s_n(n)$) e o espectro do ruído ($r(n)$), respectivamente.

C. log-MMSE

Em [1], é reportado que métricas objetivas baseadas no erro quadrático do logaritmo da magnitude espectral, entre os sinais de fala $s(t)$ e o sinal estimado $\hat{s}(t)$ (sinal tratado), deve ser mais adequado para o processamento da fala. Portanto, nesta Seção, é adotado um estimador que visa minimizar o valor esperado do erro médio quadrático do logaritmo do espectro de magnitude, como descrito pela Eq. (8)

$$erro = E\{(log|S| - log|\hat{S}|)^2\} \quad (8)$$

Onde $|S|$ e $|\hat{S}|$ denotam os espectros de magnitude do sinal original (livre de ruído) e do sinal estimado, respectivamente.

A minimização do erro na Eq. (8) pode ser calculada através do valor esperado com respeito a $p(\mathbf{S}_n; |S|)$, em que \mathbf{S}_n representa o espectro da fala com ruído ($|DFT\{s_n(t)\}|$). Assim,

$$|\hat{S}| = e^{E\{\log|abs(S)|\}|\mathbf{S}_n|} \quad (9)$$

Desenvolvendo a Equação (9), tem-se que:

$$|\hat{S}| = \frac{\xi}{\xi + 1} \exp \left\{ \frac{1}{2} \int_v^\infty \frac{e^{-t}}{t} dt \right\} Y \quad (10)$$

onde $v = \frac{\xi}{\xi + 1} \gamma$ e os termos ξ e γ referem-se às SNRs *a priori* e *a posteriori*, respectivamente (para mais detalhes, veja [1]).

III. ADIÇÃO DO RUÍDO AO SINAL DE FALA

Dado o sinal de fala original (sinal limpo) $s(n)$ ilustrado na Fig. 2.a e o sinal de ruído $r(n)$ ilustrado na Fig. 2.b, o sinal corrompido (sinal ruidoso) $s_n(t)$ é obtido através da operação fornecida pela Eq. (11):

$$s_n(t) = \alpha \cdot s(t) + \beta \cdot r(t) \quad (11)$$

Onde α e β são calculados de acordo com a SNR desejada. As figuras 2.c, 2.d e 2.f representam sinais corrompidos com SNRs de -10 dB , 0 dB e 10 dB , respectivamente. Os sinais apresentados na Fig. 2 são caracterizados nos domínios do tempo e da frequência (via spectrograma).

Nesta etapa, a função *v_addnoise.m* disponível em [3] foi utilizada para o cálculo dos parâmetros α e β da Eq. (11) e consequentemente para a obtenção do sinal ruidoso $s_n(t)$.

IV. IMPLEMENTAÇÃO PRÁTICA DAS TÉCNICAS DE REDUÇÃO DE RUÍDO

A Fig. 3 ilustra o procedimento para o tratamento dos sinais de fala corrompidos com os ruídos discutidos na Seção III. Em resumo, o sinal corrompido $s_n(t)$ (sinal de entrada) é processado, separadamente, pelas técnicas discutidas na Seção II, e como resultado, o sinal processado (sinal de saída) $\hat{s}(t)$ é uma estimativa do sinal de fala original $s(n)$:

$$F_i\{s_n(t)\} \Rightarrow \hat{s}(n)$$

A estratégia de redução de ruído, ilustrada na Fig. 3, adotada neste trabalho, utilizou os algoritmos, das técnicas discutidas na Seção II, implementados (em Matlab/Octave) e disponibilizados em [1] (a saber: *specsub.m*, *wienerit.m*, *wieneras.m* e *logmmse.m*).

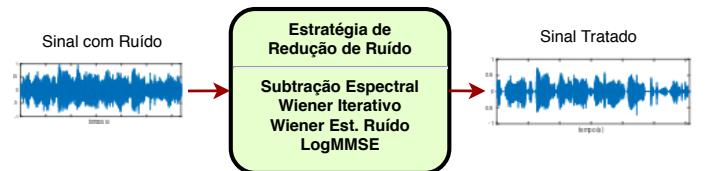


Fig. 3. Processamento dos sinais ruidosos.

V. AVALIAÇÃO DO SINAL DE FALA

A. Avaliação Objetiva do Sinal de Fala

A Fig. 4 ilustra o procedimento de avaliação objetiva do sinal estimado $\hat{s}(t)$. Aqui, a função objetivo é baseada no erro médio quadrático entre as componentes da magnitude da DFT dos sinais $s(t)$ e $\hat{s}(t)$. Inicialmente, os sinais $s(t)$ e $\hat{s}(t)$

são alinhados temporalmente a fim de garantirem que cada quadro correspondente represente o mesmo instante da fala. O procedimento de alinhamento é realizado usando a função *sigalign.m*¹ disponível em [3]. Em seguida, um procedimento de detecção de atividade de fala (*Voice activity detection - VAD*) é opcionalmente realizado. A implementação do VAD é realizada usando a função *vadsohn.m* disponível em [3].

Após os procedimentos de alinhamento e VAD, os sinais $s(t)$ e $\hat{s}(t)$ são janelados e armazenados em vetores (*buffering*). Aqui, foi utilizado o janelamento usando janela de *Hanning* de 220 amostras, correspondente a quadros (*frames*) de 20 ms de duração com sobreposição de 10 ms (50% de *overlap*). A partir desses quadros, as magnitudes das DFTs, com 512 pontos (*bins*), são calculadas. Particularmente, devido aos sinais em questão serem sinais reais, as magnitudes das DFTs são sinais pares. Por essa razão, para fins de cálculos do erro médio quadrático, apenas os primeiros 257 coeficientes das DFTs são considerados. Finalmente, os sinais de erro são obtidos através da diferença entre $|S| = |DFT\{s(t)\}|$ e $|\hat{S}| = |DFT\{\hat{s}(t)\}|$ conforme a Eq. (12):

$$\text{erro} = \frac{1}{N} \sum_{n=1}^N (|S| - |\hat{S}|)^2 \quad (12)$$

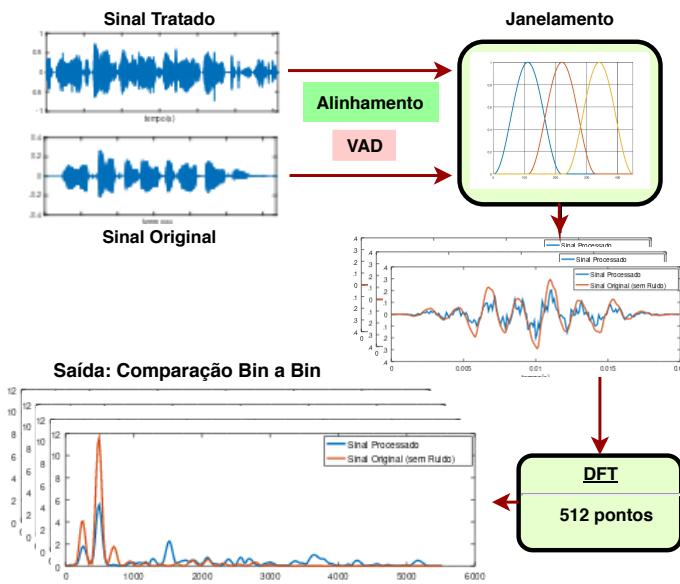


Fig. 4. Processo de avaliação objetiva do sinal estimado $\hat{s}(t)$.

A Fig. 5 apresenta as curvas de erro para os sinais provenientes das técnicas discutidas na Seção II.

B. Avaliação Subjetiva do Sinal de Fala

Nesta etapa, foram considerados testes de avaliação de acordo com as recomendações dadas em [4]. Para tal, foram coletados resultados de testes de 3 ouvintes. Os testes de audição se baseiam em avaliações da qualidade subjetiva

¹Para a técnica de filtragem de *Wiener* iterativa foi necessário um atraso de 1323 amostras, = 120 ms correspondente a 6 quadros de 20 ms necessário para estimação da potência espectral inicial do ruído [1].

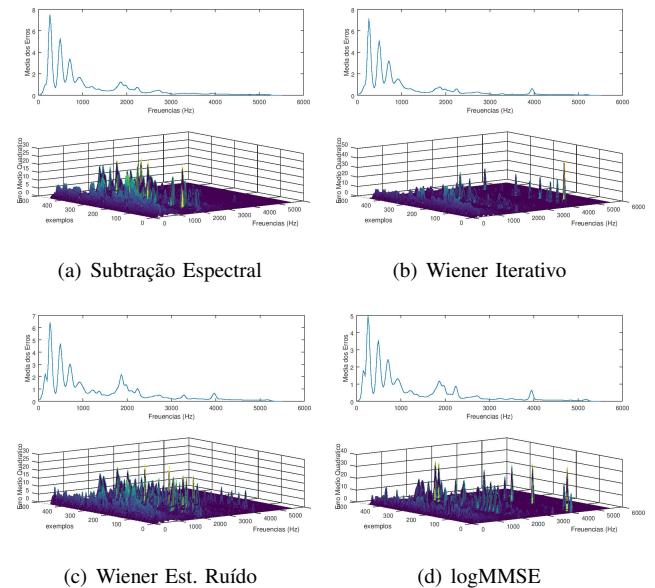


Fig. 5. Erro médio quadrático entre as magnitudes das DFTs de $s(t)$ e $\hat{s}(t)$

realizadas a partir da audição dos sinais de fala provenientes dos sistemas sob avaliação e dos sistemas de referência. Nesse contexto, os ouvintes devem receber apenas as instruções quanto à escala de avaliação a ser utilizada. Particularmente, a média aritmética dos valores obtidos nas avaliações de sinais de fala individuais é denominada “média de opiniões” (*mean opinion score - MOS*).

1) *Classificação por Categoria Absoluta:* Os testes de classificação por categoria absoluta (*absolute category rating - ACR*) se baseiam na avaliação direta da qualidade do sinal de fala, sem que o ouvinte (avaliador) disponha de material para comparação. Neste trabalho, a **qualidade subjetiva do sinal de fala** é investigada através da análise de pequenos grupos de sinais de fala de curta duração. A Tabela I define a escala de avaliação considerada.

TABELA I
ESCALA DE AVALIAÇÃO DA QUALIDADE DE AUDIÇÃO

| Pontuação | Qualidade do sinal de fala |
|-----------|----------------------------|
| 5 | Excelente |
| 4 | Bom |
| 3 | Regular |
| 2 | Ruim |
| 1 | Péssimo |

VI. RESULTADOS E ANÁLISE DE DESEMPENHO

A. Pré-processamento e preparação da base de dados

Os sinais de fala $s(t)$ foram representados através de 30 arquivos de fala com duração média de 3,43 segundos, amostrados a uma taxa de 11025 Hz. Tais arquivos correspondem a falas de frases curtas do português Brasileiro e são enunciadas por locutores do sexo feminino. Para representar os sinais de ruído $r(t)$, foram utilizados 3 arquivos contendo tipo de ruídos de máquinas elétricas domésticas (*mixer* e máquina de lavar louça). Esses arquivos apresentam duração média de 53,49 segundos e taxa de amostragem de 44100 Hz. Dessa

forma, foi necessário um procedimento de subamostragem (*downsample* usando o software *sox* [5]) nos arquivos de ruídos, resultando em arquivos com taxas de amostragem de 11025 Hz.

Após os procedimentos supracitados, 5 trechos (intervalos de curta duração, ≈ 5 s) de cada arquivo de ruído foram selecionados a fim de representarem diferentes características acústicas. Tais trechos representam o sinais $r(t)$ e são combinados com os sinais $s(t)$ (veja Seção III) visando a obtenção dos sinais contaminados $s_n(t)$ para os diferentes tipos de ruídos, considerando 8 possíveis SNRs, a saber: -20 dB, -10 dB, -5 dB, 0 dB, 5 dB, 10 dB, 20 dB e $+\infty$ dB.

B. Apresentação dos Resultados

As figuras 6, 7 e 8 apresentam, para SNRs iguais a -10 , 0 e 10 dB, as formas de onda no domínio do tempo e os espectrogramas dos sinais tratados $\hat{s}(t)$ provenientes da estimação usando as técnicas de: (a) subtração espectral, (b) filtragem de *Wiener* iterativa, (c) filtragem de *Wiener* com estimativa do ruído e (d) técnica usando o logmmse.

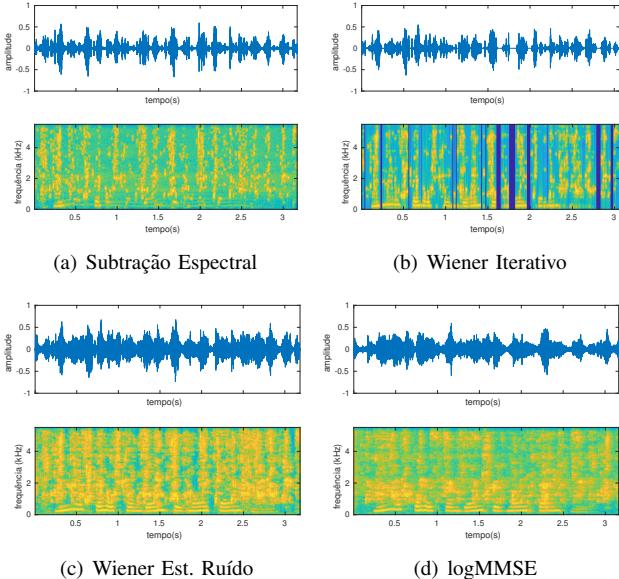


Fig. 6. Tratamento do sinal ruidoso com $SNR = -10$ dB

Avaliando os resultados apresentados aqui, verifica-se que os espectrogramas referentes aos sinais estimados através da filtragem de *Wiener* iterativa apresentam um grande número de descontinuidades. Já os sinais $\hat{s}(t)$ provenientes da filtragem de *Wiener* com estimativa do ruído e da técnica de logmmse, apresentam espectrogramas com forte intensidade para $SNR = -10$ dB e intensidade moderada e baixa para SNRs iguais a 0 e 10 dB, respectivamente.

Comparado com os demais, os sinais provenientes da subtração espectral apresentam espectrogramas com intensidade moderada para todas as SNRs apresentadas. Nesse contexto, é coerente presumir que, para intervalos originalmente de silêncio de $s(t)$, aqueles sinais estimados $\hat{s}(t)$ que apresentarem menor intensidade espectral (nesses intervalos originalmente de silêncio), apresentarão menores erros médios

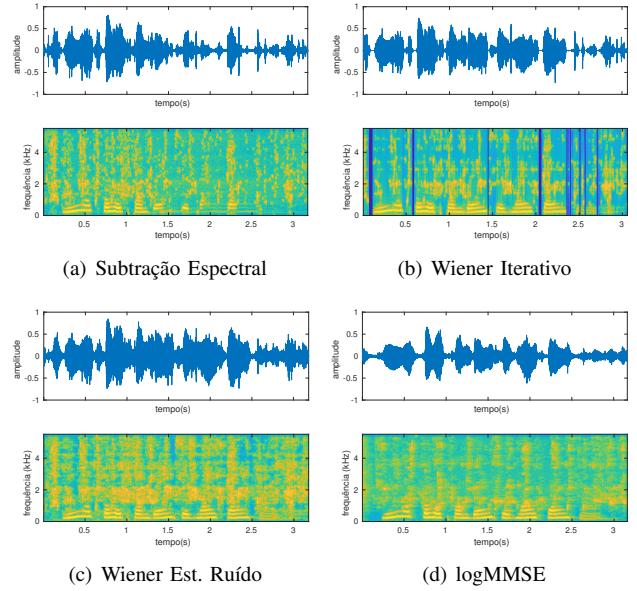


Fig. 7. Tratamento do sinal ruidoso com $SNR = 0$ dB

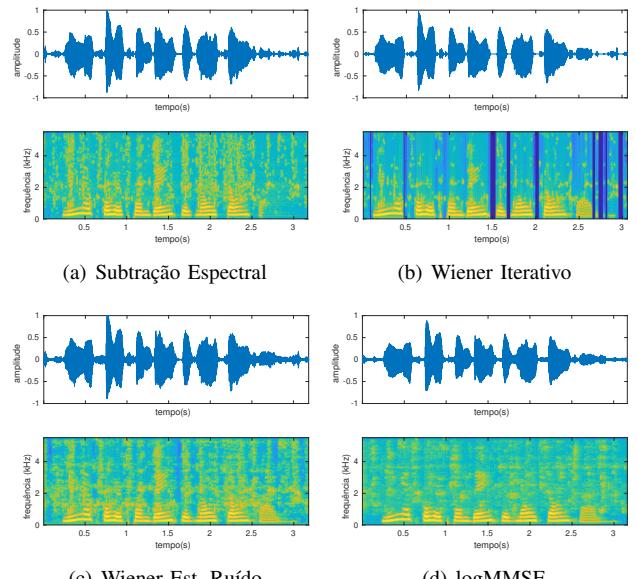


Fig. 8. Tratamento do sinal ruidoso com $SNR = 10$ dB

quadráticos de suas $|DFTs|$. No entanto, para curtos intervalos originalmente de silêncio, a permanência de um ruído de fundo nem sempre é relevante para a definição da qualidade do sinal estimado. Além disso, também é interessante analisar o comportamento das técnicas apenas para os intervalos (trechos) que são classificados como fala. Note que para intervalos de fala, $\hat{s}(t)$ provenientes da subtração espectral apresentam menos componentes harmônicos que as demais técnicas, o que pode indicar uma perda de qualidade para esses intervalos. Portanto, para a realização de uma análise distinta para os intervalos de fala e silêncio, as próximas seções apresentam resultados com e sem a utilização do VAD ilustrado na Fig. 4.

C. Avaliação Objetiva do Sinal de Fala

1) *Avaliação sem VAD*: As figuras 9 e 10 apresentam as curvas do erro médio quadrático em função de diferentes SNRs. No aspecto geral, sem a consideração isolada dos intervalos de silêncio e fala, para cenários de baixa SNR (ex.: -20dB e -10dB), a técnica de subtração espectral apresenta bons resultados quando comparados as demais técnicas. Já para valores maiores de SNR (ex.: a partir de 5dB), as técnicas de filtragem de *Wiener* iterativa e logmmse apresentam os melhores resultados. Ainda das figuras desta seção, nota-se que a técnica de filtragem de *Wiener* com estimativa de ruído apresenta os piores resultados de erro médio quadrático.

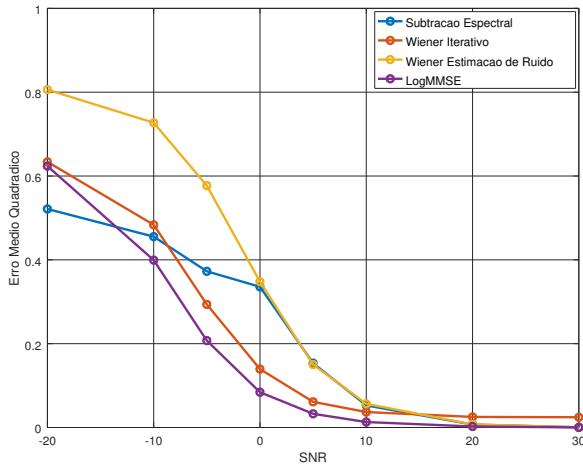


Fig. 9. Comportamento do Erro Médio Quadrático sem o VAD.

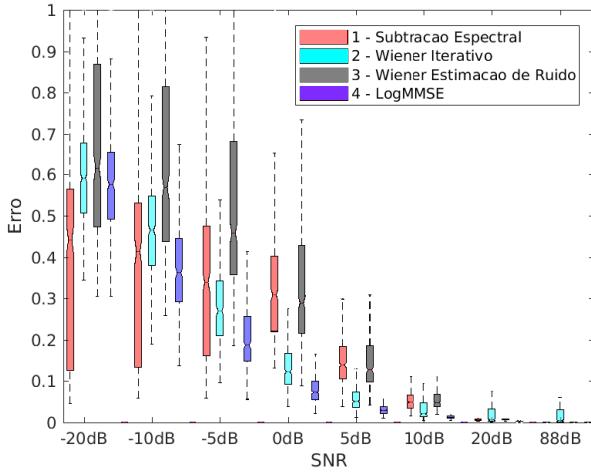


Fig. 10. Comportamento do Erro Médio Quadrático sem o VAD.

2) *Avaliação com VAD*: As figuras 11 e 12 apresentam as curvas do erro médio quadrático em função de diferentes SNRs. Nesta Seção, apenas os intervalos classificados como fala são computados para o cálculo do erro médio quadrático. Dessa forma, pode-se inferir que tal teste seja mais adequado para a avaliação da inteligibilidade dos sinais.

Nota-se que a técnica de logmmse apresenta os melhores resultados para todas as SNRs, seguido da técnica de filtragem de *Wiener* com estimativa de ruído que apresenta bons resultados para $SNRs \geq 5$ dB. Nesse contexto, a técnica de subtração espectral apresenta os piores resultados. Já a técnica de filtragem de *Wiener* iterativa apresenta resultados intermediários.

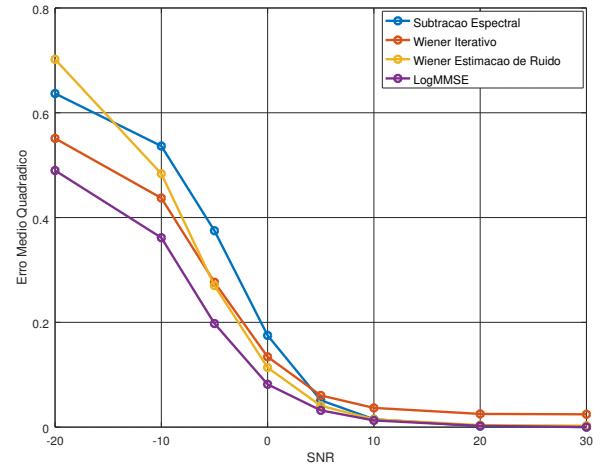


Fig. 11. Comportamento do Erro Médio Quadrático com o VAD.

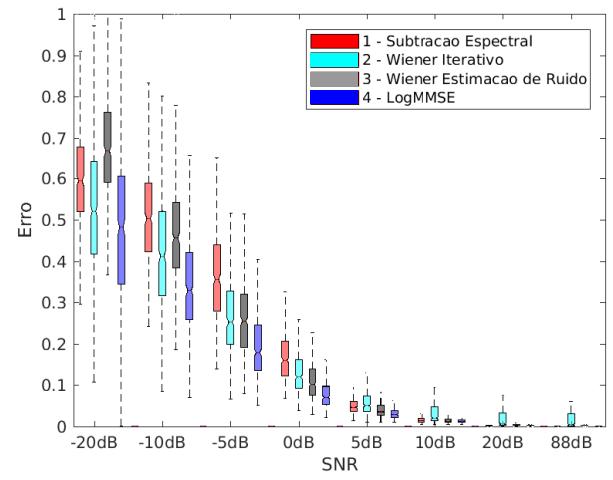


Fig. 12. Comportamento do Erro Médio Quadrático com o VAD.

Ainda para a avaliação objetiva com o VAD, as figuras 13 e 14 apresentam as curvas médias dos erros quadráticos médios comparando as diferentes técnicas e comparando as diferentes SNRs, respectivamente. Tais figuras apresentam uma análise “bin a bin” dos sinais de erro e indicam as frequências mais críticas responsáveis pelo aumento do erro. Nesse contexto, observa-se a maior diferença espectral durante o intervalo de 0 a 1000 Hz, com destaque negativo para a técnica de subtração espectral.

Para uma melhor análise da inteligibilidade dos sinais via análise objetiva, sugere-se a indicação da diferença espectral positiva e negativa, isto é, seria interessante anotar se a

diferença espectral, principalmente no intervalo de 2 a 5 kHz, é devido a baixa ou a alta intensidade dos componentes de $|\hat{S}|$. Analisando cuidadosamente os spectrogramas apresentados nas figuras 2.a, 6, 7 e 8, pode-se concluir que a diferença espectral para o intervalo de 2 a 5 kHz, para as técnicas de filtragem de Wiener e logmmse, é devido à baixa intensidade dos componentes de $|\hat{S}|$ nesse intervalo em comparação com os componentes do sinal original $|S|$. Portanto, segundo [2] (veja Fig. 1), espera-se que para as referidas técnicas os sinais resultantes sejam mais inteligíveis do que para a técnica de subtração espectral.

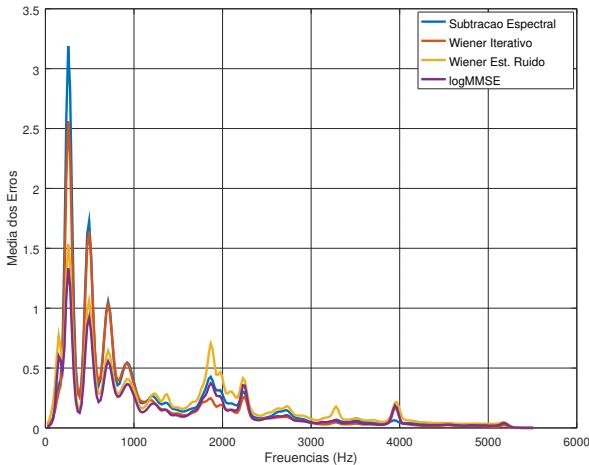


Fig. 13. Comparação das técnicas

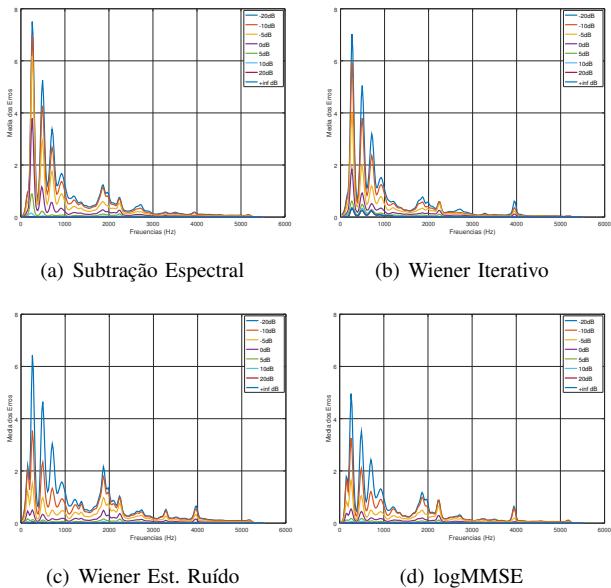


Fig. 14. Análise bin a bin

D. Avaliação Subjetiva do Sinal de Fala

A fim de analisar a qualidade subjetiva dos sinais tratados $\hat{s}(t)$, bem como suas inteligibilidades, testes subjetivos foram realizados seguindo a descrição discutida na Seção V-B

A Tabela VI-D apresenta os resultados MOS obtidos.

Da Tabela VI-D nota-se que os melhores resultados foram obtidos para os sinais provenientes da técnica de logmmse. Tais resultados confirmam as discussões apresentadas na Seção VI-C.2. Além disso, pode-se inferir que para SNRs maiores do

TABELA II
DESEMPENHO DAS TÉCNICAS SEGUNDO A AVALIAÇÃO MOS.

| SNR (dB) | Técnicas | | | |
|-----------|----------------|------------|-------------|-------------|
| | Sub. Espectral | Wiener It. | Wiener Est. | logMMSE |
| -20 | 1,00 | 1,00 | 1,06 | 1,00 |
| -10 | 1,00 | 1,00 | 1,33 | 1,33 |
| -5 | 1,28 | 1,11 | 1,72 | 1,83 |
| 0 | 1,67 | 1,56 | 2,17 | 2,56 |
| 5 | 2,50 | 2,44 | 2,67 | 2,94 |
| 10 | 2,67 | 2,61 | 2,94 | 3,33 |
| 20 | 3,33 | 4,00 | 3,50 | 4,11 |
| $+\infty$ | 4,78 | 4,78 | 4,89 | 5,00 |
| Média | 2,28 | 2,31 | 2,53 | 2,76 |

que 5 dB os resultados das quatro técnicas são marginalmente² satisfatórios.

VII. CONCLUSÕES DE COMENTÁRIOS FINAIS

Neste trabalho, diferentes técnicas de redução de ruído para o sinal de fala foram investigadas. Tais técnicas foram implementadas e testadas através de avaliações objetivas e subjetivas. A partir dos resultados obtidos pôde-se constatar a importância da análise isolada dos intervalos de silêncio e de fala. Nesse contexto, diferentes resultados são obtido para ambos os cenários. De modo geral, para fins de inteligibilidade, o intervalo de interesse é aquele que contém o sinal de fala (que contém informação linguística). Assim, uma atenção especial foi dada aos intervalos de fala, utilizando o procedimento de VAD. Dessa forma, pôde-se identificar as faixas de frequências mais afetadas ([01000] Hz) pelos processamentos das técnicas, e também analisar suas influências na qualidade objetiva dos sinais. Então, a partir da análise objetiva e da investigação dos spectrogramas dos sinais estimados, pôde-se inferir/justificar a qualidade subjetiva dos sinais provenientes das técnicas investigadas. Particularmente, notou-se que a técnica de subtração espectral gera mais artefatos nos intervalos classificados como fala e que tais artefatos são notados na faixa de frequência de 2 a 5 kHz, gerando ruídos musicais. Em contra partida, a técnica de logMMSE apresentou resultados que geraram menores quantidades de artefatos, justamente na faixa de 2 a 5 kHz (frequências mais sensíveis para o ouvido humano [2]). Além disso investigou-se o comportamento da qualidade dos sinais em função da variação da relação sinal-ruído (SNR), constatando resultados de redução de ruído satisfatórios para $SNRs \geq 5$ dB.

REFERÊNCIAS

- [1] P. C. Loizou, *Speech Enhancement: Theory and Practice*. Dallas, TX, USA: Boca Raton: CRC Press, 2007.
- [2] H. Fletcher and W. Munson, "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America*, vol. 5, pp. 82–108, Nov. 1933.
- [3] M. Brookes. (2005) Voicebox: Speech processing toolbox for matlab. [Online]. Available: <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>
- [4] ITU-T, *Recomendation P.800: Methods for subjective determination of transmission quality*. Geneve, Switzerland: Int. Telecomm. Union, 1996.
- [5] L. Norskog. (2015) Sox - sound exchange — homepage. [Online]. Available: <http://sox.sourceforge.net/>

²limite inferior para um grau de mínima qualidade subjetiva e inteligibilidade