

ESTUDO COMPARATIVO ENTRE A SUBTRAÇÃO ESPECTRAL DE MAGNITUDE E POTÊNCIA APLICADO NA REDUÇÃO DE RUÍDO EM SINAIS DE VOZ

Leandro Aureliano da Silva¹; Edna Lúcia Flôres²; Rodrigo Pinto Lemos²; Roberto Duarte Campos³; William Gigo⁴; Antonio Carlos Lemos Júnior⁵; Guilherme Henrique Alves⁶

^{1, 3, 4, 5, 6} Faculdade de Talentos Humanos - FACTHUS, Uberaba (MG), Brasil

^{1, 2, 6} Universidade Federal de Uberlândia - UFU, Uberlândia (MG), Brasil

² Universidade Federal de Goiás – UFG, Goiânia (GO), Brasil

lasilva@facthus.edu.br, edna@ufu.br, lemos@ufg.br, roberto.campos@facthus.edu.br, wgigo@facthus.edu.br, acjunior@facthus.edu.br, guilherme.henrique.alves@outlook.com

RESUMO: O presente trabalho tem a finalidade de realizar um estudo comparativo entre a subtração espectral magnitude e de potência que são técnicas utilizadas na redução de ruído em sinais de voz. Para tanto, foi usada as medidas de relação sinal/ruído segmentada e a distância de Itakura-Saito para validação dos resultados. Os estudos mostraram que a técnica de subtração espectral de potência apresenta melhores resultados quando comparado com a subtração espectral de magnitude.

PALAVRAS CHAVE: Subtração Espectral; Supressão de Ruído em sinais de voz; Distância de Itakura-Saito; Relação sinal ruído segmentada; Transformada rápida de Fourier.

COMPARATIVE STUDY BETWEEN POWER AND MAGNITUDE SPECTRAL SUBTRACTION APPLIED IN NOISE REDUCTION IN SPEECH SIGNALS

ABSTRACT: The present work aims to carry out a comparative study between the power and magnitude spectral subtraction which are techniques used in noise reduction in speech signals. For this purpose, we used segmental signal-to-noise ratio (SNRseg) and the Itakura-Saito distance for the results validation. Studies have shown that the power spectral subtraction technique offers best results when compared with magnitude spectral subtraction.

KEYWORDS: Spectral Subtraction; Noise suppression in voice signals; Distance of Itakura-Saito; Signal-to-noise ratio; Fast Fourier Transform.

INTRODUÇÃO

Em muitas aplicações de processamento de sinais de voz, a presença de ruído aditivo pode degradar a qualidade e a inteligibilidade dos sinais de voz. Basicamente, na transmissão destes sinais, vários fatores tais como: ruído ambiente, perdas devidas à codificação de enlaces digitais ou até mesmo através de outras conversações ou outras fontes de sinais podem comprometer a qualidade do sinal transmitido.

Com base nestas informações, muitas pesquisas têm sido realizadas nesta área e consequentemente muitos algoritmos para melhoria na qualidade de sinais de voz têm surgido com o objetivo de eliminar ou mesmo reduzir a intensidade do ruído aditivo. Dentre estes algoritmos, tem-se as subtrações espectrais de magnitude e potência, que são métodos utilizados para a reconstrução de sinais degradados por ruído.

Nessas técnicas, a melhoria na qualidade do sinal de voz é realizada por meio da subtração espectral (domínio da frequência) entre o sinal de voz contaminado por ruído aditivo e a estimativa média do ruído, avaliada em instantes de silêncio.

Alguns trabalhos envolvendo essas técnicas se destacam, dentre eles tem-se o artigo de Verteletskaia e Simak (2011) cuja a finalidade foi o desenvolvimento de

um método capaz de eliminar o ruído musical gerado pelo processo da subtração.

Dhivya e Justin (2014) propuseram uma técnica de redução de ruído que se baseia na combinação de Wavelet e subtração espectral. Nessa técnica, aplicam-se a subtração espectral nos coeficientes de aproximação e limiar como *soft thresholding* nos coeficientes de detalhes. Elas utilizaram cinco filtros wavelet e compararam o melhor filtro com base na relação sinal/ruído. Para verificar o desempenho da técnica que elas propuseram, utilizaram a relação sinal/ruído, coeficiente de correlação e a evolução perceptual da qualidade da fala (PESQ).

Silva (2007) em seu trabalho de redução de ruído em sinais de voz, utilizou a subtração espectral de potência como uma pré-filtragem no sinal ruidoso. Em seguida, esse sinal filtrado foi utilizado como estimativa inicial para o filtro de Kalman. Esse trabalho obteve ótimos resultados em relação a distorção espectral e relação sinal/ruído.

Karam, et al. (2014) utilizou a subtração espectral de magnitude, variando o tamanho da janela de Hanning para melhorar a relação sinal/ruído (SNR). Essa variação foi uma técnica opcional. Os resultados mostraram que o uso de quadros sobrepostos com tamanho de 128 amostras melhoraram a remoção do ruído.

Diante disso, o presente artigo tem a proposta de realizar um estudo comparativo entre a subtração espectral de magnitude e potência. Para fins de comparação será

utilizada a relação sinal/ruído segmentada e a distância de Itakura-Saito como medidas de desempenho.

SUBTRAÇÃO ESPECTRAL DE MAGNITUDE

O método de subtração espectral de magnitude foi proposto por Boll em 1979. Neste método uma estimativa que corresponde ao espectro de magnitude do sinal livre de ruído é obtida a partir da subtração entre o espectro de magnitude do sinal contaminado e o espectro médio de magnitude do ruído estimado em trechos onde não há atividade sonora.

Para que se possa desenvolver o método, assume-se que o sinal contaminado por ruído seja dado pela Equação (1).

$$y(n) = x(n) + v(n) \quad (1)$$

em que $y(n)$, $x(n)$ e $v(n)$ são: sinal contaminado por ruído, sinal livre de ruído e ruído aditivo, respectivamente.

Aplicando-se na Equação (1) a Transformada Discreta de Fourier (DFT) tem-se:

$$Y(e^{j\omega_k}) = X(e^{j\omega_k}) + V(e^{j\omega_k}) \quad (2)$$

em que $Y(e^{j\omega_k})$ é o espectro de magnitude do sinal contaminado $y(n)$. Similarmente $X(e^{j\omega_k})$ e $V(e^{j\omega_k})$ representam os espectros de magnitudes dos sinais livre de ruído, $x(n)$, e ruído aditivo, $v(n)$.

Analisando a Equação (2) nota-se que havendo o espectro de magnitude do sinal contaminado $Y(e^{j\omega_k})$ e havendo uma estimativa média do espectro de magnitude do ruído $\mu(e^{j\omega_k})$ encontrada em trechos de silêncio, determina-se a estimativa do espectro de magnitude do sinal estimado $\hat{X}(e^{j\omega_k})$, de acordo com a Equação (3).

$$\hat{X}(e^{j\omega_k}) = |Y(e^{j\omega_k})| - |\mu(e^{j\omega_k})| \quad (3)$$

Para que se tenha o sinal estimado $\hat{x}(n)$ no domínio do tempo, a fase do sinal contaminado $\theta_Y(e^{j\omega_k})$ é combinada com o espectro de magnitude do sinal estimado $\hat{X}(e^{j\omega_k})$, e em seguida, aplica-se a transformada discreta de Fourier inversa (IDFT) conforme mostra a Equação (4).

$$\hat{x}(n) = \sum_{k=0}^{N-1} |\hat{X}(e^{j\omega_k})| e^{j\theta_Y(e^{j\omega_k})} e^{-j\omega_k n} \quad (4)$$

em que: $\omega_k = \frac{2\pi}{N}k$ é a frequência discreta da transformada.

Sendo o erro espectral resultante dado pela Equação (5), Boll (1979) desenvolveu algumas modificações de

modo a reduzir o efeito audível do erro espectral resultante.

$$\varepsilon(e^{j\omega_k}) = \hat{X}(e^{j\omega_k}) - X(e^{j\omega_k}) \quad (5)$$

Essas mudanças foram:

1) Magnitude média: Nessa modificação, o valor da magnitude do sinal ruidoso $|Y(e^{j\omega_k})|$ na Equação (3) é substituído pelo seu valor médio $|\overline{Y(e^{j\omega_k})}|$, levando dessa forma a uma redução do erro de estimação. A magnitude média pode ser calculada desta forma:

$$|\overline{Y(e^{j\omega_k})}| = \frac{1}{M} \sum_{i=0}^{M-1} |Y_i(e^{j\omega_k})| \quad (6)$$

em que $Y_i(e^{j\omega_k})$ é a transformada da i -ésima janela de $y(n)$. Dessa forma, o estimador espectral utilizando a média do espectro do ruído é dado por:

$$X_\mu(e^{j\omega_k}) = \left[|\overline{Y(e^{j\omega_k})}| - |\mu(e^{j\omega_k})| \right] e^{j\theta_Y(e^{j\omega_k})} \quad (7)$$

em que $X_\mu(e^{j\omega_k})$ é o valor estimado do sinal de voz determinado a partir da subtração espectral entre o valor médio do espectro do sinal contaminado por ruído $|\overline{Y(e^{j\omega_k})}|$ e o valor médio do ruído $\mu(e^{j\omega_k})$ determinado durante os instantes de silêncio.

2) Retificação de meia onda: Devido à natureza aleatória do ruído, podem ocorrer situações em que a magnitude do espectro do sinal contaminado por ruído $Y(e^{j\omega_k})$ seja menor do que a magnitude média do ruído $\mu(e^{j\omega_k})$, o que resulta em valores negativos para subtração espectral e gera um problema conhecido como ruído musical. Nesse caso, é comum a aplicação do conceito de retificação de meia onda, que consiste em igualar a zero os valores negativos resultantes da subtração espectral de magnitude. A retificação de meia onda é dada por:

$$\left| \hat{X}(e^{j\omega_k}) \right| = \begin{cases} \hat{X}(e^{j\omega_k}) & \text{se } \hat{X}(e^{j\omega_k}) > 0 \\ 0, & \text{caso contrário} \end{cases} \quad (8)$$

3) Redução do ruído residual: O ruído residual é o ruído restante após a subtração espectral. Sua redução é feita após a retificação de meia onda. A eliminação ou diminuição deste ruído pode ser feita substituindo seu valor corrente por um valor mínimo determinado, a partir da análise de janelas adjacentes aplicadas ao sinal da seguinte forma: um valor mínimo de $\hat{X}(e^{j\omega_k})$ é determinado e usado se a magnitude espectral da

estimativa $\hat{X}(e^{j\omega_k})$ for menor que o máximo valor do ruído residual determinado durante os instantes de silêncio, caso contrário o valor permanece constante. Desta forma tem-se:

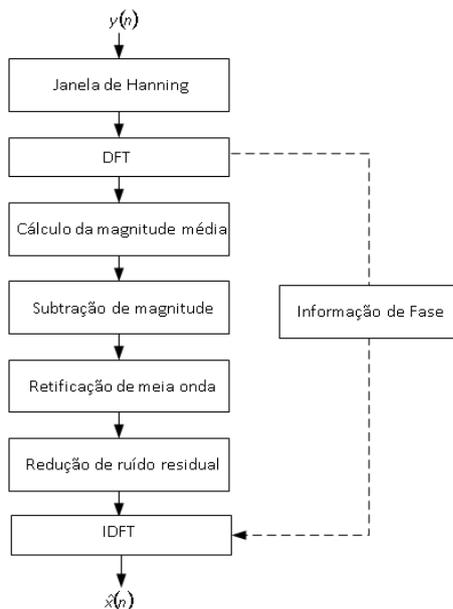
$$|\hat{X}(e^{j\omega_k})| = \begin{cases} |\hat{X}(e^{j\omega_k})|, & \text{se } |\hat{X}(e^{j\omega_k})| \geq \text{Max}|V_R(e^{j\omega_k})| \\ \text{Min}|V_R(e^{j\omega_k})|, & \text{se } |\hat{X}(e^{j\omega_k})| < \text{Max}|V_R(e^{j\omega_k})| \end{cases} \quad (9)$$

em que $|V_R(e^{j\omega_k})|$ é o ruído residual e pode ser determinado em instantes de silêncio pela Equação (10).

$$V_R(e^{j\omega_k}) = V(e^{j\omega_k}) - \mu(e^{j\omega_k}) \quad (10)$$

Assim, a Figura 1 ilustra, em forma de diagrama em blocos, todas as modificações propostas por Boll no algoritmo de subtração espectral de magnitude.

Figura 1: Diagrama em blocos da Subtração Espectral de Magnitude.



Vale ressaltar que todas essas modificações propostas por Boll comprometem a qualidade do sinal e consequentemente a inteligibilidade do mesmo. Como exemplo a magnitude média pode provocar perdas de informação em sinais de voz que apresentam variações rápidas, como início de fonemas e consoantes plosivas (p, b, t). Outro problema é a retificação de meia onda que contribui para a perda de informação de baixa amplitude. (MACIEL, 2003).

SUBTRAÇÃO ESPECTRAL DE POTÊNCIA

A subtração espectral de potência é um método proposto por Berouti et al. (1979). Consiste na subtração

do espectro de potência do sinal contaminado por ruído pelo espectro de potência do ruído estimado em trechos de silêncio. Dessa forma a subtração espectral de potência pode ser realizada conforme mostra a Equação (11).

$$|\hat{X}(e^{j\omega_k})|^2 = |Y(e^{j\omega_k})|^2 - |\mu(e^{j\omega_k})|^2 \quad (11)$$

em que $|\hat{X}(e^{j\omega_k})|^2$ é o espectro de potência estimado do sinal livre de ruído, $|Y(e^{j\omega_k})|^2$ é o espectro de potência do sinal degradado por ruído e $|\mu(e^{j\omega_k})|^2$ é o valor médio do espectro de potência do ruído avaliado em trechos de silêncio.

Para a redução do ruído musical, Berouti et al. (1979), modificou-se a Equação (11) incluindo um termo denominado fator de subtração. Esta modificação é dada pela Equação (12).

$$|\hat{X}(e^{j\omega_k})|^2 = |Y(e^{j\omega_k})|^2 - \alpha |\mu(e^{j\omega_k})|^2 \quad (12)$$

em que α é o fator de subtração. Este por sua vez, tem como finalidade o controle de redução de ruído, e é determinado empiricamente segundo Berouti et al. (1979) pela Equação (13).

$$\alpha = \alpha_0 - \frac{3}{20} \text{SNR} \quad -5\text{dB} \leq \text{SNR} \leq 20\text{dB} \quad (13)$$

em que α_0 é o valor de α encontrado quando a relação sinal/ruído (SNR) for 0dB.

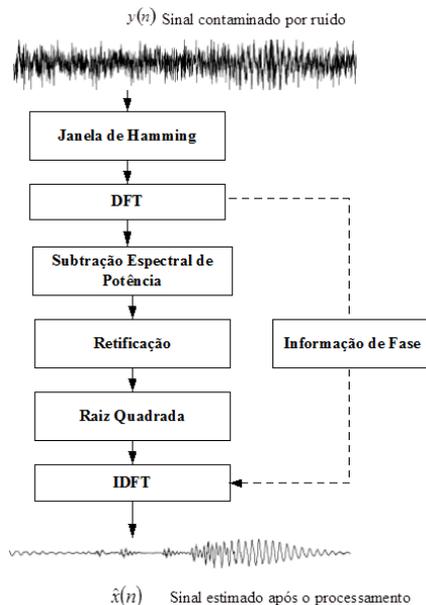
Outra mudança significativa realizada por Berouti et al. (1979) foi no processo de retificação. Nesse caso, é adicionado um limite espectral cuja função é não perder informações do sinal de baixa amplitude. Essa mudança é dada na Equação (14).

$$|\hat{X}(e^{j\omega_k})|^2 = \begin{cases} |\hat{X}(e^{j\omega_k})|^2, & \text{se } |\hat{X}(e^{j\omega_k})|^2 > \beta |\mu(e^{j\omega_k})|^2 \\ \beta |\mu(e^{j\omega_k})|^2, & \text{caso contrário} \end{cases} \quad (14)$$

em que β é o limite espectral, cuja função é mascarar o efeito das componentes do ruído musical. Segundo Berouti et al. (1979), β é encontrado a partir da relação sinal/ruído (SNR) podendo ter valores típicos que variam de 0,02 até 0,06 para uma $\text{SNR} \leq 0\text{dB}$ e valores que estão entre 0,0005 até 0,02 para $\text{SNR} > 0\text{dB}$. Neste trabalho o valor de β , foi fixado com o valor de 0,02.

A Figura 2 mostra o diagrama em blocos do processo de subtração espectral de potência.

Figura 2: Diagrama em blocos da Subtração Espectral de Potência.



RESULTADOS

Para a avaliação das técnicas apresentadas nesse trabalho, foram utilizados diferentes sinais de voz amostrados a uma taxa de 22050Hz com 16 bits. Esses sinais foram contaminados por ruído branco e aplicados em uma janela de Hamming de tamanho 512 com sobreposição de 50%.

Para avaliar e comparar os resultados de ambas as técnicas, utilizou-se o software Matlab R2013B.

Como parâmetro de medidas, foram utilizadas as medidas da relação sinal ruído segmentada (SNRseg) e a distância de Itakura Saito ($d(b,a)$).

A SNRseg é uma medida que pode ser calculada em segmentos curtos de sinal de voz de forma a equilibrar os pesos atribuídos aos trechos de maior e menor intensidades do sinal. Essa medida é calculada pela Equação (15). Deller, et al. (1993).

$$SNR_{seg} = \frac{10}{M} \sum_{j=0}^{M-1} \log_{10} \left[\frac{\sum_{n=m_j}^{m_{j+1}} x^2(n)}{\sum_{n=m_j-N+1}^{m_{j+1}} [x(n) - \hat{x}(n)]^2} \right] \quad (15)$$

onde: m_j são os limites de cada um dos M quadros de tamanho N .

Quando dois sinais possuem espectros diferentes, a SNRseg não fornece uma medida significativa para mostrar a diferença entre esses espectros. Nesse caso, a distância de Itakura Saito fornece melhores resultados. Essa distância pode ser calculada, utilizando os parâmetros de predição linear (LPC) como mostrado na Equação (16). Rabiner e Schafer (1978).

$$d(b,a) = \log \left[\frac{bRb^T}{aRa^T} \right] \quad (16)$$

onde: “a” e “b” são os vetores de coeficientes LPC dos sinais original e estimado, respectivamente e “R” é a matriz de autocorrelação do sinal original.

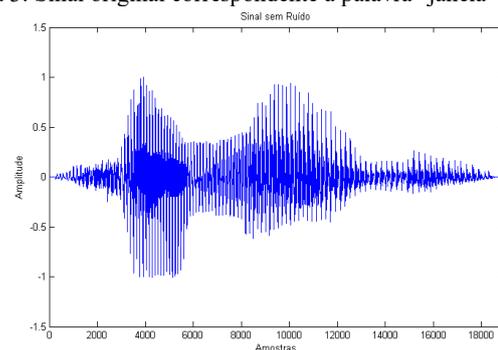
Quando o resultado da Equação (16) é aproximadamente zero, significa que os espectros dos sinais original e estimado são próximos. Se esse resultado é igual a zero, os espectros desses sinais são iguais.

No primeiro teste usou-se a subtração espectral de magnitude. Para tanto, o parâmetro α teve seu valor igual a 1 e o parâmetro β fixado em 0,02 como descrito anteriormente. Para este exemplo, a janela aplicada foi uma janela de Hamming com tamanho de 512 amostras usando sobreposição de 50%. A relação sinal/ruído segmentada de entrada (SNRI) foi de 3 dB. Sendo assim, as Figuras 3, 4 e 5, mostram os resultados obtidos nesse teste.

Observando as Figuras 4 e 5 que correspondem respectivamente, ao sinal contaminado por ruído branco e sinal após a subtração espectral de magnitude, nota-se uma redução do nível de ruído nos trechos de silêncio, o que proporcionou uma relação sinal/ruído segmentada de saída (SNRO) de 4 dB e uma distância de Itakura-Saito ($d(a,b)$), de 0,7071.

No segundo teste, utilizou-se a subtração espectral de potência. Entretanto, o parâmetro α é dado pela Equação (13) e o valor de β , o mesmo usado no teste anterior. Para este caso, adicionou-se também ao sinal de voz original um ruído branco, o qual também resultou em uma relação sinal/ruído segmentada de entrada (SNRI) de 3 dB. Sendo assim, os resultados são mostrados nas Figuras 6, 7 e 8, as quais se referenciam ao sinal original, sinal contaminado por ruído e sinal estimado após a subtração espectral de potência respectivamente.

Figura 3: Sinal original correspondente à palavra “janela”.



Analisando as Figuras 7 e 8, pode-se notar a redução de ruído envolvida neste processamento. A relação sinal/ruído segmentada de saída (SNRO) nesse caso foi de 5 dB e a distância de Itakura-Saito ($d(a,b)$) de 0,6015. Isso significa que para esse teste, a subtração espectral de potência apresenta resultados um pouco melhor do que a subtração espectral de magnitude.

Figura 4: Sinal contaminado por ruído branco.

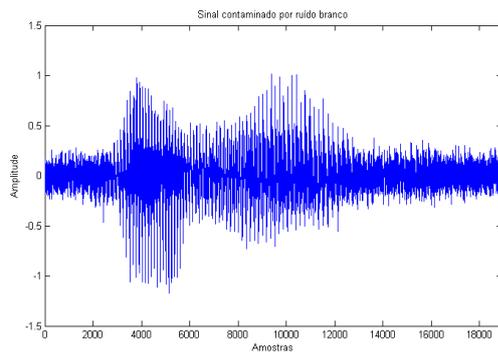


Figura 5: Sinal estimado após a Subtração Espectral de Magnitude.

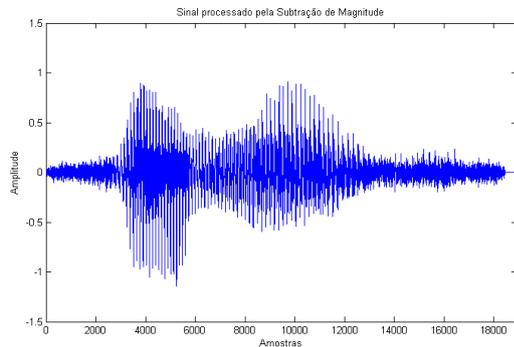
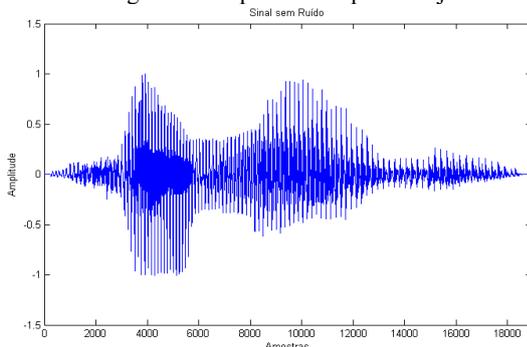


Figura 6: Sinal original correspondente à palavra “janela”.



Para mostrar que a técnica de subtração espectral de potência apresenta resultados melhores em relação a SNRO e $d(a,b)$, apresentam-se as Figuras 9 e 10 com testes de dez palavras com SNRI de 3dB.

Figura 7: Sinal contaminado por ruído branco.

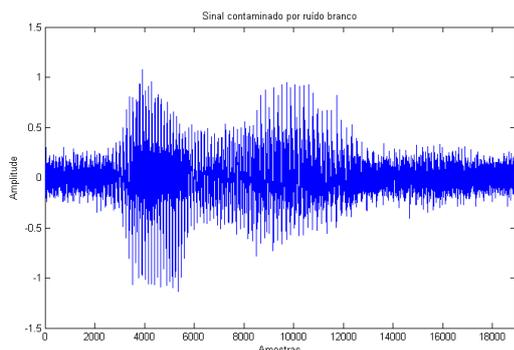
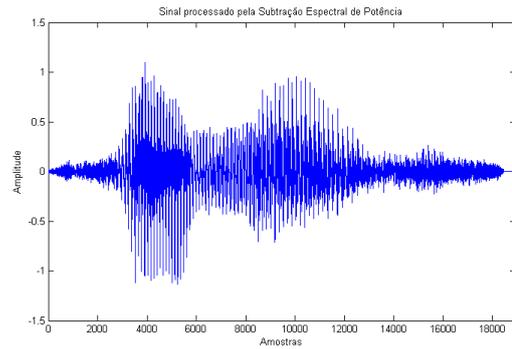


Figura 8: Sinal estimado após a Subtração Espectral de Potência.



Analisando o gráfico da Figura 9, é possível verificar que no contexto geral a subtração espectral de potência apresenta resultados melhores quando comparado com a subtração espectral de magnitude.

No entanto, quando se verifica a distância de Itakura-Saito cujo os resultados são apresentados no gráfico da Figura 10, percebe-se que esses também são melhores para a subtração espectral de potência, pois a distância espectral é menor.

Figura 9: Gráfico comparativo entre a subtração espectral de magnitude e potência para SNRI de 3 dB utilizando dez palavras.

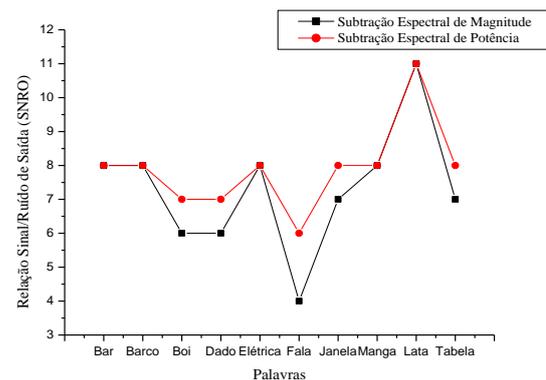
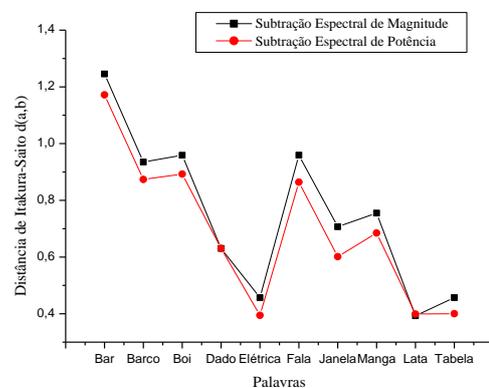


Figura 10: Gráfico comparativo entre a subtração espectral de magnitude e potência para $d(a,b)$ com SNRI 3 dB para dez palavras.



CONCLUSÃO

Este trabalho apresentou um estudo comparativo entre a subtração espectral de magnitude e potência. Foram realizados testes com essas técnicas utilizando as medidas de distância de Itakura Saito (mede a distorção espectral) e a relação sinal/ruído segmentada.

Após os testes realizados, pôde-se concluir que a subtração espectral de potência apresentou melhores resultados para a distância de Itakura Saito e para a relação sinal/ruído segmentada na reconstrução do sinal contaminado por ruído branco em comparação com a subtração espectral de magnitude.

REFERÊNCIAS

- BEROUTI, M; SCHWARTZ, R; MAKHOUL, J. Enhancement of speech corrupted by acoustic noise. *Proc. IEEE Int. conf. Acoust., Speech, Signal Process.*, pp. 208-211, Apr. 1979.
- BOLL, S. F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.27, pp. 113-120, Apr. 1979.
- DELLER, J. R; PROAKIS, J. G; HANSEN, J. H. L. **Discrete-Time Processing of Speech Signals**. Prentice Hall, 1993.
- DHIVYA, R; JUSTIN, J. A Novel Speech Enhancement Technique. *IJRET: International Journal of Research in Engineering and Technology*, Vol. 3, Special Issue: 07, May 2014, p. 98-102.
- KARAM, Marc et al. Noise Removal in Speech Processing Using Spectral Subtraction. *Jsip, Journal of Signal and Information Processing*, [s.l.], v. 05, n. 02, p.32-41, 2014. Scientific Research Publishing, Inc., <http://dx.doi.org/10.4236/jsip.2014.52006>.
- MACIEL, R. C. V. **Melhoria na qualidade de sinais de fala degradados por ruído através da utilização de sinais sintetizados**. Dissertação de mestrado. Universidade de São Paulo-USP, 2003.
- RABINNER, L. R; SCHAFER, R. W. **Digital processing of speech signals**. Prentice- Hall, 1978.
- SILVA, Leandro Aureliano da. **Filtros de Kalman no Tempo e Frequência discretos combinados com Subtração Espectral**. 2007. 127 f. Dissertação (Mestrado) - Curso de Engenharia Elétrica, Universidade Estadual de São Paulo, São Carlos, 2007.
- VERTELETSKAYA, E.; SIMAK, B. Noise Reduction Based on Modified Spectral Subtraction Method. *Iaeng International Journal Of Computer Science*, [s.l.], v. 38, 10 fev. 2011.