

Utilização dos filtros de Kalman de tempo discreto e frequência combinados com subtração espectral de potência na redução de ruído aditivo em sinais de voz contaminados por ruído colorido

Use of Kalman discrete time and frequency filters combined with spectral power subtraction in the reduction of additive noise in speech signals contaminated by colored noise

Uso de filtros discretos de tiempo y frecuencia de Kalman combinados con sustracción de potencia espectral en la reducción de ruido aditivo en señales de voz contaminadas por ruido coloreado

Recebido: 15/05/2022 | Revisado: 23/05/2022 | Aceito: 26/05/2022 | Publicado: 01/06/2022

Leandro Aureliano da Silva

ORCID: <https://orcid.org/0000-0002-2050-2249>

Faculdade de Talentos Humanos, Brasil

E-mail: leandro_aureliano@yahoo.com.br

Resumo

Este artigo tem como objetivo apresentar e comparar técnicas de redução de ruído na reconstrução de sinais de voz contaminados por ruído colorido com o objetivo de verificar a distorção espectral gerada pelo sinal processado. Os filtros de Kalman no tempo e na frequência discretos em conjunto com a técnica de subtração espectral de potência, foram utilizados para a redução do ruído na reconstrução dos sinais e comparados. Os sinais utilizados foram contaminados por ruído colorido e a avaliação do desempenho dos algoritmos foram feitas utilizando a relação sinal-ruído segmentada e a distância de Itakura-Saito que verifica a distorção espectral no sinal processado. O melhor método encontrado através dos testes realizados foi o método do filtro de Kalman de tempo discreto em conjunto com a subtração espectral de potência em relação a distorção espectral.

Palavras-chave: Filtro de Kalman; Subtração espectral; Enriquecimento da voz; Supressão de ruído.

Abstract

In This article aims to present and compare noise reduction techniques in the reconstruction of speech signals contaminated by colored noise in order to verify the spectral distortion generated by the processed signal. The discrete time and frequency Kalman filters in conjunction with the power spectral subtraction technique were used for noise reduction in signal reconstruction and compared. The signals used were contaminated by colored noise and the performance evaluation of the algorithms was performed using the segmented signal-to-noise ratio and the Itakura-Saito distance that verifies the spectral distortion in the processed signal. The best method found through the tests performed was the discrete time Kalman filter method together with the spectral power subtraction in relation to spectral distortion.

Keywords: Kalman filter; Spectral power subtraction; Speech enrichment; Noise suppression.

Resumen

Este artículo tiene como objetivo presentar y comparar técnicas de reducción de ruido en la reconstrucción de señales de voz contaminadas por ruido coloreado para verificar la distorsión espectral generada por la señal procesada. Los filtros de Kalman de tiempo discreto y frecuencia junto con la técnica de sustracción espectral de potencia se utilizaron para la reducción de ruido en la reconstrucción de la señal y se compararon. Las señales utilizadas estaban contaminadas por ruido coloreado y la evaluación del desempeño de los algoritmos se realizó utilizando la relación señal a ruido segmentada y la distancia Itakura-Saito que verifica la distorsión espectral en la señal procesada. El mejor método encontrado a través de las pruebas realizadas fue el método de filtro de Kalman de tiempo discreto en conjunto con la sustracción de potencia espectral en relación con la distorsión espectral.

Palabras clave: Filtro de Kalman; Sustracción espectral de potencia; Enriquecimiento de voz; Supresión de ruido.

1. Introdução

Durante a transmissão de sinais de voz por sistemas de telecomunicações, estes podem ser contaminados por ruído de diferentes fatores como ruído ambiente, perdas devidas à codificação em enlaces digitais, através de outras conversações ou

outras fontes de sinais. No entanto, vários estudos estão sendo realizados sobre a aplicação de algoritmos na redução de ruídos em sinais de voz. As principais técnicas utilizadas para a redução do ruído em sinais de voz são os filtros de Wiener, a subtração espectral, modificação de filtros adaptativos conforme relatam os trabalhos de Jakati e Kuntoji (2021) e Panda et al., (2022). Trabalhos com o uso de Wavelets vem sendo propostos também nos últimos anos conforme relata os autores (Silva et al., 2018; Silva et al., 2021) (Mourad, 2022).

Os autores Shao e Chang (2006) concatenaram o filtro de Kalman a um banco de filtros wavelet com um filtro de ponderação perceptual. Eles usaram uma técnica de mascaramento do modelo psicoacústico para derivar o filtro de ponderação. Segundo os autores, esse trabalho trouxe duas contribuições. O primeiro foi o modelo auditivo baseado em wavelets com um banco de filtros wavelet perceptual que mapeia a resposta em frequência do sistema auditivo humano através da decomposição em sub-bandas. O segundo foi o filtro de Kalman utilizando um modelo de espaço de estados de voz no domínio wavelet, cujo custo computacional foi reduzido quando comparado ao filtro de Kalman de tempo discreto. Eles foram capazes de reduzir o ruído em diferentes ambientes com baixa degradação do sinal.

George et al. (2018) em seu trabalho, propuseram um filtro de Kalman que ajusta dinamicamente o seu ganho a fim de minimizar os efeitos da estimação LPC na modelagem do sinal de voz. Este ajuste foi necessário para reduzir o nível de ruído residual, melhorando assim a capacidade de aprimoramento do filtro de Kalman. Eles concluíram que o desempenho de aprimoramento do ganho do filtro de Kalman, produz uma melhora significativa no ruído. Além disso, eles relatam que os resultados subjetivos dos testes, mostraram uma preferência significativa do método proposto sobre todos os outros tipos de tratamento usados na redução de ruído de voz durante as atividades de fala segundo os autores.

Mourão (2019) em seu trabalho aponta que as técnicas que usam tempo-frequência como a utilização de máscaras de Wiener e binária vem demonstrando ser técnicas importantes no enriquecimento do sinal de voz degradado por ruído. Ainda dentro dessa ótica alguns trabalhos vêm se destacando como os de Koning et al. (2018), Koning et al. (2014) e Madhu et al. (2012), sendo que estes trabalhos se concentram na redução de ruído em implantes de cocleares. Mas são técnicas utilizadas na redução de ruído aditivo em sinais de voz. Ainda em seu trabalho, Mourão usa o método de mínimo erro quadrático médio (MMSE) que segundo ele é pouco explorado. Portanto ele conclui que o método de MMSE é mais eficiente quando comparado a técnica de Wiener e máscara binária para resultados de relação sinal ruído de entrada baixa.

Contudo na maioria dos estudos, o ruído aditivo utilizado nos testes é o branco. Neste contexto, é apresentado neste trabalho um estudo de métodos de redução de ruído combinando subtração espectral de potência e filtros discretos de Kalman no tempo e frequência com a finalidade de verificar a menor distorção espectral no sinal processado quando o ruído for colorido.

Neste trabalho então são propostos dois métodos para a reconstrução de sinais de voz. Os métodos de filtragem de Kalman no tempo e o de Kalman na frequência, tendo como condição inicial para ambos os filtros, o sinal estimado através da subtração espectral de potência. Após comparações entre as reconstruções obtidas pelos dois métodos observou-se que o filtro de Kalman no tempo em conjunto com a subtração espectral de potência apresentou relações sinal/ruído segmentada próximas. Já em relação à distorção espectral, o filtro de Kalman no tempo apresenta resultados um pouco melhor.

Este trabalho foi dividido em quatro partes: a introdução descrita acima; a metodologia e descrição dos algoritmos na seção 2, demonstrando o funcionamento de ambos os algoritmos na redução de ruído nos sinais de voz. Já a seção 3, é responsável pela apresentação dos resultados que comprovaram a eficiência das técnicas propostas e por fim na seção 4, tem-se a conclusão deste trabalho.

2. Metodologia e descrição dos algoritmos

2.1 Filtro de Kalman de Tempo Discreto (FKTD)

Segundo Brown e Hwang (1997), o filtro discreto de Kalman foi desenvolvido em 1960 por Rudolf Emil Kalman que publicou o seu famoso artigo descrevendo uma solução recursiva para o problema da filtragem linear de dados discretos. Assim, com os avanços da computação digital, o filtro de Kalman tem sido uma ferramenta muito importante em áreas como: navegação, monitoração de processos, instrumentação de usinas nucleares, controle e automação, astronomia, economia e reconstrução de sinais degradados por ruído.

Sendo o objetivo deste artigo à redução de ruído em sinais de voz, inicia-se o desenvolvimento do filtro a partir da modelagem do sinal de voz, o qual é modelado por um processo auto-regressivo de ordem P ($AR(P)$) da seguinte forma:

$$x(n) = \sum_{k=1}^P a_P(k)x(n-k) + w(n), \quad (1)$$

em que: $a_P(k)$ são os coeficientes de predição LPC de ordem P , $w(n)$ ruído de excitação de entrada e $x(n)$ o sinal de voz.

Sendo a maior parte dos sinais de voz medidos na presença de ruído aditivo tem-se:

$$y(n) = x(n) + v(n), \quad (2)$$

em que: $y(n)$ é o sinal de voz contaminado por ruído e $v(n)$ o ruído aditivo.

Observando-se a equação (1), percebe-se que esta pode ser representada por um modelo de espaço de estado. Sendo assim, as equações (1) e (2) podem ser reescritas da seguinte forma:

$$x(n) = A(n-1)x(n-1) + w(n), \quad (3)$$

$$y(n) = H(n)x(n) + v(n), \quad (4)$$

nas quais:

$x(n)$ é um vetor de estado com dimensão $P \times 1$ no instante n ;

$A(n-1)$ é a matriz transição de estados com dimensão $P \times P$, que relaciona os instantes atuais n com os instantes passados $n-1$, variante no tempo;

$w(n)$ é um vetor de excitação de entrada da equação de estado de ordem $P \times 1$ modelado por um ruído branco;

$y(n)$ é o vetor de observação de dimensão $M \times 1$;

$H(n)$ é uma matriz variante no tempo de ordem $M \times P$;

$v(n)$ é um vetor de ruído branco aditivo de dimensão $M \times 1$;

Admitindo os processos $w(n)$ e $v(n)$ como sendo ruídos independentes, as equações de correlação podem ser escritas como:

$$E[\mathbf{v}(n)\mathbf{v}^T(k)] = \begin{cases} R(n), & k = n \\ 0, & k \neq n \end{cases}, \quad (5)$$

$$E[\mathbf{w}(n)\mathbf{w}^T(k)] = \begin{cases} Q(n), & k = n \\ 0, & k \neq n \end{cases}, \quad (6)$$

em que: $R(n)$ e $Q(n)$ são elementos da diagonal das matrizes de covariâncias, respectivamente relacionadas ao ruído aditivo e ruído do processo de entrada.

O filtro de Kalman de tempo discreto faz as estimativas de um processo usando uma forma de controle de realimentação: o filtro estima o estado do processo em um dado momento e então obtém a realimentação na forma de uma nova medida. Como tal, as equações para o filtro de Kalman se dividem em dois grupos: equações de atualização de tempo (predição) e equações de atualização da medida (correção). As equações de atualização de tempo (predição) são dadas por (Brown & Hwang, 1997) (Abreu, Oliveira & Fonseca Neto, 2020):

$$\hat{\mathbf{x}}^-(n) = \mathbf{A}(n-1)\hat{\mathbf{x}}(n-1). \quad (7)$$

$$\mathbf{P}^-(n) = \mathbf{A}(n-1)\mathbf{P}(n-1)\mathbf{A}^T(n-1) + \mathbf{Q}(n). \quad (8)$$

E as equações de atualização de medida (correção) são dadas por:

$$\mathbf{K}(n) = \mathbf{P}^-(n)\mathbf{H}^T(n) \left[\mathbf{H}(n)\mathbf{P}^-(n)\mathbf{H}^T(n) + R(n) \right]^{-1}. \quad (9)$$

$$\hat{\mathbf{x}}(n) = \hat{\mathbf{x}}^-(n) + \mathbf{K}(n) \left[y(n) - \mathbf{H}(n)\hat{\mathbf{x}}^-(n) \right]. \quad (10)$$

$$\mathbf{P}(n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{H}(n)]\mathbf{P}^-(n). \quad (11)$$

em que: $\mathbf{P}(n)$ é a matriz erro de covariância, $\mathbf{K}(n)$ é o ganho de Kalman, responsável pela minimização dos elementos da diagonal de $\mathbf{P}(n)$ que contém os erros médio quadráticos e $\hat{\mathbf{x}}(n)$ a estimativa do sinal da fala no instante n .

O sinal de menos (“-”) nas equações de 7 a 11 indica uma estimativa inicial antes da assimilação de uma nova medida.

2.2 Filtro de Kalman de Frequência Discreta (FKF)

Seja uma janela de comprimento l aplicada ao sinal disponível. Sendo assim, o sinal disponível pode ser reescrito desta forma:

$$y(n,l) = x(n,l) + v(n,l), \quad (12)$$

em que: $x(n, l)$ é o sinal de voz sem ruído e $v(n, l)$ é o ruído aditivo, ambos com uma janela de comprimento l .

Determinando a transformada discreta de Fourier do sinal sem ruído $X(\omega, l)$, pode-se então reescrever $y(n, l)$, em função da transformada inversa da seguinte maneira (Fujimoto & Ariki, 2000):

$$y(n, l) = \sum_{\omega=0}^{N-1} X(\omega, l) \exp\left(j2\pi \frac{\omega n}{N}\right) + v(n, l), \quad (13)$$

A equação (11) na forma matricial torna-se:

$$y(n, l) = \underbrace{\begin{pmatrix} 1 \\ \exp\left(j2\pi \frac{n}{N}\right) \\ \vdots \\ \exp\left(j2\pi \frac{(N-1)n}{N}\right) \end{pmatrix}^T}_{\mathbf{H}_n} \underbrace{\begin{pmatrix} X(0, l) \\ X(1, l) \\ \vdots \\ X(N-1, l) \end{pmatrix}}_{\mathbf{X}_l} + v(n, l), \quad (14)$$

$$y(n, l) = \mathbf{H}_n \mathbf{X}_l + v(n, l), \quad (15)$$

em que: n representa o tempo dentro do quadro l e N o número de amostras no quadro.

Na equação (15), \mathbf{X}_l é um vetor que possui um espectro complexo para cada elemento no quadro. Como o tempo n , dentro do vetor \mathbf{X}_l não tem significado, $\mathbf{X}(n, l)$, pode ser reescrito da seguinte maneira (Fujimoto & Ariki, 2000):

$$\mathbf{X}_l = \mathbf{X}(n-1, l) = \mathbf{X}(n, l). \quad (16)$$

Observando a equação (16), pode-se notar que a matriz transição de estados tem valor unitário, uma vez que o tempo não tem significado no vetor \mathbf{X}_l . Desta forma, o filtro de Kalman na frequência discreta é representado pelas seguintes equações (Fujimoto & Ariki, 2000):

$$\mathbf{K}(n, l) = \mathbf{P}(n-1, l) \mathbf{H}_n^{*T} \left[\mathbf{H}_n \mathbf{P}(n-1, l) \mathbf{H}_n^{*T} + R(n, l) \right]^{-1}. \quad (17)$$

$$\hat{\mathbf{X}}(n, l) = \hat{\mathbf{X}}(n-1, l) + \mathbf{K}(n, l) [y(n, l) - \mathbf{H}_n \hat{\mathbf{X}}(n-1, l)]. \quad (18)$$

$$\mathbf{P}(n, l) = \mathbf{P}(n-1, l) - \mathbf{K}(n, l) \mathbf{H}_n \mathbf{P}(n-1, l). \quad (19)$$

$$\hat{\mathbf{X}}(0, l) = [\hat{Y}(0), \hat{Y}(1), \dots, \hat{Y}(N-1)]. \quad (20)$$

$$P(0,l) = \begin{pmatrix} R(0,l) & 0 & \dots & 0 \\ 0 & R(1,l) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R(N-1,l) \end{pmatrix}. \quad (21)$$

O algoritmo do filtro de Kalman na frequência é representado pelas equações (17 a 21). Na equação (17) aparece o termo $*T$, este termo representa o complexo conjugado da matriz transposta. O vetor complexo do filtro de Kalman na frequência é dado por:

$$\hat{X}(n,l) = [\hat{X}(0,l), \hat{X}(1,l), \dots, \hat{X}(N-1,l)]^T, \quad (22)$$

em que: $\hat{X}(n,l)$ é o vetor complexo do espectro estimado de $X(\omega,l)$. O mesmo pode ser reconstruído pela aplicação da transformada inversa discreta de Fourier. Os valores iniciais para o filtro são dados pela subtração espectral de potência.

2.3 Subtração espectral

A subtração espectral é um método para restauração do espectro de potência ou do espectro de magnitude de um sinal observado na presença de ruído aditivo (Vaseghi, 2000) (Boll, 1979). O método é realizado por meio da subtração no domínio da frequência entre o sinal contaminado por ruído e a estimativa da média do ruído avaliado em trechos de silêncio. Na restauração do sinal no domínio do tempo, a estimativa instantânea do espectro de magnitude é combinada com a fase do sinal contaminado por ruído. Em seguida aplica-se a transformada discreta de Fourier inversa.

Para o desenvolvimento, considere que o sinal contaminado por ruído aditivo seja representado da seguinte forma:

$$y(n) = x(n) + v(n), \quad (23)$$

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

Aplicando transformada discreta de Fourier da equação (23) resulta na seguinte expressão:

$$Y(e^{j\omega}) = X(e^{j\omega}) + V(e^{j\omega}), \quad (24)$$

em que: $Y(e^{j\omega})$, $X(e^{j\omega})$ e $V(e^{j\omega})$ são respectivamente as transformadas discretas de Fourier do sinal contaminado, sinal sem ruído e ruído aditivo.

Generalizando, a subtração espectral pode ser escrita da seguinte forma conforme demonstra Berouti, Schwartz e Makhoul (1979):

$$|\hat{Y}(e^{j\omega})|^b = |Y(e^{j\omega})|^b - \alpha |\mu(e^{j\omega})|^b, \quad (25)$$

em que: $|\hat{Y}(e^{j\omega})|^b$ é uma estimativa do espectro do sinal sem ruído e $|\mu(e^{j\omega})|^b$ é a média do espectro do ruído avaliada em trechos de silêncio.

Para que se tenha à subtração espectral de magnitudes o expoente **b** deve ser igual a 1 e para o espectro de potência o expoente **b** deve ser igual a 2. Já o parâmetro α controla o quanto de ruído é subtraído do sinal contaminado.

Devido à natureza aleatória do ruído, a subtração espectral pode gerar valores negativos, que vão decrementar a relação sinal/ruído segmentada (SNRseg). Para suplantarmos este problema é usada uma retificação que consiste em mapear os valores negativos obtidos da subtração em valores positivos. O processo de retificação é representado por:

$$|\hat{Y}(e^{j\omega})|^b = \begin{cases} |\hat{Y}(e^{j\omega})|^b, & \text{se } |\hat{Y}(e^{j\omega})|^b > \beta |\mu(e^{j\omega})|^b \\ \beta |\mu(e^{j\omega})|^b, & \text{Caso Contrário} \end{cases}, \quad (26)$$

em que: $0 \leq \beta \ll 1$ é o limite espectral mínimo, $\alpha \geq 1$ é o fator de sobre-estimação.

No caso da subtração de espectral de potência o parâmetro α depende da SNRseg e pode ser calculado como relata Berouti, Schwartz e Makhoul (1979):

$$\alpha = \alpha_0 - \frac{3}{20} SNR \quad -5dB \leq SNR \leq 20dB, \quad (27)$$

em que: α_0 é o valor de α para 0dB. Neste trabalho foi utilizado α_0 igual a 4.

Uma vez processado, o sinal estimado no domínio do tempo é obtido utilizando à transformada inversa discreta de Fourier juntamente com a fase do sinal contaminado por ruído conforme abaixo.

$$\hat{y}(n) = \sum_{k=0}^{n-1} |\hat{X}(e^{j\omega})| e^{j\theta_Y(e^{j\omega})} e^{-j\frac{2\pi}{N}\omega n}, \quad (28)$$

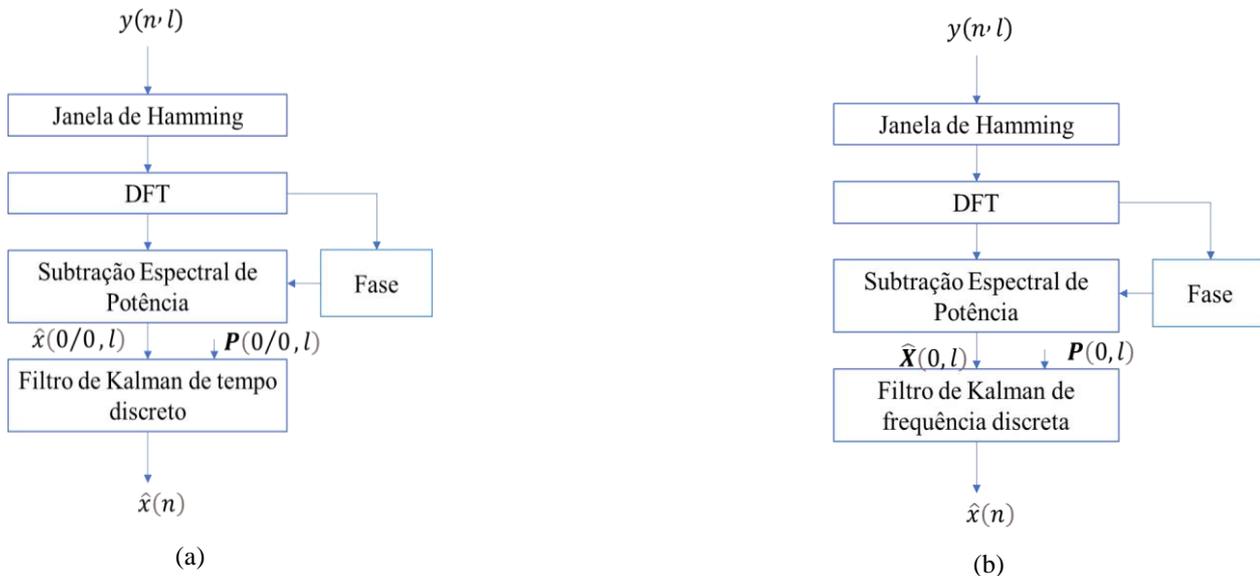
em que: $\theta_Y(\omega)$ é a fase do sinal contaminado por ruído no domínio da frequência.

Este sinal estimado na equação (28) é utilizado como condição inicial para os filtros de Kalman apresentados nas seções acima.

2.4 Método Proposto

Para um melhor entendimento da combinação das técnicas apresentadas nas seções 2.1, 2.2 e 2.3, apresenta-se o diagrama em blocos da combinação destes algoritmos na figura 1.

Figura 1: Diagrama em blocos da combinação dos algoritmos: (a) Subtração espectral de potência combinada com filtro de Kalman de tempo discreto e (b) Subtração espectral de potência combinada com filtro de Kalman de frequência discreta.



Fonte: Autor.

No diagrama apresentado na Figura 1, é aplicado ao sinal ruidoso $y(n, l)$ uma janela de Hamming de 256 amostras com sobreposição de 50%. Após, neste sinal é aplicado a transformada discreta de Fourier (DFT) separando o módulo e a fase. Em seguida aplica-se a equação (25), obtendo estimativa do espectro do sinal sem ruído com o uso da subtração espectral de potência. Contudo este sinal é retificado conforme a equação (26) e para se ter o sinal em magnitude no domínio da frequência, retira-se a raiz quadrada dele. Finalmente, aplica-se a transformada discreta inversa de Fourier (IDFT) combinando a fase, dando origem ao sinal que será utilizado como condição inicial a um dos filtros de Kalman.

Depois da obtenção do sinal de condição inicial oriundo da subtração espectral de potência, utiliza-se o algoritmo de filtro de Kalman de tempo discreto apresentado na seção 2.1 e/ou de Kalman de frequência mostrado na seção 2.2. Lembrando que no filtro de Kalman no domínio da frequência, usa-se o espectro de magnitude do sinal oriundo da subtração espectral como condição inicial. Vale ressaltar que essa combinação deu origem a um sinal sem ruído mais robusto em relação a técnica de subtração espectral no que se diz respeito a minimização do ruído residual conhecido como ruído musical.

3. Resultados e Discussão

Para avaliar os algoritmos apresentados nas seções anteriores, foram utilizados diferentes sinais de voz, amostrados numa taxa de 11025 Hz com 16 bits. A palavra usada nos testes foi “casa”. A janela aplicada nos algoritmos, foi uma janela de Hamming com 256 amostras com sobreposição espectral de 50%. No procedimento de filtragem, o sinal é primeiramente filtrado pelo algoritmo de subtração espectral de potência e em seguida aplicado a um dos filtros de Kalman (FKTD ou FKF).

Para o filtro de Kalman no tempo discreto utilizou-se um modelo auto-regressivo de ordem 10 que apresentou resultados melhores após testes.

Na subtração espectral utilizou-se um valor fixo para $\beta = 0,02$ e o valor de α foi determinado pela equação (27).

Os resultados foram analisados e comparados tendo como base as seguintes medidas: relação sinal/ruído segmentada (SNR_{seg}) e distância de Itakura.

A SNR_{seg} é uma medida de qualidade mais efetiva 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. Esta medida é calculada utilizando a equação (29) a seguir (Deller, Proakis & Hansen, 1993).

$$SNR_{seg} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log_{10} \left[\frac{\sum_{n=mj-N+1}^{mj} x^2(n)}{\sum_{n=mj-N+1}^{mj} v^2(n)} \right], \quad (29)$$

em que: mj representa o limites de cada um dos M quadros de tamanho N .

A SNR_{seg} não fornece uma medida significativa de desempenho quando dois sinais se diferem em seus espectros. Entretanto, as medidas de distância são sensíveis às variações de espectro. Neste caso, a distância de Itakura-Saito fornece melhores resultados e pode ser calculada em relação aos parâmetros de predição linear (LPC) conforme a equação (30) (Deller, Proakis & Hansen, 1993).

$$d(a,b) = \frac{(a-b)^T R(a-b)}{a^T R a}, \quad (30)$$

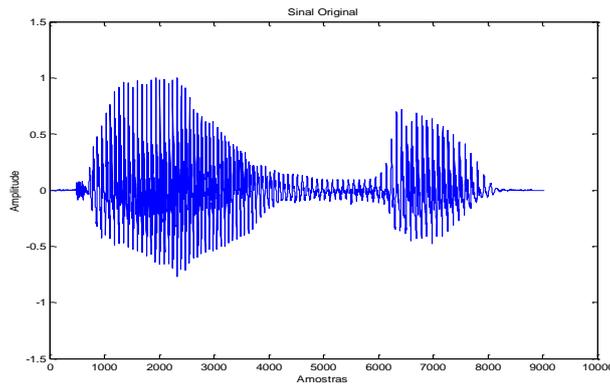
em que: “a” é o vetor de coeficientes LPC do sinal original, “R” é a matriz de autocorrelação do sinal original e “b” é o vetor de coeficientes LPC do sinal estimado. Neste trabalho utilizou-se 16 coeficientes para o preditor por ter também apresentado resultados melhores após os testes realizados.

Para gerar o ruído colorido, foi utilizado um filtro AR de ordem 8, tomando como base o trabalho de Ning, Bouchard & Goubran (2004).

$$\begin{aligned} v(n) = & -0.0851v(n-1) + 0.019126v(n-2) \\ & + 0.0458v(n-3) + 0.0229v(n-4) \\ & + 0.1057v(n-5) + 0.1553v(n-6) \\ & - 0.132v(n-7) - 0.76v(n-8) + w(n) \end{aligned} \quad (31)$$

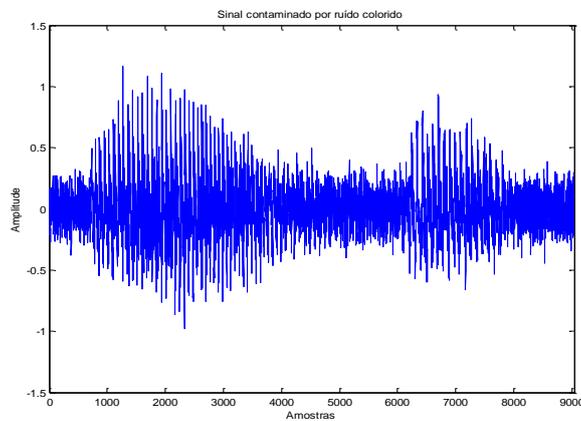
Para a execução dos testes foi utilizado o software MatlabR2013, sendo que no primeiro teste executado, o sinal de voz foi contaminado com ruído colorido uma vez que os fenômenos encontrados na natureza nem sempre possuem características estacionárias como os encontrados em ruído branco. Neste teste a relação sinal ruído de entrada segmentada (SNR_{segE}) foi de 3 dB. Para este teste, utilizou-se o algoritmo referente ao Filtro de Kalman de tempo discreto (FKTD), usando como condição inicial a subtração espectral de potência. As figuras 2, 3 e 4 mostram respectivamente, a forma de onda do sinal sem ruído, sinal contaminado por ruído colorido e sinal estimado.

Figura 2: Sinal de voz sem ruído.



Fonte: Autor.

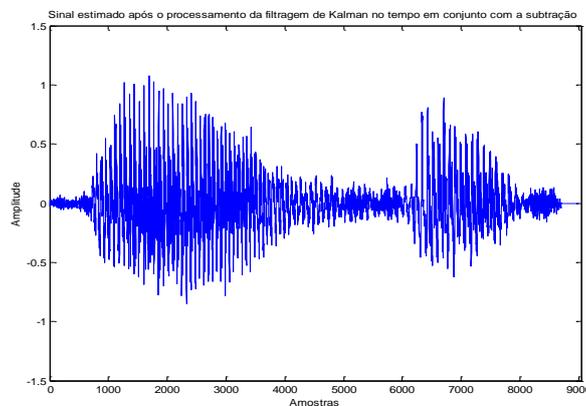
Figura 3: Sinal de voz contaminado por ruído colorido.



Fonte: Autor.

A distorção calculada utilizando a distância de Itakura resultou num valor igual a 0,21 e a relação sinal-ruído segmentada de saída (SNRsegS) obteve um aumento de 2 dB, ou seja, a SNRsegS foi de 5dB.

Figura 4: Sinal de voz estimado após o processamento por ruído colorido por FKTD.

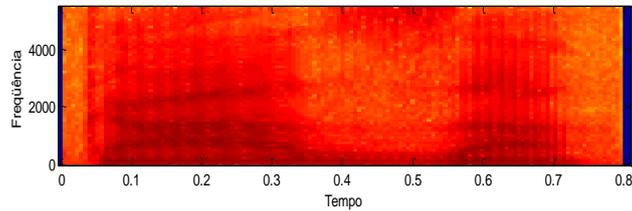


Fonte: Autor.

Além disso, uma análise do espectrograma da figura 6 (sinal estimado), permite a observação de um ruído residual conhecido como ruído musical (Loizou, 2013). Sendo assim, os espectrogramas do sinal sem ruído, com ruído colorido e sinal

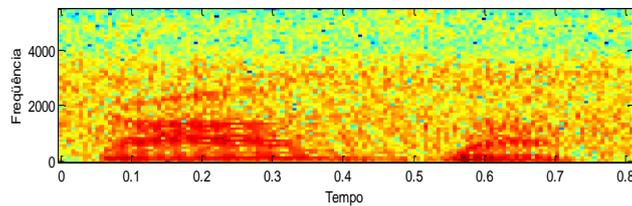
estimado são mostrados nas Figuras 5, 6 e 7, respectivamente. Por estes, percebe-se que as frequências entre 2000 e 3000Hz foram realçadas no espectrograma da Figura 7 em comparação com o espectrograma do sinal sem ruído da Figura 6.

Figura 5: Espectrograma do sinal sem ruído.



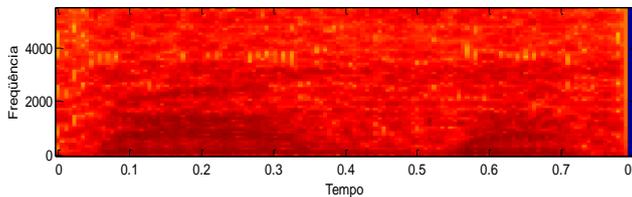
Fonte: Autor.

Figura 6: Espectrograma do sinal contaminado com ruído colorido.



Fonte: Autor.

Figura 7: Espectrograma do sinal estimado após o processamento.

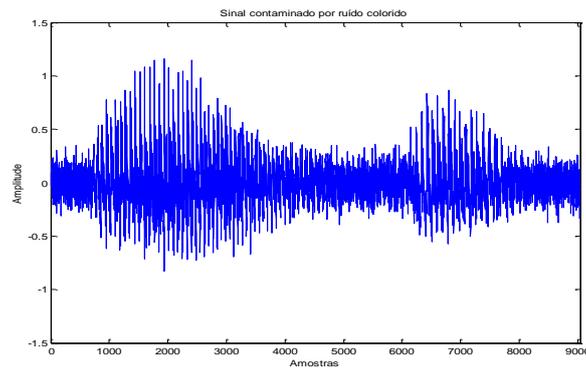


Fonte: Autor.

No segundo teste, o algoritmo usado é o filtro de Kalman na frequência (FKF), usando como condição inicial a subtração espectral de potência. Assim como no primeiro teste, foi adicionado ao sinal sem ruído o ruído colorido com uma relação sinal ruído de entrada segmentada (SNRsegE) de 3dB. Após a simulação foram obtidos resultados com uma distância de Itakura com valor de 0,24. Já este algoritmo como no primeiro teste, resultou em um aumento também na SNRseg de saída (SNRsegS) de 2 dB. Os resultados em relação aos sinais contaminados, estimados e os espectrogramas são apresentados nas Figuras 8, 9 e 10.

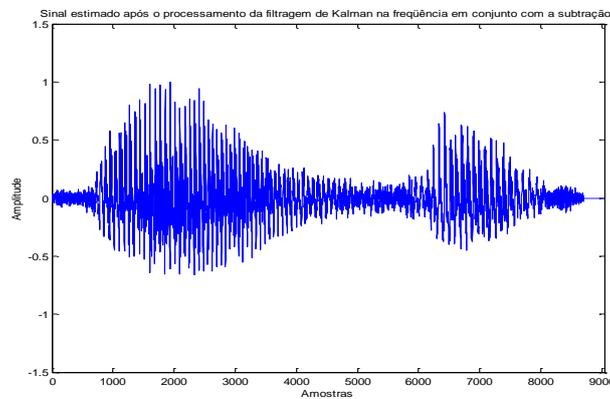
Analisando o espectrograma da Figura 10 em relação ao da Figura 5, percebe-se também que o algoritmo foi capaz de recuperar as frequências de 2000 e 3000Hz como o realizado pelo processamento de FKT, porém verifica-se também um ruído residual.

Figura 8: Sinal contaminado com ruído colorido.



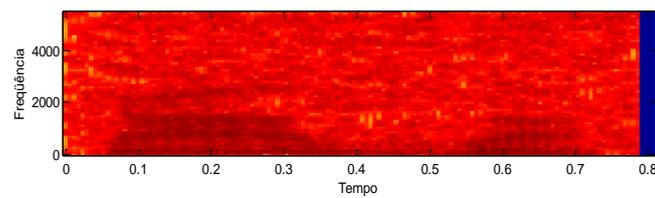
Fonte: Autor.

Figura 9: Sinal estimado após o processamento usando FKF.



Fonte: Autor.

Figura 10: Espectrograma do sinal estimado usando FKF.



Fonte: Autor.

Em suma verifica-se que a relação sinal ruído de saída segmentada é a mesma para ambos os algoritmos, porém a distorção espectral de FKF é maior em relação a técnica que usou o algoritmo FKTD. A Tabela 1, apresenta estes comparativos.

Vale ressaltar que uma distância Itakura-Saito igual a zero indica que os espectros são iguais, ou seja, quanto menor for o valor, menor é a distorção espectral.

Tabela 1: Comparativo entre FKTD e FKF para sinais contaminados com ruído colorido.

Medidas	FKTD	FKF
SNRsegE	3dB	3dB
SNRsegS	5dB	5dB
d(a,b)	0,21	0,24

Fonte: Autor.

Contudo é importante relatar aqui que outros testes foram realizados para outras relações sinal ruído de entrada e todos apresentaram o mesmo entendimento da distorção espectral, ou seja, a técnica de FKTD é melhor em relação a FKF. No entanto, com relações sinal ruído de entrada mais altas, acima de 3dB, percebe-se uma pequena melhora da técnica FKF em relação ao parâmetro de relação sinal ruído de saída, mas a distorção espectral quando comparada a técnica de FKTD se torna melhor frente a FKF como mostrado na Tabela 2.

Tabela 2: Resultados dos testes para a palavra “casa” com SNRsegE de 0 dB, 3 dB, 6 dB, 9 dB e 12 dB, para o ruído ser colorido.

SNRsegE	0 dB	3 dB	6 dB	9 dB	12 dB
FKTD – SNRsegS	1 dB	5 dB	7 dB	10 dB	12 dB
d(a,b)	0,62	0,21	0,17	0,13	0,11
FKF – SNRsegS	1 dB	5 dB	8 dB	11 dB	13 dB
d(a,b)	0,83	0,24	0,22	0,15	0,12

Fonte: Autor.

Para comprovar os resultados, foram realizados também testes subjetivos. Em ambos os sinais filtrados houve uma redução considerável da intensidade do ruído aditivo, mas percebe-se a presença de um ruído residual conhecido como ruído musical.

4. Conclusão

Neste trabalho foi apresentado o estudo e implementação de dois algoritmos. Os algoritmos foram o filtro de Kalman no tempo discreto (FKTD), usando como condição inicial a subtração espectral de potência, o filtro de Kalman na frequência (FKF) que também utilizou o espectro de magnitude da subtração como condição inicial. A comprovação da eficiência na redução de ruído dos dois algoritmos foram avaliadas pelas relações sinal ruído de saída SNRsegS, que mostraram que para relações sinal ruído de entrada de 0 e 3dB, as SNRsegS se mantiveram iguais e, para valores acima de 3dB, a SNRsegS para a técnica de FKF se mostrou um pouco melhor em relação aos seus resultados. Porém, a técnica que se utilizou do filtro de Kalman no tempo (FKTD) apresentou resultados melhores em relação a distorção espectral quando comparada com o filtro de Kalman na frequência (FKF).

Referências

- Berouti, M., Schwartz, R., & Makhoul, J. (1979, April). Enhancement of speech corrupted by acoustic noise. In ICASSP'79. *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 4, pp. 208-211). IEEE.
- Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on acoustics, speech, and signal processing*, 27(2), 113-120.
- Brown, R. G., & Hwang, P. Y. (1997). Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions. Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions.
- da Silva, L. A., Carrijo, G. A., Vasconcelos, E. S., Campos, R. D., Goulart, C. S., & Lemos, R. P. (2018). Comparative Study between the Discrete-Frequency Kalman Filtering and the Discrete-Time Kalman Filtering with Application in Noise Reduction in Speech Signals. *Journal of Electrical and Computer Engineering*, 2018.
- da Silva, L. A., Gigo, W., Costa, S. S. D. O. J., Ferreira, J. C. & Júnior, A. C. L. (2021). A transformada wavelet aplicada na redução de ruído em sinais de voz. *Jornal de Engenharia e Meio Ambiente - JETMA*, 6(2), 1-4.
- de Abreu, J. A. P., de Oliveira, R. C. L., & da Fonseca Neto, J. V. (2020). Rocket tracking impact point prediction using α - β , standard Kalman, extended, Kalman, and unscented Kalman filters: a comparative analysis. *Research, Society and Development*, 9(3), 20.
- Deller Jr, J. R. (1993). Discrete-time processing of speech signals. In *Discrete-time processing of speech signals* (pp. 908-908).

- Fujimoto, M., & Arik, Y. (2000, June). Noisy speech recognition using noise reduction method based on Kalman filter. In 2000 *IEEE international conference on acoustics, speech, and signal processing*. Proceedings (Cat. No. 00CH37100) (Vol. 3, pp. 1727-1730). IEEE.
- George, A. E., So, S., Ghosh, R., & Paliwal, K. K. (2018). Robustness metric-based tuning of the augmented Kalman filter for the enhancement of speech corrupted with coloured noise. *Speech Communication*, 105, 62-76.
- Jakati, J. S., & Kuntoji, S. S. (2021). A noise reduction method based on modified lms algorithm of real time speech signals. *WSEAS Transactions on Systems and Control*, 16, 162-170.
- Koning, R., Bruce, I. C., Denys, S., & Wouters, J. (2018). Perceptual and model-based evaluation of ideal time-frequency noise reduction in hearing-impaired listeners. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3), 687-697.
- Koning, R., Madhu, N., & Wouters, J. (2014). Ideal time–frequency masking algorithms lead to different speech intelligibility and quality in normal-hearing and cochlear implant listeners. *IEEE Transactions on Biomedical Engineering*, 62(1), 331-341.
- Loizou, P. C. (2013). *Speech enhancement: theory and practice*. CRC press.
- Ma, N., Bouchard, M., & Goubran, R. A. (2004, May). Perceptual Kalman filtering for speech enhancement in colored noise. In 2004 *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 1-717). IEEE.
- Madhu, N., Spriet, A., Jansen, S., Koning, R., & Wouters, J. (2012). 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*, 21(1), 63-72.
- Mourad, T. (2022). Speech enhancement based on stationary bionic wavelet transform and maximum a posterior estimator of magnitude-squared spectrum. In *The Stationary Bionic Wavelet Transform and its Applications for ECG and Speech Processing* (pp. 1-30). Springer, Cham.
- Mourão, G. L. (2019). Avaliação de técnicas de redução de ruído para implantes cocleares. *Master's thesis, Federal University of Santa Catarina*
- Panda, B., Nayak, S. K., & Mohanty, M. N. (2021). Noise Suppression in Non-stationary Signals Using Adaptive Techniques. In *Advances in Electronics, Communication and Computing* (pp. 261-270). Springer, Singapore.
- Vaseghi, S. V. (2008). *Advanced digital signal processing and noise reduction*. John Wiley & Sons.
- Shao, Y., & Chang, C. H. (2006, May). A Kalman filter based on wavelet filter-bank and psychoacoustic modeling for speech enhancement. In 2006 *IEEE International Symposium on Circuits and Systems* (pp. 4-pp). IEEE.