UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

GABRIEL MOCELIN GUAJARDO CUEVAS

ALGORITMO DE REDUÇÃO DE RUÍDO EM SINAIS DE FALA BASEADO EM FILTROS ESTOCÁSTICOS

CURITIBA

2024

GABRIEL MOCELIN GUAJARDO CUEVAS

ALGORITMO DE REDUÇÃO DE RUÍDO EM SINAIS DE FALA BASEADO EM FILTROS ESTOCÁSTICOS

Speech Enhancement Algorithm Based on Stochastic Filters

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia de Controle e Automação do curso de Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Victor Baptista Frencl

Coorientador: Prof. Dr. Marcelo de Oliveira Rosa

CURITIBA

2024

4.0 Internacional

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

GABRIEL MOCELIN GUAJARDO CUEVAS

ALGORITMO DE REDUÇÃO DE RUÍDO EM SINAIS DE FALA BASEADO EM FILTROS ESTOCÁSTICOS

Trabalho de conclusão de curso de graduação apresentado como requisito para obtenção do título de Bacharel em Engenharia de Controle e Automação do curso de Engenharia de Controle e Automação da Universidade Tecnológica Federal do Paraná.

Data de aprovação: 12/junho/2024

Glauber Gomes de Oliveira Brante
Doutorado
Universidade Tecnológica Federal do Paraná

Marcelo de Oliveira Rosa
Doutorado
Universidade Tecnológica Federal do Paraná

Renata Coelho Borges
Doutorado
Universidade Tecnológica Federal do Paraná

Victor Baptista Frencl
Doutorado
Universidade Tecnológica Federal do Paraná

CURITIBA

2024

RESUMO

O presente trabalho de conclusão de curso aborda o desenvolvimento de um algoritmo de redução de ruído em sinais de fala com base em filtros estocásticos, visando mitigar a interferência de diversos tipos de ruído que comprometem a qualidade do sinal. Com base na revisão da literatura, são desenvolvidos algoritmos que empregam Filtros de Kalman (KF) cujos modelos Autorregressivos (AR) são obtidos através de predição linear, aliados a outras técnicas de processamento de sinais digitais, a fim de modelar o complexo comportamento da fala. Além disso, propõe-se o aprimoramento desses algoritmos por meio da implementação do filtro de Múltiplos Modelos Interativos (IMM). Os resultados deste estudo indicam que os algoritmos desenvolvidos são eficazes na redução de ruídos gaussianos, assim como ruídos coloridos, e destaca a capacidade dos algoritmos baseados em IMM em aprimorar a precisão e inteligibilidade dos sinais em relação aos algoritmos baseados em KF.

Palavras-chave: filtro de kalman; filtro de múltiplos modelos interativos; predição linear; aprimoramento de sinais de fala.

ABSTRACT

This undergraduate research project focuses on developing a noise reduction algorithm for speech signals based on stochastic filters, aiming to mitigate the interference of various types of noise that compromise the signal quality. Based in a literature review, algorithms are developed that employ Kalman Filters (KF) whose Autoregressive (AR) models are obtained through linear prediction that together with other digital signal processing techniques are dimensioned to model the complex behavior of speech. Additionally, the improvement of these algorithms is proposed through the implementation of the Interactive Multiple Models (IMM) filter. The results of this study indicate that the developed algorithms are effective in reducing Gaussian and colored noise, and highlight the ability of IMM-based algorithms to improve the accuracy and intelligibility of signals compared to KF-based algorithms.

Keywords: kalman filter; interactive multiple models; linear prediction; speech enhancement.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

Siglas

AR	Autorregressivo
DFT	Transformada de Fourier Discreta
EKF	Filtro de Kalman Estendido
HMM	Modelo Oculto de Markov
IMM	Filtro de Múltiplos Modelos Interativos
IMMO	Filtro de Múltiplos Modelos Interativos Oráculo
IMMIt	Filtro de Múltiplos Modelos Interativos Iterativo
IMMIt	Filtro de Múltiplos Modelos Interativos Iterativo Aumentado
KF	Filtro de Kalman
KFO	Filtro de Kalman Oráculo
KFIt	Filtro de Kalman Iterativo
KFItA	Filtro de Kalman Iterativo Aumentado
LPC	Predição Linear
MA	Média Móvel
RMSE	Raiz do Erro Médio Quadrático
SNR	Razão Sinal-Ruído
STFT	Transformada de Fourier de Curto Termo
PESQ	Avaliação Perceptual de Qualidade de Fala
UTFPR	Universidade Tecnológica Federal do Paraná
WF	Filtro *Whitening*

SUMÁRIO

1 INTRODUÇÃO

1.1 TEMA

A fala caracteriza-se como uma das maneiras mais importantes de comunicação humana. Com a chegada de tecnologias de captação, gravação e reprodução de áudio, além do avanço das telecomunicações, a fala transcende suas limitações físicas iniciais, permitindo que ela seja transmitida e reproduzida independentemente do tempo e do espaço de sua emissão. Atualmente, o processo de captura de som mais usual consiste na digitalização da captura analógica de equipamentos como microfones, captadores magnéticos e piezoelétricos (ZÖLZER, 2008). A partir dessa tecnologia, viabiliza-se mídias como *podcasts*, músicas e vídeos de maneira que possam ser facilmente transmitidos e gravados.

Destaca-se, entretanto, que por mais refinado que seja o sistema de captação e conversão de áudio, este sempre estará sujeito a ruídos. Estes ruídos estão relacionados aos componentes eletrônicos dos equipamentos de áudio, a distúrbios eletromagnéticos, assim como de origem acústica no ambiente de captação (ZHENG *et al.*, 2020).

Caso haja uma grande presença de ruídos atuando no sistema em questão, o sinal de áudio fica imerso na informação das perturbações, comprometendo sua clareza ao escutá-lo. Dessa forma, torna-se interessante a aplicação de técnicas de supressão desses ruídos, de forma a permitir uma melhor inteligibilidade do sinal – não só para humanos, mas também algoritmos de reconhecimento de voz e fala (XIA; WEI, 2016).

Tradicionalmente, a redução do ruído é feita por meio da da supressão de frequências, tanto com a aplicação de filtros tradicionais – como os filtros passa-banda, passa-alta ou passa-baixa – quanto métodos mais avançados, como o método de subtração espectral (YAN *et al.*, 2020). Apesar de muitas vezes efetivas, essas técnicas podem distorcer significativamente o sinal desejado, já que tal sinal possui uma complexidade harmônica que acaba sendo também eliminada, além de elas não serem suficientes em casos em que o ruído é não estacionário (UPADHYAY; KARMAKAR, 2015). Assim, incentiva-se outras soluções para esse problema, como o desenvolvimento de algoritmos baseados no filtro de Wiener (JAISWAL; YEDURI; CENKERAMADDI, 2022) ou redes neurais artificiais (TAMMEN *et al.*, 2020; REHR; GERKMANN, 2021).

Outra possível abordagem da redução de ruídos em sinais de áudio é a utilização de filtros estocásticos, mais notavelmente o Filtro de Kalman (KF, do inglês *Kalman Filter*) e suas variações (GANNOT, 2012). O KF foi proposto por Rudolf Emil Kalman em 1960 e tem como objetivo a eliminação de ruídos gaussianos a partir de uma estimação ótima sob hipóteses de linearidade do modelo matemático e de ruídos aleatórios aditivos com distribuição gaussiana (KALMAN, 1960).

1.1.1 Delimitação do Tema

Com a utilização de medidas digitais obtidas ao longo do tempo, o KF visa a reduzir o ruído aleatório durante todo o processo de mensuração, adequando os valores captados a um sistema representado em espaço de estados previamente definido (KALMAN, 1960). Realiza-se, portanto, um processo de estimativa de estados para cada nova medida obtida, resultando em um sinal filtrado. Para cada novo valor medido do sinal, o KF se ajusta em relação a seus parâmetros de covariância, melhorando, assim, suas estimativas nos próximos instantes de tempo. Sua principal limitação, entretanto, vem do fato de ser operável somente com modelos lineares e ruídos gaussianos aditivos, os quais não são aplicáveis para todas condições de ruído existentes na prática.

O KF teve sua primeira aplicação na área de áudio em Paliwal e Basu (1987), em que o filtro fora dimensionado com o objetivo de reduzir os ruídos presentes nos sinais de fala. Nesse estudo, foi considerado que a voz pode ser aproximada por modelos Autorregressivos (AR), ou seja, o valor estimado é dado por uma combinação linear entre a amostra atual e as amostras passadas. O algoritmo proposto produziu resultados bons quando comparado a outros métodos, como o filtro de Wiener, no que diz respeito a ruídos brancos.

Esse princípio foi subsequentemente aprimorado em trabalhos como o de Gibson, Koo e Gray (1989) e em Popescu e Zeljkovic (1998), nos quais foram propostos métodos para a redução de ruídos coloridos. Modelagens não lineares também foram empregadas em conjunto com a aplicação do KF Estendido, conforme apresentado por Rigoll (1986) e Wan e Nelson (1998).

Outras aplicações do KF em áudio incluem a remoção de reverberação (SCHWARTZ; GANNOT; HABETS, 2014), reconhecimento de fala (GOH; RAVEENDRAN; GOH, 2015), sistemas de detecção de frequências fundamentais (SALOR; DEMIREKLER; ORGUNER, 2006) e recuperação de arquivos de áudio (CANAZZA; POLI; MIAN, 2009) .

Uma possibilidade de aprimorar os resultados obtidos por KF é a adoção do filtro de Múltiplos Modelos Interativos (IMM, do inglês *Interacting Multiple-Model*), que consiste em um método que utiliza de diversos filtros estocásticos em paralelo, com cada um dos filtros associados a um diferente modelo dinâmico e diferentes parâmetros de ruído (BLOM; BAR-SHALOM, 1988). Sob a hipótese de comportamento markoviano de probabilidade de transição entre os modelos, é possível ponderar a estimação de cada um dos filtros que compõem o filtro IMM, produzindo uma estimativa global a cada iteração do filtro. O objetivo do filtro IMM é permitir que diversos comportamentos possam ser analisados simultaneamente, permitindo assim uma maior versatilidade e precisão das estimativas obtidas pelo filtro.

O objetivo desse trabalho é o desenvolvimento de filtros estocásticos lineares que sejam capazes de reduzir o ruído presente em sinais de fala através da aplicação de técnicas já desenvolvidas na literatura, assim como a aplicação do IMM que visa aprimorar os resultados obtidos

pelos algoritmos baseados em KF através da aplicação do filtro para adaptação dinâmica entre janelas de tempo.

1.2 PROBLEMATIZAÇÃO

Para a realização de modelagens dos sinais, encontra-se como uma das principais dificuldades as diversas formas com as quais a voz pode ser pronunciada, além de suas variações de frequência e timbre (RABINER; SCHAFER, 2007). Portanto, é possível dizer que não se tem na voz um comportamento constante, tornando-se necessárias abordagens tal qual a identificação de sistemas, que deve ser iterada para diversos instantes de tempo. Considerando que os algoritmos são voltados pra abordar situações não ideais, a identificação de sistema é comprometida visto que não se tem acesso a um sinal ideal para performá-la. Para que os algoritmos estocásticos sejam eficientes, é necessário, portanto, encontrar e/ou elaborar técnicas que aprimorem a identificação de sistemas, desenviesando o modelo matemático e, consequentemente, o desempenho dos filtros estocásticos.

É necessário, também, sintonizar um filtro estocástico de maneira que esse possa suprimir significativamente os ruídos aleatórios imersos no modelo, além de reduzir deformações de frequência e de fase causadas nos sinais de fala durante a filtragem. Para isso, é necessário uma além de uma modelagem adequada assim como realizar a otimização dos parâmetros de covariância dos ruídos aditivos de processo e de medidas.

Para que os algoritmos desenvolvidos sejam aplicáveis em contextos reais, é essencial que eles sejam capazes de prever perturbações que vão além dos ruídos brancos, abrangendo também ruídos coloridos. Isso inclui, por exemplo, perturbações geradas pela captação de áudio em ambientes ruidosos. Para isso se tornar possível, é necessário utilizar abordagens de identificação não só relacionado à fala, mas também aos ruídos presentes no sistema, como visto em Gibson, Koo e Gray (1989) e Malik e Benesty (2013).

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Aplicar e aprimorar os algoritmos desenvolvidos em trabalhos encontrados na literatura a partir de uma revisão literária de maneira a aproveitar as suas melhores características e testar sua eficiência na redução de ruídos presentes em gravações de fala, e melhorar esses resultados através da aplicação do filtro IMM.

1.3.2 Objetivos Específicos

- Efetuar uma revisão bibliográfica que contemple os tópicos de fala e seus modelos matemáticos, filtros estocásticos e suas aplicações no tema de redução de ruído em áudio;
- Definir modelos matemáticos, de acordo com a literatura especializada adequados ao comportamento do sinal da voz humana;
- Desenvolver um filtro IMM adequado para trabalhar em conjunto com as técnicas e modelos estabelecidos;
- Avaliar o desempenho dos filtros estocásticos em sinais sujeitos a diversos tipos de ruído.
- Analisar o desempenho dos algoritmos desenvolvidos através de mensurações de razão sinal-ruído (SNR, do inglês *Signal-to-Noise Ratio*) (JOHNSON, 2006) e da métrica de avaliação perceptual de qualidade de fala (PESQ, do inglês *Perceptual Evaluation of Speech Quality*) (BEERENDS *et al.*, 2002) das estimativas.

1.4 JUSTIFICATIVA

Com um maior consumo mundial de mídias como *audiobooks* e *podcasts* nos últimos anos (SNELLING, 2021; THOMPSON; WELDON, 2022), as técnicas otimizadas de supressão de ruído tendem a ser cada vez mais requeridas.

Uma das vantagens encontradas na utilização de filtros estocásticos, tais quais o KF e o filtro IMM, tratar-se da baixa complexidade computacional. Além disso, esses filtros possibilitam a estimativa do sinal amostra a amostra, sem a necessidade de conversões para o domínio da frequência, e oferecem uma grande flexibilidade em termos de abordagens metodológicas. Levando em conta outras possíveis aplicações para tratamento de áudio encontradas na literatura, torna-se evidente o quanto o filtro estocástico é uma ferramenta que pode ser útil nessa área.

Esse trabalho visa a entender a aplicação e as limitações desses algoritmos, que, considerando os estudos recentes sobre aplicação da filtragem estocástica na área de redução de ruídos em sinais de voz (ROY; NICOLSON; PALIWAL, 2020; ROY; PALIWAL, 2021), é possível dizer que é uma área que possui grande possibilidade de aprimoramento e aprofundamento, como será o foco desse trabalho.

1.5 PROCEDIMENTOS METODOLÓGICOS

Com base na literatura especializada, serão estudadas técnicas de identificação de modelos aplicáveis para filtragem estocástica do sinal de fala e que também possam servir para

eliminação de ruídos coloridos, assim como maneiras de se desenviesar os modelos identificados e encontrar as melhores soluções adotáveis para aprimorar o desempenho dos filtros estocásticos.

Serão desenvolvidos algoritmos em MATLAB® que tendem a incorporar funcionalidades importantes vistas na literatura, assim como o desenvolvimento de um novo filtro baseado nas técnicas de IMM.

Com base no banco de dados *NOIZEUS Corpus* (HU; LOIZOU, 2007), será feita a estimação de gravações ruidosas por meio de filtro estocásticos e comparar os resultados com as gravações sem ruído com o intuito de avaliar suas performances. Essa verificação será realizada com base na análise de SNR e de PESQ, assim como nas análises gráficas dos comportamentos dos sinais de saída.

1.6 ESTRUTURA DO TRABALHO

A estrutura do TCC baseia-se em sete capítulos, sendo eles, em ordem: “Introdução”; “Filtros Estocásticos”; “Métodos de Análise de Resultados”; “Sinal de Fala e seus Modelos”; “Filtros Estocásticos para o Tratamento do Sinal de Fala”; “Resultados e Discussões”; e “Conclusões”.

Neste primeiro capítulo foi apresentado e delimitado o escopo do estudo realizado, contextualizando o tema, objetivos, justificativas, problematização, metodologia da realização do trabalho. Os dois capítulos seguintes tratarão sobre a revisão bibliográfica realizada a fim de embasar os aspectos construtivos do algoritmo proposto. em “Sinal de Fala e seus Modelos” será abordado o fenômeno da voz humana e avaliado quais os procedimentos necessários para determinação dos modelos matemáticos que serão utilizados nos métodos de filtragem estocástica. Já no capítulo “Filtros Estocásticos” serão tratados os princípios de funcionamento e os tipos de filtro estocásticos considerados para aplicação

No capítulo “Métodos de Análise de Resultados” explicará as métricas avaliativas que serão utilizadas ao longo do trabalho. Em “Filtros Estocásticos para o Tratamento do Sinal de Fala”, tratar-se-á sobre toda a implementação computacional, informando sobre as principais técnicas utilizadas, assim como a esquematização do seu funcionamento como um todo. Então, em “Resultados e Discussões” haverá a análise quantitativa da performance do algoritmo, avaliando os algoritmos, sistematicamente, em relação aos comportamentos dos sinais estimados. Por fim, o último capítulo apresentará as considerações finais e se houve um cumprimento dos objetivos estabelecidos, além de propor sugestões para trabalhos futuros.

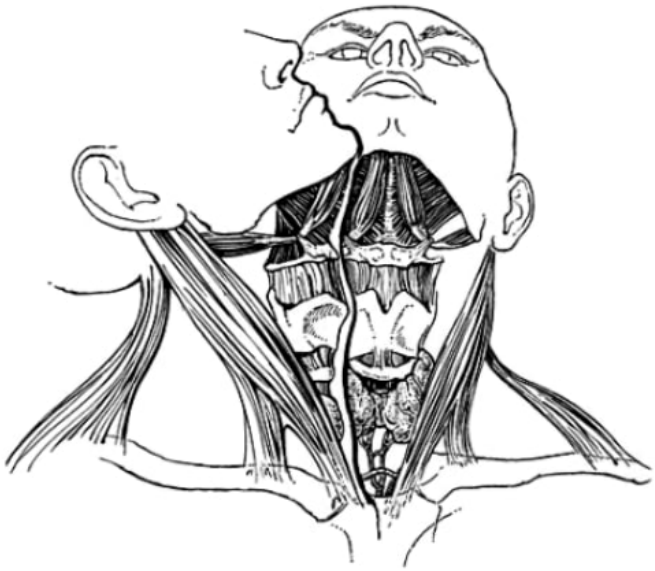
2 SINAL DE FALA E SEUS MODELOS

Neste capítulo serão contextualizados os fenômenos relacionados à produção de voz pelo corpo humano juntamente às possibilidades que se têm de modelar o fenômeno matematicamente a partir do sinal de áudio.

2.1 VOZ HUMANA

De acordo com Rosa (2002), a fala é um fenômeno produzido pelo sistema respiratório humano, como resultado do trabalho conjunto de órgãos como pulmões, traqueia e, em especial, a laringe; além de estruturas da boca como língua, lábios e dentes. A Figura 1 ilustra o posicionamento da laringe na garganta.

Figura 1 – Posição da laringe na garganta.



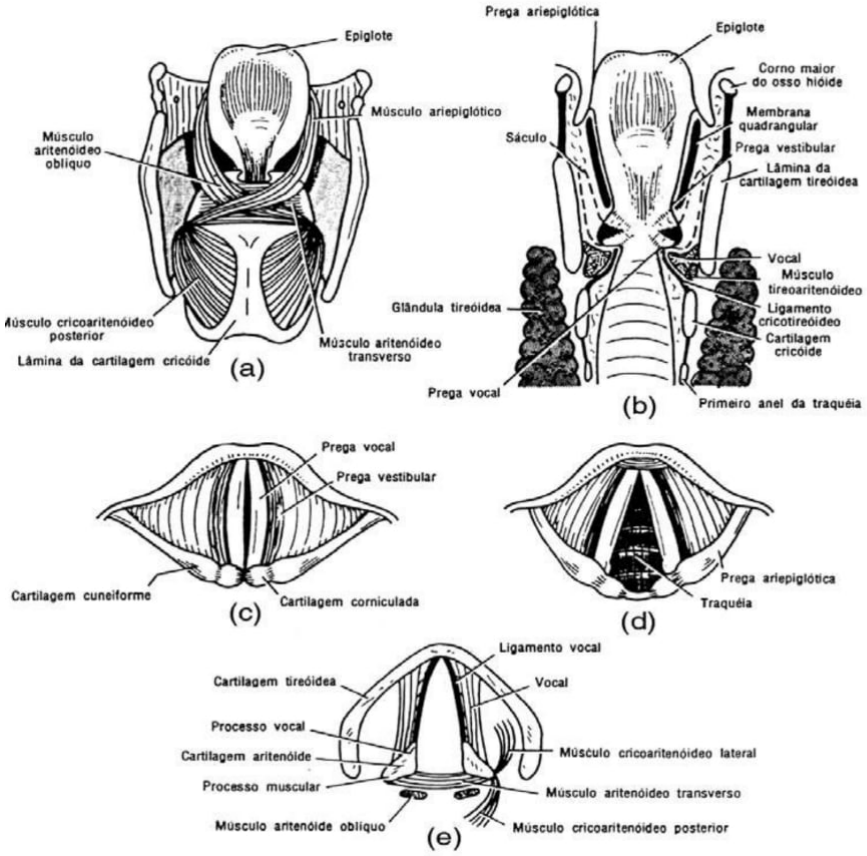
Fonte: (ROSA, 2002).

Os pulmões e a traqueia desempenham um papel crucial na produção do fluxo de ar, que fornecerá a energia necessária para a geração sonora pelo sistema vocal (ZHANG, 2016). A produção do fluxo de ar ocorre por meio da contração e relaxamento dos músculos pulmonares. Durante a fase de relaxamento, ocorre a inspiração do ar, enquanto na fase de contração, ocorre a expiração do ar. Ou seja, no processo de inspiração, realiza-se um esforço ativo dos músculos, enquanto na expiração ocorre com o relaxamento desses mesmos músculos – apesar dessa ação também poder ocorrer de maneira forçada, em casos em que há a necessidade de um maior escoamento de ar.

A laringe é considerada o local onde as principais características do sinal de voz são geradas, especialmente sua frequência fundamental. A voz é uma função secundária da laringe, cuja função primária é proteger o sistema respiratório contra corpos estranhos.

A laringe, uma estrutura muscular e cartilaginosa, é detalhada na Figura 2. Entre os músculos presentes, distinguem-se os extrínsecos, responsáveis pelo movimento do órgão como um todo, e os intrínsecos, que controlam os movimentos internos, como a abertura e fechamento da glote[1]. Dado que a laringe funciona como um controlador de vazão, essa funcionalidade é explorada para gerar a vibração conhecida como sinal glotal, a partir da qual o sinal de voz é derivado. Nesse contexto, os músculos intrínsecos, em especial o tireoaritenoideo, são os principais responsáveis pela produção da voz.

Figura 2 – Músculos e cartilagens da laringe.

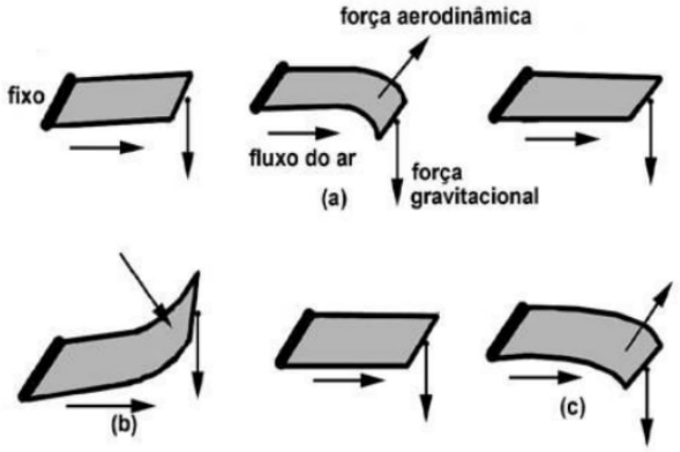


Fonte: (ROSA, 2002).

A produção vocal ocorre devido à vibração das cordas vocais verdadeiras, resultante do fluxo de ar entre suas paredes. De acordo com a teoria mioelástica-aerodinâmica, essa vibração assemelha-se ao efeito produzido quando uma folha de papel é soprada, conforme ilustrado na Figura 3. Características como o tensionamento das cordas vocais, a geometria do espaço glotal e a pressão subglotal contribuem para alterações na frequência e timbre, podendo essas alterações ser ou não controláveis (ROSA, 2002).

[1] Espaço entre as pregas vocais.

Figura 3 – Efeito da passagem de ar sob folha de papel.

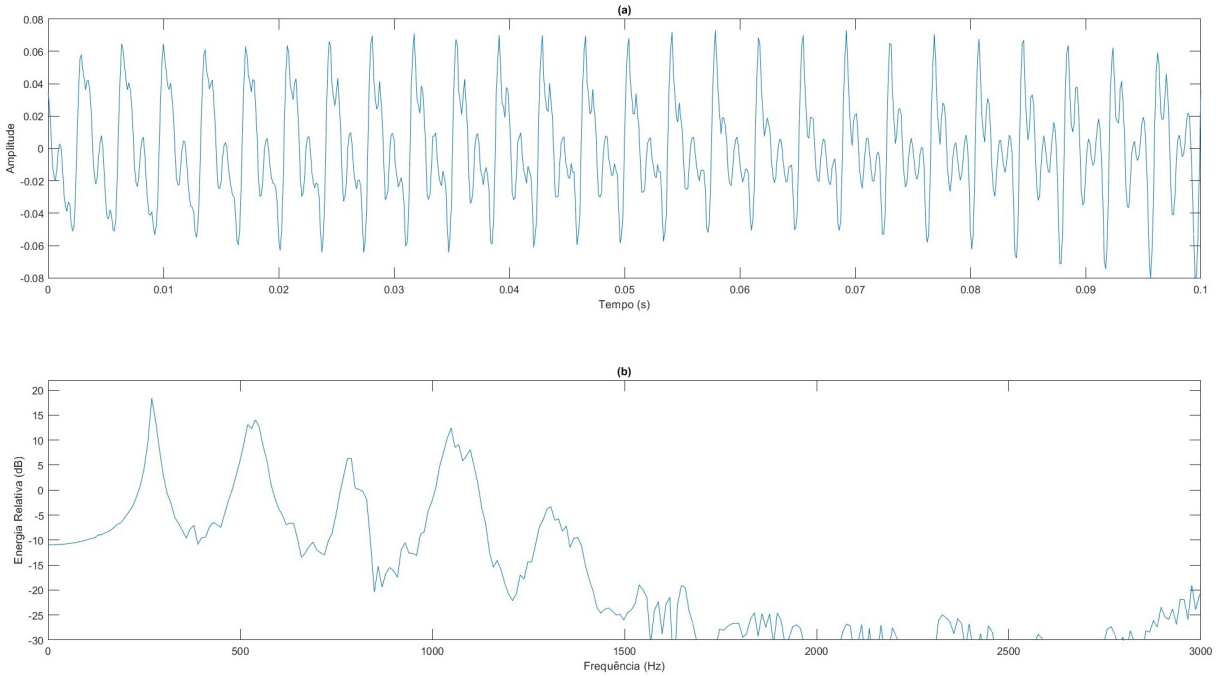


Fonte: (ROSA, 2002).

Visto que o sinal glotal é um sinal de baixa intensidade sonora, o conjunto de órgãos denominado trato vocal é responsável por amplificá-lo. Inclui-se nesse grupo boca, nariz, faringe e a cavidade torácica (DAJER, 2010). A amplificação resulta da ressonância gerada por esses órgãos, que pode ser modificada pelo movimento de língua, lábios e dentes, criando diferentes timbres para o mesmo sinal glotal. É a partir dessa movimentação, portanto, que surgem a variedade de sons dos quais as linguagens faladas são baseadas.

Os sons da fala humana podem ser categorizados em três grupos principais, conforme Rabiner e Schafer (2007): vocálicos, não vocálicos e plosivos. Os sons vocálicos, associados principalmente a vogais, são pulsos quasi-periódicos produzidos pelas cordas vocais, devido ao fluxo de ar na glote. Na Figura 4 é apresentada um trecho de um sinal vocálico representados no domínio do tempo e no domínio da frequência. É evidenciado nessas figuras o comportamento quasi-periódico desse tipo de sinal, dados os claros picos no domínio da frequência e forma de onda que se repete no domínio do tempo.

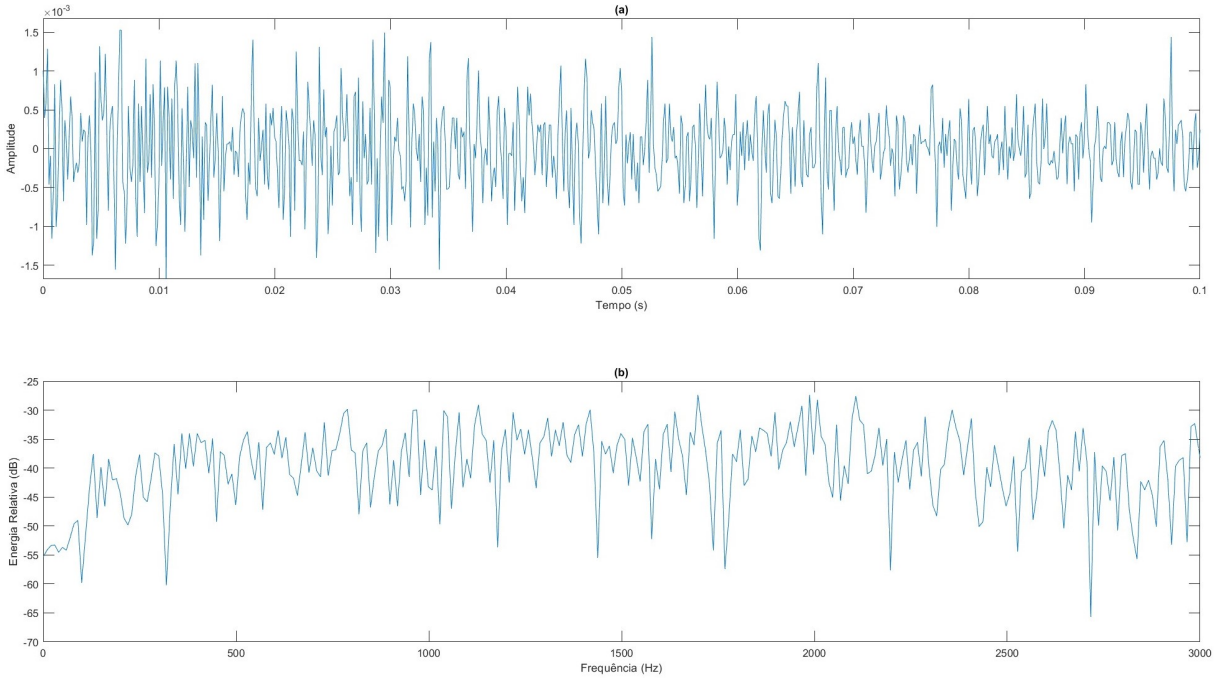
Figura 4 – Trecho de sinal de fala vocálica feminina expressa em (a) no domínio do tempo e em (b) no domínio da frequência.



Fonte: Autoria própria, amostra de áudio do banco de dados Hu e Loizou (2007).

Os sons não vocálicos diferem por não serem gerados pelas cordas vocais, mas sim pelo estreitamento da passagem de ar no trato vocal. Pronúncias de letras como “r” e “s” são exemplos desse tipo de som, cujo comportamento do sinal assemelha-se a um ruído, conforme demonstrado na Figura 5.

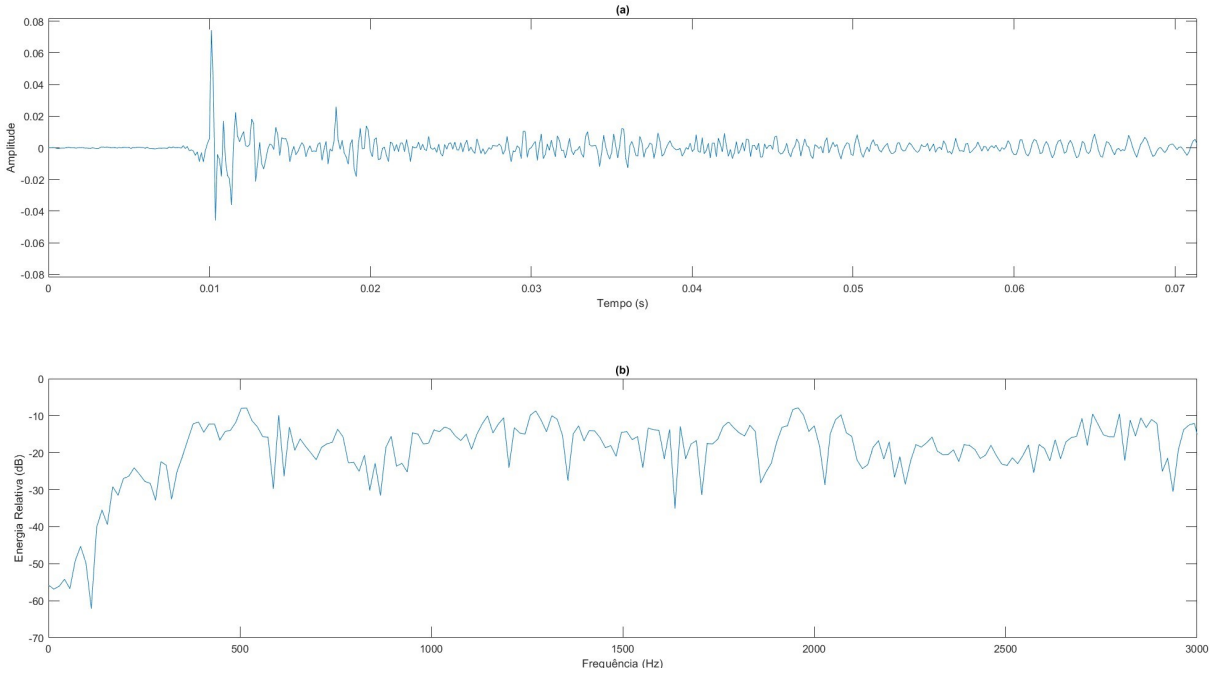
Figura 5 – Trecho de sinal de fala não-vocálica feminina expressa em (a) no domínio do tempo e em (b) no domínio da frequência.



Fonte: Autoria própria, amostra de áudio do banco de dados Hu e Loizou (2007).

Por fim, os sons plosivos resultam do rápido aumento de pressão causado pela liberação de ar após a abertura de algum elemento do trato vocal que estava impedindo a passagem de ar. Esse efeito é perceptível na pronúncia de letras como “p”, “t” e “k”, como demonstrado na Figura 6.

Figura 6 – Trecho de sinal de fala plosiva feminina expressa em (a) no domínio do tempo e em (b) no domínio da frequência.



Fonte: Autoria própria, amostra de áudio do banco de dados Hu e Loizou (2007).

2.2 ÁUDIO DIGITAL

Esta seção tem por objetivo destacar as definições fundamentais relacionadas aos sinais digitais de áudio, visando uma compreensão mais aprofundada das possibilidades e limitações dessa temática no contexto dos sinais de fala.

O termo “áudio digital” refere-se ao sinal adimensional de ondas sonoras convertido de analógico para digital ou a um sinal sintetizado diretamente com o propósito de produção sonora. Este sinal é representado por uma série temporal unidimensional, conforme exemplificado nas Figuras 4 e 5.

Por se tratar de um sinal digital, a frequência de amostragem, que representa a quantidade de amostras coletadas em um intervalo de tempo, é um fator crucial a ser considerado. Conforme estabelecido pelo teorema de Nyquist-Shannon, para a recuperação completa de um sinal analógico a partir de um sinal digital, é necessário que a frequência de amostragem seja, no mínimo, duas vezes maior que a maior frequência presente no sinal (NYQUIST, 1928). Caso contrário, ocorrerá o fenômeno de *aliasing* (TAN; JIANG, 2019).

Dado que a capacidade auditiva humana abrange uma faixa de frequência sonora audível de 20 Hz a 20 kHz, é comum que o sinal de áudio digital tenha frequências de amostragem inferiores a 48 kHz, sendo 44,1 kHz o padrão amplamente adotado na indústria do áudio (RAJAMANI; LAI; FURROW, 2000). No entanto, variações nesses valores são comuns dependendo da aplicação, como em sistemas de telefonia, que frequentemente empregam uma frequência de amostragem de 8 kHz para otimizar a transmissão de dados (RABINER; SCHAFER, 2007). Essa escolha é respaldada pelo fato de que a inteligibilidade do sinal de voz humano está concentrada na faixa de 0 a 4 kHz.

2.3 MODELAGEM DO SINAL DE VOZ

Devido à grande variabilidade de fatores que levam a um sinal de voz se distinguir tanto em relação ao indivíduo quanto em relação a própria pronúncia de palavras, dificulta-se a adoção de um modelo que generalize o comportamento desses sinais. Dessa forma, esta seção apresenta a alternativa da modelagem via autorregressão, que será importante tanto para a filtragem estocástica quanto para remoção de ruídos coloridos.

2.3.1 Modelo AR por Predição Linear

Modelos AR são aqueles que se utilizam de uma relação de amostras passadas de um sinal para predizer o seu comportamento atual. Trata-se de um tipo de modelo amplamente utilizado em áudio, dado que a natureza do sinal costuma possuir comportamentos quasi-periódicos, ou seja, comportamentos próximos a determinísticos. Modelos AR podem ser

descritos, de maneira linear, de acordo com a equação (1).

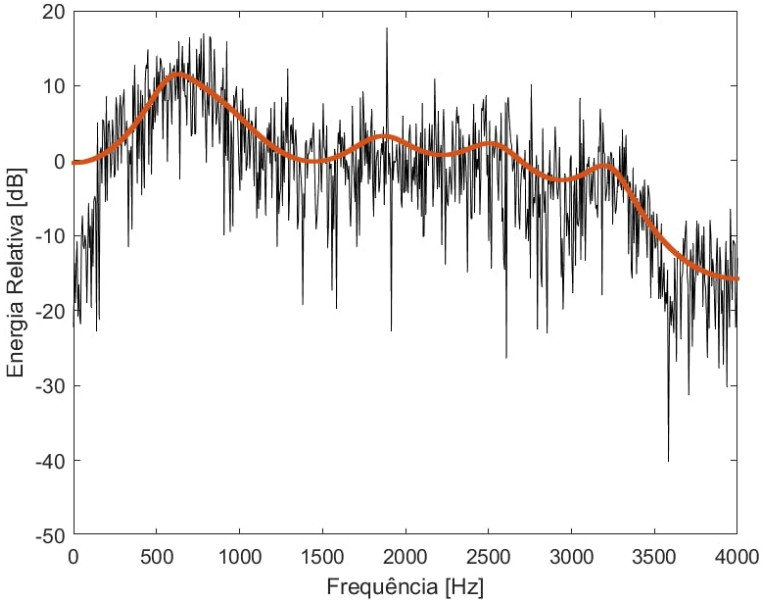
$$s(k) = \sum_{i=1}^{q} a_i s(k-i) + w(k), \tag{1}$$

sendo $s(k)$ o valor do sinal no instante k, q a ordem do modelo, $w(k)$ uma variável aleatória de ruído gaussiano e a_i os coeficientes de predição linear do modelo. Predição linear (LPC, do inglês *Linear Prediction Coding*) é uma de uma técnica que analisa uma série temporal e define os parâmetros de um modelo capaz da reconstrução dessa série e de realizar previsões de estados futuros, sob a condição da série possuir um comportamento determinístico (MAKHOUL, 1975; O'SHAUGHNESSY, 1988).

No domínio da frequência pode-se entender o modelo como uma função de transferência que sua resposta de frequência está com base na energia de cada banda de frequência do sinal analisado, como um envelope que descreve a envoltória do comportamento harmônico do sinal. O quão preciso é essa correspondência limita-se à ordem q adotada para modelagem. O modelo pode também ser entendido como um filtro digital de Resposta Infinita ao Impulso (IIR, do inglês *Infinite Impulse Response*) que, tendo como entrada um sinal de ruído branco, terá como saída um comportamento harmônico próximo à série temporal analisada – efeito que descreve o funcionamento de alguns sintetizadores VOCODER, por exemplo (GRIFFIN; LIM, 1988).

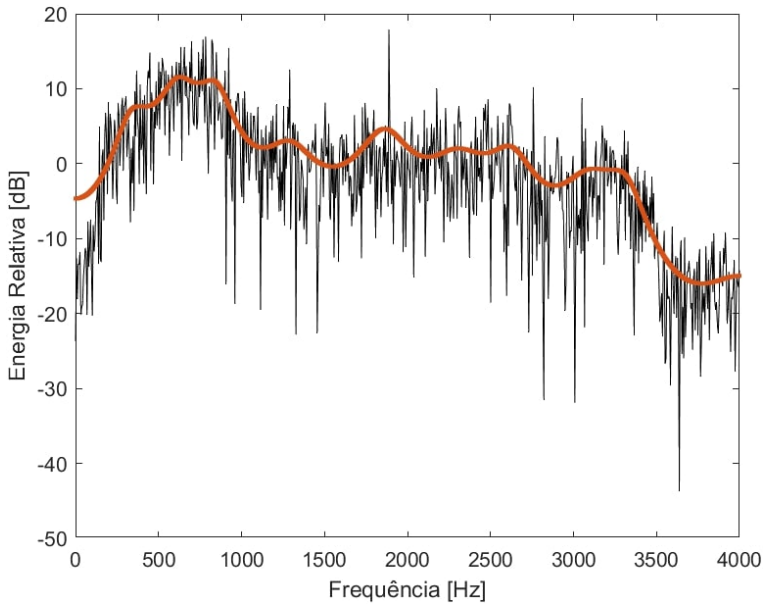
Nas Figuras 7 e 8 observa-se a resposta em frequência do modelo comparado à Transformada de Fourier Discreta (DFT, do inglês *Discrete Fourier Transform*) do sinal de referência, para duas ordens de modelo. O método de obtenção do modelo do sinal é descrito na Seção 2.3.1.1.

Figura 7 – Resposta do LPC de ordem $q = 10$ (laranja) em relação do sinal de referência no domínio da frequência (preto), cuja frequência de amostragem é de 8 kHz.



Fonte: Autoria própria.

Figura 8 – Resposta do LPC de ordem $q = 22$ (laranja) em relação do sinal de referência no domínio da frequência (preto), cuja frequência de amostragem é de 8 kHz.



Fonte: Autoria própria.

Considera-se o modelo AR como um modelo *all-poles* (do inglês, todos polos) dos coeficientes do LPC, pois sua função de transferência pode ser descrita como a função de transferência da equação (2), ou seja, sem a existência de zeros.

$$H(z) = \frac{1}{A(z)} = \frac{1}{a_0 + a_1 z^{-1} + \ldots + a_q z^{-q}}. \tag{2}$$

2.3.1.1 Obtenção dos Parâmetros

Os coeficientes e a variância do LPC possuem diversos métodos cabíveis para sua determinação. Um dos métodos mais utilizados, descrito em Makhoul (1975), é através do denominado método de autocorrelação.

Primeiramente, define-se o erro médio quadrático δ como:

$$\delta = \sum_n (s(n))^2 = \sum_n \left(s(n) - \sum_{k=1}^{q} a_k s(n-k) \right)^2. \tag{3}$$

Encontra-se o mínimo do erro com base na derivada nula em relação a a_i, ou seja:

$$\frac{\partial \delta}{\partial a_i} = 0, \quad 1 \leq i \leq q. \tag{4}$$

A partir das equações (3) e (4) obtém-se a expressão em (5).

$$-\sum_n s(n)s(n-i) = \sum_{k=1}^{q} (a_k) \left(\sum_n (s-k)s(n-i) \right), \quad 1 \leq i \leq q. \tag{5}$$

Simplifica-se a equação (5) com a utilização da definição de autocorrelação ou autocovariância obtida de acordo com a equação (6) para uma sinal com N amostras:

$$R(i) = \sum_{n=0}^{N-1-i} (s(n)s(n-i)). \tag{6}$$

A autocorrelação R pode ser entendida como uma ferramenta para se compreender a periodicidade do sinal, analisando-se a covariância de um sinal em relação a ele próprio defasado no tempo, sendo i o número de amostras defasadas.

Dessa forma, chega-se, então, à equação de Yule-Walker em (7), da qual resulta nos coeficientes a_i (KALLAS *et al.*, 2013).

$$\begin{bmatrix} R(0) & R(1) & R(2) & R(3) & ... & R(q-1) \\ R(1) & R(0) & R(1) & R(2) & ... & R(q-2) \\ R(2) & R(1) & R(0) & R(1) & ... & R(q-3) \\ R(3) & R(2) & R(1) & R(0) & ... & R(q-4) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ R(q-1) & R(q-2) & R(q-3) & R(q-4) & .... & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_q \end{bmatrix} = - \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ R(4) \\ \vdots \\ R(q) \end{bmatrix}. \tag{7}$$

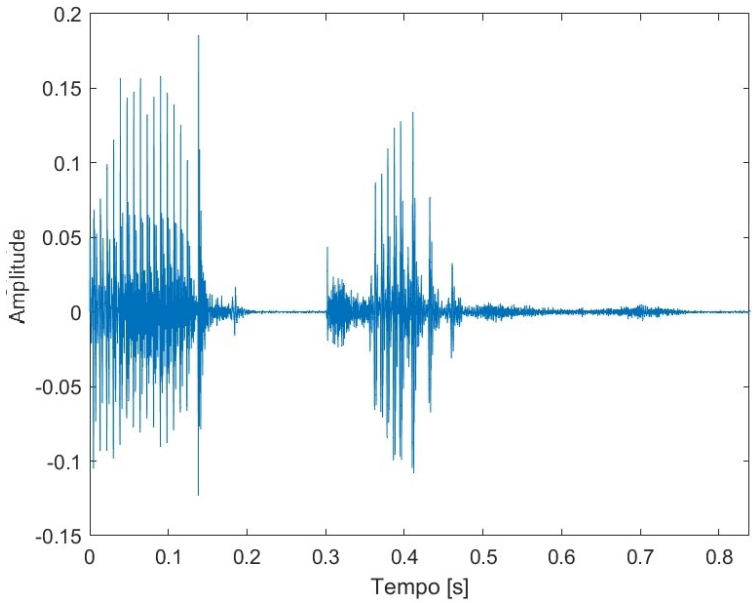
Como a matriz descrita na equação (7) possui a forma Toeplitz[2], é possível que o sistema linear seja solucionado através de métodos clássicos, como o algoritmo Levinson-Durbin (CASTIGLIONI, 2005), o qual reduz significantemente a computação da solução.

2.3.1.2 Aplicação de Janelas para Modelagem de Fala

O LPC visa a fornecer um modelo linear que descreve um sinal de características periódicas. Dessa forma, não seria possível aplicar um único modelo para a totalidade de um sinal da fala, pois mesmo que a fala possua comportamentos de frequência próximos a constantes em curtos períodos de tempo, ocorre, necessariamente, transições para outros comportamentos para que palavras sejam pronunciadas. Isso é melhor demonstrado na Figura 9, em que se é observado como o comportamento do som de uma palavra varia ao longo do tempo.

[2] Matriz em que todos os elementos em diagonal são iguais entre si.

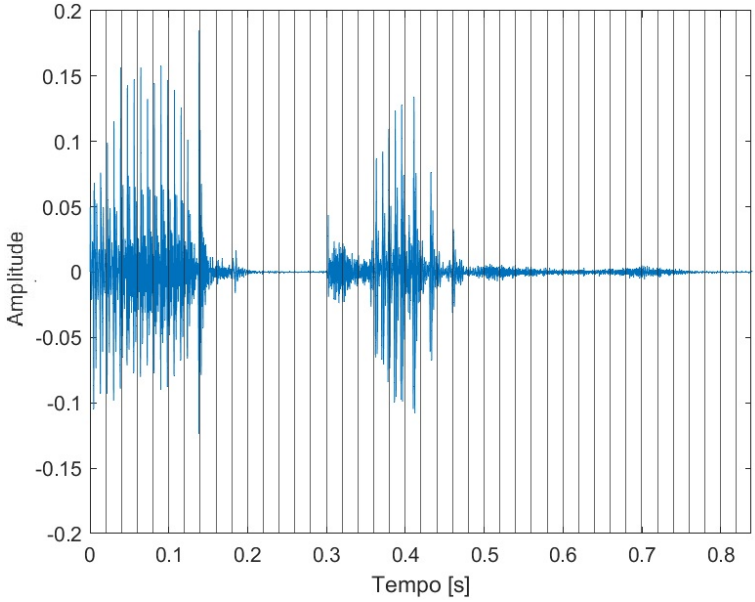
Figura 9 – Sinal de pronuncia da palavra *actress* de um sinal com frequência de amostragem de 8 kHz.



Fonte: Autoria própria, amostra de áudio do banco de dados Hu e Loizou (2007).

Portanto, uma possibilidade de aplicação do LPC para a modelagem da fala é através da divisão do áudio em janelas de tempo, como visto na Figura 10 (TIERNEY, 1980). Ou seja, para cada janela serão feitos os cálculos como os descritos pelas equações (6) e (7) para obter os modelos para cada período de tempo.

Figura 10 – Seccionamento do sinal em trechos iguais de 20 ms para produção de janelas do sinal de pronúncia da palavra *actress* com frequência de amostragem de 8 kHz.



Fonte: Autoria própria, amostra de áudio do banco de dados Hu e Loizou (2007).

Se o tamanho da janela e a ordem q estiverem devidamente dimensionados, é possível que se consigam aproximações do sinal original a partir de um processo de sintetização do sinal, como visto em McCree e Barnwell (1995).

Existem diversos tipos de janelas. Uma janela temporal é uma função matemática que multiplica o sinal de entrada ponto a ponto. Diferentes tipos de janelas temporais têm propriedades distintas, e a escolha adequada pode ter um impacto significativo na qualidade da análise de sinais. A janela retangular é a janela temporal mais simples, representada por uma função dado por:

$$w(k) = \begin{cases} 1, & 0 \leq k \leq N \\ 0, & k < 0, k > N. \end{cases} \tag{8}$$

com N sendo o tamanho da janela. Ou seja, ela não introduz nenhuma modificação nos dados de entrada. No entanto, a aplicação da janela quadrada pode causar descontinuidades nas bordas do sinal. Quando o sinal é analisado por uma transformada de Fourier, por exemplo, essa descontinuidade gera componentes de frequência adicionais que não fazem parte do sinal original, mas que aparecem no respectivo espectro como lóbulos laterais. Esses lóbulos laterais podem ser denominados como vazamento espectral. Tal fenômeno pode prejudicar a precisão da análise de frequência, tornando mais difícil identificar as frequências reais presentes no sinal (CERNA; HARVEY, 2000).

Os outros tipos de janelas temporais, portanto, são funções matemáticas aplicadas aos dados de entrada para limitar o efeito das bordas do sinal e reduzir artefatos indesejados. A janela Hamming, por exemplo, é definida como (HEINZEL; RÜDIGER; SCHILLING, 2002):

$$h(k) = \begin{cases} 0{,}54 - 0{,}46\cos(\frac{N-1}{2\pi k}), & 0 \leq k \leq N \\ 0, & k < 0, k > N. \end{cases} \tag{9}$$

Observa-se que é uma janela temporal que apresenta uma redução suave dos valores nas bordas da janela, o que a torna eficaz na redução do vazamento espectral. Ela é frequentemente usada em análises de espectro e filtragem de sinais. A desvantagem da janela Hamming é que ela introduz uma maior largura de lóbulo principal, o que pode prejudicar a resolução em frequência em comparação com outras janelas mais estreitas.

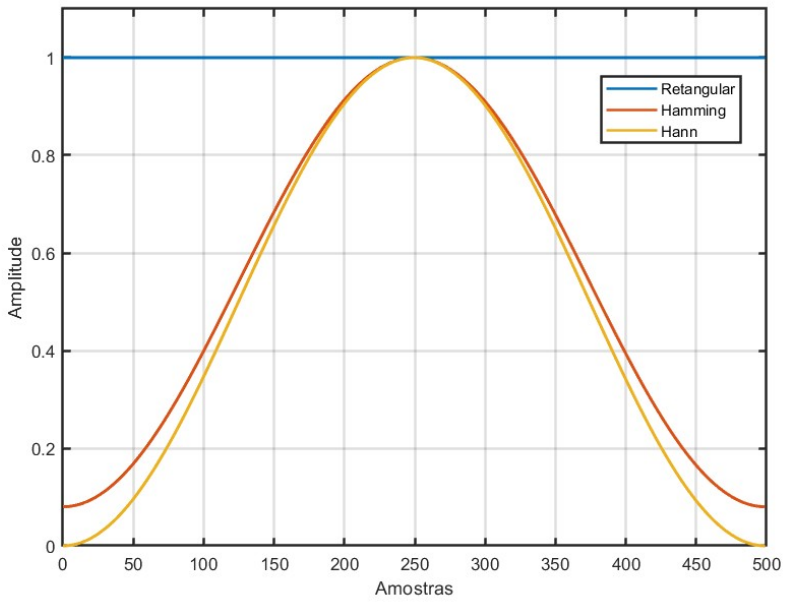
Uma janela similar é a de Hann, também muitas vezes referida como janela de Hanning, que é definida como

$$w(k) = \begin{cases} 0{,}5 - 0{,}5\cos(\frac{N-1}{2\pi k}), & 0 \leq k \leq N \\ 0, & k \leq 0, N \leq k. \end{cases} \tag{10}$$

Esta janela caracteriza-se por uma suave transição entre zero nas extremidades, diferentemente da janela de Hamming. Dessa forma, è uma solução amplamente aplicada em algoritmos de adição de janelas sobrepostas (OLA, do inglês *Overlap Add*).

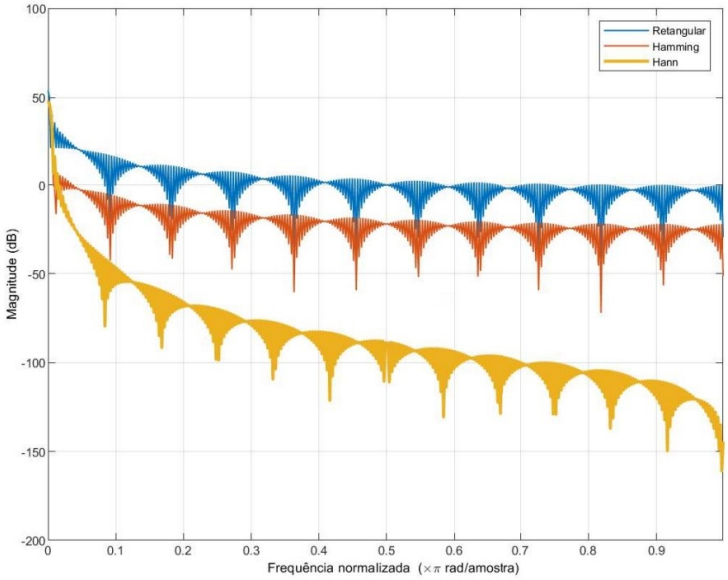
O comportamento das janelas tanto no domínio do tempo como no domínio frequência podem ser observados nas figuras Figura 11 e Figura 12.

Figura 11 – Comparação entre Janelas Retangular, de Hamming e de Hann de 500 amostras no domínio do tempo.



Fonte: Autoria própria.

Figura 12 – Comparação entre Janelas Retangular, de Hamming e de Hann de 500 amostras no domínio da frequência



.

Fonte: Autoria própria.

2.3.1.3 *Whitening*

Ao rearranjar a EquaçãoEquação 1 de maneira a isolar o ruído branco, obtém-se a Equação (11)

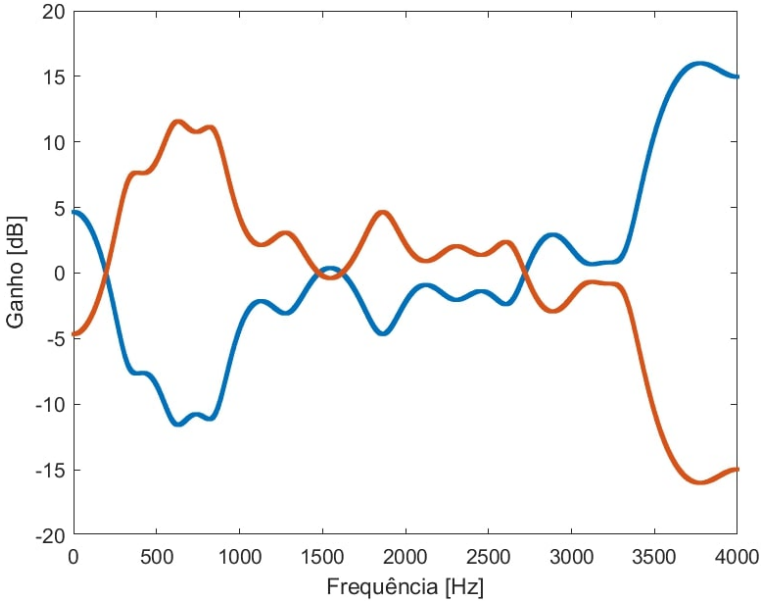
$$w(k) = s(k) - \sum_{i=1}^{q} a_i s(k-i). \tag{11}$$

Dessa forma, obtém-se um filtro de Resposta Finita ao Impulso (FIR, do inglês *Finite Impulse Response*), pois a saída depende somente da relação da entrada com as entradas anteriores. Sua função de transferência pode ser descrita pela equação (12), da qual observa-se a ausência de polos (GEORGE *et al.*, 2018).

$$W(z) = A(z) = \sum_{i=0}^{q} a_i z^i. \tag{12}$$

No domínio da frequência, portanto, o efeito é o contrário do descrito pelo filtro *all-poles*. No caso, quando se tem como entrada um ruído branco, a resposta no domínio de frequência é o resultado da remoção das bandas de frequência presentes no sinal analisado. Dessa maneira, pode se dizer que é possível utilizar os coeficientes do LPC de maneira a elaborar um filtro *whitening* (WF, do inglês *Whitening Filter*), visto que, se aplicado ao sinal analisado, o aproximará do comportamento de um ruído branco no domínio da frequência (ROY; PALIWAL, 2020). Na Figura 13 é observado a resposta de frequência do LPC do sinal analisado em relação à resposta em frequência de um WF com os mesmos coeficientes.

Figura 13 – Resposta de filtro *all-pole* obtido por LPC de $q = 22$ do sinal da Figura 7 (laranja) e filtro *all-zero* com mesmos coeficientes (azul).



Fonte: Autoria própria.

3 FILTRAGEM ESTOCÁSTICA

Neste capítulo será tratada a base teórica dos filtros estocásticos necessários para compreender o processo de redução de ruído aleatório atrelado a sinais de voz. Além disso, este capítulo também apresenta o filtro IMM para utilização simultânea de múltiplos filtros estocásticos.

3.1 CONCEITOS PROBABILÍSTICOS

Dada a natureza estatística do processo do KF, nesta seção serão brevemente abordados conceitos básicos de probabilidade e estatística utilizados nos algoritmos dos filtros estocásticos.

3.1.1 Processo estocástico

De acordo com (HINES *et al.*, 2006), um processo estocástico é um sistema no qual possui uma aleatoriedade atrelada à passagem de tempo, tornando imprevisível a determinação dos estados seguintes. Dessa forma, ao analisar cada um dos sistemas com as mesmas condições iniciais, podem ser obtidos resultados completamente discrepantes entre ambos.

Pode-se denotar o processo como uma função $f(t,\omega)$, ou seja, dependente não só do tempo t mas também de uma variável aleatória ω. Um processo estocástico pode ser tanto um processo contínuo como discreto no tempo.

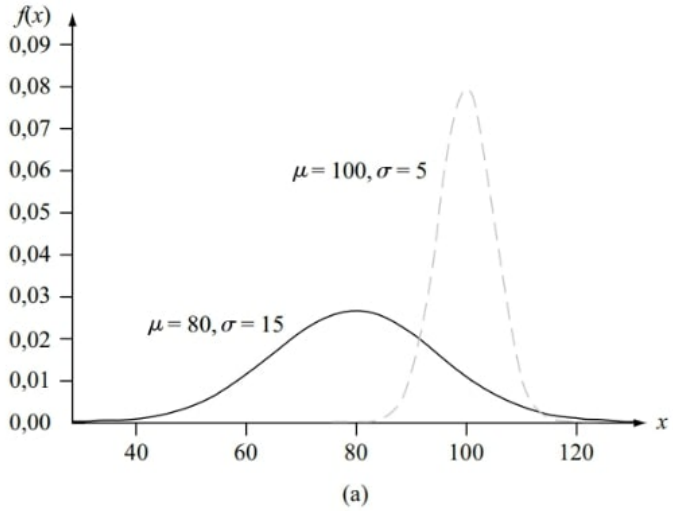
3.1.2 Distribuição Gaussiana e Ruído Branco

Uma variável aleatória que possui distribuição gaussiana ou distribuição normal possui uma Função Densidade de Probabilidade (FDP) a qual é capaz de modelar grande parte dos comportamentos aleatórios reais de medidas físicas (HINES *et al.*, 2006). Define-se através da esperança μ, em que a FDP está centrada, e do desvio-padrão σ da variável aleatória X, como descreve a equação (13)

$$g_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{13}$$

A Figura 14 ilustra o comportamento da FDP gaussiana para diferentes valores de μ e σ. Evidencia-se que os arredores da esperança possuem uma maior densidade de probabilidade e, quanto mais afastado daquele valor, menor é a densidade, indicando que o evento é menos provável de ocorrer. Quanto maior a variância, menor é a concentração da função em torno de sua média, aumentando a probabilidade de ocorrência de eventos distantes da esperança.

Figura 14 – Curvas de distribuição da FDP normal para diferentes valores de μ e σ.



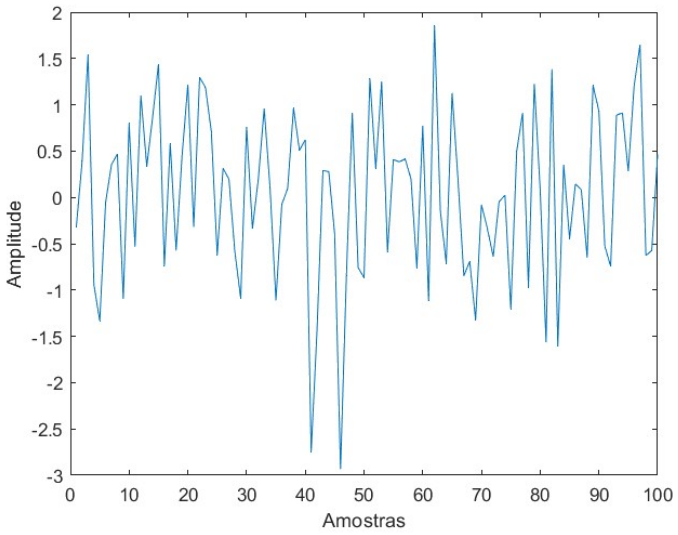
Fonte: (DEVORE; CORDEIRO, 2014).

Uma variável aleatória gaussiana, portanto, consiste em qualquer variável aleatória em que sua FDP se adeque com o descrito em (13), podendo ser resumidamente descrita como:

$$X \sim \mathcal{N}(\mu,\sigma^2). \tag{14}$$

Denomina-se ruído gaussiano ou ruído branco um processo estocástico do qual é definido unicamente por uma variável aleatória com a distribuição gaussiana com média nula, ou seja, $\mu = 0$ (BAR-SHALOM; LI; KIRUBARAJAN, 2004). A Figura 15 demonstra um exemplo de um ruído gaussiano com variância unitária.

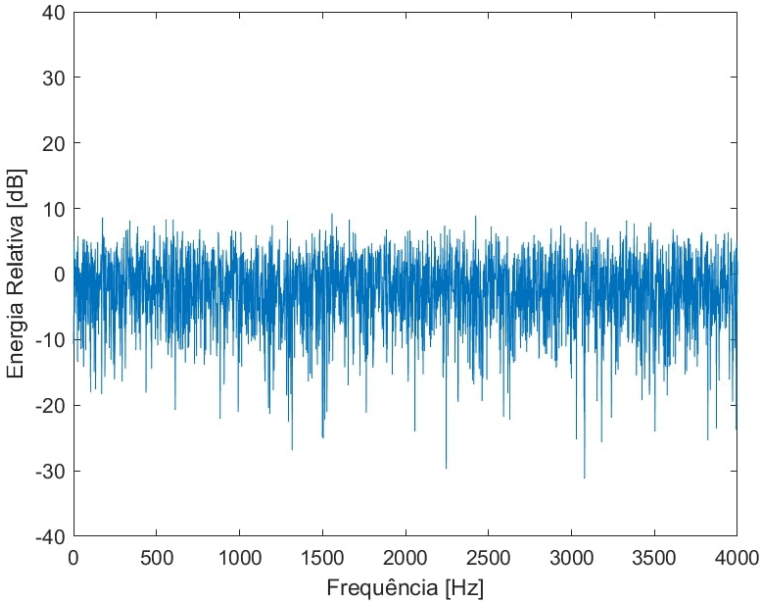
Figura 15 – Exemplo de sinal de ruído branco de variância unitária.



Fonte: Autoria própria.

O processo estocástico gaussiano tende a possuir um comportamento de energia similar para todos os valores diferentes de zero ao se analisar no domínio da frequência, como demonstrado na Figura 16. Caso o ruído não possua essa propriedade, esse pode ser classificado como um ruído colorido (VASEGHI, 2008).

Figura 16 – Exemplo de análise espectral de ruído branco de variância unitária.



Fonte: Autoria própria.

3.2 FILTRO DE KALMAN

O KF é um método recursivo de estimação de estados em sistemas lineares dos quais sofrem perturbações gaussianas aditivas. Ele possui um algoritmo computacionalmente eficiente para a redução de erro quadrático médio entre valores estimados e valores exatos. Dessa forma, é um método amplamente utilizado em controle estocástico, além de ser útil em processos de fusão de medidas, rastreamento e estimação (LI *et al.*, 2015).

O KF consiste em uma série de equações que levam em consideração as mensurações discretas ao longo do tempo, a modelagem dinâmica do sistema e a estimativa das incertezas relacionadas à medição e ao modelo matemático.

Esse filtro estocástico limita-se a sistemas com modelos lineares, que podem ser descritos como sistemas discretos em espaço de estados, como nas equações (15) e (16) (BAR-SHALOM; LI; KIRUBARAJAN, 2004).

$$\mathbf{x}_{k+1} = G_k\mathbf{x}_k + H_k\mathbf{u}_k + \mathbf{w}_k, \tag{15}$$

$$\mathbf{y}_k = C_k\mathbf{x}_k + D_k\mathbf{u}_k + \mathbf{v}_k, \tag{16}$$

sendo:

- $\mathbf{x}_k \in \mathbb{R}^m$ – vetor de estados;
- $\mathbf{u}_k \in \mathbb{R}^c$– vetor de entradas;
- $\mathbf{y}_k \in \mathbb{R}^n$ – vetor de saídas;
- $G_k \in \mathbb{R}^{m \times m}$– matriz de estados;
- $H_k \in \mathbb{R}^{m \times c}$ – matriz de entradas;
- $C_k \in \mathbb{R}^{n \times m}$ – matriz de saídas;
- $D_k \in \mathbb{R}^{n \times c}$ – matriz de transmissão direta;
- $\mathbf{w}_k \sim \mathcal{N}(0,Q_k) \in \mathbb{R}^m$ – ruído do processo;
- $\mathbf{v}_k \sim \mathcal{N}(0,\Theta_k) \in \mathbb{R}^n$ – ruído da observação;
- $Q_k \in \mathbb{R}^{m \times m}$ – matriz de covariância do ruído de processo;
- $\Theta_k \in \mathbb{R}^{n \times n}$ – matriz de covariância do ruído da observação.

Importante notar que a redução do erro se limitará a ruídos gaussianos com matrizes de covariância conhecidas, não sendo eficientes na redução de ruídos coloridos.

A filtragem estocástica, de um modo geral, consiste em duas principais etapas: a predição e a atualização. Na predição utiliza-se o modelo matemático em conjunto com uma estimativa do estado anterior (*a priori*) para realizar a previsão do estado e da saída atual do sistema. Já a atualização realiza-se uma ponderação entre as parâmetros obtidos na etapa de predição em relação às mensurações, produzindo uma estimativa do estado atual (*a posteriori*) com base na informação da nova medida obtida naquele instante de tempo. Tais etapas são descritas nas seções a seguir.

3.2.1 Predição

A previsão do estado é dada pela equação (17); nela, considera-se o estado estimado do passo anterior $\mathbf{x}_{k-1|k-1}$ e o vetor de entradas atual $\mathbf{u}_k$ para realizar uma previsão do estado atual $\mathbf{x}_{k|k-1}$ de acordo com o modelo estabelecido (WELCH; BISHOP *et al.*, 1995).

$$\mathbf{x}_{k|k-1} = G_k \mathbf{x}_{k-1|k-1} + H_k \mathbf{u}_k. \tag{17}$$

Com isso, é possível realizar a previsão da matriz de covariância do erro de previsão, a qual será atualizada na etapa seguinte, de acordo com a equação (18).

$$P_{k|k-1} = G_k P_{k-1|k-1} G_k^{\intercal} + Q_{k-1}. \tag{18}$$

visto que P_{k-1k-1} é a matriz de covariância do erro de estimativa no instante de tempo anterior.

Na expressão (19) determina-se o valor de saída previsto $\mathbf{z}_{k|k-1}$. Já nas equações (20) e (21) determina-se, respectivamente, a inovação $\tilde{\mathbf{z}}_k$ e a matriz de covariância da inovação S_k.

$$\mathbf{z}_{k|k-1} = C_k \mathbf{x}_{k|k-1} + D_k \mathbf{u}_k, \tag{19}$$

$$\tilde{\mathbf{z}}_k = \mathbf{y}_k - \mathbf{z}_{k|k-1}, \tag{20}$$

$$S_k = C_k P_{k|k-1} C_k^{\intercal} + \Theta_k. \tag{21}$$

No caso da primeira iteração do filtro, normalmente tem-se que o valor atribuído à estimativa *a priori* $\mathbf{x}_{0|0}$ é dado diretamente pelo valor da mensuração inicial $\mathbf{y}_0$. Já a matriz $P_{0|0}$ é comumente dada por uma matriz diagonal definida positiva, garantindo a incerteza atrelada a $\mathbf{x}_{0|0}$ (WELCH; BISHOP *et al.*, 1995). A matriz $P_{k|k}$ tende a zero com o passar das iterações visto um dimensionamento adequado do KF, representando, assim, uma minimização do erro atrelado à estimação.

3.2.2 Atualização

As equações (22)–(24) definem a etapa de atualização do KF.

$$K_k = P_{k|k-1} C_k^{\intercal} S_k^{-1}, \tag{22}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + K_k \tilde{\mathbf{z}}_k, \tag{23}$$

$$P_{k|k} = [I - K_k C_k] P_{k|k-1}, \tag{24}$$

em que $K_k \in \mathbb{R}^{m \times n}$ é o ganho de Kalman, tido como responsável por ponderar os valores de $\mathbf{z}_{k|k-1}$ e $\mathbf{y}_k$ para obtenção do valor estimado do processo de filtragem, e $P_{k|k}$ é a matriz de covariância de erro estimado *a posteriori*.

Tem-se com o valor de $\mathbf{x}_{k|k}$ a saída do KF analisado no instante de tempo k, que levando às hipóteses de linearidade do modelo matemático e distribuição normal das perturbações, tende a minimizar o erro em relação ao valor real da medida.

3.3 FILTRO IMM

Para os casos em que não é possível a estimação de estado com a utilização de um único modelo dinâmico do sistema, uma alternativa é a utilização do filtro IMM. Com ele, é possível a utilização simultânea de um banco composto por vários filtros estocásticos, iguais ou diferentes entre si, cada um com diferentes modelos, estes podendo ser lineares ou não lineares (FRENCL *et al.*, 2010).

Seu funcionamento está pautado na teoria de saltos markovianos. Sendo M um conjunto de n modelos matemáticos dado por $M = \{m^{(1)},m^{(2)},...,m^{(n)}\}$, um salto markoviano consiste na probabilidade de ocorrer transição para um modelo $m^{(i)}$ no instante k visto a adoção de único modelo $m^{(j)}$ no instante $k-1$.

$$P[m_k = i|m_{k-1} = j] = \rho_{i,j}. \tag{25}$$

Generalizando para todos os n modelos, a matriz de transição de probabilidade markoviana pode ser descrita como uma matriz quadrada Ω dada por:

$$\Omega = [\rho_{i,j}]_{n\times n} = \begin{bmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & \rho_{n,n} \end{bmatrix}, \tag{26}$$

em que $\Omega \in \mathbb{R}^{n\times n}$, de maneira que toda linha de Ω consista numa função de probabilidade, ou seja, com a somatória de todos elementos de uma linha sendo igual a um. No caso do filtro IMM, simplificadamente, aplica uma matriz de transição de probabilidade a fim de efetivar o "chaveamento probabilístico" entre os diferentes modelos do banco de n filtros estocásticos, assim gerando uma ponderação das estimativas geradas por esses filtros.

O algoritmo consiste em quatro etapas principais, em ordem: reinicialização condicionada por modelo, filtragem condicionada por modelo, atualização da probabilidade do modo e estimativa global. A etapa de reinicialização consiste na obtenção de uma entrada que seja coerente com as estimativas de estado e covariância dos n filtros individuais. Define-se nessa etapa a probabilidade do modelo correto ser $m^{(i)}$ visto que o modelo correto *a priori* é $m^{(j)}$, dado em (27):

$$\nu_{k-1}^{(i|j)} = \frac{\rho_{i,j}\nu_{k-1}^{(i)}}{\sum_{i=1}^{n}\rho_{i,j}\nu_{k-1}^{(i)}}. \tag{27}$$

sendo $\nu_{k|k-1}^{(i|j)}$ a probabilidade de ocorrer a transição do modelo correto $m^{(j)}$ para o modelo $m^{(i)}$, já $\nu_{k-1}^{(j)}$ representa a possibilidade de $m^{(i)}$ ser o modelo correto. As estimativas de estado e da matriz de covariância na etapa de reinicialização são dadas por:

$$\bar{\mathbf{x}}_{k-1|k-1}^{(j)} = \sum_{i=1}^{n} \mathbf{x}_{k-1|k-1}^{(i)} \nu_{k-1}^{(i|j)}, \tag{28}$$

$$\bar{P}_{k-1|k-1}^{(j)} = \sum_{i=1}^{n} \nu_{k-1}^{(i|j)} \left[P_{k-1|k-1}^{(i)} + \left(\mathbf{x}_{k-1|k-1}^{(i)} - \mathbf{x}_{k-1|k-1}^{(j)} \right) \left(\mathbf{x}_{k-1|k-1}^{(i)} - \mathbf{x}_{k-1|k-1}^{(j)} \right)^{\mathsf{T}} \right]. \tag{29}$$

Na segunda etapa, aplica-se o processo de filtragem como descrito na seção 3.2, na qual todos os filtros que compõem o banco produzem suas respectivas estimativas. A diferença aqui é que aplica-se como estimativa *a priori* $\mathbf{x}_{k-1|k-1}^{(j)}$ e $P_{k-1|k-1}^{(j)}$ as estimativas geradas com base nas equações (28) e (29), ao invés das estimativas individuais dos filtros. Além disso, faz-se necessário, nessa etapa, o cálculo da verossimilhança $\Lambda_k^{(j)}$ para cada um dos filtros, estimada como uma distribuição gaussiana cuja variância σ^2 é dada pela covariância de inovação $S_k^{(j)}$ e μ por $\mathbf{z}_{k|k-1}$.

O passo seguinte tem como finalidade determinar o quão adequado é o modelo em relação à dinâmica atual do sistema. Isso é feito a partir do cálculo de atualização de $\nu_k^{(j)}$, dado por:

$$\nu_k^{(j)} = \frac{\Lambda_k^{(j)} \sum_{i=1}^{n} \rho_{i,j} \nu_{k-1}^{(j)}}{\sum_{j=1}^{n} [\Lambda_k^{(i)} \sum_{i=1}^{n} \rho_{i,j} \nu_{k-1}^{(i)}]}. \tag{30}$$

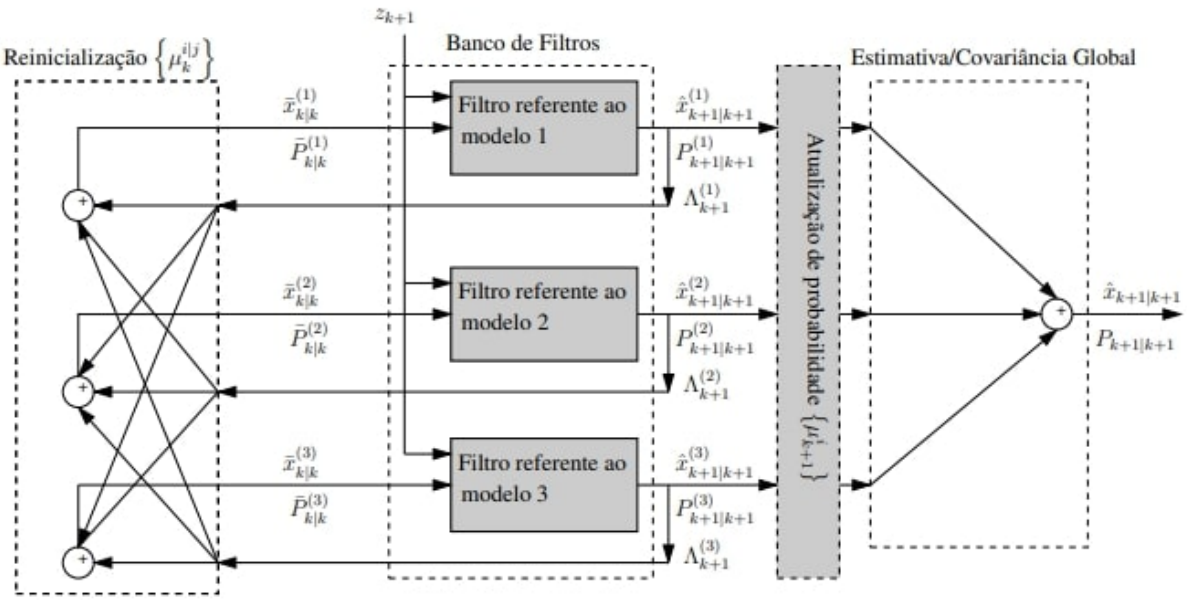
Por fim, a estimativa global é a responsável por definir a estimativa do filtro IMM para o instante de tempo k. Isso é feito de acordo com as expressões em (31) e (32).

$$\mathbf{x}_{k|k} = \sum_{j=1}^{n} \mathbf{x}_{k|k}^{(j)} \nu_k^{(j)}, \tag{31}$$

$$P_{k|k} = \sum_{j=1}^{n} \nu_k^{(j)} \left[P_{k|k}^{(j)} + \left(\mathbf{x}_{k|k}^{(j)} - \mathbf{x}_{k|k} \right) \left(\mathbf{x}_{k|k}^{(j)} - \mathbf{x}_{k|k} \right)^{\mathsf{T}} \right]. \tag{32}$$

A Figura 17 resume as iterações referentes ao algoritmo do filtro IMM.

Figura 17 – Exemplo de um ciclo IMM com um banco de três filtros.



Fonte: (FRENCL *et al.*, 2010).

4 MÉTODOS DE ANÁLISE DE RESULTADOS

Neste capítulo descrevem-se os métodos empregados para a obtenção de resultados quantitativos referentes ao desempenho das abordagens propostas.

4.1 Espectrograma

Os espectrogramas se configuram como ferramentas de análise de séries temporais, amplamente utilizado para sinais de áudio com o fim de visualizar as variações desse no domínio da frequência. É uma ferramenta que pode revelar de informações contidas em um sinal sonoro, destacando o comportamento das frequências que compõem o sinal ao longo do tempo (DENNIS; TRAN; LI, 2010).

Esse gráfico consiste em uma representação gráfica bidimensional que mapeia a distribuição de energia (PSD, do inglês *Power Spectral Density*, dado em escala logaritmica dB/Hz), de um sinal sonoro em diferentes frequências e instantes temporais, obtido através do resultado das Transformadas de Fourier de curto termo (STFT, do inglês *Short-Time Fourier Transform*) de janelas de tempo que seccionam o sinal analisado. Portanto, essa representação, similar a um mapa de calor, permite visualizar a intensidade de cada componente frequencial ao longo da duração do sinal.

Esse trabalho se utilizará de espectrogramas com o fim de tornar visual no domínio da frequência, o comportamento dos sinais estimados em comparação aos sinais limpos e ruidosos.

4.2 Relação Sinal-Ruído

Para avaliar o desempenho das técnicas propostas, uma das métricas utilizadas é a Relação Sinal-Ruído (SNR, do inglês *Signal to Noise Ratio*); uma medida quantitativa, expressa em decibéis (dB), amplamente utilizada em processamento de sinais e telecomunicações para quantificar a relação entre um sinal desejado e o ruído indesejado. Essa medida avalia a qualidade de um sinal indicando o quão bem o sinal desejado se destaca em relação ao ruído aditivo (BOSWORTH *et al.*, 2008). A SNR é calculada com base na raiz do erro quadrático médio (RMSE, do inglês *Root Mean Squared Error*) entre o sinal discreto avaliado g_k e o sinal discreto de referência f_k de N amostras:

$$SNR = 20\log_{10}\left(\frac{\sqrt{\sum_{i=1}^{N} f_i^2}}{\sqrt{\sum_{i=1}^{N}(g_i - f_i)^2}}\right). \tag{33}$$

Uma SNR maior indica uma melhor qualidade do sinal, pois o sinal desejado sobressai mais em relação ao ruído de fundo. Por outro lado, uma SNR menor sugere uma pior qualidade do sinal, com o sinal desejado sendo mais difícil de discernir do ruído.

Entretanto, é importante ressaltar que essa análise não considera a capacidade do estimador em preservar características perceptíveis do sinal, que são mais determinadas por seu comportamento no domínio da frequência do que por seu comportamento temporal.

Para obter a performance de um algoritmo na métrica considera-se a média de cinco áudios selecionados aleatoriamente por nível de SNR de entrada que serão estimados pelos algoritmos a ser comparados.

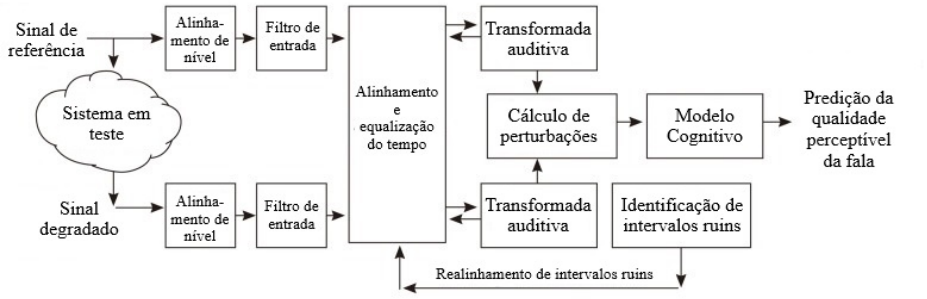
4.3 PESQ

PESQ é uma métrica amplamente usada para avaliar a qualidade perceptual da fala processada em sistemas de comunicação. Ela foi desenvolvida pela ITU-T (*International Telecommunication Union - Telecommunication Standardization Sector*) e é especificamente projetada para avaliar a qualidade da fala após passar por algoritmos de processamento, como codificação de voz, melhoria de fala e transmissão (BEERENDS *et al.*, 2002).

O PESQ é baseado na ideia de que a qualidade da fala é percebida pelo ouvinte humano com base em como o sinal analisado g_k se assemelha ao sinal de referência f_k.

Para sinais de frequência de amostragem de 8 kHz, tanto f_k quanto g_k, calcula-se a diferença de potência espectral (PSD, do inglês *Power Spectrum Density*) e de fase por meio da DFT de janelas sobrepostas após a aplicação de uma ponderação feita de acordo com um modelo psicoacústico. Isso gera a diferença perceptual que, usando uma função matemática específica definida pela ITU-T, resulta-se em uma medida de qualidade perceptual. Seu algoritmo é iustrado no diagrama de blocos da Figura 18.

Figura 18 – Ilutração de funcionamento do cálculo de PESQ



Fonte: Traduzido de Rix *et al.* (2001).

O valor PESQ final é obtido a partir da combinação das medidas de PSD e distorção de fase. Quanto maior a pontuação do PESQ indica uma maior proximidade entre os dois sinais. A pontuação PESQ varia numa escala de -0,5 a 4,5. Um valor maior indica uma qualidade de fala

percebida mais próxima da fala original não processada, enquanto valores menores indicam que o sinal fora significantemente comprometido quanto a sua inteligibilidade.

Importante denotar que, apesar do PESQ ser uma métrica objetiva que leva em consideração várias características da fala, essa ainda pode não captura completamente todas as nuances da percepção humana de qualidade de áudio.

Para medir a performance de um algoritmo, os resultados são obtidos através do cálculo da média do PESQ de cinco áudios estimados pelo algoritmo, assim como descrito na Seção 4.2.

4.4 Banco de Dados

Todos os testes conduzidos neste trabalho utilizaram o banco de dados NOIZEUS CORPUS (HU; LOIZOU, 2007). Este banco de dados contém um total de 30 arquivos de áudio em inglês amostrados a 8 kHz, que incluem uma variedade de frases pronunciadas por seis indivíduos diferentes (três homens e três mulheres).

O banco de dados também inclui variações dos áudios contaminados por diversos tipos de ruídos aditivos, os quais variam de 15 a 0 dB SNR, em intervalos de 5 dB. Dentre os tipos de ruído presentes, destacam-se:

- Ruído de carro – Este é o som ambiente dentro de um carro em movimento. Inclui o barulho do motor, pneus e outros sons mecânicos;
- Ruído de trem – Refere-se ao som dentro ou próximo a um trem em operação. Isso pode incluir o barulho das rodas nos trilhos, apitos e outros sons associados ao movimento do trem;
- Ruído de multidão – Este ruído é caracterizado pelo som de um grande número de pessoas conversando simultaneamente, típico de eventos públicos ou espaços fechados com muita gente;
- Ruído de rua – Representa o som de uma rua movimentada, com ruídos de tráfego, buzinas, conversas de pedestres e outros sons urbanos. Este tipo de ruído é bastante variado e pode incluir picos de intensidade devido a eventos inesperados, como sirenes de ambulância.

4.5 Sintetização de ruídos

Além dos ruídos já presentes no banco de dados utilizado, nesse trabalho são simuladas as condições de ruído branco, o qual se é obtido através da função `randn` do MATLAB®. Se faz de tal maneira que se produza um vetor com o mesmo número de amostras do áudio a ser

estimado. O sinal de ruído é então multiplicado pela raiz da variância correspondente ao SNR demandado para a simulação, que pode ser calculada como a Equação 34:

$$o_u^2(SNR,s_k) = \frac{\sqrt{\sum_{i=1}^{N_s} s_i^2}}{10^{\frac{SNR}{20}} N_s}, \tag{34}$$

sendo s_k o sinal de fala limpo com número de amostras N_s .

5 FILTROS ESTOCÁSTICOS PARA O TRATAMENTO DO SINAL DE FALA

Ao longo deste capítulo, serão detalhadas as etapas de implementação de cada algoritmo de filtragem descrito anteriormente: KF e IMM, e sua integração com os métodos de identificação de modelos de sinais de fala. Também será demonstrado o funcionamento dos algoritmos e os seus parâmetros utilizados. Além disso, serão abordados os processos de tratamento do sinal para emprego em sinais contaminados com ruído colorido.

5.1 Filtro de Kalman para Estimação de Sinais de Fala

Nessa seção, serão apresentados os detalhes da implementação do KF com base em um modelo AR, identificado via LPC, para a supressão de ruídos em sinais de fala assim como abordagens para aprimorar o sinal estimado.

Considerando um sinal de fala com ruído aditivo branco u_k, é possível escrevê-lo de acordo com a equação a seguir.

$$y_k = s_k + u_k, \tag{35}$$

Como discutido na Seção 2.3, pode-se aproximar o sinal em (35) de acordo com (1), sendo a_i determinado pelos modelos AR obtidos da janela de tempo a que a amostra k pertence.

A fim de realizar a estimação do sinal, é necessária determinação da variância do ruído de medida σ_u^2 a partir do cálculo da variância nos trechos de silêncio da faixa de áudio em processamento, ou seja, quando o ruído apresenta-se isolado do sinal de fala (HINES *et al.*, 2006):

$$\sigma_u^2 = \frac{\sum_{j=1}^{m}(u_j - \bar{u})^2}{m}, \tag{36}$$

sendo u_j a amostra de ruído isolado e $\bar{u}$ o valor médio das amostras e m o número de amostras desse trecho. Denota-se que esse processo também pode ser efetuado visto um conhecimento prévio da intensidade do ruído, pertinente quando o ruído é sintetizado, de acordo com a Seção 4.5.

Dessa forma adota-se a seguinte associação:

$$\Theta = \sigma_u^2. \tag{37}$$

Com base no descrito na Seção 2.3.1.2, para conseguir aplicar a modelagem via identificação LPC, é necessário realizar um seccionamento do sinal em janelas de tempo, sendo cada uma das janelas com seu próprio modelo AR. As janelas, portanto, serão utilizadas para obter sua respectiva matriz de estados A_n.

O algoritmo LPC fornecerá os coeficientes a_i do modelo AR da janela n. Adaptando para a representação em espaço de estados, é possível reescrever o modelo como mostra a (38)

$$G = A_n = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ -a_{q,n} & -a_{q-1,n} & -a_{q-2,n} & -a_{q-3,n} & \dots & -a_{1,n} \end{bmatrix}, \tag{38}$$

em que:

$$\mathbf{x}_k = \begin{bmatrix} s_{k-q} \\ \vdots \\ s_{k-2} \\ s_{k-1} \\ s_k \end{bmatrix}. \tag{39}$$

Dessa forma, adota-se $C_k = C$ como:

$$C = [0 \quad 0 \quad 0 \quad \dots \quad 1]. \tag{40}$$

Assim como os coeficientes, os valores de variância do modelo $\sigma^2_{\omega,n}$ são obtidos de acordo com o método de autocorrelação (SO; PALIWAL, 2011):

$$\sigma^2_{\omega,n} = R(0) + \sum_{i=1}^{q} a_i R(i), \tag{41}$$

$$Q_n = \sigma^2_{\omega,n} \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2_{\omega,n} C^\intercal C. \tag{42}$$

Considerando que o áudio começa em silêncio, consideram-se as seguintes condições iniciais:

$$\mathbf{x}_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{43}$$

$$P_0 = \sigma_u^2, I; \tag{44}$$

Esse algoritmo fora primeiro desenvolvido em Paliwal e Basu (1987) considerando o filtro funcionando em condições ideais isto é sem a presença de ruído para estimação do modelo, o que normalmente denomina-se de KF Oráculo (KFO).

Para esse estudo, entretanto, serão aplicadas modificações como que permitem aprimorar a performance do KFO em relação a sua versão original que serão discutidas nas seções subsequentes – essas que também serão válidas para os algoritmos em condições não ideais, ou seja, algoritmos que obtém seus modelos através de sinais ruidosos.

5.1.1 Atraso Constante de Amostras

Em Paliwal e Basu (1987) propõe-se dois métodos para reconstrução do sinal de fala estimado s'[1]: o método sem atraso de amostras e o método com atraso amostras.

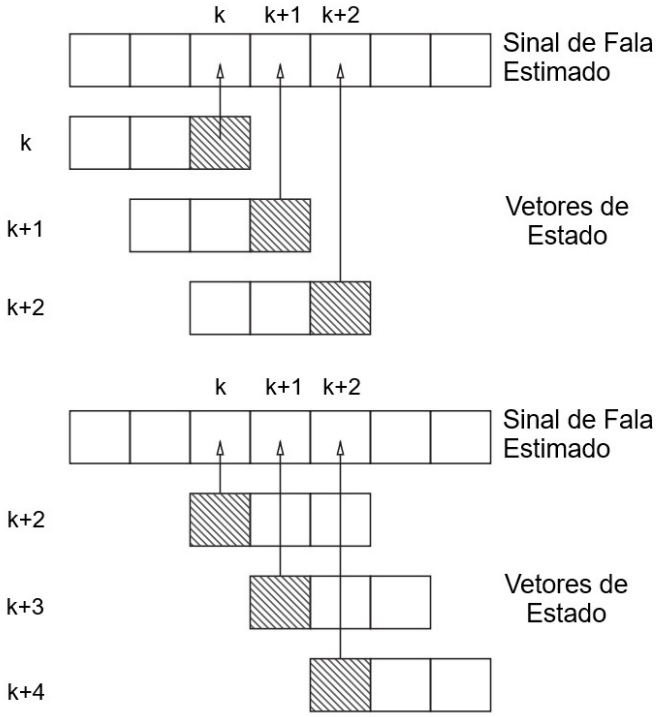
No método sem atraso, s'_k é obtido da última amostra estima do vetor vetor de estados $\mathbf{x}_{k|k}$, ou seja, seguindo como definido em (39) e (40), entende-se que

$$s'_k = C\mathbf{x}_{k|k}. \tag{45}$$

Já no método com atraso, a proposta é obter s'_k através do vetor de estados $\mathbf{x}_{k+d|k+d}$, sendo d o valor de atraso adotado. A Figura 19 ilustra ambos os métodos.

[1] A notação $'$ refere-se a sinais, vetores ou matrizes obtidos após iteração de um filtro estocástico. Dessa forma, s''_k refere-se ao sinal de fala estimado pela segunda iteração de um filtro estocástico.

Figura 19 – Ilustração de método sem atraso de amostras seguido de método com atraso de amostras considerando um modelo de $q = 3$.



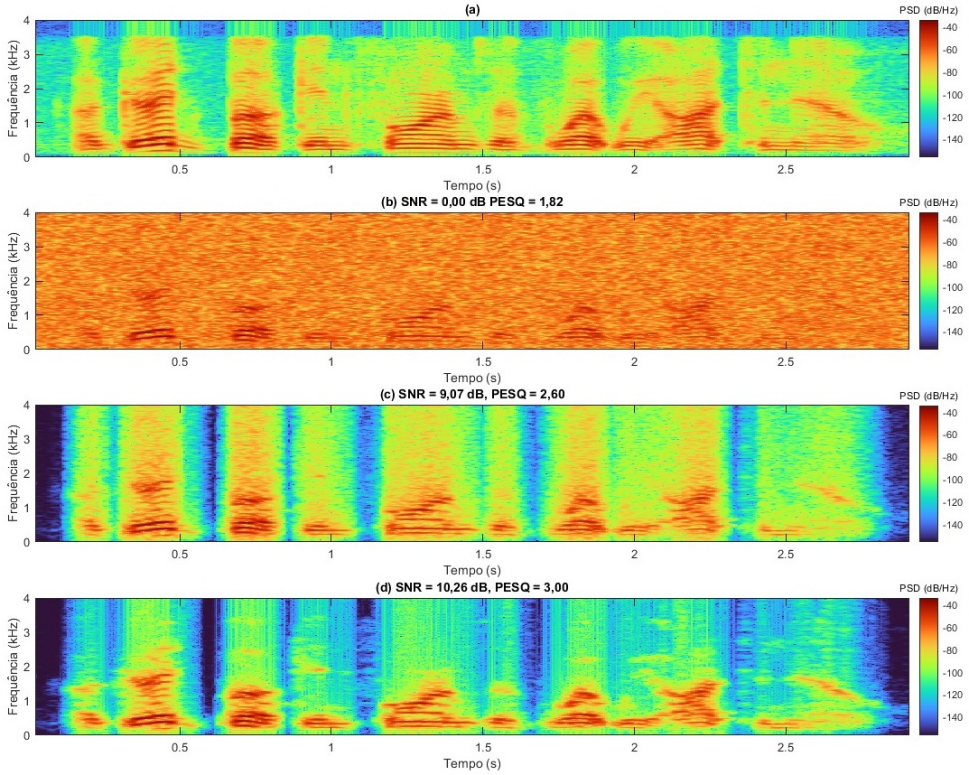
Fonte: Adaptado de So e Paliwal (2011).

De acordo com So e Paliwal (2011), o atraso de amostras é importante para a redução de ruído, porque permite que o KF tenha acesso a informações de amostras anteriores e posteriores à amostra atual, o que ajuda a melhorar a precisão da estimativa do estado do sistema.

Entende-se, portanto, que ao introduzir o atraso de amostras, o KF pode realizar uma operação de “suavização”, levando em conta informações futuras para melhorar a estimativa da amostra atual. Isso se deve ao ganho K_k que, ao longo das amostras, ajusta o valor das amostras de s_k passadas de maneira que se comportem de maneira mais similar ao ditado pelo modelo adotado.

O método com atraso constante, portanto, resulta em uma melhoria significativa na qualidade e inteligibilidade de s'_k, como exemplificado na Figura 20 na qual se compara espectralmente um KFO sem atraso e com atraso de $q - 1$ amostras ao estimar um sinal de fala.

Figura 20 – Espectrogramas de (a) sinal limpo, (b) sinal contaminado com ruido, resultados obtidos com (c) KFO sem atraso e (d) com atraso e seus respectivos valores de SNR e PESQ para $q = 40$.



Fonte: Autoria própria.

Nota-se que o KFO é capaz de amenizar a quantidade de ruído remanescente no sinal de saída, definindo melhor as frequências principais da voz, o que acaba afetando diretamente nos resultados de PESQ e SNR do algoritmo. Dessa forma, adotará-se para os algoritmos desenvolvidos nesse trabalho, um atraso de ordem $q - 1$.

Importante notar que a utilização de um algoritmo com atraso impacta diretamente na ordem utilizada pelo filtro, pois quanto maior a ordem do modelo, maior o impacto de realizar o atraso de amostras do sinal. Na seção 6.1 se realizará testes em relação ao impacto da ordem adotada nos resultados dos sinais estimados adotando um algoritmo com atraso de amostras.

5.1.2 Sobreposição de Janelas

Uma alternativa para melhorar a performance de filtros estocásticos para sinais e voz é apresentada em So e Paliwal (2008), no qual se propõe separar o sinal de entrada em janelas que possuem amostras comum entre elas. Esse tipo de técnica é de comum aplicação em métodos que envolvem processamento digital no domínio da frequência, como visto em subtração espectral (BOLL, 1979).

O intuito dessa aplicação da sobreposição vem da necessidade de se aumentar a janela para obtenção de coeficientes que melhor descrevem o modelo AR enquanto existe a necessi-

dade de janelas menores para contemplar as transições de comportamento dos sinais de fala. Ou seja, ao aplicar a sobreposição, é possível aumentar o tamanho da janela sem comprometer a precisão do algoritmo em detectar variações de comportamento.

Dessa forma, para cada janela n obtém-se suas respectivas matrizes A_n e Q_n. Considerando $\dot{k}$ como a amostra que diretamente antecede a primeira amostra de uma janela, $\mathbf{x}_{\dot{k}+1|\dot{k}+1,n}$ o vetor de estados e $P_{\dot{k}+1|\dot{k}+1,n}$ a matriz de covariância referentes à primeira amostra da janela n, entende-se que para suas estimações se faz necessário que se herde os valores referentes à amostra anterior, $\mathbf{x}_{\dot{k}|\dot{k},n}$ e $P_{\dot{k}|\dot{k},n}$, da janela que antecede. Ou seja

$$\mathbf{x}_{\dot{k}|\dot{k},n} = \mathbf{x}_{\dot{k}|\dot{k},n-1}, \tag{46}$$

$$P_{\dot{k}|\dot{k},n} = P_{\dot{k}|\dot{k},n-1}. \tag{47}$$

Dessa forma, faz-se necessário, ao estimar uma janela de tempo, armazenar os valores de $\mathbf{x}_{\dot{k}|\dot{k},n}$ e $P_{\dot{k}|\dot{k},n}$ referentes à amostra anterior ao início das amostras sobrepostas com a janela seguinte.

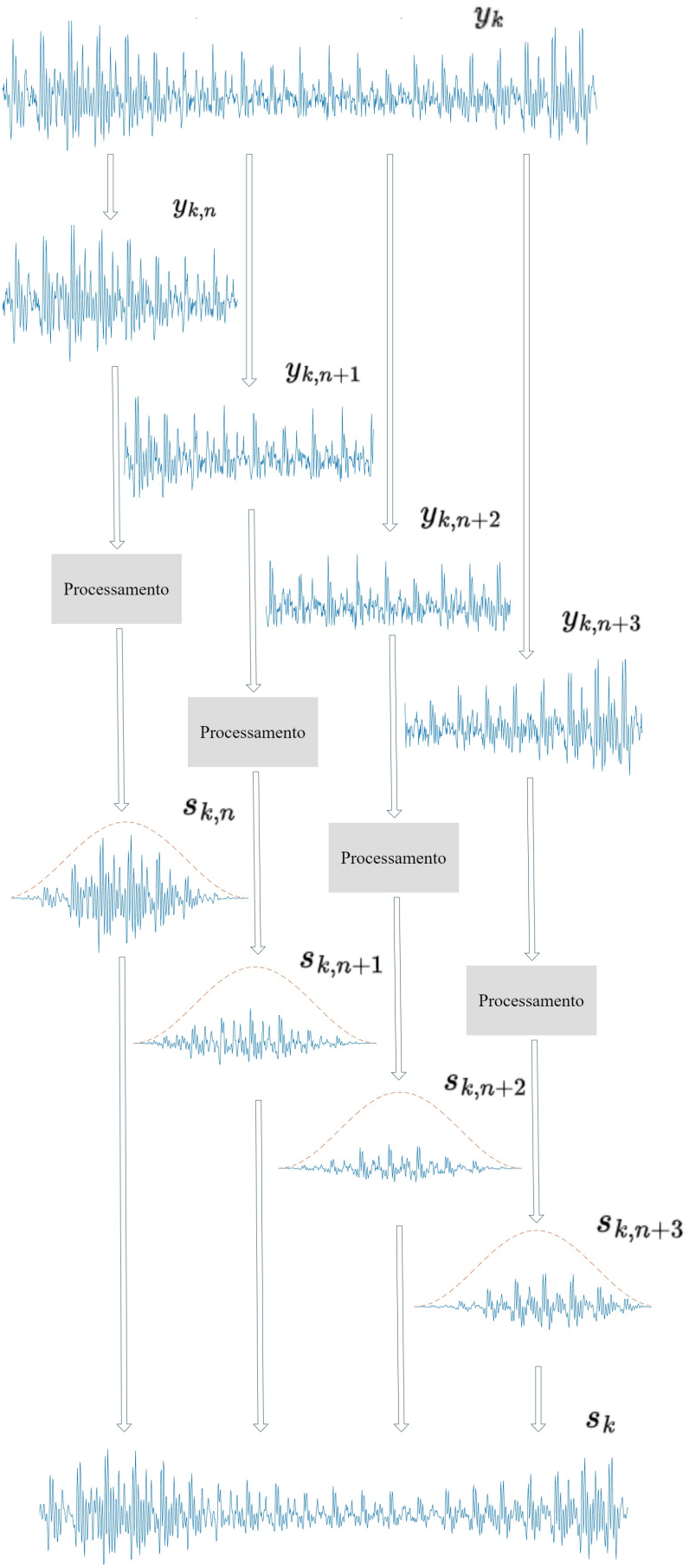
Após a realização do processo de estimação das janelas, a fim de reconstruir o sinal, é aplicado um OLA, no qual as amostras super-posicionadas de uma janela são multiplicados por uma função $L_n(k)$ que varie entre 1 a 0 nas N_w amostras super-posicionadas da janela n, enquanto as amostras super-posicionadas da janela $n+1$ são multiplicadas por $L_{n+1}(k)$. Assim, sendo $S_{n|n+z}$ o conjunto de amostras k super-posicionadas que pertencem as janelas n e $n+d$, sendo $d \in \mathbb{N}$, tem-se que:

$$\sum_{i=0}^{d} L_{n+i}(k) = 1; \qquad \forall k \in S_{n|n+1}. \tag{48}$$

Ou seja, o OLA visa a realizar uma transição suavizada do sinal da janela n para as janelas seguintes ao longo das amostras de $S_{n|n+z}$. Dessa forma, ao se realizar uma sintetização através de janelas sobrepostas, tem-se como efeito uma transição menos abrupta de uma janela para outra, visto a necessidade de se trocar de modelo para cada nova janela.

Para isso, pode-se aplicar as funções de janela como as discutidas na Seção 2.3.1.2, cuja a opção mais adotada sendo a janela de Hann visto a transição suave para zero em suas bordas. Na Figura 21 ilustra-se o processo de separação em janelas de um sinal, assim como sua reconstrução com um OLA.

Figura 21 – Ilustração de funcionamento de separação de janelas sobrepostas e reconstrução do sinal através de OLA utilizando-se de janelas de Hann.

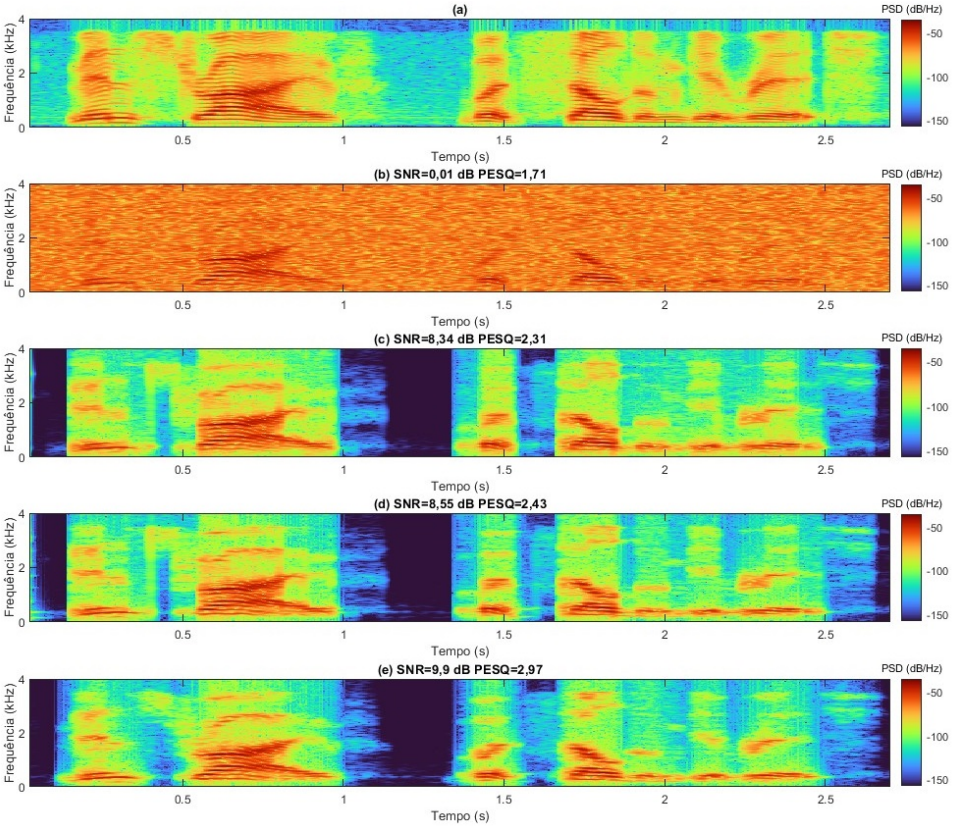


Fonte: Autoria própria.

A sobreposição $S_{n|n+1}$ adotada nesse trabalho será igual a metade do número de amostras das janelas de tempo. Dessa forma, cada amostra do sinal a ser processado pertencerá a exatas duas janelas.

Na Figura 22, demonstra-se os espectrogramas e o comportamento de um sinal estimado por um KFO sem sobreposição e janela de 30 ms, um com janela de 60 ms em relação e um KFO com a sobreposição proposta e tamanho de janela de 60 ms.

Figura 22 – Espectrogramas entre (a) sinal limpo, (b) sinal contaminado com ruido (SNR = 0 dB, PESQ = 1,36), resultados obtidos com (c) KFO sem sobreposição e $T_w = 30$ ms, (d) sem sobreposição e $T_w = 60$ ms e (e) com sobreposição e $T_w = 60$ ms e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Nota-se que as versões sem sobreposição possuem ”quebras” no comportamento espectral do sinal, isso causado pela mudança brusca de uma janela para a próxima. Ao se observar a versão com sobreposição, analisa-se que essas quebras não são mais perceptíveis e que possuem uma quantidade menor de ruído residual em seu espectro.

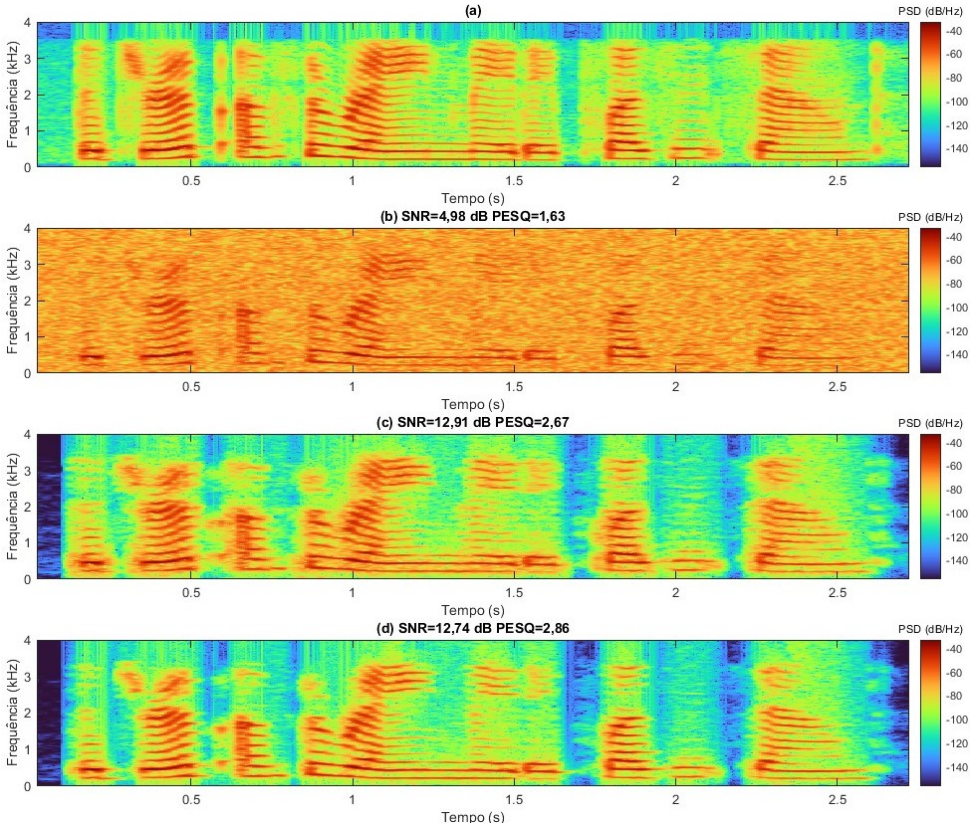
5.1.3 Janelas Temporais para Análise LPC

Uma das modificações para melhor performance do KF para estimação de fala é proposta em So e Paliwal (2011), no qual se compreende que a aplicação de janelas temporais como a de Hamming, aprimora a precisão dos modelos LPC, observando uma melhor resolu-

ção de bandas de frequência, assim como a supressão dos efeitos de borda, como discutido na Seção 2.3.1.2.

Na Figura 23 compara-se a performance de um KFO que utiliza a janela de Hamming em relação a um KFO com janela retangular. Importante notar que a aplicação das funções de janela não é feita para a estimação do KF, que segue a utilizar de uma janela retangular.

Figura 23 – Espectrogramas de (a) sinal limpo, (b) sinal contaminado com ruido, resultados obtidos com (c) KFO com janela retangular e (d) KFO com janela de Hamming e seus respectivos valores de SNR e PESQ.

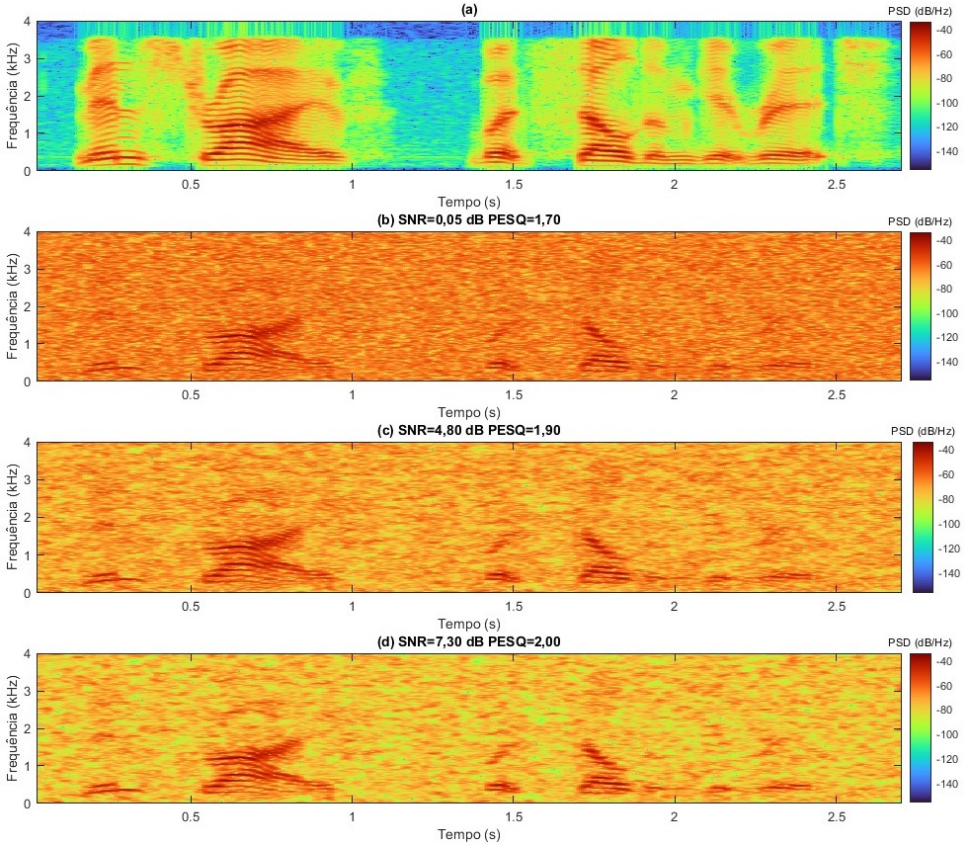


Fonte: Autoria própria.

Ao analisar o espectrograma da Figura 23 é perceptível que a versão sem a aplicação da janela de Hamming possui as frequências principais envolvidas em uma quantidade significativa de ruído, enquanto a versão que aplica a janela demonstra-se capaz de concentrar a energia espectral em frequências mais bem definidas. É observado também que há uma perda em SNR em comparação aos dois, entretanto não significativa como ganho obtido em termos de PESQ.

A mesma análise pode ser feita em relação a um KF cujo modelo é estimado por um sinal ruidoso, como exemplificado na Figura 24. Para ese caso, nota-se uma melhora significativa tanto de PESQ quanto em SNR.

Figura 24 – Espectrogramas de (a) sinal limpo, (b) sinal contaminado com ruido, resultados obtidos com (c) KF não iterativo com janela retangular e (d) KF não iterativo com janela de Hamming e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Dessa forma, entende-se que a aplicação prévia da janela de Hamming faz com que a análise LPC seja benéfica à estimação do sinal pelo KF dada a melhor definição das frequências do sinal de fala, distinguindo-as mais adequadamente dos distúrbios ruidosos para obtenção de modelo AR.

5.2 Filtro de Kalman Iterativo

Ao se estimar um modelo de um sinal ruidoso, a estimação de σ_{ω}^2, obtida através da equação(41) gera um enviesamento dos valores de $K_{q,k}$, sendo $K_{q,k}$ o valor escalar correspondente a última posição do vetor K_k, pois visto a autocorrelação da janela um sinal ruidoso $R_{y,n}$ (ROY; PALIWAL, 2021):

$$\sigma_{\omega,n}^2 = R_{y,n}(0) + \sum_{i=1}^{q} a_i R_{y,n}(i), \tag{49}$$

$$\sigma_{\omega,n}^2 = R_{y,n}(0) + \sigma_u^2 + \sum_{i=1}^{q} a_i R_{y,n}(i). \tag{50}$$

Para um ruído gaussiano, entende-se que:

$$R_w(i) \approx 0 \qquad \forall \quad i > 0, \tag{51}$$

então tem-se que:

$$\sigma^2_{\omega,n} \approx \sigma^2_u + R_{y,n}(0) + \sum_{i=1}^{q} a_i R_{y,n}(i). \tag{52}$$

Logo, ao se calcular K_k, de acordo com a Equação 22 obtém-se:

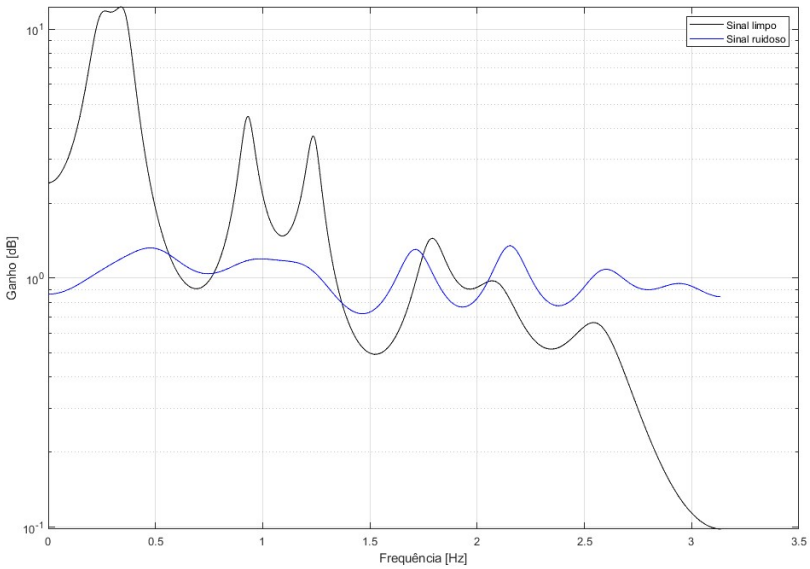
$$K_k = \frac{G_n P_{k-1|k-1} {G_n}^\intercal + Q_{k-1}}{C P_{k|k-1} C^\intercal + \sigma^2_u}, \tag{53}$$

$$K_{q,k} = \frac{G_{q,n} P_{q,k-1|k-1} {G_{q,n}}^\intercal + \sigma^2_u + R_{s,n}(0) + \sum_{i=1}^{q} a_i R_s(i)}{C P_{k|k-1} C^\intercal + \sigma^2_u}. \tag{54}$$

Dessa forma, o ruído aditivo que compromete o valor $K_{0,k}$ sempre terá um valor mínimo positivo, mesmo que em momentos de ruído isolado faça com que uma quantidade maior de ruído residual remanesça no sinal de fala estimado s'_k.

Uma possibilidade seria somente subtrair de $\sigma^2_{\omega,n}$ o valor estimado de σ^2_u. Entretanto deve-se entender que os coeficientes a_i não são completamente coerentes com o sinal limpo s_k – como demonstra a Figura 25–, e ao reduzir a incerteza, força-se que o ganho K_k aproxime-se de zero mesmo quando modelo não represente bem o sinal a ser estimado, gerando um sinal enviesado pelo ruído.

Figura 25 – Resposta em frequência de modelo obtido a partir de sinal limpo e modelo obtido a partir de sinal ruidoso em que SNR = 0 dB

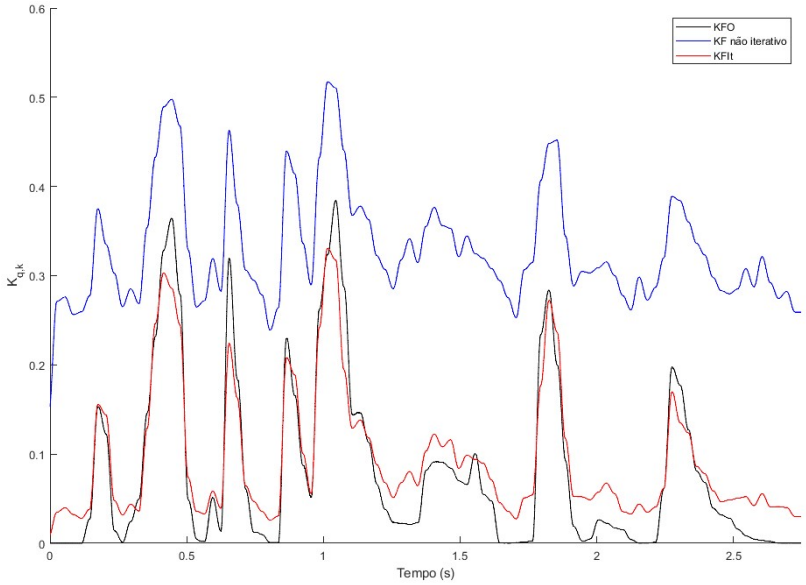


Fonte: Autoria própria.

Dessa forma, para obter um menor enviesamento do modelo, é possível utilizar o modelo de um sinal estimado pelo KF para então fazer uma nova estimação do sinal. Ou seja, o KF é iterado uma vez com o modelo de sinal ruidoso A_n e Q_n e outra vez com o modelo de sinal estimado A'_n e Q'_n. Esse algoritmo é normalmente denominado como KF Iterativo (KFIt) (GIBSON; KOO; GRAY, 1989). Importante destacar que se adota o valor de $P_{\dot{k}|\dot{k}}$ e $x_{\dot{k}|\dot{k}}$ como o mesmo da primeira iteração da janela.

O KFIt apresenta-se como uma possível solução desse problema, pois, a partir que da segunda iteração do LPC é obtido o modelo através de um sinal menos ruidoso, aprimorando a estimativa do valor de $\sigma^2_{\omega,n}$. Na Figura 26, compara-se o $K_{0,k}$ de um KFIt em relação a um KF não iterativo ao longo das amostras e comparando esses resultados com o ganho obtido através de um KFO.

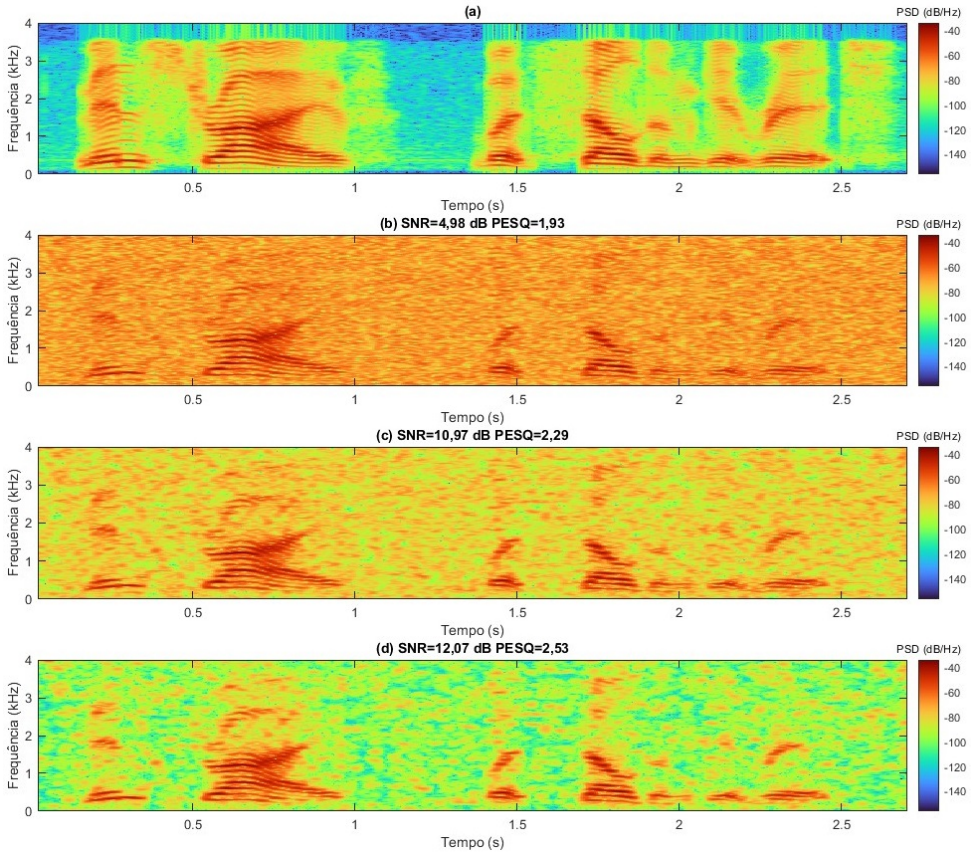
Figura 26 – Resultados de $K_{q,k}$ no domínio do tempo do KFO, KF não iterativo e KFIt, estimado de um sinal com 0 SNR.



Fonte: Autoria própria.

Na Figura 27 compara-se o desempenho de um KF não iterativo com um KFIt a partir de seus espectrogramas. Nota-se a partir Figura 27 que em ambos domínios o KFIt possui uma quantidade significativamente de ruído residual. Entretanto, ao introduzir a iteratividade no algoritmo, os níveis de ruído são consideravelmente amenizados.

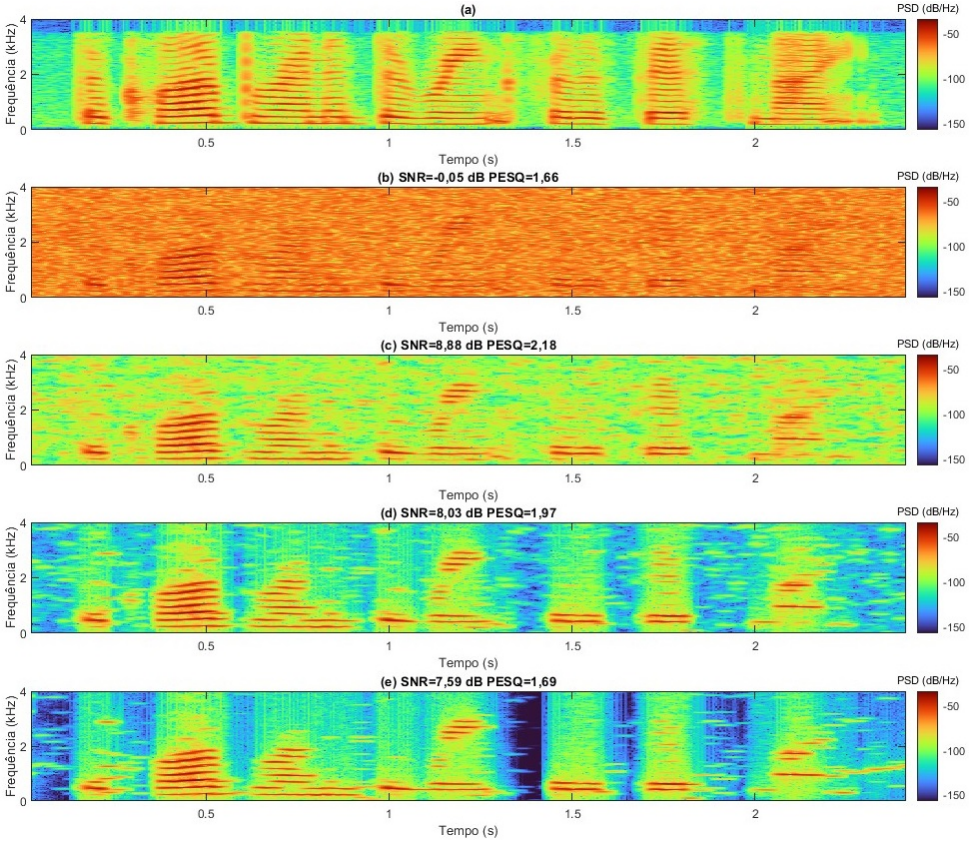
Figura 27 – Espectrogramas de (a) sinal limpo, (b) sinal contaminado com ruido, resultados obtidos com (c) KF não iterativo, e (d) KFIt e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

É possível utilizar uma lógica iterativa para ter melhores resultados, seria lógico utilizar um número maior de iterações. Entretanto, experimentalmente observa-se, como na Figura 28, que ao se iterar mais que duas vezes o sinal de fala é comprometido quanto aos valores obtidos de SNR e PESQ, apesar dos espectrogramas também indicarem uma melhor remoção de ruídos em momentos sem fala.

Figura 28 – Espectrogramas de (a) sinal limpo, (b) sinal contaminado com ruido, (c) KFIt de duas iterações, (d) KFIt de três iterações e (e) KFIt de quatro iterações e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

5.3 Filtro de Kalman Aumentado para Estimação de Ruído Colorido

A fim de expandir a usabilidade da técnica para sinais para ruídos coloridos, é necessário que se altere o modelo utilizado para o filtro estocástico de maneira que se obtenha uma estimativa relativa à parte determinística desse ruído. Originalmente, tal técnica fora proposta em Gibson, Koo e Gray (1989), no qual define-se um sinal contaminado ruído colorido, como mostra a Equação 55

$$y_k = s_k + r_k, \tag{55}$$

sendo s_k o sinal de fala e r_k o ruído aditivo, que pode ser escrito da seguinte forma:

$$r_k = \sum_{i=1}^{q} b_i r_{k-i} + v_k, \tag{56}$$

sendo b_i os coeficientes do modelo AR do ruído e q a ordem do modelo AR. Dessa forma, obtém-se um modelo AR da parte determinística do sinal, da mesma forma que é obtido um modelo de fala.

Esse coeficientes podem ser obtidos de acordo com o descrito na Seção 5.1, considerando uma janela de tempo com o ruído isolado, sem sinal de fala, para ruídos que tenham comportamento estacionário.

Assim, a estimativa de um sinal de voz sem ruído pode ser obtida via um KF, reescrevendo as equações (38)-(40) de maneira a contemplar também o modelo da parte determinística do ruído colorido. Ou seja, o vetor de estados consiste na concatenação do vetor de amostras passadas de s_k e do vetor de amostras passadas de r_k

$$\mathbf{x}_k = \begin{bmatrix} s_{k-q} \\ \vdots \\ s_k \\ r_{k-q} \\ \vdots \\ r_k \end{bmatrix}. \tag{57}$$

Dessa forma, entende-se que a matriz de estados do modelo pode ser descrita como

$$G_n = \begin{bmatrix} A_n & 0 \\ 0 & B_n \end{bmatrix}, \tag{58}$$

com B_n sendo a matriz de estados do modelo do ruído descrito como:

$$B_n = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ -b_{p,n} & -b_{p-1,n} & -b_{p-2,n} & -b_{p-3,n} & \dots & -b_{1,n} \end{bmatrix}. \tag{59}$$

No caso ideal, isto é, quando se conhece perfeitamente perfeitamente o sinal de fala, as matrizes A_n e B_n podem ser obtidas independentemente através dos sinais isolados de ruído e fala, determinando-o como KFO Aumentado. Porém, para algoritmos não ideais, o comportamento do ruído na identificação do AR do sinal de fala é comprometido com o comportamento do ruído colorido.

Portanto, sem tratamento prévio, os algoritmos de filtros estocásticos estimariam esses comportamentos determinísticos do ruído como um comportamento de fala, sendo necessária a modificação do sinal de entrada do LPC para conseguir estimar a_i sem o viés do comportamento do ruído colorido aditivo. Dessa forma, para obtenção de A_n, em trabalhos como George *et al.*

(2018) e Roy e Paliwal (2021), propõe-se a utilização de um filtro embranquecedor (WF, do inglês *Whitening Filter*), que usa b_i como coeficiente de um filtro FIR como descrito na equação (60):

$$W(z) = \sum_{i=0}^{p} b_i z^i. \tag{60}$$

Nesse caso, a ideia é aproximar o comportamento de um ruído colorido ao comportamento de um ruído branco ao se obter o modelo.

Para determinar a incerteza do ruído de observação dos filtros aumentados, é possível aproximar σ_u^2 como:

$$\sigma_u^2 \approx R_r(0) - \max[|R_r'|], \tag{61}$$

sendo $R_r' = R_r(i)$ para todo $i > 0$. Essa aproximação se deve ao fato de que para um sinal de ruído branco u_k (BOSWORTH *et al.*, 2008), o comportamento da sua autocorrelação R_u tende a ser descrito como:

$$R_u(0) >> R_u' \approx 0. \tag{62}$$

Já num sinal determinístico c_k, é possível entender que:

$$R_c(0) \approx \max[|R_c'|]. \tag{63}$$

Considerando que um ruído colorido é uma mistura entre comportamentos determinísticos e não determinísticos, logo a quantidade de ruído branco que esse sinal contém pode ser definido como

$$\sigma_u^2 = R_u(0) = R_r(0) - R_c(0), \tag{64}$$

$$R_c(0) = \max[|R_r' - R_\omega'|] \approx \max[|R_r'|]. \tag{65}$$

Logo, de (65) e (64) obtém-se o resultado apresentado na equação (61)

Para representar a (55) em espaço de estados, adota-se C como um vetor linha de dimensão $p + q$, definido como a concatenação do vetor do sinal C_s de ordem q e do ruído C_r de ordem p, ou seja

$$C_s = [0 \quad 0 \quad ... \quad 0 \quad 1], \tag{66}$$

$$C_r = [0 \quad 0 \quad ... \quad 0 \quad 1], \tag{67}$$

$$C = [C_s \quad C_r]. \tag{68}$$

Quanto aos parâmetros do KF, estes definem-se como

$$P_0 = \sigma_u^2 I, \tag{69}$$

$$\Theta = \sigma_u^2, \tag{70}$$

$$Q = \begin{bmatrix} Q_s & 0 \\ 0 & Q_r \end{bmatrix}, \tag{71}$$

sendo

$$Q_s = \sigma_\omega^2 C_s^\intercal C_s, \tag{72}$$

$$Q_r = \sigma_v^2 C_r^\intercal C_r = \left(R(0) + \sum_{i=1}^{p} b_i R_r(i) \right) C_r^\intercal C_r. \tag{73}$$

5.4 Filtro IMM para Estimação de Sinais de Fala

Nesta seção, será abordado o algoritmo do filtro IMM com modelos AR obtidos por LPC de janelas adjacentes para a supressão de ruídos em sinais de fala. Esse algoritmo se destina à redução de ruídos presentes nos sinais, empregando uma abordagem que incorpora a adaptação dinâmica de modelos ao longo do tempo.

Conforme descrito na Seção 3.3, o IMM permite a utilização simultânea de diferentes filtros que alternam para melhor se adequar à dinâmica variável do sinal. No cenário específico deste trabalho, três modelos distintos são empregados para estimação do sinal: o modelo da janela passada, o modelo da janela atual e o modelo da janela futura.

A motivação dessa aplicação desse segundo modelo vem da compreensão de que ao aumentar a janela temporal relativa à análise LPC, consequentemente possibilita-se uma maior precisão nos coeficientes obtidos e maior robustez ao ruído. Supõe-se que a análise com múltiplos modelos possibilita a detecção e adaptação de transições de comportamento, evitando-se que a estimação enviese-se em relação a um único modelo de janela. Ao se utilizar de sobreposição de janelas entende-se que o chaveamento ocorra naturalmente no começo do modelo da janela passada para o modelo da janela atual, e no fim do modelo da janela atual para o modelo da janela futura. Dessa forma, estipula-se:

$$\Omega = \begin{bmatrix} 0{,}995 & 0{,}005 & 0{,}000 \\ 0{,}005 & 0{,}990 & 0{,}005 \\ 0{,}000 & 0{,}005 & 0{,}995 \end{bmatrix}, \tag{74}$$

Depois da estimação de todas as amostras da janela n, esse processo é iterado para a próxima janela considerando a herança dos os valores de $\mathbf{x}_{k|k}$, $P_{k|k}$, assim como ν_k, referentes à amostra

k que antecede o início da janela, conforme explicado na Seção 5.1.2, de forma que

$$\nu_{\dot{k},n+1}^{(1)} = \frac{\nu_{\dot{k},n}^{(2)}}{\nu_{\dot{k},n}^{(2)} + \nu_{\dot{k},n}^{(3)}}, \tag{75}$$

$$\nu_{\dot{k},n+1}^{(2)} = \frac{\nu_{\dot{k},n}^{(3)}}{\nu_{\dot{k},n}^{(2)} + \nu_{\dot{k},n}^{(3)}}, \tag{76}$$

$$\nu_{\dot{k},n+1}^{(3)} = 0, \tag{77}$$

$$P_{k|\dot{k},n+1}^{(1)} = P_{\dot{k}|\dot{k},n}^{(2)}, \tag{78}$$

$$P_{\dot{k}|\dot{k},n+1}^{(2)} = P_{\dot{k}|\dot{k},n}^{(3)}, \tag{79}$$

$$P_{k|\dot{k},n+1}^{(3)} = P_{k|\dot{k},n}, \tag{80}$$

$$\mathbf{x}_{k|\dot{k},n+1}^{(1)} = \mathbf{x}_{k|\dot{k},n}^{(2)}, \tag{81}$$

$$\mathbf{x}_{\dot{k}|\dot{k},n+1}^{(2)} = \mathbf{x}_{\dot{k}|\dot{k},n}^{(3)}, \tag{82}$$

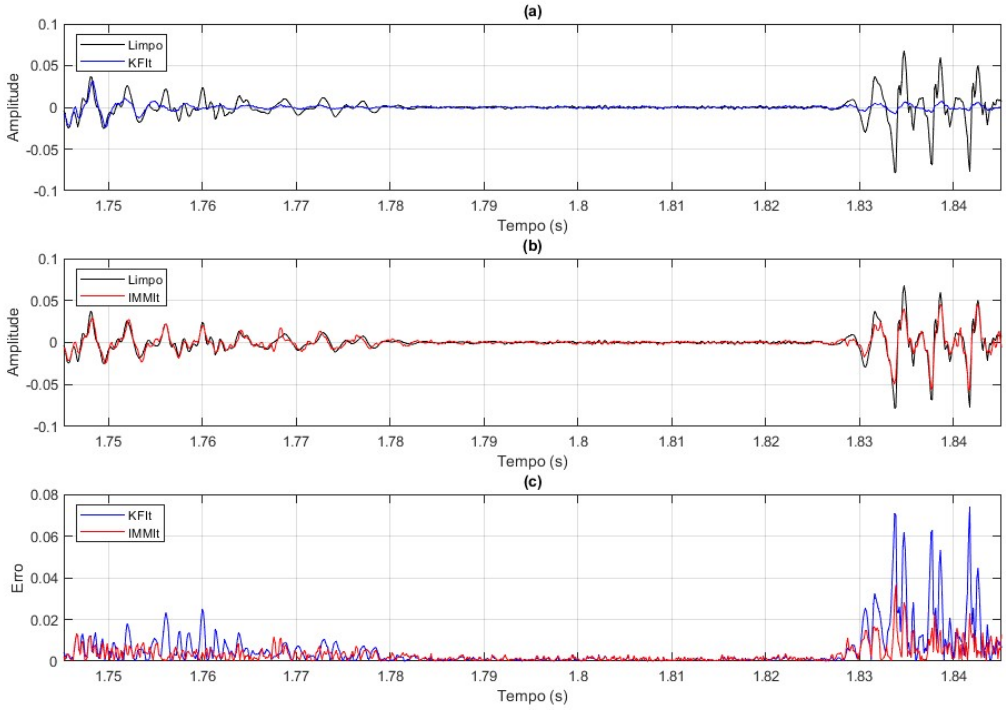
$$\mathbf{x}_{k|\dot{k},n+1}^{(3)} = \mathbf{x}_{\dot{k}|\dot{k}}. \tag{83}$$

A fim de estudar se é valido esse comportamento, nesse trabalho é desenvolvido o IMM Oráculo (IMMO) – que se utiliza do sinal limpo para obter os modelos AR – que será utilizado para comparar a performance ideal da técnica e compará-la ao KFO.

Para sua aplicação em condições não ideais, é desenvolvido também o IMM Iterativo (IMMIt), similar ao KFIt, a fim de desenviesar o efeito do ruido aditivo na obtenção do modelo, o modelo utilizado A'_n é obtido através da análise LPC de um sinal estimado por um KF. Já o IMMO Aumentado (IMMOA) e o IMMIt Aumentado (IMMItA) consiste nas versões dos filtro IMMO e IMMIt na qual se utiliza de um filtro aumentado para se estimar r_k, assim como proposto para o caso dos KFOA e KFItA.

O intuito, portanto, é que, ao limitar o modelo em um único instante temporal, resulta-se em uma representação enviesada do sinal, desconsiderando transições do comportamento do sinal que possam ocorrer dentro de uma mesma janela. Isso é agravado ao se utilizar de funções de janela para análise LPC, pois o modelo obtido desta janela tende a possuir uma maior precisão no centro de sua janela e uma precisão menor nos extremos dessa. Isso é exemplificado na Figura 29, na qual se mostra o erro na performance do KFIt de uma janela, comparada ao IMMIt proposto.

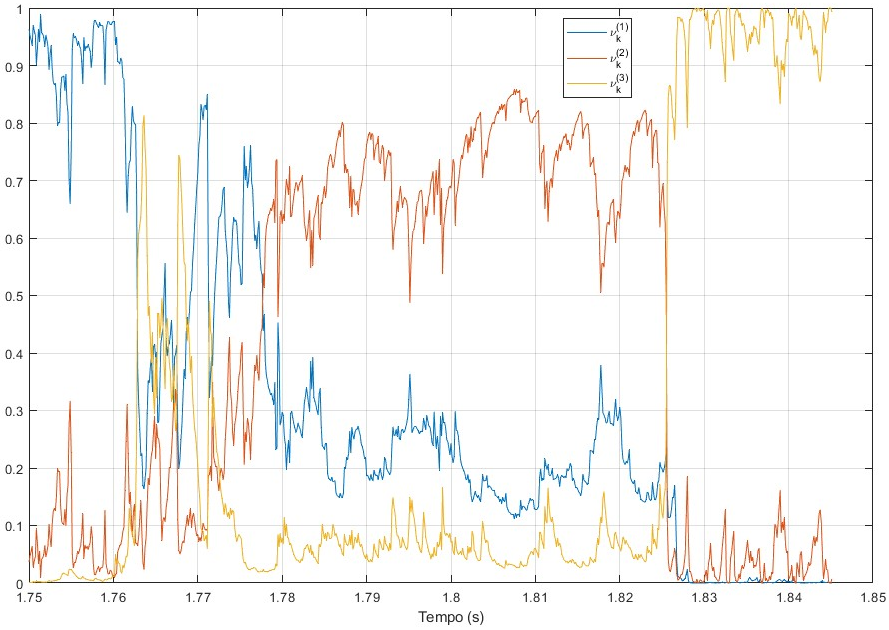
Figura 29 – Em (a) Comparação entre janela de sinal limpo e janela estimada por KFIt, em (b) comparação entre janela de sinal limpo e janela estimada por IMMIt, em (c) comparação entre o erro do KFIt e erro do IMMIt.



Fonte: Autoria própria.

Na Figura 30, para a mesma janela de tempo, exibe-se o comportamento de ν_k ao longo do tempo para o mesmo sinal que o estimado na Figura 29. Demonstra-se, assim, que o IMMIt realmente é capaz de transitar entre diferentes modelos, em ordem, em uma mesma janela considerando o Ω descrito na equação (74).

Figura 30 – Valores de ν_k ao longo das amostras de uma janela de tempo



Fonte: Autoria própria.

Entretanto, entende-se que ao sobrepor janelas como proposto na Seção 5.1.2 esse problema tende a ser mitigado, pois os extremos das janelas, quando somado às janelas adjacentes, tende-se a predominar o sinal das amostras centrais. Entretanto, mesmo que minimamente, os lóbulos laterais das janelas adjacentes podem comprometer negativamente o resultado do sinal estimado, algo que se evita ao se estimar através s'_k do método proposto.

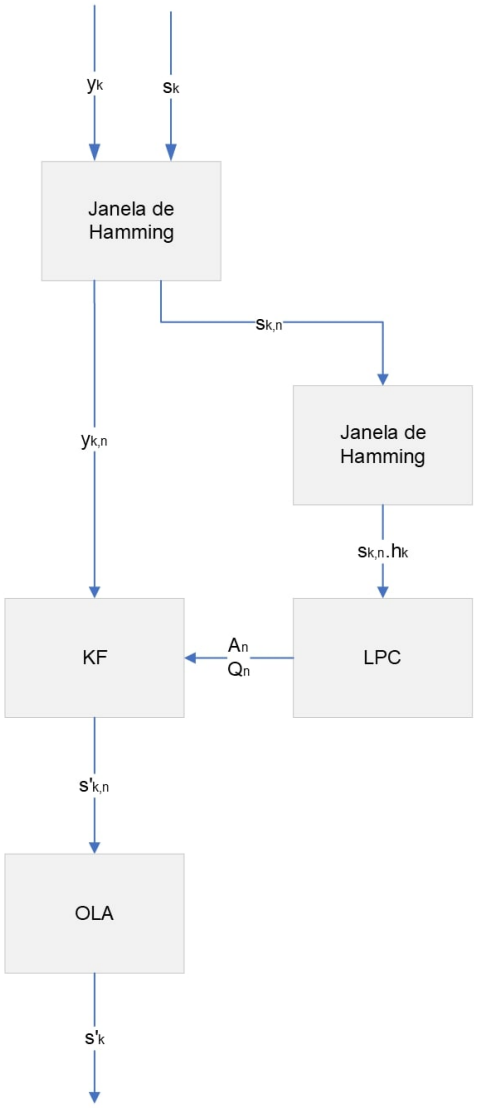
5.5 Resumo dos Algoritmos Estudados

Nesse capítulo discutiu-se diversos métodos de aprimoramento de fala via aplicação de filtros estocásticos. Essa seção tem como objetivo delimitar e resumir por meio de fluxogramas a implementação desses algoritmos.

5.5.1 Filtro de Kalman Oráculo

Utilizado como referência em relação ao desempenho dos filtros desenvolvidos no trabalho, o KFO consiste na estimação do sinal com base em modelos LPC obtidos através do sinal limpo do sinal. Ou seja, por meio desse filtro evita-se que o modelo do sinal seja enviesado pelo ruído aditivo. O algoritmo é ilustrado com base no diagrama de blocos da Figura 31.

Figura 31 – Diagrama simplificado de funcionamento do KFO.

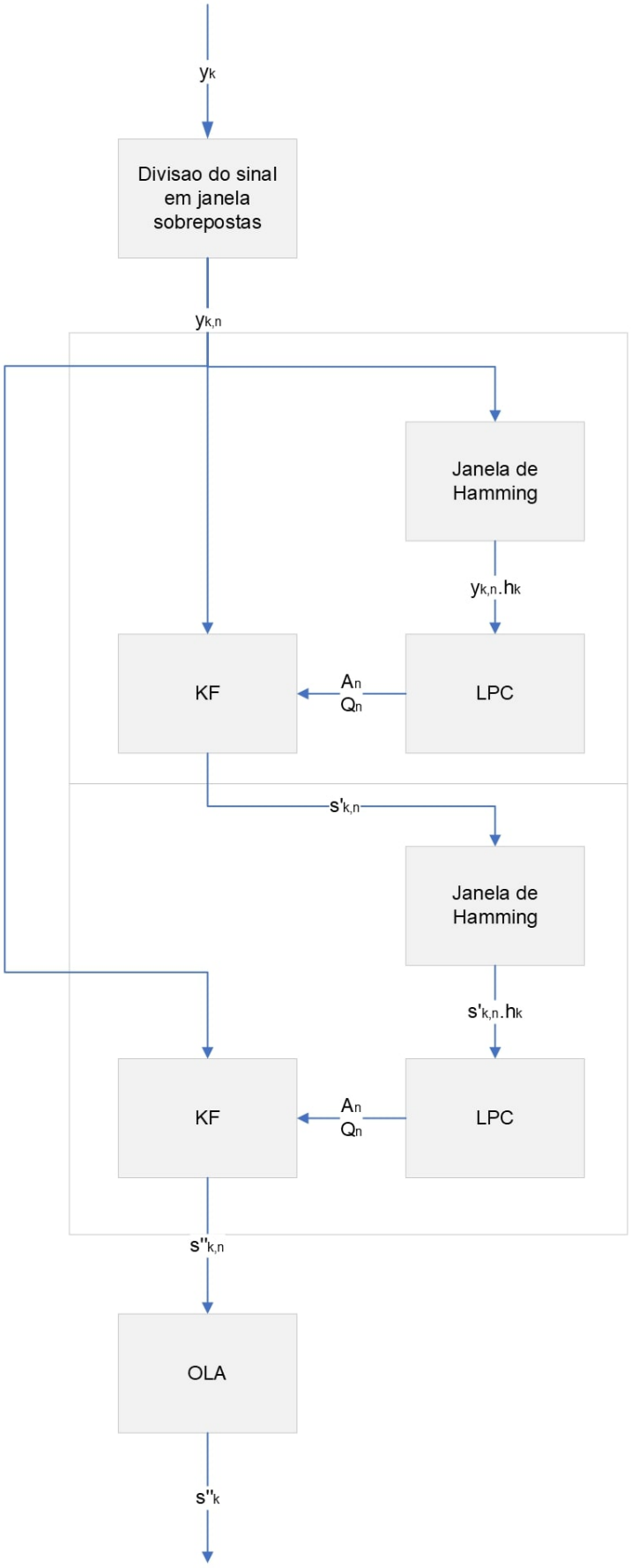


Fonte: Autoria própria.

5.5.2 Filtro de Kalman Iterativo

No FKIt, o sinal ruidoso é segmentado em janelas das quais se é estimado por meio de LPC os coeficientes que determinam o modelo AR do sinal daquele instante. Esses coeficientes são obtidos de maneira iterativa, ou seja, novos coeficientes são obtidos a partir do sinal estimado pelo FK. Denota-se que sua aplicação isolada limita-se a sinais corrompidos com ruídos brancos. Seu funcionamento pode ser resumido de acordo com o diagrama de blocos da Figura 32.

Figura 32 – Diagrama simplificado de funcionamento do KFIt.

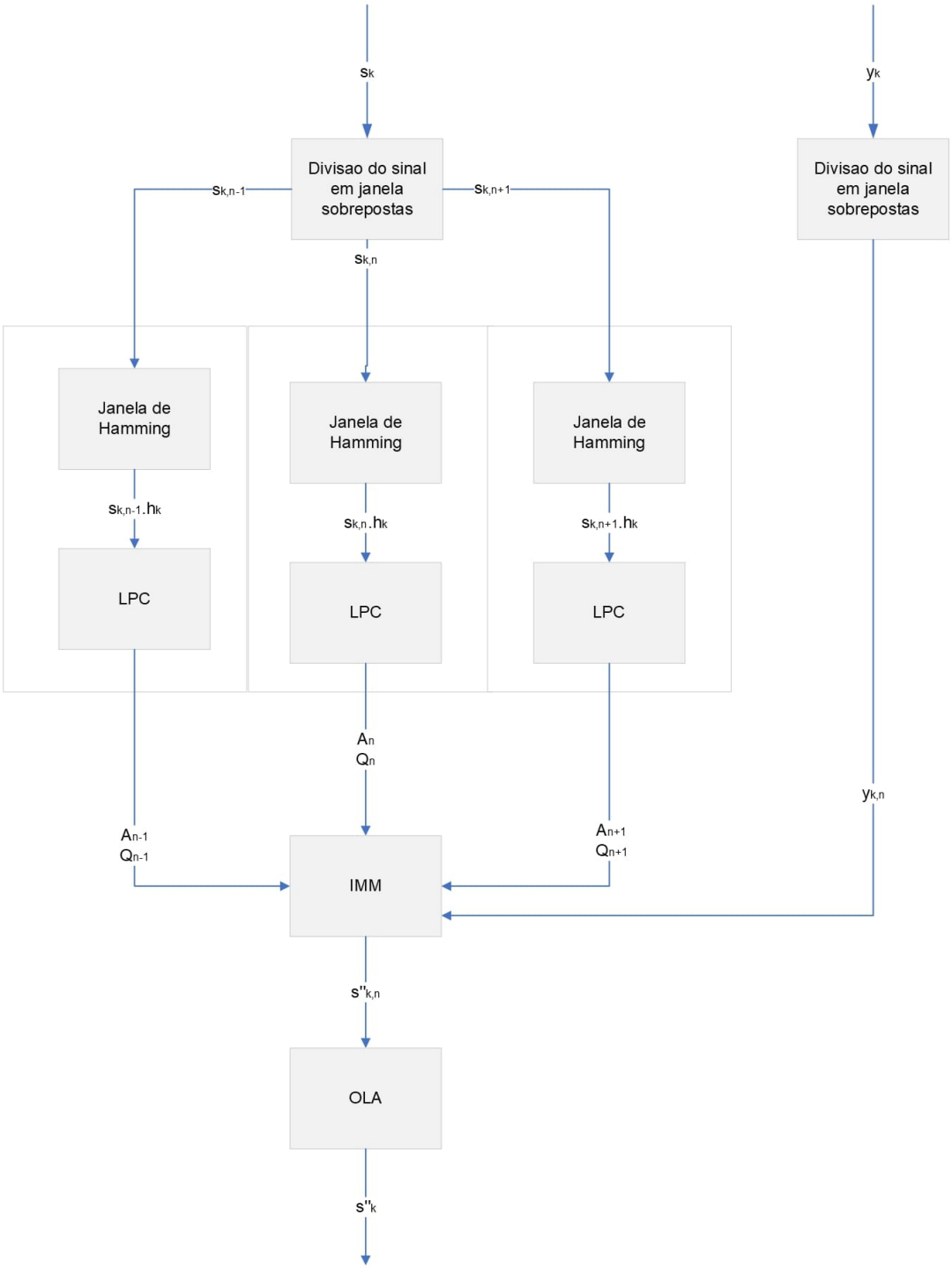


Fonte: Autoria própria.

5.5.3 Filtro IMM Oráculo

O IMMO é a versão ideal do funcionamento do IMM como um estimador do sinal de fala. Nesse utiliza-se o modelo dos sinais de fala de três de janela de tempos adjacentes a fim de se estimar o modelo ideal que então serão utilizados para estimação por um filtro IMM. Ilustra-se seu funcionamento na Figura 33

Figura 33 – Diagrama simplificado de funcionamento do IMMO.

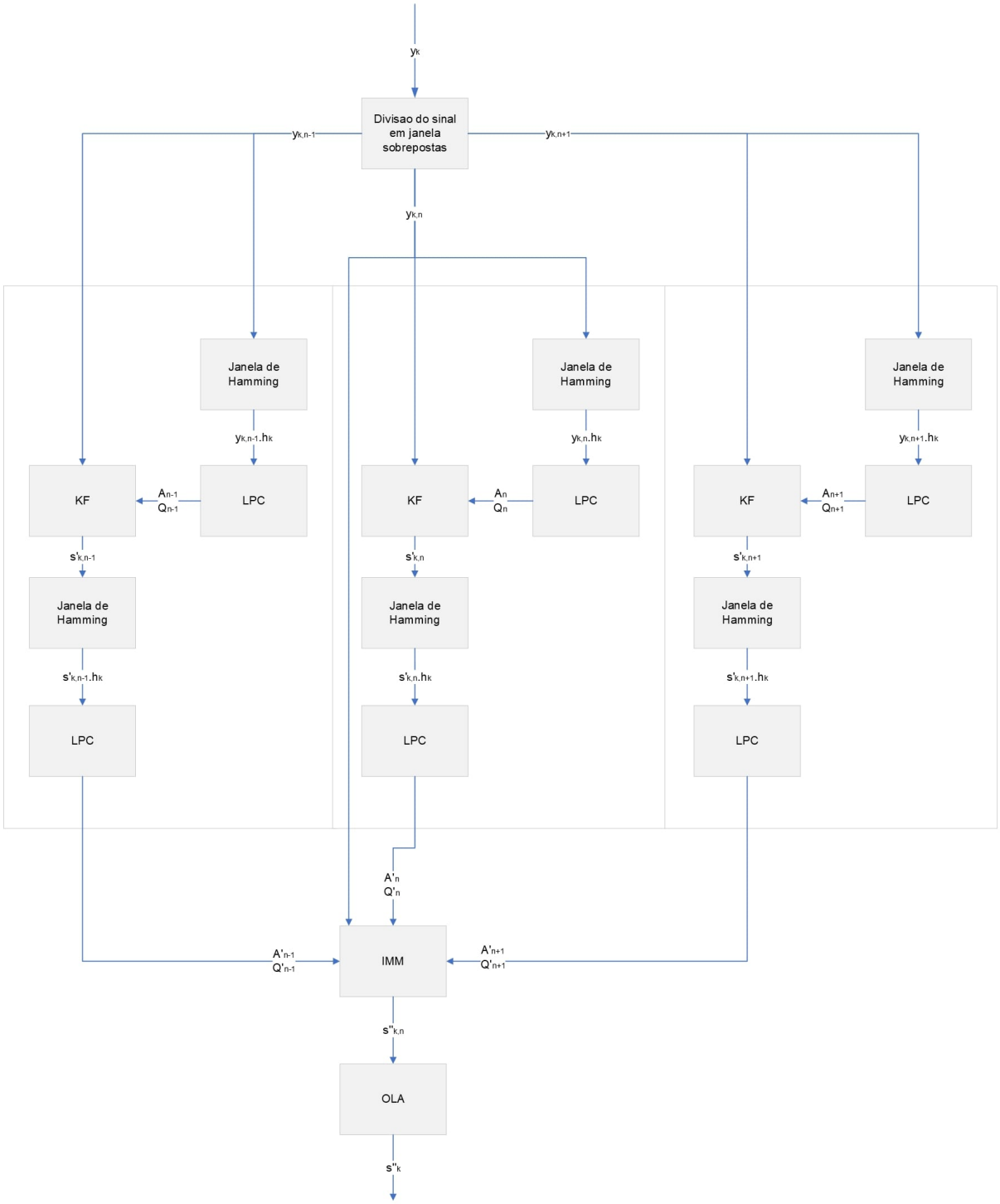


Fonte: Autoria própria.

5.5.4 Filtro IMM Iterativo

No IMM Iterativo (IMMIt) considera-se a interação entre três modelos de AR de fala: o modelo da janela anterior, atual e futura, estimados através do LPC de sinais previamente estimados por um KF. Seu funcionamento é resumido através da Figura 34.

Figura 34 – Diagrama simplificado de funcionamento do IMMIt.

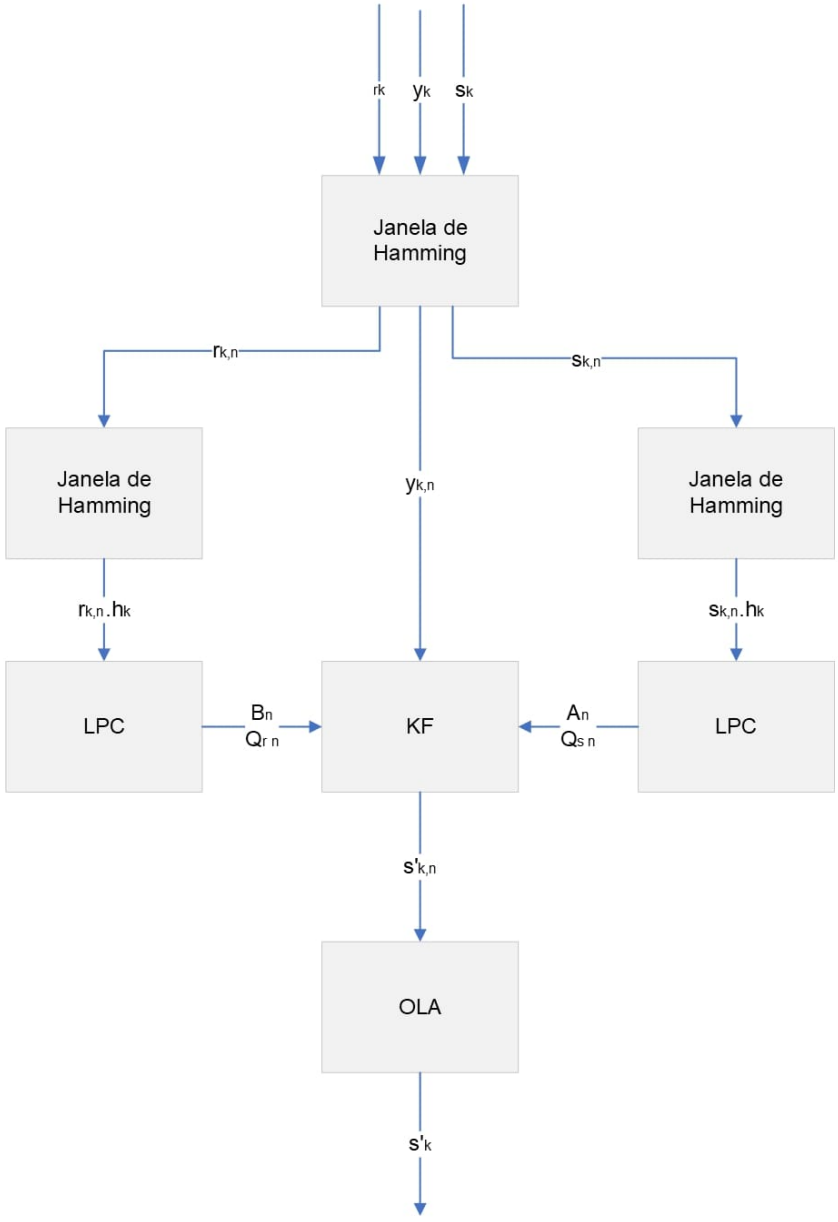


Fonte: Autoria própria.

5.5.5 Filtro de Kalman Oráculo Aumentado

O KFO aumentado (KFOA) desenvolvido para remover o efeito de ruídos coloridos na estimação dos sinal, no qual se desenvolve um KF aumentado cujo objetivo é estimar tanto o sinal de fala s_k quanto o sinal de ruído r_k. No caso do KFOA, os coeficientes AR são obtidos diretamente de um sinal de ruído isolado $r_{k,n}$ referente ao período de tempo de cada janela n. Dessa forma, é um filtro que funciona considerando o conhecimento tanto do sinal de fala quanto do sinal de ruído. O diagrama de blocos referente ao seu funcionamento é demonstrado na Figura 35.

Figura 35 – Diagrama simplificado de funcionamento do KFOA.

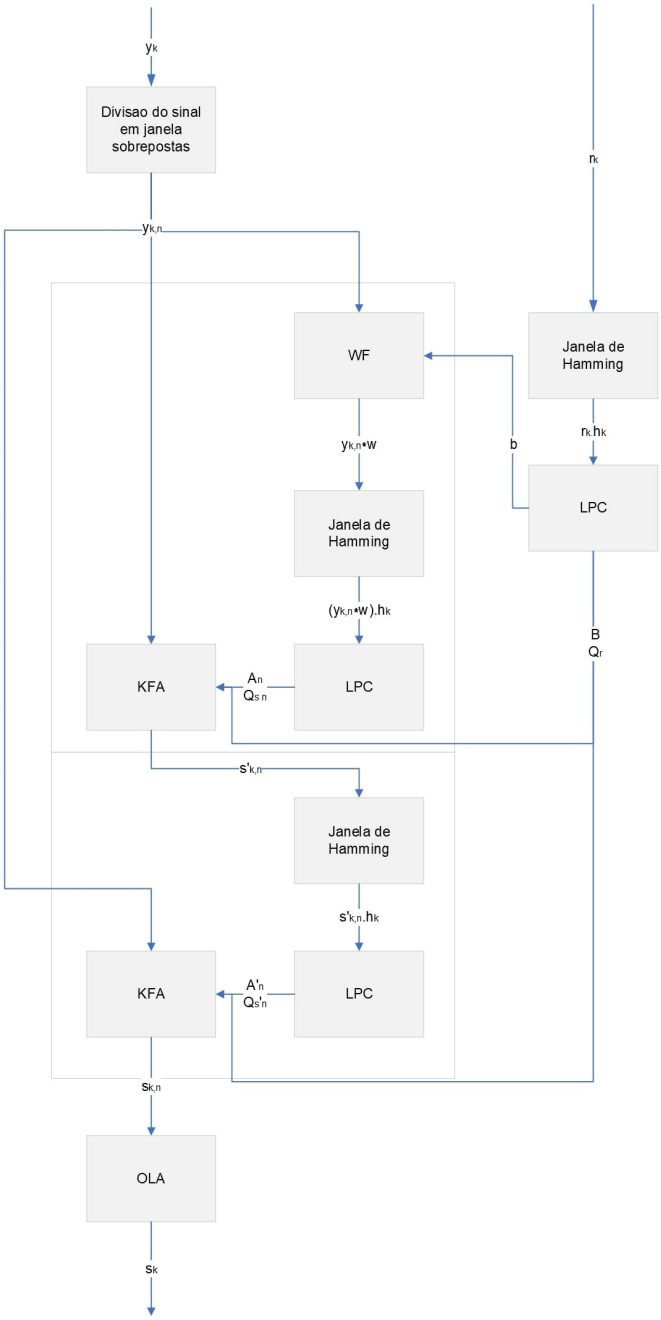


Fonte: Autoria própria.

5.5.6 Filtro de Kalman Iterativo Aumentado

O KFIt Aumentado (KFItA) se baseia na arquitetura de funcionamento de um KFIt aplicação a um filtro aumentado para estimação de estado do sinal r_k. Para isso, adota-se um WF, como descrito na equação (60), a fim de desenviesar do comportamento do sinal y_k para obtenção de a_n ao se realizar a análise LPC. Isso só é feito na primeira iteração, em que já se obtém um sinal isolado de fala s'_k que será o utilizado para se obter o modelo para a próxima iteração.

Figura 36 – Diagrama simplificado de funcionamento do KFItA

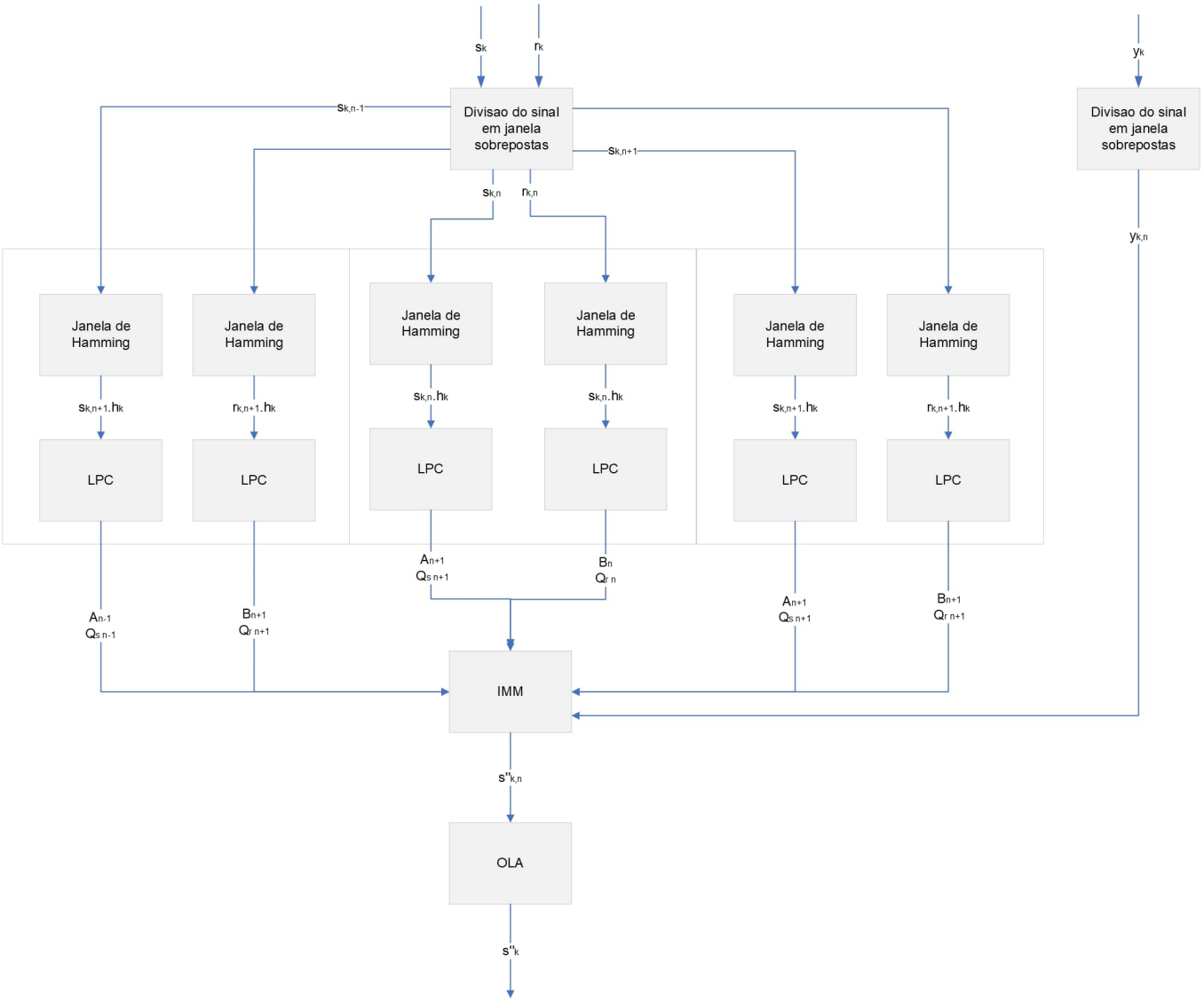


Fonte: Autoria própria.

5.5.7 Filtro IMM Oráculo Aumentado

O IMMOA consiste na implementação do mesmo modelo de ruído com a lógica desenvolvida no KFOA para um filtro IMMO. A Figura 37 demonstra um diagrama de blocos resumido do funcionamento do algoritmo.

Figura 37 – Diagrama simplificado de funcionamento do IMMOA

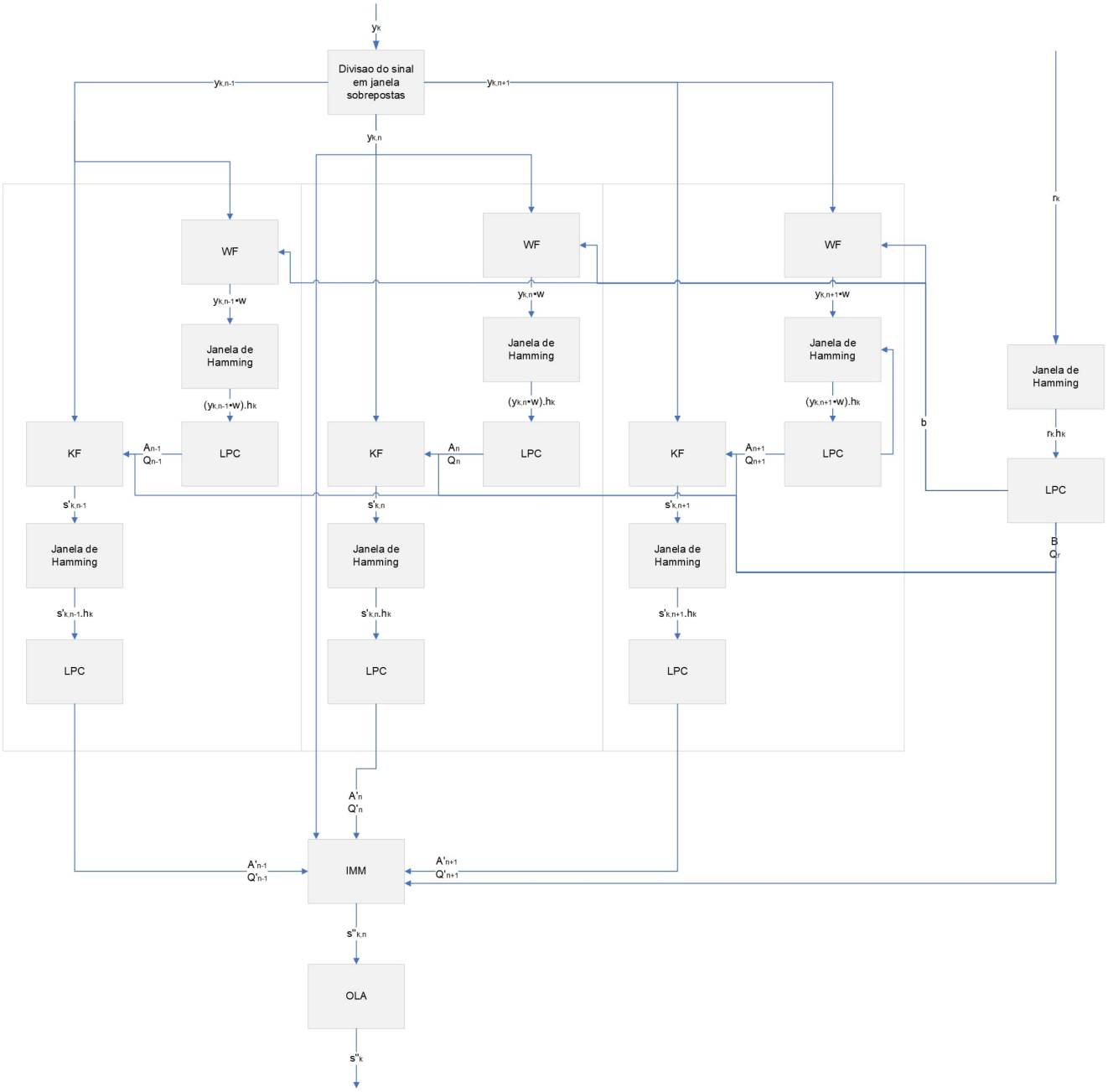


Fonte: Autoria própria.

5.5.8 Filtro IMM Iterativo Aumentado

O filtro IMM iterativo aumentado (IMMITA) trata-se da implementação do mesmo modelo de ruído a lógica desenvolvida no KFItA para um filtro IMMIt. Resume-se seu funcionamento através dpo diagrama de blocos da Figura 38.

Figura 38 – Diagrama simplificado de funcionamento do IMMItA



Fonte: Autoria própria.

6 RESULTADOS E DISCUSSÕES

Neste capítulo, apresentam-se os resultados e discussões relacionados à aplicação das técnicas de KF e IMM no contexto da redução de ruídos em sinais de fala. Este capítulo busca elucidar os impactos e implicações dos algoritmos adotados. Os códigos em MATLAB® utilizados estão disponíveis na página Github®[1] referente a esse trabalho.

6.1 Variação de ordem do modelo AR

Foi feita uma análise do impacto da ordem do modelo AR em relação ao seu desempenho, para diversos níveis de SNR de entrada de sinais contaminados com ruído branco. Na Tabela 1 são analisados os resultados obtidos quanto a essa variação utilizando-se do KFO. A janela considerada para esse estudo foi de 60 ms.

Tabela 1 – Valores de SNR e PESQ obtidos através da variação da ordem de q para o algoritmo KFO.

Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,43	2,12	1,76	1,49
$q = 8$	SNR	19,22	15,66	11,41	8,01
	PESQ	3,14	3,20	2,67	2,52
$q = 16$	SNR	19,50	16,15	11,98	8,78
	PESQ	3,17	3,25	2,78	2,65
$q = 24$	SNR	19,71	16,28	12,31	9,03
	PESQ	3,22	3,27	2,83	2,72
$q = 32$	SNR	20,10	16,49	12,73	9,31
	PESQ	3,26	3,27	2,96	2,82
$q = 40$	SNR	20,43	16,70	13,29	9,63
	PESQ	**3,29**	**3,29**	3,02	2,87
$q = 48$	SNR	20,48	16,77	13,45	9,80
	PESQ	3,28	3,28	**3,03**	**2,88**
$q = 56$	SNR	**20,46**	**16,87**	**13,48**	**9,85**
	PESQ	3,28	3,28	3,02	2,85

Fonte: Autoria própria.

Para este caso, observa-se que o aumento do tamanho do modelo está correlacionado com uma melhoria no desempenho em termos de SNR e PESQ de saída. Isso é justificável pois quanto maior o modelo, maior será o efeito de suavização causado pelo atraso de amostras, como sugerido na Seção 5.1.1. Contudo, no que tange à métrica PESQ, os resultados tendem a se estabilizar quando a ordem do modelo se aproxima de $q = 40$.

Um comportamento similar fora encontrado quando considerando um filtro não ideal. Nas tabelas Tabela 2-Tabela 3 foi testado a variação da ordem do modelo para o KFIt e o IMMIt, respectivamente .

[1] https://bit.ly/3V7ltsm

Tabela 2 – Valores de SNR e PESQ obtidos através da variação da ordem de q para o algoritmo KFIt.

Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,43	2,08	1,76	1,65
$q = 8$	SNR	18,87	14,80	10,63	6,81
	PESQ	3,11	2,82	2,51	2,15
$q = 16$	SNR	19,08	15,07	11,09	7,37
	PESQ	3,13	2,84	2,52	2,13
$q = 24$	SNR	19,24	15,26	11,20	7,63
	PESQ	3,14	2,85	2,50	2,10
$q = 32$	SNR	19,54	15,57	11,55	8,05
	PESQ	3,16	**2,86**	2,52	2,14
$q = 40$	SNR	**19,69**	**15,72**	11,86	8,37
	PESQ	**3,18**	**2,86**	**2,55**	**2,17**
$q = 48$	SNR	19,67	15,71	11,90	8,45
	PESQ	3,17	2,85	2,54	2,16
$q = 56$	SNR	19,64	**15,72**	**11,97**	**8,47**
	PESQ	3,16	2,85	2,52	2,14

Fonte: Autoria própria.

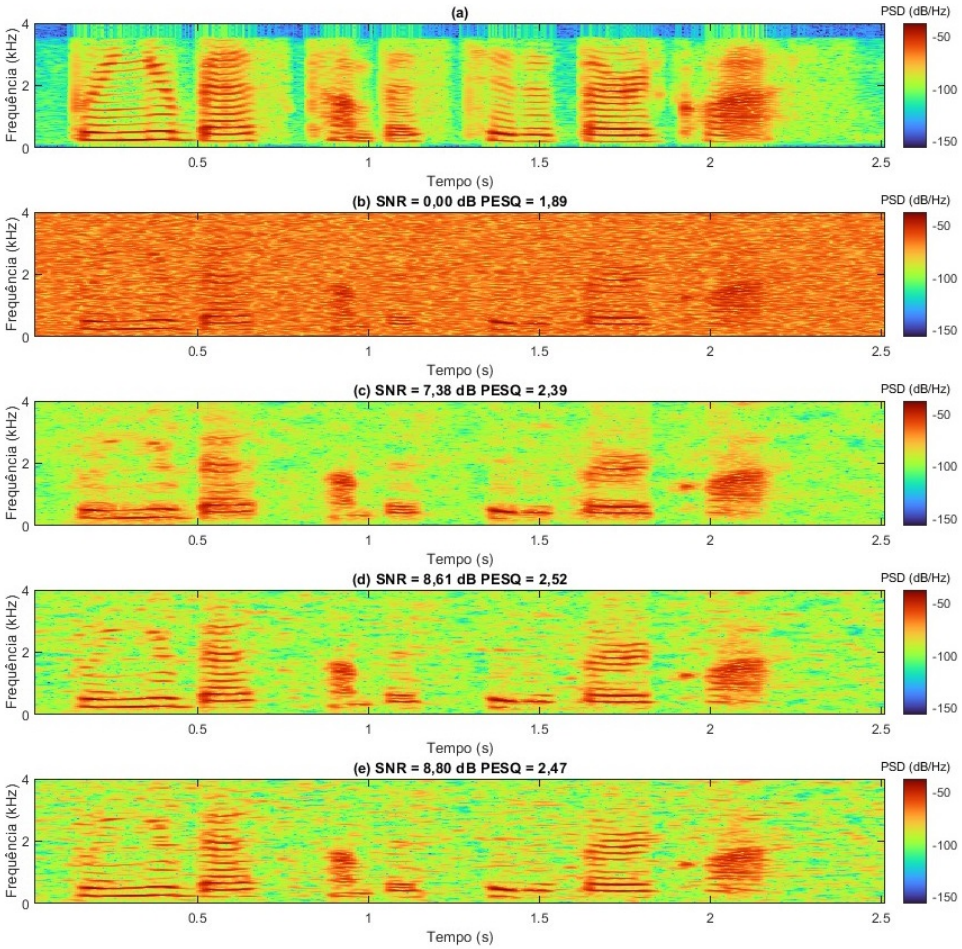
Tabela 3 – Valores de SNR e PESQ obtidos através da variação da ordem de q para o algoritmo IMMIt.

Entrada	SNR	15.00	10.00	5.00	0.00
	PESQ	2.38	2.10	1.88	1.63
$q = 8$	SNR	19.32	15.27	11.30	7.38
	PESQ	3.16	2.90	2.58	2.16
$q = 16$	SNR	19.46	15.65	11.66	7.80
	PESQ	3.22	2.95	2.64	2.18
$q = 24$	SNR	19.51	15.75	11.84	8.00
	PESQ	3.22	2.95	2.63	2.18
$q = 32$	SNR	19.73	15.89	12.15	8.43
	PESQ	3.27	2.94	2.65	2.19
$q = 40$	SNR	20.11	15.98	12.46	8.96
	PESQ	**3.30**	**2.94**	**2.66**	**2.22**
$q = 48$	SNR	20.15	16.00	12.58	9.10
	PESQ	3.27	2.90	2.63	2.21
$q = 56$	SNR	**20.18**	**16.08**	**12.72**	**9.21**
	PESQ	3.23	2.88	2.62	2.19

Fonte: Autoria própria.

Na Figura 39 se compara no espectrograma o áudio limpo com o áudio dos sinais de áudio estimados por IMMIt de diferentes ordens para um sinal de entrada com SNR de 5 dB.

Figura 39 – Espectrogramas de (a) sinal limpo, sinal contaminado com ruido (b), resultados obtidos com IMMIt (c) com $q = 16$ (d) com $q = 40$ e (e) com $q = 80$ e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Nota-se na Figura 39 que o aumento de q permite uma resolução maior das frequências que compõem o sinal fala, entretanto, nos momentos de silêncio causa-se uma grande concentração de ruídos residuais, que podem ser denominado ruído musical visto a grande concentração desses ruídos em frequências específicas. Ou seja, com o aumento da ordem, fica mais perceptível os artefatos sonoros, explicando o decréscimo de PESQ para filtros que operam em ordens maiores.

Além disso, ordens muito grandes aumentam significativamente o tempo de processamento de algoritmo, devido a necessidade de se operar com matrizes maiores.

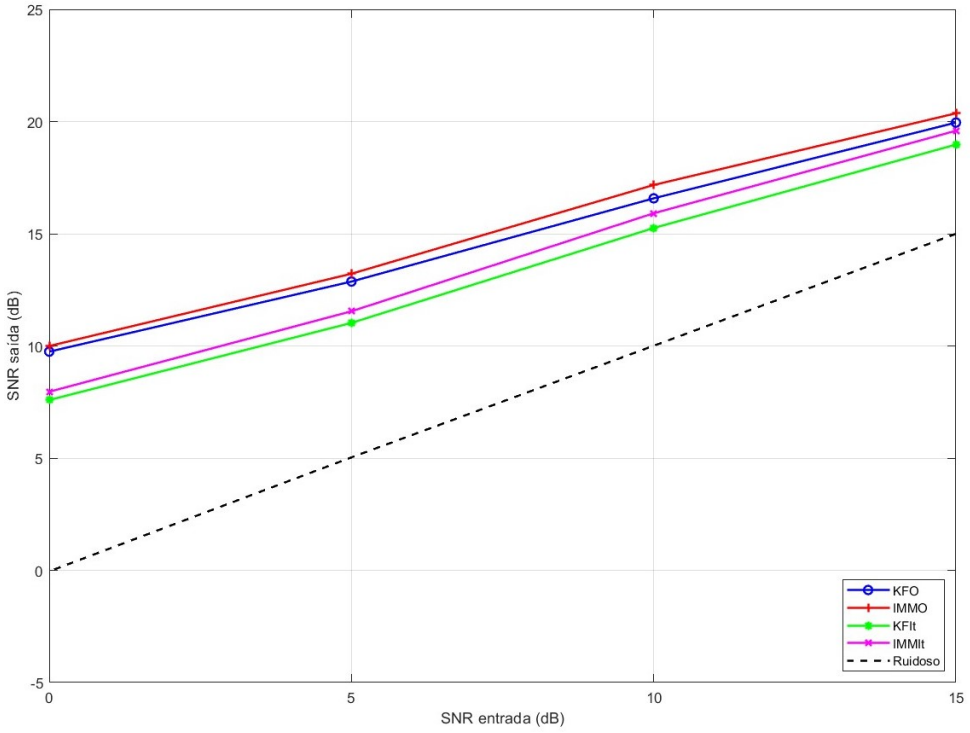
6.2 Comparação entre os Algoritmos

Nessa seção tem-se como objetivo explicitar a performance dos algoritmos propostos nesse trabalho e compara-los entre si. Dessa forma é estudado em suas respectivas seções os algoritmos para ruído branco (KFO, IMMO, KFIt, IMMIt) e para ruído colorido (KFOA, IMMOA, KFItA, IMMItA).

6.2.1 Algoritmos de Ruído Branco

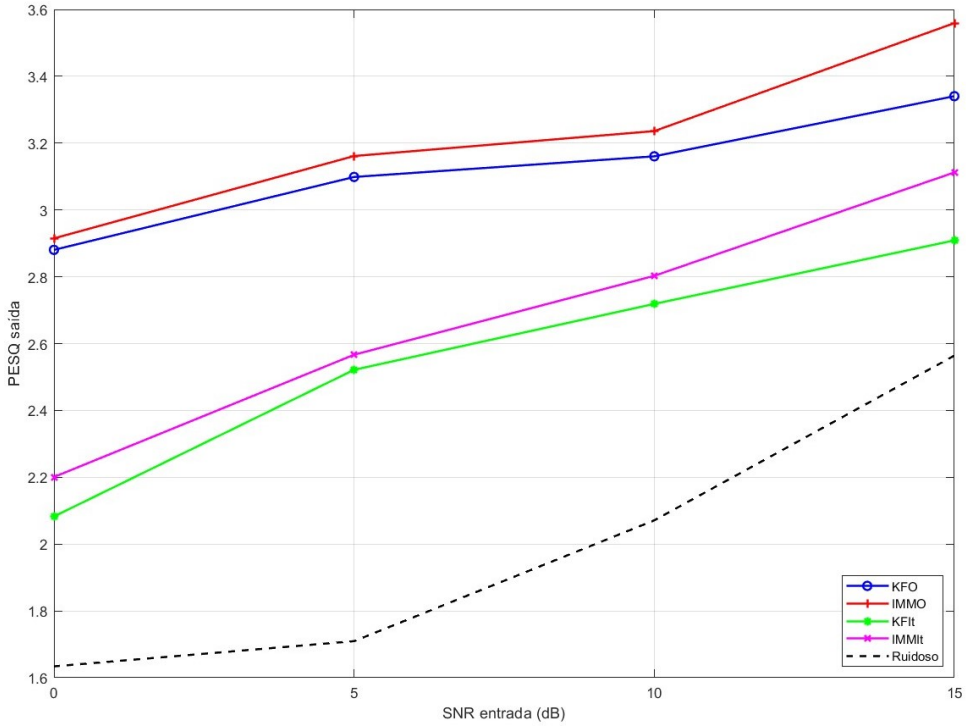
Nas Figuras 40 e 41 e na Tabela 4, comparam-se os resultados obtidos pelos algoritmos estudados nesse trabalho quando aplicado em um sinal contaminado com ruído branco, com $q = 40$ e $T_w = 60$ ms.

Figura 40 – Relação entre SNR de entrada e SNR de saída entre o sinal de entrada (tracejado) e os sinais estimados por KFO (azul), IMMO (vermelho), KFIt (verde) e IMMIt (magenta).



Fonte: Autoria própria.

Figura 41 – Relação entre SNR de entrada e PESQ de saída entre o sinal de entrada (tracejado) e os sinais estimados por KFO (azul), IMMO (vermelho), KFIt (verde) e IMMIt (magenta).



Fonte: Autoria própria.

Tabela 4 – Resultados referentes a performance dos algoritmos propostos de para ruído branco.

Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,56	2,07	1,71	1,63
KFO	SNR	19,95	16,58	12,87	9,75
	PESQ	3,34	3,16	3,10	2,88
IMMO	SNR	**20,37**	**17,17**	**13,22**	**10,00**
	PESQ	**3,56**	**3,24**	**3,16**	**2,91**
KFIt	SNR	18,97	15,26	11,03	7,59
	PESQ	2,91	2,72	2,52	2,08
IMMIt	SNR	**19,59**	**15,91**	**11,54**	**7,96**
	PESQ	**3,11**	**2,80**	**2,57**	**2,20**

Fonte: Autoria própria.

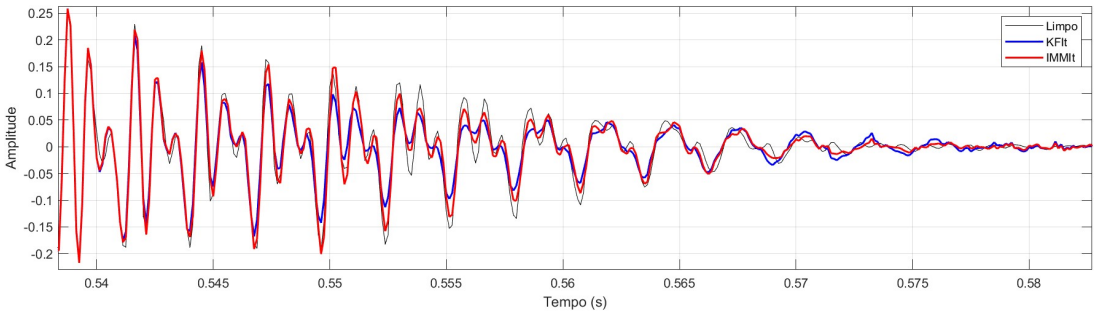
Nos resultados obtidos é perceptível que a qualidade do sinal estimado degrada de acordo com a intensidade do ruído. Além disso, nota-se que os filtros não ideais possuem comportamento similar ao KFO com SNRs mais altas, e que se distanciam desses ao aumentar o nível do SNR. Isso se explica com o impacto do maior enviesamento do modelo, causado pelo aumento do ruído.

Nota-se aqui que, idealmente, o IMMO supera tanto em PESQ quanto em SNR, a performance do KFO, principalmente quando a intensidade do ruído é menor, comprovando-se, assim, que a técnica é realmente um aprimoramento em relação ao KFO tradicional. Ao se aumentar a intensidade de ruído, entretanto, ambos tendem a ter resultados semelhantes, o que

pode ser causado pela maior distância do modelo ideal do sinal ruidoso, que provocar maior dificuldade na precisão das transições de modelo do IMM.

Da mesma forma, nota-se nesses resultados que o IMMIt se sobressai aos resultados obtidos com o KFIt, tanto em PESQ quanto em SNR, para todos os níveis de SNR de entrada estudados, deixando claro que a técnica é valida também em situações nas quais o modelo não é ideal. Essa melhora expressa-se significantemente na transição entre comportamentos, o que se é evidenciado na Figura 42.

Figura 42 – Comparação no domínio do tempo entre trecho de sinal de limpo (preto) e sinais estimados por KFIt (azul) e IMMIt (vermelho).

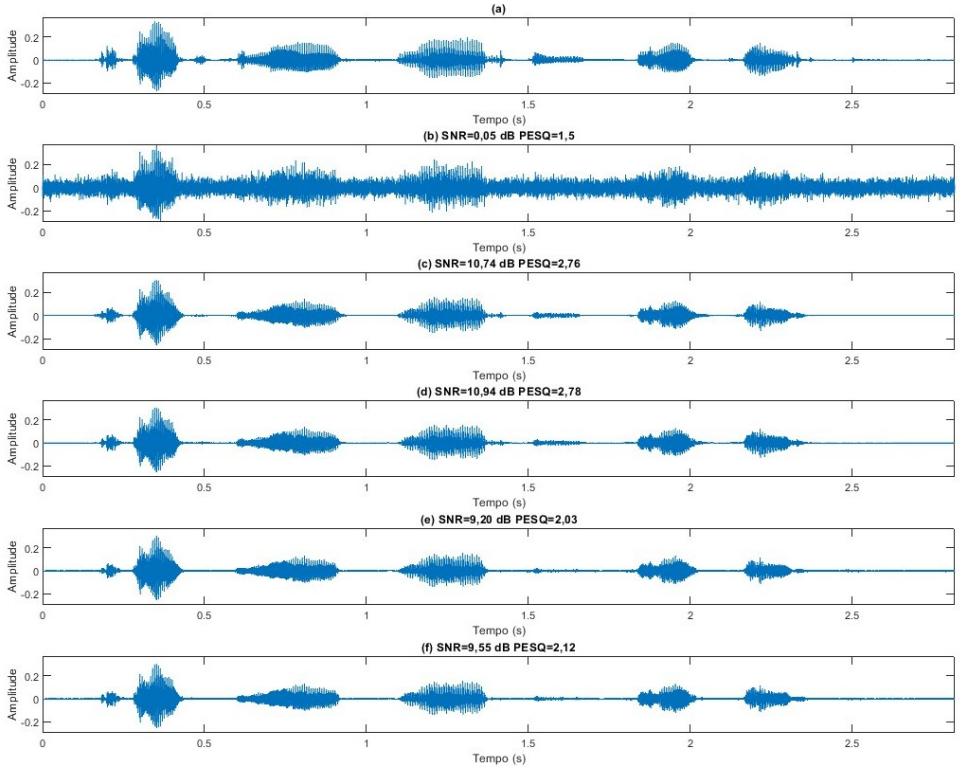


Fonte: Autoria própria.

Portanto, como sugerido na Seção 5.4, apesar da sobreposição das janelas mitigar os erros nas bordas dos sinais estimados, esse erro ainda implica numa perda no sinal pós OLA. Como o IMMIt é capaz de estimar as janelas por inteiro, resultando em erros menores, que se destacam quando existe uma mudança notável do comportamento do sinal.

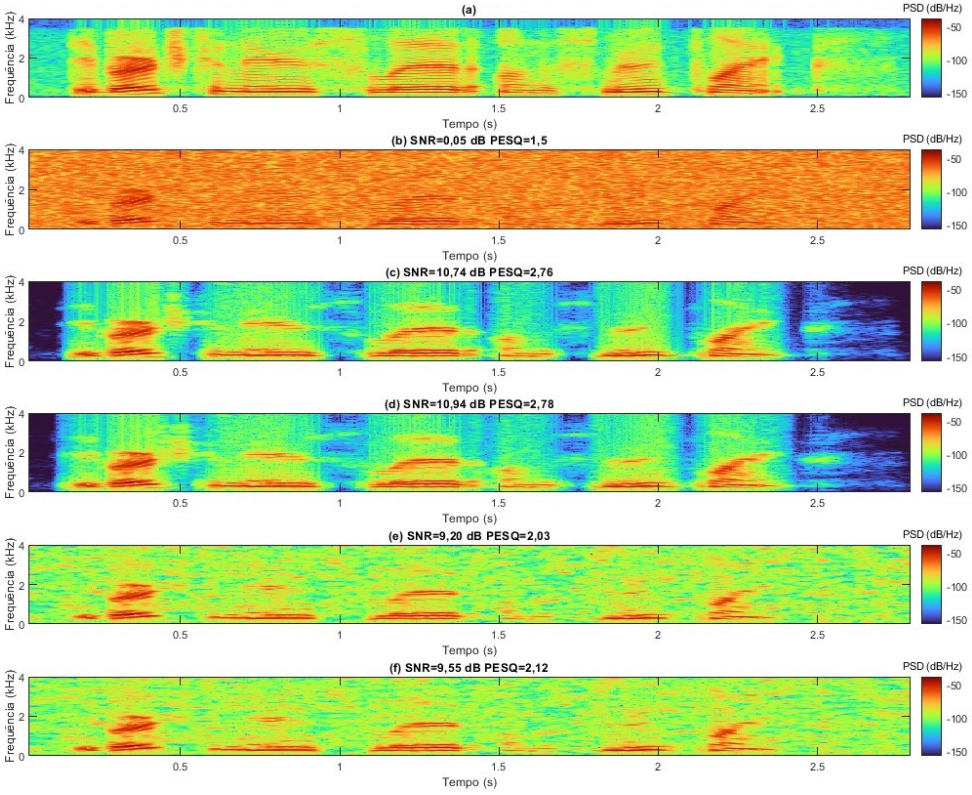
A melhora, entretanto, não é significativa, já que não se demonstra nem no domínio do tempo nem no domínio da frequência diferenças notáveis entre os dois áudios estimados. Isso é melhor evidenciado pelas Figuras 43-44, nas quais comparam os resultados obtidos no tempo e no espectrograma do sinal. Dentre as diferenças notáveis está que a intensidade do ruído musical é amenizado no IMMIt em comparação ao KFIt. Tal vantagem pode ser entendida como uma maior dificuldade do filtro IMM em estimar o sinal com um modelo único modelo enviesado através da combinação de múltiplos modelos.

Figura 43 – (a) Sinal de limpo, (b) sinal ruidoso, e os sinais estimados por (c) KFO, (d) IMMO, (e) KFIt e (f) IMMIt e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

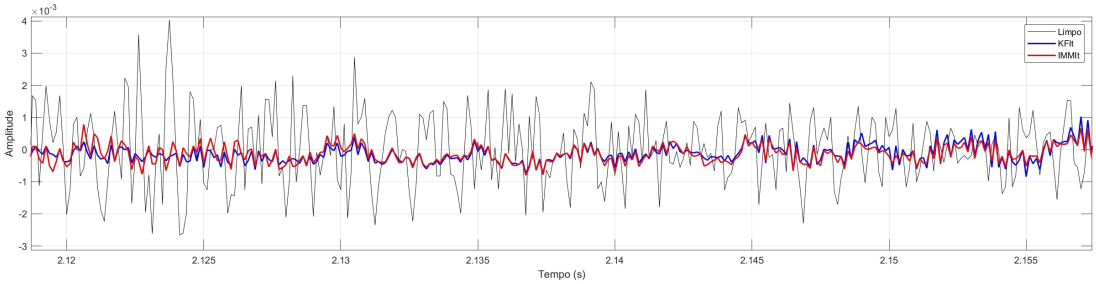
Figura 44 – Espectrograma do (a) sinal de limpo, (b) do sinal ruidoso e dos sinais estimados por (c) KFO, (d) IMMO, (e) KFIt e (f) IMMIt e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Uma característica comum aos algoritmos é que esses possuem dificuldade de estimar sinais de fala não-vocálica como mostra a figura Figura 45. Atribui-se a isso o fato de que esse tipo de sinal possuir um comportamento não periódico que se assemelha ao ruído aditivo e de baixa amplitude. Dessa forma é possível entender que esse é um tipo de informação que é difícil de se estimar através de modelos obtidos por filtragem estocástica.

Figura 45 – Comparação no domínio do tempo entre o sinal de limpo de som não vocálico e sinais estimados por KFIt e IMMIt.



Fonte: Autoria própria.

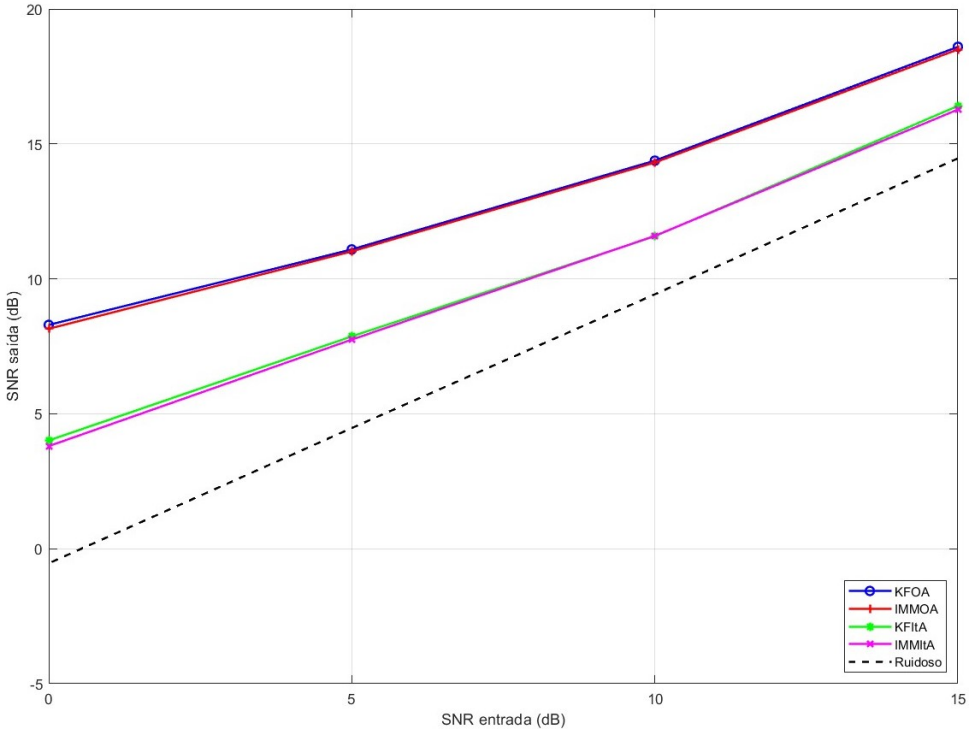
Conclui-se dessa seção que, tanto o KFIt quanto o IMMIt, possuem validade em relação a remoção de ruídos gaussianos aditivos de sinais de fala, e que, idealmente, o algoritmo de IMM como proposto nesse trabalho é capaz de superar os resultados de um KFIt. É importante notar que o processamento de um IMMIt é expressivamente maior que de um simples KFIt. Dessa forma, para aplicações práticas, o KFIt é uma algoritmo mais atrativo. Entretanto, mesmo com a pequena melhora nos índices de desempenho, ainda é possível assumir que a aplicação de um IMM para redução de ruído branco em sinais de fala é um método válido.

6.2.2 Algoritmos para Ruído Colorido

Aqui adota-se $q = 24$, $p = 24$ e $T_w = 60$ ms para todos os algoritmos. A redução da ordem de q justifica-se como uma maneira de evitar uma maior complexidade computacional do experimento com a redução dos cálculos matriciais dos KF. Os coeficientes b para os filtros KFItA e IMMItA são obtidos através de um trecho de 1000 amostras do sinal isolado de r_k.

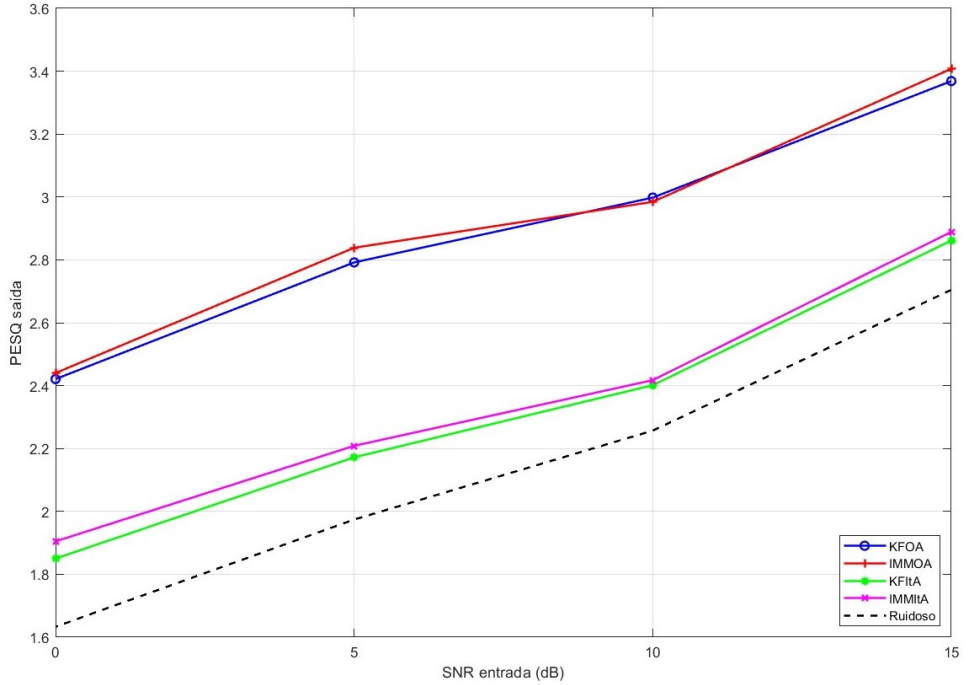
Nas Figuras 46-51 e Tabela 5-7 compara-se a performance dos algoritmos aumentados para diferentes tipos de ruído colorido.

Figura 46 – Relação entre SNR de entrada e SNR de saída entre o sinal de entrada contaminado por ruído de multidão e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



Fonte: Autoria própria.

Figura 47 – Relação entre SNR de entrada e PESQ de saída entre o sinal de entrada contaminado por ruído de multidão e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



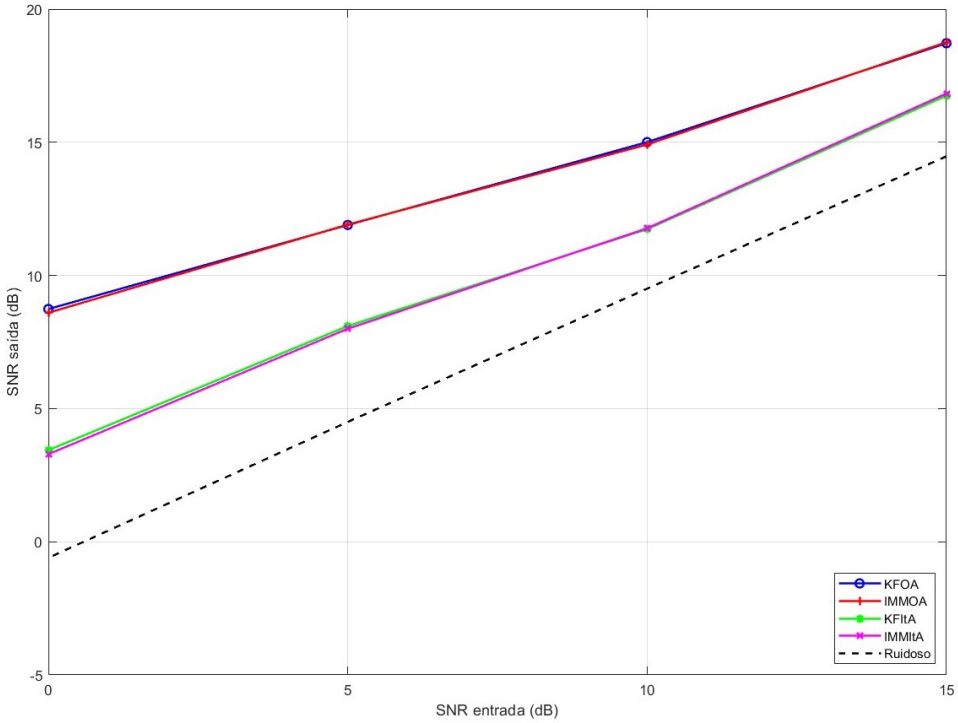
Fonte: Autoria própria.

Tabela 5 – Relação entre SNR e PESQ entre o sinal de entrada contaminado por ruído de multidão e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.

Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,59	2,14	1,90	1,61
KFOA	SNR	19,48	14,66	11,06	**8,08**
	PESQ	3,20	3,02	**2,75**	**2,44**
IMMOA	SNR	**19,53**	**14,73**	**11,08**	8,01
	PESQ	**3,29**	**3,04**	2,71	2,41
KFItA	SNR	**16,57**	12,04	**9,12**	**3,73**
	PESQ	2,79	2,39	2,22	1,69
IMMItA	SNR	16,49	**12,17**	9,08	3,52
	PESQ	**2,82**	**2,43**	**2,29**	**1,72**

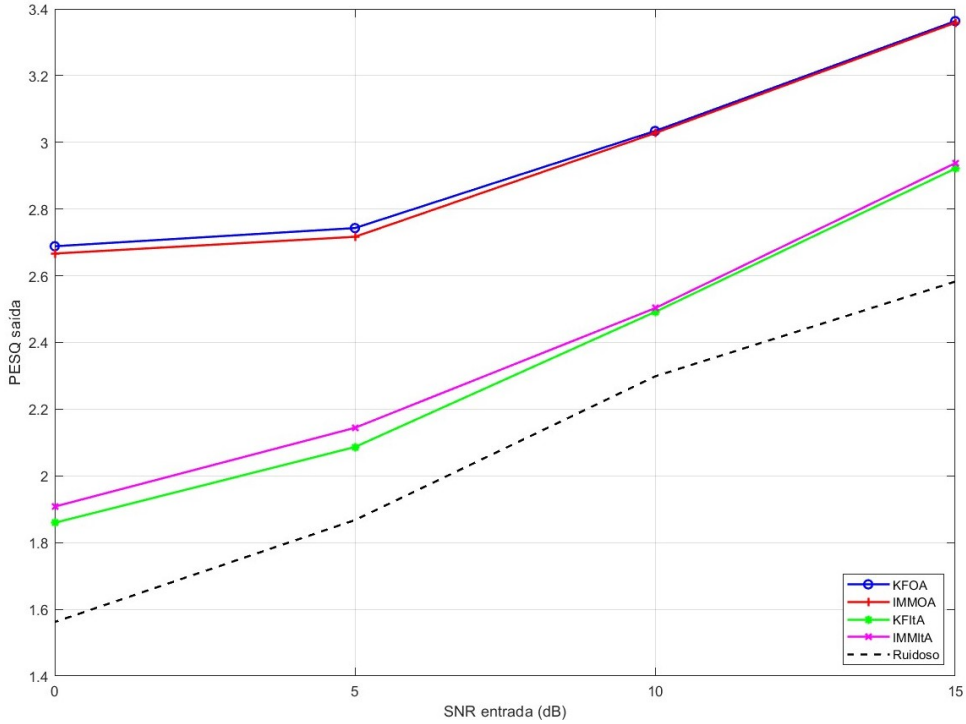
Fonte: Autoria própria.

Figura 48 – Relação entre SNR e PESQ entre o sinal de entrada contaminado por ruído de rua e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



Fonte: Autoria própria.

Figura 49 – Relação entre SNR de entrada e PESQ de saída entre o sinal de entrada contaminado por ruído de rua e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



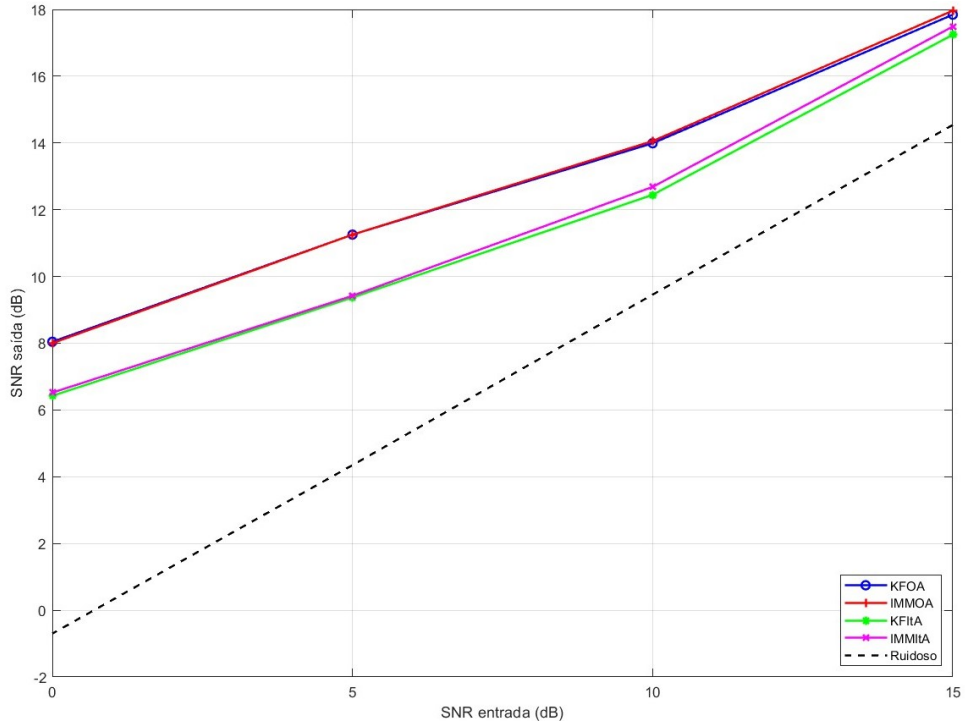
Fonte: Autoria própria.

Tabela 6 – Relação entre valores de PESQ e SNR de sinal de entrada contaminado por ruído de rua e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.

Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,59	2,14	1,90	1,61
KFOA	SNR	19,48	14,66	11,06	**8,08**
	PESQ	3,20	3,02	**2,75**	**2,44**
IMMOA	SNR	**19,53**	**14,73**	**11,08**	8,01
	PESQ	**3,29**	**3,04**	2,71	2,41
KFItA	SNR	**16,57**	12,04	**9,12**	**3,73**
	PESQ	2,79	2,39	2,22	1,69
IMMItA	SNR	16,49	**12,17**	9,08	3,52
	PESQ	**2,82**	**2,43**	**2,29**	**1,72**

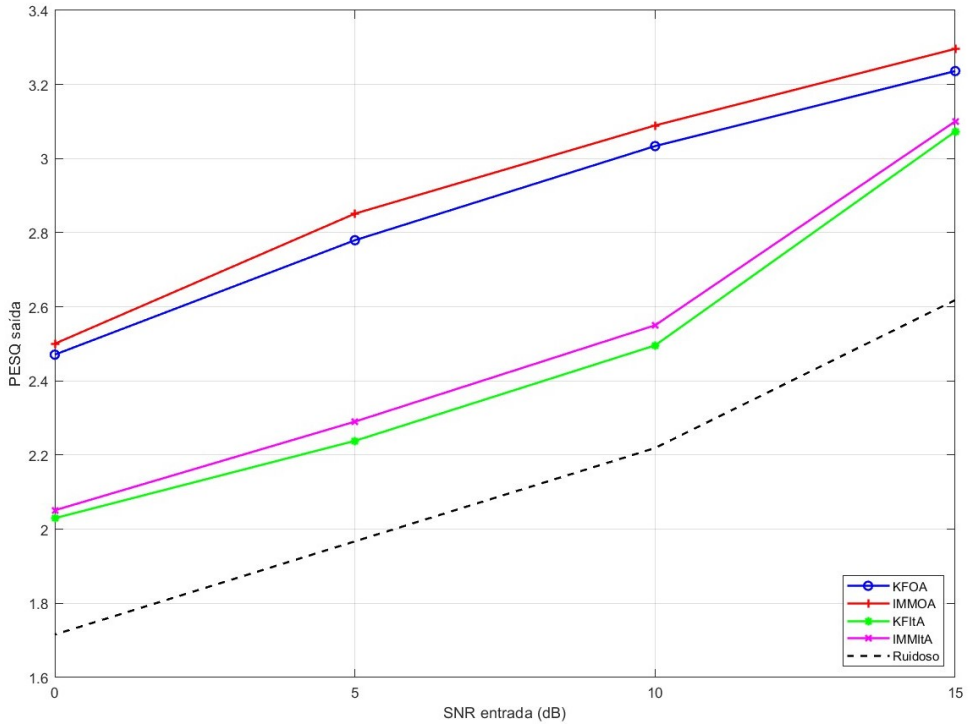
Fonte: Autoria própria.

Figura 50 – Relação entre SNR de entrada e SNR de saída entre o sinal de entrada contaminado por ruído de inteiror de carro e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



Fonte: Autoria própria.

Figura 51 – Relação entre SNR de entrada e PESQ de saída entre o sinal de entrada contaminado por ruído de carro e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.



Fonte: Autoria própria.

Tabela 7 – Relação entre valores de PESQ e SNR de sinal de entrada contaminado por ruído de carro e os sinais estimados por KFOA, IMMOA, KFItA e IMMItA.

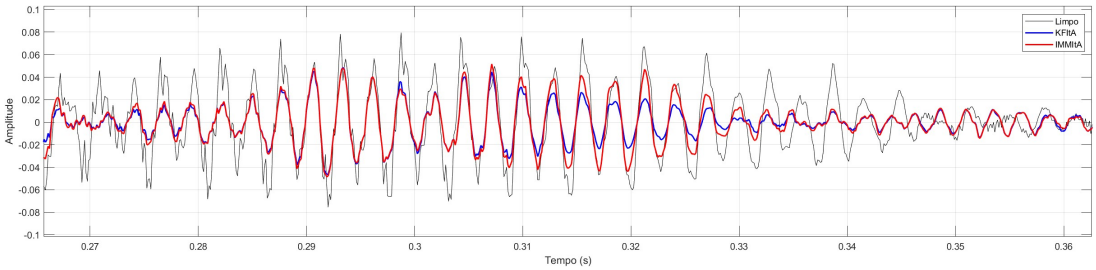
Entrada	SNR	15,00	10,00	5,00	0,00
	PESQ	2,62	2,22	1,97	1,72
KFOA	SNR	17,84	13,99	**11,25**	8,04
	PESQ	3,24	3,03	2,78	2,47
IMMOA	SNR	**17,96**	**14,06**	**11,25**	**8,00**
	PESQ	**3,30**	**3,09**	**2,85**	**2,50**
KFItA	SNR	17,24	12,45	9,37	6,42
	PESQ	3,07	2,50	2,24	2,03
IMMItA	SNR	**17,49**	**12,68**	**9,42**	**6,52**
	PESQ	**3,10**	**2,55**	**2,29**	**2,05**

Fonte: Autoria própria.

Nota-se que, no caso dos ruídos coloridos, há uma discrepância mais evidente entre os algoritmos ideais e os não ideais, o que pode ser atribuído ao fato do KFA e IMMA estimarem $r_{k,n}$ janela a janela, enquanto os não ideais adotam uma abordagem simplificada e generalista do comportamento do ruído.

Observa-se nesses resultados que os algoritmos baseados em IMM não demonstraram melhora significativa em relação aos baseados em KF em termos de SNR, como no caso de ruído branco. Entretanto, na maioria dos tipos de ruído, houve melhorias em termos de PESQ. Essa melhora pode ser atribuída ao fato desse filtro melhor se adaptar à alteração de comportamentos, assim como descrito na Seção 6.2.1.

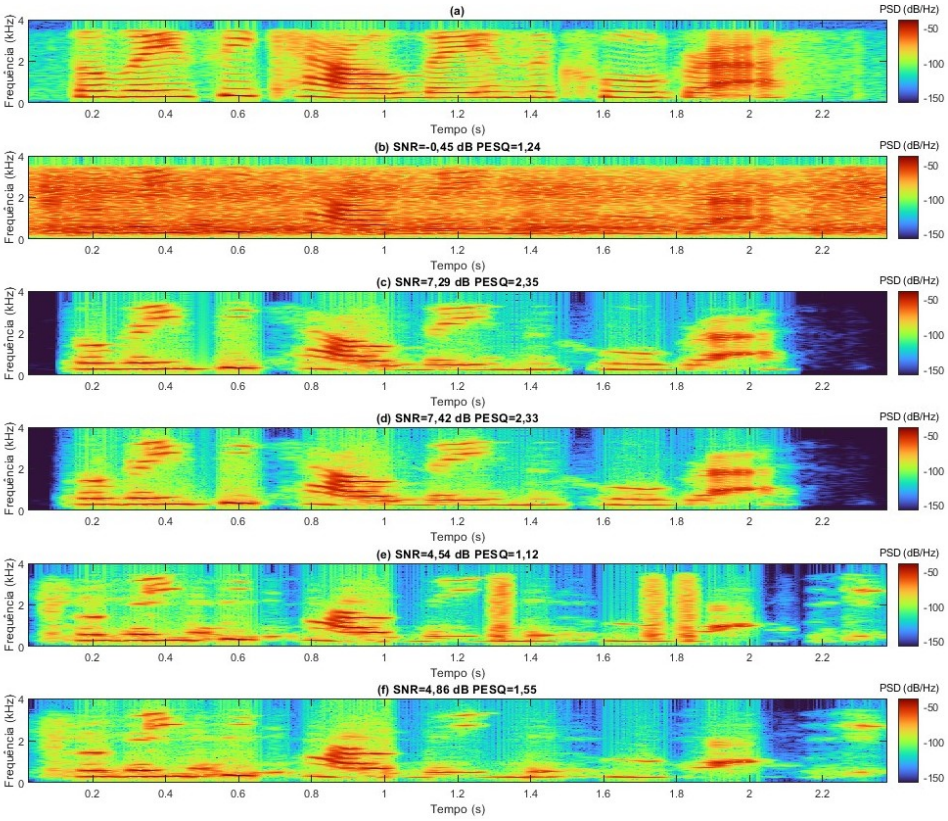
Figura 52 – Comparação no domínio do tempo entre o de trecho de sinal de limpo e sinais estimados por KFItA e IMMItA de um sinal contaminado por ruído colorido de carro.



Fonte: Autoria própria.

Outra característica que evidência a melhora trazida pelo filtro IMMItA é sua maior flexibilidade em lidar com modelos enviesados por ruído, geralmente causados por um um ruído de impulso presente no sinal de ruído, que acaba comprometendo A_n e $Q_{s,n}$. No caso do KFItA acabam comprometendo o sinal estimado por janelas inteiras, enquanto o IMMItA é capaz de evitar a predominância de um modelo enviesado. Isso é evidenciado na Figura 53, na qual nota-se que para três janelas de tempo do filtro KFItA possuem uma significante quantidade de ruído residual, enquanto no IMMItA isso não ocorre.

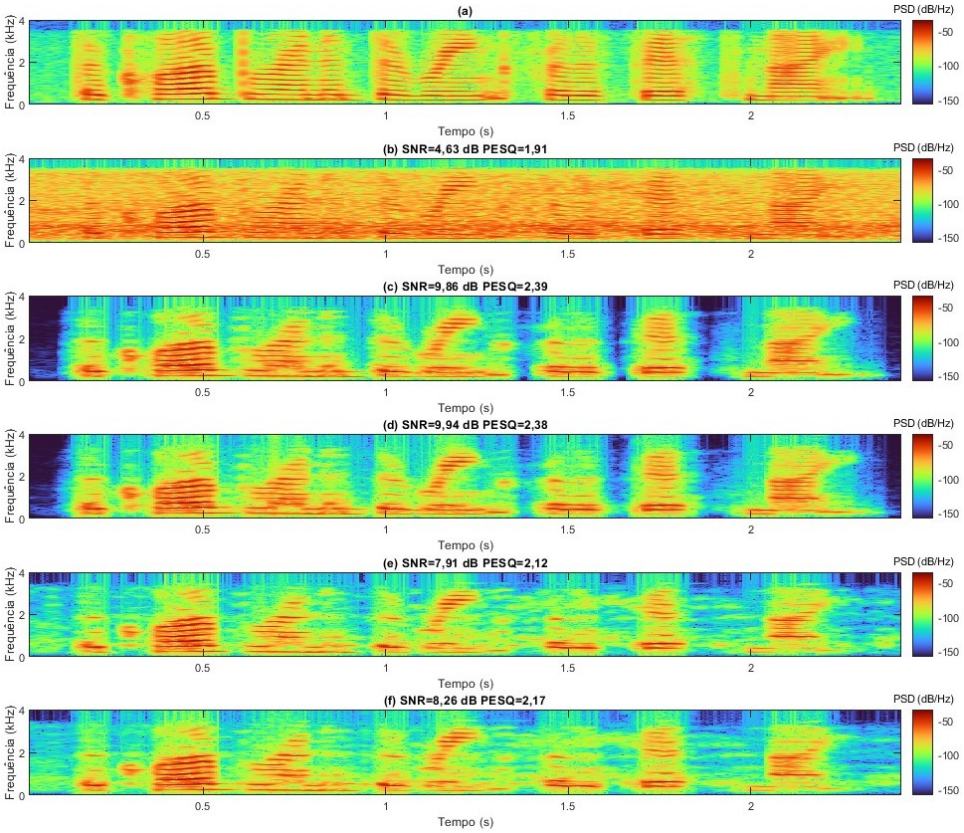
Figura 53 – Espectrograma do (a) sinal de limpo, (b) do sinal contaminado por ruido de trem e dos sinais estimados por (c) KFOA, (d) IMMOA, (e) KFItA e (f) IMMItA e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Na Figura 54 analisa-se o espectrograma dos ruídos estimados por cada um dos filtros aumentados, dado um sinal de ruído de carro. Nota-se que os algoritmos são capazes de remover do espectro grande parte das características do ruído com poucos resíduos no sinal estimado. Isso valida que as técnicas propostas na Seção 5.3 são, de fato, eficientes quando o ruído é, como nesse caso, majoritariamente estacionário.

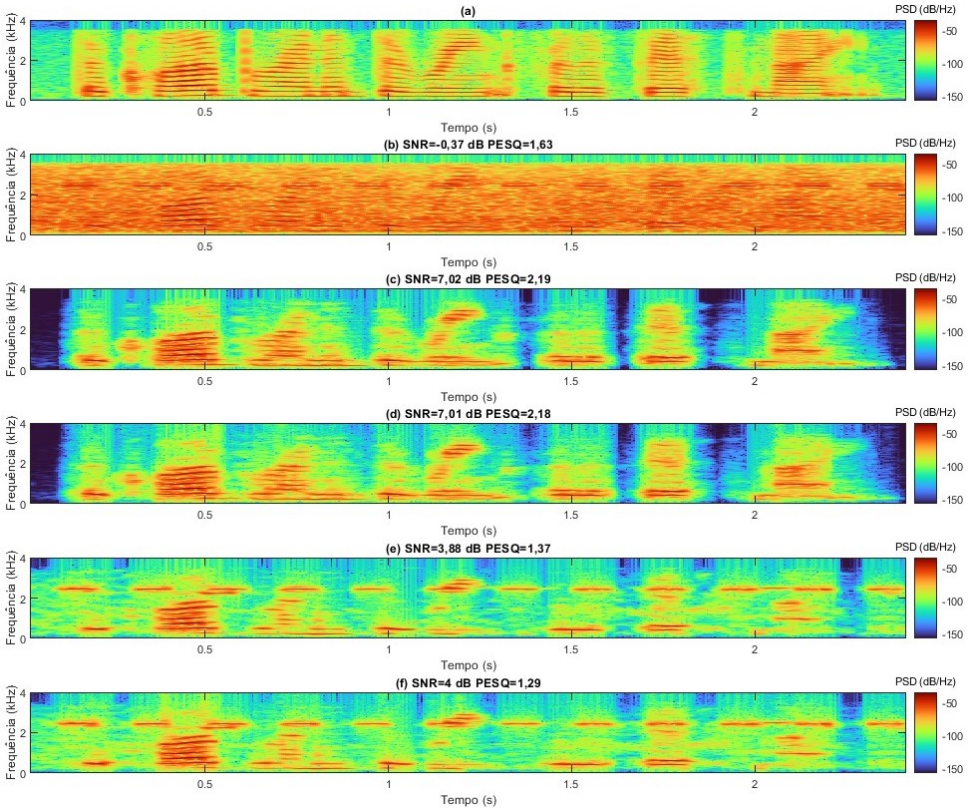
Figura 54 – Espectrograma do (a) sinal de limpo, (b) do sinal contaminado por ruido de carro e dos sinais estimados por (c) KFOA, (d) IMMOA, (e) KFItA e (f) IMMItA e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Entretanto, visto que os algoritmos não ideais limitam-se a uma única estimativa do comportamento do ruído, entende-se que, para casos em que o ruído possua alguma variabilidade em seu comportamento que não seja previsto pelo modelo de r_k obtido, tende-se a obter resultados piores. Isso é observável no espectro do ruído da Figura 55, no qual se observado o comportamento determinístico de uma sirene que ocorre no decorrer do áudio.

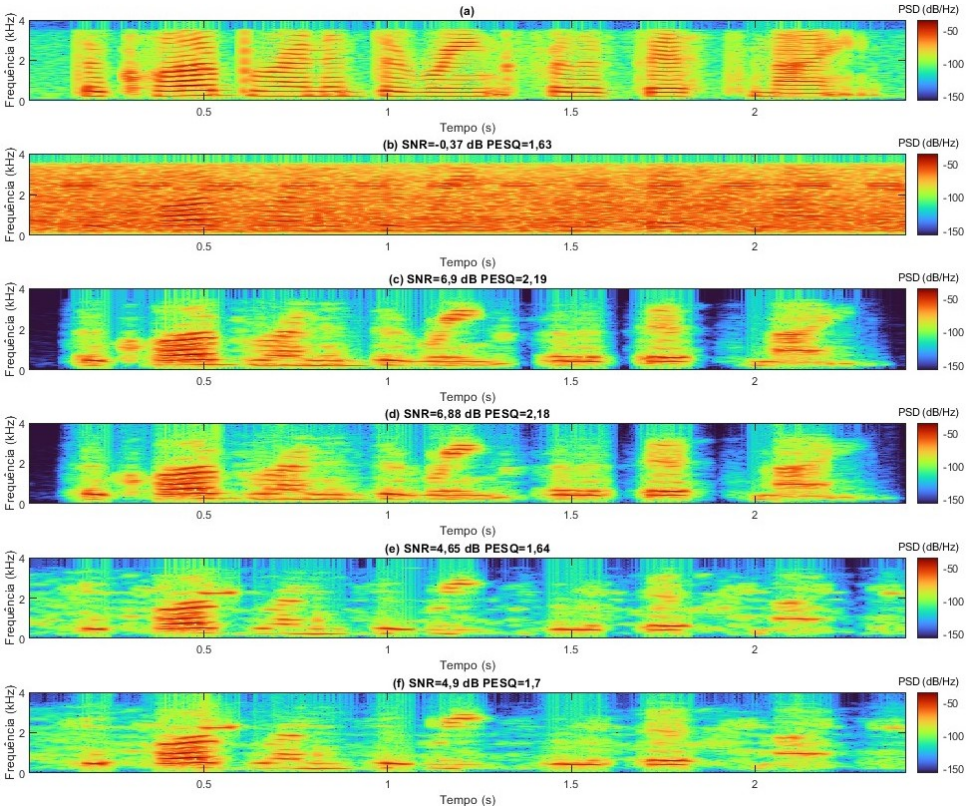
Figura 55 – Espectrograma do (a) sinal de limpo, (b) do sinal ruidoso contaminado por ruído de rua e dos sinais estimados por (c) KFOA, (d) IMMOA, (e) KFItA e (f) IMMItA e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Ao expandir a janela de tempo de r_k para um número de amostras que englobe também o comportamento da sirene, resultando em uma redução do sinal residual da sirene no sinal estimado.

Figura 56 – Espectrograma do (a) sinal de limpo, do (b) sinal ruidoso contaminado por ruído de rua e dos sinais estimados por (c) KFO, (d) IMMO, (e) KFIt e (f) IMMIt e seus respectivos valores de SNR e PESQ considerando duas mil amostras para obtenção do modelo do ruído.

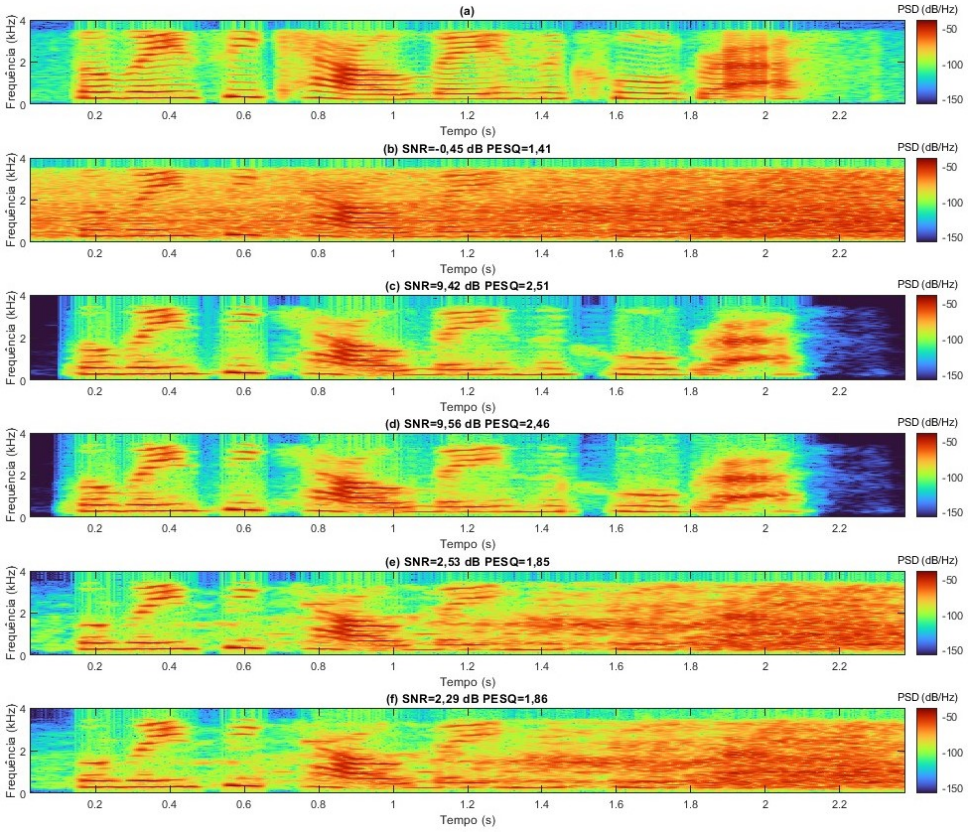


Fonte: Autoria própria.

Dessa forma, na maneira da qual fora desenvolvido a performance dos algoritmos está limitada a remoção de ruído estabelecida pela escolha de r_k – para estimar o modelo do ruído. Interessante também notar que os algoritmos são capazes de lidar até mesmo com ruído que não completamente estacionários.

Entretanto, se ocorrer no sinal estimado um ruído de elevada intensidade que não tenha sido antecipadamente previsto, a qualidade do sinal resultante pode ser significativamente comprometida. Esta situação torna-se evidente nos espectrogramas apresentados na Figura 57. No início do áudio, é possível observar uma boa estimação por parte dos algoritmos não ideais. Contudo, quando surge um ruído não estacionário causado pela passagem de um carro, a eficiência da estimação desses filtros é comprometida.

Figura 57 – Espectrograma do (a) sinal de limpo, do (b) sinal ruidoso contaminado por ruído de rua e dos sinais estimados por (c) KFO , (d) IMMO, (e) KFIt e (f) IMMIt e seus respectivos valores de SNR e PESQ.



Fonte: Autoria própria.

Reitera-se a partir da análise da performance dos algoritmos não ideais é que a abordagem adotada para remover os ruídos coloridos é eficiente em reduzir ruídos coloridos em sinais de fala, conquanto que esses sejam estacionários. Entretanto, para a abordagem adotada nesse trabalho para os filtros não ideias, essa se limita a quando o ruído colorido é previsível através de um único modelo AR. Variações no comportamento do ruído dificultam em muito a estimação comprometem a performance do KFItA e IMMItA, fazendo com que áudio seja até pior que o áudio ruidoso em termos de PESQ. Assim, para que a aplicação desses algoritmos seja de maior utilidade para casos reais faz-se necessárias abordagens mais sofisticadas para identificação do modelo de ruído quando esse possuir comportamento não-estacionário.

7 CONCLUSÕES

Conclui-se que este trabalho conseguiu atingir seus objetivos propostos, que incluíam compreender a aplicação de filtros estocásticos na estimação de sinais de fala a partir do desenvolvimento de um algoritmo baseado no filtro IMM.

Inicialmente, foi estabelecida a necessidade de modelar a fala de forma que diferentes tipos de séries temporais que variam seu comportamento ao longo do tempo possam ser adequadamente representados. A técnica de LPC foi fundamental nesse processo, permitindo a obtenção de modelos AR que descrevem o comportamento da energia do sinal de fala no domínio da frequência. Entretanto, devido à natureza variável do sinal de fala, diferentes modelos LPC devem ser aplicados em diferentes instantes de tempo. O LPC também permite que seja dimensionado um filtro de *whitening*, o que se prova útil nesse trabalho na redução do enviesamento dos ruídos coloridos na estimação do modelo do sinal nos algoritmos desenvolvidos.

Dessa forma, é proposto a implementação de dois algoritmos principais: O KFIt e o IMMIt. O KFIt consiste na aplicação do KF para segmentos de janelas, com o modelo e sua variância sendo estimados pelo LPC daquele período de tempo, para então refazer o processo numa segunda interação na qual se utiliza o sinal estimado pela primeira para se obter um modelo aprimorado para o filtro estocástico. Para melhor performance do algoritmo algumas modificações: a obtenção do sinal através de amostras atrasadas, permitindo uma suavização temporal devido às iterações de K_k; a utilização de sobreposição de janelas, que através de um OLA, que torna capaz a transição suave da troca de modelos que ocorre entre as janelas, além de permitir a utilização de janelas maiores que ajudam no desenviesamento do modelo estimado sem comprometer a capacidade do algoritmo de acompanhar as transições do comportamento da fala; a aplicação de uma janela de Hamming para auxiliar na estimação do LPC, permitindo que a reposta de frequência do AR obtido fora mais consistente com o observado. Tais modificações se demonstraram-se eficientes no quesito de melhorar a performance dos algoritmos em relação a PESQ e a SNR.

O IMMIt consiste numa adequação do que fora desenvolvido no KFIt de maneira a implementar no mesmo filtro diferentes modelos de janelas adjacentes. Essa solução tem como proposta flexibilizar o filtro de maneira que na mesma janela se consiga transicionar entre os modelos. Apesar de efeitos mínimos quanto a qualidade do áudio obtida, nesse trabalho, o algoritmo se provou como eficaz em sua proposta, e pode servir como base para utilização de diferentes modelos e filtros.

Os filtros aumentados consistem numa adaptação dos filtros desenvolvidos de maneira a conseguir reduzir os efeitos de um ruído colorido de comportamento estimável. Para tal, utiliza-se dos mesmos princípios da obtenção de modelo por meio de LPC para obter o AR do ruído, que é então tratado como um estado independente do modelo do sinal que se soma ao sinal de fala para obtenção da saída. Dessa forma, verificou-se que tanto o KFItA quanto o IMMItA são métodos eficientes na estimação de um sinal de fala contaminado com ruído colorido desde que

esse possua um comportamento estacionário, dos quais o IMMItA demonstrou-se, em grande parte, com melhores resultados que o KFItA. Comportamentos de ruídos muito variáveis demonstraram a limitação da aplicação desse tipo dos algoritmo propostos, e demandam abordagens mais sofisticadas de modelagem da fala e ruído.

Portanto, verificou-se a viabilidade da aplicação de filtros estocásticos auxiliados de identificação de sistemas para reduzir o ruído em sinais de fala, e propõem-se aqui um novo método eficaz através do IMMIt que pode servir de base para desenvolvimentos de filtros mais sofisticados. Assim, conclui-se que esse trabalho fora capaz de cumprir seus objetivos propostos como o de entender a aplicação de filtros estocásticos na estimação de sinais de fala e no desenvolvimento um algoritmo com base no IMM que aprimora, apesar de minimamente, os resultados de algoritmos baseados em KF.

7.1 Sugestões para trabalhos futuros

Visto limitações de recursos para esse trabalho, não foram realizados testes utilizando-se de ouvintes para avaliar a performance dos algoritmos qualitativamente. Dessa forma, seria interessante para estudos futuros aplicar métodos de avaliação perceptual da qualidade dos sinais estimados.

O desenvolvido por meio do filtro IMM pode servir como base para uma diversidade de aplicações. Nesse trabalho fora explorado a habilidade desse de transicionar entre diferentes modelos referentes a janelas adjacentes à janela do sinal sendo estimado. Entretanto, é possível pensar em outras abordagens que podem se aproveitar da utilização de múltiplos filtros.

Dentre essas seria a aplicação do que é proposto em trabalhos como So *et al.* (2017) e (GEORGE *et al.*, 2018) na utilização de métricas de sensibilidade e robustez (SAHA; GHOSH; GOSWAMI, 2013) para alteração de K_k de maneira a para aproxima-lo do comportamento de ganho ideal – como visto na seção 5.2. Nesses trabalhos, se é estudado a aplicação de KF através dessas métricas com o objetivo de atenuar o ruído residual e eliminar a necessidade de iterações. Em Roy e Paliwal (2021) se identifica que a métrica de robustez é mais apropriada para momentos silenciosos enquanto a modificação com a métrica de sensibilidade gera melhores resultados em amostras que contém fala, e se utiliza de identificadores de atividade de fala para alternar entre os dois filtros. Seria possível estudar implementação a mesma lógica através do filtro IMM para realizar o chaveamento entre os filtros modificados a fim de identificar a modificação mais apropriada para cada amostra do sinal.

Seria também interessante um estudo quanto a melhoria por modelos estimados do algoritmo através da aplicação de LPCs modificados, como o LPC ponderado estabilizado (SWLPC, do inglês *Stabilised Weighted LPC*) (MAGI *et al.*, 2009) a fim de conseguir modelos AR mais robustos em relação ao ruído, talvez até eliminando a necessidade a iteratividade dos filtros. Outra proposta seria a aplicação de modelos não-lineares por meio de filtros estocásticos não-lineares como o KF Estendido ou o KF *Uscented* (WAN; MERWE, 2000).

Assim como descrito na Seção 6.2.2, seria interessante metodologias para obtenção dos parâmetros B_n e Q_r janela a janela a fim de melhor estimar as possíveis variações de ruídos reais, algoritmos tais quais de redes neurais profundas como Yu, Zhu e Champagne (2020).

Por fim, seria interessante realizar estudos quanto à otimização e à aplicabilidade do filtro em tempo real, ou seja, que o sinal seja estimado ao mesmo tempo que ocorre a sua captação, dessa forma, ampliando sua usabilidade em casos práticos.

REFERÊNCIAS

BAR-SHALOM, Y.; LI, X. R.; KIRUBARAJAN, T. **Estimation with applications to tracking and navigation: theory algorithms and software**. [*S.l.*]: John Wiley & Sons, 2004.

BEERENDS, J. G. *et al.* Perceptual evaluation of speech quality (pesq) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 50, n. 10, p. 765–778, 2002.

BLOM, H.; BAR-SHALOM, Y. The interacting multiple model algorithm for systems with markovian switching coefficients. **IEEE Transactions on Automatic Control**, v. 33, n. 8, p. 780–783, 1988.

BOLL, S. Suppression of acoustic noise in speech using spectral subtraction. **IEEE Transactions on acoustics, speech, and signal processing**, IEEE, v. 27, n. 2, p. 113–120, 1979.

BOSWORTH, B. T. *et al.* Estimating signal-to-noise ratio (snr). **IEEE Journal of Oceanic Engineering**, v. 33, n. 4, p. 414–418, 2008.

CANAZZA, S.; POLI, G. D.; MIAN, G. A. Restoration of audio documents by means of extended kalman filter. **IEEE transactions on audio, speech, and language processing**, IEEE, v. 18, n. 6, p. 1107–1115, 2009.

CASTIGLIONI, P. Levinson-durbin algorithm. **Encyclopedia of Biostatistics**, John Wiley and Sons Ltd. London, UK, v. 4, 2005.

CERNA, M.; HARVEY, A. F. **The fundamentals of FFT-based signal analysis and measurement**. [*S.l.*], 2000.

DAJER, M. E. **Análise de sinais de voz por padrões visuais de dinâmica vocal**. 2010. Tese (Doutorado) — Universidade de São Paulo, 2010.

DENNIS, J.; TRAN, H. D.; LI, H. Spectrogram image feature for sound event classification in mismatched conditions. **IEEE signal processing letters**, IEEE, v. 18, n. 2, p. 130–133, 2010.

DEVORE, J. L.; CORDEIRO, M. T. A. **Probabilidade e estatística: para engenharia e ciências**. [*S.l.*]: Cengage Learning Edições Ltda., 2014.

FRENCL, V. B. *et al.* **Técnicas de filtragem utilizando processos com saltos markovianos aplicados ao rastreamento de alvos móveis**. 2010. Tese (Doutorado) — Dissertação (Mestrado)-Universidade Estadual de Campinas (UNICAMP . . . , 2010.

GANNOT, S. Speech processing utilizing the kalman filter. **IEEE Instrumentation & Measurement Magazine**, IEEE, v. 15, n. 3, p. 10–14, 2012.

GEORGE, A. E. *et al.* Robustness metric-based tuning of the augmented kalman filter for the enhancement of speech corrupted with coloured noise. **Speech Communication**, Elsevier, v. 105, p. 62–76, 2018.

GIBSON, J.; KOO, B.; GRAY, S. Filtering of colored noise for speech enhancement and coding. **IEEE Transactions on Signal Processing**, v. 39, n. 8, p. 1732–1742, 1989.

GOH, Y. H.; RAVEENDRAN, P.; GOH, Y. L. Robust speech recognition system using bidirectional kalman filter. **IET Signal Processing**, IET, v. 9, n. 6, p. 491–497, 2015.

GRIFFIN, D. W.; LIM, J. S. Multiband excitation vocoder. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 36, n. 8, p. 1223–1235, 1988.

HEINZEL, G.; RÜDIGER, A.; SCHILLING, R. Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows. 2002.

HINES, W. *et al.* Probabilidade e estatística na engenharia. 4ª edição. **Editora LTC, Rio de Janeiro**, 2006.

HU, Y.; LOIZOU, P. C. Subjective comparison and evaluation of speech enhancement algorithms. **Speech communication**, Elsevier, v. 49, n. 7-8, p. 588–601, 2007.

JAISWAL, R. K.; YEDURI, S. R.; CENKERAMADDI, L. R. Single-channel speech enhancement using implicit wiener filter for high-quality speech communication. **International Journal of Speech Technology**, Springer, v. 25, n. 3, p. 745–758, 2022.

JOHNSON, D. H. Signal-to-noise ratio. **Scholarpedia**, v. 1, n. 12, p. 2088, 2006.

KALLAS, M. *et al.* Kernel autoregressive models using yule–walker equations. **Signal Processing**, Elsevier, v. 93, n. 11, p. 3053–3061, 2013.

KALMAN, R. E. A new approach to linear filtering and prediction problems. 1960.

LI, Q. *et al.* Kalman filter and its application. *In*: **2015 8th International Conference on Intelligent Networks and Intelligent Systems (ICINIS)**. [*S.l.*: *s.n.*], 2015. p. 74–77.

MAGI, C. *et al.* Stabilised weighted linear prediction. **Speech Communication**, Elsevier, v. 51, n. 5, p. 401–411, 2009.

MAKHOUL, J. Linear prediction: A tutorial review. **Proceedings of the IEEE**, v. 63, n. 4, p. 561–580, 1975.

MALIK, S.; BENESTY, J. Variationally diagonalized multichannel state-space frequency-domain adaptive filtering for acoustic echo cancellation. *In*: **2013 IEEE International Conference on Acoustics, Speech and Signal Processing**. [*S.l.*: *s.n.*], 2013. p. 595–599.

MCCREE, A.; BARNWELL, T. A mixed excitation lpc vocoder model for low bit rate speech coding. **IEEE Transactions on Speech and Audio Processing**, v. 3, n. 4, p. 242–250, 1995.

NYQUIST, H. Certain topics in telegraph transmission theory. **Transactions of the American Institute of Electrical Engineers**, IEEE, v. 47, n. 2, p. 617–644, 1928.

O'SHAUGHNESSY, D. Linear predictive coding. **IEEE potentials**, IEEE, v. 7, n. 1, p. 29–32, 1988.

PALIWAL, K.; BASU, A. A speech enhancement method based on kalman filtering. *In*: **ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing**. [*S.l.*: *s.n.*], 1987. v. 12, p. 177–180.

POPESCU, D.; ZELJKOVIC, I. Kalman filtering of colored noise for speech enhancement. *In*: **Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)**. [*S.l.*: *s.n.*], 1998. v. 2, p. 997–1000 vol.2.

RABINER, L. R.; SCHAFER, R. W. **Introduction to digital speech processing**. [*S.l.*]: Now Publishers Inc, 2007. v. 1.

RAJAMANI, K.; LAI, Y.-S.; FURROW, C. An efficient algorithm for sample rate conversion from cd to dat. **IEEE Signal Processing Letters**, v. 7, n. 10, p. 288–290, 2000.

REHR, R.; GERKMANN, T. Snr-based features and diverse training data for robust dnn-based speech enhancement. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, IEEE, v. 29, p. 1937–1949, 2021.

RIGOLL, G. A new algorithm for estimation of formant trajectories directly from the speech signal based on an extended kalman-filter. *In*: **ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing**. [*S.l.*: *s.n.*], 1986. v. 11, p. 1229–1232.

RIX, A. W. *et al.* Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. *In*: IEEE. **2001 IEEE international conference on acoustics, speech, and signal processing. Proceedings (Cat. No. 01CH37221)**. [*S.l.*], 2001. v. 2, p. 749–752.

ROSA, M. d. O. **Laringe digital**. 2002. Tese (Doutorado) — Universidade de São Paulo, 2002.

ROY, S. K.; NICOLSON, A.; PALIWAL, K. K. Deep learning with augmented kalman filter for single-channel speech enhancement. *In*: **2020 IEEE International Symposium on Circuits and Systems (ISCAS)**. [*S.l.*: *s.n.*], 2020. p. 1–5.

ROY, S. K.; PALIWAL, K. K. Causal convolutional neural network-based kalman filter for speech enhancement. *In*: **2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)**. [*S.l.*: *s.n.*], 2020. p. 1–6.

ROY, S. K.; PALIWAL, K. K. Robustness and sensitivity tuning of the kalman filter for speech enhancement. **Signals**, Multidisciplinary Digital Publishing Institute, v. 2, n. 3, p. 434–455, 2021.

SAHA, M.; GHOSH, R.; GOSWAMI, B. Robustness and sensitivity metrics for tuning the extended kalman filter. **IEEE Transactions on Instrumentation and Measurement**, IEEE, v. 63, n. 4, p. 964–971, 2013.

SALOR, Ö.; DEMIREKLER, M.; ORGUNER, U. Kalman filter approach for pitch determination of speech signals. **St. Petersburg**, p. 4, 2006.

SCHWARTZ, B.; GANNOT, S.; HABETS, E. A. Online speech dereverberation using kalman filter and em algorithm. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, IEEE, v. 23, n. 2, p. 394–406, 2014.

SNELLING, M. The audiobook market and its adaptation to cultural changes. **Publishing Research Quarterly**, Springer, v. 37, n. 4, p. 642–656, 2021.

SO, S. *et al.* Kalman filter with sensitivity tuning for improved noise reduction in speech. **Circuits, Systems, and Signal Processing**, Springer, v. 36, p. 1476–1492, 2017.

SO, S.; PALIWAL, K. K. A long state vector kalman filter for speech enhancement. *In*: **Proc. Interspeech 2008**. [*S.l.*: *s.n.*], 2008. p. 391–394.

SO, S.; PALIWAL, K. K. Suppressing the influence of additive noise on the kalman gain for low residual noise speech enhancement. **Speech Communication**, Elsevier, v. 53, n. 3, p. 355–378, 2011.

TAMMEN, M. *et al.* Dnn-based speech presence probability estimation for multi-frame single-microphone speech enhancement. *In*: IEEE. **ICASSP 2020-2020 IEEE International conference on acoustics, speech and signal processing (ICASSP)**. [*S.l.*], 2020. p. 191–195.

TAN, L.; JIANG, J. Chapter 2 - signal sampling and quantization. *In*: TAN, L.; JIANG, J. (Ed.). **Digital Signal Processing (Third Edition)**. Third edition. Academic Press, 2019. p. 13–58.

ISBN 978-0-12-815071-9. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780128150719000026.

THOMPSON, J. D.; WELDON, J. Podcasting. *In*: **Content Production for Digital Media**. [*S.l.*]: Springer, 2022. p. 105–120.

TIERNEY, J. A study of lpc analysis of speech in additive noise. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 28, n. 4, p. 389–397, 1980.

UPADHYAY, N.; KARMAKAR, A. Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study. **Procedia Computer Science**, Elsevier, v. 54, p. 574–584, 2015.

VASEGHI, S. V. **Advanced digital signal processing and noise reduction**. [*S.l.*]: John Wiley & Sons, 2008.

WAN, E. A.; MERWE, R. V. D. The unscented kalman filter for nonlinear estimation. *In*: IEEE. **Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)**. [*S.l.*], 2000. p. 153–158.

WAN, E. A.; NELSON, A. T. Removal of noise from speech using the dual ekf algorithm. *In*: IEEE. **Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)**. [*S.l.*], 1998. v. 1, p. 381–384.

WELCH, G.; BISHOP, G. *et al.* An introduction to the kalman filter. Chapel Hill, NC, USA, 1995.

XIA, Y.; WEI, Q. An effective kalman filtering method for enhancing speech in the presence of colored noise. *In*: IEEE. **2016 International Conference on Audio, Language and Image Processing (ICALIP)**. [*S.l.*], 2016. p. 469–474.

YAN, X. *et al.* An iterative graph spectral subtraction method for speech enhancement. **Speech Communication**, Elsevier, v. 123, p. 35–42, 2020.

YU, H.; ZHU, W.-P.; CHAMPAGNE, B. Speech enhancement using a dnn-augmented colored-noise kalman filter. **Speech Communication**, Elsevier, v. 125, p. 142–151, 2020.

ZHANG, Z. Mechanics of human voice production and control. **The journal of the acoustical society of america**, Acoustical Society of America, v. 140, n. 4, p. 2614–2635, 2016.

ZHENG, B. *et al.* Analysis of noise reduction techniques in speech recognition. *In*: **2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)**. [*S.l.*: *s.n.*], 2020. v. 1, p. 928–933.

ZÖLZER, U. **Digital audio signal processing**. [*S.l.*]: John Wiley & Sons, 2008. 1–10 p.