UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Rivael Pigatto

# Desempenho de máscaras tempo-frequência para redução de ruído em aparelhos auditivos

Florianópolis
2023

Rivael Pigatto

# Desempenho de máscaras tempo-frequência para redução de ruído em aparelhos auditivos

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de mestre.
Orientador: Prof. Márcio Holsbach Costa, Dr.
Coorientador: Prof. Bruno Catarino Bispo, Dr.

Florianópolis
2023

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Pigatto, Rivael
Desempenho de máscaras tempo-frequência para redução de ruído em aparelhos auditivos / Rivael Pigatto ; orientador, Márcio Holsbach Costa, coorientador, Bruno Catarino Bispo, 2023.
78 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2023.

Inclui referências.

1. Engenharia Elétrica. 2. Máscaras tempo-frequência. 3. Redução de ruído . 4. Qualidade perceptual de fala. 5. Inteligibilidade de fala. I. Costa, Márcio Holsbach . II. Bispo, Bruno Catarino . III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. IV. Título.

Rivael Pigatto

**Desempenho de máscaras tempo-frequência para redução de ruído em aparelhos auditivos**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof. Márcio Holsbach Costa, Dr.
Instituição Universidade Federal de Santa Catarina-UFSC

Prof. Sérgio Jose Melo de Almeida, Dr.
Instituição Universidade Católica de Pelotas - UCPEL

Prof. Hans Helmut Zürn, Dr.
Instituição Universidade Federal de Santa Catarina-UFSC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre.

ICP-Edu
Documento assinado digitalmente
**Telles Brunelli Lazzarin**
Data: 08/11/2023 11:42:23-0300
CPF: ***.400.719-**
Verifique as assinaturas em https://v.ufsc.br

Coordenação do Programa de
Pós-Graduação

ICP-Edu
Documento assinado digitalmente
**Marcio Holsbach Costa**
Data: 05/11/2023 10:02:27-0300
CPF: ***.170.680-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 2023.

Este trabalho é dedicado aos meus queridos pais.

## AGRADECIMENTOS

Agradeço aos meus pais Valdir Pigatto e Ildaci M. Basso Pigatto, que foram meus primeiros professores da vida, repassando seus conhecimentos e conceitos sobre ética, caráter, humildade, ludicidade e negócios e que sempre se empenharam para me ensinar a lidar com as adversidades da vida, inspirando-me a seguir em busca por um futuro melhor, e ao meu irmão Jean Carlos Pigatto por incentivo e apoio.

Agradeço às psicólogas Marta Elisa Bringhenti, e Camila Carneiro Stedille que me auxiliaram a passar por um período "cinza", ajudando no meu autoconhecimento como pessoa e aperfeiçoando minha compreensão sobre sentimentos e emoções.

Quero agradecer também aos meus queridos senseis de Artes Marciais, principalmente a Gustavo Sagás Magalhães, que possibilitou minha iniciação no judô, seguido do sensei João Leonardo Akihito Mitsure, que proporcionou muita sabedoria, ensinamentos e apoio. Também deixo registrado meu agradecimento a Willian Pascoutto, sensei de Jiu-jutsu que me ensinou técnicas de ênfase em controle e luta no chão, (ne waza) e (katame waza), que gentilmente me chama de "Chapecó bicho brabo", e a todos os colegas e amigos que conheci no tatame.

Agradeço também ao meu orientador Márcio Holsbach Costa e coorientador Bruno C. Bispo pela disponibilidade durante as etapas do curso, Ao meu querido amigo Kalil Janvion que abdicou horas de seu lazer para me auxiliar no fechamento da parte escrita da dissertação, que parecia ser interminável.

Deixo um ato gentil de agradecimento também aos amigos e colegas do LPDs, UFSC, IFSC e SENAI os quais de algum modo, me ajudaram, com auxilio técnico, inspiração e flexibilização de tempo. E aos demais meu MUITO OBRIGADO a todos.

*"A essência do Judô não está na vitória nem na revelação do talento mas, sim, no esforço e na habilidade desprendidas para consegui-las."*
*Jigoro Kano*

## RESUMO

A deficiência auditiva é um fator limitante no desenvolvimento pessoal e profissional de um grande número de pessoas. O uso de aparelhos auditivos possibilita uma compensação adequada dessa limitação em um grande número de situações. A maior reclamação de usuários de aparelhos auditivos refere-se a dificuldades de compreensão e baixa qualidade da fala na presença de ruído acústico do ambiente. Uma estratégia muito utilizada para redução de ruído é o uso da técnica de máscaras tempo-frequência, caracterizadas por curvas de atenuação definidas em função da razão sinal-ruído do sinal captado. O objetivo deste estudo é investigar o limite de desempenho dessa estratégia de redução de ruído em diversos cenários acústicos, caracterizados por diferentes tipos de ruído e níveis de contaminação. Para tanto, são utilizados critérios objetivos de inteligibilidade e qualidade sonora perceptual. Extensivas simulações foram realizadas para diferentes sinais de fala na língua portuguesa e inglesa, com locutores masculinos e femininos. Diferentes tipos de ruído foram utilizados para a contaminação da fala utilizando razões sinal-ruído entre -10 e 20 dB. Foram avaliadas as máscaras de Wiener e Paramétrica Conformável, além de uma máscara conformável especialmente projetada para possibilitar curvas de atenuação simétricas e assimétricas. O teste ANOVA ($\rho < 0{,}05$) foi utilizado para diferenciar as distribuições dos resultados baseados nas métricas W-PESQ e NCM. Os resultados indicaram que o desempenho obtido pela máscara Paramétrica Conformável com parâmetros fixos é próximo ao limite de desempenho possível da técnica de máscaras tempo-frequência, independentemente dos diversos cenários acústicos analisados.



**Palavras-chave**: Aparelhos auditivos. Máscaras tempo-frequência. Redução de ruído. Inteligibilidade de fala. Qualidade perceptual de fala.

## ABSTRACT

Hearing impairment is a limiting factor in the personal and professional development of a significant number of individuals. The use of hearing aids allows for proper compensation of this limitation in a wide range of situations. The most common complaint from hearing aid users relates to difficulties in comprehension and low speech quality in the presence of environmental acoustic noise. A frequently employed strategy for noise reduction is the use of time-frequency masking techniques, characterized by attenuation curves based on the signal-to-noise ratio of the captured signal. The aim of this study is to investigate the performance limit of this noise reduction strategy in various acoustic scenarios, characterized by different types of noise and contamination levels. To achieve this, objective intelligibility and perceptual sound quality criteria are utilized. Extensive simulations were conducted for different speech signals in both Portuguese and English languages, with male and female speakers. Different types of noise were employed to contaminate the speech signals using signal-to-noise ratios ranging from -10 to 20 dB. The Wiener and Conformable Parametric masks were evaluated, along with a custom-designed conformable mask that enables symmetric and asymmetric attenuation curves. The ANOVA test ($\rho < 0.05$) was utilized to differentiate the result distributions based on the W-PESQ and NCM metrics. The results indicated that the performance achieved by the Conformable Parametric mask with fixed parameters is close to the potential performance limit of the time-frequency masking technique, regardless of the various analyzed acoustic scenarios.



**Keywords**: Hearing aids. Time-frequency masks. Noise reduction. Speech intelligibility. Perceptual speech quality.

## LISTA DE FIGURAS

## LISTA DE QUADROS

## LISTA DE TABELAS

**SUMÁRIO**

# 1 INTRODUÇÃO

A audição é um dos cinco sentidos fundamentais do corpo humano, permitindo-nos captar ondas sonoras do ambiente ao nosso redor e, assim, ouvir e compreender o mundo. Além de nos permitir comunicar e interagir socialmente, a audição desempenha um papel crucial na sobrevivência, alertando-nos para possíveis riscos e perigos iminentes. Portanto, perdas auditivas podem acarretar um impacto significativo na qualidade de vida das pessoas e a busca por meios para superar essa limitação é de grande interesse na sociedade atual (DE OLIVEIRA, 2017).

## 1.1 JUSTIFICATIVA

De acordo com a Organização Mundial da Saúde (OMS), em 2020, 466 milhões de pessoas eram acometidas de perda auditiva, sendo 432 milhões de adultos e 34 milhões de crianças, resultando em mais de 5% da população mundial. Em 2050 esse montante poderá alcançar mais de 900 milhões de pessoas (WORLD HEALTH ORGANIZATION, 2020). Segundo a Pesquisa Nacional de Saúde (PNS), no ano de 2019, 1,1% da população brasileira declarou possuir alguma deficiência auditiva, totalizando 2,3 milhões de pessoas (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2019).

A deficiência auditiva dificulta a comunicação, a troca de informações e o aprendizado, podendo gerar isolamento social, dificuldades profissionais e até mesmo risco à segurança pessoal, reduzindo a qualidade de vida do indivíduo (ZAHNERT, 2011). A compensação da limitação auditiva depende do tipo e do grau da perda auditiva, podendo ser realizada de diversas formas, entre elas, com a utilização de aparelho auditivo ou implante coclear (ZAHNERT, 2011).

Os aparelhos auditivos são dispositivos eletrônicos, originalmente analógicos, cuja principal função é compensar a deficiência auditiva (MARTIN; HEUTE; ANTWEILER, 2008). Atualmente, os aparelhos auditivos são digitais, e sua estrutura é composta basicamente por: microfone, conversor analógico-digital, processador de sinais, conversor digital-analógico, amplificador de som e um alto-falante (ARSINTE; LUPU; SUMALAN, 2017). As versões digitais têm como principal vantagem a alta flexibilidade, podendo possuir vários subsistemas como, por exemplo: redução de ruído, enfatização da fala, cancelamento de realimentação e redução do efeito de oclusão (MARTIN; HEUTE; ANTWEILER, 2008; BORGES; COSTA, 2016).

Apesar de sua eficácia na compensação de perdas leves a moderadas, aparelhos auditivos não são capazes de restaurar a capacidade de ouvir em situações de perdas severas a profundas, nesses casos, o uso de implante coclear (IC) é uma possível solução (TEFILI et al., 2013; WOUTERS et al., 2013). O IC é um dispositivo eletrônico que tem como objetivo estimular eletricamente as fibras nervosas ao longo

da cóclea através de eletrodos, proporcionando sensações capazes de ser interpretadas pelo cérebro como audição (TEFILI et al., 2013). Apesar de apresentarem algumas limitações na resolução dos sons transmitidos ao seus usuários, a inteligibilidade da fala se aproxima à de ouvintes normais na ausência de ruído (WILSON; DORMAN, 2007).

Apesar de aumentar a audibilidade dos sons, a capacidade dos aparelhos auditivos modernos em melhorar a inteligibilidade da fala é bastante limitada em ambientes ruidosos (MOORE, 2007; DILLON, 2001). Segundo Kates et al. (2018), a inteligibilidade média em aparelhos auditivos é de aproximadamente 90% na ausência de ruído, mas o desempenho cai para cerca de 65% e 10% para razões sinal ruído (SNRs) de 10 dB e 0 dB, respectivamente.

Similarmente, para ICs também ocorre processo semelhante de diminuição de inteligibilidade na presença de ruído (FU; NOGAKI, 2005). Segundo Hast et al. (2015), a inteligibilidade média de sentenças na ausência de ruído é de aproximadamente 75%, mas para SNR de 5 dB diminui para uma faixa entre 19% e 35%, dependendo do tipo de ruído e da idade do usuário. De acordo com Bergeron e Hotton (2016), a inteligibilidade média em ICs é de aproximadamente 80%, 50%, 40% e 20% na ausência de ruído e SNRs de 10 dB, 5 dB e 0 dB, respectivamente.

A partir do exposto, a incorporação de técnicas de redução de ruído em ICs e aparelhos auditivos se faz necessária, para melhorar a inteligibilidade e a qualidade da fala em ambientes ruidosos. Uma das principais estratégias de redução de ruído em aparelhos auditivos é o uso de máscaras tempo-frequência (CHIEA; COSTA; BARRAULT, 2019b).

## 1.2 MOTIVAÇÃO

Uma das principais estratégias de redução de ruído em aparelhos auditivos e implantes cocleares é a classe de máscaras tempo-frequência (CHIEA; COSTA; BARRAULT, 2019b). Essa técnica é caracterizada por um processo de decomposição do sinal original (realizada através de um banco de filtros ou uso da transformada de Fourier) que resulta em diferentes unidades (associadas a determinadas faixas de frequência) para janelas de tempo consecutivas. A cada unidade tempo-frequência um fator de atenuação é aplicado ao sinal de fala contaminado. Cada máscara é caracterizada por uma curva de ganho específica em função da SNR associada (CHIEA; COSTA; BARRAULT, 2019a). Após o processamento individual de cada unidade, um processo de reconstituição para um único sinal é realizado com a transformada de Fourier inversa e uma estratégia de sobreposição e soma.

Nos últimos anos muitas máscaras têm sido propostas na literatura. De forma geral, possuem origem puramente heurística ou a partir de um critério de minimização de uma função custo. As principais máscaras encontradas na literatura são a máscara

binária (WANG; BROWN, 2006), a máscara de Wiener (LOIZOU, 2013) e suas versões restrita (LOIZOU, 2013) e paramétrica (LIM; OPPENHEIM, 1979) e a máscara paramétrica conformável (FONTAINE et al., 2017). Em todas elas, a curva de ganho, em função da SNR em decibel, é simétrica em relação ao valor de meio ganho.

A máscara paramétrica conformável é capaz de gerar uma quantidade imensurável de curvas de ganho, incluindo as da máscaras binária, de Wiener, e restrita de Wiener, possuindo assim, grande flexibilidade em função da definição de seus parâmetros de conformação. No entanto, em certas situações, o seu desempenho pode não ser relevante, principalmente em termos de qualidade. Isso ocorre, principalmente, em condições de SNR baixa, nas quais o sinal processado pela máscara tempo-frequência pode apresentar alta inteligibilidade, mas ainda assim apresentar baixa qualidade.

Este trabalho tem como objetivo investigar o limite de desempenho da técnica de máscara tempo-frequência para redução de ruído e, assim, avaliar se a pesquisa por novas máscaras deve continuar sendo explorada.

## 1.3 OBJETIVOS

Nas subseções a seguir estão descritos o objetivo geral e os objetivos específicos deste trabalho.

### 1.3.1 Objetivo Geral

O objetivo geral deste trabalho é investigar os limites de desempenho das máscaras tempo-frequência para redução de ruído em aparelhos auditivos.

### 1.3.2 Objetivos Específicos

Os objetivos específicos são:

a) Estudar as máscaras tempo-frequência, a fim de avaliar sua eficácia e identificar oportunidades de otimização.

b) Propor uma máscara com alta flexibilidade para a geração de curvas de atenuação, removendo a imposição usual da simetria encontrada nas máscaras tradicionais.

c) Avaliar o desempenho da máscara proposta em várias condições de SNR e ruído.

d) Determinar a configuração da máscara proposta que se destaca em cada condição ou se há uma configuração que produza resultado satisfatório em uma ampla gama de condições.

e) Avaliar se configurações diferentes da máscara produzem variações de desempenho em vogais e consoantes.

# 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, será apresentada a fundamentação teórica do trabalho, englobando a fala e suas características, o sistema auditivo humano com ênfase nas diversas causas e graus de perda auditiva, os tipos de aparelho auditivo e a descrição matemática das máscaras tempo-frequência.

## 2.1 FALA

A fala é uma forma de comunicação que envolve a produção de sons articulados pela boca e pelas cordas vocais, permitindo que os seres humanos expressem pensamentos, sentimentos, ideias e se comuniquem uns com os outros. É uma das principais formas de linguagem utilizada pelas pessoas para se comunicar e se expressar (CORSINO, 2020).

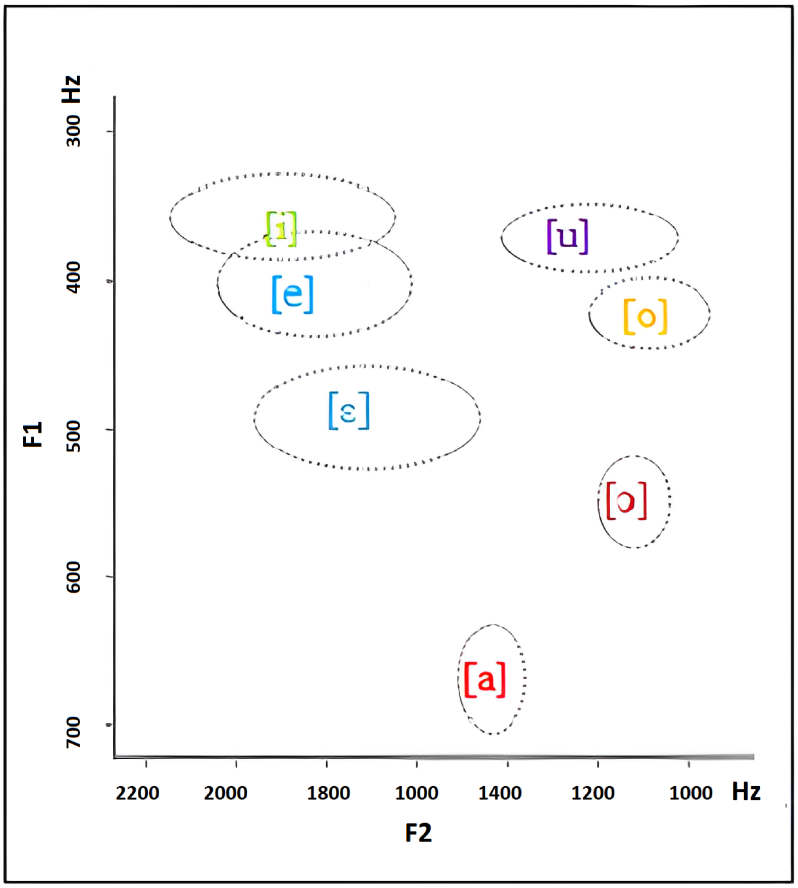
Os sons da fala podem ser classificados em duas grandes categorias: vogais e consoantes, que possuem características distintas. As vogais apresentam alta intensidade de energia em comparação com as consoantes. Para a percepção das vogais, as frequências das formantes são um fator importante. As formantes são ressonâncias do trato vocal e são responsáveis por dar às vogais suas características distintivas (LOIZOU, 2013).

A primeira formante (F1) está associada a alterações na abertura da boca. Vogais que apresentam uma F1 de alta frequência exigem uma abertura da boca mais ampla, enquanto vogais que apresentam uma abertura da boca menor têm uma F1 em frequências mais baixas. Por sua vez, a segunda formante (F2) está associada a alterações na cavidade oral, como a posição da língua e a atividade dos lábios. As semivogais geralmente apresentam padrão semelhante às vogais, porém as formantes não atingem o estado estacionário, e a percepção de mudança ocorre na F2 e na terceira formante (F3) (LOIZOU, 2013).

As frequências das formantes das vogais, quando representadas graficamente uma contra a outra, geram um padrão em elipse. A maioria dos pares de valores de F1 e F2 de uma mesma vogal pertencem a uma mesma elipse, conforme mostrado na Figura 1. Embora seja geralmente aceito que as frequências formantes em estado estacionário são as principais pistas para a percepção das vogais, elas não são as únicas informações utilizadas para sua identificação. Outras características adicionais, como duração e mudança espectral, também ajudam na identificação das vogais (LOIZOU, 2013).

As consoantes apresentam características diferentes das vogais e podem ser divididas em subgrupos (nasais, plosivas, fricativas e líquidas), cada um com suas particularidades. As nasais (/m/, /n/ e /nh/) são caracterizadas por uma ressonância na faixa de 200-500 Hz, chamada de "murmúrio " nasal. As consoantes plosivas (/p/,

Figura 1 – Padrão dos valores de F1 e F2, em Hz das vogais do Português.




Fonte – Adaptado de Pereyron e Alves (2019)

/b/, /t/, /d/, /k/ e /g/) são geradas pelo fluxo de ar restrito e podem apresentar um sinal periódico de baixa intensidade na frequência fundamental F0 (frequência produzida pela vibração das pregas vocais e seus harmônicos modificados nas cavidades supraglóticas) durante toda ou parte da restrição, que pode variar de 10 a 100 ms. As fricativas (/f/, /v/, /s/, /ch/, /z/ e /j/) apresentam picos espectrais de alta frequência na casa de kHz (PEREYRON; ALVES, 2019; LOIZOU, 2013; VIEGAS et al., 2019). Por último, as líquidas ( /r/,/l/, /l/, e /R/ ), segundo Pagan e Wertzner (2007), apresentam os valores médios de F1 por volta de 250 a 500 Hz e os de F2 entre 1250 e 1460 Hz.

## 2.2 OUVIDO HUMANO

As ondas sonoras geradas ao nosso redor são captadas, processadas e decodificadas pelo sistema auditivo humano, o qual pode ser dividido em três partes: ouvido externo, ouvido médio e interno. Cada parte serve a uma função específica na aquisição e codificação do som. O ouvido externo, também denominado pavilhão

auricular, serve para coletar o som que seguirá para o canal auditivo ou meato acústico externo. As ondas sonoras, ao percorrerem as elevações e depressões da pina (também chamada de aurícula - parte visível do ouvido em formato espiral), sofrem reflexões e atenuações, que são responsáveis por fornecer informações sobre a direção da fonte geradora do som. Após passarem pela pina, as ondas sonoras percorrem o canal auditivo e chegam até a porção inicial do ouvido médio (DURAN, 2011). Todas as partes do ouvido humano citadas acima podem ser vistas na Figura 2.

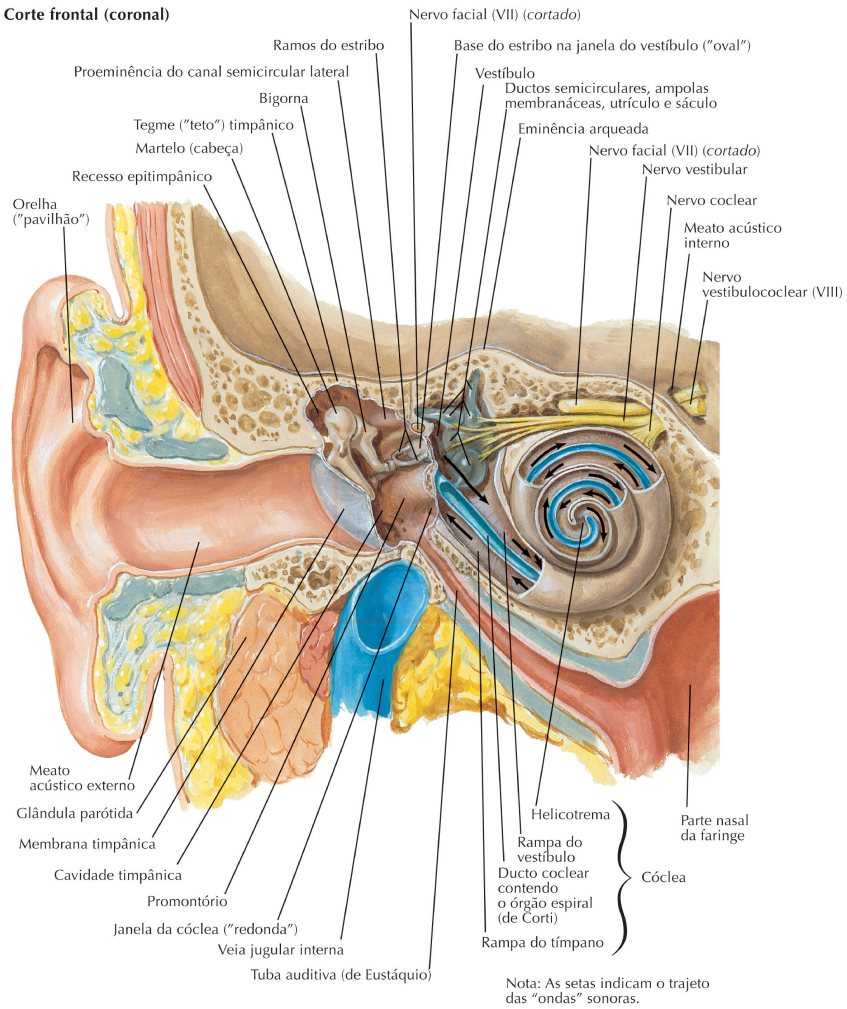



Figura 2 – Ouvido humano

Fonte – Netter (2011)

O ouvido médio consiste em uma cavidade cheia de ar com um volume de aproximadamente 2 $cm^3$, contendo três pequenos ossos: martelo, bigorna e estribo, conforme observado na Figura 2. Esses ossos têm como função realizar a transmissão do som da membrana timpânica para o ouvido interno. De forma simples, as ondas sonoras agitam a membrana timpânica, movimentando o martelo, bigorna e estribo, e transmitindo para o ouvido interno as vibrações sonoras. Segundo Garcias (2002), o ouvido médio realiza duas tarefas: fazer com que a pressão do lado interno da

membrana timpânica seja igual à pressão do lado externo; e promover um ganho mecânico a fim de que a energia da onda sonora seja suficiente para promover a vibração das linfas e membranas do ouvido interno.

O ouvido interno contém em sua estrutura a cóclea, cuja paredes são limitadas por três tubos enrolados em espiral. Os dois primeiros tubos, chamados de osso vestibular e timpânico, possuem formas cônicas e estão revestidos por uma membrana. Eles estão unidos entre si através de uma pequena abertura, chamada de janela oval, por onde entra um tendão que conecta o músculo estapedial à ossiculação. O terceiro tubo, chamado de semicanal, é constituído pela membrana basilar e pela membrana tectorial, que está fixada na janela oval. No extremo inferior desse canal está o helicotrema (OKUNO; CALDAS; CHOW, 1982).

Após a onda sonora ser amplificada pela parte do ouvido médio e interno, ela é coletada na cóclea, também conhecida como caracol. A cóclea hospeda o órgão sensitivo da audição, o órgão de Corti, que se conecta com o nervo auditivo. Esse órgão fica situado sobre uma membrana basilar com milhares de fibras que, quando movidas pela linfa, coletam a frequência específica proveniente de cada vibração, ou seja, somente algumas ondas acionam determinadas fibras de acordo com as características do som produzido e então são transmitidas ao nervo auditivo (DURAN, 2011).

A perda parcial ou total da audição é denominada deficiência auditiva. Segundo a American Speech-Language-Hearing Association (2023), a deficiência auditiva pode ter diversas causas, algumas delas são:

a) Envelhecimento: à medida que avança a idade, é comum que ocorra uma perda gradual da audição devido ao desgaste natural do sistema auditivo;

b) Exposição a ruído: a exposição frequente ou prolongada a altos níveis de ruído pode danificar as células auditivas no ouvido interno;

c) Medicamentos e agentes ototóxicos: alguns medicamentos podem ser tóxicos para as células auditivas e causar perda auditiva, como alguns antibióticos e quimioterápicos;

O grau da perda auditiva é classificado em níveis: leve, moderado, severo ou profundo. No Quadro 1, apresenta-se a classificação do grau da perda auditiva conforme o sistema classificatório dos Conselhos de Fonoaudiologia e da Sociedade Brasileira de Fonoaudiologia (SISTEMA DE CONSELHOS DE FONOAUDIOLOGIA, 2020; WORLD HEALTH ORGANIZATION, 2020).

## 2.3 APARELHOS AUDITIVOS

De acordo com os relatos históricos, por volta dos séculos XVII e XIX, os aparelhos auditivos tinham a forma de uma corneta e eram feitos a partir de madeira, metal

Quadro 1 – Classificação do grau da perda auditiva conforme a Organização Mundial da Saúde

| **Graus de perda auditiva** | **Média entre as frequências de 500, 1K, 2K e 4K Hz** | | **Capacidade auditiva** |
|---|---|---|---|
| | Criança | Adulto | |
| **Audição Normal** | 0 – 15 dB | 0 – 25 dB | Nenhuma ou pequena dificuldade; capaz de ouvir cochichos. |
| **Perda Leve** | 16 – 30 dB | 26 – 40 dB | Capaz de ouvir e repetir palavras em volume normal a um metro de distância . |
| **Perda Moderada** | 31 – 60 dB | 41 – 60 dB | Capaz de ouvir e repetir palavras em volume elevado a um metro de distância. |
| **Perda Severa** | 61 – 80 dB | 61 – 80 dB | Capaz de ouvir palavras em voz gritada próxima ao melhor ouvido. |
| **Perda Profunda** | > 81 dB | > 81 dB | Incapaz de ouvir e entender mesmo em voz gritada no melhor ouvido. |

Fonte – Adaptado de Sistema de Conselhos de Fonoaudiologia (2020)

ou até mesmo de chifres de animais (HEARING SOLUTIONS, SOUND ADVICE YOU CAN TRUST, 2023; MUSEU DO PARELHO AUDITIVO, 2023). Um exemplo é mostrado na Figura 3.

Um dos primeiros aparelhos auditivos que se tem registro é o trompete collapsible, visto na Figura 4. Com o uso deste aparelho criou-se a possibilidade de fornecer uma direcionalidade para os sons desejados e, ao mesmo tempo, dificultar que o ouvido captasse sons indesejáveis. Seguindo ainda a linha da trombeta, criou-se o tubo de conversação, que conseguia aumentar a razão sinal-ruído sem uma amplificação direta, ao fazer o som entrar por uma extremidade do tubo e seguir para dentro do canal auditivo. Os aparelhos auditivos com suprimento elétrico começaram a ser fabricados a partir do início dos anos 1920 e seus modelos foram evoluindo, acompanhando os avanços tecnológicos, até os dias de hoje (MUSEU DO PARELHO AUDITIVO, 2023; HEARING SOLUTIONS, SOUND ADVICE YOU CAN TRUST, 2023).

Existem diferentes tipos de aparelhos auditivos, os quais podem ser classificados em termos tecnológicos entre analógicos e digitais. Ambos possuem funcionamento semelhante: coletam as ondas sonoras pelo microfone, enviam-nas para a

Figura 3 – Trombetas Auditivas de Cornos


Fonte – https://museudoaparelhoauditivo.com.br/acervo-aparelhos-auditivos-nao-eletricos-trombetas-auditivas-trombetas-auditivas-de-cornos.php

Figura 4 – Trombeta Auditiva Collapsible


Fonte – https://museudoaparelhoauditivo.com.br/acervo-aparelhos-auditivos-nao-eletricos-trombetas-auditivas-trombeta-auditiva-collapsible.php

unidade de amplificação e filtragem, e então as direcionam ao alto-falante para serem reproduzidas. O aparelho analógico, devido à sua simplicidade, possui circuitos menos versáteis, com flexibilidade limitada e restrições relacionadas ao processamento do sinal no circuito analógico, apresentando baixo custo de mercado e baixo consumo de energia (ALMEIDA; IORIO, 2003). Por outro lado, os aparelhos digitais possuem unidades de processamento que possibilitam processamento mais complexo, como

de supressão de ruído. Também permitem realizar tratamento das informações recebidas por filtros de ordem elevada, amplificadores multicanais, ajustes seletivos dos níveis de saída e controle automático de ganho, para só então enviá-las ao alto-falante para realizar a reprodução do sinal sonoro (ALMEIDA; IORIO, 1996; MARTIN; HEUTE; ANTWEILER, 2008).

Os dispositivos digitais são mais versáteis e podem ser configurados individualmente, o que facilita a adequação às necessidades de cada usuário. Porém, a digitalização do sinal, por exemplo, resulta em um maior consumo de energia (DILLON, 2001). Os aparelhos auditivos possuem diversas arquiteturas e modelos diferentes, sendo os mais comuns o aparelhos auditivo no canal (ITC, do inglês, In The Canal), completamente no canal (CIC, Completely in Canal), na orelha (ITE, In The Ear) e atrás da orelha (BTE, Back The Ear) (ARSINTE; LUPU; SUMALAN, 2017). É possível encontrar aparelhos auditivos bilaterais, que são pares de aparelhos (um em cada ouvido) que operam de forma independente um do outro, e biauriculares, que são pares de aparelhos que operam de forma conjunta, compartilhando entre eles informações para melhorar os sinais sonoros de saída (DILLON, 2001).

Neste trabalho abordaremos técnicas de supressão de ruído para dispositivos digitais bilaterais mono-canais podendo ser de qualquer um dos quatro modelos citados anteriormente (ITC, CIC, ITE e BTE).

## 2.4 MÁSCARAS TEMPO-FREQUÊNCIA

Por décadas, as técnicas de aprimoramento de fala têm sido estudadas com o intuito de melhorar a qualidade, inteligibilidade e proporcionar a redução de ruídos adicionados na fala (LOIZOU, 2013). As máscaras tempo-frequência são uma das principais formas de redução de ruído em aparelhos auditivos (CHIEA; COSTA; BARRAULT, 2019b). Sua formulação matemática é apresentada a seguir.

Sejam o sinal de fala de interesse e o ruído ambiente denotados por $x(n)$ e $v(n)$, respectivamente, o sinal de fala contaminado é definido por $y(n) = x(n) + v(n)$. Assume-se que $x(n)$ e $v(n)$ são não-observáveis e descorrelacionados entre si. Neste caso, a transformada de Fourier de tempo curto (STFT) de $y(n)$, $Y(k,\lambda)$, é definida como

$$Y(k,\lambda) = X(k,\lambda) + V(k,\lambda), \tag{1}$$

onde $X(k,\lambda)$ e $V(k,\lambda)$ são as STFTs de $x(n)$ e $v(n)$, respectivamente, $k$ é o bin de frequência e $\lambda$ é o índice da janela de tempo.

A técnica de mascaramento tempo-frequência para redução de ruído consiste em, a cada unidade tempo-frequência $\{k,\lambda\}$, multiplicar $Y(k,\lambda)$ por uma máscara $M(k,\lambda)$, resultando numa estimativa de $X(k,\lambda)$ dada por (LOIZOU, 2013)

$$\hat{X}(k,\lambda) = Y(k,\lambda)M(k,\lambda). \tag{2}$$

No domínio do tempo, a estimativa $\hat{x}(n)$ do sinal de fala de interesse é reconstruída utilizando a STFT inversa e uma estratégia de sobreposição e soma (*overlap-and-add*) (CROCHIERE, 1980).

As máscaras podem ser definidas utilizando critérios objetivos ou heurísticos. Em geral, $0 \leq M(k,\lambda) \leq 1$ e $M(k,\lambda)$ é uma função da razão sinal-ruído associada à $\lambda$-ésima janela e $k$-ésimo bin de frequência, a qual é definida como

$$\xi(k,\lambda) = \frac{S_x(k,\lambda)}{S_v(k,\lambda)}, \tag{3}$$

onde $S_x(k,\lambda) = \mathrm{E}\left\{|X(k,\lambda)|^2\right\}$ e $S_v(k,\lambda) = \mathrm{E}\left\{|V(k,\lambda)|^2\right\}$ são as densidades espectrais de potência de $x(n)$ e $v(n)$, respectivamente, e $\mathrm{E}\{\cdot\}$ é o operador valor esperado. A SNR na escala decibel é dada por $\xi_{\mathrm{dB}}(k,\lambda) = 10\log_{10}\xi(k,\lambda)$.

É importante ressaltar que, quando se implementam técnicas de redução de ruído em dispositivos físicos reais, os sinais de fala $x(n)$ e ruído $v(n)$ são desconhecidos, o que torna impossível obter a SNR real *a priori* . Portanto, na prática, é necessário estimar $\xi(k,\lambda)$ por meio de algoritmos, como o método de decisão direta e suas variações (ALAM; O'SHAUGHNESSY; SELOUANI, 2008; EPHRAIM; MALAH, 1984). No entanto, em condições ideais (de simulação, utilizando sinais artificialmente somados), quando se tem conhecimento dos sinais $x(n)$ e $v(n)$, é comum utilizar os valores exatos de $\xi(k,\lambda)$ conforme a equação (3). Isso possibilita avaliar o desempenho máximo das máscaras $M(k,\lambda)$, conhecidas neste caso como máscaras ideais. A seguir, serão apresentadas as principais máscaras para redução de ruído encontradas na literatura.

## 2.5 TIPOS DE MÁSCARAS

O desenvolvimento de máscaras tempo-frequência pode ocorrer por meio de critérios de otimização ou por meio de modelos heurísticos. No primeiro caso, as máscaras são descritas como modelos ótimos, enquanto que no segundo caso elas são baseadas em regras empíricas ou conhecimento prévio do problema.

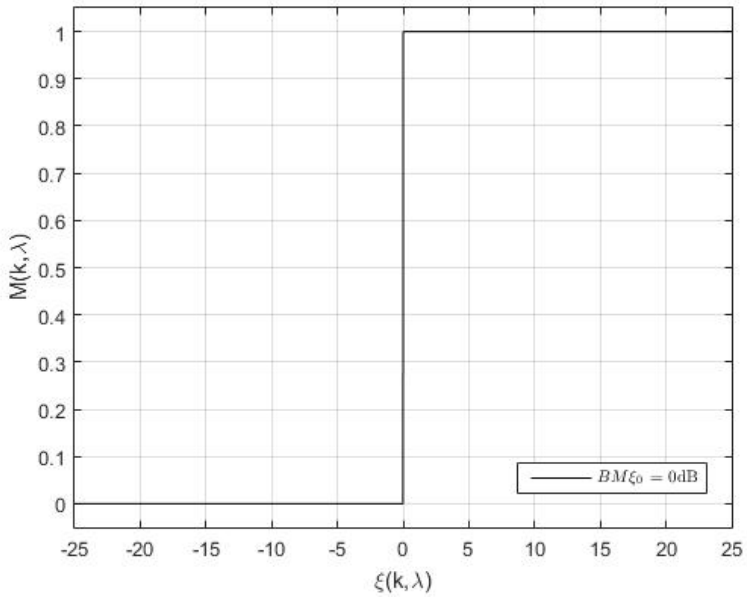
### 2.5.1 Máscara Binária

A máscara tempo-frequência mais simples encontrada na literatura é a máscara binária (BM), a qual é definida como (WANG; BROWN, 2006),

$$M(k,\lambda) = \begin{cases} 1, & \xi(k,\lambda) \geq \xi_0, \\ 0, & \xi(k,\lambda) < \xi_0, \end{cases} \tag{4}$$

onde $\xi_0$ é uma constante geralmente igual a 0 dB. Devido à descontinuidade, a BM é classificada como uma máscara dura. Por outro lado, máscaras suaves são caracterizadas por funções de ganho com transição suave entre seus valores extremos. A Figura 5 exibe a curva de ganho da BM para $\xi_0 = 0$ dB.

Figura 5 – Curva de ganho BM para $\xi_0$ = 0 dB




### 2.5.2 Máscara de Wiener

A principal máscara suave é a máscara de Wiener (WM), a qual é definida como (LOIZOU, 2013)

$$M(k,\lambda) = \frac{\xi(k,\lambda)}{\xi(k,\lambda)+1} \tag{5}$$

Sua curva de ganho, é exibida em preto na Figura 6. A WM é o filtro ótimo que minimiza, no domínio da frequência, o erro quadrático médio entre a fala e sua estimativa, isto é,

$$J(k,\lambda) = \mathrm{E}\{|\hat{X}(k,\lambda) - X(k,\lambda)|^2\} \tag{6}$$

### 2.5.3 Máscara Restrita de Wiener

Uma variação menos abrupta da WM é a máscara restrita de Wiener (WR), a qual é definida como

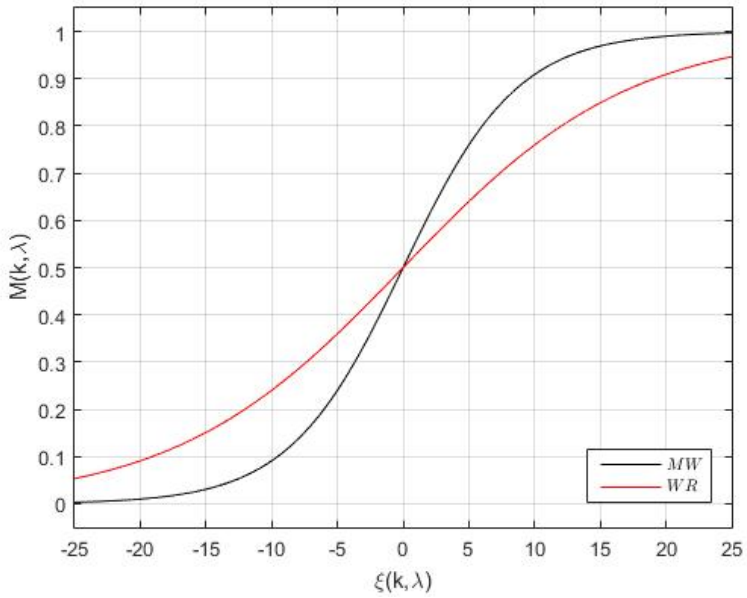
$$M(k,\lambda) = \frac{\sqrt{\xi(k,\lambda)}}{\sqrt{\xi(k,\lambda)}+1} \tag{7}$$

A WR resulta em uma atenuação mais moderada que a WM em situações de baixa SNR. A Figura 6 exibe, em vermelho, sua respectiva curva de ganho. A WR é o filtro que iguala as distorções relativas de $v(n)$ e $x(n)$, minimizando a seguinte função custo (LOIZOU, 2013)

$$J(k,\lambda) = S_{d_x}(k,\lambda) + S_{d_v}(k,\lambda), \tag{8}$$

onde $S_{d_x}(k,\lambda) = |M(k,\lambda)-1|^2 S_x(k,\lambda)$ e $S_{d_v}(k,\lambda) = |M(k,\lambda)|^2 S_v(k,\lambda)$ são as densidades espectrais de potência da distorção causada pela máscara ao sinal de fala e ao ruído, respectivamente.

Figura 6 – Curva de ganho da WM e WR




### 2.5.4 Máscara Paramétrica de Wiener

A máscara paramétrica de Wiener (WP) é uma extensão heurística da WM obtida com a introdução de dois parâmetros na função de ganho, sendo definida como (LIM; OPPENHEIM, 1979)

$$M(k,\lambda) = \left(\frac{\xi(k,\lambda)}{\xi(k,\lambda)+\eta}\right)^{\beta} \tag{9}$$

Quando $\beta = \eta = 1$, à WP equivale à WM.

### 2.5.5 Máscara Paramétrica Conformável

Em Fontaine et al. (2017), uma máscara paramétrica conformável (CM) foi proposta como uma extensão heurística da WP e definida como

$$M(k,\lambda) = \frac{\xi^{\gamma}(k,\lambda)}{\xi^{\gamma}(k,\lambda)+\mu^{\gamma}} \tag{10}$$

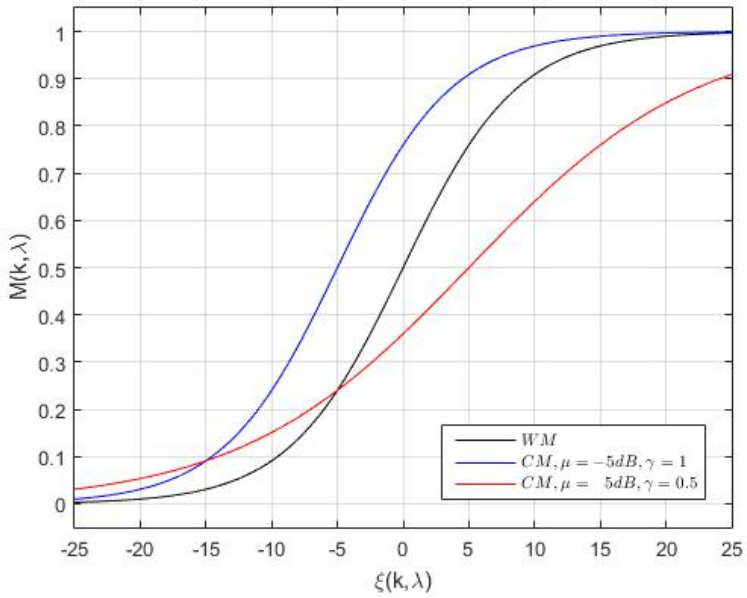
onde $\mu > 0$ e $\gamma > 0{,}5$ são parâmetros que permitem o ajuste do deslocamento lateral e da inclinação da máscara, respectivamente.

Recentemente, em Chiea, Costa e Barrault (2019b), demonstrou-se que a CM é o filtro ótimo que minimiza uma função custo dada por

$$J(k,\lambda) = S^{\delta}_{d_x}(k,\lambda) + \rho S^{\delta}_{d_v}(k,\lambda), \tag{11}$$

onde $S_{d_x}(k,\lambda) = |M(k,\lambda)-1|^2 S_x(k,\lambda)$ e $S_{d_v}(k,\lambda) = |M(k,\lambda)|^2 S_v(k,\lambda)$ são as densidades espectrais de potência da distorção causada pela máscara ao sinal de fala e ao ruído, respectivamente, $\rho = \mu^{\delta} > 0$ e $\delta = \frac{\gamma}{2\gamma-1} > 0$.

Figura 7 – Curva de ganho da CM para duas configurações distintas.




A CM apresenta uma maior maleabilidade na função de ganho em relação ao WP, permitindo seu ajuste inclusive na forma de máscaras duras. Para $\mu = \gamma = 1$, a CM equivale à WM. Para $\mu = 1$ e $\gamma = 0,5$, ela equivale à WR. E para $\mu = \xi_0$ dB e $\gamma \to \infty$, ela equivale à BM. Na Figura 7, são exibidas curvas de ganho da CM para os seguintes conjuntos de parâmetros: $\mu = \gamma = 1$ (WM), $\mu = 5$ dB e $\gamma = 0,5$, e $\mu = –5$ dB e $\gamma = 1$.

## 2.6 CRITÉRIOS OBJETIVOS DE QUALIDADE E INTELIGIBILIDADE DA FALA

Neste trabalho são utilizados dois critérios objetivos quantitativos para avaliar a qualidade e inteligibilidade dos sinais de fala processados pelas máscaras tempo-frequência.

### 2.6.1 W-PESQ

O W-PESQ (*Wideband Perceptual Evaluation of Speech Quality*) é um algoritmo para avaliação objetiva da qualidade de sinais de fala amostrados a 16 kHz (ITU-T P.862, 2001; ITU-T P.862.2, 2005). Ele compara representações psicoacústicas de um sinal de fala possivelmente degradado e sua correspondente referência não corrompida (BISPO et al., 2010). A pontuação bruta do W-PESQ pode ser mapeada para a escala 1-5 da opinião média (MOS, do inglês *Mean Opinion Score*), resultando na pontuação MOS-LQO (*Mean Opinion Score-Listening Quality Objective*) (ITU-T P.800.1, 2006). A correspondência entre a escala de cinco pontos e a classificação da categoria de degradação (DCR, do inglês *Degradation Category Rating*) é mostrada no Quadro 2. No entanto, o máximo MOS-LQO fornecida pelo W-PESQ é 4,644 quando os sinais de referência e degradados são idênticos.

Quadro 2 – Correspondência entre pontuação e DCR.

| **Pontuação** | **DCR - Qualidade da fala** |
|---|---|
| 5 | Degradação inaudível |
| 4 | Degradação audível, mas não incômoda |
| 3 | Degradação pouco incômoda |
| 2 | Degradação incômoda |
| 1 | Degradação muito incômoda |

Fonte – Adaptado de ITU-T P.800.1 (2006)

Neste trabalho, o W-PESQ é utilizado para avaliar o desempenho das máscaras em relação à qualidade sonora dos sinais processados. Para isso, os sinais $x(n)$ e $\hat{x}(n)$ são utilizados como os sinais de referência e degradado, respectivamente.

### 2.6.2 NCM

O NCM (*Normalized Covariance Metric*) é uma métrica para avaliação objetiva da inteligibilidade de sinais de fala (LOIZOU, 2013; HOLUBE; KOLLMEIER, 1996). Ela é baseada na covariância entre os envelopes temporais de sub-bandas de um sinal de fala possivelmente degradado e sua correspondente referência não corrompida (LOIZOU, 2013). Sua pontuação varia entre 0 e 1, onde valores mais altos indicam maior inteligibilidade (LOIZOU, 2013).

Neste trabalho, o NCM é utilizado para analisar o desempenho das máscaras em relação à inteligibilidade dos sinais processados. Para isso, os sinais $x(n)$ e $\hat{x}(n)$ são utilizados como os sinais de referência e degradado, respectivamente.

# 3 INVESTIGAÇÃO DO LIMITE DE DESEMPENHO DAS MÁSCARAS TEMPO-FREQUÊNCIA

Este capítulo tem o intuito de investigar o limite de desempenho da técnica de máscara tempo-frequência na redução de ruído em aparelhos auditivos e, assim, avaliar se a investigação por novas máscaras deve continuar sendo explorada. Para tanto este trabalho propõe de forma heurística uma máscara com quatro graus de liberdade. A máscara proposta apresenta flexibilidade suficiente tanto para gerar curvas de ganho simétricas (como BM, WM, WR e CM) quanto assimétricas em relação ao valor de meio ganho.

## 3.1 PROPOSTA DE MÁSCARA PARAMÉTRICA CONFORMÁVEL ASSIMÉTRICA

A máscara conformável assimétrica (ACM, do inglês *Asymmetric Conformable Mask*) é definida heuristicamente como

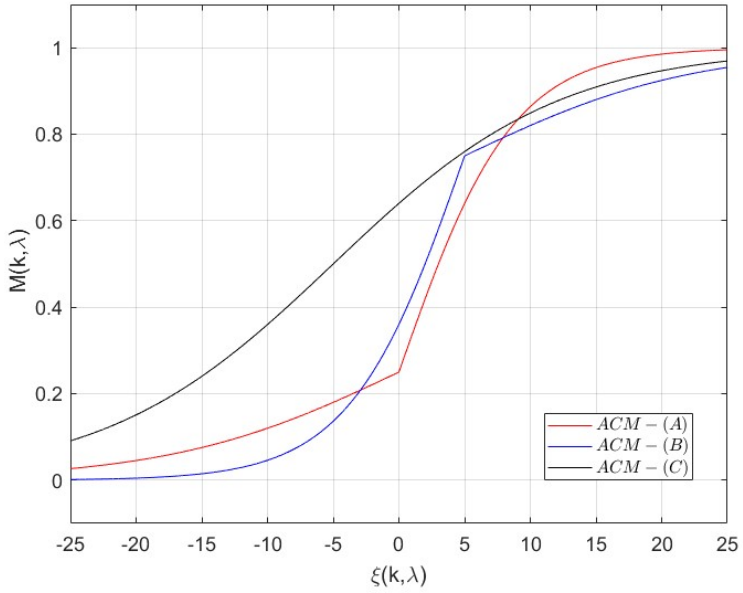
$$M(k,\lambda) = \begin{cases} \dfrac{2g}{1+e^{-\beta_1[\xi_{\mathrm{dB}}(k,\lambda)-\xi_0]}}, & \xi_{\mathrm{dB}}(k,\lambda) < \xi_0, \\ (2g-1) + \dfrac{-2g+2}{1+e^{-\beta_2[\xi_{\mathrm{dB}}(k,\lambda)-\xi_0]}}, & \xi_{\mathrm{dB}}(k,\lambda) \geq \xi_0, \end{cases} \tag{12}$$

onde $\xi_0$ representa o deslocamento lateral (em dB) da curva, $g$ é o valor de ganho ($0 < g < 1$) em relação ao ponto de meio ganho quando uma possível simetria é estabelecida, e $\beta_1$ e $\beta_2$ são parâmetros positivos que controlam o comportamento exponencial da máscara para $M(k,\lambda) < g$ e $M(k,\lambda) > g$, respectivamente. Se $\beta_1 = \beta_2$, a ACM se torna simétrica em relação a $M(k,\lambda) = g$, o qual não é obrigatoriamente igual a 0,5.

Quando $g = 0{,}5$, $\beta_1 = \beta_2 = 0{,}1\gamma \ln 10$ e $\xi_0 = 10^{0,1\mu}$, a ACM equivale à CM. Para $g = 0{,}5$, $\beta_1 = \beta_2 = 0{,}1 \ln 10$ e $\xi_0 = 0$, ela equivale à WM. Para $g = 0{,}5$, $\beta_1 = \beta_2 = 0{,}1 \ln \sqrt{10}$ e $\xi_0 = 0$, a ACM equivale à WR. Para $g = 0{,}5$ e $\beta_1 = \beta_2 \to \infty$, ela equivale à BM. A Figura 8 ilustra a máscara proposta para dois conjuntos de parâmetros que tornam a curva de ganho assimétrica e o conjunto de parâmetros $\beta_1 = \beta_2$ e $g = 0{,}5$ que gera a CM.

No conhecimento do autor, essa é a primeira máscara tempo-frequência proposta na literatura com curva de ganho com possibilidade de exibir assimetria. Por ser capaz de modelar todas as máscaras do estado-da-arte e uma infinidade de outras curvas de ganho, simétricas ou assimétricas, neste trabalho a ACM será considerada a máscara que produz o melhor desempenho possível, sendo este o limite de desempenho da técnica de máscara tempo-frequência.

Figura 8 – Curva de ganho de ACM em 3 configurações distintas: (*A*) $\mu = 0$ *dB*, $g = 0{,}25$, $\beta_1 = 0{,}1 \cdot log10(\sqrt{10})$, $\beta_2 = 0{,}1 \cdot log10(10)$. (*B*) $\mu = 5$ *dB*, $g = 0{,}75$, $\beta_1 = 0{,}1 \cdot log10(10)$, $\beta_2 = 0{,}1 \cdot log10(\sqrt{10})$. *C*) $\mu = -5\,dB$, $g = 0{,}5$, $\beta_1 = \beta_2 = 0{,}1 \cdot log10(\sqrt{10})$, $\equiv CM$,$\mu = -5dB$,$\gamma = 0{,}5$.




## 3.2 METODOLOGIA DAS SIMULAÇÕES

Essa seção descreve a metodologia empregada na aplicação e análise de desempenho da máscara ACM proposta.

### 3.2.1 Sinais de fala e ruído

Foi utilizado um conjunto de 10 sinais de fala masculina e 10 sinais de fala feminina, sendo 5 sinais por locutor. O conjunto de falas é considerado foneticamente balanceado no português (ALCAIM; SOLEWICZ; MORAES, 1992). Os sinais que estão disponíveis em Ynoguti (1999), possuem duração de aproximadamente 6 s e foram amostrados a uma taxa de 16 kHz. Os sinais de fala foram contaminados com ruído branco gaussiano (WGN, do inglês white gaussian noise) e com ruído com espectro semelhante ao da fala usando o sinal ICRA (do inglês International collegium for rehabilitative audiology) (DRESCHLER et al., 2001). A contaminação foi realizada em quatro níveis de SNR: -10 dB, 0 dB 10 dB e 20 dB.

### 3.2.2 Implementação das máscaras

As STFTs $Y(k,\lambda)$, $X(k,\lambda)$ e $V(k,\lambda)$ foram calculadas utilizando uma janela de Hamming com duração de 20 ms, sobreposição de 50 % e uma transformada discreta de Fourier (DFT, do inglês Discrete Fourier Transform) de 320 pontos (LOIZOU, 2013).

Para cada janela $\lambda$, uma estimativa de $\xi(k,\lambda)$ foi obtida como

$$\hat{\xi}(k,\lambda) = \frac{|X(k,\lambda)|^2}{|V(k,\lambda)|^2}, \quad k = 1,2,\ldots,320, \tag{13}$$

e utilizada para computar o valor de ganho $M(k,\lambda)$ de diferentes máscaras. O ganho $M(k,\lambda)$ foi por sua vez aplicado a $Y(k,\lambda)$ conforme (2), de forma a resultar na estimativa $\hat{X}(k,\lambda)$. Por fim, a estimativa $\hat{x}(n)$ do sinal de fala limpo foi construída utilizando as STFTs inversas de $\hat{X}(k,\lambda)$ e a técnica de sobreposição-e-soma.

Três máscaras foram aplicadas: WM, CM e ACM. As duas primeiras serviram de base de comparação para a ACM proposta. A WM foi utilizada por ser a máscara mais disseminada na literatura e a CM por ser o atual estado-da-arte. A implementação da CM foi feita através da ACM com $g = 0{,}5$, fazendo $\beta_1 = \beta_2$. Os parâmetros $\beta_1$ e $\beta_2$ foram analisados no intervalo $[0{,}01; 0{,}5]$ com passo de 0,05 e no intervalo $[0{,}5; 1]$ com passo de 0,1, respectivamente. Os parâmetros $g$ e $\xi_0$ foram analisados nos intervalos $[0{,}1; 1]$ e $[-15; 15]$ dB com passos de 0,1 e 1 dB, respectivamente. Esse conjunto de parâmetros resultou em 69.750 diferentes curvas de ganho, as quais foram aplicadas aos dois ruídos e nos três níveis de SNR.

### 3.2.3 Cenários de análise

Dois experimentos foram realizados para avaliar a capacidade da máscara ACM na redução de ruído. O primeiro investiga a qualidade perceptual do sinal processado $\hat{x}(n)$, buscando pela configuração de parâmetros que maximiza o valor médio do MOS-LQO. O segundo examina a inteligibilidade de $\hat{x}(n)$, buscando pela configuração que maximiza o valor médio do NCM.

Para cada combinação de ruído e SNR, os resultados foram analisados separadamente por gênero, masculino (M) e feminino (F), como também considerando todos os locutores (F & M). Além disso, para cada SNR, uma análise considerando os dois tipos de ruído (WGN & ICRA) e todos os locutores também foi realizada.

### 3.2.4 Interface gráfica de aprendizagem

Para auxiliar no estudo da técnica de máscara de tempo-frequência, foi desenvolvida uma interface que permite visualização e análise das curvas de ganho de cada uma das máscaras citadas na seção 2.5. Esta interface é constituída por três painéis: um para entrada de dados; o segundo para seleção de máscaras e ajuste de seus parâmetros, quando houver; e, por último, o painel de análise e avaliação que possibilita a realização de comparações visuais e auditivas dos áudios processados. Uma exemplificação detalhada do funcionamento da interface está disponível no Anexo A.

### 3.2.5 Testes estatísticos

O teste de Jarque-Bera ($\rho$ < 0,05), para verificar se os dados possuem uma distribuição normal (JARQUE; BERA, 1987), foi aplicado aos valores de MOS-LQO e NCM obtidos pelas máscaras ACM, CM e WM para cada combinação de ruído e SNR. A normalidade dos dados foi confirmanda em todos os casos. Então, para cada combinação de ruído e SNR, uma análise de variância (ANOVA) ($\rho$ < 0,05) foi realizada para verificar se os valores de MOS-LQO e NCM obtidos pelas máscaras são oriundos de distribuições diferentes (HOGG; LEDOLTER, 1987). Quando as distribuições não apresentaram diferenças estatisticamente significativas, o símbolo $\S$ é utilizado junto aos valores médios correspondentes nas tabelas mostradas na Seção 3.3.

## 3.3 RESULTADOS E DISCUSSÃO

Esta seção apresenta e discute os resultados obtidos nos dois experimentos realizados. Os sinais, critérios objetivos para avaliação e procedimentos descritos na Seção 3.2 foram empregados.

### 3.3.1 Resultados da Maximização do MOS-LQO

Neste experimento, buscou-se pelas configurações das máscaras CM e ACM que resultam no maior valor médio de MOS-LQO, independentemente do impacto ocasionado no NCM, sendo comparadas com a WM. As configurações ótimas em termos de qualidade da CM e ACM para SNRs de -10, 0, 10 e 20 dB são mostradas nas Tabelas 1, 2, 3 e 4, respectivamente. Os resultados de qualidade e inteligibilidade obtidos por essas configurações são apresentados nas Tabelas 5, 6, 7 e 8.

Tabela 1 – Configurações da ACM e CM que maximizam MOS-LQO para SNR de -10 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,15 | 0,4 | -1 | 0,15 | 1 |
| | M | 0,15 | 0,15 | 0,5 | -1 | 0,15 | -1 |
| | F & M | 0,15 | 0,15 | 0,5 | 0 | 0,15 | 0 |
| ICRA | F | 0,15 | 0,15 | 0,4 | 4 | 0,15 | 4 |
| | M | 0,15 | 0,1 | 0,3 | -3 | 0,15 | 4 |
| | F & M | 0,15 | 0,1 | 0,2 | -4 | 0,15 | 3 |
| WGN & ICRA | F & M | 0,15 | 0,15 | 0,3 | -1 | 0,15 | 4 |

Ao analisar as configuração das máscaras, nas quatro condições de SNR, percebe-se que $\beta_1$ de ACM e CM tende para 0,15. Conforme discutido na Seção 3.1,

Tabela 2 – Configurações da ACM e CM que maximizam MOS-LQO para SNR de 0 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,15 | 0,4 | -2 | 0,15 | 0 |
| | M | 0,15 | 0,15 | 0,4 | -3 | 0,15 | -1 |
| | F & M | 0,15 | 0,15 | 0,4 | -2 | 0,15 | 0 |
| ICRA | F | 0,15 | 0,15 | 0,5 | 3 | 0,15 | 3 |
| | M | 0,15 | 0,1 | 0,2 | -4 | 0,15 | 3 |
| | F & M | 0,15 | 0,1 | 0,2 | -4 | 0,15 | 3 |
| WGN & ICRA | F & M | 0,15 | 0,15 | 0,4 | -1 | 0,15 | 2 |

Tabela 3 – Configurações da ACM e CM que maximizam MOS-LQO para SNR de 10 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,15 | 0,6 | 2 | 0,15 | 0 |
| | M | 0,15 | 0,2 | 0,6 | 0 | 0,15 | -1 |
| | F & M | 0,15 | 0,2 | 0,6 | 1 | 0,15 | 0 |
| ICRA | F | 0,15 | 0,15 | 0,5 | 1 | 0,15 | 1 |
| | M | 0,15 | 0,1 | 0,2 | -7 | 0,15 | 1 |
| | F & M | 0,15 | 0,15 | 0,5 | 1 | 0,15 | 1 |
| WGN & ICRA | F & M | 0,15 | 0,15 | 0,5 | 0 | 0,15 | 0 |

esse parâmetro controla o comportamento exponencial da ACM para $M(k,\lambda) \leq g$. O valor 0,15 está entre os valores de $\beta_1$ das máscaras WR e WM, que são iguais a $0{,}1 \ln \sqrt{10} \approx 0{,}115$ e $0{,}1 \ln 10 \approx 0{,}23$, respectivamente. A única exceção é na CM para ruído WGN, fala masculina e SNR de 20 dB, como pode ser observado na Tabela 4, mas ainda é muito próximo ao valor de $\beta_1$ da WR.

Ao analisar as configurações de ACM e CM para SNR de -10, 0 e 10 dB, percebe-se que o parâmetro $\beta_2$ da ACM oscila entre os valores correspondentes WM e WR, enquanto $g$ e $\xi_0$ apresentam grande variabilidade. Quando $g < 0{,}5$, $\xi_0$ torna-se negativo, enquanto que para valores $g > 0{,}5$ $\xi_0$ é positivo. Na CM, observa-se pouca variabilidade de $\xi_0$ em função do gênero do locutor, as maiores diferenças ocorrem quando há a troca de ruído ou de SNR. Ademais, constata-se que na CM, o valor de $\xi_0$ tende a ser positivo, o que resulta num deslocamento da curva de ganho para a direita, conforme observado no caso de ruído WGN & ICRA com fala masculino & feminino nos quatro níveis de SNR.

Observa-se que, em relação aos valores de MOS-LQO, a ACM apresenta re-

Tabela 4 – Configurações da ACM e CM que maximizam MOS-LQO para SNR de 20 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,2 | 0,5 | 1 | 0,15 | 1 |
| | M | 0,15 | 0,15 | 0,1 | -10 | 0,2 | -3 |
| | F & M | 0,15 | 0,2 | 0,6 | 1 | 0,15 | 1 |
| ICRA | F | 0,15 | 0,15 | 0,7 | 5 | 0,15 | 1 |
| | M | 0,15 | 0,15 | 0,3 | -5 | 0,15 | 0 |
| | F & M | 0,15 | 0,15 | 0,7 | 4 | 0,15 | 1 |
| WGN & ICRA | F & M | 0,15 | 0,15 | 0,7 | 4 | 0,15 | 1 |

sultados iguais ou superiores em comparação à CM, como esperado. Entretanto, na média o ganho de desempenho da ACM em relação à CM é menor que 1%. Os valores de mediana e variância seguem um comportamento semelhante. As distinções costumam estar presentes na terceira casa decimal, levando a resultados sem diferença estatística significativa. Vale destacar que, mesmo maximizando MOS-LQO, existem aumentos do valor médio da NCM superiores ou equivalentes aos da WM, para os vários cenários de análise.

Ao avaliar diferentes tipos de ruído, WGN propicia um maior valor médio de MOS-LQO, quando comparado ao ICRA. Essa tendência também se repetiu para o gênero de fala masculina. Adicionalmente, para alguns casos, observa-se um maior valor médio da NCM mesmo maximizando a qualidade.

Finalmente, é importante destacar que, ao comparar a diferença entre os valores de MOS-LQO para ACM e CM ($MOS\text{-}LQO_{ACM} - MOS\text{-}LQO_{CM}$), para todos os cenários de análise, no caso de WGN & ICRA, com falas masculinas e femininas, observamos que o valor médio da distribuição tende a ser positivo. Este resultado pode ser visualizado na forma de diagramas de caixas na Figura 9. Note que para SNR de 10 dB, a ACM é equivalente à CM e portanto, os valores das diferenças entre MOS-LQO são nulos para todas as amostras. Isso reforça que a CM quando bem configurada se aproxima da ACM.

### 3.3.2 Resultados para Inteligibilidade

Neste experimento, buscou-se por configurações de parâmetros das máscaras CM e ACM que resultam no maior valor médio de NCM, independentemente da qualidade, resultante. Ambas as técnicas são também comparadas com a WM. As configurações das máscaras ACM e CM que obtiveram maior valor médio da inteligibilidade (utilizando NCM) para as SNRs de -10, 0, 10 e 20 dB, são mostradas nas

Tabela 5 – Resultados de MOS-LQO obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR -10 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | MOS-LQO | Média | 1,84 | **2,07**§ | **2,07**§ |
| | | | Mediana | 1,86 | 2,23 | 2,23 |
| | | | Variância | 0,0084 | 0,0115 | 0,0115 |
| | | NCM | Média | 0,876 | **0,912**§ | 0,911§ |
| | M | MOS-LQO | Média | 2,06 | **2,37**§ | **2,37**§ |
| | | | Mediana | 2,03 | 2,22 | 2,22 |
| | | | Variância | 0,014 | 0,039 | 0,037 |
| | | NCM | Média | 0,893 | **0,902**§ | **0,902**§ |
| | F & M | MOS-LQO | Média | 1,95 | **2,22**§ | **2,22**§ |
| | | | Mediana | 1,96 | 2,22 | 2,22 |
| | | | Variância | 0,0244 | 0,0542 | 0,0542 |
| | | NCM | Média | 0,886 | **0,913**§ | **0,913**§ |
| ICRA | F | MOS-LQO | Média | 1,39 | **1,50**§ | 1,48§ |
| | | | Mediana | 1,45 | 1,54 | 1,55 |
| | | | Variância | 0,0075 | 0,0127 | 0,0116 |
| | | NCM | Média | **0,885**§ | 0,883§ | 0,876§ |
| | M | MOS-LQO | Média | 1,54 | **1,64**§ | 1,63§ |
| | | | Mediana | 1,92 | 1,55 | 1,55 |
| | | | Variância | 0,023 | 0,03 | 0,031 |
| | | NCM | Média | **0,892** | 0,883§ | 0,885§ |
| | F & M | MOS-LQO | Média | 1,48§ | **1,55**§ | **1,55**§ |
| | | | Mediana | 1,46 | 1,565 | 1,55 |
| | | | Variância | 0,02 | 0,03 | 0,03 |
| | | NCM | Média | 0,888 | 0,945§ | **0,948**§ |
| WGN & ICRA | F & M | MOS-LQO | Média | 1,72 | **1,88**§ | **1,88** § |
| | | | Mediana | 1,75 | 1,92 | 1,92 |
| | | | Variância | 0,08 | 0,15 | 0,16 |
| | | NCM | Média | **0,887**§ | 0,883§ | 0,881§ |

Tabelas 9, 10, 11 e 12, respectivamente. E os resultados de qualidade e inteligibilidade obtidos por essas configurações são apresentados nas Tabelas 13, 14, 15 e 16.

Ao analisar as configurações de ACM e CM para os quatro níveis de SNR, notou-se que o parâmetro $\beta_1$ de ACM e CM tende a ser maior ou igual 0,15, indicando que o comportamento exponencial da máscara para ganhos inferiores a $g$ deve possuir uma subida com maior inclinação para maximizar a inteligibilidade.

Para as SNRs de -10, 0 e 10 dB, percebe-se que o parâmetro $\beta_2$ da ACM oscila entre 0,1 e 0,3, enquanto o parâmetro $g$ apresenta valores menores que 0,5 e $\xi_0$ é geralmente negativo. Na CM, observa-se que $\xi_0$ tende a ser negativo, resultando em curvas deslocadas à esquerda (menor SNR).

Observa-se que, em relação aos valores de NCM, a ACM apresenta resultados

Tabela 6 – Resultados de MOS-LQO obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 0 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | MOS-LQO | Média | 2,25 | **2,49**§ | **2,49**§ |
| | | | Mediana | 2,24 | 2,47 | 2,47 |
| | | | Variância | 0,01 | 0,015 | 0,015 |
| | | NCM | Média | 0,937 | **0,972**§ | 0,971§ |
| | M | MOS-LQO | Média | 2,56 | **2,98**§ | **2,98**§ |
| | | | Mediana | 2,53 | 2,95 | 2,95 |
| | | | Variância | 0,014 | 0,03 | 0,03 |
| | | NCM | Média | 0,958§ | **0,983**§ | **0,983**§ |
| | F & M | MOS-LQO | Média | 2,40 | **2,73**§ | **2,73**§ |
| | | | Mediana | 2,41 | 2,74 | 2,73 |
| | | | Variância | 0,04 | 0,08 | 0,08 |
| | | NCM | Média | 0,948 | **0,977**§ | **0,977**§ |
| ICRA | F | MOS-LQO | Média | 1,72§ | **1,79**§ | **1,79**§ |
| | | | Mediana | 1,72 | 1,77 | 1,77 |
| | | | Variância | 0,01 | 0,01 | 0,01 |
| | | NCM | Média | **0,94**§ | 0,937§ | 0,937§ |
| | M | MOS-LQO | Média | 1,93 | **2,05**§ | 2,04§ |
| | | | Mediana | 1,92 | 2,04 | 2,03 |
| | | | Variância | 0,02 | 0,04 | 0,04 |
| | | NCM | Média | **0,961** | 0,955§ | 0,955§ |
| | F & M | MOS-LQO | Média | 1,82§ | **1,92**§ | 1,91§ |
| | | | Mediana | 1,80 | 1,91 | 1,91 |
| | | | Variância | 0,03 | 0,04 | 0,04 |
| | | NCM | Média | **0,95**§ | 0,945§ | 0,948§ |
| WGN & ICRA | F & M | MOS-LQO | Média | 2,11§ | **2,32**§ | 2,31 § |
| | | | Mediana | 2,15 | 2,37 | 2,31 |
| | | | Variância | 0,12 | 0,23 | 0,23 |
| | | NCM | Média | 0,949 | 0,963§ | **0,964**§ |

iguais ou superiores em comparação com a CM, como esperado. Entretanto na média, os ganhos de desempenho da ACM em relação à CM são menores que 0,2%. Os valores de mediana e variância seguem um comportamento semelhante. As distinções costumam estar presentes na terceira casa decimal, levando a resultados sem diferenças estatísticas.

Observa-se que, para a SNR de 10 dB o valor resultante de NCM é basicamente o mesmo para as três máscaras analisadas. Destaca-se também que maximizando NCM, temos uma tendência de diminuição da qualidade máxima que pode ser obtida. Dessa forma, a WM apresentou valores de MOS-LQO superiores aos da CM e ACM para SNR de 10 e 20 dB.

Vale a pena destacar que, ao comparar a diferença do NCM entre ACM e CM

Tabela 7 – Resultados de MOS-LQO obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 10 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | MOS-LQO | Média | 3,26 | **3,49**§ | 3,47§ |
| | | | Mediana | 3,25 | 3,47 | 3,48 |
| | | | Variância | 0,006 | 0,01 | 0,01 |
| | | NCM | Média | 0,986§ | **0,988**§ | **0,988**§ |
| | M | MOS-LQO | Média | 3,56 | **3,81**§ | **3,81**§ |
| | | | Mediana | 3,57 | 3,82 | 3,82 |
| | | | Variância | 0,014 | 0,018 | 0,018 |
| | | NCM | Média | **0,988**§ | 0,987§ | 0,986§ |
| | F & M | MOS-LQO | Média | 3,40 | **3,64**§ | 3,63§ |
| | | | Mediana | 3,37 | 3,60 | 3,59 |
| | | | Variância | 0,03 | 0,04 | 0,04 |
| | | NCM | Média | 0,987§ | **0,989**§ | 0,987§ |
| ICRA | F | MOS-LQO | Média | 2,49§ | **2,57**§ | **2,57**§ |
| | | | Mediana | 2,49 | 2,58 | 2,58 |
| | | | Variância | 0,014 | 0,02 | 0,02 |
| | | NCM | Média | **0,982**§ | 0,981§ | 0,981§ |
| | M | MOS-LQO | Média | 2,88§ | **3,03**§ | 3,02§ |
| | | | Mediana | 2,88 | 3,05 | 3,05 |
| | | | Variância | 0,02 | 0,03 | 0,03 |
| | | NCM | Média | **0,988**§ | 0,987§ | 0,987§ |
| | F & M | MOS-LQO | Média | 2,69§ | **2,79**§ | **2,79**§ |
| | | | Mediana | 2,67 | 2,78 | 2,78 |
| | | | Variância | 0,06 | 0,08 | 0,08 |
| | | NCM | Média | **0,985**§ | 0,984§ | 0,984§ |
| WGN & ICRA | F & M | MOS-LQO | Média | 3,05§ | **3,21**§ | **3,21**§ |
| | | | Mediana | 3,13 | 3,32 | 3,32 |
| | | | Variância | 0,17 | 0,24 | 0,24 |
| | | NCM | Média | **0,986**§ | **0,986**§ | **0,986**§ |

($NCM_{ACM} - NCM_{CM}$), para todas as amostras na configuração de WGN & ICRA com falas masculinas e femininas, observamos que a média da distribuição tende a ser positiva. Esses resultados podem ser visualizados na forma de diagrama de caixas na Figura 10. Note que à medida que a SNR aumenta a diferença tende a zero. Isso indica que a CM quando bem configurada se aproxima da ACM.

## 3.4 CONCLUSÃO

Pode-se concluir que a ACM apresenta resultados iguais ou levemente superiores às demais máscaras encontradas na literatura para redução de ruído em sinais de fala. Isso indica o possível limite de supressão de ruído passível de obtenção através da técnica máscara tempo-frequência. A CM produz resultados muito próximos à

Tabela 8 – Resultados de MOS-LQO obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 20 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | MOS-LQO | Média | 3,99§ | **4,07**§§1 | 4,06§§1 |
| | | | Mediana | 3,98 | 4,06 | 4,05 |
| | | | Variância | 0,004 | 0,04 | 0,04 |
| | | NCM | Média | **0,991**§ | 0,99§ | 0,99§ |
| | M | MOS-LQO | Média | 4,09§ | **4,16**§ | **4,16**§ |
| | | | Mediana | 4,11 | 4,2 | 4,19 |
| | | | Variância | 0,07 | 0,08 | 0,08 |
| | | NCM | Média | **0,99**§ | 0,988§ | 0,988§ |
| | F & M | MOS-LQO | Média | 4,04§ | **4,11**§§1 | 4,10§§1 |
| | | | Mediana | 4,04 | 4,11 | 4,11 |
| | | | Variância | 0,007 | 0,007 | 0,008 |
| | | NCM | Média | **0,991** | 0,988§ | 0,988§ |
| ICRA | F | MOS-LQO | Média | 3,54§ | **3,6**§ | 3,58§ |
| | | | Mediana | 3,5 | 3,55 | 3,55 |
| | | | Variância | 0,014 | 0,016 | 0,013 |
| | | NCM | Média | **0,991**§ | **0,991**§ | **0,991**§ |
| | M | MOS-LQO | Média | 3,91§ | **3,97**§ | 3,96§ |
| | | | Mediana | 3,94 | 3,98 | 3,97 |
| | | | Variância | 0,009 | 0,01 | 0,01 |
| | | NCM | Média | **0,991**§ | **0,991**§ | **0,991**§ |
| | F & M | MOS-LQO | Média | 3,72§ | **3,79**§ | 3,78§ |
| | | | Mediana | 3,76 | 3,79 | 3,78 |
| | | | Variância | 0,04 | 0,04 | 0,04 |
| | | NCM | Média | **0,991**§ | **0,991**§ | **0,991**§ |
| WGN & ICRA | F & M | MOS-LQO | Média | 3,88§ | **3,94**§ | 3,93§ |
| | | | Mediana | 3,96 | 4 | 4 |
| | | | Variância | 0,05 | 0,05 | 0,05 |
| | | NCM | Média | **0,991**§ | 0,99§ | 0,99§ |

ACM com parâmetros otimizados, mostrando que curvas de ganho simétricas, quando bem configuradas, alcançam resultados próximos do maior desempenho encontrado. Embora haja possibilidade de otimizar a CM para cada caso, os parâmetros $\beta$ e $\xi_0$ na maximização de MOS-LQO não apresentam grande variabilidade ao considerar apenas o fator gênero. Também foi possível observar um deslocamento $\xi_0$ para o eixo positivo quando realizada a maximização para qualidade, enquanto para inteligibilidade ocorre um deslocamento para o lado oposto. Ademais, consta-se que para a inteligibilidade, os ganhos de desempenho foram pequenos devido o NCM estar próximo ao limite de escala.

Figura 9 – Diagramas de caixas das diferenças de MOS-LQO entre ACM e CM

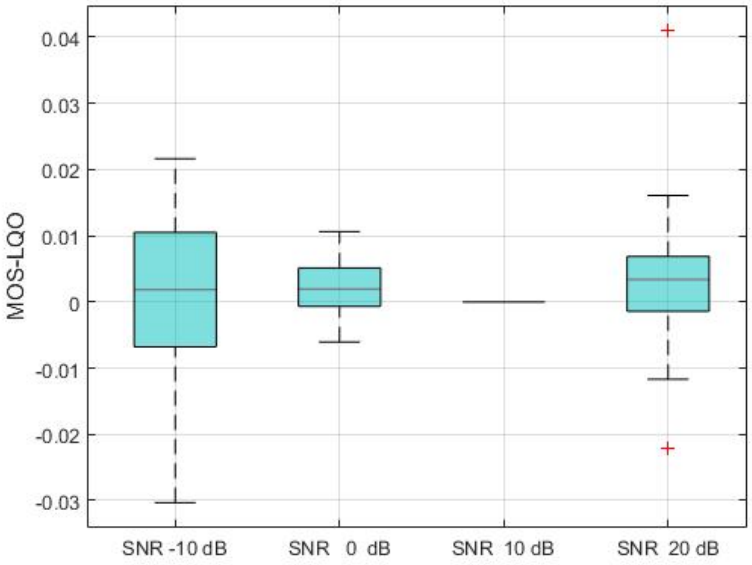



Figura 10 – Diagramas de caixas das diferenças de NCM entre ACM e CM

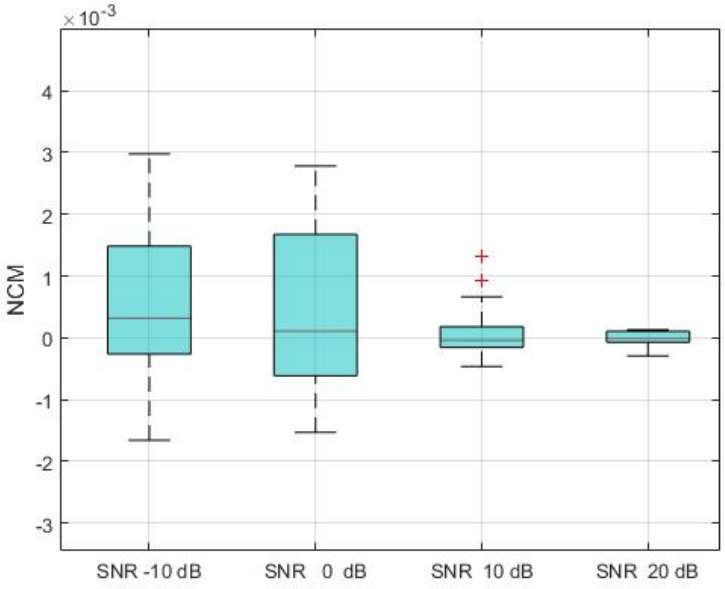


Tabela 9 – Configurações da ACM e CM que maximizam NCM para SNR de -10 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,15 | 0,4 | -6 | 0,15 | -3 |
| | M | 0,2 | 0,15 | 0,4 | -7 | 0,2 | -7 |
| | F & M | 0,2 | 0,1 | 0,2 | -12 | 0,15 | -3 |
| ICRA | F | 0,4 | 0,1 | 0,2 | -11 | 0,15 | -3 |
| | M | 0,9 | 0,1 | 0,2 | -11 | 0,2 | -3 |
| | F & M | 0,35 | 0,1 | 0,2 | -11 | 0,2 | -3 |
| WGN & ICRA | F & M | 0,2 | 0,1 | 0,2 | -12 | 0,15 | -4 |

Tabela 10 – Configurações da ACM e CM que maximizam NCM para SNR de 0 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,15 | 0,15 | 0,4 | -6 | 0,15 | -4 |
| | M | 0,2 | 0,15 | 0,4 | -8 | 0,2 | -6 |
| | F & M | 0,2 | 0,1 | 0,3 | -10 | 0,15 | -4 |
| ICRA | F | 0,35 | 0,1 | 0,2 | -11 | 0,15 | -3 |
| | M | 0,45 | 0,1 | 0,1 | -12 | 0,2 | -3 |
| | F & M | 0,45 | 0,1 | 0,2 | -11 | 0,2 | -4 |
| WGN & ICRA | F & M | 0,2 | 0,1 | 0,3 | -9 | 0,15 | -3 |

Tabela 11 – Configurações da ACM e CM que maximizam NCM para SNR de 10 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,35 | 0,15 | 0,3 | -10 | 0,25 | -6 |
| | M | 0,5 | 0,3 | 0,4 | -9 | 0,4 | -7 |
| | F & M | 0,5 | 0,15 | 0,2 | -12 | 0,3 | -6 |
| ICRA | F | 0,45 | 0,15 | 0,5 | -3 | 0,25 | -1 |
| | M | 0,25 | 0,2 | 0,2 | 5 | 0,35 | -1 |
| | F & M | 0,5 | 0,15 | 0,2 | -12 | 0,3 | -6 |
| WGN & ICRA | F & M | 0,15 | 0,15 | 0,4 | -1 | 0,15 | 1 |

Tabela 12 – Configurações da ACM e CM que maximizam NCM para SNR de 20 dB.

| Especificação | | ACM | | | | ACM = CM | |
|---|---|---|---|---|---|---|---|
| Ruído | Sexo | $\beta_1$ | $\beta_2$ | $g$ | $\xi_0$ (dB) | $\beta_1 = \beta_2$ | $\xi_0$ (dB) |
| WGN | F | 0,4 | 0,5 | 0,6 | -1 | 0,45 | -2 |
| | M | 0,45 | 0,05 | 0,7 | -3 | 0,7 | 0 |
| | F & M | 0,35 | 0,4 | 0,6 | -2 | 0,35 | -3 |
| ICRA | F | 0,1 | 0,2 | 0,2 | -1 | 0,3 | 3 |
| | M | 0,1 | 0,25 | 0,2 | 0 | 0,25 | 5 |
| | F & M | 0,1 | 0,3 | 0,3 | 0 | 0,4 | 2 |
| WGN & ICRA | F & M | 0,35 | 0,3 | 0,5 | 1 | 0,3 | 1 |

Tabela 13 – Resultados de NCM obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR -10 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | NCM | Média | 0,876 | **0,914**§ | 0,913§ |
| | | | Mediana | 0,877 | 0,920 | 0,920 |
| | | | Variância | $2{,}5\cdot10^{-4}$ | $1{,}04\cdot10^{-4}$ | $1{,}03\cdot10^{-4}$ |
| | | MOS-LQO | Média | 1,84 | 1,95§ | **1,98**§ |
| | M | NCM | Média | 0,893 | **0,920**§ | **0,920**§ |
| | | | Mediana | 0,895 | 0,921 | 0,921 |
| | | | Variância | $2{,}093\cdot10^{-4}$ | $4{,}08\cdot10^{-5}$ | $4{,}82\cdot10^{-5}$ |
| | | MOS-LQO | Média | 2,06 | **2,33**§ | 2,26§ |
| | F & M | NCM | Média | 0,886 | **0,918**§ | 0,917§ |
| | | | Mediana | 0,885 | 0,920 | 0,920 |
| | | | Variância | $2{,}951\cdot10^{-4}$ | $8{,}521\cdot10^{-5}$ | $6{,}873\cdot10^{-5}$ |
| | | MOS-LQO | Média | 1,95 | 2,17§ | **2,19**§ |
| ICRA | F | NCM | Média | 0,885 | **0,888**§ | 0,887§ |
| | | | Mediana | 0,881 | 0,895 | 0,894 |
| | | | Variância | $6{,}28\cdot10^{-4}$ | $4{,}22\cdot10^{-5}$ | $4{,}26\cdot10^{-4}$ |
| | | MOS-LQO | Média | **1,39**§ | 1,37§ | 1,37§ |
| | M | NCM | Média | 0,892§ | **0,898**§ | **0,898**§ |
| | | | Mediana | 0,894 | 0,895 | 0,895 |
| | | | Variância | $4{,}32\cdot10^{-4}$ | $3{,}22\cdot10^{-4}$ | $3{,}22\cdot10^{-4}$ |
| | | MOS-LQO | Média | 1,54§ | 1,50§ | **1,56**§ |
| | F & M | NCM | Média | 0,888§ | **0,895**§ | **0,895**§ |
| | | | Mediana | 0,892 | 0,895 | 0,895 |
| | | | Variância | $4{,}34\cdot10^{-4}$ | $2{,}96\cdot10^{-4}$ | $3{,}22\cdot10^{-4}$ |
| | | MOS-LQO | Média | 1,48§ | 1,47§ | **1,50**§ |
| WGN & ICRA | F & M | NCM | Média | 0,887 | **0,906**§ | 0,905§ |
| | | | Mediana | 0,888 | 0,907 | 0,908 |
| | | | Variância | $3{,}57\cdot10^{-4}$ | $3{,}23\cdot10^{-4}$ | $3{,}4\cdot10^{-5}$ |
| | | MOS-LQO | Média | 1,72 | **1,82**§ | **1,82**§ |

Tabela 14 – Resultados de NCM obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 0 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | NCM | Média | 0,937 | **0,978**§ | 0,977§ |
| | | | Mediana | 0,938 | 0,981 | 0,981 |
| | | | Variância | $2{,}8{\cdot}10^{-4}$ | $1{,}2{\cdot}10^{-4}$ | $1{,}2{\cdot}10^{-4}$ |
| | | MOS-LQO | Média | 2,25 | **2,40**§ | **2,40**§ |
| | M | NCM | Média | 0,958 | **0,987**§ | 0,986§ |
| | | | Mediana | 0,959 | 0,986 | 0,986 |
| | | | Variância | $1{,}8{\cdot}10^{-5}$ | $2{,}6{\cdot}10^{-5}$ | $2{,}4{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 2,56 | **2,89**§ | 2,86§ |
| | F & M | NCM | Média | 0,948 | **0,981**§ | **0,981**§ |
| | | | Mediana | 0,947 | 0,981 | 0,981 |
| | | | Variância | $3{,}3{\cdot}10^{-4}$ | $0{,}8{\cdot}10^{-4}$ | $0{,}9{\cdot}10^{-4}$ |
| | | MOS-LQO | Média | 2,40 | **2,66**§ | **2,66**§ |
| ICRA | F | NCM | Média | 0,94§ | **0,95**§ | 0,94§ |
| | | | Mediana | 0,946 | 0,95 | 0,95 |
| | | | Variância | $7{,}1{\cdot}10^{-5}$ | $4{,}8{\cdot}10^{-5}$ | $4{,}8{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **1,72**§ | 1,70§ | 1,70§ |
| | M | NCM | Média | 0,961§ | **0,966**§ | 0,965§ |
| | | | Mediana | 0,966 | 0,966 | 0,966 |
| | | | Variância | $8{,}2{\cdot}10^{-5}$ | $8{,}1{\cdot}10^{-5}$ | $8{,}1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 1,93§ | 1,90§ | **1,96**§ |
| | F & M | NCM | Média | 0,95§ | **0,958**§ | 0,957§ |
| | | | Mediana | 0,947 | 0,957 | 0,957 |
| | | | Variância | $5{,}0{\cdot}10^{-4}$ | $3{,}5{\cdot}10^{-4}$ | $3{,}4{\cdot}10^{-4}$ |
| | | MOS-LQO | Média | 1,82§ | 1,79§ | **1,82**§ |
| WGN & ICRA | F & M | NCM | Média | 0,949 | **0,969**§ | **0,969**§ |
| | | | Mediana | 0,95 | 0,97 | 0,97 |
| | | | Variância | $4{,}0{\cdot}10^{-5}$ | $3{,}6{\cdot}10^{-5}$ | $3{,}4{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 2,11§ | 2,25§ | **2,26**§ |

Tabela 15 – Resultados de NCM obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 10 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | NCM | Média | 0,986§ | **0,991**§ | 0,99§ |
| | | | Mediana | 0,988 | 0,992 | 0,991 |
| | | | Variância | $6{\cdot}10^{-5}$ | $1{,}5{\cdot}10^{-5}$ | $1{,}6{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 3,26§ | **3,27**§ | 3,25§ |
| | M | NCM | Média | 0,988§ | **0,991**§ | **0,991**§ |
| | | | Mediana | 0,989 | 0,993 | 0,993 |
| | | | Variância | $2{,}8{\cdot}10^{-5}$ | $1{,}8{\cdot}10^{-5}$ | $2{,}0{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 3,56§ | **3,57**§ | **3,57**§ |
| | F & M | NCM | Média | 0,987 | **0,991**§ | **0,991**§ |
| | | | Mediana | 0,988 | 0,99 | 0,99 |
| | | | Variância | $4{,}6{\cdot}10^{-5}$ | $1{,}5{\cdot}10^{-5}$ | $2{,}1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **3,41**§ | 3,40§ | 3,40§ |
| ICRA | F | NCM | Média | 0,982§ | **0,983**§ | 0,982§ |
| | | | Mediana | 0,988 | 0,988 | 0,988 |
| | | | Variância | $2{\cdot}10^{-5}$ | $1{,}9{\cdot}10^{-5}$ | $1{,}8{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **2,49**§ | 2,44§ | 2,48§ |
| | M | NCM | Média | **0,988**§ | **0,988**§ | **0,988**§ |
| | | | Mediana | 0,989 | 0,989 | 0,989 |
| | | | Variância | $1{,}2{\cdot}10^{-5}$ | $1{,}1{\cdot}10^{-5}$ | $1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **2,88** | 2,70§ | 2,63§ |
| | F & M | NCM | Média | 0,985 § | **0,986**§ | **0,986**§ |
| | | | Mediana | 0,988 | 0,988 | 0,988 |
| | | | Variância | $1{,}0{\cdot}10^{-4}$ | $1{,}0{\cdot}10^{-4}$ | $1{,}0{\cdot}10^{-4}$ |
| | | MOS-LQO | Média | **2,69**§ | 2,60§ | 2,60§ |
| WGN & ICRA | F & M | NCM | Média | 0,986§ | **0,988**§ | **0,988**§ |
| | | | Mediana | 0,986 | 0,988 | 0,988 |
| | | | Variância | $7{,}6{\cdot}10^{-5}$ | $6{,}6{\cdot}10^{-5}$ | $6{,}6{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | 3,05§ | **3,1**§ | 3,09§ |

Tabela 16 – Resultados de NCM obtidos pela máscara WM e pela melhor configuração das máscaras CM e ACM para SNR 20 dB.

| Especificação | | Estatística | | WM | ACM | CM |
|---|---|---|---|---|---|---|
| Ruído | Sexo | Métrica | Medidas | | | |
| WGN | F | NCM | Média | 0,991§ | **0,992**§ | **0,992**§ |
| | | | Mediana | 0,992 | 0,993 | 0,993 |
| | | | Variância | $8{\cdot}10^{-5}$ | $1{,}1{\cdot}10^{-5}$ | $1{,}1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **3,99** | 3,88§ | 3,87§ |
| | M | NCM | Média | 0,99§ | **0,991**§ | **0,991**§ |
| | | | Mediana | 0,992 | 0,993 | 0,993 |
| | | | Variância | $1{,}3{\cdot}10^{-5}$ | $2{,}1{\cdot}10^{-5}$ | $2{,}1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **4,09** | 3,09§ | 3,09§ |
| | F & M | NCM | Média | 0,99§ | **0,992**§ | **0,992**§ |
| | | | Mediana | 0,991 | 0,992 | 0,992 |
| | | | Variância | $1{\cdot}10^{-5}$ | $1{\cdot}10^{-5}$ | $1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **4,04** | 3,90§ | 3,90§ |
| ICRA | F | NCM | Média | 0,991§ | **0,992**§ | 0,991§ |
| | | | Mediana | 0,991 | 0,994 | 0,994 |
| | | | Variância | $1{,}9{\cdot}10^{-5}$ | $2{,}5{\cdot}10^{-5}$ | $2{,}4{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **3,54** | 3,39§ | 3,37§ |
| | M | NCM | Média | 0,991§ | **0,992**§ | 0,991§ |
| | | | Mediana | 0,991 | 0,993 | 0,993 |
| | | | Variância | $1{\cdot}10^{-5}$ | $1{\cdot}10^{-5}$ | $1{,}1{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **3,91** | 3,74§ | 3,67 § |
| | F & M | NCM | Média | 0,991§ | **0,992**§ | 0,991§ |
| | | | Mediana | 0,991 | 0,993 | 0,993 |
| | | | Variância | $1{,}4{\cdot}10^{-4}$ | $1{,}4{\cdot}10^{-4}$ | $1{,}4{\cdot}10^{-4}$ |
| | | MOS-LQO | Média | **3,72** | 3,56§ | 3,56§ |
| WGN & ICRA | F & M | NCM | Média | 0,991§ | **0,992**§ | 0,991 § |
| | | | Mediana | 0,991 | 0,993 | 0,993 |
| | | | Variância | $1{,}2{\cdot}10^{-5}$ | $1{,}5{\cdot}10^{-5}$ | $1{,}5{\cdot}10^{-5}$ |
| | | MOS-LQO | Média | **3,88** | 3,75§ | 3,74§ |

# 4 PARÂMETROS ÓTIMOS PARA A MÁSCARA CONFORMÁVEL

No capítulo 3, mostrou-se que, se corretamente configurada, a CM resulta em, valores de inteligibilidade e qualidade muito próximos ao máximo desempenho possível obtido por máscaras tempo-frequência. Este capítulo tem como objetivo investigar as configurações de parâmetros da CM que possam otimizar a qualidade da fala processada, visto que valores próximos ao topo da escala NCM (inteligibilidade) são naturalmente alcançados. Para tanto, foram realizadas simulações para determinar os parâmetros ótimos em diversos cenários acústicos. Comparações com a WM são também apresentadas.

## 4.1 METODOLOGIA DAS SIMULAÇÕES

A maximização da qualidade para a CM é realizada de três formas. Na primeira, buscam-se pares ($\mu$, $\gamma$) que resultem no máximo desempenho para cada tipo de ruído em um determinado nível de SNR, os quais resultam na máscaras CM ótima por ruído. Na segunda forma buscam-se pares ($\mu$, $\gamma$) que maximizam o desempenho da CM de acordo com o nível de SNR, ou seja, as configurações que apresentam os melhores resultados independentemente do ruído, resultando nas máscaras CM ótimas por SNR. E por último, busca-se o par ($\mu$, $\gamma$) único que resulte o melhor resultado na média, considerando todos os ruídos e SNRs, definido como CM ótima global.

### 4.1.1 Sinais de fala e ruído

Foram utilizadas 720 frases do corpus IEEE produzidas por um locutor masculino, descritas em Rothauser (1969). Esse conjunto de frases é foneticamente balanceado para a língua inglesa, sendo que cada uma das sentenças possui duração de aproximadamente 6 s. As falas foram amostradas a uma taxa de 16 kHz e contaminadas com 5 ruídos diferentes, sendo eles: ruído de balbúrdia (LOIZOU, 2013); ruído dentro de vagão de trem em movimento (LOIZOU, 2013); ruído de fala-sobre-fala usando o sinal ISTS (do inglês International Speech Test Signal ) (HOLUBE et al., 2010); ruído com espectro semelhante à fala usando o sinal ICRA (DRESCHLER et al., 2001); e WGN. A contaminação foi realizada com 3 níveis de SNR: -10 dB, 0 dB e 5 dB.

### 4.1.2 Implementação das máscaras

As STFTs $Y(k,\lambda)$, $X(k,\lambda)$ e $V(k,\lambda)$ foram calculadas utilizando uma janela de Hamming com duração de 20 ms, sobreposição de 50 % e transformada discreta de Fourier (DFT, do inglês Fourier Transform) de 320 pontos. Para cada janela $\lambda$, uma

estimativa de $\xi(k,\lambda)$ foi obtida como

$$\hat{\xi}(k,\lambda) = \frac{|X(k,\lambda)|^2}{|V(k,\lambda)|^2}, \quad k = 1,2,\ldots,320, \tag{14}$$

e utilizada para computar o ganho $M(k,\lambda)$ de diferentes máscaras. O ganho $M(k,\lambda)$ foi por sua vez multiplicado por $Y(k,\lambda)$, conforme (2), obtendo a estimativa $\hat{X}(k,\lambda)$. Por fim, a estimativa $\hat{x}(n)$ do sinal de fala limpo foi construída utilizando as STFTs inversas de $\hat{X}(k,\lambda)$ e a técnica de sobreposição-e-soma.

Para se ter uma boa representação dos valores de $\mu$ e $\gamma$ ótimos para cada cenário, o parâmetro $\mu$ foi variado de -30 dB até 30 dB com passos de 1 dB, e $\gamma$ variado de 0,5 até 1 em 5 passos distribuídos igualmente na escala decibel, sendo eles [0,5 0,57 0,66 0,76 0,87 1], intervalo esse obtido por simulações extensivas preliminares que indicaram a região de melhor desempenho. Este conjunto de parâmetros resultou em 366 curvas de ganho para cada tipo de ruído e nível de SNR.

### 4.1.3 Cenários de análise

Realizaram-se simulações para a determinação dos parâmetros ótimos da CM, com três objetivos distintos: CM ótima por ruído, CM ótima por SNR e CM ótima global. Utilizaram-se as distribuições estatísticas obtidas em cada situação para avaliar o resultado do MOS-LQO.

### 4.1.4 Testes estatísticos

O teste de Jarque-Bera ($\rho < 0{,}05$), para verificar se os dados possuem uma distribuição normal (JARQUE; BERA, 1987), foi aplicado nos valores MOS-LQO produzidos pelas máscaras CM e WM para cada combinação de ruído e SNR. A normalidade dos dados foi confirmanda em todos os casos. Então, para cada combinação de ruído e SNR, uma análise de variância (ANOVA) ($\rho < 0{,}05$) foi realizada para verificar se os valores de MOS-LQO obtidos pela máscara CM configurada nas três situações (ótima por ruído, SNR e global) e WM são oriundos de distribuições diferentes (HOGG; LEDOLTER, 1987). Quando as distribuições não apresentaram diferenças estatisticamente significativas, o símbolo § foi utilizado junto aos valores médios correspondentes.

## 4.2 RESULTADOS E DISCUSSÃO

Esta seção descreve os resultados obtidos para as três otimizações de CM em termos de MOS-LQO, independente da inteligibilidade. Os parâmetros ótimos em cada situação são apresentados na Tabela 17, assumindo o par de parâmetros que apresentou o maior valor médio. Os valores de MOS-LQO são apresentados nas Tabelas 18, 19 e 20.

Tabela 17 – Valores dos parâmeros $\gamma$ e $\mu$ ótimos para cada ruído, todos os ruídos, e a configuração da CM ótima global.

| Ruído | SNR | $\gamma$ | $\mu$ (dB) |
|---|---|---|---|
| WGN | -10dB | 0,66 | -1 |
| | 0dB | 0,66 | -2 |
| | 5dB | 0,66 | -1 |
| Balbúrdia | -10dB | 0,57 | 10 |
| | 0dB | 0,57 | 5 |
| | 5dB | 0,66 | 0 |
| ICRA | -10dB | 0,57 | 13 |
| | 0dB | 0,66 | 3 |
| | 5dB | 0,66 | 2 |
| Trem | -10dB | 0,57 | 24 |
| | 0dB | 0,5 | 22 |
| | 5dB | 0,5 | 14 |
| ISTS | -10dB | 0,66 | 1 |
| | 0dB | 0,66 | 1 |
| | 5dB | 0,66 | 2 |
| Todos | -10dB | 0,57 | 10 |
| | 0dB | 0,66 | 2 |
| | 5dB | 0,66 | 1 |
| Global | | 0,66 | 2 |

Ao analisar a Tabela 17, nota-se que $\gamma$ apresenta valores que aproximam a CM à WR (ao invés da WM), de acordo com os parâmetros observados na seção 2.5.5, principalmente quando o nível de SNR é de -10 dB, enquanto que $\mu$, apresenta ampla faixa de valores positivos. Valores negativos de $\mu$ são encontrados apenas para o ruído WGN.

Analisando-se, os valores de MOS-LQO presentes nas Tabelas 18, 19 e 20, observa-se que a CM apresenta menor média de MOS-LQO para o ruído ICRA e a maior para o ruído ISTS, com exceção da SNR de 5 dB na qual é obtida uma média ligeiramente maior para o ruído WGN. Entretanto, ao compararmos o desempenho da CM em relação à WM verifica-se que, para WGN, CM resulta ganhos percentuais de MOS-LQO de 13,9% (SNR -10 dB), 12,1% (SNR 0 dB) e 8,8% (SNR 5 dB), enquanto que para o ruído ISTS resultou em 5,2% (SNR 0 dB) e 3,2% (SNR 5 dB), com exceção da SNR de -10 dB na qual o melhor resultado foi para o ruído ICRA com aumento de 9,3%.

Os parâmetros de configuração obtidos para cada caso da CM ótima por ruído resultou em maior valor médio que as demais situações. A solução para o caso da CM

Tabela 18 – Valores estatísticos do MOS-LQO obtidos pela máscara WM e pelas configurações ótima de CM para ruído, nível de SNR e Global, em SNR -10 dB.

| Ruído | Estatística | Wiener | CM ótima: p/ ruído | CM ótima: p/ SNR (*) | CM ótima: Global |
|---|---|---|---|---|---|
| WGN | Média | 1,86 | **2,12**§ | 2,09§ | **2,12**§ |
| | Mediana | 1,86 | 2,10 | 2,12 | 2,11 |
| | Variância | 0,005 | 0,013 | 0,012 | 0,012 |
| | Menor valor | 1,65 | 1,71 | 1,71 | 1,73 |
| Balbúrdia | Média | 1,62 | **1,80**§ | **1,80**§ | 1,78§ |
| | Mediana | 1,62 | 1,79 | 1,79 | 1,78 |
| | Variância | 0,008 | 0,019 | 0,019 | 0,017 |
| | Menor valor | 1,35 | 1,42 | 1,42 | 1,43 |
| ICRA | Média | 1,63 | **1,78**§ | 1,77§ | 1,76§ |
| | Mediana | 1,63 | 1,77 | 1,77 | 1,76 |
| | Variância | 0,01 | 0,021 | 0,021 | 0,019 |
| | Menor valor | 1,28 | 1,34 | 1,34 | 1,34 |
| Trem | Média | 1,77 | **1,98**§ | 1,90§ | 1,85 |
| | Mediana | 1,76 | 1,98 | 1,90 | 1,85 |
| | Variância | 0,022 | 0,03 | 0,029 | 0,026 |
| | Menor valor | 1,3 | 1,39 | 1,36 | 1,34 |
| ISTS | Média | 2,05 | **2,28**§ | 2,27§ | **2,28**§ |
| | Mediana | 2,06 | 2,29 | 2,27 | 2,29 |
| | Variância | 0,02 | 0,032 | 0,032 | 0,032 |
| | Menor valor | 1,53 | 1,66 | 1,58 | 1,65 |
| Todos | Média | 1,79 | **1,97**§ | | 1,96§ |
| | Mediana | 1,77 | 1,95 | | 1,93 |
| | Variância | 0,039 | 0,058 | | 0,062 |
| | Menor valor | 1,28 | 1,34 | | 1,34 |

Informativo: CM ótima por SNR, $\gamma$ = 0,57 e $\mu$ = 10,

ótima por SNR apresenta desempenho semelhante à CM ótima por ruído, exceto para ruído de trem em SNR -10 e 0 dB. Na Figura 11, o diagrama de caixas exibe WM em azul, a CM ótima por ruído em vermelho, a CM ótima por SNR em verde e CM ótima global em magenta. A partir dessa figura fica evidente a vantagem da otimização por tipo de ruído em relação às demais.

Uma comparação entre as diferentes formas de otimização da CM indica que as diferenças médias de MOS-LQO são menores que 0,02, exceto para o ruído de trem em SNR -10 dB, onde a diferença (com significância estatística), se aproxima de 0,13. Entretanto, esses valores ainda são superiores aos da WM.

Apesar das pequenas diferenças, a configuração global da CM resulta eficiente redução dos 5 tipos de ruídos. Na Figura 12, são apresentados os resultados obtidos, para todos os 5 ruídos, nos três níveis de SNRs. Verifica-se que praticamente não há diferença visual entre o diagrama de caixas da CM ótima por ruído e a CM ótima global, mas fica evidente o maior desempenho da CM se comparada com à WM.

As Figuras 13 e 14 demostram os valores médios de MOS-LQO para ruído WGN e ruído de trem em uma SNR de -10 dB, para cada combinação de $\gamma$ e $\mu$. Observa-se que diferentes combinações de $\gamma$ e $\mu$ se aproximam das configurações

Tabela 19 – Valores estatísticos do MOS-LQO obtidos pela máscara WM e pelas configurações ótima de CM para ruído, nível de SNR e Global, em SNR 0 dB.

| Ruído | Estatística | Wiener | CM ótima: p/ ruído | CM ótima: p/ SNR (*) | CM ótima: Global |
|---|---|---|---|---|---|
| WGN | Média | 2,41 | **2,70**§ | 2,68§ | 2,68§ |
| | Mediana | 2,41 | 2,70 | 2,68 | 2,68 |
| | Variância | 0,005 | 0,01 | 0,009 | 0,009 |
| | Menor valor | 2,14 | 2,27 | 2,35 | 2,35 |
| Balbúrdia | Média | 2,09 | **2,25**§ | **2,25**§ | **2,25**§ |
| | Mediana | 2,09 | 2,25 | 2,25 | 2,25 |
| | Variância | 0,007 | 0,014 | 0,013 | 0,013 |
| | Menor valor | 1,8 | 1,8 | 1,9 | 1,9 |
| ICRA | Média | 2,07 | **2,19**§ | **2,19**§ | **2,19**§ |
| | Mediana | 2,07 | 2,19 | 2,19 | 2,19 |
| | Variância | 0,008 | 0,015 | 0,015 | 0,015 |
| | Menor valor | 1,71 | 1,78 | 1,78 | 1,78 |
| Trem | Média | 2,13 | **2,28** | 2,22§ | 2,22§ |
| | Mediana | 2,12 | 2,28 | 2,22 | 2,22 |
| | Variância | 0,016 | 0,025 | 0,021 | 0,021 |
| | Menor valor | 1,70 | 1,67 | 1,71 | 1,71 |
| ISTS | Média | 2,59 | **2,73**§ | **2,73**§ | **2,73**§ |
| | Mediana | 2,60 | 2,74 | 2,74 | 2,74 |
| | Variância | 0,019 | 0,028 | 0,028 | 0,028 |
| | Menor valor | 2,08 | 2,16 | 2,16 | 2,16 |
| Todos | Média | 2,26 | **2,41**§ | | **2,41**§ |
| | Mediana | 2,19 | 2,34 | | 2,34 |
| | Variância | 0,054 | 0,073 | | 0,073 |
| | Menor valor | 1,70 | 1,71 | | 1,71 |

Informativo: CM ótima por SNR, $\gamma$ = 0,66 e $\mu$ = 2,

ótimas por ruído, SNR e global, estando destacadas por quadrado vermelho, círculo vermelho e círculo preto, respectivamente. Isso facilita a compreensão do porquê que combinações distintas produzem valores médios semelhantes ou sem diferença estatística significativa. Para o ruído de trem em SNR de -10 dB, que é o caso que apresenta maior diferença de média entre CM ótima por ruído e global, percebe-se que a otimização resulta em valores $\mu > 10$ dB, Porém, o valor médio de MOS-LQO não varia significativamente. Para as demais especificações de ruído e SNRs o nível de MOS-LQO médio assemelha-se ao do ruído WGN em SNR de -10 dB.

## 4.3 CONCLUSÃO

Neste capítulo foram apresentadas otimizações dos parâmetros da CM para diferentes combinações de tipos de ruídos e SNRs. Verificou-se que os processos de otimização por tipo de ruído e global resultam em pontuações MOS-LQO semelhantes, mas superiores às obtidas pela WM. Essa observação indica a possibilidade de utilização de parâmetros únicos para a CM para diversos tipos de ruído e SNRs, facilitando a implementação de sistemas práticos e evitando a necessidade de sistemas de classificação de ruído.

Tabela 20 – Valores estatísticos do MOS-LQO obtidos pela máscara WM e pelas configurações ótima de CM para ruído, nível de SNR e Global, em SNR 5 dB.

| Ruído | Estatística | Wiener | CM ótima: p/ ruído | CM ótima: p/ SNR (*) | CM ótima: Global |
|---|---|---|---|---|---|
| WGN | Média | 2,76 | **3,00**§ | **3,00**§ | 2,99§ |
| | Mediana | 2,77 | 3,01 | 3,00 | 2,99 |
| | Variância | 0,005 | 0,008 | 0,008 | 0,007 |
| | Menor valor | 2,50 | 2,65 | 2,72 | 2,71 |
| Balbúrdia | Média | 2,40 | **2,56**§ | **2,56**§ | 2,55§ |
| | Mediana | 2,41 | 2,56 | 2,56 | 2,56 |
| | Variância | 0,007 | 0,012 | 0,012 | 0,012 |
| | Menor valor | 2,14 | 2,22 | 2,23 | 2,23 |
| ICRA | Média | 2,37 | **2,48**§ | 2,47§ | **2,48**§ |
| | Mediana | 2,37 | 2,48 | 2,48 | 2,48 |
| | Variância | 0,008 | 0,014 | 0,014 | 0,014 |
| | Menor valor | 2,04 | 2,07 | 2,06 | 2,07 |
| Trem | Média | 2,41 | **2,53**§ | 2,50§ | 2,50§ |
| | Mediana | 2,41 | 2,53 | 2,50 | 2,50 |
| | Variância | 0,014 | 0,02 | 0,019 | 0,019 |
| | Menor valor | 2,00 | 2,01 | 2,05 | 2,06 |
| ISTS | Média | 2,89 | **2,98**§ | **2,98**§ | **2,98**§ |
| | Mediana | 2,89 | 2,98 | 2,98 | 2,98 |
| | Variância | 0,016 | 0,022 | 0,022 | 0,022 |
| | Menor valor | 2,44 | 2,45 | 2,44 | 2,45 |
| Todos | Média | 2,57 | **2,70**§ | | **2,70**§ |
| | Mediana | 2,49 | 2,63 | | 2,63 |
| | Variância | 0,056 | 0,07 | | 0,069 |
| | Menor valor | 2,00 | 2,05 | | 2,06 |

Informativo: CM ótima por SNR, $\gamma = 0{,}66$ e $\mu = 1$,

Figura 11 – Diagrama de caixas das máscaras: Wiener em azul, CM ótima por ruído em vermelho, CM ótima por SNR em verde e CM ótima global em magenta, para o ruído de trem.

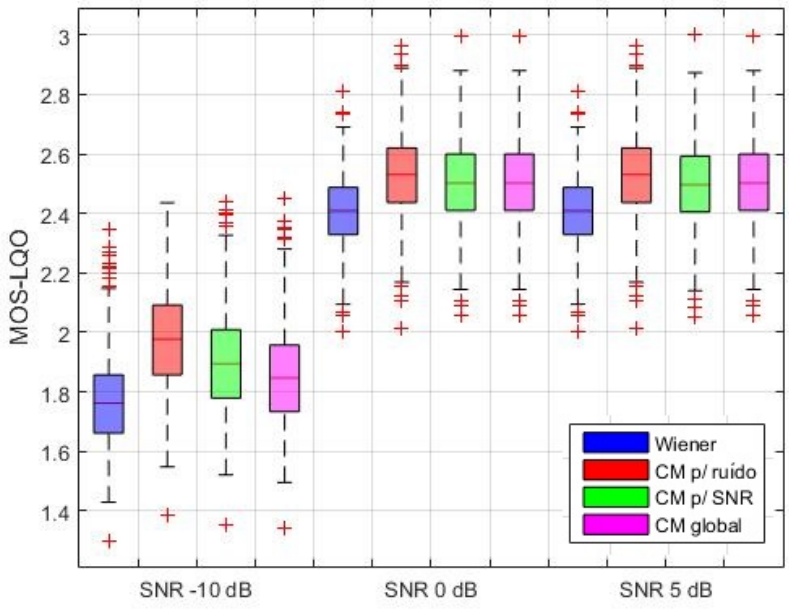


Figura 12 – Diagrama de caixas das máscaras: Wiener em azul, CM ótima por ruído em vermelho e CM ótima global em magenta, para todos os ruídos.

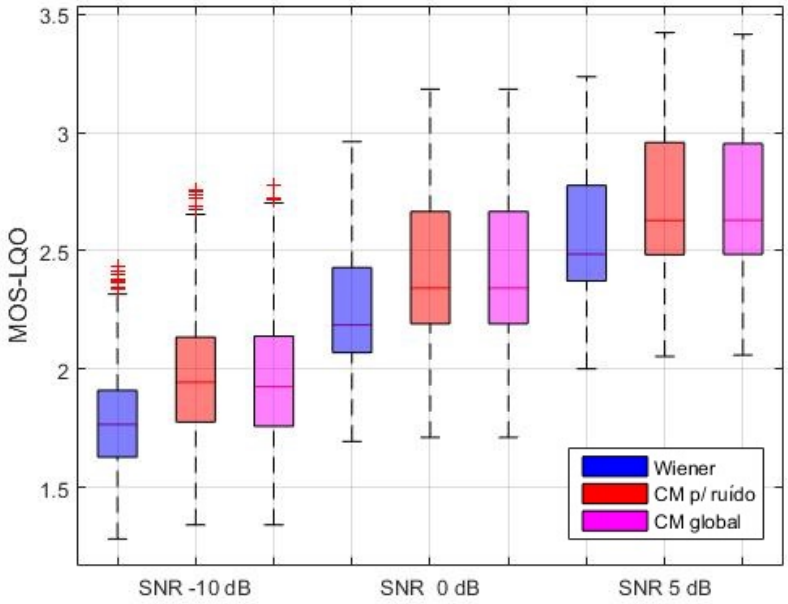



Figura 13 – Valor médio de MOS-LQO para WGN, em SNR -10 dB, para CM com diferentes configurações de $\gamma$ e $\mu$.

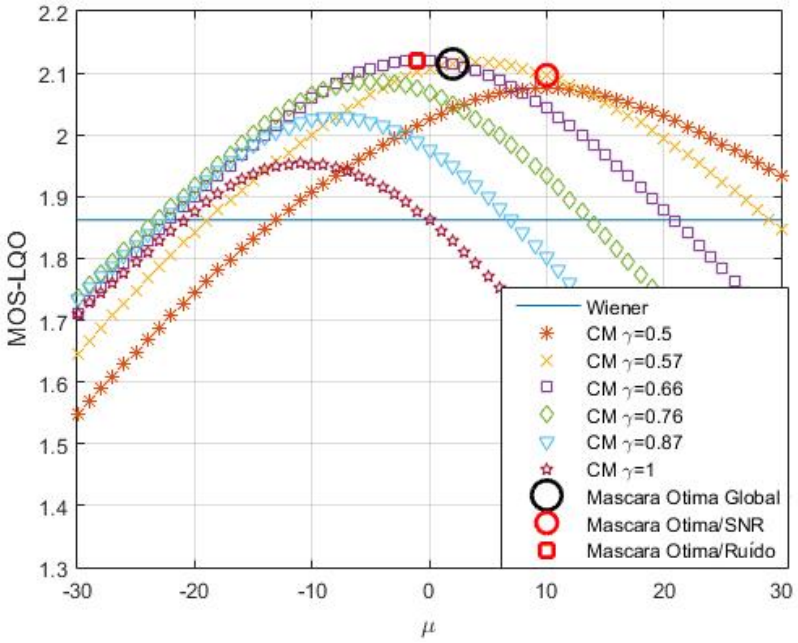
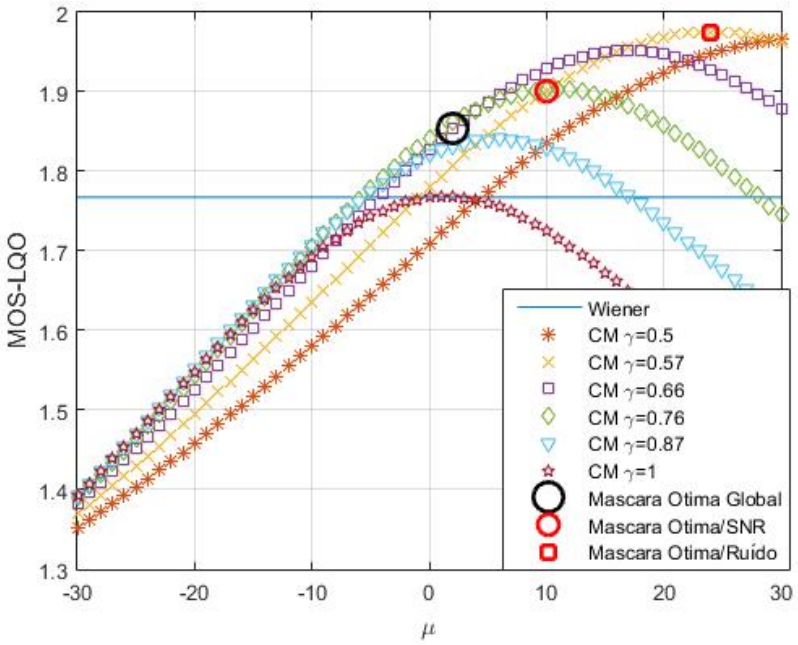


Figura 14 – Valor médio de MOS-LQO para ruído de trem, em SNR -10 dB, para CM com diferentes configurações de $\gamma$ e $\mu$.



# 5 UMA ESTRATÉGIA DE VARIAÇÃO DA MÁSCARA PARAMÉTRICA CONFORMÁVEL AO LONGO DO TEMPO

Neste capítulo, será investigada a existência de conformação diferenciada ou preferencial da curva de ganho da máscara CM, para locuções de vogais e consoantes. Em geral, as máscaras tempo-frequência são aplicadas com a mesma configuração de ganho para toda a frase ao longo do sinal de fala, implementadas de forma independente do tipo de locução, sofrendo alterações nos parâmetros de configuração apenas em situações de mudança de SNR ou a partir da sinalização de um classificador relativo ao tipo de ruído.

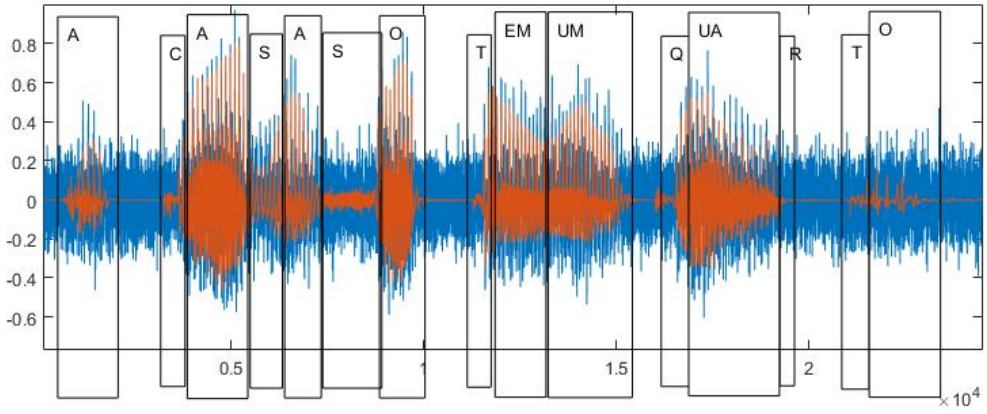
Algumas pesquisas têm examinado a contribuição das vogais e consoantes para a inteligibilidade da fala em ambientes silenciosos (OWREN; CARDILLO, 2006; KEWLEY-PORT; BURKLE; LEE, 2007). Em ambientes ruidosos, os ouvintes podem utilizar diferentes pistas acústicas e, portanto, a contribuição das vogais e consoantes pode produzir efeitos distintos na inteligibilidade da fala. Isso ocorre porque o ruído tende a mascarar as vogais e consoantes de maneira diferente. Geralmente, as consoantes apresentam baixa energia, tornando-as mais vulneráveis à interferência do ruído. Por outro lado, as vogais e semivogais possuem energia mais alta, o que pode ajudar na sua percepção em ambientes ruidosos (PHATAK; ALLEN, 2007).

Em geral, o reconhecimento de vogais em situações de baixa SNR é sempre ligeiramente superior ao reconhecimento de consoantes, mesmo em SNRs tão baixas quanto -20 dB. No trabalho de Phatak e Allen (2007), foi verificado que um grupo de consoantes (/f/,/$\theta$/,/v/,/$\eth$/,/b/ e /m/) apresentou pontuação de reconhecimento significativamente inferior (6,25%) a de qualquer vogal isolada (25%). Esses resultados destacam a importância das consoantes para a inteligibilidade da fala em situações de ruído e a necessidade de considerar a influência das diferentes classes de consoantes nesses cenários.

Li e Loizou (2008) verificaram que o impacto das informações oferecidas pelas consoantes obstrutivas na inteligibilidade da fala é notável. Considerando um cenário acústico no qual a inteligibilidade média é de 20%, a supressão dos trechos com consoantes diminui esse resultado para cerca de 10%. Entretanto, ao substituir as consoantes contaminadas pelas consoantes intactas, mantendo o restante da frase contaminada, a inteligibilidade aumenta para aproximadamente 70%.

Embora seja óbvio que substituir parte da frase corrompida pelo ruído por fala limpa aumente o entendimento, e até mesmo a qualidade do sinal, fica evidente que a contaminação das consoantes impacta diretamente na percepção da fala. Por isso, justifica-se a investigação da possibilidade de utilizar processamentos diferentes para vogais e consoantes, levando em consideração suas diferentes contribuições para a inteligibilidade da fala.

Figura 15 – Demarcação da frase "A casa só tem um quarto"




## 5.1 PROPOSTA DE VARIAÇÃO DA MÁSCARA PARAMÉTRICA CONFORMÁVEL AO LONGO DO TEMPO

Conforme discutido anteriormente, é possível perceber que a fala apresenta características diferentes para vogais e consoantes, indicando que períodos que contenham consoantes desempenham um papel importante para inteligibilidade da fala. Nota-se ainda que, em termos de melhorias de desempenho, há mais possibilidade de aumentar MOS-LQO do que NCM, uma vez que o segundo está muito próximo do limite máximo de escala.

Para buscar uma melhoria da fala, utilizou-se a máscara CM, definida pela equação (10), com duas configurações distintas. Na primeira, a máscara utiliza um par especificamente projetado para consoantes, definindo-a como $CM_C$, enquanto que na segunda situação, um outro par é utilizado para trechos de vogais isoladas e vogais com consoantes, sendo definida como $CM_V$.

Para exemplificar o problema abordado, as Figura 15, 16 e 17 apresentam três falas diferentes nas quais os períodos de existência de vogais e consoantes foram anotados manualmente, tendo sido contaminadas com ruído WGN a uma SNR de 0 dB. A fala original é apresentada em laranja e a contaminada em azul.

Note que as regiões de baixa energia da fala, caracterizadas principalmente por consoantes, são mais corrompidas pelo ruído, onde possuem SNR menor. Na Figura 18, podemos ver o nível de SNR para cada demarcação da frase "A casa só tem um quarto". É interessante notar que as demarcações com consoantes apresentam níveis de SNR negativos, enquanto as que contêm vogais apresentam níveis positivos. Com exceção do primeiro e último fonemas da frase, os demais apresentam comportamento semelhante.

Figura 16 – Demarcação da frase "A casa foi vendida sem pressa"

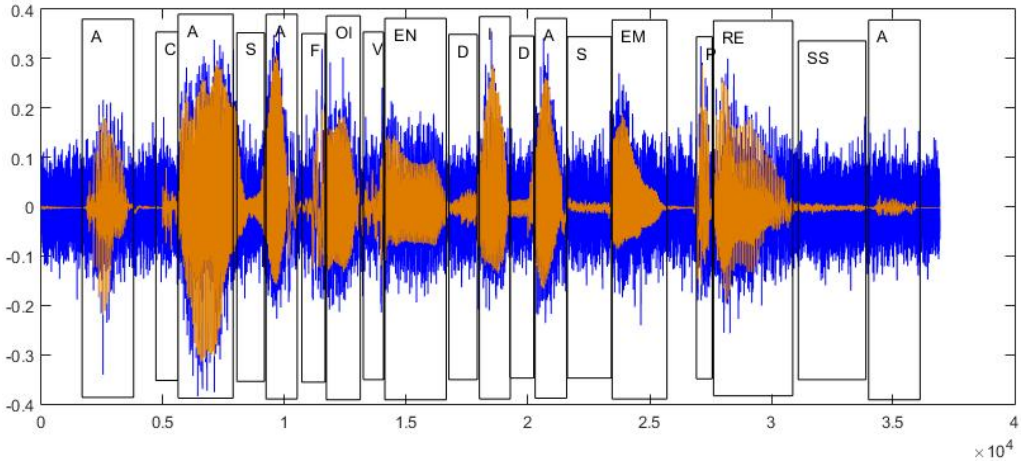
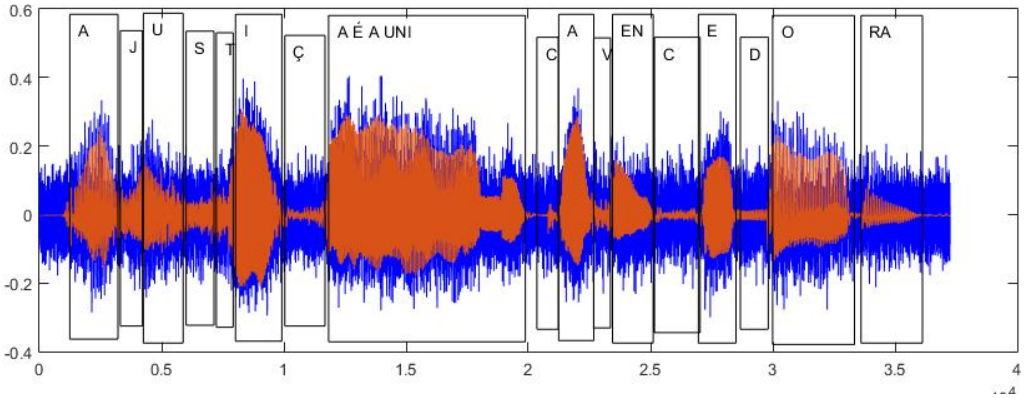



Figura 17 – Demarcação da frase "A justiça é a única vencedora"



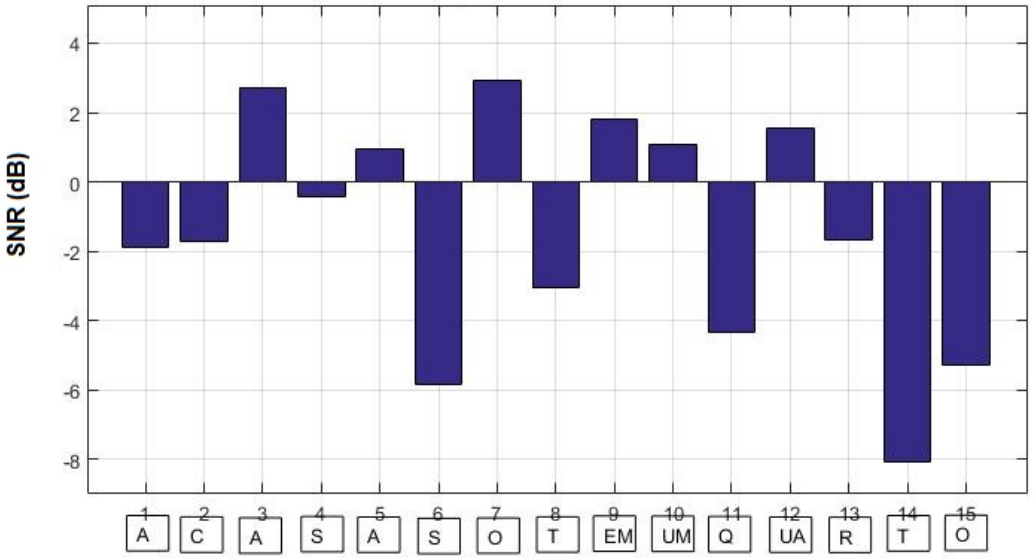

### 5.1.1 Sinais de fala e ruído

Foram utilizados três sinais de fala obtidos de Alcaim, Solewicz e Moraes (1992), disponíveis em Ynoguti (1999). Esses áudios são identificados neste trabalho pelas referências M1 (locutor masculino), F1 (locutor feminino) e F2 (locutor feminino), possuem duração de aproximadamente 6 s e foram amostrados a uma taxa de 16 kHz. Os sinais de fala foram contaminados com 5 tipos de ruídos, sendo eles: ruído de refeitório (LOIZOU, 2013); ruído dentro de vagão de trem em movimento (LOIZOU, 2013); ruído de fala-sobre-fala (HOLUBE et al., 2010); ruído com espectro semelhante à fala usando o sinal ICRA (DRESCHLER et al., 2001); e ruído WGN. A contaminação foi realizada em 2 níveis de SNR: -10 dB e 0 dB.

### 5.1.2 Implementação das máscaras

As STFTs $Y(k,\lambda)$, $X(k,\lambda)$ e $V(k,\lambda)$ foram calculadas utilizando uma janela de Hamming com duração de 20 ms, sobreposição de 50 % e uma transformada discreta

Figura 18 – SNR da demarcação da frase "A casa só tem um quarto", em condições de SNR média de 0 dB




de Fourier (DFT, do inglês Discrete Fourier Transform) de 320 pontos. Para cada janela $\lambda$, uma estimativa de $\xi(k,\lambda)$ foi obtida como

$$\hat{\xi}(k,\lambda) = \frac{|X(k,\lambda)|^2}{|V(k,\lambda)|^2}, \quad k = 1,2,\ldots,320, \tag{15}$$

e utilizada para computar o valor de ganho $M(k,\lambda)$ de diferentes máscaras. O ganho $M(k,\lambda)$ foi por sua vez aplicada a $Y(k,\lambda)$ conforme (2), obtendo a estimativa $\hat{X}(k,\lambda)$. Por fim, a estimativa $\hat{x}(n)$ do sinal de fala limpo foi construída utilizando a STFT inversa de $\hat{X}(k,\lambda)$ e a técnica de sobreposição-e-soma.

Para determinar a configuração ótima de parâmetros para os casos de vogais e consoantes foi utilizada uma grade de possibilidades para $\mu$ no intervalo de [-30, 30 ]dB com passos de 1 dB, e para $\gamma$ variando de [0,5, 1 ] em 5 passos distribuídos de forma igual em dBs, sendo eles [0,5 0,57 0,66 0,76 0,87 1].

### 5.1.3 Cenários de análise

Neste experimento, buscou-se a configuração combinada de $CM_C$ com $CM_V$ que resulta no maior valor de MOS-LQO, para cada uma das três falas. Embora se espere aumentar a inteligibilidade por meio da qualidade, a busca não leva em conta a métrica NCM. Nesta avaliação, é comparado $CM_C/CM_V$ com WM e a CM ótima global descrita no capítulo 4.

## 5.2 RESULTADOS E DISCUSSÃO

Esta seção descreve os resultados obtidos nos experimentos realizados para a maximização da qualidade utilizando a métrica MOS-LQO para $CM_C/CM_V$. Os melhores pares de $\mu$ e $\gamma$ para consoante, vogal ou vogal e consoante em cada uma das três falas para cada tipo de ruído, em níveis de SNR -10 e 0 dB são apresentados nas Tabelas 21 e 22. Os valores de MOS-LQO e NCM estão presentes nas Tabelas 23 e 24.

Percebe-se que em vários casos o parâmetro $\gamma$ é aparentemente o mesmo para $CM_C$ e $CM_V$. Note que geralmente seu valor é igual ou diferente por apenas um nível na escala, Por outro lado, o parâmetro $\mu$ apresenta maior variabilidade, sendo que em $CM_V$ praticamente todos são positivos, e em $CM_C$ é sempre menor que em $CM_V$. Isso indica que o comportamento exponencial de transição da máscara CM pode ser fixado em determinado valor $\gamma$ e apenas realizar variação em $\mu$. Vale a pena relembrar que $\mu$ está associado ao deslocamento do eixo da SNR, fornecendo indícios de que bastaria mover a curva de ganho para a esquerda em consoantes, onde SNR tende a ser negativa, e para a direita em vogais e vogais com consoantes onde as SNRs tendem a ser positivas.

Observa-se que os valores de MOS-LQO são superiores para a combinação $CM_C/CM_V$, mas não apresentam aumento de desempenho relevante para nenhum ruído, em nenhum dos dois níveis de SNR testados. Além disso, a maximização do MOS-LQO resultou em vários casos em uma perda da inteligibilidade, ao comparar-se com a máscara CM ótima global. É importante frisar que a separação de trechos de vogais e consoantes ao longo do sinal de fala não é uma tarefa simples, e portanto, sistemas automáticos estão sujeitos a erros de identificação. Então, se considerarmos o baixo ganho de qualidade, a dificuldade de seleção dos parâmetros de $CM_C$ e $CM_V$, e quando deve ser aplicada cada uma, a implementação de uma variação da máscara não parece apresentar resultados vantajosos.

## 5.3 CONCLUSÃO

Concluiu-se que a implementação de uma variação da máscara CM para trechos de consoantes e vogais ou vogais com consoante ($CM_C/CM_V$) aparentemente não apresenta resultados vantajosos. Os valores de MOS-LQO não apresentam aumentos relevantes para nenhum ruído, em nenhum dos níveis de SNR testados. Além disso, pode ocorrer uma perda da inteligibilidade em comparação à máscara CM ótima global. A dificuldade na separação de trechos de vogais e consoantes ao longo do sinal de fala torna desafiadora a implementação prática da estratégia de filtragem. Portanto, considerando o baixo aumento de qualidade e a dificuldade na seleção dos parâmetros, a implementação desta estratégia não apresenta benefícios significativos.

Tabela 21 – Configuração da $CM_C/CM_V$, para SNR de -10 dB.

| Especificação | | $CM_c$ | | $CM_v$ | |
|---|---|---|---|---|---|
| Ruído | Fala | $\gamma$ | $\mu$ [dB] | $\gamma$ | $\mu$ [dB] |
| WGN | M | 0,57 | 0 | 0,66 | 4 |
| | $F_1$ | 0,66 | -7 | 0,66 | 0 |
| | $F_2$ | 0,66 | -5 | 0,66 | 2 |
| Balbúrdia | M | 0,66 | -4 | 0,66 | 3 |
| | $F_1$ | 0,66 | 0 | 0,66 | 2 |
| | $F_2$ | 0,66 | -3 | 0,66 | 2 |
| ICRA | M | 0,66 | 6 | 0,66 | 13 |
| | $F_1$ | 0,76 | 17 | 0,87 | 3 |
| | $F_2$ | 0,76 | 15 | 0,87 | 2 |
| Trem | M | 0,57 | 20 | 0,57 | 23 |
| | $F_1$ | 0,76 | 17 | 0,57 | 24 |
| | $F_2$ | 0,57 | 16 | 0,57 | 23 |
| ISTS | M | 0,5 | 8 | 0,57 | 8 |
| | $F_1$ | 0,5 | 1 | 0,5 | 14 |
| | $F_2$ | 0,5 | 0 | 0,57 | 10 |
| Todos | | 0,87 | 11 | 0,5 | 28 |

Tabela 22 – Configuração da $CM_c/CM_v$,para SNR de 0 dB.

| Especificação | | $CM_c$ | | $CM_v$ | |
|---|---|---|---|---|---|
| Ruído | Fala | $\gamma$ | $\mu$ [dB] | $\gamma$ | $\mu$ [dB] |
| WGN | M | 0,87 | -11 | 0,66 | 3 |
| | $F_1$ | 0,76 | -8 | 0,66 | 1 |
| | $F_2$ | 0,76 | -5 | 0,66 | 3 |
| Balbúrdia | M | 1 | -10 | 0,66 | -2 |
| | $F_1$ | 0,66 | 3 | 0,66 | 2 |
| | $F_2$ | 0,66 | -2 | 0,66 | 2 |
| ICRA | M | 0,76 | 1 | 0,66 | 11 |
| | $F_1$ | 0,66 | -4 | 0,66 | 3 |
| | $F_2$ | 0,76 | -2 | 0,66 | 4 |
| Trem | M | 0,57 | 7 | 0,57 | 17 |
| | $F_1$ | 0,57 | 14 | 0,5 | 19 |
| | $F_2$ | 0,57 | 10 | 0,57 | 15 |
| ISTS | M | 0,5 | 2 | 0,57 | 8 |
| | $F_1$ | 0,5 | -4 | 0,5 | 8 |
| | $F_2$ | 0,57 | -6 | 0,57 | 5 |
| Todos | | 0,76 | 7 | 0,5 | 29 |

Tabela 23 – Valores médios de MOS-LQO e NCM para WM, $CM_c/CM_v$ e CM ótima global, em uma SNR -10 dB.

| Especificação Ruído | Fala | Métrica | WM | $CM_c/CM_v$ | CM |
|---|---|---|---|---|---|
| WGN | M | NCM | 0,842 | **0,925** | **0,925** |
| | | MOS-LQO | 1,79 | **2,05** | 2,04 |
| | $F_1$ | NCM | 0,88 | 0,94 | **0,948** |
| | | MOS-LQO | 1,58 | **1,66** | 1,65 |
| | $F_2$ | NCM | 0,89 | 0,935 | **0,938** |
| | | MOS-LQO | 1,56 | **1,63** | 1,62 |
| Balbúrdia | M | NCM | 0,82 | **0,875** | 0,802 |
| | | MOS-LQO | 1,55 | **1,72** | 1,71 |
| | $F_1$ | NCM | 0,842 | **0,913** | 0,91 |
| | | MOS-LQO | 1,34 | **1,4** | **1,4** |
| | $F_2$ | NCM | 0,891 | 0,924 | **0,925** |
| | | MOS-LQO | 1,42 | **1,64** | 1,63 |
| ICRA | M | NCM | 0,843 | 0,844 | **0,896** |
| | | MOS-LQO | 1,4 | **1,51** | 1,46 |
| | $F_1$ | NCM | 0,89 | 0,856 | **0,925** |
| | | MOS-LQO | 1,31 | **1,33** | 1,29 |
| | $F_2$ | NCM | 0,901 | 0,876 | **0,931** |
| | | MOS-LQO | 1,34 | **1,36** | **1,36** |
| Trem | M | NCM | 0,901 | 0,881 | **0,919** |
| | | MOS-LQO | 1,54 | **1,71** | 1,63 |
| | $F_1$ | NCM | 0,918 | 0,89 | **0,928** |
| | | MOS-LQO | 1,41 | **1,48** | 1,39 |
| | $F_2$ | NCM | **0,921** | 0,887 | **0,921** |
| | | MOS-LQO | 1,38 | **1,50** | 1,48 |
| ISTS | M | NCM | 0,878 | 0,88 | **0,902** |
| | | MOS-LQO | 1,64 | **2,0** | **2,0** |
| | $F_1$ | NCM | **0,978** | 0,943 | **0,978** |
| | | MOS-LQO | 1,71 | **2,14** | 2,08 |
| | $F_2$ | NCM | **0,981** | 0,976 | 0,980 |
| | | MOS-LQO | 1,74 | **2,23** | 2,22 |
| Todos | | NCM | 0,873 | 0,886 | **0,903** |
| | | MOS-LQO | 1,45 | **1,67** | 1,55 |

Tabela 24 – Valores médios de MOS-LQO e NCM para WM, $CM_c/CM_v$ e CM ótima global, em uma SNR 0 dB.

| Especificação Ruído | Fala | Métrica | WM | $CM_c/CM_v$ | CM |
|---|---|---|---|---|---|
| WGN | M | NCM | 0,933 | **0,969** | **0,969** |
| | | MOS-LQO | 2,61 | **3,02** | 2,97 |
| | $F_1$ | NCM | 0,952 | 0,976 | **0,982** |
| | | MOS-LQO | 2,17 | **2,38** | 2,36 |
| | $F_2$ | NCM | 0,945 | **0,982** | **0,982** |
| | | MOS-LQO | 2,2 | **2,45** | 2,42 |
| Balbúrdia | M | NCM | 0,924 | **0,951** | 0,945 |
| | | MOS-LQO | 2,15 | **2,40** | 2,37 |
| | $F_1$ | NCM | 0,965 | **0,97** | **0,97** |
| | | MOS-LQO | 1,78 | **1,93** | 1,93 |
| | $F_2$ | NCM | 0,951 | **0,958** | 0,957 |
| | | MOS-LQO | 1,80 | **2,02** | 2,01 |
| ICRA | M | NCM | 0,953 | 0,938 | **0,956** |
| | | MOS-LQO | 1,83 | **1,95** | 1,86 |
| | $F_1$ | NCM | 0,981 | 0,98 | **0,984** |
| | | MOS-LQO | 1,69 | **1,77** | 1,76 |
| | $F_2$ | NCM | 0,98 | 0,98 | **0,982** |
| | | MOS-LQO | 1,71 | **1,88** | 1,87 |
| Trem | M | NCM | 0,951 | 0,94 | **0,954** |
| | | MOS-LQO | 2,07 | **2,21** | 2,13 |
| | $F_1$ | NCM | 0,979 | 0,967 | **0,98** |
| | | MOS-LQO | 1,81 | **1,91** | 1,85 |
| | $F_2$ | NCM | 0,976 | **0,983** | 0,982 |
| | | MOS-LQO | 1,76 | **1,92** | 1,86 |
| ISTS | M | NCM | 0,975 | 0,958 | **0,98** |
| | | MOS-LQO | 2,37 | **2,65** | 2,63 |
| | $F_1$ | NCM | **0,991** | 0,99 | **0,991** |
| | | MOS-LQO | 2,34 | **2,71** | 2,65 |
| | $F_2$ | NCM | 0,989 | **0,992** | **0,992** |
| | | MOS-LQO | 2,40 | **2,75** | 2,62 |
| Todos | | NCM | 0,95 | 0,953 | **0,968** |
| | | MOS-LQO | 2,01 | **2,28** | 2,07 |

# 6 CONCLUSÃO

Este trabalho apresentou um estudo sobre o desempenho das máscaras tempo-frequência, tendo proposto uma máscara assimétrica para redução de ruído em aparelhos auditivos. Essa máscara indicou o possível limite de desempenho para a técnica de máscaras tempo-frequência e revelou que resultados próximos dos máximos passíveis de obtenção podem ser alcançados com máscaras simétricas quando configuradas corretamente. Foi também observado que a maximização do desempenho das máscaras tempo-frequência em termos da qualidade e da inteligibilidade apresenta evidências de que ambos os objetivos não podem ser alcançados simultaneamente. A curva de ganho otimizada para a qualidade encontra-se deslocada para a direita com o parâmetro de suavização correspondente entre WM e WR. A curva de ganho otimizada para a inteligibilidade apresenta curva de ganho deslocada para a esquerda e com parâmetro de suavização menor. No entanto, é importante ressaltar que, ao maximizar a qualidade, muitas vezes podem ocorrer melhorias no desempenho da inteligibilidade, mas o contrário nem sempre é verdadeiro.

Além disso, constatou-se que, quando ajustada corretamente, é possível encontrar uma configuração ótima de CM para cada especificação de ruído e SNR. Porém, quando configurada com os valores de $\mu$ e $\gamma$ da CM ótima global, demonstrou um desempenho consistente, independentemente do nível de SNR e do tipo de ruído, resultando na possibilidade de utilização de um par de parâmetros único para as diversas situações sem perda significativa de desempenho.

Finalmente, também observou-se que o efeito do ruído nas regiões que contêm vogais é diferente das regiões que contêm consoantes. As consoantes são mais afetadas pela contaminação de ruído. Porém, a dificuldade na separação de trechos de vogais e consoantes ao longo do sinal de fala pode tornar desafiadora a implementação prática de estratégias de filtragens para cada região separadamente, sem a evidência de aumentos significativos de qualidade ou inteligibilidade.

# REFERÊNCIAS

ALAM, Md. Jahangir; O'SHAUGHNESSY, Douglas; SELOUANI, Sid-Ahmed. Speech enhancement employing a sigmoid -type gain function with a modified a priori signal-to-noise ratio (SNR) estimator. In: CANADIAN Conference on Electrical and Computer Engineering. Niagara Falls, ON, Canada: [s.n.], mai. 2008. p. 631–635.

ALCAIM, Abraham; SOLEWICZ, Jose Alberto; MORAES, Joao Antonio de. Frequência de ocorrência dos fones e lista de frases foneticamente balanceadas no português falado no Rio de Janeiro. **Revista da Sociedade Brasileira de Telecornunicacoes**, v. 7, n. 1, p. 23–41, 1992.

ALMEIDA, Kátia; IORIO, Maria Cecília M. **Próteses Auditivas - Fundamentos e Apicações Clínicas**. São Paulo: Lovise, 1996.

ALMEIDA, Kátia; IORIO, Maria Cecília M. **Próteses auditivas – Fundamentos e aplicações clínicas**. São Paulo: Lovise, 2003.

AMERICAN SPEECH-LANGUAGE-HEARING ASSOCIATION. **Causes of Hearing Loss in Adults**. Acessado em 2023-08-17. 2023. Disponível em: https://www.asha.org/public/hearing/causes-of-hearing-loss-in-adults/.

ARSINTE, Radu; LUPU, Eugen; SUMALAN, Teodor. A Rapid Prototyping Model Concept for a DSP Based Hearing Aid, p. 337–340, jul. 2017.

BERGERON, F.; HOTTON, M. Perception in noise with the Digisonic SP cochlear implant: Clinical trial of Saphyr processor's upgraded signal processing. **European Annals of Otorhinolaryngology, Head and Neck Diseases**, v. 133, n. 1, s4–s6, jun. 2016.

BISPO, Bruno C et al. EW-PESQ: A quality assessment method for speech signals sampled at 48 kHz. **Journal of the Audio Engineering Society**, Audio Engineering Society, v. 58, n. 4, p. 251–268, 2010.

BORGES, Renata Coelho; COSTA, Márcio H. A feed forward adaptive canceller to reduce the occlusion effect in hearing aids. **Computers in Biology and Medicine**, p. 266–275, 2016.

CHIEA, Rafael Attili; COSTA, Márcio H; BARRAULT, Guillaume. Uma Comparação entre Máscaras Tempo-frequência para Redução de Ruído em Implantes Cocleares. In: PROCEEDINGS of XXXVII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais. Petrópolis, Brazil: [s.n.], out. 2019. p. 15–23.

CHIEA, Rafael Attili; COSTA, Márcio Holsbach; BARRAULT, Guillaume. New insights on the optimality of parameterized Wiener filters for speech enhancement applications. **Speech Communication**, v. 109, p. 46–54, mai. 2019.

CORSINO, Patrícia. **Educação Infantil: cotidiano e políticas**. [S.l.]: Autores Associados, 2020.

CROCHIERE, R. A weighted overlap-add method of short-time Fourier analysis/synthesis. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, v. 28, n. 1, p. 99–102, 1980.

DE OLIVEIRA, Juliana Simili. PAISAGEM SONORA ALÉM DA AUDIÇÃO: Representações sonoras urbanas das pessoas surdas., 2017.

DILLON, Harvey. **Hearing Aids**. 2nd. [S.l.]: Boomerang Press, 2001.

DRESCHLER, Wouter A.; VERSCHUURE, Hans; LUDVIGSEN, Carl; WESTERMANN, Søren. ICRA Noises: artificial noise signals with speech-like Spectral and temporal properties for hearing instrument assessment. **Audiology**, v. 40, n. 3, p. 148–157, 2001.

DURAN, JOSé ENRIQUE RODAS. **Biofísica: conceitos e aplicações**. 2. ed. São Paulo: Pearson Prentice Hall, 2011. ISBN 978-85-7605-928-8.

EPHRAIM, Yariv; MALAH, David. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. **IEEE Transactions on Acoustics, Speech and Signal Processing**, ASSP-32, n. 6, p. 1109–1121, dez. 1984.

FONTAINE, Mathieu; LIUTKUS, Antoine; GIRIN, Laurent; BADEAU, Roland. Explaining the parameterized wiener filter with alpha-stable processes. In: PROCEEDINGS of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, USA: [s.n.], out. 2017. p. 51–55.

FU, Qian-Jie; NOGAKI, Geraldine. Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing. **Journal of the Associations for Research in Otolaryngology**, v. 6, n. 1, p. 19–27, fev. 2005.

GARCIAS, Eduardo A.C. **Biofísica**. 1. ed. São Paulo: Sarvier, 2002.

HAST, Anne; SCHLÜCKER, Luisa; DIGESER, Frank; LIEBSCHER, Tim; HOPPE, Ulrich. Speech Perception of Elderly Cochlear Implant Users Under Different Noise Conditions. **Otology & Neurotology**, v. 36, n. 10, p. 1638–1643, dez. 2015.

HEARING SOLUTIONS, SOUND ADVICE YOU CAN TRUST. **The hearing aids of yesteryear: A brief history of hearing aids from then to now**. Acessado em 2023-08-17. 2023. Disponível em: `https://www.hearingsolutions.ca/the-history-and-evolution-of-the-hearing-aid/`.

HOGG, Robert V; LEDOLTER, Johannes. **Engineering statistics**. [S.l.]: Macmillan Publishing Company, 1987.

HOLUBE; FREDELAKE, Stefan; VLAMING, Marcel; KOLLMEIER, Birger. Development and analysis of an International Speech Test Signal (ISTS). **International Journal of Audiology**, v. 49, n. 12, p. 891–903, 2010.

HOLUBE, I.; KOLLMEIER, B. Speech intelligibility prediction in hearing- impaired listeners based on a psychoacoustically motivated perception model. **The Journal of the Acoustical Society of America**, v. 100, n. 3, p. 1703–1716, set. 1996.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional de Saúde**. [S.l.: s.n.], 2019.

ITU-T P.800.1. **Mean Opinion Score (MOS) Terminology**. [S.l.: s.n.], Geneva, Switzerland 2006. International Telecommunications Union.

ITU-T P.862. **Perceptual evaluation of speech quality (PESQ): objective method for end-to-end speech quality assessment of narrow band telephone networks and speech codecs**. [S.l.: s.n.], Geneva, Switzerland 2001. International Telecommunications Union.

ITU-T P.862.2. **Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs**. [S.l.: s.n.], Geneva, Switzerland 2005. International Telecommunications Union.

JARQUE, Carlos M; BERA, Anil K. A test for normality of observations and regression residuals. **International Statistical Review/Revue Internationale de Statistique**, JSTOR, p. 163–172, 1987.

KATES, James M.; AREHART, Kathryn H.; ANDERSON, Melinda C.; MURALIMANOHAR, Ramesh Kumar; JR, Lewis O. Harvey. Using Objective Metrics to Measure Hearing Aid Performance. **Ear and Hear**, v. 39, n. 6, p. 1165–1175, 2018.

KEWLEY-PORT, Diane; BURKLE, T Zachary; LEE, Jae Hee. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. **Journal of the Acoustical Society of America**, v. 122, n. 4, p. 2365–2375, out. 2007.

LI, Ning; LOIZOU, Philipos C. The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise. **Acoustical Society of America**, v. 124, n. 6, p. 3947–3958, dez. 2008.

LIM, Jae S.; OPPENHEIM, Alan V. Eancement and bandwidth compression of noisy speech. **Proceedings of the IEEE**, v. 67, n. 12, p. 1586–1604, 1979.

LOIZOU, Philipos C. **Speech enhancement: Theory and Practice**. 2nd. [S.l.]: CRC Press, 2013.

MARTIN, Rainer; HEUTE, Ulrich; ANTWEILER, Christiane. **Advances in Digital Speech Transmission**. [S.l.]: Wiley, 2008.

MOORE, Brian C. J. **Cochlear hearing loss: physiological, psychological and technical issues**. 2nd. [S.l.]: John Wiley & Sons, 2007.

MUSEU DO PARELHO AUDITIVO. **A História do Aparelho Auditivo: Conheça!** Acessado em 2023-08-17. 2023. Disponível em: `https://museudoaparelhoauditivo.com.br/publicacoes-a-historia-do-aparelho-auditivo.php`.

NETTER, Frank H. **Atlas de Anatomia Humana**. 5. ed. [S.l.]: Elsevier, 2011. ISBN 9788535237481; 8535237488.

OKUNO, Emico; CALDAS, Luiz Iberê; CHOW, Cecil. **Física para Ciências Biológicas e Biomédicas**. 1. ed. São Paulo: Harper & Row do Brasil, 1982.

OWREN, Michael J.; CARDILLO, Gina C. The relative roles of vowels and consonants in discriminating talker identity versus word meaning. **Journal of the Acoustical Society of America**, v. 119, n. 3, p. 1727–1739, mar. 2006.

PAGAN, Luciana; WERTZNER, Haydée Fiszbein. Análise acústica das consoantes líquidas do Português Brasileiro em crianças com e sem transtorno fonológico. **Revista da Sociedade Brasileira de Fonoaudiologia**, v. 12, n. 2, p. 106–13, jun. 2007.

PEREYRON, Leticia; ALVES, Ubiratã K. Descrição acústica das vogais tônicas do espanhol rioplatense e de uma variedade do português do sul do Brasil de monolíngues e bilíngues: uma discussão dinâmica sobre desenvolvimento linguístico. **Linguística**, Asociación de Linguística y Filologia de América Latina, v. 35, n. 1, p. 103–127, 2019.

PHATAK, Sandeep A.; ALLEN, Jont B. Consonant and vowel confusions in speech-weighted noise. **Journal of the Acoustical Society of America**, v. 121, n. 4, p. 2312–2326, abr. 2007.

ROTHAUSER, EH. IEEE recommended practice for speech quality measurements. **IEEE Transactions on Audio and Electroacoustics**, Institute of Electrical e Electronics Engineers (IEEE), v. 17, n. 3, p. 225–246, 1969.

SISTEMA DE CONSELHOS DE FONOAUDIOLOGIA. **Guia de Orientação na Avaliação Audiológica**. [S.l.], 2020.

TEFILI, Diego; BARRAULT, Guillaume François Gilbert; FERREIRA, Alexandre André; CORDIOLI, Júlio Apolinário; LETTNIN, Djones Vinicius. Implantes cocleares: aspectos tecnológicos e papel socioeconômico. **Revista Brasileira de Engenharia Biomédica**, v. 29, n. 4, p. 414–433, dez. 2013.

VIEGAS, Flávia; VIEGAS, Danieli; GUIMARÃES, Glaucio; SOUZA, Margareth de; LUIZ, Ronir; SIMÕES-ZENAR, Marcia; NEMR, Katia. Comparação de medidas de frequência fundamental e frequências dos formantes em duas tarefas de fala. **CEFAC**, v. 21, n. 6, 2019.

WANG, DeLiang; BROWN, Guy J. **Computational Auditory Scene Analysis: Principles, Algorithms, and Applications**. [S.l.]: Wiley-IEEE Press, 2006.

WILSON, Blake S.; DORMAN, Michael F. The Surprising Performance of Present-Day Cochlear Implants. **IEEE Transactions on Biomedical Engineering**, v. 54, n. 6, p. 969–972, jun. 2007.

WORLD HEALTH ORGANIZATION. **Deafness and hearing loss**. [S.l.: s.n.], 2020. `https://www.who.int/en/news-room/fact-sheets/detail/deafness-and-hearing-loss`.

WOUTERS, Jan; DOCLO, Simon; KONING, Raphael; FRANCART, Tom. Sound processing for better coding of monaural and binaural cues in auditory prostheses. **Proceedings of the IEEE**, v. 101, n. 9, p. 1986–1997, set. 2013.

YNOGUTI, Carlos Alberto. **Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov**. 1999. Tese (Doutorado) – Universidade Estadual de Campinas Faculdade de Engenharia Elétrica e de Computação Departamento de Comunicações.

ZAHNERT, Thomas. The differential diagnosis of hearing loss. **Deutsches Ärzteblatt International**, v. 108, n. 25, p. 433–444, jun. 2011.

# APÊNDICE A – MANUAL DO USUÁRIO PARA A INTERFACE DE APRENDIZAGEM DE MÁSCARAS DE TEMPO-FREQUÊNCIA

Este apêndice tem como objetivo apresentar uma interface gráfica para auxiliar a análise e comparação de máscaras tempo frequência.

## A.1 FUNCIONALIDADES

A interface de aprendizado de máscaras tempo-frequência é composta por três painéis principais, cada um com uma funcionalidade específica: Entrada de Dados, Seleção de Máscaras e Análise e Avaliação. Esses painéis possibilitam um procedimento simplificado para a análise e comparação entre máscaras tempo frequência. De forma secundária a interface também pode ser utilizada como ferramenta de ensino e aprendizagem.

O painel de Entrada de Dados (Figura A.1) tem como principal funcionalidade permitir ao usuário carregar sinais de áudio de interesse (fala e ruído) e contaminar o sinal de fala com o nível de ruído desejado. É possível ouvir o áudio resultante da contaminação da fala, com a proporção selecionada no nível de SNR. Essa funcionalidade permite ao usuário experimentar diferentes tipos de ruído e em vários níveis de SNR, a fim de compreender os efeitos causados pela contaminação sobre o sinal de fala original.

Ao clicar no botão 'Fala', é apresentada uma janela que permite ao usuário inserir a fala de seu interesse, como visto na Figura A.2. De maneira similar, o botão 'Ruído' tem o mesmo efeito ao ser acionado. Esse recurso é útil para personalizar a entrada de dados e experimentar diferentes tipos de fala e ruído no processo de contaminação sonora.

No mesmo painel, também são encontrados botões que permitem ao usuário ouvir o áudio de fala selecionado contaminado com o nível de SNR determinado pelo botão de controle deslizante. Os botões de alto-falantes e apresentação gráfica dos sinais são habilitados somente após a carga dos sinais e a cada definição de nível de SNR, como pode ser observado na Figura A.3. Esses recursos permitem que o usuário escute e visualize a fala contaminada de forma mais clara, o que é fundamental para o processo de aprendizado da técnica de máscara tempo frequência.

Figura A.1 – Entrada de dados


Figura A.2 – Seleção da fala de interesse

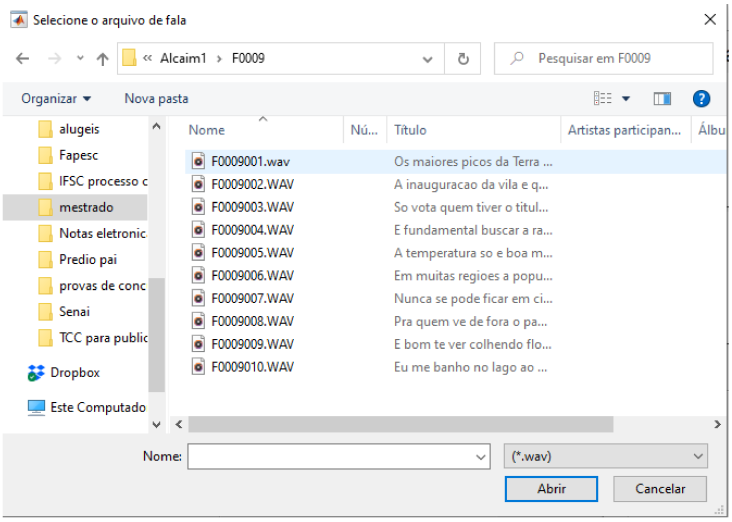
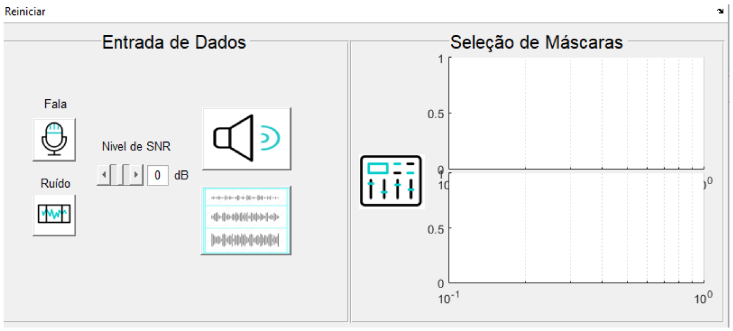
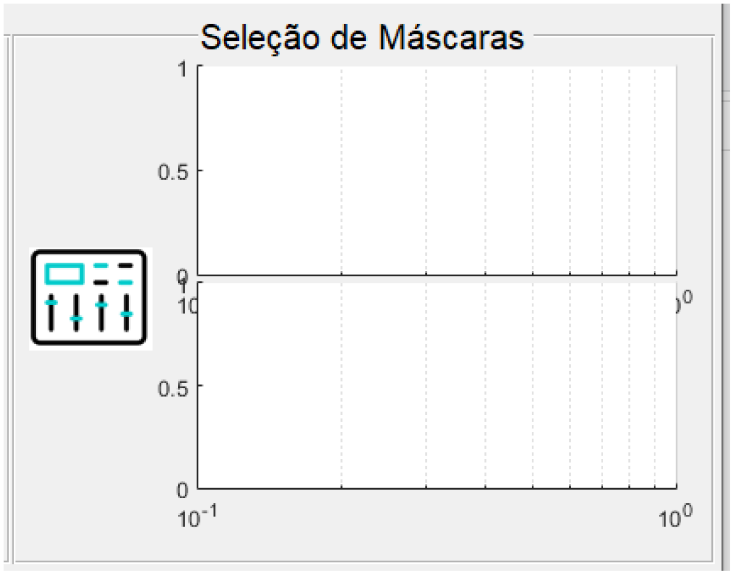


Figura A.3 – Fala e ruído carregados




O segundo painel, Seleção de Máscaras (Figura A.4) tem como objetivo permitir a seleção de diferentes tipos de máscaras, possibilitando configurá-las (quando houver parâmetros) de acordo com o interesse. Essa funcionalidade é importante para permitir que o usuário escolha a máscara mais adequada para sua aplicação específica e ajuste os parâmetros relevantes de acordo com suas necessidades.

Figura A.4 – Seleção de máscaras




A seguir, são apresentadas as telas que mostram as máscaras disponíveis para seleção, juntamente com um gráfico em tempo real da configuração, a equação matemática correspondente e uma breve descrição. Essas informações são úteis para que o usuário possa visualizar as características de cada máscara e escolher a que melhor se adequa à sua aplicação. O gráfico em tempo real também ajuda o usuário a entender como os parâmetros afetam a forma da máscara e a tomar decisões informadas sobre a configuração adequada.

Figura A.5 – Tela de seleção de máscaras

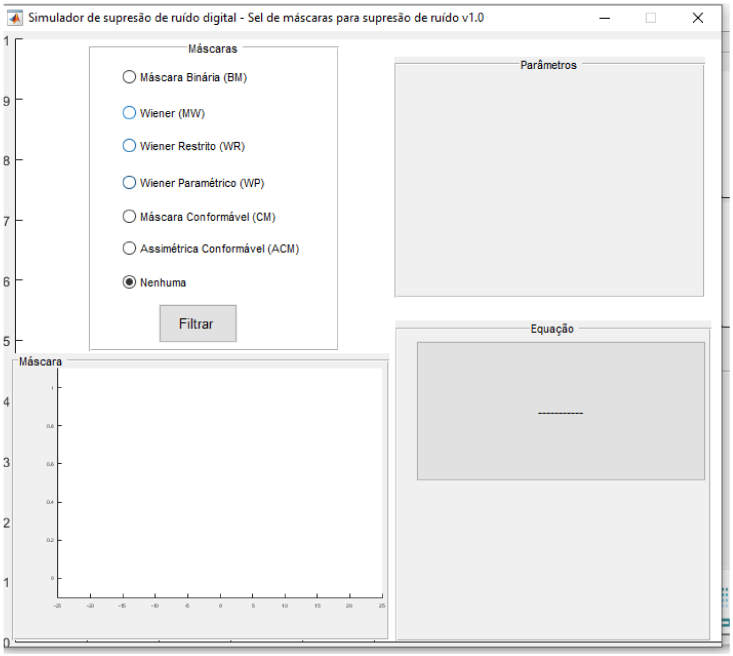



Figura A.6 – Tela de seleção da máscara BM

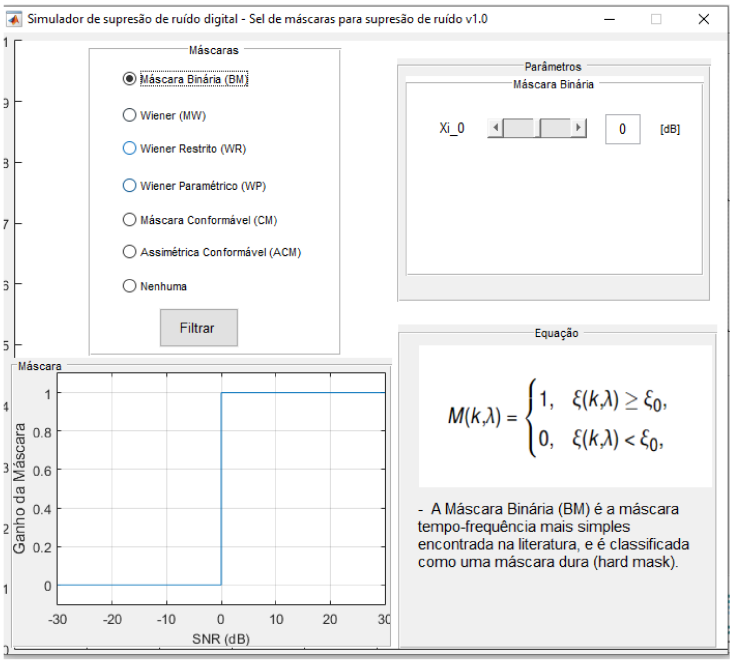


Figura A.7 – Tela de seleção da máscara Wiener

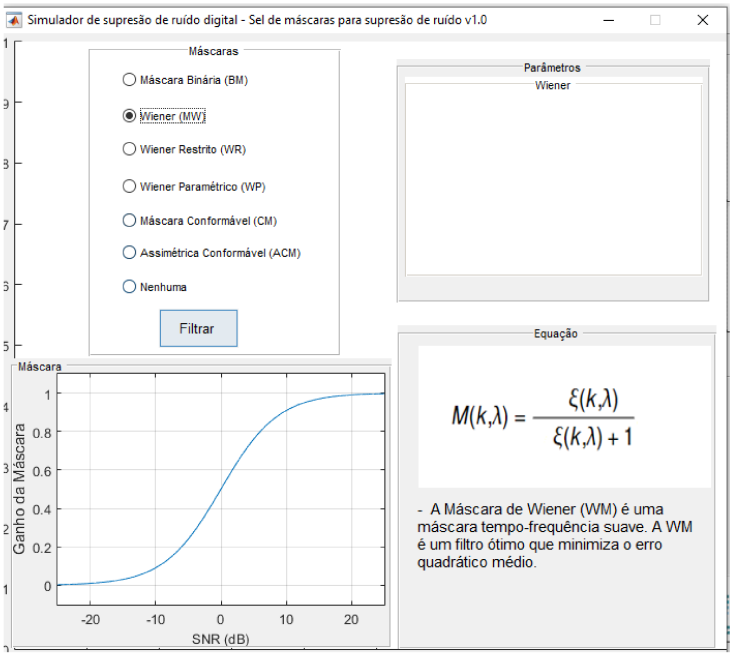



Figura A.8 – Tela da seleção de máscara WR

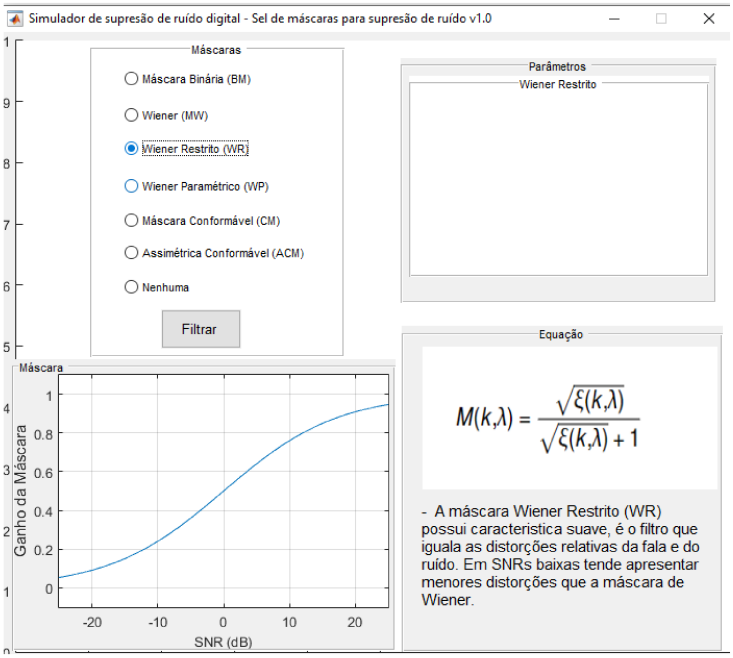


Figura A.9 – Tela de seleção da máscara WP

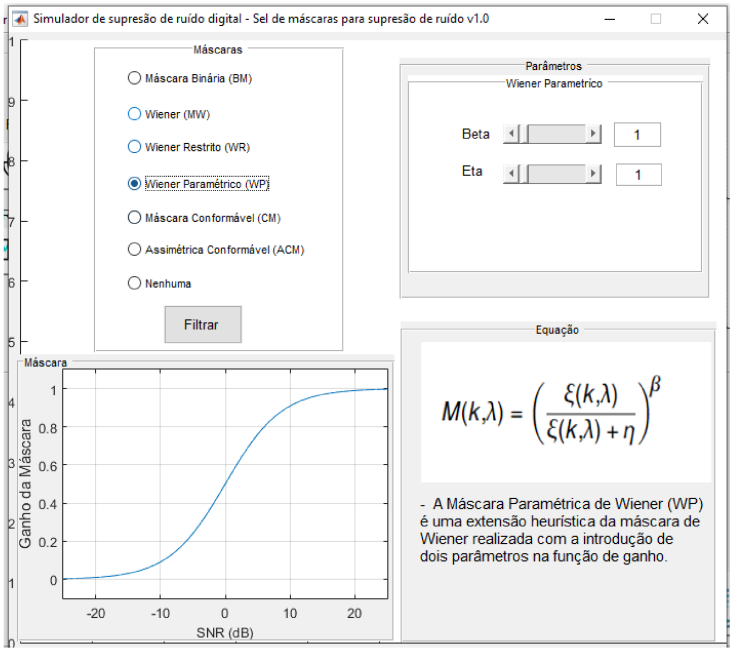



Figura A.10 – Tela de seleção de máscaras CM

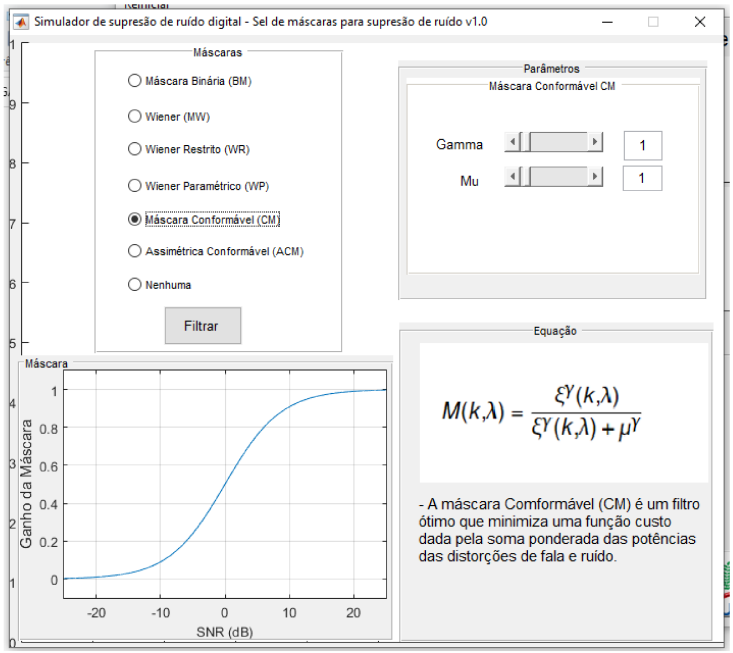
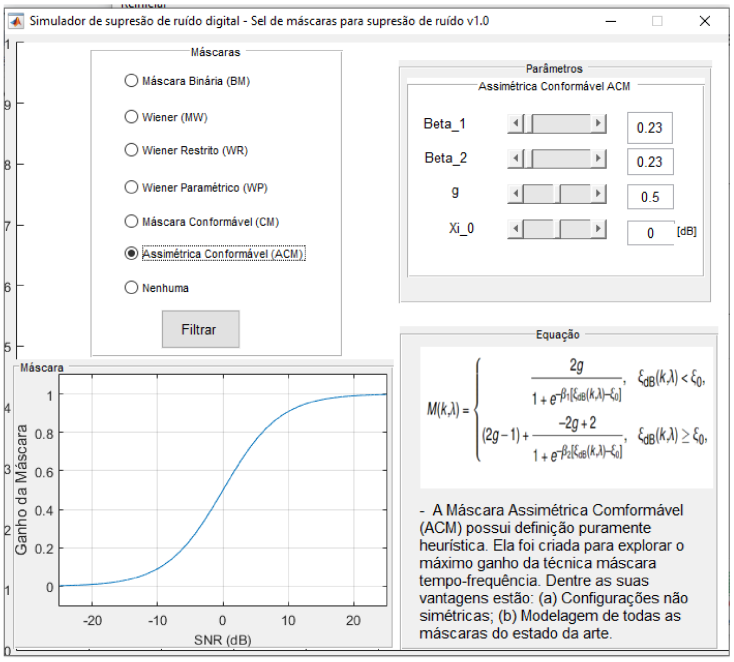


Figura A.11 – Tela de seleção de máscaras ACM




Após selecionar a máscara desejada e clicar no botão "Filtrar", a tela de seleção é fechada e a filtragem do áudio é realizada automaticamente. Os critérios objetivos de qualidade e inteligibilidade resultantes, como Pesq e NCM, são exibidos na tela de análise e avaliação (Figura A.12). Nessa tela, o usuário pode ainda visualizar o áudio filtrado, ouvi-lo ou salvá-lo para referência futura. Essa funcionalidade é essencial para permitir que o usuário avalie a eficácia da filtragem e tome decisões informadas sobre o processamento adicional do áudio.

Se o local de salvamento não for selecionado, um aviso informativo será exibido na tela, informando que nenhum arquivo foi salvo. Isso garantirá que o usuário saiba quando os dados foram salvos.

Figura A.12 – Reprodução do áudio filtrado

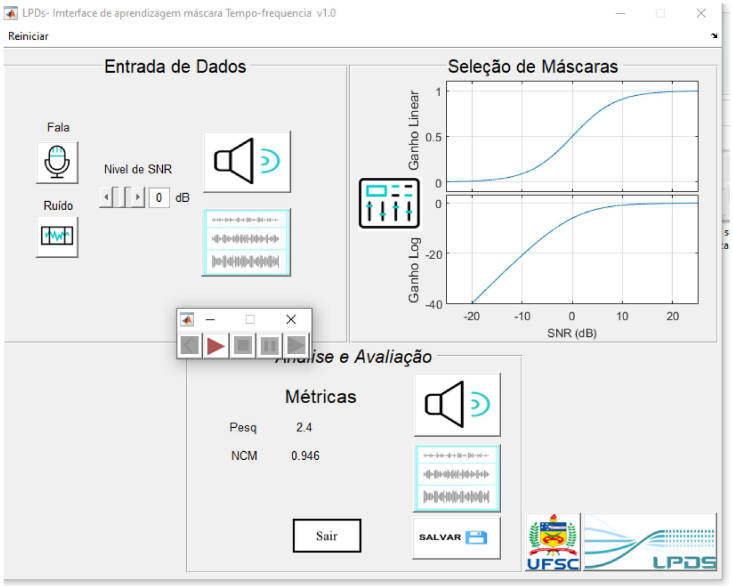



Figura A.13 – Exibição gráfica dos dados

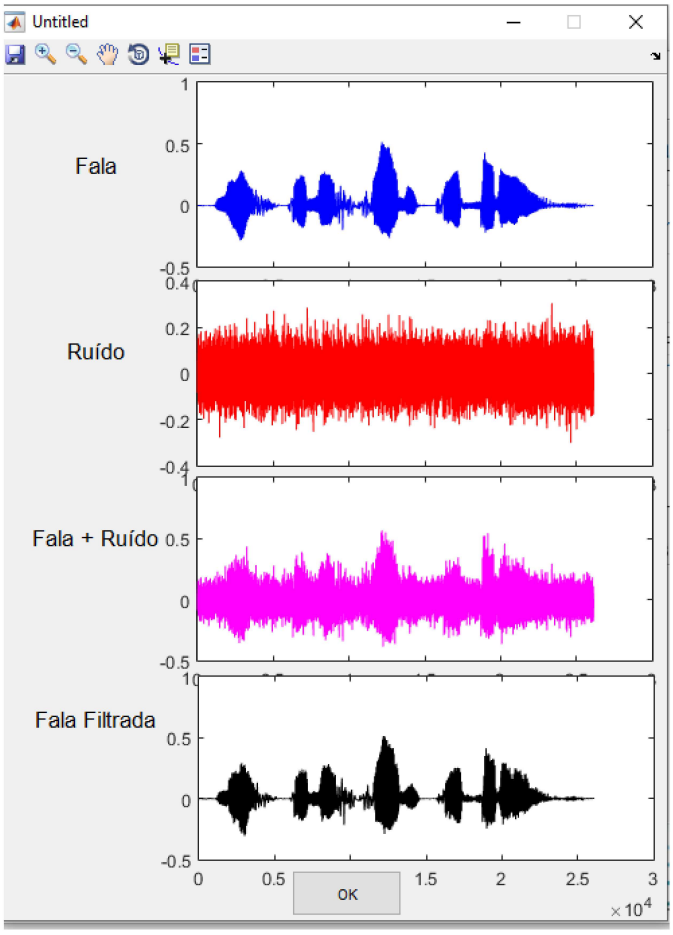


Figura A.14 – Salvar imagem

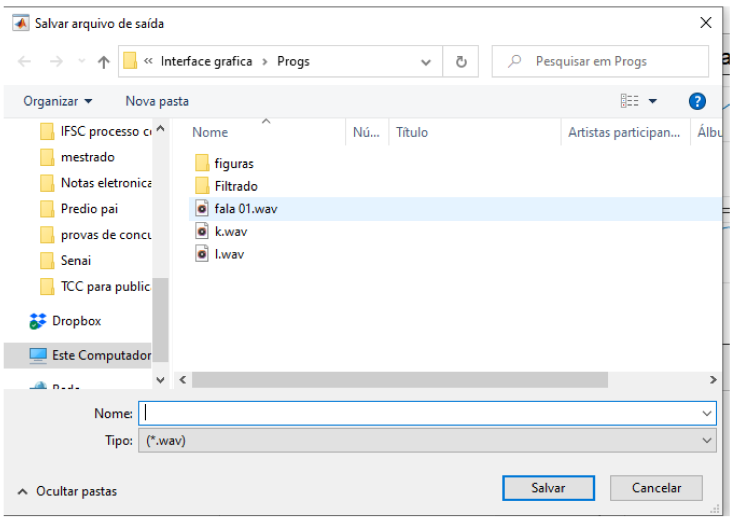
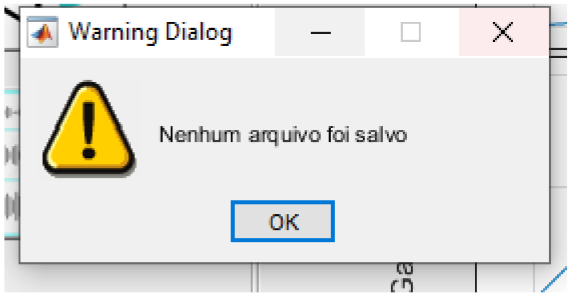



Figura A.15 – Mensagem de caminho não selecionado




## A.2 CONSIDERAÇÕES FINAIS

Este manual foi elaborado com o objetivo de esclarecer as etapas do software desenvolvido. Ele foi criado para documentar e explicar os procedimentos envolvidos.