Otimização de medidas MFCC em cenas acústicas para classificação de patologias laríngeas

Vinícius J. D. Vieira, Rafael R. Pertum e Renato Candido

***Resumo*— Neste trabalho, é realizado um ajuste fino da medida acústica MFCC (*Mel-Frequency Cepstral Coefficients*) em diferentes cenários no contexto da classificação de patologias laríngeas. As cenas acústicas consideradas são: ambiente com e sem a presença de reverberação e ruído (com efeitos isolados e misturados) e a classificação individualizada por gênero. As patologias consideradas são: edema de Reinke, carcinoma, leucoplasia, laringite, pólipos e paralisia nas pregas vocais. O classificador empregado é baseado em análise discriminante quadrática. Os resultados indicam que há uma configuração ótima dessa medida, que proporciona os maiores valores de acurácia nos experimentos realizados. Ainda, é observado que a utilização de classificadores dedicados por gênero proporciona um ganho de acurácia relevante em relação ao resultado obtido com o classificador generalista.**

***Palavras-Chave*— Processamento de sinais de voz, classificação de patologias laríngeas, MFCC, otimização**

***Abstract*— In this work, a fine-tuning of the acoustic measure MFCC (Mel-Frequency Cepstral Coefficients) is performed in different scenarios in the context of laryngeal pathologies classification. The acoustic scenes considered are: environments with and without the presence of reverb and noise (with isolated and mixed effects), and the individualized classification by gender. The pathologies considered are: Reinke's edema, carcinoma, leukoplakia, laryngitis, polyps and vocal fold paralysis. The classifier used is based on quadratic discriminant analysis. The results indicate that there is an optimal configuration of this measure, which provides the highest accuracy values in the experiments. Furthermore, it is observed that the use of a dedicated classifier by gender provides a relevant gain in accuracy in relation to the result obtained with the generalist classifier.**

***Keywords*— Speech signal processing, laryngeal pathologies classification, MFCC, optimization**

I. Introdução

A voz exerce um papel fundamental nas interações sociais humanas e, nos últimos anos tornou-se um componente crucial em diversas tecnologias emergentes [1]. Aplicações de processamento de voz, a exemplo de assistentes virtuais, estão cada vez mais presentes no cotidiano das pessoas, facilitando tarefas e proporcionando novas formas de interação com o mundo digital [2]. Para profissionais que dependem da voz, como professores, cantores, atores e operadores de telemarketing, a saúde vocal não é apenas uma questão de bem-estar, mas uma necessidade para o desempenho eficaz de suas funções [3].

A investigação de medidas acústicas para a classificação de patologias laríngeas tem sido um campo de estudo ativo [4], [5]. No entanto, ainda não existe um consenso definitivo sobre quais medidas são as mais robustas e confiáveis para discriminar diferentes tipos de problemas vocais. Com a crescente adoção de dispositivos móveis na sociedade e o acesso facilitado à internet, os mecanismos de telessaúde têm emergido como uma solução promissora para ampliar o alcance dos serviços de saúde, especialmente em regiões remotas ou com escassez de profissionais especializados [6]. A variabilidade acústica introduzida por fatores como ruído e diferentes tipos de microfones representa um desafio significativo para a padronização e a validade dessas medições.

Uma das técnicas mais comuns para extração de características acústicas de sinais de voz é a análise no domínio cepstral com MFCC (*Mel-Frequency Cepstral Coefficients*), a qual busca uma representação do áudio com base na percepção do sistema auditivo humano. Originalmente proposto em [7], tem sido utilizado em aplicações de processamento de voz, como reconhecimento de fala e de locutor [8], reconhecimento de emoções [9] e também para a identificação de patologias [10].

Apesar de ser uma técnica amplamente utilizada, existem poucos estudos sobre o impacto do ajuste dos parâmetros envolvidos em sua obtenção, como o número de coeficientes, o tamanho do quadro e o passo entre os quadros. Em [11] foi feito um estudo do efeito do tamanho do quadro no contexto de patologias da voz. Já em [12], foi feita uma extensão dessa investigação, considerando a variação do número de coeficientes e do passo entre quadros, e observou-se que o ajustes destes parâmetros traz mais robustez aos modelos de classificação. No entanto, nestes trabalhos, não foram consideradas a influência de ruído e reverberação nos sinais de áudio, efeitos comumente presentes, principalmente considerando aplicações de telessaúde.

Em [10], foi feita uma investigação sobre o impacto de distorções acústicas em características extraídas dos sinais de voz para classificação de patologias laríngeas considerando diferentes medidas. Nesse estudo, foi constatado que a medida MFCC foi a que proporcionou os melhores resultados de classificação perante às distorções acústicas consideradas que incluíram variações no cenário de ruído e de reverberação. Neste trabalho, é realizada uma extensão do estudo apresentado em [10], a partir de três perguntas norteadoras: 1) *há uma configuração ótima para os MFCCs, com a qual se obtém uma maior acurácia de classificação?* 2) *há alteração dos parâmetros ótimos dos MFCCs caso o sinal de voz seja submetido a ambientes reverberantes e ruidosos?* 3) *há influência do gênero do falante na estimação da configuração ótima para os MFCCs?* Como contribuição deste estudo, a resposta para estas perguntas pode tornar o desenvolvimento de sistemas de classificação mais robustos quando aplicados em diferentes cenas acústicas.

O restante do texto está organizado da seguinte forma: Na Seção II é apresentada a formulação do problema, juntamente com uma descrição mais detalhada dos MFCCs. Na Seção III, são apresentados os materiais e métodos utilizados para o desenvolvimento deste trabalho. Na Seção IV são apresentados os resultados obtidos nos experimentos realizados e, na Seção V, é apresentada a conclusão.

Vinícius J. D. Vieira, Rafael R. Pertum e Renato Candido, Depto. de Processamento de Voz, SiDi, Alphaville, Campinas, SP - Brasil, e-mails: {vinicius.dv, r.pertum, r.candido}@sidi.org.br.

II. Formulação do problema

A análise de patologias laríngeas por meio da voz, em geral, pode ser fundamentada em um modelo linear de produção vocal [13] e outros que sugerem medidas baseadas em um modelo não linear [14]. Por meio do modelo linear, o sinal de voz é resultado de um sistema fonte-filtro, no qual a fonte é representada pelas pregas vocais, localizadas na laringe, e o filtro é interpretado como sendo o trato vocal. A estrutura do trato vocal pode ser considerada como responsável pelo envelope do espectro de potência de curto prazo do sinal. O objetivo do MFCC é capturar essa informação e representá-la adequadamente, relacionando-a à percepção do ouvido humano, o que é feito usando uma aproximação computacional da percepção auditiva para uma escala chamada Mel [15].

Para a extração dos MFCCs, inicialmente o sinal de voz é segmentado em pequenos quadros, que são filtrados por um filtro de pré-ênfase e convertidos para o domínio da frequência por meio da transformada rápida de Fourier (*fast Fourier transform* – FFT) para a obteção dos espectros de potência. Em seguida, o espectro de potência de cada quadro passa por de um banco de filtros em escala Mel e é estimado o logaritmo da energia em cada filtro. Por fim, é aplicada a DCT (*Discrete Cosine Transform*) para a obtenção dos coeficientes c_j, de acordo com [7]:

$$c_j = \sum_{k=1}^{F} (\log\ S_k)\ \cos\left[\frac{\pi j}{F}\left(k - \frac{1}{2}\right)\right], \qquad (1)$$

para $j = 1,2,...,D$, em que D é o número de coeficientes, S_k é a energia do k-ésimo filtro e F é a quantidade de filtros na escala Mel.

Como o sinal de voz apresenta um comportamento dinâmico, sendo tipicamente não estacionário, a sua segmentação é usualmente aplicada no domínio do tempo para a extração de medidas acústicas. Neste contexto, o sinal é dividido em quadros curtos com N amostras, considerando uma sobreposição de M amostras entre os quadros, com M < N. O ajuste do comprimento do quadro pode impactar o desempenho dos sistemas. A relação entre as representações no tempo e na frequência permite enxergar mais detalhes na representação em frequência com o uso de quadros mais longos, mas acabam misturando eventos no domínio do tempo devido à não estacionariedade. Dessa forma, os ajustes adequados do tamanho do quadro N e do salto entre quadros M são fundamentais para o bom desempenho dos modelos de classificação [16].

III. Materiais e Métodos

Nesta seção é apresentado o cenário experimental desenvolvido neste trabalho.

A. Base de Dados

Para a condução deste estudo, foi utilizada a base de dados de voz conhecida como *Saarbruecken Voice Database* (SVD) [17]. Do acervo total da SVD, foram extraídos 600 registros da vogal /a/ sustentada, que é o tipo de emissão mais comum para análise acústica de patologias na laringe devido à vibração induzida nas pregas vocais [13]. Dessa seleção, 300 sinais saudáveis e 300 sinais patológicos, distribuídos entre diferentes condições laríngeas: 48 casos de edema de Reinke, 12 de carcinoma, 29 de leucoplasia, 33 de pólipos vocais, 58 de laringite e 120 de paralisia das pregas vocais.

Originalmente, os sinais da base SVD foram amostrados a uma frequência de 50 kHz. No entanto, para os propósitos desta pesquisa, optou-se por realizar uma subamostragem, ajustando a taxa para 44,1 kHz. Essa decisão visou alinhar os dados com os padrões de gravação de alta fidelidade comumente encontrados em sistemas de comunicação e gravação contemporâneos. Em ambas as categorias, saudável e patológica, a representatividade do gênero do locutor foi mantida em uma proporção de 40% de participantes do sexo masculino e 60% do sexo feminino, com idades variando entre 18 e 65 anos.

B. Extração dos MFCCs

A extração dos MFCCs foi realizada por meio da biblioteca Librosa[1]. Para fins de desenvolvimento deste estudo, foi explorada a variação de dois parâmetros: o tamanho da janela de segmentação e a quantidade de coeficientes. Assim, os tamanhos de segmento investigados são os seguintes: 20 ms, 25 ms e 30 ms, todos com sobreposição de 10 ms. Em relação à ordem dos MFCCs, são experimentadas as seguintes quantidades de coeficientes: 12, 16, 20, 24, 28 e 32. Para cada configuração, foi obtida uma matriz $Q \times D$ (Q quadros, D coeficientes), da qual foi obtido o valor médio ao longo dos quadros a fim de se aplicar o vetor $1 \times D$ por sinal na etapa de classificação. Tal investigação tem como objetivo responder à primeira pergunta norteadora deste trabalho.

C. Cenas Acústicas

Além da otimização de medidas MFCC em um ambiente sem variações acústicas (base de dados original), dois tipos de cenas acústicas são investigadas: variações no ambiente acústico e variação de gênero.

1) Variações no ambiente acústico: A investigação deste tipo de cena acústica teve como objetivo responder à segunda pergunta norteadora deste trabalho. Para tanto, foram analisados os seguintes efeitos acústicos: reverberação, ruído branco gaussiano e ruído real. Além disso, também foram analisadas a combinação desses efeitos por meio dos seguintes experimentos:

i) Variação de reverberação com o parâmetro RT60[2]: A biblioteca *pyroomacoustics*[3] foi empregada para simular uma sala *shoebox* de $36\,\mathrm{m}^3$, com comprimento, largura e altura de $3\,\mathrm{m}$, $4\,\mathrm{m}$ e $3\,\mathrm{m}$, respectivamente. Nesta sala, foram colocados um microfone, para representar o *smartphone*, e uma fonte sonora representando o locutor. O posicionamento do microfone e da fonte sonora são descritos em [10]. Então, foram geradas variações do conjunto de dados original com valores de RT60 de $0{,}1\,\mathrm{s}$ até $0{,}6\,\mathrm{s}$, com passos de $0{,}1\,\mathrm{s}$;

ii) Adição de ruído branco gaussiano: Foram geradas variações do conjunto de dados original adicionando ruído branco gaussiano (*additive white Gaussian noise* – AWGN), com relação sinal ruído (*signal-to-noise ratio* – SNR) de $0\,\mathrm{dB}$ até $40\,\mathrm{dB}$, com passos de $5\,\mathrm{dB}$ [10];

[1] https://librosa.org/doc/latest/index.html, acesso em 03/2025

[2] RT60: tempo para a pressão sonora cair 60 dB após a emissão do áudio.

[3] https://github.com/LCAV/pyroomacoustics, acesso em 06/2024

iii) Adição de ruído real: Foram geradas variações do conjunto de dados original com a adição de períodos aleatórios de ruídos da base ESC-50 [18], com valores de SNR de $0\,\mathrm{dB}$ até $40\,\mathrm{dB}$, com passos de $5\,\mathrm{dB}$;

iv) Combinação do efeito de reverberação com ruído branco gaussiano e com ruído real: Foram adicionados os efeitos ii) e iii), separadamente, aos dados com reverberação gerados em i).

2) Variações de gênero: Para verificar se a otimização dos MFCCs afeta a classificação de vozes masculinas e vozes femininas de maneira distinta (relacionada à terceira pergunta norteadora deste trabalho), dois tipos de variações são investigadas:

- Classificação com treino utilizando ambos os gêneros, mas com testes separando-os.
- Classificação com treino e teste separando os gêneros.

D. Etapa de Classificação

Neste estudo, a tarefa de classificação é binária, distinguindo entre indivíduos com vozes saudáveis e aqueles com alguma patologia na laringe. Para realizar essa discriminação, é utilizado um classificador baseado em QDA (*quadratic discriminant analysis*), que é uma técnica de aprendizado de máquina reconhecida por ser uma extensão da LDA (*linear discriminant analysis*), oferecendo maior flexibilidade ao modelar as fronteiras de decisão por meio de funções quadráticas, o que permite capturar relações não lineares nos dados [19]. Para assegurar a robustez e a generalização dos resultados obtidos, foi implementado o método de validação cruzada k-*fold* [20], com o valor de k fixado em 10.

As medidas de acurácia, sensibilidade e especificidade são empregadas para analisar o desempenho dos classificadores. Essas medidas estão relacionadas à capacidade de um classificador em diagnosticar uma doença (sensibilidade), diagnosticar um estado saudável (especificidade), bem como medir seu desempenho global (acurácia) [21].

IV. Resultados

Nesta seção são apresentados os resultados obtidos dos experimentos visando responder às perguntas norteadoras deste estudo. Dessa forma, o desempenho de classificação foi medido em termos da otimização dos MFCCs sem variações acústicas e nas variações de cenas acústicas e de gênero do locutor. A medida de acurácia foi escolhida como a métrica predominante nos resultados a seguir, por se tratar de uma representação global do acerto.

Para fins de comparação, é considerado como *resultado base*, para cada cenário experimental, aquele obtido com a configuração de MFCCs utilizada em [10]: 12 coeficientes, estimados em quadros de 25 ms com sobreposição de 10 ms.

A. Otimização de MFCCs sem variações acústicas

Na Tabela I são apresentados os valores percentuais de acurácia da classificação considerando o banco de dados original (ambiente sem variações acústicas), para todas as configurações de MFCC utilizadas nos experimentos. Há dois valores destacados: o *resultado base* e a que pode ser considerada como a *configuração ótima global* dos MFCCs (32 coeficientes, 30 ms). Ao considerar apenas o valor médio da acurácia, pode-se observar uma melhora de mais de 6 pontos percentuais (p.p.) em relação ao *resultado base*. Tal melhoria também ocorre nas médias de sensibilidade (73,36% do *resultado base* contra 85,75% do melhor resultado da otimização) e especificidade (83,38% do *resultado base* contra 83,71% do melhor resultado da otimização). Note que a otimização teve como consequência uma melhora de mais de 10 p.p. no acerto dos casos patológicos (sensibilidade). Outro ponto de destaque está na variação dos parâmetros aplicados, em que pode-se observar que a acurácia é mais sensível ao aumento da quantidade de coeficientes do que ao aumento do tamanho do quadro. Por exemplo, considerando 12 MFCCs, há uma variação de apenas 1 p.p. na média da acurácia (79,33% $\leftrightarrow$ 78,33%) entre o maior e o menor tamanho de quadro testados.

TABELA I

Acurácia (%) da classificação considerando variação de tamanho de segmento e quantidade de MFCCs.

Quantidade de MFCCs	**Tamanho do Quadro**		
	20 ms	**25 ms**	**30 ms**
12	79,33 ± 0,72	**78,50 ± 0,76**	78,33 ± 1,08
16	81,00 ± 1,01	80,50 ± 1,03	80,17 ± 1,01
20	82,33 ± 0,89	81,83 ± 1,12	82,33 ± 1,11
24	80,33 ± 1,39	81,83 ± 0,98	82,33 ± 0,95
28	82,50 ± 1,42	82,83 ± 0,82	81,83 ± 1,42
32	82,83 ± 1,81	83,17 ± 1,50	**84,67 ± 1,33**

B. Otimização de MFCCs com variações acústicas

1) Resultados com efeitos isolados: Na Figura 1 são apresentados os resultados de acurácia considerando cada um

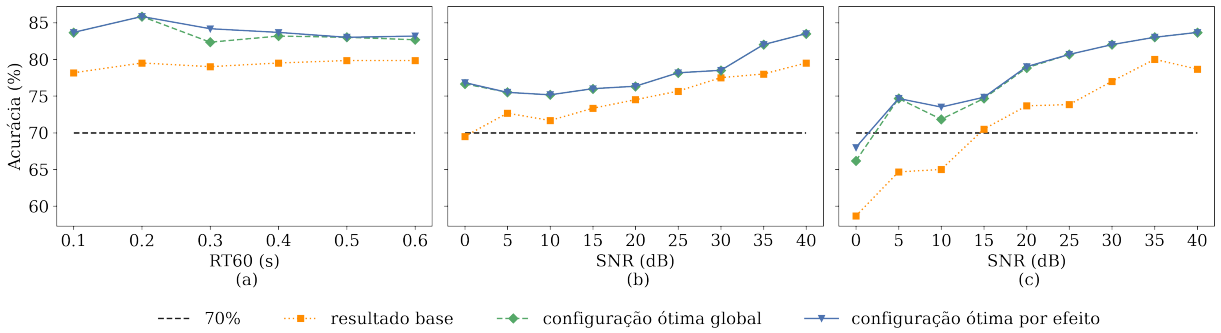

Fig. 1. Comparação dos resultados de classificação do banco de dados variando: (a) RT60; (b) SNR com ruído branco gaussiano; (c) SNR com ruído real.

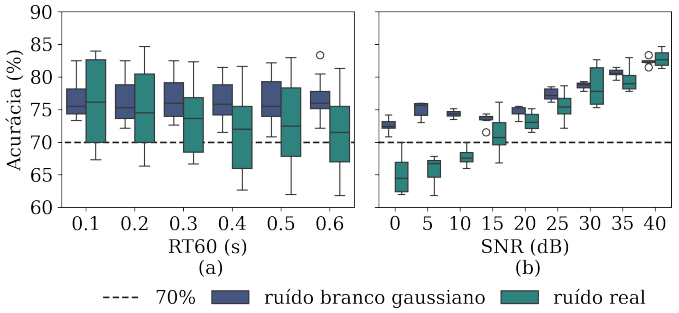

Fig. 2. *Boxplots* com os resultados da classificação considerando mistura de efeitos: (a) perspectiva do RT60; (b) perspectiva da SNR.

dos efeitos aplicados na base de dados (reverberação com RT60, ruído branco gaussiano e ruído real com SNR). Para cada efeito, há o *resultado base*, o resultado obtido com a *configuração ótima global* obtida no ambiente sem variação acústica (32 coeficientes, 30 ms), o melhor resultado obtido do ajuste fino para cada efeito (ou seja, *configuração ótima por efeito*) e, ainda, é traçada a linha no limiar de 70% de acurácia, que é frequentemente considerado um indicador de desempenho aceitável para sistemas de análise acústica que visam auxiliar na avaliação vocal [22]. Em todos os cenários, há um aumento da acurácia em relação ao *resultado base*. Outro ponto importante de destaque é a coerência apresentada pelos MFCCs mesmo com as variações acústicas. Ou seja, na grande maioria dos parâmetros acústicos testados, a configuração ótima de MFCCs foi a mesma que leva o classificador ao melhor desempenho no ambiente sem variações acústicas. Por exemplo, no caso do ruído branco gaussiano, em todos os valores de SNR o maior valor de acurácia foi atingido com 32 coeficientes e 30 ms de quadro. Nos demais casos, quando ocorre alguma diferença de acurácia, ela está relacionada ao tamanho do quadro. Ainda, pode-se observar que, no contexto do limiar de 70% de acurácia, a otimização dos MFCCs é importante nos casos de ruído. Por exemplo, no caso de ruído real, enquanto no *resultado base* o limiar só é ultrapassado com SNR de 15 dB, a medida otimizada consegue superá-lo com SNR de 5 dB.

2) Resultados com combinação de efeitos: Para facilitar a observação dos resultados com a combinação de efeitos, eles são apresentados, na Figura 2, em *boxplots* sob duas perspectivas: distribuição da acurácia para os valores de SNR de ruído considerando cada valor de RT60 individualmente; e a distribuição da acurácia para os valores de RT60 considerando cada valor de SNR individualmente. Nesse contexto, é utilizada a *configuração ótima global* dos MFCCs. Então, é possível observar, por exemplo, que para o ruído real há casos de acurácia abaixo de 70% para todos os valores considerados de RT60 (Figura 2a). Estes, por sua vez, estão associados a valores de SNR entre 0 e 15 dB (Figura 2b). Note que, considerando o ruído branco gaussiano, o desempenho do classificador atinge valores de acurácia acima de 70% para todas as variações dos efeitos combinados. Por outro lado, no que diz respeito à presença do ruído real, é possível observar que a acurácia pode estar mais condicionada ao grau de severidade da variação acústica, ou seja, o classificador apresenta um melhor desempenho (acurácia > 70%) em casos com menor RT60 e maior SNR.

C. Otimização de MFCCs com variação do gênero do locutor

Os experimentos realizados neste contexto buscam verificar o comportamento do classificador com diferentes configurações de MFCCs e, ainda, diferenças que podem existir entre treinar o classificador de maneira generalizada e treiná-lo separando por gênero. Na Figura 3 são apresentados valores de acurácia obtidos do ajuste fino de MFCCs em quatro diferentes casos: treino com todos e teste com vozes masculinas (Figura 3a), treino com todos e teste com vozes femininas (Figura 3b), treino e teste com vozes masculinas (Figura 3c) e treino e teste com vozes femininas (Figura 3d). Para tanto, foi utilizado o conjunto de dados original (sem reverberação e sem ruído). Note que há diferenças entre a acurácia obtida com a configuração do *resultado base* e as demais configurações experimentadas.

A utilização de um classificador generalista (treino com todos) proporciona valores de acurácia mais baixos do que aqueles obtidos com classificadores individualizados para vozes masculinas e femininas. No caso de vozes femininas, é possível notar que um desajuste dos parâmetros MFCC em relação à configuração ótima causa mais prejuízo em termos de desempenho do que no caso de vozes masculinas. Outro ponto de destaque está na utilização de classificadores dedicados por gênero (treino e teste). No caso do gênero masculino, percebe-

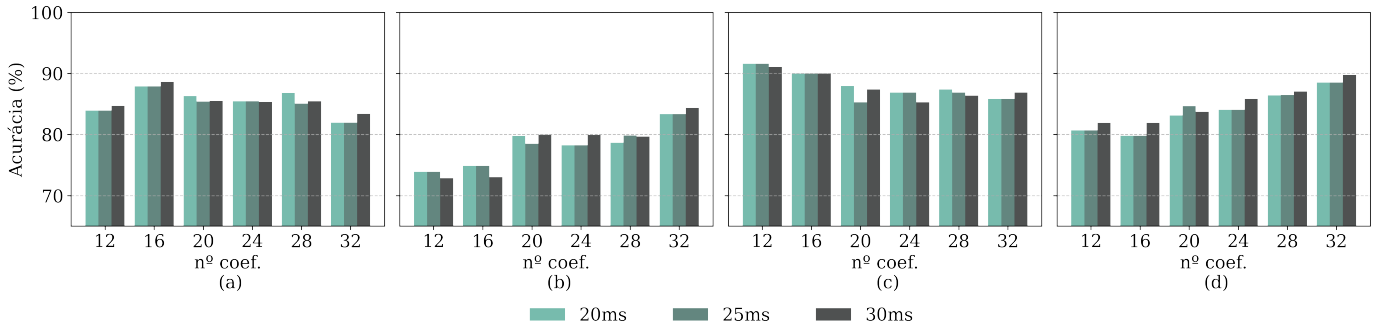

Fig. 3. Desempenho do classificador utilizando: (a) treino com todos e teste com vozes masculinas; (b) treino com todos e teste com vozes femininas; (c) treino e teste com vozes masculinas; (d) treino e teste com vozes femininas.

se que os maiores valores de acurácia são obtidos com uma quantidade menor de MFCCs, 12. Por outro lado, para atingir a máxima acurácia nos experimentos com gênero feminino, é necessário utilizar 32 coeficientes.

A fim de estudar o efeito do uso de um classificador dedicado por gênero, na Tabela II são apresentados os resultados dos experimentos envolvendo a *configuração ótima global*. Em relação aos valores médios de acurácia e especificidade, pode-se observar a melhora no desempenho do classificador quando ele é treinado com um único gênero. Por exemplo, a especificidade para o teste com homens aumenta em aproximadamente 10 p.p. com a utilização de um classificador dedicado. Por outro lado, a sensibilidade cai em aproximadamente 4 p.p.. Contudo, a partir desses experimentos, nota-se que pode ser utilizada a configuração de MFCCs do *resultado base* para um classificador dedicado ao gênero masculino e a *configuração ótima global* no caso do gênero feminino, o que mostra a influência do gênero na otimização dessa medida.

TABELA II
DESEMPENHO DA CLASSIFICAÇÃO COM A CONFIGURAÇÃO ÓTIMA DE MFCCS NOS EXPERIMENTOS COM VOZES MASCULINAS (M) E FEMININAS (F).

Experimento	**Ac. (%)**	**Sens. (%)**	**Esp. (%)**
treino: todos, teste: M	83,39 ± 1,93	88,55 ± 2,65	78,96 ± 4,18
treino: todos, teste: F	84,35 ± 1,80	84,41 ± 2,15	85,43 ± 2,95
treino: M, teste: M	86,87 ± 1,87	84,58 ± 3,99	88,70 ± 2,80
treino: F, teste: F	89,76 ± 1,30	89,67 ± 1,70	89,93 ± 2,48

V. CONCLUSÕES

Neste trabalho, foi realizada uma série de experimentos de ajuste fino na configuração da medida MFCC no contexto de classificação de patologias laríngeas em diferentes cenas acústicas. A classificação, realizada com QDA, mostrou que há uma configuração ótima de MFCCs quando é considerado o banco de dados original. No ajuste fino em ambientes reverberantes e ruidosos, foi observado que esta configuração (32 coeficientes extraídos em quadros de 30 ms) prevaleceu entre os valores mais altos de acurácia. Ou seja, é possível utilizar uma configuração de MFCCs que tenha potencial discriminativo em ambientes livres de variações acústicas e naqueles com presença de reverberação e ruído. Ainda, pôde-se verificar a influência do gênero nos experimentos realizados, de maneira que a configuração ótima, encontrada com o banco de dados original, foi mais robusta no contexto do classificador dedicado ao gênero feminino. Portanto, os resultados deste estudo revelaram que a otimização da medida MFCC na análise de desordens vocais em diferentes cenários é uma tarefa promissora. Ainda, tais achados demonstraram que o uso de classificadores dedicados por gênero pode tornar o sistema ainda mais robusto. Em trabalhos futuros, pretende-se: expandir o domínio dos parâmetros usados na análise, incluindo testes estatísticos para validar a significância das diferenças observadas; classificar patologias laríngeas separadamente; combinar MFCCs com outras medidas acústicas (tais como delta-MFCC e delta-delta-MFCC); e verificar o desempenho de classificação em outras cenas acústicas.

AGRADECIMENTOS

Os resultados apresentados neste artigo foram desenvolvidos como parte de um projeto do SiDi, financiado pela Samsung Eletrônica da Amazônia Ltda., com o apoio da Lei Federal de Informática no. 8248/91.

REFERÊNCIAS

[1] E. R. Manfio, "Avaliação de dispositivos acionados por voz e texto para o português brasileiro," Tese de doutorado, Universidade Estadual de Londrina, 2024.

[2] A. S. Tulshan and S. N. Dhage, "Survey on virtual assistant: Google assistant, siri, cortana, alexa," in *Intern. symposium on signal processing and intelligent recognition systems*. Springer, 2018, pp. 190–201.

[3] P. Oliveira, *et al.*, "Prevalence of work-related voice disorders in voice professionals: systematic review and meta-analysis," *Journal of Voice*, 2022.

[4] S.-S. Wang, *et al.*, "Continuous speech for improved learning pathological voice disorders," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 3, pp. 25–33, 2022.

[5] S. R. de Abreu, *et al.*, "Performance of acoustic measures for the discrimination among healthy, rough, breathy, and strained voices using the feedforward neural network," *Journal of Voice*, 2022.

[6] N. V. Mallipeddi, A. Mehrotra, and J. H. Van Stan, "Telepractice in the treatment of speech and voice disorders: What could the future look like?" *Perspectives of the ASHA Special Interest Groups*, vol. 8, no. 2, pp. 418–423, 2023.

[7] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, p. 357–366, Aug. 1980.

[8] P. Bansal, S. A. Imam, and R. Bharti, "Speaker recognition using mfcc, shifted MFCC with vector quantization and fuzzy," in *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*. IEEE, 2015, pp. 41–44.

[9] M. S. Fahad, *et al.*, "DNN-HMM-based speaker-adaptive emotion recognition using MFCC and epoch-based features," *Circuits, Systems, and Signal Processing*, vol. 40, pp. 466–489, 2021.

[10] V. J. D. Vieira, R. R. Pertum, and R. Candido, "Sobre a influência de distorções acústicas na classificação de patologias laríngeas," in *Anais do XLII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, Sociedade Brasileira de Telecomunicações, 2024.

[11] S. Tirronen, S. R. Kadiri, and P. Alku, "The effect of the MFCC frame length in automatic voice patholyogy detection," *Journal of Voice*, vol. 38, no. 5, p. 975–982, Sep. 2024.

[12] Y. Yan, *et al.*, "Optimizing MFCC parameters for the automatic detection of respiratory diseases," *Applied Acoustics*, vol. 228, Jan. 2025.

[13] S. L. do Nascimento Cunha Costa, "Análise acústica, baseada no modelo linear de produção da fala, para discriminação de vozes patológicas," Tese de doutorado, Programa de Pós-Graduação em Engenharia Elértica, Universidade Federal de Campina Grande, 2008.

[14] V. J. Vieira *et al.*, "Exploiting nonlinearity of the speech production system for voice disorder assessment by recurrence quantification analysis," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, p. 085709, 2018.

[15] D. O'shaughnessy, *Speech communication: human and machine*. Universities press, 1987.

[16] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*. Pearson Higher Education, 2010.

[17] B. Woldert-Jokisz, "Saarbruecken voice database," 2007.

[18] K. J. Piczak, "ESC: Dataset for Environmental Sound Classification," in *Proceedings of the 23rd Annual ACM Conference on Multimedia*. ACM Press, pp. 1015–1018.

[19] T. Hastie, R. Tibshirani, J. Friedman, "The elements of statistical learning," 2009.

[20] T.-T. Wong and P.-Y. Yeh, "Reliable accuracy estimates from k-fold cross validation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 8, pp. 1586–1594, 2019.

[21] L. Lopes, V. Vieira, and M. Behlau, "Performance of different acoustic measures to discriminate individuals with and without voice disorders," *Journal of Voice*, vol. 36, no. 4, pp. 487–498, 2022.

[22] L. W. Lopes, J. d. N. Alves, D. d. S. Evangelista, F. P. França, V. J. D. Vieira, M. F. B. d. Lima-Silva, and L. d. A. Pernambuco, "Acurácia das medidas acústicas tradicionais e formânticas na avaliação da qualidade vocal," in *CoDAS*, vol. 30. SciELO Brasil, 2018, p. e20170282.