

Avaliação de Medidas Acústicas na Detecção da Depressão Por Meio da Fala

Laura Sousa, Laura Simões, Fabiana Macêdo, Rodrigo Silva, Felipe Nascimento e Suzete Correia

Resumo— Este trabalho avalia o potencial discriminativo de medidas acústicas para detecção do Transtorno Depressivo Maior, em mulheres jovens e idosas, por meio da fala. Foram analisadas oito medidas relacionadas à prosódia, qualidade vocal e parâmetros cepstrais. Após remoção automática de pausas, as características acústicas foram extraídas e submetidas a testes estatísticos. Os resultados indicaram que o envelhecimento vocal modifica o comportamento acústico associado à depressão. As medidas de instabilidade vocal *jitter* e *shimmer* apresentaram maior relevância principalmente nas vozes de mulheres jovens.

Palavras-Chave— *Análise Acústica, Análise estatística, Processamento de Sinais.*

Abstract— This work evaluates the discriminative potential of acoustic measures for depression detection in young and elderly women through speech analysis. Eight measures related to prosody, vocal quality, and cepstral parameters were analyzed. After automatic pause removal, the acoustic features were extracted and subjected to statistical tests. The results indicated that vocal aging modifies the acoustic behavior associated with depression. The vocal instability measures *jitter* and *shimmer* showed greater relevance mainly in the voices of young women.

Keywords— *Acoustic analysis, statistical analysis, depression, signal processing.*

I. INTRODUÇÃO

A fala se destaca por ser um marcador fisiológico obtido de maneira não invasiva, rápida e de fácil coleta [1]. Estudos indicam que no Transtorno Depressivo Maior (depressão) mudanças neurofisiológicas podem modificar a dinâmica da laringe e o controle da vibração das pregas vocais [2]. Além disso, alterações cognitivas e fisiológicas observadas em pessoas deprimidas interferem diretamente no processo de produção vocal, ocasionando fala lenta e pausada, com baixa intensidade e hesitações frequentes [2,3].

Uma das alternativas para se avaliar de maneira objetiva as alterações vocais relacionadas à depressão é por meio da análise acústica da voz. Essa técnica emprega métodos computacionais capazes de quantificar diferentes propriedades do sinal de voz, permitindo a identificação de parâmetros acústicos que refletem alterações no padrão vocal de indivíduos acometidos pelo transtorno [1].

Podem ser encontrados trabalhos que empregam medidas associadas à prosódia, que refletem o ritmo e entonação, tais como frequência fundamental, taxa de cruzamento por zero, duração no tempo e energia. Medidas baseadas no espectro, relacionadas à amplitude/energia em diferentes frequências,

como HNR (*Harmonic to Noise Ratio*), centróide e declínio espectral. Medidas cepstrais, relacionada à análise do Cepstrum como CPPS (*Cepstral Peak Prominence Smoothed*) e coeficientes MFCC (*Mel Frequency Cepstral Coefficient*). Como também medidas que medem informações sobre a qualidade vocal, como formantes, *jitter*, *shimmer* e GNE (*Glottal to Noise Excitation*) [1-4].

Diversos estudos têm examinado a viabilidade e a validade da avaliação automatizada da depressão por meio de medidas acústicas. No entanto, a maioria das pesquisas analisam a relação entre depressão e as medidas sem considerar diferenças de idade ou sexo, incorporando-as em um único modelo [2]. Poucos trabalhos abordam o desempenho em sistemas baseados na fala para depressão em idosos, bem como para homens e mulheres, separadamente. Há mudanças na frequência fundamental de vozes masculinas e femininas, como também com a idade a voz se torna mais fraca e trêmula, devido à atrofia da musculatura da laringe e redução da elasticidade [5].

Lee et al. [4] usaram dois conjuntos de medidas acústicas, AVEC 2013 e eGeMAPS, para a detecção da depressão na voz de pessoas idosas. O algoritmo de aprendizado de máquina AdaBoost foi empregado na classificação. As características acústicas que apresentaram desempenho discriminatório significativo para os homens foram as espectrais e as relacionadas à energia, alcançando uma acurácia de 86%. Para as mulheres, destacaram-se as características prosódicas, com acurácia de 77%.

Neste trabalho foi feita uma avaliação de oito medidas acústicas, frequência fundamental, energia, três primeiros formantes, *jitter*, *shimmer* e CPPS, para identificação da depressão a partir da fala de mulheres jovens e idosas. Em seguida, foi conduzida uma análise estatística baseada em testes de hipótese para investigar quais dessas medidas apresentam potencial discriminativo na identificação da depressão em vozes de mulheres jovens e na terceira idade.

II. MEDIDAS ACÚSTICAS

O modelo linear da produção da fala consiste em duas etapas principais: uma fonte sonora, gerada na glote, que emite um trem de impulsos unitários para sinais sonoros, cuja periodicidade é determinada pelo período de *pitch* T_0 ; e um filtro acústico (ressoador), que corresponde ao trato vocal, responsável por moldar o espectro da voz [2].

Características prosódicas descrevem variações de tom, intensidade, duração e ritmo da fala, podendo ser influenciadas por condições neurológicas e estados psicológicos [5]. O tom e a intensidade são as medidas, comumente usadas. A percepção de quão agudo ou grave é um som pode ser obtida em termos de frequência fundamental F_0 , que é a componente de frequência mais baixa da vibração das pregas vocais, correspondendo ao

Laura Sousa, Laura Simões, Fabiana Macêdo, Rodrigo Silva, Felipe Nascimento e Suzete Correia, Unidade Acadêmica 3, João Pessoa, IFPB, e-mails: macaiba.laura@academico.ifpb.edu.br, laura.wanderley@academico.ifpb.edu.br, macedo.fabiana@academico.ifpb.edu.br, silva.serafim@academico.ifpb.edu.br, felipe-nascimento.fn@academico.ifpb.edu.br, suzete@ifpb.edu.br. Este trabalho foi parcialmente financiado por IFPB.

inverso do período de *pitch* T_0 . Para mensurar a intensidade sonora, a energia do sinal de voz pode ser empregada, a qual é calculada através da soma do quadrado de cada amostra ao longo do tempo [2].

As características da qualidade vocal refletem a configuração do trato vocal. As frequências formantes são as frequências de ressonância e correspondem aos picos nos quais a energia acústica de um sinal de fala está concentrada, caracterizando aspectos articulatórios da fala. O formante F1 reflete a altura da língua e abertura da boca, o F2 a posição horizontal da língua e o F3 está associado à geometria da cavidade oral e ao arredondamento labial [2,5].

Durante a fonação, as pregas vocais vibram periodicamente produzindo ciclos sucessivos de abertura e fechamento. O *jitter* mede a variação ciclo a ciclo do período fundamental da voz, enquanto o *shimmer* mede a variação ciclo a ciclo da sua amplitude. Ambas as medidas refletem as instabilidades na vibração das pregas vocais [2].

O parâmetro CPPS, refere-se a Proeminência do Pico Cepstral Suavizado, calculado a partir do cepstro do sinal de fala. É usado para avaliar o grau de periodicidade e organização da voz, indicando quão forte e regular é a vibração das pregas vocais. O conceito matemático é baseado na diferença entre a amplitude do pico cepstral principal e uma linha de regressão do cepstro [1].

III. MATERIAIS E MÉTODOS

A Figura 1 ilustra o diagrama em blocos da metodologia empregada.



Fig. 1. Diagrama em blocos da metodologia.

Os sinais de voz empregados neste estudo são provenientes de uma base de dados disponibilizada na plataforma Kaggle, composta por 800 áudios em língua inglesa, distribuídos igualmente em quatro grupos: 200 vozes de mulheres jovens sem depressão, 200 vozes de mulheres jovens com depressão, 200 vozes de mulheres idosas sem depressão e 200 vozes de mulheres idosas com depressão. Os arquivos encontram-se no formato wav e foram padronizados com uma taxa de amostragem de 24,4 kHz e 16 bits/amostra.

Para garantir que as amostras de áudio refletissem apenas a dinâmica das pregas vocais e não fossem influenciados por períodos de silêncio, as pausas foram removidas por meio de um processo de segmentação automática do sinal. Configurou-se um limiar de silêncio de intensidade -25dB, com duração mínima de pausa e duração mínima de som de 0,2 e 0,1 segundos, respectivamente.

A extração das medidas acústicas de frequência fundamental F_0 , energia, formantes F1, F2 e F3, *jitter*, *shimmer* e CPPS foi realizada no software Praat [6], empregando os parâmetros padrão de cada algoritmo de análise. Cada medida acústica foi

estimada por meio de um procedimento específico, utilizando janelas de análise com diferentes durações, tipos e níveis de sobreposição, de acordo com a metodologia implementada pelo software.

A fim de verificar a existência de possíveis diferenças estatísticas entre os grupos com e sem depressão, para mulheres jovens e na terceira idade, foram realizados testes de hipótese, admitindo um nível de significância de 0,05. Para tanto, empregou-se a ferramenta Jamovi [7].

IV. RESULTADOS

Os *boxplots* apresentados na Figura 2, representam a distribuição das medidas acústicas para as vozes jovens e idosas, com e sem depressão. Os valores 50% centrais da distribuição (intervalo entre o primeiro e o terceiro quartil) estão representados por uma caixa e a mediana (segundo quartil) está representada por uma linha horizontal dentro dessa caixa.

O *jitter* apresentou valores maiores, de forma mais evidente, no grupo de mulheres jovens depressivas, entretanto, nas vozes idosas o aumento não foi tão significativo, o que pode ser devido à perda do tônus muscular da laringe. Já o *shimmer* teve um aumento de valor na fala de mulheres jovens com depressão. Observou-se que nas idosas sem depressão, o *shimmer* apresenta valor elevado em relação às mulheres jovens na mesma condição, mantendo-se inalterado no quadro depressivo.

Por sua vez, a medida de energia teve um decréscimo em mulheres depressivas, sendo mais expressivo nas mulheres jovens, o que pode estar relacionado à lentidão na fala. O CPPS se apresentou como uma das medidas com maior potencial, apresentando redução visual nas mulheres depressivas, independentemente da idade.

A frequência fundamental sofreu um aumento nas vozes com depressão, devido à fonação mais tensa. O formante F1 apresentou redução de valor para as mulheres jovens deprimidas devido a uma articulação menos expansiva. O formante F2 não apresentou mudanças significativas e o formante F3 apresentou aumento no quadro depressivo, tanto para as jovens quanto para as idosas.

A fim de verificar a existência de possíveis diferenças estatísticas significativas entre os grupos de mulheres jovens e na terceira idade, definidos por classe (normal vs. depressão), foi realizado testes de hipótese. Inicialmente, foi realizada a análise de normalidade por meio do teste de Kolmogorov-Smirnov, considerando que todos os grupos apresentaram tamanho amostral superior a 50 observações. As variáveis que apresentaram distribuição aproximadamente normal foram analisadas por meio do teste paramétrico t para amostras independentes, enquanto as variáveis não normais foram submetidas ao teste não paramétrico U de Mann-Whitney.

Para o grupo de mulheres jovens, observou-se distribuição normal para F1, F2, F3 e CPPS, enquanto F_0 , *jitter*, *shimmer* e energia apresentaram distribuição não normal. Nas mulheres idosas, F1, F2, CPPS, F_0 e *shimmer* apresentaram distribuição normal, enquanto F3, *jitter* e energia apresentaram desvio da normalidade.

Os testes comparativos evidenciaram diferenças estatisticamente significativas entre mulheres com depressão e mulheres saudáveis em praticamente todas as variáveis analisadas. No grupo jovem, as variáveis F1, F3, CPPS, F_0 , *jitter*,

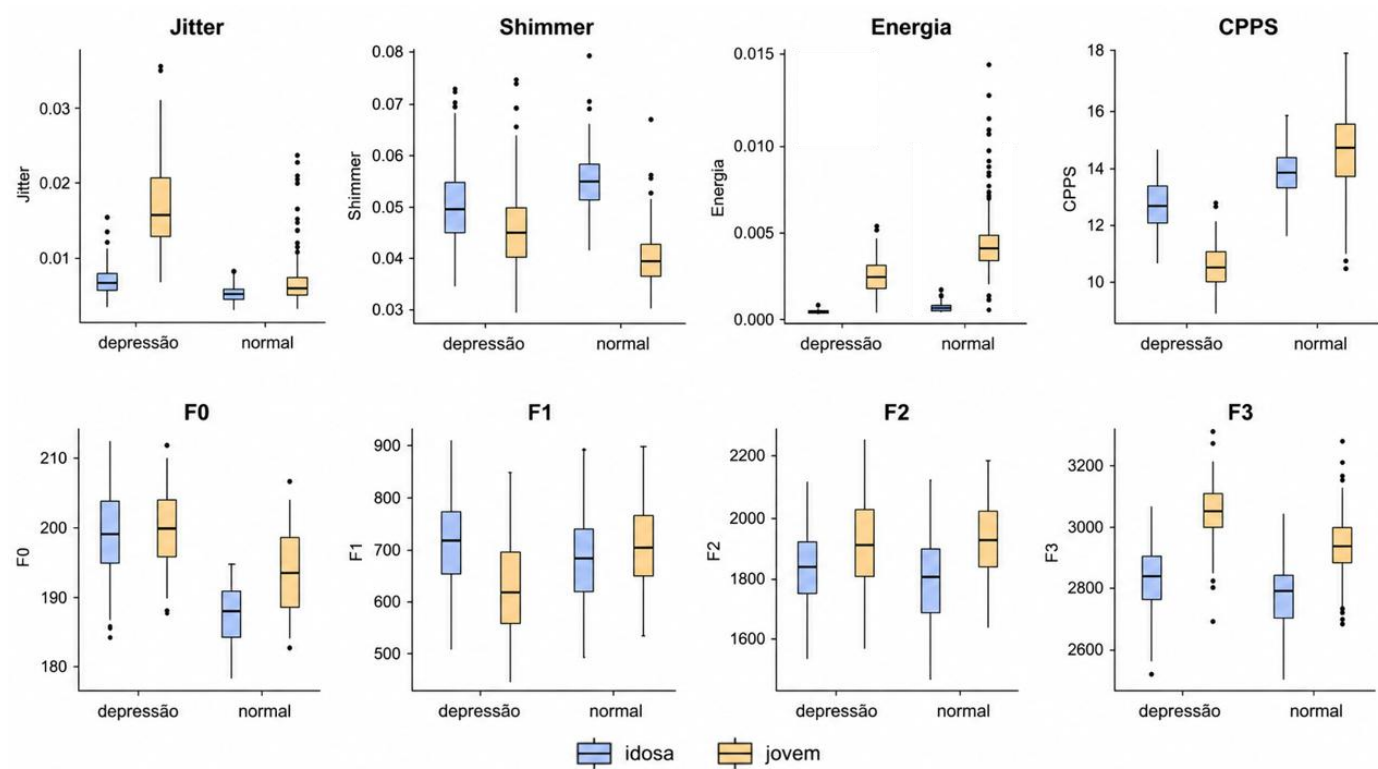


Fig. 2. Boxplots das medidas acústicas.

shimmer e energia apresentaram diferenças significativas entre os grupos com $p < 0,05$, enquanto apenas o formante F2 não apresentou diferença significativa. Já no grupo idoso, todas as variáveis analisadas apresentaram diferenças estatisticamente significativas.

A análise descritiva demonstrou diferenças importantes entre jovens e idosas. Mulheres jovens com depressão apresentaram média de F_0 mais elevada (200 Hz) quando comparadas às jovens saudáveis (193 Hz), além de maiores valores médios de *jitter*. Nas mulheres idosas, observou-se redução da frequência fundamental média nas participantes saudáveis (186 Hz) em comparação às idosas com depressão (199 Hz), indicando que o envelhecimento vocal pode modificar o comportamento acústico associado aos sintomas depressivos.

V. CONCLUSÕES

Os resultados obtidos sugerem que a depressão influencia significativamente as características acústicas da voz. Além disso, observou-se que o envelhecimento vocal modifica parte desses parâmetros, indicando que mulheres jovens e idosas com depressão apresentam comportamentos acústicos distintos. Das medidas avaliadas, as mais promissoras para detectar depressão em mulheres idosas são energia, CPPS, F3 e F_0 . Além dessas, para as mulheres jovens podem ser consideradas as medidas de perturbação *jitter* e *shimmer*.

Em trabalhos futuros, medidas baseadas nos coeficientes MFCC serão consideradas, bem como a classificação empregando aprendizado de máquina.

AGRADECIMENTOS

Ao IFPB pelo suporte financeiro.

REFERÊNCIAS

- [1] W. Silva, L. Lopes, M. Galdino e A. Almeida, "Voice Acoustic Parameters as Predictors of Depression", *Journal of Voice*, v. 38, pp. 77-85, Janeiro, 2024.
- [2] S. Almaghrabi, S. Clark e M. Baumert, "Bio-acoustic features of depression: A review", *Biomedical Signal Processing and Control*, v. 85, 2023.
- [3] E. Rejaibi, A. Komaty, F. Meriaudeau, S. Agrebi e A. Othmani, "MFCC-based Recurrent Neural Network for automatic clinical depression recognition and assessment from speech", *Biomedical Signal Processing and Control*, v. 71, 2022.
- [4] S. Lee et al., "Screening major depressive disorder using vocal acoustic features in the elderly by sex", *Journal of Affective Disorders*, v. 291, 2021.
- [5] M. Behlau, *Voz: o livro do especialista*. Rio de Janeiro: Revinter, 2001.
- [6] P. Boersma, D. Weenink e A. Schupak, *Praat: doing phonetics by computer* [Computer program]. Version 6.4.67.
- [7] THE JAMOVİ PROJECT. jamovi (Version 2.6) [Software de computador]. Disponível em: <https://www.jamovi.org>. Acesso em: 26 de maio de 2026.