

Avaliação da Qualidade Vocal Empregando a Decomposição Variacional de Modos

Catarina L. L. Pedrosa, Máisa L. F. Santos, Silvana C. Costa, Suzete E. N. Correia e Vinícius J. D. Vieira

Resumo— Este trabalho emprega a Decomposição Variacional de Modos na avaliação da qualidade vocal. Duas medidas são obtidas das Funções de Modo Intrínseco em oito níveis de resolução. Os resultados apontam que a energia nas baixas frequências e a dispersão da frequência instantânea, nas altas frequências, permitem a separação entre sinais de vozes saudáveis e com desvios vocais de rugosidade e sopro.

Palavras-Chave— Análise Acústica, Qualidade vocal, Desvios vocais, Decomposição Variacional de Modos.

Abstract— This work uses Variational Mode Decomposition to evaluate vocal quality. Two measurements are obtained from the Intrinsic Mode Functions at eight levels of resolution. The results indicate that the energy in low frequencies and the instantaneous frequency dispersion, in high frequencies, allow the separation between healthy voice signals and those with vocal deviations of roughness and breathiness.

Keywords— Acoustic Analysis, Vocal quality, Vocal deviations, Variational Mode Decomposition.

I. INTRODUÇÃO

Voz, articulação e linguagem são os elementos mais importantes da produção da fala humana. Quando um distúrbio relacionado a qualquer desses elementos se desenvolve, a habilidade de comunicação verbal fica comprometida [1].

No ambiente clínico, as alterações vocais mais frequentes podem ser classificadas como: vozes rugosas, quando a impressão auditiva é de irregularidade fonatória, com rouquidão e aspereza; e sopro, quando há presença de ar na emissão vocal, que corresponde fisiologicamente a abertura entre as pregas vocais. A identificação desses desvios vocais é uma tarefa difícil em determinados casos, tendo em vista que podem ocorrer de forma conjunta na fala de um mesmo paciente [2].

Ao analisar vozes disfonias, usando a análise tempo-frequência, pesquisadores buscam por características que indiquem irregularidades no conteúdo da frequência, variações no tom e na intensidade. Dragomiretskiy e Zosso [3] propuseram uma técnica de processamento, baseada na Decomposição Empírica de Modos (EMD – *Empirical Mode Decomposition*), denominada Decomposição Variacional de Modos (VMD – *Variational Mode Decomposition*), que determina as frequências relevantes do sinal, de forma mais precisa, por decompor o sinal em um conjunto de subsinais (modos) quase ortogonais com banda limitada. Esse método é particularmente útil para análise de sinais que apresentam comportamentos complexos ao longo do tempo e irregularidades como a presença

Catarina L. L. Pedrosa; Máisa L. F. Santos; Silvana C. Costa e Suzete E. N. Correia, Unidade Acadêmica de Indústria, Instituto Federal da Paraíba (UA3/IFPB). Vinícius J. D. Vieira, Depto. de Processamento de Voz, SiDi, Alphaville, Campinas, SP - Brasil; E-mails: {catarinalenicelp@gmail.com, maisa.lauriane@academico.ifpb.edu.br, silvana@ifpb.edu.br, suzete@ifpb.edu.br, vinicius.dv@sidi.org.br}

de ruídos. Na classificação de sinais de voz, a VMD tem sido empregada na detecção da doença de Parkinson [4], na análise de vozes com resfriado [5] e no reconhecimento de emoções por meio da fala [6].

O presente trabalho visa avaliar a aplicabilidade do método da Decomposição Variacional de Modos para a análise acústica de sinais de vozes saudáveis e com desvios de rugosidade e sopro.

II. DECOMPOSIÇÃO VARIACIONAL DE MODOS

A VMD decompõe um sinal em um número finito de modos, denominados Funções de Modo Intrínseco (IMFs – *Intrinsic Modal Functions*), que correspondem a componentes de uma modulação AM-FM. O k -ésimo modo é definido como:

$$s_k(t) = A_k(t)\cos(\phi_k(t)), \quad (1)$$

em que A_k é a amplitude instantânea, ϕ_k a amplitude instantânea e sua derivada w_k a frequência instantânea. Como cada IMF está centrada em uma frequência instantânea distinta, obtém-se com essa decomposição um conjunto de subsinais com banda limitada, que correspondem a um banco de filtros [3].

III. MATERIAIS E MÉTODOS

Os sinais utilizados são provenientes de um banco de vozes sintetizadas, disponibilizado em [7]. Tratam-se de emissões da vogal sustentada /ε/, amostradas a uma taxa de 44.100 Hz e quantizadas com 16 bits por amostra. Foram selecionados 30 sinais, sendo 10 considerados saudáveis, 10 com o desvio de rugosidade e 10 com o desvio de sopro.

A VMD foi aplicada a cada um dos sinais de voz, empregando-se a decomposição em 8 IMFs. Foram extraídas duas características de cada IMF, uma temporal e uma espectral. A característica temporal extraída foi a energia, que representa a sua intensidade sonora:

$$E_k = \sum_{n=0}^{N-1} s_k^2(n), \quad (2)$$

em que N representa o número de amostras e $s(n)$ o sinal em cada k IMF.

A característica espectral obtida foi a Dispersão da Frequência Instantânea (DFI), que calcula a distância de uma frequência instantânea w_k para o seu valor esperado [2]:

$$DFI_k = \frac{1}{N} \sum_{n=0}^{N-1} |w_k(n) - \overline{w_k}|^2. \quad (3)$$

IV. RESULTADOS

As Figuras 1 e 2 apresentam a distribuição das medidas de energia, dos sinais de vozes saudáveis e com distúrbios de

rugosidade e soprosidade, respectivamente, para cada IMF. Observa-se que as IMFs 7 e 8 se destacam, fornecendo faixa de valores de energia distintos, para sinais saudáveis e com os desvios vocais. Sinais desviados tem irregularidades na fala, podendo ter intensidades diminuídas, o que justifica menores valores de energia.

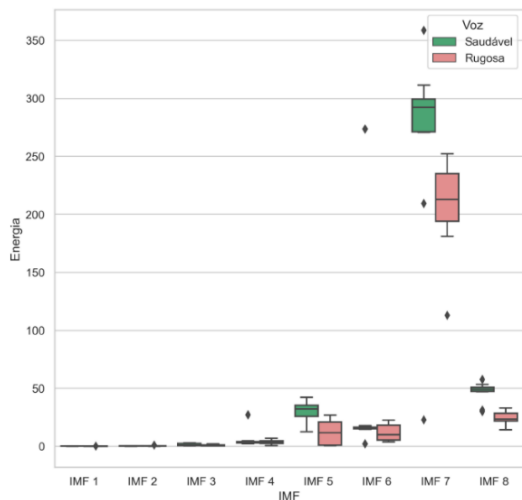


Fig. 1. Energia sinais de vozes saudáveis e com desvio de rugosidade.

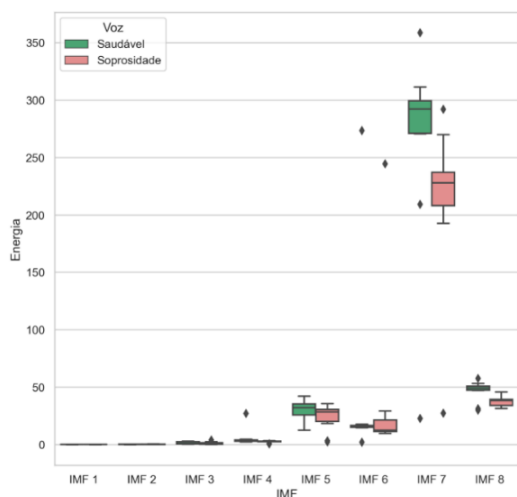


Fig. 2. Energia sinais de vozes saudáveis e com desvio de soprosidade.

A variabilidade da medida DFI para os sinais de vozes saudáveis e com distúrbios de rugosidade e soprosidade podem ser vistos nas Figuras 3 e 4, respectivamente. É possível identificar que as medianas são maiores, para a primeira IMF, responsável pela frequência mais alta, em ambos os distúrbios. Para o desvio soprosidade, há uma separação entre as classes, o que pode ser devido à presença de ruídos de alta frequência, nesse tipo de disfonia.

V. CONCLUSÕES

A Decomposição Variacional de Modos mostrou-se capaz de detectar as disfonias de rugosidade e soprosidade, apresentando potencial para ser usada como ferramenta na análise acústica. A medida de energia, nas baixas frequências, foi capaz de prover a separação entre as classes de sinais saudáveis e com desvios vocais. A medida da dispersão da frequência instantânea, permitiu a separação entre as vozes saudáveis e soprosas.

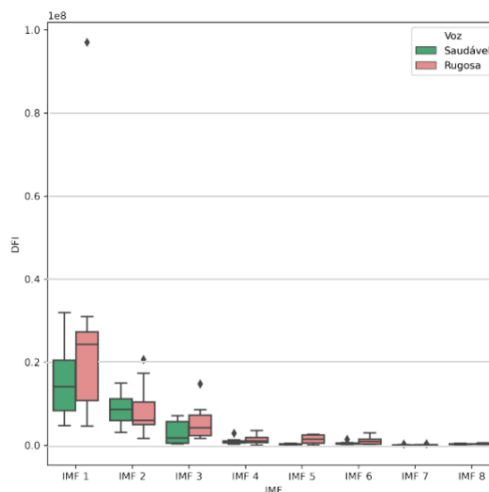


Fig. 3. DFI sinais de vozes saudáveis e com desvio de rugosidade

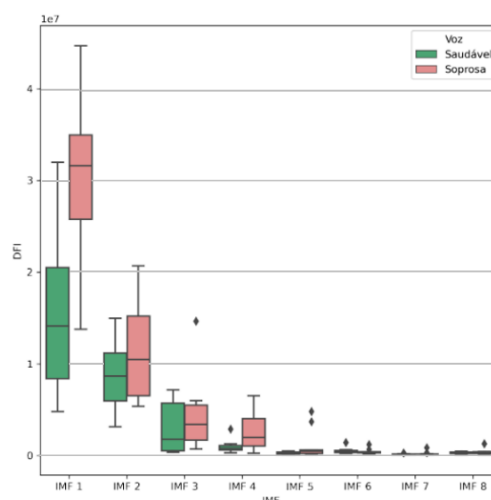


Fig. 4. DFI sinais de vozes saudáveis e com desvio de soprosidade.

AGRADECIMENTOS

Este trabalho não teria sido possível sem o apoio do Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) e do Instituto SiDi.

REFERÊNCIAS

- [1] J. C. Stemple; N. Roy; B. K. Klaben, *Clinical voice pathology: Theory and management*. Plural Publishing, 2018.
- [2] Behlau et al., "Reducing the GAP between science and clinic: lessons from academia and professional practice - part A: perceptual-auditory judgment of vocal quality, acoustic vocal signal analysis and voice self-assessment". In: *Codas*, 2022.
- [3] K. Dragomiretskiy; D. Zosso, "Variational mode decomposition". In: *IEEE Transactions on Signal Processing*, 2013.
- [4] B. Karan.; S. S. Sahu, "An improved framework for Parkinson's disease prediction using Variational Mode Decomposition-Hilbert spectrum of speech signal". In: *Biocybernetics and Biomedical Engineering*, 2021.
- [5] S. Deb; S. Dandapat; J. Krajewski, "Analysis and Classification of Cold Speech Using Variational Mode Decomposition". In: *IEEE Transactions on Affective Computing*, 2020.
- [6] S. Mishra; P. Warule; S. Deb, "Variational mode decomposition based acoustic and entropy features for speech emotion recognition". In: *Applied Acoustics*, 2023.
- [7] SimuVox. "https://lucerojc.github.io/simuvox/" (acessado Maio. 24, 2024).

