Decomposição Tempo-Frequência Adaptativa de Sinais Acústicos com Variações Emocionais

V. Vieira, R. Coelho e F. M. de Assis

***Resumo*— Este trabalho apresenta uma análise acústica baseada na decomposição empírica de modos (EMD - *Empirical Mode Decomposition*) em sinais de voz com presença de variações provocadas por estados emocionais. O estudo investiga o índice de não-estacionariedade e emprega a distância Bhattacharyya como medida de separabilidade destes sinais. Os resultados demonstraram que a EMD é uma interessante ferramenta para análise de sinais com variações acústicas emocionais. Este conceito permitiu evidenciar a separação acústica dos sinais em função de suas faixas de densidade espectral. Além disso, os resultados da análise evidenciaram as componentes das diferentes variações acústicas emocionais.**

***Palavras-Chave*— Análise acústica, Decomposição Empírica de Modos, Variações emocionais.**

***Abstract*— This paper presents an acoustic analysis based on empirical mode decomposition (EMD) of speech signals with the presence of variations caused by emotional states. The study investigates the index of non-stationarity and employs the Bhattacharyya distance as a separability measure of these signals. The results show that EMD is an interesting tool for analyzing signals with acoustic emotional variations. This concept has highlighted the acoustic separation of the signals according to their spectral density ranges. Moreover, the results of the analysis showed the different components of acoustic emotional variations.**

***Keywords*— Acoustic analysis, Empirical Mode Decomposition, Emotional variations.**

I. Introdução

A voz, considerada como sendo o meio mais natural do ser humano, contém informações a respeito do locutor, seja em relação à sua identidade ou ao conteúdo linguístico (ou sentido literal) da sentença pronunciada. Recentemente, aspectos prosódicos têm sido observados na emissão sonora, fazendo com que as variações emocionais presentes nos sinais de voz fossem estudadas [1-4]. Dentre as aplicações biométricas relacionadas à análise acústica de emoções estão: sistemas de segurança; sistemas de bordo de veículos nos quais a informação do estado mental do condutor é útil para a sua própria segurança; sistemas de tradução automática em conversações, nas quais é útil que se entenda o estado emocional de ambas as partes; sistemas de *call centers*; entre outras [1].

O estudo a respeito das variações emocionais faz com que a análise acústica volte-se não apenas para questões como "quem fala?" ou "o que é falado?", mas também procura saber "como é falado". O estado emocional do locutor constitui de uma experiência subjetiva, que pode ou não influenciar em seu reconhecimento, o que faz da tarefa de análise de emoções por meio de sinais de voz mais um desafio da biometria, que, para tanto, deve levar em consideração questões como: variabilidade acústica introduzida pela existência de diferentes sentenças, locutores e estilos de fala; presença de mais de uma emoção em uma mesma sentença pronunciada; quando há a presença de mais de uma emoção, uma delas pode durar mais que outra; entre outros [1].

Um dos principais pontos de partida para a análise acústica de emoções é observar aspectos fisiológicos estreitamente relacionados a variações emocionais. Atividade muscular, diâmetro da pupila, pressão sanguínea e batimentos cardíacos são alguns destes aspectos. Emoções como raiva e felicidade, por exemplo, acarretam aumento na pressão sanguínea, nos batimentos cardíacos e alteração nos movimentos respiratórios [1, 2]. No contexto da produção vocal, alterações deste tipo provocam mudanças no fluxo glotal [5], tendo como consequência alterações na densidade espectral de potência do sinal de voz resultante. Sinais de voz que não apresentam emoção (estado neutro), apresentam uma densidade espectral de potência com queda de 12dB/oitava. A variação de 3dB/oitava (para mais ou para menos) do estado neutro define os estados emocionais como sendo de alta e de baixa ativação (9dB/oitava e 15dB/oitava, respectivamente). Sinais de voz de alta ativação apresentam maior variação e, portanto, verifica-se concentração superior de energia nas altas frequências. Para a ocorrência de estados emocionais de baixa ativação, o sinal apresenta pouca mudança e desta forma exibe maior concentração de energia nas baixas frequências [6, 7].

Fatores inerentemente relacionados à voz podem ser observados em conjunto com os aspectos fisiológicos para tornar mais robusta a análise das variações emocionais introduzidas na fala. Entre estes fatores estão: as diferentes componentes de frequência (cada uma com sua própria energia) que, do fluxo glotal, tomam forma no trato vocal [8, 9]; e a não-estacionariedade dos sinais de voz em trechos superiores a 40 ms [10, 11].

No contexto dos fatores supracitados, se faz necessário que sejam empregadas técnicas de análise acústica que melhor representem as variações emocionais. A Decomposição Empírica de Modos (EMD - *Empirical Mode Decomposition*) foi proposta como uma forma não-linear e adaptativa para análise tempo-frequência de sinais não-estacionários [12], e tem sido utilizada em análise de ruídos e realce de sinais de voz [13-15]. O método EMD permite a observação do sinal em diferentes componentes

V. Vieira, doutorando no Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande (PPgEE/UFCG) e bolsista CNPq; R. Coelho, Departamento de Engenharia Elétrica, Instituto Militar de Engenharia (IME); F. M. de Assis, Departamento de Engenharia Elétrica, UFCG, Brasil. E-mails: {vinicius.vieira@ee.ufcg.edu.br, coelho@ime.eb.br, fmarcos@dee.ufcg.edu.br}. Este trabalho foi parcialmente financiado pelo CNPq 307866/2015-7.

de frequência, as chamadas Funções Intrínsecas de Modo (IMFs - *Intrinsic Mode Functions*), que combinadas garantem a perfeita reconstrução do sinal. Essa é uma vantagem em relação a métodos tradicionais, a exemplo da análise wavelets, na qual a decomposição é realizada utilizando funções base fixas, que não necessariamente conseguem representar o sinal de voz [16]. Além disso, na decomposição wavelet é preciso que seja feita a escolha das funções base que são mais apropriadas para um determinado tipo de sinal. Já no método EMD, as IMFs obtidas da decomposição são totalmente dependentes do sinal.

Neste trabalho, a EMD é empregada em sinais de voz com 5 estados emocionais: Raiva, Felicidade, Tédio, Tristeza e Neutro. São analisadas as concentrações de energia ao longo de 6 IMFs. A separação entre as emoções é quantificada por meio da distância de Bhattacharyya, e o índice de não-estacionariedade dos sinais é calculado. Os resultados dos experimentos indicam a eficiência da técnica EMD em dividir os sinais de voz em faixas de frequência, algumas das quais permitem a separação das variações emocionais de acordo com a concentração de energia. As medidas de distância e do grau de não-estacionariedade dos sinais de voz comprovam o potencial de discriminação dos estados emocionais por meio da análise tempo-frequência.

O restante deste trabalho está organizado da seguinte forma: Na Seção II é apresentada uma breve descrição teórica do método da decomposição empírica de modos. Na Seção III são apresentados os parâmetros empregados na análise acústica deste trabalho. Na Seção IV são apresentados os resultados obtidos nos experimentos realizados, e na Seção V são apresentadas as considerações finais.

II. Decomposição Empírica de Modos

A ideia principal da decomposição empírica de modos é analisar um sinal $x(t)$ contendo dois extremos consecutivos (máximo ou mínimo) nos pontos t_- e t_+, definindo uma componente de altas frequências do sinal, também chamada de detalhe $d(t)$, e uma componente de tendência local, ou resíduo $r(t)$, tal que

$$x(t) = d(t) + r(t)\,,\; t_- \leq t \leq t_+\,. \tag{1}$$

O conjunto dos detalhes locais, obtidos de todos os extremos consecutivos de $x(t)$, compõe a primeira IMF (primeiro modo da decomposição). A separação entre componentes de altas e de baixas frequências é repetida de forma iterativa sobre o sinal residual $r(t)$, chegando-se a um conjunto de IMFs e a um resíduo de baixas frequências.

Para a realização da decomposição empírica de modos sobre um sinal $x(t)$, os seguintes passos são seguidos [12, 13]:

1) Identificar todos os extremos de $x(t)$, ou seja, os pontos de máximo $x_{max}(t)$ e mínimo $x_{min}(t)$ locais;
2) Obter as envoltórias $e_{max}(t)$ e $e_{min}(t)$, interpolando-se os pontos de máximo e de mínimo, respectivamente. Para isto, adota-se a interpolação polinomial de terceiro grau utilizando o método de *splines*;
3) Calcular o resíduo como a média entre as envoltórias: $r(t) = (e_{min}(t) + e_{max}(t))\,/2$;
4) Extrair as componentes de detalhes: $d(t) = x(t) - r(t)$;
5) Repetir a iteração sobre o sinal residual $r(t)$.

Para que o passo (5) possa ser realizado, a componente de detalhe $d(t)$ obtida no passo (4) deve obedecer a duas propriedades: a primeira está relacionada à diferença entre a quantidade de extremos e a quantidade de cruzamento em zero, a qual deve ser nula ou igual a um; a outra propriedade diz respeito ao valor médio definido pelas envoltórias dos seus máximos e mínimos, o qual deve ser nulo. Caso contrário a estas propriedades, os passos (1-4) são novamente efetuados, com $d(t)$ no lugar de $x(t)$. Este processo, denominado *sifting*, é repetido até garantir que a nova função $d(t)$ seja considerada uma IMF [12].

O algoritmo do método EMD assegura que qualquer sinal $x(t)$ pode ser decomposto em um número finito (K) de iterações, e pode ser escrito como

$$x(t) = \sum_{k=1}^{K} \text{IMF}_k(t) + r(t)\,, \tag{2}$$

em que $\text{IMF}_k(t), 1 \leq k \leq K$, são as funções de detalhes $d(t)$ obtidas no passo (4) de cada iteração, e $r(t)$ é o sinal residual final decorrente da última iteração.

Exemplos da decomposição empírica de modos são apresentados na Figura 1, com $K = 3$. São observados trechos de 96 ms de sinais de voz no estado Neutro e com as emoções Raiva e Tristeza. A redução no número de extremos de um modo para o próximo implica que, localmente, as primeiras IMFs possuem oscilações mais rápidas (altas frequências) que as IMFs de maior índice. A primeira IMF referente à Raiva possui amplitudes superiores àquelas dos outros estados emocionais, enquanto que na terceira IMF este papel passa a ser da emoção Tristeza. Esta análise tempo-frequência da concentração de energia em cada um dos modos da EMD é detalhada neste trabalho e, nas Seções seguintes, são apresentados os aspectos referentes aos experimentos realizados e as discussões pertinentes.

III. Análise dos Sinais Acústicos

Esta Seção apresenta os principais parâmetros e medidas utilizados na análise da aplicação da EMD nos sinais de voz: o índice de não-estacionariedade, Energia Segmental e, finalmente, a Distância de Bhattacharyya.

A. Índice de Não-Estacionariedade

O índice de não-estacionariedade (INS - *index of nonstationarity*) é um método tempo-frequência que determina, de forma objetiva, o grau de não-estacionariedade de sinais e ruídos [18]. Esta avaliação é realizada a partir da comparação das componentes espectrais do sinal e de referenciais estacionários (*surrogates*). Dada a transformada discreta de Fourier, $X[k]$, do sinal $x(t)$,

$$X[k] = A[k]\exp(i\,\phi[k])\,, \tag{3}$$

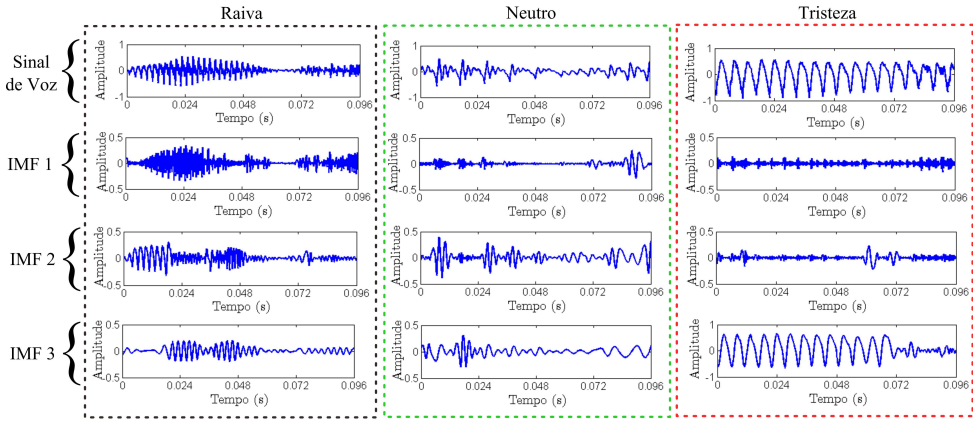


Fig. 1. Decomposição empírica de modos, em 3 IMFs, de trechos de 96 ms de sinais de voz dos estados emocionais raiva, neutro e tristeza (sinais extraídos da base EMO-DB [17]).

em que $A[k]$ e $\phi[k]$ são a magnitude e a fase de $x(t)$, respectivamente. Os referenciais, $\tilde{x}(t)$, são obtidos aplicando-se a transformada inversa de Fourier de $\tilde{X}[k]$:

$$\tilde{X}[k] = A[k]\exp(i\,\psi[k])\,, \tag{4}$$

em que $\psi[k]$ é uma sequência aleatória com amostras independentes e uniformemente distribuídas no intervalo $[-\pi, \pi]$.

O espectrograma obtido com um janelamento multi-ortogonal (*multitaper*) de $x(t)$ é comparado com os referenciais estacionários obtidos, utilizando a distância de Kullback-Leibler (KL) [19]. O INS é então definido como a razão entre os valores de KL obtidos do sinal analisado e dos seus referenciais estacionários. Após a representação de KL por meio de uma distribuição Gama, é definido um limiar (γ) para o teste de não-estacionariedade considerando um valor de precisão de 95%. Desta forma, o sinal é considerado não-estacionário se o valor de INS estiver acima deste limiar. Ou seja,

$$\text{INS} \begin{cases} \leq \gamma & , x(t) \text{ é estacionário;} \\ > \gamma & , x(t) \text{ não é estacionário.} \end{cases} \tag{5}$$

B. Energia Segmental

Neste estudo, a enegia segmental (E - Energia) foi estimada de seis IMFs ao longo de trechos de 96 ms dos sinais de voz. Esta medida calcula a intensidade sonora de um sinal acústico ao longo do tempo [11]:

$$E = \int_{t=t_1}^{t_2} x(t)^2\, dt, \tag{6}$$

em que $t_1 < t < t_2$ pode representar toda a duração do sinal $x(t)$ ou apenas um trecho (segmento). Uma forma mais usual de analisar a energia de um sinal é empregando a escala decibel (dB), que permite a observação da variação das amplitudes de forma mais eficiente [11]:

$$E(\text{dB}) = 10\log_{10}(E). \tag{7}$$

Uma análise quantitativa da distribuição dos valores de energia para cada uma das IMFs obtidas da EMD é realizada por meio da distância de Bhattacharyya, descrita a seguir.

C. Distância de Bhattacharyya

Dadas duas distribuições de probabilidade, $p_1(x)$ e $p_2(x)$, a distância de Bhattacharyya (Bd - *Bhattacharyya distance*) mede a similaridade entre elas [20]. Para o cálculo dessa distância, primeiramente é medido o chamado coeficiente de Bhattacharyya, dado por:

$$\rho = \int \sqrt{p_1(x)p_2(x)dx}, \tag{8}$$

em que $0 < \rho < 1$ quantifica quão sobrepostas estão as duas distribuições. Assim, a distância de Bhattacharyya é calculada como sendo:

$$Bd = -\ln\ \rho, \tag{9}$$

em que $0 < Bd < \infty$.

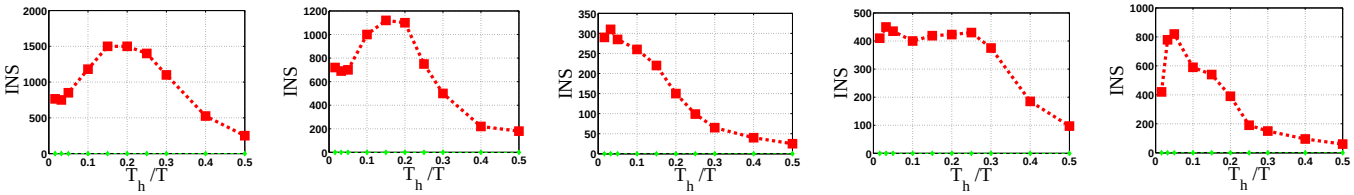
IV. Experimentos e Resultados da Análise

Esta Seção apresenta os principais resultados obtidos nos experimentos. Assim, são discutidos os valores obtidos da análise do INS, dos histogramas da energia segmental das IMFs, média da energia segmental por IMF e a Bd.

A base de dados utilizada é a EMO-DB [17], da qual foram utilizados 401 sinais de voz de 10 locutores (5 homens e 5 mulheres) na língua alemã. As emoções analisadas foram: Raiva (127 sinais), Felicidade (64 sinais), Neutro (78 sinais), Tédio (79 sinais) e Tristeza (53 sinais). A taxa de amostragem utilizada foi de 8 kHz.

A. INS

O grau de não-estacionariedade dos sinais de voz com as variações emocionais analisadas neste trabalho está ilustrado na Figura 2. Os valores de INS são apresentados para diferentes tamanhos de segmento (T_h) em relação ao tamanho total do sinal (T). Os valores máximos de INS para as emoções Raiva e Felicidade são acima de 1000. Este alto grau de não-estacionariedade reflete a natureza de alta ativação destes sinais. Assim, estes possuem maior concentração de energia nas oscilações mais rápidas (primeira IMF). A Tristeza tem



(a) Raiva (b) Felicidade (c) Neutro (d) Tédio (e) Tristeza

Fig. 2. Índice de não-estacionariedade (INS) de sinais de voz com variações emocionais.

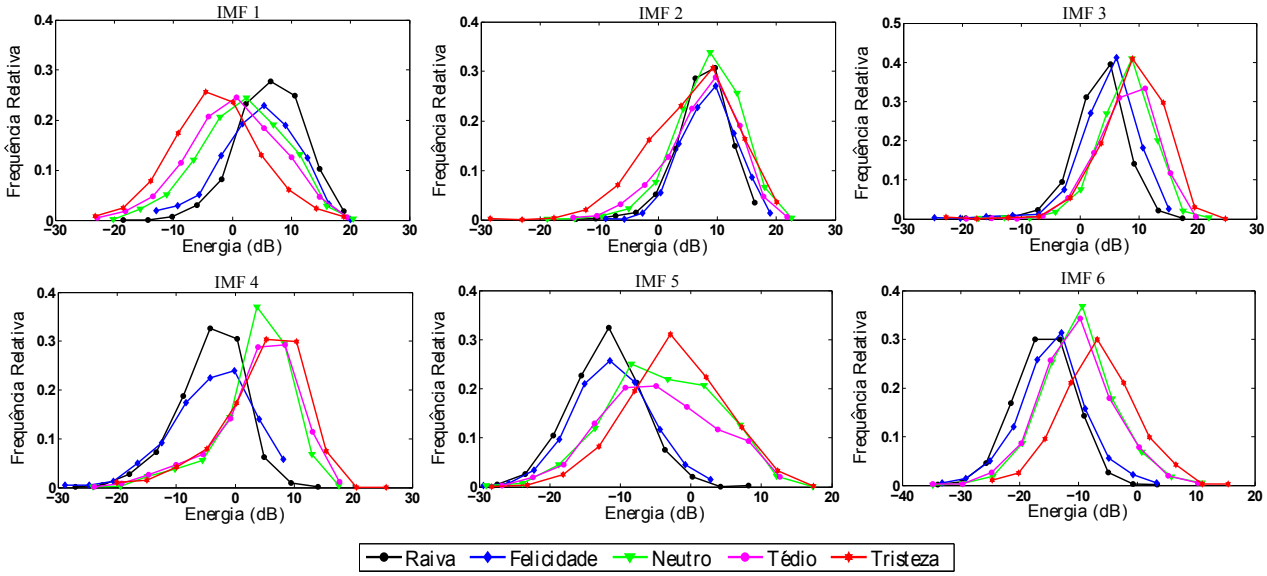


Fig. 3. Distribuição dos valores da energia medida nos sinais de cada emoção, para cada uma das seis IMFs observadas.

seus maiores valores de INS entre 400 e 800. Isto comprova, também, a separação entre as emoções de alta e baixa ativação. O estado Neutro, por sua vez, apresenta menores INS (valor máximo $\approx$ 300). Porém, comportamento semelhante é observado na emoção Tédio.

B. Histogramas da Energia Segmental

Os histogramas da energia segmental foram obtidos dos valores de 6 IMFs calculados em segmentos consecutivos de 96 ms dos sinais de voz. Estes histogramas estão apresentados na Figura 3. Note que na primeira IMF é evidente a separação das emoções por suas distintas concentrações de energia em relação ao estado neutro (energia centrada próximo a 0 dB), que, nesta faixa de frequência, pode ser considerado um limiar que separa as emoções de alta e baixa ativação. A proximidade das emoções, observada na segunda IMF, pode ser explicada por meio da análise dos histogramas seguintes. Na transição para faixas de frequência menores (da segunda IMF em diante), emoções como Raiva e Felicidade tendem a apresentar menor energia se comparadas a Tédio e Tristeza. A partir da terceira IMF essas diferenças começam a ser realçadas e ficam mais claras na quarta e quinta IMF, embora o estado Neutro apresente também valores elevados de energia. É possível observar que na sexta IMF as múltiplas emoções voltam a se confundir, o que pode ser uma garantia que estas seis faixas de frequência são suficientes para analisar a separação das emoções.

C. Média da Energia Segmental por IMF

Na Figura 4, os valores médios de energia mostram que, nas frequências mais elevadas (primeira IMF), a emoção Raiva (alta ativação) apresenta a maior concentração de energia quando comparada à demais emoções. Por outro lado, a Tristeza (baixa ativação) tem este comportamento nas frequências mais baixas (sexta IMF). Além disso, é possível notar que a partir da terceira IMF fica mais evidente a separabilidade entre as emoções de alta e baixa ativação.

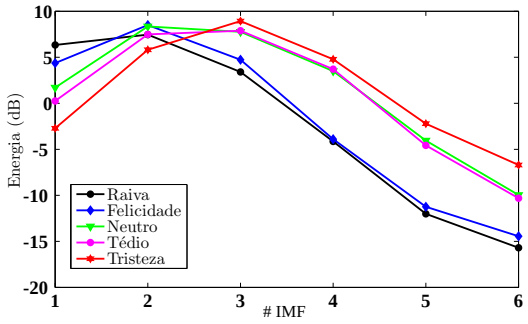


Fig. 4. Média dos valores de Energia (dB), por emoção, para cada uma das seis IMFs observadas.

D. Distância de Bhattacharyya entre os Histogramas

A distribuição dos valores de energia para cada IMF e suas respectivas médias permitem, até então, uma análise subjetiva a respeito da diferença entre as concentrações de energia dos estados emocionais nas diferentes faixas de frequência dos sinais de voz. A distância de Bhattacharyya foi calculada para quantificar essa separação, tornando a análise mais objetiva.

Na Tabela I, diferentes cenários são apresentados com seus respectivos valores da distância de Bhattacharyya. Os números em destaque nas IMFs 1 e 5 implicam em duas observações: a primeira diz respeito à maior distância entre a emoção Tristeza e duas outras de alta ativação (Raiva e Felicidade), como pode ser notado na Figura 3; a outra observação está relacionada ao fato de que, nestes casos de discriminação, os valores máximos de distância são verificados na quinta IMF, na qual a tristeza, por sua vez, apresenta os maiores valores de energia. Os menores valores de distância na sexta IMF indicam que as diferenças entre os estados emocionais são realçadas até a quinta IMF, como é observado nos histogramas. É possível verificar, ainda, que os valores centrais de energia obtidos para o estado Neutro, podem ser referenciais de distância entre as emoções Raiva e Tristeza, com valores de Bd de 0,077 e 0,052, respectivamente. Este referencial fica ainda mais evidente na IMF 5, quando obtem-se os valores 0,205 e 0,015 de Bd para este mesmo cenário. Além disso, conforme esperado, confirma-se que o sinal de voz Neutro se confunde naturalmente com a emoção Tédio, o que é comprovado com os valores mínimos (≈ 0) de distância em todas as IMFs.

TABELA I

DISTÂNCIA DE BHATTACHARYYA CALCULADA ENTRE HISTOGRAMAS DE DIFERENTES EMOÇÕES.

Cenário	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6
Raiva ↔ Tristeza	**0,252**	0,053	0,192	0,305	**0,317**	0,307
Raiva ↔ Felicidade	0,020	0,014	0,019	0,017	0,009	0,012
Felicidade ↔ Tristeza	**0,145**	0,064	0,109	0,258	**0,268**	0,209
Neutro ↔ Raiva	0,077	0,022	0,129	0,265	0,205	0,134
Neutro ↔ Felicidade	0,022	0,009	0,050	0,185	0,170	0,070
Neutro ↔ Tristeza	0,052	0,028	0,020	0,026	0,015	0,034
Neutro ↔ Tédio	0,008	0,006	0,006	0,009	0,005	0,001

V. CONCLUSÃO

Este trabalho apresentou a análise de sinais acústicos com variações emocionais baseada na aplicação do método EMD. No estudo foram analisados sinais de cinco tipos de emoção (Raiva, Felicidade, Tédio, Tristeza e Neutro). A análise considerou a concentração da energia sonora ao longo de seis faixas de frequência (IMFs). Os experimentos mostram que há IMFs em que a separação entre os múltiplos estados emocionais são realçadas. Além disso, é observado que o grau de não-estacionariedade dos sinais de voz difere de um estado emocional para outro. Os resultados indicam que o método EMD é promissor na discriminação de estados emocionais, podendo ser empregado juntamente a outras técnicas de análise acústica para fins de análise e identificação/classificação de emoções.

REFERÊNCIAS

[1] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, and W. Fellenz, "Emotion recognition in human-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, pp. 32–80, January 2001.

[3] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll, and A. Wendemuth, "Acoustic emotion recognition: A benchmark comparison of performances," in *Automatic Speech Recognition & Understanding, 2009. ASRU 2009. IEEE Workshop on*, pp. 552–557, IEEE, 2009.

[4] L. Zao, D. Cavalcante, and R. Coelho, "Time-frequency feature and ams-gmm mask for acoustic emotion classification," *Signal Processing Letters, IEEE*, vol. 21, pp. 620–624, 2014.

[5] G. Zhou, J. H. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress," *Speech and Audio Processing, IEEE Transactions on*, vol. 9, no. 3, pp. 201–216, 2001.

[6] J. M. Pickett, "The sounds of speech communication," *Baltimore, MD: University Park*, 1980.

[7] T. F. Quatieri, *Discrete-time speech signal processing: principles and practice*. Pearson Education, 2002.

[8] H. M. Teager, "Some observations on oral air flow during phonation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 5, pp. 599–601, 1980.

[9] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "Energy separation in signal modulations with application to speech analysis," *Signal Processing, IEEE Transactions on*, vol. 41, no. 10, pp. 3024–3051, 1993.

[10] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*, vol. 100. Prentice-hall Englewood Cliffs, 1978.

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

[12] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, March 1998.

[13] P. Flandrin, G. Rilling, and P. Gonçalvès, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, pp. 112–114, February 2004.

[14] P. Flandrin, P. Gonçalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," *Proceedings of the European Signal Processing Conference (EUSIPCO 2004)*, pp. 1581–1584, September 2004.

[15] R. Coelho and L. Zão, "Empirical mode decomposition theory applied to speech enhancement," in *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering and Processing* (R. Coelho, V. Nascimento, R. Queiroz, J. Romano, and C. Cavalcante, eds.), pp. 123–153, Boca Raton, FL: CRC Press, 2015.

[16] K. Khaldi, A.-O. Boudraa, A. Bouchikhi, and M. T.-H. Alouane, "Speech enhancement via emd," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 1–8, 2008.

[17] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of german emotional speech," in *Proceedings of the Interspeech 2005, Lissabon, Portugal*, pp. 1517–1520, 2005.

[18] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Transactions on Signal Processing*, vol. 58, pp. 3459–3470, July 2010.

[19] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 03 1951.

[20] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *Communication Technology, IEEE Transactions on*, vol. 15, no. 1, pp. 52–60, 1967.