

# Análise de Variações Acústicas Não Estacionárias em Sinais de Voz Gerados em Condições de Estresse

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

**Resumo**— Este trabalho apresenta uma análise da não estacionariedade de sinais com variações acústicas produzidas pelos indivíduos quando submetidos a diferentes condições de estresse. O estudo investiga o índice de não estacionariedade (*index of non-stationarity* – INS) das condições de estresse Médio, Alto e Grito. A análise inclui os espectrogramas e cocleogramas dos sinais. A decomposição empírica de modos (*empirical mode decomposition* – EMD) é utilizada para estimar o INS e a energia nas subbandas. A distância de Bhattacharyya ( $Bd$ ) é empregada como medida de separabilidade destes sinais. Os resultados demonstraram que o INS é promissor na discriminação das variações acústicas.

**Palavras-Chave**— Variações acústicas, Índice de não estacionariedade, Decomposição Empírica de Modos, Condições de estresse.

**Abstract**— This work presents an analysis of the non-stationarity of signals with acoustic variations produced by individuals when submitted to different stress conditions. The study investigates the index of non-stationarity (INS) of the Medium, High and Screaming stress conditions. The analysis includes the spectograms and cochleagrams of the signals. The empirical mode decomposition (EMD) is used to estimate the INS and the energy in the subbands. The Bhattacharyya distance ( $Bd$ ) is employed as a separability measure of these signals. The results showed that INS is promising in discriminating the acoustic variations.

**Keywords**— Acoustic variations, Index of non-stationarity, Empirical Mode Decomposition, Stress conditions.

## I. INTRODUÇÃO

Reconhecimento de condições de estresse e estados emocionais é um interessante campo de estudo que vem recebendo atenção nos últimos anos [1], [2], [3], [4]. Muitas aplicações tem buscado ferramentas de análise acústica para reconhecer estados afetivos, tais como sistemas de segurança, *call-centers*, sistemas de comunicação, tradução automática e interação homem-robô [5], [6]. Condições de estresse estão presentes no cotidiano dos seres humanos, influenciando a cognição, a percepção, o aprendizado e a comunicação. Um evento inesperado, por exemplo, pode ser motivado por algo que gere a sensação de felicidade. Por outro lado, uma surpresa negativa pode ocorrer, levando o indivíduo à sensações como medo e estresse, o que afetaria, entre outras coisas, tomadas de decisão de curto prazo.

V. Vieira, doutorando no Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Campina Grande (PPGEE/UFPG) e bolsista CNPq; R. Coelho, Laboratório de Processamento de Sinais Acústicos, Instituto Militar de Engenharia (IME); F. M. de Assis, Departamento de Engenharia Elétrica, UFPG, Brasil. E-mails: {vinicius.vieira@ee.ufcg.edu.br, coelho@ime.br, fmarcos@dee.ufcg.edu.br}. Este trabalho foi parcialmente financiado pelo CNPq 140816/2014-3 e 307866/2015-7.

Para que seja desenvolvido um sistema homem-máquina robusto para detecção de condições de estresse e até mesmo estados emocionais, se faz necessário buscar as características mais significativas presentes nos sinais acústicos [2]. Alguns estudos utilizam correlatos com a percepção do ouvido humano [3], [7], outros com a não linearidade da produção da voz [8] e com a densidade espectral de potência da voz sob estados afetivos [4].

Um fator que pode ser levado em consideração é a não estacionariedade dos sinais de voz em trechos superiores a 40 ms [9]. Apesar disso, diferentes variações acústicas, quando introduzidas na voz, podem acarretar em diferentes graus de não estacionariedade. Em [10], foi mostrado que diferentes fontes acústicas possuem diferentes graus de não estacionariedade. Em [11], foi empregado o índice de não estacionariedade (*index of non-stationarity* – INS) para analisar cinco diferentes estados emocionais. Os autores mostraram que o INS varia de uma emoção para outra, e que esta medida está relacionada com a densidade espectral de potência das variações acústicas emocionais. Além disso, foi utilizada em [11] a decomposição empírica de modos (*empirical mode decomposition* – EMD) para uma análise adaptativa das variações acústicas não estacionárias.

Neste trabalho, o INS é estimado em sinais com quatro diferentes condições de estresse: Neutro, Médio, Alto e Grito. A análise destes sinais inclui ainda os seus respectivos espectrogramas e cocleogramas. Além disso, a EMD é empregada para a decomposição dos sinais acústicos em seis funções intrínsecas de modo (*intrinsic mode functions* – IMFs). Então, o INS é estimado em cada uma das componentes de frequência. A energia segmental também é estimada em cada IMF. A separação entre as condições de estresse é quantificada por meio da distância de Bhattacharyya ( $Bd$ ). A análise mostra que as condições de estresse possuem características espectrais particulares, o que é observado nos espectrogramas e cocleogramas. Estas diferenças espectrais são quantificadas pelo INS. Os resultados dos experimentos indicam o potencial de discriminação das condições de estresse por meio da análise tempo-frequência com o INS e a EMD. Por meio da  $Bd$ , é possível observar que o INS é mais eficiente em separar as variações acústicas nas suas IMFs do que a concentração de energia delas.

O restante deste trabalho está organizado da seguinte forma: Na Seção II são apresentadas as principais técnicas utilizadas neste trabalho para a avaliação das variações acústicas: INS e EMD. Na Seção III são apresentados os materiais e métodos

utilizados neste estudo. Na Seção IV são apresentados os resultados obtidos nos experimentos realizados, e na Seção V são apresentadas as considerações finais.

## II. TÉCNICAS DE ANÁLISE DE SINAIS NÃO ESTACIONÁRIOS

Nesta Seção, são apresentadas duas técnicas para a análise das variações acústicas produzidas por diferentes condições de estresse. A primeira delas é o índice de não estacionariedade (*index of non-stationarity* – INS). Esta medida é capaz de analisar o grau de não estacionariedade do sinal ao longo do tempo. A segunda técnica é a decomposição empírica de modos (*empirical mode decomposition* – EMD). A EMD faz uma análise tempo-frequência de sinais não estacionários baseada nas próprias características do sinal.

### A. INS

O INS é uma medida tempo-frequência que analisa objetivamente a não estacionariedade de um sinal [12]. Seus autores definem um sinal como sendo estacionário em relação a uma escala de observação se o seu espectro local de tempo curto em diferentes instantes de tempo for estatisticamente similar ao seu espectro global. O teste de estacionariedade é realizado pela comparação de componentes espectrais do sinal com referenciais estacionários, chamados *surrogates*, obtidos do próprio sinal. Para tanto, os espectrogramas do sinal e dos *surrogates* são obtidos por meio da Transformada de Fourier de tempo curto (*short time Fourier transform* – STFT). Então, a distância Kullback-Leibler (KL) [13] é usada para medir a divergência entre o espectro de curto tempo do sinal analisado e seu espectro global. Finalmente, o INS é dado pela razão entre esta distância e a KL correspondente aos valores obtidos dos referenciais estacionários. Em [12], os autores consideram que a distribuição dos valores da KL são aproximados por uma distribuição Gamma. Por isso, para cada janela de tempo  $Th$ , um limiar  $\gamma$ , com 95% de precisão, pode ser definido para o teste de estacionariedade. Desta forma, o sinal é considerado não estacionário se o valor de INS estiver acima deste limiar. Ou seja,

$$\text{INS} \begin{cases} \leq \gamma & , x(t) \text{ é estacionário;} \\ > \gamma & , x(t) \text{ não é estacionário.} \end{cases} \quad (1)$$

### B. EMD

O método EMD foi proposto em [14] como uma técnica não linear de decomposição adaptativa tempo-frequência de sinais não estacionários em diferentes modos de oscilação. A ideia geral é analisar localmente um sinal  $x(t)$  entre dois extremos consecutivos (mínimo ou máximo). Então, duas partes são definidas: oscilações locais rápidas, também conhecidas como componentes de detalhes,  $d(t)$ , e o resíduo local  $a(t)$ , tal que  $x(t) = d(t) + a(t)$ . Detalhes locais,  $d(t)$ , obtidos de  $x(t)$ , correspondem ao primeiro modo (IMF). Esta primeira IMF extraída consiste da componente de frequência mais alta do sinal, e as componentes seguintes são iterativamente obtidas a partir do resíduo  $a(t)$  da IMF anterior. O procedimento para

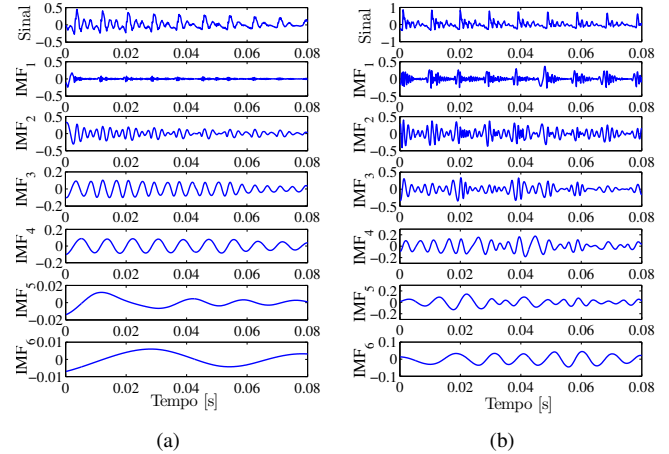


Fig. 1. Primeiras 6 IMFs obtidas de sinais de: (a) estado Neutro e (b) Alto estresse.

extrair cada IMF é conhecido como *sifting*, o qual pode ser resumido nos seguintes passos:

- 1) Identificar todos os extremos locais (mínimos e máximos) de  $x(t)$ ;
- 2) Conectar todos os extremos locais por meio de interpolação<sup>1</sup> a fim de se obter as envoltórias superior ( $e_{max}(t)$ ) e inferior ( $e_{min}(t)$ );
- 3) Calcular o resíduo como a média entre as envoltórias:  $a(t) = (e_{min}(t) + e_{max}(t)) / 2$ ;
- 4) Extrair as componentes de detalhes:  $d(t) = x(t) - a(t)$ .

Para que uma sequência decomposta seja considerada uma IMF, duas condições devem ser satisfeitas: a diferença entre as quantidades de máximos e de cruzamentos por zero em toda a componente deve ser nula ou igual a um; e a sequência deve ter média zero. Se a componente de detalhes não segue estas propriedades, as etapas (1-4) são repetidas com  $d(t)$  no lugar de  $x(t)$  até que um novo  $d(t)$  possa ser considerado uma IMF. Para a próxima IMF, o procedimento de *sifting* é aplicado no resíduo  $a(t) = x(t) - d(t)$ .

O método EMD possui uma propriedade de integrabilidade, a qual garante que qualquer sinal  $x(t)$  que seja decomposto em uma quantidade finita de IMFs, possa ser expresso por

$$x(t) = \sum_{m=1}^M \text{IMF}_m(t) + r(t), \quad (2)$$

em que  $m$  é o índice de cada componente tempo-frequência, e  $r(t)$  é o resíduo final.

Exemplos da aplicação do método EMD são apresentados na Figura 1. Os sinais decompostos têm duração de 80 ms e constituem da voz em estado Neutro e de uma condição de Alto estresse. Nas primeiras IMFs as oscilações são mais rápidas que as seguintes. Isto significa que, em cada intervalo de tempo, a EMD realiza uma separação entre IMFs de alta e de baixa frequência. Uma maior energia pode ser notada na primeira IMF do sinal de Alto estresse, em relação ao estado Neutro. Ainda, a variação acústica introduzida pelo Alto estresse provoca oscilações mais irregulares que no

<sup>1</sup>Geralmente, adota-se a interpolação polinomial de terceiro grau utilizando o método de *splines*.

caso da voz em estado Neutro, o que pode ser observado principalmente na terceira e na quarta IMF.

### III. EXPERIMENTOS REALIZADOS

Os materiais e métodos empregados neste trabalho são apresentados nesta Seção. A análise das variações acústicas causadas por diferentes condições de estresse é realizada com uma base de sinais gravados na língua inglesa. Além do INS e da EMD, os experimentos são realizados com espectogramas, cocleagramas, energia segmental e distância de Bhattacharyya.

#### A. Base de Dados

A SUSAS (*Speech Under Simulated and Actual Stress*) [15] é uma base de dados composta por 3593 sinais captados a uma taxa de 8.000 amostras/s, em condições reais de estresse e medo. Os sinais foram obtidos com 7 locutores (4 homens e 3 mulheres), submetidos a duas situações distintas: montanha-russa e queda livre. Um total de 35 comandos de curta duração foram gravados na língua inglesa, tais como *break*, *hello* e *help*. A base SUSAS aborda as situações de Alto estresse, Médio estresse e Grito, além do estado Neutro. Antes da etapa de análise dos sinais acústicos, eles foram submetidos a uma etapa de seleção dos trechos sonoros. Para a reconstrução dos sinais, foram mantidos os quadros de 16 ms com alta energia e baixa taxa de cruzamentos por zero. A fim de balancear os experimentos, foi mantida a mesma duração (considerando todos os trechos sonoros) para todas as condições de estresse: 146 s.

#### B. Métodos Empregados

1) *Espectogramas e Cocleagramas*: A análise das variações acústicas não estacionárias é primeiramente realizada traçando um paralelo entre o INS e dois outros métodos de análise tempo-frequência: espectogramas e cocleagramas. O espectrograma [9] é uma das formas mais clássicas de análise tempo-frequência, e é gerado a partir da STFT calculada sobre o sinal. O cocleograma, diferente do espectrograma, não possui suas componentes uniformemente espaçadas. Cocleagramas utilizam o banco de filtros Gammatone, o qual modela a filtragem ocorrida na coclea humana, que percebe frequências mais elevadas de forma mais espaçada que frequências mais baixas [16], [17].

2) *Energia Segmental*: Esta medida calcula a intensidade sonora de um sinal acústico ao longo do tempo [9]:

$$E = \int_{t=t_1}^{t_2} x(t)^2 dt, \quad (3)$$

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),

$$E(\text{dB}) = 10\log_{10}(E). \quad (4)$$

Neste estudo, a energia segmental ( $E$  - Energia) foi estimada de seis IMFs ao longo de trechos de 80 ms dos sinais de voz, sobrepostos 50%.

3) *Distância de Bhattacharyya*: A Distância de Bhattacharyya (*Bhattacharyya distance* -  $Bd$ ) foi proposta em [18], inicialmente para aplicações de seleção de sinais. Esta é uma medida da dissimilaridade entre duas distribuições de probabilidade,  $p_1(x)$  and  $p_2(x)$ . Assim,  $Bd$  pode ser definida como

$$Bd = -\ln \left( \int \sqrt{p_1(x)p_2(x)} dx \right), \quad (5)$$

em que  $0 < Bd < \infty$ .

A medida  $Bd$  é empregada neste trabalho para medir a dissimilaridade entre os histogramas obtidos dos valores de média, desvio padrão e máximo do INS estimado em seis IMFs. Também é analisada a  $Bd$  entre os histogramas da energia segmental calculada após a EMD.

### IV. RESULTADOS DA ANÁLISE E DISCUSSÃO

Nesta Seção, são apresentados os resultados das análises com INS em relação a espectogramas e cocleagramas. Também, são apresentados os resultados obtidos da  $Bd$  utilizando INS e energia estimados de IMFs.

#### A. Análise do INS versus Espectogramas e Cocleagramas

Os valores do INS, o espectrograma e o cocleograma de um trecho de 1 s de um sinal de voz no estado Neutro são apresentados na Figura 2. A escala  $Th/T$  refere-se a razão entre as janelas de observação ( $Th$ ) do INS e o tamanho todo ( $T$ ) do sinal. Os pontos em verde representam o limiar de estacionariedade, enquanto que os pontos em vermelho representam os valores de INS encontrados em cada escala de tempo. O INS é aproximadamente constante em quase todo o intervalo de observação (Figura 2a). Isto reflete o comportamento apresentado em seu espectrograma (Figura 2b), no qual pode ser observado um padrão ao longo do tempo. Ou seja, a não estacionariedade é quase a mesma em todas as escalas de observação. O cocleograma (Figura 2c), por sua vez, apresenta componentes em frequências mais elevadas. Isto porque as subbandas não são uniformemente espaçadas no domínio tempo-frequência pelo banco de filtros Gammatone [17]. Embora apresente diferenças na distribuições das frequências do sinal, o cocleograma tem comportamento similar ao espectrograma. Note que, ao longo do tempo, a energia tem aproximadamente o mesmo padrão de distribuição.

Nas Figuras 3, 4 e 5 são apresentados os valores de INS, espectogramas e cocleagramas para Médio estresse, Alto estresse e Grito, respectivamente. As condições de estresse apresentam diferentes comportamentos em relação à sua não estacionariedade. Apesar de os sinais acústicos afetados por estresse apresentarem uma maior quantidade de componentes em frequências mais elevadas que o estado Neutro, cada condição possui sua particularidade. Médio estresse (Figura 3a) atinge um valor máximo de INS mais elevado que Alto estresse (Figura 4a). Para estas duas condições de estresse, diferenças podem ser observadas em relação aos seus espectogramas. Porém, fica ainda mais claro quando são observados os seus cocleagramas. Médio estresse apresenta uma maior descontinuidade da distribuição de energia

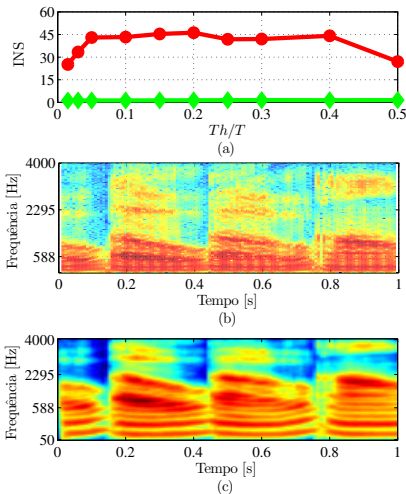


Fig. 2. Sinal do estado Neutro: (a) INS; (b) espectrograma; (c) cocleograma.

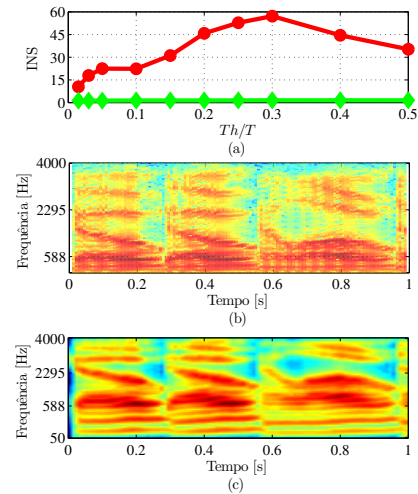


Fig. 4. Sinal de Alto estresse: (a) INS; (b) espectrograma; (c) cocleograma.

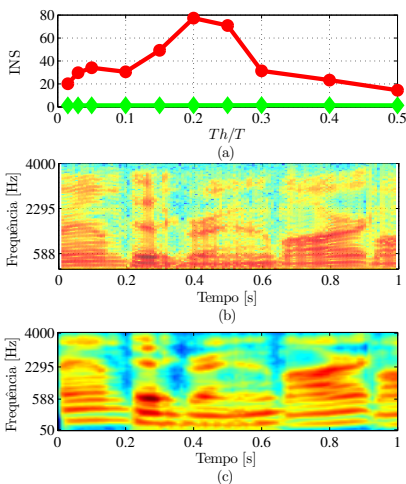


Fig. 3. Sinal de Médio estresse: (a) INS; (b) espectrograma; (c) cocleograma.

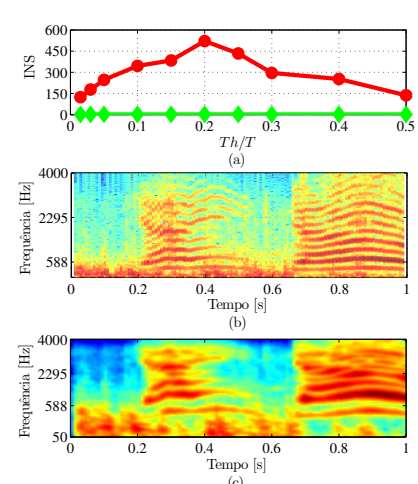


Fig. 5. Sinal de Grito: (a) INS; (b) espectrograma; (c) cocleograma.

no domínio tempo-frequência (Figura 3c) em relação ao Alto (Figura 4c). O Grito (Figura 5), por sua vez, atinge valores de INS bem mais elevados que as demais condições observadas. Na observação espectral (Figuras 5b e 5c), é possível notar que há trechos em que há mais concentração de energia nas mais altas frequências.

### B. *Bd* de Condições de Estresse

Na Tabela I estão apresentados os valores da *Bd* calculada entre os histogramas obtidos das estatísticas analisadas do INS estimado em seis IMFs, obtidas em trechos de 800 ms. Na maioria dos cenários analisados os maiores valores de *Bd* estão concentrados nas primeiras duas IMFs e nas duas últimas. No cenário Neutro (Neu.) versus todos os tipos de estresse juntos (Est.), o maior valor de *Bd* é obtido na primeira IMF para todas as estatísticas analisadas, exceto para o desvio padrão do INS, que tem maior *Bd* na segunda IMF. Os valores em destaque para os demais cenários discriminativos indicam que as condições de estresse podem ser melhor separadas quando analisadas individualmente, uma vez que os valores de

*Bd* são mais altos. Alto estresse (Alt.) é a condição de estresse que tem a maior *Bd* em relação ao estado Neutro ( $Bd = 0,147, IMF_1$ ), considerando a média do INS. Porém, Neutro versus Grito (Gri.) apresenta  $Bd > 0,1$  para mais de uma IMF em todas as estatísticas analisadas. O Grito, por sua vez, é a mais separável das condições de estresse. Por exemplo, no cenário Médio estresse (Med.) versus Grito, a *Bd* atinge um valor de 0,221 na sexta IMF, considerando a média do INS. O valor máximo do INS foi importante principalmente na discriminação entre Alto estresse e Grito, em que foi atingida uma *Bd* de 0,353 na primeira IMF. Por outro lado, a média do INS foi a estatística que apresentou a maior *Bd* na maioria dos cenários de discriminação.

No contexto da energia segmental, os valores da *Bd* calculada entre os histogramas desta medida são apresentados na Tabela II. Na maioria dos cenários, as últimas IMFs apresentaram maior *Bd* para esta medida. Assim como observado para o caso das estatísticas do INS (Tabela I), uma maior separabilidade entre as condições de estresse ocorre quando elas são analisadas individualmente. A discriminação entre

TABELA I

DISTÂNCIA DE BHATTACHARYYA CALCULADA ENTRE HISTOGRAMAS DE CONDIÇÕES DE ESTRESSE USANDO INS

Cenário	Média do INS por IMF					
	IMF <sub>1</sub>	IMF <sub>2</sub>	IMF <sub>3</sub>	IMF <sub>4</sub>	IMF <sub>5</sub>	IMF <sub>6</sub>
Neu.↔Est.	<b>0,056</b>	0,031	0,011	0,028	0,028	0,036
Neu.↔Gri.	0,134	0,061	0,122	0,084	0,029	<b>0,142</b>
Neu.↔Alt.	<b>0,147</b>	0,085	0,034	0,038	0,039	0,078
Neu.↔Med.	<b>0,092</b>	0,059	0,033	0,039	0,039	0,059
Med.↔Alt.	0,081	<b>0,126</b>	0,045	0,027	0,045	0,028
Med.↔Gri.	0,096	0,097	0,166	0,058	0,066	<b>0,221</b>
Gri.↔Alt.	<b>0,287</b>	0,163	0,167	0,071	0,042	0,256

Cenário	Desvio Padrão do INS por IMF					
	IMF <sub>1</sub>	IMF <sub>2</sub>	IMF <sub>3</sub>	IMF <sub>4</sub>	IMF <sub>5</sub>	IMF <sub>6</sub>
Neu.↔Est.	0,031	<b>0,055</b>	0,013	0,009	0,034	0,039
Neu.↔Gri.	0,112	0,061	0,088	0,061	0,053	<b>0,133</b>
Neu.↔Alt.	<b>0,101</b>	0,087	0,049	0,035	0,034	0,051
Neu.↔Med.	0,063	0,069	0,031	0,054	<b>0,081</b>	0,056
Med.↔Alt.	<b>0,078</b>	0,071	0,062	0,021	0,032	0,057
Med.↔Gri.	0,103	0,067	0,099	0,096	0,076	<b>0,198</b>
Gri.↔Alt.	<b>0,266</b>	0,158	0,151	0,093	0,059	0,136

Cenário	Valor Máximo do INS por IMF					
	IMF <sub>1</sub>	IMF <sub>2</sub>	IMF <sub>3</sub>	IMF <sub>4</sub>	IMF <sub>5</sub>	IMF <sub>6</sub>
Neu.↔Est.	<b>0,061</b>	0,044	0,019	0,028	0,034	0,021
Neu.↔Gri.	0,098	0,084	0,113	0,064	0,038	<b>0,131</b>
Neu.↔Alt.	<b>0,134</b>	0,081	0,043	0,039	0,037	0,048
Neu.↔Med.	<b>0,077</b>	0,056	0,038	0,026	0,052	0,028
Med.↔Alt.	0,071	<b>0,123</b>	0,077	0,017	0,037	0,038
Med.↔Gri.	0,101	0,085	0,133	0,041	0,058	<b>0,191</b>
Gri.↔Alt.	<b>0,353</b>	<b>0,214</b>	0,191	0,072	0,039	0,186

Grito e as demais condições de estresse é mais acentuada do que nos demais cenários ( $Bd > 0, 1$ ), principalmente nas duas últimas IMFs. Alto estresse tem maior  $Bd$  em relação a Neutro e Médio estresse na primeira IMF. Os menores valores de  $Bd$  são observados para o cenário de discriminação entre Neutro e Médio estresse.

TABELA II

DISTÂNCIA DE BHATTACHARYYA CALCULADA ENTRE HISTOGRAMAS DE CONDIÇÕES DE ESTRESSE USANDO ENERGIA

Cenário	IMF <sub>1</sub>	IMF <sub>2</sub>	IMF <sub>3</sub>	IMF <sub>4</sub>	IMF <sub>5</sub>	IMF <sub>6</sub>
Neu.↔Est.	0,021	0,013	0,006	0,023	<b>0,041</b>	0,037
Neu.↔Gri.	0,035	0,021	0,036	0,051	<b>0,187</b>	<b>0,224</b>
Neu.↔Alt.	<b>0,065</b>	0,026	0,014	0,038	0,015	0,007
Neu.↔Med.	0,011	0,011	0,014	<b>0,018</b>	0,008	0,001
Med.↔Alt.	<b>0,041</b>	0,017	0,023	0,031	0,011	0,004
Med.↔Gri.	0,042	0,009	0,039	0,017	<b>0,177</b>	<b>0,229</b>
Gri.↔Alt.	0,062	0,037	0,071	0,071	<b>0,138</b>	<b>0,210</b>

A partir dos resultados obtidos com a  $Bd$ , é possível notar que as estatísticas do INS foram mais discriminativas que a energia na maioria dos cenários. A energia apresentou maiores valores de  $Bd$  para os casos de discriminação do Grito com Neutro e com Médio estresse. Por outro lado, variações como Neutro versus Médio estresse e Médio estresse versus Alto estresse obtiveram valores de  $Bd$  com INS bem acima daqueles obtidos com a medida de energia nestes cenários.

V. CONCLUSÃO

Este trabalho apresentou a análise de sinais acústicos com diferentes condições de estresse baseada na sua não estacionariedade. O INS e a energia segmental foram estimados em seis faixas de frequência (IMFs) obtidas pelo

método EMD. As variações acústicas analisadas apresentaram diferentes níveis de não estacionariedade. Os experimentos mostraram que o INS é mais discriminativo que a energia. Além disso, foi observado com a  $Bd$  que as diferentes condições de estresse são melhor separadas do estado Neutro quando consideradas individualmente. Os resultados indicam que a não estacionariedade das variações acústicas estudadas é um importante aspecto a ser considerado no reconhecimento de estados afetivos. Os métodos INS e EMD são promissores e podem ser empregados juntamente a outras técnicas para fins de análise e classificação de condições de estresse.

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, W. Fellenz, and J. G. Taylor, "Emotion recognition in human-computer interaction," *Signal Processing Magazine, IEEE*, vol. 18, no. 1, pp. 32–80, 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. Zão, D. Cavalcante, and R. Coelho, "Time-frequency feature and AMS-GMM mask for acoustic emotion classification," *IEEE Signal Processing Letters*, vol. 21, pp. 620–624, May 2014.
- [5] J. Deng, X. Xu, Z. Zhang, S. Frühholz, and B. Schuller, "Exploitation of phase-based features for whispered speech emotion recognition," *IEEE Access*, vol. 4, pp. 4299–4309, 2016.
- [6] M. Tahon and L. Devillers, "Towards a small set of robust acoustic features for emotion recognition: Challenges," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 1, pp. 16–28, 2016.
- [7] M. J. Alam, Y. Attabi, P. Dumouchel, P. Kenny, and D. D. O'Shaughnessy, "Amplitude modulation features for emotion recognition from speech," in *Proc. INTERSPEECH, 2013*, pp. 2420–2424, 2013.
- [8] G. Zhou, J. H. Hansen, and J. F. Kaiser, "Nonlinear feature based classification of speech under stress," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 3, pp. 201–216, 2001.
- [9] D. O'Shaughnessy, *Speech communication: human and machine*. IEEE Press, 2000.
- [10] A. Venturini, L. Zao, and R. Coelho, "On speech features fusion,  $\alpha$ -integration gaussian modeling and multi-style training for noise robust speaker classification," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1951–1964, 2014.
- [11] V. Vieira, R. Coelho, and F. M. de Assis, "Decomposição tempo-frequência adaptativa de sinais acústicos com variações emocionais," in *Anais do XXXIV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrt 2016)*, pp. 65–69, 2016.
- [12] 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.
- [13] M. Basseville, "Distance measures for signal processing and pattern recognition," *Signal processing*, vol. 18, no. 4, pp. 349–369, 1989.
- [14] 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.
- [15] J. H. Hansen, S. E. Bou-Ghazale, R. Sarikaya, and B. Pellom, "Getting started with susas: a speech under simulated and actual stress database," in *Proc. EUROSPEECH, 1997*, no. 4, pp. 1743–46, 1997.
- [16] G. Hu and D. Wang, "Monaural speech segregation based on pitch tracking and amplitude modulation," *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 1135–1150, 2004.
- [17] B. Gao, W. L. Woo, and L. Khor, "Cochleagram-based audio pattern separation using two-dimensional non-negative matrix factorization with automatic sparsity adaptation," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1171–1185, 2014.
- [18] T. Kailath, "The divergence and bhattacharyya distance measures in signal selection," *IEEE Transactions on Communication Technology*, vol. 15, no. 1, pp. 52–60, 1967.