Verificação Acústica de Emoções Baseada no Modelo Estocástico α-GMM-UBM

G. Zucatelli e R. Coelho

***Resumo*— Este trabalho apresenta uma nova solução para a verificação acústica de emoções baseada no modelo α-GMM-UBM. Esta proposta permite a definição dos valores de α dos modelos do referencial de decisão (UBM) e de cada emoção. Para a avaliação do sistema α-GMM-UBM foram considerados 5 emoções com diferentes índices de não-estacionariedade da base EMO-DB. Os sistemas GMM-UBM e SRV (Sparse Representation Verification) foram adotados como referência. Os resultados demonstraram que o sistema α-GMM-UBM apresentou as menores taxas médias de erro para todos os experimentos realizados.**

***Palavras-Chave*— verificação acústica de emoções, GMM-UBM, α-GMM, representação esparsa.**

***Abstract*— This paper presents a study of a new solution for acoustic emotion verification based on the α-GMM-UBM model. This approache allows different values for α used to model the baseline UBM and each emotion. For the evaluation of the α-GMM-UBM system, 5 emotional states with distict values of the index of non-stationarity were used from the EMO-DB base. Two other systems, GMM-UBM and SRV (Sparse Representation Verification) were adopted for comparison. The result shows that the α-GMM-UBM system presented the best mean of emotion verification accuracy for all experiments conducted.**

***Keywords*— acoustic emotion verification, GMM-UBM, α-GMM, sparse representation.**

I. Introdução

Os estados emocionais dos seres humanos podem ser compreendidos como necessária ferramenta evolutiva [1]. Apesar de não haver um consenso quanto a definição de emoção [2], seu reconhecimento possui importante aplicação em diferentes cenários dos dias atuais. Por exemplo, em sistemas de emergência, sistemas embarcados de automóveis para prevenção de acidentes, como ferramenta auxiliar em diagnóstico terapêutico, em *callcenters* como controle de estresse e aplicações forense.

Os diferentes estados emocionais (desgosto, felicidade, medo, raiva, tédio, tristeza) podem ser acompanhados de mudança na taxa de respiração, no batimento cardíaco e na tensão das cordas vocais o que provoca alterações no sinal de voz [3]. Além disso, a fala é considerada o meio mais natural de comunicação entre os seres humanos, o que motiva sua utilização em sistemas de verificação acústica de emoções.

Geralmente, a classificação em eixos (ativação e valência) é utilizada para discriminação dos estados emocionais [4]. A ativação está relacionada a energia necessária para expressar determinada emoção e a valência refere-se a experiência emocional, classificando-a como positiva ou negativa. As variações emocionais provocam mudanças no fluxo glotal que acarretam em alterações na densidade espectral de potência desses sinais [12] [13]. O estado neutro (ausência de emoção) apresenta uma densidade espectral de potência com queda de 12dB/oitava. Esta tendência pode ser atenuada (queda de 9dB/oitava) ou acentuada (queda de 15dB/oitava) na presença de emoções de alta e baixa ativação, respectivamente. Emoções de alta ativação ocasionam maiores mudanças nos sinais de voz e, portanto, verifica-se concentração superior de energia nas altas frequências. Em contrapartida, estados emocionais de baixa ativação provocam pouca variação no sinal de voz e, desta forma, exibem maior concentração de energia nas baixas frequências. Em [14] um novo atributo acústico foi proposto para discriminar as diferentes emoções.

O sistema de verificação de locutores composto pelo atributo MFCC (*Mel-Frequency Cepstral Coeficient*) [8] e o modelo GMM (*Gaussian Mixture Model*) [5] é comprovadamente uma referência de bom desempenho na literatura [6] [7]. Neste contexto, este sistema é avaliado para classificação de diferentes sinais acústicos (áudio, emoção, biosinais). Esta abordagem estocástica utiliza um elemento base para decisão de autenticação denominado UBM (*Universal Background Model*) [23]. Uma outra abordagem baseia-se na representação esparsa das características dos sinais como os *i-vectors* [17] e *eigenvoice* [16]. Usualmente, emprega-se o mapeamento dos parâmetros dos modelos GMM em vetores [15] [17] os quais representam as características dos elementos em análise.

Neste artigo, uma nova solução para verificação acústica de emoções é proposto baseado no modelo α-GMM-UBM. Esta abordagem é fundamentada na utilização do modelo de misturas gaussianas α-integráveis α-GMM (α-*integrated GMM*) [9]. Em [10], os autores demonstraram a eficiência deste classificador para verificação de locutores robusta a diversos ruídos acústicos. Além disso, a utilização do modelo α-GMM é incentivada por sua capacidade de adaptação proporcionada pelo parâmetro α [11]. O classificador α-GMM será aplicado para a representação do UBM e de cada estado emocional. Além disso, o estudo também investiga o índice de não-estacionariedade INS (*Index of Nontationarity*) [18] dos sinais de voz sob as variações acústicas emocionais como forma de discriminação. O sistema de verificação de emoções proposto (α-GMM-UBM), o referencial (GMM-UBM) e o sistema de representação esparsa SRV (*Sparse Representation Verification*) [20] são avaliados para diversas configurações de testes. Os resultados demonstram que o método proposto obteve as menores taxas de erro médio de verificação para todos os sinais com variações acústicas emocionais analisados quando comparado aos métodos competitivos.

O restante deste artigo está organizado da seguinte forma.

G. Zucatelli e R. Coelho*, Laboratório de Processamento de Sinais Acústicos (lasp.ime.eb.br), Instituto Militar de Engenharia (IME), Rio de Janeiro, Brasil, E-mails: zucatelli@ime.eb.br, coelho@ime.eb.br. *Este trabalho foi parcialmente financiado pelo CNPq/307866/2015-7.

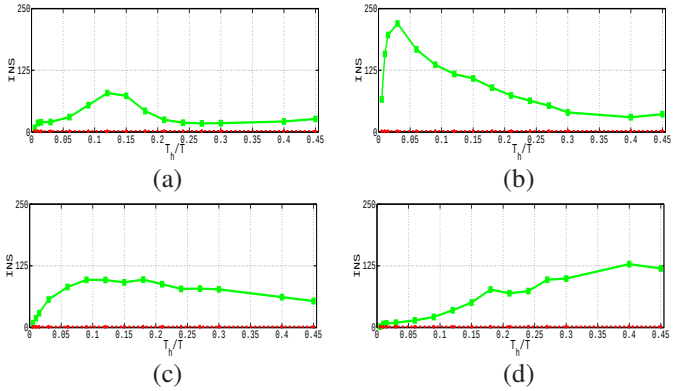

Fig. 1. INS de sinais de voz de 1,0 s utilizados para teste no estado neutro (a) e sob variações emocionais raiva (b), felicidade (c) e tristeza (d).

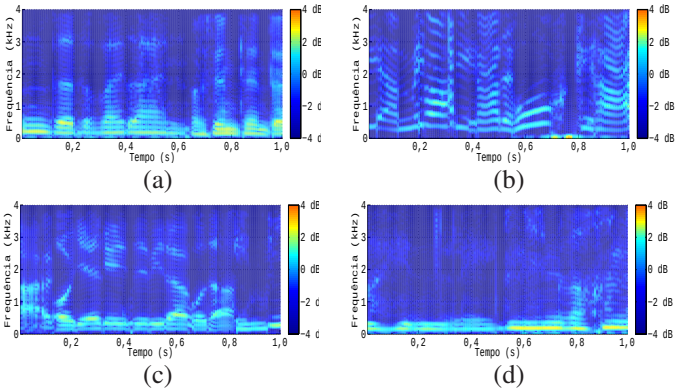

Fig. 2. Espectograma de sinais de voz de 1,0 s utilizados para teste no estado neutro (a) e sob variações emocionais raiva (b), felicidade (c) e tristeza (d).

Na seção II, são descritos os aspectos gerais dos sistemas de verificação acústica de emoções e as características dos sinais sob efeitos acústicos emocionais. Na seção seguinte, o sistema α-GMM-UBM é retratado em conjunto com as modelagens α-GMM e SRV. Posteriormente, apresenta-se os experimentos e resultados na seção IV e a conclusão na seção V.

II. Verificação Acústica de Emoções

A principal função da tarefa de verificação acústica de emoções é averiguar se o sinal em observação refere-se, ou não, ao estado emocional a ser autenticado. De modo geral, um sistema de verificação é composto de duas fases: treinamento e teste. Durante a fase de treinamento, é extraído um conjunto de atributos acústicos para, em seguida, serem construídos os modelos de cada emoção que serão armazenados para emprego na fase de teste. Nesta fase, os atributos extraídos do sinal de teste são comparados com os modelos das emoções para tomada de decisão. Considerando a matriz de atributos $\mathbf{F}$ do sinal de teste ϕ e uma emoção E, a função dos sistemas de verificação consiste em decidir a veracidade de uma das hipóteses de teste

$$\begin{cases} H_0: & \mathbf{F} \text{ pertence a } E, \\ H_1: & \mathbf{F} \text{ não pertence a } E. \end{cases} \tag{1}$$

O critério de decisão para que uma locução ϕ pertença a $\mathbf{F}$ é, geralmente, definido pela razão de verossimilhança

$$\frac{p(\mathbf{F}|\lambda_E)}{p(\mathbf{F}|\lambda_{UBM})} \begin{cases} \geq \theta: & \text{aceita } H_0 \\ < \theta: & \text{rejeita } H_0, \end{cases} \tag{2}$$

onde θ é o limiar de decisão, λ_E representa o modelo da emoção E, λ_{UBM} o modelo do UBM e $p(\mathbf{F}|\lambda)$ é a distribuição de probabilidade condicional de $\mathbf{F}$ dado o modelo λ.

A abordagem por reconstrução esparsa do sinal mapeia os parâmetros das gaussianas obtidas nos modelos GMM em vetores coluna. O critério de decisão destes sistemas é baseado no erro decorrente da recomposição das informações do sinal teste a partir de uma combinação linear destes vetores. O desempenho dos sistemas de verificação acústica de emoções é geralmente baseado nos resultados dos testes de falsa aceitação (FA) e falsa rejeição (FR). O primeiro refere-se a reconhecer como correta uma emoção falsa e o segundo em averiguar como falsa uma emoção correta. A escolha do limiar de decisão θ para a abordagem estocástica é uma relação de compromisso entre os erros de FA e FR. Estas probabilidades são frequentemente avaliadas a partir de curvas DET (*Detection Error Tradeoff*) [22]. A análise desta curva possibilita a definição da taxa de erro EER (*Equal Error Rate*) obtida para o valor do limiar θ em que ambos os testes de FA e FR apresentam a mesma probabilidade. Outra análise do desempenho pode ser estimada com a auxílio da função de decisão de custo DCF (*Decision Cost Function*) definida como uma combinação das probabilidades de ocorrer FA, $P(\text{FA})$, e de ocorrer FR, $P(\text{FR})$.

A. Indice de Não-Estacionariedade

Cada estado emocional se manifesta de forma única nos seres humanos. Consequentemente, os sinais de voz resultantes apresentam diferentes características temporais (distribuição da amplitude), estacionariedade, grau de correlação e natureza espectral. A seguir, apresenta-se a definição do INS e analisa-se as características das variações acústicas emocionais com relação a estacionarieadade e natureza espectral.

Por definição, um sinal é considerado estacionário em uma determinada escala quando seus espectros em tempo-curto (T_h) para qualquer instante são estatisticamente semelhantes ao espectro global. O indice de não-estacionariedade (INS - *Index of Nonstationarity*) [18] é uma medida objetiva de não estacionarieadade baseada na análise tempo-frequência. A divergência de Kullback-Leibler (KL) é utilizada para medir a distância entre os espectros de tempo-curto e o espectro global. A medida INS é dada pela relação entre a distância e o valor correspondente de KL obtidos em relação a referenciais estacionários denominados *surrogates*. Os autores em [18] aproximam a distribuição dos valores de KL como uma distribuição Gamma. Desta maneira, para cada segmento T_h, um limiar γ pode ser definido com um grau de confiança de 95%. Então

$$\text{INS} \begin{cases} \leq \gamma: & \text{sinal estacionário}, \\ > \gamma: & \text{sinal não-estacionário}. \end{cases} \tag{3}$$

A Fig. 1 ilustra os valores de INS obtidos de sinais de voz no estado neutro e sob o efeito de 3 diferentes estados emocionais. O limiar γ está representado em vermelho. Note que as variações acústicas emocionais provocam alterações no valor do INS em relação ao sinal neutro. O maior valor de INS (INS=250) é alcançado com a presença da emoção raiva seguida da emoção tristeza (INS=125). Os espectogramas dos mesmos estados emocionais são apresentados na Fig. 2. Os

sinais que exibem maiores variações espectrais correspondem àqueles com maior valor de INS ao passo que as de pouca variação espectral possuem menor valor de INS. Como os sinais acústicos sob efeito emocional indicam valores de INS superiores a 40, neste trabalho são considerados altamente não-estacionários.

III. Sistema α-GMM-UBM para Emoções

O sistema α-GMM-UBM proposto neste trabalho para verificação acústica de emoções apresenta uma abordagem estocástica. O emprego do modelo α-GMM é incentivada por sua capacidade de adaptação às distribuições dos atributos acústicos de cada estado emocional. Além disso, foi demonstrado em [11] que o uso do classificador α-GMM na identificação de emoções é capaz de aprimorar as taxas de acerto dos diversos estados emocinais segundo diferentes valores de α.

A. Treinamento do α-GMM-UBM

O limiar de referência da decisão bayesiana, aplicado na autenticação dos estados emocionais é definido em função da representação de um UBM. Desta forma, na fase de treinamento do sistema de verificação é criado o UBM. Este modelo de referência deve conter características bem distintas dos sinais acústicos emocionais a serem autenticados. Assim sendo, a construção do UBM para os sistemas GMM-UBM e α-GMM-UBM foi realizada pela união das emoções medo e desgosto de dois locutores excluídos da etapa de teste. Desta maneira, a mistura diferencia-se de todos os demais sinais a serem verificadas. Na fase de treinamento, também são elaborados os modelos de cada estado emocional. Os histogramas gerados pelos modelos (UBM e emoções) são armazenados para emprego na fase de teste na etapa final da decisão do sistema de verificação.

O modelo de misturas gaussianas α-integráveis α-GMM [9] foi inicialmente proposto para sistemas de reconhecimento de locutores onde os modelos são contruídos pelo processo de α-integração de funções distribuição de probabilidade. O modelo GMM convencional (obtido para α=-1) de uma emoção E é definido como uma soma ponderada de distribuições gaussianas,

$$p(\vec{x}|\lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \tag{4}$$

onde $\vec{x}$ é um vetor de atributos acústicos emocionais de dimensão D e $b_i(\vec{x})$ são as funções densidades com vetores média ($\vec{\mu_j}$) e matrizes covariância K_j, as quais podem ser escritas na forma

$$b_i(\vec{x}) = \frac{\exp\left(-\frac{1}{2}(\vec{x_i}-\vec{\mu_i})^T K_i^{-1}(\vec{x_i}-\vec{\mu_i})\right)}{(2\pi)^{\frac{D}{2}}\sqrt{det\, K_i}}. \tag{5}$$

Supondo um número K de funções densidade de probabilidade $p_i(s)$ e pesos w_i a função α-integrável $q(s)$ é definida como

$$q(s) \;=\; c\, f_\alpha^{-1}\left\{\sum_{i=1}^{K} w_i\; f_\alpha[p_i(s)]\right\}, \tag{6}$$

onde c é uma constante tal que $q(s)$ é uma função densidade de probabilidade. Desta forma, o modelo α-GMM, $p_\alpha(\vec{x}|\lambda_\varepsilon)$, é dado por

$$p_\alpha(\vec{x}|\lambda_\varepsilon) = \begin{cases} c\left(\sum_{j=1}^{M} p_j(b_j(\vec{x}))^{\frac{1-\alpha}{2}}\right)^{\frac{2}{1-\alpha}} & \text{para } \alpha \neq 1, \\ c e^{\sum_{j=1}^{M} p_j log(b_j(\vec{x}))} & \text{para } \alpha = 1. \end{cases} \tag{7}$$

Uma importante característica deste modelo é observada ao utilizar $\alpha < -1$, desta maneira o modelo enfatiza as maiores valores de probabilidade e diminiu a relevância dos menores valores.

B. SRV

Em sistemas de verificação acústicos de representação esparsa, as características das matrizes de atributos são geralmente utilizadas para criação de supervetores GMM de tamanho fixo. O supervetor GMM é um vetor coluna obtido pelo mapeamento ordenado dos parâmetros de um modelo GMM. Usualmente utiliza-se as médias normalizadas em função dos pesos e variâncias. A exposição dos supervetores em forma matricial (**M**) compõe um dicionário que integra as características individuais de cada classe em análise. Desta forma, o supervetor GMM **y** de um sinal de teste pode ser obtido a partir de uma combinação linear das colunas de **M**. Ou seja, $\mathbf{y} = \mathbf{Mx}$, onde **x** representa um vetor de coeficientes esparsos. De maneira ideal, este sistema possuiria solução **x** com todos os coeficientes não nulos relativos a mesma variação emocional presente no sinal teste. Geralmente, a solução esparsa **x** é obtida por minimização de sua norma l_1. Para o sistema SRV adotado utiliza-se um conjunto de 25 supervetores para cada classe de forma que a solução do sistema pode ser aproximada, considerando o erro de recontrução ε, por $\hat{\mathrm{x}} = argmin||\mathrm{x}||_1$, dado $||\mathbf{y} - \mathbf{Mx}||_2 < \varepsilon$ [20]. Assim, calcula-se o erro residual referente a cada classe C como $r_C(\mathbf{y}) = ||\mathbf{y} - \mathbf{M}\delta_C(\hat{\mathbf{x}})||_2$, onde $\delta_C(\hat{\mathbf{x}})$ é um vetor com elementos nulos a exceção dos que correspondem a classe C. No sistema SRV adotado neste trabalho o erro de reconstrução é modelado por uma distribuição gaussiana de média nula e variância 2, a qual é utilizada no critério de decisão dos testes de falsa aceitação e falsa rejeição.

IV. Experimentos e Resultados

Nesta Seção, são expostos os principais resultados obtidos de diferentes experimentos da avaliação comparativa dos sistemas proposto (α-GMM-UBM), referencial (GMM-UBM) e SRV. A base de língua alemã EMO-DB [21], amplamente utilizada na literatura, foi adotada para os testes experimentais neste trabalho. Esta base é composta de cerca de 500 locuções de 10 atores (5 homens e 5 mulheres) selecionadas subjetivamente que simulam as emoções desgosto, felicidade, medo, raiva, tédio e tristeza, além do estado neutro. Todos os sinais foram subamostradas a 8 kHz para a realização dos experimentos. Como as informações de variações acústicas emocionais estão presentes nos sinais sonoros, os sons surdos foram suprimidos. Em seguida, os sinais foram concatenados aleatoriamente e resultaram em sinais de 40 s. Para a etapa de treinamento foram utilizados 25 s. Os modelos do UBM e dos estados emocionais foram elaborados com 1024 gaussianas a partir da matriz de atributos MFCC com vetores de 12 coeficientes extraídos de janelas de Hamming de 20 ms e 50% de sobreposição. Os sinais em estado de felicidade,

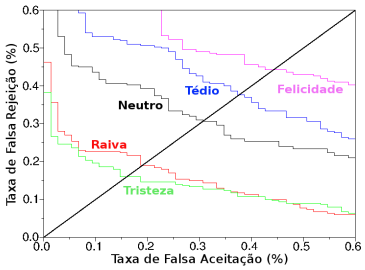

Fig. 3. Curva DET referente ao sistema α-GMM-UBM com testes de 1,0 s e α_{UBM}=-2.

neutro, raiva, tédio, tristeza foram utilizados para teste, obtidos pela segmentação de 15 s em tamanhos de 1,0 s e 0,5 s. A técnica *Round Robin* foi utilizada para criação do UBM proporcionando novas configurações de teste e uma soma total de 1875 testes (375 de falsa rejeição e 1500 de falsa aceitação) para os casos de testes de 1,0 s. Os resultados da avaliação comparativa incluem o valor médio de EER e DCFmin para sinais de teste com tamanhos de 1,0 s e 0,5 s.

A. EER: UBM proposto (α-GMM) e referencial (GMM)

O primeiro experimento deste trabalho avalia o desempenho da verificação de cada estado emocional em função dos valores de α para a modelagem do UBM. O objetivo é a definição de modelo α-GMM para o referencial que obteve os menores valores de EER para todos os cenários de teste. O parâmentro α utilizado para a modelagem do UBM (α_{UBM}) foi variado de -2 a -8. Testes com sinais de 1,0 s foram realizados para o modelo referencial (GMM) e o proposto (α-GMM). A Fig. 3 ilustra a curva DET do sistema α-GMM-UBM ao adotar α_{UBM}=-2. Nesta abordagem as menores taxas de EER ocorrem para as emoções raiva e tristeza. Os resultados individuais de EER alcançados para as diversas configurações estão presentes na Tab. I.

TABELA I

EER(%) OBTIDO PARA TESTES DE 1,0 S EM FUNÇÃO DE DIFERENTES VALORES DE α_{UBM}.

Emoções	GMM ($\alpha=-1$)	α_{UBM}			
		$\alpha=-2$	$\alpha=-4$	$\alpha=-6$	$\alpha=-8$
Raiva	21,3	**19,0**	29,0	29,3	36,0
Felicidade	**42,0**	44,3	46,7	60,0	50,0
Neutro	38,0	**30,7**	40,0	44,0	49,7
Tédio	**33,3**	37,7	41,3	45,3	52,0
Tristeza	16,7	**16,0**	24,0	23,7	21,3

Note que, o modelo proposto obteve menores taxas de EER para os casos raiva (19,0%), neutro (30,7%) e tristeza (16,0%) quando adotado α_{UBM}=-2. Além disso, as emoções raiva e tristeza por apresentarem os maiores valores de INS obtiveram menores taxas de erro para todas as configurações. No caso do estado neutro, o valor de EER é reduzido de 7,3 p.p. (pontos percentuais). O sistema baseado no classificador referencial GMM apresentou menores valores de erro para os demais estados. Este resultado comprova a eficiência do modelo α-GMM para classificação do UBM.

B. EER: modelagem α-GMM para emoções

Neste segundo experimento, espera-se a definição dos sistemas de verificação de emoções ao empregar o modelo α-GMM tanto para o UBM quanto para os distintos estados emocionais. Os testes foram realizados com sinais de 1,0 s. Os valores de α utilizados para o UBM (α_{UBM}) e para as emoções (α_{EMO}) foram variados de -1 a -8. Este experimento foi realizado em duas etapas. A primeira consiste em fixar um valor para α_{UBM} (representado por diferentes cores na Fig. 5). Na segunda etapa observou-se o comportamento do sistema conforme a variação do parâmetro α para modelagem das emoções (eixo α_{EMO}). Note que a composição α_{UBM}=α_{EMO}=-1 equivale ao sistema referência (GMM-UBM). A Fig. 5 ilustra os valores de EER médios obtidos para todas as configurações.

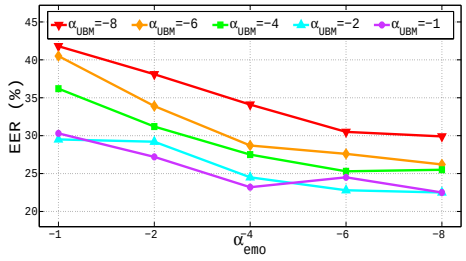

Fig. 5. EER médio para sistemas de verificação com α_{UBM} fixo em função da variação de α_{EMO}.

Pela análise da Fig. 5, quanto mais α_{EMO} se aproxima de -8 menores são as taxas de erro dos sistemas. De forma geral, as composições obtidas com α_{UBM}=-2 exibem as menores taxas de erro médio. Estes sistemas estão representados na pela cor azul. A composição que apresenta a menor taxa de erro médio é formada por α_{UBM}=-2 e α_{EMO}=-8. Os resultados obtidos por emoção ao fixar α_{UBM}=-2 em conjunto com os testes correspondentes ao sistema referencial e de representação esparsa SRV são exibidos na Tab. II. A partir dos resultados apresentados observa-se que os menores valores de EER de cada emoção são alcançados ao empregrar o modelo α-GMM. No caso em que α_{EMO}=-8 obtem-se as menores taxas de erro para 3 estados: raiva (11,0%), felicidade(25,3%) e neutro(28,0%). O emprego de α_{EMO}=-6 acarreta no menor valor de EER para as emoções tédio (30,0%) e tristeza (13,0%). A classificação SRV quando comparado ao referencial GMM-UBM apresentou menor taxa de erro para a emoção felicidade (32,4%) além do estado neutro (32,3%). Os sistemas expostos foram empregados para verificação de sinais altamente não-estacionários. A partir dos resultados apresentados, observa-se que o α-GMM-UBM além de apresentar as menores taxas de erro individuais também exibe as menores taxas de erro médio para α_{EMO}=-8 (22,7%) e α_{EMO}=-6 (22,8%).

TABELA II

EER(%) PARA TESTES DE 1 S NOS CASOS SRV, GMM-UBM E α-GMM-UBM COM α_{UBM}=-2 E α_{EMO} VARIADO.

Emoções	GMM	α_{EMO}				SRV
		$\alpha=-2$	$\alpha=-4$	$\alpha=-6$	$\alpha=-8$	
Raiva	21,3	22,7	11,7	14,0	**11,0**	31,4
Felicidade	42,0	42,7	29,3	28,7	**25,3**	32,4
Neutro	38,0	32,0	30,3	28,3	**28,0**	32,3
Tédio	33,3	32,3	35,7	**30,0**	31,0	48,6
Tristeza	16,7	16,3	15,3	**13,0**	18,0	34,3
Média	**30,3**	**29,2**	**24,5**	**22,8**	**22,7**	**35,8**

C. Curva DET: Sistemas SRV, GMM-UBM e α-GMM-UBM

Neste trabalho, também foram realizados experimentos de verificação de emoções utilizando sinais de tamanho 0,5 s.

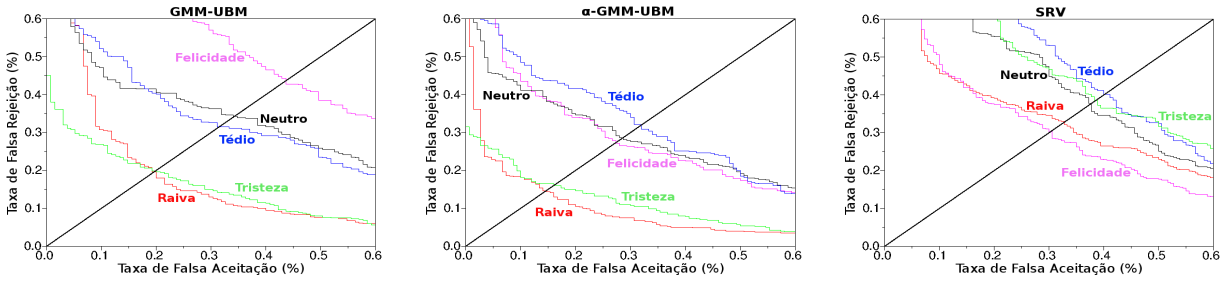

Fig. 4. Curvas DET obtida para testes de 0,5 s nos sistemas SRV, GMM-UBM e α-GMM-UBM (α_{UBM}=−2 e α_{emo}=−6).

As curvas DET dos sistemas SRV, GMM-UBM e α-GMM-UBM (α_{UBM}=−2 e α_{EMO}=−6) são apresentadas na Fig. 4. De modo geral, os sistemas obtiveram as menores taxas de erro para emoções raiva e tristeza. Novamente, o sistema α-GMM-UBM exibiu os menores valores de EER para a maioria das emoções: raiva (14,7%), felicidade (27,3%), neutro (28,7%) e tristeza (16,0%). A proposta deste trabalho apresenta uma redução de 16,4 p.p. no valor de EER para a emoção raiva quando comparada ao GMM-UBM e 22,0 p.p. para a emoção tristeza quando confrontada ao SRV.

D. Análise Comparativa:

Uma análise comparativa do desempenho dos sistemas de verificação foi realizada pela comparação dos valores médios de EER e valores mínimos de DCF obtidos em cada teste. Estes resultados estão apresentados na Tab. III.

TABELA III
VALORES DE EER MÉDIO DE DCFMIN OBTIDOS NOS EXPERIMENTOS PARA OS SISTEMAS GMM-UBM, α-GMM-UBM E SRV.

Sistema	1,0 s		0,5 s	
	EER (%)	DCFmin	EER (%)	DCFmin
α-GMM-UBM	**22,7**	0,247	**23,7**	0,258
GMM-UBM	30,3	0,330	29,7	0,324
SRV	35,8	0,390	35,8	0,390

Em média a proposta α-GMM-UBM apresenta uma redução de 7,6 p.p. quando comparado ao GMM-UBM e 13,1 p.p. quando comparado ao SRV para testes de 1,0 s. A redução equivalente aos testes de 0,5 s são de 6,0 p.p. e 12,1 p.p. respectivamente. Além disso, a função custo DCF apresenta valor mínino para a proposta para ambos os tamanhos de teste.

V. CONCLUSÃO

Este artigo apresentou a solução α-GMM-UBM para a tarefa de verificação acústica de emoções. Os resultados demonstraram que o sistema proposto apresenta menor taxa de erro para cada emoção para os testes de 1,0 s quando comparado ao referencial GMM-UBM e o SRV. O mesmo ocorre para os testes de 0,5 s a exceção do estado emocional tédio. A utilização do sistema α-GMM-UBM leva a uma redução média de 7,6 p.p. e 13,1 p.p. quando comparado ao GMM-UBM e SRV para os testes de 1,0 s. Além disso, o sistema proposto apresenta redução de 6,0 p.p. e 12,1 p.p. quando utiliza-se sinais de teste de 0,5 s.

REFERÊNCIAS

[1] C. Darwin, "The Expression of Emotions in Man and Animals", John Murray, Ed., 1872. Reprinted by Univ. Chicago Press, 1965.
[2] R. Cowei, E. Douglas-Cowei, N. Tsapatsoulis, G. Votsis, S.Kollias, W. Fellenz and J.G.Taylor, "Emotion recognition in humam-computer interaction," *IEEE Signal Processing Magazine*, vol. 18, Jan 2001.
[3] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll and A. Wendemuth, "Acoustic emotion recognition: A benchmark comparison of performances," *Automatic Speech Recognition & Understanding*. IEEE Workshop on, pp. 552-557, 2009.
[4] M. E. Ayadi, M. S. Kamel and F. Karray, "Survey on speech recognition: resources, features and methods", *Pattern Recognition*, Mar 2011.
[5] D. A. Reynolds and R. C. Rose. "Robust text-independent speaker identification using Gaussian mixture speaker models." *Speech and Audio Processing, IEEE Transactions*, 1995.
[6] D. Reynolds, T. Quatieri and R. Dunn, "Speaker verification using adapted Gaussian mixture models", *Digital Signal Process.*, 2000.
[7] F. Bimbot, J. F. Bonaste, C. Fredoulli, G. Gravier, M. I. Chagnolleau, S. Meignier, T. Merlin, O. J. Garcia, P. Delacretez and D. Reynolds, "A tutorial on text-intependent speaker verification", *EURASIP J. Appl. Signal Process.*, vol.4, pp. 430-451, 2004.
[8] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Trans. on Acoust., Speech, and Sig. Proc.*, Aug 1980.
[9] D. Wu, J. Li e H. Wu, "α-GMM mixture modelling for speaker recognition," *Pattern Recognition Letters*, vol. 30, no. 6, 2009.
[10] A. Venturini, L. Zão and R. Coelho, "On Speech Features Fusion, α-Integration Gaussian Modeling and Multi-Style Training for Noise Robust Speaker Classification", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, Dec 2014.
[11] D. Cavalcante and R. Coelho, "Atributos Acústicos Baseados na Simetria Glotal e no Classificador Alfa-GMM para Identificação de Emoções e Locutor", Anais do XXX Simp. Bras. de Telec., 2012.
[12] J. M. Pickett "The sounds of speech communication." Baltimore, MD: University Park (1980).
[13] T. F. Quatieri, "Discrete-Time Speech Signal Processing", Upper Saddle River, Prentice-Hall, 2001.
[14] L. Zão, D. Cavalcante e R. Coelho, "Time-Frequency Feature and AMS-GMM Mask for Acoustic Emotion Classification", *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 897-909, May 2014.
[15] W.Campbell, S. Sturim, D. Reynolds and A. Solomonoff, "Svm based speaker verification using gmm supervector kernel and nap variability compensation", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006.
[16] P. Kenny, G. Boulianne, and P. Domouchel, "Eigenvoice modeling with sparse training data," *IEEE Trans. Speech Audio Process.*, vol.13, 2005.
[17] N. Dehak, P. Kenny, R. Dehak, P. Domouchel, and P. Ouellet, "Front-end factor analisys for speaker verification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no.4, May 2011.
[18] P. Borgnat, P. Flandrin, P. Honeine, C. Richard and J. Xiao, "Testing Stationarity With Surrogates: A Time-Frequency Approach," in IEEE Transactions on Signal Processing, vol. 58, no. 7, Jul 2010.
[19] "Compressed sensing", *IEEE Trans. on Info. Theory*, 2006.
[20] J. C. Wang, Y. H. Chin, B. W. Chen, C. H. Lin and C. H. Wu, "Speech Emotion Verification Using Emotion Variance Modeling and Discriminant Scale-Frequency Maps," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, Oct 2015.
[21] F. Burkhardt, A. Paetchke, M. Rolfes, W. Sendlmeier and B. Weiss, "A database of german emotional speech", *Proc. of the Inters.*, 2005.
[22] A. Martin et al., "The DET Curve in Assessment of Detection Task Performance", *Proceedings of the Eurospeech*, vol. 4, Sep 1997.
[23] D. Reynolds, "Comparison of Background Normalization Methods for Text-Independent Speaker Verification", *Proc. Eurospeech*, 1997.