

Classificação de Sinais de Sensores Acústicos Seletivos de *Drones*

R. Alcântara e R. Coelho

Resumo— Este artigo apresenta um estudo de classificação de sinais de fonte acústicas captadas por *drones*. As principais contribuições são o uso de sinais *surrogates* e do classificador α -GMM (GMM α -integráveis) neste cenário. Experimentos foram conduzidos com seis fontes ambientais e duas ego-interferências com diferentes graus de estacionariedade. Os resultados mostraram que o α -GMM e sinais *surrogates* incrementaram separadamente a acurácia de classificação. Além disso, foi demonstrado que o uso conjunto do classificador e dos modelos *surrogates* superou a acurácia obtida individualmente, alcançando 73,8% de média de taxa de acerto.

Palavras-Chave— Classificação de fontes acústicas, *surrogates*, aprendizagem adaptativa, VANT.

Abstract— This article presents a study on the classification of acoustic source signals captured by *drones*. The main contributions are the use of surrogate signals and the α -GMM (α -integrated GMM) classifier in this scenario. Experiments were conducted with six environmental sources and two ego-noises with different degrees of stationarity. The results showed that α -GMM and surrogate signals separately provided an increase in classification accuracy. Furthermore, it was demonstrated that the joint use of the classifier and surrogate models surpassed the accuracy of the techniques individually, achieving an average accuracy of 73.8%.

Keywords— Acoustic source classification, *surrogates*, adaptive learning, UAV.

I. INTRODUÇÃO

A classificação de fontes acústicas possui crescente relevância na área de processamento de sinais acústicos [1]-[3]. Diversas aplicações no âmbito de navegação robótica [4], monitoramento de pessoas idosas [5] e *smart homes* utilizam esta tarefa. Em particular, a classificação de fontes acústicas a partir de sinais captados por sensores embarcados em *drones* é um desafio. O motivo é a proximidade destes sensores a hélices e rotores do veículo, que geram a chamada ego-interferência [6][7]. Os sinais captados desta maneira geralmente apresentam baixos valores de SNR (*Signal-to-Noise Ratio*). Isto faz com que o desempenho de sistemas de detecção e classificação seja reduzido drasticamente quando embarcados em *drones* [8].

Soluções usuais para classificação de fontes requerem grande quantidade de dados para treinamento [9]. Alternativamente, a técnica ALSS (*Adaptive Learning with Surrogate Assistance* – Aprendizagem Adaptativa Assistida de *Surrogates*) foi proposta recentemente para cenários com poucas amostras

disponíveis [10]. A ideia consiste em gerar *surrogates* que representem as características dos sinais de treinamento. Em seguida, os modelos *surrogates* mais discriminativos são selecionados para substituir os modelos originais, aumentando a acurácia da classificação. O uso da técnica ALSS se mostrou capaz de incrementar a taxa média de acerto de classificação em mais de 10 pontos percentuais (p.p.). Além disso, foi demonstrado em [11] que a técnica foi eficaz em recuperar as taxas de acerto também em ambientes reverberados.

O presente artigo apresenta um estudo de classificação de sinais de sensores acústicos seletivos embarcados em *drones*. Os experimentos foram realizados com seis fontes acústicas ambientais e dois sinais de ego-interferências. Inicialmente, experimentos foram conduzidos utilizando a técnica ALSS e o classificador α -GMM (α -integrated GMM – GMM α -integráveis) [12] separadamente neste cenário. Os resultados mostraram que, em ambos os casos, foram obtidas maiores taxas de acerto na classificação. Por fim, os resultados demonstraram que o incremento da acurácia foi maior quando a técnica ALSS foi utilizada em conjunto com o α -GMM. Neste caso, a acurácia média de classificação alcançou aumento de 10,9 p.p..

As principais contribuições deste trabalho são:

- Analisar a classificação de sinais de sensores acústicos captados por *drones* e a eficiência da técnica ALSS neste cenário.
- Avaliar a eficácia do classificador α -GMM para a tarefa de classificação de fontes acústicas.
- Demonstrar que técnica de ALSS, em conjunto com o classificador α -GMM, é capaz de superar o uso das técnicas separadamente.

O restante deste artigo está organizado da seguinte maneira: A Seção II descreve a captação de sinais acústicos por *drones* e realiza uma análise dos sinais de ego-interferência utilizados neste trabalho. Na Seção III, é mostrada a técnica ALSS e a obtenção dos modelos *surrogates*. A Seção IV apresenta um resumo classificador α -GMM. Na Seção V, é avaliada a efetividade do uso dos modelos α -GMM da técnica ALSS para classificação de sinais captados por *drones*. Além disso, é avaliado o desempenho das técnicas em conjunto. Por fim, na Seção VI é apresentada a conclusão.

II. CAPTAÇÃO DE SINAIS ACÚSTICOS POR DRONES

A captação de sinais acústicos através sensores embarcados em *drones* é acompanhada da forte presença da ego-interferência. Este ruído é gerado pelos rotores e hélices dos *drones* e degradam significativamente o sinal captado, com valores de SNR abaixo de 20 dB [13]. Como mostrou [14], o

R. Alcântara é doutorando do Programa de Pós-Graduação em Engenharia de Defesa do Instituto Militar de Engenharia (IME). O trabalho dos autores R. Alcântara e R. Coelho é desenvolvido no Laboratório de Processamento de Sinais Acústicos (LASP/IME) e parcialmente financiado pelo CNPq (305488/2022-8) e pela FAPERJ (200518/2023). E-mails: {raoni, coelho}@ime.eb.br.

espectro da ego-interferência depende da velocidade de rotação dos rotores do *drone*, que pode variar ao longo do tempo. Além disso, fatores exteriores como vento influenciam nas componentes que formam este sinal.

Neste trabalho, dois tipos de sinais de ego-interferências são consideradas fontes a serem identificadas entre fontes acústicas ambientais, captadas sem a presença destes ruídos. A base de sinais dos experimentos utilizou duas fontes de ego-interferências obtidas da base AVQ [7], aqui chamadas de Ego (E) (estacionário) e Ego (NE) (não-estacionário). Os sinais destas fontes foram capturados por um sensor acústico selecionado aleatoriamente de um arranjo circular de microfones embarcado em um quadricóptero 3DR IRIS. Os sinais estacionários e não-estacionários foram obtidos, respectivamente, com potência constante e potência variável dos seus motores do *drone*.

A Figura 1 ilustra os espectrogramas e respectivos valores de INS (*Index of Non-Stationarity* – Índice de Não-Estacionariedade) [15] de dois sinais de ego-interferência utilizados neste trabalho. Os valores de INS máximo ($INS_{\text{máx}}$) e limiares de estacionariedade mostram que estes sinais apresentam graus de estacionariedade distintos. O sinal em (a) possui valores de INS abaixo do limiar de estacionariedade, portanto é considerado estacionário ($INS_{\text{máx}} = 0,6$). O mesmo não ocorre para o sinal em (b), sendo classificado como não-estacionário ($INS_{\text{máx}} = 34,0$). Esta diferença também pode ser observada na variação dos espectrogramas dos sinais no tempo.

III. APRENDIZAGEM ADAPTATIVA ASSISTIDA POR SURROGATES

A técnica ALSS foi proposta para classificação de fontes acústicas em cenários com poucas amostras disponíveis para treinamento [10]. O seu principal objetivo é selecionar, para cada classe de sinais acústicos c , um modelo λ_c com maior separabilidade dos demais. Dado um conjunto de sinais de treinamento $\{\Phi_c^0 | c = 1, 2, \dots, C\}$ disponíveis para C classes diferentes. Para cada $c \in \{1, 2, \dots, C\}$, é extraída uma matriz de atributos \mathbf{Y}_c^0 de Φ_c^0 , que é utilizada para a obtenção dos modelos acústicos iniciais λ_c^0 . Em seguida, são gerados M sinais *surrogates* $\{\Psi_c^m | m = 1, 2, \dots, M\}$ utilizando as estatísticas de Kurtosis, densidade espectral de potência e INS dos sinais de treinamento. Depois, um conjunto de matrizes $\{\mathbf{Y}_c^m | m = 1, 2, \dots, M\}$ é estimado destes sinais para obtenção de um novo conjunto de modelo $\{\lambda_c^m | m = 1, 2, \dots, M\}$. Finalmente, estes modelos gerados a partir dos *surrogates* são comparados com os λ_c^0 para a definição do modelo mais informativo λ_c .

A geração dos *surrogates* é realizada a partir de um sinal de referência $\{x(t)\}$. Após a sua divisão em Q quadros com 50% de sobreposição, para cada quadro q , são executados os seguintes passos:

- 1) Sinais aleatórios $\{y_q(t)\}$ de amostras descorrelacionadas são gerados com amplitude distribuída a partir da Kurtosis (K) do quadro $\{x_q(t)\}$ de acordo com [16].
- 2) Um filtro de resposta ao impulso finita $\{h(t)\}$, baseado na DEP (densidade espectral de potência) de $\{x_q(t)\}$

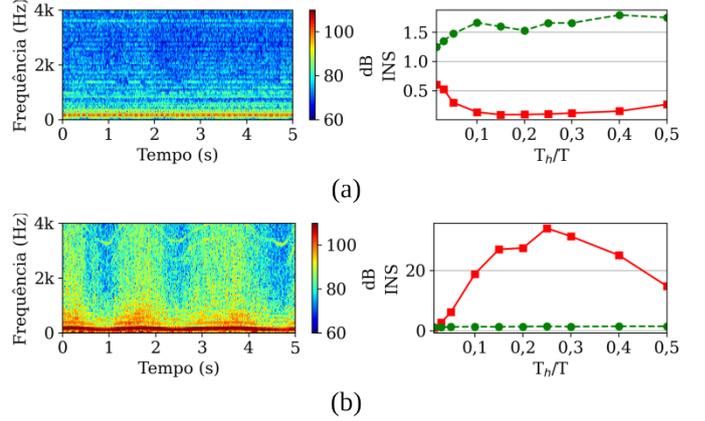


Fig. 1. Espectrogramas e INS dos sinais de ego-interferência (a) estacionário e (b) não-estacionário. Os valores de INS estão representados pela linha vermelha contínua, enquanto o limiar de estacionariedade γ está ilustrado na linha verde pontilhada.

[17]-[19], é aplicado na sequência $\{y_q(t)\}$ para a obtenção das amostras $\{\bar{y}_q(t)\}$.

- 3) O sinal é sintetizado a partir dos quadros $\{\bar{y}_q(t)\}$ e considerado um candidato a *surrogate*. Em seguida, o INS deste sinal é calculado e obtida a razão r entre o INS do sinal original (INS_{original}) e INS do candidato. Caso r se distancie, o candidato a *surrogate* é multiplicado por r^2 e o processo é repetido.

A seleção dos modelos mais representativos para cada classe $c \in \{1, \dots, C\}$ ocorre entre os modelos λ_c^0 e os modelos obtidos através dos *surrogates* λ_c^M . Inicialmente, a classificação é realizada com λ_c^0 , $c \in \{1, 2, \dots, C\}$ e um conjunto de sinais de avaliação. A taxa de acerto destes testes é dada por R . Em seguida, uma matriz de parâmetros \mathbf{Y}_c^m é extraída de cada *surrogate* Φ_c^m , dos quais são obtidos modelos acústicos $\{\lambda_c^m | c = 1, \dots, C; m = 1, \dots, M\}$. Estes modelos são utilizados para se obter os valores R_c^m , que equivalem às taxas de acerto de classificação com a substituição do modelo λ_c^0 por λ_c^m . Os modelos mais representativos são então escolhidos entre os que proporcionaram maior aumento da acurácia de classificação para cada classe c . Por fim, os modelos acústicos λ_c são considerados mais informativos que λ_c^0 se $R_c^m > R$. Neste caso, λ_c é escolhido para substituir os modelos originais λ_c^0 . Após a seleção final dos *surrogates*, os testes de classificação são realizados novamente com um novo conjunto de amostras de teste e os modelos selecionados, obtendo-se assim os resultados de classificação.

IV. O MODELO DE MISTURAS GAUSSIANAS α -INTEGRÁVES

O modelo α -GMM foi proposto em [12] como uma generalização do classificador GMM (*Gaussian Mixture Model* – Modelo de Mistura Gaussiana) [20]. Seja um modelo acústico λ_c , composto por G densidades de probabilidade Gaussianas $b_i(\mathbf{x}), i = 1, \dots, G$, a α -integração das densidades é definida por:

$$p(\mathbf{x}|\lambda_c) = C \left[\sum_{i=1}^G \pi_i b_i(\mathbf{x})^{\frac{1-\alpha}{2}} \right]^{\frac{2}{1-\alpha}}, \quad (1)$$

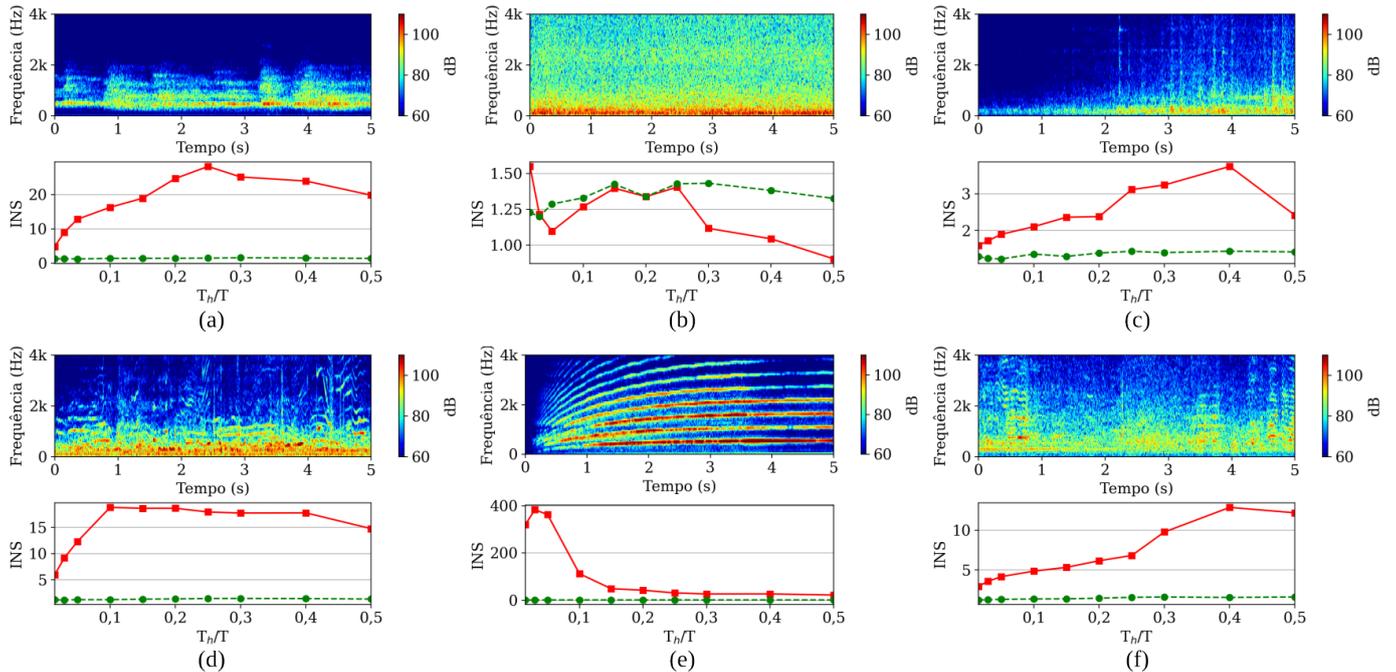


Fig. 2. Espectrogramas e INS dos sinais de treinamento das fontes: (a) Cães, (b) Chuva, (c) Metrô, (d) Parque, (e) Sirene, (f) Trânsito. Os valores de INS estão representados pela linha vermelha contínua, enquanto o limiar de estacionariedade γ está ilustrado na linha verde pontilhada.

em que π_i são os pesos das componentes gaussianas tais que $\sum_{i=1}^G \pi_i = 1$ e C representa uma constante de normalização. Quando $\alpha = -1$, a probabilidade se torna uma combinação linear de $b_i(\mathbf{x})$ e o modelo corresponde ao GMM tradicional.

Um modelo λ_c é parametrizado por um vetor de média, uma matriz de covariância e o peso de cada densidade Gaussiana. Estes parâmetros são obtidos no treinamento utilizando uma adaptação do algoritmo EM (expectativa-maximização – *expectation-maximization*) para maximizar a função verossimilhança

$$p(\mathbf{X}|\lambda_c) = \prod_{t=1}^Q p(\mathbf{x}_t|\lambda_c), \quad (2)$$

onde $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_Q]$ representa a matriz com os vetores de características extraídos de um segmento Φ_L de uma fonte acústica. Esta expressão da máxima verossimilhança é utilizada também como critério de decisão nos testes de classificação.

Esta generalização do GMM tradicional foi proposta para reconhecimento de locutor e ressalta maiores valores de probabilidade quando usados valores negativos de α . Em [21], o α -GMM também demonstrou melhorar o desempenho da classificação de estados emocionais e condições de estresse.

V. EXPERIMENTOS, RESULTADOS E DISCUSSÃO

Experimentos de classificação de fontes acústicas foram conduzidos com sinais de seis classes (Cães, Chuva, Metrô, Parque, Sirene e Trânsito) obtidas da base freeSFX. Estes sinais foram escolhidos por serem originados de fontes urbanas comumente captadas por *drones*. De cada classe, foram utilizados 3 sinais de 5 s de duração e taxa de amostragem de 22050 Hz para treinamento, avaliação dos modelos *surrogates*

TABELA I

ACURÁCIA DA CLASSIFICAÇÃO DE SINAIS (%) COM MFCC + GMM.

fonte verdadeira	classificação							
	Cães	Chuva	Ego (E)	Ego (NE)	Metrô	Parque	Sirene	Trânsito
Cães	31,8	0,0	0,0	0,0	0,4	24,8	7,4	35,6
Chuva	0,0	96,2	0,0	3,0	0,6	0,2	0,0	0,0
Ego (E)	0,0	1,8	53,4	43,6	0,0	0,2	1,0	0,0
Ego (NE)	0,0	1,0	14,8	84,2	0,0	0,0	0,0	0,0
Metrô	0,0	0,8	0,0	1,0	90,0	5,6	0,6	2,0
Parque	0,0	5,2	0,0	1,0	7,6	49,8	13,8	22,6
Sirene	0,0	0,0	0,0	0,0	0,0	3,6	94,6	1,8
Trânsito	0,0	0,4	0,0	1,6	82,2	13,0	0,2	2,6
Acurácia média: 62,8								

e testes, respectivamente. A Figura 2 apresenta os espectrogramas e os valores de INS dos sinais de treinamento de cada fonte acústica. De acordo com $INS_{m\acute{a}x}$, a fonte Sirene ($INS_{m\acute{a}x} = 383$) é altamente não-estacionária. As fontes Cães ($INS_{m\acute{a}x} = 28$), Parque ($INS_{m\acute{a}x} = 18$) e Trânsito ($INS_{m\acute{a}x} = 12$) são não-estacionárias. Por fim, as fontes Chuva ($INS_{m\acute{a}x} = 1,5$) e Metrô ($INS_{m\acute{a}x} = 3,7$) são consideradas estacionárias por terem seus valores de INS abaixo ou próximas do limiar de estacionariedade.

Para os experimentos de classificação de fontes acústicas, os sinais de 5 s de duração foram divididos em quadros de 20 ms com 50% de sobreposição. Em seguida, foram extraídos, de cada quadro, 12 componentes MFCC (*Mel-frequency cepstral coefficients* – Coeficientes Mel-cepstrais) [22], formando matrizes de tamanho 12×500 por sinal. Modelos GMM com 5 componentes foram gerados a partir das matrizes de coeficientes extraídas dos sinais de treinamento e de seus *surrogates*. A quantidade total de testes realizada para cada resultado de classificação foi de $500 \times 8 = 4000$. A Tabela I apresenta os resultados de classificação com MFCC

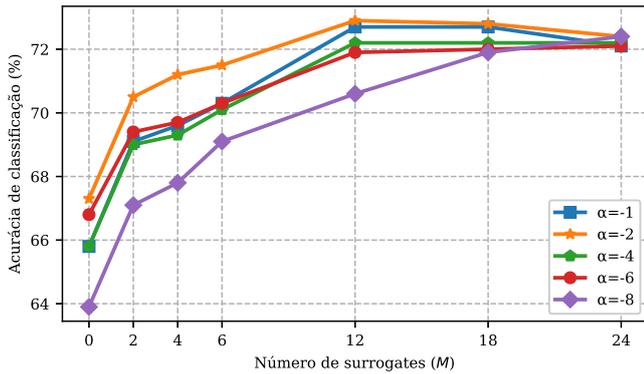


Fig. 3. Acurácia da classificação obtida com α -GMM e ALSS para múltiplos valores de α e quantidades de *surrogates* (M).

+ GMM. Neste cenário, a acurácia da classificação foi de 62,8%. Note que a fonte estacionária Chuva obteve maior acurácia, de 96,2%, enquanto a não-estacionária Trânsito teve o menor valor, de 2,6%. Adicionalmente, os quadros de Ego (E) foram classificados erradamente como Ego (NE) em 43,6% dos casos, um valor próximo da acurácia da fonte estacionária, de 53,4%.

Para a técnica ALSS, um total de 24 *surrogates* foram gerados para cada sinal disponível para treinamento. Os parâmetros utilizados para a geração foram definidos em [10]. Além de experimentos com ALSS e GMM, o α -GMM foi avaliado com diferentes valores de α : -2 , -4 , -6 e -8 . Os valores negativos de α foram escolhidos para ressaltar as maiores probabilidades na α -integração. A Figura 3 mostra as acurácias da classificação de fontes acústicas com e sem a técnica ALSS. Nela, estão representados os resultados para diferentes quantidades de *surrogates* (M) e múltiplos valores de α . Estes resultados foram obtidos a partir de experimentos com o conjunto de sinais de avaliação. Os valores para $M = 0$ indicaram que o melhor resultado de α -GMM sem a ALSS ocorreu para $\alpha = -2$, com taxa de acerto de 67,3%. Este resultado representa um incremento de 1,5 p.p. em relação ao GMM tradicional ($\alpha = -1$). Além de $\alpha = -2$, os modelos α -GMM com $\alpha = -6$ também apresentaram melhores desempenhos em relação ao GMM. O uso dos modelos *surrogates* aumentou a acurácia da classificação para todos os valores de α . O maior ganho ocorreu para $\alpha = -8$, de 8,5 p.p. com $M = 24$. Note que a curva de $\alpha = -2$ se manteve acima das demais para todos os valores de M e superando o GMM e alcançando acurácia de 72,9%. Em $M = 12$ foram obtidas as maiores taxas de acerto para $\alpha = -2$ e GMM. Portanto, os experimentos seguintes de classificação com α -GMM foram realizados com $\alpha = -2$ e os experimentos que utilizam a ALSS usaram $M = 12$ *surrogates*.

A Figura 4 ilustra a acurácia da classificação no cenário de testes realizados com α -GMM e ALSS separadamente. Nota-se que, em média, ocorreu aumento na taxa de acerto nos dois cenários. O classificador α -GMM foi capaz de incrementar a acurácia de 4 das 8 fontes acústicas, superando a ALSS+GMM em Chuva, *Ego-noise* E, Metrô e Sirene. Além disso, o α -GMM alcançou uma taxa de acerto média de 63,8%, 1,0

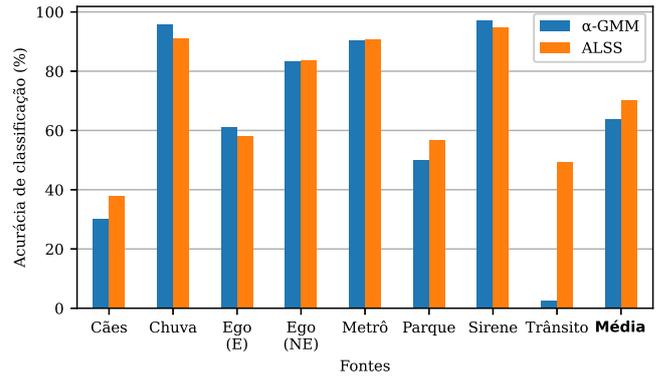


Fig. 4. Acurácias de classificação de cada fonte obtida α -GMM e ALSS separadamente.

TABELA II

ACURÁCIA DA CLASSIFICAÇÃO DE SINAIS (%) OBTIDAS COM MFCC + α -GMM E ALSS.

fonte verdadeira	classificação							
	Cães	Chuva	Ego (E)	Ego (NE)	Metrô	Parque	Sirene	Trânsito
Cães	39,8	0,0	0,0	0,0	0,2	42,6	17,4	0,0
Chuva	0,0	92,2	0,2	3,0	0,6	0,2	0,0	3,8
Ego (E)	0,0	0,2	53,8	20,8	0,0	0,0	0,0	25,4
Ego (NE)	0,0	0,6	16,0	83,4	0,0	0,0	0,0	0,0
Metrô	0,0	0,6	0,0	0,6	91,4	6,2	0,4	0,8
Parque	0,0	2,4	0,0	1,0	8,2	52,2	16,6	19,6
Sirene	0,0	0,0	0,0	0,0	0,0	1,2	98,8	0,0
Trânsito	0,0	0,2	0,0	2,2	0,2	18,4	0,2	78,8
Acurácia média: 73,8								

p.p. acima da acurácia com GMM. Os resultados com a técnica ALSS+GMM apresentaram um aumento significativo da acurácia de fonte Trânsito de 46,8 p.p., atingindo taxa de acerto média de 49,4%. Isto contribuiu para que a acurácia média com o emprego da técnica ALSS alcançasse 70,3%. Este resultado superou a taxa de acerto média da classificação sem ALSS com GMM em 7,5 p.p. e o emprego do α -GMM em 6,5 p.p..

Por fim, foram conduzidos testes de classificação de fontes acústicas com α -GMM e ALSS simultaneamente. A matriz de confusão relativa a estes experimentos está apresentada na Tabela II. O uso do α -GMM com os modelos *surrogates* alcançou maior acurácia média da classificação que os outros cenários, com 73,8%. Seis fontes acústicas tiveram maior acurácia neste cenário em relação aos resultados da Tabela I. Entre elas, destacou-se a fonte Trânsito, com melhora da taxa de acerto média de 76,0 p.p.. A fonte Chuva permaneceu entre as de maior acurácia, alcançando 92,2%. Também foi demonstrado que o uso do α -GMM e ALSS em conjunto diminuiu a capacidade dos modelos se confundirem com fontes diferentes. Os resultados mostram que a maior taxa de classificação errada de uma fonte foi de 42,6%. Este valor foi consideravelmente abaixo do máximo de 82,2% encontrado na Tabela I. Além disso, a matriz de confusão da Tabela II apresentou somente dois valores de classificação entre fontes diferentes acima de 20%, dois a menos que na Tabela I. Em relação à confusão entre os sinais de ego-interferência, com α -GMM e ALSS a taxa de classificação errada de Ego (E) como

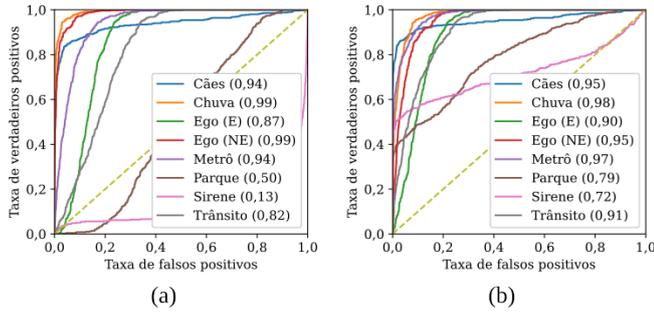


Fig. 5. Curvas ROC e respectivas AUC para classificação com: (a) GMM sem ALSS e (b) α -GMM com ALSS.

Ego (NE) foi de 20,8%. Este valor representa uma diminuição de 22,8 p.p. quando comparado ao cenário com GMM.

Na Figura 5 estão ilustradas as curvas ROC (*Receiver operator characteristic* – Característica de operação do receptor) para classificação com GMM sem ALSS e α -GMM com ALSS. Além disso, estão indicadas as AUC (*Area under the ROC curve* – Área abaixo da curva ROC) para cada fonte acústica. As curvas ROC de cada fonte foram levantadas considerando experimentos multi-classe entre a própria fonte contra as outras sete. Valores de AUC próximos de 1 indicam que o modelo da fonte possui maior separabilidade dos demais. Os resultados de AUC em (a) e (b) indicam que as fontes Parque e Sirene apresentaram ROCs com menor área nos dois cenários apresentados. Entretanto, estas fontes apresentaram níveis de acurácia de classificação distintos, de 49,8% e 94,6% com GMM sem ALSS, respectivamente. As curvas ROC mostram que a AUC média aumentou de 0,77 para 0,90 com a utilização conjunta do classificador α -GMM e ALSS. Os incrementos mais significativos dos valores de AUC ocorreram para as fontes Parque e Sirene de, respectivamente, 0,29 e 0,59. No total, seis fontes apresentaram aumento da AUC, indicando que a combinação de α -GMM e ALSS fez com que estes modelos se tornassem mais discriminativos.

VI. CONCLUSÃO

Este artigo apresentou um estudo de classificação de sinais de fonte acústicas captadas por *drones* com o uso do classificador α -GMM e aprendizagem adaptativa com *surrogates*. Experimentos foram realizados com seis fontes ambientais e duas ego-interferências. Os sinais das fontes utilizadas possuem diferentes graus de estacionariedade e foram avaliados através do INS. Os resultados dos experimentos mostraram que ambas as soluções separadamente proporcionaram maior acurácia de classificação. Além disso, foi demonstrado que o melhor desempenho ocorreu para a utilização do α -GMM e ALSS em conjunto. Neste caso, os resultados superaram o uso do GMM tradicional em 10,9 p.p., alcançando 73,8% de taxa de acerto média. Os resultados de classificação foram complementados com as curvas ROC. As curvas indicaram que o emprego de α -GMM e ALSS em conjunto incrementou a AUC de seis fontes acústicas. Isto demonstrou que a sua utilização fez com que fossem gerados e selecionados modelos das fontes acústicas mais representativos e discriminativos.

REFERÊNCIAS

- [1] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017.
- [2] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Piscataway, NJ: IEEE Press, 2006.
- [3] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. D. Plumbley, "Detection and classification of acoustic scenes and events: An ieeea asp challenge," in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4, 2013.
- [4] S. Chu, S. Narayanan, C.-c. J. Kuo, and M. J. Mataric, "Where am i? scene recognition for mobile robots using audio features," in *2006 IEEE International Conference on Multimedia and Expo*, pp. 885–888, 2006.
- [5] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun, "Towards robotic assistants in nursing homes: Challenges and results," *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 271–281, 2003. Socially Interactive Robots.
- [6] J. Martinez-Carranza and C. Rascon, "A review on auditory perception for unmanned aerial vehicles," *Sensors*, vol. 20, no. 24, 2020.
- [7] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, "Audio-visual sensing from a quadcopter: dataset and baselines for source localization and sound enhancement," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5320–5325, 2019.
- [8] C. Rascon, O. Ruiz-Espitia, and J. Martinez-Carranza, "On the use of the air-uas corpus to evaluate audio processing algorithms in unmanned aerial systems," *Sensors*, vol. 19, no. 18, 2019.
- [9] S. Sigtia, A. M. Stark, S. Krstulović, and M. D. Plumbley, "Automatic environmental sound recognition: Performance versus computational cost," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2096–2107, 2016.
- [10] G. Zucattelli, R. Coelho, and L. Zão, "Adaptive learning with surrogate assisted training models for acoustic source classification," *IEEE Sensors Letters*, vol. 3, no. 6, pp. 1–4, 2019.
- [11] G. Zucattelli and R. Coelho, "Adaptive learning with surrogate assisted training models using limited labeled acoustic sample sequences," in *2021 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 21–25, 2021.
- [12] D. Wu, J. Li, and H. Wu, " α -gaussian mixture modelling for speaker recognition," *Pattern Recognition Letters*, vol. 30, no. 6, pp. 589–594, 2009.
- [13] L. Wang and A. Cavallaro, "Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles," in *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 152–158, 2016.
- [14] L. Wang and A. Cavallaro, "Microphone-array ego-noise reduction algorithms for auditory micro aerial vehicles," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2447–2455, 2017.
- [15] 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, no. 7, pp. 3459–3470, 2010.
- [16] R. Webster, "A random number generator for ocean noise statistics," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, pp. 134–137, 1994.
- [17] M. A. Al-Alaoui, "Novel digital integrator and differentiator," *Electronics Letters*, vol. 29, pp. 376–378(2), Feb. 1993. Publisher: Institution of Engineering and Technology.
- [18] L. Zão and R. Coelho, "Generation of coloured acoustic noise samples with non-Gaussian distributions," *IET Signal Processing*, vol. 6, pp. 684–688(4), Sept. 2012. Publisher: Institution of Engineering and Technology.
- [19] R. Santana and R. Coelho, "Low-frequency ambient noise generator with application to automatic speaker classification," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, p. 175, Aug. 2012.
- [20] D. Reynolds and R. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.
- [21] V. Vieira, R. Coelho, and F. M. de Assis, "Hilbert–huang–hurst-based non-linear acoustic feature vector for emotion classification with stochastic models and learning systems," *IET Signal Processing*, vol. 14, no. 8, pp. 522–532, 2020.
- [22] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, 1980.