Detecção automática dos sinais acústicos de delfinídeos obtidos por monitoramento acústico passivo

Celso Jose Munaro*, Lilian Sander Hoffmann, Agnaldo Martins**, Beatriz Araujo Farias**, Helder Antonio Medina Avila****

**Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo/UFES, Vitória 29075-900, Brasil, (celso.munaro@ufes.br)*

***Laboratório de Nectologia, Departamento de Oceanografia e Ecologia, Universidade Federal do Espírito Santo/UFES, Vitória 29075-900, Brasil (liliansander@gmail.com, agnaldo.martins@ufes.br, beatriz.araujo.farias@gmail.com, helder.avila@edu.ufes.br)*

Abstract: Cetaceans, comprising whales, dolphins, and porpoises, are integral components of marine ecosystems, playing pivotal roles in maintaining ecological balance. Acoustic monitoring emerges as a crucial tool in understanding cetacean behavior, distribution, and population dynamics. Equipment that captures and records these signals over days generates a large amount of data that requires a great effort to extract only the intervals in which there are signals to be analyzed. This work aims to present a new methodology for automatically detecting the presence of delphinid vocalizations in audio signals. A test statistic is applied directly to the audio signals generating the intervals in which there are whistles or clicks and burst pulses. In the case of whistles, the spectrogram is computed and the contour showing the frequency over time is obtained, allowing its use in classifiers to indicate dolphin species and activities. The methodology was applied to two case studies and for detecting whistles, an accuracy greater than 89% and precision greater than 97% were achieved. When detecting clicks or burst pulses, the performance was 86% for accuracy and 81% for precision. The algorithms require low computational effort and can significantly reduce the effort required to extract captured signals, thus allowing an increase in the number of collection points and their periods.

Resumo: Os cetáceos, incluindo baleias, golfinhos e botos, são componentes integrantes dos ecossistemas marinhos, desempenhando um papel fundamental na manutenção do equilíbrio ecológico. O monitoramento acústico surge como uma ferramenta crucial na compreensão do comportamento, distribuição e dinâmica populacional dos cetáceos. Equipamentos que captam e gravam estes sinais ao longo de dias geram uma grande quantidade de dados que requerem um grande esforço para extrair apenas os intervalos nos quais há sinais a serem analisados. Este trabalho visa apresentar uma nova metodologia para detecção automática da presença das vocalizações de delfinídeos em sinais de áudio. Uma estatística de teste é aplicada diretamente aos sinais de áudio gerando os intervalos nos quais há assobios ou cliques e sons pulsantes. No caso de assobios, o espectrograma é computado e o contorno que mostra a frequência no tempo é obtido, permitindo seu uso em classificadores para indicar espécies e atividades dos golfinhos. A metodologia foi aplicada para dois estudos de caso e para detecção de assobios obteve-se uma acurácia maior que 89% e precisão maior que 97%. Na detecção de cliques ou sons pulsantes, o desempenho foi de 86% para acurácia e 81% para precisão. Os algoritmos requerem baixo esforço computacional e podem reduzir significativamente o esforço para sua extração de sinais capturados, permitindo assim aumentar o número de pontos de coleta e seus períodos.

Keywords: Bioacoustics; Dolphins; Whistles; Echolocation; Spectrogram; Classification; Patterns Recognition.

Palavras-chaves: Bioacústica; Golfinhos; Assobios; Ecolocalização; Espectrograma; Classificação; Reconhecimento de Padrões.

1. INTRODUÇÃO

Cetáceos são considerados sentinelas ambientais, e sua presença é um excelente bioindicador da qualidade do ecossistema onde vivem (Winship et al., 2023). Devido ao fato de passarem quase todo tempo embaixo da água, e pelo comportamento críptico e hábitos oceânicos de muitas espécies, uma das melhores ferramentas é o monitoramento acústico de suas vocalizações, pois permite a aquisição contínua de dados, o que não seria possível com observadores em campo (Korkmaz et al., 2022). As emissões acústicas produzidas pelos cetáceos possuem duas funções básicas: ecolocalização e comunicação. A ecolocalização (ou biossonar) destina-se a localizar e analisar objetos, bem como definir sua própria orientação, onde o animal emite cliques e detecta os ecos refletidos em objetos. A ecolocalização está fortemente desenvolvida no grupo conhecido como Odontocetos (golfinhos, botos e cachalotes). Os cliques são pulsos sonoros de banda ampla (até 150 kHz), com duração de 10 a 200μsec (Diercks *et al*., 1973). Estas altas frequências resultam em maior eficiência, por apresentarem comprimentos de onda mais curtos e maior poder de resolução (Marten, 2000). Existem ainda os sons pulsantes explosivos (*burst pulsed sounds*), são sequências de cliques com intervalo menor que 3ms (Lammers et al., 2003), sendo produzidos em certos contextos sociais (Herzing, 1996) ou para atordoamento de presas.

A categoria de sons relacionada à comunicação ocorre em toda a Ordem Cetacea. Os sons chamados de assobios, registrados em várias espécies de golfinhos, são tipicamente sons de frequência modulada de banda estreita, considerados sinais usados para organização do grupo (Janik & Slater, 1998). Os assobios apresentam a frequência fundamental usualmente abaixo de 20 kHz e harmônicas até 100 kHz (Lammers et al., 2003; Steiner, 1981), e durações entre 0,05 e 3,2s (Bazúa-Durán & Au, 2002). Golfinhos produzem um grande número de diferentes contornos de assobios que, juntos, compreendem o repertório da espécie e/ou população (Janik & Slater, 1998).

Para obtenção dessas vocalizações, são utilizados hidrofones, que se caracterizam por ser um método não invasivo (Sousa-Lima et al. 2013, Rio et al., 2024), não interferindo em seu comportamento natural. Atualmente, e com o avanço das tecnologias, uma das abordagens mais utilizadas vem sendo o monitoramento acústico passivo (MAP). Essa técnica consiste na instalação de equipamentos autônomos na coluna d'água ou fundo marinho, e que possibilita a aquisição de dados em grandes escalas temporais e espaciais. Contudo, o MAP gera, também, uma grande quantidade de informação, o que dificulta seu pós-processamento. A análise e interpretação de todas essas informações é dispendiosa e demanda muito tempo quando realizada de forma não automatizada. Por conta disso, tentativas de automatização vêm sendo desenvolvidas, para maximizar essas análises, de forma a indicar a presença de vocalizações e interpretar os sinais gerados (Korkmaz et al., 2022).

A literatura mostra avanços em relação a estas análises, concentrada na categoria dos assobios. Um estudo categorizou os tipos existentes de assobios aplicando filtros wavelet de Gabor aos espectrogramas (Esfahanian et al., 2014). Outro estudo (Kohlsdorf et al., 2016) relacionou o comportamento de golfinhos com os sons produzidos por eles durante diferentes atividades. Ainda, alguns autores sugeriram classificar as espécies de golfinhos a partir dos seus assobios (Gannier et al., 2020). Já pesquisas com redes neurais (Korkmaz et al., 2022) conseguiram identificar assobios em gravações de áudio, treinando redes convolucionais para diferenciar as espécies, baseado nas imagens geradas pelos espectrogramas.

Sabe-se que nem sempre os assobios são os sinais mais abundantes, por serem usados somente em determinados contextos, mas ainda assim a grande maioria dos estudos de monitoramento passivo, que visam a detecção dos animais, não fazem a utilização dos três tipos de vocalização. Os cliques são usados mais frequentemente, visto estarem relacionados à necessidade constante de orientação do animal e atividades de forrageamento. Com isso, o desenvolvimento de uma ferramenta de reconhecimento e detecção automática que levasse em consideração os três tipos de sinais (assobios, cliques e sons pulsantes), maximizaria a eficiência dos esforços de pesquisa, contribuindo ainda para o conhecimento e conservação das espécies. Neste estudo, pretende-se apresentar uma técnica para detectar, em um processo automatizado, a presença dos três tipos de vocalizações e também extrair destas os contornos dos assobios, para fins de parametrização espectral e temporal, embora não faça parte do escopo deste artigo.

Este artigo é organizado da seguinte forma: na seção de Metodologia discute-se o registro e a preparação dos dados, os métodos de detecção dos sinais de interesse e a extração dos contornos para o caso de assobios. Na seção de Resultados e Discussões os dados utilizados são apresentados bem como, a metodologia usando dados de teste, e as métricas de desempenho são calculadas usando dois estudos de caso. O artigo é finalizado com a seção de Conclusões resumindo a proposta e destacando suas vantagens.

2. METODOLOGIA

2.1 Registro das vocalizações

Para o estudo em questão, foram utilizados dados acústicos gerados a partir de gravações, em ambiente natural, de uma população do golfinho-nariz-de-garrafa (*Tursiops gephyreus*) no sul do Brasil (Hoffmann, 2004). O registro acústico das vocalizações dos golfinhos foi realizado através de um hidrofone C54XRS da Cetacean Research Technology (www.cetaceanresearch.com), imerso a uma profundidade de aproximadamente 3m, conectado a um gravador portátil Fostex FR-2 (taxa de amostragem de 48kHz e 24 bits resolução), resultando numa resposta linear de até 24 kHz.

2.2 Preparação dos dados

Os sinais selecionados para avaliação das rotinas de extração automática possuíam uma boa relação S/R, clareza no contorno, sem sobreposição de outros sons e sem corte nos limites de frequência. A análise dos sinais foi feita por espectrogramas, calculados pelo aplicativo Raven Pro 1.6 (Cornell Laboratory of Ornithology), com FFTs de 512 a 1024 pontos, janelamento Hanning e 50% de sobreposição. Foi realizada a distinção auditiva e visual dos sons, separando-os primariamente em assobios, cliques e sons pulsantes. Os intervalos com a presença de assobios, cliques e sons pulsantes em cada arquivo dos sinais de áudio foram assim rotulados para avaliar os algoritmos de detecção propostos. Esse ambiente requer um especialista para analisar os espectrogramas para concluir sobre a presença de sinais e seu tipo.

2.3 Detecção das vocalizações

No ambiente para detecção automática proposto, a primeira etapa da metodologia é detectar a presença de sons no sinal de áudio, podendo ser eles cliques, assobios ou sons pulsantes. Neste artigo, não distinguiremos cliques de sons pulsantes, de modo que ao nos referimos à detecção de cliques os sons pulsantes também são detectados. Um dos métodos da literatura é o uso do desvio padrão (Mallawaarachchi, 2008), que é alterado na presença de sons. A média não é alterada (ver Figura 1). Existem muitas formas de detectar alterações como estas em variáveis aleatórias (Montgomery et al., 2010). Utilizaremos neste artigo, como estatística de teste, o range móvel, calculado para cada amostra de sinal e definido por:

$$r_i = \left| y_i - y_{i-1} \right| \quad (1)$$

Para m observações, a média do range móvel é dada por:

$$\bar{r} = \frac{1}{m-1} \sum_{i=2}^{m} \left| y_i - y_{i-1} \right| \quad (2)$$

Uma estimativa do desvio padrão é:

$$\sigma = \frac{\bar{r}}{1{,}128} \quad (3)$$

Para monitoramento, usa-se a curva central definida por $\bar{r}$ e o limiar superior dado por (Montgomery et al., 2010):

$$UCL = 3{,}267 \times \bar{r} \quad (4)$$

Um conjunto de dados inicial com as m amostras deve ser fornecido para treinamento, ou seja, calcular $\bar{r}$. Com isto, o sinal pode ser monitorado e a cada nova amostra calculando r_i. Valores $r_i > UCL$ indicam a presença de sons no sinal monitorado. Ocorre que devido a aleatoriedade do sinal, pelo menos α% das amostras devem ultrapassar o limiar, onde α é o nível de confiança. A estratégia comum nestes casos é definir uma janela de comprimento L amostras e verificar se ao menos β% dos valores de $r_i > UCL$. A análise das N amostras do sinal y definirá desta forma intervalos nos quais há evidências de que há algum som a ser analisado afeta o tamanho da janela L usada para detectar os intervalos. Uma janela L grande não permite detectar os cliques por serem de curta duração. O parâmetro β também deve ser maior para assobios, pela maior quantidade de amostras tais que $r_i > UCL$ do que para cliques. Obtidos os intervalos, eles são selecionados de acordo com sua duração em amostras. Intervalos de assobios têm maior duração, e aqueles menores do que um parâmetro I_{L1} são eliminados. Intervalos de cliques têm menor duração e aqueles maiores que um parâmetro I_{L2} são eliminados. Por fim, agrupar intervalos próximos é uma estratégia muito útil para fornecer um único intervalo ao invés de muitos intervalos de pequena duração e próximos. Eles ocorrem naturalmente devido à variação da amplitude dos sinais de áudio (ver Figura 1). Usa-se para isso o parâmetro B_1 para assobios e B_2 para cliques. Pode-se então enumerar os passos para o algoritmo de detecção dos sons nos arquivos de áudio:

1. Selecionar as amostras do sinal de áudio tais que $r_i > UCL$.
2. Para assobios:
   a. Selecionar intervalos de L_1 amostras tais que pelo menos β_1% das amostras atendem $r_i > UCL$.
   b. Manter apenas os intervalos com mais de I_{L1} amostras
   c. Agrupar intervalos que têm menos de B_1 amostras entre si
3. Para detectar cliques e sons pulsantes:
   a. Selecionar intervalos de L_2 amostras tais que pelo menos β_2% das amostras atendem $r_i > UCL$.
   b. Manter apenas os intervalos com menos de I_{L2} amostras.
   c. Agrupar intervalos que têm menos de B_2 amostras entre si

Discutiremos agora a obtenção dos parâmetros utilizados nos passos acima. Eles podem ser obtidos com um ou mais arquivos de áudio que contenham: intervalos sem sons a detectar, intervalos com assobios, cliques e sons pulsantes, não ocorrendo essas diferentes situações simultaneamente. O

primeiro passo é gerar a estatística r_i a partir de y usando a equação (1), como mostrado nos dois gráficos superiores da Figura 1. Os valores de $\bar{r}$ e UCL são calculados usando apenas amostras com ruído. O gráfico inferior mostra os 120 intervalos selecionados de r_i tais que ao menos β% dos valores de $r_i > UCL$, usando $\beta = 0.1$ e $L = 50$. Observa-se intervalos de curta duração que caracterizam cliques e sons pulsantes e de maior duração que caracterizam assobios. Observa-se na Figura 1 que isto gera intervalos menores para cliques, que podem assim ser discriminados dos assobios. Os intervalos de cliques vão até 0.45s e os intervalos de assobios de 0,45s a 1,7s. O parâmetro β_1 para assobios é obtido da proporção de valores em que $r_i > UCL$ em uma janela de tamanho L_1 no trecho de assobios 0.5s a 1.7s na Figura 1. Para cliques, obtém-se β_2 de forma similar no trecho de 0s a 0.45s na mesma figura usando uma janela de tamanho L_2. O tamanho da janela L é menor para cliques e sons pulsantes (L_2) e é maior para assobios (L_1), pois maiores valores de L não permitem a detecção dos primeiros. Valores pequenos de L tendem a gerar muitos intervalos na detecção de assobios. Assim, β_1 e L_1são ajustados em conjunto de forma a indicar a presença dos assobios no conjunto de dados de teste. O mesmo é feito com β_2 e L_2 para os cliques. A largura dos intervalos permite escolher os seguintes parâmetros: duração máxima de cliques: $I_{L2} = 250$ e duração mínima dos assobios $I_{L1} = 1000$ amostras. A distância entre intervalos também é obtida com os dados desta figura. O valor de B_1 para agrupar intervalos de assobios é escolhido usando o intervalo de dados de 0,45s a 1,7s. O valor de B_2 para agrupar intervalos de cliques é escolhido avaliando a distância dos mesmos de 0 a 0.45s.

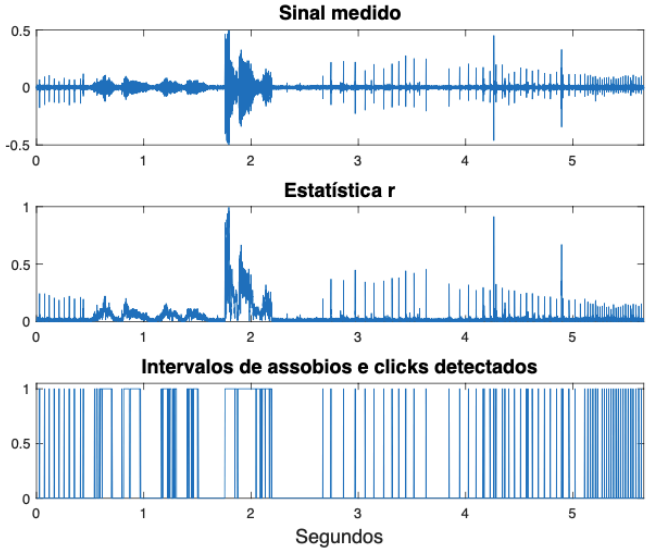

Figura 1. Passos iniciais da metodologia

2.4 Extração de contornos

Neste trabalho a detecção de cliques e sons pulsantes visa apenas informar sobre a presença dos golfinhos produzindo este tipo de som. Os intervalos de assobios são processados para gerar os contornos do espectrograma, feito sobre os intervalos de amostras correspondentes. São identificadas as faixas de frequências do espectro para cada instante de tempo, de forma similar a aplicações para detectar sinais oscilatórios em processos industriais (Dambros et al., 2019). A diferença aqui é que a frequência muda rapidamente no tempo, gerando os contornos. A presença de oscilações gera estreitas faixas no espectro que podem ser detectadas. Para isto, é preciso definir um limiar γ acima do qual estas faixas serão buscadas, o que é feito usando um trecho inicial do sinal em que há apenas ruído. A frequência no instante t é a média das frequências que ultrapassam o limiar, de forma similar ao que é feito em Gannier et al. (2020). Na construção do espectrograma é comum ocorrerem instantes em que não há frequências acima do limiar, produzindo trechos sem informação nos contornos. Esses trechos são os mesmos que produzem os gaps entre intervalos já discutidos e são facilmente reconstruídos usando técnicas de interpolação lineares ou quadráticas (Erdogan et al., 2013). Essa metodologia é agora aplicada a um conjunto de registros acústicos das vocalizações dos golfinhos.

Todos os algoritmos descritos foram escritos usando o software Matlab R2019b e a biblioteca de processamento digital de sinais.

3. RESULTADOS E DISCUSSÕES

3.1 Exemplo de aplicação

Os dados mostrados na Figura 1 são utilizados para ilustrar a aplicação da metodologia. Esse ajuste foi realizado com o conjunto de dados de teste da Figura 1. Os valores de $\beta_1 = 0.75$ e $\beta_2 = 0.13$ foram selecionados da mesma figura verificando o percentual de amostras tal que $r_i > UCL$ para janelas de largura $L_1 = 300$ para assobios e largura $L_2 = 50$ para cliques. Verificou-se que a sensibilidade a estes parâmetros é baixa. O comprimento máximo dos intervalos de cliques foi avaliado em $I_{L2} = 400$ e o comprimento mínimo dos intervalos dos assobios foi de $I_{L1} = 800$ com os dados de teste. Finalmente, foram ajustados os parâmetros utilizados para realizar o agrupamento de intervalos próximos. Intervalos de assobios são agrupados somente se houver menos de $B_1 = 800$ amostras entre eles. No caso de cliques, os intervalos são maiores, sendo $B_2 = 10000$.

Os resultados da aplicação dos algoritmos de detecção de assobios e cliques usando estes parâmetros são mostrados na Figura 2. Os gráficos da estatística $r_i > UCL$ para assobios e

cliques são diferentes devido aos valores de janelas L_1 e L_2 bem como os limiares β_1 *e* β_2. Foram detectados 3 intervalos de assobios entre 0,55 e 2,2s, e dois intervalos de cliques entre 0 e 0,54s e outro de 2,67 até o final dos dados. A observação do sinal medido dá indicações que as detecções estão corretas, pela presença de variações de pequena duração durante os cliques e de maior duração durante os assobios. O espectrograma mostrado na Figura 3 confirma estes intervalos. Os assobios têm contornos que aumentam e diminuem em frequência, enquanto os cliques se comportam como impulsos de curta duração. Os intervalos detectados podem ser então usados para as análises subsequentes. Usaremos o algoritmo proposto para produzir o contorno a partir do espectrograma, usando para isto os dois primeiros intervalos. O parâmetro γ para detectar as faixas de frequência é obtido a partir da média da FFT dos 5 primeiros instantes de tempo, ou seja, a média das 5 primeiras colunas do espectrograma, resultando em γ=0.68. O cálculo da média das frequências que ultrapassam o limiar gera o contorno, mostrado na Figura 4. Os trechos sem pontos são aqueles para as quais não havia valores do espectro com amplitude maior que γ, e que podem posteriormente serem substituídos por valores interpolados. Contornos como o mostrado são então utilizados para treinar classificadores para identificar diversas características dos delfinídeos que os produziram (Korkmaz et al, 2022).

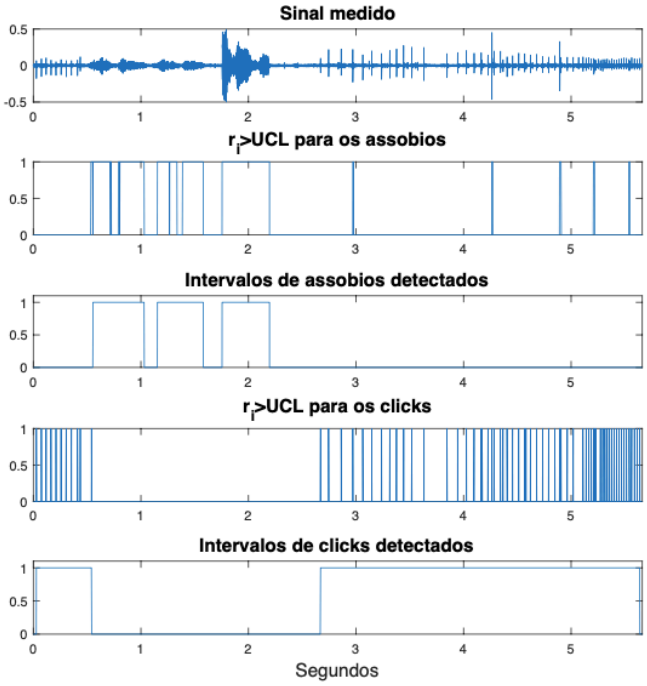

Figura 2. Resultado da aplicação dos algoritmos

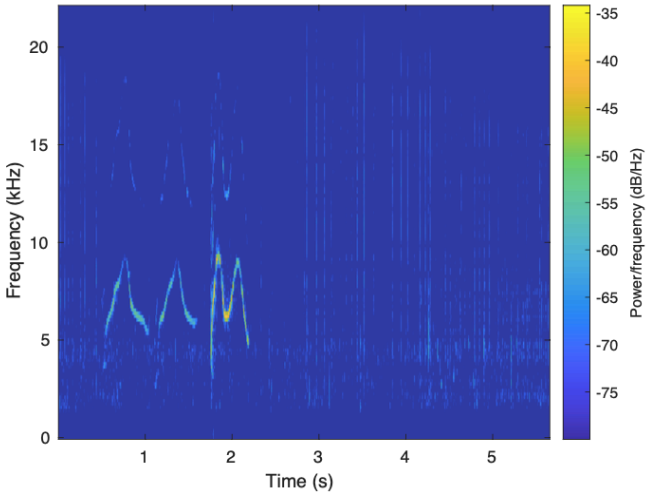

Figura 3. Espectrograma da sequência de assobios.

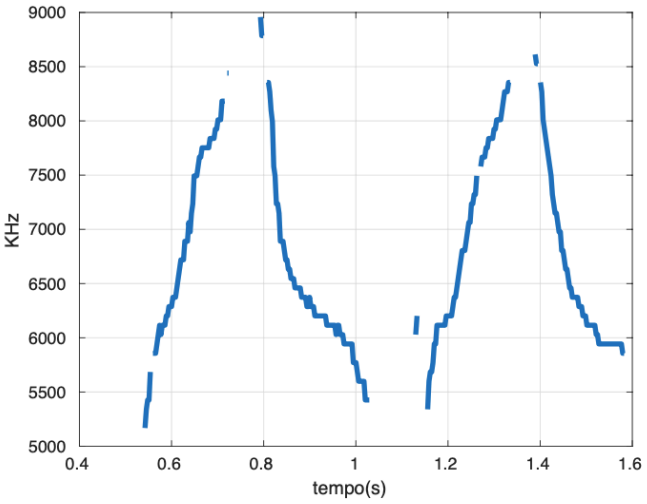

Figura 4. Contorno de dois assobios.

3.2 Avaliação estatística da metodologia

Para os dois casos, o desempenho foi calculado usando as métricas de precisão e acurácia. Esses conjuntos foram rotulados às cegas por um especialista, avaliando o espectrograma e o sinal de áudio. Um vetor com o mesmo número de amostras do sinal avaliado recebe rótulo igual a 1 na presença de assobios e rótulo igual a 2 na presença de cliques ou sinais pulsantes, e zero na ausência de sinais (apenas ruído).

As estatísticas usadas são a precisão P e a acurácia A dadas por

$$P = \frac{VP}{VP+FP} \quad (5)$$

$$A = \frac{VP+VN}{VP+FP+VN+FN} \quad (6)$$

onde VP são os verdadeiros positivos, FP os falsos positivos, VN os verdadeiros negativos, e FN os falsos negativos. Os mesmos parâmetros usados anteriormente para detecção são usados agora. Uma revisão destes valores deve ser feita caso os dados sejam originados de situações diversas e/ou na presença de diferentes espécies.

O caso 1 é mostrado na Figura 5, no qual foram rotulados 4 intervalos de assobios. Os retângulos marcados em vermelho foram aqueles em que assobios foram detectados pela metodologia, sendo as métricas calculadas a cada amostra. Obteve-se uma acurácia de 89% e uma precisão de 97%. A avaliação não pode ser feita para detecção de cliques neste caso, pois VP=0.

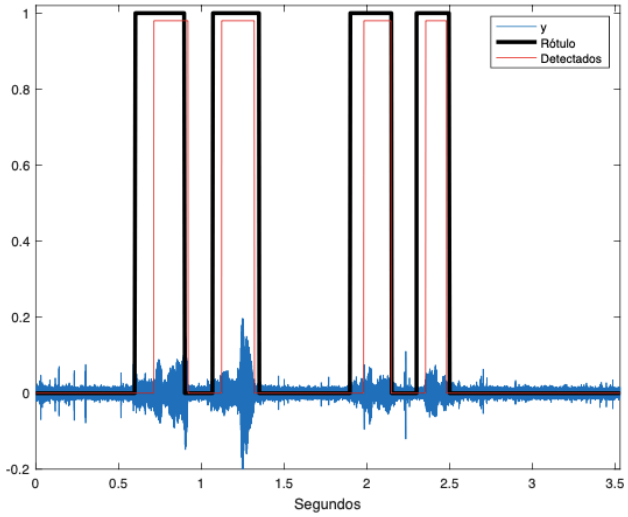

Figura 5. Caso 1 - 4 assobios

Um segundo estudo de caso foi avaliado, contendo dois intervalos de assobios e um intervalo de cliques rotulados. Para a detecção de assobios obteve-se acurácia de 98% e precisão de 99%, claramente visualizados na Figura 6. Para a detecção de cliques o desempenho foi menor, de 86% para acurácia e 81% para precisão (Figura 7).

Quando comparado com métodos da literatura tais como Korkmaz et al. (2022) usando redes neurais convolucionais para detecção de assobios, o desempenho em termos de precisão e acurácia típicos foi superior com uma simplicidade muito maior.

Observa-se que, apesar dos intervalos de assobios e cliques rotulados não coincidirem com os detectados, eles foram detectados todas as vezes que ocorreram. A metodologia foi aplicada a outros conjuntos de dados com bom desempenho, não sendo apresentados devido a limitação de espaço.

A escolha dos parâmetros para as detecções feitas é um procedimento similar ao usado em qualquer algoritmo de aprendizado de máquinas, onde são fornecidos conjuntos de dados de treino rotulados e os algoritmos extraem deles os hiperparâmetros necessários. A vantagem aqui é que a função de cada parâmetro é bem definida, e eles podem ser obtidos de conjuntos de dados rotulados como assobios ou cliques e sons pulsantes. Obviamente, dados coletados de regiões ou espécies diferentes podem requerer um novo ajuste de parâmetros usando novos dados de treino.

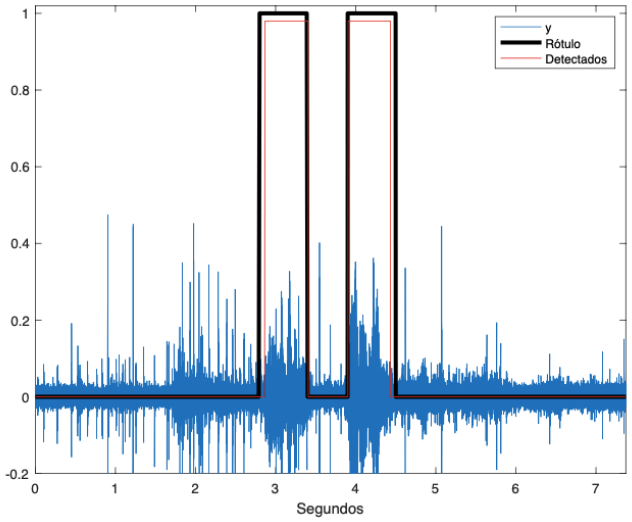

Figura 6. Caso 2 - Assobios

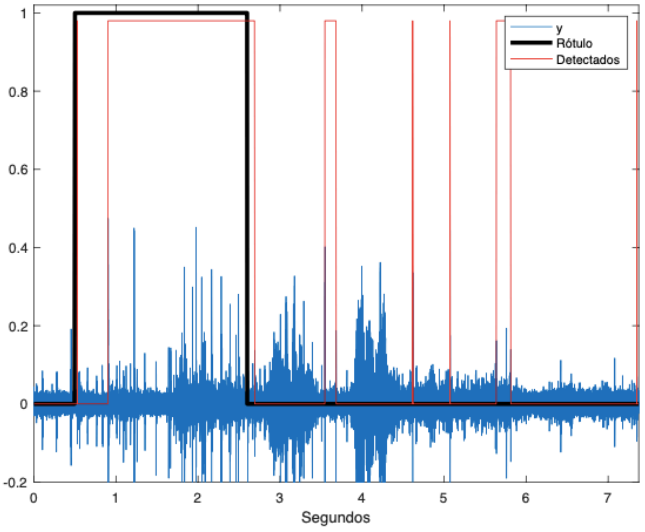

Figura 7. Caso 2 - Clicks

6. CONCLUSÕES

Uma metodologia para a detecção de assobios, cliques e sons pulsantes em sinais de áudio obtidos via monitoramento acústico de delfinídeos foi proposta. São usadas estatísticas de fácil compreensão, e todos os parâmetros utilizados nos algoritmos podem ser obtidos a partir de dados de treinamento rotulados. Os algoritmos classificam os sons detectados em duas classes: assobios ou cliques e sons pulsantes, diferentemente da maioria de métodos que detectam apenas os assobios. No caso de assobios, uma metodologia é proposta para extrair seus contornos. Os

métodos propostos foram ilustrados usando sinais de áudio rotulados por especialistas. Métricas de desempenho foram calculadas e comprovaram resultados animadores, que permitem a aplicação da metodologia em situações gerais, reduzindo muito o esforço de especialistas ao processar grandes quantidades de dados. Como trabalho futuro será verificada a possibilidade de treinar classificadores conhecidos usando os parâmetros deste artigo. Algoritmos de busca poderão fazer sua seleção de forma a otimizar a detecção e classificação das vocalizações dos sinais de áudio.

AGRADECIMENTO

Os autores agradecem à FAPES pelo apoio dado na forma de bolsas (PROFIX , Edital 15/2022 e CPID 2.0).

REFERÊNCIAS

Bazúa-Durán, C., and Au, W.W.L. (2002) The whistles of Hawaiian spinner dolphins. *The Journal of the Acoustical Society of America*, 112 (6), 3064-3072.

Dambros, J. W., Trierweiler, J. O., and Farenzena, M. (2019). Oscillation detection in process industries–Part I: Review of the detection methods. *Journal of Process Control*, 78, 108-123.

Diercks, K. J., Trochta, R.T., and Evans, W. E. (1973) Delphinid sonar: measurement and analysis. *The Journal of the Acoustical Society of America*, 54 (1), 200-204.

Erdogan, K. A. Y. A. (2013). Spline interpolation techniques. *Journal of Technical Science and Technologies*, 47-52.

Esfahanian, M., Zhuang, H., and Erdol, N. (2014) A new approach for classification of dolphin whistles. *Institute of Electrical and Electronics Engineers,* 10.1109/ICASSP.2014.6854763.

Gannier, A., Fuchs, S., Gannier, A., Fernandez, M., and Azevedo, J. M. N. (2020). Dolphin whistle repertoires around São Miguel (Azores): Are you common or spotted? *Applied Acoustics,* 161, 107169.

Herzing, D. L. (1996) Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*. *Aquatic Mammals*, 22 (2), 61-79.

Hoffmann, L. S. (2004) Um estudo de longa duração de um grupo costeiro de golfinhos *Tursiops truncatus* (Montagu, 1821) (Cetacea, Delphinidae) no sul do Brasil: aspectos de sua biologia e bioacústica. 2004. 273 f. Tese (Doutorado em Biologia Animal). Instituto de Biociências, PPGBAN, UFRGS, Porto Alegre.

Janik, V. M. and Slater, P. J. B. (1998) Context-specific use suggests that bottlenose dolphins' signature whistles are cohesion calls. *Animal Behaviour*, 56, 829-838.

Kohlsdorf, D., Herzing, D., and Starner, T. (2016) Feature learning and automatic segmentation for dolphin communication analysis. *Proc. Interspeech 2016*, 2621-2625, 10.21437/Interspeech.2016-748.

Korkmaz, B. N. Diamant, R., Danino, G., and Testolin, A. (2022). Automated detection of dolphin whistles with convolutional networks and transfer learning. *Frontiers in Artificial Intelligence.* arXiv:2211.15406v1 [cs.SD].

Lammers, M. O., Au, W.W., and Herzing, D. L. (2003) The broadband social acoustic signaling behavior of spinner and spotted dolphins. *The Journal of the Acoustical Society of America*, 114 (3), 1629-1639.

Mallawaarachchi, A. "Spectrogram denoising for the automated extraction of dolphin whistle contours." , Master Thesis, Singapore, (2008).

Marten, K. (2000) Ultrasonic analysis of pigmy sperm whale (*Kogia breviceps*) and Hubbs' beaked whale (Mesoplodon carlhubbsi) clicks. *Aquatic Mammals*, 26 (1), 45-48.

Montgomery, D. C., and Runger, G. C. (2010). Applied statistics and probability for engineers. John Wiley & Sons.

Rio, R., Romero-Mujalli, D., Cobarrubia-Russo, S., Nanduca, H. R., Vieites, F. M., Freitas, T. R. O., and Hoffmann, L. S. (2024). Macro-spatial comparison of common bottlenose dolphin whistles in tropical waters. *Marine Mammal Science,* e13122.

Sousa-Lima, R. S., Norris, T. F., Oswald, J. N., and Fernandes, D.P. (2013). A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals. *Aquatic Mammals*, 39(1) 1-9.

Steiner, W. W. (1981) Species-specific differences in pure tonal whistle vocalizations of five western north Atlantic dolphin species. *Behavioral Ecology and Sociobiology*, 9, 241-246.

Winship, K. A., and Jones, B. L. (2023). Acoustic monitoring of professionally managed marine mammals for health and welfare insights. *Animals*, 13(13), 2124.