

CLASSIFICAÇÃO DE SINAIS DE SONAR PASSIVO NO DOMÍNIO DO TEMPO E DA FREQUÊNCIA UTILIZANDO REDES NEURAIS PROFUNDAS

Classification of passive sonar signals in the time and frequency domains using deep neural networks

Marlon Jovenil de Souza¹ , Natanael Nunes de Moura Júnior² , José Manoel de Seixas³ , Caio César Marques Pereira de Alcântara⁴ , Fábio Oliveira Baptista da Silva⁵ , Vinícius Ávila da Silva⁶ 

Resumo: O interesse crescente pelo ambiente submarino aumenta a necessidade de compreendê-lo e explorá-lo através da acústica submarina. Sinais provenientes de embarcações podem ser detectados e classificados por diversos sistemas. Os sistemas de sonar são uns desses sistemas, eles têm sido amplamente utilizados tanto em aplicações militares como civis. Em particular, os sistemas de sonar passivo desempenham um papel importante nas operações submarinas na Marinha de qualquer nação. Neste trabalho, é proposto um classificador baseado em redes neurais profundas, mais especificamente as redes convolucionais e recorrentes. O classificador é alimentado a partir das informações espectrais e das amplitudes dos sinais no domínio do tempo. O modelo proposto tem como objetivo classificar os sinais provenientes de 24 classes de navios militares, que foram organizados em 4 superclasses baseadas em conhecimento especialista. Os resultados de um classificador baseado em redes neurais perceptron multicamadas serão utilizados como *baseline* deste estudo. Entre todos os modelos avaliados, o modelo LSTM alimentados pelos dados no domínio do tempo apresentou o melhor resultado ($85,55\% \pm 3,28\%$) em comparação com os outros modelos.

Palavras-chave: Aprendizado profundo, Rede neural artificial. Sistema de sonar passivo. Treinamento supervisionado.

Abstract: The rise of interest in underwater environment prompts the need to understand and explore it, using underwater acoustics technologies. Underwater signals radiated from vessels can be detected and classified for several systems. An example of this type of system is the sonar system. It is used in both military and civilian applications. In particular, passive sonar systems play an important role in submarine warfare for any navy across the world. In this paper, a classifier based on deep neural networks is proposed, more specifically, a recurrent and convolutional neural network. The input of the classifier is the spectral information and the time series amplitudes of the signals. The proposed model aims to classify the underwater signals radiated by 24 classes of warships that were grouped into 4 superclasses by specialized knowledge. The results obtained from the multilayer perceptron classifier are used as a baseline study. For all models analyzed in this paper, the LSTM model with time series input showed the best result ($85.55 \pm 3.28\%$) in benchmarking with the other models.

Keywords: Passive sonar system. Deep learning. Passive sonar system. Supervised training.

1. Encarregado da Divisão de Processamento de Sinais, Grupo de Sistemas Acústicos Submarinos (GSAS), Instituto de Pesquisas da Marinha, Rio de Janeiro, RJ - Brasil. E-mail: jovenil@marinha.mil.br

2. Professor adjunto do Laboratório de Processamento de Sinais, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (Coppe), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ - Brasil. E-mail: natmourajr@lps.ufrj.br

3. Professor titular do Laboratório de Processamento de Sinais, Coppe-UFRJ, Rio de Janeiro, RJ - Brasil. E-mail: seixas@lps.ufrj.br

4. Ajudante da Divisão de Processamento de Sinais, GSAS, Instituto de Pesquisas da Marinha, Rio de Janeiro, RJ - Brasil. E-mail: caio.alcantara@marinha.mil.br

5. Ajudante da Divisão de Processamento de Sinais, GSAS, Instituto de Pesquisas da Marinha, Rio de Janeiro, RJ - Brasil. E-mail: fabio.oliveira@marinha.mil.br

6. Ajudante da Divisão de Processamento de Sinais, GSAS, Instituto de Pesquisas da Marinha, Rio de Janeiro, RJ - Brasil. E-mail: vinicius.avila@marinha.mil.br

1. INTRODUÇÃO

O ambiente marinho tem sido explorado em diferentes níveis devido a sua inquestionável importância para a economia de qualquer país, por ser a principal rota do comércio exterior, por sua diversidade de recursos naturais, suas reservas de petróleo, entre outros aspectos importantes (VAGTS, 1956). Portanto são necessários navios, submarinos, entre outros equipamentos de patrulha, para a defesa desses recursos. O submarino militar utiliza os sistemas Sound Navigation and Ranging (Sonar) para navegação, comunicação, extração de características ambientais, detecção de outros navios, classificação de ameaças, entre outras tarefas (MOURA JUNIOR; SEIXAS, 2015). O sistema Sonar utiliza as propriedades da propagação do som na água do mar e pode ser dividido em dois principais tipos de tecnologias: passivo, que não emite nenhum sinal e utiliza a emissão acústica do alvo a ser identificado; e ativo, que opera emitindo pulsos específicos e analisando o sinal de eco correspondente para detectar o alvo (MOURA JUNIOR; SEIXAS, 2015). O sistema Sonar é comumente usado em ambientes militares e civis para diferentes aplicações como: sonar de pesca, sonares de eco, exploração mineral e monitoramento de portos e é também um dos principais extratores de características subaquáticas a bordo de submarinos e outros navios (LI, 2012).

A tarefa de classificar os sinais de sonar passivo é desafiadora devido ao grande número de classes a serem tratadas e ao ruído ambiente no meio subaquático (KOMARI ALAIE; FARSI, 2018). Diferentes modelos de aprendizagem de máquinas têm sido aplicados no contexto de sonar passivo (YUE et al., 2017), e as Redes Neurais Artificiais (ANN) têm mostrado bons resultados nesse contexto (SOARES FILHO; SEIXAS; CALOBA, 2001). Entre as técnicas ANN, a abordagem de aprendizado profundo se destaca em trabalhos recentes (NEUPANE; SEOK, 2020). Em Souza Filho e Seixas (2016), foi abordado um esquema de classificação baseado em modelos classe especialista. Em (KOMARI ALAIE; FARSI, 2018), foi usado um novo método para detecção de alvo em sonar passivo usando limiar adaptativo. O sinal de alvo (som) é processado nos domínios de tempo e frequência.

Um modelo de Memória de Curto e Prazo Longo (LSTM) para previsão de tendência de séries temporais foi implementado em Liu et al. (2019), e os resultados superaram os resultados apresentados nos modelos convencionais

de regressão automática. Em Komari Alaie e Farsi (2018), é proposto um novo método para a detecção de alvos usando limiar adaptativo. Este método realiza a fusão das informações no domínio da frequência e do tempo, realizando uma classificação bayesiana. O método melhorou a taxa de detecção cerca de 24% quando comparado a outros métodos de detecção. Em Zhang et al. (2021), uma rede neural integrada para o reconhecimento de alvos subaquáticos através do aprendizado de fusão de características foi proposta. As tecnologias usadas para a extração de características foram *Demon*, *Low Frequency Analysis and Recording (Lofar)* e *Mel Frequency Cepstrum Coefficient (MFCC)*. A rede era composta por três modelos convolucionais, e a fusão era realizada por um algoritmo alimentado pela decisão dos modelos convolucionais. O modelo proposto apresentou maior precisão de reconhecimento em relação a métodos tradicionais.

O presente artigo propõe o desenvolvimento de um classificador responsável por fundir as informações espectrais e temporais do sinal do sistema de sonar passivo, além de explorar o uso das técnicas de aprendizado profundo, especificamente as redes LSTM. Para fins de comparação, um modelo MLP também é projetado usando dados no domínio da frequência. Os dados militares utilizados foram fornecidos pelo Instituto de Pesquisa Marinha (IPQM) e compreenderam 24 classes de sinais, que foram organizadas hierarquicamente utilizando os conhecimentos de militares especialistas na área. Os resultados experimentais do modelo proposto serão avaliados usando a figura de mérito índice de Soma-Produto (SP), especificada na Subseção 4.3.

O restante do trabalho está organizado da seguinte forma: a Seção 2 descreve, de forma básica, o sistema de sonar passivo e sua cadeia de processamento de sinais. A Seção 3 apresenta brevemente as redes neurais recorrentes. Na Seção 4, o método desenvolvido é explicado. Os resultados experimentais são analisados na Seção 5, e as conclusões são descritas na Seção 6.

2. SISTEMA DE SONAR PASSIVO

O sistema de sonar passivo é um conjunto de técnicas que utilizam as propriedades da propagação do som nas águas do mar. Os ruídos acústicos irradiados por diversas fontes, tais como vida marinha, navios, fenômenos sísmicos e vários

outros, são adquiridos por n transdutores chamados hidrofones, que compõem a parte eletrônica inicial do sistema de sonar passivo. Após a aquisição dos sinais, eles são processados por um conformador de feixes, que consiste em uma filtragem espacial dos sinais provenientes dos transdutores para obter a direção de chegada (DoA) (MOURA JUNIOR; SEIXAS, 2015). Os sinais conformados são analisados tanto no domínio do tempo quanto no domínio da frequência com o objetivo de extrair características que ajudem na identificação e classificação do alvo (MALIK et al., 2010; SOUZA FILHO; SEIXAS, 2016). Os modelos de aprendizagem da máquina podem ser alimentados com base nessas características com o intuito de classificar embarcações (SOUZA FILHO; SEIXAS, 2016).

2.1. ANÁLISE NO DOMÍNIO DO TEMPO E DA FREQUÊNCIA

No domínio da frequência, o sistema de sonar passivo utiliza como extração de características uma análise de banda larga que processa o ruído de máquina, chamada de análise Lofar (SOARES FILHO; SEIXAS; CALOBA, 2001). O diagrama de blocos da análise Lofar é mostrado na Figura 1. Os sinais provenientes do DoA são decimados por um fator L , a fim de fazer a seleção da banda de interesse do sistema de sonar passivo. Após a decimação, o sinal processado passa por uma janela *Hanning* para reduzir o efeito de *Aliasing* após a aplicação de uma Transformada de *Fourier* de Tempo Curto (STFT). Após a STFT, o módulo de espectro é normalizado e o sinal passa por uma função TPSW (*Two-Pass Split Window*) para reduzir o ruído de fundo do ambiente subaquático (SOARES FILHO; SEIXAS; CALOBA, 2001; MOURA JUNIOR; SEIXAS, 2015).

Após o processamento pela análise Lofar, o sinal pode ser visto pelo chamado Lofargrama (Figura 2). Nesse gráfico, temos no eixo horizontal os *bins* de frequência e, no eixo vertical, o tempo. Além disso, esse gráfico é, comumente, visto como um *waterfall display*, ou seja, à medida que o tempo avança, mais janelas são processadas, o que faz com que seja atualizado resultando na adição de mais linhas à parte superior do gráfico.

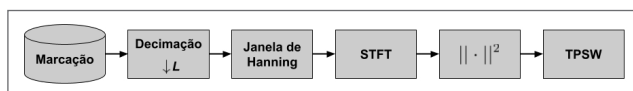


Figura 1. Diagrama de blocos da análise Lofar.

A análise no domínio do tempo deriva da análise Lofar. A Figura 3 mostra o diagrama de blocos dessa análise. As amplitudes do sinal de tempo discreto sofrem uma decimação por um fator L . Após essa etapa, o sinal passa por um filtro passa-baixa para selecionar as faixas de interesse. A passagem através do filtro é extremamente importante, pois os sinais provenientes de navios têm uma quantidade significativa de energia na faixa de baixa frequência (MOURA JUNIOR; SEIXAS, 2015). Por fim, os sinais são normalizados em magnitude, mantendo sua média zero e variância unitária (SOUZA; SEIXAS, 2021).

Após o processamento do sinal pela análise Temporal, é preciso segmentar o sinal para tentar obter um processo mais localmente estacionário possível. Alguns valores foram testados para o tamanho da janela do sinal (1.024 e 2.048). Esses valores são os mesmos utilizados no tamanho da janela da STFT e foram escolhidos por apresentarem bons resultados em alguns trabalhos na literatura (SOUZA FILHO; SEIXAS, 2016; MOURA JUNIOR, 2018; GOLTZ, 2019).

3. ARQUITETURA DAS REDES NEURAS ARTIFICIAIS

As ANN são estruturadas e compostas por unidades de processamento não lineares, neurônios, totalmente conectadas baseadas em redes neurais biológicas. Elas vêm sendo amplamente utilizadas como ferramenta de análise e classificação de sinal de sistema sonar (HAYKIN, 2009). Como mencionado anteriormente, este trabalho abordará três tipos de redes neurais, as MLP (Perceptron Multicamadas), CNN (Rede Neural Convolutiva) e as recorrentes (RNN), mais especificamente as LSTM. As redes CNN e LSTM são redes chamadas de profundas. Aprendizado profundo é o nome dado a uma classe de algoritmos de aprendizado de máquinas que utilizam várias camadas de processamento (GOODFELLOW; BENGIO; COURVILLE, 2016). A ideia da aprendizagem profunda é descobrir múltiplos níveis de representação com a expectativa de que recursos de alto nível representem uma abstração mais detalhada dos dados. Vale ressaltar que as redes serão utilizadas também com dados de séries temporais. De maneira sucinta, uma série temporal é uma realização de um processo estocástico (PEEBLES JR., 2001). Ela difere de um *iid* (independente e identicamente distribuído), uma

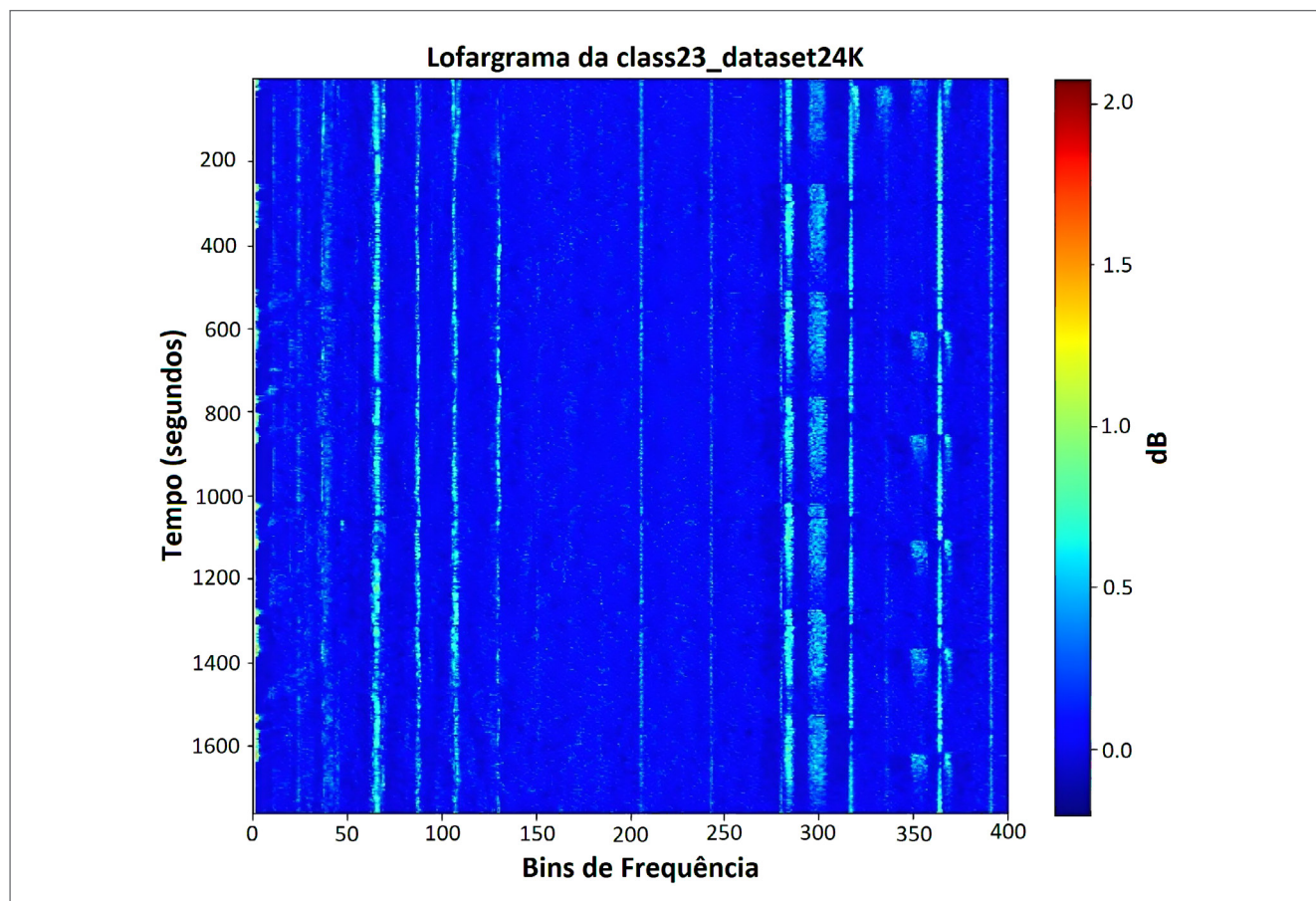


Figura 2. LOFARgrama.

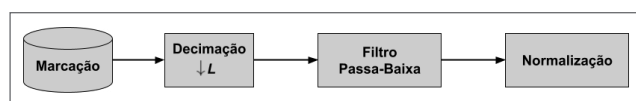


Figura 3. Diagrama de blocos de análise Temporal.

vez que existe uma dependência de seus valores. Os processos estocásticos baseados na premissa de estarem em um determinado estado particular de equilíbrio estatístico são chamados de estacionários (BLOUIN, 2017).

As redes MLP são constituídas de uma camada que recebe os sinais de entrada, uma ou mais camadas intermediárias e uma camada de saída. As camadas intermediárias são amplamente conhecidas como camadas escondidas (HAYKIN, 2009). As MLP são ainda caracterizadas pelas elevadas possibilidades de aplicações em diversos tipos de problemas relacionados com as mais diferentes áreas de conhecimento, sendo consideradas uma das arquiteturas mais versáteis quanto à aplicabilidade (HAYKIN, 2009).

3.1. REDE NEURAL CONVOLUCIONAL

As redes CNN foram criadas originalmente para tarefas envolvendo imagens e dados em duas dimensões. O nome é devido ao fato da rede utilizar operações matemáticas chamadas de convolução, entre os dados da entrada por filtros (chamados de *Kernels*), no lugar de usar multiplicação matricial em, pelo menos, uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016). A CNN pode ser dividida em duas partes: extração de características e uma rede neural tradicional (GOODFELLOW; BENGIO; COURVILLE, 2016). A Figura 4 mostra o processo da rede. Os filtros operam sobre a matriz de entrada. Essa operação ocorre diversas vezes em diferentes regiões dos dados de entrada até que todos os dados sejam abordados. O resultado dessa operação é chamado de *feature map* (mapas de parâmetros). Pode ocorrer também uma operação chamada de *Pooling* após a convolução. Essa operação consiste em uma transformação com a

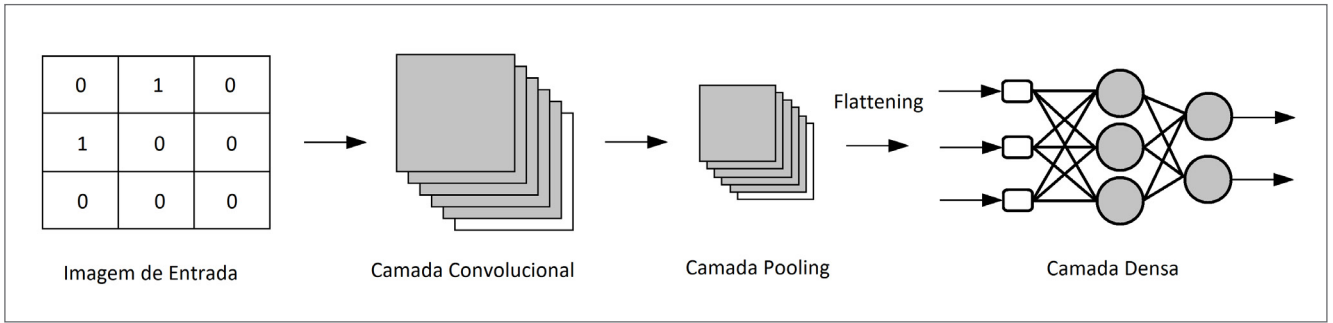


Figura 4. Arquitetura CNN.

redução do tamanho da dimensão dos dados resultantes. Por fim, o *feature map* alimenta uma rede neural totalmente conectada (*fully connected*) para a classificação/previsão dos dados. Na análise de séries temporais, os dados são unidimensionais, portanto pode-se utilizar uma convolução unidimensional (1D) operando com vetores.

3.2. REDES DE MEMÓRIA DE CURTO E PRAZO LONGO

As RNN podem processar sequências de comprimento variável (LIU et al., 2019). Elas têm conexões dentro da mesma camada, mantendo seu estado interno através de nós de contexto (GOODFELLOW; BENGIO; COURVILLE, 2016). De uma maneira geral, podemos considerar a RNN como uma rede com “memória”, que mantém informações sobre o que foi calculado até o momento atual. Um dos problemas dessa rede é aprender as dependências de longo prazo, uma vez que os gradientes propagados por muitas etapas tendem a desaparecer ou explodir (GOODFELLOW; BENGIO; COURVILLE, 2016).

As arquiteturas de redes LSTM são capazes de aprender dependências de longo prazo. Para isso, a LSTM introduz *loops* automáticos para fazer caminhos nos quais os gradientes podem fluir por longas durações (GOODFELLOW; BENGIO; COURVILLE, 2016). Uma célula LSTM tem um estado que funciona como uma esteira transportadora em que algumas operações lineares acontecem (memória de longo prazo). Além disso, uma memória de curto prazo contém a entrada e a saída do instante anterior. O controle das informações da célula é realizado por três portas: *Forget gate*, *Input gate* e *Output gate*. Elas são responsáveis respectivamente por remover informações inúteis do estado da célula, adicionar informações úteis ao estado da célula e extrair

informações úteis do estado atual da célula para apresentar na saída (ZEBIN et al., 2018). A Figura 5 mostra três células LSTM conectadas.

Das Equações 1 a 6, w são os pesos associados a cada *gate*, h são os estados ocultos, σ (Função *Sigmoid*) e \tanh são opções para as funções de ativação.

A *input gate* controla a extensão na qual um novo valor flui para a célula e pode ser escrita como:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (1)$$

A *forget gate* realiza uma operação semelhante:

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (2)$$

A candidata a memória da célula é atualizada da mesma forma:

$$C_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (3)$$

combinando diferentes valores internos, a memória interna de longo prazo ou a memória da próxima célula é gerada como:

$$c_t = f_t \times c_t + i_t \times C_t \quad (4)$$

Com base nisso, a saída da célula é gerada pelo *output gate* para controlar até que ponto o valor na célula é usado para calcular a ativação da saída, fazendo uma operação de leitura.

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (5)$$

Por fim, a saída oculta da célula é:

$$h_t = o_t \times \tanh(c_t) \quad (6)$$

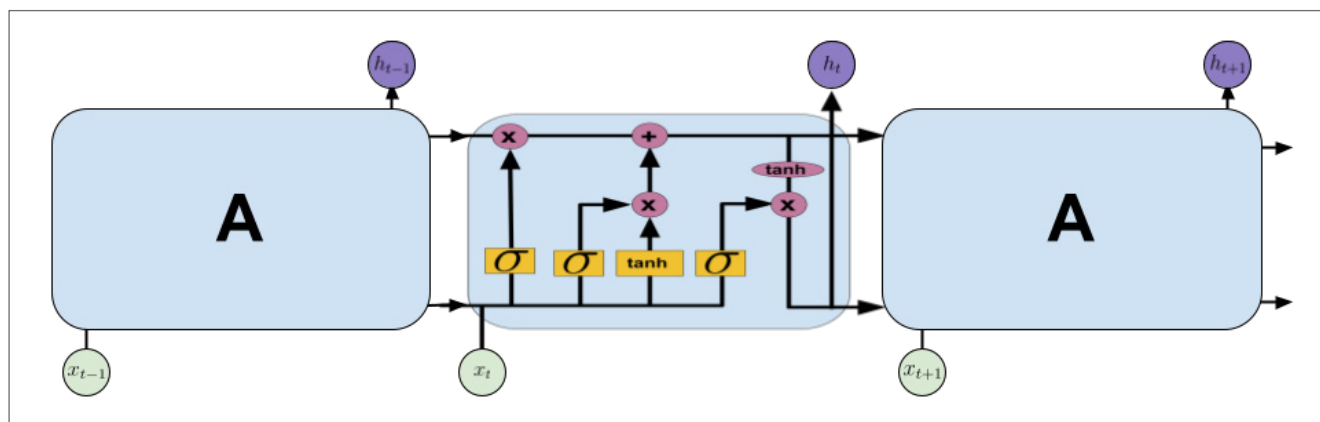


Figura 5. Célula LSTM.

4. MÉTODO PROPOSTO

Esta seção apresenta o método proposto e as ferramentas utilizadas durante o desenvolvimento do sistema de classificação.

Para a realização deste trabalho, foram utilizados três modelos de classificadores (MLP, CNN e LSTM). Os dados processados pela análise Lofar alimentam os modelos MLP e CNN. Já os dados processados pela análise Temporal alimentam o modelo LSTM. Os resultados de classificação de cada modelo são avaliados com o intuito de avaliar a capacidade de classificação dos modelos profundos.

4.1. TREINAMENTO DO MODELO

Após a etapa de processamento do sinal (ver Subseção 2.1), os sinais resultantes da análise Lofar alimentam os modelos MLP e CNN e os sinais resultantes da análise Temporal alimentam o modelo LSTM. Vale ressaltar que a análise Lofar utiliza como entrada o mesmo trecho de sinal utilizado pela análise Temporal. Na análise Lofar, a dimensão de entrada da rede está associada à quantidade de *bins* de frequência utilizada. Na análise Temporal, a dimensão de entrada da rede corresponde ao comprimento da janela de aquisição dos sinais. Na camada intermediária, foram usadas funções de ativação do tipo *ReLU* nos neurônios da rede MLP/CNN e tangente hiperbólica para as células LSTM. Na camada de saída, as funções dos neurônios são do tipo *Softmax*. A função *Softmax* é uma generalização da função *Sigmoid* para casos não binários e é normalmente utilizada na saída de problema de classificação de múltiplas classes, indicando a possibilidade de a instância

pertencer a uma determinada classe (GOODFELLOW; BENGIO; COURVILLE, 2016). As redes também aplicam o algoritmo de retropropagação (*backpropagation*) ao longo da fase de treinamento (DA SILVA; SPATTI; FLAUZINO, 2010).

Com o objetivo de estimar as flutuações estatísticas a partir do conjunto de dados, foi aplicada a validação cruzada (ver Subseção 4.2). Todo o processo de treinamento, processamento de dados e características da camada de entrada e saída das três redes são similares. A única diferença é a quantidade de neurônios/células da camada oculta. Vale ressaltar que, na rede CNN, alguns parâmetros extras devem ser incluídos (*Kernel*, camada *fully connected*).

Em suma, o processo de treinamento ocorre da seguinte forma:

- os sinais acústicos do sistema de sonar passivo são processados, utilizando as análises Lofar e Temporal;
- são implementados e treinados três modelos de redes neurais, MLP, CNN e LSTM;
- os melhores hiperparâmetros para cada modelo são definidos de acordo com os resultados obtidos do índice SP sobre o conjunto de validação durante o processo de treinamento.

4.2. VALIDAÇÃO CRUZADA

O método utilizado é baseado no *jackknife* (SUND, 2016) e consiste em dividir o conjunto de dados aleatoriamente em partes de K do mesmo tamanho (K -folds; $K = 10$, no nosso caso). O processo de treinamento utiliza amostras de partes $K-1$, e o restante é utilizado para avaliar a generalização do processo

de treinamento. Para cada *fold*, o processo de validação é realizado com base em uma das partes do subconjunto ($K-1$) da amostra, de modo que o restante dos dados é usado para adaptar o modelo durante o treinamento. Essa validação também é variada para cada *fold*, a fim de melhor avaliar a generalização do treinamento. No final, a incerteza estimada do processo pode ser feita através do cálculo das figuras de mérito descritas na Subseção 4.3 ao longo dos *fold*s. A Figura 6 mostra os blocos de validação cruzada e suas subdivisões.

4.3. FIGURAS DE MÉRITO

O desempenho dos modelos de classificação é avaliado com as seguintes figuras de mérito clássicas: Acurácia, Precisão e *Recall* (GOODFELLOW; BENGIO; COURVILLE, 2016). Entretanto, para casos de classificação com muitas classes e classes desequilibradas, essas figuras de mérito podem não fornecer informações adequadas sobre a capacidade de discriminação do classificador. Uma alta taxa de acurácia para uma classe majoritária às custas de uma classe minoritária representa um problema na identificação de navios militares. É aconselhável que a classificação seja a mais equilibrada possível. Para resolver esse problema, uma outra figura de mérito chamada de índice SP, descrita na Equação 7, foi utilizada.

Nela, θ_i representa a porcentagem de amostras corretamente classificadas na i -ésima classe. Esse índice foi adotado em trabalhos anteriores, pois fornece informações valiosas sobre o quão equilibrada é a eficiência da classificação para um grande número de classes (GOLTZ, 2019). Na verdade, o índice SP tende a colapsar quando a baixa eficiência é atingida para um número de classes e pode ser usado em valores percentuais.

$$SP = \sqrt{\left(\frac{\sum_{i=1}^N \theta_i}{N}\right) \times \left(\sqrt[N]{\prod \theta_i}\right)} \quad (7)$$

5. RESULTADOS E DISCUSSÃO

O conjunto de dados utilizado neste trabalho é oriundo de gravações de sinais acústicos coletados de sonares passivos de submarinos em operação, contendo 24 classes. O conjunto de dados foi gravado com resolução de 16 bits, mas com diferentes taxas de amostragem (t_s). Sendo 22.050 e 50.000 Hz o valor mais baixo e mais alto de t_s , respectivamente.

A primeira etapa no processamento dos dados foi a realização de uma nova amostragem dos sinais de 24 classes para uma única taxa, a fim de permitir a comparação dos resultados



Figura 6. Processo de validação cruzada.

entre as classes. Todos foram reamostrados para a menor t_s , 22.050 Hz. Os sinais de sonar passivos têm suas informações de interesse concentradas em frequências mais baixas (MOURA JUNIOR; SEIXAS, 2015), portanto uma decimação de baixa ordem foi usada nas análises Lofar e Temporal. O valor escolhido foi $L=3$ como tem sido usado em trabalhos anteriores obtendo bons resultados (MOURA JUNIOR; SEIXAS, 2015). Isto resultou na redução da t_s para 7.350 Hz. O filtro passa-baixa utilizado possui ordem 8 e foi projetado com uma frequência de corte de 2.870 Hz. Outro fator importante durante o processamento dos dados é o tamanho da janela de interesse. Dois valores foram testados, 1.024 e 2.048. Esses valores foram escolhidos pois apresentaram bons resultados em alguns trabalhos na literatura (FERNANDES; JUNIOR; SEIXAS, 2007; MOURA JUNIOR, 2018; GOLTZ, 2019). Ao analisar o índice SP no conjunto de validação, o melhor parâmetro é escolhido. Essa definição será descrita mais adiante neste artigo, após a definição do processo de treinamento.

Este trabalho utiliza o arranjo hierárquico dos dados desenvolvido por Goltz (2019). O primeiro nível é chamado “Super”, no qual todas as classes envolvidas no problema foram reagrupadas em novas classes. Os critérios para agrupar os navios são de acordo com características operacionais comuns. O segundo nível da hierarquia foi composto de novos classificadores criados a partir da saída dos classificadores hierárquicos do nível acima. Esse processo de criação de nível é repetido até que não haja características comuns relevantes entre as classes. Determinado esse conhecimento especializado sobre os dados, o novo reagrupamento do conjunto de dados de 24 classes é: “Super”, A, B, C, D, AA, AB, AC, DA e DB. Este documento abordará apenas o nível “Super” do conjunto de dados. A Tabela 1 mostra o nível abaixo do elemento “Super” e suas respectivas classes agrupadas.

5.1. DEFINIÇÃO DO TREINAMENTO E TOPOLOGIA DO MODELO

Para encontrar o melhor modelo de classificação, os métodos descritos na Seção 4 foram usados para treinamento e avaliação das arquiteturas MLP, CNN e LSTM. Os nós da camada de entrada têm a dimensão correspondente ao tamanho da janela de sinal, e a camada de saída da rede tem quatro neurônios. Na camada intermediária foram realizadas variações no número de células de 25 a 100 com um passo de 25 na camada LSTM. Na camada MLP, a variação foi de 10 a

Tabela 1. Nível abaixo do elemento “Super” e suas respectivas classes agrupadas.

Elemento	Classes
A	9, 10, 13, 14, 16, 23, 1, 2, 22, 21
B	4, 6, 8, 12, 17, 19
C	11, 24
D	5, 7, 15, 3, 18, 20

100 com um passo de 10 neurônios. O modelo CNN variou de uma a duas camadas convolucionais com 8 e 16 filtros, uma camada *fully connected* com variação de 10, 20 e 30 neurônios e os *kernels* foram 3 e 6. A validação cruzada descrita na Subseção 4.2 foi utilizada durante o processo de treinamento dos classificadores. Para isso, o conjunto de dados durante o processo foi dividido em 10 *folds*, ou seja, para cada uma das 10 iterações, 80% do conjunto foi alocado para treinamento, 10% para validação e 10% para testes. O índice SP no conjunto de validação foi utilizado na avaliação de cada classificador e na escolha dos hiperparâmetros.

A função de perda utilizada foi a entropia cruzada, e o método de otimização para calcular a taxa de aprendizagem adaptativa para cada parâmetro foi Adam (GOODFELLOW; BENGIO; COURVILLE, 2016), com taxa de aprendizagem $\eta = 0,001$; e as taxas de decaimento para o primeiro e segundo momentos estimados são $\beta_1 = 0,9$ e $\beta_2 = 0,999$. A fim de evitar *overfitting* nos modelos, foi usado *early stopping* para reduzir o tempo de treinamento. O modelo foi treinado com 200 épocas, mas com um parâmetro de paciência de 30 épocas, ou seja, o erro do conjunto de validação foi monitorado durante o treinamento e, caso o erro não apresente melhora em 30 épocas, o treinamento é interrompido e os melhores parâmetros de treinamento (pesos e *biases*) são restaurados.

5.2. AVALIAÇÃO DOS CLASSIFICADORES

A Tabela 2 mostra os valores do índice SP no conjunto de validação de acordo com a variação dos neurônios na camada oculta do MLP. Os dados foram processados pela análise Lofar. Os dois valores de janela são representados, 1.024 e 2.048. É possível observar uma melhoria no índice SP com o aumento do número de neurônios. Considerando o valor médio e a margem de erro, a janela com 1.024 pontos mostrou melhores resultados.

A variação de parâmetros da arquitetura convolucional gerou 32 modelos distintos. A Tabela 3 apresenta os melhores resultados obtidos do índice SP pela variação de modelos CNN alimentados por dados processados pela análise Lofar. As posições da nomenclatura (16,8,10,3) na coluna modelos significam: a 1ª posição é o número de filtros da primeira camada convolucional; a 2ª é o número de filtros da segunda camada convolucional; a 3ª é número de neurônios da camada densa; e 4ª é a quantidade de *kernel*.

Os resultados do índice SP pela variação da quantidade de células da rede LSTM alimentada por dados processados pela análise Temporal podem ser observados na Tabela 4.

Os resultados do índice SP apresentado nas Tabelas 2 e 4 são bem próximos, sendo assim os diagramas de violino das Figuras 7 e 8 nos ajudam a ter uma melhor compreensão da diferença estatística dos resultados obtidos. Para a rede MLP,

Tabela 2. Resultados do índice SP para o modelo MLP (conjunto de validação).

Neurônios	SP (%) - 1.024	SP (%) - 2.048
10	70,44 ± 2,05	75,33 ± 2,86
20	73,98 ± 2,44	76,84 ± 2,91
30	76,30 ± 2,77	77,99 ± 3,00
40	77,29 ± 2,86	78,57 ± 3,32
50	76,42 ± 2,68	79,11 ± 3,68
60	77,33 ± 2,99	79,15 ± 3,48
70	77,53 ± 2,52	79,90 ± 3,11
80	77,53 ± 2,94	80,20 ± 3,09
90	77,21 ± 2,75	80,64 ± 3,31
100	77,54 ± 2,51	80,70 ± 3,69

Tabela 3. Resultados do índice SP para o modelo CNN (conjunto de validação).

Células	SP (%) - 1.024	SP (%) - 2.048
(16,16,10,3)	78,93 ± 3,47	83,73 ± 3,49
(16,16,10,3)	81,97 ± 3,15	85,36 ± 4,67
(8,8,10,3)	79,17 ± 3,50	83,67 ± 4,11
(8,8,20,3)	80,86 ± 2,90	84,03 ± 4,30
(8,-,10,6)	75,71 ± 2,91	79,56 ± 3,96

é possível observar que, a partir de 30 neurônios, não há uma melhora significativa dos resultados. Ressalta-se a importância de manter a parcimônia em relação a escolha da quantidade de neurônios/parâmetros do modelo, dado que o conjunto de dados tem um total de 50.160 amostras (Classe A: 33.120;

Tabela 4. Resultados do índice SP para o modelo LSTM (conjunto de validação).

Células	SP (%) - 1.024	SP (%) - 2.048
25	79,41 ± 2,59	64,41 ± 5,03
50	82,71 ± 4,59	72,37 ± 9,63
75	85,80 ± 3,51	79,73 ± 6,65
100	86,18 ± 3,04	80,64 ± 5,99

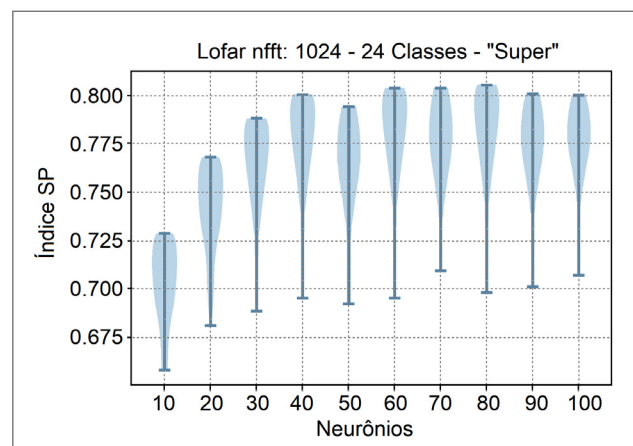


Figura 7. Gráfico de violino do desempenho do modelo MLP.

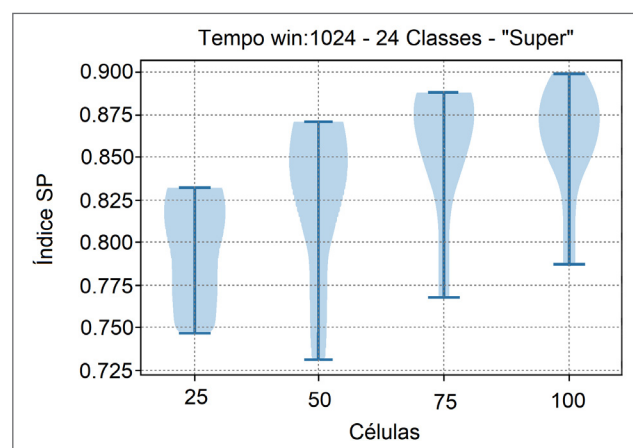


Figura 8. Gráfico de violino do desempenho do modelo LSTM.

Classe B: 5.470; Classe C: 4.260 e Classe D: 7.310) e apenas 80% é destinada ao treinamento. No diagrama de violino da rede LSTM, é possível observar um bom desempenho com o modelo que apresenta 75 células na camada oculta. O modelo Lofar MLP com 30 neurônios na camada oculta e o modelo Temporal LSTM com 75 células na camada oculta são os modelos com os melhores desempenho.

Os resultados do índice SP dos modelos convolucionais da Tabela 3 são próximos. Sendo assim, a quantidade de parâmetros influencia na escolha do modelo. Dado que o *dataset* possui 40.128 amostras destinadas ao treinamento, o modelo escolhido foi o modelo com duas camadas convolucionais com 8 filtros cada, 20 neurônios na camada densa e *kernel* 3, por apresentar 32.336 parâmetros. Essa quantidade é compatível com o número de amostras disponíveis para treinamento. O índice SP desse modelo é de 80,86% \pm 2,90%.

Aplicando os modelos escolhidos ao conjunto de teste obtemos os valores de 76,42 \pm 2,69%, 80,64% \pm 2,81% e 85,55% \pm 3,28% para os modelos MLP, CNN e LSTM respectivamente.

6. CONCLUSÕES E TRABALHOS FUTUROS

O sistema de sonar passivo é importante em muitas áreas da acústica submarina. Neste artigo, exploramos o uso de modelos MLP, CNN e LSTM para classificar sinais processados pela análise Lofar e Temporal.

A rede LSTM com os dados temporais apresentou resultado superior em relação às redes MLP e CNN no domínio da frequência. Esse resultado comprova a eficiência das redes recorrentes no tratamento de dados sequenciais. A rede CNN com os dados espectrais apresentou melhoras no valor médio do índice SP em relação à rede MLP, mas, devido à incerteza estatística, não é possível afirmar que ela foi superior aos resultados apresentados pela rede MLP, pois os valores se sobrepõem. Em Goltz (2019), o índice SP para todo o comitê hierárquico é 73,00% \pm 1,00%. Neste artigo, foi alcançado 85,55% \pm 3,28% com a rede LSTM no nível “Super”, o que indica uma possível melhora nos resultados.

Como trabalho futuro, seria interessante desenvolver um classificador baseado na arquitetura de memória de longo prazo (LSTM) para sinais no domínio de frequência, a fim de comparar os resultados. Seria interessante também utilizar as redes convolucionais bidimensionais com os dados espectrais no intuito de tentar explorar a correlação espacial dos dados. Por fim, em razão dos resultados obtidos, seria relevante aplicar o método proposto nos outros níveis da hierarquia.

AGRADECIMENTOS

Agradeço ao Programa de Engenharia Elétrica da Coppe/UFRJ, a Faperj e ao CNPq, pela oportunidade de realizar esta pesquisa, bem como ao IPqM (Instituto de Pesquisas da Marinha), pela disponibilização da base de dados e estrutura para realização deste trabalho.

REFERÊNCIAS

BLOUIN, S. Simultaneous design of underwater acoustic sensor and communication networks. In: INTERNATIONAL CONFERENCE ON UNDERWATER NETWORKS & SYSTEMS, 2017. *Proceedings...* 2017. p. 1-5.

DA SILVA, I. N.; SPATTI, D. H.; FLAUZINO, R. A. *Redes neurais artificiais para engenharia e ciências aplicadas: curso prático*. São Paulo: Artliber, 2010.

FERNANDES, J. de C. V.; JUNIOR, N. N.; SEIXAS, J. M. de. Análise de Amostras Sintéticas de Sinais de Sonar Passivo Geradas por Redes Neurais Generativas Adversariais. 2007. <https://doi.org/10.21528/CBIC2019-64>

GOLTZ, G. A. M. *Classificação de navios com uso de sinais de sonar passivo*. 2019. Tese (Doutorado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 2019.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. Cambridge: MIT Press, 2016.

HAYKIN, S. *Neural Networks and Learning Machines*. 3. ed. Ontário: Prentice Hall, 2009.

KOMARI ALAIE, H.; FARSI, H. Passive sonar target detection using statistical classifier and adaptive threshold. *Applied Sciences*, v. 8, n. 1, p. 61, 2018. <https://doi.org/10.3390/app8010061>

LI, Q. *Digital sonar design in underwater acoustics: principles and applications*. [S.l.]: Springer Science & Business Media, 2012.

LIU, Y.; SU, Z.; LI, H.; ZHANG, Y. An LSTM based classification method for time series trend forecasting. In: IEEE CONFERENCE ON INDUSTRIAL ELECTRONICS AND APPLICATIONS (ICIEA), 2019, 2019. *Anais...* 2019. p. 402-406.

- MALIK, S. A.; SHAH, M. A.; DAR, A. H.; HAQ, A.; KHAN, A. U.; JAVED, T.; KHAN, S. A. Comparative analysis of primary transmitter detection based spectrum sensing techniques in cognitive radio systems. *Australian Journal of Basic and Applied Sciences*, v. 4, n. 9, p. 4522-4531, 2010.
- MOURA JUNIOR, N. N. de. *Deteção de Novidade para Sistemas de Sonar Passivo*. 2018. Tese (Doutorado em Ciência da Computação) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.
- MOURA JUNIOR, N. N. de; SEIXAS, J. M. de. Novelty detection in passive sonar systems using support vector machines. In: LATIN AMERICA CONGRESS ON COMPUTATIONAL INTELLIGENCE (LA-CCI), 2015., 2015. Anais... 2015. p. 1-6.
- NEUPANE, D.; SEOK, J. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics*, v. 9, n. 11, p. 1972, 2020. <https://doi.org/10.3390/electronics9111972>
- PEEBLES JR., P. Z. *Probability, random variables, and random signal principles*. [S.l.]: McGraw-Hill, 2001.
- SOARES FILHO, W.; SEIXAS, J. M. de; CALOBA, L. P. Principal component analysis for classifying passive sonar signals. In: IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (CAT. NO. 01CH37196), 2001. Anais... 2001. v. 3, p. 592-595.
- SOUZA, M. N. J.; SEIXAS, J. Classificação de Sinais Temporais de Sonar Passivo utilizando Redes Neurais Convolucionais. In: CONGRESSO BRASILEIRO DE INTELIGÊNCIA COMPUTACIONAL, 15., 2021, Joinville, SC. Anais... SBIC, 2021.
- SOUZA FILHO, J. B.; SEIXAS, J. M. de. Class-modular multi-layer perceptron networks for supporting passive sonar signal classification. *IET Radar, Sonar & Navigation*, v. 10, n. 2, p. 311-317, 2016. <https://doi.org/10.1049/iet-rsn.2015.0179>
- SUND, R. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science* Bradley Efron and Trevor Hastie Institute of Mathematical Statistics Monographs. Cambridge: Cambridge University Press, 2016.
- VAGTS, A. *Defense and Diplomacy*. Colúmbia: Columbia University Press, 1956.
- YUE, H.; ZHANG, L.; WANG, D.; WANG, Y.; LU, Z. The classification of underwater acoustic targets based on deep learning methods. *Advanced Intelligence System and Research*, v. 134, p. 526-529, 2017. <https://doi.org/10.2991/caai-17.2017.118>
- ZEBIN, T.; SPERRIN, M.; PEEK, N.; CASSON, A. J. Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. In: IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC), 2018., 2018. Anais... 2018. p. 1-4.
- ZHANG, Q.; DA, L.; ZHANG, Y.; HU, Y. Integrated neural networks based on feature fusion for underwater target recognition. *Applied Acoustics*, v. 182, p. 108261, 2021. <https://doi.org/10.1016/j.apacoust.2021.108261>