SISTEMA DIGITAL PARA LOCALIZAÇÃO DE SONS APLICADO A GRAVAÇÕES DA FAUNA BRASILEIRA

Palavras-Chave: ARRAY DE MICROFONES, BIOACÚSTICA, REDES NEURAIS

Autores:

EDUARDO KOJI TAKESHITA JUNIOR, IA – UNICAMP

Prof. Dr. JOSÉ AUGUSTO MANNIS (orientador), IA - UNICAMP

INTRODUÇÃO:

Há um esforço global para catalogar as espécies e entender o ecossistema, a fim de preservar e restaurar as florestas. Entender o relacionamento entre espécies de pássaros, sua tendência populacional, padrões migratórios, e saúde do ecossistema é essencial para desenvolver planos eficientes de conservação e prevenção de extinção de espécies ameaçadas (Baowaly et al., 2024).

Nessa pesquisa, tivemos como objetivo desenvolver um software capaz de detectar a posição de uma fonte sonora a partir de um arranjo de microfones. Integrando com softwares de reconhecimento de timbres e redes neurais, identificando pontos de interesse em cada cena tomada e registrada. Inferindo dados sobre a fauna e seu ambiente a partir dos dados obtidos e, assim, reconstruir a cena original disponibilizando informações sobre a posição espacial das principais fontes sonoras observadas.

METODOLOGIA:

De forma semelhante ao método aplicado por Cruz (2016) o procedimento para localizar uma fonte sonora foi dividido de três fases: Gravação do som, pré-processamento e extração dos atributos. O código foi construído na linguagem de programação *Python*, com a utilização da biblioteca *TensorFlow* para a implementação de uma rede neural, e a biblioteca *Librosa* para o processamento dos áudios.

Gravação:

Para a gravação, confeccionamos um arranjo hexagonal, com um microfone cardioide em cada vértice. Possuindo raio de 60 centímetros e distancia do chão de um metro e meio. Gravamos arquivos de teste em locais abertos além de sintetizar cenários ideias no computador, através de Digital work Stations (DAWS).

Pré-processamento:

Para realizar o reconhecimento do canto dos pássaros, utilizamos uma rede neural de convolução(CNN), treinado com 2020 cantos de pássaros, totalizando 210 espécies diferentes, e 120

sons de floresta e sons ambiente. As gravações foram coletadas de acervos online, como o iNaturalist, Xeno-canto, Colombian Sound Archive e da Fonoteca Neurotropical Jacques Vielliard (FNJV).

Seguindo o modelo proposto por Silvestre (2022) para otimização do treinamento da máquina, sessionamos os áudios de treino em trechos de três segundos aproximadamente, e convertemos em mel-espectrogramas (figura 1). O espectrograma compreende o movimentos do som em três dimensões, na escala da frequência, pressão sonora e tempo. Cada som possui sua assinatura espectrográfica que revela exatamente o que o caracteriza (Everest, 2001). Silvestre (2022) e Baowaly (2024) mostram que a utilização de mel-espectrogramas no treinamento do modelo, traz melhores resultados do que a utilização dos arquivos de áudio bruto:

> Mel spectrogram images offer a valuable approach for classifying bird species through audio data due to their ability to represent frequency patterns effectively. By breaking down audio signals into distinct frequency components over time, mel-spectrograms serve as a powerful feature extraction technique. (Baowaly et al., 2024)

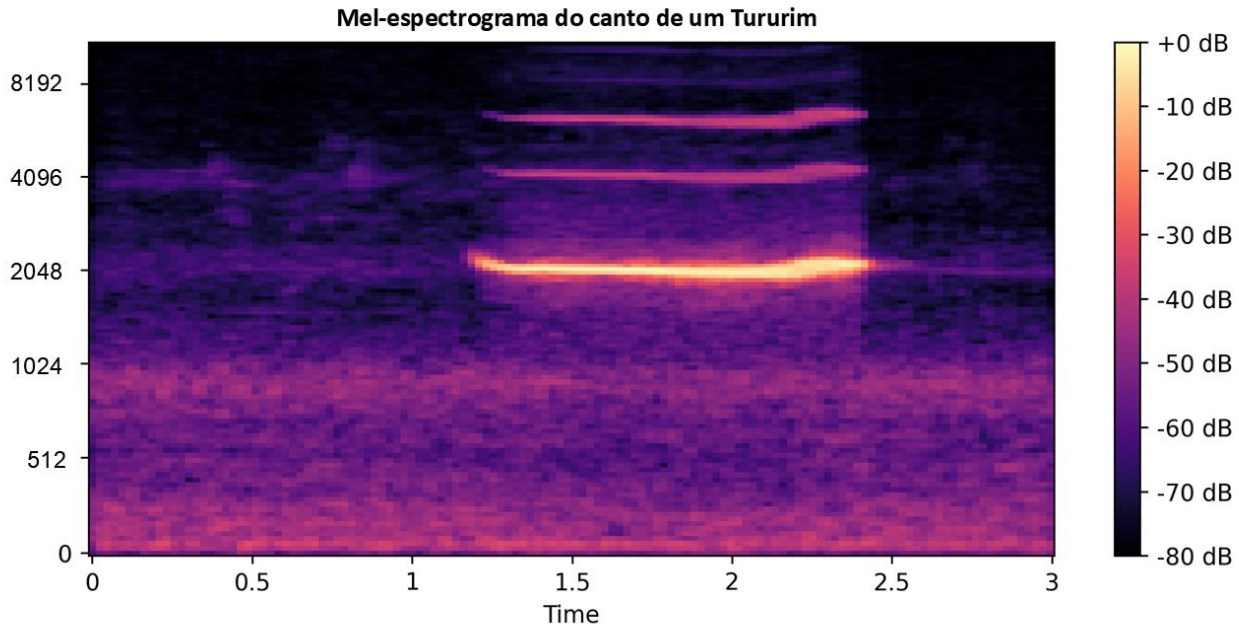

Figura 1-Mel espectrograma Gerado a partir do canto de uma ave da espécie Tururim. Gravação retirada do acervo FNJV, e processada com a biblioteca *Librosa*.

Extração dos dados:

As gravações realizadas com o dispositivo hexagonal são convertidas em Mel-espectrogramas, e escaneado por padrões que se assemelham aos assimilados pelo modelo durante a etapa de treinamento. O arquivo de áudio a ser analisado é segmentado em janelas de meio segundo. Se em uma dessas janelas for detectado o som de pássaro, utilizamos o algoritmo de correlação cruzada para identificar a diferença de fase com que esse som chegou em cada microfone.

Considerando a distância máxima entre dois microfones sendo de um metro e vinte centímetros, e considerando a velocidade do som conforme a ISO 9613-1 à 21ºc, o tempo máximo que um som demora a chegar de um microfone ao outro, é de 0,003491 segundos. Dado a dificuldade de se detectar

o *onset*[1] devido a presença de ruídos nas gravações, e pela necessidade de uma precisão de milissegundos, optamos pelo método de correlação cruzada, utilizando a biblioteca Scikit image, que analisa os padrões do espectrograma de um dos microfones e busca o mesmo padrão nos áudios captados pelos demais microfones, retornando o tempo de atraso com que o som chega em cada capsula.

Dado a direcionalidade dos microfones cardioides, também podemos contar com a diferença de intensidade com que o som chega em cada canal, corroborando os resultados obtidos por TDOA (Time Difference of Arrival).

RESULTADOS:

O vasto repositório na área e as pesquisas no assunto facilitaram o desenvolvimento do nosso programa, bem como uma base extensa de dados e gravações. O uso de redes neurais agiliza o processo, mas a qualidade de seus resultados é proporcional ao número de amostras no processo de treinamento do modelo.

O programa apresentou acuracidade nos testes controlados, com sons planejados digitalmente sem a presença de ruídos de fundo. Mas numa gravação real, considerando as reflexões do som, de outros objetos sonoros e do ruído ambiente, tivemos problemas em precisar os *Onsets* de cada som. A comparação da diferença de fase dos mel-espectrogramas trouxe maior precisão. No último mês de pesquisa, trabalharemos para a otimização do programa, e para averiguar a acuracidade do programa em contextos reais.

BIBLIOGRAFIA

CRUZ, Daylson C.V. **Plataforma para recolha e identificação automática de espécies de pássaros utilizando registros de áudio.** Tese (mestrado em Engenharia Informática) – Especialização em Sistemas Computacionais, ISEP. Porto, 2016.

BAOWALY, Mrinal K. et al. **Deep transfer learning-based bird species classification using mel spectrogram images**. Saudi, Arabia, 2024.

SILVESTRE, Carvalho; GOMES, Elsa F. **Automatic Classification of bird sounds: Using MFCC and Mel Spectrogram Features with deep learning**. Vietam Journal of computer science.

EVEREST, A. **The Master handbook of acoustics**. 4. ed. New York: McGraw Hill, 2001.

VOS, Joos. RASCH, Rudolf. **The perceptual onset of musical tones**. Perception & psychophysics, 29(4):323–335, 1981.

[1] Segundo Rasch e Vos (1981), *onset* pode ser definido como o momento em que o primeiro estimulo sonoro daquele som é percebido