

RESUMO

COSTA, Daniel Ferreira da. *Classificação de Distúrbios em Áudios de Reportagens Bro-adcast com CNNs e Transferência de Aprendizado*. 88 f. Dissertação (Mestrado em Engenharia Eletrônica) - Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2025.

A garantia da qualidade do áudio em transmissões jornalísticas é um desafio recorrente no ambiente *broadcast*, especialmente em captações externas, nas quais interferências, ruídos, ausência de sinal e problemas de *loudness* são frequentes. A detecção manual desses problemas, geralmente realizada por especialistas de áudio, além de demandar esforço especializado, é suscetível a limitações operacionais, subjetividade e falta de escalabilidade. Neste contexto, o presente trabalho propõe uma metodologia baseada em *Deep Learning* para classificar a qualidade de áudio em reportagens externas, utilizando redes neurais convolucionais (CNNs) aplicadas a espectrogramas. Foi produzida uma base de dados com 571 amostras reais de áudio captadas em situações operacionais, classificadas em: áudio adequado (*OK*), ausência de áudio (*NOK_Ausencia*), presença de ruído (*NOK_Ruido*) e problemas de *loudness* (*NOK_Loudness*). A metodologia emprega transferência de aprendizado com a arquitetura VGG16, originalmente treinada no ImageNet, adaptada para interpretar padrões espectro-temporais característicos dos sinais convertidos em espectrogramas. O treinamento utilizou validação cruzada estratificada ($k=5$), assegurando a valiação robusta do desempenho. O modelo alcançou acurácia final de 77,78% na etapa de teste. O desempenho foi robusto nas classes *OK* e *NOK_Ausencia*. Para a classe *NOK_Loudness*, os resultados foram intermediários, satisfatórios, mas sujeitos a confusões ocasionais com *OK*. A principal limitação concentrou-se na classe *NOK_Ruido*, que apresentou a menor taxa de acerto devido à sobreposição espectro-temporal com outras categorias. Esse comportamento evidencia os desafios da detecção de ruídos em condições de captação externa no ambiente *broadcast*.

Palavras-chave: Qualidade de Áudio; Espectrograma; Redes Neurais Convolucionais; VGG16; Transferência de Aprendizado.

ABSTRACT

COSTA, Daniel Ferreira da. *Classification of Audio Disturbances in Broadcast News Reports Using CNNs and Transfer Learning*. 88 f. Dissertação (Mestrado em Engenharia Eletrônica) - Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2025.

Ensuring audio quality in news live broadcasting remains a recurring challenge within the broadcast environment, particularly outdoors, where interference, noise, signal loss, and loudness inconsistencies are frequently encountered. Manual detection of these problems, typically performed by audio engineers, is labor-intensive, subject to operational limitations and human subjectivity, and difficult to scale. In this context, this work proposes a Deep Learning-based methodology for automated audio quality classification in outdoor news gathering, employing convolutional neural networks (CNNs) applied to spectrogram representations of audio signals. A custom dataset was developed with 571 real audio samples captured under operational conditions, labeled into four categories: adequate audio (OK), signal loss (NOK_Ausencia), background noise presence (NOK_Ruido), and loudness inconsistencies (NOK_Loudness). The proposed approach leverages transfer learning from the VGG16 architecture, originally trained on ImageNet, adapting it to capture spectro-temporal patterns of audio spectrograms. Model training used stratified 5-fold cross-validation, ensuring a robust performance assessment. The model achieved a final test accuracy of 77.78%. Performance was robust in the OK and NOK_Ausencia classes, which showed high recognition rates. For NOK_Loudness, the results were intermediate, satisfactory but still subject to occasional confusions with OK. The main limitation was observed in NOK_Ruido, which presented the lowest recognition rate due to spectro-temporal overlap with other categories. This highlights the inherent challenges of noise detection in external audio capture conditions within the broadcast environment.

Keywords: Audio Quality; Spectrogram; Convolutional Neural Networks; VGG16; Transfer Learning.