

# Desenvolvimento de base de dados para treinamento de redes neurais de reconhecimento de voz através da geração de áudios com resposta ao impulso simuladas por técnicas de data augmentation [RESUMO]

Bruno Machado Afonso

18 de abril de 2021

## 1 Introdução

Com o avanço das tecnologias de automação residencial, assistentes pessoais nos smartphones e comunicação online, o estudo de técnicas de processamento de áudio (nesse caso específico, voz), torna-se cada vez mais relevante para a sociedade. Junto a isso, houve grandes avanços no âmbito do aprendizado de máquina, fornecendo alternativas para os métodos tradicionais de processamento de áudio.

Modelos de arquitetura necessitam de um grande volume de dados para que sejam treinados e aprimorados, e um dos mais recentes desafios é de capturar essa extensa quantidade de gravações de áudio, pois é uma tarefa alto custo tanto financeiro e temporal, necessitando de equipamento especializado e diversos locais com sons de fundo e pessoas para amostras de voz.

A proposta dessa tese é de gerar uma base de dados, contendo amostras de voz reverberadas em diversos ambientes, para treinamento de redes neurais, partindo de amostras de voz e resposta ao impulso de salas já existentes. A tese propõe em utilizar técnicas de data augmentation para gerar um grande volume de vozes reverberadas à partir de um pequeno conjunto de amostras.

## 2 Resposta ao Impulso de uma sala (RIR) e técnicas de Data Augmentation

[Falar sobre RIR]

[Falar sobre artigo de DA no RIR]  
[Falar sobre artigo de ruídos em voz]

[1] [2]

## Referências

- [1] N. J. Bryan, “Impulse response data augmentation and deep neural networks for blind room acoustic parameter estimation,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2020.
- [2] T. Ko, V. Peddinti, D. Povey, M. L. Seltzer, and S. Khudanpur, “A study on data augmentation of reverberant speech for robust speech recognition,” in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5220–5224, 2017.