



# **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**

Bruno Machado Afonso

`bruno.ma@poli.ufrj.br`

**Departamento de Engenharia Eletrônica e de Computação - Escola Politécnica**

**Universidade Federal do Rio de Janeiro**

11 de julho de 2021

# Sumário

**1** Motivação

**2** Metodologia

**3** Resultados

**4** Conclusão

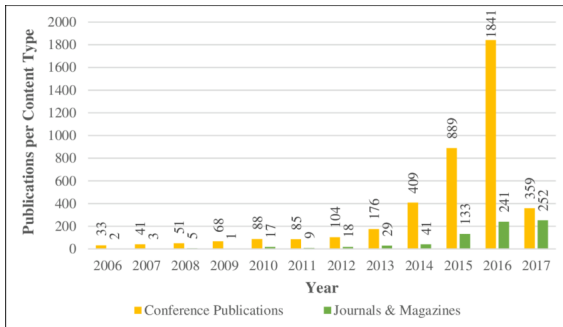
# Motivação

Crescimento no número de aplicações de algoritmos de processamento de áudio.

- Detecção e reconhecimento de voz
  - Smartphones
  - Automação residencial
  - Comunicação online
- Cancelamento de eco
- Separação de fontes

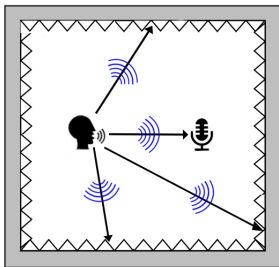
# Deep Learning

Aumento no número de artigos que envolvem *deep learning* publicados em grandes conferências.

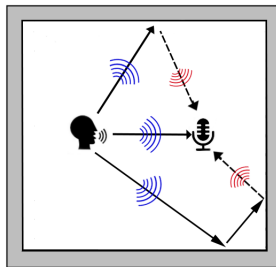


# Amostra de Voz em Campo Distante (AVCD)

Sinal de voz anecóico que é corrompido pela reverberação do ambiente fechado e ruído.



(a) Sala anecóica



(b) Sala reverberante

# Amostra de Voz em Campo Distante (AVCD)

$$Y(t) = s(t) * h(t) + n(t)$$

$Y(t) \rightarrow$  AVCD

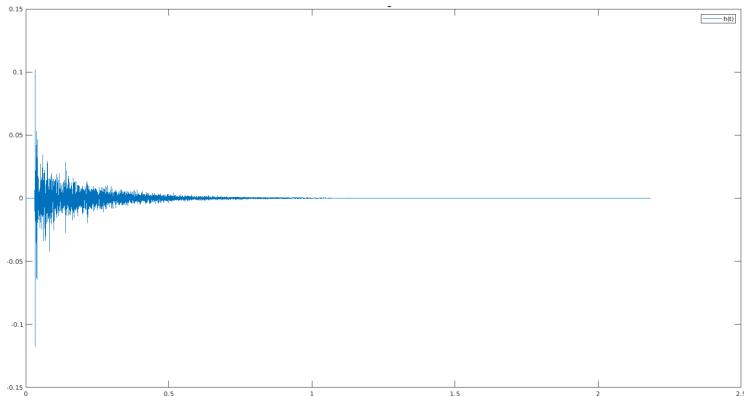
$s(t) \rightarrow$  Amostra de Voz Anecóica

$h(t) \rightarrow$  Resposta ao Impulso de Sala (RIR)

$n(t) \rightarrow$  Sinal de Ruído

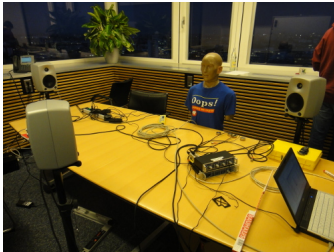
# Resposta ao Impulso de Sala (RIR)

Representa um modelo acústico de um ambiente para um par fonte/-receptor.



# Desafios

- Baixa quantidade e variedade de bases de dados contendo RIRs anotadas para treinamento de redes de *deep learning*.
- Dificuldade para realizar gravações de RIRs (equipamentos especializados, variedade de ambientes, etc.)





# Conclusões

- Em grande parte, os resultados alcançados estão condizentes com os valores esperados.
- Discrepância nos valores de T60 podem ser explicados pelas diferenças de implementação entre este projeto e [1].
- Avaliação empírica das sensações subjetivas de “distância” e “eco” condizentes com as modificações esperadas.

# Trabalhos Futuros

- Implementação de uma metodologia de *data augmentation* de T60 mais próxima à usada no artigo [1].
- Comparação entre as RIRs geradas com a metodologia implementada e RIRs geradas através de programas de simulação acústicas (RAIOS [2]).
- Proposta de um modelo de rede de *deep learning* para estimação de T60 e DRR em AVCDs para observação da eficácia das RIRs como aprimoradoras do treinamento de redes neurais.

# Obrigado!

# Referências

- [1] N. J. Bryan. “Impulse Response Data Augmentation and Deep Neural Networks for Blind Room Acoustic Parameter Estimation”. Em: **ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. 2020, pp. 1–5. DOI: 10.1109/ICASSP40776.2020.9052970.
- [2] Roberto Tenenbaum et al. “Hybrid method for numerical simulation of room acoustics: Part 2-validation of the computational code RAIOS 3”. Em: **Journal of the Brazilian Society of Mechanical Sciences and Engineering** 29 (abr. de 2007). DOI: 10.1590/S1678-58782007000200013.