
Universidade Federal
do Rio de Janeiro
Escola Politécnica

# 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

Projeto de Graduação apresentado ao Curso de Engenharia Eletrônica e de Computação da Escola Politécnica, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Engenheiro.

Orientador: Mariane Rembold Petraglia

Rio de Janeiro

Julho de 2021

# 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

PROJETO DE GRADUAÇÃO SUBMETIDO AO CORPO DOCENTE DO CURSO DE ENGENHARIA ELETRÔNICA E DE COMPUTAÇÃO DA ESCOLA POLITÉCNICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO ELETRÔNICO E DE COMPUTAÇÃO

Autor:

_____

Bruno Machado Afonso

Orientador:

_____

Profª. Mariane Rembold Petraglia, Ph. D.

Examinador:

_____

Prof. A SER DEFINIDO, D. Sc.

Examinador:

_____

Prof. A SER DEFINIDO, D. E.

Rio de Janeiro

Julho de 2021

Declaração de Autoria e de Direitos

Eu, *Bruno Machado Afonso* CPF *136.151.347-02*, autor da monografia *título da monografia*, subscrevo para os devidos fins, as seguintes informações:

1. O autor declara que o trabalho apresentado na disciplina de Projeto de Graduação da Escola Politécnica da UFRJ é de sua autoria, sendo original em forma e conteúdo.

2. Excetuam-se do item 1. eventuais transcrições de texto, figuras, tabelas, conceitos e ideias, que identifiquem claramente a fonte original, explicitando as autorizações obtidas dos respectivos proprietários, quando necessárias.

3. O autor permite que a UFRJ, por um prazo indeterminado, efetue em qualquer mídia de divulgação, a publicação do trabalho acadêmico em sua totalidade, ou em parte. Essa autorização não envolve ônus de qualquer natureza à UFRJ, ou aos seus representantes.

4. O autor pode, excepcionalmente, encaminhar à Comissão de Projeto de Graduação, a não divulgação do material, por um prazo máximo de 01 (um) ano, improrrogável, a contar da data de defesa, desde que o pedido seja justificado, e solicitado antecipadamente, por escrito, à Congregação da Escola Politécnica.

5. O autor declara, ainda, ter a capacidade jurídica para a prática do presente ato, assim como ter conhecimento do teor da presente Declaração, estando ciente das sanções e punições legais, no que tange a cópia parcial, ou total, de obra intelectual, o que se configura como violação do direito autoral previsto no Código Penal Brasileiro no art.184 e art.299, bem como na Lei 9.610.

6. O autor é o único responsável pelo conteúdo apresentado nos trabalhos acadêmicos publicados, não cabendo à UFRJ, aos seus representantes, ou ao(s) orientador(es), qualquer responsabilização/ indenização nesse sentido.

7. Por ser verdade, firmo a presente declaração.

_____

Bruno Machado Afonso

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Escola Politécnica - Departamento de Eletrônica e de Computação
Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária
Rio de Janeiro - RJ CEP 21949-900

## AGRADECIMENTO

Sempre haverá. Se não estiver inspirado, aqui está uma sugestão: dedico este trabalho ao povo brasileiro que contribuiu de forma significativa à minha formação e estada nesta Universidade. Este projeto é uma pequena forma de retribuir o investimento e confiança em mim depositados.

# RESUMO

Inserir o resumo do seu trabalho aqui. O objetivo é apresentar ao pretenso leitor do seu Projeto Final uma descrição genérica do seu trabalho. Você também deve tentar despertar no leitor o interesse pelo conteúdo deste documento.

Palavras-Chave: trabalho, resumo, interesse, projeto final.

## ABSTRACT

Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here.

Key-words: word, word, word.

## SIGLAS

RIR - Resposta ao Impulso de Ambiente Acústico

UFRJ - Universidade Federal do Rio de Janeiro

# Sumário

**1 Introdução** **1**

1.1 Tema . . . . . 1

1.2 Delimitação . . . . . 1

1.3 Justificativa . . . . . 1

1.4 Objetivos . . . . . 2

1.5 Metodologia . . . . . 3

1.6 Descrição . . . . . 4

**2 Reconhecimento de Voz e seus Desafios** **5**

2.1 Histórico da Pesquisa de Reconhecimento de Voz . . . . . 5

2.2 Desafios do Reconhecimento de Voz em Campo Aberto . . . . . 5

**3 Data Augmentation da Resposta ao Impulso do Ambiente** **6**

3.1 Razão Direct-to-Reverberant (DRR) . . . . . 6

3.2 Tempo de Reverberação (T60) . . . . . 6

3.3 Comparação entre RIR real e simulada . . . . . 6

**4 Desenvolvimento de Sinais de Voz Reverberadas Simuladas com Ruídos** **7**

4.1 Simulação de fala em campo distante . . . . . 7

**5 Resultados Experimentais** **8**

5.1 Configuração dos parâmetros . . . . . 8

5.2 Resultados . . . . . 8

**6 Conclusões** **9**

**Bibliografia** **10**

**A O que é um apêndice?** **11**

**B Encadernação do Projeto de Graduação** **12**

**C O que é um anexo?** **14**

# Lista de Figuras

B.1 Encadernação do projeto de graduação. . . . . 13

# Lista de Tabelas

# Capítulo 1

# Introdução

Neste capítulo, será introduzido os principais tópicos do projeto, além de mostrar sua relevância para o escopo da engenharia moderna e as metodologias que são usadas para alcançar seus objetivos. Ao final é descrito a estrutura organizacional do texto.

## 1.1 Tema

O tema do trabalho é sobre o estudo de uma forma de simular Respostas ao Impulso de Ambientes Acústicos (RIR) com parametrizações diferentes a partir de amostras de RIR gravadas em ambientes reais, e ainda usar a RIR para gerar amostras de áudio em locais simulados a partir de gravações de voz reais.

## 1.2 Delimitação

O estudo é focado em inferir uma técnica de reforço de dados tanto em amostras reais de RIR quanto nas gravações de voz. Este trabalho está delimitado em apenas modificar amostras reais de áudio, e não gerar amostras simuladas sem uma gravação de base.

## 1.3 Justificativa

Com o avanço das tecnologias de automação residencial, assistentes pessoais nos smartphones e comunicação online, o estudo de técnicas de processamento de

áudio (no caso específico deste trabalho, relacionados a voz), tornou-se mais relevante para a sociedade. Uma das características mais importantes a ser detectada no processamento de áudio é a Resposta ao Impulso do ambiente, que representa o modelo acústico do ambiente, pois através desta é possível extrair informações pertinentes do local em que o áudio foi gravado e também detectar a posição de fontes sonoras e as isolar para reconhecimento. No âmbito da área de reconhecimento de voz, a fala reverberante, ou seja, o sinal de fala combinado com o modelo acústico do ambiente é um dos desafios encontrados para a detecção da voz, tornando a identificação do RIR de vital importância para o reconhecimento de fala [1].

Junto a isso, houve avanços no âmbito do aprendizado de máquina, fornecendo alternativas para os métodos tradicionais de processamento de áudio [2]. Modelos de arquitetura de redes neurais necessitam de um grande volume de dados para que sejam treinados e aprimorados, e um dos mais recentes desafios nessa área é o fato das bases de RIR não serem extensas, conforme esclarecidas no artigo [3], pois capturar essa extensa quantidade de gravações de áudio é uma tarefa alto custo tanto financeiro e temporal, necessitando de equipamento especializado e diversos locais com características de modelo sonoro diferentes e pessoas diversas para amostras de voz.

## 1.4 Objetivos

O objetivo deste trabalho é desenvolver um algoritmo capaz de gerar amostras de RIR simuladas para diferentes ambientes a partir de uma RIR real e gerar um banco de dados de amostras de voz convoluídas com as RIR simuladas e com ruídos para uso em treinamento de redes neurais. Dessa forma, têm-se como objetivos específicos:

1. Propor um algoritmo que altere as características da RIR para simular diferentes ambientes com RIR diferentes.

2. Elaborar um algoritmo que faça o acréscimo de ruídos pontuais ou ruídos de fundo em uma amostra de voz.

3. Desenvolver um sistema computacional que aplique ambos os algoritmos an-

teriores em sequência para gerar amostras de voz em Ambientes ruidosos.

## 1.5 Metodologia

Um sinal de voz gravado em um ambiente pode ser interpretado como a junção de três partes; uma amostra de voz pura, sem nenhum fator externo ou reverberação envolvido, convoluída com a Resposta ao Impulso da sala (RIR) onde ocorre a gravação, somada a um sinal de ruído, podendo este ser pontual ou um ruído de ambiente. A RIR representa um modelo acústico do ambiente, que define como um receptor acústico irá receber caso o áudio seja gerado e percebido de dentro deste ambiente. Uma definição de Resposta ao Impulso é a de uma função que registra a pressão sonora temporalmente em um ambiente fechado após uma excitação extremamente curta e cheia de energia (dirac).

Neste trabalho é proposto uma forma de gerar RIR simuladas partindo de uma RIR real, ou seja, gravando um áudio que representa um impulso em um ambiente fechado real, e alterando suas propriedades. Reproduz-se o que foi proposto no artigo de data augmentation para respostas ao impulso para estimação do modelo acústico [4], onde é gerado RIR simuladas modificando as propriedades de Tempo de Decaimento (T60) e de razão entre áudio direto e reverberado (DRR). Através dessas duas propriedades, defini-se praticamente todos os RIR possíveis de serem gravados artificialmente.

Para gerar as amostras de vozes reverberadas que compõe a base de dados, acompanha-se o que é proposto no artigo de estudo de data augmentation em vozes reverberadas [5], onde são convoluídos sinais de voz puros com os RIR simulados que foram gerados anteriormente. Além disso, é acrescentado a essa sinal de voz reverberado ruídos diversos, que são caracterizados de duas formas: ruídos pontuais e de ambiente. Os ruídos pontuais são amostras de áudio curta que podem ser introduzidos em qualquer momento da fala, já os ruídos de ambiente são sons constantes ao fundo da gravação para simular um ambiente externo. Os ruídos foram extraídos da biblioteca MUSAN [6].

Através desses dois passos, são gerados vários sinais de vozes reverberados artificialmente. A simulação do RIR tem por objetivo colocar a amostra de voz

em vários ambientes fechados, e já os ruídos ajudam drasticamente no treinamento de redes neurais impedindo que as redes fiquem viciadas em características muito específicas da fala durante o treinamento, pois eles tendem a simular os fatores externos que podem estar envolvidos em uma gravação real.

## 1.6 Descrição

O capítulo 2 apresenta uma breve história sobre as principais aplicações do tema e os desafios que este trabalho auxilia na solução.

No capítulo 3 será descrito a metodologia usada para fazer a *data augmentation* de uma RIR já existente.

No capítulo 4 explica-se a metodologia usada para gerar sinais de voz aleatórios a partir de RIRs simuladas anteriormente e da adição de ruídos pontuais ou de fundo.

O capítulo 5 é focado em exibir os resultados obtidos através dos métodos anteriores e demonstrar sua eficácia.

Por fim, o capítulo 6 trata das conclusões que são tiradas sobre este projeto, além de mostrar trabalhos futuros.

# Capítulo 2

# Reconhecimento de Voz e seus Desafios

## 2.1 Histórico da Pesquisa de Reconhecimento de Voz

## 2.2 Desafios do Reconhecimento de Voz em Campo Aberto

# Capítulo 3

# Data Augmentation da Resposta ao Impulso do Ambiente

## 3.1 Razão Direct-to-Reverberant (DRR)

## 3.2 Tempo de Reverberação (T60)

## 3.3 Comparação entre RIR real e simulada

# Capítulo 4

# Desenvolvimento de Sinais de Voz Reverberadas Simuladas com Ruídos

## 4.1 Simulação de fala em campo distante

# Capítulo 5

# Resultados Experimentais

## 5.1 Configuração dos parâmetros

## 5.2 Resultados

# Capítulo 6

# Conclusões

Tratam-se das considerações finais do trabalho, mostrando que os objetivos foram cumpridos e enfatizando as descobertas feitas durante o projeto. Em geral reserva-se um ou dois parágrafos para sugerir trabalhos futuros.

Observe que neste modelo a conclusão é numerada pelo numeral 3, mas o projeto não tem a obrigatoriedade de possuir apenas 3 capítulos. Alias, espera-se que tenha mais que isso.

# Referências Bibliográficas


[1] HAEB-UMBACH, R., HEYMANN, J., DRUDE, L., *et al.*, "Far-Field Automatic Speech Recognition", *Proceedings of the IEEE*, v. 109, n. 2, pp. 124–148, 2021.

[2] MOKGONYANE, T. B., SEFARA, T. J., MODIPA, T. I., *et al.*, "Automatic Speaker Recognition System based on Machine Learning Algorithms". In: *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, pp. 141–146, 2019.

[3] XIONG, F., GOETZE, S., MEYER, B., "Joint Estimation of Reverberation Time and Direct-To-Reverberation Ratio from Speech Using Auditory-Inspired Features". In: *ACE Challenge Workshop, satellite event of IEEE-WASPAA*, 2015.

[4] Bryan, N. J., "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.

[5] Ko, T., Peddinti, V., Povey, D., *et al.*, "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.

[6] SNYDER, D., CHEN, G., POVEY, D., "MUSAN: A Music, Speech, and Noise Corpus", 2015, http://www.openslr.org/17/, visitado última vez em 07/06/2021.

# Apêndice A

# O que é um apêndice?

Elemento que consiste em um texto ou documento elaborado pelo autor, com o intuito de complementar sua argumentação, sem prejuízo do trabalho. São identificados por letras maiúsculas consecutivas e pelos respectivos títulos.

# Apêndice B

# Encadernação do Projeto de Graduação

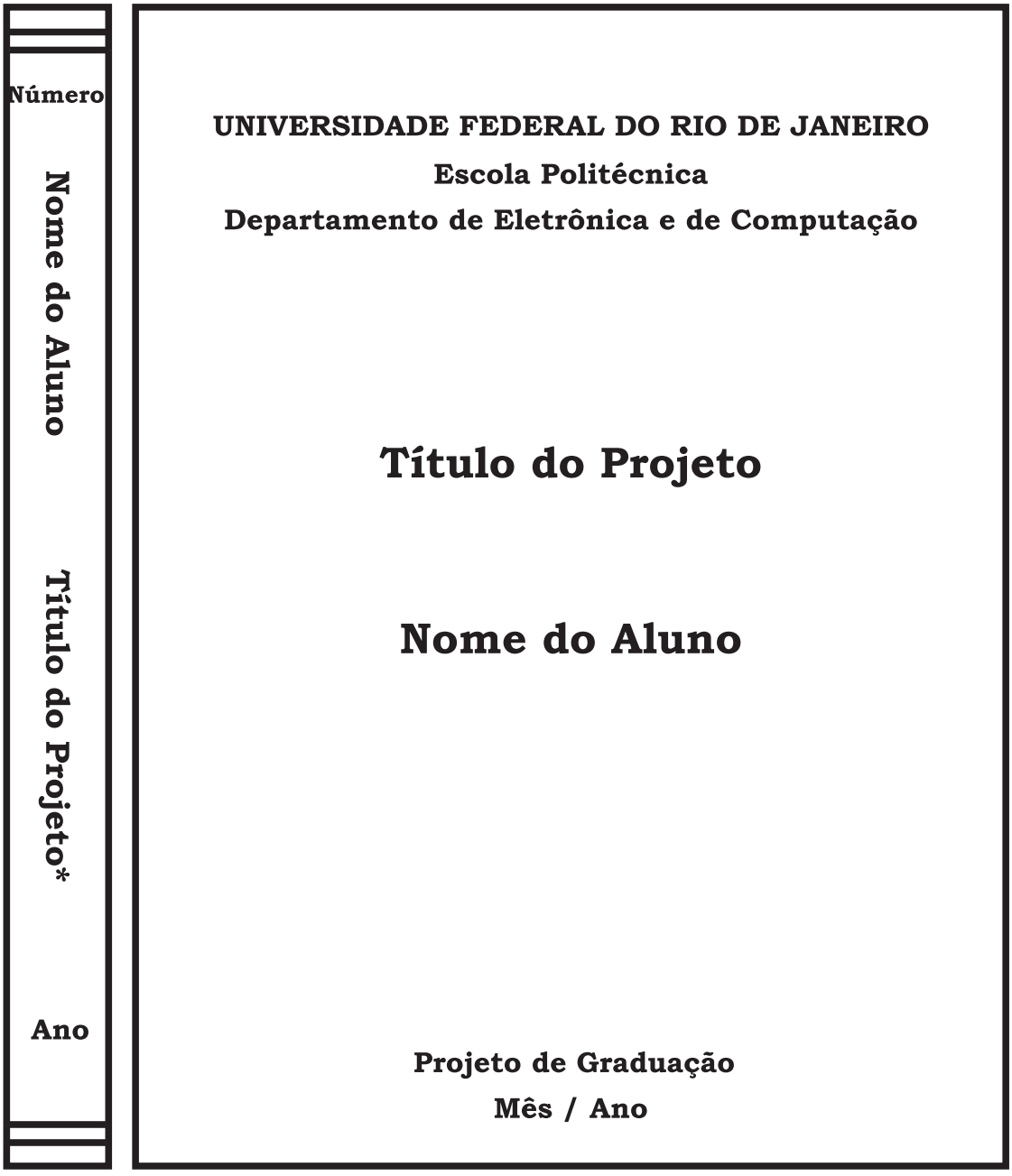


Figura B.1: Encadernação do projeto de graduação.

# Apêndice C

# O que é um anexo?

Documentação não elaborada pelo autor, ou elaborada pelo autor mas constituindo parte de outro projeto.