CENTRO UNIVERSITÁRIO DA FEI

ARIEL GRAÇA FERREIRA

ESTIMATIVA DE FAIXA ETÁRIA PELO PROCESSAMENTO DO SINAL DE VOZ UTILIZANDO FEEP LEARNING

SÃO BERNARDO DO CAMPO - SP

2023

ARIEL GRAÇA FERREIRA

ESTIMATIVA DE FAIXA ETÁRIA PELO PROCESSAMENTO DO SINAL DE VOZ UTILIZANDO FEEP LEARNING

Dissertação de Mestrado, apresentada ao Centro Universitário da FEI para obtenção do título de Mestre em Engenharia Elétrica. Orientado pelo Prof. Dr. Carlos Eduardo Tomaz.

SÃO BERNARDO DO CAMPO

2023

**Ficha Catalográfica**

**Folha de aprovação**

**Dedicatória**

Texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória texto de dedicatória.

**AGRADECIMENTOS**

Texto de agradecimentos....

“Faça as coisas o mais simples possível, mas não  
mais simples.”

Albert Einstein

“Inteligência é a capacidade de se adaptar à mu-  
dança.”

Stephen Hawking

**RESUMO**

No decorrer dos últimos anos tem sido possível acompanhar a expansão e popularização do conceito de Internet das Coisas, onde os mais diversos objetos do cotidiano estão sendo interconectados. De forma similar, o progresso na área de Inteligência Artificial tem proporcionado uma interação cada vez maior entre as pessoas e esses objetos, formando assim uma grande rede que tende a crescer, e se desenvolver, cada vez mais. Visando o avanço tecnológico dentro desse contexto, é imprescindível que se busque uma interação homem-máquina mais refinada, e um dos principais meios para estabelecer tal interação é a voz humana. O presente trabalho propõe estudar os sistemas de estimação automática de faixa etária pela voz com Redes Neurais Convolucionais (RNC's), porém não é proposta apenas a busca pela compreensão e melhorias na arquitetura desse tipo de modelo, mas de se fazer também um contraponto com modelos mais tradicionais, que utilizem outras técnicas, como por exemplo Gaussian Mixture Models (GMM). Nos ensaios iniciais realizados com uma RNC, o modelo apresentou precisão de 60\%, aproximadamente, e a partir desses resultados deve-se desenvolver o estudo proposto.

**ABSTRACT**

**LISTA DE ILUSTRAÇÕES**

|  |  |
| --- | --- |
| Figura 1 – Título do desenho | 21 |
| Figura 2 – Título da gravura | 22 |
| Figura 3 – Título do fluxograma | 23 |
| Figura 4 – Título da fotografia | 35 |
| Figura 5 – Título do gráfico | 42 |
| Figura 6 – Título do mapa | 48 |
| Figura 7 – Título do organograma | 57 |
| Figura 8 – Título da planta | 62 |
| Figura 9 – Título do quadro | 64 |

**LISTA DE TABELAS**

|  |  |
| --- | --- |
| Tabela 1 – Título da tabela | 24 |

**LISTA DE SÍMBOLOS**

|  |  |
| --- | --- |
| Símbolo 1 – Título do símbolo | 23 |

**LISTA DE ABREVITURAS E SIGLAS**

ABNT – Associação Brasileira de Normas Técnicas.

**SUMÁRIO**

[**1 INTRODUÇÃO** 14](#_Toc125210919)

[**2 REFERENCIAL TEÓRICO** 15](#_Toc125210920)

[**3 METODOLOGIA** 17](#_Toc125210921)

[**4 RESULTADOS E DISCUSSÃO** 18](#_Toc125210922)

[**5 CONSIDERAÇÕES FINAIS** 19](#_Toc125210923)

[**REFERÊNCIAS** 20](#_Toc125210924)

[**APÊNDICES** (se houver) 21](#_Toc125210925)

[**ANEXOS** (se houver) 22](#_Toc125210926)

# **1 INTRODUÇÃO**

Com o advento das tecnologias baseadas em Inteligência Artificial (IA), aliado a evolução científica e produção acadêmica das últimas décadas, é notável o estímulo que existe na sociedade a fim de criar e aperfeiçoar a interação homem-máquina, revolucionando assim a forma como os indivíduos realizam atividades, sejam elas cotidianas, corriqueiras, ou com propósitos mais complexos.

A estimação de faixa etária pela voz, possui diversas aplicações que visam criar uma interface homem-máquina para utilização em áreas como transporte, segurança, medicina e automação residencial.

Um exemplo bastante presente no cotidiano das pessoas atualmente, seriam as Inteligências Artificiais (IA) como a *Alexa*, desenvolvida pela Amazon, e o *Google Assistant*, criado pelo Google.

São dispositivos espalhados em milhares de casas ao redor do mundo, e que recebem os mais diversos comandos de voz para execução de várias tarefas, que podem ser desde tocar uma música, até realizar uma complexa rotina de acionamento em conjunto com outros elementos e/ou equipamentos inteligentes, interconectados e espalhados pela casa de um indivíduo, ou fora dela. Para citar apenas uma aplicação, a estimação de faixa etária poderia funcionar como uma validação de segurança, evitando assim acesso ou acionamento indevido por alguma criança que tenha acesso ao dispositivo, mas não possa ter acesso a todas as funcionalidades disponíveis.

Tomando como base esse cenário de crescimento tecnológico e expansão das aplicações de  *Internet of Things* (IoT) e IA, este trabalho tem por finalidade estudar os sistemas de estimação de faixa etária por sinais de voz que utilizem técnicas de aprendizado profundo, como Redes Neurais Convolucionais (RNC). Busca-se entender e mostrar evidências do desempenho desse tipo de sistema, considerando a utilização de tais técnicas.

Através do sinal da voz de um indivíduo, é possível obter muitas informações sobre quem produz o som, como gênero, emoções e idade, porém essa tarefa também traz desafios. A incidência de distorções na fonte do sinal, geradas pelo próprio indivíduo (estado de saúde, emoção), ou geradas pelo ambiente (ruído externo), deformam o sinal de tal forma que a etapa de extração das características acústicas *features* fica prejudicada e, consequentemente, a etapa seguinte, a classificação do dado, também vai sofrer com a degradação.

Nesse sentido, as técnicas de aprendizado de máquina, mais especificamente aprendizado profundo, com base em pesquisas realizadas na literatura atual, mostram-se mais robustas para implementação desse tipo de sistemas que trabalham com sinais acústicos. Por essa razão, este trabalho busca estudar e avaliar a utilização de RNC's no problema de estimação automática de idade pela voz.

Para efeitos de comparação e posterior análise dos resultados, será utilizado como referência a pesquisa FAPESP desenvolvida pelo Prof. Dr. Ivandro, que utiliza técnicas mais tradicionais de aprendizado de máquina, como o *Gaussian Mixture Models* (GMM) e *i-Vector*. Visto que essa pesquisa faz uso do mesmo banco de dados que será utilizado no desenvolvimento do presente trabalho com redes neurais, haverá oportunidade de comparar e discutir os resultados obtidos com ambos os sistemas.

Em relação ao banco de dados, será adotado um corpus com sinais de chamadas telefônicas que pertence a *Deutsche Telekom Laboratories*, de Berlim na Alemanha.

# **2 REFERENCIAL TEÓRICO**

Texto de revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão.

**Exemplo de Citação Direta:**

Fávero (1983 apud CABRAL, 1998, p. 40) assim definiu cultura:

Cultura é tudo o que o homem acrescenta à natureza; tudo o que não está inscrito no determinismo da natureza e que aí é incluído pela ação humana. Distinguem-se na cultura seus produtos: instrumentos, linguagem, ciência, a vida em sociedade; e os modos de agir e pensar comuns a uma determinada sociedade, que tornam possíveis a essa sociedade a criação da cultura.

**Exemplo de Ilustração:**

Figura 1 – Título do gráfico

Fonte: Elaboração própria, 2019.

Texto de revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão texto revisão.

# **3 METODOLOGIA**

## 3.1 Pesquisas Anteriores como base de referência para esse trabalho.

O foco desse trabalho está na análise dos resultados obtidos com um sistema de estimação de faixa etária utilizando técnicas modernas de aprendizagem de máquina, e então comparando tais dados com os que podem ser obtidos através de um sistema que possui a mesma finalidade, mas que faz uso de ferramentas estatísticas que já são utilizadas há muito tempo, ou seja, que já foram exaustivamente testadas e tiveram sua capacidade comprovada.

Para isso, toma-se como base de referência para este trabalho, a pesquisa realizada pelo prof. Ivandro Sanches (**REF**), na qual ele utiliza o modelo estatístico de mistura de gaussianas (Gaussian Mixture Model), em conjunto com a técnica de i-vectors, estimação de pitch, e o algoritmo de Viterbi para estimação da faixa etária de um indivíduo a partir do processamento da sua voz. Durante os trabalhos, foram implementados métodos para extração das características acústicas e classificação dos dados analisados.

Tal pesquisa é composta por dois relatórios científicos. No primeiro (REF), são descritos os métodos para extração das características acústicas do sinal, a base de dados utilizada, a implementação do modelo, o sistema de métricas para avaliação de desempenho de classificação, e por fim os resultados obtidos. Já no segundo relatório, propõe-se a adição de uma nova característica biométrica aos vetores que alimentam o sistema e representam os sinais de fala. Trata-se do *pitch*, que é definido como um parâmetro psicoacústico

ue podem ser encontradas ao se utilizar técnicas nos resultados oriundos de um sistema de estimação processamento do sinal de voz de ferramentas matemáticas e tecnologias em uma aplicação dExiste um trabalho de pesquisa liderado pelo prof. Dr. Ivandro Sanches, dois Relatórios Científicos que compõem uma pesquisa realizada pelo Prof. Dr. Ivandro Sanches em \citeonline{IvandroRel1, IvandroRel2}, foram utilizados como ponto de partida para este trabalho. Ele investiga e faz a análise das capacidades de estimação de faixa etária de um certo indivíduo utilizando o modelo estatístico Gaussian Mixture Model (GMM) e a técnica de i-vector. O modelo é alimentado por vetores que representam cada sinal e são compostos por coeficientes mel-cepstrais (MFCC) obtidos a partir da decomposição da fala. Tais vetores podem ser representados da seguinte forma:

\[ X = \{X\_{1},X\_{2},...,X\_{T}\} \]

# **4 RESULTADOS E DISCUSSÃO**

Texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão.

Texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão texto dos resultados e discussão.

# **5 CONSIDERAÇÕES FINAIS**

Texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais.

Texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais texto das considerações finais.

# **REFERÊNCIAS**

FRIGOTTO, Gaudêncio (Org.). **A cidadania negada**: políticas de exclusão na educação e no trabalho. 5. ed. São Paulo: Cortez, 2011. p. 35-48.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

BACCAN, Nivaldo et al. **Química analítica quantitativa elementar**. 3. ed. rev. ampl. e reest. São Paulo: Edgard Blücher, 2001.

BEGON, Michael; TOWNSEND, Colin R.; HARPER, John L. **Ecologia**: de indivíduos a ecossistemas. Tradução de Adriano Sanches Melo et al. 4. ed. Porto Alegre: Artmed, 2007.

BEGON, Michael; TOWNSEND, Colin R.; HARPER, John L. **Ecology**: individuals, populations and communities. 3nd. ed. Massachusetts: Blackwell, 1996.

BELLUSCI, Silvia Meirelles. **Doenças profissionais ou do trabalho**. 11. ed. São Paulo: Senac, 1996.

CARRAHER, Terezinha Nunes (Org.). **Aprender pensando.** Petrópolis: Vozes, 1989.

CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UFPE, 21., 2013, Recife. **Anais eletrônicos**... Recife: UFPE, 2013. Disponível em: <http://www.ufpe.br/conic/images/stories/anais/2013/ANAIS.html>. Acesso em: 2 set. 2014.

MINTZBERG, Henry; AHLSTRAND, Bruce; LAMPEL, Joseph. **Safári de estratégia**: um roteiro pela selva do planejamento estratégico. 2. ed. Porto Alegre: Bookman, 2010.

# **APÊNDICES**

# **ANEXOS**