Sistema de Identificação de Locutores Baseado em Texto

Thais Rodrigues G. Neiva[1,2] | Álvaro C. Lemos Neto[1] | Arlindo Follador Neto[1,4] | Hani Camille Yehia[1,3]
[1]thaisrgneiva@gmail.com | alvarolemos@gmail.com | arlindo.neto@ict.ufvjm.edu.br | hani@cefala.org
[2]Bolsista PIBIC/CNPq | [3]Professor orientador do Departamento de Engenharia Eletrônica/UFMG
[4]Co-orientador e doutorando do PPGEE/UFMG | Professor do Instituto de Ciência e Tecnologia/UFVJM
Escola de Engenharia – Universidade Federal de Minas Gerais (UFMG)

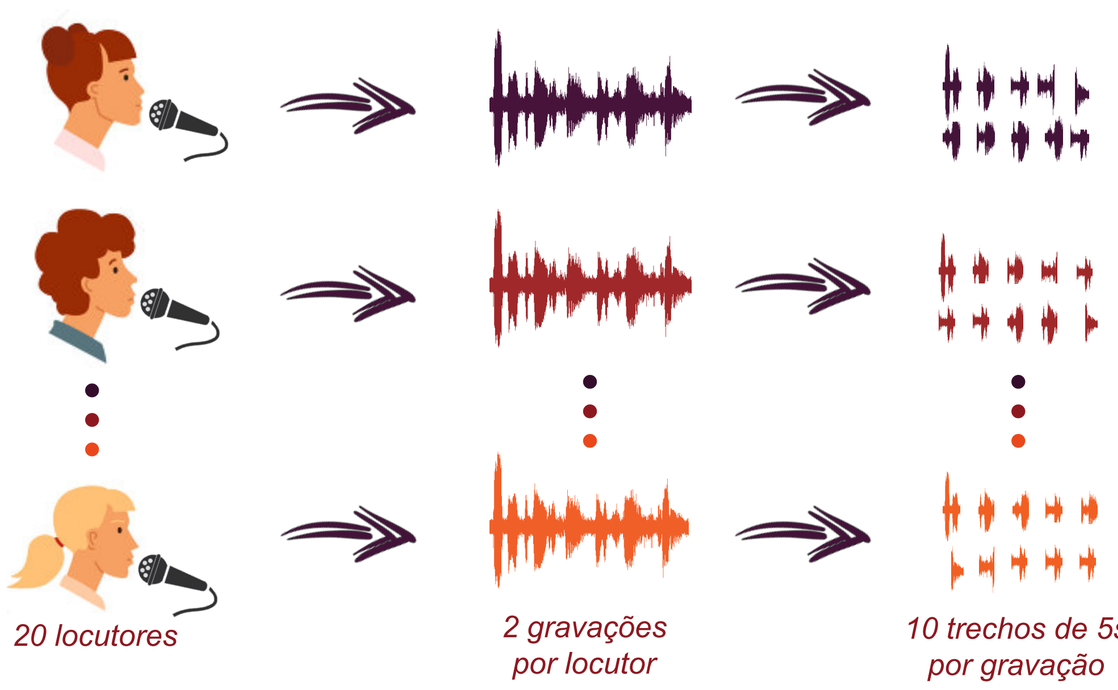
INTRODUÇÃO E OBJETIVOS

O Processamento da Fala é uma área do conhecimento que possui diversas aplicações. Uma delas consiste na extração automática de informações do sinal da fala relacionadas ao locutor com o objetivo de identificá-lo.

O objetivo desse projeto é de criar um Sistema de Identificação de Locutores Baseado em Texto, isto é, espera-se que todos os registros de fala processados pelo sistema sejam referentes a uma mesma frase pré-determinada.

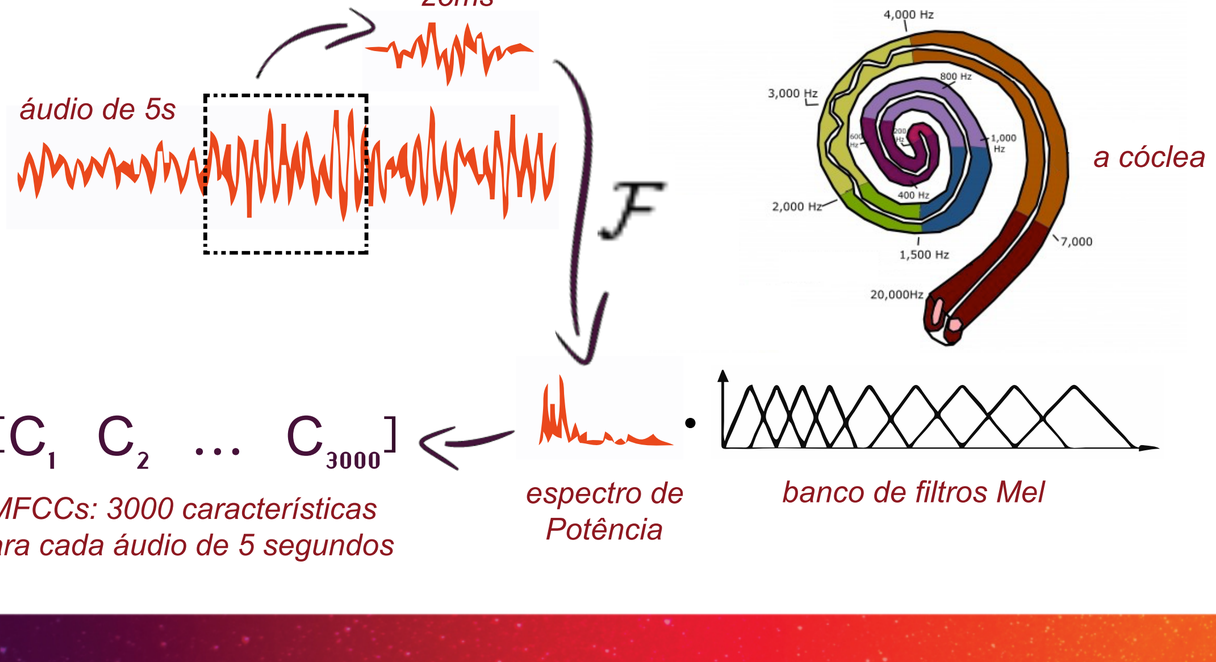
METODOLOGIA

Foram realizadas gravações das vozes de 20 locutores. Cada um deles gravou a leitura de dois textos: um pré-fixado e um não pré-fixado. Depois disso, cada gravação foi separada em 10 trechos de 5 segundos cada.

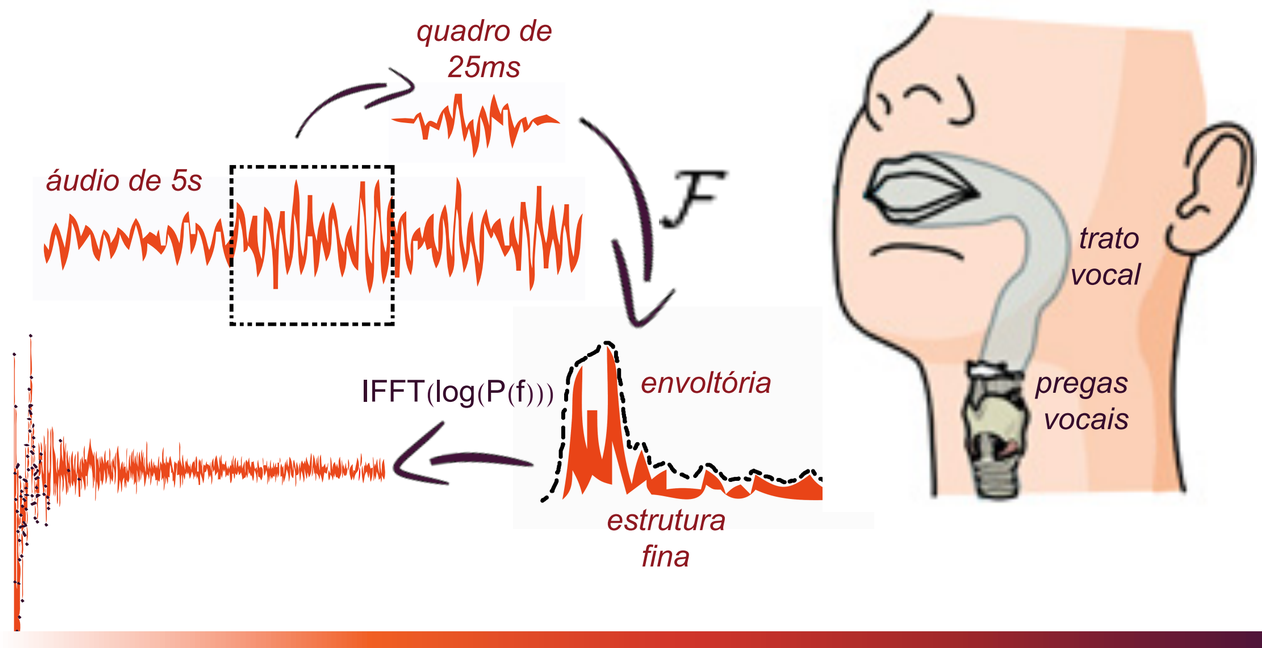


De posse das gravações, a etapa seguinte foi a de extrair de cada uma delas as características que identificam os seus respectivos locutores. Para isso, foram utilizados dois métodos:

CEPSTRUM NA ESCALA MEL DE FREQUÊNCIAS

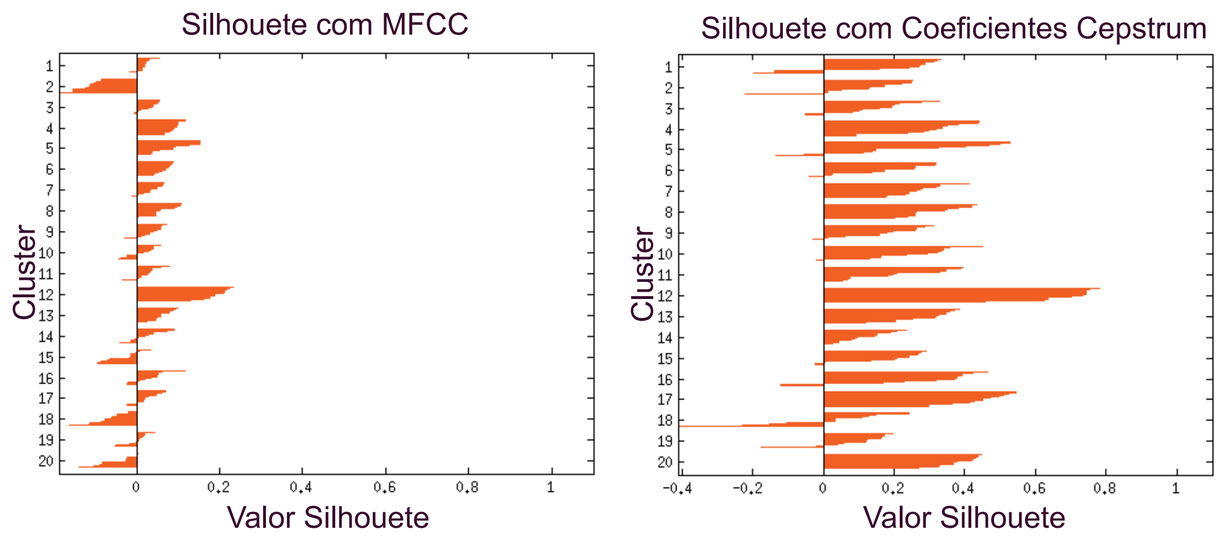


ANÁLISE CEPSTRAL



RESULTADOS

Foi utilizada a interpretação Silhouete para avaliar quão bem as características extraídas de cada gravação identificam o seu locutor. Os resultados podem ser vistos nos gráficos abaixo:



CONCLUSÕES

A análise cepstral, tanto na escala linear, quanto na escala mel de frequências, forneceu um índice de acertos superior a 80%. Os valores de Silhouete negativos observados são minoria, confirmando a eficácia das características utilizadas para a tarefa de identificação de locutor.

Os trabalhos futuros visarão aumentar a precisão do sistema. Para isso, criaremos uma metodologia de gravação mais rigorosa, revisaremos os códigos das ferramentas de extração de características e utilizaremos outras ferramentas novas, como o LPCC (*Linear Prediction Cepstral Coefficients)*.

REFERÊNCIAS BIBLIOGRÁFICAS

X. Huang et al. **Spoken Language Processing: A Guide to Theory, Algorithm and System Development**. 1st ed, 2001.

L. Rabiner, R. Schafer. **Digital Processing of Speech Signals**, 1978.

R. O. Duda et al. **Pattern Classification and Scene Analysis: Pattern Classification**, 1999.

Rousseeuw, P. J. **Finding Groups in Data: An Introduction to Cluster Analysis**, 1990.