

Trabalho 1 - EG940

Fernanda Esteves
(fernandaestevescc@gmail.com)
Mariane Martins
(marianemartins1997@gmail.com)






Informação buscada e público alvo




Identificador de pássaros por meio de suas vocalizações







Literatura científica (artigos)



- **Large-scale identification of birds in audio recordings,**
Mario Lasseck
https://www.tierstimmenarchiv.de/RefSys/LifeCLEF2014/LifeCLEF2014_WorkingNotes.pdf
- **Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge, Dan Stowell; Michael D. Wood; Hanna Pamula; Yannis Stylianou; Hervé Glotin**
<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13103>
- **Audio signal classification using Linear Predictive Coding and Random Forests - IEEE Conference Publication, Lacrimioara Grama; Corneliu Rusu**
<https://ieeexplore.ieee.org/document/7990431>
- **Audio Classification of Bird Species: A Statistical Manifold Approach - IEEE Conference Publication, Forrest Briggs; Raviv Raich; Xiaoli Z. Fern**
<https://ieeexplore.ieee.org/document/5360230>



Estratégias para reconhecer e classificar o som selecionado





**Experimento a ser
realizado**



Base de dados

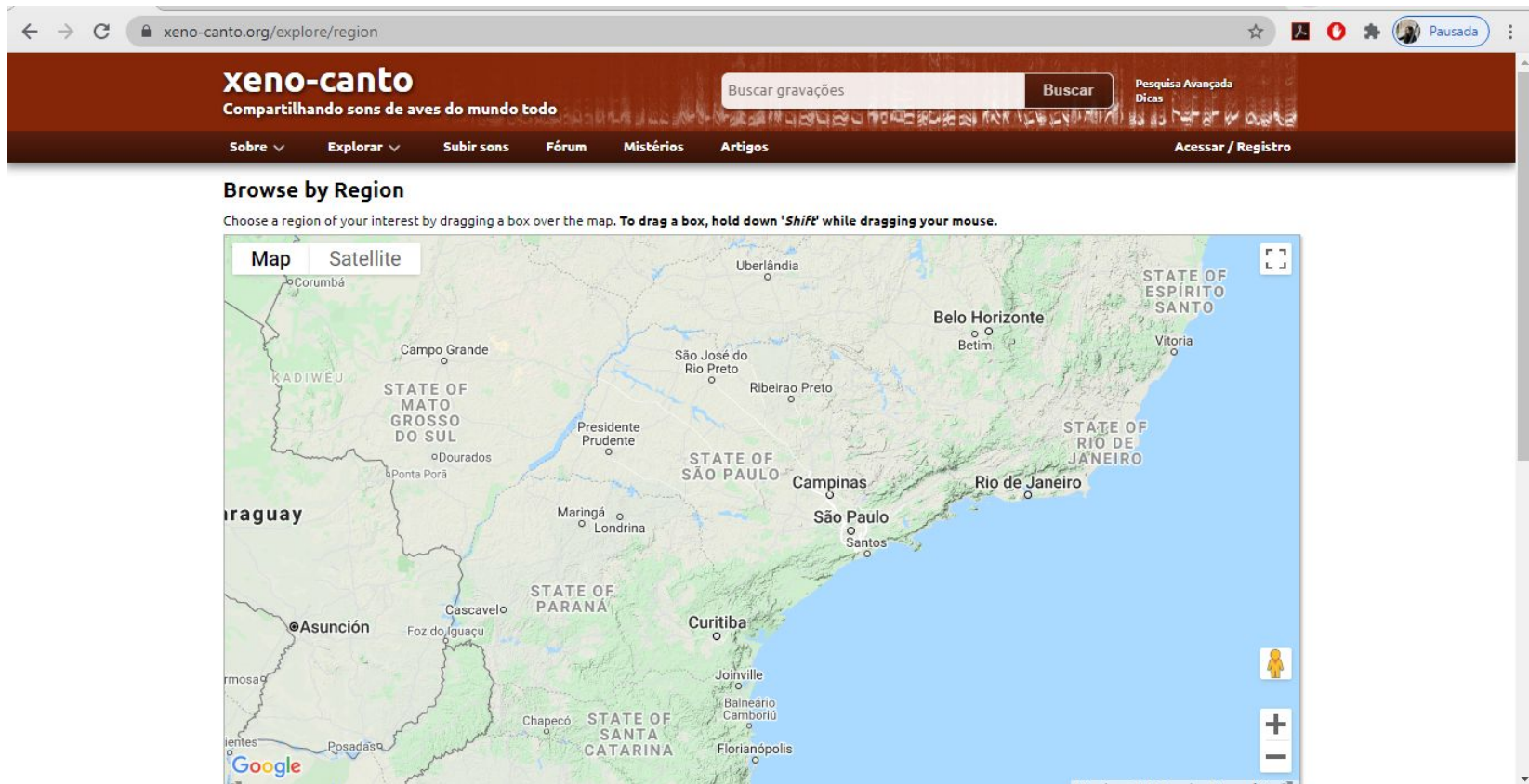


Imagem retirada do site: [Browse by Region :: xeno-canto](https://xeno-canto.org/explore/region)

Banco de sons: [Sons](#)

Features espectrais e rítmicas

Feature extraction

Spectral features

<code>chroma_stft</code> <code>([y, sr, S, norm, n_fft, ...])</code>	Compute a chromagram from a waveform or power spectrogram.
<code>chroma_cqt</code> <code>([y, sr, C, hop_length, fmin, ...])</code>	Constant-Q chromagram
<code>chroma_cens</code> <code>([y, sr, C, hop_length, fmin, ...])</code>	Computes the chroma variant "Chroma Energy Normalized" (CENS)
<code>melspectrogram</code> <code>([y, sr, S, n_fft, ...])</code>	Compute a mel-scaled spectrogram.
<code>mfcc</code> <code>([y, sr, S, n_mfcc, dct_type, norm, lifter])</code>	Mel-frequency cepstral coefficients (MFCCs)
<code>rms</code> <code>([y, S, frame_length, hop_length, ...])</code>	Compute root-mean-square (RMS) value for each frame, either from the audio samples <code>y</code>
<code>spectral_centroid</code> <code>([y, sr, S, n_fft, ...])</code>	Compute the spectral centroid.

imagem retirada da documentação do librosa: <https://librosa.org/doc/main/feature.html>



Features estatísticas

Averages and variances

<code>median(a[, axis, out, overwrite_input, keepdims])</code>	Compute the median along the specified axis.
<code>average(a[, axis, weights, returned])</code>	Compute the weighted average along the specified axis.
<code>mean(a[, axis, dtype, out, keepdims])</code>	Compute the arithmetic mean along the specified axis.
<code>std(a[, axis, dtype, out, ddof, keepdims])</code>	Compute the standard deviation along the specified axis.
<code>var(a[, axis, dtype, out, ddof, keepdims])</code>	Compute the variance along the specified axis.

imagem retirada da documentação do numpy:

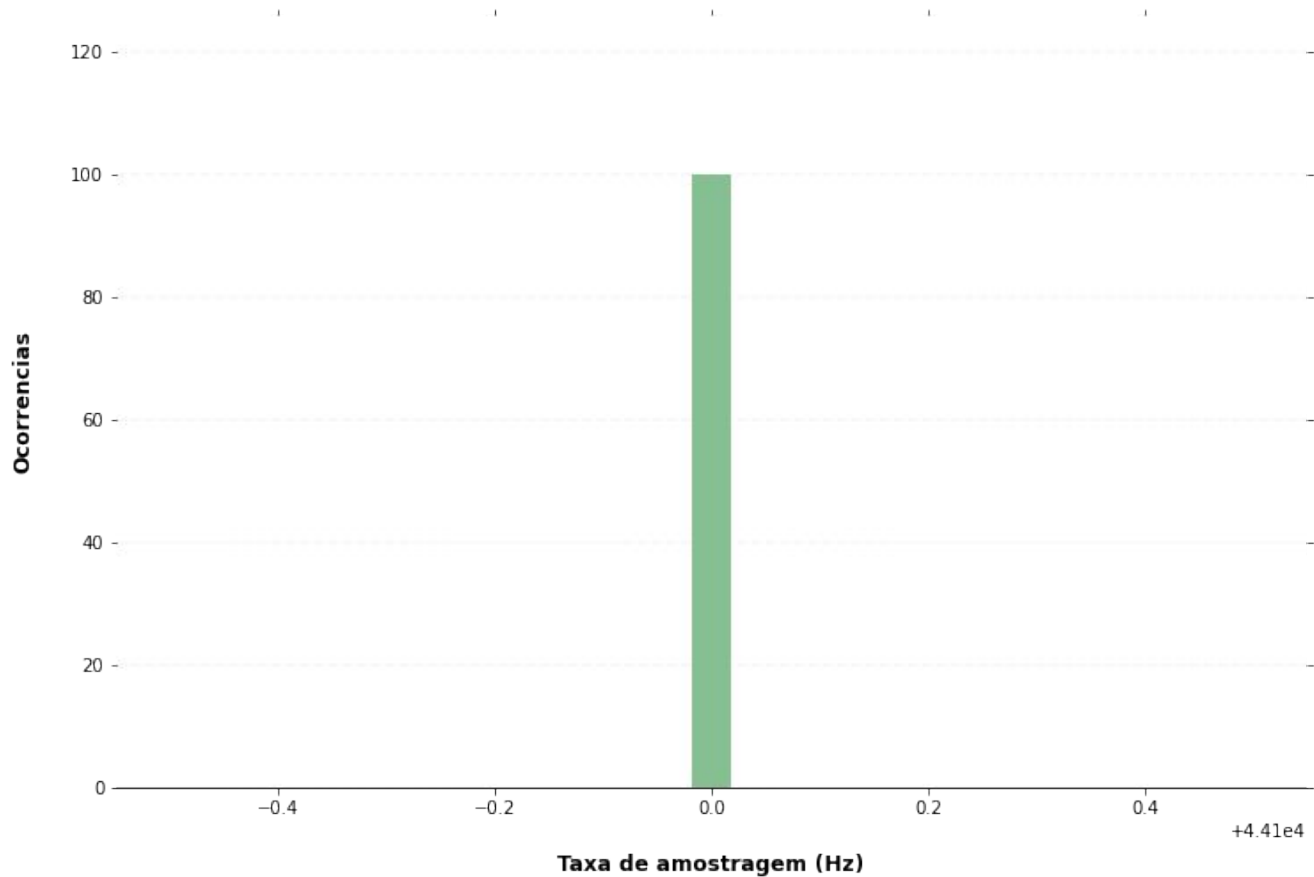
<https://numpy.org/doc/stable/reference/routines.statistics.html>



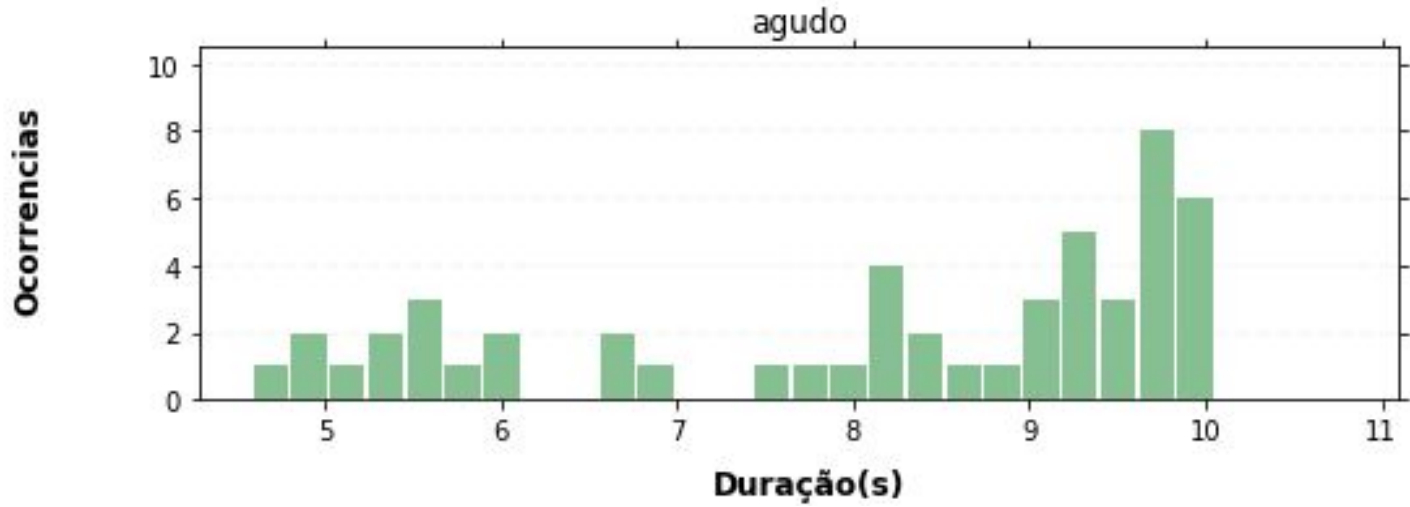
Resultados obtidos



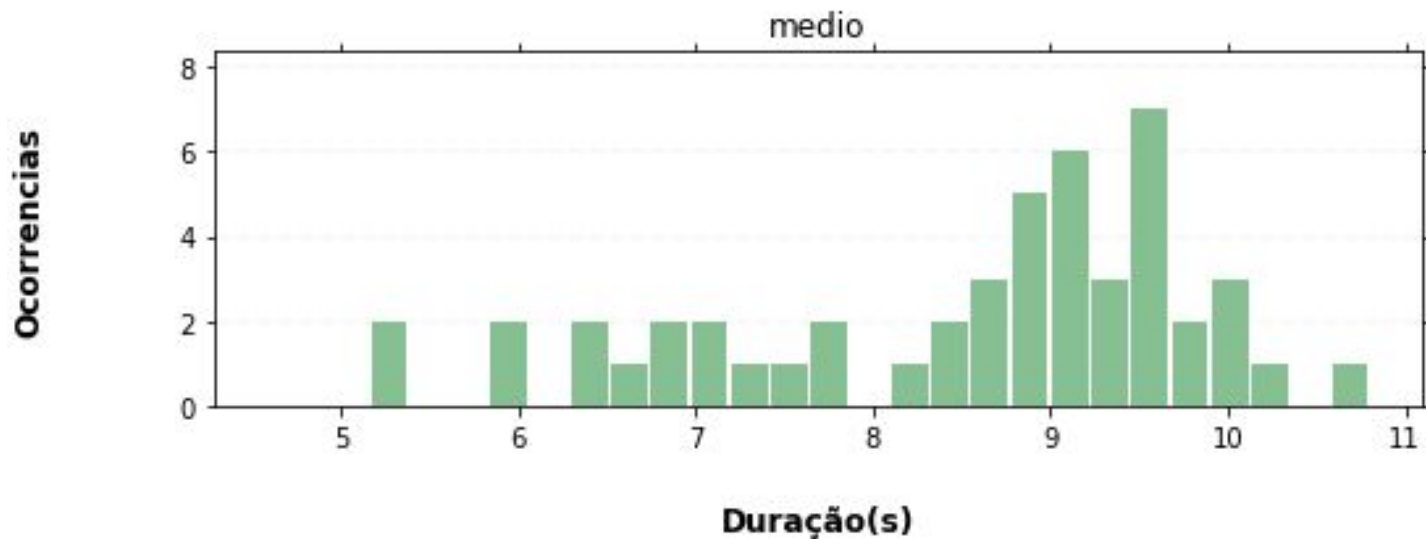
Ocorrência da taxa de amostragem



Quantidade de sons classificados como agudos e tempo de duração



Quantidade de sons classificados como médio e tempo de duração

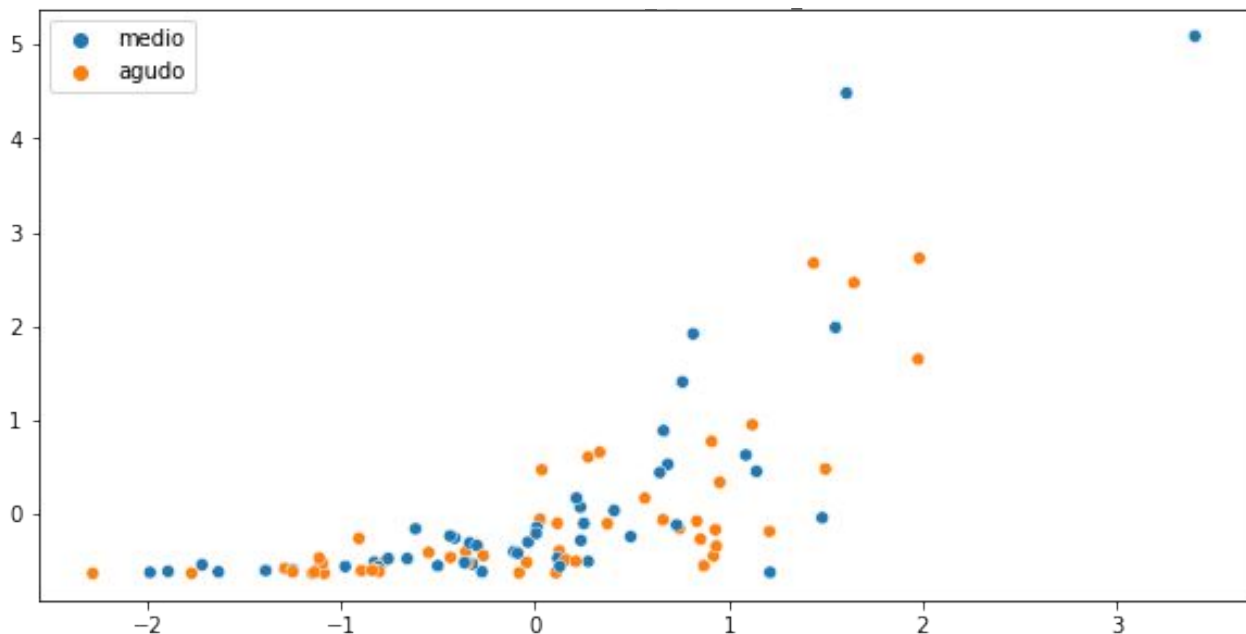


**Teste 1 - Features espectrais: “flatness”
(planicidade) espectral, rmse, centróide
espectral
feature estatística: mediana**

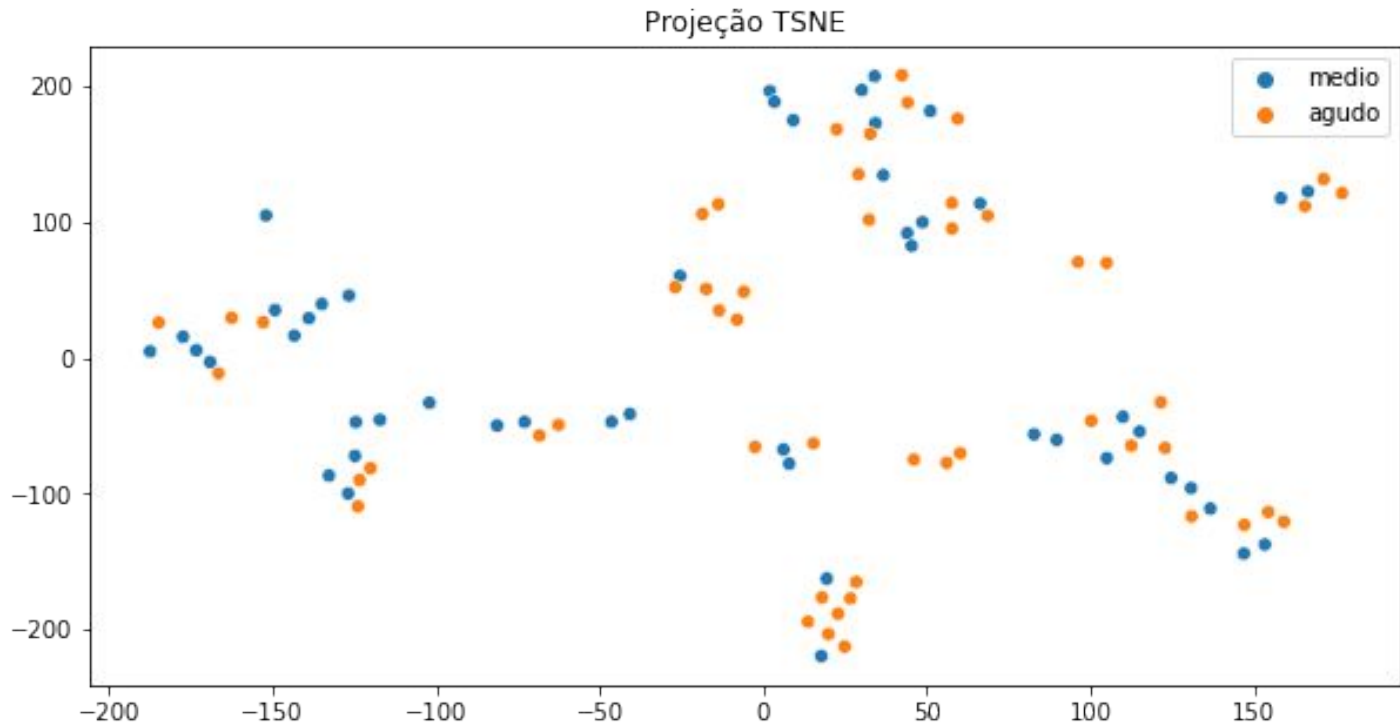


João bobo

Visualização dos dados: média do centróide espectral normalizada x média do “flatness” espectral



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

joão-bobo.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.46	0.46	0.46	13
medio	0.42	0.42	0.42	12
accuracy			0.44	25
macro avg	0.44	0.44	0.44	25
weighted avg	0.44	0.44	0.44	25

Classificação do áudio de saída (banco de dados): médio

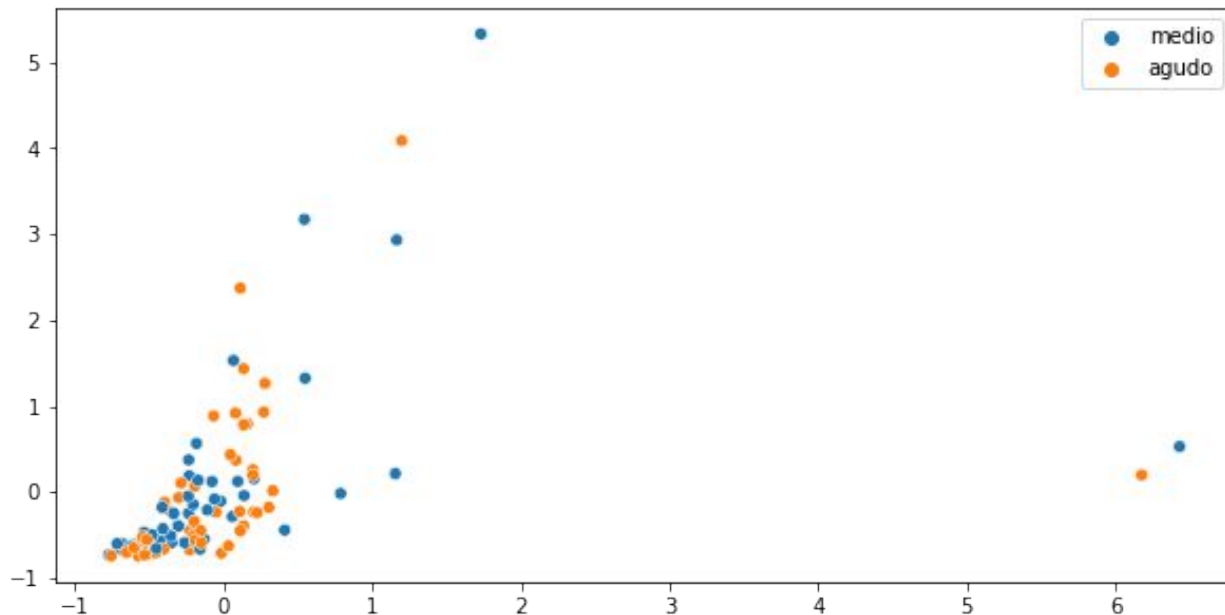
Classificação do áudio (classificador automático): agudo

Teste 2 - Features espectrais: “flatness” espectral, rmse, centróide espectral feature estatística: desvio padrão

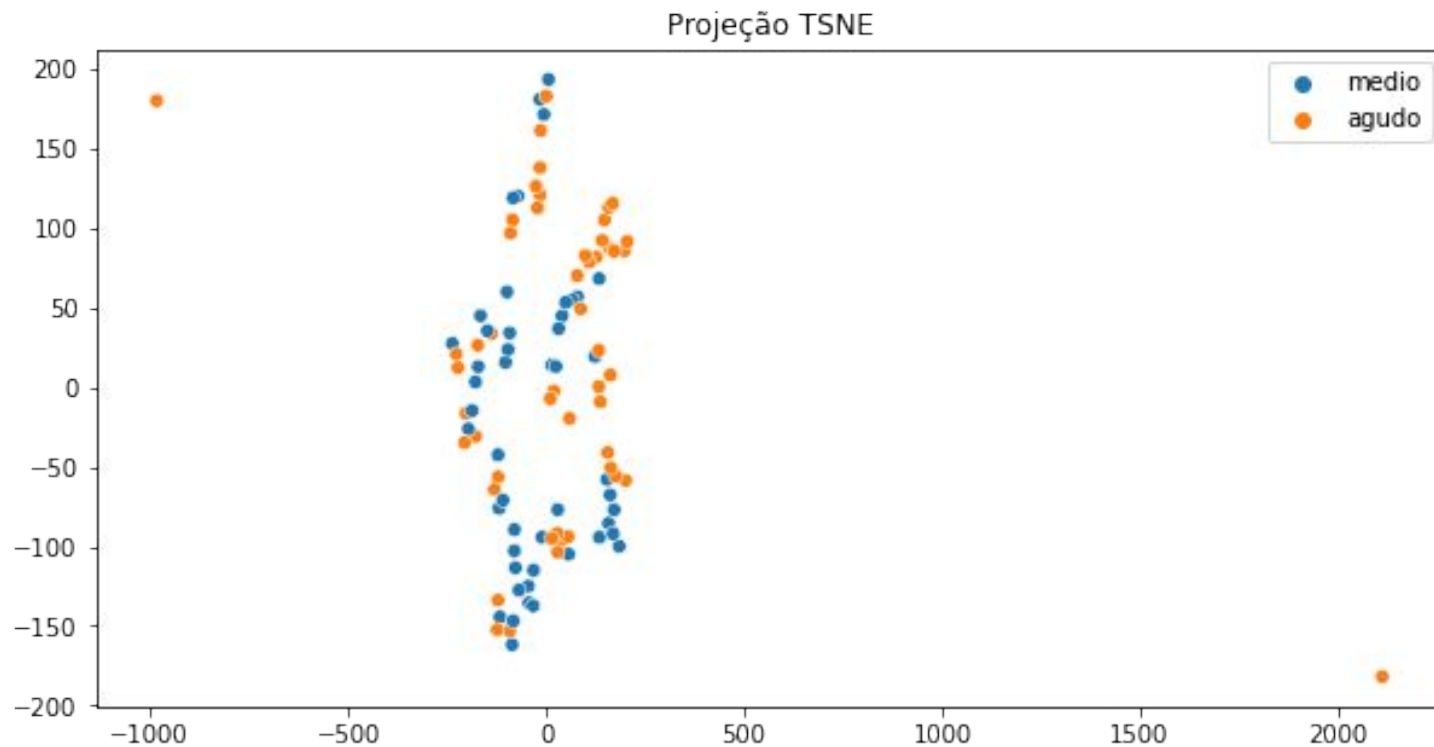


coró-coró

Visualização dos dados: média do centróide espectral normalizada x média do “flatness” espectral normalizada



Projeção TSNE



Classificador

Entrada:

coró-coró_teste.wav



Saída:

coró-coró.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.53	0.69	0.60	13
medio	0.50	0.33	0.40	12
accuracy			0.52	25
macro avg	0.51	0.51	0.50	25
weighted avg	0.52	0.52	0.50	25

Classificação do áudio de saída (banco de dados): médio

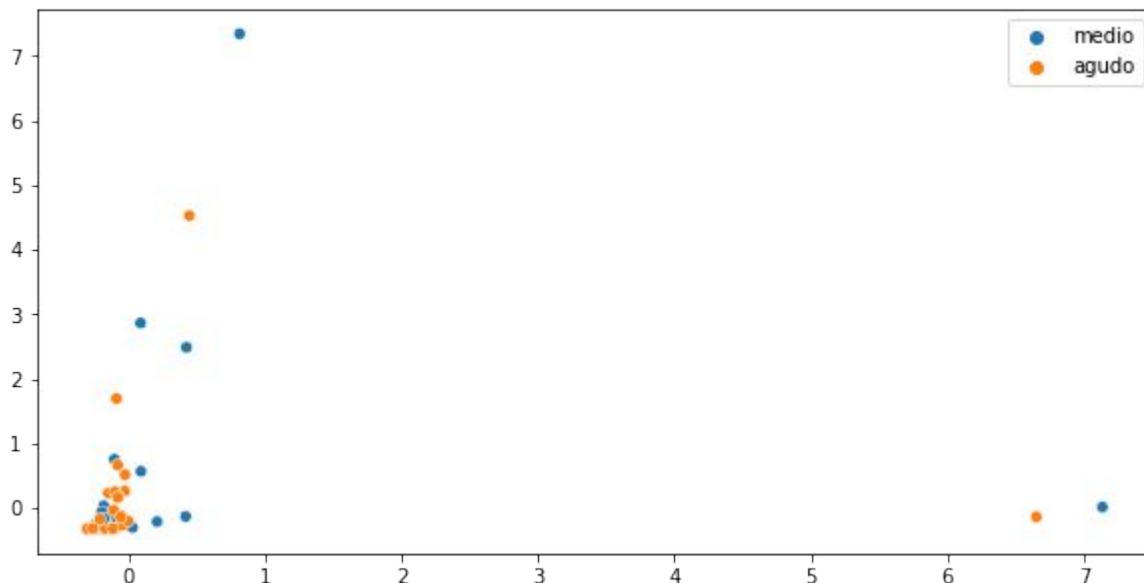
Classificação do áudio (classificador automático): médio

Teste 5 - Features espectrais: “flatness” espectral, rmse, centróide espectral features estatísticas: variância e desvio padrão

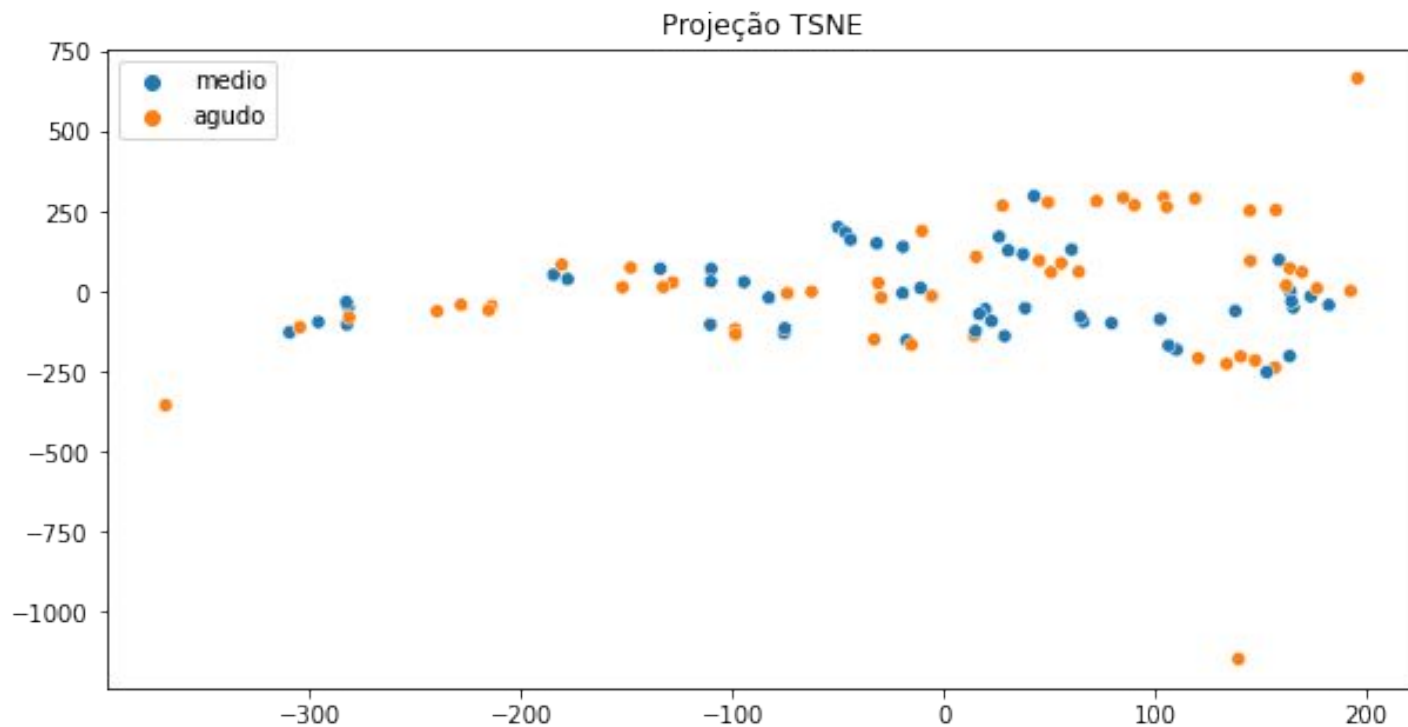


coró-coró

Visualização dos dados: média do centróide espectral normalizada x média do “flatness” espectral normalizada



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

coró-coró.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.53	0.69	0.60	13
medio	0.50	0.33	0.40	12
accuracy			0.52	25
macro avg	0.51	0.51	0.50	25
weighted avg	0.52	0.52	0.50	25

Classificação do áudio de saída (banco de dados): médio

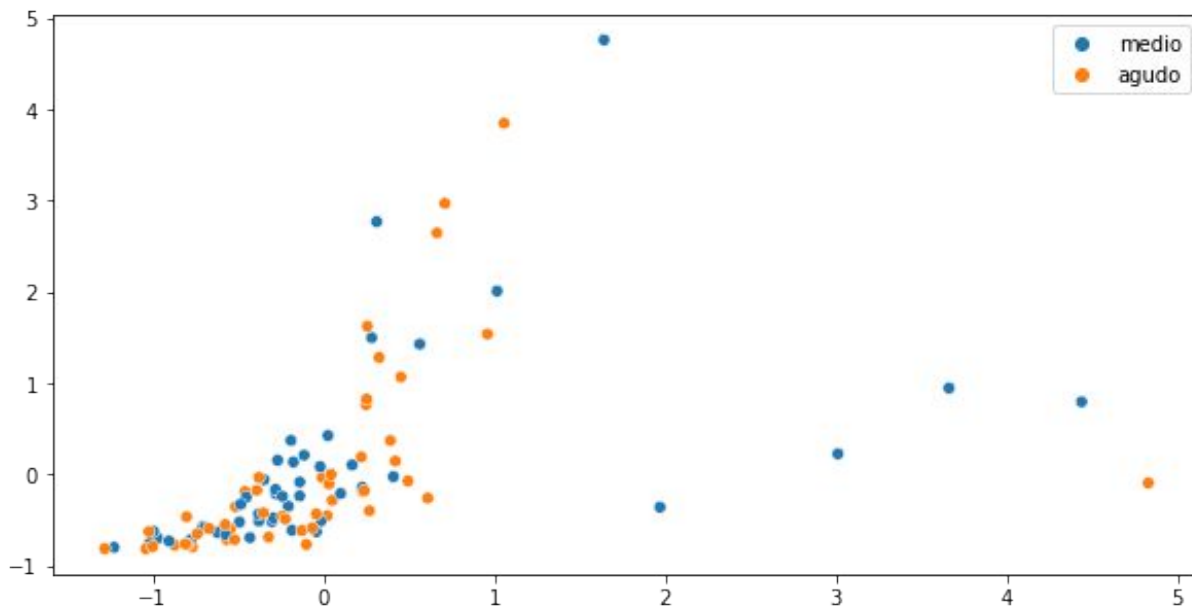
Classificação do áudio (classificador automático): médio

Teste 7 - Features espectrais: “flatness” espectral, rmse, centróide espectral feature estatística: máximo

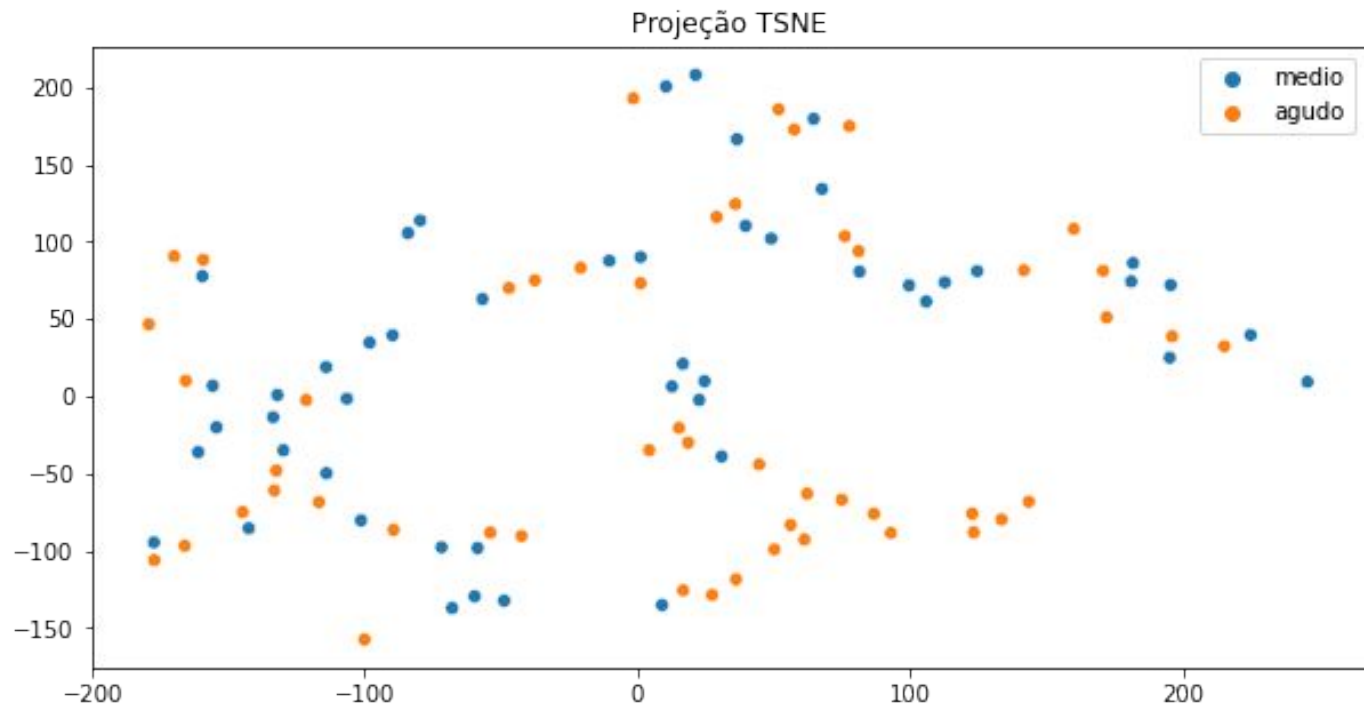


falcão-caburé

Visualização dos dados: média do centróide espectral normalizada x média do "flatness" espectral normalizada



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

falcão-caburé.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.67	0.77	0.71	13
medio	0.70	0.58	0.64	12
accuracy			0.68	25
macro avg	0.68	0.68	0.68	25
weighted avg	0.68	0.68	0.68	25

Classificação do áudio de saída (banco de dados): médio

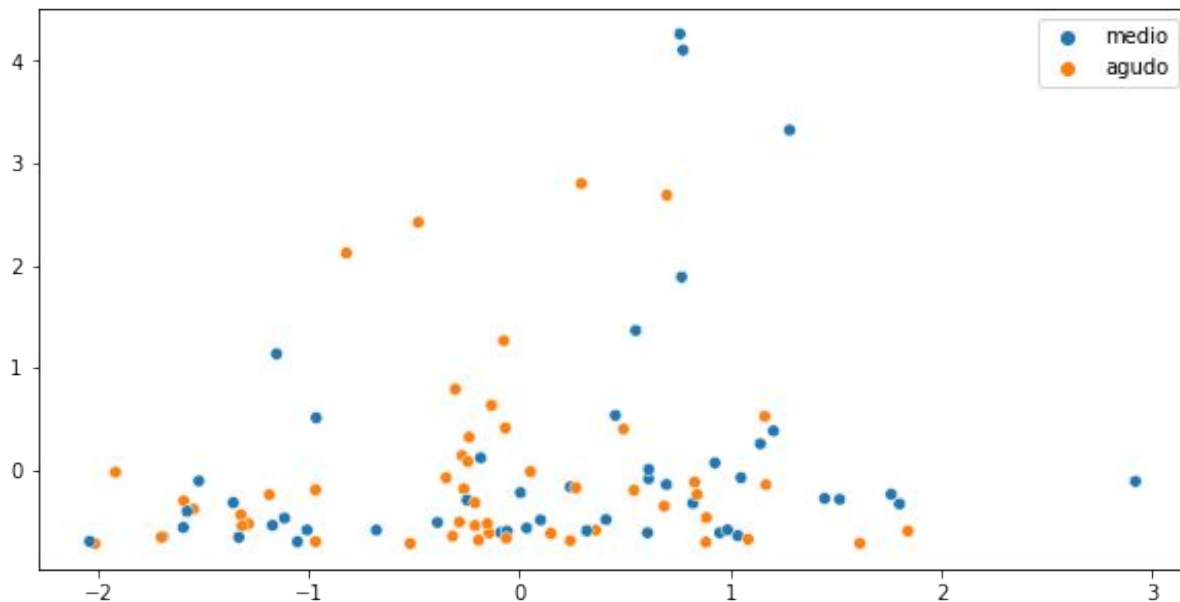
Classificação do áudio (classificador automático): médio

**Teste 8 - Features espectrais: “flatness”
espectral, rmse, chroma stft
feature estatística: mediana**

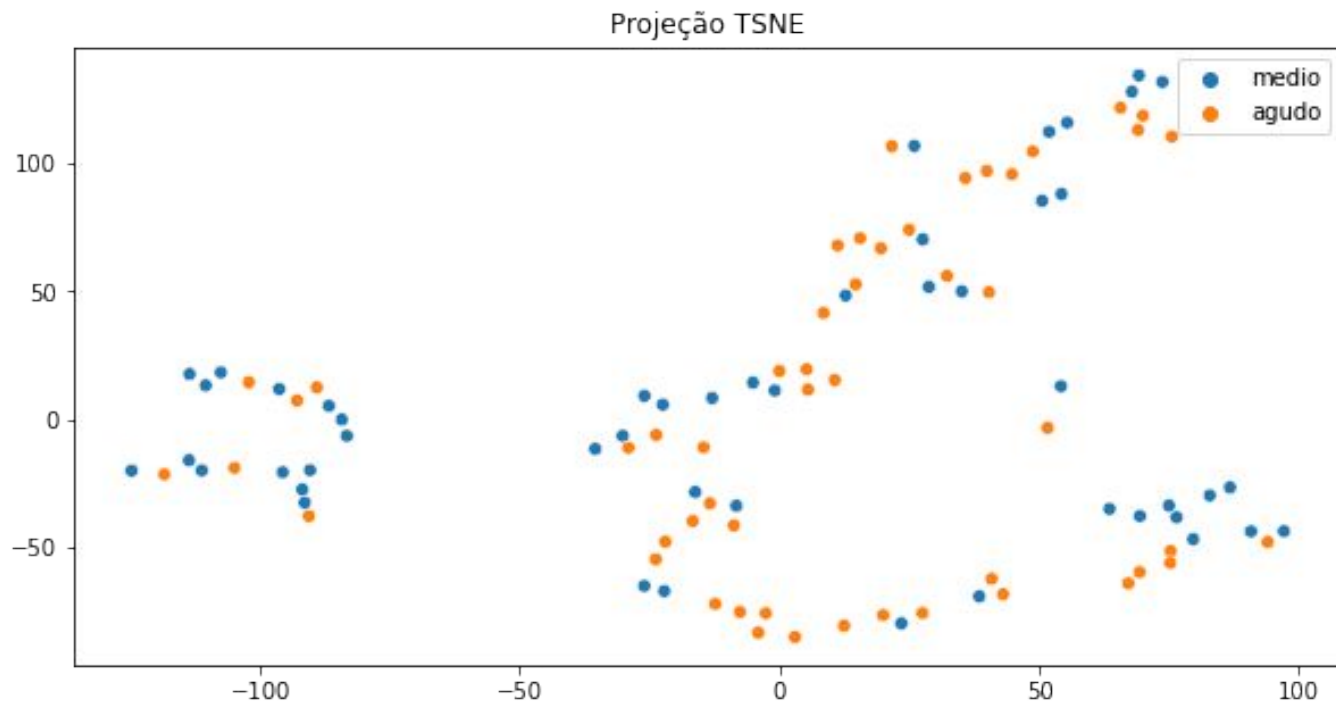


falcão de coleira

Visualização dos dados: média do chroma stft x média do “flatness” espectral normalizado



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

falcão-de-coleira.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.71	0.77	0.74	13
medio	0.73	0.67	0.70	12
accuracy			0.72	25
macro avg	0.72	0.72	0.72	25
weighted avg	0.72	0.72	0.72	25

Classificação do áudio de saída (banco de dados): médio

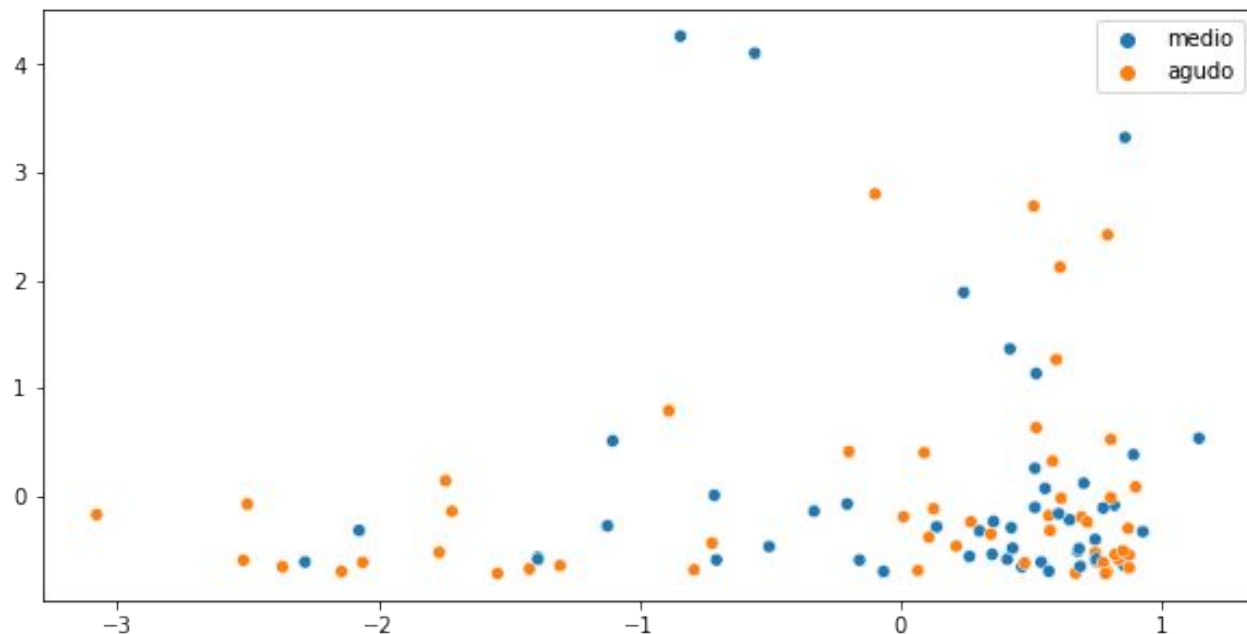
Classificação do áudio (classificador automático): médio

**Teste 10 - Features espectrais: “flatness”
espectral, rmse
feature rítmica: tempograma de fourier
feature estatística: mediana**

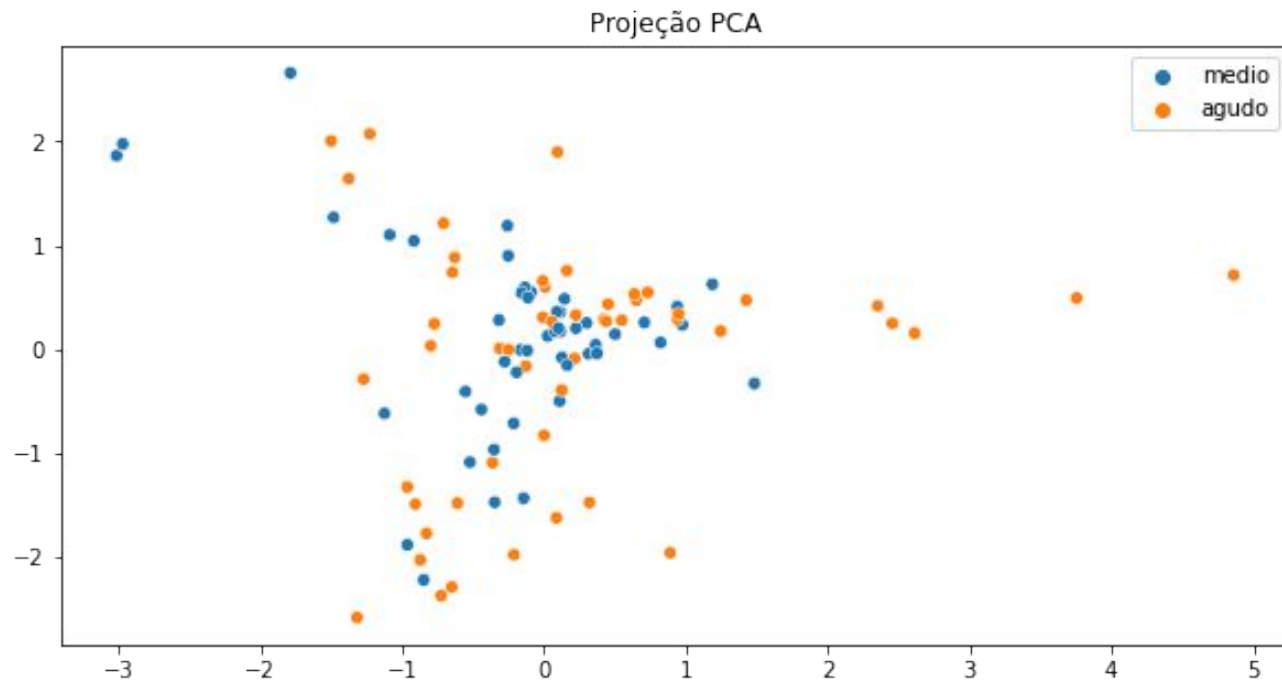


apuim-de-costas-pretas

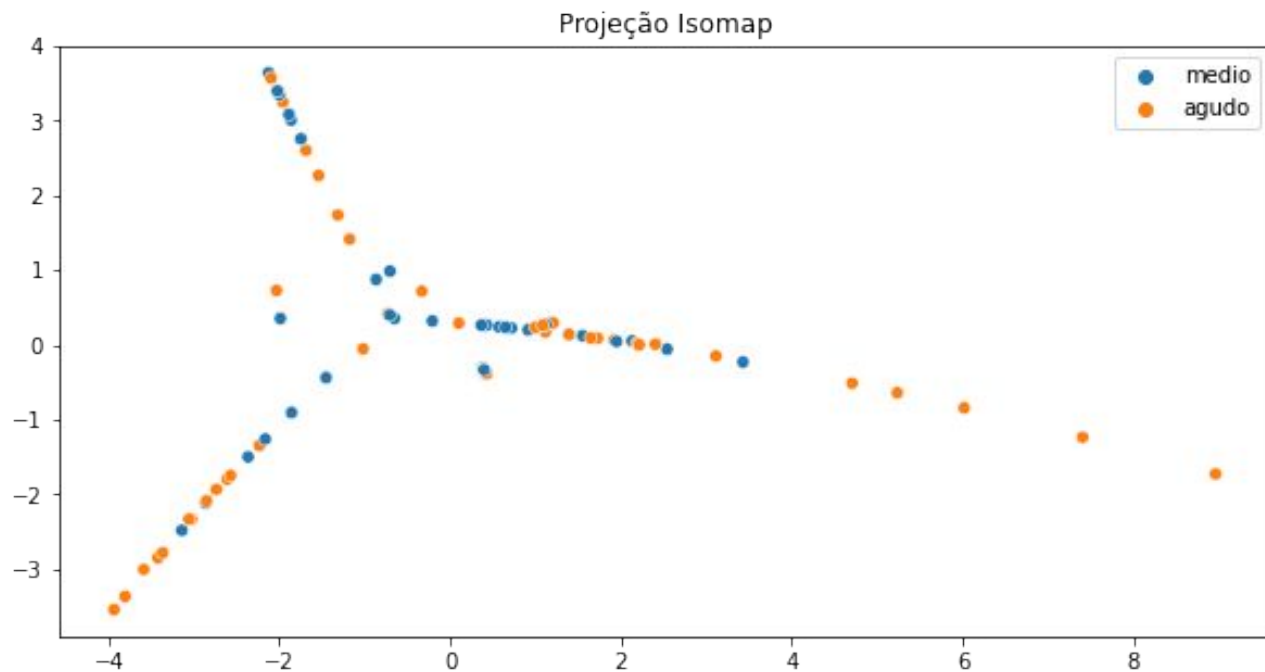
Visualização dos dados: média tempograma de fourier normalizado x média do “flatness” espectral normalizada



Projeção PCA



Projeção Isomap





Classificador

Entrada:

coró-coró_teste.wav



Saída:

apuim-de-costas-pretas.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.50	0.54	0.52	13
medio	0.45	0.42	0.43	12
accuracy			0.48	25
macro avg	0.48	0.48	0.48	25
weighted avg	0.48	0.48	0.48	25

Classificação do áudio de saída(banco de dados): médio

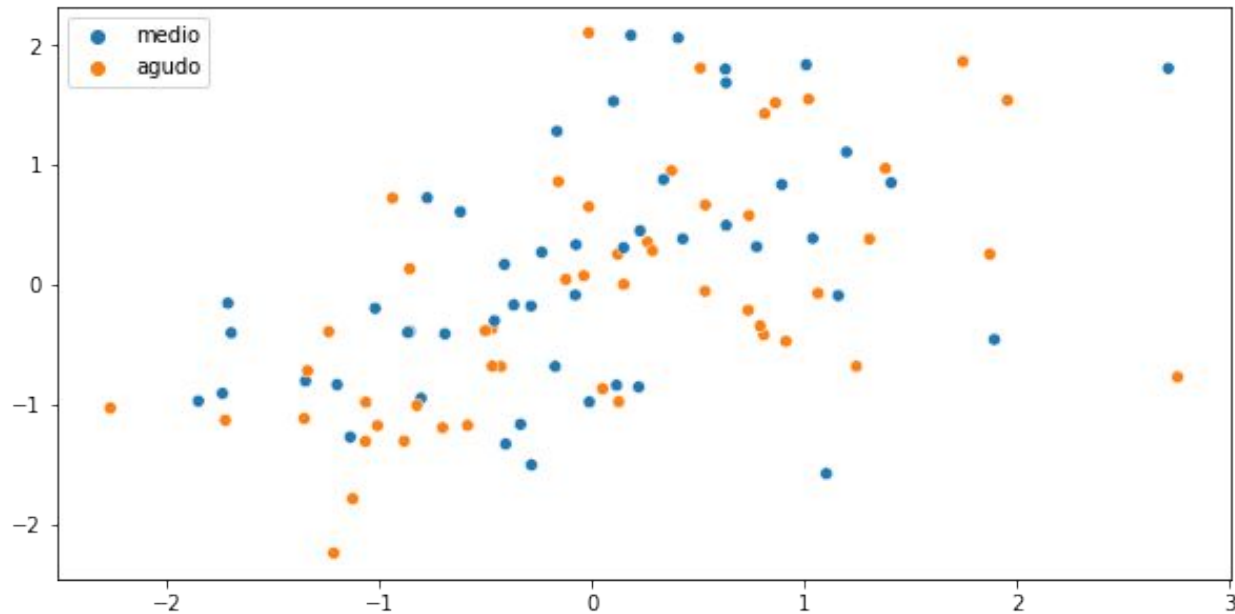
Classificação do áudio (classificador automático): agudo

**Teste 11 - Features espectrais: centroide
espectral, rmse, largura de banda
espectral
feature estatística: mediana**

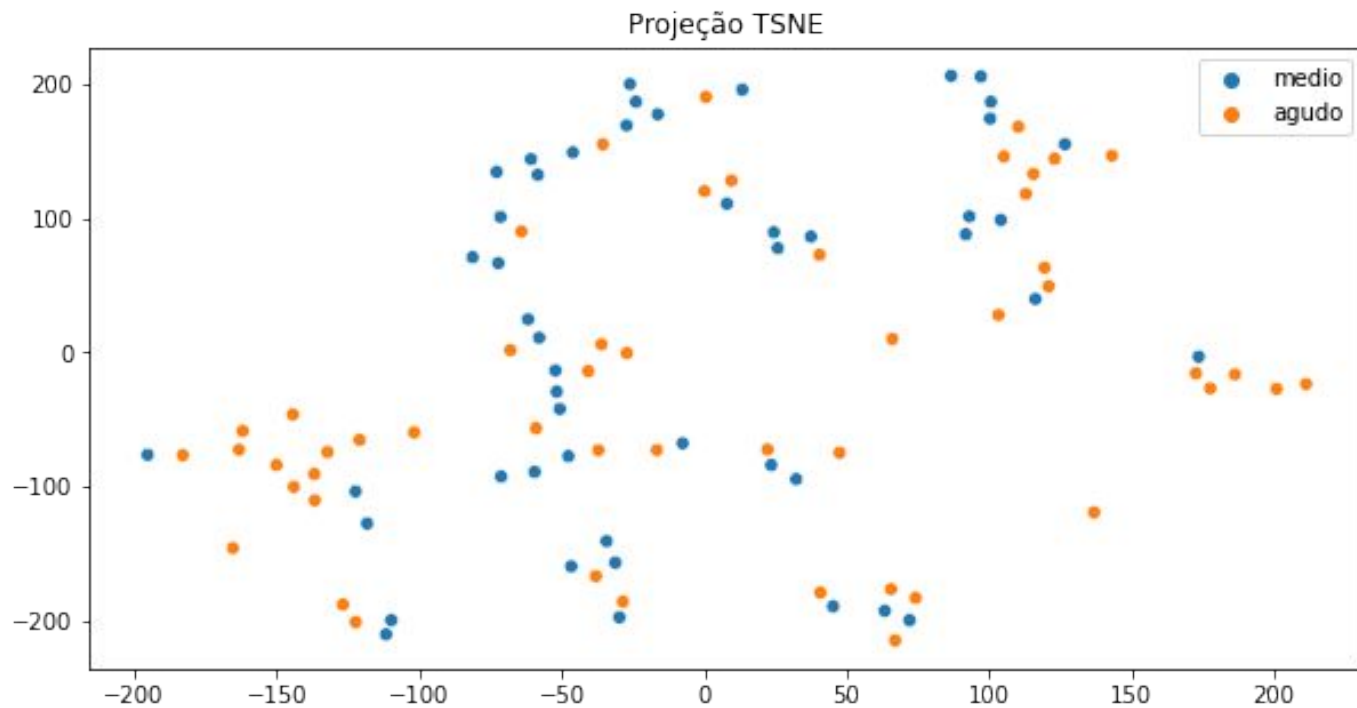


uru

Visualização dos dados: média da largura de banda espectral normalizada x média do “flatness” espectral normalizada



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

uru.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.80	0.62	0.70	13
medio	0.67	0.83	0.74	12
accuracy			0.72	25
macro avg	0.73	0.72	0.72	25
weighted avg	0.74	0.72	0.72	25

Classificação do áudio de saída (banco de dados): médio

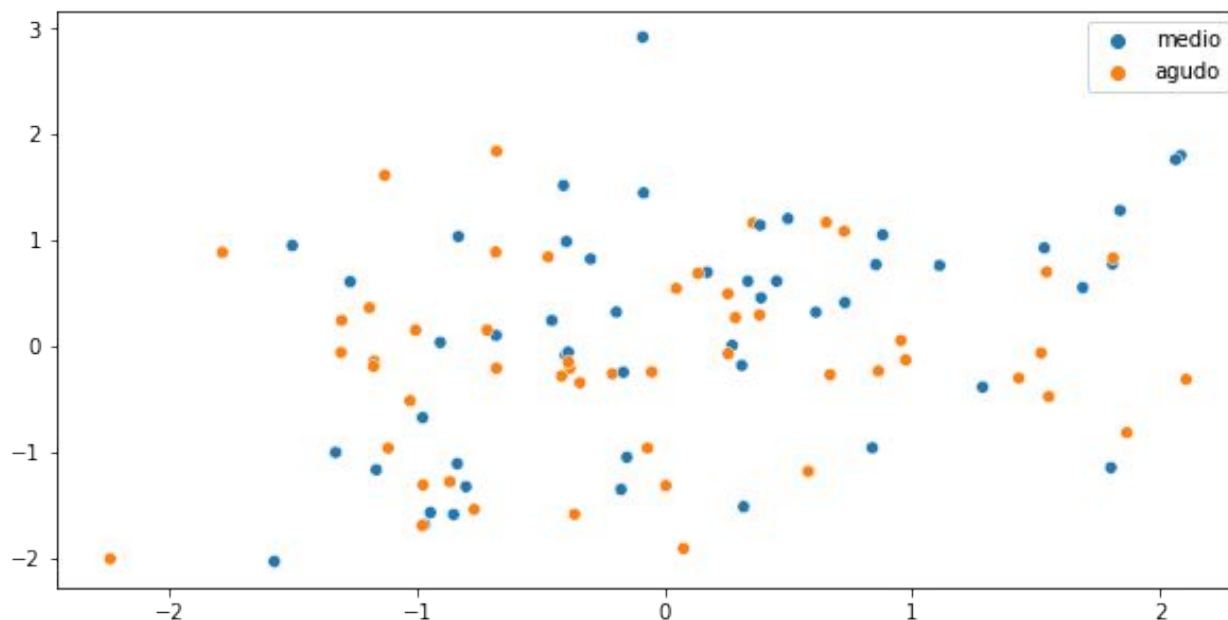
Classificação do áudio (classificador automático): médio

**Teste 12- Features espectrais: largura de
banda espectral, rmse, croma stft
Feature estatística: média**

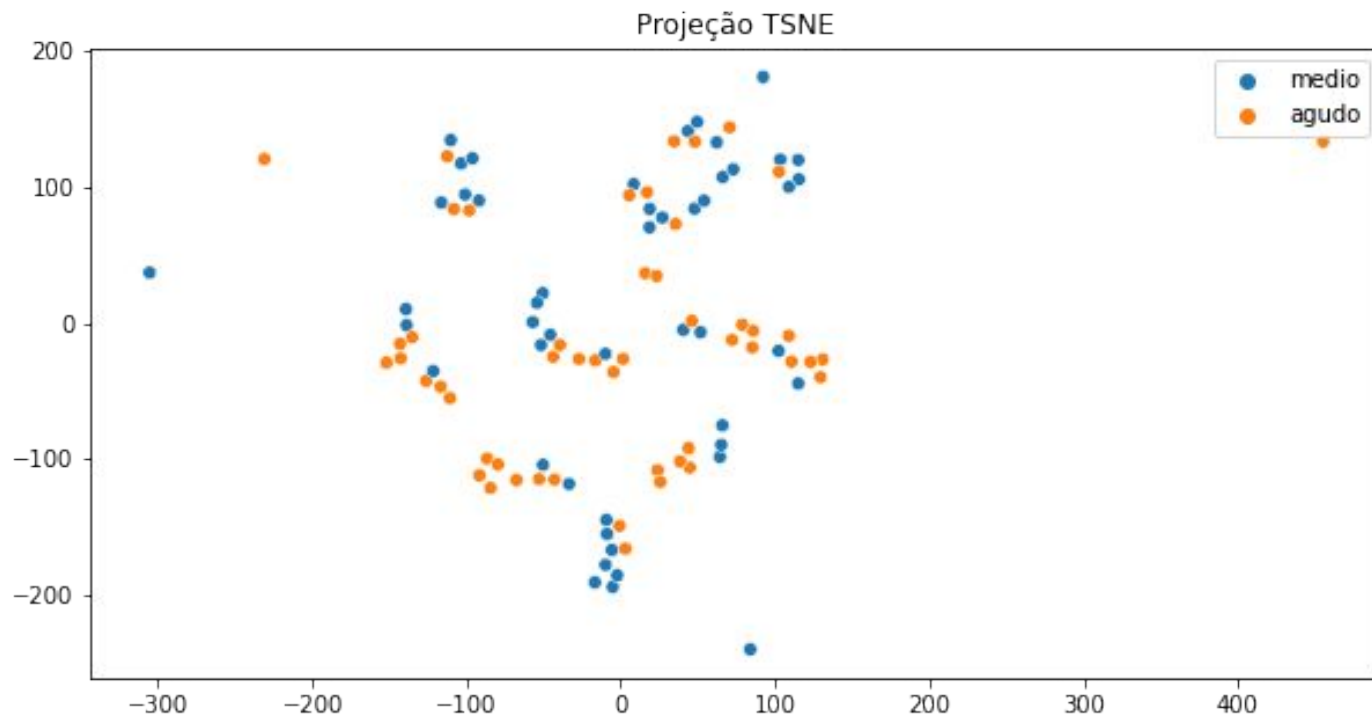


garça-branca

Visualização dos dados: média da largura de banda espectral x média do croma stft normalizada



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

garça-branca.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.75	0.69	0.72	13
medio	0.69	0.75	0.72	12
accuracy			0.72	25
macro avg	0.72	0.72	0.72	25
weighted avg	0.72	0.72	0.72	25

Classificação do áudio de saída (banco de dados): médio

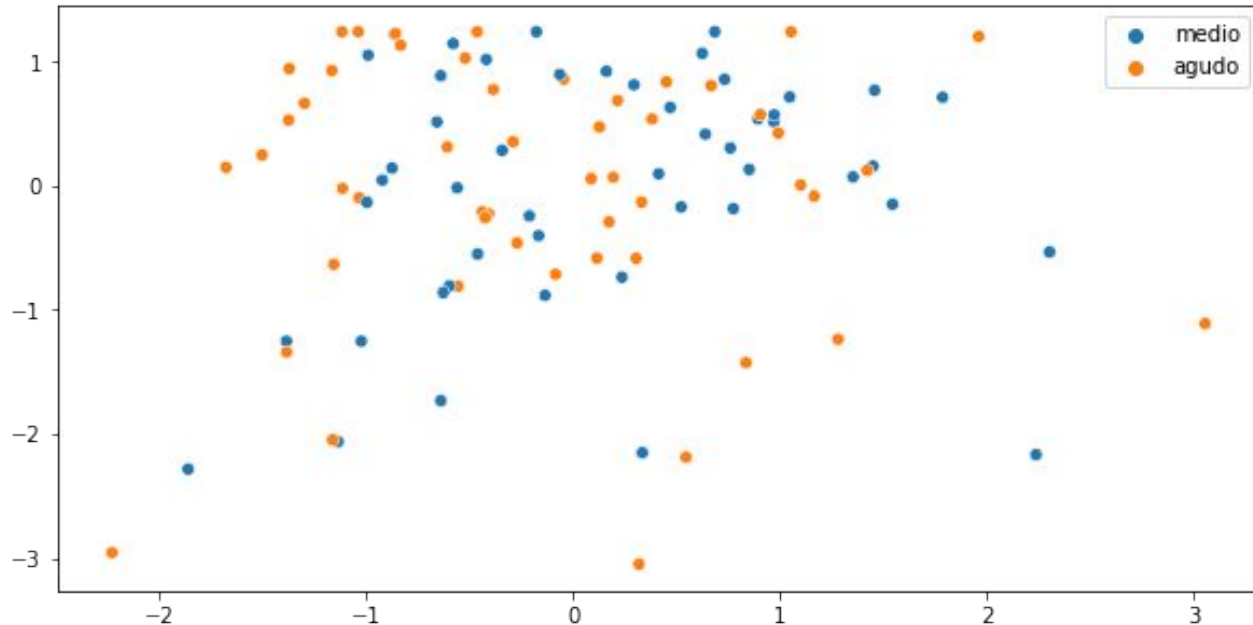
Classificação do áudio (classificador automático): agudo

**Teste 13- Features espectrais: largura de
banda espectral, rmse, croma stft
feature estatística: máximo**

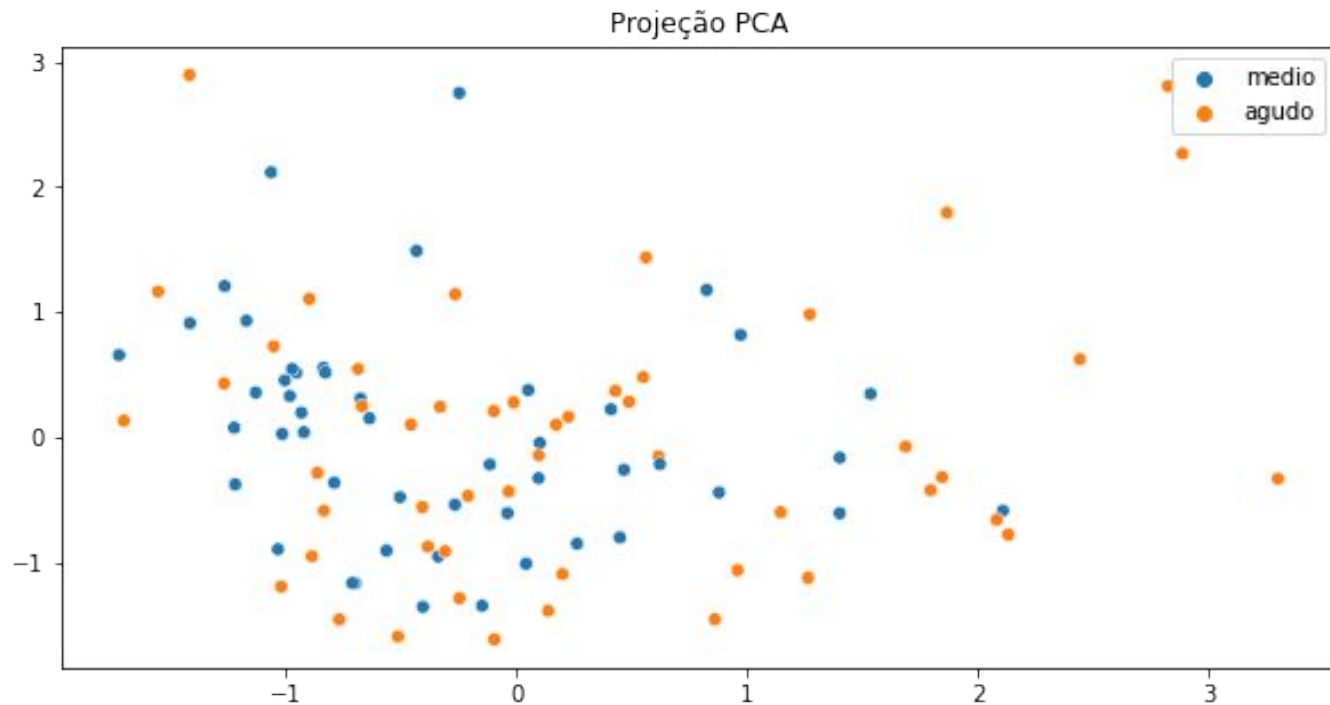


tucano de bico verde

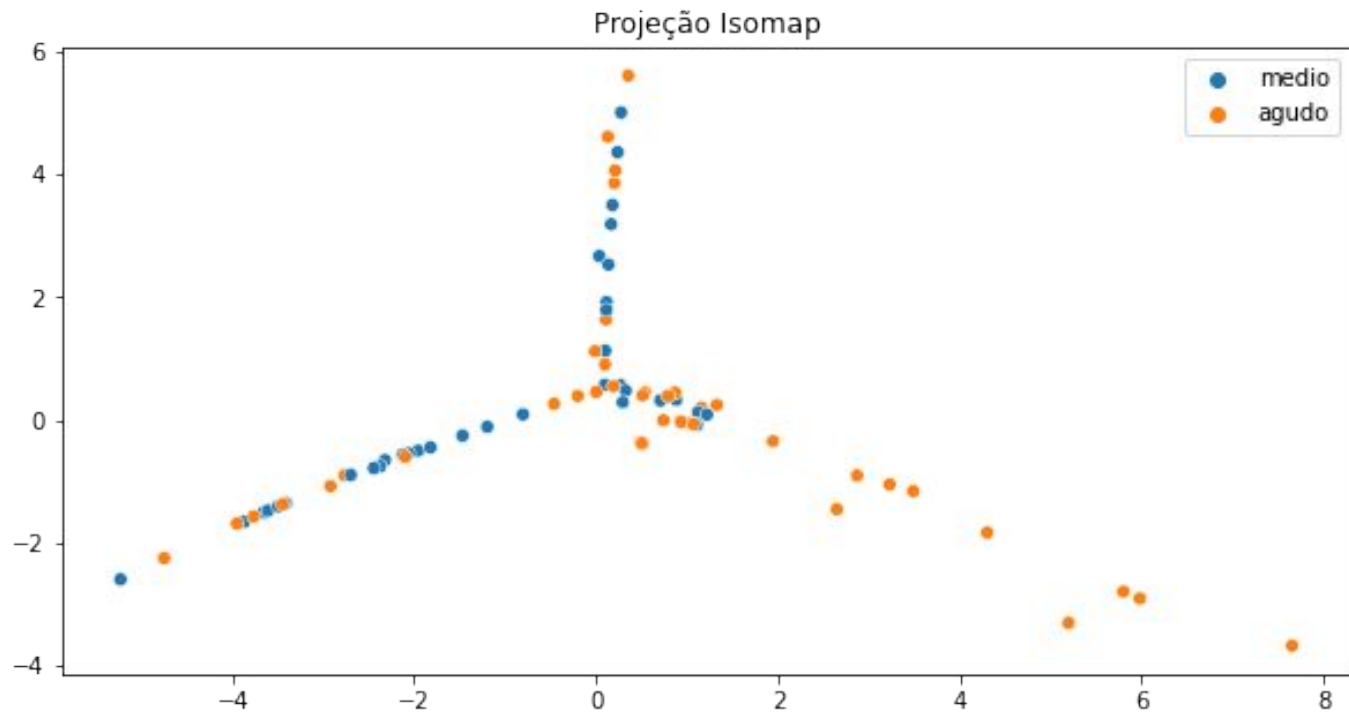
Visualização dos dados: média da largura de banda espectral x média do croma stft normalizada



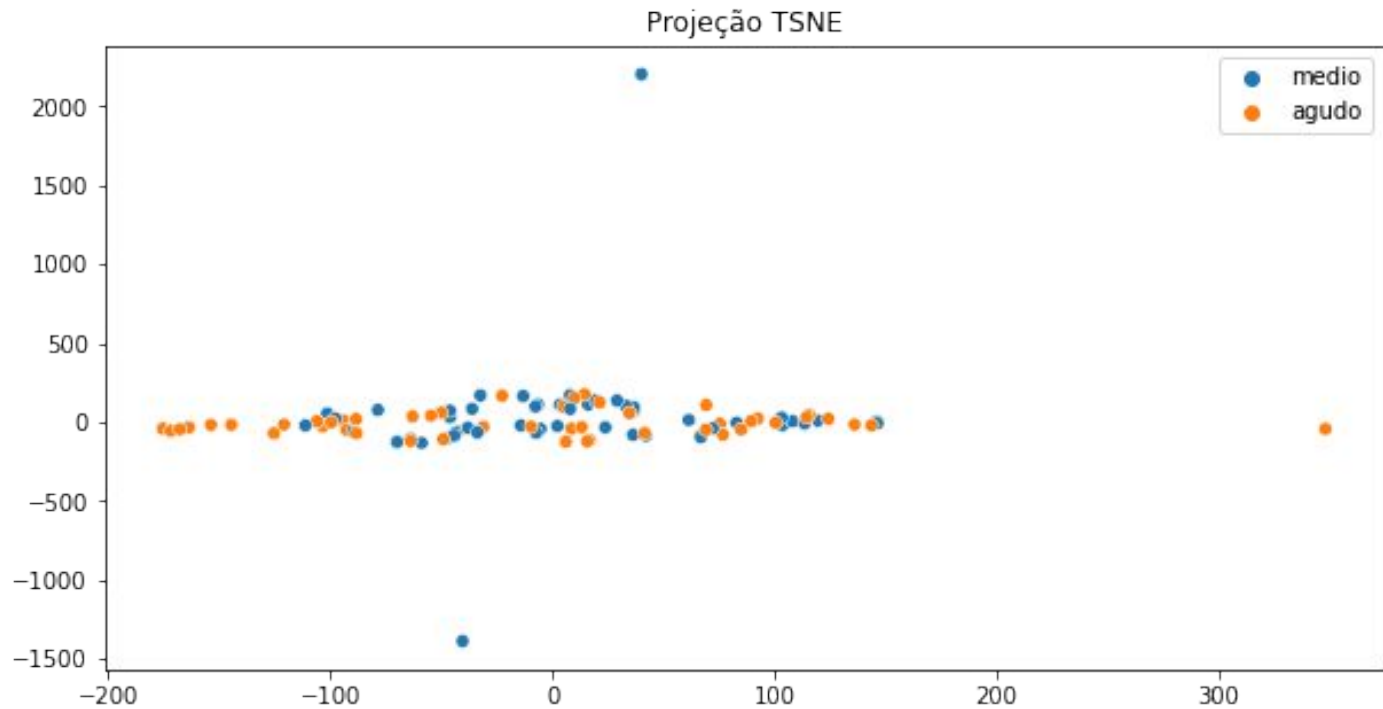
Projeção PCA



Projeção Isomap



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

tucano-de-bico-verde.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.60	0.46	0.52	13
medio	0.53	0.67	0.59	12
accuracy			0.56	25
macro avg	0.57	0.56	0.56	25
weighted avg	0.57	0.56	0.56	25

Classificação do áudio de saída (banco de dados): médio

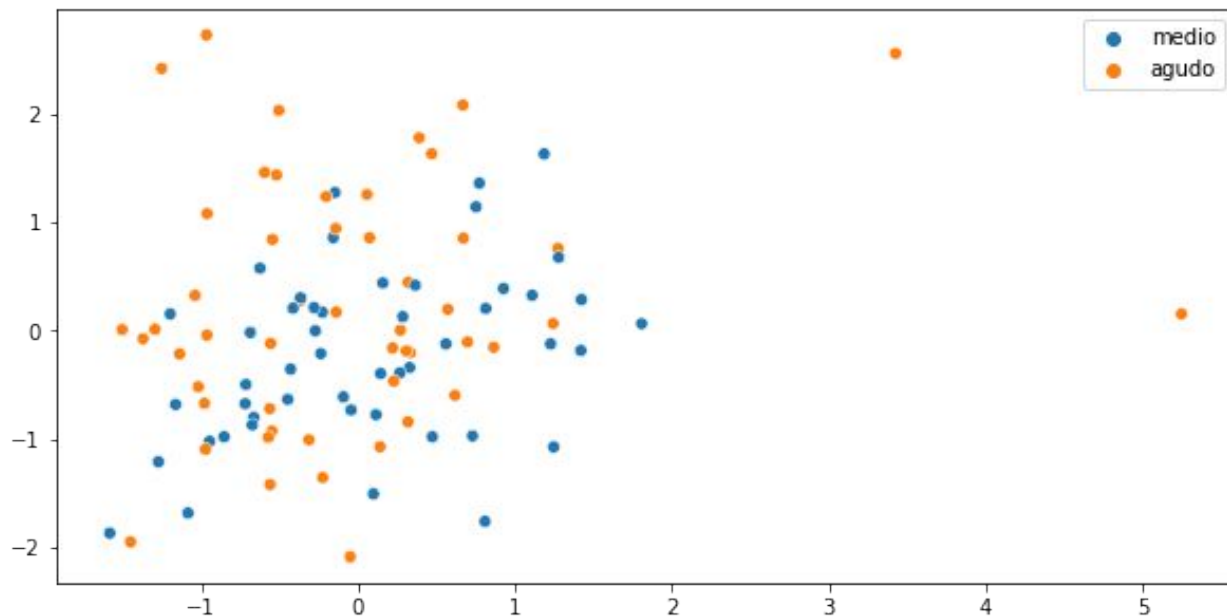
Classificação do áudio (classificador automático): agudo

**Teste 14 - Features espectrais: largura de
banda espectral, rmse, croma stft
features estatísticas: desvio padrão e
variância**

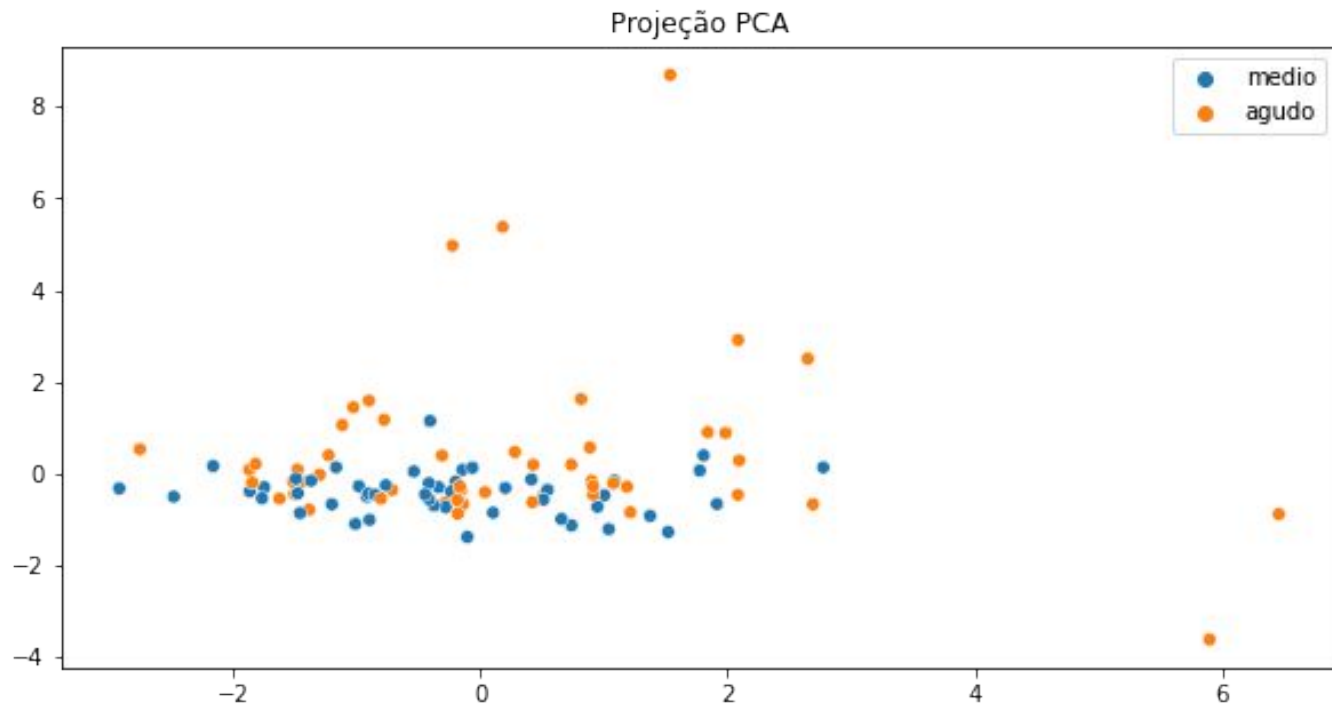


beija-flor cinza

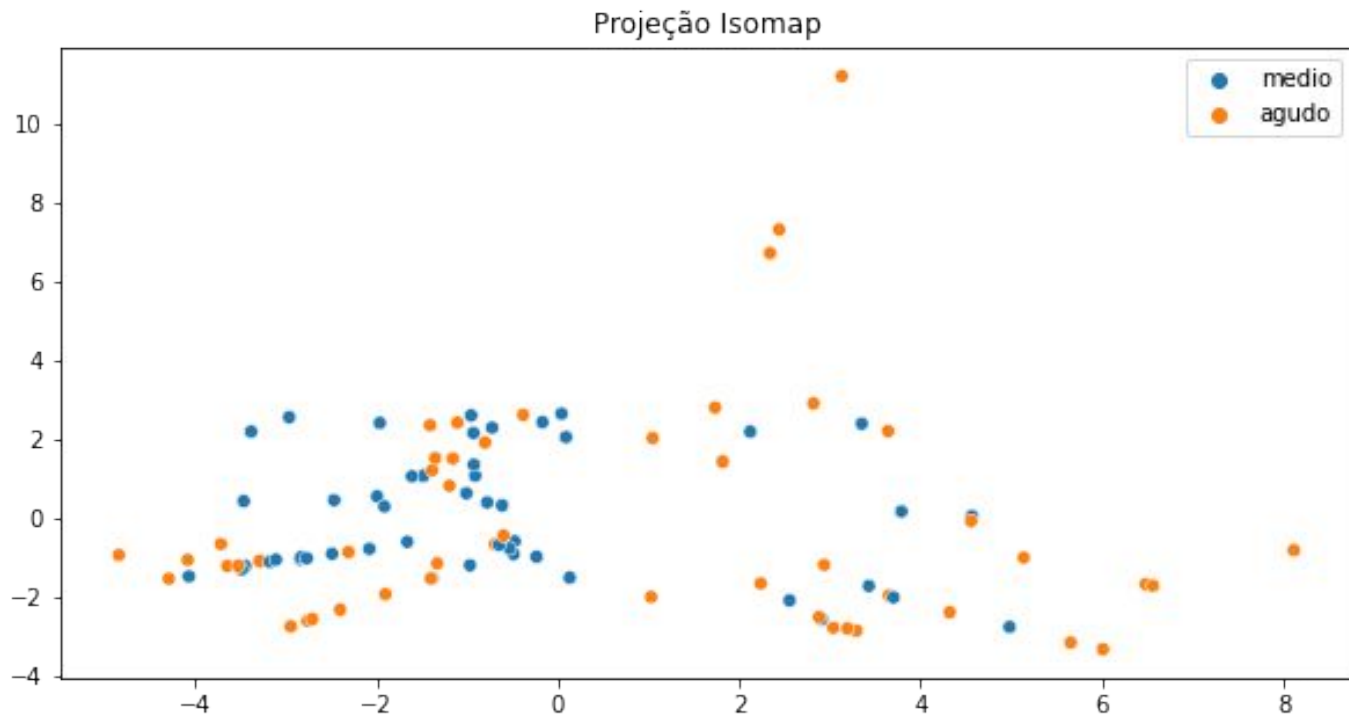
Visualização dos dados: média da largura de banda espectral x média do croma stft normalizada



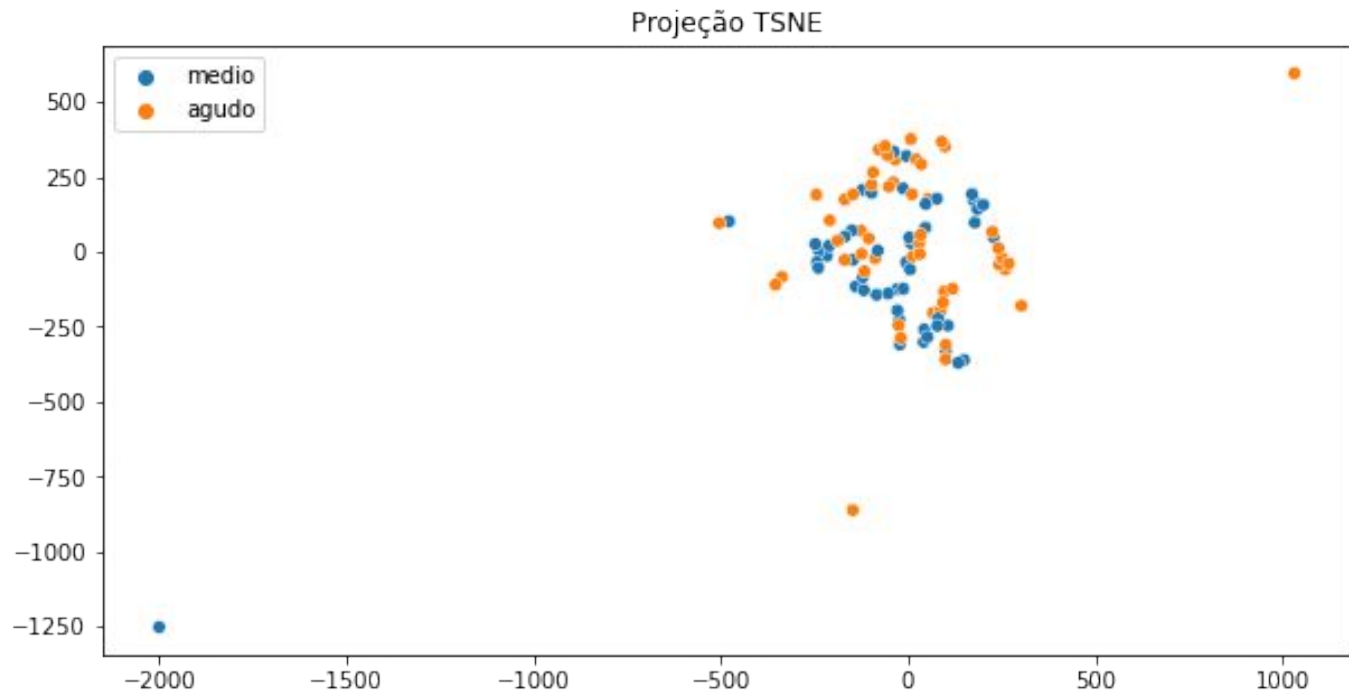
Projeção PCA



Projeção Isomap



Projeção TSNE





Classificador

Entrada:

coró-coró_teste.wav



Saída:

beija-flor-cinza.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.56	0.38	0.45	13
medio	0.50	0.67	0.57	12
accuracy			0.52	25
macro avg	0.53	0.53	0.51	25
weighted avg	0.53	0.52	0.51	25

Classificação do áudio de saída (banco de dados): médio

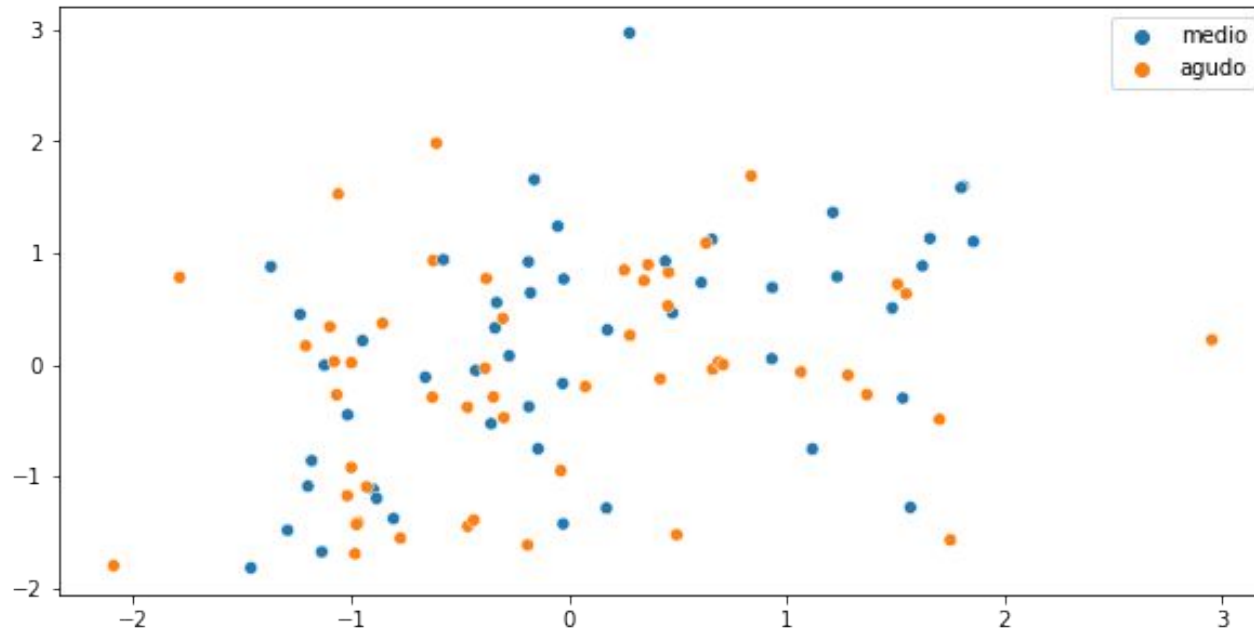
Classificação do áudio (classificador automático): agudo

**Teste 15 - Features espectrais: largura de
banda espectral, rmse, croma stft
feature estatística: mediana**

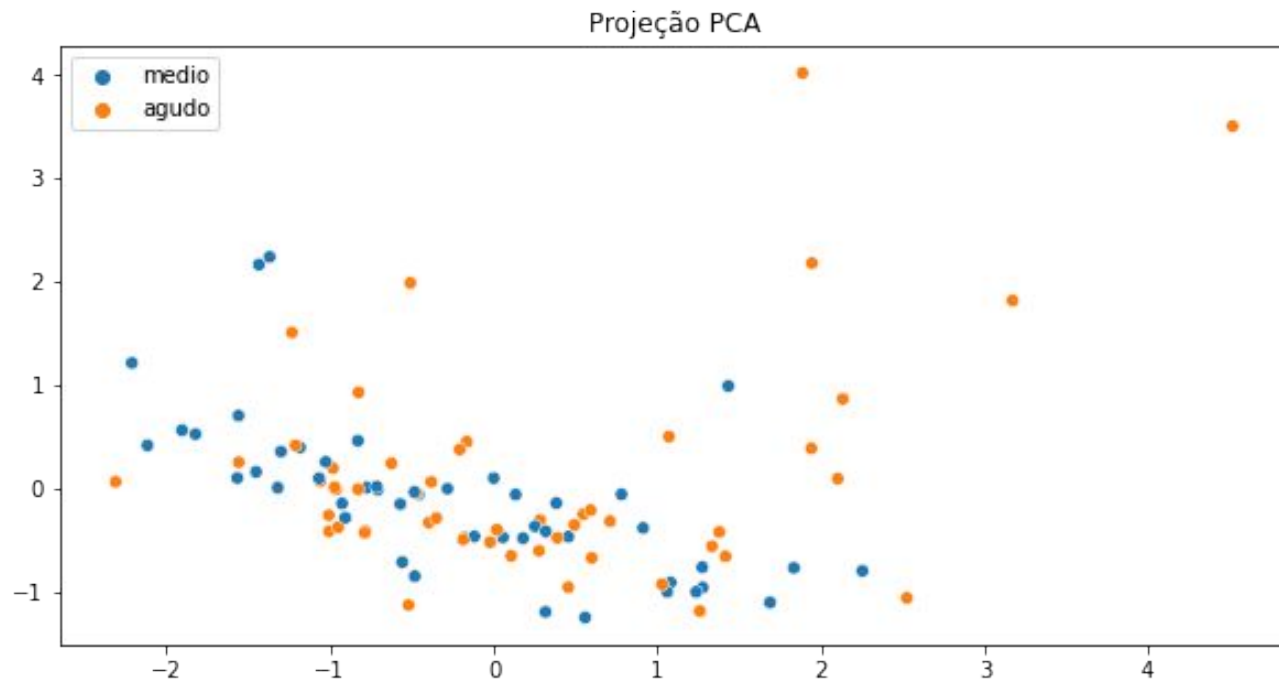


quero-quero

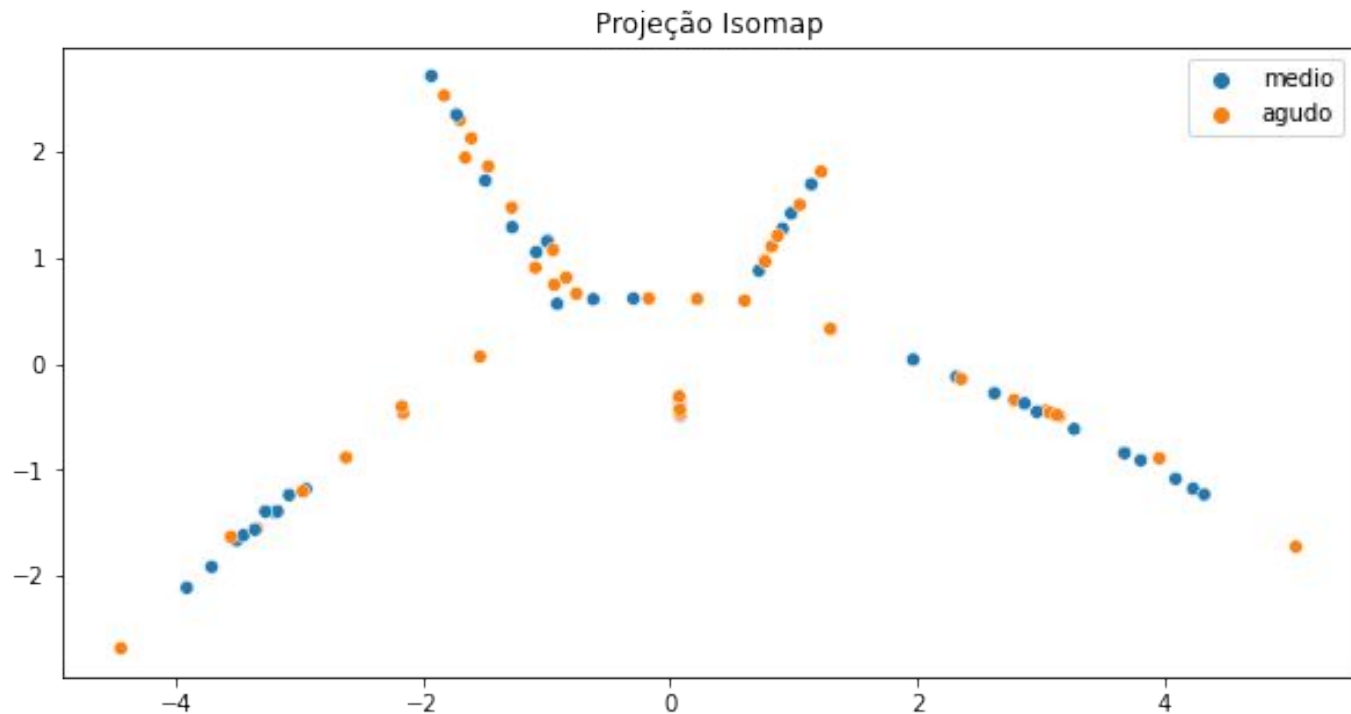
Visualização dos dados: média da largura de banda espectral x média do croma stft normalizada



Projeção PCA



Projeção Isomap





Classificador

Entrada:

coró-coró_teste.wav



Saída:

quero-quero.wav





Principais métricas de classificação

	precision	recall	f1-score	support
agudo	0.64	0.69	0.67	13
medio	0.64	0.58	0.61	12
accuracy			0.64	25
macro avg	0.64	0.64	0.64	25
weighted avg	0.64	0.64	0.64	25

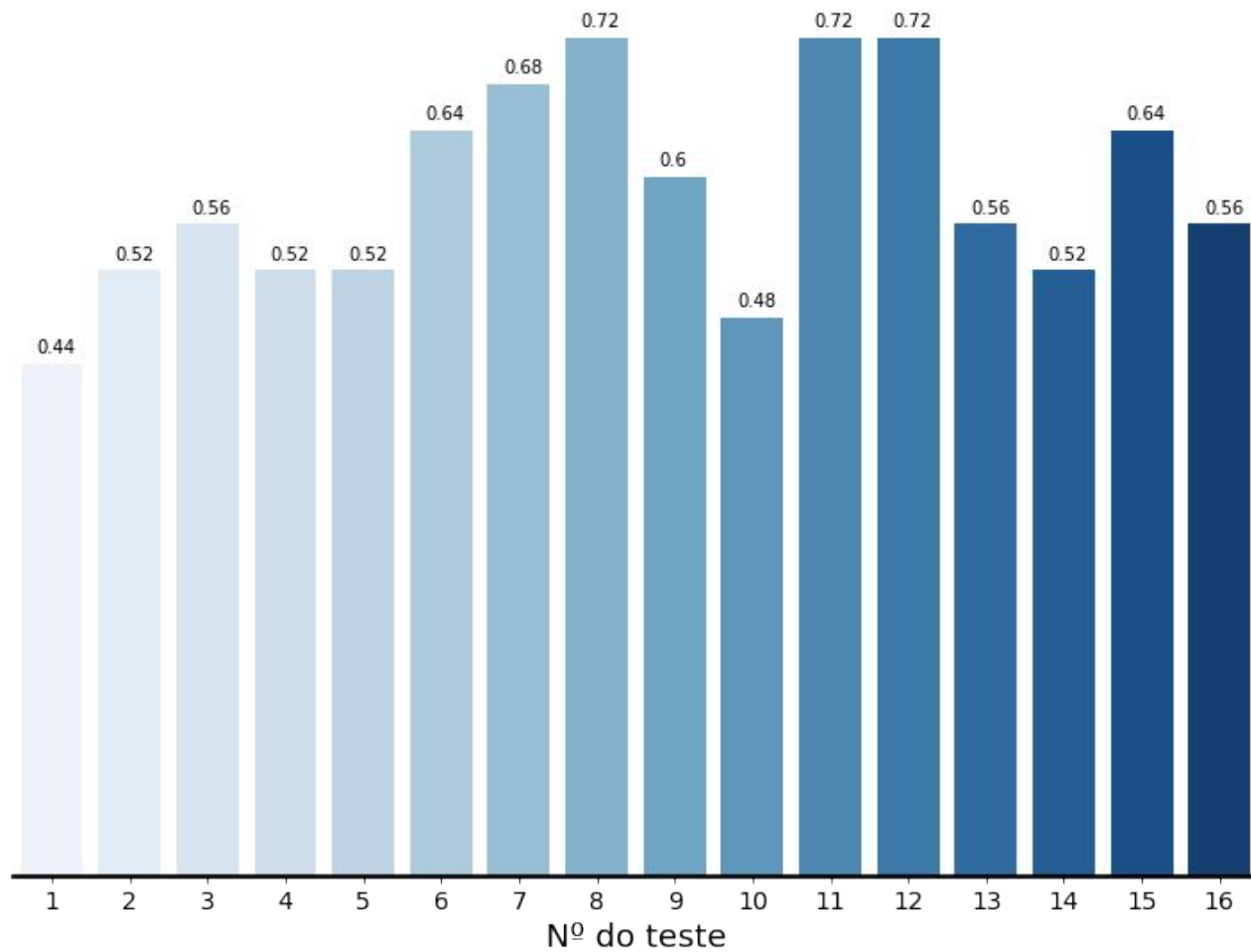
Classificação do áudio de saída (banco de dados): agudo

Classificação do áudio (classificador automático): médio

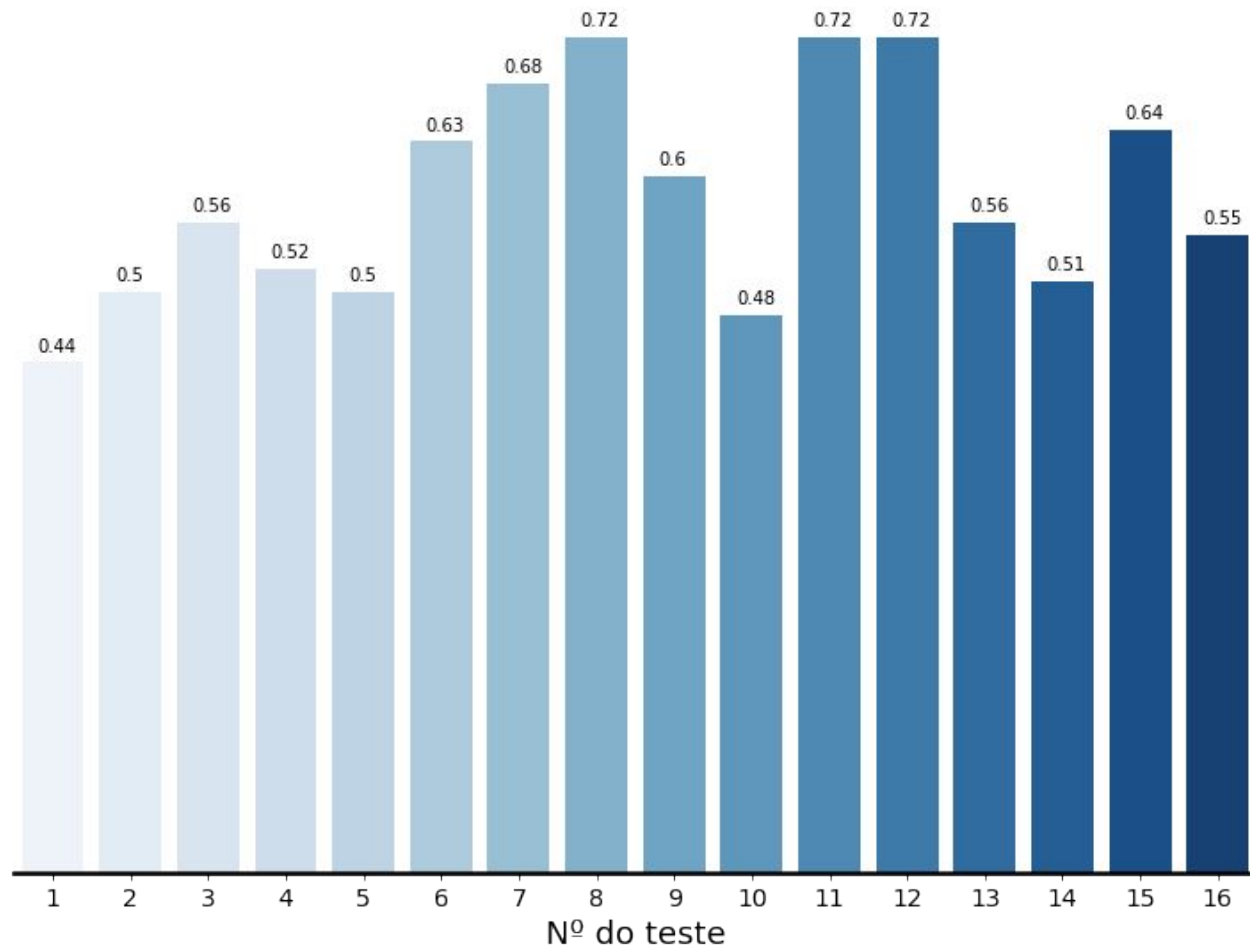


Resumo

Acurácia de f1-score em cada teste



Macro média (macro average) em cada teste



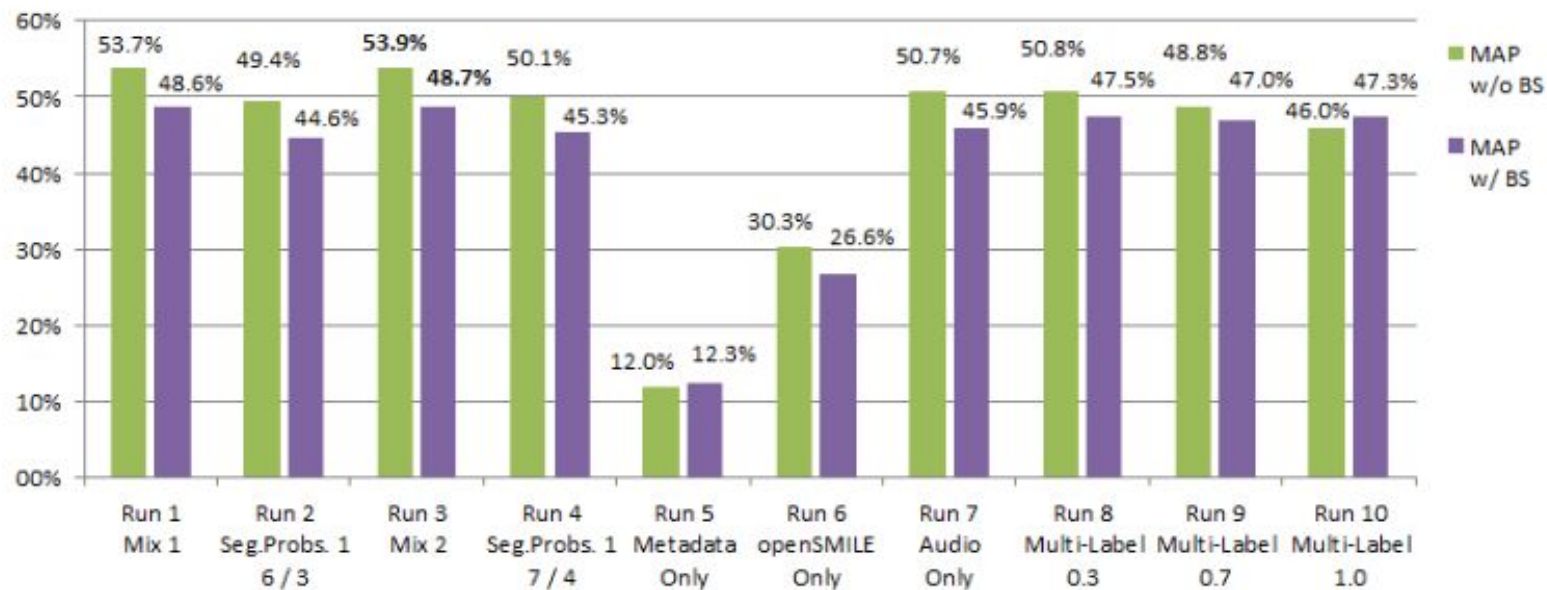


Fig. 4. Mean Average Precision (MAP) of Runs

Imagem extraída do artigo “Large-scale identification of birds in audio recordings”, de Mario Lasseck^[1]

Conclusões