UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA APLICADA

Descritores de frequência para a classificação de instrumentos musicais

por

Marco Cantergi

Dissertação submetida como requisito parcial
para a obtenção do grau de

Mestre em Matemática Aplicada

Prof. Dr. Fábio Souto De Azevedo
Orientador

Prof. Dr. Rodrigo Schramm
Coorientador

Porto Alegre, agosto de 2019.

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Cantergi, Marco

Descritores de frequência para a classificação de instrumentos musicais / Marco Cantergi.—Porto Alegre: PPGMAp da UFRGS, 2019.

46 p.: il.

Dissertação (mestrado) —Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Matemática Aplicada, Porto Alegre, 2019.
Orientador: De Azevedo, Fábio Souto; Coorientador: Schramm, Rodrigo

Dissertação: Matemática Aplicada,
Features, Timbre, Identificação de instrumentos musicais

Descritores de frequência para a classificação de instrumentos musicais

por

Marco Cantergi

Dissertação submetida ao Programa de Pós-Graduação em Matemática Aplicada do Instituto de Matemática da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do grau de

Mestre em Matemática Aplicada

Orientador: Prof. Dr. Fábio Souto De Azevedo

Coorientador: Prof. Dr. Rodrigo Schramm

Banca examinadora:

Prof. Dr. Flávio Luiz Schiavoni
PPGCC-UFSJ

Prof. Dr. Eduardo Horta
IME-UFRGS

Prof. Dr. Marcelo de Oliveira Johann
INF-UFRGS

Dissertação apresentada e aprovada em agosto de 2019.

Prof. Dr. Esequia Sauter
Coordenador

"The knower of the mistery of sound
knows the mistery of the whole universe."
— AZRAT INAYAT KHAN

AGRADECIMENTO

Agradeço ao professor Dr.Fábio Souto de Azevedo, o qual me permitiu desenvolver um tema na minha dissertação de mestrado que difere um pouco das linhas de investigação tradicionais do Departamento de Matemática Aplicada da Universidade.

Agradeço ao professor Dr. Rodrigo Schramm, que sempre esteve disponível para me ajudar e incentivar a explorar esse campo de estudo de intersecção entre matemática, computação e música. Seu conhecimento e sua *expertise* foram extremamente valiosos para o meu entendimento do assunto.

Sumário

Lista de Figuras

Lista de Tabelas

RESUMO

Diferentemente da altura e intensidade, a caracterização do timbre como qualidade sonora não é bem delimitada; todavia, é essa qualidade que abarca os elementos que nos permitem identificar a natureza de um som: sabemos, de algum modo, diferenciar uma flauta de um violino sem olharmos para o instrumento. Na realidade da rede mundial de computadores e de uma quantidade quase que infinita de músicas disponíveis, no entanto, temos a possibilidade de treinar sistemas automatizados para buscar músicas as quais contenham os instrumentos cujo som desejamos ouvir; nesse sentido, intenta-se encontrar as variáveis *features* (descritores) que melhor os identifiquem. Treze descritores do domínio de frequências, a partir da Transformada de Fourier, foram testados diante de bases de dados de notas isoladas de dezoito instrumentos musicais, assim como de trechos musicais mais complexos de piano, violino, clarinete, violão, fagote e flauta, trechos esses tanto monofônicos como em *duo*. A importância das variáveis na tarefa de identificação de instrumentos musicais foi determinada a partir do classificador supervisionado *TreeBagger* implementado no *software* MATLAB, para diversas combinações destes instrumentos; obteve-se assim, para cada experimento um grau de acerto e, ao cabo de todos os experimentos, um *ranking* para os descritores.

Palavras-chave: *features*, timbre, identificação de instrumentos musicais.

ABSTRACT

Unlike pitch and intensity, the characterization of timbre as a sound quality is not well defined; nevertheless, it is this quality that embraces the elements that allow us to identify the nature of a sound: we certainly know how to differentiate a flute from a violin without looking at the instrument. In the realm of the World Wide Web and of an almost infinite amount of online available music, however, we are able to train automated systems to search for songs containing the instruments whose sound we wish to hear; in this sense, we intend to find the feature variables that best identify them. Following the application of the Fourier Transform, thirteen frequency domain features were tested against a database composed of isolated musical notes from eighteen musical instruments as well as more complex musical excerpts from piano, violin, clarinet, guitar, bassoon and flute, for both monophonic and duo snippets. The importance of variables in the task of identifying musical instruments was determined from the TreeBagger supervised classifier implemented in the MATLAB software for various combinations of these instruments; thus, for each experiment, a degree of correctness and a ranking for the features were obtained.

Keywords: Features, Timbre, Musical instrument identification

1 INTRODUÇÃO

Quando um som é executado em um instrumento musical, as variações de pressão produzidas no meio se propagam ao ouvido humano o qual conta com um sofisticado sistema de reconhecimento e classificação da fonte emissora da onda. Tal capacidade é em nós moldada por anos de experiência como ouvintes de diversas fontes sonoras, de forma que conseguimos aprender a distingui-los razoavelmente, e melhor o fazemos quanto maior nossa exposição a diversas fontes musicais e *expertise*. Das qualidades do som, o timbre nos permite distinguir duas fontes sonoras de mesma frequência e intensidade. Segundo [2], *"Timbre exists at the confluence of the physical and the perceptual, and due to inconsistencies between these frames, it is notoriously hard to describe".*[1]

O desenvolvimento da capacidade computacional, por outro lado, vem permitindo que possamos desenvolver métodos automatizados de reconhecimento destes padrões instrumentais e construir aplicações no mundo real a partir dessa tecnologia. Tal tecnologia permite, dentre outros, a classificação e busca automática de determinados tipos de áudio na rede mundial de computadores.

A presente dissertação tem por objetivo definir os descritores de frequências mais robustos no reconhecimento e classificação do timbre de instrumentos musicais, qualidade sonora que lhes atribui justamente essa textura que nos permite reconhecê-los, como se fosse uma impressão digital. Estes instrumentos serão analisados isoladamente e em um conjunto duo polifônico. Inicialmente, se fará uma breve revisão teórica da análise um sinal no tempo como decomposição em suas frequências fundamentais (análise espectral) através da Transformada de Fourier. Como segunda etapa, será feita uma exposição dos descritores (*features)* de frequência mais empregados na classificação e reconhecimento instrumental automatizado na ten-

[1] O timbre existe na confluência entre o físico e sensorial e, devido a inconsistências entre *frames*, é reconhecidamente difícil de descrever (tradução do autor).

tativa de se estabelecer o estado da arte do problema. Em seguida, se seguirá a descrição do classificador *Bagged Trees* utilizado na dissertação.

Como fechamento, serão descritos os experimentos conduzidos no treinamento de uma base de dados e seus resultados de capacidade de reconhecimento instrumental perante a amostras de notas individuais de dezoito instrumentos e de certos trechos musicais sonoros mono e polifônicos. Será dada ênfase mormente aos descritores de frequência utilizados na classificação aos instrumentos de corda: piano, violino e violão e de sopro: clarinete e flauta, em seu som puro e em duo conjunto, justificada esta escolha pela sua presença importante em sonatas na literatura musical e sua consequente disponibilidade para treinamento.

A presente dissertação se encerra num campo de estudo bastante abrangente e atualmente em franca evolução conhecido como MIR *(Music Information Retrieval)*; uma ciência com interfaces na musicologia, psicoacústica, processamento de sinais, matemática, *machine learning*, dentre outros. Para mais informações, visite a *International Society for Music Information Retrieval* (https://www.ismir.net).

2 BASE TEÓRICA

A presente dissertação tem como fundamento teórico básico a decomposição espectral de um sinal (função) como combinação linear de seus componentes harmônicos. Esse é um processo de síntese que tem sua base no estudo da Série de Fourier.

Um sinal periódico real $f(t)$ pode ser representado como a soma discreta de senóides de diferentes amplitudes (A_k), frequências (k) e fases (ϕ_k). Inicialmente, para um sinal de período unitário, temos:

$$f(t) = \sum_{k=1}^{N} A_k \sin(2\pi kt + \phi_k) = \sum_{k=1}^{N} a_k \sin(2\pi kt) + b_k \cos(2\pi kt)$$

Utilizando a relação de Euler, podemos ainda escrever:

$$f(t) = \sum_{k=-n}^{n} C_k \mathrm{e}^{2\pi ikt}$$

onde C_k são complexos com propriedades de simetria, de forma que essa soma seja real:

$$C_{-k} = \bar{C}_k$$

Supondo que possamos de fato escrever esta soma, o coeficiente C_m pode ser isolado:

$$C_m = f(t).\mathrm{e}^{-2\pi imt} - \sum_{k=-n}^{n} C_k \mathrm{e}^{2\pi i(k-m)t}, (k \neq m).$$

Aplicando-se a integração num intervalo de período unitário, temos:

$$\begin{aligned}\int_0^1 C_m dt = C_m &= \int_0^1 f(t)\mathrm{e}^{-2\pi ikt}dt - \int_0^1 \sum_{k\neq m} C_k \mathrm{e}^{2\pi i(k-m)t}dt \\ &= \int_0^1 f(t)\mathrm{e}^{-2\pi ikt}dt - \sum_{k\neq m} C_k \int_0^1 \mathrm{e}^{2\pi i(k-m)t}dt \\ &= \int_0^1 f(t)\mathrm{e}^{-2\pi ikt}dt - \sum_{k\neq m} \frac{C_k}{2\pi i(k-m)}(\mathrm{e}^{2\pi i(k-m)} - 1) = \int_0^1 f(t)\mathrm{e}^{-2\pi ikt}dt.\end{aligned}$$

Denominando este coeficiente como $\hat{f}(k)$, temos que:

$$f(t) = \sum_{k=-n}^{n} \hat{f}(k)\mathrm{e}^{2\pi ikt}$$

onde

$$\hat{f}(k) = \int_0^1 f(t)\mathrm{e}^{-2\pi ikt}.$$

O problema desta definição é que esta soma não pode ser escrita para todo f, já que a soma é contínua e infinitamente diferenciável (é a soma de funções contínuas e infinitamente diferenciáveis). Uma onda triangular, por exemplo, tem pontos não diferenciáveis evidentes. O que ocorre é que precisamos de altas frequências para formar esses cantos, o que nos leva a definir a série de Fourier do sinal como:

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{2\pi ikt} \qquad \text{(série de Fourier)}$$

É possível provar a existência dos coeficientes $\hat{f}(k)$ caso a função seja tal que

$$f \in L^2([(0,1)]) \Longleftrightarrow \int_0^1 |f(t)|^2 dt < \infty$$

o que é comumente referido na literatura como um sinal de energia finita. Nesse contexto, a identidade de Rayleigh (Parseval) (abaixo) evidencia que a soma quadrática

dos coeficientes de Fourier vai justamente fornecer essa energia total do sinal, já que as exponenciais complexas são bases ortonormais (para as quais os coeficientes são projeções ortogonais). Essa relação de energia é fundamental, pois indica que podemos calcular a energia de um sinal no domínio do tempo ou da frequência.

$$\int_0^1 |f(t)|^2 dt = \sum_{k=-\infty}^{\infty} |\hat{f}(k)|^2 \qquad \text{(Parseval)}$$

onde

$$f(t) = \sum_{k=-\infty}^{\infty} \langle\, f(t), \mathrm{e}^{2\pi ikt} \rangle\ \mathrm{e}^{2\pi ikt}$$

Ainda, verifica-se que a convergência da série se dá no sentido tal que (*mean square convergence*):

$$\int_0^1 \left| \sum_{k=-n}^{n} \hat{f}(k)\mathrm{e}^{2\pi ikt} dt - f(t) \right|^2 \longrightarrow 0, n \longrightarrow \infty$$

De forma geral, tomando agora T como período, teremos a série escrita como:

$$f(t) = \sum_{k=-\infty}^{\infty} \hat{f}(k)\mathrm{e}^{2\pi ikt/T}$$

em que o coeficiente agora será dado por:

$$\hat{f}(k) = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\mathrm{e}^{-2\pi ikt/T}$$

Fazendo a passagem de uma função periódica para não-periódica através de um processo de limite da série quando o período T tende ao infinito, chegaremos à definição da transformada de Fourier do sinal $f(t)$ como generalização dos

coeficientes de Fourier:

$$\mathscr{F}(\nu) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-2\pi i\nu t}dt \qquad \text{(Transformada de Fourier)}$$

A transformada de Fourier, em suas aplicações em sinais, pode ser entendida como uma função que permite transportar um função do domínio do tempo para o domínio das frequências. À representação gráfica do plano da intensidade ou energia (módulo do número complexo, pois os coeficientes são números imaginários) em função das frequências dá-se o nome de **espectro ou envelope espectral**.

É bastante elucidativo tomarmos inicialmente o caso mais simples de uma nota musical isolada. Pela própria definição da série de Fourier, as frequências geradas por essa transformação, chamadas de **harmônicos ou parciais**, serão sempre múltiplos inteiros da frequência fundamental do sinal (ν_0). Uma nota musical A4 (notação de cifras), por exemplo, possuirá frequência fundamental de 440 Hz independentemente do instrumento musical fonte; entretanto, a mesma gerará uma distribuição linear de harmônicos (440 Hz, 880 Hz, 1320 Hz,...) distinta, com intensidades próprias do instrumento gerador. A Figura 2.1 ilustra essa situação para uma mesma notas em dois instrumentos diferentes, evidenciando as diferenças de amplitudes (energia) nos mesmos harmônicos. Nesse exemplo, como temos a mesma frequência fundamental para ambos instrumentos e os n parciais/harmônicos aparecem nas mesmas posições, poderiam ser tratados como vetores de n dimensões, e então o problema de classificação seria proposto a partir dos cossenos diretores. No entanto, aumentando um pouco a complexidade ao tocarmos um acorde de dó maior (3 notas) de violão e violoncelo, como ilustra a Figura 2.2, a situação se torna mais difícil de distinguir, e assim precisamos de descritores mais robustos. Quando adicionamos mais instrumentos, claro que o problema se tornará bastante mais complexo. Justamente estas distribuições de energia, consequências do fenômeno de ressonância da onda sonora na constituição física do instrumento (e.g.formato, matéria-prima),

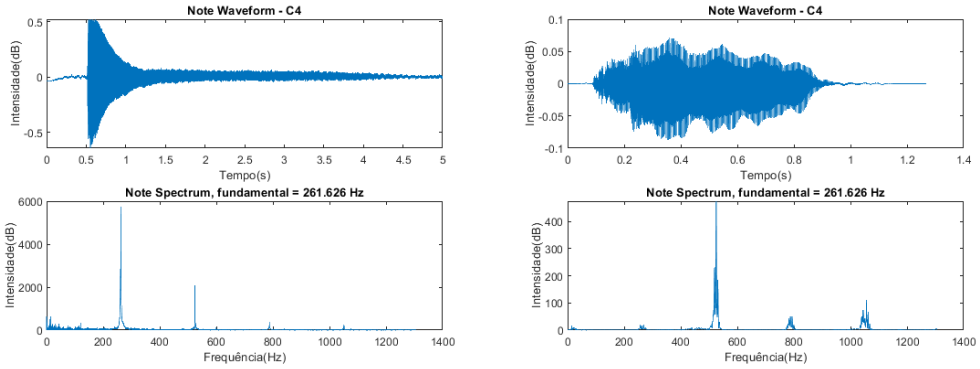


Figura 2.1: Envelopes temporais e espectrais da Nota Dó de frequência fundamental 262 Hz para piano(esquerda) e violino.

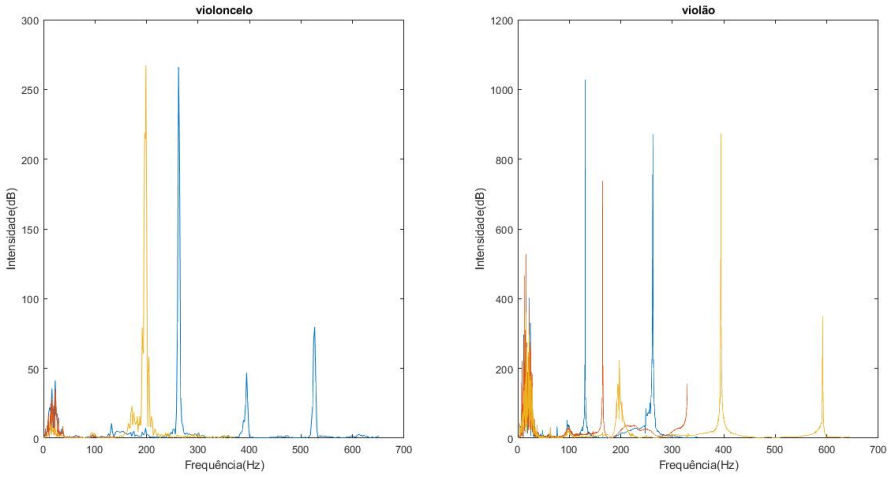


Figura 2.2: Espectro de acordes de Do maior de Violoncelo e Violão

são objetos de modelagem para fins de identificação do instrumento musical que originou o sinal.

O processo de quantização e discretização por amostragem é certamente necessário para a análise computacional de um sinal contínuo, de modo que como resultado tenhamos vetores que representem o sinal num espaço vetorial. Comumente vemos definido na literatura matemática o espaço vetorial $\mathbb{S}$ dos sinais em tempo discreto; um sinal nesse espaço é uma função definida apenas em $\mathbb{Z}$ e visualizada como uma sequência numérica x_n ou $x[n]$. Uma ilustração dessa passagem

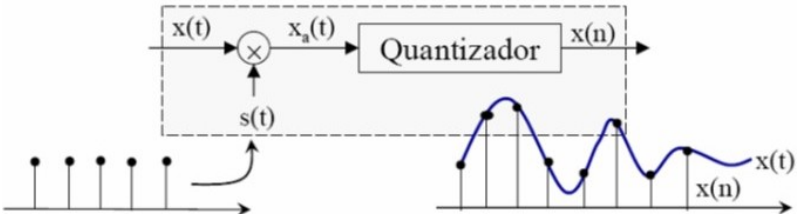


Figura 2.3: Processo de quantização e discretização do sinal

se encontra na Figura 2.3, mas os detalhes fogem ao escopo do trabalho e não serão abordados.

Neste contexto, a Transformada Discreta de Fourier (DFT) foi utilizada para o tratamento automatizado finito dos dados; o software Matlab, para fins de eficiência, utiliza como algoritmo a *Fast Fourier Transform* (FFT) para o cálculo da DFT, como segue:

$$X[k] = \sum_{n=0}^{N-1} x[n] \mathrm{e}^{-2\pi i n k/N}$$

onde $x[n]$ representa um bloco de sinal discreto para a amostragem n e para o k-ésimo *bin* de frequência, a partir de uma frequência de amostragem f_s *(sampling frequency)* do sinal original. A frequência correspondente do *bin* k é dada por:

$$f(k) = k.f_s/N$$

Uma propriedade importante da transformada de Fourier é de que o sinal pode ser reconstruído a partir dos coeficientes $\hat{f}(k)$: basta que sejam sobrepostos os sinusóides em todas as suas possíveis frequências ponderadas pelos respectivos coeficientes e fases. A ideia é que as duas representações, tanto temporais quanto espectrais, contenham a mesma quantidade de informação. No entanto, temos que garantir que, no caso de amostragens de sinais contínuos, a recomposição do sinal original seja viável; pelo teorema de Nyquist-Shannon, tal reconstrução é possível

se f_s for superior ao dobro da máxima frequência do sinal (frequência de Nyquist). Tipicamente e também nesse trabalho, a amostragem foi feita com f_s=44,1 KHz, o que corresponde a uma frequência de Nyquist de 22,05 KHz.

Apesar da DFT ser bastante útil, ela encripta o elemento temporal no seu conteúdo de fase do espectro. No sentido de melhor visualizar esse elemento, um procedimento bastante comum, denominado *Short-time Fourier Transform (STFT)*, provê uma matriz que descreve como as frequências do sinal evoluem ao longo do tempo. Computacionalmente, esse método corresponde a tomar segmentos do sinal de áudio (*frames*) ao longo do tempo e aplicar a FFT em cada um deles; certamente, de acordo com a velocidade com que a janela se deslocará *(hopsize)* em relação ao próprio tamanho do *frames*, poderemos ter uma superposição de dados (o que reduz efeitos de fronteira).

Usualmente, quando aplicamos os procedimentos STFT/FFT, os *frames* passam por um segundo processo denominado janelamento *(windowing)* em que o sinal é multiplicado por uma função suavizadora $w(t)$ que seja não-nula apenas da região do frame em análise; a *Hanning Window* é umas das funções mais utilizadas para tal fim. Pelas propriedades da Transformada, uma multiplicação no domínio temporal corresponde a uma convolução do espectro do sinal com o espectro da janela. No entanto, por padrão, o procedimento do FFT do Matlab não faz janelamento, e logo podemos assumir $w[n]$ simplesmente como uma função retangular (às vezes referida como função característica).

O processo STFT é resumido pela equação que segue:

$$X[k,m] = \sum_{n=-\infty}^{\infty} w[n - m.R]x[n]\mathrm{e}^{-2\pi ink/N}, 1 \leq k \leq N-1.$$

onde R representa o *hopsize*. Nesse trabalho, o tamanho da janela *(winsize)* foi definido como 1024, e *hopsize* como 4096.

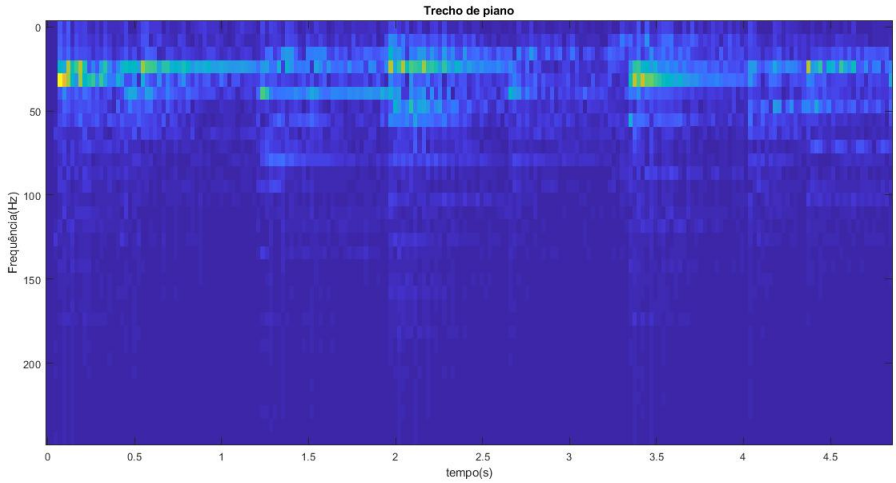


Figura 2.4: Espectrograma de um trecho da Sonata n.5 em Fá menor de J.S.Bach

A magnitude da STFT (espectrograma) de um trecho da Sonata n.5 em Fá menor, de J.S.Bach, é representada na Figura 2.4, sendo as cores mais leves/claras aquelas de maior magnitude (energia).

3 DESCRITORES (*FEATURES*) DE FREQUÊNCIA

Aqui pretende-se caracterizar os descritores utilizados no domínio de frequência utilizados no trabalho, em consonância com o utilizado na literatura revisada posteriormente:

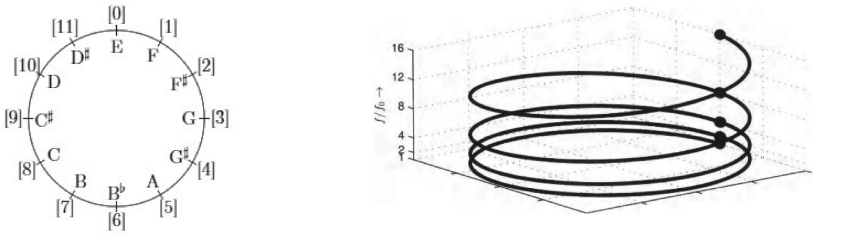
Mel Frequency Cepstral Coefficients (MFCCs): O MFC (Mel frequency cepstrum) é um descritor compacto da forma do envelope espectral, composto de coeficientes MFCCs, os quais são calculados através da *Discrete Cosine Transform (DCT)*, ou seja, tais coeficientes correspondem à parte real de uma Transformada de Fourier, como segue:

$$v^{j}_{MFCC}(m) = \sum_{k=1}^{\kappa} \log\left(|X^{*}[k,m]|\right) \cdot \cos\left(j \cdot \left(k - \frac{1}{2}\right)\frac{\pi}{\kappa}\right),$$

Esta Transformada é aplicada a partir do logaritmo da magnitude do espectro conhecido como Mel, o qual, por sua vez, corresponde a um mapeamento das amplitudes do espectro de modo que as bandas de frequências sejam igualmente espaçadas. A escala Mel (Stevens, Volkmann, and Newman), acrônimo de *melody*, foi construída considerando o aspecto sensorial para que, ao nossos ouvidos, os intervalos de frequência sejam igualmente espaçados. Ocorre que, a partir de 500 Hz, julgamos intervalos cada vez maiores como igualmente espaçados; assim, na escala Mel há uma modelagem logarítmica para o fenômeno como consequência da resposta do nosso próprio sistema de audição:

$$f^{*} = 2595 \log\left(1 + \frac{f}{700}\right)$$

Por essa característica sensorial, muitas vezes este descritor é classificado como perceptual; no entanto, neste contexto, por se tratar também uma representação espec-



(a) círculo com 12 tonalidades de uma escala

(b) hélice de frequencias com as tonalidade como classes de equivalência

Figura 3.1: Representação gráfica do descritor *Pitch Chroma Vector*

tral, apenas aplicada sobre uma escala não-linear, aqui ela será apresentada como um descritor de frequência.

Segundo [3], os MFCCs tem sido largamento utilizados no campo de processamento de sinais de voz desde a sua introdução em 1980; ainda, mais relevante, comenta que no contexto de classificação de áudios, um subconjunto entre 4 e 20 coeficientes MFCCs usualmente já contém as informações principais para tal fim. No presente estudo, seguindo a própria formulação de A.Lerch, serão utilizados 13 coeficientes.

Pitch Chroma Vector ou Vetor Tonal (VT): descritor vetorial de doze dimensões que representa a magnitude da energia acumulada para cada uma das doze notas-base da representação ocidental da música (classes de equivalência de 0 a 11), como na Figura 3.1. A ideia é de não haver distinção no caso repetição de uma mesma nota em escala superior. Apesar deste descritor ser mais comumente usado na detecção da tonalidade (*key*) do áudio, eles será aqui utilizado por se basear puramente no domínio de frequências; no entanto, não se espera obter grande significância, a não ser no caso de notas isoladas.

Centroide Espectral (CEN): O conceito do centróide se assemelha ao de centro de gravidade (baricentro); trata-se, assim, da seguinte média ponderada:

$$v_{CEN}(m) = \frac{\sum_{k=0}^{\kappa/2-1} k \cdot |X(k,m)|^2}{\sum_{k=0}^{\kappa/2-1} |X(k,m)|^2}$$

Claro que esse resultado deve estar em algum *bin* entre 0 e $\kappa/2-1$ (o qual pode ser convertido para frequência). Na literatura, há vários indicativos de que esse descritor é um dos mais importantes para caracterização do timbre, o que aqui buscamos.

Fator espectral da crista (FEC): Esta medida compara a magnitude máxima do espectro com a soma de todas das magnitudes do espectro; pela sua construção, o resultado estará sempre entre $2/\kappa$ e 1:

$$v_{FEC}(m) = \frac{\max_{0\geq k\geq \kappa/2-1} |X(k,m)|}{\sum_{k=0}^{\kappa/2-1} |X(k,m)|}$$

Resultados baixos do FEC indicam que há uma distribuição mais homogênea de energia entre os *frames*, de modo que há uma tendência da distribuição ser mais uniforme, enquanto que resultados próximos da unidade reforçam a existência de frames com energia dominante ou de picos de energia no espectro, o que aponta mais para uma distribuição tipo sinusoide.

Decrescimento Espectral (DE): este descritor estima a velocidade de decrescimento (taxa média de variação) do envelope espectral em relação à frequência inicial zero, como segue:

$$v_{DE}(m) = \frac{\sum_{k=1}^{\kappa/2-1} \frac{1}{k}.(|X(k,m)| - |X(0,m)|)}{\sum_{k=1}^{\kappa/2-1} |X(k,m)|}$$

Em função da normalização deste estimador, temos que $v_{DE}(m) \leq 1$.

Spectral Flatness (SF) (tonality coefficient, Wiener entropy): razão entre as médias geométrica e aritmética da magnitude do espectro:

$$v_{SF}(m) = \frac{\kappa}{2} \cdot \frac{(\prod_{k=0}^{\kappa/2-1} |X(k,m)|)^{2/k}}{\sum_{k=0}^{\kappa/2-1} |X(k,m)|}$$

Nessa formulação, resultados próximos de zero indicam uma distribuição mais plana, e aqueles maiores apontam para a direção contrária, de um espectro de mais alto ruido.

Fluxo Espectral (FE): medida da variação média entre frames consecutivos; assim, intenta medir a variação da forma do envelope espectral.

$$v_{FE}(m) = \frac{2}{\kappa}. \sum_{k=0}^{\kappa/2-1} (|X(k,m)| - |X(k,m-1)|)^2$$

De maneira informal, relaciona-se este descritor com a sensação de textura "aspereza" *(roughness)* do áudio.

Curtose Espectral (CE): a curtose é um momento de quarta ordem transladado à média, ligado usualmente ao achatamento de uma distribuição probabilística. No caso das magnitudes do espectro, temos:

$$v_{CE}(m) = \frac{2}{\kappa.\sigma^4_{|X|}}. \sum_{k=0}^{\kappa/2-1} (|X(k,m)| - \mu(|X|))^4 - 3$$

Rolloff Espectral (RE): esse descritor procura o bin abaixo do qual as magnitudes acumuladas atingem certo percentual determinado, comumente 85% ou 95%. No presente trabalho utilizou-se 85%. Claro, esse bin pode novamente ser convertido para Hertz.

Assimetria Espectral (AE)(skewness): mede a simetria em relação à média aritmética, relacionada assim ao momento de terceira ordem, como segue:

$$v_{AE}(m) = \frac{2}{\kappa.\sigma^3_{|X|}}. \sum_{k=0}^{\kappa/2-1} (|X(k,m)| - \mu(|X|))^3$$

Declividade Espectral (DE): mede o coeficiente angular da reta estimada por uma regressão linear modelada a partir do espectro, assim teremos:

$$v_{DE}(m) = \frac{\sum_{k=0}^{\kappa/2-1}(k-\mu_k)(|X(k,m)|-\mu_{|X|})}{\sum_{k=0}^{\kappa/2-1}(k-\mu_k)^2}$$

Espalhamento (spread) Espectral (EE):descreve a concentração de energia em torno do centróide. Segundo Lerch, há indicativos na literatura de que esse descritor tenha relevância na percepção do timbre, fato a ser aqui investigado.

$$v_{EE}(m) = \sqrt{\left(\frac{\sum_{k=0}^{\kappa/2-1}(k-v_{CEN}(m))^2|X(k,m)|^2}{\sum_{k=0}^{\kappa/2-1}|X(k,m)|^2}\right)}$$

Tonal Power Ratio (TPR): Estimativa da energia relativa dos componentes tonais, em oposição a ruídos: medidas da qualidade do som. Pela sua construção, naturalmente esse descritor está sempre entre 0 e 1.

$$v_{TPR}(m) = \frac{E_T(m)}{\sum_{k=0}^{\kappa/2-1}|X(k,m)|^2}$$

Na tentativa de estabelecer o estado da arte no uso dos descritores de frequência da identificação/classificação de instrumentos musicais, a Tabela 3.1 foi construída.

Tabela 3.1: Uso de descritores de frequência em recentes trabalhos da área de MIR

Ref	**Descritores**	**Classificadores**	**Acurácia**
[4]	CEN, coeficientes de Cepstrum	Gaussiano e KNN*(k-nearest neighbors)*	80.6% para instrumentos individuais
[5]	razões odd-even e harmônicas, CEN normalizado	Rede neural Feed Forward	87.5% para redes simples
[6]	PFCC e MFCC (*Pitch and Mel frequency Cepstral Coefficients*)	GMM e Random Forest	90-100% para notas individuais
[7]	*Cepstral features(Mel, primeira derivada)*, FE, CEN, *Harmonic spectral deviation, harmonic spectral spread, harmonic spectral variation*	PNMF (variação do NNMF*(non-negative matrix factorization)*	média de 87.9% para 11 instrumentos
[8]	MFCCs	Multilayer perceptron e PCA *(Principal Component Analysis)*	92-96%, para piano, violino e flauta
[9]	Autocorrelação no espectrograma	Não aplicável	Não aplicável
[10]	CEN, inarmonicidade(distância cumulativa entre os parciais estimados e seus valores teóricos),percentual de energia nos quatro primeiros parciais, bandwidth (variações entre parciais e o centróide), harmonic energy skewness.	Canonic Discriminant Analysis, Quadrant Discriminant Analysis, KNN (k=1,3,5,7) e *Support Vector Machine*	o melhor resultado foi 92.81% para 27 instrumentos solos usando Quadrant Discriminant Analysis
[11]	Pitch, Brightness(energia acima de 1500 Hz) e CEN	Linear Discriminant Analysis e Random Forest	melhor indice 82.1% usando Random Forest para 12 instrumentos individuais
[12]	RE, AE, CE, *Brightness, SF, Root Mean square Energy*, MFCC	Redes Neurais	89.17%
[13]	CEN, FE, spread espectral, AE, MFCC, delta MFCC	KNN e SVM(Support Vector Machine)	60.43% para KNN, 73.73% para SVM

[14]	MFCC, LPCC	Redes Neurais	70.15% para MFCC, 71.73% para LPCC
[15]	CEN, spectral brightness, AE, SF, FE,timbre zero cross, dynamic root mean square	Redes Neurais	melhor atingiu 83% para trompete
[16]	Transformada Q, coeficientes cepstrais, CEN	KNN	máxima acurácia de 87%
[17]	média, desvio-padrão, CEN	*Linear Distinction Analysis*, KNN, SVM e Random Forest	Melhor distinção entre Violino e Viola (mesma nota) e com Random Forest e agregando features temporais: 79.6%
[18]	MFCC, PLP(perceptual Linear Prediction), RASTA-PLP	HMM e KNN	MCC teve melhor desempenho na identificação de flauta, violão, piano e violino (85.5% para mesma nota, 75% para notas diferentes)

Os trabalhos citados são bastante heterogêneos em termos de técnica de classificação, base de dados e objetivos, de modo que a comparação entre eles e mesmo com o resultado do presente trabalho tem que ser feita com muito cuidado: temos, por exemplo, que [17] se utiliza de diversos métodos de classificação (análise discriminante, KNN, SVM e *Random Forest*) e apenas dois instrumentos (violino e viola) com uma nota apenas, enquanto [8] fez uso de redes neurais *(Multilayer perceptron)* para trechos de piano, violino e flauta. Essa variabilidade seria de se esperar dada a vastidão do *MIR* como campo de estudo; mesmo assim, procurou-se selecionar os descritores que frequentemente foram utilizados nas referências.

4 CLASSIFICADOR

De acordo com [19], em problemas de classificação, é sabido que cada caso (entrada) corresponde a uma única classe dentre um número finito de classes e, dados um conjunto de variáveis (descritores), deseja-se predizer corretamente a qual classe este pertence. Assim, um classificador é uma regra que atribui uma classe a um determinado conjunto de medidas $\boldsymbol{x} = (x_1, x_2, .., x_M)$, sendo o espaço X o conjunto de todas as possibilidades destas medidas; seja ainda $C = \{c_1, c_2, ..., c_J\}$ o espaço de classes possíveis, o classificador pode ser entendido como uma função $X \rightarrow C$. De outro modo, o classificador corresponde a uma partição de X em conjuntos disjuntos $B_1, ..., B_J$, de modo que uma entrada particular será atribuída a c_j se o vetor $\boldsymbol{x} \in B_j$. Normalmente, é desejável que utilizemos a experiência de dados já conhecidos; como esperado, as amostras de instrumentos musicais já classificadas *a priori* na base de dados do estudo servirão justamente para que o classificador aprenda a fazer novas previsões.

Segundo [19], classificadores estruturados em árvores se constituem em métodos não-paramétricos computacionalmente extensivos e de recente popularidade, especialmente no contexto de *data mining*, podendo ser utilizados para uma enorme quantidade de casos (entradas) e de variáveis.

De acordo com [1], uma árvore de decisão/classificação é um processo de estágios múltiplos, em que uma decisão binária deve ser tomada a cada estágio. A ideia de um classificador binário é dividir o espaço em partições cada vez mais refinadas em termos de classificação: ou seja, queremos que a maior parte dos membros da classe c_j efetivamente pertençam a esta classe no classificador.

Cada árvore é formada por ramos e nós, sendo estes últimos ainda classificados em internos (se originam descendentes) ou terminais. A cada nó terminal é associada uma classe, e as entradas que porventura recaiam sobre ele serão associ-

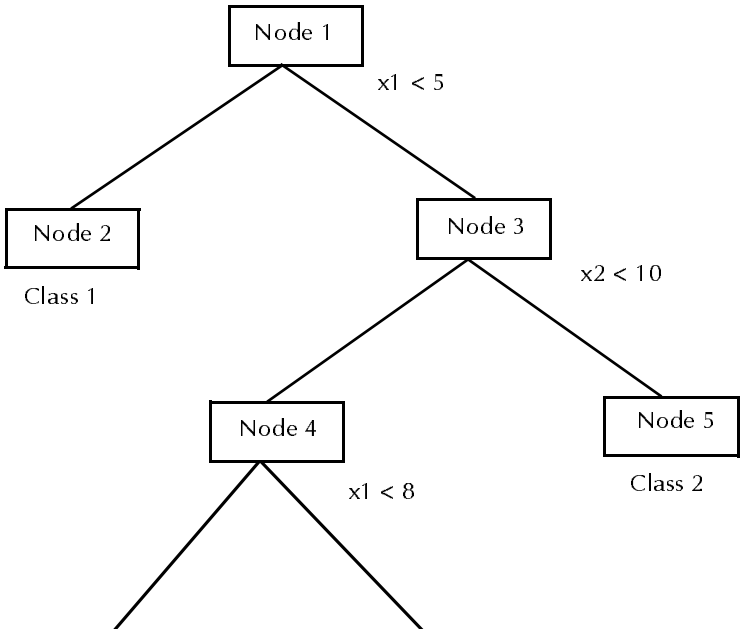


Figura 4.1: Exemplo de árvore simples com duas variáveis, x_1 e x_2; referência [1].

adas a essa classe, como evidencia a Figura 4.1 em que os nodos 2 e 5 são terminais e delimitam duas classes.

Temos certamente de ter critérios para o crescimento das árvores no que se refere à maneira como a divisão em cada nó é feita e ao momento de parar seu desenvolvimento. Para tanto, temos que estudar com mais profundidade a implementação deste algoritmo no software MATLAB, aqui utilizado para todos os processamentos. Segundo [1], quando temos uma divisão em um nodo, o objetivo será encontrar uma partição que de algum modo reduza a "impureza" local; seja o nodo t, teremos uma medida de impureza $\rho(t)$ para este nodo t. Implementada por *Breiman et al. [1984]*, uma dessas medidas (a qual é utilizada na implementação do algoritmo do software) é denominada **Índice de diversidade de Gini**.

$$\rho(t) = \sum_{i \neq j} p(c_i|t).p(c_j|t) = 1 - \sum_{j=1}^{J} p^2(c_j|t) \qquad \text{(Índice de diversidade de Gini)}$$

onde $p(c_j|t)$ denota a probabilidade de que, dado um nó t, uma observação pertença à classe c_j e é calculada como:

$$p(c_j|t) = \frac{p(c_j, t)}{p(t)},$$

sendo $p(c_j, t)$ a probabilidade conjunta de pertencer à classe j e estar no nodo t e $p(t)$ a probabilidade de que a observação (entrada) esteja no nodo t; estas, por sua vez, são dadas por, respectivamente:

$$p(t) = \sum_{j=1}^{J} p(c_j, t)$$

e

$$p(c_j, t) = \frac{\hat{\pi}_j n_j(t)}{n_j}$$

O estimador $\hat{\pi}_j$ é computado *a priori* diretamente da base de dados de treinamento D pela razão entre o número de observações n_j que pertencem à classe c_j e o número total de observações n da base D.

Ainda resta a questão do momento dessa medida de impureza. Para um nodo determinado, temos de encontrar o melhor descritor dentre M possibilidades para a separação nos ramos esquerdo e direto; limita-se o problema pela seguinte convenção: para todos os vetores (entradas) da nossa amostra, procuramos a melhor separação para nosso k-ésimo descritor propondo como separatriz a média entre cada dois valores consecutivos para esse descritor. Neste momento então que medimos o **Índice de diversidade de Gini** e decidimos por aquele descritor (associado a uma determinada média) que resulta em um decrescimento mais elevado do Índice.

Segundo [1], de forma estruturada, seguem as etapas para o crescimento de uma árvore:

1. Determinação do número máximo de observações permitido no nodo terminal;

2. Determinação das probabilidades $\hat{\pi}_j$ *a priori* de que uma entrada pertença a cada classe c_j, estimada a partir do *dataset* D;

3. Se um nodo terminal da árvore contém mais do que o máximo permitido de observações com observações de várias classes, então procura-se a melhor separação; para cada descritor k, temos:

   (a) Colocação em ordem ascendente dos dados de entrada para cada descritor (variável) no nodo em análise, de modo a termos os valores ordenados $x_{(i)k}$;

   (b) Determinação de todos as separatrizes S_k para o descritor k através de:

   $$S_{(i)k} = (x_{(i)k} + (x_{(i)k} - x_{(i+1)k})/2$$

   (c) Para cada separação proposta, avaliação da impureza $\rho(t)$;

   (d) Escolha daquela que resulta em um maior decrescimento de impurezas.

4. Dos k descritores, separação do nodo por aquele que melhor resultado geral tem na separação do item 3;

5. Para a separação proposta no item anterior (4), determinação de quais observações irão para os ramos esquerdo e direto;

6. Repetição dos passos 3 a 5 até que a regra de parada seja atendida (tenha observações de apenas uma classe e número máximo permitido de entradas naquele nodo).

No presente estudo, utilizou-se um classificador *TreeBagger* (ou *Bagged Trees*), o qual equipa o classificador em árvores de decisão já descrito com um conjunto de algoritmos do tipo *Bagging* (*Bootstrap aggregating*).

Proposto inicialmente por *Leo Breiman* em 1996, os algoritmos do tipo *Bagging* têm por objetivo, através da redução de variabilidade, aumentar a estabilidade e acurácia de diversos métodos de classificação e aprendizado. A sua ideia principal é bastante simples: tomar diversas amostras (subconjuntos) de mesmo tamanho da base de dados original (*dataset*), com possibilidade de repetição de entradas (e de que, inclusive, certos dados simplesmente não sejam selecionados), e em cada uma delas aplicar um método de classificação associado. Ao final, os resultados são combinados através de uma média de resultados individuais (para fins de regressão) ou de simples votação dos resultados para fins de classificação; a Figura 4.2 ilustra o processo.

O motivo de algoritmos do tipo *bagging* reduzirem a variabilidade pode ser explicado pela simples demonstração abaixo: seja $\phi(x)$ o preditor para a entrada x de um método de classificação particular e $\mu(x) = E(\phi(x))$ o valor esperado da distribuição subjacente às amostras de treinamento. Então tem-se:

$$\begin{aligned} E(Y_x - \phi(x))^2 &= E(Y_x - \mu(x)) + (\mu(x) - \phi(x))^2 \\ &= E(Y_x - \mu(x)^2 + 2E[(Y_x - \mu(x).(\mu(x) - \phi(x))] + E(\mu(x) - \phi(x))^2 \\ &= E((Y_x - \mu(x))^2 + E(\mu(x) - \phi(x))^2 \\ &= E((Y_x - \mu(x))^2 + Var(\phi(x)) \\ &\geq E((Y_x - \mu(x))^2 \end{aligned}$$

Assim, se pudermos usar $\mu(x) = E(\phi(x))$ como preditor, sua variabilidade será igual ou inferior à de $\phi(x)$.

Apesar do *Bagging* poder ser utilizado com outros métodos de classificação e regressão, como de Redes Neurais, ele é comumente associado a métodos

de árvores (*Trees*), como aqui empregado. Outro algoritmo bastante utilizado, semelhante ao *Bagged Trees*, é o chamado *Random Forests*; a diferença fundamental entre estes consiste em que, o último, além de trabalhar com subconjuntos amostrais da base de dados original, conta apenas um subconjunto $m < M$ de descritores os quais são aleatoriamente selecionados na divisão de um nó (nodo) da árvore; para o primeiro, todos os descritores são considerados para a divisão em cada nodo.

Segundo [20], são os seguintes os passos básicos que os algoritmos do tipo *Bagging* empregam no contexto de classificação por árvores; para o treinamento, temos:

1. Inicialização do *dataset* de treinamento D.
2. Para cada iteração $i \in \{1, 2, \ldots, T\}$
   (a) Criação de um novo *dataset* D_i de mesmo tamanho —D— por amostragem aleatória com reposição de D.
   (b) Treinamento do classificador de árvores baseado no subconjunto Di;

Para o teste, simplesmente inicializamos todas as bases de treinamento e predizemos as classes pela combinação dos T modelos treinados por voto de maioria simples.

Finalmente, juntando os entendimentos sobre os conceitos das árvores de decisão e de *Bagging*, temos que o conjunto de algoritmos do *TreeBagger* cria diversas árvores binárias ao mesmo tempo, de modo que em cada nodo (nó) o classificador procura a variável que melhor separa os dados naquele nodo em particular.

O aspecto fundamental na escolha desse classificador para o objetivo de determinar a importância dos descritores do timbre de instrumentos musicais é a de que ele permite que consigamos determinar a importância relativa (votação) de

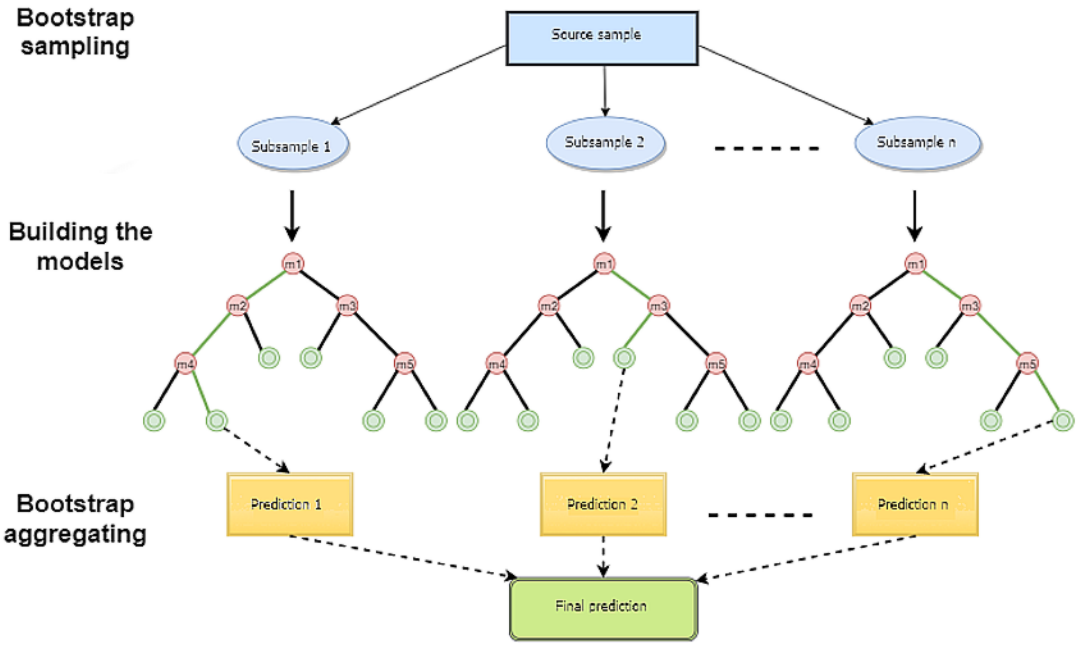


Figura 4.2: *Bootstrap aggregation*
Fonte: https://www.mql5.com/en/articles/3856

cada variável e a consequente criação de um *ranking* de sua importância. Segundo a documentação do software MATLAB, referido em [21], a importância das variáveis (descritores) é calculada somando as variações dos erros quadráticos médios (EQM) das separações de cada uma delas, dividida pelo número de nodos, incluindo ramos descendentes. Os EQMs são calculados pelo erro ponderado pela probabilidade associada a cada nodo.

No sentido de evitar o fenômeno de *overfitting*, trabalhou-se com um subconjunto de treinamento ainda menor através de um processo de *cross-validations* (validação cruzada), em que um dos subconjuntos é utilizado para teste e os demais para treinamento. De forma sintética, o processo de *k-fold cross validation* permite que dividamos aleatoriamente D em k grupos de aproximadamente mesmo tamanho; a primeira partição é separada para teste/validação e o modelo é treinado $k-1$ vezes, tendo sua performance medida através do grupo de teste; este procedimento é então repetido k vezes. Dessa forma, garantimos que, a cada iteração, cada vetor

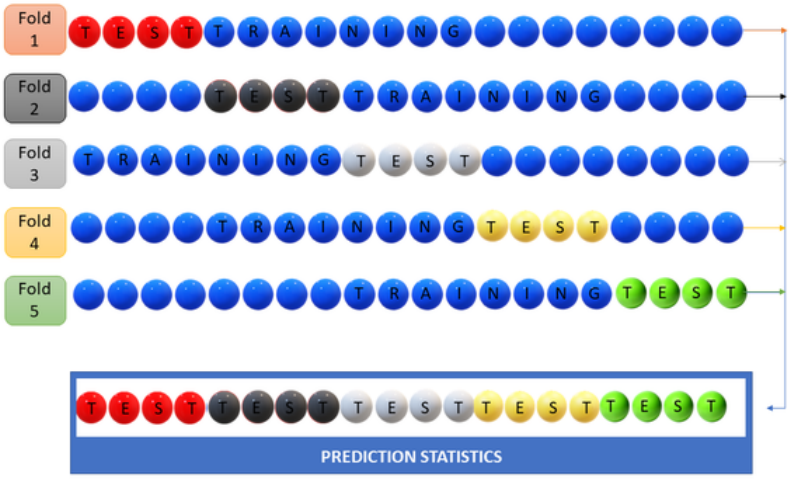


Figura 4.3: Uma iteração no processo *5-fold cross-validation*
Fonte:
https://www.datasciencecentral.com/profiles/blogs/cross-validation-in-one-picture

de entrada esteja num subconjunto de validação uma única vez, mas sirva como treinamento k-1 vezes, o que aumenta a efetividade e reduz o viés das estimativas, conforme ilustra a Figura 4.3.

5 BASE DE DADOS

A amostra do experimento inicial com notas individuais conta com 10.619 notas musicais distribuídas por dezoito instrumentos, coletadas através dos referidos *websites* [22] e [23]. Na Tabela 5.1 se encontra a distribuição de frequência dessas notas, juntamente com o número de dados de entrada *(frame inputs)* gerados pelo procedimento de STFT para cada instrumento que servirá como a base D de treinamento; essa informação é fundamental, e a correlação desta frequência com o percentual de identificação do instrumento será analisada nas considerações finais do trabalho. Como princípio empírico, qualquer tipo de treinamento (e não seria diferente para um treinamento automatizado) tende a ser mais eficiente quanto mais elevado é o número de informações disponíveis. É importante, no entanto, distinguirmos o conceito de informação do de dado: podemos ter *inputs* de dois vetores linearmente dependentes, são de fato dois conjuntos de dados, e no entanto não constituem informações no sentido de que não informam nada de novo ao processo de treinamento.

Nos experimentos subsequentes, foram utilizadas amostras de trechos musicais as quais foram adquiridas através de [24–28], distribuídos como na Tabela 5.2.

Ainda, na tarefa de identificação de instrumentos, é interessante termos uma classe que contenha os elementos musicais que não se encaixaram em nenhum dos critérios de separação das nossas árvores de decisão; no mundo real, muitas vezes queremos buscar na rede de computadores determinadas músicas de instrumentos (digamos, piano e flauta) e precisamos de uma classe que responda pela negativa desta procura e não simplesmente do encaixe em uma dualidade de subespaços em que tudo é piano ou flauta. A referida classe foi chamada de "NOT CLASS", e corresponde a uma base de dados anotada chamada *IRMAS*, apresentada no *13th International Society for Music Information Retrieval Conference (ISMIR 2012),*

Tabela 5.1: Distribuição amostral dos instrumentos musicais: notas individuais

Instrumento	Família	amostras	% amostras	frames	% frames
banjo	cordas	74	1	8.099	1
violoncelo	cordas	732	7	34.165	6
contrabaixo	cordas	817	8	49.082	8
violão	cordas	110	1	21.795	4
bandolim	cordas	80	1	8.878	2
piano	cordas	260	2	75.140	13
violino	cordas	1.074	10	43.848	8
clarinete baixo	sopro	854	8	28.290	5
fagote	sopro	640	6	24.734	4
clarinete	sopro	764	7	42.549	7
contrafagote	sopro	623	6	35.892	6
corne inglês	sopro	652	6	32.636	6
flauta	sopro	778	7	34.891	6
trompa	sopro	568	5	35.800	6
oboé	sopro	551	5	24.274	4
trombone	sopro	763	7	33.149	6
trompete	sopro	406	4	25.614	4
tuba	sopro	859	8	24.287	4
Total		10.605	100	583.123	100

Fonte: Autor.

Porto, Portugal,2012 ([29]) e muito utilizada em trabalhos no contexto do estudo de *MIR*; sua distribuição amostral encontra-se na Tabela 5.3.

Nos experimentos que se seguem, foram extraídas as STFT's, como já explicado anteriormente, de modo que todos os descritores foram calculados a partir do espectrograma do sinal sonoro.

Tabela 5.2: Distribuição amostral dos instrumentos musicais: trechos musicais, solos e duos

Instrumento(s)	amostras	% amostras	frames	% frames
violão	17	3	67.728	3
piano	240	45	888.434	44
violino	70	13	393.412	19
fagote	3	1	129.487	6
clarinete	7	1	75.918	4
flauta	3	1	14.717	1
piano e violino	149	28	235.953	12
piano e clarinete	34	6	106.230	5
violão e flauta	8	1	35.063	2
Total	535	100	2.030.592	100

Fonte: Autor.

Tabela 5.3: Distribuição amostral dos instrumentos musicais: base anotada IRMAS

Instrumento principal	amostras	% amostras	frames	% frames
violoncelo	388	6	48.888	6
clarinete	505	8	63.630	8
flauta	451	7	56.826	7
violão	637	10	80.262	10
guitarra	760	11	95.760	11
órgão	682	10	85.392	10
piano	721	11	90.846	11
saxofone	626	9	78.876	9
trompete	577	9	72.702	9
violino	580	9	73.080	9
voz	778	12	98.028	12
Total	6.705	100	844.290	100

Fonte: Autor.

6 EXPERIMENTOS E RESULTADOS

Em todos os experimentos, utilizou-se o classificador *TreeBagger* com 30 árvores e *5-fold cross-validation*; 36 variáveis abrangem a totalidade do espaço dos descritores já definidos. Além da matriz de confusão para cada um dos treinamentos, será apresentada a importância de cada descritor normalizada, destacando-se em cores aqueles cuja importância relativa seja superior a 70 %. Uma matriz de confusão $M(a, b)$ tem como seus elementos de formação o percentual de reconhecimento do elemento a como elemento b; assim, a diagonal principal dessa matriz representará o índice de acerto dos experimentos.

Para o descritor k de um dado experimento:

$$\text{Importância relativa}(k) = \text{Importância}(k)/\text{máx}(\text{Importância})$$

Como um primeiro experimento, a base de dados D contou com notas individuais dos 18 instrumentos com a distribuição anteriormente apresentada. A matriz de confusão respectiva está na Figura 6.1, tendo atingido uma média de acerto de 80,2 %.:

Aqui destaca-se um ponto já esperado (conforme Figuras 6.2 e 6.3), de que os coeficientes do descritor *Pitch Chroma* poderiam ter alguma relevância

	Banjo	Violoncelo	Contrabaixo	Violão	Bandolim	Piano	Violino	Clarinete baixo	Fagote	Clarinete	Contrafagote	Corne Inglês	Flauta	Trompa	Oboe	Trombone	Trompete	Tuba
Banjo	28,83%	1,52%	6,20%	20,76%	10,14%	0,04%	12,21%	0,42%	1,28%	0,77%	1,78%	3,95%	4,82%	3,03%	1,26%	1,86%	0,78%	0,37%
Violoncelo	0,17%	63,42%	6,64%	0,46%	0,11%	0,01%	5,54%	1,40%	0,67%	1,15%	6,42%	1,28%	2,12%	1,97%	1,40%	1,78%	0,94%	4,50%
Contrabaixo	0,45%	3,40%	77,85%	2,07%	0,24%	0,01%	2,47%	1,30%	0,23%	0,47%	3,61%	0,78%	0,55%	1,02%	0,19%	1,01%	0,29%	4,05%
Violão	2,82%	1,24%	5,63%	73,12%	3,24%	0,04%	1,63%	0,48%	0,26%	0,31%	4,19%	0,38%	2,49%	2,53%	0,07%	0,36%	0,02%	1,19%
Bandolim	9,57%	1,98%	3,78%	18,13%	29,92%	0,14%	15,74%	0,34%	0,25%	0,73%	0,53%	4,72%	9,48%	0,43%	2,28%	1,50%	0,45%	0,03%
Piano	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
Violino	0,75%	1,94%	2,36%	2,07%	1,05%	0,04%	77,28%	0,33%	0,96%	1,25%	0,93%	2,44%	2,00%	1,52%	1,04%	2,27%	1,68%	0,10%
Clarinete baixo	0,01%	2,80%	3,86%	0,01%	0,01%	0,00%	0,26%	85,82%	0,21%	1,31%	2,70%	0,32%	0,18%	0,27%	0,08%	0,82%	0,27%	1,08%
Fagote	0,10%	0,85%	0,89%	0,17%	0,08%	0,00%	1,18%	0,26%	84,24%	0,23%	1,79%	1,13%	0,91%	5,38%	0,39%	1,77%	0,42%	0,20%
Clarinete	0,01%	0,59%	1,19%	0,05%	0,01%	0,00%	1,29%	1,19%	0,20%	90,21%	0,99%	0,48%	0,79%	0,48%	0,88%	0,61%	0,66%	0,37%
Contrafagote	0,30%	2,25%	6,37%	0,82%	0,05%	0,01%	1,65%	0,57%	0,98%	0,26%	75,81%	1,29%	0,40%	2,83%	0,69%	0,91%	0,84%	3,96%
Corne Inglês	0,55%	0,38%	1,45%	0,53%	0,49%	0,03%	3,81%	0,32%	0,36%	0,55%	1,10%	86,15%	1,66%	0,20%	0,58%	1,19%	0,62%	0,04%
Flauta	0,54%	1,05%	1,50%	3,12%	0,81%	0,07%	3,62%	0,75%	0,29%	1,73%	0,93%	2,16%	76,74%	0,90%	3,15%	1,44%	1,04%	0,13%
Trompa	0,23%	0,96%	1,74%	2,77%	0,14%	0,01%	1,63%	0,20%	2,38%	0,77%	4,52%	0,34%	0,96%	79,56%	0,68%	1,79%	0,65%	0,68%
Oboe	0,14%	1,05%	0,35%	0,15%	0,27%	0,00%	2,78%	0,09%	0,05%	2,97%	0,99%	1,17%	4,76%	0,48%	82,32%	0,82%	1,61%	0,01%
Trombone	0,05%	1,46%	1,88%	0,09%	0,01%	0,01%	2,51%	0,74%	3,07%	1,17%	2,73%	0,75%	0,45%	3,94%	0,17%	79,27%	1,35%	0,36%
Trompete	0,12%	1,07%	1,07%	0,04%	0,05%	0,01%	4,77%	0,41%	0,44%	2,33%	1,19%	1,52%	1,44%	0,93%	2,37%	4,60%	77,61%	0,02%
Tuba	0,14%	1,35%	12,87%	0,96%	0,01%	0,00%	0,46%	0,36%	0,25%	0,15%	8,74%	0,17%	0,12%	1,34%	0,25%	0,37%	0,21%	72,25%

Figura 6.1: Matriz de confusão para os dezoito instrumentos: notas individuais

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	7,41	0,77	SpectralMfccs	8,35	0,87	SpectralPitchChroma	2,35	0,25
SpectralMfccs	3,30	0,34	SpectralPitchChroma	3,52	0,37	SpectralCentroid	2,69	0,28
SpectralMfccs	4,23	0,44	SpectralPitchChroma	2,84	0,30	SpectralCrestFactor	2,90	0,30
SpectralMfccs	4,11	0,43	SpectralPitchChroma	2,73	0,29	spectraldecrease	2,79	0,29
SpectralMfccs	2,92	0,31	SpectralPitchChroma	1,95	0,20	SpectralFlatness	4,86	0,51
SpectralMfccs	3,19	0,33	SpectralPitchChroma	2,79	0,29	SpectralFlux	9,36	0,98
SpectralMfccs	3,41	0,36	SpectralPitchChroma	7,16	0,75	SpectralKurtosis	3,09	0,32
SpectralMfccs	6,21	0,65	SpectralPitchChroma	2,61	0,27	SpectralRolloff	2,78	0,29
SpectralMfccs	3,67	0,38	SpectralPitchChroma	8,05	0,84	SpectralSkewness	6,11	0,64
SpectralMfccs	3,94	0,41	SpectralPitchChroma	2,98	0,31	SpectralSlope	2,23	0,23
SpectralMfccs	5,26	0,55	SpectralPitchChroma	8,65	0,90	SpectralSpread	4,05	0,42
SpectralMfccs	9,57	1,00	SpectralPitchChroma	9,48	0,99	SpectralTonalPowerRatio	4,52	0,47

Figura 6.2: Importância dos descritores: notas individuais

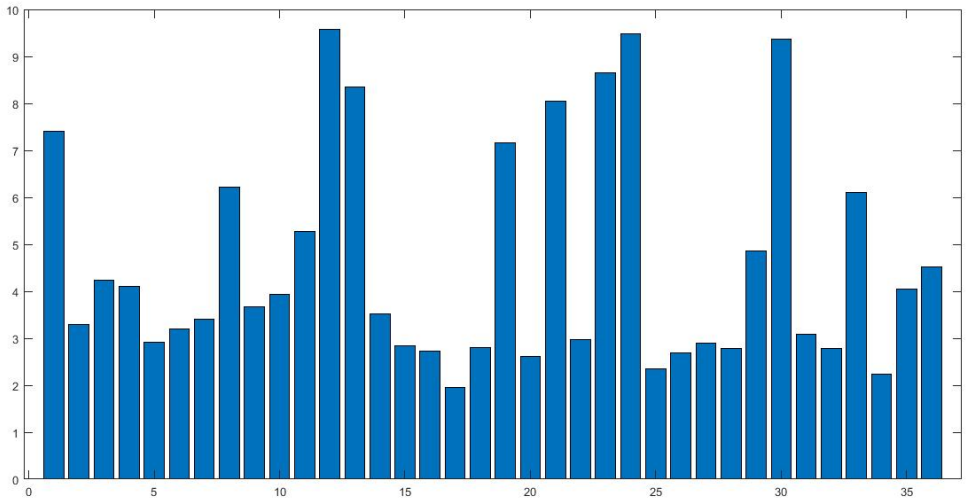


Figura 6.3: Gráfico de barras da importância absoluta: notas individuais

na identificação de notas individuais, e não necessariamente do instrumento-fonte; assim, apesar de haver um número alto de notas dispostas em muitas das 88 notas para cada instrumento, ainda podemos ter dúvidas quanto a sua representatividade na classificação de trechos de maior complexidade (ou ainda polifônicos). Destacam-se aqui também os descritores Espectrais de Mel (MFCC) e o Fluxo Espectral.

Este primeiro experimento serviu como um primeiro passo no entendimento do processo de classificação de instrumentos musicais por descritores de frequência. No entanto, as limitações desse primeiro classificador ficam evidentes na extrapolação para o mundo real: quando realizamos buscas seletivas de músicas, por exemplo, gostaríamos de não apenas encontrar determinar instrumentos que tocam um única nota; e, mais importante do que isso, o contexto prático de busca *online* é de um espaço amostral composto de um universo quase que infinito de trechos monofônicos mais complexos e, muito mais comumente, polifônicos. Desse modo, a partir desse segundo experimento, como já destacado na descrição de amostras de treinamento, serão realizadas tentativas de identificação de trechos musicais de determinados instrumentos em contraste a uma miríade de combinações de sons que existem na base *IRMAS* ("NOT CLASS").

Assim, como segundo experimento, criou-se uma base de dados composta por trechos de músicas de piano a serem separadas nas classes PIANO e "NOT CLASS", cujos resultados se encontram na Figura 6.4. Como alguns dos dados da "NOT CLASS" continham piano como instrumento dominante, este foram retirados do *database*. O mesmo será feito para os demais experimentos, sendo sempre retirados da base de testes *IRMAS* ("NOT CLASS") aqueles instrumentos cuja anotação de dominância (já que muitas vezes há instrumentação de fundo) seja igual à do instrumento em treinamento.

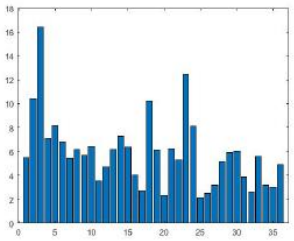
Neste treinamento ficou evidente a importância dos descritores de *Mel* número 3 e *Pitch Chroma* número 10. O índice geral de acerto do experimento ficou em **90,5%**.

	Piano	Not class
Piano	94,0%	6,0%
Not class	13,6%	86,4%

(a) Matriz de confusão do classificador para trechos musicais: trechos de piano solo

descritores	importância	Imp. Relativa	descritores	importância	Imp. Relativa	descritores	importância	Imp. Relativa
SpectralMfccs	5,46	0,33	SpectralMfccs	6,17	0,38	SpectralPitchChroma	2,12	0,13
SpectralMfccs	10,39	0,63	SpectralPitchChroma	7,27	0,44	SpectralCentroid	2,50	0,15
SpectralMfccs	16,41	1,00	SpectralPitchChroma	6,35	0,39	SpectralCrestFactor	3,19	0,19
SpectralMfccs	7,09	0,43	SpectralPitchChroma	4,03	0,25	spectraldecrease	5,19	0,32
SpectralMfccs	8,16	0,50	SpectralPitchChroma	2,68	0,16	SpectralFlatness	5,88	0,36
SpectralMfccs	6,77	0,41	SpectralPitchChroma	10,23	0,62	SpectralFlux	6,01	0,37
SpectralMfccs	5,45	0,33	SpectralPitchChroma	6,09	0,37	SpectralKurtosis	3,84	0,23
SpectralMfccs	6,18	0,38	SpectralPitchChroma	2,33	0,14	SpectralRolloff	2,61	0,16
SpectralMfccs	5,71	0,35	SpectralPitchChroma	6,24	0,38	SpectralSkewness	5,60	0,34
SpectralMfccs	6,44	0,39	SpectralPitchChroma	5,27	0,32	SpectralSlope	3,17	0,19
SpectralMfccs	3,53	0,21	SpectralPitchChroma	12,48	0,76	SpectralSpread	3,03	0,18
SpectralMfccs	4,70	0,29	SpectralPitchChroma	8,12	0,49	SpectralTonalPowerRatio	4,92	0,30

(b) importância dos descritores: trechos de piano solo



(c) Gráfico de barras da importância absoluta:piano solo

Figura 6.4: Segundo experimento: piano

Do mesmo modo como foi feito o treinamento e classificação para o piano, na Figura 6.5 se encontra o resultado para o instrumento violino e base IRMAS (sem trechos com dominância de violino). Para tal experimento, o resultado geral de acerto ficou em **87,6%**, com diversos descritores importantes de *Mel* e *Pitch Chroma*.

Na Figura 6.6 se encontra o resultado para o instrumento clarinete e base IRMAS (sem trechos com dominância de clarinete). Para este terceiro experimento, o resultado geral de acerto ficou em **96,4%** e com, novamente, apenas descritores significativos de *Mel* e *Pitch Chroma*.

Seguindo-se a mesma lógica, na Figura 6.7 apresentamos o resultado para o instrumento fagote e a base IRMAS (sem trechos com dominância de fagote); aqui o resultado geral de acerto ficou em **97,9%** e com descritores significativos de *Mel* números 3 e 4.

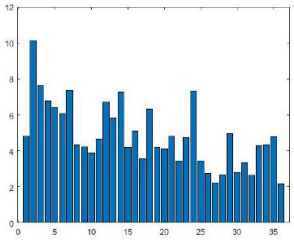
Segue-se na Figura 6.8 o resultado para o instrumento violão e base IRMAS (sem trechos com dominância de violão). Este último experimento monofônico atingiu acerto geral de **93,4%** e com descritores significativos de *Mel*, *Pitch Chroma*,

	Violino	Not class
Violino	76,3%	23,7%
Not class	6,7%	93,3%

(a) Matriz de confusão do classificador para trechos musicais: trechos de violino solo

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	4,80	0,47	SpectralMfccs	5,82	0,58	SpectralPitchChroma	3,46	0,34
SpectralMfccs	10,12	1,00	SpectralPitchChroma	7,26	0,72	SpectralCentroid	2,75	0,27
SpectralMfccs	7,65	0,76	SpectralPitchChroma	4,17	0,41	SpectralCrestFactor	2,22	0,22
SpectralMfccs	6,78	0,67	SpectralPitchChroma	5,08	0,50	spectraldecrease	2,66	0,26
SpectralMfccs	6,40	0,63	SpectralPitchChroma	3,57	0,35	SpectralFlatness	4,95	0,49
SpectralMfccs	6,05	0,60	SpectralPitchChroma	6,32	0,62	SpectralFlux	2,79	0,28
SpectralMfccs	7,35	0,73	SpectralPitchChroma	4,19	0,41	SpectralKurtosis	3,34	0,33
SpectralMfccs	4,31	0,43	SpectralPitchChroma	4,12	0,41	SpectralRolloff	2,63	0,26
SpectralMfccs	4,22	0,42	SpectralPitchChroma	4,81	0,48	SpectralSkewness	4,30	0,42
SpectralMfccs	3,86	0,38	SpectralPitchChroma	3,44	0,34	SpectralSlope	4,33	0,43
SpectralMfccs	4,62	0,46	SpectralPitchChroma	4,74	0,47	SpectralSpread	4,78	0,47
SpectralMfccs	6,69	0,66	SpectralPitchChroma	7,35	0,73	SpectralTonalPowerRatio	2,17	0,21

(b) importância dos descritores: trechos de violino solo



(c) Gráfico de barras da importância absoluta:violino solo

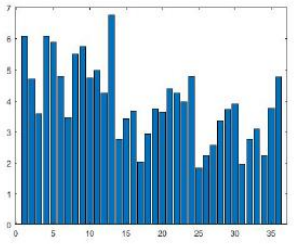
Figura 6.5: Terceiro experimento: violino

	Clarinete	Not class
Clarinete	63,6%	36,4%
Not class	0,5%	99,5%

(a) Matriz de confusão do classificador para trechos musicais: clarinete solo

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	6,07	0,90	SpectralMfccs	6,74	1,00	SpectralPitchChroma	1,83	0,27
SpectralMfccs	4,71	0,70	SpectralPitchChroma	2,76	0,41	SpectralCentroid	2,24	0,33
SpectralMfccs	3,59	0,53	SpectralPitchChroma	3,42	0,51	SpectralCrestFactor	2,56	0,38
SpectralMfccs	6,06	0,90	SpectralPitchChroma	3,68	0,55	spectraldecrease	3,37	0,50
SpectralMfccs	5,86	0,87	SpectralPitchChroma	2,02	0,30	SpectralFlatness	3,73	0,55
SpectralMfccs	4,81	0,71	SpectralPitchChroma	2,92	0,43	SpectralFlux	3,91	0,58
SpectralMfccs	3,46	0,51	SpectralPitchChroma	3,75	0,56	SpectralKurtosis	1,94	0,29
SpectralMfccs	5,48	0,81	SpectralPitchChroma	3,65	0,54	SpectralRolloff	2,76	0,41
SpectralMfccs	5,73	0,85	SpectralPitchChroma	4,39	0,65	SpectralSkewness	3,09	0,46
SpectralMfccs	4,75	0,70	SpectralPitchChroma	4,25	0,63	SpectralSlope	2,23	0,33
SpectralMfccs	4,99	0,74	SpectralPitchChroma	3,97	0,59	SpectralSpread	3,76	0,56
SpectralMfccs	4,25	0,63	SpectralPitchChroma	4,80	0,71	SpectralTonalPowerRatio	4,78	0,71

(b) importância dos descritores: trechos de clarinete solo



(c) Gráfico de barras da importância absoluta:clarinete solo

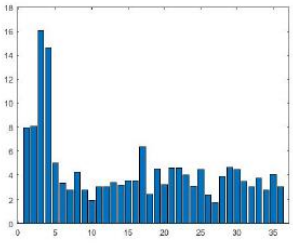
Figura 6.6: Quarto experimento: clarinete

	Fagote	Not class
Fagote	89,9%	10,1%
Not class	0,9%	99,1%

(a) Matriz de confusão do classificador para trechos musicais: fagote solo

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	7,92	0,49	SpectralMfccs	3,34	0,21	SpectralPitchChroma	4,44	0,28
SpectralMfccs	8,03	0,50	SpectralPitchChroma	3,13	0,20	SpectralCentroid	2,39	0,15
SpectralMfccs	16,01	1,00	SpectralPitchChroma	3,49	0,22	SpectralCrestFactor	1,75	0,11
SpectralMfccs	14,61	0,91	SpectralPitchChroma	3,48	0,22	spectraldecrease	3,83	0,24
SpectralMfccs	4,98	0,31	SpectralPitchChroma	6,34	0,40	SpectralFlatness	4,58	0,29
SpectralMfccs	3,28	0,21	SpectralPitchChroma	2,41	0,15	SpectralFlux	4,43	0,28
SpectralMfccs	2,80	0,18	SpectralPitchChroma	4,51	0,28	SpectralKurtosis	3,43	0,21
SpectralMfccs	4,22	0,26	SpectralPitchChroma	3,16	0,20	SpectralRolloff	2,94	0,18
SpectralMfccs	2,79	0,17	SpectralPitchChroma	4,52	0,28	SpectralSkewness	3,72	0,23
SpectralMfccs	1,93	0,12	SpectralPitchChroma	4,51	0,28	SpectralSlope	2,80	0,17
SpectralMfccs	2,95	0,18	SpectralPitchChroma	3,99	0,25	SpectralSpread	4,00	0,25
SpectralMfccs	2,98	0,19	SpectralPitchChroma	3,00	0,19	SpectralTonalPowerRatio	2,94	0,18

(b) importância dos descritores: trechos de fagote solo



(c) Gráfico de barras da importância absoluta:fagote solo

Figura 6.7: Quinto experimento: fagote

Spectral Flatness e *Spectral Tonal Power Ratio*. No entanto, este índice não reflete o acerto do instrumento violão em si, que teve apenas acerto de **36,7%**

Posteriormente, como seguimento dos experimentos, para termos um ideia do que acontece quando a complexidade aumenta, foi usado o mesmo método já descrito de classificação numa base de dados composta por trechos instrumentais em que dois instrumentos também estão presentes em duo.

Para o resultado da classificação para os instrumentos piano e violino com respectiva base IRMAS (sem estes instrumentos com dominância), apresentamos a Figura 6.9. No caso do duo piano e violino, ficou clara a dificuldade na classificação com aumento de complexidade: apesar do índice geral de acerto ficar em **77%**, o acerto dessa dupla ficou em apenas **31%**. Os descritores importantes foram *Mel* e *Pitch Chroma*.

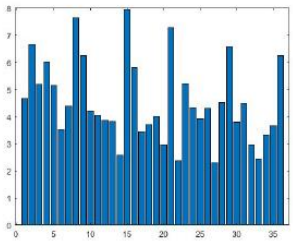
Do mesmo modo, foram feitas classificações para os dois instrumentos piano e clarinete, cujos resultados estão na Figura 6.10. Para esta dupla, o mesmo comportamento foi percebido, com apenas **31%**, apesar do índice geral de acerto ser de **86,3%**. Os descritores importantes seguiram sendo *Mel* e *Pitch Chroma*, além de neste caso o descritor *Tonal Power Ratio*.

	Violão	Not class
Violão	36,7%	63,3%
Not class	0,8%	99,2%

(a) Matriz de confusão do classificador para trechos musicais: violão solo

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	4,65	0,59	SpectralMfccs	3,82	0,48	SpectralPitchChroma	3,90	0,49
SpectralMfccs	6,66	0,84	SpectralPitchChroma	2,58	0,33	SpectralCentroid	4,29	0,54
SpectralMfccs	5,20	0,66	SpectralPitchChroma	7,94	1,00	SpectralCrestFactor	2,29	0,29
SpectralMfccs	6,00	0,76	SpectralPitchChroma	5,80	0,73	spectraldecrease	4,52	0,57
SpectralMfccs	5,14	0,65	SpectralPitchChroma	3,45	0,43	SpectralFlatness	6,57	0,83
SpectralMfccs	3,50	0,44	SpectralPitchChroma	3,70	0,47	SpectralFlux	3,78	0,48
SpectralMfccs	4,36	0,55	SpectralPitchChroma	4,00	0,50	SpectralKurtosis	4,50	0,57
SpectralMfccs	7,64	0,96	SpectralPitchChroma	2,95	0,37	SpectralRolloff	2,95	0,37
SpectralMfccs	6,25	0,79	SpectralPitchChroma	7,25	0,91	SpectralSkewness	2,44	0,31
SpectralMfccs	4,20	0,53	SpectralPitchChroma	2,37	0,30	SpectralSlope	3,32	0,42
SpectralMfccs	4,05	0,51	SpectralPitchChroma	5,21	0,66	SpectralSpread	3,65	0,46
SpectralMfccs	3,86	0,49	SpectralPitchChroma	4,33	0,55	SpectralTonalPowerRatio	6,23	0,79

(b) importância dos descritores: trechos de violão solo



(c) Gráfico de barras da importância absoluta:violão solo

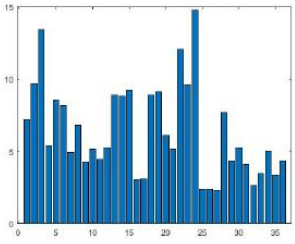
Figura 6.8: Sexto experimento: violão

	Piano	Piano e Violino	Violino	Not class
Piano	92%	2%	2%	4%
Piano e Violino	38%	31%	21%	10%
Violino	14%	6%	69%	11%
Not class	14%	2%	6%	78%

(a) Matriz de confusão do classificador para trechos musicais: duo piano e violino

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	7,20	0,49	SpectralMfccs	8,89	0,60	SpectralPitchChroma	2,36	0,16
SpectralMfccs	9,69	0,66	SpectralPitchChroma	8,83	0,60	SpectralCentroid	2,40	0,16
SpectralMfccs	13,40	0,91	SpectralPitchChroma	9,24	0,63	SpectralCrestFactor	2,28	0,15
SpectralMfccs	5,39	0,36	SpectralPitchChroma	3,06	0,21	spectraldecrease	7,65	0,52
SpectralMfccs	8,55	0,58	SpectralPitchChroma	3,10	0,21	SpectralFlatness	4,36	0,30
SpectralMfccs	8,14	0,55	SpectralPitchChroma	8,89	0,60	SpectralFlux	5,24	0,35
SpectralMfccs	4,91	0,33	SpectralPitchChroma	9,11	0,62	SpectralKurtosis	4,08	0,28
SpectralMfccs	6,81	0,46	SpectralPitchChroma	6,11	0,41	SpectralRolloff	2,63	0,18
SpectralMfccs	4,28	0,29	SpectralPitchChroma	5,13	0,35	SpectralSkewness	3,48	0,24
SpectralMfccs	5,13	0,35	SpectralPitchChroma	12,06	0,82	SpectralSlope	5,00	0,34
SpectralMfccs	4,42	0,30	SpectralPitchChroma	9,58	0,65	SpectralSpread	3,35	0,23
SpectralMfccs	5,24	0,35	SpectralPitchChroma	14,78	1,00	SpectralTonalPowerRatio	4,35	0,29

(b) importância dos descritores: trechos de piano e violino em duo



(c) Gráfico de barras da importância absoluta: piano e violino

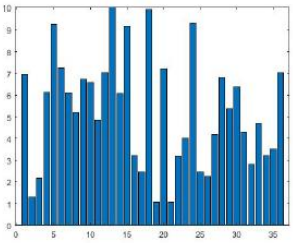
Figura 6.9: Sétimo experimento: piano e violino

	Piano	Piano e Clarinete	Clarinete	Not class
Piano	95%	1%	0%	5%
Piano e Clarinete	48%	38%	1%	13%
Clarinete	33%	3%	53%	11%
Not class	13%	1%	0%	86%

(a) Matriz de confusão do classificador para trechos musicais: duo piano e clarinete

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	6,96	0,70	SpectralMfccs	9,98	1,00	SpectralPitchChroma	2,47	0,25
SpectralMfccs	1,29	0,13	SpectralPitchChroma	6,07	0,61	SpectralCentroid	2,26	0,23
SpectralMfccs	2,17	0,22	SpectralPitchChroma	9,16	0,92	SpectralCrestFactor	4,19	0,42
SpectralMfccs	6,15	0,62	SpectralPitchChroma	3,22	0,32	spectraldecrease	6,80	0,68
SpectralMfccs	9,26	0,93	SpectralPitchChroma	2,45	0,25	SpectralFlatness	5,39	0,54
SpectralMfccs	7,25	0,73	SpectralPitchChroma	9,91	0,99	SpectralFlux	6,37	0,64
SpectralMfccs	6,11	0,61	SpectralPitchChroma	1,07	0,11	SpectralKurtosis	4,30	0,43
SpectralMfccs	5,19	0,52	SpectralPitchChroma	7,21	0,72	SpectralRolloff	2,82	0,28
SpectralMfccs	6,72	0,67	SpectralPitchChroma	1,07	0,11	SpectralSkewness	4,70	0,47
SpectralMfccs	6,58	0,66	SpectralPitchChroma	3,19	0,32	SpectralSlope	3,22	0,32
SpectralMfccs	4,84	0,48	SpectralPitchChroma	4,01	0,40	SpectralSpread	3,51	0,35
SpectralMfccs	7,04	0,71	SpectralPitchChroma	9,31	0,93	SpectralTonalPowerRatio	7,05	0,71

(b) importância dos descritores: trechos de piano e clarinete em duo



(c) importância dos descritores: trechos de piano e clarinete

Figura 6.10: Oitavo experimento: piano e clarinete

Encerrando-se essa fase dos experimentos, foram feitas classificações para os instrumentos violão e flauta (Figura 6.11). Exatamente os mesmos descritores do experimento anterior foram importantes.

Ultimando-se os experimentos, várias classes anteriores foram reunidas para aumentar ainda mais a complexidade da separação e entender como isto afeta o desempenho do classificador. Aqui, está-se fazendo a separação entre combinações dos instrumentos: clarinete, violino e piano. Em função de ausência de disponibilidade amostral, foi omitida classe do duo clarinete e violino. Os resultados se encontram na Figura 6.12. O índice de acerto geral ficou em **74,8%**, com descritores importantes *Mel* e *Pitch Chroma*.

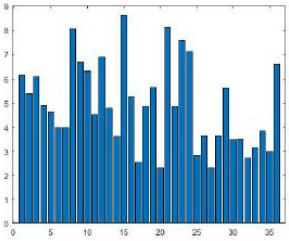
De tudo posto, finalmente destacaram-se ao longo dos experimentos (aqui não se inclui o primeiro experimento com notas individuais), os descritores dispostos nas Tabelas 6.1 (36 variáveis dispostas individualmente) e 6.2 (*ranking* agregado dos 13 descritores).

	Violão	Violão e Flauta	Flauta	Not class
Violão	40%	0%	0%	59%
Violão e Flauta	5%	29%	0%	66%
Flauta	1%	1%	59%	39%
Not class	1%	0%	0%	99%

(a) Matriz de confusão do classificador para trechos musicais: duo violão e flauta

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	6,17	0,72	SpectralMfccs	4,78	0,56	SpectralPitchChroma	2,83	0,33
SpectralMfccs	5,40	0,63	SpectralPitchChroma	3,61	0,42	SpectralCentroid	3,63	0,42
SpectralMfccs	6,09	0,71	SpectralPitchChroma	8,61	1,00	SpectralCrestFactor	2,30	0,27
SpectralMfccs	4,89	0,57	SpectralPitchChroma	5,26	0,61	spectraldecrease	3,64	0,42
SpectralMfccs	4,64	0,54	SpectralPitchChroma	2,54	0,29	SpectralFlatness	5,64	0,65
SpectralMfccs	3,98	0,46	SpectralPitchChroma	4,87	0,57	SpectralFlux	3,48	0,40
SpectralMfccs	3,98	0,46	SpectralPitchChroma	5,66	0,66	SpectralKurtosis	3,51	0,41
SpectralMfccs	8,04	0,93	SpectralPitchChroma	2,30	0,27	SpectralRolloff	2,72	0,32
SpectralMfccs	6,68	0,78	SpectralPitchChroma	8,11	0,94	SpectralSkewness	3,13	0,36
SpectralMfccs	6,31	0,73	SpectralPitchChroma	4,86	0,56	SpectralSlope	3,83	0,45
SpectralMfccs	4,52	0,52	SpectralPitchChroma	7,58	0,88	SpectralSpread	2,97	0,34
SpectralMfccs	6,85	0,80	SpectralPitchChroma	7,09	0,82	SpectralTonalPowerRatio	6,58	0,76

(b) importância dos descritores: trechos de violão e flauta em duo



(c) importância dos descritores: trechos de violão e flauta

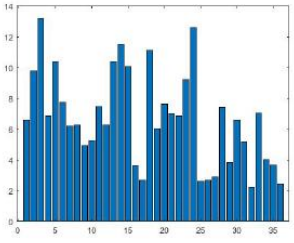
Figura 6.11: Nono experimento: violão e flauta

	Piano	Violino	Clarinete	Piano e Violino	Piano e Clarinete	Not class
Piano	92,5%	1,8%	0,1%	2,0%	0,6%	3,1%
Violino	14,1%	70,1%	0,0%	6,3%	0,4%	9,1%
Clarinete	31,8%	3,7%	52,2%	1,4%	2,8%	8,0%
Piano e Violino	38,7%	21,7%	0,0%	31,5%	0,6%	7,4%
Piano e Clarinete	47,0%	4,1%	0,5%	2,2%	37,3%	8,8%
Not class	13,2%	6,2%	0,1%	2,1%	0,7%	77,6%

(a) Matriz de confusão do classificador para trechos musicais: violino, piano e clarinete

descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa	descritores	Importância	Imp. Relativa
SpectralMfccs	6,59	0,50	SpectralMfccs	10,41	0,79	SpectralPitchChroma	2,60	0,20
SpectralMfccs	9,77	0,74	SpectralPitchChroma	11,48	0,87	SpectralCentroid	2,68	0,20
SpectralMfccs	13,20	1,00	SpectralPitchChroma	10,05	0,76	SpectralCrestFactor	2,88	0,22
SpectralMfccs	6,82	0,52	SpectralPitchChroma	3,62	0,27	spectraldecrease	7,41	0,56
SpectralMfccs	10,38	0,79	SpectralPitchChroma	2,69	0,20	SpectralFlatness	3,83	0,29
SpectralMfccs	7,72	0,58	SpectralPitchChroma	11,12	0,84	SpectralFlux	6,58	0,50
SpectralMfccs	6,20	0,47	SpectralPitchChroma	6,02	0,46	SpectralKurtosis	5,17	0,39
SpectralMfccs	6,26	0,47	SpectralPitchChroma	7,63	0,58	SpectralRolloff	2,22	0,17
SpectralMfccs	4,93	0,37	SpectralPitchChroma	7,00	0,53	SpectralSkewness	7,03	0,53
SpectralMfccs	5,23	0,40	SpectralPitchChroma	6,84	0,52	SpectralSlope	4,01	0,30
SpectralMfccs	7,44	0,56	SpectralPitchChroma	9,19	0,70	SpectralSpread	3,66	0,28
SpectralMfccs	6,26	0,47	SpectralPitchChroma	12,59	0,95	SpectralTonalPowerRatio	2,44	0,18

(b) importância dos descritores: combinações de violino, piano e clarinete



(c) importância dos descritores: violino, piano e clarinete

Figura 6.12: Nono experimento: violino, piano e clarinete

Tabela 6.1: *Ranking* geral dos descritores em todos os experimentos (visão individual)

rank	Descritor	média	desvio-padrão	% coeficiente variação
1	Spectral Mfccs3	0,753	0,263	35,0
2	Spectral Pitch Chroma11	0,708	0,268	37.8
3	Spectral Pitch Chroma2	0,648	0,289	44,7
4	Spectral Mfccs2	0,647	0,257	39,7
5	Spectral Mfccs5	0,643	0,198	30,8
6	Spectral Mfccs4	0,637	0,168	26,4
7	Spectral Mfccs13	0,621	0,268	43,2
8	Spectral Pitch Chroma10	0,594	0,196	32,9
9	Spectral Pitch Chroma5	0,589	0,256	43,5
10	Spectral Mfccs8	0,581	0,260	44,7
11	Spectral Mfccs1	0,577	0,153	26,5
12	Spectral Mfccs6	0,522	0,168	32,3
13	Spectral Mfccs9	0,521	0,259	49,7
14	Spectral PitchChroma8	0,515	0,295	57,3
15	Spectral PitchChroma1	0,510	0,221	43,3
16	Spectral Mfccs12	0,509	0,201	39,5
17	Spectral Flatness	0,477	0,196	41,0
18	Spectral Mfccs10	0,474	0,211	44,6
19	Spectral Mfccs7	0,464	0,169	36,4
20	Spectral TonalPowerRatio	0,460	0,282	61,3
21	Spectral PitchChroma9	0,455	0,193	42,5
22	Spectral Decrease	0,453	0,153	33,9
23	Spectral Mfccs11	0,442	0,168	38,0
24	Spectral Pitch Chroma6	0,441	0,182	41,3
25	Spectral Flux	0,430	0,134	31,2
26	Spectral Pitch Chroma3	0,406	0,197	48,5
27	Spectral Pitch Chroma7	0,405	0,171	42,2
28	Spectral Skewness	0,374	0,112	30,0
29	Spectral Kurtosis	0,349	0,110	31,6
30	Spectral Spread	0,347	0,119	34,2
31	Spectral Slope	0,328	0,088	26,7
32	Spectral Pitch Chroma4	0,289	0,085	29,3
33	Spectral Centroid	0,273	0,136	49,9
34	Spectral Pitch Chroma12	0,272	0,102	37,5
35	Spectral Rolloff	0,259	0,091	35,3
36	Spectral Crest Factor	0,250	0,106	42,2

Fonte: Autor.

Tabela 6.2: *Ranking* geral dos descritores em todos os experimentos (visão geral)

rank	Descritor
1	Spectral Mfccs
2	Spectral Pitch Chroma
3	Spectral Flatness
4	Spectral Tonal PowerRatio
5	Spectral Decrease
6	Spectral Flux
7	Spectral Skewness
8	Spectral Kurtosis
9	Spectral Spread
10	Spectral Slope
11	Spectral Centroid
12	Spectral Rolloff
13	Spectral Crest Factor

Fonte: Autor.

7 CONSIDERAÇÕES SOBRE OS RESULTADOS

Algumas considerações finais sobre os resultados se fazem necessárias. Um primeiro ponto é perceber que, confirmando a literatura, os coeficientes de Mel são significativos na classificação de timbres, como era esperado. No entanto, ao contrário do previsto inicialmente, também foi fundamental para a maioria dos experimentos o descritor *Spectral Pitch Chroma*. Uma possível explicação, a ser investigada em trabalhos futuros, é de que, para alguns instrumentos, foi utilizado um número pequeno de músicas na composição da base do treinamento; assim, apesar de haver, para quase todos os instrumentos, um número importante de *frames* ou de amostras (já que algumas músicas foram divididas em mais de uma amostra com vistas a eliminar elementos sonoros indesejados), um menor número de músicas, as quais conservam tonalidades determinadas, poderá ter por consequência repetições mais frequentes de suas notas tônicas, o que pode estar se refletindo na importância do descritor cromático.

Um segundo ponto interessante é que, no primeiro experimento em que a base IRMAS não estava presente e em que tínhamos de apenas classificar notas individuais de dezoito instrumentos, o fluxo espectral foi um dos mais importantes descritores. Esse fato se repetiu em diversos experimentos que não foram aqui descritos por escaparem do objetivo do trabalho (mas que foram analisados pelo autor), em que foi feita separação direta entre instrumentos dois a dois sem a presença da base IRMAS. No entanto, nos experimentos aqui descritos com tal base, o descritor de fluxo espectral teve apenas *rank* geral na sexta posição, o que não era esperado de acordo com a literatura consultada.

O terceiro ponto reforça uma aspecto bastante intuitivo: verifica-se a redução do índice geral de acerto em acordo com o aumento do número de ins-

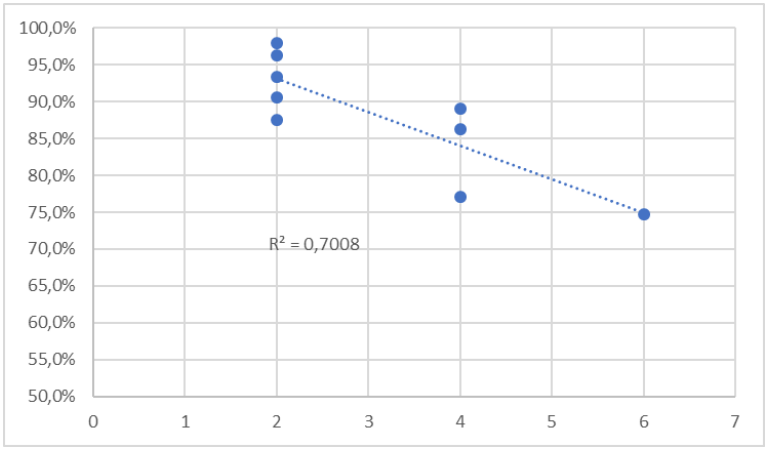


Figura 7.1: Gráfico de dispersão dos experimentos: eixo horizontal se refere ao número de classes, eixo vertical ao índice geral de acertos, com coeficiente de correlação linear de 0,84 e p-valor=$1,5.10^{-5}$

trumentos (refletido no aumento do número de classes), de acordo com a Figura 7.1.

Como último ponto, coloca-se a importância do tamanho amostral na determinação do sucesso na classificação; não é de se estranhar que um número maior de dados de entrada no classificador resultem em uma maior eficiência de classificação. Como é abaixo exposto na Tabela 7.1 e na Figura 7.2 para o primeiro experimento de notas individuais, essa relação de fato é estatisticamente significativa, o que não implica causalidade mas procura apontar para um dos fatores que parece influenciar o resultado dos experimentos, o que deve ser mais profundamente investigado no futuro.

A ideia final do presente trabalho é de que, dada uma situação de restrição de memória ou capacidade de processamento, possamos escolher aqueles descritores que mais facilmente (ou com menor custo computacional) realizariam o trabalho de reconhecimento de instrumentos musicais. No entanto, temos que pensar ainda que talvez estas *features* possam funcionar bem justamente em conjunto, cada uma contribuindo com determinado aspecto do envelope espectral. Assim como numa distribuição probabilística, não só a média e o desvio-padrão (com

Tabela 7.1: Distribuição amostral das notas individuais versus índice de acerto

Instrumento	% frames	% acerto
banjo	1,39	28,3
violoncelo	5,86	63,42
contrabaixo	8,42	77,85
violão	3,74	73,12
bandolim	1,52	29,92
piano	12,89	100
violino	7,52	77,28
clarinete baixo	4,85	85,82
fagote	4,24	84,24
clarinete	7,30	90,21
contrafagote	6,16	75,81
corne inglês	5,6	86,15
flauta	5,98	76,74
trompa	6,14	79,56
oboé	4,16	82,32
trombone	5,68	79,27
trompete	4,39	77,61
tuba	4,16	72,25

Fonte: Autor.

exceção de uma distribuição perfeitamente normal) seriam suficientes para determinar o formato da função de densidade de probabilidade; por isso há momentos de maior ordem, como curtose e assimetria. Fica essa investigação para trabalhos futuros, em que podemos tomar subconjuntos dos melhores descritores e comparar o resultado dos treinamentos frente à mesma base de dados, podendo ainda lançar mão de técnicas como PCA *(principal component analysis)*, a qual tenta reduzir a dimensionalidade das variáveis frente a grandes *datasets*.

Também se poderia pensar em outros aspectos que já se mostraram relevantes em alguns trabalhos, como o ataque de uma nota. Outro ponto de investigação seria a utilização de filtros de Kalman para melhorar a predição de duos baseados nas predições individuais.

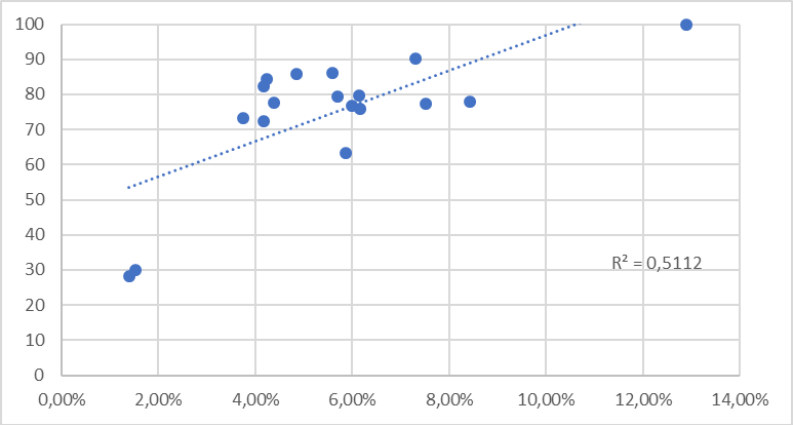


Figura 7.2: Gráfico de dispersão dos experimentos: eixo horizontal se refere à representatividade dos instrumentos no experimento 1; eixo vertical ao índice de acerto desse mesmo experimento (em %), com coeficiente de correlação linear de 0,71, p-valor=0,008

De tudo posto, nessa tentativa de estabelecer os melhores descritores de um instrumento musical, caminhamos para que talvez possamos desenvolver ferramentas para entender nossa própria lógica intrínseca de classificação dessa qualidade do som a que chamamos de timbre.

Referências Bibliográficas

[1] MARTINEZ W.; MARTINEZ, A. *Computational Statistics Handbook with MATLAB*.

[2] WALLMARK ZACHARY; KENDALL, R. A. *Describing Sound: The Cognitive Linguistics of Timbre*. 2018.

[3] LERCH, A. *An Introduction to Audio Content Analysis*. 2012.

[4] ERONEN A.; KLAPURI, A. *Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features*. 2000.

[5] WICAKSANA H;HARTONO, S. *Recognition of Musical Instruments*. 2006.

[6] HU Y;LIU, G. *Dynamis Characteristics of Musical Note for Musical Instrument Classification*.

[7] RUI R;BAO, C. *Projective Non-negative matrix factorization with bregman divergence for musical instrument classification*. 2012.

[8] LOUGHRAM L.;WALKER, J. M. M. *The use of Mel-frequency Cepstral Coefficients in Musical Instrument Identification*. 2013.

[9] BANCHHOR S.;KHAN, A. *Musical Instrument Recognition using Spectrogram and Autocorrelation*. 2012.

[10] AGOSTINI C.;LONGARI, M. E. *Musical instrument timbres classification with spectral features*. 2001.

[11] TAKAHASHI Y; KONDO, K. *Comparison of two classification methods of musical instrument identification*. 2014.

[12] MASOOD S.; GUPTA, S. *Novel Approach for Musical Instrument Identification Using Neural Network*. 2015.

[13] KOTHE R.S.; BHALKE, D. G. P. *Musical Instrument Rocognition using K-Nearest Neighbour and Support Vector Machine*. 2016.

[14] RAVI N..; BHALKE, D. *Musical Instrument Information Retrieval using Neural Network*. 2016.

[15] ANDERSON, T.-A. *Musical Instrument Classification Utilizing a Neural Network*. 2017.

[16] BAHRE, S. e. a. *Machine Learning Based Classification of Violin and Viola Instrument Sounds for the Same Notes*. 2017.

[17] ARICAN K.;POLAT, K. *Novel Audio Feature Set for Monophonic Musical Instrument Classification*. 2018.

[18] JEYALAKSHMI C.; MURUGESHWARI, B. M. *HMM and K-NN based Automatic Musical Instrument Recognition*. 2018.

[19] SUTTON, C. D. *Classification and Regression Trees, Bagging, and Boosting*. 2004.

[20] MACHOVA K; BARCAK, F. B. P. *A Bagging Method using Decision Trees in the Role of Base Classifiers*. 2006.

[21] MATHWORKS documentation page. Disponível em: ⟨https://www.mathworks.com/help/stats/compactregressionensemble.predictorimportance.html⟩.

[22] UNIVERSITY OF IOWA ELETRONIC MUSIC STUDIOS. Disponível em: ⟨http://theremin.music.uiowa.edu/MIS.html⟩.

[23] THE PHILHARMONIA ORCHESTRA (UK). Disponível em: ⟨http://www.philharmonia.co.uk/explore⟩.

[24] SOUND ON SOUND. Disponível em: ⟨https://www.soundonsound.com⟩.

[25] JURG HOCHWEBER: Composer of Guitar Music. Disponível em: ⟨https://www.hochweber.ch/⟩.

[26] THE CLASSICAL CLARINET PAGE. Disponível em: ⟨https://www.classiccat.net⟩.

[27] LIBER LIBER. Disponível em: ⟨https://www.liberliber.it⟩.

[28] MUSICIAN'S PAGE. Disponível em: ⟨https://www.musicianspage.com⟩.

[29] IRMAS: a dataset for instrument recognition in musical audio signals. Disponível em: ⟨https://zenodo.org/record/1290750l⟩.

[30] REUTER, C. *Formants Position of musical instruments and recommendations for ensemble scoring in orchestration treatises*. 1995.

[31] WRIGHT, D. *Mathematics and Music*. 2009.

[32] NAGAWADE M.; RATNAPARKHE, V. *Musical Instrument Identification using MFCCs*. 2017.