



SENAI CIMATEC

**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM
COMPUTACIONAL E TECNOLOGIA INDUSTRIAL**
Mestrado em Modelagem Computacional e Tecnologia Industrial

Dissertação de mestrado

**Estudo de Flutuações de Sinais de Áudio
Classificados por Gênero Musical**

Apresentada por: Dirceu de Freitas Piedade Melo
Orientador: Prof. Dr. Gilney Figueira Zebende

Agosto de 2012

Dirceu de Freitas Piedade Melo

Estudo de Flutuações de Sinais de Áudio Classificados por Gênero Musical

Dissertação de mestrado apresentada ao Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial, Curso de Mestrado em Modelagem Computacional e Tecnologia Industrial do SENAI CIMATEC, como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Área de conhecimento: Interdisciplinar

Orientador: Prof. Dr. Gilney Figueira Zebende
SENAI CIMATEC

Salvador
SENAI CIMATEC
2012

Nota sobre o estilo do PPGMCTI

Esta dissertação de mestrado foi elaborada considerando as normas de estilo (i.e. estéticas e estruturais) propostas aprovadas pelo colegiado do Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial e estão disponíveis em formato eletrônico (*download* na Página Web http://portais.fieb.org.br/portal_faculdades/dissertacoes-e-teses-mcti.html ou solicitação via e-mail à secretaria do programa) e em formato impresso somente para consulta.

Ressalta-se que o formato proposto considera diversos itens das normas da Associação Brasileira de Normas Técnicas (ABNT), entretanto opta-se, em alguns aspectos, seguir um estilo próprio elaborado e amadurecido pelos professores do programa de pós-graduação supracitado.

SENAI CIMATEC

Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial

Mestrado em Modelagem Computacional e Tecnologia Industrial

A Banca Examinadora, constituída pelos professores abaixo listados, leram e recomendam a aprovação [com distinção] da Dissertação de mestrado, intitulada “Estudo de Flutuações de Sinais de Áudio Classificados por Gênero Musical”, apresentada no dia (dia) de (mês) de (ano), como requisito parcial para a obtenção do título de **Mestre em Modelagem Computacional e Tecnologia Industrial**.

Orientador:

Prof. Dr. Prof. Dr. Gilney Figueira Zebende
SENAI CIMATEC

Membro externo da Banca:

Prof. Dr. Pedro Ribeiro Kroger Júnior
Universidade Federal da Bahia

Membro externo da Banca:

Prof. Dr. José Garcia Vivas Miranda
Universidade Federal da Bahia

Membro interno da Banca:

Prof. Dr. Hernane Borges de Barros Pereira
SENAI CIMATEC

Dedico este trabalho a duas pessoas espetaculares :
meu pai Gino Frey (*in memoriam*) por ter me ensinado o amor pela música e a busca
incessante pelo conhecimento;
minha mãe Esmeralda, que tem sido para mim e outras tantas pessoas, um referencial
de virtude, sabedoria e humanidade.

Agradecimentos

Agradeço a Deus por sua graça inexplicável e por seu amor incondicional.

Aos meus pais e irmãos, cujos laços de afetividade me dão abrigo e segurança.

À minha esposa e à minha filha por me acrescentarem alegria e motivação para viver.

Aos professores do MCTI por seu compromisso e dedicação.

Ao meu orientador Prof. Dr. Gilney Figueira Zebende por seu trabalho de orientação, e por seu empenho em realizar o sucesso do Mestrado em Modelagem Computacional do SENAI CIMATEC.

Salvador, Brasil
10 de Agosto de 2012

Dirceu de Freitas Piedade Melo

Resumo

Os descritores musicais são modelos computacionais que buscam prever, a partir de uma representação numérica, determinadas características musicais de um sinal de áudio. Os Modelos de extração de características são ferramentas fundamentais na realização da classificação automática de arquivos. Ultimamente, a extração automática de informações musicais tem ganhado muita importância pois consiste em uma forma de estruturar e organizar o crescente número de arquivos de música disponíveis digitalmente na Web. O estabelecimento de hierarquias de gênero, geralmente criadas pelo trabalho manual de especialistas, é atualmente uma das maneiras mais utilizadas para estruturar conteúdos de música na internet. A Classificação automática de gênero musical pode potencialmente automatizar esse processo e criar uma nova alternativa para realizar a manipulação e organização destes arquivos. Deste modo, o estudo e a criação de novos descritores musicais podem colaborar no processo de classificação automática de arquivos de áudio. Neste trabalho, será realizado o estudo de um descritor musical derivado do DFA (Detrended Fluctuation Analysis), proposto por [JENNINGS \(2004\)](#), denominado de DVFE (Detrended Variance Fluctuation Exponent) ou DFA expoente, aplicado a dois bancos de dados com arquivos musicais classificados por gênero musical. Esta dissertação tem como objetivo avaliar o potencial do DVFE como descritor na caracterização de um conjunto de arquivos classificados em gêneros musicais, propor um banco de dados com música brasileira para ser utilizado em pesquisa MIR (Music Information Retrieval) e explorar a utilização do expoente DFA em novas taxonomias.

Palavras-chave: Gêneros Musicais, Detrended Fluctuation Analysis, Detrended Variance Fluctuation Exponent, Séries Temporais.

Abstract

The musical descriptors are computer models that seek to predict, from a numerical representation, certain musical characteristics of an audio signal. Models feature extraction tools are critical in achieving the automatic sorting of files. Lately the automatic extraction of musical information has gained much importance because it consists of a way to structure and organize the growing number of music files available digitally on the Web. The establishment of genre hierarchies, usually created by manual labor specialists, are currently a most used way to structure the music content on the Internet. The Automatic classification of musical genre can potentially automate this process and create a new alternative to perform the manipulation and organization of these files. Thus, study and development of new musical descriptors can collaborate in the process for automatic audio files classification. This work will be carried out to study a musical descriptor derived from DFA (Detrended Fluctuation Analysis), proposed by (JENNINGS, 2004), called DVFE (Detrended Fluctuation Variance Exponent) or DFA Exponent, applied to the two musical genre databases. This thesis aims to evaluate the potential of DVFE as a descriptor for characterizing a set of files classified into genres, suggest a database of Brazilian music to be used in research MIR (Music Information Retrieval) and explore the use of DFA exponent of new taxonomies.

Sumário

1	Introdução	1
1.1	Importância da pesquisa	2
1.2	Limites e limitações	2
1.3	Aspectos metodológicos	3
1.4	Organização da Dissertação de mestrado	4
2	Áudio, Extração de Parâmetros e Classificação em Gêneros Musicais	5
2.1	O Paradigma dos Quatro Universos	5
2.2	Áudio Analógico	6
2.3	Áudio Digital	6
2.4	Teorema de Nyquist-Shannon (Teorema da Amostragem)	8
2.5	PCM (Pulse Code Modulation)	9
2.6	Formatos no Áudio Digital	9
2.7	Loudness	12
2.8	Percepção Rítmica e Tonal	13
2.9	Descritores Musicais	16
2.10	Extração de Parâmetros	17
2.11	Gêneros Musicais	19
3	Séries Temporais	21
3.1	Definições de Séries Temporais	21
3.2	Análise de Séries Temporais	24
3.2.1	Autocorrelação	24
3.3	Random Walk	27
3.4	Expoente Hurst	29
3.5	Método DFA	32
4	Implementação e resultados	35
4.1	Modelo Utilizado	35
4.1.1	Base de Dados	35
4.1.2	Cálculo do DFA Expoente	36
4.1.3	Interpretação do DFA expoente	40
4.2	Resultados	44
4.2.1	Banco GTZAN	44
4.2.2	Banco BRMUSIC	49
4.2.3	Subgêneros	51
4.2.4	Aplicação do DFA expoente em Outras Categorias	52
4.2.5	DFA expoente de Arquivos Editados	56
4.2.6	<i>Outliers</i>	57
4.2.7	Comparação de Resultados	60
5	Considerações finais	63
5.1	Conclusões	63
5.2	Contribuições	64
5.3	Atividades Futuras de Pesquisa	64

referencias

66

Lista de Tabelas

4.1	GTZAN GENRE COLLECTION	35
4.2	BRMUSIC GENRE COLLECTION	36
4.3	Resultados da função DFA e da função $\alpha(i)$ de um sinal tipo <i>white noise</i> , a partir do arquivo random16. O valor do DFA expoente é $\approx 0,5$. Este resultado é obtido através da média aritmética de todos os valores de $\alpha(i)$ encontrados. Fonte: o autor.	41
4.4	Resultados da função DFA e da função $\alpha(i)$ a partir do arquivo mbeat1 do banco BR-MUSIC. O valor do DFA expoente é 0,474, que corresponde à média aritmética de todos os valores de $\alpha(i)$ encontrados. É importante observar que os valores de t são incrementados através da multiplicação por uma constante, o que resulta num espaçamento igual quando aplicada a escala logarítmica. Fonte: o autor.	43
4.5	Valor médio do DFA expoente, seguido de seus respectivos desvios-padrão. Esses valores foram calculados para cada grupo de cem arquivos musicais pertencentes a um gênero musical do banco GTZAN. Fonte: o autor. . . .	45
4.6	Valores médios do DFA expoente do banco BRMUSIC e seus respectivos desvios-padrão. Fonte: o autor.	49

Lista de Figuras

2.1	Paradigmas de abstração no estudo de sinais (CARVALHO, 2009)	6
2.2	Sinal Contínuo (em vermelho) e seu equivalente Sinal Digital (em azul). Fonte: (GUADARRAMA, 2010)	7
2.3	Digitalização, quantização e reconstrução de um sinal analógico num sistema de conversão analógico-digital (CAD) e digital-analógico (CDA). $f(n)$: sinal analógico; $f[n]$: sinal digital; $f_r(n)$: sinal reconstruído. Fonte: o autor.	7
2.4	Efeito de <i>aliasing</i> . Fonte: IAZZETA (2011a).	8
2.5	No estágio <i>sampling</i> é realizada a amostragem e o sinal analógico (<i>analog signal</i>) se transforma no sinal PAM (<i>Pulse Amplitude Modulation Wave</i>). No segundo estágio (<i>quantizing</i>) é realizada a quantização do sinal, e no terceiro, a codificação (<i>encoding</i>) resultando finalmente no sinal digital. Fonte: OLIVEIRA (2012).	9
2.6	A técnica PCM foi introduzida nos EUA na década de 1960, quando as empresas de telefonia começaram a conversão de voz para o sistema digital, que seria utilizada na transmissão em troncos interurbanos. Fonte: FARLEX (2011).	10
2.7	Comparação entre decibéis e pascais. Fonte: AUDIOLIST (2011).	13
2.8	Curvas de igual intensidade perceptual sonora para sons simples. Observe-se na figura que o ouvido é mais sensível à intensidade sonora dos parciais com frequências medianas, entre 1 kHz a 5 kHz, pois esta é a região de frequência relacionadas aos parciais gerados pela voz humana, o que garante um melhor entendimento das linguagens. Para sons complexos, a percepção da intensidade sonora é correspondente ao loudness resultante de todos os seus parciais (FORNARI, 2010).	14
2.9	Quadro comparativo entre níveis de intensidade e potências acústicas, variando do limiar da audição ao limiar da dor. Aqui são mostrados sons de objetos e cenários que são facilmente identificados pelo censo comum, e seus respectivos níveis de intensidade em db e watts. (IAZZETA, 2011b)	15
2.10	A fronteira entre o domínio do contexto rítmico e o contexto tonal encontra-se aproximadamente estendida entre 0,05 e 0,1 s. Em torno de 50ms de período, ou 20 Hz de frequência, a percepção do intervalo de tempo entre eventos sonoros passa de rítmica (domínio do tempo) para tonal (domínio da frequência) (FORNARI, 2010).	16
2.11	Estrutura da taxonomia de gêneros musicais adotada em BARBETO J.; LOPES (2007).	20
3.1	(a) Série temporal dos montantes anuais de precipitação pluviométrica registrado em Los Angeles, Califórnia, ao longo de mais de 100 anos (CRYER; CHAN, 2008); (b) Série Temporal de um processo químico - industrial. A variável medida aqui é uma propriedade da cor em consecutivos banhos durante o processo (CRYER; CHAN, 2008); (c) Sinal de áudio classical1 do banco GTZAN, sampler de 50 seg. Fonte: o autor.	21
3.2	(a) Realização de um <i>Binary Process</i> para os estados de um disjuntor (CHATFIELD, 1995); (b) Realização de um <i>point process</i> para desastres ferroviários. Fonte: CHATFIELD (1995).	22

3.3	A figura mostra a série das medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970, juntamente com as tendências estimadas. Para isso foram utilizados polinômios de graus 3 e 6, ajustados por mínimos quadrados. Fonte: (EHLERS, 2009)	24
3.4	(a) 50 observações de uma série não-estacionária simuladas segundo um <i>random walk</i> ; (b) 20 primeiras autocorrelações amostrais dessa série. Fonte: EHLERS (2009)	27
3.5	A figura mostra um <i>random walk</i> simulado onde os e 's foram selecionados de uma distribuição normal padrão. Fonte CRYER e CHAN (2008)	29
3.6	Simulação de um <i>random walk</i> bidimensional feita com um milhão de passos. Fonte PEREIRA (2009b).	30
3.7	Simulação de um <i>random walk</i> tridimensional com 1 milhão de passos. Fonte PEREIRA (2009b).	31
3.8	O modelo do reservatório de Harold Edwin Hurst, utilizado para determinar o coeficiente R/S, calculado pela razão entre a amplitude R e o desvio-padrão das vazões de água. Fonte: (FEDER, 1988)	32
3.9	O primeiro gráfico mostra uma série gaussiana. Os três gráficos seguintes mostram o perfil dessa série e o cálculo do R/S para valores distintos de τ . Fonte: GALHARDO (2010).	33
3.10	O primeiro gráfico mostra a série original $x(t)$; o segundo gráfico mostra o perfil $y(t)$ da função $x(t)$; no terceiro gráfico é mostrado o perfil após retiradas as tendências. Fonte: GALHARDO (2010).	34
4.1	Janela de interesse, fronteira rítmico-melódica e intervalos de tempo na investigação rítmica citados por diversos autores. Fonte: o autor.	39
4.2	Os quatro passos para a determinação do DFA expoente, aplicados ao arquivo mbeat1, ilustrados através do: (a) gráfico da série $V(j)$, que representa o desvio-padrão da série original. Este gráfico está relacionado com o <i>loudness</i> do sinal musical; (b) gráfico da série $Y(j)$, resultante da integração de $V(j)$; (c) gráfico da função DFA; (d) gráfico \log x \log da função DFA, cuja inclinação determina o valor do α_{DFA} . Fonte: o autor.	42
4.3	Função $\alpha(t)$ expoente do arquivo mbeat1 do banco BR-MUSIC. Fonte: o autor.	42
4.4	Variações na função $\alpha(t)$ para as peças Classic, Jazz e Hiphop. Uma estabilidade maior na escala é apresentada pela peça Clássica e um comportamento de grande instabilidade é mostrado na peça Hiphop, devido a presença de eventos percussivos com ataque mais pronunciado e padrões de batida mais fortes e regulares. Já a peça musical Jazz apresenta um comportamento intermediário. Fonte: o autor.	44
4.5	Gráfico do DFA expoente médio do banco GTZAN. As barras de erro correspondem ao desvio-padrão. Fonte: o autor.	45
4.6	Boxplot dos gêneros musicais do banco GTZAN. Fonte: o autor.	46
4.7	Histograma da distribuição de frequências de α_{DFA} nos gêneros do banco GTZAN. Fonte: o autor.	46
4.8	Comparação de médias para o banco GTZAN, segundo o teste <i>Tukey</i> . Os pares de gêneros Hiphop-Disco, Metal-Reggae, Metal-Blues, Reggae-Blues, Reggae-Pop, Blues-Pop, Blues-rock, Pop-Rock, Rock-Jazz, não possuem diferenças significativas em suas médias. Isto significa que nove pares de gêneros não tem diferenças significativas entre seus $\langle \alpha_{DFA} \rangle$. Fonte: o autor.	47

4.9	O gráfico mostra o expoente α em função do tempo para dez <i>tracks</i> dos gêneros Classic, Blues e Hiphop do banco GTZAN. Pode-se notar diferenças no comportamento gráfico para cada grupo, associadas às características rítmicas de cada gênero. No gráfico (a) os eventos percussivos são geralmente mais suaves, resultando um valor médio de alfa superior aos outros dois gêneros em questão. Em (b), observa-se uma grande instabilidade de escala, associado a eventos percussivos mais intensos e a batidas mais regulares que caracterizam o Hiphop. Em (C) nota-se um comportamento intermediário no gênero Blues, pois apresenta eventos percussivos mais intensos que no Classic, porém mais moderados que no Hiphop. Fonte: o autor.	48
4.10	DFA expoente dos gêneros Classic e Hiphop. Fonte: o autor.	49
4.11	DFA expoente médio do banco BRMUSIC. As barras de erro correspondem ao desvio-padrão. Fonte: o autor.	50
4.12	Boxplot dos gêneros musicais do banco BRMUSIC. Fonte: o autor.	50
4.13	Histograma do banco BRMUSIC. Fonte: o autor.	51
4.14	Comparação de médias do banco BRMUSIC. Segundo o teste <i>Tukey</i> vinte e quatro pares de gêneros não tem diferenças significativas entre seus $< \alpha_{DFA} >$. Fonte: o autor.	52
4.15	DFA expoente dos gêneros Bossa-Nova e Mangue-Beat ordenados em função dos seus respectivos tracks. Fonte: o autor.	53
4.16	DFA expoente médio com desvio-padrão dos subgêneros do Axé Music. Fonte: o autor.	53
4.17	DFA expoente dos subgêneros do gênero Pagode. Fonte: o autor.	54
4.18	DFA expoente dos subgêneros do gênero Forró. Fonte: o autor.	54
4.19	DFA expoente da programação musical de rádios FM. Fonte: o autor.	55
4.20	DFA expoente das faixas do CD Alfagamabetizado, de Carlinhos Brown. Fonte: o autor.	56
4.21	DFA expoente das faixas do CD Caçador de Mim, de Milton Nascimento. Fonte: o autor.	57
4.22	DFA expoente das faixas do CD Afrociberdelia, de Chico Science. Fonte: o autor.	58
4.23	DFA expoente das faixas do CD João Voz e Violão. Fonte: o autor.	59
4.24	DFA expoente das faixas do CD Song of songs, do grupo Stile Antico. Fonte: o autor.	60
4.25	Expoente α_{DFA} médio em dos 20 tracks do gênero Forró relativo aos <i>tracks</i> 1, 3, 5, 9, (a) antes da edição e (b) após a edição. Fonte: o autor.	61
4.26	Função $\alpha(t)$ dos tracks 9, 5, 3 e 1 e seus expoentes α_{DFA} , antes e depois da edição.	62

Lista de Siglas

AIFF	Audio Interchange File Format
AIFF	Audio Interchange File Format
CAD	Conversor Analógico-Digital
CD	Compact Disc
CDA	Conversor Digital-Analógico
CODEC	Contração das palavras coder / decoder
DFA	Detrended Flutuuation Analysis
DVFE	Detrended Variance Fluctuation Exponent
GTZAN	George Tzanetakis
IBM	International Business Machines Corporation
IFF	Interchange File Format
LLD	Low-Level Descriptors
MFCC	Mel Frequency Cepstral
MIR	Music Information Retrieval
MP3	MPEG-1/2 Audio Layer 3
MPB	Música Popular Brasileira
MPEG-7	Moving Picture Experts Group Phase 7
PC	Personal Computer
PCM	Pulse Code Modulation
PPGMCTI ..	Programa de Pós-graduação em Modelagem Computacional e Tecnologia Industrial
RIFF	Resource Interchange File Format
RMS	Root Mean Square
STFT	Short Time Fourier transform
WAV	Wave Form Audio Format
WAVE	Wave Form Audio Format
WWW	World Wide Web

Introdução

Devido à necessidade de classificação, busca, recuperação e tratamento do crescente número de arquivos musicais na internet, o estudo da extração de características de sinais de áudio tem ganhado, nos últimos tempos, grande importância. A execução destas tarefas manualmente, pode resultar em uma grande demanda de tempo e recursos financeiros. Portanto, é necessário desenvolver ferramentas computacionais que sejam capazes de manipular um grande conjunto de arquivos de música em formato digital automaticamente, sem a necessidade da interferência humana. Os modelos de classificação automática se baseiam na extração de parâmetros ou vetores de características, que devem fornecer o máximo possível de informação relativa aos sinais de áudio. Esses parâmetros são utilizados na classificação dos sinais musicais de acordo com um critério pré-estabelecido (PEREIRA, 2009a). A extração de características consiste em computar uma representação numérica compacta que pode ser usada como característica de um segmento de áudio. Para isso são utilizados vários tipos de descritores musicais, que são modelos computacionais que buscam prever aspectos musicais a partir de arquivos de áudio. Uma vez que as características são extraídas, podem ser utilizadas técnicas de aprendizado de máquina, utilizando os vetores característicos a fim de realizar a mineração de dados em bancos de arquivos de áudio.

O desenvolvimento de descritores musicais conta com diversas abordagens de pesquisa. FORNARI (2010) destaca: descritores de baixo nível, ou LLD (Low-Level Descriptors), utilizados industrialmente pelo padrão MPEG-7 (FORNARI, 2010 apud MANJUNATH, 2002); descritores de aspectos emocionais e afetivos, investigados por campos de pesquisa como o MIR (Music Information Retrieval), que combina musicologia, psicologia, processamento de sinais e aprendizado de máquina; descritores que visam a classificação de gêneros musicais (FORNARI, 2010 apud TZANETAKIS; COOK, 2002); e até descritores de alto nível que estudam de aspectos gestuais na performance musical (FORNARI, 2010 apud LEMAN, 2004). Muitos trabalhos têm dedicado especial atenção aos aspectos rítmicos maiores detalhes em BEROIS (2008), (CARVALHO, 2009 apud FROOTE; COOPER, 2001), (CARVALHO, 2009 apud ROY P.; PACHET; KRAKOWSKI, 2007). Uma novo descritor para esse campo de pesquisa, denominado de Detrended Variance Fluctuation Exponent, foi introduzido por JENNINGS (2004), e ao quantificar as propriedades de escala de variâncias locais em séries não-estacionárias (que são sinais de áudio neste caso), foram encontradas diferenças quantitativas, associadas a gêneros musicais. Essas diferenças têm relação com propriedades rítmicas do sinal musical.

Esta dissertação se propõe a avaliar o potencial do Detrended Variance Fluctuation Ex-

ponent como um vetor característico unidimensional na caracterização de arquivos pré-organizados em gêneros musicais, bem como verificar a consistência dos resultados obtidos na literatura, onde geralmente se utilizam bancos de dados com os mais diversos critérios de classificação e padronização técnica. Para isso os seguintes objetivos específicos são definidos: criar um banco de dados para pesquisa com arquivos digitais de música brasileira; aplicar o modelo computacional do DFA expoente em dois bancos de dados (um inédito com arquivos de música brasileira e outro já utilizado em pesquisa MIR); analisar os resultados obtidos da aplicação do algoritmo nos dois bancos de dados e comparação com resultados obtidos na literatura; apresentar propostas de utilização do DFA expoente em um novo contexto de agrupamento para arquivos musicais; comparação dos resultados obtidos com outros encontrados na literatura.

1.1 Importância da pesquisa

A pesquisa relacionada à extração de características de sinais de áudio é ainda recente no contexto mundial e ainda há muito o que se explorar. As técnicas de extração de parâmetros ainda estão se consolidando e existem poucos bancos de dados públicos disponíveis, específicos para pesquisa, com características técnicas padronizadas para se testar ou calibrar os modelos. Este trabalho contribui com este tipo de pesquisa, na medida em que apresenta o estudo de um descritor que utiliza a análise de séries temporais, ao invés de transformadas de Fourier, como em muitos descritores, e que ainda é um modelo pouco usado na extração de vetores característicos. A dissertação apresenta também para o contexto da pesquisa um banco de dados inédito com música brasileira, podendo contribuir para a evolução de outras pesquisas e teste de outros modelos.

1.2 Limites e limitações

O estudo de extração de características de um determinado conjunto de dados é geralmente feito a partir de vetores de características com n dimensões ($n > 1$). Usualmente um vetor de características unidimensional não é suficiente para descrever corretamente um conjunto de sinais. Nesta pesquisa, o DFA expoente é utilizado como um vetor unidimensional para extração de características, porém essa aplicação é feita em um agrupamento de arquivos pré-classificados por gêneros. Se por outro lado, considerarmos os arquivos dos bancos de dados sem nenhum rótulo e aplicarmos o algoritmo, não será possível identificar a que gênero musical ele pertence, pois assim esse vetor unidimensional não seria suficiente. Outro aspecto que deve-se levar em conta é que, como estamos falando de método estatístico, é possível a ocorrência de *outliers* dentro de um determinado agrupamento.

Como os gêneros musicais não têm definições rigorosas entre os pesquisadores de música e não existe também definição de um esquema de classificação de gêneros puramente para pesquisa MIR, as interpretações dos números que o algoritmo retorna relativos ao gênero musical que aquele arquivo pertence podem variar um pouco de pesquisador para pesquisador. Em outras palavras, um determinado arquivo que um pesquisador rótula como samba, outro pesquisador pode considerar por exemplo como MPB (Música Popular Brasileira).

1.3 Aspectos metodológicos

O desenvolvimento da pesquisa teve como base o algoritmo derivado do DFA (Detrended Fluctuation Analysis), proposto por [JENNINGS \(2004\)](#), denominado de DVFE (Detrended Variance Fluctuation Exponent) ou DFA expoente. Este algoritmo quantifica as propriedades de escala de variância local em séries temporais não-estacionárias. Neste trabalho as séries temporais são os sinais armazenados em arquivos digitais de música. O primeiro passo é calcular o desvio-padrão da série original, que vai gerar uma nova série $v(j)$. A variância $v(j)^2$ representa a intensidade média do som ou loudness percebido pelo nosso sistema auditivo. A partir de $v(j)$ é calculada uma função $\alpha(t)$ que quantifica correlações na escala de tempo t .

A pesquisa foi desenvolvida seguindo as seguintes etapas:

1. Implementação do algoritmo:

- Revisão da literatura;
- Programação utilizando a linguagem C.

2. Criação do banco BRMUSIC:

- Definição dos gêneros a serem trabalhados dentro da música brasileira;
- Coleta e seleção dos arquivos de áudio dos gêneros a serem trabalhados;
- Conversão dos arquivos para formato WAV, caso necessário;
- Edição dos arquivos.

3. Aplicação do algoritmo e cálculo do expoente DFA nos bancos BRMUSIC e GTZAN

- *Downsampling* dos arquivos;
- Mudança da taxa de quantização;
- Conversão do formato WAV para txt;
- Cálculo do Expoente DFA.

4. Análise e descrição estatística dos resultados

1.4 Organização da Dissertação de mestrado

- **Capítulo 1 - Introdução:** Contextualiza o âmbito, no qual a pesquisa proposta está inserida. Apresenta, portanto, a definição do problema, objetivos e justificativas da pesquisa e como esta dissertação de mestrado está estruturada;
- **Capítulo 2 - Áudio, Gêneros Musicais e Extração de Parâmetros:** Aborda conceitos básicos de áudio analógico e digital; percepção rítmica e tonal; gêneros musicais e MIR.
- **Capítulo 3 - Análise de Séries Temporais:** Apresenta definições de séries temporais e exemplos variados, aborda sobre os principais tópicos da análise de séries temporais e apresenta o método DFA;
- **Capítulo 4 - Implementação e Resultados:** Mostra o processo de implementação do algoritmo utilizado, bem como os resultados da pesquisa;
- **Capítulo 5 - Considerações finais:** Apresenta conclusões, contribuições e atividades futuras de pesquisa;
- **Capítulo 5 - Considerações Finais:** Apresenta as conclusões, contribuições

Áudio, Extração de Parâmetros e Classificação em Gêneros Musicais

Os sinais sonoros representam um papel social muito importante. Através deles, podemos nos comunicar e interagir com o meio-ambiente, extraindo informações e dando respostas a vários tipos de estímulos advindos dessa fonte. Uma vez que recebemos externamente o estímulo sonoro através da audição e processamos internamente através da nossa cognição, podemos considerar os sinais sonoros como informação multidimensional, cujos aspectos podem ser estudados em categorias como afeto, percepção e cognição (FORNARI, 2010). A representação desses sinais no formato digital se tornou muito importante graças ao crescimento da influência e do poder da computação digital. Através desta representação, é possível realizar, entre outros processos, a análise, o processamento e a síntese de arquivos digitais. Neste capítulo, veremos alguns princípios que são importantes para compreender como se processa a representação digital do sinal sonoro e algumas de suas influências na cognição humana. Além disso, serão abordados alguns aspectos com respeito à classificação de músicas por gêneros musicais.

2.1 O Paradigma dos Quatro Universos

Segundo CARVALHO (2009), a compreensão do estudo e análise de sinais sonoros podem ser realizados sob um paradigma de abstração denominado de *paradigma dos quatro universos*. Os sinais do mundo físico necessitam de uma descrição matemática a fim de se construir representações discretas dos sinais, e assim, permitir que se realizem implementações de técnicas de síntese, análise e processamento computacionais. Com base no paradigma dos quatro universos, temos três níveis de abstração conhecidos como: sinais contínuos, sinais discretos e sinais codificados. As mudanças de um nível para o outro são efetuadas através das operações: discretização, codificação, decodificação e reconstrução (Figura 2.1). A *codificação* consiste em obter, a partir do sinal discreto, uma representação que pode ser descrita por um conjunto finito de símbolos organizados em concordância com uma estrutura de dados. A *decodificação* é o processo que permite o acesso às informações codificadas no formato discreto. Quando ocorre do sinal codificado ser igual ao sinal discretizado existe aí uma codificação sem perdas, caso contrário, temos uma codificação com perdas. Outros dois problemas importantes na codificação e decodificação são o tempo de processamento e o espaço ocupado pelo código. *Discretização* é o processo onde o sinal contínuo é convertido em uma representação discreta. O processo inverso é chamado de *reconstrução* e consiste em obter um sinal contínuo a partir de uma

representação discreta.



Figura 2.1: Paradigmas de abstração no estudo de sinais (CARVALHO, 2009)

2.2 Áudio Analógico

Um sinal de áudio é uma representação elétrica de uma energia acústica produzida pelo som. Este tipo de energia é causado por variações de pressão, contínuas no tempo, em um meio físico (usualmente o ar) (GUADARRAMA, 2010 apud DAVIS G.; JONES, 1989). Os sinais de áudio podem ser representados através de um modelo funcional

$$f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n.$$

O subespaço constituído por essas funções f é chamado de espaço de sinais, que é um espaço vetorial em relação às operações de soma e multiplicação por um escalar, e cada função f é chamada de sinal contínuo. No campo da engenharia os sinais contínuos são chamados de sinais analógicos. Em termos computacionais o conjunto dos números reais tem sua representação através de um conjunto finito que utiliza aritmética de ponto flutuante (CARVALHO, 2009).

2.3 Áudio Digital

Nos dias de hoje, a maior parte do processamento e análise de sinais é feita com o uso de equipamentos programáveis baseados em circuitos digitais. Para realizar esse tipo de processamento é necessário que a informação esteja na forma de uma série temporal numérica. Desta forma, o sinal contínuo precisa de uma representação discreta ou digital. Essa representação é então realizada através um sistema de conversão analógico-digital (CAD), que executa as operações de *amostragem* (ou *sampleamento*) e *quantização* do sinal contínuo (Figura 2.2) (GUADARRAMA, 2010 apud ROBERTS, 2004). *Samplear* significa capturar os valores de um sinal contínuo em pontos discretos no tempo. Por outro lado, *quantização* significa ajuste de valores de amplitude de sinais discretos para valores fixos denominados de níveis. Esses níveis de quantização estarão em uma extensão de -2^{n-1} a $2^{n-1} - 1$ onde n é o número de bits de quantização. Usualmente, essa extensão é normalizada entre -1 e 1. São geralmente utilizados no processo de quantização os valores 16 e 24 bits (GUADARRAMA, 2010).

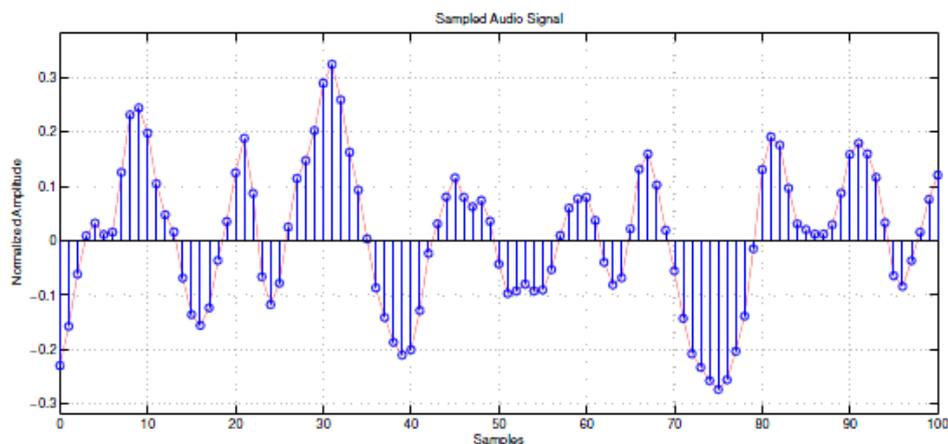


Figura 2.2: Sinal Contínuo (em vermelho) e seu equivalente Sinal Digital (em azul). Fonte: (GUADARRAMA, 2010)

Para restaurar o sinal de áudio original, é necessário que os sinais digitais sejam reconvertidos em sinais analógicos por meio de um conversor digital-analógico (CDA). No mundo ideal o processo de reconstrução deveria ser capaz de trazer de volta o sinal contínuo original, porém, de um modo geral, isso não se realiza. Daí surge um dos problemas mais importantes do processamento de sinais: a obtenção de métodos que tornem o sinal reconstruído o mais próximo do sinal original (CARVALHO, 2009).

Como exemplo consideremos um CD de áudio que é, geralmente, amostrado com uma frequência de 44.100 Hz e quantizado com 16 bits. A frequência de 44.100 Hz corresponde a um intervalo de amostragem (*sampling interval*) de aproximadamente $23\mu s$. Assim para cada 1 min do arquivo tem-se 2.646.000 samples. Já os 16 bits significam que é possível usar 32.767 níveis de tensão para ajustar o sinal discretizado ao sinal original. A Figura 2.3 mostra o processo de digitalização, quantização e reconstrução de um sinal analógico.

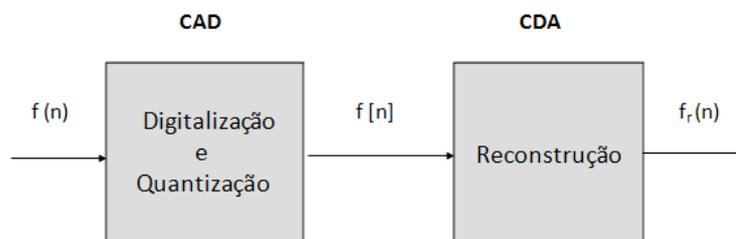


Figura 2.3: Digitalização, quantização e reconstrução de um sinal analógico num sistema de conversão analógico-digital (CAD) e digital-analógico (CDA). $f(n)$: sinal analógico; $f[n]$: sinal digital; $f_r(n)$: sinal reconstruído. Fonte: o autor.

Do ponto de vista matemático, o processo de amostragem de um sinal consiste em calcular a função f , associada a um sinal de áudio analógico, em um número finito de amostras

p_1, p_2, \dots, p_k do domínio de f , enquanto que no processo de quantização, a discretização acontece no contradomínio. A reconstrução do sinal consiste em utilizar uma técnica de interpolação a fim de se obter o sinal original f , ou uma aproximação f_r , a partir dos valores $f(p_1), f(p_2), \dots, f(p_k)$ de f nas amostras (CARVALHO, 2009).

2.4 Teorema de Nyquist-Shannon (Teorema da Amostragem)

Sejam f_n (frequência de Nyquist), a maior frequência presente em um sinal contínuo definido, e f_s a frequência de amostragem (ou frequência de sampleamento). O teorema de amostragem estabelece que, se um sinal contínuo é sampleado com $f_s \geq 2f_n$, o sinal contínuo original pode ser reconstruído a partir de suas amostras. Por haver frequências contidas acima de f_n , um fenômeno conhecido como *aliasing* ocorre, onde frequências mais altas do que $f_s/2$ são reconstruídas com valores de frequência mais baixos.

Se considerarmos o espectro sonoro audível do ser humano na faixa de 20 a 20.000 Hz, o f_s mínimo para sinais de áudio é de 40.000 Hz. No entanto, um valor de f_s de 20.000 Hz também é válido para os sinais de áudio musicais. Instrumentos de música tradicional produzem sons definidos chamados notas. Cada nota é caracterizada por possuir uma frequência fundamental, que é percebida pelo ouvido humano como *pitch* (tom). Essas frequências fundamentais, para instrumentos tradicionais, estão abaixo de 10.000 Hz. Samples em estúdios de áudio profissional possuem 96.000 Hz, mas são *downsampleados* a 22.050 Hz ou 44.100 Hz ao serem transferidos para os formatos de CD ou MP3. Assim, o teorema da amostragem de Nyquist-Shannon diz que a taxa de amostragem deve ser pelo menos duas vezes maior que a frequência que se deseja registrar, e esse valor é conhecido como *Frequência de Nyquist*.

A Figura 2.4 representa uma onda de 17.500 Hz (em amarelo) digitalizada com uma taxa de amostragem de 20.000 Hz. Cada amostra é representada pelos pontos verdes. A onda em azul é a onda resultante do efeito de *aliasing* (IAZZETA, 2011a).

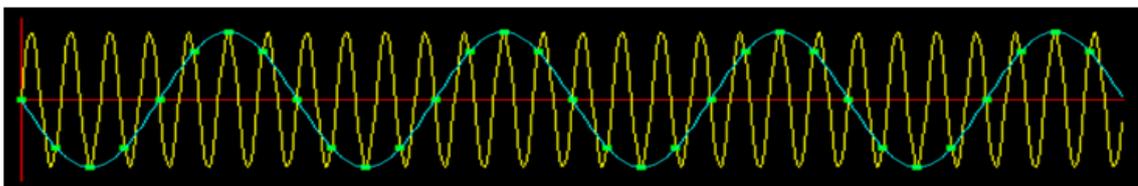


Figura 2.4: Efeito de *aliasing*. Fonte: IAZZETA (2011a).

2.5 PCM (Pulse Code Modulation)

Pulse Code Modulation (PCM), ou Modulação por Código de Pulso, é um método utilizado para realizar a representação digital de um sinal analógico. Neste método a magnitude do sinal é amostrada em intervalos de tempo iguais, onde cada amostra é quantizada para o valor mais próximo dentro de um intervalo finito de amplitudes, e finalmente o sinal é codificado digitalmente (FIBEROPTICS, 2012). A Figura 2.5 apresenta uma ilustração das três etapas que constituem a técnica PCM.

O sistema PCM foi patenteado em 1937 pelo engenheiro britânico Alec Reeves e foi utilizado primeiramente em sistemas de telefonia, onde o sinal de voz é digitalizado pela central telefônica utilizando a técnica PCM, e transmitido no formato digital entre as centrais, sendo novamente convertido em sinal analógico ao chegar na central telefônica do destinatário (Figura 2.6). O uso da informação digitalizada através do sistema PCM tem sido utilizada atualmente em diversas aplicações como: áudio CD, DVD, vídeo telefonia, voz e imagem na internet, telefonia digital fixa e celular, TV digital, etc. (MOECKE, 2004).

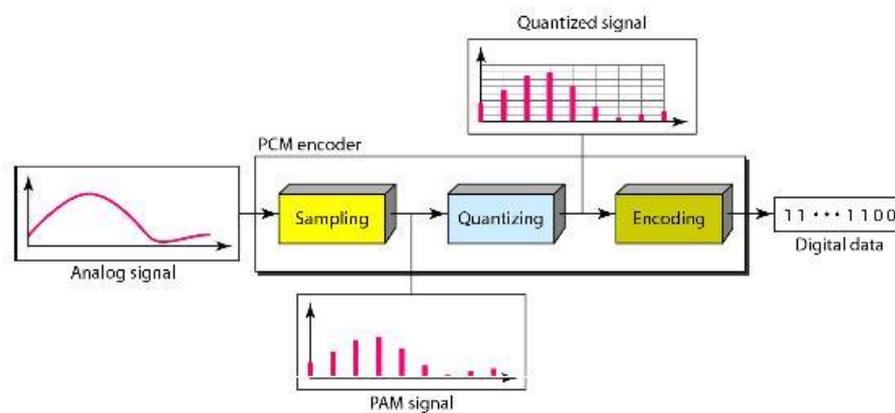


Figura 2.5: No estágio *sampling* é realizada a amostragem e o sinal analógico (*analog signal*) se transforma no sinal PAM (*Pulse Amplitude Modulation Wave*). No segundo estágio (*quantizing*) é realizada a quantização do sinal, e no terceiro, a codificação (*encoding*) resultando finalmente no sinal digital. Fonte: OLIVEIRA (2012).

2.6 Formatos no Áudio Digital

Arquivos de áudio digital são uma coleção de amostras que podem ser armazenadas e editadas segundo um formato padrão. Portanto, essa coleção pode ser armazenada em drives de computadores, transferida para outros computadores ou *samplers*, compartilhada na internet para ser baixada, ou ser utilizada em jogos em tempo real. Os arquivos de áudio são diferentes de faixas de CD, que na sua maioria contêm apenas dados *raw* ou itens

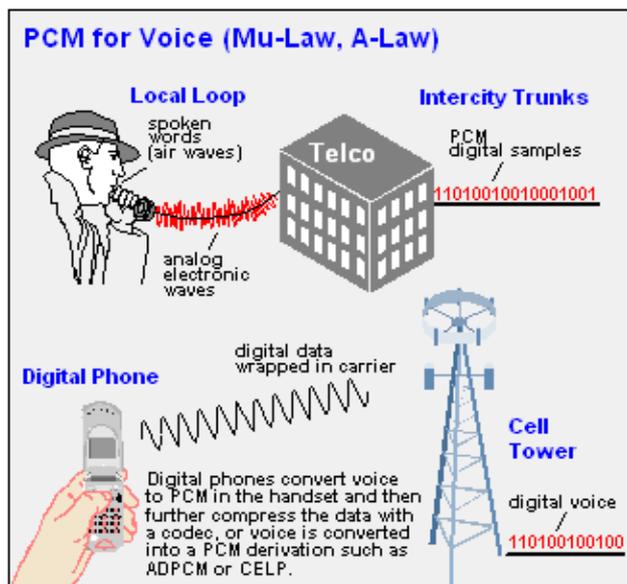


Figura 2.6: A técnica PCM foi introduzida nos EUA na década de 1960, quando as empresas de telefonia começaram a conversão de voz para o sistema digital, que seria utilizada na transmissão em troncos interurbanos. Fonte: FARLEX (2011).

inúteis para um computador, tais como correção de erros e dados de subcode para auxiliar a leitura de laser do CD, etc. Por isso é que uma faixa de CD deve ser extraída para um formato de arquivo de áudio a ser usado por um aplicativo de computador. Quando o objetivo é economizar espaço em disco e minimizar o tempo de *download*, pode-se buscar a utilização de formatos que usem alguma técnica de compressão, a fim de diminuir o tamanho do arquivo (INDIANA, 2012). A seguir serão apresentados alguns dos principais formatos de arquivos de áudio:

Formatos Sem Compressão:

- WAVE (*Wave Form Audio Format*) é um formato de áudio digital baseado no *Resource Interchange Format* (RIFF). O WAVE é um formato nativo do Windows, mas agora é utilizável com a maioria dos programas de áudio em Machintosh ou PC. Usando a extensão do arquivo .wav, amostras de 8 bit ou 16 bit podem ser recolhidas com taxas de 11.025 Hz, 22.050 Hz e 44.100 Hz. A qualidade mais elevada, isto é, 16 bit com 44.100 Hz, correspondem à taxa de quantização e amostragem de um CD de áudio e usa 88kb de armazenamento por segundo.

O formato WAV é amplamente utilizado como meio de áudio para gravação e edição profissional. Para criar CDs de música, arquivos WAV são convertidos para o formato de áudio CD-DA, e ambos arquivos (WAV e CD-DA) ocupam a mesma quantidade de espaço de armazenamento em disco (PCMAG, 2012b).

- AIFF (*Audio Interchange File Format*), é o formato de arquivo de áudio padrão

usado pela Apple. Transita bem entre quase todos os computadores e softwares, inclui informações de cabeçalho, como nome do arquivo, taxa de amostragem, número de nota MIDI para samplers, pontos de loop, número de bytes no arquivo. Também capaz de uma resolução de 24 bit e 32 bit. O AIFF Poderia ser considerado como o equivalente da Apple para o wave.

Formatos Com Compressão:

- MP3 (MPEG Layer I-3). Em 1987, o instituto de pesquisas britânicas *Fraunhofer II* começou a trabalhar na codificação de áudio perceptiva no âmbito do projecto EU-REKA. Em uma cooperação conjunta com a Universidade de Erlangen, o *Fraunhofer II* finalmente concebeu um algoritmo muito poderoso que é padronizado como ISO-MPEG Audio Layer-3. Com os codecs adequados, podem ser conseguidas taxas de compressão de até 24 vezes, com qualidade próxima a de um CD. Uma característica do MP3 é que se percebe a qualidade em função do tamanho. Outra característica é também sua grande capacidade de ser baixado e carregado na memória flash de tocadores de MP3. Quando arquivos são codificados em certas taxas de bits, para atingir velocidades de download, podem atingir uma boa qualidade com, por exemplo, 160 kbps de codificação.
- WMA (*Windows Media Áudio*) é um método de compressão de áudio da Microsoft. Conhecido originalmente como MSAudio, este formato proprietário concorre com os métodos de MP3 e AAC (Advanced Audio Coding). Vale observar que o WMA, assim como o MP3, é um formato *lossy*, isto é, ocorre nele a remoção de determinados harmônicos e frequências contidas no fluxo de áudio original durante o processo de compressão.
- FLAC (*Codec Free Lossless Audio*) é uma tecnologia de compressão de áudio da Fundação Xiph.Org (www.xiph.org). É um codec livre de royalties usado para comprimir arquivos de CD de áudio para pelo menos 50% do seu tamanho original (2:1). Ao contrário do MP3, que proporciona uma compressão muito maior, mas elimina conteúdo, o FLAC restaura a música para o PCM original com uma taxa de amostragem 44kHz e 16-bit, padrão de um CD de áudio ([PCMAG, 2012a](#)).
- AAC (*Advanced Audio Coding*) é um esquema de codificação para compressão com perda de dados de som digital. Projetado para ser o sucessor do formato MP3, o AAC geralmente consegue melhor qualidade de som do que o MP3 em bitrates similares ([INDIANA, 2012](#)).
- AU é definido como um formato de áudio simples, e foi introduzida pela mesma empresa que criou a linguagem de programação Java - a *Sun Microsystems*. Os usuários do sistema operacional NeXT e páginas da web em seu período inicial podem estar

familiarizados com a extensão do arquivo da AU, uma vez que os arquivos de áudio AU foram muito utilizados por eles. Às vezes, este formato de arquivo é chamado de *uLaw*, já que ele é simplesmente um de 8-bit *u-law-encoded data* a uma taxa de cerca de 8000 Hz. A extensão do arquivo da AU é considerada simples, porque, em comparação com os mais recentes formatos de arquivos de áudio, este não é usado com 8192 Hz, com um cabeçalho, ou com palavras de 32 bits. Uma desvantagem sobre a extensão de arquivo da AU é que novos *players* de áudio digital não conseguem reconhecê-lo, uma vez que é mais velho do que os mais utilizados atualmente. É por isso que os usuários deste formato de arquivo muitas vezes precisam de conversores para deixar seus arquivos com a extensão da AU compatíveis com o seu leitor de áudio digital. Além da codificação com perdas o formato AU também admite codificação sem perdas (MAZZONI, 2012).

2.7 Loudness

Loudness é uma medida subjetiva de caráter fisiológico que descreve a sensação de volume ou intensidade sonora. Filtros de ponderação, levando em conta a pressão sonora e frequência são utilizados para tentar ajustar estas medidas às intensidades percebidas pelo ser humano. Enquanto medida subjetiva, a *loudness* depende de fatores como por exemplo, sensibilidade do indivíduo, tipo de sinal sonoro, frequência e envoltória ADSR (Ataque, Decaimento, Sustentação, e Repouso) do sinal. O nível mínimo de variação de pressão (nível mínimo que se pode escutar), chamado de limiar da percepção, é de $20\mu Pa$, correspondente a $0dB$ de pressão acústica. O nível máximo, ou limiar da dor (onde a intensidade sonora é tão grande que provoca sensação de desconforto ou dor), corresponde a $120dB$ (FORNARI, 2010). A Figura 2.7 mostra a equivalência entre níveis em dB (Decibéis) e Pa (Pascais).

A unidade de loudness é chamada de *phon* e as curvas cujo loudness se mantém constante foram chamadas de curvas isofônicas, conforme mostradas na Figura 2.8. Estes experimentos também ajudaram a mapear os limites de percepção sonora para sons simples, no que concerne a sua intensidade, em dB SPL e frequência em Hz, que determinam, como limites aproximados da audição humana, intensidades sonoras variando entre 0 e 120 dB e frequências entre 20 e 20 kHz (FORNARI, 2010). A Figura 2.9 mostra os níveis de intensidade sonora de objetos e cenários conhecidos, variando do limiar da audição ao limiar da dor.

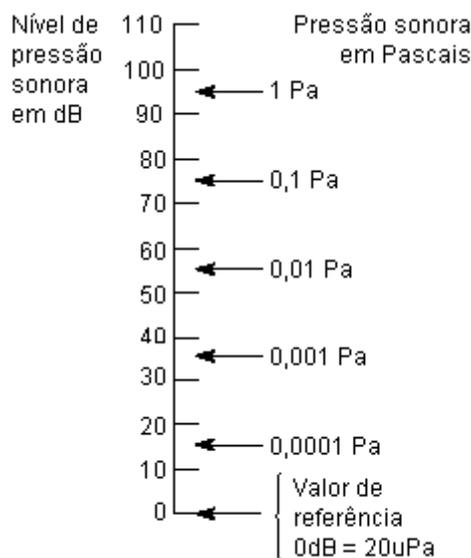


Figura 2.7: Comparação entre decibéis e pascais. Fonte: [AUDIOLIST \(2011\)](#).

2.8 Percepção Rítmica e Tonal

A palavra *ritmo* tem origem no termo grego *rhythmos* e significa agrupamento de valores de tempo, combinados de maneira que marquem com regularidade uma sucessão de sons fortes e fracos, de maior ou menor duração, conferindo a cada trecho características especiais; marcação de tempo própria de cada forma musical: ritmo de marcha, de valsa, de samba ([FERREIRA, 1986](#)). Segundo [CARVALHO \(2009\)](#), se adotada uma perspectiva ingênua, pode-se definir música como uma distribuição de sons no tempo. Ritmo então seria o fenômeno que emerge dessa distribuição temporal. Contudo, vale à pena pontuar que a experiência musical deve ser vista como um todo e não isoladamente, assim, parâmetros unicamente temporais não são suficientes para o entendimento do conceito de ritmo. [CARVALHO \(2009\)](#) também define alguns conceitos, baseados em pressupostos da música de concerto europeia ocidental do período do século XVII ao XIX (denotada aqui como música clássica), cuja teoria musical é utilizada para análise de peças do período e que podem ajudar a entender a estruturação do fenômeno musical independente do período histórico. São esses conceitos:

- **Grupamento:** Ao escutar um sinal musical os seres humanos tem a capacidade de segmentá-lo em notas, que se agrupam em motivos, temas, seções e, por último em peças musicais completas. Essa capacidade é fundamentada pelos princípios da *gestalt*.
- **Métrica:** Dentro de uma frase musical algumas notas são mais acentuadas que as outras. Essas notas são quase sempre a primeira da cada compasso e ocorrem pe-

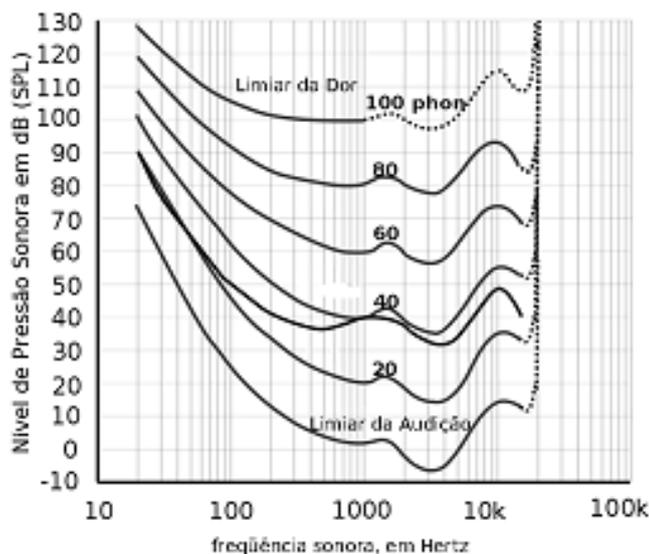


Figura 2.8: Curvas de igual intensidade perceptual sonora para sons simples. Observa-se na figura que o ouvido é mais sensível à intensidade sonora dos parciais com frequências medianas, entre 1 kHz a 5 kHz, pois esta é a região de frequência relacionadas aos parciais gerados pela voz humana, o que garante um melhor entendimento das linguagens. Para sons complexos, a percepção da intensidade sonora é correspondente ao loudness resultante de todos os seus parciais (FORNARI, 2010).

riodicamente. Além disso, a duração de cada nota é normalmente uma subdivisão em partes iguais do período de acentuação. A estrutura métrica é, portanto, uma estrutura hierárquica formada por vários níveis de subdivisão, do compasso até as notas musicais.

- Pulso ou Tactus: Série de estímulos isocrônicos que, preferencialmente, tem um período próximo de 600 ms. Informalmente, seria a frequência com que o maestro move a batuta ou a frequência que batemos o pé ao ouvirmos música. Em relação a estrutura métrica, seria um nível especial de referência, onde os eventos mais frequentes que o pulso seriam vistos como subdivisões deste e os menos frequentes, como grupamentos.

A experiência musical e a sua representação escrita mais utilizada formalmente, a partitura, se distanciam no que se refere a uma característica denominada de *timing* ou expressividade. Quando um intérprete executa uma música através da leitura de uma partitura, ele depende de suas possibilidades motoras, cognitivas e de suas influências culturais. Com isso, por exemplo, o fluxo rítmico de uma música obtido por um indivíduo, pode se diferenciar do fluxo obtido por outro, graças a características pessoais. Isso ocorre também devido a indicações escritas na partitura que dão ao intérprete a liberdade de execução, não quantificada exatamente pela partitura. Nesse sentido o *timing* pode ser considerado como um desvio da notação, um erro, uma discrepância, que torna a música

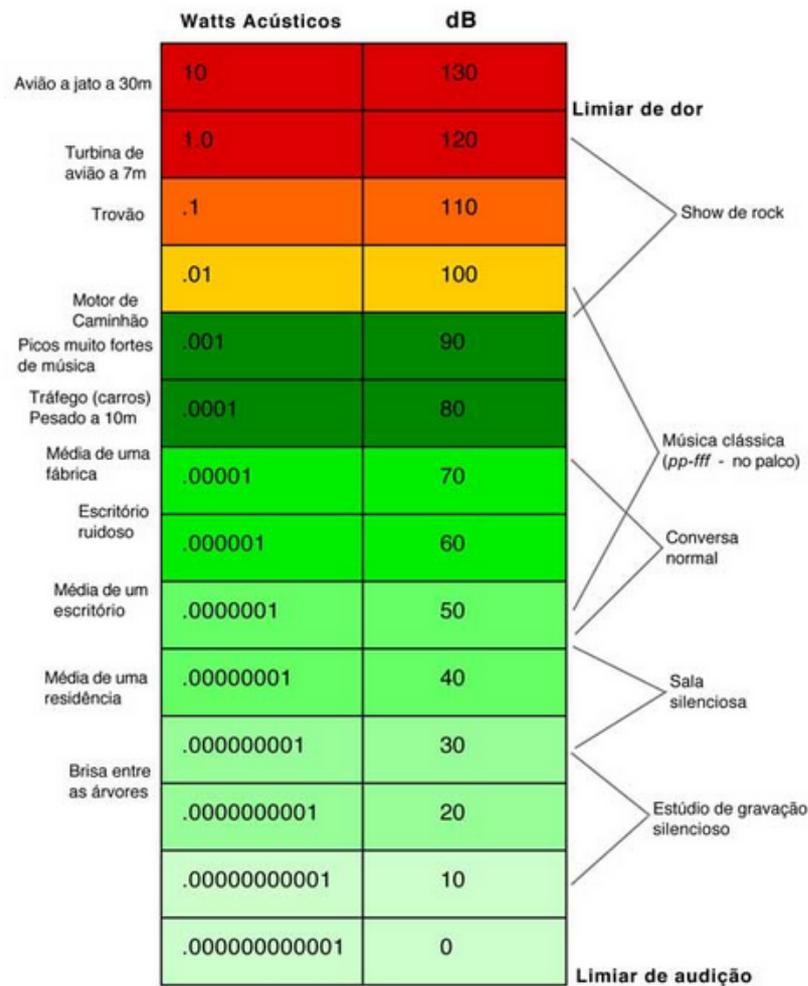


Figura 2.9: Quadro comparativo entre níveis de intensidade e potências acústicas, variando do limiar da audição ao limiar da dor. Aqui são mostrados sons de objetos e cenários que são facilmente identificados pelo senso comum, e seus respectivos níveis de intensidade em db e watts. (IAZZETA, 2011b)

mais humana. Em FRAISSE (1982) é ressaltado o estudo sobre a importância do valor de 600 milissegundos, chamado de valor de indiferença, um limiar que separa qualitativamente o conjunto de segmentos temporais e que serviria como valor preferencial para a marcação física do fluxo musical, como ressaltado na definição de pulso. PARNCUTT (1994) também enfoca de maneira metódica a questão de encontrar o pulso em uma sequência rítmica periódica de estímulos. Mais uma vez o valor de 700 ms foi ressaltado como preferencial (em acordo com os 600 ms de Fraise).

A psicoacústica é uma área da ciência que estuda os aspectos que descrevem a maneira como a informação sonora é captada como estímulo acústico pelo sistema binaural. As características psicoacústicas ocorrem em pequenos intervalos de tempo, anteriores à formação de um modelo de memorização sonora. Esses aspectos estão associados a um

intervalo de tempo de aproximadamente 0,1 s de duração, conhecido como intervalo de persistência auditiva (FORNARI, 2010 apud FELDMAN MARTIN; BENDER, 1972). Isto quer dizer que a predição de eventos sonoros de uma mesma fonte, distanciados no tempo por intervalos maiores que a persistência auditiva, podem apresentar o contexto de ritmo, por outro lado os eventos sonoros separados por intervalos de tempo menores que o da persistência rítmica apresentam apenas o contexto da altura (tom ou pitch) (Figura 2.10).

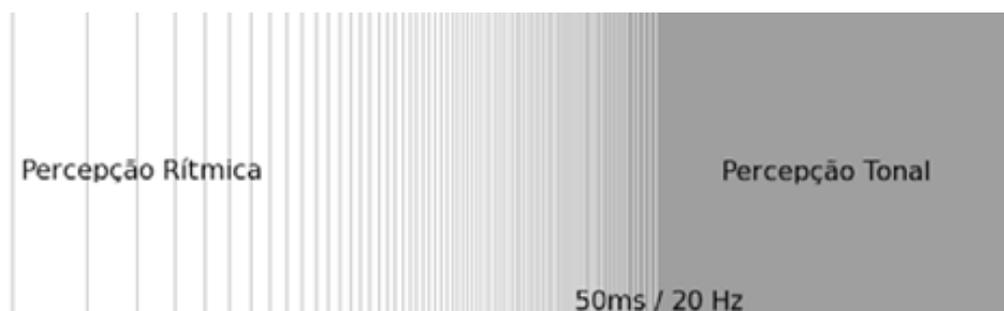


Figura 2.10: A fronteira entre o domínio do contexto rítmico e o contexto tonal encontra-se aproximadamente estendida entre 0,05 e 0,1 s. Em torno de 50ms de período, ou 20 Hz de frequência, a percepção do intervalo de tempo entre eventos sonoros passa de rítmica (domínio do tempo) para tonal (domínio da frequência) (FORNARI, 2010).

2.9 Descritores Musicais

Segundo FORNARI (2010), *descritores musicais* são modelos computacionais que têm o objetivo de prever aspectos ou características musicais a partir de arquivos de áudio. Tais modelos procuram simular a capacidade cognitiva humana de perceber e identificar tais aspectos. O desenvolvimento de descritores musicais conta com diversas abordagens de pesquisa. FORNARI (2010) também destaca : descritores de baixo nível, ou LLD (Low-Level Descriptors), utilizados industrialmente pelo padrão MPEG-7 (FORNARI, 2010 apud MANJUNATH, 2002) ; descritores de aspectos emocionais e afetivos, investigados por grupos como o MIR (Music Information Retrieval), que combina musicologia, psicologia, processamento de sinais e aprendizado de máquina; descritores que visam a classificação de gêneros musicais (FORNARI, 2010 apud TZANETAKIS; COOK, 2002); até descritores de alto nível que estudam de aspectos gestuais na performance musical (FORNARI, 2010 apud LEMAN, 2004). Dentro desse contexto surgem também trabalhos onde o foco principal é o aspecto rítmico, como detecção e descrição de eventos percussivos em sinais polifônicos (BEROIS, 2008), visualização das auto-similaridades de um sinal de áudio (CARVALHO, 2009 apud FROOTE; COOPER, 2001), ou ainda estudos que usam programação genética aplicada a construção de funções características para instrumentos percussivos (CARVALHO, 2009 apud ROY P.; PACHET; KRAKOWSKI, 2007). JENNINGS

(2004) introduz uma nova contribuição para esse campo de pesquisa quando, ao quantificar as propriedades de escala de variâncias locais em arquivos de áudio, foram encontradas diferenças quantitativas, associadas aos gêneros musicais aos quais esses arquivos são classificados.

A percepção musical está relacionada com a audição dos aspectos que descrevem características psicoacústicas do material escutado. Esses são chamados conhecidos como aspectos de baixo-nível, ou não-contextuais. Os aspectos de alto-nível são contextuais como, por exemplo, aspectos afetivos e cognitivos. A terminologia baixo-nível refere-se às características do sinal acústico em si, que ocorrem na escala do limiar da persistência auditiva e definem a percepção do som pelo sistema auditivo. Dessa maneira, percepção sonora e musical são a mesma coisa, uma vez que estes ocorrem antes da formação da memória, numa escala temporal pequena. Os LLD (Low-Level Descriptors) ou descritores de baixo-nível são algoritmos computacionais que têm como objetivo emular aspectos psicoacústicos parecidos com o processamento fisiológico da audição dos seres humanos. Um exemplo de utilização industrial dessas descritores é o padrão MPEG-7, que utiliza 17 LLDs para a descrição de áudio (FORNARI, 2010 apud MANJUNATH, 2002). Este padrão consiste em descritores no domínio do tempo e da frequência, como ataque, intensidade, frequência fundamental, densidade espectral, etc (FORNARI, 2010).

2.10 Extração de Parâmetros

Segundo TZANETAKIS e COOK (2002) o fundamento de qualquer sistema de análise automática de dados é o processo de extração de vetores de características. Este processo consiste em computar uma representação numérica compacta que pode ser usada como característica de um segmento de áudio. Uma vez que os recursos são extraídos, podem ser aplicadas técnicas de aprendizado de máquina, que podem utilizar os vetores característicos a fim de realizar a mineração de dados em bancos de arquivos de áudio. Em TZANETAKIS e COOK (2002) são apresentados alguns dos parâmetros mais usados na extração de informação dos sinais de áudio.

1. **Centróide Espectral:** é considerado o centro de gravidade da magnitude do espectro da STFT (short time Fourier transform). Grandes valores do Centróide Espectral estão relacionadas a texturas mais brilhantes, pois apresentam uma maior quantidade de energia nas componentes de alta frequência.
2. **Rollof Espectral:** calcula a frequência R_i abaixo da qual se concentra 85% da área do espectro de magnitude. Sinais musicais tendem a apresentar valores maiores para este parâmetro que os sinais de voz.

3. Fluxo Espectral: caracteriza a rapidez em que se processam as mudanças no espectro do sinal ao longo dos quadros. Os sinais de voz normalmente apresentam maiores valores para o fluxo espectral do que os sinais musicais, pois apresentam seqüências de fonemas surdos e sonoros.

4. Taxa de Passagem pelo Zero: quantifica o número de cruzamentos pelo zero do sinal no domínio do tempo. Sons inaudíveis ou fricativos possuem baixas quantidades energia, mas altas taxas de passagem pelo zero, enquanto o silêncio ou ruído de fundo, possuem baixa energia e baixas taxas de passagem pelo zero. Sinais musicais tendem a apresentar uma taxa de cruzamentos por zero diretamente relacionada às frequências fundamentais dos instrumentos que estão sendo tocados e normalmente os valores não variam muito ao longo do sinal.

5. MFCC (Mel Frequency Cepstral Coefficients): As medidas cepstrais fornecem informação a respeito do formato do espectro. O termo cepstro é usado para designar a transformada de Fourier inversa do logaritmo do espectro de potência de um sinal. Estes coeficientes são comumente utilizados para aplicações de tratamento de voz. [TZANETAKIS e COOK \(2002\)](#) utilizou cinco dos treze coeficientes típicos MFCC em conjunto com outras características para processos de classificação automática de gêneros musicais, obtendo resultados satisfatórios.

6. Low-Energy Feature: é a única característica que baseia-se na janela de textura em vez da janela de análise. É definida como a porcentagem de janelas de análise que têm menos energia RMS do que a energia média de RMS da janela de textura.

7. Características do conteúdo rítmico: a maioria dos sistemas de detecção automática de batida (ou pulso) fornecem uma estimativa de execução da batida principal e uma estimativa de sua intensidade. Além desses recursos, outras informações de conteúdo rítmico de uma peça podem ser utilizadas para caracterizar gêneros musicais. A regularidade da batida, a relação do batida principal com suas subdivisões, e a força relativa das subdivisões em relação a batida principal, são alguns exemplos de características interessantes para se representar através de vetores de características. Uma das estruturas comuns de um detector automático de batida consiste em uma decomposição usando banco de filtros, seguido por uma etapa de extração de envelope e, finalmente, um algoritmo de detecção de periodicidade (que é usado para detectar o atraso em que envelope do sinal é mais semelhante a si mesmo). O processo de detecção automática da batida lembra detecção de tons com períodos maiores (cerca de 0,5 - 1,5 s para batidas, comparadas com 2 ms a 50 ms para tons). O cálculo de recursos para representar a estrutura rítmica da música é baseada na transformada wavelet, que é uma técnica de análise de sinais que foi desenvolvido como uma alternativa à Transformada Curta de Fourier para superar seus problemas resolução. ([TZANETAKIS; COOK, 2002](#))

2.11 Gêneros Musicais

TZANETAKIS e COOK (2002) definem gêneros musicais como rótulos criados e utilizados por seres humanos para categorizar e descrever o vasto universo da música. Para esses autores, os gêneros musicais não têm definições e limites rigorosos, principalmente por se tratar de um tema que exigiria uma complexa interação entre especialistas, público, marketing e fatores histórico-culturais. Este fato tem levado alguns pesquisadores a sugerirem a definição de um novo esquema de classificação de gêneros musicais puramente para fins de pesquisa MIR (TZANETAKIS; COOK, 2002 apud PACHET F.;CAZALY, 2000b). Os autores BARBETO J.; LOPES (2007) abordam em sua pesquisa que, apesar de não haver unanimidade nesse tipo de classificação, os membros de um determinado gênero compartilham características em comum como estrutura rítmica, instrumentação, conteúdo tonal e outros. O grau de arbitrariedade e incoerência na classificação da música em gêneros foi discutido por PACHET F.;CAZALY (2000a), onde foram comparadas três diferentes taxonomias de gênero da internet: <http://www.allmusic.com> (531 gêneros), <http://www.amazon.com> (719 gêneros), e <http://www.mp3.com> (430 gêneros), chegando a basicamente três conclusões:

- Existem apenas 70 palavras comuns a todas as taxonomias, portanto não existe unanimidade sobre a nomenclatura dos gêneros.
- Nem mesmo nomes bastante usados como Rock e Pop, denotam o mesmo conjunto de músicas.
- As estruturas hierárquicas das três taxionomias são bem diferentes.

BARBETO J.; LOPES (2007) afirma que outro fator importante a ser considerado na discussão sobre classificação de gêneros musicais, é que grande parte das canções produzidas na atualidade possuem elementos de mais de um gênero musical, e que para lidar com esse problema pode-se utilizar uma divisão básica de gêneros em uma série de subgêneros capazes de abranger classes intermediárias. A Figura 2.11 mostra uma estrutura de taxonomia que tem quatro camadas hierárquicas e um total de 29 gêneros musicais nas camadas mais baixas.

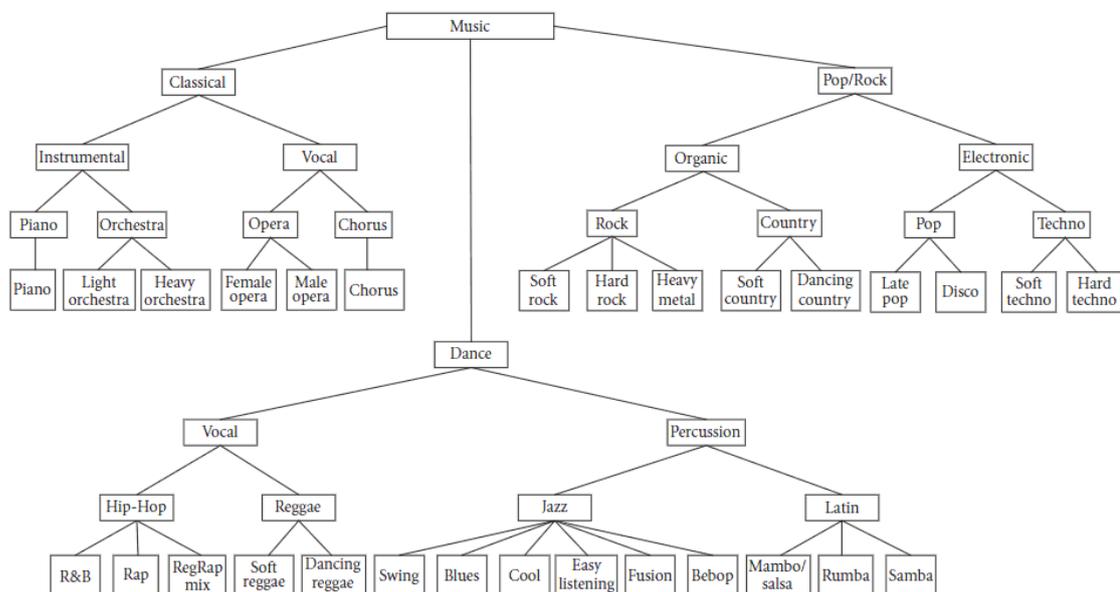


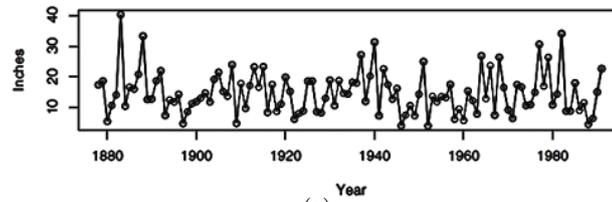
Figura 2.11: Estrutura da taxonomia de gêneros musicais adotada em [BARBETO J.; LOPES \(2007\)](#).

Séries Temporais

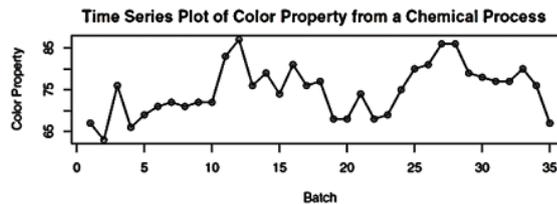
Neste capítulo trataremos de alguns tópicos que envolvem séries temporais. O estudo da análise de séries temporais tem ganhado uma grande abrangência e diversidade. Muitos trabalhos científicos têm sido desenvolvidos nas mais diversas áreas do conhecimento, como a física ZEBENDE G. F.; OLIVEIRA (1998), cardiologia GALHARDO (2010), mercado acionário MUELLER (1996), climatologia BLAIN (2010), entre outras.

3.1 Definições de Séries Temporais

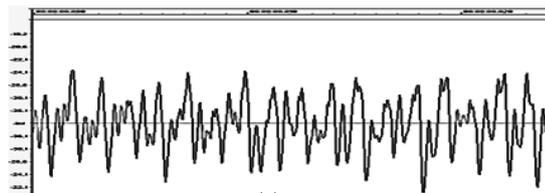
Uma série temporal é um conjunto de informações estatísticas, compiladas, registradas ou observadas em diferentes datas ou períodos de tempo e organizadas cronologicamente (BACHMANN, 2006). A Figura 3.1 mostra alguns exemplos dessas séries.



(a)



(b)



(c)

Figura 3.1: (a) Série temporal dos montantes anuais de precipitação pluviométrica registrado em Los Angeles, Califórnia, ao longo de mais de 100 anos (CRYER; CHAN, 2008); (b) Série Temporal de um processo químico - industrial. A variável medida aqui é uma propriedade da cor em consecutivos banhos durante o processo (CRYER; CHAN, 2008); (c) Sinal de áudio classical1 do banco GTZAN, sampler de 50 seg. Fonte: o autor.

Existem também séries temporais especiais, como por exemplo, a série chamada de *Binary Process*, que pertence ao campo da teoria das comunicações. Esta série surge quando observações são feitas a partir de dois únicos valores, geralmente denotados por 0 e 1, que representam, por exemplo, o estado de um disjuntor desligado ou ligado (CHATFIELD, 1995). Outro tipo de séries temporais acontece quando consideramos uma série de eventos ocorrendo randomicamente no tempo. Um exemplo é o registro de datas dos maiores desastres ferroviários. Uma série de eventos desse tipo são frequentemente chamadas de *Point Process* (CHATFIELD, 1995). Na Figura 3.2 temos uma ilustração grafica do *Binary Process* e no *Point Process*.

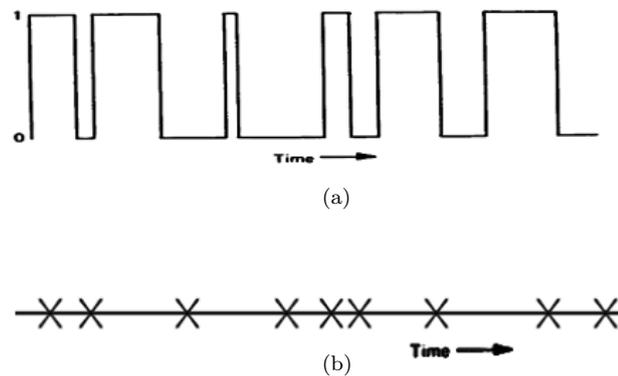


Figura 3.2: (a) Realização de um *Binary Process* para os estados de um disjuntor (CHATFIELD, 1995); (b) Realização de um *point process* para desastres ferroviários. Fonte: CHATFIELD (1995).

A representação de uma série temporal é dada através do conjunto $\{W(t), t \in T\}$, que é uma coleção de observações da variável W , onde T é um conjunto de índices (tempo, espaço, etc.). Uma série temporal é contínua quando as observações são feitas continuamente no tempo, e desse modo, T é um intervalo finito. Quando T é um conjunto finito de pontos, onde as observações são feitas em tempos específicos e, em geral, igualmente espaçados, temos uma série temporal discreta. O termo "discreta" é usado para séries temporais deste tipo mesmo quando a variável medida é uma variável contínua (CHATFIELD, 1995). Segundo CHATFIELD (1995), existem várias formas de se conceber uma série temporal discreta. Uma delas é quando, a partir de uma série temporal contínua, podemos digitalizar os valores em intervalos de tempo iguais com o objetivo de obter séries discretas, denominadas de séries amostradas. Um outro tipo de série discreta pode surgir quando a variável não possui valor instantâneo, porém pode-se agregar ou acumular os valores sobre intervalos de tempo iguais.

As séries temporais podem ser multivariadas ou multidimensionais. Neste caso o conjunto das observações $\{W(t), t \in T\}$ tem na variável W um vetor $k \times 1$ (série multivariada) ou no índice t um vetor $p \times 1$ (série multidimensional). Assim, por exemplo, o conjunto de observações $\{W_1(t), W_2(t); t \in T\}$, onde a pressão uterina da parturiente $W_1(t)$ e as batidas do coração do feto $W_2(t)$ em trabalho de parto, configura uma série temporal

multivariada. Por outro lado, no conjunto $\{W(t), t = (n, k) \in T\}$, onde W representa o número de casos de homicídios dolosos na semana (n), ocorridos no bairro (k) da cidade de Salvador, temos uma série temporal multidimensional (PEREIRA, 1984).

PRIESTLY (1988) constata que uma das mais importantes características da vasta maioria das séries temporais, é que valores registrados em diferentes pontos do tempo são todos influenciados, pelo menos em parte, por algum mecanismo randômico. O registro de temperaturas diárias, por exemplo, exibirá padrões irregulares de variação que não podem ser adequadamente descritos por uma simples fórmula matemática; no caso de medições de voltagem feitas em um circuito elétrico simples, podemos pensar que aplicando as leis do eletromagnetismo poderíamos, a princípio, conceber uma fórmula matemática explícita para a voltagem em função do tempo. De qualquer forma, é importante lembrar que voltagens registradas terão inevitavelmente erros, e esses erros apresentam por natureza, padrões estocásticos, isto é, tais variáveis são influenciadas por elementos randômicos. A análise de séries temporais pode ser efetuada basicamente no domínio temporal e no domínio de frequências. No domínio temporal os modelos propostos são paramétricos, como por exemplo, os modelos ARIMA (modelos auto-regressivos integrados de médias móveis). No domínio das frequências os modelos são não-paramétricos, como a análise espectral, que consiste em decompor a série dada em componentes de frequência, e é caracterizado fundamentalmente pela existência do espectro. (MORETTIN; TOLOI, 2004).

Um importante passo no estudo de séries temporais é o de identificarmos tendências. Logo, podemos aqui definir *tendência* como mudança a longo prazo no nível médio da série. Essa definição, todavia, não é um consenso entre os autores. A forma mais simples de *tendência* é conhecida como tendência linear + ruído, onde a observação no tempo t é dada pela variável randômica X_t (CHATFIELD, 1995).

$$X_t = \alpha + \beta t + \epsilon_t. \quad (3.1)$$

As constantes α e β são valores a serem estimados e ϵ é um erro aleatório com média zero. O nível médio da série varia no tempo t , e é dado por $m_t = (\alpha + \beta t)$. Alguns autores chamam a inclinação β de tendência, uma vez que ela é a mudança de nível da série por unidade de tempo. A tendência na Equação 3.2, também chamada de *tendência global*, é uma função determinística do tempo (CHATFIELD, 1995).

$$X_t = \beta_0 + \beta_1 t + \dots + \beta_t^p + \epsilon_t. \quad (3.2)$$

Existem vários tipos de curvas que podem realizar esse ajuste. Seja qual for a curva utilizada, a função ajustada fornece uma medida de tendência da série, por outro lado os resíduos fornecem uma estimativa das flutuações locais. A Figura 3.3 mostra como exemplo medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970, ajustadas

com polinômios de 3^o e 6^o graus (EHLERS, 2009).

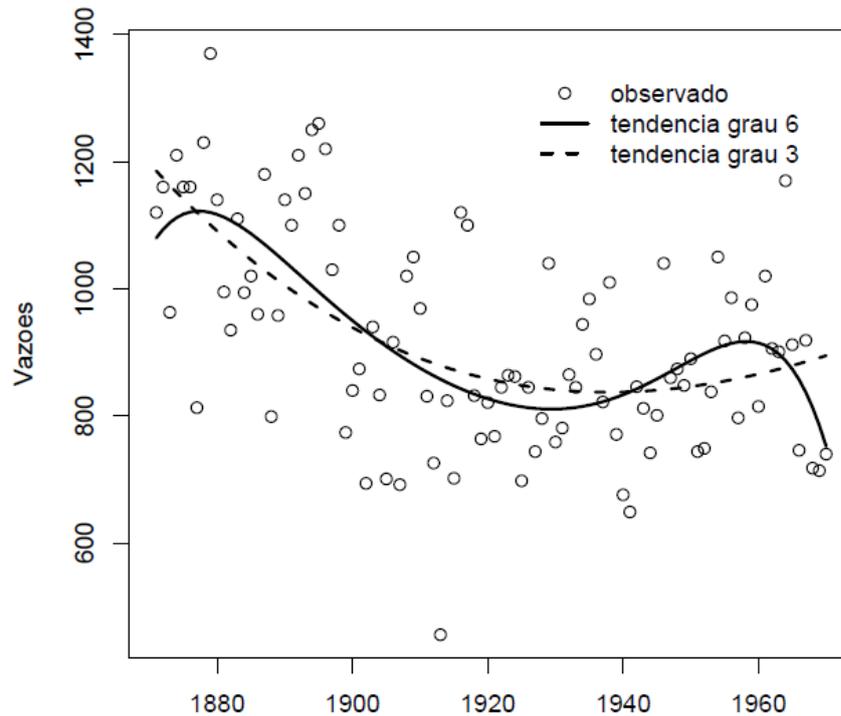


Figura 3.3: A figura mostra a série das medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970, juntamente com as tendências estimadas. Para isso foram utilizados polinômios de graus 3 e 6, ajustados por mínimos quadrados. Fonte: (EHLERS, 2009)

3.2 Análise de Séries Temporais

3.2.1 Autocorrelação

Uma importante ferramenta matemática para analisar as propriedades de uma série temporal consiste do chamado coeficiente de autocorrelação amostral. Esse coeficiente frequentemente fornece informações a respeito do modelo de probabilidade que gerou o conjunto de dados. Aqui, no entanto, queremos medir a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo, isto é, correlações entre observações defasadas de 1, 2, ... períodos de tempo. Assim, dadas n observações x_1, \dots, x_n de uma série temporal discreta podemos formar os pares $(x_1, x_2), \dots, (x_{n-1}, x_n)$. Considerando x_1, \dots, x_{n-1} e x_2, \dots, x_n como duas variáveis, o coeficiente de correlação entre x_t e

x_{t+1} é dado por

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_2)^2 \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_2)^2}}, \quad (3.3)$$

onde

$$\bar{x}_1 = \frac{\sum_{t=1}^{n-1} x_t}{n-1} \quad (3.4)$$

e

$$\bar{x}_2 = \frac{\sum_{t=2}^n x_t}{n-1} \quad (3.5)$$

São as médias aritméticas de x_1 e x_2 . Como o coeficiente dado pela Equação 3.3 mede a correlação entre sucessivas observações ele é chamado de *coeficiente de correlação* ou *coeficiente de correlação serial* (CHATFIELD, 1995).

Por questões de simplificação, e usando o argumento de que $\bar{x}_1 \approx \bar{x}_2$ a Equação 3.3 pode ser escrita da seguinte forma

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(n-1) \sum_{t=1}^{n-1} (x_t - \bar{x})^2 / n}, \quad (3.6)$$

onde a média de todas as observações é dada por

$$\bar{x} = \frac{\sum_{t=1}^n x_t}{n} \quad (3.7)$$

Alguns autores retiram termo $n/n-1$ da Equação 3.6, uma vez que $\lim_{n \rightarrow \infty} (n/n-1)$ é igual a 1, ficando do assim

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^{n-1} (x_t - \bar{x})^2}. \quad (3.8)$$

Do mesmo modo podemos encontrar a correlação entre observações defasadas de k períodos

de tempo com

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (3.9)$$

que é denominado de *coeficiente de autocorrelação de ordem k* , cuja variação está no intervalo $-1 < r_k < 1$

Em termos práticos calcula-se primeiro os *coeficientes de autocovariância*

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (3.10)$$

com $k = 0, 1, \dots$ e depois os coeficientes de autocorrelação $r_k = c_k/c_0$

Para interpretar os valores de r_k pode-se considerar as seguintes situações:

- $0 < r_k \leq 1$ a série possui autocorrelação positiva
- $-1 \leq r_k < 0$ a série possui autocorrelação negativa
- $r_k = 0$ não existe autocorrelação na série

Se a série for estacionária, o decaimento da função de autocorrelação determinará se a série possui ou não correlação de longo alcance. Um decaimento exponencial,

$$c_k \approx \exp \frac{-k}{t_x} \quad (3.11)$$

com tempo característico t_x , indica que a série possui autocorrelação de curto alcance. Por outro lado, uma função de autocorrelação com decaimento seguindo uma lei de potência,

$$c_k \propto k^{-\gamma}, \text{ onde } 0 < \gamma < 1 \quad (3.12)$$

indica que a série possui autocorrelação de longo alcance com o expoente de correlação γ . Se uma série temporal possui tendência, os valores de r_k não decairão para zero a não ser numa situação de grandes defasagens. Isto ocorre pois uma observação de um lado da média tende a ser seguida por um grande número de observações do mesmo lado, por causa da tendência. Neste caso, pouca ou nenhuma informação pode ser extraída do correlograma, já que a tendência dominará outras características. A função de autocorrelação só tem um significado para séries estacionárias, sendo assim qualquer tendência deve ser removida antes do cálculo de r_k (CHATFIELD, 1995). A Figura 3.4 mostra uma série temporal com 50 observações geradas segundo o modelo $x_t = x_{t-1} + \epsilon_t$, juntamente com o seu

correlograma. Note que a não estacionariedade da série fica evidenciada no correlograma já que as autocorrelações amostrais decaem muito lentamente (EHLERS, 2009).

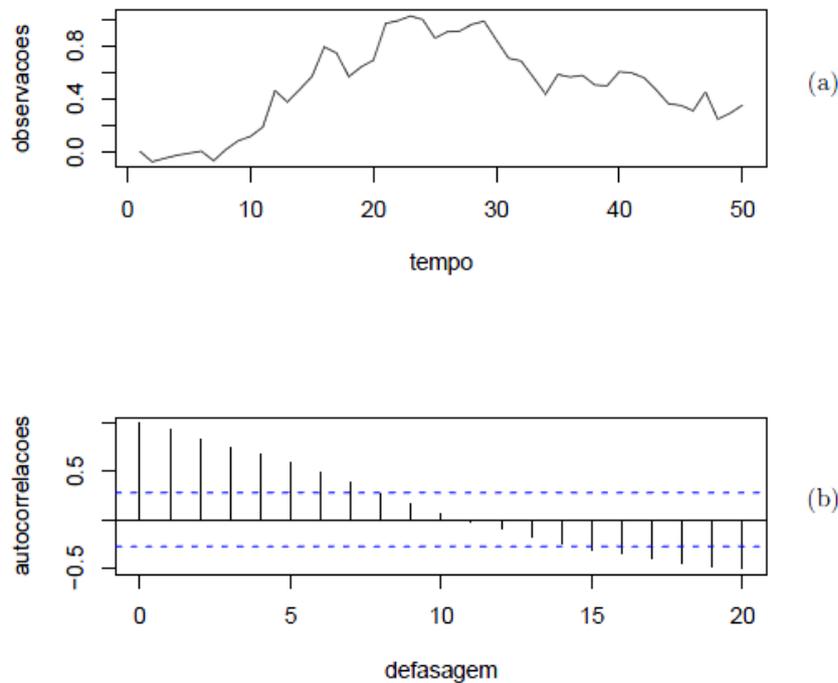


Figura 3.4: (a) 50 observações de uma série não-estacionária simuladas segundo um *random walk*; (b) 20 primeiras autocorrelações amostrais dessa série. Fonte: EHLERS (2009)

3.3 Random Walk

O problema do passeio aleatório, proposto por Karl Pearson, apareceu pela primeira vez no início do século XX, publicado na revista Nature (PEARSON, 1905). O problema foi proposto por Pearson no caso de duas dimensões e foi reconhecido por Rayleigh como análogo a um problema molecular que ele havia resolvido alguns anos antes. Não muito tempo depois, uma analogia do passeio aleatório foi feita com a difusão gasosa (1912), assim como os trabalhos para a generalização em 3 dimensões. Em 1921, Polya mostrou que o passeio aleatório comum unidimensional e bidimensional é recorrente, porém não o é a partir da terceira (PEREIRA, 2009b).

De CRYER e CHAN (2008) temos a seguinte dedução: Sejam e_1, e_2, \dots uma sequência de variáveis aleatórias independentes e idênticamente distribuídas, cada uma com média nula e variância σ_e^2 . A série temporal observada, $Y_t : t = 1, 2, \dots$ é construída da seguinte

maneira:

$$\begin{cases} Y_1 = e_1 \\ Y_2 = e_1 + e_2 \\ \vdots \\ Y_t = e_1 + e_2 + \dots + e_t \end{cases} \quad (3.13)$$

que pode ser escrita de forma alternativa como

$$Y_t = Y_{t-1} + e_t \quad (3.14)$$

com condição inicial $Y_1 = e_1$. Uma vez que os e 's são interpretados como o tamanho do passo dado (para frente ou para trás) ao longo de uma linha de números, então Y_t é a posição do *caminhante aleatório* (random walker) no tempo t . A partir da Equação 3.13 se obtém o valor esperado

$$\mu_t = E(Y_t) = E(e_1 + e_2 + \dots + e_t) = E(e_1) + E(e_2) + \dots + E(e_t) = 0 + 0 + \dots + 0 \quad (3.15)$$

logo,

$$\mu_t = 0 \text{ para todo } t \quad (3.16)$$

Temos também

$$Var(Y_t) = Var(e_1 + e_2 + \dots + e_t) = Var(e_1) + Var(e_2) + \dots + Var(e_t) = \sigma_e^2 + \sigma_e^2 + \dots + \sigma_e^2$$

daí

$$Var(Y_t) = t\sigma_e^2 \quad (3.17)$$

Note que a variância cresce com o tempo.

Para investigar a função de covariância, vamos tomar $1 \leq t \leq s$, então tem-se

$$\gamma_{t,s} = Cov(Y_t, Y_s) = Cov(e_1 + e_2 + \dots + e_t, e_1 + e_2 + \dots + e_t + e_{t+1} + \dots + e_s) \quad (3.18)$$

de onde vem

$$\gamma_{t,s} = \sum_{i=1}^s \sum_{j=1}^t Cov(e_i, e_j) \quad (3.19)$$

De qualquer modo, essas covariâncias serão iguais a zero a menos que $i = j$, neste caso

$Var(e_i) = \sigma_e^2$. Uma vez que $\gamma_{t,s} = \gamma_{s,t}$, isso especifica a função de autocovariância para todos os pontos t e s e podemos escrever

$$\gamma_{t,s} = t\sigma_e^2 \quad (3.20)$$

A função de correlação do *random walk* é então calculada através de

$$\rho_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t} \cdot \gamma_{s,s}}} = \sqrt{\frac{t}{s}} \quad (3.21)$$

para $1 \leq t \leq s$

Um *random walk* simulado é mostrado na Figura 3.5 onde os e 's foram selecionados de uma distribuição normal padrão. Um processo simples de *random walk* fornece um bom modelo para fenômenos distintos, como o movimento do preço das ações ordinárias e a posição de pequenas partículas suspensas em um fluido, chamado *movimento browniano* (CRYER; CHAN, 2008). Podemos obter também simulações de *random walk* em duas dimensões Figura (3.6) e três dimensões Figura (3.7).

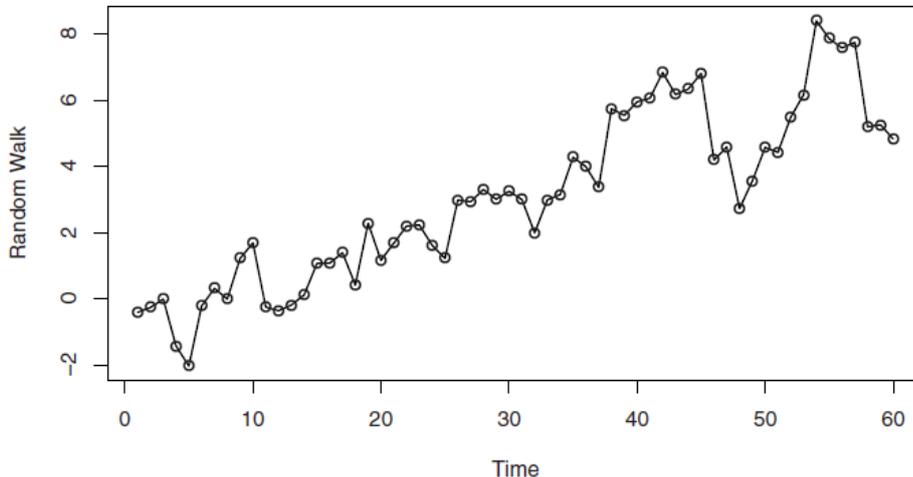


Figura 3.5: A figura mostra um *random walk* simulado onde os e 's foram selecionados de uma distribuição normal padrão. Fonte CRYER e CHAN (2008)

3.4 Expoente Hurst

Em seus estudos de hidrologia, ao investigar problemas de dimensionamento de represas, o britânico Harold Edwin Hurst (1880-1978), desenvolveu um método de análise que possibilita quantificar correlações de longo alcance em séries temporais. Hurst mediu

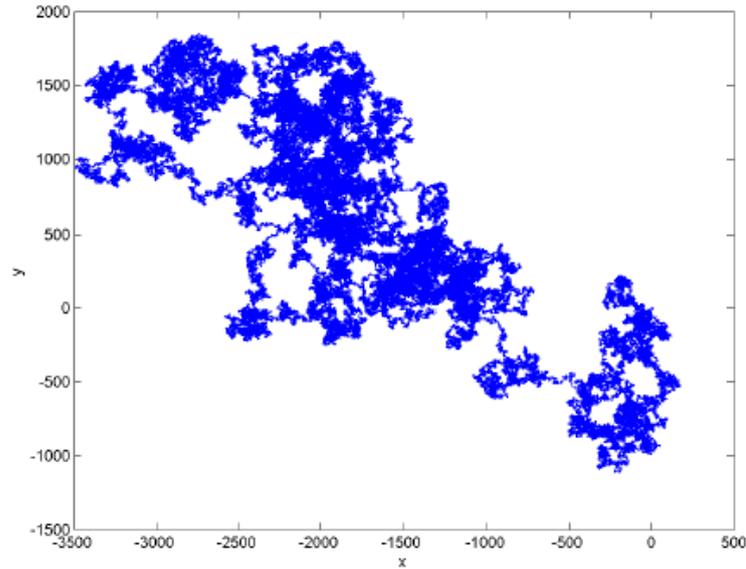


Figura 3.6: Simulação de um *random walk* bidimensional feita com um milhão de passos. Fonte PEREIRA (2009b).

os volumes máximos e mínimos de um reservatório de água e calculou o desvio-padrão dos fluxos em um determinado período de tempo, de modo que R/S seria a razão entre a amplitude R e o desvio padrão das vazões de água. Ao repetir esse procedimento para a mesma série, em diversos períodos de tempo, percebeu que havia uma função relacionando este resultado com o número de observações que faziam parte do cálculo (Figura(3.8)). Mais tarde, ele verificou que essa relação também era válida para outros fenômenos naturais (SOUZA S.; TABAK, 1998).

Através de GALHARDO (2010) e SOUZA S.; TABAK (1998) temos a seguinte dedução: Seja $\xi(t)$ uma série temporal e $\langle \xi \rangle_\tau$ a sua média em um período de τ intervalos de tempo, isto é,

$$\langle \xi \rangle_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t), \quad (3.22)$$

então o perfil é definido como:

$$X(t, \tau) = \sum_{u=1}^t [\xi(u) - \langle \xi \rangle_\tau]; \quad (3.23)$$

Aqui definimos $R(\tau)$ (alcance) através da diferença

$$R(\tau) = \max_{1 \leq x \leq \tau} X(t, \tau) - \min_{1 \leq x \leq \tau} X(t, \tau). \quad (3.24)$$

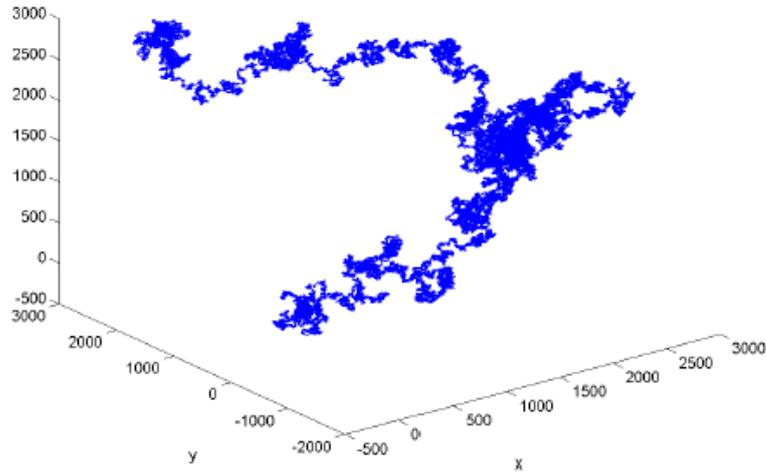


Figura 3.7: Simulação de um *random walk* tridimensional com 1 milhão de passos. Fonte PEREIRA (2009b).

Logo após, calculamos o desvio-padrão em uma janela τ através da equação

$$S(\tau) = \left[\frac{1}{\tau} \sum_{t=1}^{\tau} (\xi(t) - \langle \xi \rangle_{\tau})^2 \right]^{\frac{1}{2}} \quad (3.25)$$

Calcula-se a estatística R/S para cada i -ésima janela:

$$(R/S)_i = \frac{1}{S_i} \left[\max_{1 \leq t \leq \tau} \sum_{k=1}^t (\xi(t) - \langle \xi \rangle_{\tau}) - \min_{1 \leq t \leq \tau} \sum_{k=1}^t (\xi(t) - \langle \xi \rangle_{\tau}) \right] \quad (3.26)$$

Determina-se a média dos valores $(R/S)_i$, associando-se ao tamanho da janela τ :

$$(R/S)_{\tau} = \frac{1}{n} \sum_{i=1}^n (R/S)_i, \quad (3.27)$$

Ao serem concluídos os cálculos da estatística R/S para diversos valores de τ , temos a relação R/S dada por

$$R/S = (\tau/2)^H. \quad (3.28)$$

O valor de H pode ser obtido através da relação entre R/S e τ , em um gráfico $\log x \log$.

Os valores do expoente de Hurst variam entre 0 e 1, e têm a seguinte interpretação:

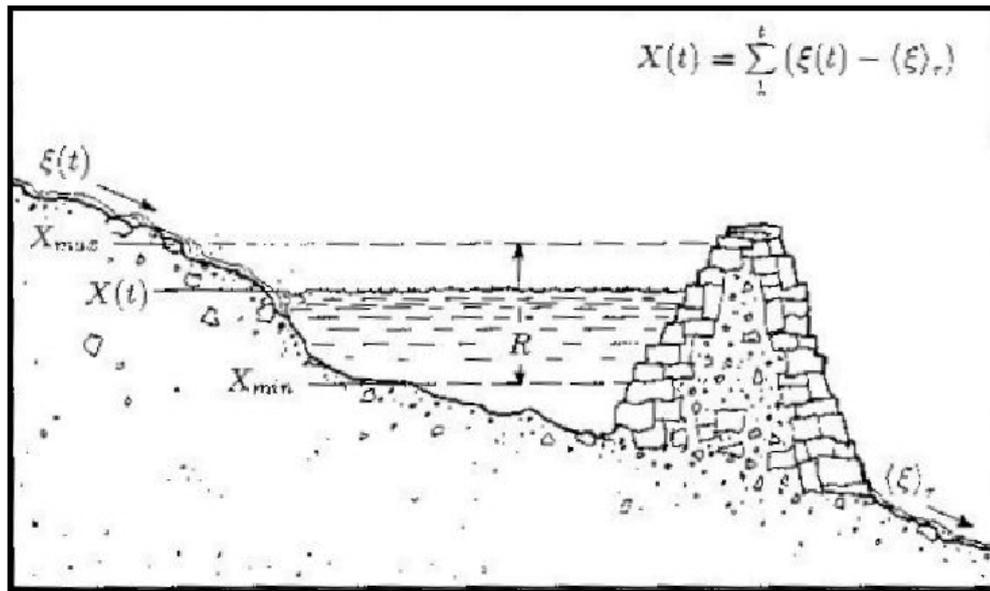


Figura 3.8: O modelo do reservatório de Harold Edwin Hurst, utilizado para determinar o coeficiente R/S , calculado pela razão entre a amplitude R e o desvio-padrão das vazões de água. Fonte: (FEDER, 1988)

- se $H = 0,5$ a série é aleatória ou decorrelacionada (*random walk*);
- se $0 < H < 0,5$ a série tem um comportamento antipersistente;
- se $0,5 < H < 1,0$ a série é persistente.

Na Figura 3.9 é mostrado graficamente o cálculo do R/S em uma série gaussiana para vários tamanhos de janela (GALHARDO, 2010).

3.5 Método DFA

O DFA (*Detrended Fluctuation Analysis*) é um método estatístico que permite eliminar a tendência de uma série temporal em diferentes escalas, analisando as flutuações intrínsecas da série. Essas flutuações podem ser entendidas como medida de variabilidade do sinal associada a variância de cada segmento da série em diferentes escalas. O método DFA foi proposto por PENG (1994) e tem se consolidado como uma ferramenta de grande importância na detecção de correlações de longo alcance em séries temporais não-estacionárias. Esse método é baseado na teoria de caminhos aleatórios (*random walk theory*) (MANDELBROT, 1983); (ADDISON, 1997), e é um aperfeiçoamento do método Flutuatuion Analysis (FA) (PENG, 1994).

O algoritmo do DFA é composto das seguintes etapas. Seja $\{x(t)\}$ uma série temporal

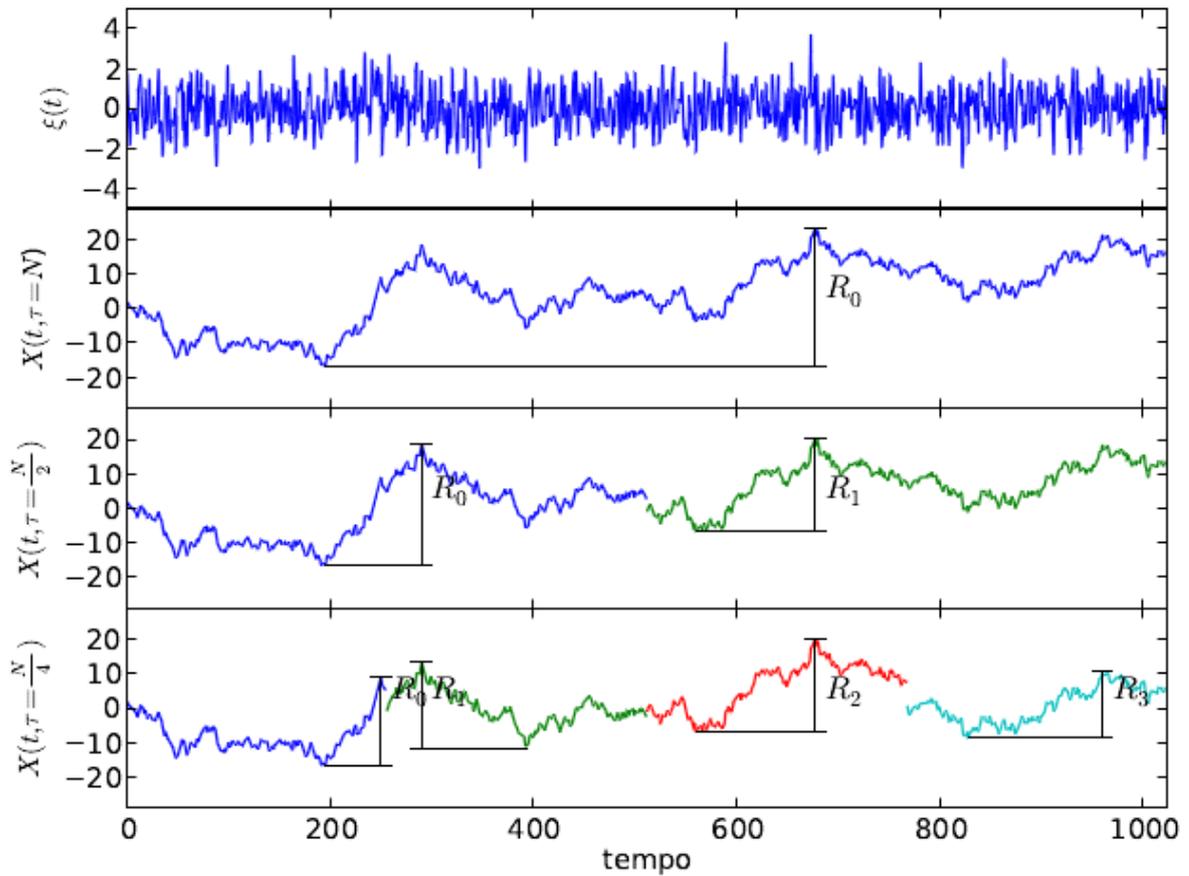


Figura 3.9: O primeiro gráfico mostra uma série gaussiana. Os três gráficos seguintes mostram o perfil dessa série e o cálculo do R/S para valores distintos de τ . Fonte: GALHARDO (2010).

de tamanho N e $\langle x \rangle$ sua média. Em primeiro lugar efetua-se a remoção da média e em seguida a sua integração. O resultado desta operação é denominado de perfil e é estabelecido pela equação:

$$y(t) = \sum_{k=1}^N (x(k) - \langle x \rangle) \quad (3.29)$$

O *perfil* é então dividido em caixas de tamanho n . Para cada caixa, é removida a tendência, subtraindo um polinômio de grau k , y_n^k , através do método dos mínimos quadrados. Calcula-se então a flutuação do sinal

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(t) - y_n^k(t))^2}, \quad (3.30)$$

Se a série original apresenta correlações de longo alcance, então os valores de $F(n)$ seguem

uma lei de potência:

$$F(n) \propto n^\alpha, \quad (3.31)$$

Através de um ajuste linear em escala log x log podemos calcular o expoente α , que nos dá informação a respeito dos padrões de correlação da série. A Figura 3.10 mostra graficamente este processo (GALHARDO, 2010).

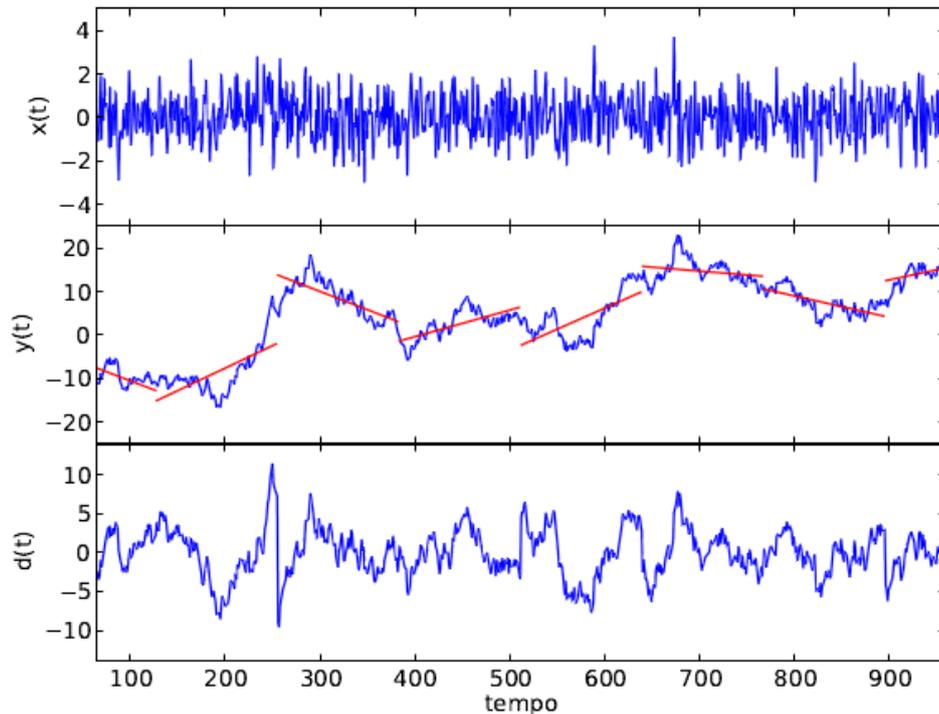


Figura 3.10: O primeiro gráfico mostra a série original $x(t)$; o segundo gráfico mostra o perfil $y(t)$ da função $x(t)$; no terceiro gráfico é mostrado o perfil após retiradas as tendências. Fonte: GALHARDO (2010).

Através da faixa dinâmica de α , podemos obter a seguinte classificação :

- $\alpha < 0,5$, anticorrelacionado ou anti-persistente;
- $\alpha = 0,5$, sem correlação ou sem memória;
- $\alpha > 0,5$, o sinal é persistente ou apresenta correlações de longo alcance;

A proposta inicial do DFA foi a análise em sequências de DNA (PENG, 1994), contudo este método tem encontrado aplicações na análise de séries temporais em diversas áreas como: sistemas astrofísicos (NADA, a), pressão arterial (GALHARDO, 2010), imagens de radar em exploração de petróleo (FREITAS R.; GENOVEZ, 2009), climatologia (BLAIN, 2010) entre outras aplicações.

Implementação e resultados

4.1 Modelo Utilizado

O modelo computacional utilizado na pesquisa foi o mesmo adotado em JENNINGS (2004), que utiliza uma variante da *Detrended Fluctuation Analysis* e determina uma função chamada de Flutuação DFA e a partir dela se obtém o DFA expoente. Este algoritmo também foi utilizado em trabalhos científicos como em TERMENS (2009), STREICHS (2006) e BEROIS (2008). O algoritmo foi implementado utilizando a linguagem de programação C e um computador com processador intel Core 2 Duo e 2MB de ram, utilizando o sistema operacional Windows XP. A seguir são apresentados os bancos de dados utilizados e os passos do algoritmo implementado para o cálculo do DFA expoente.

4.1.1 Base de Dados

Foram utilizados dois bancos de dados para a realização da pesquisa. Estes bancos consistem em arquivos de música digitalizados, que no texto deste capítulo aparecem como representantes do sinais de áudio analógicos. Portanto, o termo sinal original será utilizado para se referir ao arquivo digitalizado da coleção escolhida para o estudo aqui realizado.

Um dos Bancos de dados utilizado foi o GTZAN Genre Collection http://marsyas.info/download/data_sets. Esse banco possui 1000 arquivos de música no total tem sido utilizados em outros trabalhos, como em TZANETAKIS e COOK (2002) e LAMBROU T.; KUDUMAKIS e LINNEY (1988). A Tabela 4.1 mostra as principais características do banco GTZAN.

Tabela 4.1: GTZAN GENRE COLLECTION

formato	au
taxa de amostragem	22.050 Hz
taxa de quantização	16 bit
modo	mono
# gêneros	10
# tracks/gênero	100
# total de tracks	1000
# duração/track	30 s

Existem 10 gêneros nessa base de dados: Blues, Classical, Country, Disco, Hip-hop, Jazz,

Metal, Pop, Reggae e Rock. Cada gênero possui 100 músicas com trechos de 30 segundos cada, extraídos de CDs, rádio e microfone, com o objetivo de ter a representação de varias condições de gravação. Os *tracks* ou faixas, originalmente no formato WAVE em 16 bit mono com 22050 Hz, tiveram seus *sample rates* mudados para 11.025 Hz, por questões de melhoria de desempenho computacional

O Segundo banco de dados é o banco BRMUSIC Genre Collection <http://www.4shared.com/folder/KsxAP6fL/brmusicdatabase.html>. Ele é constituído de arquivos com música popular brasileira, extraídos da coleção particular do autor. Os critérios definidos para a escolha das músicas e sua classificação dentro de cada gênero foram realizadas pelo autor deste trabalho (Tabela 4.2).

Tabela 4.2: BRMUSIC GENRE COLLECTION

formato	wave
taxa de amostragem	44.100 Hz
taxa de quantização	16 bit
modo	stereo
# gêneros	10
# tracks/gênero	20
# total de tracks	200
# duração/track	completo

O banco BRMUSIC é constituído por 10 gêneros da música popular brasileira, com um total de 200 tracks: Axé Music, Funk Carioca, Samba, Forró, Mangue Beat, Rock Nacional, Clube da Esquina, Tropicalismo, Vanguarda Paulistana e Bossa-Nova. Foi utilizado também um *dowsampling* com 11.025 Hz e 8 bit. Neste banco foram preservadas as durações originais de cada música e foram capturadas diretamente do formato digital

4.1.2 Cálculo do DFA Expoente

O procedimento computacional adotado para implementar a função $\alpha(t)$, que calcula o *Detrended Variance Fluctuation Expoente*, ou simplesmente DFA Expoente segundo STREICH e HERRERA (2005), segue a metodologia descrita em JENNINGS (2004).

Esse método segue os seguintes passos:

- (i) Determinação da série $V(j)$, a partir da série $U(i)$, onde V é o desvio-padrão. O *loudness* sonoro está relacionado com a variância do sinal musical JENNINGS (2004);
- (ii) Integração de $V(j)$, gerando a série $Y(j)$;
- (iii) Cálculo da Função DFA;

(iv) Cálculo do DFA Expoente.

A série $U(i)$ representa o sinal de áudio, com $i = 1 \cdots N$. O número de total de pontos N é uma função

$$N(t) = S_r \cdot t, \quad (4.1)$$

onde S_r é o *sample rate*, cujo valor é aproximadamente 11 Khz e o tempo t é dado em segundos.

Em primeiro lugar o conjunto $\{U(i)\} = \{U(1), \dots, U(N)\}$ é segmentado em m blocos ou caixas não-sobrepostas de tamanho λ . Nesta pesquisa foi adotado um tamanho de caixa λ igual 110 *samples*, que está associado a um tempo de 10ms. Deste modo cada caixa j é formada por um subconjunto de $U(i)$, tomando-se $\{U(110(j-1)+1), \dots, U(110j)\}$, com j variando de 1 a m e i variando de $110(j-1)+1$ a $110j$. Escrevendo de forma mais detalhada temos,

$$\text{caixa } j=1: \{U(1), U(2), \dots, U(110)\},$$

$$\text{caixa } j=2: \{U(111), U(112), \dots, U(220)\},$$

...

$$\text{caixa } j=m: \{U(110(m-1)+1), \dots, U(110m)\}, \text{ onde } N = 110m.$$

Para cada caixa $j = 1 \cdots m$ é calculado o desvio-padrão

$$V(1) = \sqrt{\frac{\sum_1^\lambda (U(i) - \bar{U}_1)^2}{\lambda - 1}}$$

$$V(2) = \sqrt{\frac{\sum_{\lambda+1}^{2\lambda} (U(i) - \bar{U}_2)^2}{\lambda - 1}}$$

$$V(3) = \sqrt{\frac{\sum_{2\lambda+1}^{3\lambda} (U(i) - \bar{U}_3)^2}{\lambda - 1}}$$

...

$$V(m) = \sqrt{\frac{\sum_{(m-1)\lambda+1}^{m\lambda} (U(i) - \bar{U}_m)^2}{\lambda - 1}}$$

Então para a j -ésima caixa temos:

$$V(j) = \sqrt{\frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} (U(i) - \bar{U}_j)^2}{\lambda - 1}}, \quad (4.2)$$

onde a média é dada por

$$\bar{U}_j = \frac{\sum_{(j-1)\cdot\lambda+1}^{j\lambda} (U(i))}{\lambda} \quad (4.3)$$

Com isso cria-se uma nova série $V(j) = \{V(1), V(2), \dots, V(m)\}$, com N/λ amostras. Esse conjunto se assemelha a séries temporais não-estacionárias limitadas (*bounded time series*) que têm relação com a intensidade média do som em cada bloco. Efetua-se então a integração de $V(j)$, com o objetivo de obter uma série temporal ilimitada (*unbounded time series*)

$$Y(m) = \sum_{j=1}^m V(j), \quad (4.4)$$

que pode ser interpretada como um *random walk* unidimensional (JENNINGS, 2004).

O processo de integração é fundamental na computação do processo DFA. Em séries limitadas, o DFA expoente seria sempre igual a zero quando temos escalas de tempo de grande porte. Maiores detalhes em (STREICH; HERRERA, 2005 apud PENG,).

A sequência gerada por $Y(m)$ é também dividida em subsequências sobrepostas de tamanho τ . Cada subsequência é deslocada de uma única amostra em relação à subsequência anterior (JENNINGS, 2004).

Para cada bloco de comprimento τ é removida a tendência linear \hat{y}_k , através da regressão

$$\hat{y}_k = b_1(x) + b_0, \quad (4.5)$$

onde,

$$b_0 = \bar{y} - b_1 \bar{x} = \frac{1}{\tau} \cdot (\sum (y_i) - b_1 (\sum (x_i))) \quad (4.6)$$

e

$$b_1 = \frac{\sum(x_i - \bar{x})(y_1 - \bar{y})}{\sum(x_i - \bar{x})^2}. \quad (4.7)$$

Em seguida é computada a média do quadrado residual para cada bloco k (STREICH; HERRERA, 2005),

$$D(k, \tau) = \frac{1}{\tau} \cdot \sum_{m=0}^{\tau-1} (y(k+m) - \hat{y}_k(m))^2 \quad (4.8)$$

Finalmente a flutuação DFA é dada por:

$$F(\tau) = \sqrt{\frac{1}{K} \sum_{i=1}^k (D(k, \tau))} \quad (4.9)$$

A flutuação DFA é uma função de τ , ou seja, é função da variável de tempo em foco, denominada de janela ou intervalo de interesse. O objetivo do DFA é revelar propriedades de correlação em diferentes escalas de tempo. O processo então é repetido sobre diferentes valores de τ inseridos na janela de interesse. Essas escalas de tempo estão relacionadas com o sinal musical, abrangendo desde pulsações de alto nível até padrões rítmicos mais simples (STREICH; HERRERA, 2005). Deste modo a janela de interesse irá computar variações de *loudness* dentro de janelas temporais que têm relação com a atividade rítmica do sinal estudado. A Figura 4.1 mostra a relação da janela de interesse adotada nesse trabalho e outras janelas de tempo usadas para investigação rítmica presentes na literatura em TZANETAKIS e COOK (2002), FORNARI (2010), CARVALHO (2009) e PARCUTT (1994).

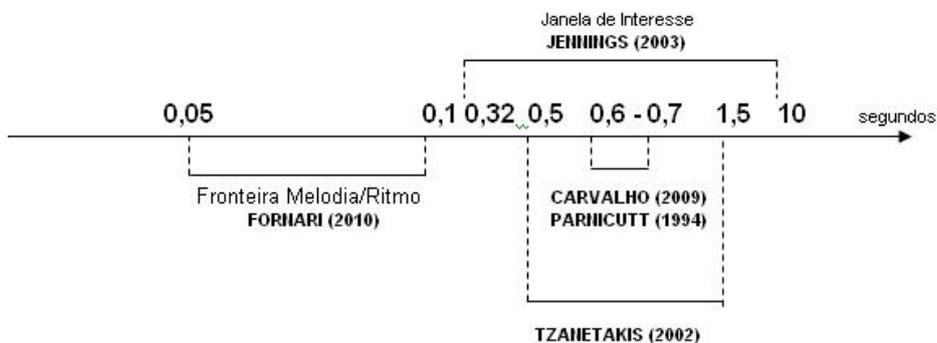


Figura 4.1: Janela de interesse, fronteira rítmico-melódica e intervalos de tempo na investigação rítmica citados por diversos autores. Fonte: o autor.

Nesta pesquisa foi utilizada uma janela de interesse com uma variação de τ de 31 a 909.

Isso corresponde a um intervalo de tempo 310 ms a 9,09 seg e a uma faixa de frequências de 0,11Hz a 3,22 Hz. O DFA expoente é definido como a inclinação em um gráfico $\log x \log$ de F sobre τ . Para calcular o DFA expoente primeiramente foi adotada a Equação 4.10, utilizada em JENNINGS (2004), que representa a taxa de variação de F em função de t para cada valor de i . No cálculo de α para valores pequenos de τ é necessário um ajuste no denominador. A proporção que τ cresce, a influencia da correção torna-se desprezível (STREICH; HERRERA, 2005).

$$\alpha(i) = \frac{\log_{10} F(\tau_{i+1}) - \log_{10} F(\tau_i)}{\log_{10}(\tau_{i+1} + 3) - \log_{10}(\tau_i + 3)} \quad (4.10)$$

Finalmente, o DFA expoente é calculado através da Equação 4.11.

$$\alpha_{DFA} = \frac{\sum_{i=1}^{32} \alpha(i)}{32} \quad (4.11)$$

O cálculo de $F(t)$ e do α_{DFA} para um sinal do tipo *white noise*, simulado a partir de um gerador de números pseudo-aleatórios é mostrado na Tabela 4.3. Ao ser realizado o cálculo do valor do DFA expoente, o valor obtido é $\approx 0,5$. É importante observar que os valores de t são incrementados através da multiplicação por uma constante, o que resulta em um espaçamento igual quando aplicada a escala logarítmica.

A Figura 4.2 ilustra os quatro passos para o cálculo do DFA expoente, aplicados ao arquivo mbeat1. Na Tabela 4.4 estão os resultados obtidos no cálculo do DFA e de cada $\alpha(i)$. O DFA expoente é a média aritmética dos $\alpha(i)$, que resulta neste caso 0,474. Por último a Figura 4.3 mostra o gráfico de $\alpha(t)$.

4.1.3 Interpretação do DFA expoente

Os sinais musicais não possuem propriedades de escala estáveis, pois em música existe muito mais variância em $\alpha(i)$ do que em outros tipos de séries temporais. Ainda assim, podemos constatar que música com saltos repentinos de intensidade geralmente resultam em um menor nível de α_{DFA} do que músicas cuja série tem uma variação suave em seus valores de intensidade.

Tabela 4.3: Resultados da função DFA e da função $\alpha(i)$ de um sinal tipo *white noise*, a partir do arquivo random16. O valor do DFA expoente é $\approx 0,5$. Este resultado é obtido através da média aritmética de todos os valores de $\alpha(i)$ encontrados. Fonte: o autor.

i	t	$\alpha(i)$
1	0,31	0,550
2	0,34	0,549
3	0,39	0,544
4	0,44	0,533
5	0,49	0,517
6	0,55	0,499
7	0,62	0,487
8	0,69	0,481
9	0,78	0,478
10	0,87	0,479
11	0,97	0,483
12	1,08	0,496
13	1,21	0,517
14	1,35	0,534
15	1,51	0,537
16	1,68	0,530
17	1,88	0,519
18	2,09	0,501
19	2,33	0,495
20	2,59	0,492
21	2,88	0,483
22	3,20	0,480
23	3,56	0,479
24	3,96	0,467
25	4,40	0,461
26	4,88	0,466
27	5,42	0,471
28	6,01	0,456
29	6,67	0,448
30	7,40	0,465
31	8,20	0,506
32	9,09	0,551

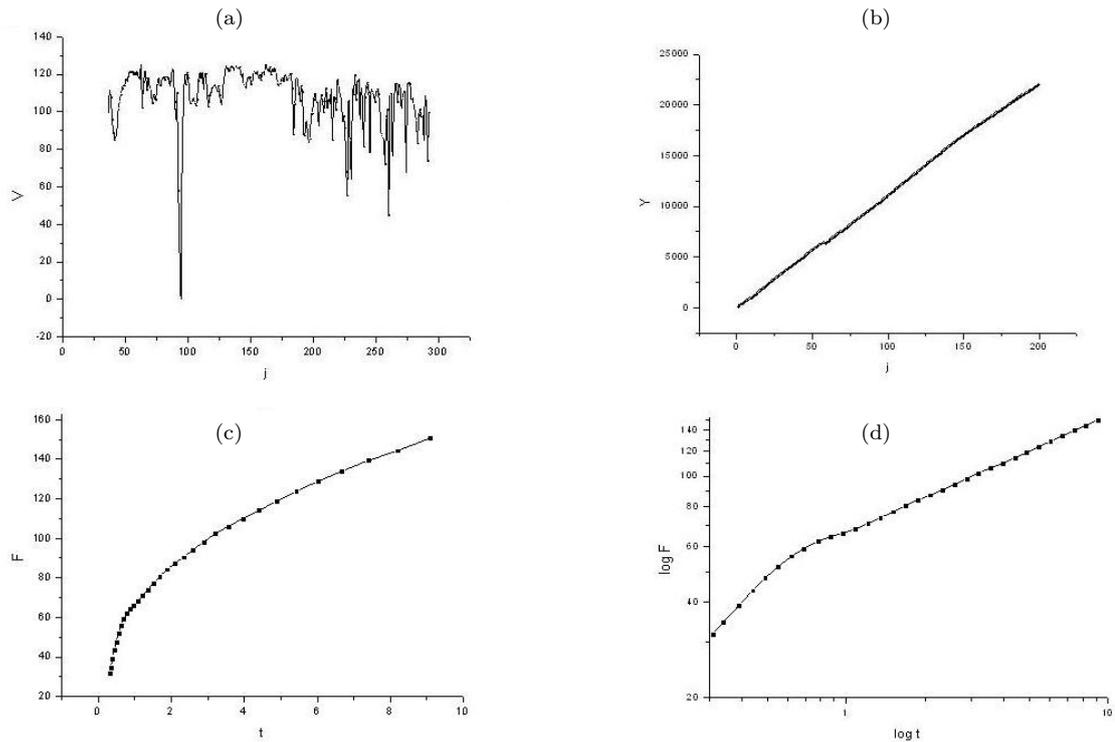


Figura 4.2: Os quatro passos para a determinação do DFA expoente, aplicados ao arquivo mbeat1, ilustrados através de: (a) gráfico da série $V(j)$, que representa o desvio-padrão da série original. Este gráfico está relacionado com o *loudness* do sinal musical; (b) gráfico da série $Y(j)$, resultante da integração de $V(j)$; (c) gráfico da função DFA; (d) gráfico $\log \times \log$ da função DFA, cuja inclinação determina o valor do α_{DFA} . Fonte: o autor.

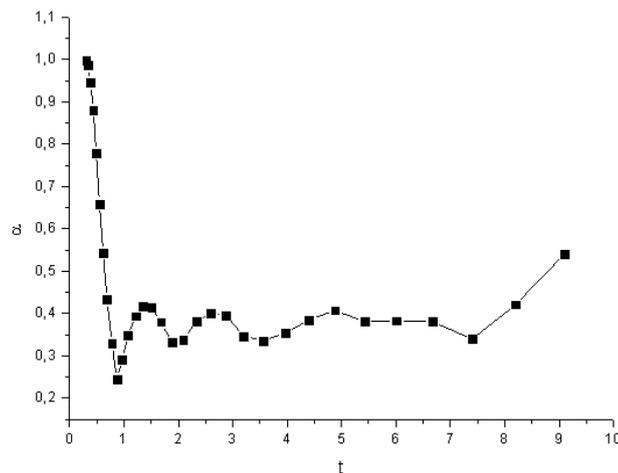


Figura 4.3: Função $\alpha(t)$ expoente do arquivo mbeat1 do banco BR-MUSIC. Fonte: o autor.

O *beat* ou batida periódica da música, considerada abstratamente como uma superposição de tendências periódicas do sinal acústico, ocasiona desvios significantes na escala da lei de potência, naquela escala de tempo (JENNINGS, 2004). Assim, quanto menor correlação de longo alcance das flutuações do *loudness*, menor será a estabilidade de escala, devido a

Tabela 4.4: Resultados da função DFA e da função $\alpha(i)$ a partir do arquivo mbeat1 do banco BR-MUSIC. O valor do DFA expoente é 0,474, que corresponde à média aritmética de todos os valores de $\alpha(i)$ encontrados. É importante observar que os valores de t são incrementados através da multiplicação por uma constante, o que resulta num espaçamento igual quando aplicada a escala logarítmica. Fonte: o autor.

i	t	F(t)	α
1	0,31	31,805	0,997
2	0,34	34,604	0,988
3	0,39	39,220	0,946
4	0,44	43,624	0,881
5	0,49	47,690	0,780
6	0,55	51,929	0,659
7	0,62	55,978	0,544
8	0,69	59,182	0,433
9	0,78	62,279	0,329
10	0,87	64,474	0,245
11	0,97	66,160	0,291
12	1,08	68,198	0,348
13	1,21	70,875	0,394
14	1,35	73,926	0,416
15	1,51	77,379	0,415
16	1,68	80,819	0,380
17	1,88	84,289	0,331
18	2,09	87,255	0,338
19	2,33	90,480	0,381
20	2,59	94,158	0,400
21	2,88	98,196	0,395
22	3,20	102,331	0,346
23	3,56	106,142	0,336
24	3,96	109,976	0,354
25	4,40	114,121	0,385
26	4,88	118,728	0,407
27	5,42	123,882	0,381
28	6,01	128,824	0,383
29	6,67	134,038	0,381
30	7,40	139,423	0,340
31	8,20	144,348	0,421
32	9,09	150,727	0,540

um maior desvio da escala da lei de potência. Se temos um maior desvio, teremos também uma maior predominância de tendências periódicas. Com isso teremos um menor valor de alfa em sinais de áudio que apresentem atividade percussiva mais regular. Esses sinais musicais estão associados a gêneros mais dançantes, ou com maior *danceability* segundo STREICH e HERRERA (2005).

Por outro lado quanto maior a correlação nas flutuações de *loudness*, menor será a predominância de tendências periódicas, resultando em um maior valor para o DFA expoente.

Consequentemente, sinais musicais com menor *danceability* estão associados a um maior valor do DFA expoente.

Como pode ser vista na Figura 4.4, uma estabilidade maior de escala está associada à música clássica. Um comportamento intermediário é apresentado no Jazz, e no Hip-hop apresenta uma grande instabilidade, isto porque existe neste gênero uma grande tendência periódica em uma peça musical, faz com que as propriedades de escala se deteriorem, causando uma queda significativa no valor de α_{DFA} . Em peças de música Pop ou no Jazz existe uma batida regular, porém menos dominante como no Hip-hop ou no Forró.

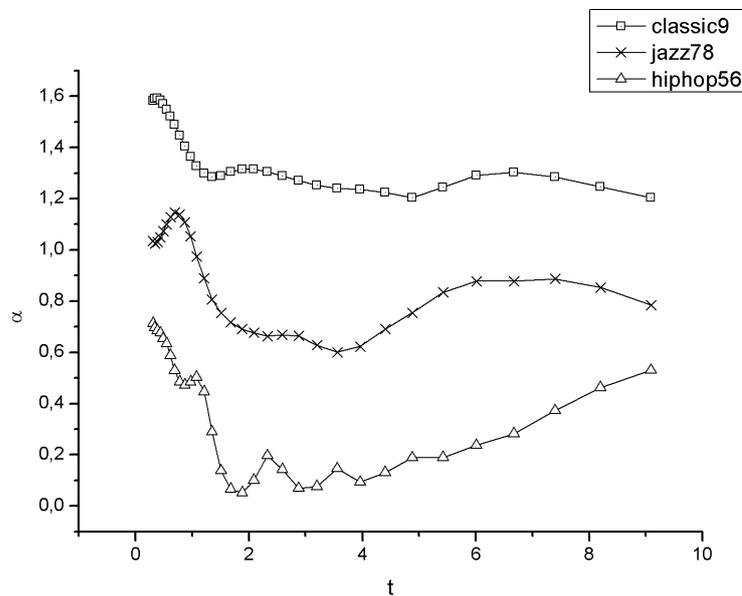


Figura 4.4: Variações na função $\alpha(t)$ para as peças Classic, Jazz e Hip-hop. Uma estabilidade maior na escala é apresentada pela peça Clássica e um comportamento de grande instabilidade é mostrado na peça Hip-hop, devido a presença de eventos percussivos com ataque mais pronunciado e padrões de batida mais fortes e regulares. Já a peça musical Jazz apresenta um comportamento intermediário. Fonte: o autor.

4.2 Resultados

4.2.1 Banco GTZAN

O algoritmo para o cálculo do DFA expoente foi aplicado em 1000 arquivos, subdivididos em 10 gêneros do banco GTZAN. Posteriormente, foi calculada a média aritmética do DFA expoente para os 100 tracks contidos em cada gênero. A Tabela 4.5 e a Figura 4.5,

apresentam os $\langle \alpha_{DFA} \rangle$, calculados para todos os gêneros do Banco GTZAN, com seus respectivos desvios-padrão.

Tabela 4.5: Valor médio do DFA expoente, seguido de seus respectivos desvios-padrão. Esses valores foram calculados para cada grupo de cem arquivos musicais pertencentes a um gênero musical do banco GTZAN. Fonte: o autor.

gênero	$\langle \alpha \rangle$	σ
Hiphop	0,610	0,143
Disco	0,633	0,133
Metal	0,744	0,140
Reggae	0,775	0,146
Blues	0,781	0,138
Pop	0,811	0,140
Rock	0,840	0,140
Jazz	0,878	0,162
Country	0,970	0,152
Clássico	1,070	0,152

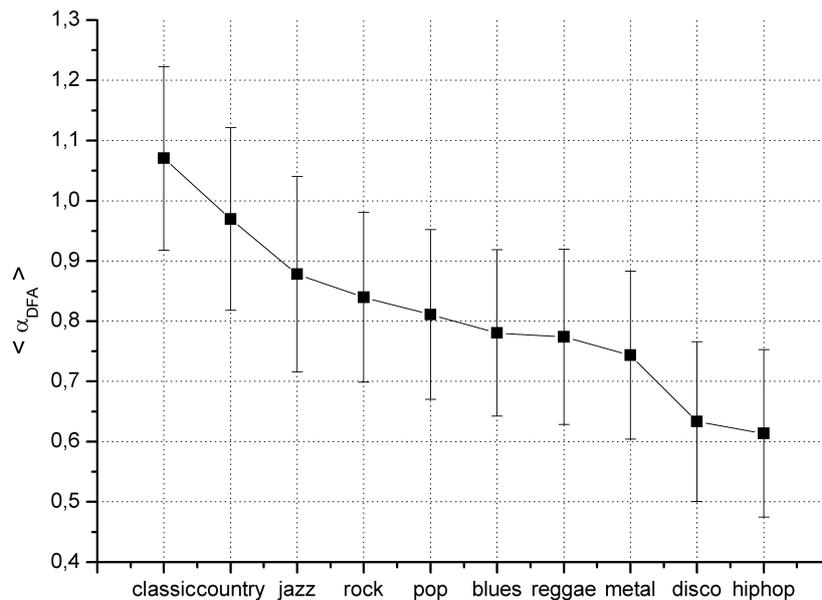


Figura 4.5: Gráfico do DFA expoente médio do banco GTZAN. As barras de erro correspondem ao desvio-padrão. Fonte: o autor.

A Figura 4.6 mostra os resultados obtidos através de boxplot.

Uma vez calculados os valores do DFA expoente para todos os arquivos do banco, foi realizada a verificação de aderência à distribuição normal, utilizando o teste de *Shapiro – Wilk*. Este procedimento constatou que, a um nível de confiança de 95%, todos os dez gêneros provêm de uma distribuição normal. A Figura 4.7 mostra os histogramas das

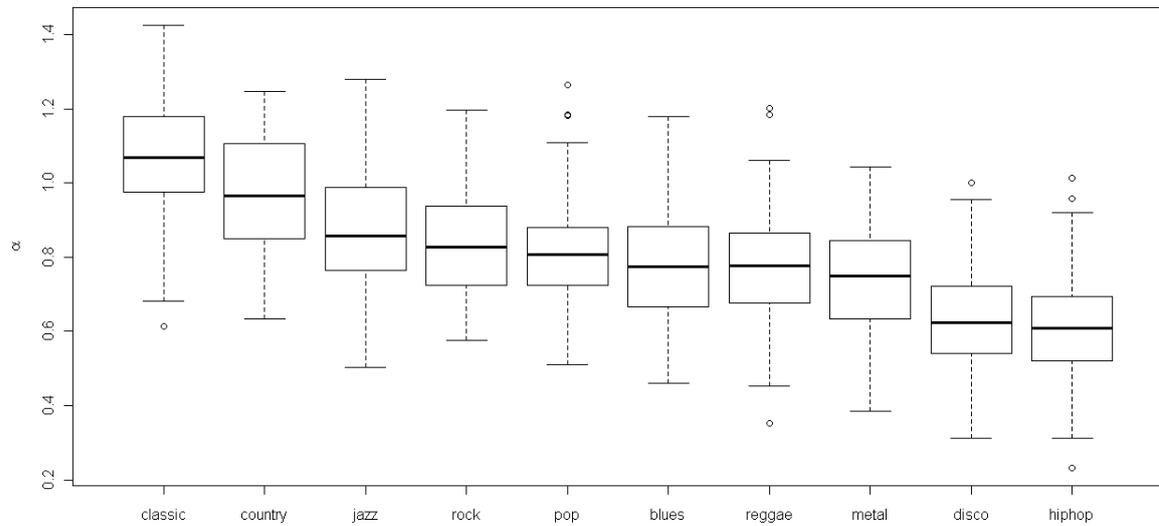


Figura 4.6: Boxplot dos gêneros musicais do banco GTZAN. Fonte: o autor.

distribuições de frequências.

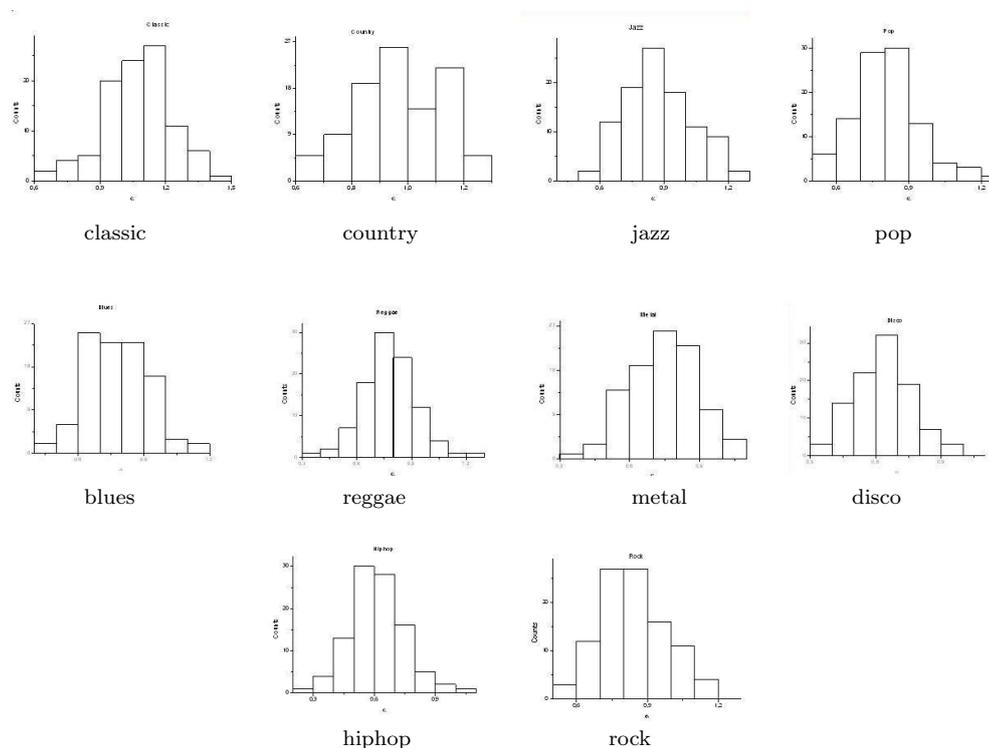


Figura 4.7: Histograma da distribuição de frequências de α_{DFA} nos gêneros do banco GTZAN. Fonte: o autor.

Uma vez verificada que as distribuições relativas ao DFA expoente de cada gênero, são aproximadamente normais, aplicou-se a Análise de Variância (ANOVA). Este procedimento foi utilizado, a fim de avaliar se o DFA expoente médio ($\langle \alpha_{DFA} \rangle$), seria repre-

sentativo ou não para estabelecer diferenças entre os gêneros. Primeiramente foi testada a homogeneidade de variância do conjunto de dados através do teste *Levene*, que indicou, com 95% de confiança, que as variações populacionais não são significativamente diferentes. Depois foi utilizado o teste *Tukey* para comparação de médias, que revelou, com 95% de confiança, que nove pares de gêneros não possuem diferenças significativas em seu $\langle \alpha_{DFA} \rangle$. O resultado deste teste é mostrado graficamente na Figura 4.8.

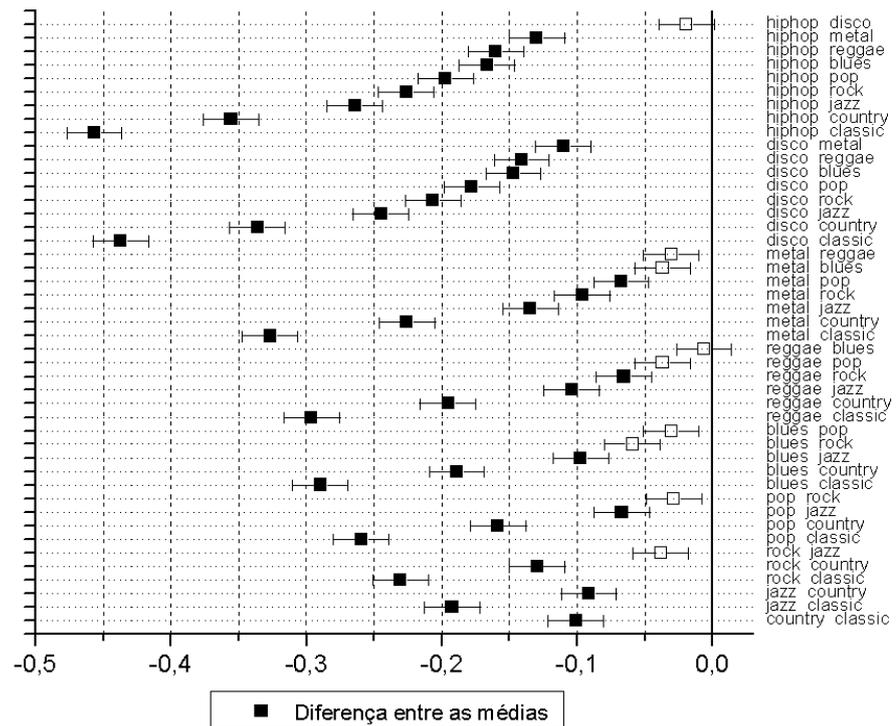


Figura 4.8: Comparação de médias para o banco GTZAN, segundo o teste *Tukey*. Os pares de gêneros Hiphop-Disco, Metal-Reggae, Metal-Blues, Reggae-Blues, Reggae-Pop, Blues-Pop, Blues-rock, Pop-Rock, Rock-Jazz, não possuem diferenças significativas em suas médias. Isto significa que nove pares de gêneros não tem diferenças significativas entre seus $\langle \alpha_{DFA} \rangle$. Fonte: o autor.

A Figura 4.9 gráficos da função $\alpha(t)$ de três gêneros musicais. Este gráfico ilustra as diferenças na flutuação DFA em cada um dos gêneros. Pode-se associar essas diferenças de flutuação às peculiaridades rítmicas de cada gênero.

A Figura 4.10 mostra os valores do DFA expoente para 100 tracks do gêneros Classic e Hiphop.

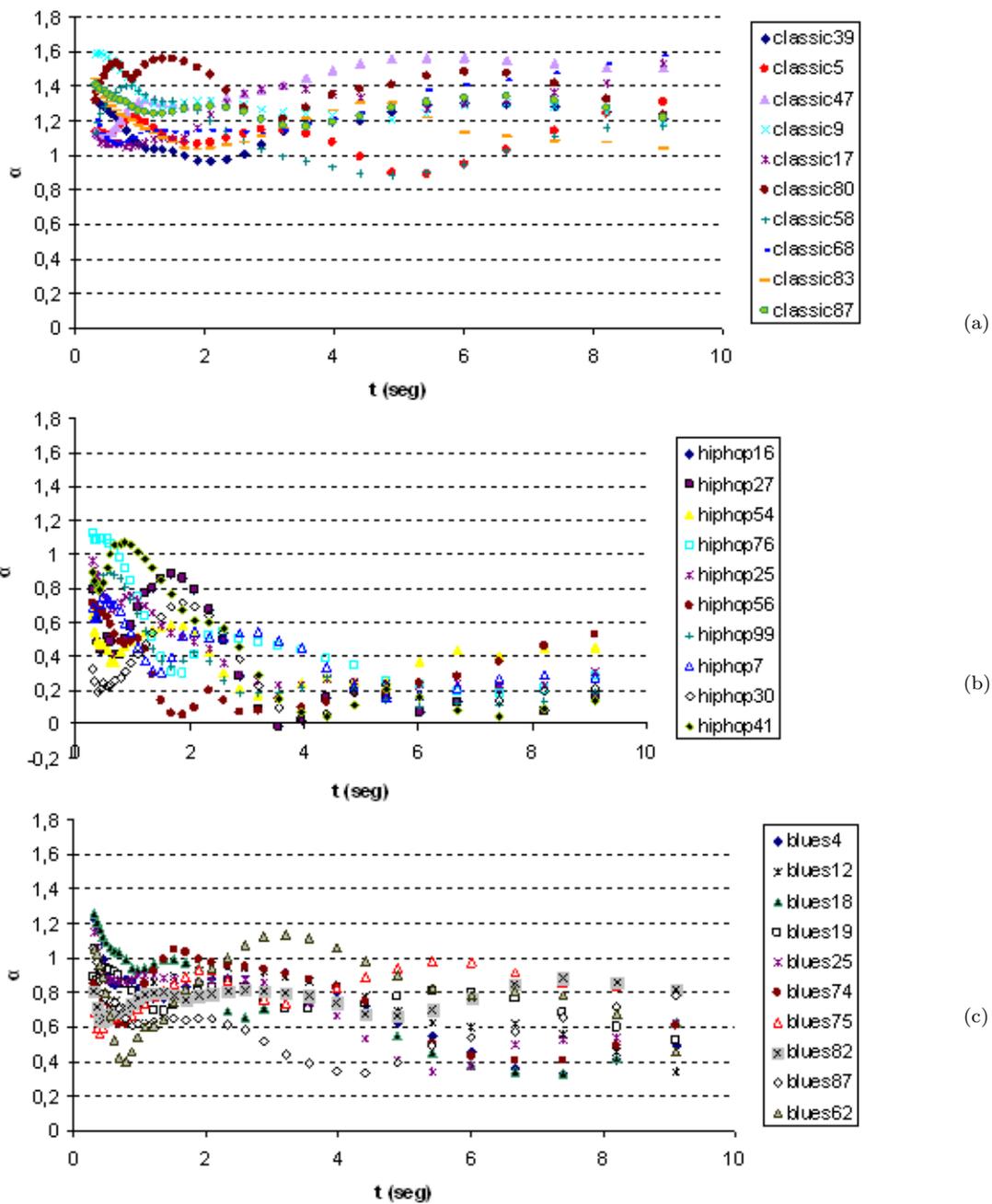


Figura 4.9: O gráfico mostra o expoente α em função do tempo para dez *tracks* dos gêneros Classic, Blues e Hip-hop do banco GTZAN. Pode-se notar diferenças no comportamento gráfico para cada grupo, associadas às características rítmicas de cada gênero. No gráfico (a) os eventos percussivos são geralmente mais suaves, resultando um valor médio de alfa superior aos outros dois gêneros em questão. Em (b), observa-se uma grande instabilidade de escala, associado a eventos percussivos mais intensos e a batidas mais regulares que caracterizam o Hip-hop. Em (C) nota-se um comportamento intermediário no gênero Blues, pois apresenta eventos percussivos mais intensos que no Classic, porém mais moderados que no Hip-hop. Fonte: o autor.

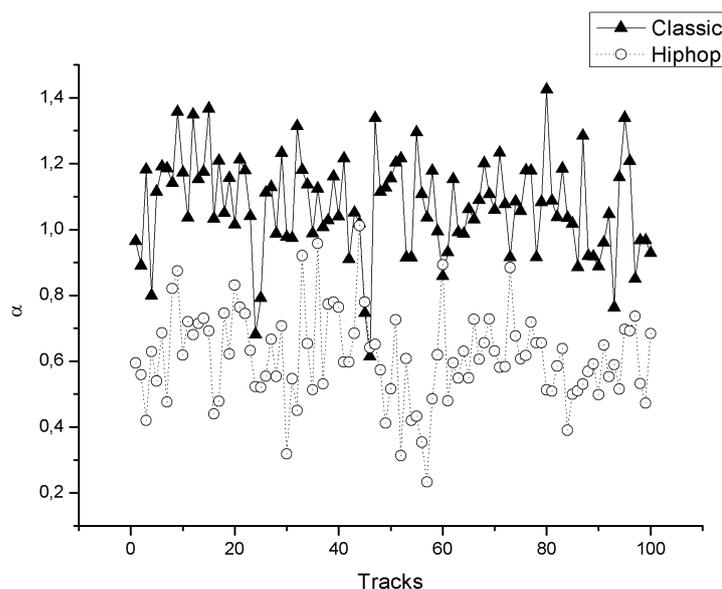


Figura 4.10: DFA expoente dos gêneros Classic e Hiphop. Fonte: o autor.

4.2.2 Banco BRMUSIC

Para os gêneros do banco BRMUSIC foram calculados o DFA expoente, bem como o $\langle \alpha_{DFA} \rangle$ e os seus respectivos desvios-padrão. A Tabela 4.6 e as Figuras 4.11 e 4.12 mostram esses resultados.

Tabela 4.6: Valores médios do DFA expoente do banco BRMUSIC e seus respectivos desvios-padrão. Fonte: o autor.

	$\langle \alpha \rangle$	σ
Axé Music	0,677	0,094
Funk Carioca	0,688	0,096
Samba	0,695	0,050
Forró	0,702	0,119
Mangue Beat	0,754	0,178
Rock Nacional	0,811	0,181
Clube da Esquina	0,831	0,088
Tropicalismo	0,883	0,163
Vanguarda Paulistana	1,041	0,175
Bossa-Nova	1,114	0,080

Foi aplicado então o teste *Shapiro – Wilk* e foi verificado que, com um nível de 95% de confiança, os agrupamentos do banco atenderam as condições de aderência à distribuição normal. As distribuições de frequência podem ser vistas na Figura 4.13.

A Análise de Variância (ANOVA way1) foi então realizada, e o resultado, segundo o teste

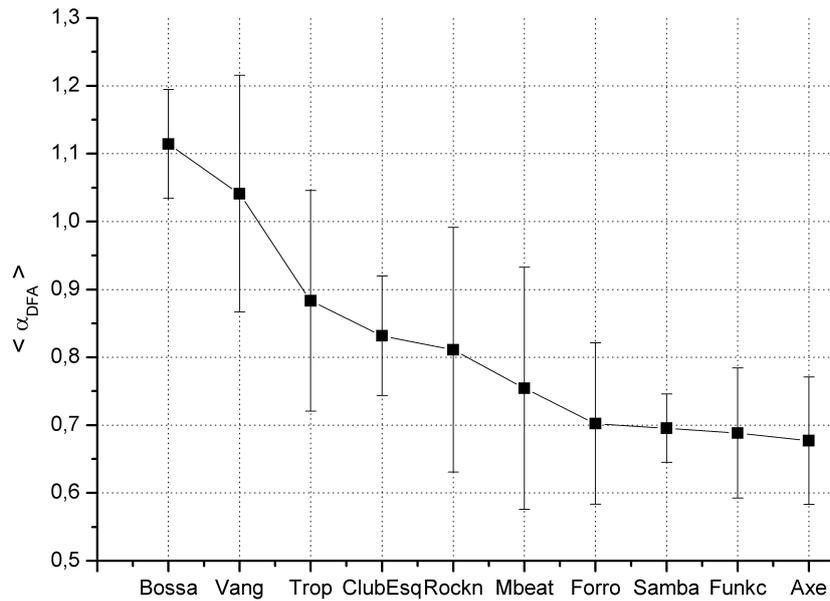


Figura 4.11: DFA expoente médio do banco BRMUSIC. As barras de erro correspondem ao desvio-padrão. Fonte: o autor.

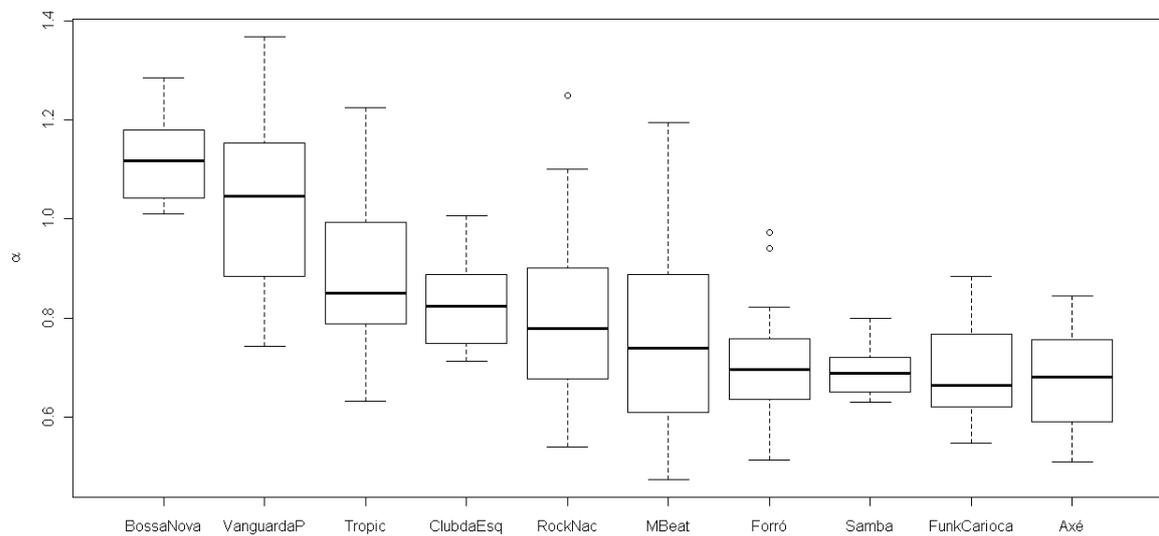


Figura 4.12: Boxplot dos gêneros musicais do banco BRMUSIC. Fonte: o autor.

Tukey, a um nível de 95% de confiança, foi que a diferença entre as médias de vinte e quatro pares de gêneros, não são significantes. O gráfico 4.14 mostra as diferenças entre as médias, segundo o teste *Tukey* e os resultados apresentados na Figura

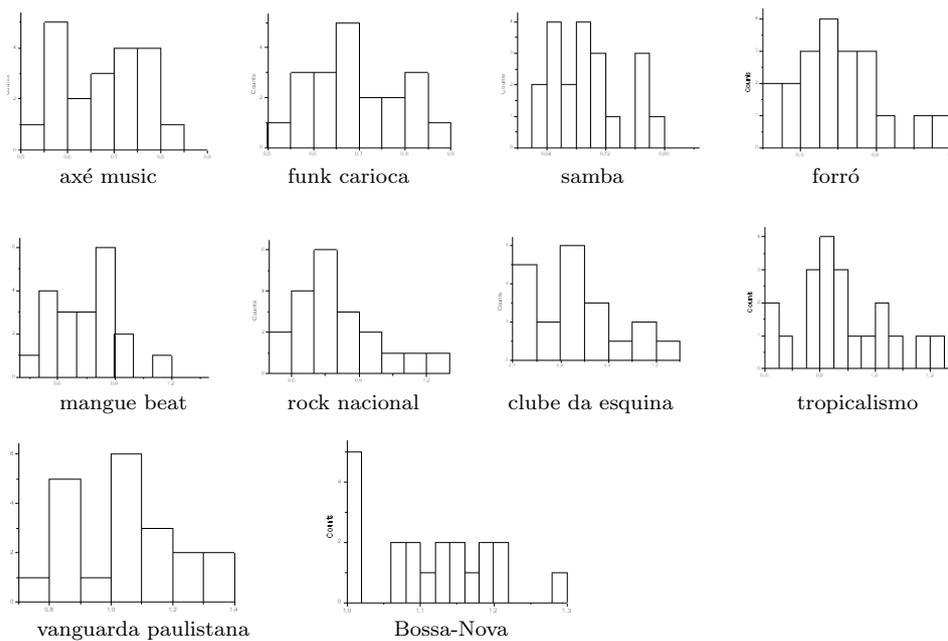


Figura 4.13: Histograma do banco BRMUSIC. Fonte: o autor.

Neste banco, a Bossa-Nova se destacou com o maior coeficiente DFA médio. Uma vez contatata a representatividade do DFA expoente para os gêneros Bossa-Nova e Mangue Beat, pode-se ver na Figura 4.15, os valores de α para cada um 20 tracks. Pode-se considerar que os valores mais elevados para Bossa-Nova estão associados a uma maior estabilidade de escala para cada track.

4.2.3 Subgêneros

Foi também investigado o comportamento de subgêneros, onde foi aplicada a Análise de Variância, a fim de identificar se haviam diferenças quantitativas entre os valores médios do DFA expoente.

Os subgêneros submetidos à comparação foram:

- Axé Afro e Axé Pop. (Figura 4.16).
- Pagode Baiano e Pagode Romântico. (Figura 4.17)
- Forró tradicional e Forró Eletrônico. (Figura 4.18)

Os dois primeiros pares de subgêneros, segundo o método post-hoc *Tukey*, não apresentam diferenças entre seus $\langle \alpha_{DFA} \rangle$, a um nível de confiança de 95%. Como os subgêneros

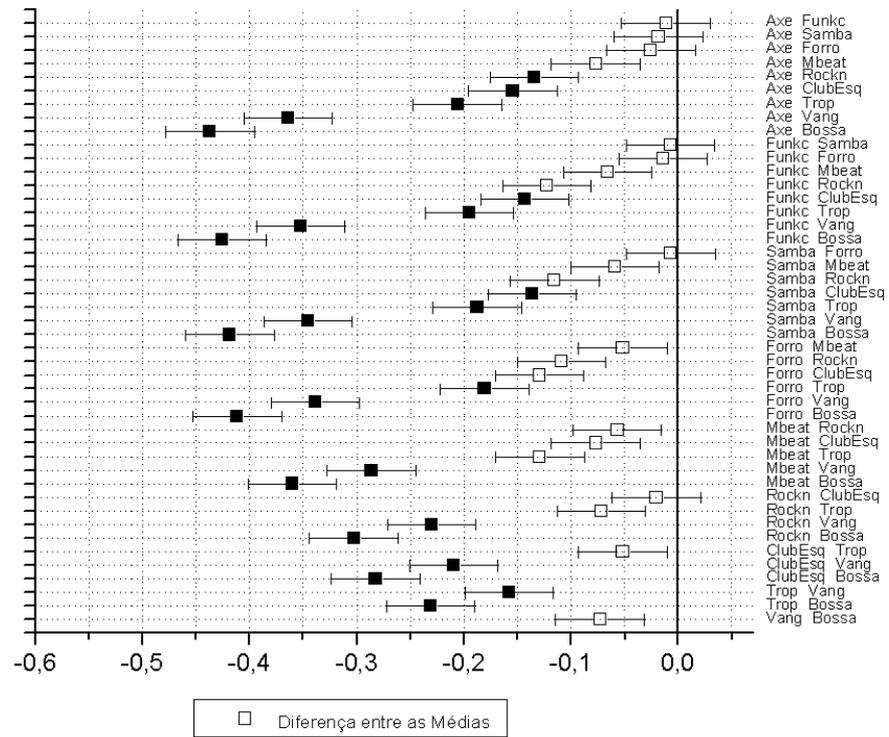


Figura 4.14: Comparação de médias do banco BRMUSIC. Segundo o teste *Tukey* vinte e quatro pares de gêneros não tem diferenças significativas entre seus $\langle \alpha_{DFA} \rangle$. Fonte: o autor.

Forró Eletrônico e Forró Tradicional, não apresentam distribuições aderentes à normalidade, foi utilizado o teste não-paramétrico *Kruskal Wallis*, que indicou a não existência de diferenças significativas entre as médias, com uma confiança de 95%. Aqui nenhum dos subgêneros estudados mostrou diferença significativa, o pode indicar, uma vez que estes apresentam características rítmicas bem próximas, que o DFA expoente médio não seja suficiente para realizar a distinção entre eles.

4.2.4 Aplicação do DFA expoente em Outras Categorias

O algoritmo para o cálculo do DFA expoente foi aplicado em agrupamentos musicais com dois tipos de rótulos. O primeiro foi criado utilizando o nome de estações de rádio, onde a idéia é estabelecer uma relação entre o nome das estações de rádio e a característica predominante de sua seleção musical, através do DFA expoente. Uma vez que a grade da programação musical de uma rádio é pensada para atender um conjunto de interesses como público-alvo, patrocinadores e indústria fonográfica, o repertório a ser escolhido

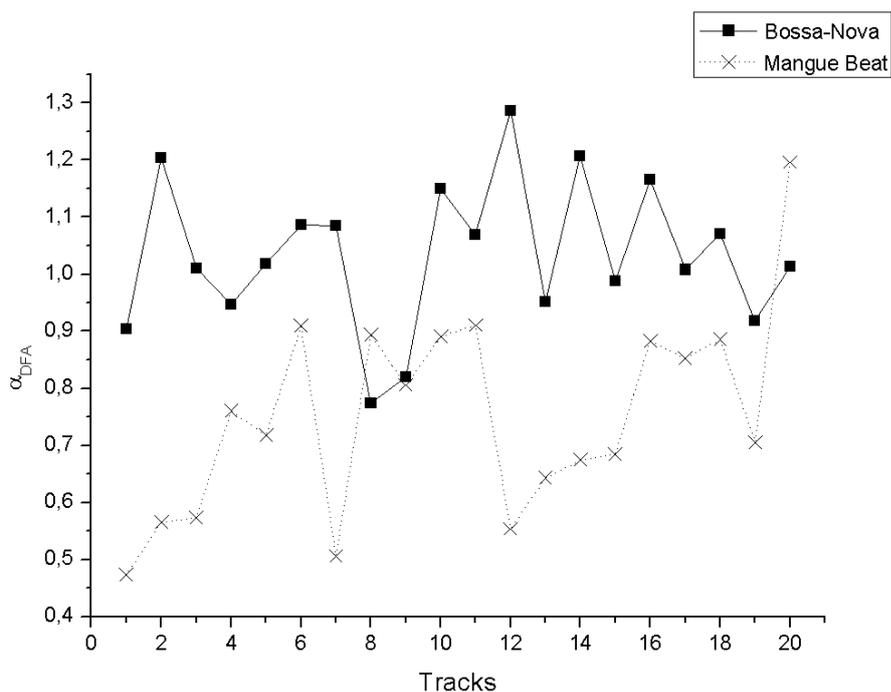


Figura 4.15: DFA expoente dos gêneros Bossa-Nova e Mangue-Beat ordenados em função dos seus respectivos tracks. Fonte: o autor.

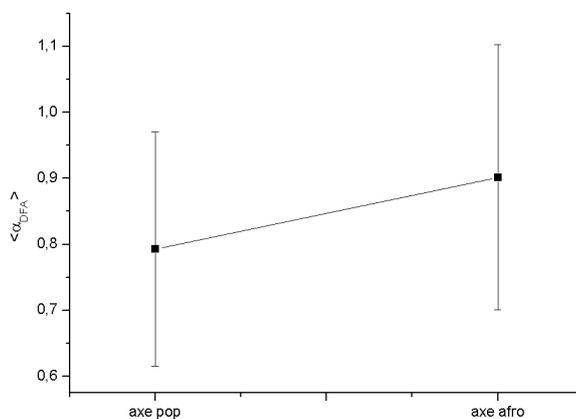


Figura 4.16: DFA expoente médio com desvio-padrão dos subgêneros do Axé Music. Fonte: o autor.

tem um papel importante no cumprimento dos objetivos de uma determinada rádio. Daí pode-se pensar que o rótulo "rádio FM X" pode estar fortemente associada à sua seleção musical. Para investigar tal aspecto foram feitas aproximadamente duas horas de gravação da programação de cada rádio, e depois disso, foram retiradas as locuções, vinhetas e propagandas, e foi separada apenas a seleção musical feita pelos programadores

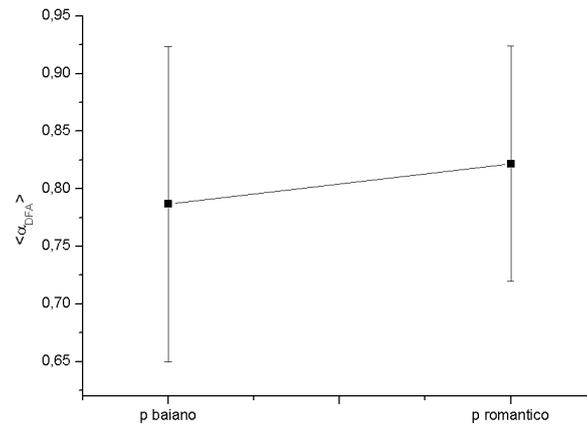


Figura 4.17: DFA expoente dos subgêneros do gênero Pagode. Fonte: o autor.

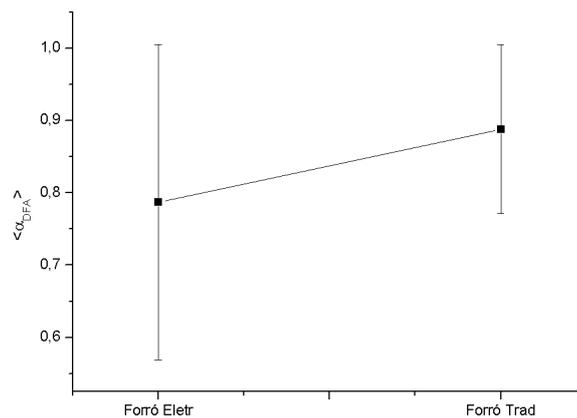


Figura 4.18: DFA expoente dos subgêneros do gênero Forró. Fonte: o autor.

das rádios escolhidas.

As rádios FM utilizadas nesta pesquisa foram gravadas através das versões online, nos endereços a seguir:

- Bahia FM: <http://www.radios.com.br/aovivo/Radio-Bahia-88.7-FM/10351>
- Transamerica FM: <http://www.radios.com.br/aovivo/Radio-Transamerica-Pop-100.1-FM/15712>
- Band News: <http://www.radios.com.br/aovivo/Radio-BandNews-99.1-FM/10403>
- Cultura FM: <http://www.radios.com.br/aovivo/Radio-Cultura-103.3-FM/11335>

A Figura 4.19 mostra os resultados obtidos.

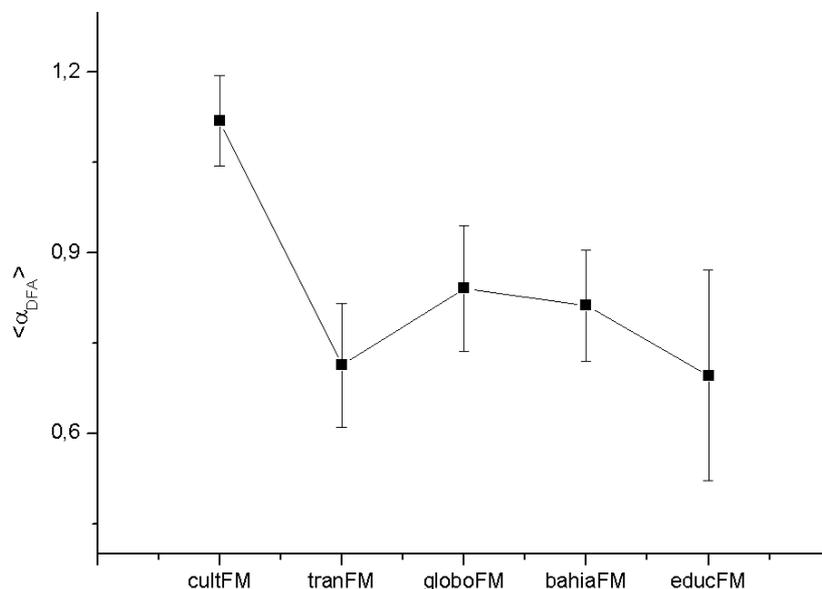


Figura 4.19: DFA expoente da programação musical de rádios FM. Fonte: o autor.

O algoritmo foi também aplicado a um agrupamento de peças musicais de um CD, onde se calculou o DFA expoente de cada música. O objetivo aqui foi analisar o comportamento de cada música dentro do contexto da seleção musical do CD como um todo. Além disso procurou-se comparar o coeficiente DFA de cada música com as constatações encontradas na literatura, quando se procura dar uma interpretação a esse coeficiente, relacionando-o a aspectos rítmicos. Foram escolhidos CD's que pudessem apresentar aspectos distintos de instrumentação, execução e atividade percussiva, e que pudessem ser reconhecidos pelo DFA expoente, expressando o seu nível de *danceability*. Então foram escolhidos o CD João Voz e Violão (<http://joaogilberto.org/disco.htm>), de João Gilberto (Figura 4.23), que traz uma execução bem suave tanto da voz quanto do violão; o Stile Antico (<http://www.stileantico.co.uk/SongofSongs.php>), apenas com música vocal do período da renascença (Figura 4.24); o Caçador de Mim (http://pt.wikipedia.org/wiki/Cacador_de_Mim), de Milton Nascimento (Figura 4.21), que apresenta orquestrações suaves, e também peças com percussão mais ativa; Afrociberdelia (<http://pt.wikipedia.org/wiki/Afrociberdelia>), de Chico Science e Nação Zumbi (Figura 4.22), que apresenta uma forte presença do maracatu e do rock, e uma grande atividade percussiva de tambores com batidas de música eletrônica; e o CD Alfacamabetizado (<http://pt.wikipedia.org/wiki/Alfacamabetizado>), de Carlinhos Brown (Figura 4.20), que mostra uma forte presença da percussão baiana contemporânea.

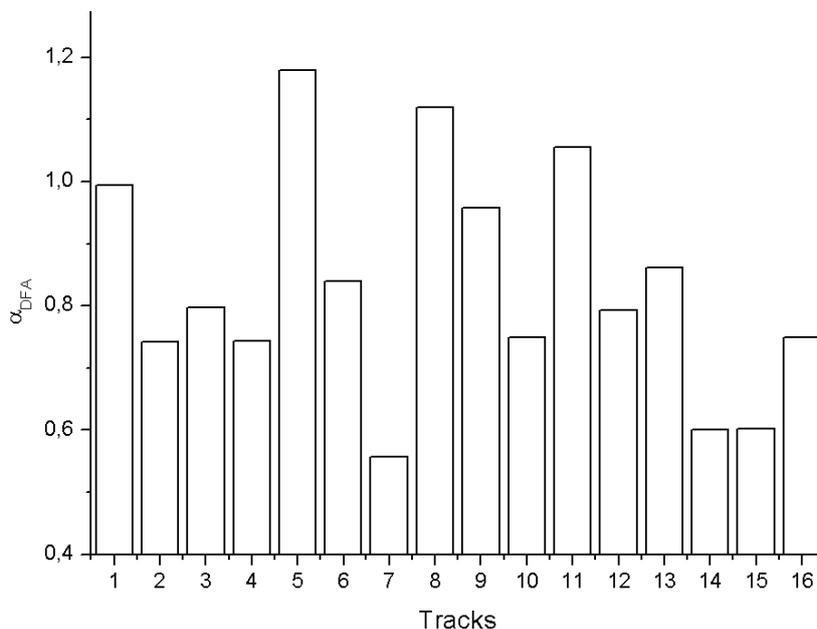


Figura 4.20: DFA expoente das faixas do CD Alfacamabetizado, de Carlinhos Brown. Fonte: o autor.

4.2.5 DFA expoente de Arquivos Editados

Apesar de terem sido usados na pesquisa arquivos em sua íntegra, foi realizada uma pequena investigação a respeito da influência da edição de arquivos de áudio, no valor do DFA expoente.

Foram escolhidos os vinte tracks do gênero forró, e dentre esses foram editados os arquivos com DFA expoente acima de um. Observou-se que os *tracks* 1, 3, 5, 9 (Figura 4.25) estavam acima deste valor. Usando o recurso de ouvir cada um desses *tracks*, foi observado que eventos como pausas bruscas, trechos sem percussão ou com predominância de instrumentos não-percussivos, geralmente na introdução, no final da música ou em trechos de duração importante, estavam aumentando o valor do DFA expoente. Isto se constatou pois, ao excluir essas partes, e aplicar novamente o algoritmo, valor do expoente α_{DFA} caiu significativamente, alcançando valores menores que um (Figura 4.26). Existe uma indicação inicial da influência de eventos não característicos da batida característica de um determinado gênero. Pode ser então que para fins de estudo de recuperação de informações musicais, de treinamento de máquina e classificação de gêneros musicais, seja interessante pensar na possibilidade de um banco de treinamento, em que os arquivos obedeçam alguns critérios de edição.

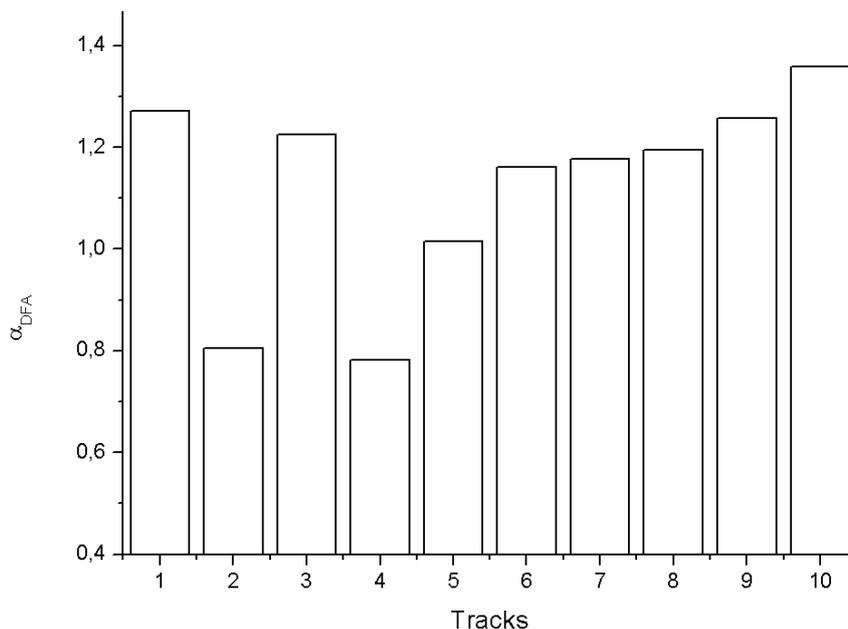


Figura 4.21: DFA expoente das faixas do CD Caçador de Mim, de Milton Nascimento. Fonte: o autor.

4.2.6 *Outliers*

A partir da representação dos DFA expoentes de cada gênero, nas duas bases de dados utilizadas nesta pesquisa, através do boxplot (Figuras 4.6 e 4.12), pode-se observar a presença de *outliers*. Uma vez que o DFA expoente busca estabelecer uma relação entre o arquivo musical e o seu nível de *danceability*, através das flutuações de *loudness*, podemos estabelecer também essa relação, através de um julgamento humano, que é passível de certa variação. Assim, por exemplo, espera-se intuitivamente que músicas com orquestração suave e sem presença percussão repetitiva, tenha um DFA expoente maior do que aquelas com percussão mais pronunciada e repetitiva. A seguir serão apontados alguns aspectos, julgados a partir da audição de cada música, e que podem explicar porque alguns arquivos aparecem estatisticamente como *outliers*.

No banco GTZAN aparecem como *outliers* os arquivos :

- hiphop36 ($\alpha_{DFA} = 0,957$): existem neste arquivo a presença de espaços em branco ou pausas nos primeiros 17", o que pode ser responsável pelo DFA expoente acima da média do gênero hiphop para o banco GTZAN ($0,610 \pm 0,143$), conforme a observação feita na seção 4.2.5. Após esse período a música segue o padrão do hiphop, porém os trechos inicialmente citados elevaram o valor final do DFA expoente.

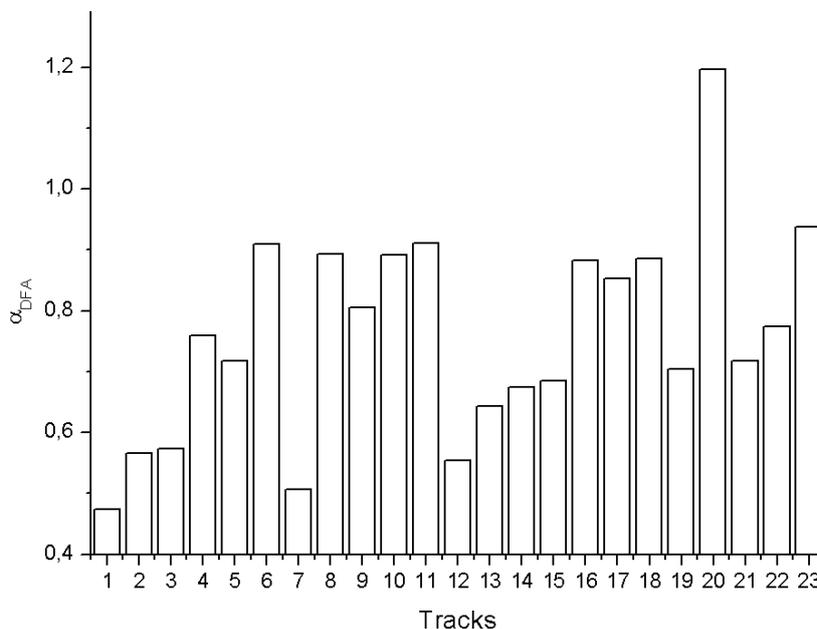


Figura 4.22: DFA expoente das faixas do CD Afrociberdelia, de Chico Science. Fonte: o autor.

- hiphop44 ($\alpha_{DFA} = 1,012$): sem presença da marcante percussão eletrônica característica do gênero hiphop, conta basicamente com duas vozes e alguns efeitos sonoros. Nos primeiros 14" ocorre a narração de uma voz feminina, sem acompanhamento musical, dos 14" aos 24" essa voz executa uma canção em ritmo de uma salsa lenta, de 24" a 30" a narração de uma voz masculina misturada a alguns efeitos de mixagem. Temos aqui um *outlier* superior, portanto o algoritmo reconheceu esse arquivo com baixa *danceability* para o conjunto que forma o gênero hiphop do banco estudado.
- hiphop57 ($\alpha_{DFA} = 0,233$): neste caso temos um *outlier* inferior, com um valor que indica uma *danceability* muito alta, que é ocasionada pela batida forte, marcante e constante desta música.
- disco42 ($\alpha_{DFA} = 0,935$): é um *outlier* acima da média ($0,633 \pm 0,133$). Pode-se observar que, até aproximadamente os 11" de música, acontecem eventos sonoros, onde predominam na mixagem um côro crescente de vozes e metais, e após esses 11", há uma mudança, onde predomina uma percussão mais característica do gênero disco. Daí a primeira parte citada influenciou para o aumento do DFA expoente segundo o que se verificou em 4.2.5.
- reggae83 ($\alpha_{DFA} = 1,202$): é uma balada em ritmo country, com uma bateria bem leve ao fundo e um órgão fazendo acompanhamento, isto é, possui uma batida não característica do gênero reggae. O número encontrado para o DFA expoente reflete

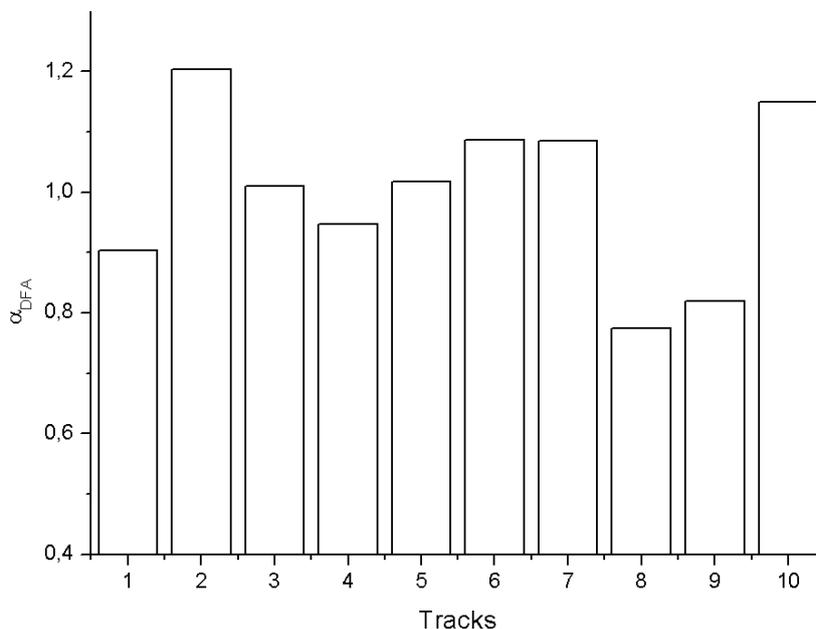


Figura 4.23: DFA expoente das faixas do CD João Voz e Violão. Fonte: o autor.

uma baixa *danceability*.

- reggae40 ($\alpha_{DFA} = 1,183$): essa música apresenta predominância de voz e corno na mixagem, sem a presença do baixo e da bateria na sessão rítmica. Uma percussão leve marca o ritmo. Podemos interpretar o resultado numérico como um indicador de uma *danceability* muito abaixo da esperada para o gênero em questão.
- reggae51 ($\alpha_{DFA} = 0,353$): neste *outlier* inferior, existe uma batida forte e repetitiva, de um loop de bateria, bem semelhante ao encontrado em música eletrônica, hip hop e dance music. O DFA expoente encontrado indica uma *danceability* bem maior que a calculada para a média do grupo ($0,775 \pm 0,146$).
- pop53 ($\alpha_{DFA} = 1,185$): este arquivo representa um *outlier* superior, para o grupo do gênero pop, que possui uma média de $0,811 \pm 0,140$. A canção é a tradicional *amazing grace*, cantada por uma solista e um corno *gospel acapella*. O resultado obtido para o DFA expoente indica uma *danceability* abaixo da esperada para o gênero pop.
- pop11 ($\alpha_{DFA} = 1,264$): é uma balada suave, cuja instrumentação não apresenta bateria, nem percussão. Nesta música um piano e um naipe de cordas, executados suavemente, acompanham uma voz feminina. Daí o DFA expoente calculado refletiu as características desta gravação através de um número que representa uma baixa *danceability* para o grupo pop.
- classical 46 ($\alpha_{DFA} = 0,614$): uma atividade percussiva intensa e constante nos

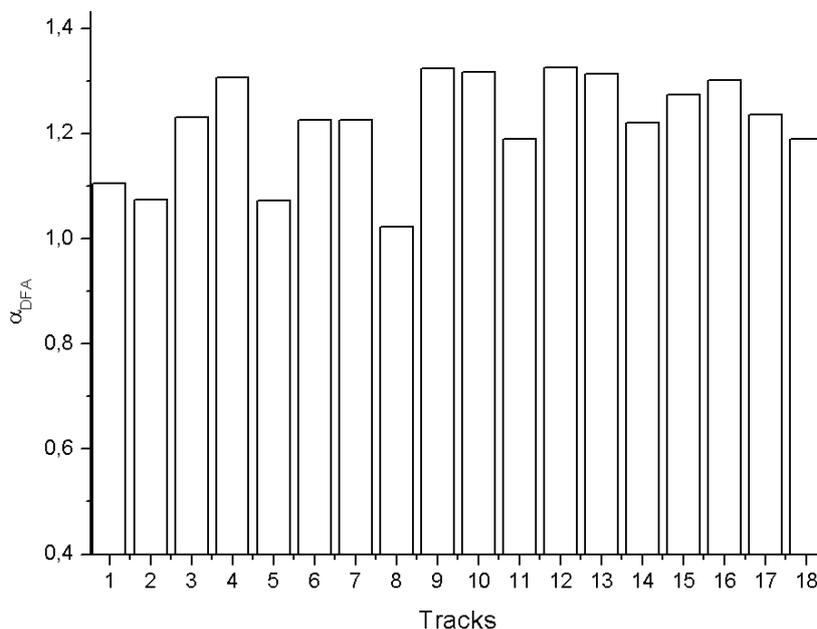


Figura 4.24: DFA expoente das faixas do CD Song of songs, do grupo Stile Antico. Fonte: o autor.

10" iniciais, pode ter determinado a diminuição do DFA expoente, em relação aos outros arquivos do gênero classic. O DFA expoente médio para o gênero classic é $1,070 \pm 0,152$.

No banco BRMUSIC aparecem como *outliers* os arquivos :

- forro3 ($\alpha_{DFA} = 0,973$): nesse arquivo aparecem pausas que podem ter influenciado no aumento do α_{DFA} , conforme visto em 4.2.5.
- forro6 ($\alpha_{DFA} = 0,941$): nos primeiros 27" existe uma introdução instrumental, e logo após, uma voz acompanhada apenas por contrabaixo, teclado e guitarra. A bateria em ritmo de forró só aparece depois desses 27". Daí esse período de tempo inicial influenciou no aumento de α_{DFA} para esse arquivo.

4.2.7 Comparação de Resultados

A partir dos resultados obtidos, foi realizada a comparação com os resultados apresentados em JENNINGS (2004), TERMENS (2009), STREICHS (2006).

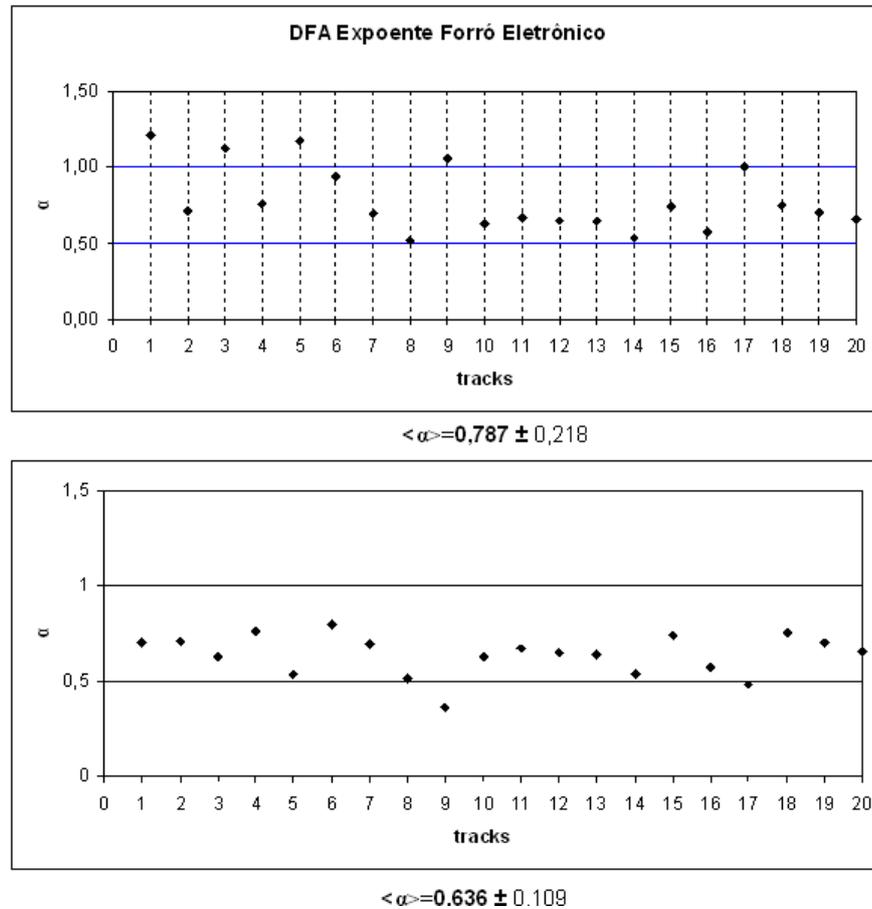


Figura 4.25: Expoente α_{DFA} médio em dos 20 tracks do gênero Forró relativo aos *tracks* 1, 3, 5, 9, (a) antes da edição e (b) após a edição. Fonte: o autor.

No artigo *Variance fluctuations in nonstationary time series: a comparative study of music genres* JENNINGS (2004), o primeiro trabalho a publicar a respeito do DFA expoente ou *Detrended Variance Flutuation Exponent*, não são explicitados os métodos utilizados para mostrar o DFA expoente médio, como índice de distinção entre os gêneros. Também não mostram uma descrição estatística que permita uma comparação mais apurada. Não se sabe também como foram editados os trechos dos arquivos escolhidos. Ainda assim, pode-se notar a semelhança de resultados quando se trata da classificação dos gêneros, quanto às características de atividades percussivas ou *danceability*. Assim os gêneros que possuem alta *danceability*, como Forró e *Techno-Dance* (em JENNINGS (2004)), e Axé Music, Funk, Hiphop e Disco (nesta dissertação) apresentaram os menores valores de $\langle \alpha_{DFA} \rangle$; os de média *danceability* como *Brazilian Pop*, *Rock'n Roll* e *Jazz* (em JENNINGS (2004)), e Rock Nacional, Clube da Esquina e Blues (nesta dissertação), estão associados a uma faixa intermediária; e gêneros de baixa *danceability* como *Javanese Music* e *New Age* (em JENNINGS (2004)), e Bossa-Nova, Vanguarda Paulistana e Classic (nesta dissertação), apresentam os maiores valores de $\langle \alpha_{DFA} \rangle$.

Na tese *Audio Content Processing for Automatic Music Classification: Descriptors, Da-*

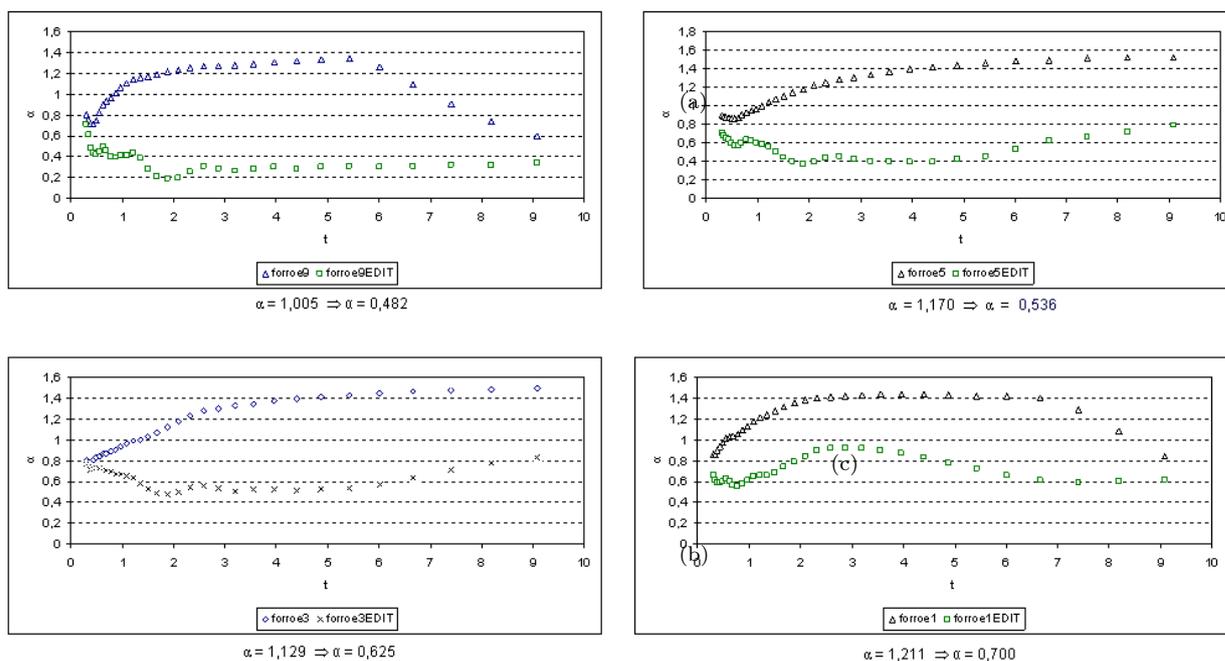


Figura 4.26: Função $\alpha(t)$ dos tracks 9, 5, 3 e 1 e seus expoentes α_{DFA} , antes e depois da edição.

tabases, and Classifiers (TERMENS (2009)), é aplicado o cálculo do DFA expoente no banco GTZAN, um dos bancos usados nesta dissertação. Aqui também não é realizada uma apresentação estatística que permita uma melhor comparação entre os estudos, contudo os resultados apresentados indicam valores do DFA expoente médio bem próximos aos encontrados aqui nesta pesquisa. Se colocarmos os valores do $\langle \alpha_{DFA} \rangle$ em ordem crescente, teremos a uma ordenação de gêneros bem semelhante em (TERMENS (2009)) e nesta dissertação.

Em relação aos resultados apresentados em *Music Complexity: A Multi-faceted Description Of Audio Content* (STREICHS (2006)), foi possível notar uma grande semelhança no comportamento gráfico da função DFA ($F(t)$) e da função $\langle \alpha(t) \rangle$, encontrados nos dois trabalhos.

Considerações finais

A proposta desta dissertação foi realizar o estudo do DFA expoente ou *Detrended Variance Fluctuation Exponent*, sobre dois bancos de dados, sendo um deles inédito, avaliando os resultados encontrados em relação àqueles divulgados por outros pesquisadores, e também estudar o uso do DFA expoente em outros tipos de agrupamento, diferentes da classificação por gênero.

Durante o processo de realização deste trabalho, muitos desafios foram sendo apresentados, dentre eles: a implementação computacional, que em primeira mão exigiu uma compreensão e interpretação rigorosa do algoritmo; a resolução de problemas no processamento de séries geradas pelos arquivos de áudio, pois foi exigida uma manipulação de grandes séries numéricas (na ordem de milhões de pontos); devido a existência de poucos núcleos de pesquisa relacionadas ao tema, houve uma dificuldade de encontrar uma vasta literatura científica relacionada ao tema.

5.1 Conclusões

Sobre os resultados obtidos com a aplicação do algoritmo do DFA expoente aos Bancos GTZAN e BRMUSIC, pode-se concluir que:

- O DFA expoente médio se mostrou um bom descritor na caracterização de gêneros musicais, pois consegue dar informações importantes a respeito de características rítmicas do sinal de áudio. Contudo ele não é auto-suficiente, pois é necessário que exista uma comparação estatística entre as médias dos outros gêneros, a fim de averiguar se existem diferenças significativas entre elas.
- a janela de interesse e o critério distanciamento entre as caixas utilizados neste trabalho funcionaram satisfatoriamente, uma vez que os resultados obtidos pelo algoritmo, refletiram o que se esperava, tendo como base os trabalhos correlatos.
- é necessário um tratamento preliminar das amostras, quando se trata da criação de um banco de dados para classificação em gêneros musicais, pois foi constatado que existe uma influência de alguns eventos atípicos que aparecem nas gravações, e que não são característicos de um determinado gênero, influenciando no DFA expoente calculado.

- ao analisar subgêneros constatou-se que as diferenças numéricas no $\langle \alpha_{DFA} \rangle$ são pouco importantes, ou seja não se consegue fazer distinção entre elas através do DFA expoente médio, segundo o teste estatístico usado neste trabalho.
- é possível criar novas categorias ou rótulos utilizando o DFA expoente, onde podem ser identificadas diferenças e particularidades das categorias em questão. Pode-se então criar um rótulo, como nomes de estações de rádio, e associá-los a um índice numérico que reflete características da seleção musical escolhida para o seu público-alvo.
- ao trabalhar o rótulo "CD de um Artista", os resultados encontrados para cada track podem trazer alguns enganos de interpretação, se comparados ao que se espera através de um julgamento intuitivo. Isto se deve à influência de certos eventos da mixagem na estatística do DFA expoente, ou por outros fatores que ainda não foram suficientemente esclarecidos.

5.2 Contribuições

Esta dissertação espera colaborar com as atividades de pesquisa relacionadas a recuperação de informações musicais e também outras áreas, uma vez que realizou:

- detalhamento matemático de etapas do algoritmo do DFA expoente.
- o estabelecimento de uma relação conceitual entre da janela de interesse do algoritmo para o cálculo do DFA expoente e várias outras janelas de tempo usadas na literatura para estudar percepção rítmica.
- a proposição de um banco de dados inédito para a pesquisa MIR.
- apresentação de um estudo do DFA expoente médio, como expoente na classificação de gêneros musicais, com o auxílio de um método estatístico de análise de variância.
- a apresentação de novas possibilidades de agrupamentos usando o DFA expoente.

5.3 Atividades Futuras de Pesquisa

Com o intuito de dar continuidade à pesquisa utilizando DFA expoente, pretendo realizar:

- a utilização de uma janela deslizante no cálculo da série $V(j)$ (série obtida a partir do desvio-padrão da série original).

- o estudo de outro método de cálculo do DFA expoente.
- a inclusão no algoritmo computacional da importação de arquivos no formato WAVE e sua transformação em uma série numérica.
- a determinação de intervalos de DFA expoente para cada gênero.
- o aprendizado de máquina e a mineração de dados dos bancos GTZAN e BRMUSIC, utilizando o DFA expoente como uma das componentes do vetor de características.
- a definição de novos critérios de edição dos arquivos do banco BRMUSIC a fim de melhorar as estatísticas.
- a ampliação do número de arquivos e de gêneros do Banco BRMUSIC.
- a investigação da influência de novas janelas de interesse sobre os resultados.
- a aplicação de um método estatístico que seja mais apropriado na avaliação da diferença entre as médias, para o caso da classificação em gêneros musicais.

Referências Bibliográficas

PENG, C.K. et al.

ADDISON, P. S. *Fractals and Chaos*. [S.l.]: CRC Press, 1997.

AUDIOLIST. *SPL - Sound Pressure Level*. 2011. URL: [//audiolist.org/forum/lexicon.php?letter=SPL](http://audiolist.org/forum/lexicon.php?letter=SPL); acessado em 12 de outubro de 2011.

BACHMANN, K. Time series analysis as a method to characterize musical structures. In: *International Conference on Music Perception and Cognition, 9th*. Bologna, Italy: [s.n.], 2006. p. 1338–1346.

BARBETO J.; LOPES, A. Automatic genre classification of musical signals. *Journal on Advances in Signal Processing*, v. 2007, n. 64960, p. 12, 2007.

BEROIS, M. *Detecting and describing percussive events in polyphonic music*. Dissertação (Master Thesis in Information, Communication and Audiovisual Media Technologies) — Universitat Pompeu Fabra, Pompeu Fabra, Barcelona, 2008.

BLAIN, G. *Precipitação Fluvial e Temperatura do Ar no Estado de São Paulo: periodicidades, probabilidades associadas, tendências e variações climáticas*. Tese (Tese de Doutorado) — USP, São Paulo, 2010.

CARVALHO, P. C. e. a. *Métodos Matemáticos e Computacionais em Música*. São Carlos, SP: SBMAC, 2009.

CHATFIELD, C. *Analysis of time series an introduction*. London: Chapman and Hall, 1995.

CRYER, J.; CHAN, K. *Time series analyses. With applications in R*. Stanford, CA: Springerand, 2008.

DAVIS G.; JONES, R. *The Sound Reinforcement Handbook*. [S.l.]: Hal Leonard Publishin, 1989.

EHLERS, R. *Análise de Séries Temporais*. São Paulo: Notas de Aula, 2009.

FARLEX. *The Free Dictionary by Farlex/Encyclopedia*. 2011. URL: encyclopedia2.thefreedictionary.com/PCM; acessado em 8 de outubro de 2011.

FEDER, J. *Fractals*. New York: Plenum Press, 1988.

FELDMAN MARTIN; BENDER, M. J. L. persistence of visual or auditory percepts as symptoms of irritative lesions of the cerebrum of man. *Journal of Neurology*, v. 203, n. 3, p. 211–218, 1972.

- FERREIRA, A. B. d. H. *Novo Dicionário Aurélio da Língua Portuguesa*. 2. ed. Rio de Janeiro: Nova Fronteira, 1986.
- FIBEROPTICS. *Fiber Optic Training & Tutorials*. 2012.
URL:www.fiberoptics4sale.com/wordpress/what-is-pulse-code-modulation-pcm/;
acesso em 5 de julho de 2012.
- FORNARI, J. Percepção, cognição e afeto musical. In: KELLER, D.; BUDAS, Z. (Ed.). *Criação musical e tecnologias: teoria e prática interdisciplinar*. Goiânia: ANPPOM, 2010. p. 6–21.
- FRAISSE, P. *Rhythm and Tempo*. New York: D. Deutsch, 1982.
- FREITAS R.; GENOVEZ, P. B. M. R. R. S. Y. Classification of audio signals using statistical features on time and wavelet transform domains. In: *XIV Simpósio Brasileiro de Sensoriamento Remoto*. Seattle, E.U.A.: INPE, 2009. p. 6463–6470.
- FROOTE, J.; COOPER, M. Visualizing musical structure and rhythm via self-similarity. In: *Proceedings of International Computer Music Conference*. Havana, Cuba: [s.n.], 2001.
- GALHARDO, C. *Análise e modelagem estocástica do barorreflexo através de séries temporais de pressão arterial sistólica*. Dissertação (Tese de Doutorado em Física) — Universidade Federal Fluminense, Rio de Janeiro, 2010.
- GUADARRAMA, C. *Nonlinear Audio Recurrence Analysis with Application to Music Genre Classification*. Dissertação (Master Thesis in Sound and Music Computing) — Universitat Pompeu Fabra, Pompeu Fabra, Barcelona, 2010.
- IAZZETA, F. *Audio Digital*. outubro 2011.
URL:www.eca.usp.br/prof/iazzetta/tutor/audio/adigital/adigital.html.
- IAZZETA, F. *Intensidade Sonora*. outubro 2011.
URL:[//www.eca.usp.br/prof/iazzetta/tutor/acustica/intensidade/db.html](http://www.eca.usp.br/prof/iazzetta/tutor/acustica/intensidade/db.html).
- INDIANA. *Digital Audio Formats*. *Jacobs School of Music*. *Indiana University*. 2012.
URL:[//www.indiana.edu/~music/etext/digital_audio/chapter5_file2.shtml](http://www.indiana.edu/~music/etext/digital_audio/chapter5_file2.shtml); acessado em 9 de julho de 2012.
- JENNINGS, H. e. a. Variance fluctuations in nonstationary time series: a comparative study of music genres. *Physica A*, n. 336, p. 585–594, 2004.
- LAMBROU T.; KUDUMAKIS, P. S. R. S. M.; LINNEY, A. Classification of audio signals using statistical features on time and wavelet transform domains. In: *International Conference on Acoustic, Speech and Signal Processing*. Seattle, E.U.A.: [s.n.], 1988. v. 6, p. 3621–3624.

- LEMAN, M. e. a. Correlation of gestural musical audio cues and perceived expressive qualities. In: *Gesture-Based Communication in Human-Computer Interaction. International Gesture Workshop, 5th*. Berlim: [s.n.], 2004. p. 40–54.
- MANDELBROT, B. *The Fractal Geometry of Nature*. New York: Freeman and Company, 1983.
- MANJUNATH, B. S. e. a. *Introduction to MPEG-7: multimedia content description interface*. New York: [s.n.], 2002.
- MAZZONI, D. *File Extension AU*. 2012. URL://file-extension.paretologic.com/detail.php/File-Extension-au.
- MOECKE, M. *Telefonia Digital: Modulação por Código de Pulso - Notas de Aula*. Santa Catarina: cefet-sc, 2004.
- MORETTIN, P.; TOLOI, C. *Análise de séries temporais*. London: Blutcher, 2004.
- MUELLER, A. *Uma Aplicação De Redes Neurais Artificiais Na Previsão Do Mercado Acionário*. Dissertação (Mestrado em Engenharia) — Universidade Federal de Santa Catarina, Florianópolis, 1996.
- OLIVEIRA, A. P. *Pulse Code Modulation*. 2012. URL://paginapessoal.utfpr.edu.br/alessandro/disciplinas-do-semester/principios-de-comunicacao; acessado em 5 de julho de 2012.
- PACHET F.;CAZALY, D. A classification of musical genre. In: *Content-Based Multimedia Information Access Conference*. Paris, França: RIAO, 2000.
- PACHET F.;CAZALY, D. A taxonomy of musical genres. In: *Proceedings of the 6th Conference on Content-Based Multimedia Information Access*. Paris, França: RIAO, 2000.
- PARNCUTT, R. A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, v. 11, n. 72, p. 409–464, 1994.
- PCMAG. *Free Lossless Audio Codec*. 2012. URL://www.pcmag.com/encyclopedia_term/0,1237,t=FLAC&i=43249,00.asp; acessado em 9 de julho de 2012.
- PCMAG. *Wave Audio Format*. 2012. URL://www.pcmag.com/encyclopedia_term/02/C1237/2Ct/3DWAV/i/3D59636/2C00.asp; acessado em 9 de julho de 2012.
- PEARSON, C. The problem of random walk. *Nature*, v. 294; 318; 342, n. 72, p. 419–427, 1905.
- PENG, C. e. a. *Fractal Mechanisms in Neural Control: Human Heartbeat and Gait Dynamics in Health and Disease*. URL: <http://www.physionet.org/tutorials/fmnc/>.

- PENG, C. e. a. Mosaic organization of dna nucleotides. *Physical Review*, n. E49, p. 1685–16895, 1994.
- PEREIRA, B. Séries temporais multivariadas. In: *VI Simpósio Nacional de Probabilidade e Estatística*. Rio de Janeiro, Brasil: [s.n.], 1984.
- PEREIRA, E. M. *Estudos Sobre uma Ferramenta de Classificação Musical*. Tese (Dissertação de Mestrado) — Faculdade de Engenharia Elétrica e de Computação - UNICAMP, São Paulo, 2009.
- PEREIRA, F. A. *Passeio Aleatório*. 2009. URL: fge.if.usp.br/~ttome/cursos/seminarios2/resumo_Felipe.pdf; acessado em 22.03.2012.
- PRIESTLY, M. *Non-linear and non-stationary nime series analysis*. London: Academic Press Limited, 1988.
- ROBERTS, M. *Signals and Systems*. [S.l.]: McGraw Hill, 2004.
- ROY P.;PACHET, F.; KRAKOWSKI, S. Improving the classification of percussive sounds with analytical features: a case study. In: *Proc. of the 8th ISMIR Conference*. Vienna, Austria: [s.n.], 2007.
- SOUZA S.; TABAK, B. C. D. Investigação da memória de longo prazo na taxa de câmbio no brasil. *Revista Brasileira de Economia*, v. 60, p. 193–209, 1998.
- STREICH, S.; HERRERA, P. Detrended fluctuation analysis of music signals: Danceability estimation and further semantic characterization. In: KELLER, D.; BUDAS, Z. (Ed.). *Audio Engineering Society Convention, 118th*. Barcelona, Spain: [s.n.], 2005.
- STREICHS, S. *Music Complexity: A Multi-faceted Description Of Audio Content*. Tese (Doctor in Computer Science And Digital Communication) — Universitat Pompeu Fabra, Barcelona, 2006.
- TERMENS, E. G. *Audio Content Processing for Automatic Music Classification: Descriptors, Databases, and Classifiers*. Tese (Doctor in Computer Science And Digital Communication) — Universitat Pompeu Fabra, Barcelona, 2009.
- TZANETAKIS, G.; COOK, P. Musical genre classification of audio signals. *IEEE Transactions On Speech And Audio Processing*, v. 10, n. 5, p. 293–302, 2002.
- ZEBENDE G. F.;OLIVEIRA, P. M. T. Long-range correlations in computer diskettes. *Physical Review E*, v. 57, n. 3, p. 3311–3314, 1998.

Estudo de Flutuações de Sinais de Áudio Classificados por Gênero Musical

Dirceu de Freitas Piedade Melo

Salvador, Agosto de 2012.