



UNIVERSIDADE ESTADUAL DE CAMPINAS  
Instituto de Artes

**Vinícius César de Oliveira**

**Estratégias composicionais voltadas ao emprego de  
*Machine Listening e Music Information Retrieval* no  
contexto da *live-electronics***

Campinas  
2022

**Vinícius César de Oliveira**

**Estratégias composicionais voltadas ao emprego de *Machine Listening*  
e *Music Information Retrieval* no contexto da *live-electronics***

Dissertação apresentada ao Instituto de Artes da  
Universidade Estadual de Campinas como parte  
dos requisitos exigidos para a obtenção do título  
de Mestre em Música na área de Música: Teoria,  
Criação e Prática.

**Orientador: Prof. Dr. José Henrique Padovani Velloso**

ESTE TRABALHO CORRESPONDE À  
VERSÃO FINAL DE DISSERTAÇÃO  
DEFENDIDA PELO ALUNO VINÍCIUS CÉSAR  
DE OLIVEIRA E ORIENTADA PELO PROF.  
DR. JOSÉ HENRIQUE PADOVANI VELLOSO

Campinas  
2022

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Artes  
Sílvia Regina Shiroma - CRB 8/8180

OL41e Oliveira, Vinícius César, 1987-  
Estratégias composicionais voltadas ao emprego de Machine Listening e Music Information Retrieval no contexto da live-electronics / Vinícius César de Oliveira. – Campinas, SP : [s.n.], 2022.

Orientador: José Henrique Padovani Velloso.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Artes.

1. Composição musical. 2. Música eletroacústica. 3. Música e tecnologia. 4. Música contemporânea. I. Velloso, José Henrique Padovani, 1981-. II. Universidade Estadual de Campinas. Instituto de Artes. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Compositional strategies aimed at the use of Machine Listening and Music information Retrieval in the context of live-electronics

**Palavras-chave em inglês:**

Music composition

Electroacoustic music

Music and technology

Contemporary music

**Área de concentração:** Música: Teoria, Criação e Prática

**Titulação:** Mestre em Música

**Banca examinadora:**

José Henrique Padovani Velloso [Orientador]

Jônatas Manzolli

Ivan Eiji Yamauchi Simurra

**Data de defesa:** 28-01-2022

**Programa de Pós-Graduação:** Música

**Identificação e informações acadêmicas do(a) aluno(a)**

- ORCID do autor: <https://orcid.org/0000-0002-7711-394>

- Currículo Lattes do autor: <http://lattes.cnpq.br/4983693580063613>

# COMISSÃO EXAMINADORA DA DEFESA DE MESTRADO

VINÍCIUS CÉSAR DE OLIVEIRA

ORIENTADOR: JOSÉ HENRIQUE PADOVANI VELLOSO

## **MEMBROS:**

1. PROF. DR. JOSÉ HENRIQUE PADOVANI VELLOSO
2. PROF. DR. JÔNATAS MANZOLLI
3. PROF. DR. IVAN EIJI YAMAUCHI SIMURRA

Programa de Pós-Graduação em Música do Instituto de Artes da Universidade Estadual de Campinas.

A ata de defesa com as respectivas assinaturas dos membros da comissão examinadora encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

DATA DA DEFESA: 28.01.2022

# Agradecimentos

Ao Prof. Dr. José Henrique Padovani, pela dedicação, confiança e apoio ao longo do desenvolvimento da pesquisa.

Aos meus pais, Mara e Roberto, e à minha irmã Leticia, por estarem sempre ao meu lado e por estenderem as suas mãos nos momentos que mais precisei.

À Tatiany, ao Benjamim e à Beatriz, por terem feito parte da minha vida e me acolhido como membro da família durante um período do desenvolvimento desta pesquisa.

À minha grande amiga Sarah Migliori, pela amizade sincera, companherismo, pelas conversas, e por estar comigo sempre que precisei

Aos meus amigos do CACOM, André, Angelo, Djalma, Paulão, Tales, e Thales, pelo companherismo que se estende desde os anos de graduação.

Ao violonista Gabriel Angelo, pelas contribuições e sugestões ao longo da composição de *Do corpo ao pó*.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Processo nº 88887.388098/2019-00.

# Resumo

O presente trabalho consiste em um estudo acerca de estratégias composicionais no emprego de técnicas de *Machine Listening* (ML) e *Music Information Retrieval* (MIR) aplicadas à *live-electronics*, propondo um estudo técnico, contextual e criativo. A pesquisa parte de uma fundamentação teórica que busca discutir e contextualizar o surgimento e desenvolvimento das áreas de ML e MIR enquanto campos de pesquisa interdisciplinares, assim como aborda conceitos ligados aos *sistemas musicais interativos*, *música interativa* e *live-electronics*. A fundamentação teórica ainda se estende à revisão de um conjunto de técnicas, processos e ferramentas, visando elucidar os principais conceitos na implementação dos métodos de ML/MIR. Posteriormente, o estudo investigativo abrange algumas observações gerais acerca das principais características da escrita musical na *música interativa/live-electronics* associadas ao emprego das ferramentas estudadas. Com isso apontamos algumas aplicações práticas e abordagens composicionais em um conjunto de peças do repertório. Por fim, a última parte do trabalho discute o processo criativo de uma peça para violão e *live-electronics*, abordando estratégias particulares de emprego das técnicas de ML/MIR.

**Palavras-chave:** Composição musical; Live-electronics; Machine listening; Music information retrieval; Computação musical.

# Abstract

The present work consists of a study about compositional strategies in the use of *Machine Listening* (ML) and *Music Information Retrieval* (MIR) techniques applied to *live-electronic music*, proposing a technical, contextual, and creative study. The research starts from a theoretical foundation that seeks to discuss and contextualize the emergence and development of the ML and MIR as interdisciplinary research fields, as well as approaches concepts related to *interactive musical systems*, *interactive music* and *live-electronics*. The theoretical foundation also extends to the review of a set of techniques, processes, and tools, aiming to elucidate the main concepts in the implementation of ML/MIR methods. Subsequently, the investigative study covers some general observations about the main characteristics of musical writing in *interactive music/live-electronics* associated with the use of the studied tools. With that, we point out some practical applications and compositional approaches in a set of pieces from the repertoire. Finally, the last part of the work discusses the creative process of a composition for acoustic guitar and *live-electronics*, approaching particular strategies for employing ML/MIR techniques.

**Keywords:** Music composition; Live-electronics; Machine listening; Music information retrieval; Computer music.

# Lista de Figuras

2.1	Curva de Fletcher, Munson (FLETCHER; MUNSON, 1933) para níveis de <i>loudness</i> de 0 à 120 <i>phons</i> (LOY, D. G., 2006, p. 168) . . . . .	42
2.2	Sinal periódico e sua envoltória de energia . . . . .	46
2.3	Estrutura de um filtro FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 46) . . . . .	47
2.4	Resposta ao impulso de um filtro FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 46) . . . . .	48
2.5	Estrutura de um filtro IIR (ZÖLZER; AMATRIAIN, X., 2002, p. 23) . . . . .	49
2.6	Resposta ao impulso de um filtro IIR . . . . .	49
2.7	Representação da atenuação de amplitude nos quatro tipos de filtros (ZÖLZER; AMATRIAIN, X., 2002, p. 32) . . . . .	50
2.8	Representação da resposta frequência dos filtros <i>shelving</i> e <i>peaking</i> (ZÖLZER; AMATRIAIN, X., 2002, p. 32) . . . . .	52
2.9	Resposta frequência do filtro <i>comb</i> baseado no sistema FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 64) . . . . .	53
2.10	Resposta frequência do filtro <i>comb</i> baseado no sistema IIR (ZÖLZER; AMATRIAIN, X., 2002, p. 65) . . . . .	54
2.11	Representação da janela triangular e sua aplicação em um espectro de frequência . . . . .	57
2.12	Representação da janela <i>hann</i> e de sua resposta frequência . . . . .	58
2.13	Representação da janela <i>hamming</i> e de sua resposta frequência . . . . .	59
2.14	Representação da janela <i>blackman</i> e de sua resposta frequência . . . . .	59
2.15	Demonstração do <i>hop size</i> em um sinal segmentado em sucessivas janelas sobrepostas . . . . .	60
2.16	Espectro de <i>Fourier</i> de um sinal periódico (gráfico superior) e sua representação no domínio do <i>quefreny</i> (gráfico inferior) . . . . .	66
2.17	Representação das bandas de frequência do banco de filtros <i>Mel</i> . . . . .	67
2.18	Representação gráfica dos MFCCs extraídos a partir de um $l_{\Lambda_4}$ tocado por uma flauta . . . . .	68
2.19	Representação das bandas de frequência do banco de filtros <i>Bark</i> . . . . .	69
2.20	Representação gráfica dos BFCCs extraídos a partir de um $l_{\Lambda_4}$ tocado por uma flauta . . . . .	69
3.1	Função de sinal $y(K)$ e função resultante do produto de $y(K) \cdot y(K)$ (TEMPELAARS, 2016, p. 261) . . . . .	77
3.2	Função de auto-correlação de um acorde arpejado no violão (acima) e de uma onda senoidal com frequência de 440 Hz (abaixo) . . . . .	78
3.3	Curva de <i>centróide spectral</i> de um mi 4 tocado no violão (acima) e de um mi 4 tocado em um crotales (abaixo) . . . . .	79
3.4	Análise do <i>fluxo spectral</i> (preto) e detecção de <i>onsets</i> (vermelho) do violão tocando a nota Mi4 em <i>tenuto</i> e em seguida executando acordes em <i>rasgueado</i> . . . . .	80

3.5	Análise do <i>espalhamento espectral</i> de uma onda senoidal de 440Hz (esquerda) e de um violino executando vibrato (direita) . . . . .	81
3.6	Diferentes valores de <i>obliquidade espectral</i> de acordo com o comportamento do espectro (PEETERS, 2004, p. 13) . . . . .	82
3.7	<i>Inclinação espectral de um espectro de frequências</i> (PEETERS, 2004, p. 14) . . . .	83
3.8	Curva de <i>roll-off</i> de um <i>pizzicato Bartok</i> (acima) e de um harmônico natural tocados no violão . . . . .	84
3.9	Análise da <i>irregularidade espectral</i> do clarinete (vermelho) e do trompete (azul) executando a nota Mi4 . . . . .	85
3.10	Análise da <i>razão entre componentes ímpares e pares</i> do violino (esquerda) e clarinete (direita) . . . . .	86
3.11	Diagrama com os três valores do <i>tristimulus</i> (YSTAD, 1998, p. 80) . . . . .	87
3.12	Esquema geral de implementação de <i>detectores de onset</i> (BELLO et al., 2005, p. 1036) . . . . .	91
3.13	Análise comparativa de métodos de detecção de <i>onsets</i> (vermelho) a partir do <i>fluxo espectral</i> (superior) e do <i>superflux</i> (inferior) . . . . .	92
3.14	Comportamento rotacional (esquerda) e rotacional (direita) das alturas em diferentes oitavas . . . . .	93
3.15	Chromagrama com as magnitudes de cada altura de um acorde arpejado no violão . . . . .	93
3.16	Curva de <i>aspereza</i> entre dois sons puros (PLOMP; LEVELT, 1965, p. 556) . . .	94
4.1	<i>Patch</i> em Max/MSP de <i>Cortazar, ou quarto com caixa vazia</i> . . . . .	101
4.2	Primeiros compassos de <i>Cortazar, ou quarto com caixa vazia</i> . . . . .	103
4.3	Módulo responsável pela extração da amplitude do sinal de entrada (esquerda) e <i>patch</i> do processo de síntese granular (direita) de <i>Proceratophrys boiei</i> . . . .	104
4.4	Trecho da seção final de <i>Proceratophrys boie</i> . . . . .	105
4.5	Rascunho da trajetória de cada grupo de ouvintes em <i>Ps. I will be home soon!</i> (EINARSSON, 2017, p. 59) . . . . .	106
4.6	Trecho com <i>tremolos</i> de vogais referente aos compassos 122 à 130 de <i>Da una crepa</i> . . . . .	108
4.7	Módulo de análise, extração e armazenamento de atributos sonoros do <i>patch</i> em Max/MSP de <i>Da una crepa</i> . . . . .	109
4.8	Compassos 12 (superior) e 74 (inferior) de <i>Decostructing Dowland</i> . . . . .	112
4.9	Início da seção <b>VC</b> de <i>Tensio</i> . . . . .	114
4.10	Início da seção <b>ID</b> de <i>Jupiter</i> . . . . .	118
4.11	Arquitetura geral do <i>antescofo</i> (CONT, 2008, p. 3) . . . . .	121
4.12	Exemplo da partitura do <i>antescofo</i> e da partitura instrumental do início da seção <b>IA</b> de <i>Do Corpo ao Pó</i> . . . . .	122
4.13	Trecho da seção final de <i>Raggi di Stringhe</i> . . . . .	125
4.14	<i>Setup</i> de <i>Dispersion de trajectoires</i> . . . . .	126
4.15	Exemplo da partitura do <i>antescofo</i> e da partitura instrumental referente aos compassos 251 e 252 de <i>Dispersion de trajectoires</i> . . . . .	127
4.16	<i>Setup</i> de <i>Estremo d'ombra</i> (MORCIANO, 2018, p. 97) . . . . .	128
4.17	Trecho da partitura do <i>antescofo</i> (acima) e da partitura instrumental (abaixo) referente ao compasso 188 de <i>Estremo d'ombra</i> . . . . .	129
4.18	Engendramento de estruturas harmônicas na seção <b>IV</b> de <i>Tensio</i> . . . . .	134

4.19	Arquitetura da implementação das <i>redes harmônicas</i> associadas ao cálculo da tensão harmônica, e ao mapeamento da amplitude do violoncelo em <i>Traverser le réseau</i> (TOFFOLO, 2014, p. 106) . . . . .	138
5.1	Primeira seção de <i>Do corpo ao pó</i> . . . . .	142
5.2	Exemplo de <i>rasgueado</i> combinado e intercalado pela <i>tambora</i> , <i>arpejos</i> e <i>notas ligadas</i> na segunda seção de <i>Do corpo ao pó</i> . . . . .	143
5.3	Exemplo de textura polifônica complexa na seção três (acima) e de textura rarefeita na seção quatro (abaixo) de <i>Do corpo ao pó</i> . . . . .	144
5.4	Exemplo de notação métrica (compasso 63 ao 65) e proporcional na terceira seção de <i>Do corpo ao pó</i> . . . . .	145
5.5	Exemplo de notação aberta na segunda seção de <i>Do corpo ao pó</i> . . . . .	145
5.6	Exemplo de uso de tablatura para indicar o sentido da mão direita na execução do <i>rasgueado</i> em <i>Do corpo ao pó</i> . . . . .	146
5.7	Exemplo de notação do disparos eventos relacionados à eletrônica em tempo-real em <i>Do corpo ao pó</i> . . . . .	147
5.8	Interface do <i>patch</i> de <i>Do corpoa ao pó</i> . . . . .	149
5.9	Exemplo de <i>Patch</i> do <i>Spectrum scale</i> . . . . .	150
5.10	Implementação do <i>Delay spectral</i> . . . . .	150
5.11	Diagrama do funcionamento da <i>síntese concatenativa</i> (STURM, B., 2006, p. 51) . . . . .	151
5.12	Implementação do <i>ataque palindromo</i> em <i>Pure Data</i> . . . . .	152
5.13	Coordenadas esféricas utilizadas na codificação do campo sonoro (THOMAZ, 2007, p. 63) . . . . .	153
5.14	Exemplo de reconhecimento de acordes para controle da espacialização . . . . .	155
5.15	Implementação do <i>patch</i> para detecção de harmônicos e <i>pizzicato Bartok</i> . . . . .	156
5.16	Emprego de estratégias de detecção de harmônicos e <i>pizzicato Bartok</i> no fim da seção <b>III</b> e no início da seção <b>IV</b> de <i>Do corpo ao pó</i> . . . . .	157
5.17	Reconhecimento dos acordes acentuados (vermelho) na seção <b>II</b> . . . . .	157
5.18	Implementação do <i>cepstrum</i> em <i>Pure Data</i> e do <i>patch</i> responsável por ler os dados extraídos . . . . .	158
5.19	Implementação da SNAC em <i>Pure Data</i> . . . . .	159
5.20	Exemplo de utilização do <i>cepstrum</i> e da <i>Special Normalized Autocorrelatoin</i> para controlar processos na seção <b>I</b> . . . . .	159
5.21	<i>Utilização do 3f synthesis para gerar estruturas sonoras/musicais na seção IV</i> . . . . .	160
5.22	<i>Patch de síntese concatenativa retirado dos exemplos do timbreID</i> . . . . .	161
5.23	Curvas de <i>roughness</i> no eixo vertical, e intervalos em <i>cents</i> no eixo horizontal (PORRES, A.; PIRES, 2009, p. 193) . . . . .	162
5.24	Extração da curva de <i>roughness</i> de um acorde para derivar intervalos escalares na seção <b>IIc</b> . . . . .	162

# Sumário

<b>Introdução</b>	<b>13</b>
<b>1 Contextualização da pesquisa</b>	<b>16</b>
1.1 <i>Música mista, live-electronics e sistemas interativos</i>	17
1.1.1 <i>Sistemas musicais interativos no contexto das novas ferramentas computacionais</i>	22
1.2 Contextualização dos métodos de <i>Machine Listening</i> e <i>Music Information Retrieval</i>	23
1.2.1 <i>Machine Listening</i>	24
1.2.2 <i>Music Information Retrieval</i>	27
1.3 Emprego no contexto da criação musical	29
1.3.1 Problemáticas e perspectivas na exploração criativa	31
1.4 Abordagem da pesquisa	34
<b>2 Conceitos, escalas, medidas e técnicas fundamentais</b>	<b>36</b>
2.1 Escalas e medidas de amplitude e frequência	37
2.1.1 Escalas e medidas relacionadas à amplitude	37
2.1.2 Escalas e medidas relacionadas à frequência	41
2.2 Envoltória de energia	45
2.3 Filtros	46
2.3.1 Filtros equalizadores	50
2.3.2 Filtros <i>comb</i>	52
2.3.3 <i>Vocoder</i> de fase	54
2.4 Janelamento	56
2.4.1 Janela triangular	57
2.4.2 Janela <i>Hann</i>	58
2.4.3 Janela <i>hamming</i>	58
2.4.4 Janela <i>Blackman</i>	58
2.4.5 <i>Overlap</i> e <i>hop size</i>	59
2.5 Análise de <i>Fourier</i>	60
2.5.1 Série de <i>Fourier</i>	61
2.5.2 Transformada de <i>Fourier</i>	62
2.5.3 Transformada de <i>Fourier</i> para sinais discretos	63
2.5.4 Transformada de <i>Fourier</i> de curto-termo	64
2.6 <i>Cepstrum</i>	65
2.6.1 <i>Mel Frequency Cepstral Coefficients</i>	66
2.6.2 <i>Bark Frequency Cepstral Coefficients</i>	68

<b>3</b>	<b>Processos de <i>Music Information Retrieval</i> e <i>Machine Listening</i></b>	<b>70</b>
3.1	Categorias e taxonomia geral dos descritores de áudio . . . . .	71
3.2	Descritores temporais de baixo-nível . . . . .	75
3.2.1	<i>Centróide temporal</i> [ <i>Temporal centroid</i> ] . . . . .	75
3.2.2	<i>Taxa de cruzamento por zero</i> [ <i>Zero-crossing rate</i> ] . . . . .	75
3.2.3	<i>Auto-correlação</i> [ <i>Autocorrelation function</i> ] . . . . .	76
3.3	Descritores espectrais de baixo-nível . . . . .	78
3.3.1	<i>Centróide espectral</i> [ <i>Spectral centroid</i> ] . . . . .	78
3.3.2	<i>Fluxo espectral</i> [ <i>Spectral flux</i> ] . . . . .	79
3.3.3	<i>Espalhamento espectral</i> [ <i>Spectral spread</i> ] . . . . .	81
3.3.4	<i>Obliquidade espectral</i> [ <i>Skewness</i> ] . . . . .	82
3.3.5	<i>Inclinação espectral</i> [ <i>Spectral slope</i> ] . . . . .	83
3.3.6	<i>Roll-off espectral</i> [ <i>Spectral roll-off</i> ] . . . . .	83
3.3.7	<i>Irregularidade espectral</i> [ <i>Spectral irregularity</i> ] . . . . .	84
3.3.8	<i>Razão entre componentes ímpares e pares</i> [ <i>Odd-to-even ratio</i> ] . . . . .	85
3.3.9	<i>Tristímulus</i> . . . . .	86
3.4	Descritores de alto-nível . . . . .	88
3.4.1	<i>Estimativa da frequência fundamental</i> . . . . .	88
3.4.2	<i>Detector de onset</i> . . . . .	90
3.4.3	<i>Chroma</i> . . . . .	92
3.4.4	<i>Aspereza</i> [ <i>Roughness</i> ] . . . . .	93
<b>4</b>	<b>Estratégias de aplicação composicional na <i>live-electronics</i></b>	<b>95</b>
4.1	Características gerais da <i>live-electronics</i> no contexto das técnicas de ML e MIR	96
4.2	Paradigmas composicionais associados ao emprego de técnicas de ML e MIR .	98
4.2.1	Mapeamento . . . . .	99
4.2.2	Gatilho ( <i>Trigger</i> ) . . . . .	109
4.2.3	Notação . . . . .	114
4.2.4	Procedimento . . . . .	130
<b>5</b>	<b>Processo composicional: <i>Do corpo ao pó</i></b>	<b>139</b>
5.1	Corpo e territorialidade: a poética no discurso da peça . . . . .	140
5.2	Sobre <i>Do corpo ao pó</i> . . . . .	141
5.3	Escrita musical . . . . .	144
5.4	Estruturação e funções formais da eletrônica em tempo-real . . . . .	147
5.5	Estratégias de aplicação das técnicas de ML/MIR . . . . .	153
	<b>Considerações Finais</b>	<b>163</b>
	<b>Referências bibliográficas</b>	<b>168</b>
<b>A</b>	<b>Apêndice</b>	<b>185</b>

# Introdução

Nas últimas décadas, com o desenvolvimento e expansão das técnicas de processamento digital de sinais (*Digital Signal Processing* - DSP), observou-se um interesse crescente de um conjunto de áreas em desenvolver métodos de análise e extração de informações sonoras/musicais. Dentre esse métodos destacam-se a concepção de técnicas voltadas à *Music Information Retrieval* (MIR) e *Machine Listening* (ML).

ML e MIR são campos de pesquisa interdisciplinares e envolvem um conjunto de áreas, como computação; engenharia elétrica; acústica, e música. Concentram-se em desenvolver técnicas e mecanismos capazes de recuperar informações quantitativas, qualitativas e/ou simbólicas de um sinal de áudio digital. Isso se dá a partir de uma série de métodos de processamento, análise de sinal e modelização de capacidades cognitivas/psicoacústicas. Eventualmente, os dados extraídos pelas técnicas de ML/MIR podem ser correlacionados à características semânticas ligadas à princípios da percepção e da cognição musical, como altura, intensidade, aspereza, etc.

Enquanto os métodos de ML trabalham com a extração de informações simbólicas a partir do uso do protocolo MIDI, os métodos voltados à MIR operam na recuperação de dados diretamente do sinal de áudio digital. A MIR enquanto campo de pesquisa surgiu de uma necessidade em desenvolver ferramentas voltadas à criação de sistemas de busca, recomendação de música, e à manipulação de grandes acervos digitais de música. Embora apresentem diferenças quanto aos seus métodos e arcabouços técnicos, ML e MIR possuem alguns pontos de convergência, quando exploradas conjuntamente no contexto da criação musical possibilitam a criação de uma grande paleta de possibilidades criativas. Com o advento dos ambientes de computação musical em tempo-real, compositores têm cada vez mais incorporado técnicas de ML/MIR em suas práticas composicionais. Estas técnicas passam a ser incorporadas como parte de sistemas interativos em composições musicais que exploram diversos métodos para gerar, espacializar e processar sons instrumentais/vocais.

Neste contexto, o presente trabalho propõe um estudo técnico, contextual e criativo voltado às estratégias composicionais de aplicação de técnicas de ML/MIR no contexto da *música interativa/live-electronics*. Além de possibilitar o entendimento de particularidades técnicas na implementação desses processos, o trabalho busca desenvolver ferramentas técnico-criativas para elaboração de estratégias composicionais particulares. A dissertação está dividida em cinco capítulos.

No capítulo 1 realizamos uma contextualização da pesquisa, primeiramente, abordando os termos *música interativa*, *sistemas interativos*, e *live-electronics* em perspectiva às suas ferramentas tecnológicas. Com isso buscamos circunscrever e delimitar um recorte específico acerca dos aspectos técnicos e poéticos que circundam estas práticas musicais. Em seguida introduzimos as áreas de ML/MIR enquanto campos de pesquisa interdisciplinares, contextualizando seu surgimento e desenvolvimento. Para finalizar o capítulo, trazemos ML/MIR para um contexto de exploração criativa ao levantarmos questões relativas à problemática do emprego destas ferramentas na criação musical.

No capítulo 2 apresentamos uma revisão dos conceitos, escalas, medidas e técnicas fundamentais para a implementação dos métodos de ML/MIR. Primeiramente introduzimos algumas escalas e medidas de amplitude e frequência, em seguida métodos de implementação de filtros, técnicas de janelamento e transformada de *fourier*, e por fim descrevemos algumas medidas psicoacústicas. Além de fundamentar e instrumentalizar o capítulo 3, esta revisão técnica contribui para o entendimento de procedimentos que são amplamente utilizadas na elaboração de processos voltados à *live-electronics*.

O capítulo 3 apresenta uma revisão técnica dos métodos e modelos voltados à ML/MIR. Definimos o conceito de descritores de áudio e os classificamos em dois grupos: descritores de baixo-nível e alto-nível. Esse agrupamento proposto foi realizado considerando os métodos de implementação e a natureza das informações que cada um deles recupera. Dessa forma, descrevemos todos os descritores abordados ao longo do trabalho — tanto aqueles discutidos no capítulo 4, como aqueles empregados no projeto composicional — apresentando a definição e os cálculos empregados na implementação de cada um.

No capítulo 4 realizamos uma investigação a respeito das estratégias de aplicação das técnicas de ML/MIR na *live-electronics*. Em primeiro lugar, descrevemos as características gerais da *live-electronics* no contexto das técnicas de ML/MIR, e em seguida propomos uma categorização das estratégias de emprego composicionais comumente utilizadas em quatro

paradigmas: *mapeamento*; *gatilho*; *notação*, e *procedimento*, ao mesmo tempo ilustramos sua aplicação a partir de exemplos de onze peças do repertório: Cortazar, ou quarto com caixa vazia (1999), de Silvio Ferraz; Proceratophrys boiei(2016/2019), de Danilo Rossetti; Ps. I will be home soon!(2012), de Anna Einarsson; Da una crepa (2011-2012), de Marta Gentilucci; Deconstructing Dowland(2009), de Natasha Barrett; Jupiter (1987) e Tensio (2010), de Philippe Manoury; Raggi di Stringhe(2011) e Estremo d’ombra (2015), de Lara Morciano; Dispersion de trajectoires (2014), de José Miguel fernandez, e Traverser le réseau (2014), de Rael Toffollo. Com isso forneceremos uma compreensão procedimental geral das realizações técnicas, das concepções poéticas, e das abordagens composicionais que vêm sendo empregadas por compositores na exploração criativa das técnicas de ML/MIR, criando subsídios que instrumentalizaram o desenvolvimento do projeto criativo.

Finalmente, o capítulo 5 aborda a peça *Do corpo ao pó*, projeto composicional fruto da pesquisa criativa. A peça concentra uma gama variada de processos que foram estudados no capítulo 2 e 3, de estratégias empregadas no repertório abordado no capítulo 4, e de planos e métodos composicionais individuais. Neste capítulo descrevemos a poética, os processos composicionais gerais, a escrita instrumental, a concepção dos procedimentos de transformação e geração sonora, e as estratégias de emprego das técnicas de ML/MIR que foram utilizados na peça.

# Capítulo 1

## Contextualização da pesquisa

Neste capítulo introdutório buscamos contextualizar o surgimento e o desenvolvimento das técnicas de *Machine Listening* e *Music Information Retrieval* de maneira a pontuar suas implicações no contexto da criação musical. Para isso, o capítulo está dividido em quatro seções.

A seção 1.1 realizará uma reflexão sobre os termos *música mista*, *live-electronics* e *sistemas musicais interativos/música interativa*, de maneira a pontuar as transformações tecnológicas que moldaram tais práticas criativas. Contudo, a definição destes termos é motivada, evidentemente, não por uma questão puramente terminológica e conceitual, mas sim por uma tentativa de delinear um determinado conjunto de ferramentas/tecnologias associadas à essas práticas, e com isso introduzir o ferramental e o arcabouço técnico que nos são de interesse específico neste trabalho. Não pretendemos abordar tais termos de maneira historicista. Embora possam estar relacionadas à práticas composicionais que nos remetem à certas estéticas e períodos históricos específicos, essas abordagens não são de maneira alguma restritivas quanto ao emprego fora de seu contexto original.

Na seção 1.2 definiremos os conceitos relacionados às áreas de *Machine Listening* e *Music Information Retrieval* enquanto campo de pesquisa interdisciplinar, abordando seu emprego em diferentes frentes de atuação, e contextualizando seu surgimento e desenvolvimento em perspectiva às transformações tecnológicas relacionadas ao ferramental tecnológico e ao arcabouço técnico implicados no seu uso.

A seção 1.3 abordará as implicações técnicas, poéticas e éticas do emprego dessas ferramentas na criação. Com isso pretendemos aproximá-las de um contexto exploratório e criativo dentro do escopo da *live-electronics* na atualidade, e discutir em certa medida maneiras de analisar e interpretar os dados extraídos do sinal de áudio por meio dessas

técnicas tendo em vista uma perspectiva de organização e concepção de procedimentos composicionais. Ainda nos é de interesse nesta seção, abarcar questões de caráter ético envolvendo a abordagem e o uso extensivo desses tipos de mecanismos e algoritmos na atualidade.

Por fim, na seção 1.4, definiremos o escopo do trabalho a partir dos pontos levantados e discutidos nas seções anteriores, e assim, delinearemos e descreveremos as abordagens, ferramentas, e os procedimentos empregados na pesquisa, assim como os aspectos técnicos, analíticos e criativos que serão desenvolvidos ao longo do texto.

## **1.1 *Música mista, live-electronics e sistemas interativos***

*Música mista, live-electronics e sistemas musicais interativos* são termos que adquiriram uma certa polissemia ao longo dos anos. São frequentemente utilizados de maneira intercambiável e não sistematizada para designar de forma geral toda a prática musical que de alguma maneira utilize escrita instrumental e sons eletroacústicos — sejam estes gravados, transformados em tempo-real ou resultantes de processos interativos.

Em inglês, o termo “*live-electronic music*” frequentemente é utilizado para se referir tanto à música produzida como executada através de processos em tempo-real, e também à música que combina performance ao vivo e *tape*. Em francês, a música que faz uso de processos em tempo-real é chamada de *traitements en temps réel*, apesar de também ser utilizado o termo *musique mixte*, que antes era designado para qualquer prática musical que combinasse performance ao vivo com sons sob suporte fixo (EMMERSON, P. S., 2013, p. 104). De fato, há um caráter circunstancial no uso de tais termos.

A despeito da realização de um certo léxico com o intuito de delinear as características intrínsecas dessas abordagens, não é de nosso interesse levantar questões acerca do uso destes termos, salientamos novamente que o objetivo aqui é de circunscrever uma série de ferramentas e aportes técnicos específicos que são de nosso interesse.

Para tanto, percorremos de alguma maneira quatro momentos pontuados por grandes mudanças tecnológicas que levaram ao desenvolvimento dessas práticas: (1) surgimento e estabelecimento dos estúdios e centros de pesquisa; (2) adaptação dos aparatos tecnológicos de seu contexto técnico original (estúdio) para aquele da performance (palco); (3) utilização de instrumentos e interfaces digitais baseados em mensagens de controle (MIDI) para tratar

dados com o objetivo de controlar módulos de síntese e processamento do som em tempo-real, e (4) uso extensivo de computadores de maneira a sintetizar, processar e analisar áudio digital em tempo-real.

O termo *música mista* está diretamente ligado às práticas musicais voltadas à combinação de sons provenientes de fontes acústicas/mecânicas (instrumentos tradicionais/voz) e sons de origem eletroacústica (gerados/manipulados em estúdios) registrados em suporte fixo como por exemplo, fita magnética, CD, HD ou qualquer tipo de armazenamento.

Esse tipo de abordagem eventualmente está relacionada à criação enquanto processo que manipula de maneira paralela — embora apresente uma certa inter-relação —, os (1) suportes de escrita e representação simbólica, como a partitura musical, ou, qualquer tipo de suporte que permita representar eventos distribuídos no tempo e, (2) de geração, manipulação, e registro de sons mediados por recursos técnicos. Portanto, o *misto* aqui nos remete à noção de que paralelamente aos eventos notados na partitura instrumental, coexistem e se mesclam — tanto na criação como na realização — eventos sonoros concebidos e registrados por meio de aparatos tecnológicos como microfones, gravadores, computadores e sintetizadores.

O surgimento da *música mista* remonta ao final da década de 1950 e início da década de 1960. Nesse período, os rápidos avanços da tecnologia decorrentes da Segunda Guerra Mundial e o aumento do interesse de muitos setores em novas técnicas sólidas e um clima econômico expansionista, forneceram subsídios suficientes para o estabelecimento dos estúdios como grandes centros de pesquisa e criação (MANNING, 2004, p. 19), logo, propiciando um ambiente favorável para a inserção dessas ferramentas no contexto da exploração criativa.

Foi durante estes anos, com a contínua atividade de estúdios como aqueles de Colônia, Paris e Milão, que começaram a surgir peças significativas combinando os sons “ao vivo” dos instrumentos musicais com aqueles eletroacústicos previamente concebidos e registrados em suporte (MANNING, 2004, p. 135). Peças como *Musica sue due Dimensioni* (1952-58), para flauta e *tape* de Bruno Maderna — considerada a primeira peça do gênero —; *Rimes pour Différentes Sources Sonores* (1958), para orquestra e *tape*, de Henri Pousseur; *Kontakte* (1958-60), para piano, percussão e *tape*, de Karlheinz Stockhausen; *La fabbrica illuminata* (1964), para soprano e *tape*, de Luigi Nono, e *Laborintus II* (1965), para duas

sopranos, contralto, narrador, coro, grupo de câmara e *tape*, de Luciano Berio, são algumas das obras que ilustram as primeiras tentativas composicionais envolvendo sons concebidos em estúdio aliados à escrita instrumental.

A constante exploração criativa possibilitada pelas tecnologias e aparatos disponíveis nos estúdios nesta época, além de resultar em uma vasta produção musical, acabou gerando um conjunto de ferramentas e arcabouços técnicos que são amplamente utilizados nos mais diversos contextos criativos. Técnicas como cortes, eco, reverberação, *loops*, *delay*, sons reversos e manipulações na velocidade de reprodução da fita magnética foram constantemente transpostas e adaptadas a outras situações, impulsionando assim novas práticas criativas.

As primeiras incursões no campo da *live-electronics* surgem deste contexto de adaptação tecnológica. Motivados por esse desejo de transferir os procedimentos analógicos do estúdio para a situação de palco e incorporá-los nas performances ao vivo, surgem as primeiras experimentações voltadas ao uso de aparelhos de síntese e processamento para gerar/transformar áudio em tempo-real (MANNING, 2004, p. 157) — i.e., sem a intermediação primordial de um suporte fixo, tal como ocorre nas práticas de *música mista*.

Neste sentido, *Imaginary Landscape n° 1* (1939), para piano, prato e dois toca-discos, de John Cage pode ser considerada uma obra embrionária, uma vez que utiliza o toca-discos em contexto de performance e o explora como um “instrumento”. Cage, em vez de usar instrumentos eletrônicos da época, como o *Ondes Martenot* ou o *Theremin*, optou por um aparelho acessível, que nunca antes foi pensado como um "instrumento", e que teoricamente poderia ser tocado por qualquer pessoa (COLLINS, 1988, p. 1).

Apesar de *Imaginary Landscape n° 1* apresentar um caráter inovador no que diz respeito a inserção de aparatos de estúdio na performance, foi somente em meados da década de 1960 que os compositores começam de fato explorar a interação desses aparelhos e dispositivos com instrumentos acústicos. Com isso a *live-electronics* incorpora em sua prática criativa a ideia de extensão técnica instrumental, noção essa que foi explorada composicionalmente e poeticamente em grande escala nas produções a partir desse período.

Segundo Padovani e Ferraz (2011) , essa nova conjuntura permitiu que a noção de técnica estendida fosse repensada a partir de uma perspectiva em que as tecnologias passam a ampliar as possibilidades técnicas, sonoras e expressivas dos instrumentos musicais tradicionais, transformando assim a paleta criativa da composição musical. Essa série de

processos técnicos e poéticos possibilitam ampliar, deformar e/ou transformar as possibilidades sonoras de instrumentos musicais associadas aos mais diversos processos de transformação/geração sonora.

*Transición II* (1959), para piano, percussão e dois gravadores de fita, de Mauricio Kagel, é considerada uma das primeiras obras a explorar essas novas abordagens. Nessa peça, enquanto um gravador reproduz material pré-gravado, outro registra trechos da performance e reproduz *loops* dos gestos instrumentais tocados ao vivo permitindo que o pianista toque com seus próprios ecos do passado. Processos como esses despertaram um grande interesse durante a década de 1960, e diversos compositores envolvidos com música eletroacústica passaram a explorar essas novas possibilidades.

Peças como *Mikrophonie I* (1964), *Mixtur* (1964) e *Mikrophonie II* (1965), de Karlheinz Stockhausen, são marcantes quanto a utilização desses aparelhos e técnicas de maneira a criar processos temporais (*delay* e *feedback*), modulatórios (modulação em anel) e timbrísticos (modulação em anel, filtragens), estimulando concomitantemente a busca por novos aportes poéticos, criativos e técnicos.

É igualmente notável que uma peça como *Solo* (1966), também de Stockhausen, explore processos temporais de *delay* e *feedback*, a partir da leitura e gravação de fitas magnéticas durante a performance, ao mesmo tempo que busca uma estruturação bastante flexível e aberta da forma. As possibilidades oferecidas pela *live-electronics* com relação à reprodução e ao deslocamento temporal do material gravado durante a performance, e de sua retroalimentação parecem ter influenciado a própria concepção da estrutura temporal da peça, estimulando a busca por uma sintaxe composicional que viesse a dialogar de maneira análoga ao procedimento técnico explorado, com o tempo e a forma.

No final da década de 1980 e início da década de 1990, com desenvolvimento do MIDI (*musical instrument digital interface*) enquanto protocolo de comunicação para controlar e interconectar sintetizadores, *samplers* e módulos de processamento, as possibilidades colocadas à *live-electronics* foram transformadas consideravelmente (HOLMES, 2008, p. 227). Trabalhando com a informação paramétrica do som, passível de ser processada em tempo-real de maneira a controlar ao mesmo tempo e de forma mais fina e contextual os processos sonoros, o uso do protocolo MIDI marcou o desenvolvimento da computação musical em tempo-real ao introduzir os primeiros métodos de *Machine Listening*.

O uso do MIDI associado às técnicas de escuta de máquina contribuiu para o surgimento e desenvolvimento dos *sistemas musicais interativos*. Esses sistemas são *softwares* desenhados especialmente para interagir com performances musicais ao vivo, sendo caracterizados pela capacidade de mudar seu comportamento em resposta à entrada musical (CHADABE, 1984).

Através de dispositivos controladores originalmente concebidos para enviar dados numéricos de controle ou a partir do uso de sensores acoplados a instrumentos tradicionais, tornou-se possível captar as ações do instrumentista em tempo-real, estruturando processos cuja natureza está diretamente ligada às características da performance, tanto de música notada quanto em contexto de improvisação (ROWE, 1993). Essa abordagem trouxe para a prática criativa da música mediada por tecnologias a possibilidade de conceber sistemas interativos/conjuntos técnicos, os quais apresentam uma certa autonomia em relação ao instrumento a eles acoplados, aproximando-se muitas vezes ao comportamento de um *performer*.

Rowe (1993) classifica os *sistemas musicais interativos* a partir de três perspectivas que os caracterizam e os diferenciam quanto às suas finalidades musicais: (1) sistemas que respondem a eventos musicais prescritos em uma partitura/roteiro, ou aos parâmetros mais gerais relacionados à performance, como densidade e regularidade; (2) sistemas voltados à transformação, geração e sequenciação de material sonoro/musical, e (3) sistemas que de alguma maneira estendem as características dos instrumentos tradicionais ou que buscam aproximar seu comportamento ao de um instrumentista.

Além de contribuir com reflexões acerca do uso dos *sistemas musicais interativos* em contexto de improvisação (LEWIS, 1999) (LEWIS, 2007), George Lewis levanta questões importantes quanto à concepção e estruturação destes sistemas/conjuntos técnicos. Ao descrever os processos empregados em seu ambiente musical interativo voltado à improvisação — *Voyager* (1986-88) —, Lewis discute sobre como debates políticos a respeito da natureza e função da música e dos corpos estão inevitavelmente embutidos na estrutura desses *softwares*. No caso particular de *Voyager*, a origem cultural afro-americana da composição corrobora com uma identificação de *multidominância*<sup>1</sup> nos níveis tanto da estrutura lógica do sistema quanto de sua articulação performativa (LEWIS, 2000, p. 34).

---

<sup>1</sup>A noção de *multidominância* está presente em (DOUGLAS, 1991), e busca formalizar uma estética afro-americana de forma a integrar elementos visuais e musicais.

Philippe Manoury foi um dos compositores pioneiros a incorporar os *sistemas musicais interativos* em suas composições. *Jupiter* (1987), para flauta e *live-electronics*<sup>2</sup>, e *Pluton* (1988), para piano e *live-electronics*, são emblemáticas neste aspecto. Nestas peças, por meio de instrumentos originalmente construídos para extrair e enviar dados MIDI — e mais tarde com o auxílio de sensores que captam o sinal dos instrumentos e convertem altura, ataque e dinâmica em uma representação simbólica (MIDI) —, foi possível desenvolver procedimentos voltados à *score following*, em que os processos computacionais são controlados e disparados automaticamente em resposta à performance. Esses métodos são discutidos no capítulo 4.

Peças importantes do mesmo período e que também fazem uso dessas estratégias são *Rhizomes* (1993), para dois pianos, percussão e eletrônica, de Michael Jarrell; *D'un geste apprivoisé...* (1995), para fagote e eletrônica, de José-Luis Campanae, e *Anthèmes II* (1997), para violino e eletrônica, de Pierre Boulez (LEMOUTON; MANOURY, 1998, p. 2). Muitos dos processos desenvolvidos em peças compostas neste período com a colaboração técnica de Miller Puckette, no IRCAM (*Institut de Recherche et Coordination Acoustique/Musique*), levaram ao amadurecimento de ambientes de programação musical voltados ao processamento digital de sinais em tempo-real (CONT, 2011, p. 2).

### **1.1.1 Sistemas musicais interativos no contexto das novas ferramentas computacionais**

As possibilidades abertas pelos avanços tecnológicos ocorridos nas duas últimas décadas levaram ao desenvolvimento de um conjunto de novas ferramentas computacionais que vêm sendo constantemente transformadas e integradas às novas práticas composicionais. *Softwares* como Max/MSP, *Pure Data* (PUCKETTE, 2002) e *SuperCollider* (WILSON; COTTLE; COLLINS, 2011) surgiram a partir da segunda metade da década de 1990, impulsionados pela expansão dos métodos de processamento de áudio digital, e com a popularização dos computadores pessoais. Segundo (MANNING, 2004, p. 221), isso ocorre em consequência da revolução do microprocessador, iniciada na década de 1970, bem como da emergência de um mercado doméstico e do desenvolvimento e produção em larga escala de processadores 16-*bits* na década seguinte.

---

<sup>2</sup>Apesar dessas peças apresentarem fortes características do uso de *sistemas musicais interativos*, o compositor utiliza o termo *live-electronics* para se referir à elas

É evidente que a difusão desses *softwares* e *hardwares* impulsionou em grande escala a prática da *live-electronics*, uma vez que permitiu maior acesso da comunidade musical à essas ferramentas — que até então eram muito restritas a artistas vinculados aos grandes centros como IRCAM e GRM (*Groupe de recherches musicales*) —, ao mesmo tempo em que ambientes baseados em programação visual como o próprio Max/MSP e *Pure Data* permitiam a exploração por artistas sem especialização em práticas informáticas de programação.

O desenvolvimento dessas ferramentas abriu novas possibilidades levando à concepção de técnicas mais elaboradas de extração/mensuração de informações sonoras desenvolvidas pelo campo da *Music Information Retrieval* a partir dos anos 2000. Se os processos de *Machine Listening* dependiam até então da conversão das informações extraídas em dados simbólicos, as ferramentas voltadas à *Music Information Retrieval* passam a permitir a extração e análise de dados e informações diretamente do sinal de áudio. Essas informações podem estar relacionadas ao próprio comportamento do sinal, ou aos modelos psicoacústicos.

A incorporação dessas novas ferramentas às práticas musicais no contexto dos *sistemas musicais interativos* aponta em direção à uma abordagem criativa que cada vez mais busca um uso extensivo do computador, de maneira a permitir uma maior complexidade e refinamento das estratégias no que diz respeito à integração entre escrita instrumental/vocal, performance e procedimentos computacionais.

Nas últimas décadas, novas técnicas envolvendo inteligência artificial vêm ganhado maior notoriedade no contexto dos *sistemas musicais interativos*. Através desses processos é possível treinar o computador de maneira a gerar estruturas sonoras/musicais que irão estabelecer relações mais complexas e elaboradas com o músico ao vivo, a partir da análise das informações geradas pela própria performance humana ou de dados pré-analisados. No contexto de sua integração com técnicas de *Machine Listening* e *Music Information Retrieval*, processos de aprendizado de máquina e inteligência artificial podem vir a possibilitar a estruturação de sistemas mais complexos de tratamento de dados de áudio, viabilizando estratégias e processos composicionais com características ainda pouco exploradas.

## 1.2 Contextualização dos métodos de *Machine Listening* e *Music Information Retrieval*

*Machine Listening* e *Music Information Retrieval* são campos de pesquisa interdisciplinares que envolvem áreas como computação, engenharia elétrica, acústica, música, entre outras, e se propõe a estudar e desenvolver modelos algorítmicos voltados à extração e análise de dados e informações relacionadas a aspectos do comportamento do som.

Tais modelos são descritos como mecanismos que, a partir de um conjunto de técnicas específicas voltadas à análise e processamento digital de sinais, recuperam informações quantitativas, qualitativas e/ou simbólicas de um fluxo de áudio, permitindo que através de uma correlação associemos tais informações às características semânticas relacionadas a princípios da percepção e da cognição musical, tais como altura, andamento, intensidade, harmonicidade/inarmonicidade, brilho, etc.

Segundo Collins (2006) , grande parte desses algoritmos são baseados em uma modelagem hipotética das habilidades cognitivas que a percepção humana possui, e buscam de certa forma mimetizar fenômenos psicoacústicos. Contudo, salientamos que a ML e MIR não buscam realizar uma engenharia reversa do sistema auditivo humano e de seu comportamento, muito menos reconstruí-los computacionalmente. Nesse sentido, existem outras frentes de pesquisas voltadas à neurociência, neuro-física e à bioacústica, que realmente dedicam-se à realização de uma modelagem mais precisa baseada na fisiologia do sistema auditivo.

Embora haja uma convergência e integração entre as áreas de ML e MIR, elas se diferem quanto ao ferramental técnico e à abordagem empregada na implementação de seus algoritmos. A *Machine listening* originalmente está ligada à extração de informações musicais simbólicas associadas ao uso do protocolo MIDI e de seu emprego na concepção de *sistemas musicais interativos*, enquanto a *Music Information retrieval* se volta à análise e extração de atributos sonoros diretamente do áudio digital. É evidente que estes dois métodos são empregados de forma combinada e integrada no contexto da *live-electronics* na atualidade, e são responsáveis pela concepção de processos voltados à recuperação de informações sonoras e musicais bastante eficazes e com grande potencial criativo.

### 1.2.1 *Machine Listening*

*Machine Listening* pode ser definida como a capacidade que as máquinas podem apresentar em emular a escuta e as habilidades musicais humanas (COLLINS, 2011a, p. 439). Uma das principais motivações para se implementar algoritmos com essas propriedades, é a possibilidade que esses mecanismos oferecem em mimetizar/gerar de certa maneira materiais musicais, os quais podem estabelecer relações em diferentes níveis com o contexto musical em que estão inseridos. De acordo com (ROWE, 1993)<sup>3</sup>, a modelagem de processos cognitivos humanos na construção de algoritmos/sistemas de escuta de máquina pode ser frutífera, pois, eventualmente, as estruturas musicais que emergirão, embora geradas procedimentalmente, entrarão em ressonância com os processos cognitivos humanos aos quais se destinam.

No entanto, é importante salientar que não há uma teoria sistematizada e definitiva sobre como o sistema auditivo humano recebe e interpreta sinais musicais. Existem vários estudos sendo desenvolvidos por diferentes frentes de pesquisa voltadas à cognição musical, e que apresentam abordagens e resultados os mais diversos. Logo, no contexto dos processos de *Machine Listening*, construir algoritmos e sistemas com essas capacidades exige, sobretudo, fazer escolhas e lidar com as consequências dessas escolhas em situações de performance ao vivo (ROWE, 1993). De fato, a concepção desses algoritmos, bem como as interações realizadas com eles, tendem a refletir as características da comunidade, do pensamento e da cultura que os produziu (LEWIS, 2000, p. 34).

No que diz respeito ao seu aporte técnico, os métodos de extração implementados pela *Machine Listening*, a princípio, estão fortemente ligados a processos voltados à abstração e transmissão de determinadas informações sonoras. Essa abstração é realizada através da representação de informações recuperadas a partir do áudio em dados binários organizados em oito *bits* (um *byte*), embora, frequentemente apenas sete *bits* sejam usados, o que resulta em um intervalo binário que vai de 0 a 127.

A conversão e transmissão dessas informações é realizada pelo protocolo MIDI. O MIDI foi concebido com o propósito básico de conectar e controlar sintetizadores e computadores através de protocolos de comunicação. Esses protocolos são divididos em duas grandes categorias: *mensagens de canal*, e *mensagens de sistema*.

As *mensagens de canal* direcionam um comando (informação) MIDI para qualquer canal disponível. Essas mensagens podem conter informações e parâmetros associados ao início e

---

<sup>3</sup>[https://wp.nyu.edu/robert\\_rowe/text/interactive-music-systems-1993/chapter5/](https://wp.nyu.edu/robert_rowe/text/interactive-music-systems-1993/chapter5/)

fim de notas (*note on* e *note off*); *velocity* — que indica com qual velocidade uma nota foi tocada, portanto, é uma medida associada à força (dinâmica) —, e elementos expressivos, como vibratos, envoltória de amplitude e *portamentos* (HOLMES, 2008, p. 230).

Já as *mensagens de sistema* consistem na transmissão de informações para todos os dispositivos MIDI conectados em rede. Essas informações são utilizadas para controlar eventos temporais, sequências de dados e efeitos, e se dividem em três categorias: *mensagens de sistema comum*, que englobam informações de tempo usadas para sincronizar sistemas MIDI e outros equipamentos, como reprodutores de áudio ou vídeo; *mensagens de sistema exclusivo*, que carregam informações específicas do fabricante do dispositivo MIDI que está recebendo a mensagem — o fabricante pode enviar uma mensagem que apenas dispositivos construídos por ele podem receber —; e *mensagens de sistema em tempo-real*, usadas para sincronizar todos os dispositivos controlados pelo relógio MIDI e que estão conectados em uma mesma rede (HOLMES, 2008, p. 233).

Devido às características intrínsecas do próprio arcabouço técnico empregado nos métodos de *Machine Listening*, é evidente que há perda de informações ao longo do processo de extração e representação. Uma vez que a abstração das informações não permite a realização de um processo reverso que possa reconstruir o sinal original — tal como as técnicas de processamento digital de sinais possibilitam. A conversão de informações sonoras e sua representação simbólica realizada pelo protocolo MIDI descarta uma grande quantidade de informações ligadas a diversos aspectos do comportamento do som.

Dada a natureza simbólica dos processos relacionados à *Machine Listening*, Rowe (2004) os classifica em dois grupos: (1) *processos simbólicos* e (2) *processos sub-simbólicos*. Os *processos simbólicos* são aqueles baseados na representação de objetos e na relação e manipulação dessas representações de acordo com um sistema pré-estabelecido (ROWE, 2004, p. 17), ou seja, informações simbólicas, que representam atributos musicais são analisadas por algoritmos de maneira a inferir suas relações com base em um conjunto de regras específicas. Processos como a identificação e classificação de acordes e a dedução de tonalidade são exemplos de *processos simbólicos*.

Já os *processos sub-simbólicos* são responsáveis por mapear as informações de entrada em um conjunto de classificações. Esses processos podem identificar padrões a partir das informações que entram no algoritmo sem se basear em um sistema de organização específico. O que os difere dos *processos simbólicos* é a capacidade de entender de certa forma

o comportamento do material musical, e também prever o comportamento dos materiais subsequentemente (ROWE, 2004, p. 93).

As *redes neurais* são os processos *sub-simbólicos* mais aplicados em música. Esses algoritmos são inspirados em modelos de estruturas dos neurônios, sendo capazes de detectar relações entre entradas e saídas de dados de maneira a mapear seu comportamento. Com esse tipo de processo é possível estimar a tonalidade de um dado excerto musical, prever o comportamento melódico, mapear a quantidade de eventos simultâneos e sucessivos (ROWE, 1993), quantificar estruturas rítmicas (DESAIN; HONING, 1989), entre outros.

Portanto, em virtude da capacidade de extrair, manipular e transmitir informações paramétricas do som a aparelhos, sintetizadores e computadores interconectados em uma rede, as técnicas empregadas nos métodos de *Machine Listening* estão fortemente atreladas ao surgimento dos *sistemas musicais interativos*, sendo a principal base do desenvolvimento desses conjuntos técnicos, como já relatado na seção 1.1. De fato, a *Machine Listening* marcou de forma decisiva a pesquisa em torno da recuperação de informações sonoras, seja impulsionando em certa medida o surgimento de novos métodos e abordagens, ou na sua incorporação à outras frentes de estudo, como é o caso da *Music Information retrieval*.

### 1.2.2 *Music Information Retrieval*

*Music Information Retrieval* é um campo de pesquisa interdisciplinar em constante expansão que abarca um conjunto de áreas ligadas à ciência da computação, música, processamento digital de sinais, cognição e psicoacústica (FUTRELLE; DOWNIE, J., 2003, p. 121). Seu interesse é voltado ao estudo e desenvolvimento de métodos de análise de áudio digital com o objetivo de extrair e mensurar informações relacionadas aos atributos sonoros e musicais.

Este campo de pesquisa surgiu da necessidade de um conjunto de áreas distintas em desenvolver ferramentas que pudessem automatizar a manipulação de grandes bancos de dados musicais digitais (BYRD; FINGERHUT, 2002). De acordo com J. Stephen Downie (2003), a motivação inicial era conceber mecanismos de indexação de arquivos digitais de música, sistemas de busca, recuperação e recomendação de música baseados em conteúdo, e interfaces que permitissem o fácil acesso a grandes acervos *online*.

A ideia de implementar técnicas de recuperação automática de informações aplicadas à música vem sendo pensada desde meados da década de 1960 (KASSLER, 1966). Mas foi

recentemente, a partir do início da década de 2000, que a *Music Information Retrieval* cresceu e se estabeleceu enquanto campo de pesquisa. O aumento do poder de processamento dos computadores, além da constante expansão e implementação de novas técnicas de processamento digital de sinais que ocorreram a partir da segunda metade da década de 1990, foram decisivas para que técnicas mais elaboradas de análise de áudio digital começassem a ser empregadas extensivamente (SCHEDL; GÓMEZ; URBANO, 2014, p. 4).

Em virtude desse avanço tecnológico, houve um crescente aumento no interesse por acervos de arquivos musicais em formato digital. Serviços *online* como *Napster*, *Spotify*, entre outros, surgiram desse contexto e foram impulsionados pelo desenvolvimento de tecnologias de compressão de áudio, como *mp3* (*MPEG-1/2 Audio Layer 3*), e pela capacidade cada vez maior de armazenamento digital e de largura de banda (FUTRELLE; DOWNIE, J., 2003, p. 121). Em suma, a demanda comercial gerada por essas novas tecnologias foi o fator decisivo que impulsionou o surgimento e a expansão da MIR.

De fato, há uma linha contínua que liga o campo da *Music Information Retrieval* às pesquisas desenvolvidas pela *Machine Listening*. Uma vez que, ao implementar ferramentas voltadas à análise e recuperação de informações do sinal de áudio, bem como buscar transpor alguns métodos de extração para o contexto digital, a MIR acaba por expandir e ampliar, de certa maneira, as ferramentas desenvolvidas até então, incorporando os elementos da ML em suas pesquisas e dando continuidade ao desenvolvimento de técnicas voltadas à recuperação de dados sonoros/musicais cada vez mais eficientes.

Ao desenhar algoritmos que trabalham diretamente com os dados contidos no sinal de áudio digital, os métodos de *Music Information Retrieval* extraem informações de naturezas distintas, as quais estão associadas a aspectos do comportamento do sinal, e/ou à uma apreensão mais qualitativa do som. Logo, essas informações podem ser separadas em dois grupos: atributos de baixo-nível e alto-nível – no capítulo 3 é realizada uma taxonomia desses algoritmos considerando a natureza das informações que cada um deles busca recuperar.

Os atributos de baixo-nível são calculados a partir do sinal de áudio de uma forma direta (*envoltória de energia*, *taxa de cruzamento por zero*, etc.), ou a partir de uma derivação – i.e. por meio da representação do sinal no domínio da frequência através da *transformada de Fourier* (*envoltória espectral*, *centróide espectral*, etc.). As informações de baixo-nível têm pouco significado para os usuários, ou seja, são medidas quantitativas que descrevem o

comportamento físico do sinal de áudio, logo, não apresentam uma correlação direta com atributos ligados à percepção ou à cognição musical. O processo de extração dos atributos de baixo-nível exerce papel fundamental na preparação do sinal de áudio para o cálculo dos atributos de alto-nível (SCHEDL; GÓMEZ; URBANO, 2014, p. 25). Estes, por sua vez, estão correlacionados a modelos baseados na cognição musical e em medidas psicoacústicas que buscam mimetizar o funcionamento do sistema auditivo humano (*pitch*, *chroma*, *Mel Frequency Cepstral Coefficient*, entre outros), e representam medidas qualitativas que estão associadas às características semânticas da percepção. O capítulo 2 é inteiramente dedicado a descrever em detalhes as medidas, escalas, técnicas e cálculos utilizados na implementação desses algoritmos.

No sentido de criar uma taxonomia e classificação, o trabalho desenvolvido pelo projeto *CUIDADO (Content-based Unified Interfaces and Descriptors for audio/music databases available online)* (PEETERS, 2004), busca definir e unificar uma certa padronização organizacional desses algoritmos sob uma perspectiva dos seus processos de cálculo, natureza dos dados recuperados e características intrínsecas.

A *Music Information Retrieval* compreende um vasto e diversificado conjunto de frentes de estudo, cujo escopo abarca assuntos de interesse de várias disciplinas. (SCHEDL; GÓMEZ; URBANO, 2014, p. 16). As principais frentes que compõe as pesquisas em torno da MIR são a engenharia de áudio, que se dedica ao desenvolvimento de métodos de processamento digital de sinais e análise de áudio digital; a ciência da computação e recuperação de informações, que buscam elaborar algoritmos, ferramentas, e *softwares* voltados à recuperação de dados — Chowdhury (2010) traça um panorama do desenvolvimento em torno dessas pesquisas —; musicologia e etnomusicologia, voltadas à análise musical com suporte computacional (ROSSETTI; TEIXEIRA; MANZOLLI, 2018) (NUTTALL et al., 2019), (KRAUSE et al., 2020), (JU et al., 2020), (DAVIES et al., 2020), e a ciência cognitiva, a qual estuda modelos de percepção musical e suas implicações no desenvolvimento de ferramentas (DANNENBERG, R., 2001), (ESLING; CHEMLA-ROMEU-SANTOS; BITTON, 2018), (OFNER; STOBBER, 2020).

### 1.3 Emprego no contexto da criação musical

Em seu emprego voltado à criação, a *Machine Listening*, juntamente com a *Music Information Retrieval*, vem transformando e moldando a poética, a estética, as ferramentas e as estratégias disponíveis à prática musical associada ao suporte computacional. Seja através da concepção de novas ferramentas, ou, a partir de trabalhos que buscam realizar uma reflexão acerca das práticas musicais voltadas à exploração criativa desses algoritmos e sistemas.

Dentro do escopo do desenvolvimento de ferramentas tecnológicas, os trabalhos realizados por pesquisadores, como Pampalk (2006) e Bogdanov et al. (2013), impulsionaram o surgimento de novos ambientes computacionais voltados à extração de dados relacionados às diversas categorias do comportamento do sinal de áudio, como atributos espectrais/harmônicos, temporais e aqueles relacionados à altura e ritmo (SIMURRA, 2016, p. 33). Uma série de ferramentas de suporte à pesquisa no campo da ML e MIR têm sido criadas e utilizadas em diversos contextos e projetos. Isso inclui *softwares* como *Sonic visualizer* (CANNAM et al., 2006) e os *plugins Vamp Pluglins*, desenvolvidos pelo projeto *Mazurka* (CANNAM, 2009)<sup>4</sup>; o ambiente *MARSYAS* (TZANETAKIS; COOK, P., 2002), e as bibliotecas *MIRToolbox* (LARTILLOT; TOIVIAINEN, 2007) e *Librosa* (MCFEE et al., 2015), para *Matlab* e *python*, respectivamente.

No âmbito da modelização e implementação de ferramentas específicas voltadas à criação musical, trabalhos como os de Bullock (2007); Malt, Jourdan et al. (2008); Brent (2009a); Collins (2011b), e Monteiro (2012) concentram-se na concepção de algoritmos desenvolvidos para *Pure Data*, *Max/MSP* e *SuperCollider*, com o intuito de extrair atributos de baixo e alto-nível em tempo-real e diferido, possibilitando seu uso modular. Dessa forma, a utilização dessas ferramentas possibilita a criação de estratégias de detecção de eventos musicais/sonoros, mapeamento paramétrico dos dados extraídos, e classificação e ordenação de timbres.

No contexto da composição assistida por computador, Roger Reynolds utilizou métodos de MIR e ML para estruturar os materiais composicionais de sua peça *The Angel of Death* (1998-2001), para piano, *ensemble* e eletrônica, a partir da correlação de tais materiais a aspectos da percepção auditiva de sua própria audiência (REYNOLDS, 2004). Simurra (2016) faz uso da biblioteca *PDescriptors* (MONTEIRO, 2012), para desenvolver seu ambiente computacional *SSC*

<sup>4</sup><https://www.vamp-plugins.org/guide.pdf>

(*Sound-Shizuku-Composition*) voltado à orquestração assistida por computador. Esse ambiente desenvolvido em *Pure Data* fornece sugestões de combinações orquestrais a partir da análise de bancos de dados sonoros. Esling e Agon (2010) conceberam um sistema voltado à composição assistida por computador em *OpenMusic*, o qual trabalha com extração de atributos temporais, permitindo a detecção de início e decaimento de eventos sonoros.

No campo da performance e dos *sistemas musicais interativos*, encontramos trabalhos como os de Stowell (2010), o qual utiliza mecanismos de extração de informações da voz para aplicação em performances ao vivo, e, de Stowell e Mclean (2013), voltado ao uso de algoritmos de MIR e ML em contexto de *live-coding*. Neste mesmo campo, os trabalhos de Monteiro (2012) e Rimoldi (2018) exploram a criação e performance com *sistemas interativos* e instrumentos musicais digitais/instrumentos expandidos a partir de estratégias de recuperação de informações sonoras e da captura gestual instrumental. Collins (2012) e (ROWE, 2004) abordam o desenvolvimento de algoritmos aplicados à formalização de processos generativos musicais automáticos empregados em *sistemas musicais interativos*. George Lewis (LEWIS, 2007), (LEWIS, 1999), (LEWIS, 2000) implementa *sistemas musicais interativos* aplicados à improvisação, e Cont (2008) implementa métodos de *score following*, voltados ao rastreamento e sincronização de performances à partituras eletrônicas em tempo-real.

Na composição dentro do escopo da *live-electronics* atual, o compositor Daniele Ghisi utiliza a biblioteca *zsa.descriptors* no contexto das técnicas estendidas, para desenvolver algoritmos de detecção de *key clicks* em sua peça *Comment pouvez vous lire à present ? Il fait nuit* (2009), para saxofone alto e *live-electronics* (MALT; JOURDAN, 2009). O compositor Carmine-Emanuele Cella utilizou mecanismos computacionais voltados à MIR e ML para analisar a orquestra durante a performance de sua peça, e para agrupar classes de sons que são usadas para gerar hibridizações sonoras em tempo-real (CELLA; BURRED, 2013).

A ML e MIR abriram uma enorme frente de pesquisa voltada à exploração criativa, resultando em produções bibliográficas que trazem consigo novos aportes técnicos, novas estratégias e uma produção artística diversificada. Através desta breve revisão bibliográfica é possível observar algumas aproximações criativas, as quais nos permitem vislumbrar o emprego de paradigmas na utilização dos métodos de ML e MIR em composição. Processos como mapeamento, concepção de estruturas algorítmicas, e detecção de eventos musicais, são alguns dos exemplos encontrados nos trabalhos citados — no capítulo 4 realizamos uma

reflexão aprofundada acerca das estratégias composicionais que vêm sendo exploradas na *música interativa*. A seguir, apresentamos algumas problemáticas envolvidas na aplicação da ML e MIR em contexto de criação, tangendo de alguma forma as problemáticas e as perspectivas futuras voltadas ao uso criativo dessas ferramentas.

### 1.3.1 Problemáticas e perspectivas na exploração criativa

De fato, a ML e MIR são áreas de pesquisa já bem consolidadas. Em duas décadas de desenvolvimento, tais campos vêm crescendo exponencialmente não só no âmbito acadêmico, como também no cenário industrial/comercial. Esse crescimento fica evidente ao observarmos a principal conferência da área, *International Symposium of Music Information retrieval* (ISMIR). O número de pesquisadores aumentou mais de cinco vezes, contando com menos de cem participantes no início dos anos 2000, e chegando a quinhentos em 2019. Houve também um aumento de 40% no número de trabalhos publicados desde a primeira edição da conferência (LOW et al., 2019, p. 218).

No entanto, apesar das pesquisas nestes campos estarem em constante expansão, transformando as possibilidades colocadas às práticas musicais de maneira a impulsionar pesquisas com vieses criativos bem diversificados, ainda há uma escassez de trabalhos preocupados em realizar reflexões acerca da exploração composicional dessas ferramentas. Isso se deve pelo caráter interdisciplinar da área e pelo fato de que grande parte da comunidade acadêmica — composta em sua maioria por engenheiros e programadores — impulsionada majoritariamente por interesses comerciais, está focada em desenvolver ferramentas ligadas à música *mainstream*, logo, não estão necessariamente preocupados com problemas específicos do campo da composição, principalmente em relação às problemáticas colocadas à música atual (CELLA, 2020, p. 130).

Segundo Low et al. (2019), embora a indústria da música já fosse considerada um nicho com grande potencial para as pesquisas em ML/MIR desde o início da década de 2000, com o desenvolvimento de aplicativos e plataformas como *Shazam* (WANG, A., 2003) e *Last.fm*, foi recentemente que a comunidade ligada a essa área intensificou a relação acadêmico-industrial a partir de demandas geradas por empresas como *Spotify*, *Google*, *Pandora*, *Deezer*, etc., estimulando o desenvolvimento de ferramentas em uma escala ainda maior.

É evidente que os meios para se analisar e extrair informações do som estão bem estabelecidos e consolidados, porém, nos parece uma questão central e, sobretudo, atual a reflexão acerca de como manipular, interpretar e/ou categorizar os dados extraídos de maneira que seja possível organizar procedimentos e estratégias composicionais, e compreender como esses mecanismos e algoritmos podem dialogar ao mesmo tempo com concepções poéticas, escolhas estéticas, e soluções técnicas.

Logo, se torna viável circunscrever estes questionamentos de maneira particular às perspectivas de um conjunto de necessidades específicas da composição, visto que a tecnologia incorpora um papel fundamental nas diversas etapas dos processos criativos dentro do escopo da *música interativa*. De fato, as minúcias técnicas ligadas à construção de processos de transformação/geração sonora e ou interação, os quais se alicerçam sobre métodos de escuta de máquina e recuperação de informações sonoras, impactam diretamente a criação, colocando-se assim como um elemento de fundamental importância no processo composicional como um todo.

O caráter interdisciplinar e comercial da ML/MIR ainda pode se desdobrar apontando para outra problemática. As técnicas e parâmetros empregados neste campo derivam de, e também refletem, as áreas comercialmente dominantes da música *mainstream* e da tradição musical ocidental como um todo. No entanto, essas técnicas e ferramentas são aplicadas em diversos contextos como se fossem universais (BORN, 2020, p. 195). Nem todas as ferramentas desenvolvidas por este viés suprem as necessidades e dificuldades específicas que compositores podem vir a enfrentar no processo criativo, não só no contexto da *música interativa* atual, mas também em práticas musicais não ocidentais, por exemplo. Técnicas de mensuração de estruturas rítmicas, ou aquelas voltadas à *beat track*, nem sempre fazem sentido quando aplicadas em músicas nas quais a linguagem rítmica se desenvolve em torno de diferentes conceitos de tempo. Da mesma forma, medidas relacionadas à altura, como *chroma* podem não dialogar com práticas musicais baseadas em outros sistemas de afinação (CELLA, 2020, p. 130).

De fato, existem trabalhos recentes voltados à implementação de técnicas de ML/MIR em práticas musicais não ocidentais, como o projeto *CompMusic*<sup>5</sup>, que trabalha com abordagens computacionais para descrever gravações musicais não ocidentais, focando em cinco culturas do Magrebe, China, Turquia, e Norte e Sul da Índia (SERRA, 2014), e o trabalho

---

<sup>5</sup><https://compmusic.upf.edu/>

de Cornelis et al. (2013) , que visa analisar como abordagens de estimativa de tempo automatizadas operam no contexto da música centro-africana. Entretanto, como pensar o uso dessas ferramentas em situações criativas as mais diversas, ainda parece ser uma questão que necessita de um maior aprofundamento reflexivo e técnico.

Outra problemática que incita reflexões quando se busca conceber métodos para recuperar, extrair e manipular dados, é a ética na implementação e uso de algoritmos. A ética sociotécnica da computação (JOHNSON, D. G., 2008) defende que um *software*/algoritmo não é um objeto isolado, mas sim uma compatibilização de arranjos humanos, artefatos técnicos e práticas sociais em um sistema sociotécnico (HUGUES, 1994) que envolve uma interação entre tecnologia e sociedade, na qual ambas se moldam.

Esta noção nega a ideia de que o desenvolvimento tecnológico pode ser ponderado apenas através de uma perspectiva interna, de forma a objetivar o progresso em direção à concepção de artefatos tecnológicos objetivamente aprimorados (HOLZAPFEL; STURM, B. L.; COECKELBERGH, 2018, p. 46). Motivado por tal perspectiva mais ampla, a não neutralidade e a importância da transparência da tomada de decisão algorítmica é um assunto que deve ser colocado em discussão, principalmente quando há algoritmos sendo treinados com vieses, os quais refletem comportamentos intolerantes quanto às identidades racial e de gênero (NOBLE, 2018).

Outro desdobramento ético no uso de algoritmos diz respeito à exposição de dados a qual estamos sujeitos constantemente, uma vez que mecanismos voltados à recuperação e ao tratamento de informações vêm sendo empregados extensivamente na coleta de dados pessoais. Isso tem ocorrido em virtude do advento do *Big Data*, impulsionado pela enorme quantidade de dados gerados por usuários conectados à *internet*, bem como pela capacidade cada vez maior de processamento dessas informações por algoritmos.

Quanto mais nossas vidas se refletem em uma realidade virtual e são registradas, mais nosso presente e passado se tornam quase que completamente transparentes e acessíveis (ZWITTER, 2014, p. 3). Redes sociais se tornaram a principal ferramenta geradora de dados em grande escala. Por exemplo, o *Facebook*, responsável pelo maior site de rede social com cerca de 2 bilhões de usuários,<sup>6</sup> processa 4 *petabytes* de dados coletados de seus usuários por dia<sup>7</sup> (WIENER; BRONSON, 2014).

---

<sup>6</sup><https://acodez.in/facebook-users-worldwide/>

<sup>7</sup><https://research.fb.com/blog/2014/10/facebook-s-top-open-data-problems/>

Além de informações diretamente observáveis, as redes sociais também podem rastrear facilmente os registros e padrões de navegação, busca ou dados secundários que fornecem informações sobre orientação sexual, pontos de vista políticos e religiosos, raça, uso de substâncias, inteligência e personalidade geral, estado mental, pontos de vista individuais, entre outros (KOSINSKI et al., 2014). Estas informações geralmente são utilizadas visando fins econômicos, políticos e comerciais.

Neste contexto, a apropriação e o uso criativo de ferramentas dessa natureza, nos parece um argumento bastante válido enquanto resposta ética à maneira em que inteligências artificiais e mecanismos de recuperação e extração de dados vêm sendo empregados na atualidade. Portanto, este trabalho busca pensar a composição musical por uma perspectiva que, tanto quanto possível, aborde as questões aqui destacadas, de forma a contribuir com reflexões acerca do fazer musical e do uso de tecnologias na composição.

## 1.4 Abordagem da pesquisa

A pesquisa tem como objetivo a realização de um estudo técnico e composicional voltado ao emprego de métodos de *Music Information Retrieval* e *Machine Listening* dentro do recorte da *música interativa* na atualidade. A partir da breve contextualização realizada é possível entrever que, apesar da relevância e do potencial que as áreas de ML e MIR vêm apresentando em geral, ainda há a necessidade de trabalhos que desenvolvam e reflitam sobre os desafios técnicos e poéticos enfrentados no emprego dessas ferramentas quando incorporadas pelo campo da composição. A reflexão acerca das problemáticas do emprego criativo de tais ferramentas mostra que, apesar da constante expansão das pesquisas nestas áreas, há ainda um potencial criativo pouco explorado.

Nesse sentido, a pesquisa visa contribuir, com este estudo técnico e conceitual dos métodos e ferramentas de recuperação de informações sonoras/musicais e de escuta de máquina, ao mesmo tempo que busca apontar para algumas estratégias composicionais que dialoguem com a arquitetura desses algoritmos por meio da investigação do uso específico em peças do repertório, e ainda, da concepção de estratégias individuais na composição de uma nova peça.

O estudo técnico caminha de forma paralela à fundamentação teórica acerca dos processos e mecanismos de interesse deste trabalho e à elucubração de soluções

técnico-poéticas, tanto no repertório abordado, como no processo composicional desenvolvido.

Apesar da pesquisa abordar algumas peças dentro do recorte da *live-electronics*, não se trata de um trabalho com viés analítico, o que se busca é observar como compositores vêm pensando e explorando as potencialidades da ML e MIR nas últimas décadas, de modo que, através do apontamento de certas aproximações criativas, seja possível organizá-las em *paradigmas* composicionais.

Devido a variedade poética das peças escolhidas, essa abordagem permitirá, além de um mapeamento geral da aplicação criativa dessas ferramentas, também possibilitará a discussão de certas peculiaridades dos processos criativos dos compositores em questão. Neste contexto, as peças abordadas procuram demonstrar uma diversidade, seja em termos estéticos/poéticos, seja em termos de escolhas e uso de ferramentas e técnicas.

Por fim, o trabalho busca discutir de maneira pessoal as estratégias, escolhas e soluções técnico-poéticas no processo composicional da peça *Do corpo ao pó*(2021), para violão e *live-electronics*. Essa etapa reflete a convergência e o empenho na apreensão de um conjunto de aspectos ferramentais associados à escrita instrumental integrada a processos interativos, à implementações de ferramentas e à aplicabilidade de estratégias criativas em uma nova peça. Trata-se de um processo composicional que tange os pontos discutidos ao longo da pesquisa, visando abordar aspectos práticos da escrita musical em conjunto a elementos poéticos e estéticos de interesse pessoal.

Essa estrutura, a qual dialoga com as três frentes da pesquisa — a saber: fundamentação teórica/estudo técnico, investigação e mapeamento de estratégias criativas, e processo composicional —, irá conduzir a organização geral do texto. Dessa forma, o estudo técnico instrumentalizará as demais etapas, permitindo a compreensão do amplo panorama de estratégias de forma a viabilizar sua adaptação a situações artísticas particulares, bem como à concepção de novas abordagens e ferramentas composicionais. Assim, o trabalho embasa a sua contribuição dentro do escopo técnico e criativo nesse recorte específico da prática da *música interativa*.

## Capítulo 2

# Conceitos, escalas, medidas e técnicas fundamentais

Este capítulo tem por objetivo introduzir medidas, escalas, técnicas e conceitos relacionados aos métodos de Processamento Digital de Sinais que são empregados na implementação de processos de *Machine Listening* e *Music Information Retrieval*. Tendo em vista que a mecânica operacional destes processos depende de conceitos, medidas, escalas e técnicas de processamento específicas, o intuito, aqui, é aquele de possibilitar uma melhor organização da revisão técnica que se faz necessária para uma compreensão mais global das especificidades e de emprego de técnicas de MIR/ML em peças do repertório da música interativa e da *live-electronics*.

Dado a grande quantidade de técnicas e conceitos relacionados ao Processamento Digital de Sinais e a especificidade de aplicação de apenas alguns desses processos no escopo de interesse desta pesquisa, não serão abordados, aqui, conceitos mais específicos relacionados, por exemplo, a digitalização de sinais de áudio, às diferentes implementações de filtros, a processos modulatórios ou a processos de transformação sonora mais específicos. Enquanto tais métodos podem ser vistos em livros como o de Zölzer e Xavier Amatriain (2002), Gareth Loy (2007), Tempelaars (2016), e Roads (1996), por exemplo, o interesse aqui é estritamente voltado à compreensão de elementos técnicos que constituem processos de ML e MIR que serão abordados no capítulo 3.

O capítulo está organizado da seguinte maneira: a primeira parte trata de escalas e sistemas de medidas de amplitude e frequência, e envoltória de energia; e a segunda parte aborda conceitos e técnicas voltados aos processos espectrais.

## 2.1 Escalas e medidas de amplitude e frequência

As técnicas de processamento digital de sinais no contexto da extração de características do sinal de áudio implicam na medição de determinadas grandezas físicas inerentes ao comportamento do som. Os processos de *Music Information Retrieval* e *Machine Listening* trazem a necessidade de organizar e expressar, de diversas maneiras, grandezas como amplitude e frequência de forma que possam ser ordenadas em uma escala de valores. Os métodos de quantificação e organização dessas grandezas variam de acordo com a perspectiva a partir da qual se observa o fenômeno sonoro e, em alguns casos, podem combinar mais de uma forma de representação visando estabelecer relações entre escalas, unidades de medida ou grandezas diferentes.

Para mensurar informações relacionadas a amplitude podemos abordá-la em termos de energia, ou a partir de aspectos psicoacústicos. Por outro lado, com relação ao campo das frequências, além da representação mais tradicional das oscilações harmônicas em termos de ciclos por segundos, ou a partir de diferentes unidades baseadas em distâncias intervalares, existem medidas e escalas construídas a partir de modelos psicoacústicos que são ostensivamente empregadas na mecânica de descritores de áudio. Dada a importância dessas medidas para a implementação de processos de MIR e ML, apresentamos nas seções seguintes as premissas envolvidas na medição e na organização dos valores de grandezas ligadas à amplitude e frequência.

### 2.1.1 Escalas e medidas relacionadas à amplitude

Amplitude é uma variável intrínseca a todos os movimentos oscilatórios, e pode ser definida como uma medida escalar da diferença máxima de quantidade perturbada em relação ao seu valor médio (AUDI, R.; AUDI, P. R., 2017)<sup>1</sup>. Para uma onda sonora, a amplitude é a medida escalar da diferença entre o pico de pressão da onda e a pressão atmosférica ou ambiental padrão (LOY, D. G., 2006, p. 117). Em termos sonoros, essa medida está correlacionada ao quão forte ou fraco um som é percebido pela audição, o que está diretamente relacionado à característica do aparelho auditivo humano ser especialmente sensível a pequenas variações de pressão (HELLER, 2013, p. 3)

---

<sup>1</sup>Livro eletrônico, sem numeração única de páginas/localizações. Ver verbete “wave”.

A amplitude está diretamente relacionada com a *potência*, com a *pressão sonora*, com a *intensidade sonora* e com o *loudness*, isto é, com a sensação subjetiva do quão forte um som é. Para mensurar tais parâmetros, são utilizadas diferentes medidas e conceitos, que serão abordados a seguir.

## Logaritmos e Decibel

Para facilitar comparações, geralmente valores relacionados ao nível de *potência*, *pressão* e *intensidade sonora* são expressos em uma escala logarítmica chamada *decibel* (dB) (PUCKETTE, 2007, p. 4). O *decibel* é uma unidade logarítmica usada para descrever proporções de valores físicos.

Sempre que uma quantidade pode variar por fatores de milhões ou bilhões e, ainda assim, continuar significativa em toda a extensão, nós buscamos a ajuda dos logaritmos. O *logaritmo* mede o *expoente* que nos retorna um número. A *base* do logaritmo é o número que estamos elevando à potência, e a *potência* é o logaritmo do número. (HELLER, 2013, p. 44, t.n.)

A vantagem do uso da escala dos *decibéis* é que ela permite a visualização de altos picos de amplitude justapostos aos valores baixos, o que seria dificultado no caso de se utilizar uma representação linear. Assim, dada uma amplitude de referência  $a_{ref}$ , e uma amplitude a ser mensurada  $a$ , define-se a relação entre estes dois valores pela fórmula:

$$dB = 10 \cdot \log_{10} \left( \frac{a}{a_{ref}} \right) \quad (2.1)$$

## Potência sonora e Nível de potência sonora

A *potência acústica* ou *potência sonora* é definida como a taxa de energia acústica emitida por uma fonte sonora em um determinado período de tempo. Trata-se de uma medida intrinsecamente relacionada à capacidade de irradiação de energia acústica de uma fonte sonora e, portanto, não depende do contexto ou da distância entre a fonte e o ouvinte (HENRIQUE, 2002, p. 245; HOWARD; ANGUS, 2017, p. 22). O *nível de potência sonora*, por outro lado, mede a razão entre a *potência sonora* a ser medida e uma *potência sonora* de referência.

De acordo com o sistema internacional, a *potência sonora* é expressa em *watts* (W) e, por sua vez, o *nível de potência sonora* é medido em *decibéis*. O *nível de potência sonora* é designado pelos acrônimos SWL ou PWL e é expresso pela fórmula:

$$dB_{SWL} = 10 \log \left( \frac{W}{W_{ref}} \right) \quad (2.2)$$

onde  $W$  é a potência da fonte sonora e  $W_{ref}$ , a potência de referência. Essa potência de referência é dada pelo valor de  $10^{-12} W$ , que é considerado o nível mais baixo que pode ser percebido pelo sistema auditivo humano e, portanto, equivalente a  $0 \text{ dB}_{SWL}$ .

### Intensidade sonora e Nível de intensidade sonora

*Intensidade sonora* é a energia por unidade de tempo que flui através de uma superfície. Está relacionada com a energia das oscilações que são provocadas no ouvido do observador (LOY, D. G., 2006, p. 118). A passagem de uma onda sonora através de um meio fluido gera flutuações de pressão que se manifestam a partir de alterações da pressão e da velocidade das partículas. Assim, a *intensidade sonora* pode ser definida como o produto da pressão pela velocidade, que equivale à potência recebida por unidade de área (HENRIQUE, 2002, p. 247).

$$I = \frac{F}{a} \cdot \frac{d}{t} = \frac{E}{a \cdot t} = \frac{P}{a} \quad (2.3)$$

logo,

$$I = \frac{W}{a} \quad (2.4)$$

Portanto, se a potência é medida em *watt*, e a área em  $m^2$ , então a intensidade sonora é expressa em  $W/m^2$

É através da *intensidade* que percebemos sons fortes e sons fracos, ela está relacionada à amplitude sonora, à pressão efetiva e à energia transportada. Esta grandeza é popularmente conhecida como volume e pode variar de acordo com a distância devido a natureza da emissão sonora intrínseca de cada fonte. Já o *nível de intensidade sonora* é dado em *decibel* por:

$$dB_{SIL} = 10 \log \left( \frac{I}{I_{ref}} \right) \quad (2.5)$$

sendo  $I_{ref}$  a intensidade sonora de referência correspondente ao limiar da audição ( $10^{-12} W/m^2$ ). O *nível de intensidade sonora* mede a razão entre um valor de *intensidade sonora* a ser medido e o valor de *intensidade sonora* de referência.

Em uma onda esférica ou em uma onda plana — que se propaga em um tubo —, o nível de intensidade e pressão sonora são aproximadamente iguais. Entre outros fatores, a geometria da fonte sonora influencia na relação entre os valores dos níveis de intensidade e pressão sonora (HENRIQUE, 2002, p. 248).

### **Pressão sonora e Nível de pressão sonora**

*Pressão sonora* é a quantidade de força de um som dividida pela área sobre a qual essa força é aplicada, seu valor é medido em *newtons/m<sup>2</sup>*, e de acordo com o sistema internacional sua unidade é dada em *pascal* (Pa), portanto,  $1 Pa = 1 N/m^2$ .

$$p = \frac{F}{a} \quad (2.6)$$

A faixa de *pressão sonora* audível para a frequência de 1000 Hz corresponde a uma amplitude de pressão sonora de  $2 \cdot 10^{-5} N/m^2$  para o limiar da audição (pressão sonora mínima audível). O sistema auditivo humano tem a capacidade de detectar flutuações de pressão na ordem de  $10^{-9}$  do valor da pressão atmosférica. Expressando em valores de *pressão sonora*, o limiar da dor é um milhão de vezes maior do que o limiar da audibilidade (HENRIQUE, 2002, p. 248).

A gama de variações dos valores de pressão sonora entre o som imperceptível e o mais intenso é da ordem de  $5 \cdot 10^6 \mu Pa$ , torna-se, portanto, necessário o uso de uma escala condensada que permita a fácil leitura dos valores (TEMPELAARS, 2016, p. 22). Por essa razão utiliza-se a escala logarítmica para calcular o nível de pressão sonora (*sound pressure level* - SPL) expresso em *decibel*.

$$dB_{SPL} = 20 \log \left( \frac{P}{P_{ref}} \right) \quad (2.7)$$

sendo o valor da pressão de referência  $P_{ref} = 2 \cdot 10^{-5} Pa$ .

### **Loudness e phon**

*Loudness* corresponde à impressão subjetiva da magnitude de um som, trata-se de um atributo da percepção auditiva em termos dos quais os sons podem ser ordenados em uma escala que varia do suave ao mais intenso (ROSSING, 2007, p. 468). Apesar de haver uma correlação com a intensidade, o *loudness* é limitado ao alcance da capacidade de percepção

do ouvido humano, enquanto a intensidade é ilimitada (LOY, D. G., 2006, p. 166). Embora nossa percepção de *loudness* seja aproximadamente proporcional à intensidade, ela também é influenciada pela faixa de frequência e pela presença ou ausência de outras frequências. Assim, três dimensões se relacionam: *loudness*, frequência, e intensidade.

Se pretendemos relacionar esses três atributos, é preciso uma medida capaz de indicar a igualdade de valores de *loudness* em frequências diferentes, para isso utiliza-se o *phon*. *Phon* é a unidade que mede o nível de *loudness* de um som. A escala *phon* identifica *loudness* iguais em todas as frequências e intensidades perceptíveis. Essa escala consiste em um conjunto de curvas de *loudness* iguais que estabelecem uma relação entre a intensidade em uma determinada região de frequência, com a intensidade necessária para alcançar o mesmo valor de *loudness* em outra região de frequência.

O valor de *phon* corresponde ao mesmo valor medido pelo nível de intensidade sonora para uma frequência de 1000 Hz, essa mesma frequência no limiar da audição ( $10^{-12} \text{ W/m}^2$ ), equivale a 0 *phon*. Logo, um nível de intensidade sonora de 10 dB é igual a 10 *phons* para uma frequência de referência igual a 1000 Hz. Comparando determinadas frequências e intensidades a um conjunto de intensidades a partir de uma frequência de 1000 Hz, define-se a escala *phon*.

Em suma, considerando uma frequência  $f$ , busca-se descobrir qual valor de intensidade  $I$  é preciso para que  $f$  tenha o mesmo valor de *loudness*  $L$  da frequência de referência (1000 Hz). Se  $\varepsilon$  é o critério do *loudness*, então tem-se:

$$I = \varepsilon(L, f) \quad (2.8)$$

Portanto, se uma frequência com determinada intensidade causar a sensação de ser tão alta quanto a frequência de referência com um nível de intensidade sonora igual à 45 dB, então o nível de *loudness* dessa frequência será de 45 *phons* (ROSSING, 2007, p. 468). A figura 2.1 mostra as curvas de *loudness* igual.

### 2.1.2 Escalas e medidas relacionadas à frequência

No contexto dos movimentos periódicos, definimos as frequências pelo número de ciclos realizados por unidade de tempo, entretanto, do ponto de vista psicoacústico também podemos correlacioná-las à percepção de altura. A altura é um dos aspectos ou atributos

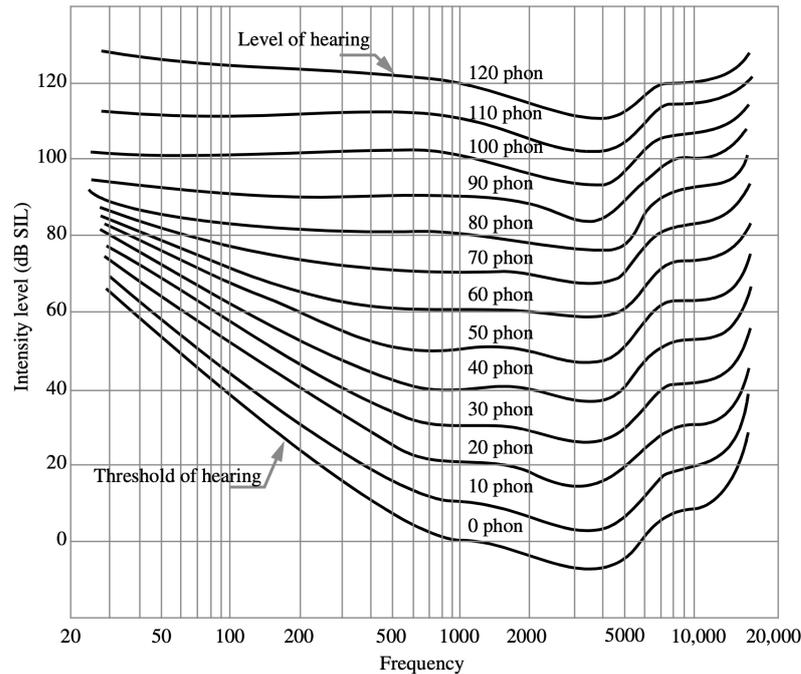


Figura 2.1: Curva de Fletcher, Munson (FLETCHER; MUNSON, 1933) para níveis de *loudness* de 0 à 120 *phons* (LOY, D. G., 2006, p. 168)

psicoacústicos de um som. É uma das dimensões que somos capazes de distinguir em termos de sensações auditivas (STEVENS; VOLKMANN, 1940, p. 329), e que estão diretamente ligadas às características timbrísticas, melódica e harmônica. Dessa forma, as frequências podem ser representadas em termos de espectro e de altura, e para cada uma dessas categorias existem diversos meios de expressar seus valores.

Desde o comportamento da fundamental de um espectro harmônico até a mensuração de suas parciais, os processos de cálculo voltados à medição das frequências nos auxiliam na organização e manipulação de seus valores de forma a permitirem que se estabeleçam relações entre outros aspectos do som. Descreveremos a seguir alguns métodos voltados à quantificação de unidades de frequências e à criação de sistemas de organização e ordenação.

### **Hertz**

*Hertz* (Hz) é a unidade de medida que expressa a frequência em termos de ciclos por segundo. Dada a natureza dos movimentos periódicos das ondas, há a necessidade de definir uma grandeza que quantifique no tempo sua periodicidade. Podemos caracterizar esses movimentos medindo o número de vezes que a trajetória se repete dentro de um intervalo de tempo. Portanto, a frequência pode ser definida pelo número de ciclos realizados por unidade

de tempo dada em segundos (HENRIQUE, 2002, p. 47). A unidade de um ciclo por segundo é equivalente a um *hertz*. Dessa forma, se o período é o intervalo de tempo para que ocorra um ciclo, então a frequência em *hertz* é dada por:

$$f = \frac{1}{T} \text{ e } T = \frac{1}{f} \quad (2.9)$$

Conclui-se, portanto, que ao passo em que a frequência aumenta o período diminui, e vice-versa. Devido a sua relação com a percepção de altura, quanto maior for seu número de ciclos por segundo, mais agudo será o som resultante. Assim, a percepção de altura está relacionada à frequência fundamental do som. Embora haja casos em que a fundamental não é audível, ela pode ser inferida devido ao fenômeno psicoacústico causado pelas relações harmônicas entre as parciais de um espectro (ROSSING, 2007, p. 541).

### Afinação igual

Na música ocidental, organizamos as alturas dividindo a oitava em doze partes iguais. Esse sistema de afinação é chamado de afinação igual, visto que todos os intervalos são determinados a partir de um intervalo uniforme de semitom. Considerando  $f_x = f_R \cdot 2^x$ , sendo  $x$  um intervalo que inclui todos os números reais, e  $f_R$  a frequência de referência, é possível calcular a frequência  $f_x$  correspondente à afinação igual. Para frequências de referência  $f_R$ , obtemos a frequência  $f_K$  de qualquer intervalo referente à afinação igual na primeira oitava, a partir de:

$$f_K = f_R \cdot 2^{K/12} \quad (2.10)$$

Logo, considerando que a frequência equivalente a um semitom acima de  $f_R = 440\text{Hz}$  é  $f_1 = f_R \cdot 2^{1/12} \cong 466,16\text{Hz}$ . Então, o tamanho do intervalo de semitom correspondente à afinação igual pode ser calculado por:

$$\frac{2^{1/12}}{1} = \frac{\sqrt[12]{2}}{1} \cong 1,05946 \quad (2.11)$$

Dessa forma, a escala cromática pode ser obtida multiplicando a frequência anterior pela constante 1,05946. Para calcular a frequência de intervalos maiores que uma oitava a partir do sistema de afinação igual, partimos de:

$$f_{K,v} = f_R \cdot 2^{v+K/12} \quad (2.12)$$

onde  $f_{K,v}$  é a frequência do intervalo  $K$  equivalente à afinação igual na oitava  $v$ , sendo  $K$  um número inteiro entre 0 e 11, e  $v$  qualquer número inteiro. É importante destacar que o valor de  $v$  corresponde à oitava em relação à  $f_R$ , ou seja, se  $v$  é menor que 0, a frequência calculada estará localizada oitava abaixo da frequência de referência  $f_R$ , se  $v$  for maior que 0, a frequência estará oitava acima de  $f_R$  (LOY, D. G., 2006, p. 40). Tomando 261, 6256 Hz (equivalente ao Dó central) como frequência de referência, e desejando obter a frequência correspondente ao intervalo de nona, tem-se:

$$f_{K,v} = 261,6256 \text{ Hz} \cdot 2^{1+2/12} = 587,3295 \text{ Hz} \quad (2.13)$$

### Escala *cents*

A dificuldade apresentada na comparação de intervalos medidos em *hertz*, e o interesse no estudo de intervalos menores que o semitom levaram à criação da escala *cents*. Esse sistema é capaz de comparar o tamanho de intervalos similares a partir de diferentes sistemas de afinação. Enquanto o temperamento igual divide a oitava em 12 partes, a escala em *cents* — como definida por John Ellis (1885) — divide em 1200 resultando numa resolução de afinação cem vezes maior do que a do temperamento igual (LOY, D. G., 2006, p. 45). É utilizado em situações em que pretende-se quantificar intervalos ou medir pequenas diferenças entre afinações (HENRIQUE, 2002, p. 928).

Seguindo a mesma lógica do cálculo do semitom no temperamento igual mostrado na equação 2.11, pode-se definir o intervalo de 1 *cents* a partir da equação:

$$2^{\frac{1}{1200}} = 1.0005778 \quad (2.14)$$

Dessa forma, conseqüentemente 100 *cents* equivale a 1 semitom no temperamento igual. Portanto, se  $r$  é um intervalo, seu tamanho em *cents* se dá por:

$$c = 1200 \cdot \left( \frac{\log r}{\log 2} \right) \quad (2.15)$$

Considerando  $r = 2/1$  correspondente à razão do intervalo de oitava, calcula-se seu tamanho em *cents* a partir de:

$$c = 1200 \cdot \left( \frac{\log 2}{\log 2} \right) = 1200 \quad (2.16)$$

### Escala *Mel*

O comportamento do sistema auditivo humano pode ser modelado a partir de uma série de filtros de banda crítica, como aquele baseado na escala de frequência *Mel*. A escala *Mel* Volkman, Stevens e Newman (1937) é uma escala criada para representar alturas a partir de aspectos psicoacústicos. A maneira como o sistema auditivo humano percebe as alturas não mantém uma relação direta com as frequências expressadas em *hertz*, com a escala *Mel* podemos converter as bandas de frequência e aproximá-las do modo como nosso ouvido as percebe. Essa escala é linear em frequências abaixo de 1000 Hz e logarítmica em frequências acima desse valor (PEETERS, 2004, p. 4). Assim, os valores da escala de frequência *Mel* correspondem aproximadamente à distribuição real de frequências da relação logarítmica e, portanto, mais próximo do funcionamento do ouvido humano (XU; ZHANG, X.; JIA, 2012). Os valores de frequência expressos em *hertz* convertidos para *Mel*, podem ser obtidos por:

$$m = 2595 \log \left( 1 + \frac{f}{700} \right) \quad (2.17)$$

onde  $m$  é a frequência expressa em *Mel*, e  $f$  a frequência em *hertz*.

### Escala *Bark*

Apesar da escala *Mel* ser popular por conta do seu amplo uso em pesquisas desenvolvidas na área de reconhecimento de fala, a escala *Bark* apresenta uma melhor aproximação ao comportamento do sistema auditivo humano (PEETERS, 2004, p. 5). Assim como a escala *Mel*, a escala *Bark* também é uma escala psicoacústica, que apresenta comportamento logarítmico acima de 500 Hz, e abaixo se torna cada vez mais linear. Proposta por Zwicker (1961), a escala *Bark* varia seu alcance de 1 à 24 bandas, correspondendo às 24 bandas críticas da escuta. A conversão das frequências em *hertz* para *Bark* pode ser realizada por:

$$B = 13 \operatorname{atan}(0,00076f) + 3.5 \operatorname{atan} \left( \frac{f}{7500} \right)^2 \quad (2.18)$$

Onde  $B$  é a frequência dada em *Bark*, e  $f$  frequência em *hertz*.

## 2.2 Envoltória de energia

A *envoltória de energia* é uma representação simples que está associada à intensidade do sinal e busca mostrar como sua energia evolui no tempo. A partir de seu processo de cálculo podemos descrever comportamentos como aumento e decréscimo de energia, picos, momentos de estabilidade, e dissipação. A figura 2.2 mostra a representação da forma de onda de um sinal periódico (em azul), e sua *envoltória de energia* (em laranja).

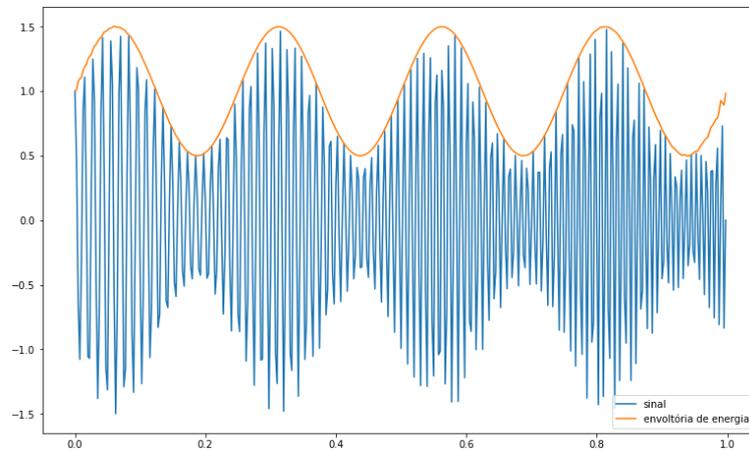


Figura 2.2: Sinal periódico e sua envoltória de energia

Existem vários processos para se extrair a energia de um sinal sonoro, no entanto, uma implementação eficiente está ligada ao cálculo do RMS (*root mean square*) (PEETERS, 2004, p. 3). Por definição a energia é igual ao quadrado dos valores da função do sinal, ou seja, a energia é sempre um valor positivo. O nível de energia de um sinal é caracterizado pela raiz quadrada da energia média, denominada média quadrática da raiz (RMS) (TEMPELAARS, 2016, p. 17). O RMS é o cálculo da média quadrática dos valores de amplitude em uma janela de amostras. Os valores extraídos a partir desse processo podem descrever o contorno da *envoltória de energia* de um sinal e seu cálculo é definido pela equação:

$$\mathbf{RMS} = \sqrt{\frac{1}{N} \sum_{i=0}^N [x]^2} \quad (2.19)$$

sendo  $x$  os valores de amplitude na  $n$ ésima amostra da  $i$ ésima janela do sinal, e  $N$  é o número de amostras da  $i$ ésima janela.

## 2.3 Filtros

Em geral, podemos definir filtro como qualquer sistema capaz de selecionar elementos com características específicas a partir de um sinal sonoro, e isolá-los a fim de que se possa operar diferentes tipos de manipulação. De fato, o filtro irá realizar uma seleção de acordo com as faixas de frequências que se deseja rejeitar, reter ou enfatizar. Dessa forma, irá atuar destacando ou atenuando a amplitude das frequências ou faixas de frequência selecionadas.

A implementação e a utilização de filtros é um assunto vasto devido ao fato de que a sua ampla gama de aplicações sugere uma grande variedade de processos e métodos específicos de implementação. Em certas situações um filtro deve obedecer exatamente uma resposta de frequência predefinida, em outras, a resposta de fase ou a otimização do tempo de processamento são aspectos fundamentais, já em outras, há a necessidade de manipular seus parâmetros rapidamente ao longo tempo (PUCKETTE, 2007, p. 223). Enfim, para cada tipo de aplicação há um tipo de filtro adequado. No escopo deste trabalho, os filtros são ferramentas importantes que auxiliam nos métodos de análise e processamento dos sinais, na recuperação de informações musicais/sonoras, na simulação do sistema auditivo humano, e na implementação de processos de escuta de máquina.

É possível classificar os filtros de acordo com a sua natureza, tipo de aplicação, e/ou método de implementação. Logo, partindo dessa premissa podemos separá-los em dois grupos: LTI (*Linear Time-Invariant*) – geralmente utilizado nas aplicações em tempo diferido, uma vez que, seu processo não gera componentes espectrais adicionais, e não submete o sinal a distorções –; (III SMITH, 2006)<sup>2</sup>; e os filtros variáveis no tempo (*Time-Variant*) – filtros que apresentam parâmetros que podem ser controlados pelo usuário, geralmente utilizados em aplicações em tempo real (WISHNICK, 2014, p. 1).

Os filtros podem ser implementados a partir de funções tanto no domínio do tempo como no domínio da frequência. Para especificá-los no domínio do tempo utiliza-se a resposta ao impulso gerada pelo processo de filtragem. De acordo com Tempelaars (2016, p. 198), a especificação no domínio do tempo não é padronizada, portanto, não existe um sistema de classificação para tais filtros, exeto pela distinção entre dois tipos de sistemas: FIR (*Finite Impulse Response*) e IIR (*Infinite Impulse Response*).

<sup>2</sup>[https://ccrma.stanford.edu/~jos/fp/Linear\\_Time\\_Invariant\\_Digital\\_Filters.html](https://ccrma.stanford.edu/~jos/fp/Linear_Time_Invariant_Digital_Filters.html)

Os filtros FIR são filtros cuja resposta ao impulso apresenta duração finita, pois se estabiliza em zero em um determinado intervalo de tempo. A estrutura dos filtros FIR pode ser vista na figura 2.3.

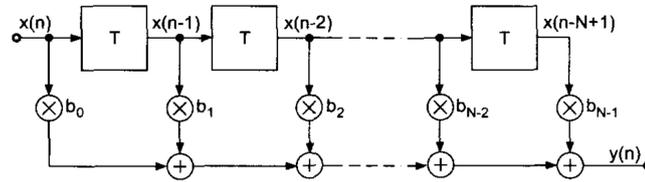


Figura 2.3: Estrutura de um filtro FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 46)

A relação entre a saída e entrada do filtro FIR representado na figura 2.3 é descrito pela equação:

$$y(n) = \sum_{i=0}^{N-1} b_i \cdot x(n-i) = b_0 x(n) + b_1 x(n-1) + \dots + b_{N-1} x(n-N+1) \quad (2.20)$$

onde  $x(n)$  é o sinal de entrada,  $y(n)$  o sinal de saída,  $N$  é a ordem do filtro, e  $b_i$  é o valor da resposta ao impulso no  $i$ -ésimo instante para  $0 \leq i \leq N$ . Se o sinal de entrada é um impulso unitário  $\zeta(n)$ , onde tem-se 1 para  $n = 0$  e 0 para  $n \neq 0$  (ZÖLZER; AMATRIAIN, X., 2002, p. 46), obtemos a resposta ao impulso a partir de:

$$h(n) = \sum_{i=0}^{N-1} b_i \cdot \zeta(n-i) \quad (2.21)$$

Uma representação gráfica da resposta ao impulso de um filtro FIR é mostrada na figura 2.4.

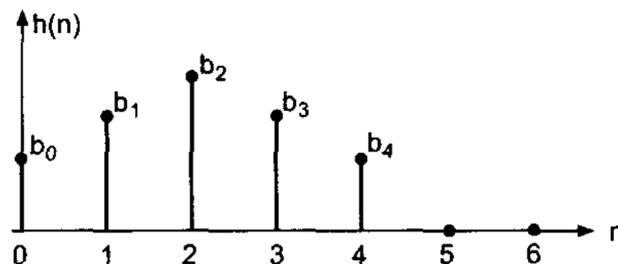


Figura 2.4: Resposta ao impulso de um filtro FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 46)

Os filtros IIR são sistemas recursivos que apresentam uma resposta ao impulso  $h(t)$  contínua e infinita, pois não se estabiliza em zero após um certo intervalo de tempo. O sinal de saída  $y(n)$  é realimentado por meio de elementos de atraso, e uma soma ponderada dessas saídas atrasadas é adicionada ao sinal de entrada  $x(n)$ . A maioria dos filtros lineares variantes

no tempo são implementados a partir de um sistema IIR. Podemos descrever a estrutura do filtro IIR (representando na figura 2.4) pela equação:

$$y(n) = x(n) - a_1 y(n-1) - a_2 y(n-2) \quad (2.22)$$

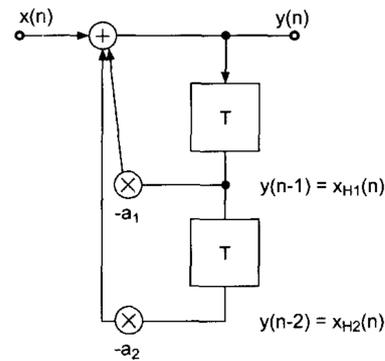


Figura 2.5: Estrutura de um filtro IIR (ZÖLZER; AMATRIAIN, X., 2002, p. 23)

Uma representação gráfica da resposta ao impulso de um filtro IIR é mostrada na figura 2.6.

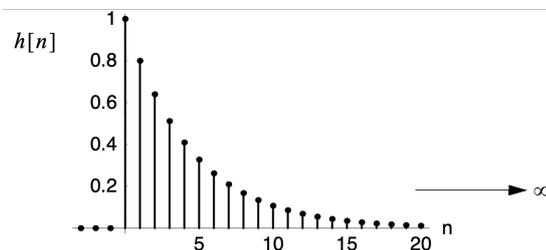


Figura 2.6: Resposta ao impulso de um filtro IIR

Para especificar os filtros no domínio da frequência, partimos das características da amplitude e/ou fase, ou seja, o efeito causado no espectro em termos de atraso de fase e atenuação da amplitude das frequências. Ainda de acordo com Tempelaars (2016, p. 196), há três formas de representar os filtros a partir das características e comportamentos da fase: (1) a partir da variação de fase  $\Delta\phi(f)$  dado por:

$$\Delta\phi(f) = \phi_{entrada}(f) - \phi_{sa-da}(f) \quad (2.23)$$

(2) a partir do atraso de fase, ou seja, o atraso de tempo de cada componente espectral. Pode ser derivado a partir da equação 2.23 como a mudança de fase em relação ao sinal de entrada, assim  $\phi_{sa-da} - \phi_{entrada} = -\Delta(f)$ :

$$\sin(\pi f t - \Delta\phi(f)) = \sin\left(\pi f t \frac{\Delta\phi(f)}{2\pi f}\right) = \sin 2\pi f \left(t - \frac{\Delta\phi(f)}{2\pi f}\right) = \sin 2\pi f (t - \tau_f) \quad (2.24)$$

logo:

$$\tau_f = \frac{\Delta\phi(f)}{2\pi f} \quad (2.25)$$

(3) a partir do atraso de grupo, medida baseada no princípio da fase estacionária. Com esse princípio, o centro de gravidade da função de um sinal pode ser derivado do espectro de fase:

$$t_p(-1/2\pi) \cdot d\phi df \quad (2.26)$$

O tempo de atraso  $\tau_{gr}$  desse centro de gravidade é calculado por:

$$\tau_{gr} = t_{p,sa-da} - t_{p,entrada} = \frac{1}{2\pi i} \frac{d(\phi_{sa-da} - \phi_{entrada})}{df} = \frac{1}{2\pi} \frac{d}{df} \quad (2.27)$$

Já as características da amplitude são determinadas a partir de dois parâmetros elementares: *frequência de corte*, e *curva de atenuação*. A *frequência de corte* é aquela frequência em que os valores de amplitude começam a sofrer atenuação. Comumente, a relação entre a *curva de atenuação* (expressa em dB) e a distância de frequência (intervalo) é linear, isto é, a *curva de atenuação* pode ser expressa em dB/oitava (TEMPELAARS, 2016, p. 195).

Apesar de ser possível implementar filtros no domínio do tempo, geralmente são representados como operações no domínio da frequência. Dessa forma, considerando o comportamento espectral causado pela atenuação da amplitude das frequências selecionadas, podemos definir quatro tipos básicos de filtros de acordo com o caráter da atenuação: (1) filtro passa-baixa (LP), que atenua frequências acima da *frequência de corte*, mantendo as frequências baixas; (2) filtro passa-alta (HP), que atenua frequências abaixo da *frequência de corte*, mantendo as frequências altas; (3) filtro passa-banda (BP), que atenua as frequências que estão fora do intervalo entre dois valores de *frequência de corte*; e (4) filtro rejeita-banda (BR), que atenua as frequências que estão dentro do intervalo entre dois valores de *frequência de corte*. A figura 2.7 mostra o efeito de atenuação das amplitudes dos quatro tipos de filtros.

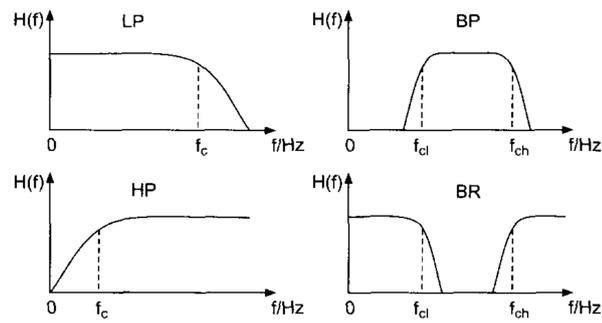


Figura 2.7: Representação da atenuação de amplitude nos quatro tipos de filtros (ZÖLZER; AMATRIAIN, X., 2002, p. 32)

### 2.3.1 Filtros equalizadores

Em alguns casos — como nos processos de equalização —, o objetivo não é manter certas frequências enquanto elimina outras completamente, mas fazer ajustes controláveis de maneira que possa aumentar ou atenuar a amplitude de determinadas faixas de frequência. Esses tipos de filtros são chamados de *equalizadores* (PUCKETTE, 2007, p. 227). Para esse tipo de processo é utilizado o filtro *shelving*. Os filtros *shelving* atuam no aumento e/ou na atenuação de frequências em regiões graves ou agudas, são utilizados em pares (um filtro para as regiões extremo-graves, e outro para regiões extremo-agudas). Seus parâmetros de controles são a *frequência de corte* e nível de ganho aplicado nas frequências selecionadas (ZÖLZER; AMATRIAIN, X., 2002, p. 50). É comumente utilizado como botões de controle grave e agudo em equipamentos como *home theaters* e amplificadores. Um filtro *shelving* passa-alta terá a resposta frequência dada por:

$$|H(f)|^2 = \frac{1 + (f/f_z)^2}{1 + (f/f_p)^2} \quad (2.28)$$

onde  $f_p$  e  $f_z$  são denominados polo e frequência zero, respectivamente. Para diminuir a presença de frequência agudas, o valor de  $f_z$  aumenta e  $f_p$  diminui, dessa forma, atenuando as frequências acima de  $f_p$ . Para destacar frequências agudas, o valor de  $f_p$  aumenta e  $f_z$  diminui, assim as frequências maiores que  $f_z$  são reforçadas. Se  $f_z = f_p$ , logo  $|H(f)|^2 = 1$ , ou seja, o filtro não tem efeito. Assim como o filtro *shelving* passa-alta, a resposta frequência do filtro *shelving* passa-baixa é dada por:

$$|H(f)|^2 = (f_p/f_z)^2 \frac{1 + (f/f_z)^2}{1 + (f/f_p)^2} \quad (2.29)$$

Outro filtro utilizado nos processos de equalização é o filtro *peaking*. Esse filtro é capaz de aumentar ou atenuar determinadas bandas de frequência ao redor de uma frequência central. O filtro *peaking* é parametrizado em termos de *pico de ganho* ( $G_{dB}$ ), frequência central, dada por  $f_c = \omega \cdot f_s / (2\pi)$ , e um parâmetro  $Q$ , que é inversamente proporcional à largura de banda. A frequência central, a largura de banda — que juntas determinam a faixa de frequência afetada — e os valores de ganho são ajustáveis separadamente (PUCKETTE, 2007, p. 227). O filtro *peaking* pode ser definido por:

$$H_{pk}(z) = C_{pk} \left( \frac{1 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \right) \quad (2.30)$$

onde os parâmetros são dados por:

$$C_{pk}(z) = \left( \frac{1 + k_q \mu}{1 + k_q} \right) \quad (2.31)$$

$$b_1 = \left( \frac{-2 \cos(\omega_c)}{1 + k_q \mu} \right) \quad (2.32)$$

$$b_2 = \left( \frac{1 - k_q \mu}{1 + k_q} \right) \quad (2.33)$$

$$a_1 = \left( \frac{-2 \cos(\omega_c)}{1 + k_q} \right) \quad (2.34)$$

$$a_2 = \left( \frac{1 - k_q}{1 + k_q} \right) \quad (2.35)$$

$$k_q = \left( \frac{4}{1 + \mu} \right) \tan \left( \frac{\omega_c}{2Q} \right), \quad \mu = 10^{\frac{G_{dB}}{20}} \quad (2.36)$$

### 2.3.2 Filtros *comb*

O filtro *comb* ou filtro em pente consiste em um processo de filtragem do sinal pela resposta frequência em formato de pente. Isso ocorre quando o sinal é somado a uma cópia sua deslocada no tempo, assim, gerando interferências construtivas e destrutivas. As frequências que serão filtradas são determinadas pela quantidade de deslocamento aplicada ao sinal de

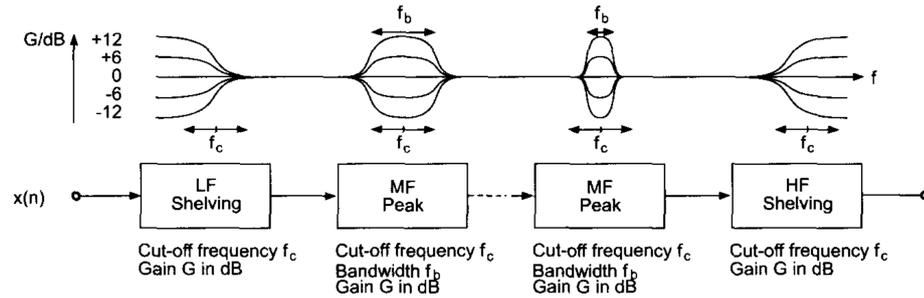


Figura 2.8: Representação da resposta frequência dos filtros *shelving* e *peaking* (ZÖLZER; AMATRIAIN, X., 2002, p. 32)

entrada (BURG; ROMNEY; SCHWARTZ, 2016)<sup>3</sup>. Considerando um atraso de  $t$  segundos entre duas cópias idênticas de um sinal, as frequências  $f_i$  que serão filtradas podem ser determinadas por:

$$f_i = \frac{2i + 1}{2t} \text{ para todos inteiros } i \geq 0 \quad (2.37)$$

Assim como todo tipo de filtro, o *comb* pode ser implementado no domínio do tempo (através do uso de linhas de atraso), no domínio do tempo-frequência (*vocoder* de fase) e no domínio da frequência. Quando implementado como um filtro do tipo *Time-Variant*, o filtro *comb* é chamado de *phasing*. Nesse processo a quantidade de atraso entre os dois sinais varia no tempo. Este atraso variável — que pode ser controlado por um LFO (oscilador de baixa frequência) — resulta em diferentes frequências e seus harmônicos são filtrados em diferentes momentos no tempo.

Segundo (ZÖLZER; AMATRIAIN, X., 2002, p. 63), existem três maneiras de se implementar o filtro *comb*: a partir de um sistema FIR, a partir de um sistema IIR, e a partir da combinação dos dois sistemas, chamado de *universal*. A implementação baseada no sistema FIR simula apenas uma linha de atraso, logo, o sinal de entrada é atrasado por uma determinada duração de tempo e somado à sua versão não processada. O efeito dessa filtragem será percebido apenas quando o sinal processado for adicionado ao sinal de entrada. Essa implementação apresenta dois parâmetros de ajuste: a quantidade de atraso de tempo  $\tau$ , e a amplitude relativa do sinal atrasado em relação ao sinal de referência. O filtro *comb* baseado no sistema FIR pode ser definido por:

$$y(n) = x(n) + g \cdot x(n - m) \quad (2.38)$$

<sup>3</sup><http://digitalsoundandmusic.com/7-3-7-comb-filters-and-delays/>

onde  $y(n)$  é o sinal filtrado,  $x(n)$  o sinal de entrada,  $m$  é a quantidade de atraso, e  $g \leq 1$ . Para valores positivos de  $g$ , o filtro amplifica todas as frequências que são múltiplos de  $1/\tau$  e atenua todas as frequências intermediárias. Para valores negativos de  $g$  o filtro atenua as frequências que são múltiplos de  $1/\tau$  e amplifica aquelas intermediárias (ver figura 2.9)

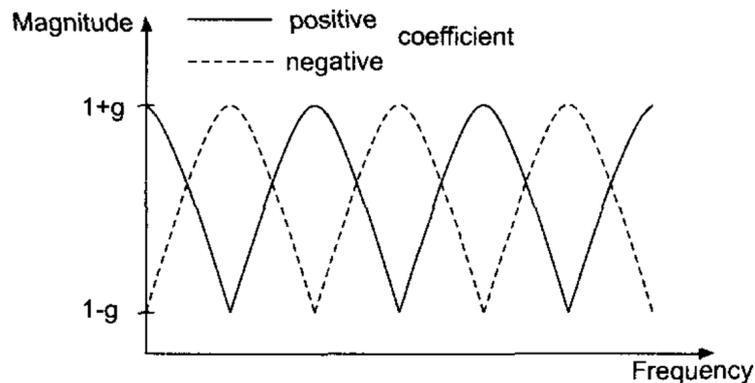


Figura 2.9: Resposta frequência do filtro *comb* baseado no sistema FIR (ZÖLZER; AMATRIAIN, X., 2002, p. 64)

Na implementação do filtro *comb* baseado em um sistema IIR, o sinal de entrada circula em uma linha de atraso que retorna à entrada do filtro, assim, produzindo uma série infinita de respostas  $y(n)$  para uma entrada  $x(n)$ . Essa implementação pode ser definida pela equação:

$$y(n) = x(n) + g \cdot y(n - m) \quad (2.39)$$

onde  $y(n)$  é o sinal filtrado,  $x(n)$  o sinal de entrada,  $m$  é a quantidade de atraso, e  $g \leq 1$ . Cada vez que o sinal passa pela linha de atraso é atenuado por  $g$  — devido ao seu valor  $\leq 1$ . Em virtude deste processo de *feedback*, o sinal anterior continua a ter efeito nos sinais posteriores, porém com intensidade cada vez que menor (BURG; ROMNEY; SCHWARTZ, 2016)<sup>4</sup>. A figura 2.10 mostra a resposta frequência do filtro *comb* do tipo IIR.

### 2.3.3 Vocoder de fase

*Vocoder* de fase é um método voltado ao processamento de sinais amplamente utilizado em processos de análise e ressíntese. Foi originalmente desenvolvido na área de telecomunicações por Flanagan e Golden (1966) equanto método de codificação voltado à redução da quantidade de dados envolvidos na transmissão de sinais sonoros provenientes da voz. A palavra *vocoder* é uma contração do termo "*voice coder*" [codificador de voz].

<sup>4</sup><http://digitalsoundandmusic.com/7-3-7-comb-filters-and-delays/>

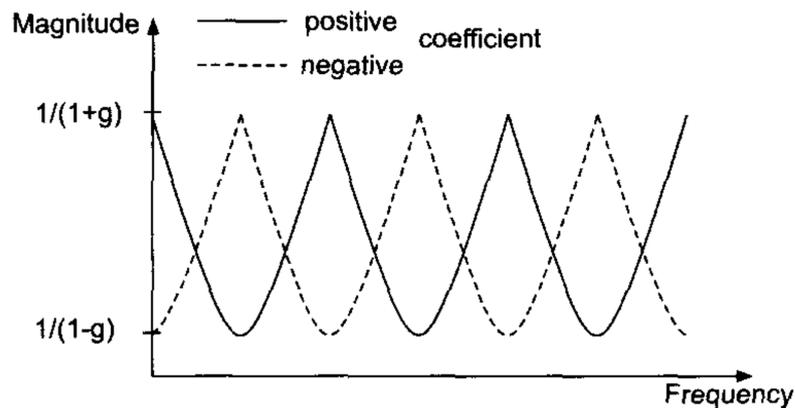


Figura 2.10: Resposta frequência do filtro *comb* baseado no sistema IIR (ZÖLZER; AMATRIAIN, X., 2002, p. 65)

No contexto da *live-electronics*, o *vocoder* de fase se mostra uma técnica com grande potencial criativo, podendo atuar na manipulação de sons no domínio do tempo e da frequência de maneira independente, dessa forma, fornecendo ao compositor um controle arbitrário e individual em termos de frequência e duração do som. Quando aplicado aos processos de *Music Information Retrieval* e *Machine listening*, o *vocoder* de fase atua no pré-cálculo dos descritores de áudio preparando o sinal no domínio do tempo e da frequência para que seja possível a extração das informações musicais e sonoras do sinal.

A premissa do *vocoder* de fase parte da idéia de operar um processo de análise de um sinal de entrada, e resintetizá-lo de maneira a produzir um sinal de saída idêntico – ou uma versão modificada, caso houver algum tipo de manipulação nos parâmetros. Para isso, o sinal a ser analisado é modelado como uma soma de senoides com amplitudes e frequências variáveis no tempo. Logo, a etapa de análise consiste em determinar os valores desses parâmetros, enquanto a etapa de resíntese trata de reconstruir o sinal de entrada aplicando as modificações realizadas.

Segundo Dolson (1986, p. 15), o *vocoder* de fase pode ser implementado de duas maneiras. O primeiro método consiste na aplicação de um conjunto de filtros passa-banda idênticos centralizados em torno de frequências igualmente espaçadas que variam de 0 Hz à  $fs/2$ , onde  $fs$  é a frequência de amostragem. A saída de cada filtro passa-banda retorna a magnitude e a frequência do sinal dentro de cada banda de frequência – vale ressaltar que as características da resposta frequência de cada filtro passa-banda devem ser constantes e idênticas ao longo do sinal, exceto em suas frequências centrais. Depois de manipular esses valores, o resultado é enviado a um conjunto de osciladores (um oscilador por banda) que

irão ressamplear o sinal e representá-lo novamente no domínio do tempo (GRÜNWARD, 2010, p. 8).

O segundo método parte da aplicação da transformada de *Fourier*. Primeiramente multiplica-se o sinal por uma função de janelamento (*Hamming*, *Hann*, *Blackman*, etc.), e então é aplicada a transformada *Fourier*, representando o sinal no domínio da frequência. Assim como na implementação através do banco de filtros, nessa etapa de pós-análise é possível operar manipulações no sinal antes de ressampleá-lo através da transformada inversa de *Fourier*. Para um determinado valor de amostra  $n$ , cada valor (bin de frequência) retornado pela operação da transformada de *Fourier*  $x(k, n)$  corresponde à magnitude e fase de um filtro passa-banda com frequência central  $kfs/N$ , assim como descrito no primeiro método. Neste processo a função de janelamento acaba apresentando o mesmo papel que a resposta ao impulso no método baseado em banco de filtros (e.g., tornar a resposta ao impulso mais curta aumenta o valor de sobreposição das janelas) (DOLSON, 1986, p. 21).

Esses dois métodos de implementação do *vocoder* de fase sugerem duas interpretações igualmente divergentes do processo de ressample. Na implementação baseada no banco de filtros, o processo de ressample consiste em um procedimento de síntese aditiva com controles de amplitude e frequência que variam no tempo para cada um dos osciladores. Já no método baseado na transformada de *Fourier*, a ressample é realizada através da adição de sucessivas transformadas inversas, sobrepondo-as no tempo de maneira que correspondam à sobreposição das janelas de análise.

## 2.4 Janelamento

Os sinais reais que lidamos no contexto dos processos de análise e extração de informações musicais e sonoras não apresentam caráter periódico e são finitos no tempo. Devido à não periodicidade é preciso segmentá-los em janelas e tratá-las como se cada uma fossem um período de um sinal periódico. Isso ajuda na visualização do espectro, limitando a análise a segmentos curtos (geralmente com menos de um segundo) que resultará em poucos pontos, assim, possibilitando localizar no tempo com maior precisão onde ocorre cada frequência. Esse processo se dá através do janelamento, método que consiste na seleção de uma pequena parte do sinal. A porção do sinal que corresponde ao intervalo selecionado é

multiplicado por 1, enquanto a porção que está fora é multiplicado por 0 (ROADS, 1996, p. 1100). Esse processo resulta numa função de janelamento retangular, que é definida por:

$$W(n) = 1, \quad 0 \leq n < N \quad (2.40)$$

onde  $N$  é o número de amostras contidas na janela e  $n$  é a amostra atual.

O método de janelamento é aplicado na transformada de *Fourier*, que é uma operação realizada no domínio do tempo, resultando em uma representação do sinal no domínio da frequência — como veremos na subseção 2.5 —, e também em processos no domínio do tempo. Vale ressaltar que o processo de janelamento implementado no domínio do tempo já é suficiente para extração dos descritores temporais instantâneos (MONTEIRO, 2012, p. 26).

Um dos principais fatores envolvidos no processo de janelamento é a duração ou tamanho da janela. Esse parâmetro é responsável por limitar a segmentação do sinal a um tamanho específico. O tamanho da janela está diretamente relacionado à resolução da análise no contexto dos processos no domínio da frequência. As frequências são espaçadas por múltiplos inteiros da razão entre a taxa de amostragem — corresponde ao número de amostras por segundo contidas em um sinal —, e o tamanho da janela. Por exemplo, se a taxa de amostragem equivale à 50000 Hz e o tamanho da janela corresponde à 1000 amostras, a análise das frequências estará espaçada por 50 Hz (ROADS, 1996, p. 1099).

No entanto, ao aplicar o processo de janelamento em um sinal lidamos com o surgimento de descontinuidades nas extremidades devido ao recorte realizado pela janela. A descontinuidade gera ruído no sinal, espalhando energia nas frequências adjacentes e afetando o resultado da análise. Esse fenômeno é chamado de vazamento (TEMPELAARS, 2016, p. 145). Para lidar com o problema do vazamento, cria-se uma função com o mesmo tamanho da janela de análise atenuando seu início e fim. Multiplicando o sinal por essa função, os efeitos da descontinuidade são minimizados (LOY, G., 2007, p. 140). Tais função são chamadas de função de apodização ou função de afunilamento. Existem vários tipos de funções de apodização com diferentes formatos, a seguir descreveremos as principais utilizadas no processamento de sinais.

### 2.4.1 Janela triangular

A função de afunilamento mais simples é a função triangular. Consiste em um aclave no início e um declive no fim resultando em uma janela com formato triangular. Sua função se dá por:

$$T(n) = 1 - \left| \frac{n - \frac{N}{2}}{\frac{L}{2}} \right| \quad (2.41)$$

onde  $N$  é o comprimento da janela e  $n$  é a amostra atual. A figura 2.11 mostra a janela triangular e seu efeito de atenuação quando aplicado a um espectro de frequências.

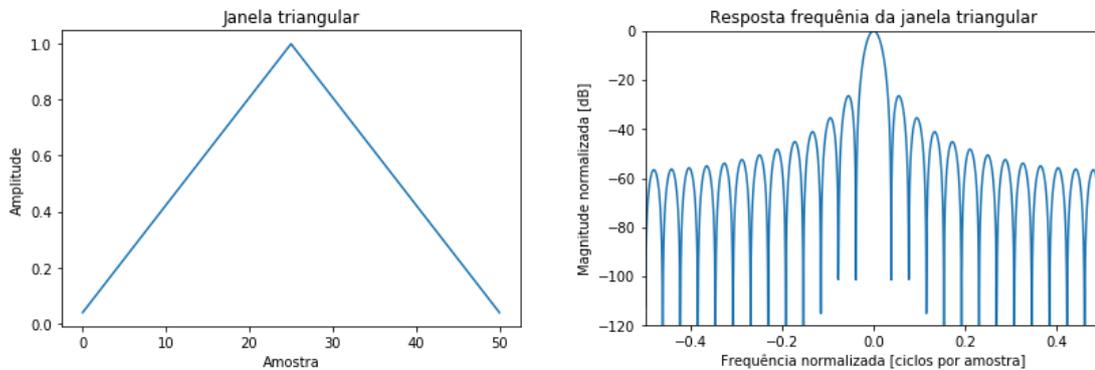


Figura 2.11: Representação da janela triangular e sua aplicação em um espectro de frequência

### 2.4.2 Janela Hann

A janela *Hann*, a qual recebe o nome do meteorologista austríaco Julius von Hann, é implementada através da função *hann*. Ao contrário da janela *hamming* (como veremos logo em seguida), suas extremidades começam em zero (TEMPELAARS, 2016, p. 147). A figura 2.12 mostra o formato da janela *Hann* e seu efeito quando aplicado em um espectro de frequência. Sua equação se dá por:

$$h(n) = a \left[ 1 - a \left( \frac{(2\pi n)}{N} \right) \right] \quad (2.42)$$

sendo  $a = 0.5$ .

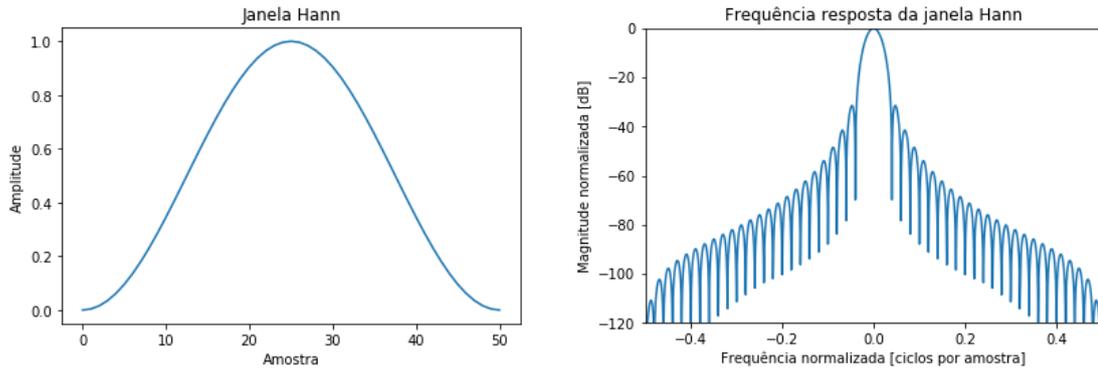


Figura 2.12: Representação da janela *hann* e de sua resposta frequência

### 2.4.3 Janela *hamming*

A equação da janela de *hamming* (nome do matemático Richard W. Hamming) é a mesma equação da janela *hann* (equação 2.42), com diferença no valor de  $a$ , que passa a valer 0.54. O valor de  $a$  ligeiramente elevado faz com que a janela *hamming* deixe entrar um pouco de energia nas extremidades, o que a difere da janela *Hann* (LOY, G., 2007, p. 144). A figura 2.13 apresenta o formato da janela e sua aplicação no espectro de frequências.

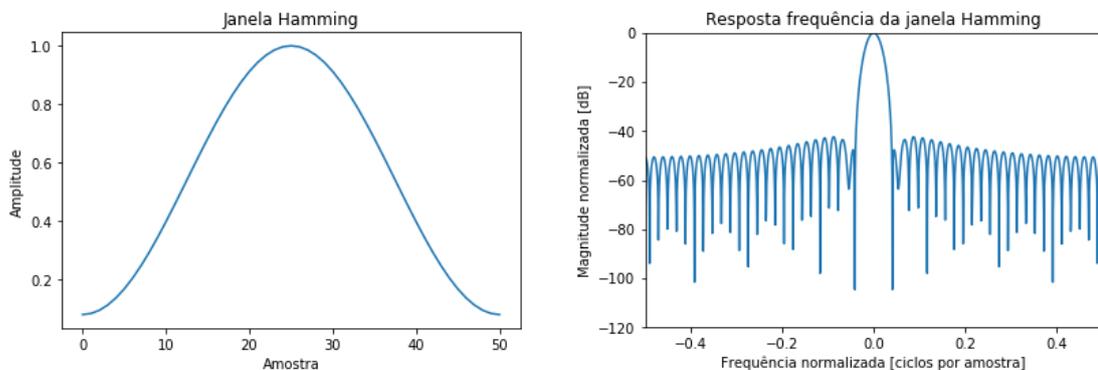


Figura 2.13: Representação da janela *hamming* e de sua resposta frequência

### 2.4.4 Janela *Blackman*

A janela de *Blackman* é uma função formada pelos três primeiros termos de uma soma de cossenos. Essa janela foi projetada para ter o mínimo de vazamento possível. É bastante utilizada em processamento de sinais e principalmente aplicada em sinais instáveis e ruidosos. A figura 2.14 mostra o formato da janela e seu efeito aplicado a um espectro de frequências. Sua função é definida por:

$$b(n) = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N}\right) + 0.08 \cos\left(\frac{4\pi n}{N}\right) \quad (2.43)$$

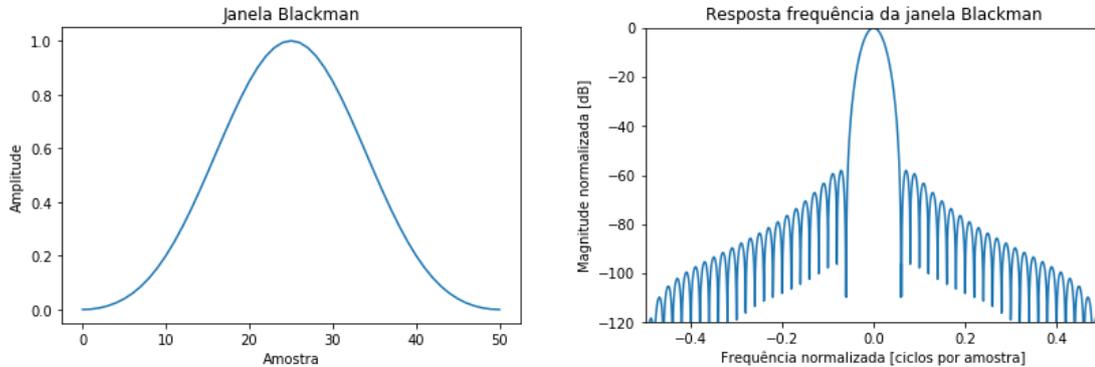


Figura 2.14: Representação da janela *blackman* e de sua resposta frequência

### 2.4.5 *Overlap e hop size*

Com o método de janelamento, há uma perda de informação no início e fim da janela. Para compensar essa perda devido atenuação das extremidades das janelas, é realizado um processo chamado *overlap*. Ao aplicar esta técnica, as janelas adjacentes são sobrepostas de maneira que o fim de uma seja sobreposto ao início da próxima. A forma como as janelas são sobrepostas está relacionada com a taxa de sobreposição em porcentagem, um valor de *overlap* igual 2 corresponde a uma sobreposição de 50%. Em geral, mais sobreposição resultará em mais pontos de análise e, portanto, resultados mais atenuados ao longo do tempo, porém o gasto computacional é proporcionalmente maior.

Um parâmetro importante envolvido no processo de *overlap* é o *hop size*. O *hop size* está relacionado ao tamanho da janela e corresponde ao número de amostras entre uma análise e outra. Define-se o valor de *hop size* pela razão entre o tamanho da janela e o *overlap*. Considerando uma janela com tamanho de 512 amostras, e *overlap* 2, então o *hop size* corresponderá à 256 amostras. Quanto menor for seu valor, mais janelas sucessivas serão sobrepostas (ROADS, 1996, p. 567). Na figura 2.15 é possível ver um sinal segmentado em janelas sobrepostas de tamanho  $K$ , e *hop size* equivalente à  $Q$ .

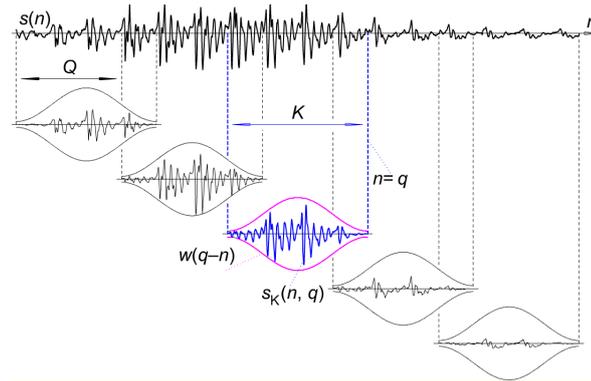


Figura 2.15: Demonstração do *hop size* em um sinal segmentado em sucessivas janelas sobrepostas

## 2.5 Análise de *Fourier*

A análise de *Fourier* consiste em um conjunto de operações matemáticas que englobam a série de *Fourier* e as transformadas de *Fourier*. Tais operações são voltadas à descrição do comportamento dos sinais em termos de suas componentes frequenciais. Desenvolvida a partir de estudos sobre a teoria analítica do calor pelo matemático Jean Baptiste Joseph Fourier, descobriu-se a possibilidade de descrever uma função arbitrária enquanto a soma de um conjunto de senoides (TEMPELAARS, 2016, p 118). Dessa maneira, é possível decompor o sinal sonoro em um conjunto de frequências, cada qual com valores individuais de amplitude e fase. Em virtude disso, tal processo de decomposição implica na transformação do sinal representado no domínio do tempo, para uma representação no domínio da frequência (COOK, P. R., 2002, p. 52).

A análise de *Fourier* é uma ferramenta central no que diz respeito às técnicas de processamento e análise de sinais digitais. Esse processo é o ponto de partida para a realização de análises, bem como para operar transformações diversas no sinal. No contexto dos processos de *Music Information Retrieval* e *Machine Listening*, a análise de *Fourier* é amplamente utilizada na implementação de diversos tipos de descritores de áudio de baixo-nível e alto-nível, atuando na recuperação de informações espectrais e harmônicas — em alguns casos também pode ser aplicada na extração de atributos temporais. A aplicação dessas técnicas variam de acordo com a natureza do sinal analisado. Para sinais periódicos comumente utiliza-se a série de *Fourier*, e para sinais não periódicos aplica-se a transformada de *Fourier*. Existem alguns tipos de transformada de *Fourier* que são específicas para se

analisar sinais discretos no tempo. A seguir descreveremos os princípios de atuação de cada uma dessas operações.

### 2.5.1 Série de *Fourier*

Através das séries de *Fourier* sinais periódicos que compreendem à uma soma ponderada de senoides podem ser representados completamente no domínio da frequência em termos de amplitude e ângulo de fase de seus componentes individuais (ANTONIOU, 2005, p. 30). Essa periodicidade se dá quando ao adicionarmos qualquer múltiplo inteiro  $m$  de  $T$  à variável de tempo  $t$ , a função mantém o mesmo valor como no tempo  $t$  (ROSSING, 2007, p. 505). Ou seja:

$$x(t + mT) = x(t), \text{ para } m = 1, 2, 3, \dots\infty \quad (2.44)$$

onde  $x(t)$  é uma função periódica, e  $T$  é uma constante denominada de período do sinal, representada por:

$$x(t) = \sum_{r=-\infty}^{\infty} x(t + mT) \quad (2.45)$$

Como  $m$  pode ser positivo, negativo e infinito,  $x$  é periódico no futuro e no passado infinitos. Logo, podemos representar  $x$  como uma série de *Fourier* a partir de:

$$x(t) = A_0 + \sum_{n=1}^{\infty} [A_n \cos(\omega_n t) + B_n \sin(\omega_n t)] \quad (2.46)$$

Todas as senoides que compõe o sinal têm frequência angular  $\omega_n$ , que são múltiplos inteiros (i.e., harmônicos) de uma frequência angular fundamental  $\omega_0$ .

$$\omega_n = n\omega_0 = 2\pi n/T \quad (2.47)$$

onde  $n$  é um número inteiro.

Para aplicações práticas da série de *Fourier*, não é suficiente apenas saber que é possível decompor um sinal em um conjunto de senoides, mas também é necessário determinar a amplitude e os ângulos de fase iniciais dessas componentes senoidais. Para isso é preciso definir os coeficientes de *Fourier* (TEMPELAARS, 2016, p. 122) — são os termos  $A_0$ ,  $A_n$ , e  $B_n$  presentes na equação 2.46.

O coeficiente  $A_0$  é uma constante que desloca a função  $x(t)$  para cima ou para baixo, sendo o único termo na série de *Fourier* que pode apresentar valores diferentes de zero quando calculada a média de um período. Portanto, para encontrar  $A_0$  projetamos a função  $x(t)$  em um cosseno de frequência zero que leva ao valor médio de  $x$  (ROSSING, 2007, p. 506). O termo  $A_0$  pode ser determinado pela equação:

$$A_0 = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt x(t) \quad (2.48)$$

Já os coeficientes  $A_n$  e  $B_n$  podem ser calculados projetando a função  $x(t)$  nas funções seno e cosseno das frequências harmônicas  $\omega_n$ . Ou seja, integrando o produto de  $x(t)$  e uma função seno ou cosseno ao longo de uma duração de tempo igual a um período de  $x(t)$ . As equações para os coeficientes  $A_n$  e  $B_n$  são:

$$A_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt x(t) \cos(\omega_n t) \quad \text{para } n < 0 \quad (2.49)$$

$$B_n = \frac{2}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} dt x(t) \sin(\omega_n t) \quad \text{para } n < 0 \quad (2.50)$$

## 2.5.2 Transformada de *Fourier*

Assim como a série de *Fourier*, a transformada de *Fourier* opera na decomposição de uma função representando suas componentes em função da frequência. Pode ser considerada uma extensão da série de *Fourier* voltada à análise de sinais contínuos e não periódicos — uma vez que a série de *Fourier* só é aplicável a sinais periódicos. Em suma, a transformada de *Fourier* consiste em uma função complexa com partes reais e imaginárias, representando frequências de 0 Hz à valores infinitos com imagem espelhada nas frequências negativas (ROADS, 1996, p. 1086). Portanto, é simétrica por volta de 0 Hz. A transformada de *Fourier* de um sinal  $x(t)$  é dada pela integral:

$$X(\omega) = F[x(t)] = \int dt e^{-i\omega t} x(t) \quad (2.51)$$

A aplicação da transformada de *Fourier* implica no processamento do sinal em dois estágios. Primeiramente decompõe-se o sinal e o descreve em relação à frequência — nesta fase do processo é possível operar manipulações na frequência e na amplitude e fase de cada uma delas. O segundo estágio consiste na reconstrução do sinal decomposto, passando do

domínio da frequência novamente para o domínio do tempo. Essa operação é realizada pela transformada inversa de *Fourier*, que é dada por:

$$x(t) = \frac{1}{2\pi} \int d\omega e^{i\omega t} X(\omega) \quad (2.52)$$

### 2.5.3 Transformada de *Fourier* para sinais discretos

Um sinal é discreto quando a variável frequência e a variável tempo são amostradas (COOK, P. R., 2002, p. 54), ou seja, variam de forma discreta e espaçada uniformemente. Um sinal discreto é obtido ao capturar o valor de uma função contínua  $x(t)$  com intervalos de tempo uniformemente distribuídos. Em situações nas quais o sinal a ser analisado é periódico e discreto, aplica-se a transformada discreta de *Fourier* (*Discrete Fourier Transform* - DFT). Calcula-se a transformada discreta através da equação:

$$X(m) = \sum_{i=0}^{N-1} x(n) [\cos(2\pi mn/N) - j\text{sen}(2\pi mn/N)] \quad (2.53)$$

onde  $m$  representa a frequência amostrada,  $n$  o tempo, e  $N$  é o tamanho do sinal analisado. A transformada inversa se dá por:

$$x(n) = \frac{1}{N} \sum_{m=0}^{N-1} X(m) [\cos(2\pi mn/N) + j\text{sen}(2\pi mn/N)] \quad (2.54)$$

Já a transformada de *Fourier* de tempo discreto (*Discrete-Time Fourier Transform* - DTFT), é a transformada aplicada à análise de sinais discretos não periódicos. Pode ser descrita por:

$$X_{2\pi}(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \quad (2.55)$$

e sua transformada inversa é dada por:

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X_{2\pi}^{\omega} \cdot e^{i\omega n} d\omega \quad (2.56)$$

### 2.5.4 Transformada de *Fourier* de curto-termo

A transformada de *Fourier* de curto termo (*Short-time Fourier Transform* - STFT) é um processo proposto por Gabor (1946), que se dá através da aplicação de uma sequência de

transformadas de *Fourier* em um sinal segmentado. Logo, seu processo consiste primeiramente na segmentação do sinal em janelas sucessivas com tamanho invariável, para então aplicar a transformada de *Fourier* em cada segmento individualmente.

Podemos relacionar a visualização de um sinal à uma fotografia em que se captura um *frame*. Se capturarmos vários *frames* com janelas sucessivas teremos algo que se assemelha ao registro de um filme do sinal. Portanto, a transformada de *Fourier* de curto-termo aplicada sucessivamente nestes *frames* revela a evolução espectral do sinal sonoro no tempo. Esse processo fornece informações localizadas, e é utilizado em situações nas quais os dados que se pretende extrair variam no tempo (KEHTARNAVAZ, 2008, p. 176). Ao selecionar o tamanho da janela e o *hop size* de maneira a ser perceptualmente relevante, a STFT pode ser usada como uma aproximação ao funcionamento do sistema auditivo humano em termos de percepção de áudio. (COOK, P. R., 2002, p. 60).

Dessa forma, sendo  $x : [0 : L - 1] := 0, 1, \dots, L - 1 \rightarrow \mathbb{R}$  um sinal discreto no tempo de tamanho  $L$  obtido por amostragens equidistantes em relação a uma taxa de amostragem fixa  $f_s$  dada em *Hertz*;  $w : [0 : N - 1] \rightarrow \mathbb{R}$  uma função de janela de tamanho  $N \in \mathbb{N}$  — no caso de uma janela retangular temos  $w(n) = 1$  para  $n \in [0 : N - 1]$  —;  $N$  a duração das seções que equivale a  $N/f_s$  dado em segundos, e  $H \in \mathbb{N}$  o *hopsiz*e. A STFT do sinal  $x$  é dado por:

$$X(m, k) := \sum_{n=0}^{N-1} x(n + mH)w(n)\exp(-2\pi i kn/N) \quad (2.57)$$

onde  $M := \lfloor \frac{L-N}{H} \rfloor$  é o número máximo de *frames* — assim o intervalo de tempo da janela está totalmente contido no intervalo de tempo do sinal —  $K = \frac{N}{2}$  é o índice de frequência correspondente à frequência *Nyquist* (assumindo que  $N$  seja um número par),  $x(m, k)$  caracteriza o  $k$ ésimo coeficiente de *Fourier* para o  $m$ ésimo *frame*. Nota-se que para cada intervalo de tempo fixo  $m$  obtém-se um vetor espectral de tamanho  $K + 1$  dado pelos coeficientes  $m, k$  para  $k \in [0 : K]$ <sup>5</sup>.

## 2.6 Cepstrum

O *Cepstrum* é uma operação matemática originalmente proposta por Bogert, Healt e Tukey (1963) para detectar ecos em movimentos sísmicos. Essa operação consiste em uma cadeia de processos que envolve a aplicação da transformada de *Fourier*, o cálculo do logaritmo

<sup>5</sup>[https://www.audiolabs-erlangen.de/resources/MIR/FMP/C2/C2\\_STFT-Basic.html](https://www.audiolabs-erlangen.de/resources/MIR/FMP/C2/C2_STFT-Basic.html)

da magnitude do espectro, e a transformada inversa de *Fourier*. A palavra *cepstrum* é um anagrama que se origina a partir da inversão das quatro primeiras letras da palavra *spectrum*.

Em virtude de sua operação atuar na transformação inversa de um sinal no domínio da frequência, o *cepstrum* resulta em uma função de uma variável semelhante ao tempo (ROSSING, 2007, p. 517). Quando calculamos o logaritmo da magnitude do espectro no domínio da frequência, as amplitudes são atenuadas tornando a sequência de picos quase periódica. Ao aplicarmos a transformada inversa de *Fourier* no logaritmo da magnitude do espectro obtemos algo como uma frequência, porém “invertida”. Essa frequência “invertida” é a coordenada semelhante ao tempo chamada de *quefrequency*, um anagrama de *frequency*. Logo, a magnitude, fase, harmônico e período no contexto do *cepstrum* são denominadas *gamitude*, *saphe*, *rahmonics*, e *repiod* respectivamente. A figura 2.16 mostra o espectro de *Fourier* e a representação no domínio do *quefrequency* de um sinal periódico composto por vinte componentes frequenciais.

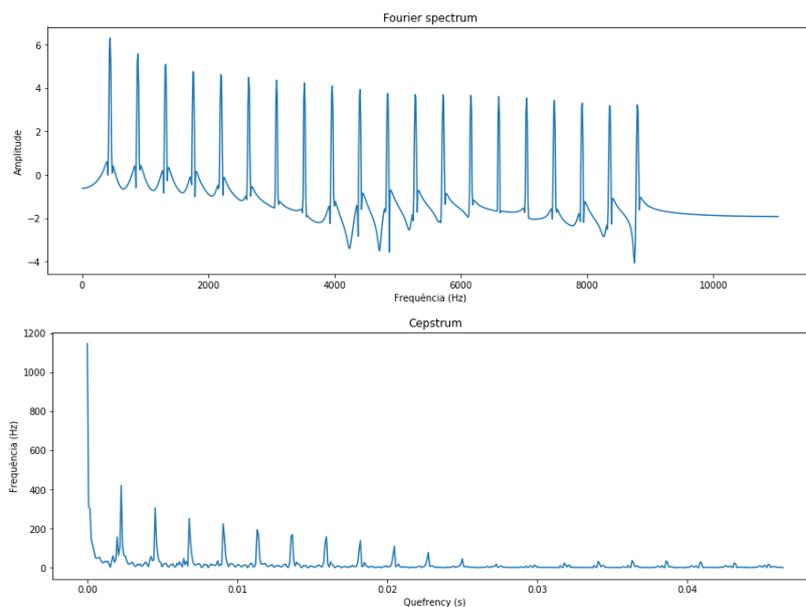


Figura 2.16: Espectro de *Fourier* de um sinal periódico (gráfico superior) e sua representação no domínio do *quefrequency* (gráfico inferior)

De acordo com Brent (2009b, p. 3), o *cepstrum* real é definido por:

$$x_{RC}(n) = \Re [If T \ln |X(k)|] \quad (2.58)$$

onde  $X(k)$  é a representação do domínio da frequência de um sinal  $x(n)$ , e  $\Re$  corresponde a porção real da transformada inversa de *Fourier*

Segundo Roads (1996, p. 516), a análise *cepstral* tende a separar o componente que está fortemente relacionado à altura dos outros elementos do espectro. Ou seja, aplicando tal operação é possível separar o impulso — excitação que coloca um sistema em vibração — da parte ressonante — a parte filtrada de um som produzida pela caixa de ressonância/corpo do instrumento ou o trato vocal. Em decorrência dessa particularidade o processo de *cepstrum* é amplamente utilizado no processamento da voz, uma vez que a produção de sons vocais consiste na convolução de impulsos glotais com a ressonância da cavidade oral. As técnicas de *cepstrum* são amplamente utilizadas em aplicações musicais — podendo atuar em conjunto com outras técnicas — voltadas a processos de *pitch tracking* em sons vocais.

### 2.6.1 Mel Frequency Cepstral Coefficients

*Mel Frequency Cepstral Coefficient* (MFCC) são coeficientes obtidos a partir da multiplicação da magnitude do espectro por um banco de filtros derivado da escala de frequência *Mel* antes da extração do *cepstrum*. De fato, ao implementar o processo de *cepstrum* a partir de filtros de banda crítica — como *Mel* e *Bark* — tal processo passa a atuar mais efetivamente como um descritor de características timbrísticas (BRENT, 2009b, p. 3).

O filtro de banda crítica baseado na escala de frequência *Mel* consiste em uma série de filtros triangulares sobrepostos de maneira uniforme e espaçados de acordo com a escala de frequência *Mel* (ver figura 2.17). Os parâmetros que definem esse banco de filtros são o número de filtros, frequência mínima, e frequência máxima (frequências de corte) (KOPPARAPU; LAXMINARAYANA, 2010, p. 2). O banco de filtro *Mel* pode ser definido por:

$$H(k, m) = \begin{cases} 0, & \text{para } f(k) < f_c(m-1) \\ \frac{f(k)-f_c(m-1)}{f_c(m)-f_c(m-1)}, & \text{para } f_c(m-1) \leq f(k) < f_c(m) \\ \frac{f(k)-f_c(m+1)}{f_c(m)-f_c(m+1)}, & \text{para } f_c(m) \leq f(k) < f_c(m+1) \\ 0, & \text{para } f(k) \geq f_c(m+1) \end{cases} \quad (2.59)$$

onde cada filtro calcula a média do espectro em torno da frequência central dada em *Mel*, resultando em diferentes larguras de banda. Assim, quanto maior for a frequência, maior será a largura da banda.

O processo de cálculo dos MFCCs se dá primeiramente com a segmentação do sinal a ser analisado através de uma função de janelamento e da aplicação da transformada discreta

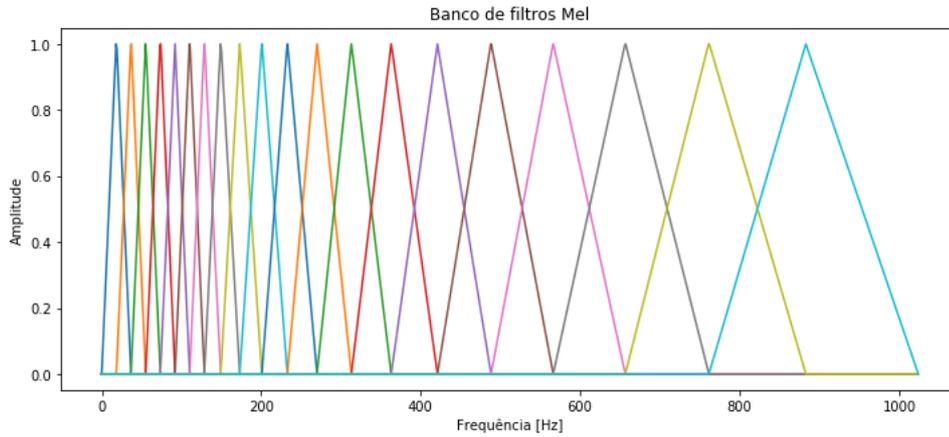


Figura 2.17: Representação das bandas de frequência do banco de filtros *Mel*

de *Fourier* para cada janela. Após aplicar-se a DFT, a magnitude do espectro passa pelo banco de filtros *Mel*. Em seguida calcula-se o logaritmo da saída do banco de filtros — que consiste no produto do banco de filtros *Mel* e da magnitude do espectro. E por fim, aplicando-se a transformada discreta de cosseno (CDT) tem-se os MFCCs, que são definidos matematicamente por:

$$MFCC = \sum_{k=0}^{N-1} X_k \cos \left[ i \left( k + \frac{1}{2} \right) \frac{\pi}{N} \right]; \quad i = 0, 1, \dots, M - 1 \quad (2.60)$$

onde  $M$  é o número de coeficientes que se deseja obter,  $N$  é o número de filtros, e  $X_k$  é o logaritmo da saída do  $k$ -ésimo filtro (BRENT, 2009b, p. 3). A figura 2.18 mostra a representação gráfica dos MFCCs extraídos de um som.

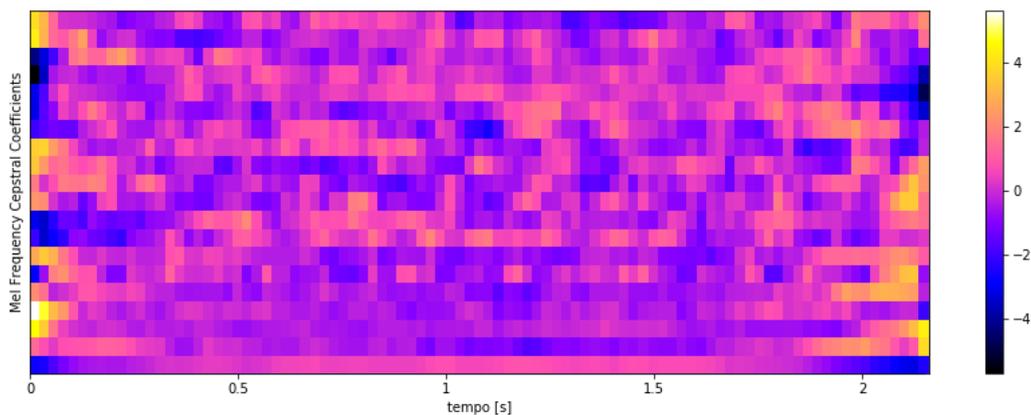


Figura 2.18: Representação gráfica dos MFCCs extraídos a partir de um  $l_4$  tocado por uma flauta

### 2.6.2 Bark Frequency Cepstral Coefficients

Assim como o MFCC, os *Bark Frequency Cepstral Coefficients* (BFCC) são coeficientes que resultam da aplicação de um banco de filtros *Bark* ao processo de cálculo do *cepstrum*. A única diferença no processo de extração do BFCC em relação ao MFCC, é a implementação do banco de filtros. Para o cálculo do BFCC utiliza-se um banco de filtros derivado da escala de frequência *Bark* – as frequências centrais são definidas pela escala *Bark*. Segundo Cheng, Abdulla e Salcic (2005, p. 7), define-se matematicamente o banco de filtros *Bark* por:

$$H(k) = \begin{cases} 0, & \text{para } f(k) - f_c < -2.5 \\ 10^{(f(k)-f_c+0.5)}, & \text{para } -2.5 \leq f(k) - f_c \leq -0.5 \\ 1, & \text{para } 0.5 < f(k) - f_c < 0.5 \\ 10^{-2.5(f(k)-f_c-0.5)} & \text{para } 0.5 \leq f(k) - f_c \leq 1.3 \\ 0, & \text{para } f(k) - f_c > 1.3 \end{cases} \quad (2.61)$$

onde  $f_c$  é a frequência central do filtro dada em *Bark*,  $H(k)$  é o peso do filtro em  $f(k)$ . A figura 2.19 mostra uma representação gráfica das bandas de frequência do banco de filtros *Bark*.

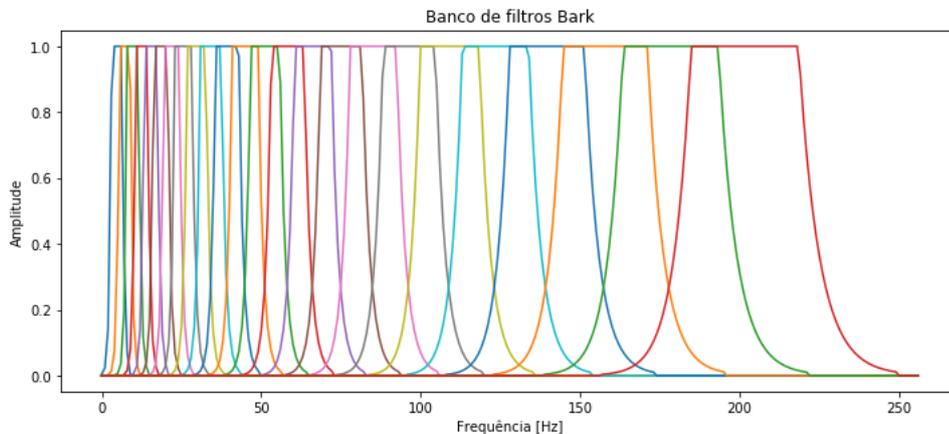


Figura 2.19: Representação das bandas de frequência do banco de filtros *Bark*

Calculando o logaritmo da saída do banco de filtros *Bark* – magnitude do espectro multiplicada pelo banco de filtros – e em seguida aplicando a transformada discreta inversa de cosseno, tem-se os 24 BFCCs. Liang et al. (2020) define os BFCCs matematicamente por:

$$BFCC = c(i) \sum_{m=1}^M \lg [e(m)] \cos \left[ \frac{(m - 0.5)\pi i}{M} \right] \quad (2.62)$$

$$\text{onde } c(i) = \begin{cases} \sqrt{\frac{1}{M}}, & i = 0 \\ \sqrt{\frac{2}{M}}, & i \neq 0 \end{cases}$$

Aqui, o número de bandas  $M$  é igual à 24. A figura 2.20 mostra a representação gráfica dos BFCCs extraídos a partir da mesma amostra sonora utilizada na representação da 2.18.

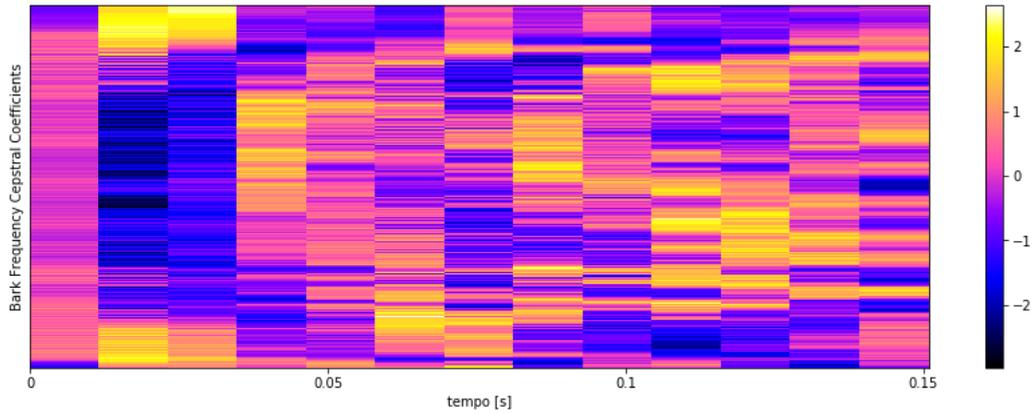


Figura 2.20: Representação gráfica dos BFCCs extraídos a partir de um  $l_{\wedge 4}$  tocado por uma flauta

%

## Capítulo 3

# Processos de *Music Information Retrieval* e *Machine Listening*

O presente capítulo tem o propósito de apresentar os modelos e técnicas voltados à ML/MIR que serão de alguma forma abordados neste trabalho, seja no contexto das observações analíticas realizadas no capítulo 4, ou no desenvolvimento do projeto composicional que será descrito no capítulo 5. Para tanto, faremos uma taxonomia de tais modelos de maneira a classificá-los de acordo com o tipo de processamento digital do sinal necessário para implementá-los, e em perspectiva à natureza das informações que são extraídas.

Essa organização tem como ponto de partida a revisão bibliográfica voltada aos campos da ML e MIR buscando ao mesmo tempo apontar para sua aplicabilidade em contexto criativo. Assim, o objetivo aqui não é aquele de oferecer uma classificação exaustiva e definitiva desses modelos ou de sua correlação com determinados atributos sonoros, mas o de reunir arcabouços técnicos necessários à compreensão dos próximos capítulos.

O capítulo segue a seguinte estrutura: na primeira parte será realizada uma taxonomia geral dos modelos e técnicas voltados à ML/MIR através da correlação com os processos de cálculo utilizados na recuperação das informações e atributos sonoros/musicais; e a segunda parte aborda de maneira descritiva os modelos e técnicas que serão abordados ao longo do trabalho.

### 3.1 Categorias e taxonomia geral dos descritores de áudio

Há uma grande variedade de informações sonoras e musicais que podem ser extraídas a partir dos métodos de DSP associados às técnicas de ML/MIR. Visando a recuperação de tais informações, e conseqüentemente, a possibilidade de descrever características ligadas ao comportamento físico e perceptível do som, há uma série de modelos matemáticos em constante desenvolvimento pelas áreas da ML e MIR, denominados *descritores de áudio* (PEETERS, 2004, p. 1).

Os *descritores de áudio* são um conjunto de algoritmos/modelos computacionais voltados à extração de informações quantitativas que podem ser correlacionadas à uma ampla gama de aspectos e características ligadas ao comportamento sonoro, tais como altura, dinâmica, brilho, harmonicidade, entre outros. Além de extrair informações estatísticas relacionadas ao comportamento físico do sinal de áudio, alguns desses descritores também podem ser estruturados computacionalmente a partir da modelização de fenômenos psicoacústicos e cognitivo-musicais (SIMURRA, 2016, p. 7).

Visando uma compreensão tipomorfológica que possa oferecer arcabouços técnicos, é proposto nesta sub-seção uma taxonomia geral dos descritores de áudio que serão abordados ao longo do trabalho. Tal categorização apresenta como critério de agrupamento a natureza dos processos de cálculo e os tipos de informações que cada um dos modelos extrai a partir do sinal de áudio.

Ao propor uma taxonomia, é importante ressaltar que tal processo sempre estará sujeito a escolhas arbitrárias devido às diferentes possibilidades e perspectivas que podem ser consideradas como critérios no processo de categorização, critérios esses que podem apresentar variados graus de subjetividade. Se por um lado deve-se atentar para o fato de que o procedimento taxonômico apresenta um caráter arbitrário, por outro, a classificação dos descritores, sobretudo do ponto de vista da natureza de seus processos de cálculo pode mostrar-se eficaz no que diz respeito à um estudo técnico. Contudo, não temos como objetivo o esgotamento das possibilidades, nem a criação de uma categorização absoluta, mas sim fornecer subsídios para uma compreensão técnica.

De modo geral, dividiremos os descritores em duas “famílias”: descritores de baixo-nível, e descritores de alto-nível. Os descritores de baixo-nível englobam aqueles que são representados tanto no domínio do tempo como no da frequência, e são responsáveis por extrair informações que correspondem à características quantitativas ligadas diretamente ao

comportamento do próprio sinal sonoro. Já os descritores de alto-nível extraem informações correlacionadas à características semânticas, qualitativas e perceptivas do som (AMATRIAIN, 2005, p. 249), e são implementados a partir dos descritores de baixo-nível associados a processos de modelização de habilidades cognitivas e auditivas humanas (MONTEIRO, 2012, p. 24).

A “família” dos descritores de baixo-nível se organiza em dois grupos de acordo com o domínio em que são representados: descritores temporais, e descritores espectrais. Os descritores temporais são calculados no domínio do tempo e estão ligados ao comportamento temporal do sinal digital, já os descritores espectrais são calculados no domínio da frequência, e estão relacionados ao comportamento espectral do sinal.

Os descritores computados no domínio do tempo podem ser divididos em dois subgrupos: global e instantâneo. O subgrupo temporal global reúne os descritores que extraem informações ligadas à envoltória de energia de todo o sinal ou em regiões pré-segmentadas. Este tipo de descritor fornece dados estatísticos extraídos a partir da evolução da energia do sinal em relação ao tempo, dessa forma permitindo a classificação de aspectos físico-temporais que se correlacionam com a própria envoltória de energia — transientes de ataque; centróide temporal; energia instantânea, por exemplo. O subgrupo temporal instantâneo abarca os descritores computados em pontos específicos do sinal a partir de pequenas janelas de amostras. Eles são responsáveis por extrair informações relacionadas à energia em um dado *frame*/janela. Logo, o resultado numérico obtido corresponde ao valor do cálculo realizado para cada janela que se pretende analisar.

Já os descritores espectrais são calculados no domínio da frequência, ou seja, para extrair as informações primeiramente é preciso decompor a função temporal (sinal) e representá-la no domínio da frequência. Para isso aplica-se a transformada discreta de *Fourier* — em análise de sinal é comum aplicar a transformada de *Fourier* em janelas de amostras sucessivas com tamanho invariável (*Short-Time Fourier Transform* — STFT) (MONTEIRO, 2012, p. 123).

Assim como os descritores temporais, as informações recuperadas pelos descritores espectrais são de caráter quantitativo e estatístico, e correspondem ao comportamento espectral do sinal — centróide espectral, espalhamento espectral, decaimento espectral, entre outros —, e/ou à características do comportamento harmônico extraídas a partir de modelos

que buscam descrever as relações harmônicas entre as frequências de um espectro — por exemplo, harmonicidade/inarmonicidade, frequência fundamental, *tristimulus*, entre outros.

Enquanto os descritores de baixo-nível recuperam informações relacionadas à estrutura e ao comportamento do sinal, os descritores de alto-nível buscam extrair informações que podem ser correlacionadas a atributos sintáticos e semânticos como padrões e modelos musicais, e/ou classificações de características sonoras do ponto de vista da interpretação perceptiva e cognitiva do ouvinte. Apesar de não extraírem dados que correspondem diretamente à estrutura interna do sinal sonoro, esses descritores são implementados a partir da combinação do cálculo dos descritores de baixo-nível (tanto espectrais como temporais) e de modelos perceptíveis/cognitivos. (AMATRIAIN, 2005, p. 251).

Um dos modelos perceptivos utilizados na construção de alguns descritores psicoacústicos, é a banda crítica (FLETCHER; MUNSON, 1933; FLETCHER, 1940). Trata-se de um modelo que descreve a largura da banda de frequência de um filtro hipotético criado pela cóclea, e tem como objetivo estabelecer coordenadas de posição na membrana basilar de maneira que o espectro de frequência possa ser dividido em partes iguais em perspectivas a fenômenos auditivos. A partir disso, pode-se calcular os Coeficientes Cepstrais por Frequências *Mel* (*Mel Frequency Cepstral Coefficients* – MFCC) e *Bark* (*Bark Frequency Cepstral Coefficients* – BFCC). Com tais modelos é possível, por exemplo implementar o descritor *perceptual sharpness*, que é o equivalente perceptivo da centróide espectral, porém calculado usando *loudness* específico das bandas *Bark* (ZWICKER, 1977; PEETERS, 2004)

Os descritores de alto-nível também são capazes de extrair informações relacionadas à eventos musicais simbólicos. Por exemplo, os intervalos entre *onsets* — ou *Inter-onset Interval* - IOI —, podem ser correlacionados com durações de eventos musicais (BROWN, 1993, p. 1954). O intervalo entre *onsets* identifica a densidade de *onsets* de um dado excerto. Logo, uma grande quantidade de *onsets* resulta em uma alta densidade de eventos, e em baixo índice de IOI's. Ao contrário, trechos com uma baixa quantidade de *onset* terão valor alto para IOI (SIMURRA, 2016, p. 47).

Vale ressaltar que embora possa-se traçar algumas correlações entre os dados extraídos pelos descritores de baixo/alto-nível a critérios tipomorfológicos e qualificações da escuta, não há uma correlação direta que possa associá-los. Os processos de cálculo realizados pelos algoritmos dos descritores de áudio retornam resultados de natureza

quantitativa, ou seja, são uma representação estatística que concerne à medição de valores ou quantidades relacionadas ao comportamento do sinal digital analisado.

Apesar de reduzir a complexidade do fenômeno sonoro a uma representação numérica, em certa medida, uma correlação entre as informações recuperadas pelos descritores de áudio e semânticas utilizadas na descrição verbal de características sonoras, como brilho, rugosidade e maciez é possível. Por exemplo, os trabalhos de Freire, Padovani e Campos (2021b,a) buscam classificar sons percussivos a partir de uma associação dos descritores de áudio de baixo/alto-nível aos critérios tipomorfológicos de Pierre Schaeffer<sup>1</sup>.

Os processos de recuperação de informações executados pelos algoritmos dos descritores de áudio envolvem uma série de técnicas e procedimentos que abrangem desde a segmentação e representação do sinal, até a simulação de propriedades psicoacústicas. Como vimos anteriormente, os tipos de processos variam de acordo com o domínio em que cada descritor é representado e podem combinar diferentes tipos de procedimentos.

Segundo Peeters et al. (2011), os processos de cálculo e implementação dos descritores, bem como o formato das informações extraídas por eles estão relacionados a duas propriedades: extensão temporal, e representação do sinal. A extensão temporal está ligada à duração do segmento do sinal analisado. Um descritor pode extrair informações a partir de todo o evento sonoro (um ataque, ou um segmento com maior duração), ou a partir de um segmento específico de curta duração (janela) dentro do evento sonoro. No caso da representação do sinal temos três situações: (1) representação pela envoltória de energia, utilizada na implementação de descritores temporais; (2) pela segmentação em janelas e pela transformada de *Fourier*, utilizada no cálculo dos descritores espectrais instantâneos; e (3) pela modelização de propriedades perceptivas, aplicada no cálculo dos descritores psicoacústicos.

Cada uma dessas situações envolve a aplicação de uma cadeia de procedimentos específicos que são responsáveis por preparar o sinal e representá-lo de forma adequada ao tipo de informação que se pretende extrair. No capítulo 2 foram introduzidos alguns dos procedimentos necessários para estruturação de tais ferramentas. Dessa forma, para compreender os aspectos técnicos relacionados aos métodos de ML e MIR, um estudo que contemple técnica e conceitualmente os descritores de áudio se faz necessário. Para isso, nas

---

<sup>1</sup>A tipomorfologia de Schaeffer é uma metodologia prática que visa a descrição de objetos sonoros considerando aspectos perceptivos relacionados a altura, duração e intensidade divididos em sete critérios tipomorfológicos: *massa*; *dinâmica*; *timbre harmônico*; *perfil melódico*; *perfil de massa*; *grão* e *allure* (SCHAEFFER, 1966)

subseções seguintes iremos definir e descrever cada um dos descritores que serão abordados neste trabalho, de maneira a dividí-los em dois grupos: baixo-nível e alto-nível.

## 3.2 Descritores temporais de baixo-nível

### 3.2.1 *Centróide temporal* [*Temporal centroid*]

O *centróide temporal* calcula a energia ponderada do sinal sobre o tempo, sendo definida como o baricentro da envoltória de energia. Seu cálculo retorna um valor no eixo temporal dividindo o sinal analisado em duas partes que possuem a mesma quantidade de energia, a somatória da energia no trecho (MONTEIRO, 2012, p. 119). Para extrair a *centróide temporal* utiliza-se a potência instantânea, que é definida por  $P(t) = x^2(t)$ . Esta é a potência no momento  $t$  – i.e. a potência medida em tempo-real – e não depende do que acontece com o sinal antes ou depois desse momento (ROSSING, 2007, p. 510). O cálculo do *centróide temporal* se dá pela somatória da potência instantânea nos instantes  $t_1$  à  $t_2$  (instante inicial e final do sinal analisado, respectivamente), multiplicada pelo tempo e dividida pela somatória da energia em todos os instantes do sinal. Definimos matematicamente por:

$$CT = \frac{\sum_{t=t_1}^{t_2} te(t)}{\sum_{t_1}^{t_2} e(t)} \quad (3.1)$$

O *Centróide Temporal* é comumente utilizado na distinção entre sons percussivos e sons sustentados. A envoltória de energia de sons percussivos tendem a ter um acúmulo de energia próximo ao início do som, resultando em valores de *centróide temporal* baixos, e sons sustentados apresentam uma distribuição da energia mais simétrica no tempo, retornando valores mais centralizados no que diz respeito à duração do som (MONTEIRO, 2012, p. 119).

### 3.2.2 *Taxa de cruzamento por zero* [*Zero-crossing rate*]

Assim como a envoltória de energia, a *taxa de cruzamento por zero* (*zero-crossing rate*) é uma forma simples de representação do sinal. Esse método é responsável por medir no domínio do tempo em cada janela de amostras o número de vezes que a amplitude de um sinal passa pelo valor zero. O *cruzamento por zero* ocorre se sucessivas amostras têm diferentes sinais algébricos. (VALIATI, 2000, p. 36).

Sinais periódicos apresentam uma *taxa de cruzamento por zero* baixa, enquanto sinais ruidosos tendem a retornar um valor alto. Em sinais simples a *taxa de cruzamento por zero* pode ser utilizada no reconhecimento de alturas, já aplicada em sinais complexos, resulta em valores que podem ser correlacionados ao grau de ruidosidade do som. Tal cálculo é utilizado na detecção de blocos com sons surdos (como a consoante “s”), sonoros (vogais) e consoantes fricativas (como a consoante “f”), retornando valores baixos para vogais e altos para consoantes (RABINER; SCHAFER, 1978, p. 132). Seu processo de cálculo é definido formalmente pela equação:

$$TCZ = \frac{1}{2} \sum_{n=1}^{Na-1} |\text{sgn}[x(n)] - \text{sgn}[x(n-1)]| \quad (3.2)$$

em que:

$$\text{sgn}[x(n)] = \begin{cases} 1, & \text{se } x(n) \geq 0 \\ -1, & \text{se } x(n) < 0 \end{cases} \quad (3.3)$$

Em janelas que apresentam elevado número de amostras de valor nulo, esta definição não corresponde à realidade, pois, na transição de uma amostra negativa para uma amostra nula será contado um cruzamento por zero mesmo que a próxima amostra também seja negativa (PASSOS et al., 2006, p. 28).

### 3.2.3 *Auto-correlação [Autocorrelation function]*

A *auto-correlação (autocorrelation function)* trata-se de um processo temporal instantâneo que mede a similaridade de um sinal (ou de formas de ondas) com ele mesmo. Este cálculo retorna um vetor no domínio do tempo que representa a distribuição espectral do sinal (MONTEIRO, 2012, p. 122). No processo de cálculo, todas as amostras de um segmento de sinal são multiplicadas por pontos com amostras do mesmo segmento deslocadas no tempo, esses resultados são somados para que se obtenha os coeficientes de *auto-correlação*.

Esse método é baseado na determinação da função de auto-correlação  $R$ . Para obter a função de *auto-correlação*, partimos da função de um sinal finito discreto no tempo  $y(K)$  (figura 3.1)

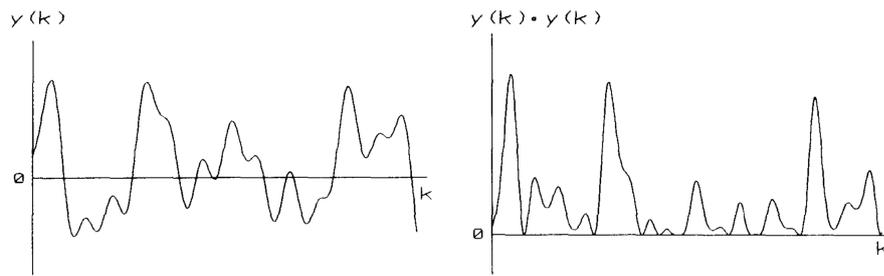


Figura 3.1: Função de sinal  $y(K)$  e função resultante do produto de  $y(K) \cdot y(K)$  (TEMPELAARS, 2016, p. 261)

Primeiramente multiplica-se a função  $y(K)$  por ela mesmo, como mostra a figura 3.1, em seguida soma-se todos os valores resultantes, resultando em um valor de soma que será designado como  $R(0)$ . Portanto:

$$R(0) = \sum_{K=1}^N y(K) \cdot y(K) \quad (3.4)$$

Em seguida, repetimos todo o processo, porem, multiplicando o sinal por ele mesmo deslocado uma amostra. O valor resultante da soma é designado como  $R(1)$ .

$$R(1) = \sum_{K=1}^N y(K) \cdot y(K + 1) \quad (3.5)$$

O valor de  $R(1)$  será um pouco menor que  $R(0)$ , porque os produtos  $y(K) \cdot y(K)$  são sempre positivos, enquanto no caso dos produtos  $y(K) \cdot y(K + 1)$  pode acontecer de dois valores de  $y$  apresentarem sinais diferentes, o que resultará em um produto negativo, portanto, contribuindo com um valor negativo na soma (TEMPELAARS, 2016, p. 261). Sendo assim, o índice de  $R$  fornece a defasagem em amostras aplicadas sobre o sinal. A equação geral para o processo de cálculo da função de *auto-correlação* é:

$$R(m) = \sum_{K=1}^N y(K) \cdot y(K + m) \quad (3.6)$$

No entanto, o cálculo descrito na equação 3.6 exige muita capacidade de processamento e pode ser otimizado implementando-o no domínio da frequência. Considerando que a operação de multiplicação no domínio do tempo corresponde a operação de convolução no domínio da frequência (e vice-versa) pode-se calcular a função de auto-correlação elevando o sinal representado no domínio da frequência à potência de dois.

(MONTEIRO, 2012, p. 122). Tal processo é chamado de *auto-correlação rápida* e se dá pela equação:

$$ACR = IFT \{ |x_i[K]|^2 \} \quad (3.7)$$

Sendo  $|x_i[K]|$  a magnitude da  $k$ ésima componente espectral do  $i$ ésimo quadro do sinal, e IFT é a transformada rápida de *fourier*. A figura 3.2 mostra a função de auto-correlação calculada a partir do sinal de um acorde arpejado no violão (gráfico de cima), e de uma senóide com frequência de 440 Hz (gráfico de baixo).

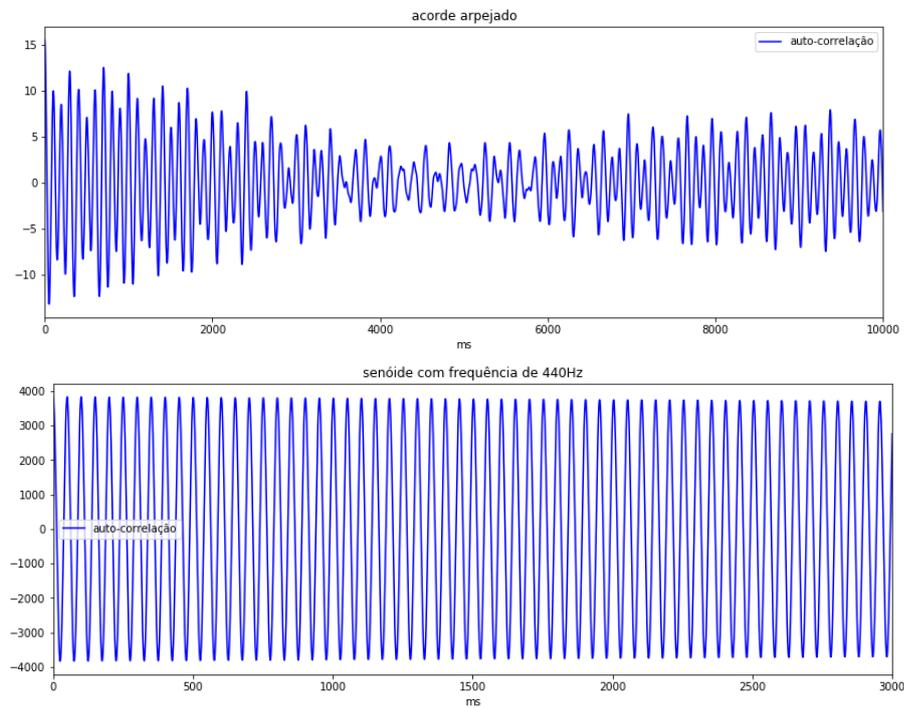


Figura 3.2: Função de auto-correlação de um acorde arpejado no violão (acima) e de uma onda senoidal com frequência de 440 Hz (abaixo)

### 3.3 Descritores espectrais de baixo-nível

#### 3.3.1 *Centróide espectral* [*Spectral centroid*]

O *centróide espectral* (*spectral centroid*) calcula o centro de massa (energia) ou baricentro do espectro de frequências. Em outras palavras, este descritor retorna um valor em *Hertz* que corresponde a região da envoltória espectral em que encontra-se a média da distribuição energética de todo o espectro. O cálculo do *centróide espectral* se dá por:

$$CE = \frac{\sum_{k=1}^{k=f} |x_i[k]|^2}{\sum_{k=1}^k -1 |x_i[k]|^2} \quad (3.8)$$

onde,

$$f[k] = \frac{\text{taxa de amostragem}}{2k} k \quad (3.9)$$

Apesar de se tratar de um descritor que extrai informações quantitativas a respeito do comportamento do espectro de frequências, o *centróide espectral* pode ser correlacionado à percepção de brilho do sinal sonoro analisado (PEREIRA, 2009, p. 16). Assim, quanto maior for o valor retornado pelo *centróide espectral*, maior será a concentração de energia nas regiões de alta frequência e maior será a sensação de brilho. Da mesma forma, valores baixos indicam concentração de energia em frequências baixas, portanto, causando a sensação de um som com pouco brilho. Contudo, o *centróide espectral* não está diretamente correlacionado ao registro ou à altura do som. (SIMURRA, 2016, p. 41). A figura 3.3 mostra a curva do *centróide espectral* da nota mi 4 tocada no violão (gráfico de cima) e em um crotales (gráfico de baixo). Nota-se que apesar de se tratarem da mesma altura (mi 4), o gráfico de baixo mostra uma curva com valores de *centróide espectral* maior.

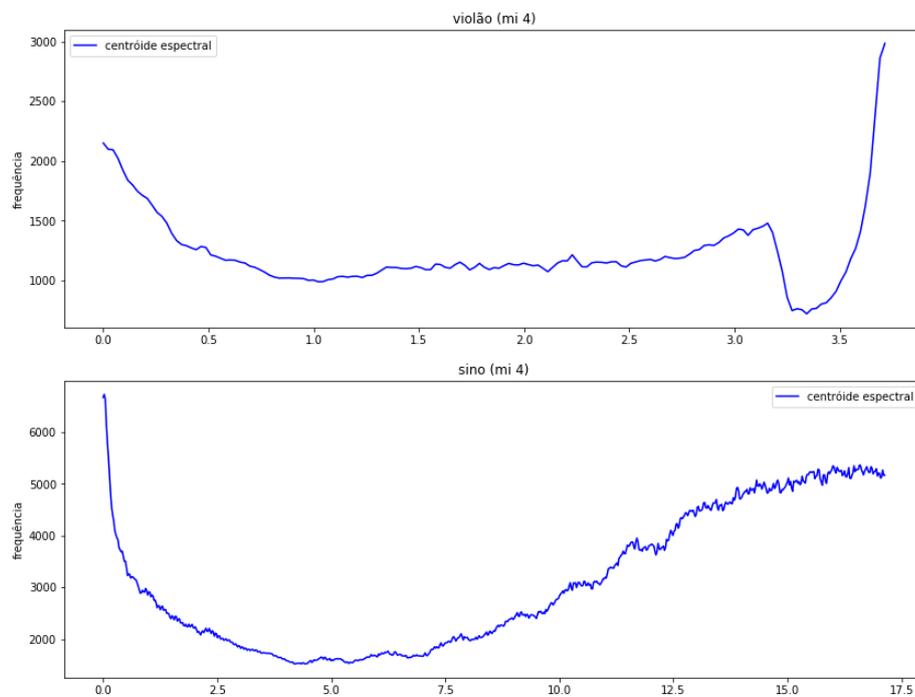


Figura 3.3: Curva de *centróide espectral* de um mi 4 tocado no violão (acima) e de um mi 4 tocado em um crotales (abaixo)

### 3.3.2 Fluxo espectral [Spectral flux]

*Fluxo espectral* ou *Variação espectral* (*Spectral variation*) mede a variação do espectro de frequências no tempo, é uma medida que descreve o quão rápido sucessivos *frames* do sinal variam ao longo do tempo. O *Fluxo espectral* retorna valores entre zero e um. Quanto mais próximo a zero, menor a similaridade entre os *frames*. Quanto mais próximo a um, maior a similaridade entre os espectros dos *frames* sucessivos. Utiliza-se esse descritor na implementação de métodos de detecção de *onset* – i.e. identificação de ataques e inícios de eventos sonoros –, e também na identificação de variação de articulação sonora, como vibrato e tremolo, por exemplo. De acordo com Wengen Wang et al. (2012), o cálculo do *fluxo espectral* é realizado pela distância euclidiana entre *frames* espectrais sucessivos:

$$FE = |X_k(n) - X_k(n-1)|_2 = \left( \sum_k |X(n, k) - X(n-1, k)|^2 \right)^{\frac{1}{2}} \quad (3.10)$$

Onde  $X(k, n)$  é a magnitude do espectro no *bin*  $k$  e no *frame*  $n$

A figura 3.4 representa a análise do *fluxo espectral* com a forma de onda do violão tocando a nota Mi4 em *tenuto* e em seguida executando acordes com a técnica do *rasgueado*. Ao observar a curva do *fluxo espectral* (preto) é possível distinguir o momento em que há a mudança de articulação do som *tenuto* para o *rasgueado*, pois a curva passa a apresentar um caráter mais difuso.

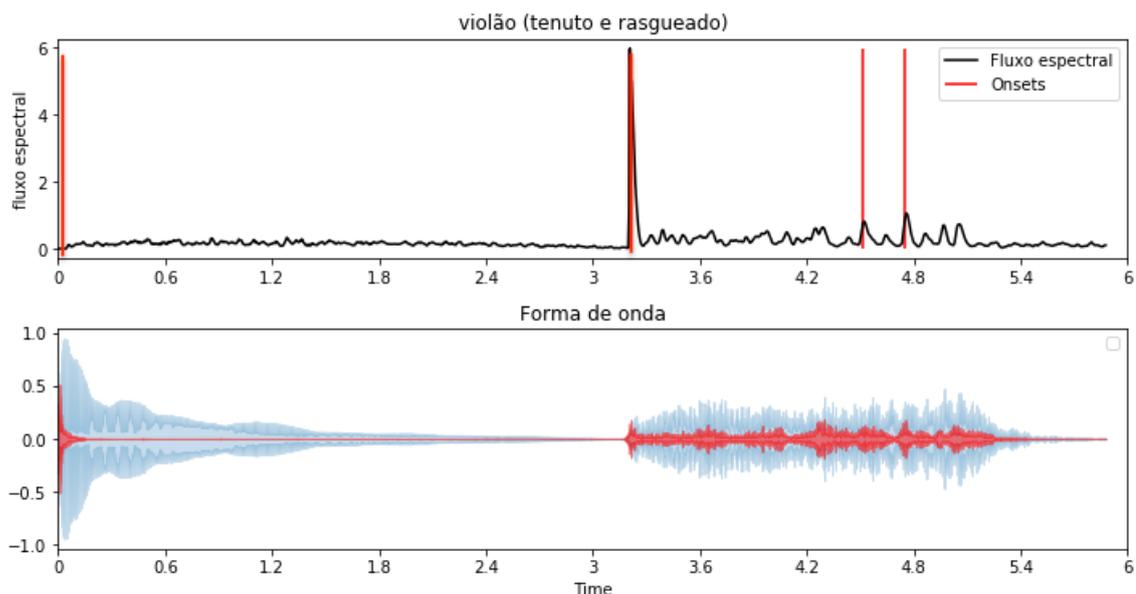


Figura 3.4: Análise do *fluxo espectral* (preto) e detecção de *onsets* (vermelho) do violão tocando a nota Mi4 em *tenuto* e em seguida executando acordes em *rasgueado*

### 3.3.3 *Espalhamento espectral [Spectral spread]*

O *espalhamento espectral* (*Spectral spread*), ou desvio padrão do espectro (*Spectral standard-deviation*) mede a dispersão ou variância da energia de um dado espectro ao redor do seu centro de massa — i.e., em torno do seu centróide —, e retorna valores medidos em *Hz*. O *espalhamento espectral* representa a “largura de banda instantânea” do espectro, e é usado como uma indicação da predominância de uma altura — e.g., uma senóide terá um valor de espalhamento muito baixo, pois toda a energia do seu espectro está concentrada em um só lugar. Por outro lado, o ruído apresentará um valor elevado. O *espalhamento espectral* é obtido a partir do cálculo do segundo momento estatístico.

$$\theta^2 = \int (x - \mu)^2 \cdot p(x) \sigma x \quad (3.11)$$

Na figura 3.5 podemos observar a análise do *espalhamento espectral* de uma onda senoidal de 440Hz (esquerda) e de um violino articulando vibrato (direita)<sup>2</sup>. Nota-se que a curva de *espalhamento espectral* da senóide apresenta um comportamento estável em relação ao *espalhamento espectral* do violino, uma vez que na senóide a energia está concentrada em um único lugar.

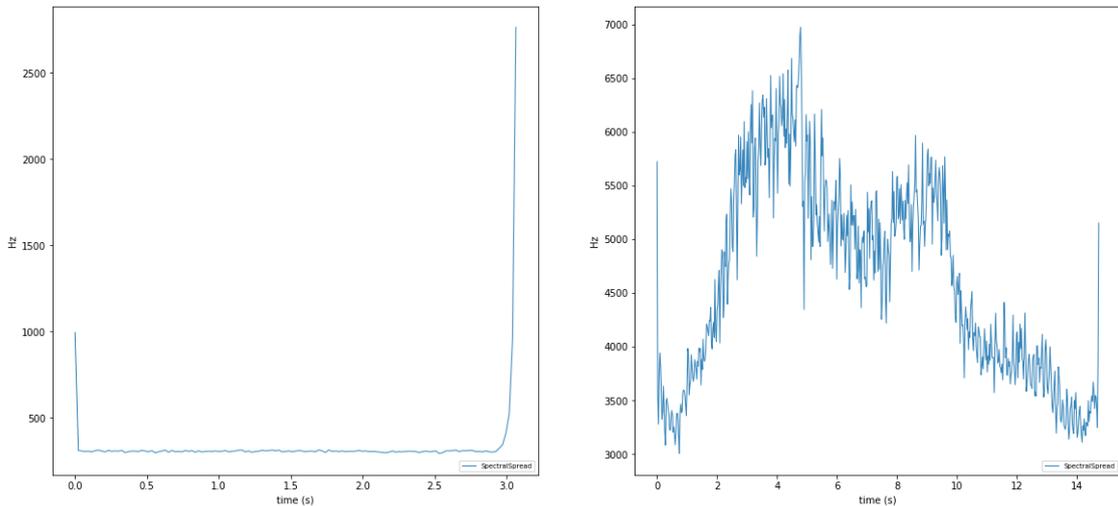


Figura 3.5: Análise do *espalhamento espectral* de uma onda senoidal de 440Hz (esquerda) e de um violino executando vibrato (direita)

<sup>2</sup>Esta análise foi realizada utilizando a biblioteca *Iracema* desenvolvida em *python* (MAGALHAES; BARROS; LOUREIRO, 2020)

### 3.3.4 *Obliquidade espectral [Skewness]*

*Obliquidade espectral (Skewness)* é uma medida de simetria/assimetria ou distorção que se relaciona com uma curva *gaussiana*. Esse descritor calcula o grau de assimetria da distribuição energética em torno da *centróide espectral*, e é computada a partir do terceiro momento estatístico (PEETERS, 2004, p. 13) (ver equação 3.12)

$$m_3 = \int (x - \mu)^3 \cdot p(x) \sigma x \quad (3.12)$$

onde,  $\mu$  é o *centróide espectral* e  $\sigma$  é o desvio padrão. Dessa forma, a *obliquidade espectral* é obtida por:

$$\gamma_1 = \frac{m_3}{\sigma^3} \quad (3.13)$$

Uma vez que este descritor mede a assimetria da distribuição de energia do espectro de frequências, um valor de *obliquidade espectral* igual a zero corresponde a um espectro simétrico. Valores menores do que zero indicam maior distribuição de energia no lado direito, portanto, predominância de frequências graves — podendo se correlacionar à percepção de sonoridades “arredondadas” e “quentes”. Já valores maiores do que zero indicam maior concentração de energia no lado esquerdo, consequentemente, indica predominância de frequência nas regiões agudas — relacionando-se com a percepção de sonoridades “brilhantes” (BRENT, 2010, p. 131). A figura 3.6 mostra diferentes valores de *obliquidade espectral* — esquerda = 0, centro < 0 e direita > 0.

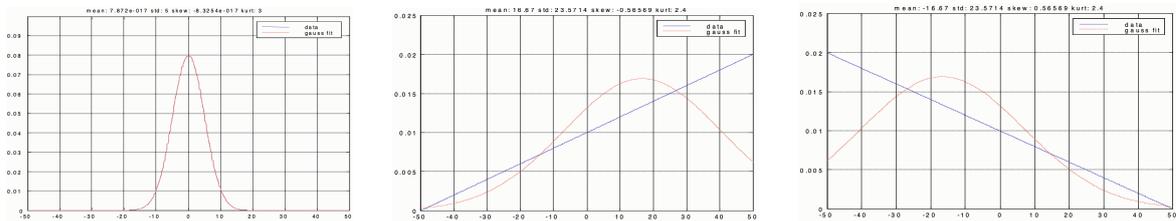


Figura 3.6: Diferentes valores de *obliquidade espectral* de acordo com o comportamento do espectro (PEETERS, 2004, p. 13)

### 3.3.5 Inclinação espectral [Spectral slope]

A *Inclinação espectral* é uma medida que consiste na estimativa do decrescimento da energia da envoltória espectral, e é calculada por regressão linear<sup>3</sup> da energia espectral. Seu cálculo é realizado através da equação:

$$IE = \frac{1}{\sum_{k=1}^{K-1} |x_i[k]|} \frac{N \sum_{k=1}^k f[k] |x_i[k]| - \sum_{k=1}^k f[k]}{\sum_{k=1}^k |x_i[k]| N \sum_{k=1}^{K-1} f[k]^2 - (\sum_{k=1}^{K-1} f[k])^2} \quad (3.14)$$

Sendo  $|x_i[k]|$  a magnitude da *késima* componente espectral da *inésima* janela,  $k$  é metade do número de amostras da janela de análise, e  $f[k]$  é a frequência central do componente  $k$  medida em Hz.

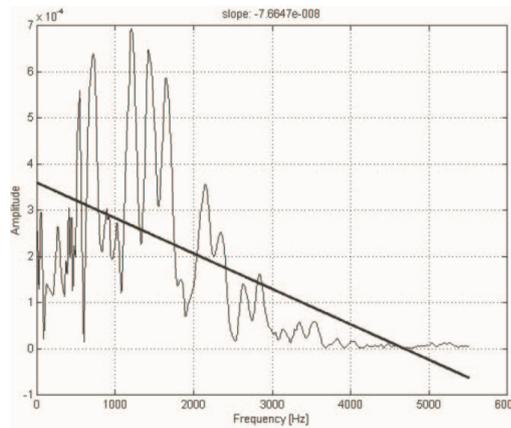


Figura 3.7: *Inclinação espectral de um espectro de frequências* (PEETERS, 2004, p. 14)

### 3.3.6 Roll-off espectral [Spectral roll-off]

O *Roll-off espectral* foi inicialmente proposto por Scheirer e Slaney (1997) , e calcula o ponto (frequência) do espectro abaixo do qual uma certa porcentagem da energia total da envoltória espectral – e.g.,85 – está contida. Isto é, o *roll-off espectral* está correlacionado ao ponto de declive que antecede o pico de frequência que contém a maior energia do espectro de frequências (SIMURRA, 2016, p. 49). Seu cálculo se dá por:

$$\sum_{f=0}^{f_c(tm)} a_f^2(tm) = 0.95 \sum_{f=0}^{\frac{sr}{2}} a_f^2(tm) \quad (3.15)$$

<sup>3</sup>regressão linear gera uma equação que descreve a relação estatística entre uma ou mais variáveis preditoras e a variável resposta. A regressão linear encontra a linha que melhor representa as variáveis de entrada com a variável de saída

Sendo  $f_c(tm)$  a frequência de *roll-off*, e  $\frac{sr}{2}$  a  $n$ ésima frequência.

Pode-se utilizar o *roll-off* na distinção entre sons harmônicos (abaixo do *roll-off*) e ruidosos (acima do *roll-off*), e também entre a voz falada e voz não falada. — a voz não falada tem uma alta proporção de energia contida na faixa de alta frequência do espectro. (SCHEIRER; SLANEY, 1997, p. 1). A figura 3.8 mostra a curva de *roll-off* de um *pizzicato* *Bartok* tocado no violão (gráfico acima) e de um harmônico natural também tocado no violão (gráfico abaixo).

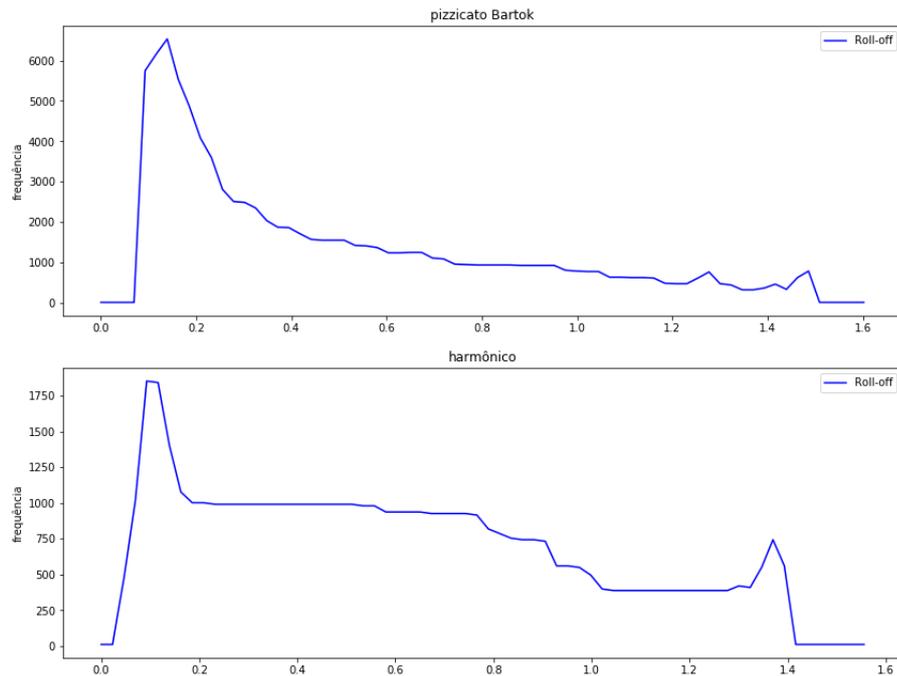


Figura 3.8: Curva de *roll-off* de um *pizzicato* *Bartok* (acima) e de um harmônico natural tocados no violão

### 3.3.7 Irregularidade espectral [*Spectral irregularity*]

A *irregularidade espectral*, ou *Desvio espectral harmônico* (*Harmonic spectral deviation*) é responsável por extrair o grau de variação espectral, e mede o desvio padrão da magnitude das frequências presentes na envoltória espectral. Segundo Krimphoff, Mcadams e Winsberg (1994) a irregularidade espectral é um importante atributo perceptivo utilizado na identificação e distinção de sonoridades e timbres. Envoltórias espectrais, com pouca variação de picos apresentam baixo índice de irregularidade, caracterizando timbres “aveludados” e “lisos”. Já envoltórias espectrais com maior variação de picos espectrais

apresentam alto índice de *irregularidade espectral*, indicando sonoridades “rugosas” e “estriadas” (SIMURRA, 2016, p. 48). O cálculo da *irregularidade espectral* é realizado por:

$$IE = \frac{1}{H} \sum_h (a(h) - SE(h)) \quad (3.16)$$

Sendo  $H$  o número total de harmônicos,  $a(h)$  a amplitude no harmônico  $h$  e  $SE(h)$  a amplitude da envoltória espectral na frequência  $f_h$  (PEETERS et al., 2011, p. 2909).

A figura 3.9 mostra a curva de *irregularidade espectral* de dois instrumentos, clarinete em Sib e trompete, executando um Mi4 com a dinâmica forte. É possível observar que a curva de irregularidade do clarinete apresenta poucos picos, por outro lado a curva do trompete apresenta um comportamento mais difuso com maior número de picos. Dessa forma, comparando as duas análises conclui-se que o clarinete possui um som mais “aveludado” em relação ao trompete, mais “rugoso”.

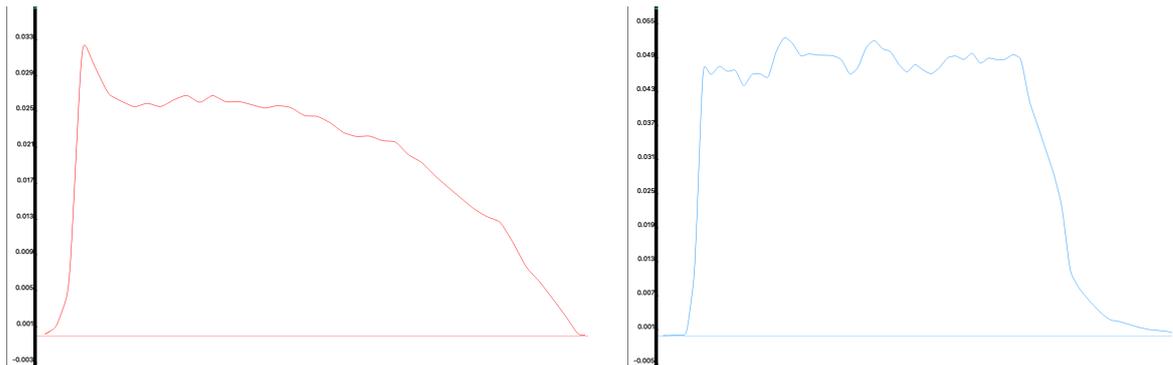


Figura 3.9: Análise da *irregularidade espectral* do clarinete (vermelho) e do trompete (azul) executando a nota Mi4

### 3.3.8 Razão entre componentes ímpares e pares [*Odd-to-even ratio*]

Este descritor calcula a razão entre a energia de componentes harmônicos ímpares e pares e é utilizado para distinguir sons que apresentam predominância de energia nos harmônicos ímpares de sons que possuem maior energia nos componentes pares (PEETERS et al., 2011, p. 2909). Valores  $> 1$  indicam maior energia nos componentes ímpares, enquanto valores  $< 1$  indicam predominância de energia nos componentes pares (SIMURRA, 2016, p. 51). O cálculo da *razão entre componentes ímpares e pares* é realizado por:

$$RCIP_{(tm)} = \frac{\sum_{h=1}^{H/2} a_{2h-1}^2(tm)}{\sum_{h=1}^{H/2} a_{2h}^2(tm)} \quad (3.17)$$

A figura 3.10 apresenta a *razão entre componentes ímpares e pares* do violino (esquerda) e do clarinete (direita), ambos com dinâmica *fortissimo*. A análise do som do clarinete retornou valores maiores que um, indicando predominância de energia nos componentes harmônicos ímpares. Já a análise do som do violino retornou valores menores que um, indicando predominância de energia nos componentes harmônicos pares.

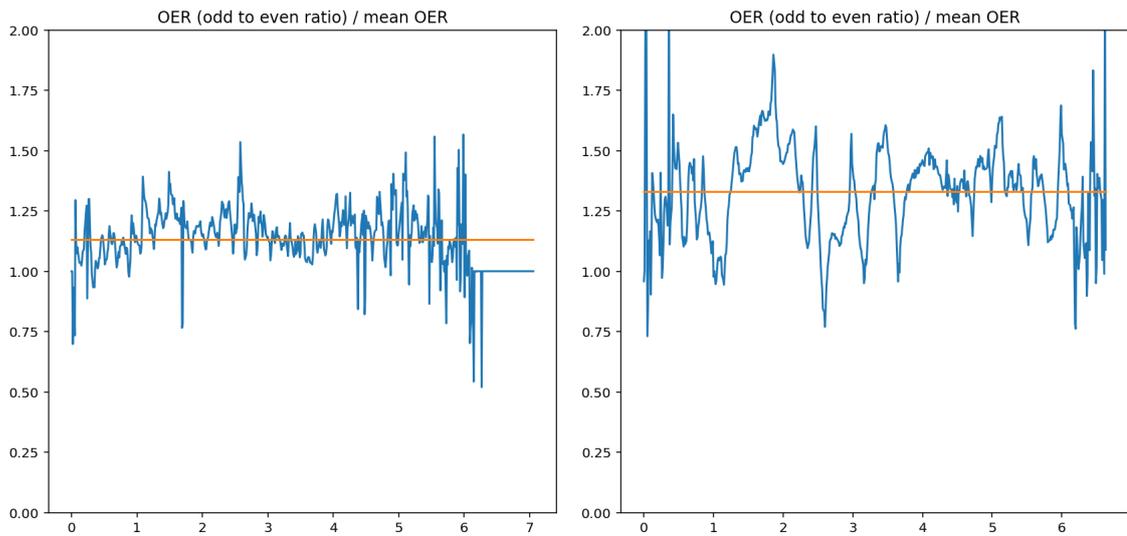


Figura 3.10: Análise da *razão entre componentes ímpares e pares* do violino (esquerda) e clarinete (direita)

### 3.3.9 *Tristimulus*

O *Tristimulus* consiste em três valores correspondentes a energia de três regiões diferentes do espectro de frequências. Foi inicialmente proposto por Pollard e Jansson (1982) como um equivalente timbrístico a um sistema homônimo que busca combinar visualmente cores sob condições padronizadas a partir das três cores primárias — vermelho, verde e azul. No âmbito do processamento digital de áudio os valores do *tristimulus* são calculados a partir da extração das medidas de *loudness* de três regiões específicas do espectro de frequências. Essa divisão foi escolhida por Pollard e Jansson (1982) considerando as bandas críticas, e pela importância que certos parciais ímpares têm na classificação de timbres (YSTAD, 1998, p. 79).

Dessa forma, o primeiro valor do *tristimulus* corresponde ao *loudness* da frequência fundamental; o segundo é referente à soma dos *loudness* dos parciais dois à quatro; e o terceiro consiste na soma dos *loudness* dos parciais restantes. O cálculo dos valores do *tristimulus* pode ser realizado através da fórmula de Steven:

$$N_i^n = 0.85N_{MAX} + 0.15 \sum_i^n N_i \quad (3.18)$$

Onde,  $N_i^n$  corresponde ao *loudness* de cada conjunto de parciais,  $N_{MAX}$  é a parte com maior valor de *loudness* de cada grupo de parciais, e  $\sum_i^n N_i$  é o *loudness* total de cada grupo. Somando o *loudness* de cada um dos três grupos obtemos o *loudness* total  $N$  do som.

$$N = N_1 + N_2^4 + N_5^n \quad (3.19)$$

Uma vez que cada um dos três valores do *tristimulus* corresponde a uma coordenada  $(x, y, z)$ , é possível representá-los em um diagrama tridimensional (figura 3.11).

$$x = \frac{N_1}{N} \quad y = \frac{N_2^4}{N} \quad z = \frac{N_5^n}{N} \quad (3.20)$$

No entanto, se  $x + y + z = 1$  apenas duas coordenadas são suficientes para criar o diagrama. Assim, o diagrama do *tristimulus* pode representar as propriedades tanto de sons sustentados como de ataques que variam no tempo. Quando os valores de  $x$  e  $y$  são baixos, há uma forte presença da fundamental, já um alto valor de  $y$  indica forte presença de parciais de média frequência, enquanto um valor alto de  $x$  significa forte presença de parciais de alta frequência.

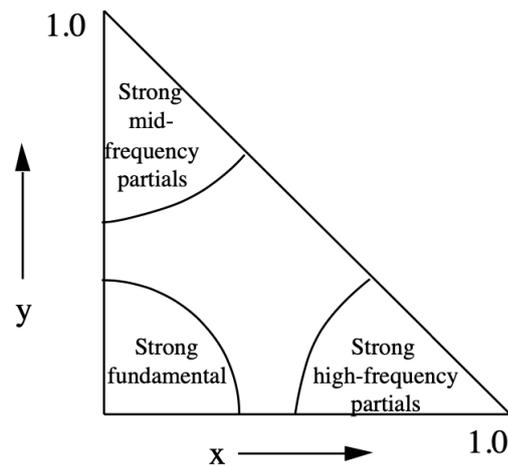


Figura 3.11: Diagrama com os três valores do *tristimulus* (YSTAD, 1998, p. 80)

## 3.4 Descritores de alto-nível

### 3.4.1 Estimativa da frequência fundamental

As técnicas de extração da frequência fundamental ( $f_0$ ), rastreamento de alturas (*pitch tracking*), e rastreamento de parciais (*partials tracking*) são utilizadas em larga escala. Há uma grande variedade de métodos de implementação desses algoritmos. A fim de ilustrar alguns desses métodos destacaremos aqui o [fiddle~] (PUCKETTE; APEL; ZICARELLI, 1998), [sigmund~] — ambos desenvolvidos por Miller Puckette para *Pure Data* e Max/MSP —, e *yin* (DE CHEVEIGNÉ; KAWAHARA, 2002).

O [fiddle~] realiza uma estimativa da altura e da amplitude de um sinal de entrada retornando opcionalmente uma lista de parciais que são usadas na determinação da frequência fundamental. A implementação deste algoritmo parte das ideias de Noll (1969) e Rabiner e Schafer (1978) e consiste primeiramente na obtenção das parciais e de sua amplitude. Para isso o sinal é dividido em  $N$  amostras, sendo  $N$  uma potência de dois — geralmente entre 256 e 2048. O sinal é analisado a cada  $\frac{N}{2}$  amostras, para cada análise é obtida uma DFT com janela retangular. Em seguida é feita uma convolução no domínio da frequência para obter uma DFT com janela *Hanning*:

$$x_H[k] = \frac{x[k]}{2} - \left( \frac{x[k+2] + x[k-2]}{4} \right) \quad (3.21)$$

Sendo  $x[k]$  a DFT com janela retangular. Então busca-se por picos de frequência no espectro sem janelamento utilizando um *phase vocoder*:

$$\omega = \frac{\Pi}{N} \left( k + re \left[ \frac{x[k-2] - x[k+2]}{2x[k] - x[k-2] - x[k+2]} \right] \right) \quad (3.22)$$

Já para estimar a frequência fundamental, é utilizado um esquema de “probabilidade”. A função de “probabilidade” é uma função não negativa  $\iota(f)$  sendo  $f$  a frequência. A presença de picos nos parciais múltiplos de  $f$  ou ao seu redor aumenta  $\iota(f)$  de tal maneira que depende dos picos de amplitude e frequência como mostra a equação 3.23:

$$\iota(f) = \sum_{i=0}^k a_i k_i n_i \quad (3.23)$$

Onde  $k$  é o número de picos no espectro;  $a_i$  é um fator que depende do *inésimo* pico;  $t_i$  depende do quão perto o *inésimo* pico está dos múltiplos de  $f$ , e  $n_i$  depende se o pico está mais perto dos múltiplos mais baixos ou mais altos.

Para a estimativa de sons monofônicos o [fiddle~] retorna o valor de  $f$  que apresenta a maior probabilidade. Já para sons polifônicos o algoritmo pega sucessivamente os valores de  $f$  com maior probabilidade que não são múltiplos nem submúltiplos do anterior (PUCKETTE; APEL; ZICARELLI, 1998, p. 2).

O [sigmund~] possui as mesmas funções do [fiddle~]. O algoritmo recebe um sinal de áudio de entrada, realiza uma transformada de *Fourier*, retornando a altura do som analisado, sua envoltória de amplitude e os parciais mais proeminentes. Puckette sugere o uso do [sigmund~] como alternativa ao [fiddle~].

Uma abordagem diferente no processo de estimativa da frequência fundamental é utilizado pelo algoritmo *yin*<sup>4</sup>, desenvolvido por Cheveigné e Kawahara (2002). O algoritmo é implementado no domínio do tempo a partir de um modelo de sinal periódico que pode ser estendido para trabalhar com várias formas de aperiodicidade que ocorrem em situações particulares, e utiliza o método de *auto-correlação* com uma série de modificações que se combinam com o intuito minimizar erros, tais como a identificação de picos de sub-harmônicos como candidatos a frequência fundamental.

A auto-correlação – responsável por identificar picos através da variação dos atrasos –, pode selecionar erroneamente picos quando o valor de  $\tau$  utilizado é próximo a 0. Visando evitar esses erros é utilizada a diferença entre elementos (DALTER, 2016, p. 26), como apresentado na equação 3.24.

$$d_t(\tau) = \sum_{j=1}^W (x[j] - x[j + \tau])^2 \quad (3.24)$$

Quando  $\tau = 0$  a função diferença apresenta valor zero, portanto, utiliza-se a média normalizada da função, como mostra a equação 3.25.

$$d'_t(\tau) = \begin{cases} 1, & \text{se } \tau = 0 \\ d_t(\tau) / [(1/\tau) \sum_{j=1}^{\tau} d_t(j)], & \text{se } \tau \neq 0 \end{cases} \quad (3.25)$$

<sup>4</sup>O algoritmo *yin* foi implementado como objeto externo para Max/MSP e está disponível como parte da biblioteca *Max Sound Box* <https://forum.ircam.fr/projects/detail/max-sound-box/>

O próximo passo é definir um limite absoluto e escolher o menor valor de  $\tau$  que forneça um mínimo de  $d'$  dentro desse limite. Caso nenhum for encontrado, o mínimo global será escolhido. O mínimo selecionado é considerado uma estimativa do período do sinal. Este período é refinado utilizando uma interpolação parabólica e finalmente convertido na frequência fundamental correspondente.

### 3.4.2 *Detector de onset*

O *detector de onset* consiste em um algoritmo responsável por encontrar início de notas ou eventos musicais em um sinal de áudio. A detecção automática de eventos musicais é extensivamente utilizada no contexto dos *sistemas interativos*, na classificação e indexação de músicas em bases de dados, e na análise e processamento de sinal sonoro. Há uma diversidade de abordagens na implementação de *detectores de onsets*, contudo, o método mais utilizado é aquele que busca por mudanças bruscas na envoltória de energia, e no comportamento do espectro nas regiões de alta frequência. (BRENT, 2011, p. 1). Bello et al. (2005) descreve algumas dessas implementações e fornece uma visão geral das abordagens e métodos de *detecção de onsets*.

Uma vez que sinais musicais geralmente são compostos por uma sobreposição de sons, a identificação de tais comportamentos do espectro requer um pré-processamento do sinal, a fim de melhor prepará-lo para a localização dos *onsets*. Este pré-processamento implica na transformação do sinal original para acentuar ou atenuar alguns aspectos que não são relevantes para a detecção dos *onsets*. Assim, a partir do sinal de áudio original, uma função de detecção é obtida, e em seguida um algoritmo de seleção de picos é aplicado para localizar os *onsets* (BELLO et al., 2005, p. 1036). A figura 3.12 mostra um esquema geral de implementação padrão de *detectores de onset*.

Usar informações relacionadas à envoltória de energia permite que os algoritmos detectem a maioria dos *onsets* de forma confiável. Embora funcione em sons percussivos, quando aplicados em instrumentos como violino, violoncelo, clarinete, ou flauta, por exemplo, esse método pode estar suscetível às variações de energia decorrentes do vibrato e do tremolo, e também às particularidades do comportamento sonoro de cada um deles — uma vez que estes instrumentos podem produzir ataques longos com um lento aumento de energia.

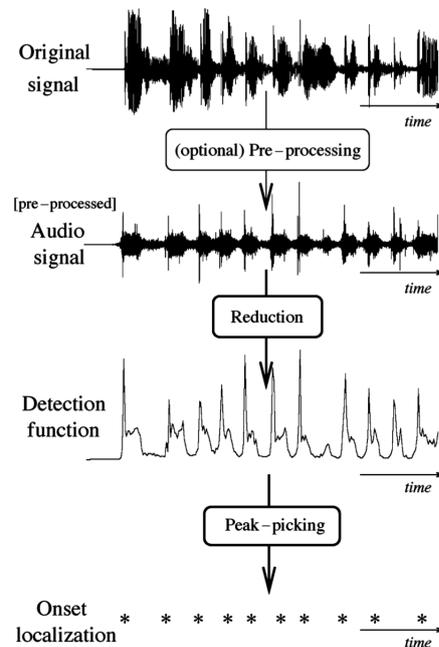


Figura 3.12: Esquema geral de implementação de *detectores de onset* (BELLO et al., 2005, p. 1036)

Existem alguns trabalhos que focam em desenvolver *detectores de onsets* que consideram as flutuações causadas pela presença de vibratos ou tremolos. Collins (2005b) implementa uma etapa de supressão de vibrato em seu método de detecção de *onsets* baseado em alturas. Primeiramente são identificadas regiões do sinal de áudio com presença de vibrato de no máximo um semitom em torno da frequência fundamental. A região é expandida gradualmente no tempo para cobrir toda a duração do vibrato. Este método é baseado no conceito de alturas estáveis e utiliza a mudança de frequência fundamental como indicador novos *onsets*.

Schleusing, B. Zhang e Y. Wang (2008) desenha um sistema baseado na correlação inversa de  $N$  *frames* espectrais consecutivos. Regiões com altura estável levam a valores de correlação inversa baixos, e mudanças de altura resultam em picos na função de detecção. Para suprimir o vibrato é empregado uma compensação que cancela pequenas mudanças de altura dentro de uma certa janela sem alterar os *onsets* reais.

Já Böck e Widmer (2013a) propuseram um novo método de supressão de vibrato e tremolo para detecção de *onset* baseado no algoritmo *SuperFlux* (BÖCK; WIDMER, 2013b), uma versão adaptada do *fluxo espectral*, que é menos sensível às variações de frequência causadas pelo vibrato. O método ainda adiciona uma técnica de ponderação baseada em atraso de grupo local (Local group delay – LGD) à etapa de cálculo, o que torna o algoritmo ainda mais eficaz

contra variações de energia em sons estáveis, dessa forma reduzindo em até 60 por cento o número de detecções de falsos positivos.

A figura 3.13 mostra uma análise comparativa dos métodos de detecção de *onset* baseados no *fluxo espectral* e no *Superflux*. A análise foi realizada a partir do violino executando uma nota longa com muito vibrato. Comparando as duas análises, vemos que o método que utiliza o *Superflux* (azul) possui uma maior precisão na detecção dos *onsets*, uma vez que não sofre uma grande interferência decorrente da variação de energia característica do vibrato. Por outro lado, o método baseado no *fluxo espectral* (preto) está mais suscetível aos efeitos do vibrato, assim retornando um maior número de falsos positivos.

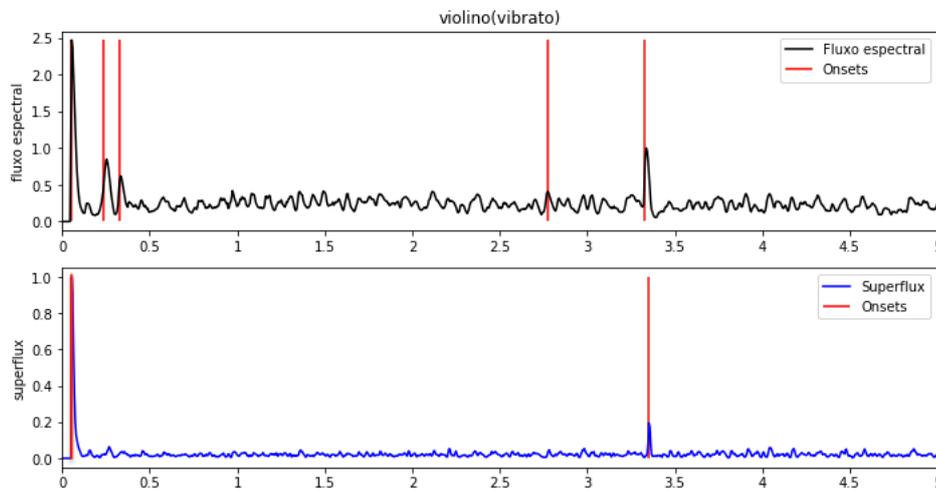


Figura 3.13: Análise comparativa de métodos de detecção de *onsets* (vermelho) a partir do *fluxo espectral* (superior) e do *superflux* (inferior)

### 3.4.3 Chroma

*Chroma* ou *Pitch Class Profile* (PCP) fornece o grau de presença de cada uma das classes de alturas da escala cromática. Este descritor é concebido a partir de um modelo psicoacústico baseado no trabalho de Shepard (SHEPARD, 1982) que apoia-se em dois aspectos relacionados ao reconhecimento de hierarquias de alturas: (1) a configuração rotacional denominada de *chroma* ou *Pitch Class*, que trata-se da equivalência da percepção das alturas independente do seu registro ou oitava; (2) a posição vertical do índice do *chroma* que Shepard denomina por altura ou *height*, representada por uma espiral (SIMURRA, 2016, p. 56) (figura 3.14)

O cálculo do *chroma* foi introduzido originalmente por Fujishima (FUJISHIMA, 1999), e é realizado através da comparação da energia espectral (magnitude) concentrada em



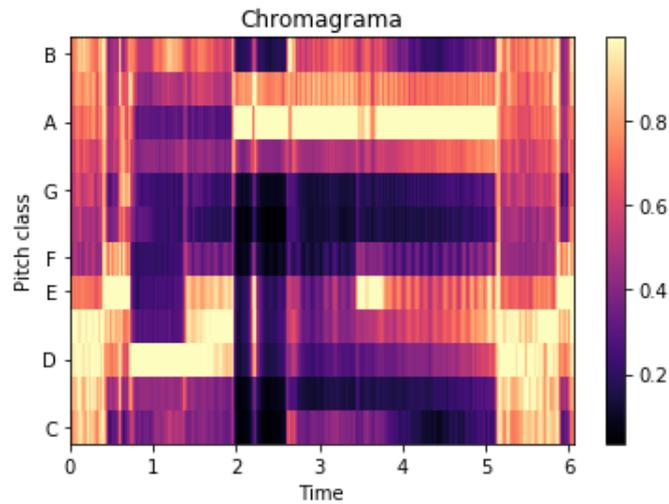


Figura 3.15: Chromagrama com as magnitudes de cada altura de um acorde arpejado no violão

batimentos acima de 20 Hz como flutuações de amplitude, é possível ter a sensação de “aspereza” causada pelas flutuações rápidas (PORRES, A. T., 2012). O estudo de (PLOMP; LEVELT, 1965) mostra que a sensação mais forte de aspereza entre dois sons puros acontece quando eles estão em um intervalo correspondente a um quarto da banda crítica. A figura 3.16 mostra a curva de *aspereza* entre dois sons puros em função da diferença de frequência medida em *bark*.

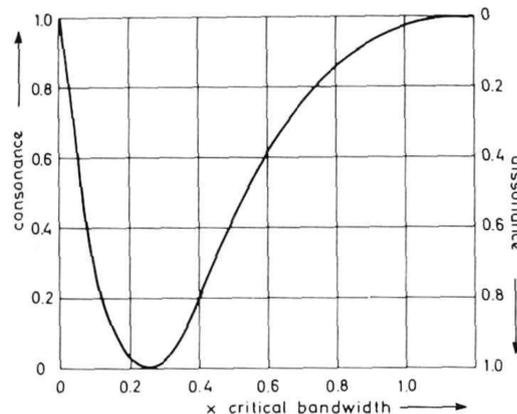


Figura 3.16: Curva de *aspereza* entre dois sons puros (PLOMP; LEVELT, 1965, p. 556)

A curva de *aspereza* pode ser calculada, segundo (BIGAND; PARNCUTT; LERDAHL, 1996a) pela equação:

$$g(x) = \left[ \left( e^{\frac{b}{0.25}} \right) \exp\left( \frac{-b}{0.25} \right) \right]^2 \quad (3.26)$$

Onde  $b$  é a diferença entre as frequências nas bandas *bark* e  $R = 0$  se  $b$  for  $>1.2$ .

A *aspereza* de sons complexos pode ser calculada somando o resultado de cada combinação de pares de parciais. Da mesma forma intervalos entre sons complexos também podem ser medidos a partir da soma dos valores de cada combinação de parciais de ambos os sons.

## Capítulo 4

# Estratégias de aplicação composicional na *live-electronics*

Nas últimas duas décadas, com a expansão e desenvolvimento das técnicas de ML e MIR, a busca por novas estratégias no emprego exploratório e criativo de tais ferramentas na composição de peças que envolvem instrumentos e processos de transformação, síntese e espacialização sonora em tempo-real vem expandindo e se diversificando. Como resultado desse crescente interesse pela exploração das possibilidades que tais métodos oferecem no campo da criação, tem-se uma vasta produção artística que engloba as mais variadas correntes estéticas, abordagens técnicas e estratégias composicionais.

Este capítulo abordará de forma geral as principais características da escrita musical no contexto do emprego de ferramentas e processos de ML e MIR na *live-electronics*. Isso se dará a partir de um estudo centrado na observação e na investigação de estratégias composicionais que vêm sendo exploradas por compositores desde as primeiras experimentações com escuta de máquina e recuperação de informações musicais. Assim, a partir do apontamento de aproximações criativas, buscamos circunscrever uma série de abordagens de forma a organizá-las em *paradigmas* composicionais, ao mesmo tempo em que procuramos ilustrá-las através de peças do repertório específico de nosso interesse.

Vale ressaltar que a organização em paradigmas aqui proposta não representa uma classificação absoluta das inúmeras possibilidades composicionais que o emprego dessas ferramentas podem oferecer. Há uma grande diversidade de abordagens e estratégias de aplicação das técnicas de ML e MIR no campo da música interativa/*live-electronics*, o que torna uma classificação exaustiva inviável tendo em vista o escopo deste trabalho. Não

pretendemos criar termos e conceitos extremamente delimitados, mas sim contribuir com reflexões teóricas acerca das práticas composicionais de modo a abarcar maneiras de pensar (concepções poéticas) e de realizar (arcabouços técnicos). Também é importante destacar que não se trata de um estudo com viés analítico. O objetivo é investigar e apontar estratégias e modos individuais de pensar o uso criativo e exploratório dessas ferramentas no contexto da criação. Portanto, não temos a pretensão de realizar uma análise exaustiva e detalhada das peças abordadas do ponto de vista formal/estrutural, .

O presente capítulo está estruturado da seguinte maneira: a primeira parte dedica-se a descrever as principais características da escrita musical no contexto do emprego da ML e MIR na *live-electronics*; e a segunda parte aborda a organização de certas estratégias em *paradigmas* composicionais a partir de exemplos de peças do repertório.

## **4.1 Características gerais da *live-electronics* no contexto das técnicas de ML e MIR**

Ao integrar as técnicas de ML e MIR às práticas musicais, o pensamento composicional no contexto da *live-electronics* ganhou uma nova dimensão no que diz respeito à complexidade e refinamento das estratégias de interação entre o ambiente acústico (instrumento/voz) e o ambiente eletroacústico (procedimentos de tratamento sonoro em tempo-real). Os diversos métodos de ML e MIR tornaram possível a criação de processos interativos altamente reativos, em que dados obtidos podem controlar ou influenciar mecanismos algorítmicos complexos de processamento, síntese e espacialização em tempo-real.

Os dados extraídos por meio das técnicas de ML/MIR podem ser correlacionados a atributos sonoros característicos, tais como brilho, inarmonicidade e rugosidade. Uma vez que há a possibilidade da correlação entre dados quantitativos e qualidades sonoras — que buscam descrever verbalmente as variadas características perceptíveis que o som pode apresentar —, a introdução dessas novas técnicas no contexto da criação abriu a possibilidade da concepção de ferramentas profícuas para se trabalhar com as diversas nuances timbrísticas e com as sonoridades não convencionais que a prática composicional atual geralmente incorpora ao vocabulário instrumental/vocal.

Com o advento da *live-electronics*, e sobretudo, com as inúmeras possibilidades de agenciamento dos processos interativos viabilizados pela incorporação das técnicas de

ML/MIR à prática criativa, procedimentos composicionais como manipulações de ordem estrutural; rítmica; melódica, e harmônica, e especialmente aqueles voltados à organização e transformação do material sonoro pela perspectiva do comportamento espectral, passaram a integrar a própria performance musical (CHADABE, 1984). Desse forma, tais procedimentos podem se adaptar às mais diferentes situações musicais, e conseqüentemente aproxima os mecanismos algorítmicos de tratamento sonoro a instrumentos musicais, instrumentistas virtuais, ou à agentes autônomos.

Uma vez que os processos interativos são geridos por dados e informações sonoras/musicais extraídas diretamente da performance musical, seu comportamento passa estar sujeito à uma certa abertura que se assemelha àquela apresentada pela relação notação-interpretação. De fato, a partitura não fornece toda a informação sonora que um evento musical produz. As alturas notadas em uma partitura de piano, por exemplo, podem ser interpretadas com precisão (valores absolutos), no entanto, a dinâmica se encontra dentro de regiões mais ou menos definidas (valores relativos), possibilitando uma certa flexibilidade na sua execução.

Da mesma forma, sabemos a natureza dos parâmetros que serão extraídos e processados, mas nem sempre o valor exato que eles receberão no momento em que serão recuperados. Parte desses parâmetros serão fixos e não mudarão de uma interpretação para outra, outra parte estará sujeita à eventos externos — desde a captação do sinal do instrumento até as particularidades na execução do interprete (MANOURY, 2008).<sup>1</sup>

Neste contexto, há uma relação direta entre aspectos sonoros/musicais específicos que estão em jogo em uma performance — como variações no andamento; timbre; estrutura rítmica; articulações, etc —, e as estruturas geradas pelos processos de transformação/síntese. A escrita na *live-electronics* no contexto dos processos de ML/MIR acaba lidando com estruturas sonoras e musicais emergentes, que se constroem a partir de uma inter-relação entre gestos e ideias notadas na partitura; nuances na interpretação e execução de cada *performer*; características intrínsecas dos instrumentos/vozes; e as transformações e gerações de sons pelos sistemas computacionais à eles acoplados. Logo, essas relações podem ser exploradas em diferentes níveis de interação a partir de dois fluxos básicos de operação: (1) o *performer* controla os processos eletrônicos e, (2) os processos eletrônicos afetam a performance do instrumentista (GARNETT, 2001, p. 24).

---

<sup>1</sup><http://www.philippemanoury.com/?p=319>

A paleta de possibilidades de agenciamento dos processos de interação, transformação, geração e espacialização sonora que o emprego das ferramentas de ML/MIR trazem para a prática da *live-electronics* são as mais variadas, diferenciando-se quanto à abordagem técnica, poética e musical. Assim, a implementação de processos de geração sonora como síntese concatenativa e *morphing*; detecção de eventos musicais complexos — tal como sons instrumentais ruidosos (MALT; JOURDAN, 2009) —, mapeamento, classificação e ordenação de timbres (BRENT, 2009a), e mapeamento de informações extraídas a partir de vibratos produzidos pela voz como controle de parâmetros de processos em tempo-real (EINARSSON; FRIBERG, 2015a) são algumas das estratégias de aplicação que vêm sendo explorados por compositores na *live-electronics* nas últimas décadas. Na próxima seção organizaremos algumas dessas estratégias em *paradigmas composicionais*, e discutiremos os aspectos técnicos e poéticos envolvidos no emprego dessas ferramentas a partir de exemplos de aplicação prática em peças do repertório.

## 4.2 Paradigmas composicionais associados ao emprego de técnicas de ML e MIR

Para uma compreensão procedimental geral das realizações técnicas, das concepções poéticas, e das abordagens composicionais que vêm sendo empregadas por compositores na exploração criativa do conjunto de ferramentas aqui estudado, buscamos circunscrever de maneira a organizar as diferentes formas de pensar sua aplicação em quatro grupos, a saber: (1) *paradigma do mapeamento*; (2) *paradigma do gatilho*; (3) *paradigma notacional*, e (4) *paradigma do procedimento*. Esses *paradigmas* podem ser identificados quanto ao modo de implementação da arquitetura dos algoritmos, à organização e tratamento dos dados extraídos, e à finalidade de uso dessas técnicas na concepção de processos computacionais os mais diversos no âmbito da *live-electronics*.

A identificação desses *paradigmas* não é, evidentemente, desvinculada de uma interpretação arbitrária e subjetiva que visa compreender diferentes abordagens poéticas quanto ao uso de ferramentas específicas na criação musical. Ao mesmo tempo, é relevante salientar que a noção de *paradigma* empregada neste trabalho diz respeito, justamente, à identificação de determinadas aproximações criativas quanto ao uso das ferramentas disponíveis à criação musical no contexto da *live-electronics*, objetivando menos a segregação

de estéticas ou escolas composicionais específicas do que a compreensão analítica de como se dá a inserção desse novo ferramental técnico em práticas poéticas heterogêneas.

Ressaltamos, assim, que tanto as peças abordadas neste texto quanto outras criações envolvendo o uso das ferramentas que nos são de interesse específico não devem ser compreendidas como representantes de um ou outro *paradigma* que lhes antecederia como modelo, categoria ou grupo estético. Pelo contrário, é a manifestação de determinadas particularidades de tais criações que torna possível identificar certas características específicas e abordagens que, por razões didáticas e analíticas, buscamos organizar a partir de *paradigmas composicionais*.

Neste contexto, o que se espera, é, por um lado que seja possível compreender como a inserção desse novo conjunto de técnicas e ferramentas vem transformando a prática e o pensamento criativo da atualidade. Por outro, que permita vislumbrar novas possibilidades de emprego dessas ferramentas, assim, abrindo a possibilidade de uma reflexão dialética sobre seu uso dentro do recorte da *live-electronics* na atualidade.

#### 4.2.1 Mapeamento

Identificamos como *paradigma do mapeamento* todo e qualquer processo que de alguma forma estabelece uma correlação entre interface de controle (informações de entrada) e processos/algoritmos de geração e transformação sonora. Essa correlação é realizada por meio de um sistema de mediação que é encarregado de definir os caminhos que o fluxo das informações de entrada percorrerão e as funções que irão adquirir dentro da arquitetura de um dado processo (WANG, S.; WANDERLEY; SCAVONE, 2020, p. 107). O mapeamento consiste na transferência de informações e na sua conversão em dados e parâmetros que irão controlar processos.

O mapeamento é um procedimento característico da *lutheria digital*, i.e., *design* de instrumentos digitais/expandidos e interfaces musicais digitais (*Digital Music Interfaces*)(MIRANDA; WANDERLEY, 2006; MONTEIRO, 2012; RIMOLDI, 2018). Se nos instrumentos acústicos a interface de execução está diretamente ligada ao mecanismo de produção sonora — a corda de um violino é parte integrante do mecanismo de controle, ao mesmo tempo em que está acoplada fisicamente ao gerador de som —, nos instrumentos eletrônicos a interface geralmente está separada do dispositivo gerador de som, portanto, a

relação entre eles necessita ser definida, o que ocorre através do processo de mapeamento (HUNT; DD et al., 2002, p. 1).

Segundo Miranda e Wanderley (2006) , uma vez que o processo de mapeamento é considerado uma etapa importante no *design* de um instrumento, há duas estratégias principais no que diz respeito a implementação de mecanismos de mapeamento: (1) uso de técnicas de aprendizado de máquina, análise de dados, e reconhecimento de padrões (LEE; WESSEL, 1992; FELLS; HINTON, 1995; MODLER, 2021); (2) estratégias pré-definidas de correlação entre dados e parâmetros (BOWLER et al., 1990; ROVAN; WANDERLEY; DUBNOV, 1997; HUNT; KIRK, 2000).

Enquanto a primeira utiliza lógicas circulares, as quais fornecem uma estratégia de mapeamento em que o próprio sistema se autorregula modificando-se dinamicamente, a segunda abordagem utiliza estratégias pré-definidas de mapeamento que estabelecem de forma direta as relações entre os dados de entrada e a sua parametrização.

Em geral os processos de mapeamento que são concebidos a partir de estratégias pré-definidas, as quais relacionam os parâmetros de um conjunto para o outro podem ser divididos em quatro categorias: (1) um dado de entrada controla apenas um parâmetro do processo (*one-to-one*); (2) um dado de entrada controla simultaneamente mais de um parâmetros do processo (*one-to-many*); (3) vários dados de entrada controlam simultaneamente um parâmetro específico do processo (*many-to-one*); (4) combinação de todas as categorias, na qual vários dados de entrada controlam mais de um parâmetro do processo (*many-to-many*)(IAZZETTA, 2000, p. 20).

A natureza dos processos de *design* de instrumentos musicais digitais implica na implementação de uma interface de controle, como teclas; cordas; chaves; sensores de captura de movimento, entre outros mecanismos e interfaces de controle gestuais que são acoplados à esses instrumento. É justamente o acionamento desses mecanismos que gera os dados de entrada que irão ser mapeados.

Uma alternativa aos controladores físicos como fonte de dados de controle é usar informações sonoras captadas de uma apresentação instrumental ao vivo (BULLOCK, 2008, p. 19). Isto é, as informações de entrada são geradas pelos processos de recuperação de informações musicais/sonoras através das técnicas de ML/MIR. Por exemplo, informações correlacionadas à atributos sonoros são mapeadas e traduzidas para parâmetros que por sua vez poderão controlar diferentes tipos de processos de síntese, transformação, espacialização

sonora, entre outros. Ou seja, o som passa a se apresentar como sua própria interface de controle, manipulando processos que irão transformar a si mesmo ou gerar novos sons.

A despeito do mapeamento estar fortemente ligado à *lutheria digital*, há uma crescente propensão em direção ao entrecruzamento entre a prática do desenvolvimento de instrumentos digitais e a criação musical com suporte tecnológico, principalmente no contexto da *música interativa* na atualidade. Neste sentido, no que diz respeito ao emprego de técnicas de ML e MIR aplicadas à prática composicional, sobretudo aquelas voltadas ao mapeamento dos dados recuperados através de tais técnicas, a implementação de um sistema de mediação acaba tornando-se parte integrante do processo criativo.

Sendo assim, descreveremos a seguir algumas aplicações práticas de processos de mapeamento ilustrando-as a partir de peças do repertório da *música interativa*. Buscaremos discutir pensamentos e estratégias composicionais, de forma a apontar a particularidade de cada compositor na concepção dos processos de mapeamento.

#### **a) Cortazar, ou quarto com caixa vazia (1999)**

Um exemplo de estratégia simples de mapeamento direto pode ser observada em *Cortazar, ou quarto com caixa vazia* (1999), para piano e *live-electronics*, de Silvio Ferraz. O título da peça, além de homenagear o escritor argentino Julio Cortázar, nos remete à imagem de um pianista, que trancado em um quarto com uma caixa vazia quebra as paredes e liberta o som do piano dos padrões musicais ocidentais. Para o compositor, a caixa vazia é uma metáfora para o ambiente de computação musical Max/MSP utilizado na concepção dos processos de transformação e espacialização sonora (PESTOVA, 2008, p. 26).

Nesta peça Ferraz utiliza três procedimentos diferentes de transformação e síntese sonora: *pitch shifter* com um controle randômico de rampas para criar oscilações microtonais nas transposições; síntese aditiva a partir de um conjunto de oito parciais que podem variar de acordo com oito *presets*, e um processo de modulação em anel.

Nosso interesse aqui é observar como Ferraz utiliza o próprio sinal captado do piano como uma espécie de interface dinâmica para controlar o processo de modulação em anel aplicado à síntese aditiva. Vale ressaltar que, embora o compositor tenha utilizado o sinal bruto do piano, tal como é captado pelo microfone sem nenhum tratamento específico para se extrair informações, ainda assim atributos que compõem e descrevem o comportamento físico

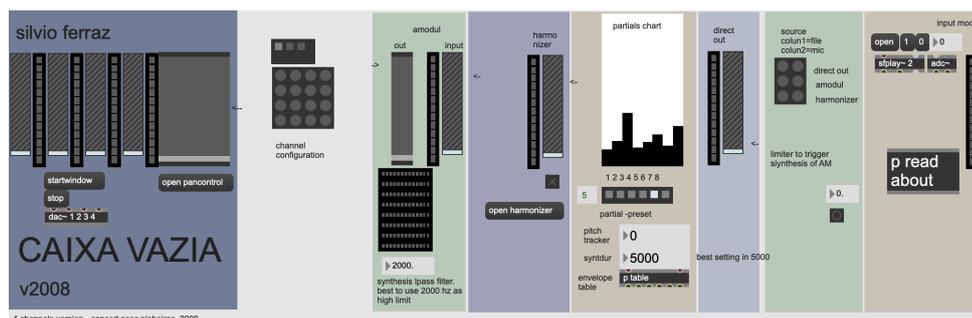


Figura 4.1: Patch em Max/MSP de Cortazar, ou quarto com caixa vazia

do sinal no domínio do tempo (envoltória de amplitude e forma de onda) são utilizados no mapeamento do processo em questão.

De acordo com Ferraz (1999), o processo de modulação em anel foi concebido tomando-se por base a criação de um módulo de síntese aditiva, que por sua vez obedece um sistema de expansão e contração de acordes e um determinado grau de inharmonicidade gerado a partir da multiplicação da frequência fundamental extraída em tempo-real do piano, por oito parciais pré estabelecidas — esse princípio harmônico também guia a escrita da parte instrumental. O resultado dessa síntese é então multiplicado pela envoltória de amplitude do sinal do piano, resultando em uma sonoridade com um certo grau de inharmonicidade e distorção do espectro.

A modulação em anel é uma técnica de processamento de sinais herdada da área da telecomunicações na qual a amplitude de um determinado sinal ou frequência (portadora) varia de acordo com a oscilação de outra frequência (moduladora). Como resultado as frequências das duas entradas se somam e se subtraem. Por exemplo, se considerarmos duas ondas senoidais puras, uma portadora e a outra moduladora de 500 Hz e 100 Hz, teremos como resultado a sobreposição de duas ondas, 600 Hz (adição) e 400 Hz (subtração).

Logo, a medida em que o sinal de qualquer uma das duas entradas se torna mais complexo tem-se como resultado final uma maior sobreposição de frequências, uma vez que o processo de adição e subtração passa a se aplicar a cada uma das componentes do sinal. Por exemplo, uma nota de piano possui um espectro com inúmeras componentes frequenciais que apresentam comportamentos, em certa medida, imprevisíveis e independentes. Portanto, cada uma delas passará pelo processo de adição e subtração (EMMERSON, S., 1977, p. 14), gerando assim, sobreposições. Esse processo é um recurso bastante utilizado na música eletrônica em geral, sobretudo na *live-electronics*, uma vez que permite a contração e a expansão intervalar de alturas e a elaboração de colorações timbrísticas diversas.

Em *Cortazar, ou quarto com caixa vazia*, o próprio sinal digital do piano é mapeado diretamente e utilizado como parâmetro de controle para modular a amplitude do sinal de saída da síntese aditiva. Um detector de ataque é responsável por acionar o processo de modulação. Com isso, o compositor estabelece uma correlação direta entre o sinal do piano (dados de entrada), e a frequência moduladora (parâmetro de controle) do processo de modulação em anel em uma relação *one-to-one*.

De fato, com esse processo de mapeamento direto Ferraz cria um mecanismo automático e dinâmico de controle, no qual o som torna-se a sua própria interface de controle, manipulando processos que irão engendrar novas estruturas sonoras. Desse modo, ao passo que a amplitude do ataque do piano aumenta, maior é a curvatura espectral aplicada à síntese aditiva, portanto, resultando em uma sonoridade mais inarmônica.

A figura 4.2 mostra os primeiros compassos da peça, onde tal procedimento pode ser observado. A princípio, apenas o som direto do piano pode ser ouvido. Com o aumento da presença de ataques e acordes acentuados na parte instrumental, o processo de modulação torna-se gradualmente mais perceptível.

Figura 4.2: Primeiros compassos de *Cortazar, ou quarto com caixa vazia*

### b) *Proceratophrys boiei* (2016/2019)

Em *Proceratophrys boiei* (2016/2019)<sup>2</sup>, para saxofone tenor e *live-electronics*, de Danilo Rossetti, encontramos outro exemplo de utilização de atributos sonoros recuperados do sinal digital e aplicados ao *paradigma do mapeamento*. Nesta peça, o compositor utiliza o nível de

<sup>2</sup>A partitura de *Proceratophrys boiei* foi publicada Funarte (2021), e pode ser encontrada online em: <https://www.funarte.gov.br/partituras-brasileiras-online/>

amplitude extraído diretamente do sinal do instrumento como parâmetro que controlará o tamanho dos grãos (em milissegundos), de um processo de síntese granular.

Para isso, é utilizado um módulo em Max/MSP, criado pelo compositor Alain Bonardi ([niveauinstrument]), que é responsável por extrair os picos de amplitude do sinal digital do instrumento através do objeto [peakamp~]<sup>3</sup>, e mapeá-los para o processo de síntese granular. Uma vez que a amplitude do sinal é extraída, calcula-se a raiz quadrada a fim de obter a energia instântanea, que em seguida é convertida para decibél e então escalonada de maneira que apresente valores entre 50 e 400 – valores que determinarão o tamanho dos grãos em milissegundos. A figura 4.3 mostra a implementação dos módulos de extração da amplitude e do processo de síntese granular.

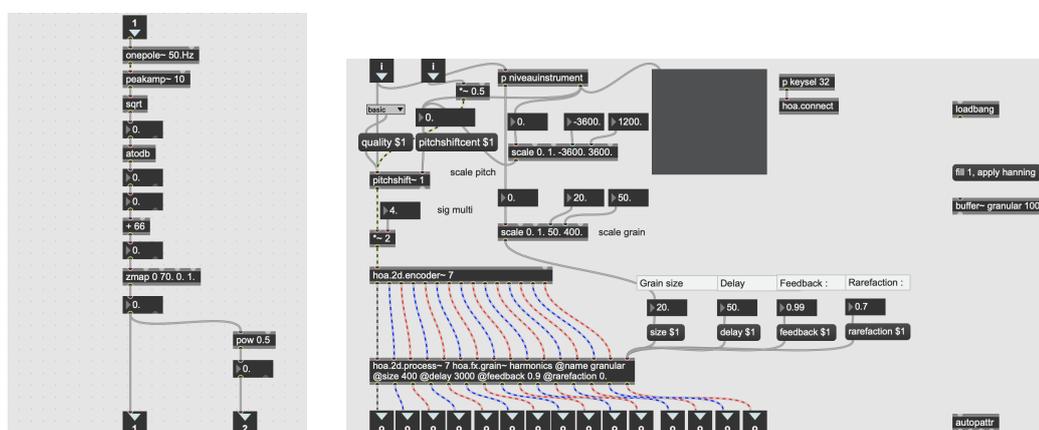


Figura 4.3: Módulo responsável pela extração da amplitude do sinal de entrada (esquerda) e *patch* do processo de síntese granular (direita) de *Proceratophrys boiei*

Com essa estratégia, Rossetti cria uma correlação direta entre a amplitude do instrumento e a densidade da textura que emergirá a partir do processo de síntese granular. Isto é, há uma proporcionalidade entre a dinâmica executada pelo saxofonista e o parâmetro que controla os tamanhos dos grãos.

De fato, ao passo em que o instrumentista executa eventos e gestos musicais com baixa dinâmica – e.g., compasso 67 (ver figura 4.4) – os grão apresentarão tamanhos pequenos, portanto gerando uma textura densa. Por outro lado, havendo a presença de gestos instrumentais com alta dinâmica – e.g., compasso 78 – o tamanho que os grão apresentarão será maior, consequentemente, dando origem à uma textura mais rarefeita. Essa estratégia pode ser observada a partir do compasso 66 da peça (figura 4.4).

<sup>3</sup>Objeto de Max/MSP que retorna a amplitude máxima do sinal analisado

The image displays a musical score for the final section of *Proceratophrys boie*, consisting of three staves of music. The first staff begins at measure 65 and includes a box labeled '9 parar pvoc' and a tempo marking '♩ = 70'. It features dynamic markings such as *mf*, *f*, *ff*, *mp*, *mf*, *p*, *mp*, *mf*, *mp*, *mf*, and *mf*. Annotations include 'frullato', 'v. aumentando', 'v. freq.', and 'simile'. The second staff starts at measure 71 and includes a box labeled '10 conv. 1/grain 1'. It features dynamic markings *mf*, *f*, *mf*, *mf*, *f*, *mp*, and *mf*. Annotations include 'frullato', 'gliss.', 'v. som eólico', and 'bisbigliando'. The third staff starts at measure 75 and includes a box labeled '11 conv. 1/grain 2'. It features dynamic markings *mp*, *mf*, *mp*, *f*, *mp*, *mf*, and *f*. Annotations include 'frullato', 'slap tongue', and 'molto vibrato'. The score includes various time signatures and complex rhythmic patterns.

Figura 4.4: Trecho da seção final de *Proceratophrys boie*

A voz é um dos instrumentos mais versáteis, e apresenta uma grande variedade de características e particularidades. Devido à sua grande flexibilidade articulatória e, conseqüentemente, à sua grande variabilidade timbrística, a recuperação de atributos a partir da sua vasta gama de expressividade sonora pode se mostrar um terreno fértil no que diz respeito ao emprego criativo das técnicas de ML/MIR. Logo, isso torna o uso de seus atributos como dados de controle de processos uma estratégia com grande potencial criativo.

O processo de mapeamento associado aos atributos sonoros recuperados a partir da voz pode oferecer um conjunto amplo de possibilidades e estratégias, ao passo em que sonoridades complexas, e procedimentos mais elaborados de extração de informações sonoras podem vir a gerar uma grande variedade de dados. Uma vez que tais dados são recuperados, mapeados e associados à controles de parâmetros de processos, a expressividade musical e sonora inerente à voz é potencializada.

Neste contexto, abordaremos algumas estratégias de aplicação de técnicas de ML/ MIR à voz cantada e, conseqüentemente, discutiremos procedimentos de mapeamento de seus atributos sonoros como controle de processos de transformação e geração sonora. Isso se dará a partir de exemplos práticos observados em duas peças: (1) *Ps. I will be home soon!* (2012) para quatro vozes, *ensemble* e *live-electronics*, de Anna Einarsson; (2) *Da una crepa* (2011-2012) para coro, solista, *ensemble* e *live-electronics*, de Marta Gentilucci.

c) *Ps. I will be home soon!*

*Ps. I will be home soon!* (2012), é uma ópera de câmara para seis instrumentos, quatro vozes e *live-electronics*, composta pela compositora Anna Einarsson como encomenda da companhia de ópera *Malmö Opera* na Suécia. O libreto foi escrito por Maria Sundqvist, e se baseia na vida do marinheiro suéco, Calle Pettersson abordada no livro *Kung Kalle av Kurrekurreduddön* (REGIUS; LANGER, 2002).

A narrativa e a forma de *Ps. I will be home soon!* não é linear, sua estrutura pode ser montada de diversas maneiras, dessa forma, o foco está na experiência individual de cada ouvinte. O público é dividido em dois grupos, cada grupo é guiado através das cenas por um ator (a figura 4.5 mostra um rascunho da compositora com a trajetória de cada grupo). Uma vez que não há cadeiras, os ouvintes podem se mover livremente pelas cenas ou sentar-se como desejarem. Entretanto, segundo a compositora, não se trata de um teatro musical, e nem de uma instalação, na qual o público entra e sai. O fato de a obra ter sido realizada dentro de uma companhia de ópera enfatiza que há um discurso musical na peça (EINARSSON, 2015).

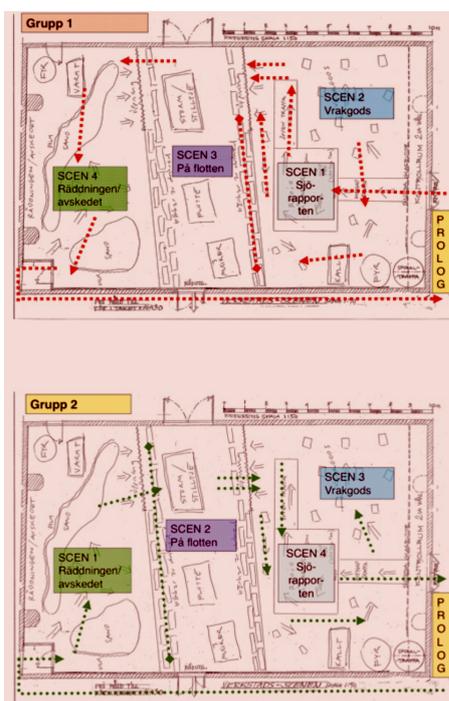


Figura 4.5: Rascunho da trajetória de cada grupo de ouvintes em *Ps. I will be home soon!* (EINARSSON, 2017, p. 59)

No que diz respeito aos aspectos técnicos e musicais, Anna Einarsson buscou desenvolver ferramentas e procedimentos visando a extração de informações musicais/espectrais da voz cantada, bem como mapear os dados e informações recuperadas

para controlar os processos eletrônicos da peça. De acordo com a compositora, a idéia desde o início do processo composicional era desenvolver mecanismos de controle dinâmico, sem o uso de interfaces ou sensores acoplados aos *performers* (EINARSSON, 2017, p. 58).

Em colaboração com o pesquisador Anders Friberg, Einarsson implementou em *Ps. I will be home soon!*, um método de extração de atributos relacionados à extensão intervalar de vibratos produzidos pela voz cantada. Tal método é baseado no algoritmo CUEX (*Cue-extractor*) desenvolvido em *matlab* e posteriormente em *Pure Data*, por Friberg (FRIBERG; SCHOONDERWALDT; JUSLIN, 2007). As informações extraídas da voz são mapeadas e utilizadas para controlar quatro sintetizadores polifônicos, cada um contendo filtros de ressonância e bancos de osciladores, os quais são excitados por ruído (EINARSSON; FRIBERG, 2015b, p. 19).

O vibrato produzido pela voz cantada é um elemento técnico importante do ponto de vista expressivo da interpretação musical. Basicamente consiste em uma modulação da frequência fundamental que é caracterizada por uma taxa de oscilação e por uma extensão intervalar entre as frequências. Esta modulação, por sua vez, também gera modulações de amplitude, que pode estar em fase e/ou fora de fase em relação à modulação de frequência (HORII, 1989; PRAME, 1998).

O processo de extração consiste em cinco passos: (1) extração da frequência fundamental e mensuração da amplitude do sinal, através do algoritmo *yin* (DE CHEVEIGNÉ; KAWAHARA, 2002); (2) implementação de um *gate*, utilizando um determinado *range* de alturas e um limite de amplitude. Somente as alturas que estiverem dentro de tais limites são direcionadas para análise; (3) o sinal é dividido em duas partes, sendo a primeira parte filtrada por um filtro passa-baixa, e a segunda filtrada por um filtro médio com uma janela correspondente à um ciclo de vibrato — considerando uma taxa de vibrato de 6 Hz; (4) as duas partes são subtraídas de maneira que o sinal resultante será zero quando não houver vibrato; (5) é calculado a média do valor absoluto do sinal através de um filtro médio com o dobro do tamanho de um ciclo de vibrato, assim, o sinal resultante será proporcional à extensão do vibrato dado em *cents* (FRIBERG; SCHOONDERWALDT; JUSLIN, 2007, p. 413), (EINARSSON; FRIBERG, 2015b, p. 19).

Em *Ps. I will be home soon!* o processo de mapeamento aborda a noção de *mapeamento metafórico*, tal qual descrito por Mark Johnson (2013) . Logo, o que se transfere não são os atributos, mas as relações entre os domínios da fonte (origem) e do alvo (destino). Segundo

(EINARSSON; FRIBERG, 2015b), nesta peça, há uma relação de forças empregada no método de mapeamento. A densidade expressa a força no domínio do alvo, que representa uma estrutura dinâmica de agrupamentos de alturas, e o vibrato é a expressão de força no domínio da fonte, ou seja, a voz cantada. Dessa forma, quanto maior é a presença de vibrato, mais denso será o agrupamento de alturas resultantes do processo de síntese.

#### **d) *Da una crepa***

Outra peça que ilustra o emprego de atributos sonoros extraídos da voz para controlar parâmetro de processos, é *Da una crepa* (2011-2012), para coro, soprano, *ensemble* e *live-electronics*, de Marta Gentilucci. Composta a partir de um poema homônimo, da escritora italiana Elisa Biagin, esta peça contou com a colaboração técnica de Mikhail Malt na concepção dos processos interativos, e foi estreada em 2012 no *Centre Pompidou*, em Paris.

Alguns dos aspectos composicionais centrais de *Da una crepa*, surgem do desejo de Gentilucci em explorar estruturas rítmicas, tanto no nível macro – controlar e desenvolver estruturas rítmicas complexas ao longo da composição –, quanto no nível micro – relações entre as estrutura temporais internas do som, e a possibilidade do uso dessas estruturas como interface de controle para gerar e/ou manipular outros sons.

Neste contexto, surge a ideia de correlacionar as estruturas rítmicas geradas por *tremolos* de vogais extraídos da voz cantada – i.e., a alternância repetida de vogais sobre uma mesma nota sustentada, por exemplo, *a-e-a-e* –, a parâmetros de processos que pudessem manipular alturas. A alternância de vogais é aplicada a um determinado som com o intuito de controlar sua variação de transposição. Com isso, a compositora pretendia conceber um processo que pudesse modular a altura de uma nota em *tenuto* (sustentada) através da estrutura rítmica proveniente da articulação da voz (MALT; GENTILUCCI, 2015, p. 3).

Para isso, partindo de uma análise prévia de amostras sonoras vocais a partir de cinquenta e dois descritores, Gentilucci avaliou qual descritor apresentou melhor resultado na detecção das oscilações para cada combinação de vogal (*i-e-i-e*, *a-i-a-i*, etc.) e para cada tipo de voz (*soprano*, *mezzo-soprano*, *tenor*, *barítono* e *baixo*). Dessa forma, a compositora chegou em uma combinação de quatro descritores, a saber: *spectral deviation*; *odd to even*; *tristimulus*, e *total energy*. Tal processo de avaliação dos descritores foi realizado por meio do método PCA (*Principal Component Analysis*), e é descrito com detalhes em (MALT; GENTILUCCI, 2015).

The image shows a musical score for the piece 'Da una crepa', specifically measures 122 to 130. The score is written for five vocal parts: Bass (Bs), Male (Ma), Soprano (S), Tenor (T), and Baritone (Br), along with a Solo line. The vocal parts feature complex rhythmic patterns, including tremolos, and are marked with various dynamics such as *mf*, *mp*, and *mf*. The Solo line is marked with *mf* and *mf*. The score is written in a standard musical notation with a treble clef and a common time signature.

Figura 4.6: Trecho com *tremolos* de vogais referente aos compassos 122 à 130 de *Da una crepa*

Assim, o processo de extração e mapeamento dos atributos empregados em *Da una crepa* é realizado em três etapas: ao longo da peça trechos específicos são gravados em *buffers*, esses trechos passam pela análise dos descritores e os resultados são armazenados no objeto [coll]<sup>4</sup>. Por fim os dados armazenados são mapeados e enviados como parâmetros que controlarão a transposição de dois módulos de *pitch shifter* e um *harmonizer*.

Entretando, nem todos os dados extraídos são armazenados no objeto [coll], há um *gate* implementado a partir do descritor *total energy*, com um *threshold* padrão de 0.005. Desse modo, os dados extraídos são enviados ao objeto [coll] somente se a energia total do sinal analisado for maior ou igual ao valor do *threshold*. A figura 4.7 mostra o módulo de extração e armazenamento dos atributos.

#### 4.2.2 Gatilho (*Trigger*)

O *paradigma do gatilho* ou *paradigma do trigger* está ligado ao emprego das técnicas de ML/MIR na concepção de sistemas e algoritmos voltados à detecção de eventos sonoros/musicais e o disparo de processos. Esta estratégia lida com procedimentos de natureza condicional, i.e. os algoritmos analisam o sinal digital buscando por características específicas e pontuais que podem descrever determinados eventos, como *pizzicatos*; variação de dinâmica; *trillo*; *vibrato*, entre outros, e somente responderão caso tais eventos forem detectados.

Ao contrário do *paradigma do mapeamento*, esta estratégia não tem como objetivo final o controle paramétrico de processos, tampouco implica no registro de eventos sob um suporte

<sup>4</sup>Objeto de Max/MSP que permite o armazenamento, organização, edição e recuperação de mensagens

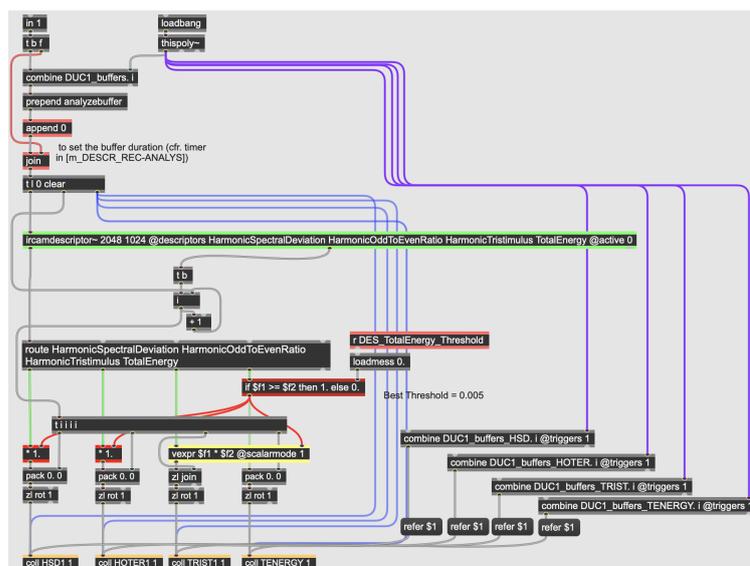


Figura 4.7: Módulo de análise, extração e armazenamento de atributos sonoros do *patch* em Max/MSP de *Da una crepa*

para que possam ser detectados — embora se assemelhe, em certa medida, ao emprego do *paradigma notacional* (abordado em 4.2.3), no qual o computador rastreia a performance e reage à presença de certos eventos prescritos em uma partitura ou roteiro de ações. O emprego do *paradigma do gatilho* está relacionado ao disparo de processos condicionado pela detecção automática de eventos específicos.

O exemplo mais comum de implementação de algoritmos tendo em perspectiva o emprego do *paradigma do gatilho* são aqueles concebidos a partir de métodos de rastreamento da envoltória de energia (*envelope follower*), tais como detectores de ataque, *onset*, e *offset* — i.e., algoritmos voltados à detecção de ataque, início e fim de uma nota (BELLO et al., 2005; COLLINS, 2005a).

Entretanto, a escrita musical atual envolve o uso de técnicas instrumentais estendidas e/ou o emprego de instrumentos expandidos, logo, suscitando uma ampla paleta de sonoridades que são incorporadas à prática musical atual. Evidentemente, muitos dos eventos musicais constituídos por essas sonoridades podem, eventualmente, apresentar um comportamento bastante complexo do ponto de vista espectral, o que nos coloca em uma situação na qual o desafio já não é mais apenas a detecção de envoltórias de energia, mas a mensuração de um conjunto de atributos sonoros que, correlacionados, descrevem um determinado evento sonoro.

Neste sentido, Collins (2006) e Malt e Jourdan (2009) descrevem a implementação de alguns métodos de detecção de eventos sonoros/musicais a partir de um conjunto de

descritores de baixo e alto nível. Equanto Collins implementa processos de análise e detecção de eventos sonoros/musicais baseados em modelos psicoacústicos para o desenvolvimento de *sistemas musicais interativos autônomos*, Malt e Jourdan criam funções de detecção de eventos sonoros complexos, a partir de descritores de baixo nível para aplicação no contexto das técnicas estendidas na *live-electronics*.

O *paradigma do gatilho* é amplamente utilizado dentro do escopo da *live-electronics*, assim como na concepção de *sistemas musicais interativos* ou como parte da arquitetura de agentes autônomos, enquanto estratégia direta de interação entre a máquina e o *performer*. Ao “ouvir” o instrumentista e detectar a presença de eventos pré-definidos, o computador dispara processos, estabelecendo assim, uma relação direta entre as ações gestuais instrumentais e a resposta da máquina à estes impulsos.

A seguir demonstraremos alguns exemplos de aplicação composicional do *paradigma do gatilho* em duas peças do repertório da *live-electronics* compostas nas últimas décadas: (1) *Deconstructing Dowland* (2009), para violão e *live-electronics*, de Natasha Barrett, e (2) *Tensio* (2010), para quarteto de cordas e *live-electronics*, de Philippe Manoury.

#### a) *Deconstructing Dowland*

*Deconstructing Dowland* (2009), para violão e *live-electronics*, de Natasha Barrett — dedicada ao violonista Stefan Östersjö —, ilustra de forma simples e didática a aplicação prática do *paradigma do gatilho*. Com o desejo de unir a música antiga com a prática musical atual, a compositora toma como ponto de partida, as três galhardas publicadas no primeiro livro de canções de John Dowland.

A desconstrução sugerida pelo título remete a processos de desmontagem; transgressão; distorção e ruptura. Logo, essas estratégias de manipulação de materiais, aliadas à extensa variedade de possibilidades sonoras que os procedimentos computacionais oferecem, atuam na criação de novos materiais musicais que, ao aludirem às três galhardas de John Dowland são ressignificados de dentro para fora ganhando um novo sentido em um novo contexto musical.

Em *Deconstructing Dowland*, o *paradigma do gatilho* é empregado através do uso de um detector de ataques para disparar de forma automática e precisa pequenas amostras sonoras concebidas em estúdio e que são armazenadas em *buffers*. Para isso, a compositora utiliza o objeto [bonk~], implementado por Miller Puckette (PUCKETTE; APEL; ZICARELLI, 1998) para *Pure Data* e *Max/MSP*.

O objeto [bonk~] analisa um sinal de entrada através de um banco de filtros limitado, e assim, detecta ataques, os quais podem ser comparados à um conjunto de modelos espectrais que por sua vez irão predizer quais dos possíveis tipos de ataques ocorreram. Esse objeto oferece a possibilidade de ser “treinado” com intuito de detectar ataques com determinadas características a partir de instrumentos específicos, e pode carregar modelos de ataques pré-definidos armazenados em um arquivos de texto.

Enquanto os métodos de detecção de ataques mais populares baseados em *envelope followers* procuram por aumentos repentinos de energia — o que os tornam sensíveis às várias mudanças que eventualmente a envoltória pode apresentar —, o método implementado pelo [bonk~] procura por mudanças relativamente bruscas que não afetam a envoltória de energia como um todo, evitando assim, a detecção de ataques indesejados (PUCKETTE; APEL; ZICARELLI, 1998, p. 3).

O processo de detecção é empregado no compasso 12 e no trecho que compreende do compasso 74 ao 76 de *Deconstructing Dowland* (ver figura 4.8). Nestes dois trechos, o objeto [bonk~] responde aos *pizzicatos bartok* (compasso 12), e aos ataques produzidos pelos harmônicos artificiais (compassos 74 ao 76), disparando as amostras sonoras pré-gravadas. Embora tais trechos possuam estruturas rítmicas relativamente simples, a utilização da estratégia de detecção automática de ataques resulta em uma sincronia mais precisa entre a ação do violonista e o disparo das amostras. Logo, proporcionando uma amálgama mais eficiente entre os timbres do instrumento e dos sons pré-gravados, de maneira que os sons disparados atuam como uma prolongação e expansão espectral da ressonância do violão.

The figure displays two musical staves. The top staff, labeled 'Gtr.', shows measures 11 and 74-76. Measure 11 includes a *pizzicato* section with *sul pont* and *p* dynamics. Measures 74-76 feature artificial harmonics with *sfz* dynamics and a *pull-off Blue-Tack* instruction. The bottom staff, labeled 'cmp', shows the corresponding computer triggers: T2 (Activate auto trigger pre-prep sound) at measure 11, and T29, T30, T31, and T32 (Manual trigger pre-prep sound, 5 seconds) at measures 74, 75, 76, and 77 respectively. A tempo marking of  $\text{♩} = 50$  is present at the start of the bottom staff.

Figura 4.8: Compassos 12 (superior) e 74 (inferior) de *Deconstructing Dowland*

## b) *Tensio* (2010)

Uma aplicação mais complexa do *paradigma do gatilho* pode ser observada em *Tensio* (2010), para quarteto de cordas e *live-electronics*, de Philippe Manoury. O título da peça remete à palavra “tensão”. Segundo o compositor, essa tensão é a tensão física das cordas dos instrumentos do quarteto que se proliferam pela parte eletrônica. “Me parecia benéfico recorrer àquela imagem primordial de uma corda tensionada entre dois pontos, e de fazê-la ser tocada de maneiras [régimes] extremas, que só a tecnologia nos permite entrever” (MANOURY, 2010)<sup>5</sup>.

Em *Tensio*, o compositor utiliza em conjunto uma combinação de estratégias e técnicas de ML/MIR na construção de mecanismos que trabalham paralelamente na detecção de diversos aspectos de eventos musicais específicos. Isso acontece na seção **Vc** da peça, onde há uma textura instrumental construída a partir de gestos compostos por sons rápidos em *sul ponticello*, e por sons com duração média tocados com o arco na posição natural. Com o objetivo de detectar separadamente tais eventos, visando o acionamento de um módulo de síntese aditiva e de um processo de espacialização, Manoury utiliza o objeto [sigmund~]<sup>6</sup> associado à extração do terceiro *tristimulus*, e ao cálculo da potência do sinal realizado pelo descritor *total energy*.

Desta forma, o [sigmund~] é configurado para reportar a frequência fundamental do sinal analisado acima de 56 rms, em um intervalo de 200 milisegundos. Quando detectado eventos com tais características, é acionado um processo de cálculo gerando um conjunto de frequências e amplitudes que irão alimentar um banco de osciladores, produzindo assim, um espectro sintético (descreveremos em detalhes o processo de cálculo desses espectros na seção 4.2.4.

Por outro lado, para detectar sons rápidos em *sul ponticello*, é utilizada uma combinação entre [sigmund~], terceiro *tristimulus*, e o cálculo da energia total do sinal. Enquanto o [sigmund~] é responsável pela detecção dos sons rápidos — intervalo de 7 milisegundos entre cada nota, —, o *tristimulus* e o cálculo da potência são responsáveis pela detecção do timbre gerado pela excitação da corda com o arco em *sul ponticello*. Uma vez que a potência do sinal for 1 — os valores da potência do sinal estão escalonadas de 0 à 1 —, é retornado o valor do terceiro *tristimulus*. Os eventos em *sul ponticello* são detectados quando o terceiro *tristimulus*

<sup>5</sup>Tradução nossa.

<sup>6</sup>O objeto analisa um sinal de áudio de entrada e o decompõe em componentes senoidais, que podem ser retornadas individualmente ou combinadas para realizar uma estimativa da frequência fundamental.

reteronar um valor acima de 0.5. Logo, quando houver a detecção de um evento que apresente essas características, um processo de espacialização é acionado. A figura 4.9 mostra a textura instrumental composta por notas com duração média (azul) e notas rápidas em *sul ponticello* na seção Vc de *Tensio*.

Figura 4.9: Início da seção Vc de *Tensio*

Portanto, considerando que o trecho em questão está escrito com fórmula de compasso 4/4, e com um andamento que corresponde à semínima igual à 50, o detector configurado para reportar a frequência fundamental a cada 200 milissegundos, só irá reagir à figuras rítmicas com valor acima de uma fusa, aproximadamente. Logo, reagindo aos eventos destacados em vermelho na figura 4.9, o que resulta em um espectro gerado pela síntese aditiva, que apresenta uma estrutura rítmica que reflete àquela do quarteto de cordas. Enquanto o outro detector, configurado para reportar a fundamental a cada 7 milissegundos irá reagir a figuras rítmicas próximas à *appoggiaturas* tocadas em *sul ponticello*, assim, permitindo que somente os gestos com essas características sejam submetidos ao processo de espacialização.

### 4.2.3 Notação

O *paradigma da notação* enquanto modo de pensar e conceber o uso criativo das ferramentas aqui estudadas abarca de forma ampla procedimentos, soluções técnicas e abordagens composicionais voltadas ao emprego de sistemas e processos de *score following*, ou seja, métodos voltados ao rastreo da performance musical (fluxo de áudio) e sua sincronização à uma partitura — ou a qualquer tipo de suporte capaz de registrar eventos distribuídos no tempo. Tais métodos operam através de uma representação simbólica dos eventos musicais notados na partitura, e a partir da extração de dados sonoros e musicais,

mapeia o fluxo de áudio que entra no sistema de maneira a decodificar em tempo-real a posição atual dos eventos musicais realizados pelo *performer* em relação à partitura.

A concepção de sistemas e métodos de *score following* encontrou nas áreas de ML/MIR um arcabouço técnico propício que permitiu seu desenvolvimento concomitantemente ao avanço das pesquisas nessas áreas. De fato, a noção de *score following* é um desdobramento direto do uso extensivo do MIDI e do desenvolvimento dos *sistemas musicais interativos*. Suas primeiras implementações utilizavam técnicas preliminares de escuta de máquina que se davam a partir da recuperação de informações simbólicas transmitidas através de sensores acoplados aos instrumentos (DANNENBERG, B. R., 1984; VERCOE, 1984).

Foi na década de 1990, com o desenvolvimento de ambientes de computação musical que a extração de informações e dados sonoros diretamente do sinal de áudio foi possível. Consequentemente, técnicas de *pitch e beat tracking* começaram a ser utilizadas, e segundo Cont (2010, p. 976), modelos estatísticos também passaram a integrar as implementações de tais métodos (GRUBB; DANNENBERG, R., 1997; RAPHAEL, 1999).

Esses sistemas se tornaram bastante populares nas últimas décadas, e vêm sendo utilizados em um conjunto amplo de aplicações, como a implementação de editor de partituras (DANNENBERG, R. B., 2007); criação de banco de dados musicais (MCNAB et al., 1996); desenvolvimento de aplicativos que auxiliam a leitura da partitura em performances (ARZT; WIDMER; DIXON, 2008); sincronização e visualização da partitura em concertos ao vivo (ARZT; FROSTEL et al., 2015) (PROCKUP et al., 2013); acompanhamento e sincronização automática do computador à um *performer* ao vivo (CONT, 2010; RAPHAEL, 2009), entre outras aplicações.

No âmbito da *live-electronics* a aplicação mais evidente é a utilização como estratégia de sincronização de ações computacionais prescritas em uma partitura eletrônica à performance instrumental/vocal ao vivo. Se por um lado o surgimento da computação musical e, conseqüentemente, das técnicas de processamento digital de sinais possibilitaram uma maior flexibilidade temporal do instrumentista em relação aos eventos eletroacústicos, por outro, a instantaneidade trazida pelo uso extensivo dessas ferramentas no campo da criação trouxe à tona questões técnicas e poéticas relevantes quanto à interatividade e reatividade entre *performer* e computador. Era necessário que os processos eletrônicos estivessem inseridos no mesmo fluxo temporal que o ambiente instrumental. Nesse contexto,

o *paradigma da notação* trouxe novas possibilidades técnicas e poéticas não só para o âmbito composicional, mas também para o da performance.

Ao viabilizar a possibilidade de determinar e prescrever com certa precisão as ações que o computador deverá executar e sincronizar a um músico ao vivo, a concepção das morfologias resultantes da interação entre instrumento/voz e computador passa a operar no mesmo domínio do processo da escrita instrumental. Uma vez que a notação musical apresenta um caráter simbólico e abstrato, permite que um gesto musical possa ser representado e registrado graficamente sobre um suporte — neste caso, a partitura.

Isto é, um “dó4” escrito para violino em *sul ponticello*; com duração igual à uma semibreve, e dinâmica que cresce gradualmente de “*pp*” à *ff*, além de fornecer indicações que permitirão ao intérprete realizar o gesto notado, permite que o compositor crie uma imagem sonora do resultado final. Portanto, da mesma maneira, a concepção de ideias e conceitos musicais sobre suporte computacional pode emergir e se desenvolver a partir desse trabalho específico de notação e registro de eventos distribuídos no tempo (MANOURY, 1990).

A seguir descreveremos algumas ferramentas voltadas ao uso de *score following* e suas aplicações práticas em peças do repertório. Ilustraremos também como o emprego do *paradigma da notação* afeta a prática criativa, trazendo os princípios da notação musical para o contexto da *live-electronics*, permitindo assim, o registro de eventos e morfologias musicais/sonoras — seja sobre uma partitura, um roteiro de ações, ou qualquer tipo de suporte notacional que viabilize a grafia de eventos distribuídos no tempo.

#### **a) *Jupiter* (1987)**

Uma das primeiras peças a explorar criativamente esse conjunto de estratégias que classificamos aqui como *paradigma da notação*, é *Jupiter* (1987), para flauta e *live-electronics*, de Philippe Manoury. Essa é a primeira peça de um ciclo chamado *Sonus ex machina* compostas em colaboração com Miller Puckette entre 1987 e 1991, e se destaca no repertório da *música interativa*, pois estabeleceu novos paradigmas no que diz respeito ao uso de escuta de máquina e sistemas interativos como estratégias de sincronização e interação entre processos eletrônicos e performance instrumental (MAY, 2005, p. 147).

O mecanismo de *score following* implementado em *Jupiter* utiliza o objeto [scofo~], capaz de ler as alturas em MIDI prescritas em um arquivo de texto (partitura de acompanhamento) e compará-las com aquelas extraídas a partir do sinal de áudio da flauta.

Para cada nota contida na partitura de acompanhamento o [scofo~] atribui uma “teoria”. A teoria é responsável por descrever a força da evidência de que a altura mais recente tocada pela flauta corresponde à altura mais recente lida na partitura de acompanhamento, e caso elas se correspondam o [scofo~] retorna o número do evento à ela vinculado.

Dessa forma, para cada seção da peça há uma partitura de acompanhamento contendo as durações em milissegundos e as alturas em MIDI de cada nota que a flauta deve tocar — assim como aparece na partitura instrumental —, e um outro arquivo com a sequência de mensagens carregadas pelo objeto [qlist] <sup>7</sup> que irá disparar e controlar os processamentos correspondentes à cada evento.

Uma vez que a altura extraída da flauta ao vivo coincide com a altura prescrita na partitura de acompanhamento — que pode ou não estar vinculada à um evento —, uma mensagem de controle contida no [qlist] é enviada aos módulos de transformação e síntese, e um processo é disparado. Se não há nenhum evento vinculado à aquela nota, o sistema apenas receberá a altura e continuará acompanhando o fluxo de dados extraídos da flauta.

Em sua primeira versão tal processo se dava a partir da recuperação de informações simbólicas através de uma flauta MIDI — construída pelo flautista Larry Beauregard, a quem a peça foi dedicada —, que enviava os dados para uma versão preliminar do ambiente Max/MSP de Puckette, onde o sistema de *score following* acionava os processos — que também eram realizados dentro do Max/MSP —, bem como controlava o sintetizador 4x de Giuseppe DiGiugno (FAVREAU et al., 1986). Em 1992, já sem a necessidade do uso da flauta MIDI, a peça foi implementada em Max/FTS, um *software* de controle de processamento de sinais que operava dentro do *Ircan Signal Processing Work Station (ISPW)* (LINDEMANN et al., 1991). Miller Puckette ainda realizou uma versão em *Pure Data* <sup>8</sup>, utilizando o objeto [fiddle~] (PUCKETTE; APEL; ZICARELLI, 1998) para extrair os dados referentes às alturas.

Um dos principais desdobramentos do emprego de *score following* em *Jupiter* reflete na escolha do compositor em representar as estruturas e morfologias sonoras/musicais emergentes da interação flauta-computador através da notação musical (representação de alturas, durações e dinâmicas). Uma vez que os parâmetros dos processos computacionais

<sup>7</sup> Objeto do *Pure Data* que lê arquivos de texto contendo mensagens com marcação de tempo, as quais podem ser sequenciadas automaticamente ou manualmente

<sup>8</sup> A versão em *Pure Data* está disponível em <http://msp.ucsd.edu/pdrp/pdrp-13/>

estão prescritos de forma precisa no [qlist], é possível prever o comportamento aproximado que irão apresentar quando acionados, logo, possibilitando que sejam representados por meio do suporte de notação tradicional.

Esse processo de notação dos eventos eletrônicos na partitura musical entra em ressonância com a noção de *solfejo de modelos*. Este conceito consiste em conhecimentos, habilidades intelectuais e cognitivas adquiridas pelo compositor ao conceber determinados processos, o que lhe permite controlar e dominar tanto as estruturas musicais resultantes de modelos generativos, quanto associar representações gráficas e/ou textuais de determinados *softwares* e sistemas musicais à um resultado musical final (MALT, 2004, p. 1).

Dessa forma, o emprego de *score following* atrelado à ideia de *solfejo de modelos* permitiram que, a partir de um conhecimento prévio da natureza e dos parâmetros dos processos de transformação/geração sonora envolvidos na peça, o compositor pudesse entrever seus comportamentos quando integrados aos gestos da flauta. Isso nos leva a acreditar que uma etapa importante do processo composicional de *Jupiter* – e que parece ser uma das características mais presentes na escrita de Manoury – diz respeito, justamente, à concepção da parte eletrônica de forma aproximada àquela da escrita instrumental – i.e., imaginando as estruturas e morfologias musicais resultantes da amálgama entre processos computacionais e gestos instrumentais, e reduzindo sua complexidade sonora à uma abstração simbólica que pudesse ser representada pela notação tradicional.

A figura 4.10 representa o início da seção **ID**, de *Jupiter*, e mostra a representação da morfologia resultante da interação entre a flauta (pauta inferior) e quatro módulos de transposição de alturas associados a um *reverb* (pautas superiores). Cada módulo recebe um valor específico de transposição em *midicents* e um valor de atraso em milissegundos no seu acionamento, gerando intervalos melódicos e durações rítmicas (em vermelho), enquanto algumas das alturas geradas pelo módulo de transposição passam por um *reverb* infinito prolongando sua duração (em azul).

Há uma dicotomia entre os aspectos determinísticos da notação musical e aqueles indeterminísticos inerentes à performance. Logo, trazer esses elementos intrínsecos à interpretação musical para o contexto da *música interativa* parece ser uma das preocupações de Manoury. É a partir dessa ideia que surge a noção de *partitura virtual*. Segundo o compositor, *partitura virtual* é uma representação em que os parâmetros constituintes são conhecidos previamente, porém seu comportamento e manifestação concreta exata estão

Figura 4.10: Início da seção **ID** de *Jupiter*

sujeitos à variações. A *partitura virtual* apresenta como premissa, um conceito que parece estar ligado ao princípio do que se denomina “tempo-real” e que se baseia na possibilidade de detectar, analisar, e processar eventos produzidos pela performance (MANOURY, s.d.).<sup>9</sup>

Essa parece ser uma característica intrínseca à escrita musical associada ao *paradigma da notação*, uma vez que a organização da *partitura virtual* – assim como a dos processos de *score following* – se fundamenta em dois princípios básicos: o reconhecimento e a monitorização de eventos musicais. Em suma, estes dois aspectos possibilitam uma representação paramétrica determinística (notação), ao mesmo tempo em que está sujeita à uma certa abertura ocasionada por fatores indeterminísticos (interpretação) (MANOURY, s.d.).

Outro desdobramento do emprego do *paradigma da notação* em *Jupiter*, é a geração de processos estáticos, nos quais as estruturas sonoras geradas pelo computador tendem a replicar o comportamento gestual apresentado pela flauta numa espécie de “pergunta e resposta”. Logo, estratégias como dilatação temporal dos gestos do flautista, atrasos no disparo dos eventos, bem como a manipulação de crescendo e diminuendos são algumas das estratégias adotadas em *Jupiter* para criar uma certa independência entre os materiais e criar uma interação mais fluida e menos imitativa entre processos eletrônicos e a performance instrumental ao vivo.

Cada tipo de processamento apresenta uma estratégia específica de interação com o flautista. Segundo May (2005, p. 152), essas estratégias de interação podem ser divididas em quatro grupos, (1) extensão temporal dos gestos instrumental; (2) criação de estruturas rítmicas independentes do comportamento da flauta; (3) extensão do timbre instrumental, e (4) geração de harmonias que complementam ou contrastam com a parte instrumental.

As primeiras experiências na implementação de métodos de *score following* fundamentavam-se na recuperação de informações simbólicas, influenciados fortemente pelo

<sup>9</sup><http://www.philippemanoury.com/?p=340>

uso extensivo do protocolo MIDI e pelo emprego dos *sistemas musicais interativos*. Ao passo em que novas ferramentas surgiram, impulsionadas pelo desenvolvimento e expansão dos métodos de processamento digital de sinais, e pelo avanço das pesquisas nas áreas de ML e MIR, modelos psicoacústicos, estocásticos e métodos de aprendizado de máquina começaram a integrar a arquitetura dos sistemas de *score following*. Assim, essas novas ferramentas, mais complexas e robustas, passam a trabalhar não só com informações acerca da altura, mas também a decodificar aspectos temporais da performance, permitindo prever o andamento do interprete e se adequar à sua agógica.

Em virtude da natureza temporal da performance musical, a possibilidade de rastrear e sincronizar eventos temporais constitui um importante mecanismo dos sistemas de *score following*. Uma vez que tais sistemas atuam em situações que, de certa maneira, não são previsíveis, os modelos probabilísticos e o aprendizado de máquina se mostraram uma solução eficaz no rastreamento da performance ao vivo. Logo, essas novas ferramentas quando empregadas na *música interativa* transformam profundamente a escrita musical, abrindo novas possibilidades criativas, ao mesmo tempo em que trazem à tona novas problemáticas.

### ***Antescofo***

Dentre essas novas ferramentas destaca-se o *antescofo* (CONT, 2008), um sistema de *score following* que vêm sendo utilizado extensivamente nas produções mais recentes. Nesse sentido, discutiremos algumas concepções poéticas e soluções técnico-musicais, e ilustraremos algumas das implicações que o uso específico dessa ferramenta acarreta na escrita instrumental e na concepção dos processos eletrônicos, a partir de exemplos de algumas peças do repertório da *música interativa* atual.

*Antescofo* é um sistema de *score following* modular que combina métodos de escuta de máquina com uma linguagem de programação reativa e síncrona que opera por meio da correlação entre partitura instrumental e eletrônica e da sua sincronização com a performance musical humana. Seu emprego permite a representação simultânea e o reconhecimento de diferentes atributos sonoros — ao contrário dos sistemas de *score following* tradicionais que utilizam apenas reconhecimento de altura —, controle sobre várias escalas de tempo, bem como a capacidade de rastrear e sincronizar ações à uma performance.

Essa ferramenta começou a ser desenvolvida em 2007, por Arshia Cont em colaboração com o compositor Marco Stroppa, no IRCAM, com o objetivo de conceber uma peça para saxofone e *sistema interativo*. Em 2011, os pesquisadores Jean-Louis Giavitto, Florent Jacquemard, e José Echeveste se juntaram à equipe. O *antescofo* pode ser utilizado como um objeto *external* para Max/MSP e *Pure Data*<sup>10</sup> (JEAN-LOUIS GIAVITTO; CONT; ECHEVESTES, s.d.), o que permite seu uso de forma modular, podendo se adequar aos mais variados projetos artísticos.

Sua arquitetura geral é composta por dois agentes: de áudio e temporal, ambos atuam concomitantemente com o objetivo de rastrear e sincronizar o fluxo de áudio em tempo-real à sequência de estados mais provável contida em um modelo probabilístico concebido com base na partitura instrumental. O agente temporal opera na escala de tempo do evento [*event-time scale*] – i.e., tempo baseado na duração relativa ou absoluta de eventos sucessivos – a partir de um modelo psicoacústico de estrutura métrica musical desenvolvido por Large (2001) e assim, fornece previsões das informações temporais com base na entrada de áudio e na partitura musical. Já o agente de áudio opera de acordo com o fluxo contínuo de áudio, e atribui valores probabilísticos à estados relevantes contidos no modelo. Em suma, o que se tem é um sistema antecipatório que implica em um modelo preditivo de si mesmo, e que lhe permite mudar de estado em um instante de acordo com previsões baseadas em eventos passados (CONT, 2008, p. 2).

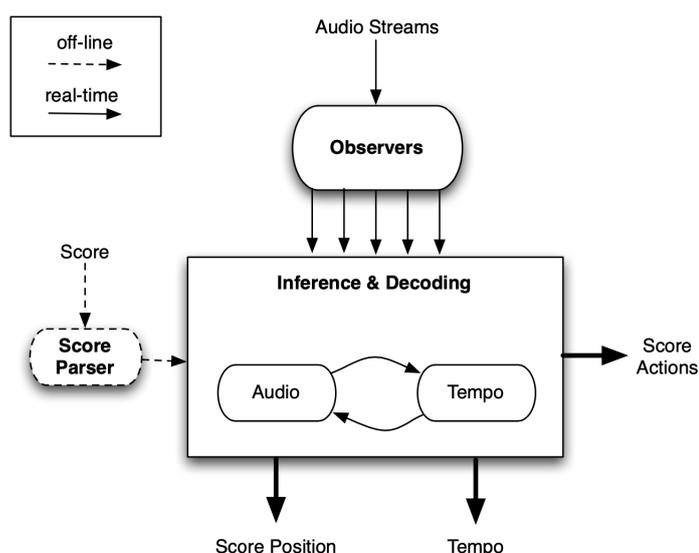


Figura 4.11: Arquitetura geral do *antescofo* (CONT, 2008, p. 3)

<sup>10</sup><https://forum.ircam.fr/projects/detail/antescofo/>

Um dos princípios do *antescofo* é sua abordagem temporal na sincronização da partitura com a performance ao vivo — em grande parte inspirado pelo efeito de sincronização de pendulos, como observado por Huygens (1888) . Logo, alguns fundamentos musicais, de certa maneira, influenciaram a sua modelização tendo em perspectiva uma aplicação prática, como a noção de evento atemporal e temporal. Segundo Xenakis (1992) , são eventos atemporais aqueles que possuem sua própria estrutura temporal interna independente da estrutura temporal geral de uma peça. Trazendo para o contexto do *antescofo*, um evento atemporal é aquele que ocupa um espaço físico na partitura, mas não contribui para a noção de tempo na estrutura geral (CONT, 2010, p. 3). São exemplos de eventos atemporais, notas de *apogiatura*, *fermata*, *trilos* ou trechos de improvisação livre. Já os eventos temporais são aqueles que evoluem ao longo do tempo e são caracterizados pela sua envoltória de energia.

A despeito de ambos terem presença física, quando considerados individualmente como eventos isolados eles não contribuem para a noção de tempo. É justamente a natureza relativa, no caso dos eventos atemporais, e a estrutura geral no tempo, no caso dos eventos temporais que contribuem para a noção de tempo de uma forma global.

Outro conceito relacionado ao aspecto temporal que inspirou a concepção dos mecanismos presentes no *antescofo* é a noção de *tempo liso* e *tempo estriado* (BOULEZ, 1964). O *tempo liso* consiste em uma noção contínua do tempo como um fluxo ininterrupto de informação, enquanto o *tempo estriado* é baseado em regularidades temporais que são recorrentes (CONT, 2008, p. 3).

O *antescofo* possui uma linguagem de programação textual própria para representar os eventos musicais e definir os processos vinculados à tais eventos. Sua sintaxe representa de forma simples elementos musicais comuns ao repertório da música ocidental, além de permitir a importação de partituras nos formatos mais comuns utilizados por *softwares* de edição musical, como MIDI e *xml*, de maneira que possam ser convertidas para o formato do *antescofo*.

Dessa forma, a partitura do *antescofo* possui dois elementos principais: *eventos* — elementos musicais que serão reconhecidos pelo computador representados por palavras-chaves como NOTE, CHORD, TRILL e MULTI —, e *ações* — processos a serem realizados uma vez que os eventos ou condições correspondentes sejam detectados (JEAN-LOUIS GIAVITTO; CONT; ECHEVESTE, s.d.). As ações no *antescofo* funcionam de

forma análoga ao objeto [qlist] do *Pure Data*, porém estendendo algumas das suas funcionalidades.

A sintaxe básica para representar um evento na linguagem do *antescofo* segue o seguinte formato: <float> (representação do evento) <duração> (duração do evento) <nome opcional> (nome do evento). A figura 4.12 mostra a partitura instrumental e a partitura do *antescofo* referentes ao primeiro compasso da seção **IA** de *Do Corpo ao Pó*.

The figure displays two parts: a musical score on the left and a Pure Data code block on the right. The musical score is for a 4/4 time signature with a tempo of 40 BPM. It features a treble clef and a key signature of one sharp (F#). The first measure contains a complex chord structure with dynamics ranging from *f* (forte) to *p* (piano). A box labeled 'IA' is placed above the first measure, and a box labeled '1' is placed below the first measure. The Pure Data code block contains the following lines:

```

26 BPM 40
27
28 VARIANCE 0.5
29
30 //Compasso 1
31
32 NOTE F#2 0.4 metro bang ; inicia a randomização
33 CHORD (A2 D#3 A#3 D4 E4) 2.8 pitch-in 0 300 ; abre entrada de audio
34   shifter_event bang Annotation "dispara pitch shifter"
35 CHORD (A2 D#3 A#3 D4 E4) 0.4 pitch-in -128 300 ; fecha entrada de audio
36   metro stop ; para a randomização
37 CHORD (A2 D#3 A#3 D4 E4) 0.4 ; sem evento

```

Figura 4.12: Exemplo da partitura do *antescofo* e da partitura instrumental do início da seção **IA** de *Do Corpo ao Pó*

## b) *Raggi di Strighe* (2011)

O desejo de aproximar a concepção dos processos eletrônicos à natureza notacional da escrita instrumental nos parece ser uma questão central que permeia os processos composicionais que exploram o *paradigma da notação*, como mostra alguns aspectos do processo composicional de *Jupiter*. Isso fica ainda mais acentuado ao observarmos alguns pensamentos composicionais empregados em peças que exploram o uso do *antescofo* enquanto ferramenta atuante na integração entre processos interativos e escrita instrumental. De fato, o emprego desta ferramenta está ligada à uma estética e à um modo de pensar a concepção da eletrônica, que é particular dos compositores ligados ao IRCAM.

*Raggi di Stringhe* (2011), para violino e *live-electronics* da compositora Lara Morciano além de ilustrar tal pensamento, exemplifica algumas possibilidades e implicações composicionais do emprego do *antescofo* e consequentemente nos serve como campo reflexivo para abordarmos o emprego do *paradigma da notação*. Essa peça foi composta no IRCAM em colaboração com o diretor de informática musical José Miguel Fernandez. A ideia básica, segundo a compositora, era compor uma peça na qual os processos eletrônicos pudessem ser desenvolvidos de maneira bem articulada e fossem concebidos como um complemento ao virtuosismo da escrita instrumental, estabelecendo certa relação com o

violino através de um foco especial nas reações, e nas transformações sonoras e gestuais do instrumento. (MORCIANO, s.d.).

Uma das questões cruciais na base da minha reflexão artística, residia no desejo de produzir uma obra concebida de forma unitária e coerente no que diz respeito à escrita e aos materiais utilizados, bem como à concepção e realização da parte eletrônica tendo em vista a expressividade e a liberdade interpretativa; as diferentes problemáticas ligadas ao desenvolvimento temporal da ação musical no momento da execução, e a sincronização entre intérprete e máquina<sup>11</sup> (MORCIANO, 2018, p. 36).

Destacamos aqui, que essa exigência de definir e detalhar minuciosamente no tempo os processos de síntese e transformação sonora desta peça — numa tentativa de emular a natureza da escrita instrumental —, parece ter sido motivada, em certa medida, por este desejo de derivar e correlacionar as morfologias sonoras decorrentes dos processos eletrônicos à escrita instrumental, de maneira a torná-la parte integrante do discurso musical. Logo, o emprego e a exploração das possibilidades oferecidas pelo *antescofo* é uma escolha técnica que vem a contribuir diretamente na realização das premissas poéticas e musicais de *Raggi di Stringhe*.

Com o intuito de encontrar uma correspondência direta entre modos de tocar específicos, estruturas rítmicas e gestos instrumentais, os sons eletrônicos são trabalhados por vários tipos de transformações rítmicas e timbrísticas, evidenciando a influência que a escrita instrumental exerce sobre a concepção dos processos de transformação e síntese sonora na peça como um todo. Por outro lado, é evidente que ao empregar o *antescofo* há o movimento contrário, no qual o uso de tal ferramenta acaba por transformar a própria escrita instrumental numa tentativa de adaptá-la ao emprego do *paradigma da notação* e de integrar ambos ambientes sonoros.

Segundo Morciano, foi necessário questionar certos elementos da sua escrita rítmica, como a *apoggiatura*, por exemplo, que é frequentemente utilizada para antecipar figuras rítmicas, aumentando também a duração das células a que se referem ou apresentando a função de acelerar certos gestos rítmicos dentro de pulsações regulares estabelecidas (MORCIANO, 2018, p. 83).

Esse processo pode ser observado na seção final de *Raggi di Stringhe* (figura 4.13). Neste trecho a compositora utiliza *appoggiaturas* (em azul) no final dos *glissandos* como estratégia para estabelecer um ritmo e uma altura de referência, permitindo que o *antescofo* detecte o início e

---

<sup>11</sup>Tradução nossa.

fim de cada evento e acompanhe os gestos instrumentais com maior precisão — uma vez que o *glissando* não fornece uma regularidade temporal recorrente que possa guiar o *antescofo*. Essa estratégia de escrita permitiu manter uma fluidez rítmica de caráter irregular e imprevisível ao mesmo tempo em que permite que uma partitura eletrônica escrita de forma detalhada possa ser executada precisamente em resposta às ações do instrumentista.

The image shows a handwritten musical score for the piece 'Raggi di Stringhe'. It consists of ten staves of music, each with various annotations and markings. Key features include:

- Event Markers:** Red boxes containing event codes such as E-007, E-008, E-009, E-010, E-012, E-013, E-014, E-019, E-020, E-030, E-031, and E-042.
- Performance Instructions:** Words like 'SEGUI' and 'SUBITO' are written in blue ink across the staves.
- Handwritten Notes:** Various notes in blue and black ink provide additional context, such as 'Sul clarinetto', 'Sul violoncello', and 'Sul violino'.
- Mathematical Notations:** Some event markers include mathematical expressions, such as E-014 with the formula  $(=15, +15+15+12 \cdot 18)$ .
- Dynamic Markings:** Symbols like *f* (forte) and *ff* (fortissimo) are present.
- Structural Elements:** Brackets and lines indicate phrasing and structural divisions.

Figura 4.13: Trecho da seção final de *Raggi di Stringhe*

### c) *Dispersion de trajectoires*(2014)

Um exemplo mais complexo de estratégia composicional no que diz respeito ao uso do *antescofo* pode ser observado em *Dispersion de trajectoires* (2014), para saxofone barítono e *live-electronics*, de José Miguel Fernández. Nesta peça o *antescofo* exerce um papel central na cadeia de processos, uma vez que é responsável por concectar, monitorar e coordenar processos em diferentes dispositivos e ambientes computacionais.

De fato, assim como em *Jupiter* e em *Raggi di Stringhe*, há em *Dispersion de trajeroires* um pensamento composicional, que em certa medida implica no emprego de algum tipo de suporte ou aparato tecnológico visando a concepção e o registro de eventos sonoros paralelamente à uma partitura instrumental – mais uma vez numa tentativa de centralizar a concepção dos processos computacionais/interativos e trazê-los para o domínio da escrita instrumental. Isso fica evidenciado pela arquitetura do conjunto técnico criado pelo compositor, o que dá ao *antescofo* um protagonismo na coordenação dos processos ao longo da peça.

Logo, como estratégia técnica para trabalhar com a escrita instrumental e a concepção dos processos interativos de forma centralizada, Fernández desenvolveu uma biblioteca chamada *antescollider*. Essa biblioteca integra o *antescofo* ao *SuperCollider*, permitindo a concepção de uma partitura eletrônica centralizada. As informações do processamento de áudio são distribuídas aos servidores do *SuperCollider* utilizando o protocolo *Open Sound Control* (OSC) (WRIGHT; FREED, 1997) sob a supervisão e coordenação de uma partitura/roteiro concebida na linguagem do *antescofo*. Com o *antescollider* os processamentos de áudio podem ser distribuídos por vários servidores do *SuperCollider*, assim, maximizando a eficiência da CPU (FERNANDEZ; GIAVITTO; DONAT-BOUILLUD, 2019).

A configuração e a interconexão entre tais dispositivos e os processos em *Dispersion de trajeroires* consiste em dois módulos: (1) módulo de controle geral (*antescofo*), e módulo de análise e extração de atributos musicais/sonoros implementados em Max/MSP, e (2) módulo de síntese e processamento sonoro instanciados no *Supercollider*. Dessa forma, os dados extraídos da performance instrumental são mapeados e enviados como parâmetros de controle dos processos de síntese e transformação sonora que estão implementados no *Supercollider*, o fluxo desse mecanismo está sujeito ao acionamento do *antescofo*, que por sua vez irá reagir de acordo com a performance instrumental.

Segundo Fernández, a partitura centralizada se mostra relevante para a execução da performance, uma vez que esse modelo de emprego do *antescofo* e do *antescollider* cria um suporte que viabiliza a escrita das interações instrumento-computador de forma explícita e favorece uma abordagem dinâmica na gestão dos processamentos e sínteses de áudio – as cadeias de áudio podem ser facilmente criadas e destruídas dinamicamente em resposta direta às ações do *antescofo*.

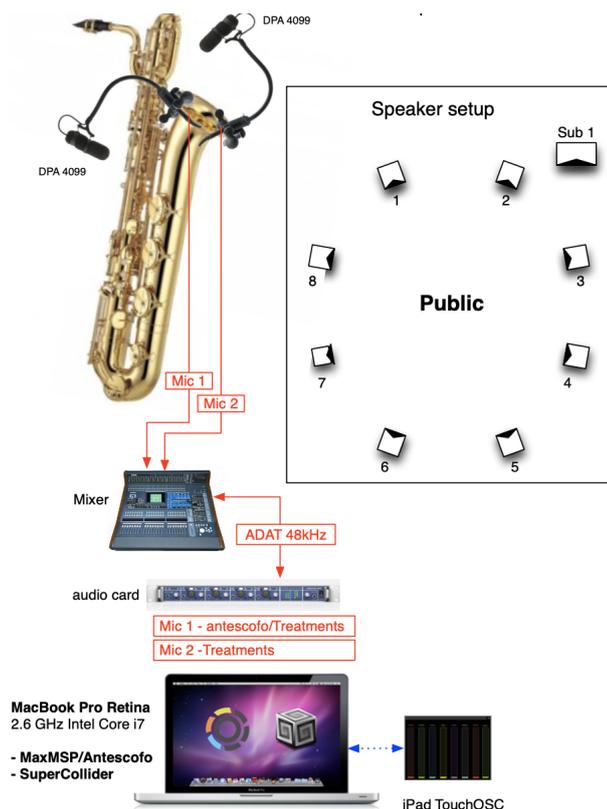


Figura 4.14: *Setup de Dispersion de trajectoires*

A figura 4.15 ilustra a dinâmica entre o *antescofo*, *antescollider*, e os processos instanciados no *SuperCollider* nos compassos 251 e 252 de *Dispersion de Trajectoires*. A nota “ré4”<sup>12</sup> dispara o evento 1 (vermelho) que aciona dezesseis instâncias de um modulador de frequência, (*shift\_rev\_env*) cada um com um valor específico de duração, frequência moduladora, envoltória e posição do alto falante. A primeira pausa do compasso seguinte dispara o evento 2 (azul) que desativa três instâncias de um sintetizador granular (*abort \$play\_gran\_seq\_mes252*), quatro instâncias de um sequenciador (*abort \$pitch\_seq\_mes252*), e aciona uma amostra sonora contendo a nota “ré2”, que está armazenada em um *buffer* (*group play\_re*). A terceira e a última pausa disparam respectivamente os eventos 3 (verde) e 4 (roxo), que também são responsáveis por acionarem a mesma amostra sonora (*group play\_re*).

#### d) *Estremo d’ombra* (2015)

Outro exemplo de uso do *antescofo* aliado ao *antescollider* é *Estremo d’ombra* (2015), para cinco solistas e *live-electronics*, da compositora Lara Morciano. A concepção desta peça

<sup>12</sup>A partitura do *antescofo* mostra as notas que soam enquanto a partitura instrumental está transposta em um intervalo de décima terceira acima da nota que realmente soa.

Figura 4.15: Exemplo da partitura do *antescofo* e da partitura instrumental referente aos compassos 251 e 252 de *Dispersion de trajectoires*

surge da vontade de explorar elementos virtuosísticos do solistas e combiná-los à uma escrita eletrônica muito articulada e detalhada. O dualismo entre escrita solo e densidade instrumental, heterogeneidade e homogeneidade, expressividade interpretativa e interação com o ambiente computacional em tempo-real caracteriza a gênese a poética, e a concepção técnica da escrita eletrônica de *Estremo d'ombra* (MORCIANO, 2018, p. 91).

A arquitetura do sistema criado para esta peça é similar ao empregado por Fernández em *Dispersion de trajectoires*, no qual temos o *antescofo* como a gente central coordenando todos os processos computacionais instanciados no *Supercollider* intermediado pela biblioteca *antescollider*. O sistema computacional de *Estremo d'ombra* ainda permite o controle e gerenciamento de cinco monitores de vídeo colocados no palco para visualização das partituras dos músicos, através de um aplicativo – também controlado pelo *antescofo* – que rastreia o número de compasso e as pulsações em tempo-real, permitindo a sincronização dos intérpretes entre si, e entre os intérpretes e os processamentos sem a necessidade de um regente. A figura 4.16 mostra a configuração e as conexões entre os dispositivos do sistema implementado em *Estremo d'ombra*.

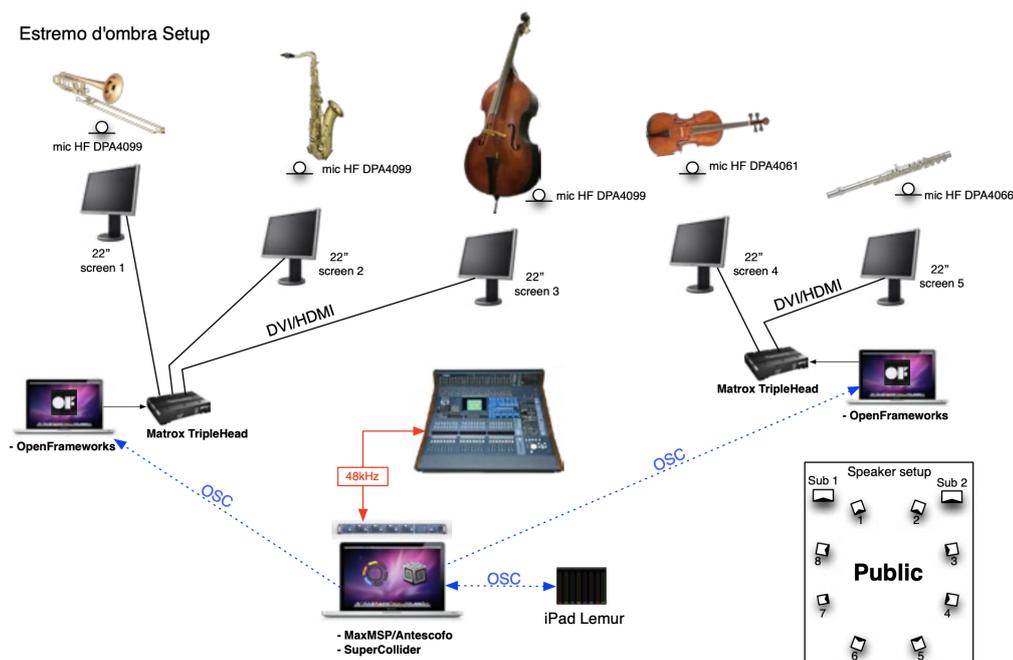


Figura 4.16: *Setup de Estremo d'ombra* (MORCIANO, 2018, p. 97)

Destacamos aqui uma particularidade intrínseca da escrita musical em *Estremo d'ombra* que corrobora com o emprego do *paradigma da notação*. Aqui, o *antescofo* adquire a função de “maestro virtual”, monitorando e coordenando todos os processos envolvidos na peça, o que garantiu que Morciano pudesse explorar diversos aspectos camerísticos da escrita instrumental, valorizando o virtuosismo de cada músico e possibilitando uma escrita eletrônica bastante minuciosa e dinâmica, ao mesmo tempo viabilizando a livre expressão e movimentação dos solistas – que em muitos momentos da peça mudam de posição no palco buscando diversas configurações instrumentais, espaciais e combinações timbrísticas, por meio da estratégia do uso de partituras digitais interativas. Dessa forma, as estruturas e morfologias que se engedram desse processo dinâmico tendem a se comportar de forma bastante precisa e independente, mantendo uma correlação com a escrita instrumental.

A figura 4.17 exemplifica o papel de “maestro virtual” adquirido pelo *antescofo* na gerência dos processos interativos da peça. O exemplo mostra trecho da partitura eletrônica (acima) e da partitura instrumental (abaixo) referente ao compasso 188 da peça. Observa-se que na linha 38 da partitura eletrônica há um comando que abre uma porta *OSC* permitindo a comunicação entre o *antescofo* e o *SuperCollider*, dessa forma, ao detectar o dó4 tocado pelo saxofone barítono, (destacado na partitura instrumental), o *antescofo* aciona o evento *measure188* (linha 42 à linha 54 da partitura eletrônica) que consiste em três congelamentos granulares aplicados à flauta, saxofone e viola (linhas 47 à 49 da partitura eletrônica) e três

*low frequency oscillators* randômicos (linhas 51 à 53 da partitura eletrônica), que por sua vez controlam a posição panorâmica do congelamento no espaço.

The image displays a Pure Data patch window at the top, showing code for a 'freeze' group. The code includes audio input processing for three channels (flute, saxophone, and violin) and LFO-based modulation of parameters like pan and amplitude. Below the patch window is a musical score for the piece 'Estremo d'ombra', specifically measure 188. The score includes a vocal line and piano accompaniment. Annotations include 'pp' (pianissimo), 'Sord. wawa ou plunger', 'tast. N.V.', and 'ord'. A red circle highlights a note in the vocal line, and an arrow points to a note with the annotation 'changer le timbre'.

Figura 4.17: Trecho da partitura do *antescofo* (acima) e da partitura instrumental (abaixo) referente ao compasso 188 de *Estremo d'ombra*

De fato, a relação estabelecida entre *antescofo* e os diversos processos presentes na peça permite criar, destruir e modificar processamentos e síntese, de acordo com uma partitura/roteiro de ações pré-definidos na própria linguagem do *antescofo*. Logo, fica evidente que a concepção poética da peça antecede o emprego do *paradigma da notação*, ou seja, é justamente por conta da natureza poética e do caráter camerístico que a compositora pretendia imprimir em sua peça, que foi adotado o emprego do *antescofo* enquanto ferramenta. Associado ao uso do *antescollider*, foi possível sincronizar o ambiente eletrônico e o instrumental de forma satisfatória, respeitando a liberdade interpretativa dos músicos e capturando todas as informações necessárias para a elaboração de processos de tratamentos e síntese (MORCIANO, 2018, p. 90).

#### 4.2.4 Procedimento

Nomeamos como *Paradigma do procedimento* as estratégias que fazem uso das técnicas de ML/MIR visando a modelização computacional de procedimentos algorítmicos e

composicionais implementados para operarem em tempo-real. Esses procedimentos podem ser definidos como uma sequência finita de regras, raciocínios ou operações que, uma vez aplicadas passo-a-passo a um conjunto de dados, é capaz de engendrar estruturas sonoras/musicais como produto final.

Observando pelo prisma da prática da escrita instrumental e vocal ocidental, pode-se constatar que a noção de algoritmo já está presente nos procedimentos composicionais para além da inserção do computador na prática criativa. As regras e fórmulas que guiam a escrita contrapontística; a tabela de Guido d'Arezzo, que gera melodias para um dado texto; os jogos de dados musicais, ou mesmo a composição de séries dodecafônicas, por exemplo, se aproximam muito dos conceitos que estão por trás da definição de algoritmo.

Dessa forma, esse conjunto de regras composicionais pressupõe uma certa formalização, ou seja, um processo de identificação de uma problemática e a escolha por uma solução. No entanto, é evidente que essa formalização não isenta o compositor da busca por soluções pessoais visando alcançar um determinado resultado estético. Ainda que coexista a necessidade de um pensamento de ordem não racional, tais procedimentos visam a solução de impasses musicais através de um conjunto de etapas lógicas, que demandam a concretização de um pensamento congruente e bem formatado.

Logo, a aproximação entre algoritmos e procedimentos composicionais é válida quando estes prevêm uma representação simbólica e quantitativa da música, e esta é submetida às operações lógicas. Ou seja, os procedimentos são definidos por uma formalização que demanda a elaboração e a definição de representações matemáticas, e de operações automatizadas sobre essas representações, que por sua vez manipularão um conjunto abstrações musicais paramétricas (PADOVANI, 2009a, p. 22).

Quando integradas à modelização computacional de procedimentos composicionais, as técnicas de ML/MIR viabilizam uma série de possibilidades de aplicações criativas, as quais apontam para a elaboração e exploração de estratégias de interação entre computador e *performer* que se dão em diversos níveis de complexidade. Neste contexto, ML/MIR podem atuar na (1) representação simbólica dos dados musicais recuperados, os quais se pretende submeter às operações, e (2) modelização computacional de habilidades cognitivas e de fenômenos psicoacústicos que integrarão a arquitetura dos algoritmos, visando a mímese de estruturas musicais, assim como aquelas produzidas por músicos humanos.

De fato, uma vez que as informações extraídas pelos métodos de ML/MIR são utilizadas como dados de entrada nos algoritmos, a implementação computacional de procedimentos composicionais passa a envolver também estratégias ligadas à processos de mapeamento. Isto é, as informações de entrada são mapeadas e convertidas em parâmetros que serão submetidas à uma série de operações. Assim, ao passo em que informações correlacionadas à atributos sonoros/musicais extraídas em tempo-real passam a ser utilizadas como dados de entrada dos algoritmos, o emprego do *paradigma do procedimento* se torna capaz de engendrar estruturas sonoras/musicais que derivam, em certa medida, do próprio material musical gerado pelo instrumentista/cantor durante a performance.

Eventualmente, essas estratégias trazem consigo a possibilidade da utilização de aprendizado de máquina, processos estocásticos, bem como outros métodos de manipulação de dados no tratamento das informações de entrada. Como consequência, essa estratégia pode aproximar o emprego do *paradigma do procedimento* daquele voltado à implementação de agentes musicais autônomos, ou seja, algoritmos que atuam na geração automática de estruturas musicais — sem supervisão humana, ou com uma supervisão limitada (MORREALE; DE ANGELI, 2017).

Contudo, no contexto da criação dentro do escopo da *live-electronics* o emprego do *paradigma do procedimento* pode implicar no engendramento não só de estruturas musicais provenientes de regras e operações lógicas baseadas em preceitos musicais — ou em determinados procedimentos de organização dos materiais do ponto de vista rítmico, melódico ou harmônico —, mas também na criação de estruturas timbrísticas, as quais são geradas a partir de procedimentos de manipulação do material sonoro no nível espectral.

Portanto, combinados com métodos de aprendizado de máquina, dentre outros processos de tratamento de dados, essas estratégias podem viabilizar a geração automatizada de gestos musicais que se auto-regulam e se adaptam de acordo com a natureza da performance humana. Essas estratégias criativas serão exemplificadas através da aplicação prática em duas peças: *Tensio* (2010), para quarteto de cordas e *live-electronics*, de Philippe Manoury, e *Traverser le réseau* (2014), para violoncelo e *live-electronics*, de Rael Toffolo.

#### a) *Tensio* (2010)

Novamente utilizaremos as estratégias composicionais empregadas em *Tensio*, para ilustrar as aplicação criativas dos paradigmas composicionais que buscamos circunscrever

aqui. Nesta peça, o *paradigma do procedimento* é empregado através da concepção de algoritmos/procedimentos responsável pelo cálculo e geração de estruturas harmônicas e melódicas a partir das frequências fundamentais e das durações dos sons recuperadas diretamente do quarteto de cordas.

É importante ressaltar, que assim como algumas das outras peças abordadas neste trabalho, no exemplo ilustrado a seguir é possível identificar estratégias composicionais que combinem dois ou mais dos paradigmas descritos aqui, desse modo, podendo servir como exemplo de um ou outro paradigma. No entanto, nos restringiremos à mostrar como o compositor utilizou métodos de recuperação de informações musicais/sonoras na elaboração e modelização de procedimentos composicionais. Portanto, o foco será no emprego do *paradigma do procedimento*.

Em um primeiro momento, o emprego do *paradigma do procedimento* em *Tensio*, se dá por meio de um algoritmo desenvolvido por Miller Puckette em 2006, denominado *3F synthesis* (MANOURY, 2020). Esse método calcula um conjunto de valores de frequência e amplitude a partir da entrada de três frequências principais. Os valores retornados pelo cálculo são enviados para um banco de osciladores que irão gerar espectros com densidade e graus de harmonicidade variados.

As três frequências de entrada são recuperadas em tempo-real a partir dos instrumentos do quarteto de cordas, de modo que, uma vez extraídas, cada uma das três frequências são submetidas à combinações lineares. Isto é, aplica-se todas as possibilidades de soma e subtração umas às outras. Por exemplo, dada três frequências A, B e C, tem-se:

$$2A \quad (A+B) \quad (A-B) \quad (A+C) \quad (A-C)$$

$$2B \quad (B+A) \quad (B-A) \quad (B+C) \quad (B-C)$$

$$2C \quad (C+A) \quad (C-A) \quad (C+B) \quad (C-B)$$

$$3A \quad (2A+B) \quad (2A-B) \quad (2A+C) \quad (2A-C)$$

$$3B \quad (2B+A) \quad (2B-A) \quad (2B+C) \quad (2B-C)$$

$$3C \quad \dots$$

Considerando somente os números absolutos, teremos como resultado uma ampla gama de valores de frequências. Como os espectros criados dessa forma podem se tornar excessivamente densos, esse aglomerado de frequências passa por um procedimento de organização baseado em uma lógica probabilística, cujos valores podem variar de 0,03 à 1. Imaginando que deseja-se gerar um espectro composto por oito frequências, visto que o

índice de probabilidade seja 1, o algoritmo selecionará em ordem crescente as primeiras oito frequências calculadas.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28...

Ao passo em que o índice de probabilidade diminui, o algoritmo seleciona aleatoriamente frequências mais distantes umas das outras até chegar ao total de oito componentes.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28...

Assim, é estabelecida uma relação proporcional entre o índice de probabilidade e o grau de harmonicidade que o espectro irá adquirir. Embora tais espectros sejam únicos do ponto de vista da combinação de frequências — visto que um índice de probabilidade menor que 1 sempre irá gerar uma combinação de frequências diferente a cada cálculo, ainda que o valor do índice de probabilidade não seja alterado —, é possível reconhecê-los como pertencentes a uma mesma classe (MANOURY, 2010, p. 68). Logo, tais estruturas espectrais podem criar uma certa amálgama espectral com os sons instrumentais, dos quais são derivados.

No início da seção **IVA** de *Tensio* pode-se observar a aplicação desse procedimento — em conjunto com o *paradigma do gatilho*, como demonstrado na seção 4.2.2. Cada uma das três frequências de entrada são extraídas a partir de um instrumento diferente, i.e., a primeira frequência é recuperada a partir do violino 2; a segunda da viola, e a terceira do violoncelo. Contudo, o processo é realizado apenas quando sons curtos forem detectados — os sons mais longos são ignorados.

Esse procedimento resulta em uma estrutura harmônica que se comporta como uma espécie de “subconjunto virtual”, cujo o conteúdo harmônico/espectral e rítmico deriva do próprio material instrumental. Uma vez que o número de componentes frequenciais calculado neste trecho da peça é doze, e o índice de probabilidade 0,19, o cálculo do *3F synthesis* acaba gerando uma estrutura sonora com um grau de inarmonicidade relativamente alto. O exemplo 4.18 mostra a camada resultante deste processo, em verde, sobreposta ao quarteto de cordas, em azul.

Outra estratégia composicional que envolve o emprego do *paradigma do procedimento* em *Tensio*, consiste em contruir automaticamente linhas melódicas atribuindo valores de probabilidade de sucessões entre os elementos musicais que as constituirão. Esse procedimento é realizado através da recuperação de informações relacionadas à altura, intensidade e duração — extraídas da performance pelo objeto [sigmund~], e convertidas

The image displays a musical score for section IV of the piece 'Tensio'. It features multiple staves, including a string quartet (Violins I and II, Violas, and Cellos/Double Basses) and woodwinds (Flutes, Oboes, and Clarinets). The score is marked with 'IVa' and includes various performance instructions such as 'pp' (pianissimo) and 'mf' (mezzo-forte). Two specific sections of the score are highlighted with colored boxes: a green box at the top and a blue box at the bottom, both indicating harmonic structures. The score is numbered 33 in the top right corner.

Figura 4.18: Engendramento de estruturas harmônicas na seção IV de *Tensio*

em valores MIDI — e do cálculo da probabilidade de ocorrência de cada uma das informações recuperadas a partir de processos estocásticos.

Para isso, Manoury implementa computacionalmente um procedimento denominado como cadeia de *Markov*: processo aleatório capaz de realizar previsões futuras com base apenas no estado atual, portanto independentemente dos eventos ocorridos no passado. Ou seja, condicionalmente ao estado atual de um dado sistema, suas evoluções futuras e passadas são independentes.

Dessa forma, a partir do cálculo da matriz de transição — responsável por atribuir a cada elemento de entrada um valor que descreve a probabilidade de ser precedido por outro elemento específico —, é gerada aleatoriamente uma nova sequência de valores MIDI (*pitch*, *velocity* e *duração*). Uma vez gerada, essa sequência de valores pode passar por alguns procedimentos de manipulação que influenciarão as características que as estruturas melódicas emergentes apresentarão: (1) andamento; (2) contrações e expansões intervalares; (3) controle da duração; (4) transposição; (5) faixa de amplitude mínima e máxima; (6) multiplicação intervalar que permite a sobreposição de fluxos melódicos paralelamente.

Em *Tensio*, Manoury emprega o material gerado pelo processo markoviano de duas formas. A primeira é através do mapeamento da sequência de alturas geradas como uma das três frequências principais para o cálculo do espectro gerado pelo *3F synthesis* — enquanto as outras duas frequências permanecem com valores fixos pré-determinados. Os demais valores retornados pela cadeia de *Markov* controlam a duração e a amplitude dos sons gerados pelo

banco de osciladores. Assim, essa estratégia resulta em estruturas espectrais/timbrísticas que se transformam a cada conjunto de valores retornados pela matriz de *Markov*.

A segunda estratégia consiste no mapeamento das sequência de valores gerados pelo processo markoviano para acionar e controlar um banco de dados de amostras sonoras de instrumentos musicais chamado *Synful Orchestra* desenvolvido por Eric Lindenmann. O *Synful Orchestra* é baseado em um processo denominado como *Reconstructive Phrase Modeling* (RPM) que...

...utiliza uma base de dados de frases de instrumentos orquestrais (...), passagens musicais completas que apresentam todos os tipos de articulações e fraseado. (...) *Synful Orchestra* utiliza um conjunto de algoritmos avançados para buscar na base de dados de frases do RPM por fragmentos que podem ser combinados para formar uma [nova] frase. (SYNFUL, 2019)

Desse modo, o compositor utiliza o *Synful Orchestra* como um performer virtual (“quarteto de cordas virtual”) que executa as informações musicais paramétricas resultantes do processo markoviano e as transformam em estruturas melódicas. Com essa estratégia, o quarteto de cordas é expandido por meios computacionais no nível *idiomático*. Isto é, o mecanismo do *Synful Orchestra* consiste precisamente nessa utilização de trechos instrumentais pré-gravados que são re combinados de maneira a criar novas frases e articulações sonoras. Explora, portanto, materiais sonoros típicos da literatura desses instrumentos, e uma vez que executam estruturas melódicas geradas a partir de um processo estocástico e de manipulações intervalares e temporais — reordenando sequências de alturas, intensidades e durações —, viabilizam a criação de rearranjos que podem estender tecnicamente e idiomáticamente os instrumentos do quarteto de cordas.

#### **b) *Traverser le réseau* (2014)**

*Traverser le réseau* (2014), para violoncelo e *live-electronics*, de Rael Toffollo ilustra outro tipo de estratégia no emprego da modelização computacional de procedimentos composicionais a partir do uso de técnicas de ML/MIR. Segundo Toffolo (2014), o título da peça [*atravessar a rede*] faz alusão à exploração criativa das *redes harmônicas* de Henri Pousseur, que é o ponto central da poética que guia tanto a escrita instrumental, quanto a concepção dos processos de transformação, espacialização e geração sonora presentes nesta peça.

Em *Traverser le réseau*, o compositor busca modelizar computacionalmente as *redes harmônicas* tais quais descritas por Henri Pousseur em (POUSSEUR, 2009), visando a geração de material para a criação de estruturas harmônico-espectrais as quais se engendrarão em tempo-real a partir da interação violoncelo-computador.

As *redes harmônicas* de Pousseur consiste em um procedimento composicional de caráter serial voltado à organização, manipulação e geração de material intervalar (harmônico e melódico), e consiste na projeção de um dado conjunto de alturas distribuídas em eixos que se caracterizam cada um como uma cadeia de um único intervalo (POUSSEUR, 1998, p. 249). Através dessa projeção de determinado aglomerado intervalar, é possível aplicar processos de “deformação” (transposição intervalar) nos eixos de maneira a gerar um novo conjunto de alturas.

Segundo Pousseur, a questão central que o levou à concepção das *redes harmônicas* foi:

como conseguir fazer “rimar” numa mesma composição uma citação de Gluck ou Monteverdi com uma de Webern (dois domínios gramaticais que me pareciam até então exatamente opostos e praticamente incompatíveis), como conseguir “conjugá-los”, encontrar-lhes funções comuns, e, para começar, estabelecer entre eles uma série de tipos intermediários suscetíveis de convencer o ouvido musical de que pertenceriam a um mesmo domínio mais geral? (POUSSEUR, 2009, p. 194)

Neste contexto, a fim de criar um método para classificação do material gerado pelas *redes harmônicas*, Toffollo implementa um algoritmo responsável por calcular o nível de tensão harmônica a partir da categorização de intervalos enquanto polar e apolar (COSTÈRE, 1954, 1962), e de modelos psicoacústicos voltados à percepção de intervalos consonantes e dissonantes baseados em Helmholtz (1954), Plomp e Levelt (1965) e Bigand, Parncutt e Lerdahl (1996b).

Para tanto, o compositor utiliza a extração dos níveis de rugosidade para construir o cálculo da taxa de tensão harmônica de cada conjunto de alturas. O modelo, consiste em calcular a taxa de rugosidade resultante de todos os parciais pertencentes a um ou mais sons e pode ser resumido em três passos: (1) gera-se as séries harmônicas de cada nota do aglomerado; (2) calcula-se a rugosidade entre cada um dos componentes da série harmônica de forma combinada; (3) por fim, os valores de rugosidades entre todos os pares de frequências são somados e dividido pela amplitude geral do aglomerado, resultando na taxa de tensão harmônica (TOFFOLO, 2014, p. 77).

Ao longo de *Traverser le réseau*, tal procedimento é utilizado de diversas maneiras, e operam tanto na criação de estruturas harmônicas/melódicas, quanto na geração de massas e texturas sonoras complexas a partir da interação do instrumentista com o computador. Salientamos novamente que este trabalho não tem o objetivo de realizar uma análise exaustiva das peças abordadas, o foco é investigar estratégias composicionais que façam uso de técnicas de ML/MIR. Desse modo, a seguir descreveremos brevemente como essa implementação é empregada na terceira parte da peça. Para uma explicação detalhada do funcionamento das *redes harmônicas*, da implementação dos algoritmos e da estruturação geral da peça, ver (TOFFOLO, 2014).

Na terceira parte de *Traverser le réseau*, Toffollo parte de um acorde retirado do movimento *Rondes Printanières* da *Sagração da primavera* (1913), de Igor Stravinsky que consiste na sobreposição de um acorde de Fá maior à um Lá bemol menor.<sup>13</sup> Esse acorde é projetado nas *redes harmônicas* com todas combinações intervalares de eixos possíveis, gerando assim, uma série de aglomerados harmônicos.

A partir do cálculo da taxa de tensão harmônica, esses aglomerados são ordenados gradualmente do menos tenso para o mais tenso. Uma vez classificados, cada um desses aglomerados é associado à uma faixa de valor da amplitude captada do violoncelo. Logo, ao passo em que a amplitude é extraída da performance instrumental em tempo-real, um aglomerado de alturas — que está correlacionado ao valor de amplitude extraído — é selecionado e enviado a um banco de osciladores que por sua vez sintetiza uma textura espectral. Segundo (TOFFOLO, 2014), com essa estratégia é possível criar objetos/massas sonoras com forte grau de variabilidade e com grande densidade espectral. A figura 4.19 mostra a representação da arquitetura deste procedimento.

---

<sup>13</sup>O acorde foi transposto uma terça maior abaixo para se adaptar à tectura do violoncelo (TOFFOLO, 2014, p. 99).

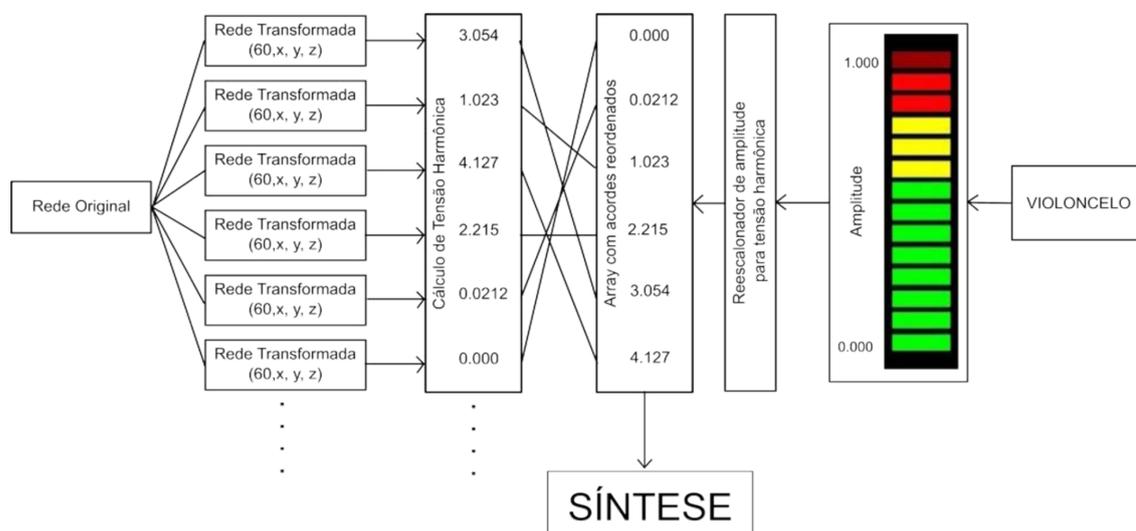


Figura 4.19: Arquitetura da implementação das *redes harmônicas* associadas ao cálculo da tensão harmônica, e ao mapeamento da amplitude do violoncelo em *Traverser le réseau* (TOFFOLO, 2014, p. 106)

## Capítulo 5

### Processo composicional: *Do corpo ao pó*

O presente capítulo tem como objetivo apresentar o processo criativo da principal peça composta ao longo do período de pesquisa, *Do corpo ao pó* para violão e *live-electronics*. Descreverei de forma mais pessoal algumas estratégias particulares de exploração criativa das ferramentas aqui estudadas, e também realizarei alguns apontamentos reflexivos/conceituais associados ao processo criativo da peça. Uma vez que a pesquisa apresenta um caráter criativo, é de fundamental importância discutir não só os aspectos técnicos, mas também o processo composicional e as implicações poéticas, técnicas e práticas envolvidas no projeto.

Com isso, evidencia-se uma outra frente desta pesquisa, para a qual tanto o estudo técnico das ferramentas de ML/MIR quanto a investigação de paradigmas na aplicação prática/criativa em peças do repertório da *live-electronics* serviram como subsídios para a realização de *Do corpo ao pó*. É necessário ressaltar que, juntamente aos estudos realizados nas etapas anteriores, as demandas poéticas envolvidas no projeto composicional desta peça também exerceram papel fundamental no estudo dos recursos técnicos relacionados à pesquisa, guiando o desenvolvimento e a busca por soluções técnicas que suprissem determinadas necessidades relativas à escrita instrumental e à estruturação da eletrônica em tempo-real.

O objetivo geral deste capítulo é proporcionar um entendimento sobre *Do corpo ao pó* de forma a contribuir com pesquisas dentro do campo criativo que busquem explorar, de alguma forma as ferramentas aqui estudadas. Com isso, espera-se também estimular outros compositores a encontrarem estratégias e soluções particulares às suas necessidades criativas. Tendo em vista que o intuito aqui é apresentar conceitos, estratégias e ferramentas

relacionados à peça, uma análise exaustiva está fora do escopo deste capítulo. Para a observação e uma eventual análise mais detalhada, a partitura completa de *Do corpo ao pó* encontra-se em anexo.

O capítulo está dividido da seguinte forma: a seção 5.1 aborda a poética da peça; a seção 5.2 descreve as principais características e introduz os princípios fundamentais que guiaram a estruturação de *Do corpo ao pó*; a seção 5.3 trata de aspectos relacionados à escrita musical; em *sec:5.4* abordamos a estruturação, elaboração e as funções formais da eletrônica em tempo real, e por fim, na seção 5.5 descrevemos as estratégias de aplicação das técnicas de ML/MIR.

## 5.1 Corpo e territorialidade: a poética no discurso da peça

*Do corpo ao pó*, para violão e *live-electronics* — dedicada ao violonista Gabriel Angelo — foi composta durante o ano de 2021 visando a exploração e a aplicação criativa de ferramentas específicas voltadas à escuta de máquina e à recuperação de informações musicais como parte desta pesquisa. A peça tem duração aproximada de 13:00 min, e busca através do uso criativo das ferramentas e técnicas de MIR/ML estabelecer um diálogo entre instrumento e computador no nível temporal, espacial, e timbrístico-morfológico. Esta seção irá introduzir os princípios básicos da peça, tais como poética e a sua relação com a estruturação e elaboração dos materiais que a compõe.

O título da peça faz referência à dissertação e ao livro *Do corpo ao pó: crônicas da territorialidade Kaiowá e Guarani nas adjacências da morte*, do antropólogo Bruno Morais (MORAIS, 2015). Este texto traz reflexões acerca do genocídio sofrido pelos povos *Kaiowá* e *Guarani*, no Mato Grosso do Sul no contexto dos conflitos fundiários, a partir de suas relações com o mundo, sendo a terra e o corpo, vivo ou morto, indissociáveis nesse debate (MARCURIO, 2018). A partir de reflexões sobre a violência, e a morte, Morais busca percorrer dois modelos de territorialidade; (1) a partir de concepções próprias dos indígenas das categorias territoriais; (2) a partir da expressão *tekoha* como contingências políticas do Estado que historicamente vincularam os *Kaiowá* e *Guarani* a um território específico.

A imagem poética que permeia a peça surge a partir deste debate em torno da territorialidade que se direciona à uma crítica *kaiowá* e *guarani* à colonização de suas terras. Esta crítica é traçada a partir do delineamento de uma genealogia própria do corpo, e de uma escatologia.

... durante a vida as atividades da pessoa vão sediando o corpo no terreno, vão mesclando-o com a terra, até o ponto em que ele se confunde com ela mesma. O corpo “faz parte do local”, ao mesmo tempo em que seu pó escatológico, sua poeira, se mistura com a terra, e em alguma dimensão é a terra ela mesma. A princípio, essa confusão material da pessoa com o espaço só está dada na relação com os parentes e familiares — o avô, o tio, o primo. Mas à medida que vive, come, dorme, casa, faz filho e, principalmente, morre, a pessoa vai se misturando progressivamente ao terreno (MORAIS, 2021, p. 10)

A vida e sobretudo a morte, são responsáveis por registrar as pessoas na terra até o ponto em que acabam se confundindo à ela. Por fim, o sujeito está definitivamente inscrito no espaço — o corpo passa a integrar o terreno de fato. Em outras palavras, a terra enquanto espaço é composta pelos corpos e pelas relações dos corpos com ela e dos corpos entre si, na vida e na morte — trocas materiais e imateriais entre os corpos e a terra.

Trazendo essa imagem poética para o contexto da criação, buscou-se evocar metáforas que remetesse à essas relações, bem como à noção de territorialidade descrita acima. Seja nas relações estabelecidas entre os processos interativos e os gestos instrumentais, ou na estruturação e desenvolvimento dos materiais musicais no nível micro e macro, diferentes desdobramentos dessa metáfora podem ser notados na concepção do projeto composicional como um todo. Nas próximas seções descreveremos os principais aspectos composicionais envolvidos tanto na concepção da escrita instrumental, quanto dos processos interativos de *Do corpo ao pó*. Ao mesmo tempo, buscaremos, tanto quanto possível, relacionar a metáfora da territorialidade com a escrita musical, mostrando o desenvolvimento e a busca por ferramentas e soluções técnico-musicais no processo criativo.

## 5.2 Sobre *Do corpo ao pó*

O processo de elaboração dos materiais e a estruturação do discurso musical em *Do corpo ao pó* se deu, primeiramente, pela seleção de um conjunto de elementos/motivos/unidades musicais e gestuais intrínsecas ao idiomatismo do violão. Essas unidades foram definidas de acordo com princípios musicais, como ritmo; acorde; perfil melódico, e/ou técnicas violinísticas, tais como *arpejo* e *rasgueado*<sup>1</sup>, deixando outros parâmetros musicais e técnicos em aberto. Essa estratégia tornou o material musical

---

<sup>1</sup>*Rasgueado* é uma técnica violonística característica da música flamenca, e é executada com a mão direita em padrões ritmicamente precisos e frequentemente rápidos.

maleável e com um maior grau de variabilidade e conectividade, permitindo sua inserção em diversos contextos musicais.

Uma vez elencadas, essas unidades foram justapostas e combinadas entre si de maneira a engendrar novas morfologias, conseqüentemente moldando a micro e macro forma da peça. Isto permitiu que as pequenas unidades morfológicas fossem utilizadas como blocos constituintes de ideias musicais mais extensas. Dependendo do grau de atomização dessas unidades, tem-se uma maior ou menor possibilidade de recombina-las e, com isso, obter estruturas mais complexas e variadas.

Este processo — apesar de não ser empregado de forma sistemática nesta peça — se assemelha de certa maneira com as *gramáticas musicais generativas*<sup>2</sup>, método de escrita utilizado por Philippe Manoury em muitas de suas peças (MANOURY, 2012). A partir deste processo, baseado nos sistemas-L<sup>3</sup> de Aristid Lindenmayer (LINDENMAYER, 1968), Manoury entrevê uma maneira de elaborar formas temporais que são baseadas na constituição de blocos morfológicos claramente definidos.

A primeira seção (figura 5.1) ilustra de forma simplificada esse aspecto da escrita instrumental de *Do corpo ao pó*. Neste trecho foram trabalhados os seguintes elementos e unidades musicais: *acordes*; *tambora*<sup>4</sup>; *apogiatura*; *arpejo* e *harmônico natural*. A medida que a unidade *acorde* — elemento em destaque desta seção — vai sendo reiterado se combina e se justapõe a outros elementos. É justamente a partir desse processo de justaposição e fusão das unidades que as morfologias sonoras/musicais vão emergindo.

Cada seção de *Do corpo ao pó* explora diferentes combinações e diferentes elementos enquanto blocos formadores de morfologias. De fato, a forma temporal de toda a peça é elaborada através deste processo de parametrização e incrustação de ideias musicais atomizadas, o que viabilizou a circunscrição desses blocos morfológicos de tal maneira que um certo espaço/repositório fosse delineado.

Na segunda seção os gestos que foram trabalhados anteriormente (seção I), são de certa forma diluídos com a inserção de dois novos elementos: *rasgueado* e *notas ligadas*. Ao longo de

---

<sup>2</sup>Vale ressaltar que, neste contexto, a *gramática generativa* explorada pelo compositor não guarda uma relação direta com a teoria homônima do linguista Noam Chomsky (2015) voltando-se antes à geração algorítmica de estruturas do que à análise de algo pré-existente como, por exemplo, uma linguagem (MANOURY, 2013, p.89).

<sup>3</sup>Processo algorítmico que permite descrever padrões recorrentes a partir de um conjunto restrito de parâmetros e regras, originalmente desenvolvido para modelar processos de crescimento e desenvolvimento de plantas.

<sup>4</sup>Técnica violonística que consiste em golpear as cordas com o polegar da mão direita na região da ponte ou da boca do instrumento, produzindo um som percussivo.

### Do corpo ao pó

para Gabriel Angelo

Vinicius Cesar

Figura 5.1: Primeira seção de *Do corpo ao pó*

toda a seção o *rasgueado* é explorado e apresentado reiteradamente com variações de dinâmica e estruturas rítmicas variadas. Este elemento é intercalado por alguns daqueles que foram desenvolvidos na seção anterior — e.g., acordes arpejados com harmônico, e a *tambora* —, resultando em morfologias que ao serem combinadas e justapostas compõe uma textura com alto grau de movimento e energia, mas que ainda mantém relações com as ideias que foram desenvolvidas na seção I (ver figura 5.2).

Figura 5.2: Exemplo de *rasgueado* combinado e intercalado pela *tambora*, *arpejos* e *notas ligadas* na segunda seção de *Do corpo ao pó*

Já na terceira seção a elaboração dos materiais se dá pela sobreposição de elementos — *acordes*; *arpejos*; *harmônicos*, e *pizzicato Bartok* —, juntamente ao desenvolvimento gestual das morfologias resultantes a partir de um processo de variação em que um simples gesto

periódico adquire um novo perfil através de uma ruptura na periodicidade (KOZU, 2021, p. 5). Isto é, cada linha paramétrica é tratada individualmente por um conjunto de transformações e manipulações, como inserção de pausas; mudanças de registro no perfil melódico; variações de dinâmica, articulação (modos de ataque), e estrutura rítmica. Esse processo acaba resultando em gestos com alto grau de complexidade rítmica e melódica. Desse modo, essa textura altamente densa constituída pela sobreposição de elementos aos poucos vai se rarefazendo até o ponto em que restam somente gestos compostos por *acordes longos* e *harmônicos* encerrando a peça na quarta seção.

Figura 5.3: Exemplo de textura polifônica complexa na seção três (acima) e de textura rarefeita na seção quatro (abaixo) de *Do corpo ao pó*

As noções de corpo e territorialidade serviram como uma imagem poética responsável por guiar os processos de elaboração e concepção dos materiais musicais presentes em *Do corpo ao pó*. Esses conceitos dialogam diretamente com os procedimentos de escrita musical descritos acima ao estabelecer relações de fusão e combinação das unidades musicais entre si. Com este processo de amálgama; superposição; justaposição, e diluição dos elementos musicais em jogo buscou-se evocar a metáfora das relações dos corpos entre si e entre a terra, ao mesmo tempo que possibilitou entrever maneiras de circunscrever e sedimentar um espaço formal constituído pelos elos criados na elaboração dos materiais, assim remetendo ao conceito de territorialidade segundo os *Kaiowá*.

### 5.3 Escrita musical

A natureza de alguns gestos instrumentais demandou a elaboração de diferentes estratégias de notação de tal maneira que possibilitassem expressar com a maior clareza possível as ações que o instrumentista deveria executar no violão. A discussão em torno dos

aspectos notacionais da peça se dará a partir de exemplos que enfatizem algumas das particularidades de escrita e ilustrem suas soluções de notação.

Partindo do prisma das organizações temporais destacamos que a peça faz uso de três estratégias diferentes de notação: (1) métrica; (2) proporcional; (3) aberta. A notação métrica é predominante ao longo da peça sendo aplicada em construções formais mais rigorosas — como por exemplo as seções III e IV. Em certa medida, esse tipo de notação reflete o próprio caráter dos processos de elaboração e cálculo das estruturas musicais desenvolvidas nestas seções, e imprime um certo rigor na execução rítmica.

Por outro lado, o uso da notação proporcional tem o objetivo de proporcionar uma certa liberdade rítmica/temporal ao instrumentista na construção de certos gestos, o permitindo explorar de forma mais particular as estruturas temporais que lhe parecem mais interessantes nestes momentos específicos. A figura 5.4 mostra um exemplo de notação métrica — compasso 63 ao 65 —, e de notação proporcional — a partir da indicação *non messuare*. Embora haja indicações de durações rítmicas bem definidas nos trechos com notação proporcional, vale ressaltar que o instrumentista não tem a necessidade de tocá-las de forma precisa, essas figuras são sugestões de durações que servirão apenas como referência.

Figura 5.4: Exemplo de notação métrica (compasso 63 ao 65) e proporcional na terceira seção de *Do corpo ao pó*

Já a notação aberta visa conceder mais liberdade de interpretação do que a notação proporcional — expandindo a liberdade interpretativa para outros parâmetros além do temporal, como timbre; articulação; dinâmica; modos de ataque, etc. Nestes trechos específicos, fica a cargo do instrumentista como os gestos serão tocados. Por essa razão, tais trechos sempre apresentam indicações textuais sugestivas que buscam direcionar o instrumentista na construção das morfologias sonoras, mas sem manter os parâmetros da performance fixos. A figura 5.5 exemplifica o uso da notação aberta na peça. Neste trecho há uma pausa que deve durar aproximadamente dez segundos, logo em seguida um acorde que deve ser tocado dez vezes variando a dinâmica de *pianíssimo* à *fortíssimo* com durações que variam entre semibreve e colcheia, timbre de *sul tasto* à *sul ponticello*, e alternando entre a

nota Mi e o harmônico natural si — ambas tocadas na primeira corda. Com isso, baseado nas indicações textuais o instrumentista decide com total liberdade como o gesto será desenhado.

Figura 5.5: Exemplo de notação aberta na segunda seção de *Do corpo ao pó*

Outra particularidade na notação de *Do corpo ao pó* é o uso de tablaturas, que muitas vezes servem apenas como um guia para especificar digitações de uma forma mais precisa, e em outras atuam como um complemento que indica particularidades técnicas na execução de certos gestos instrumentais com precisão — por exemplo, o sentido da mão direita na execução dos *rasgueados*. No trecho exemplificado pela figura 5.6, a notação com a seta indicando o sentido do *rasgueado* no gesto em *ralentando* (final do sistema) induz o instrumentista a destacar o caráter de desaceleração do gesto, enfatizando o acorde em *fff*. Portanto, há aqui uma direcionalidade latente proporcionada pelo uso deste tipo de notação que conduz o intérprete de forma natural na construção do gesto a medida em que os acordes são tocados da corda mais grave para mais aguda.

Figura 5.6: Exemplo de uso de tablatura para indicar o sentido da mão direita na execução do *rasgueado* em *Do corpo ao pó*

No que diz respeito a notação da parte eletrônica, optou-se pela indicação apenas dos números referentes ao disparo dos eventos relacionados aos processos de transformação e geração sonora. Devido ao caráter interativo que os processos baseados nos métodos de escuta de máquina e recuperação de informações musicais apresenta, a notação das morfologias resultantes da interação entre os gestos instrumentais e o processamento em tempo-real se tornam, em certa medida, inviáveis. Sabemos a natureza das informações musicais/sonoras que serão utilizadas na estruturação dos processos, contudo, muitas vezes não é possível prever o comportamento que apresentarão. Há uma dificuldade em escrever

com notação musical tais sonoridades, faltam-nos recursos. Assim, da mesma forma quando trabalhamos com instrumentos musicais, não é a sonoridade que notamos, mas dispositivos de disparo de sons, com o que nossas partituras traduzem muito pouco o som que estamos procurando realizar (FERRAZ, 2013, p. 58). A figura 5.7 mostra a notação dos números correspondentes a cada evento da eletrônica em tempo-real.

Figura 5.7: Exemplo de notação do disparos eventos relacionados à eletrônica em tempo-real em *Do corpo ao pó*

## 5.4 Estruturação e funções formais da eletrônica em tempo-real

Em *Do corpo ao pó*, a eletrônica em tempo-real foi concebida pensando na criação de um diálogo bilateral com a escrita instrumental. Esta inter-relação se dá por meio de uma correspondência estabelecida entre os atributos morfológicos do violão e a própria natureza dos procedimentos de manipulação sonora. A modelização destes processos associados a diferentes estratégias de emprego das técnicas de ML/MIR, bem como a criação de uma correlação simbólica com os conceitos de *corpo* e *territorialidade* foram os principais elementos que guiaram a realização da eletrônica.

Neste sentido, a correspondência entre os materiais musicais advindos da escrita instrumental e dos processos interativos é trabalhada a partir da criação de uma morfologia da interação. Esta estrutura interativa se dá através da fusão e contraste entre os materiais a partir da exploração de estágios transitórios entre estes dois fenômenos.

A fusão e o contraste fundamentam-se em aspectos estruturais do material sonoro, isto é, para haver fusão é necessário que haja a manutenção e transferência de atributos de um material ao outro, enquanto o contraste é fundamentado pela ausência de transferência — ou uma transferência que mantém uma mínima relação entre os materiais. Destacamos que em *Do corpo ao pó* essa relação de transferência não está necessariamente vinculada apenas a aspectos timbrístico (espectrais), mas também à outras características estruturais do comportamento

sonoro. Dessa forma, apontamos três dimensões exploradas nos procedimentos de fusão e contraste: (1) timbrístico; (2) temporal; (3) espacial.

As relações estabelecidas na dimensão do timbre ocorrem através do emprego tanto de processos transformativos, quanto de procedimentos generativos. Com isso buscamos explorar variados graus de aproximação e distanciamento das características espectrais do violão. Os processos utilizados no tratamento dos materiais por este viés se dão pela manipulação das relações entre as componentes frequenciais; pela correlação de atributos extraídos do violão a um corpus sonoro pré-analisado, e pela síntese sonora a partir de informações recuperadas da performance. Essas relações acontecem de forma predominante nas seções I, III, e IV.

No nível temporal as manipulações ocorrem exclusivamente através da criação de linhas de atraso (seções I e III), de estruturas rítmicas geradas a partir da reprodução reiterada de pequenos trechos gravados em tempo-real (seções I e II); e de processos palindrômicos (seções II e III). O principal objetivo é estabelecer associações bilaterais entre as estruturas temporais do violão e dos sons provenientes da eletrônica, e explorá-las através de diferentes modos de manipulação/transformação — defasagem, reiteração, e inversão.

Já a dimensão espacial busca, a partir de um sistema de difusão sonora projetar as morfologias emergentes da interação violão-eletrônica em espaços sonoros distintos. Para cada processamento empregado na peça há tipos de projeção espacial que ora estão em relação de proximidade entre si, ora em total oposição. Com isso ressaltamos a dissociação espacial entre um gesto e a percepção de sua resultante sonora, o que muitas vezes pode induzir à dúvidas em relação a causalidade da ação do instrumentista.

É importante destacar que em muitos momentos da peça há o emprego de processos que operam manipulações em mais de uma das dimensões descritas acima. Por exemplo a criação de linhas de atraso para cada *bin* do espectro de frequência, e a defasagem na projeção da trajetória espacial.

#### **a) Estrutura do *patch***

A eletrônica de *Do corpo ao pó* foi programada como um *patch* no ambiente de computação musical *Pure Data*, e consiste em seis tipos de processamento associados a um sistema de difusão espacial. Todos os controles relacionados à performance são automatizados, desde o acionamento dos processamentos até a configuração dos parâmetros

de cada evento. A entrada do sinal do microfone (captação do violão) e a saída de cada um dos processamentos são enviadas separadamente para o módulo de espacialização, assim cada processamento representa uma fonte sonora na difusão espacial.

Os eventos da eletrônica, indicados pelo seu respectivo número na partitura instrumental, estão configurados para serem disparados acionando a barra de espaço do teclado do computador. Os comandos dos *presets* contendo os valores paramétricos relativos aos eventos estão armazenados em uma “partitura eletrônica” que consiste em um conjunto de mensagens. Cada seção da peça possui uma “partitura” individual, o último evento de cada seção configura e inicia a “partitura” da próxima seção tornando assim a transição automatizada. Apesar de todos os comandos do *patch* serem executados automaticamente, os sinais de entrada e saída do microfone, bem como a saída de cada processamento possuem controles que podem ser manipulados manualmente durante a performance, caso haja necessidade.

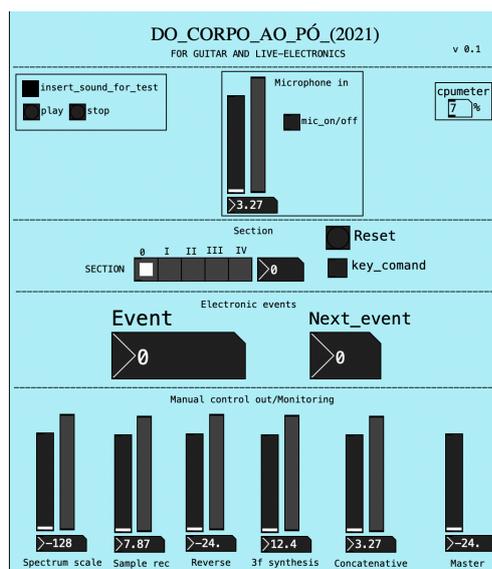


Figura 5.8: Interface do *patch* de *Do corpo ao pó*

## b) Processamentos

As técnicas de transformação e geração sonoras utilizadas em *Do corpo ao pó* consistem em seis tipos de processamento: *spectrum scale*; *delay espectral*; *síntese concatenativa*; *gravador e reproduzidor de amostras*; *ataque palíndromo*; *3f synthesis* e *espacialização*, sendo que alguns destes são implementados de forma combinada. Nesta seção

descreveremos apenas a estruturação e o funcionamento desses processos, as estratégias de emprego das técnicas de ML/MIR serão abordadas separadamente na seção ??.

O *spectrum scale* consiste em um processamento espectral que re-escala o espectro de frequências a partir da manipulação de um fator de multiplicação. Primeiramente o sinal é transformado e representado no domínio da frequência; em seguida as partes imaginária e real são multiplicadas por um fator escalar, e por fim é realizada a transformada inversa de *Fourier*. O *spectrum scale* foi utilizado na peça com o intuito de criar diferentes níveis de contraste timbrístico. Quanto maior for o fator de multiplicação maior será o contraste em relação ao timbre original. A implementação deste processamento utiliza objetos da biblioteca *Pd Spectral Toolkit*<sup>5</sup> (BAKER; ERBE, 2013). A figura 5.9 mostra um exemplo de implementação deste processamento.

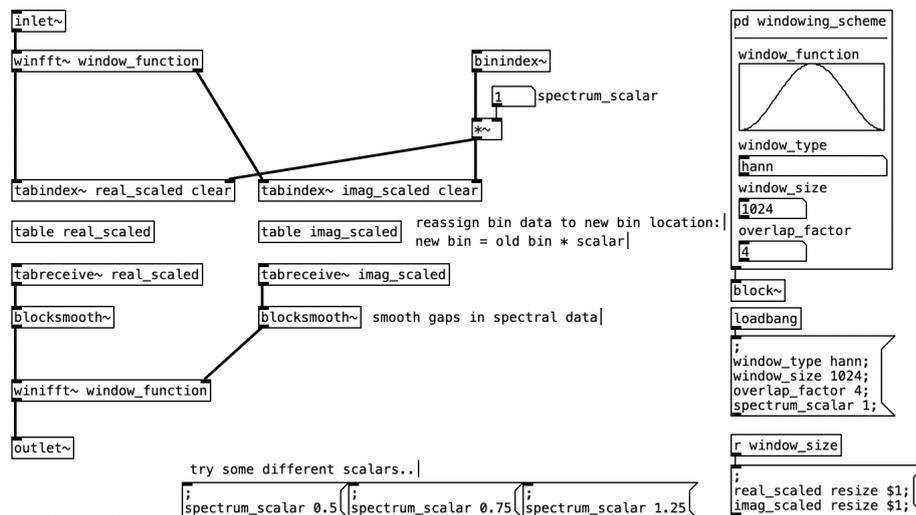


Figura 5.9: Exemplo de Patch do *Spectrum scale*

O *delay spectral* também é um processo realizado no domínio da frequencial e trata-se da criação de linhas de atraso para cada *bin* de frequência. Em *Do corpo ao pé*, este processo é empregado de forma combinada com o *spectrum scale* em momentos específicos onde há a presença de acordes longos, por exemplo. Quando empregado em conjunto, o *delay spectral* recebe o sinal do *spectrum scale* antes de passar pela transformada inversa, aplica as linhas de atraso para cada *bin* e só então é realizada a transformada inversa do sinal. A figura 5.10 mostra o *patch* do *delay spectral* implementado em *Pure Data*.

*Síntese concatenativa* é uma técnica voltada à geração sonora realizada a partir da relação de similaridade entre segmentos de dois sons. O resultado é um som sintetizado

<sup>5</sup>A biblioteca e exemplos de *patches* estão disponíveis em: <https://github.com/cooperbaker/Pd-Spectral-Toolkit>

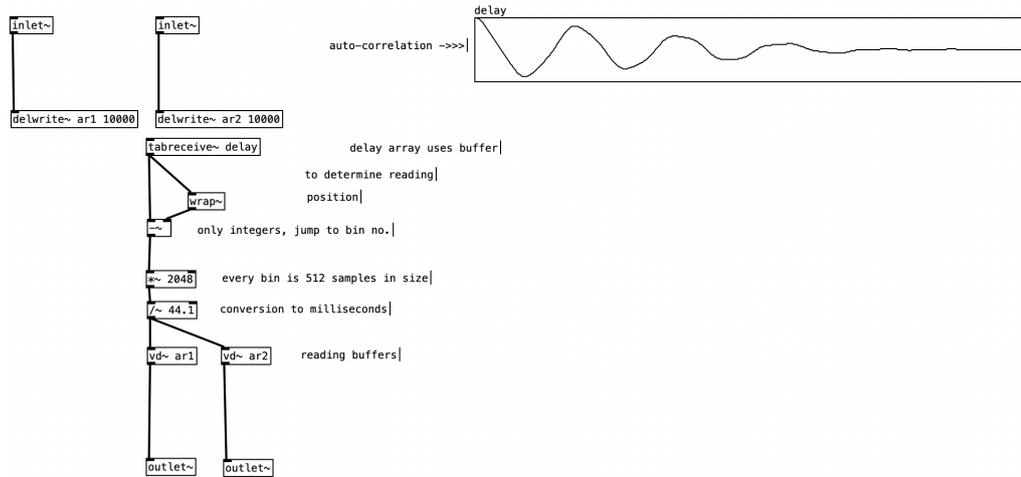


Figura 5.10: Implementação do *Delay espectral*

constituído pela concatenação de grãos<sup>6</sup> de áudio selecionados de um corpus sonoro de acordo com a semelhança entre seus atributos e aqueles do “som-alvo”. Este processo opera através da segmentação do corpus sonoro e do “som-alvo” em pequenos grãos que são analisados por meio de técnicas de ML/MIR. (GUATIMOSIM, J.; PADOVANI; GUATIMOSIM, C., 2021, p. 3). Em *Do corpo ao pó*, esse processo é implementado utilizando objetos da biblioteca *timbreID*<sup>7</sup> A figura 5.11 mostra uma representação da implementação deste processo.

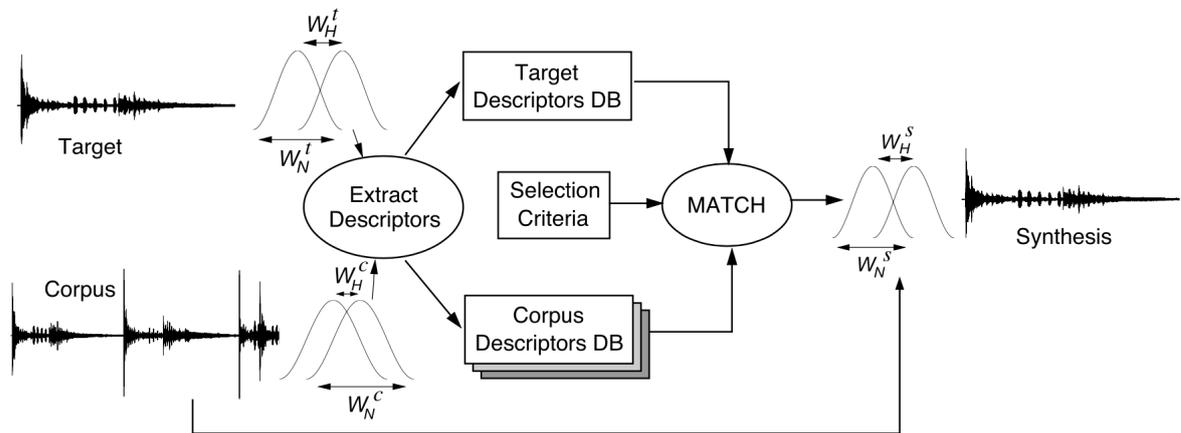


Figura 5.11: Diagrama do funcionamento da *síntese concatenativa* (STURM, B., 2006, p. 51)

O *gravador e reproduzidor de amostras* é um processo de ordem temporal e consiste na gravação em tempo-real de amostras sonoras da própria performance e na sua reprodução de

<sup>6</sup>Embora se assemelhe à síntese granular, a *síntese concatenativa* geralmente sequencia unidades de som de longa duração de forma monofônica, enquanto a síntese granular se concentra na criação de texturas através da sobreposição de diferentes densidades de grãos com durações curtas (STURM, B., 2006, p. 48).

<sup>7</sup>O *patch* que foi utilizado como modelo para implementação da *síntese concatenativa* está disponíveis em: <https://github.com/mxa/timbreID-examples>.

acordo com alguns parâmetros de controle. Este processo foi implementado, com algumas modificações, a partir do módulo *Trevor sampler*<sup>8</sup>, utilizado por Philippe Manoury em *Pluton*, para piano e *live-electronics*. A reprodução das amostras gravadas é acionada a partir de três metrônomos, cada metrônomo é configurado com intervalos regulares distintos produzindo um efeito de polirritmia em três níveis. Há também um parâmetro de controle que recebe valores em *midicents*, esses valores estão ligados à transposição das amostras (CLARKE; DUFEU; MANNING, 2020).

O *ataque palíndromo* é outro processo transformativo realizado na dimensão temporal e trata-se da gravação de trechos da performance e da sua reprodução no sentido inverso. Sua implementação é simples, há um parâmetro que define em milissegundos qual o intervalo de tempo a amostra deverá ser reproduzida após o acionamento do processo, e.g., se o intervalo for de 2000 milissegundos, significa que o trecho a ser reproduzido de trás para frente terá duração de 2000 milissegundos contados a partir do disparo do processo.

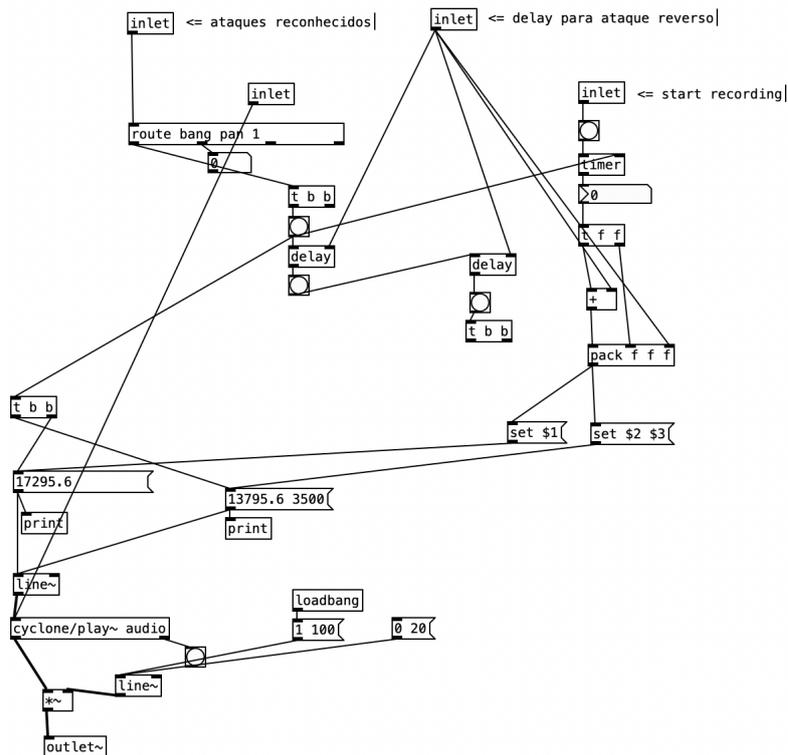


Figura 5.12: Implementação do *ataque palíndromo* em *Pure Data*

O *3f synthesis* é um processo de síntese realizado a partir de três frequências bases. Este procedimento já foi descrito no capítulo 4, contudo vale ressaltar que para a realização de *Do corpo ao pó* este processo foi adaptado e implementado como um objeto externo para

<sup>8</sup>O *patch* de *Pluton* está disponível em: <http://msp.ucsd.edu/pdrp/pdrp-13/>

*Pure Data*, por José Henrique Padovani<sup>9</sup>. As modificações realizadas consistem no aumento do número de frequência máxima do espectro para até 64, na alteração dos valores mínimo e máximo das frequências geradas (-20000Hz até 20000Hz), e na inclusão de um parâmetro que controla o decaimento da amplitude em dB para cada banda de 1kHz.

A *espacialização* é implementada por meio de um método desenvolvido por Gerzon (1972) denominado *Ambisonics*. Trata-se de um sistema voltado à simulação de campos sonoros (pantonía e perifonia)<sup>10</sup>, a partir de uma representação física/psicoacústica do espaço sonoro no domínio dos harmônicos (funções de decomposição do espaço), e da ondas planas (modelo de projeção e decomposição do campo sonoro).

A operação de codificação fornece as informações espaciais necessárias para que um som possa ser percebido em um certo ponto no espaço. Após os processos realizados no domínio dos harmônicos, o campo sonoro pode ser processado no domínio das ondas planas e decodificado para um conjunto de alto-falantes. Uma vez que estas duas operações são independentes, é possível realizar vários processos entre elas. Assim, viabilizando a adaptação do campo sonoro a diferentes configurações de alto-falantes. Neste sentido, para cada ordem do sistema *Ambisonics* há um número mínimo de canais necessários. Para o espaço circular (2d) temos: primeira ordem, 3 canais, e segunda ordem, 5 canais. Para o espaço esférico (3d) temos: primeira ordem 4 canais; segunda ordem 9 canais, e terceira ordem 16 canais (FARIA, 2005, p. 36). Em *Do corpo ao pó*, a espacialização em *Ambisonics* foi implementada utilizando os objetos da biblioteca *CEAMMC*.<sup>11</sup>

## 5.5 Estratégias de aplicação das técnicas de ML/MIR

As técnicas de ML/MIR foram aplicadas de forma extensiva na eletrônica de *Do corpo ao pó* abarcando desde a implementação dos processos de transformação, síntese, e espacialização, até a manipulação paramétrica de procedimentos. Através do emprego de tais ferramentas, buscou-se alcançar uma interatividade que valorizasse as particularidades da performance, principalmente nos momentos em que a escrita instrumental proporciona uma maior liberdade ao intérprete. Também buscou-se a modelização de procedimentos de tal maneira que fosse possível estabelecer um diálogo com os próprios processos composicionais

<sup>9</sup>O objeto está disponível em: <https://github.com/zepadovani/tresf>

<sup>10</sup>Pantonía refere-se ao campo sonoro circular (2d), e perifonia ao campo esférico (3d) (FARIA, 2005, p. 34)

<sup>11</sup>A biblioteca está disponível em: <https://github.com/ceammc>

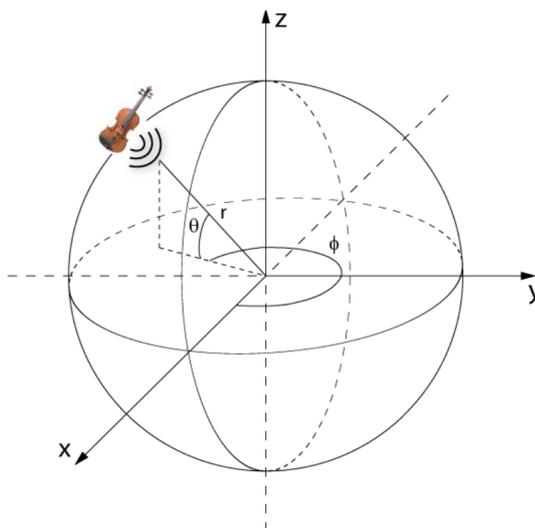


Figura 5.13: Coordenadas esféricas utilizadas na codificação do campo sonoro (THOMAZ, 2007, p. 63)

empregados na peça. Com isso, as estratégias de aplicação das técnicas de ML/MIR vieram a contribuir de maneira decisiva nas soluções técnicas relacionadas a fusão e contraste entre os gestos instrumentais e as estruturas emergentes da eletrônica.

A fim de melhor descrever as estratégias aplicadas em *Do corpo ao pó* as dividiremos em três frentes: (1) disparo de processos condicionado pelo reconhecimento de estruturas e eventos musicais específicos; (2) controle de parâmetros dos processos; (3) modelização de algoritmos e procedimentos generativos. É válido destacar que esta divisão foi realizada apenas para organizar as estratégias composicionais empregadas na peça de forma mais sistematizada, essas três frentes podem ocorrer concomitantemente e de forma fluida.

#### a) Reconhecimento de eventos musicais e disparo de processos

Em *Do corpo ao pó* o reconhecimento de eventos e estruturas musicas foi implementado com o auxílio do objeto [timbreID], presente na biblioteca homônima desenvolvida para *Pure Data* por William Brent. Este objeto funciona como um mecanismo de armazenamento e roteamento que pode agrupar e ordenar dados relacionados a atributos sonoros salvos na sua memória (BRENT, 2009a, p. 2). O uso do objeto [timbreID] foi empregado de duas formas na peça, no reconhecimento de acordes e na detecção de harmônico natural e *pizzcato Bartok*.

No processo de reconhecimento de acordes foram utilizados os descritores *chroma*; *spectral spread*, e os coeficientes *bark*. Os dados relacionados a estes descritores foram extraídos de todos os acordes presentes na peça, e enviados ao objeto [timbreID] para que pudesse

ser treinado. Assim, foi possível clusterizar os acordes em cinco grupos e associá-los cada um à uma posição no espaço.

A figura 5.14 ilustra a aplicação desta estratégia na subseção **IB** de *Do corpo ao pó*. O evento 6 da eletrônica notado em algarismo arábico abaixo da partitura, aciona o *gravador e reprodutor de amostras* — que grava o acorde presente no compasso 11 (destacado em vermelho) e o reproduz reiteradamente com três intervalos temporais distintos. Cada acorde tocado pelo violão ao longo da seção é analisado pelo conjunto de descritores, os dados recuperados são comparados com aqueles armazenados previamente pelo [timbreID], que por sua vez classifica os acordes como pertencentes a um dos cinco grupos. Dessa forma, cada grupo de acordes é responsável por posicionar a amostra sonora reproduzida pelo *gravador e reprodutor de amostras* em um ponto específico do espaço sonoro circular — atrás; esquerda; frente; direita e atrás.

The image shows a musical score for guitar and electronics. The score is divided into two systems. The first system contains measures 9 through 13. Measure 11 is highlighted with a red box and labeled 'IB' and '6'. The second system contains measures 16 through 19. The score includes various chords (VI, VII, VIII, IX, I, II, III, IV, V), dynamics (p, mf, f, mp), and articulation (T, L.v., harm., @). The time signature is 6/4.

Figura 5.14: Exemplo de reconhecimento de acordes para controle da espacialização

O emprego desta estratégia busca realizar uma aproximação com a poética da peça ao estabelecer uma correlação entre as classes de acordes e os vários pontos do espaço sonoro. A associação entre a estratégia composicional e a poética está justamente na circunscrição de um espaço sonoro que se constitui através das relações estabelecidas entre cada um dos acordes e entre os acordes e o próprio espaço. Tal qual a noção de *territorialidade*, as atividades do instrumentista vão circunscrevendo as estruturas sonoras geradas pela eletrônica no espaço, e em certa medida passa a ser incorporada a ele.

Já na detecção dos harmônicos e do *pizzicato Bartok* foi utilizado uma combinação dos descritores *spectral centroid*; *spectral slope*; *spectral spread*; *spectral skewness*, e *spectral roll-off*. Da mesma forma que o procedimento de reconhecimento dos acordes, os dados extraídos

foram enviados ao [timbreID] e clusterizados em dois grupos: harmônicos e *pizzcato Bartok*. A figura 5.15 mostra o *patch* em *Pure Data* que detecta os dois eventos musicais.

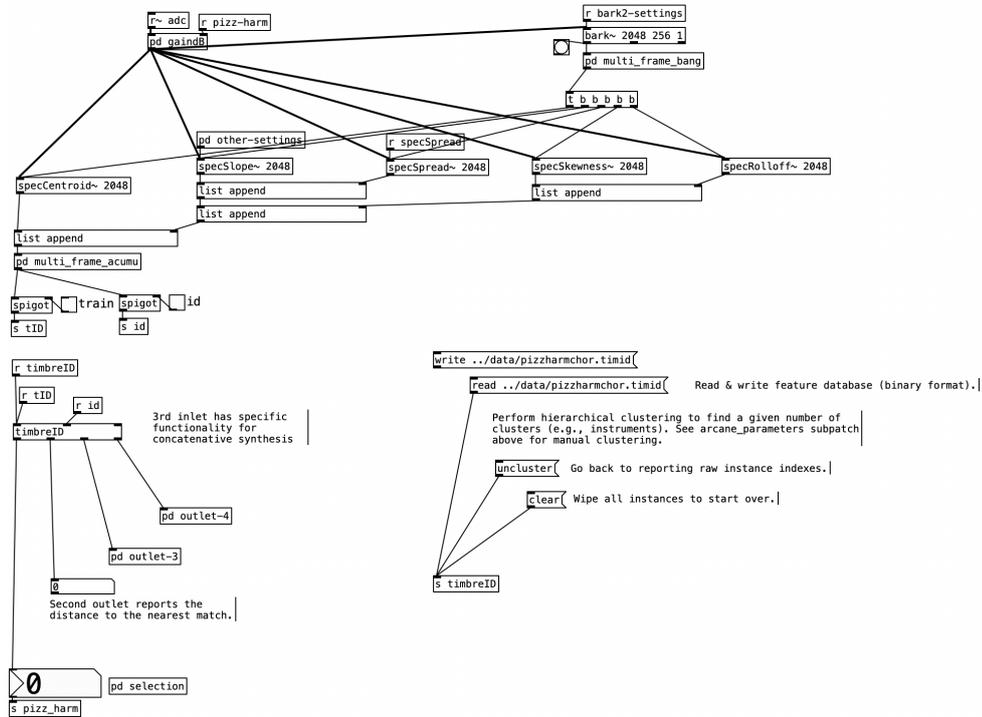


Figura 5.15: Implementação do *patch* para detecção de harmônicos e *pizzcato Bartok*

Esta estratégia é empregada no fim da seção **III** e na seção **IV**. O *3f synthesis* é disparado a cada vez que o [timbreID] reconhece os harmônicos naturais e os *pizzcato Bartok*. Entretanto, nem todos os harmônicos e *pizzcatos* detectados disparam o processo de síntese. O acionamento deste processo está condicionado a um índice de probabilidade que varia em uma escala de 0 a 1. Por exemplo, se o índice de probabilidade for igual a 1 todos harmônicos e *pizzcatos* detectados irão disparar o processo, por outro lado, nenhum dos eventos detectados irão disparar a síntese se o índice de probabilidade for igual a 0.

A figura 5.16 mostra um exemplo de emprego desta estratégia. Na subseção **IIIc** o índice de probabilidade é igual a 0.3, isto é, apenas 30% dos eventos detectados irão disparar o cálculo da síntese. Ao longo de **IIIc** e **IIIId** o índice de probabilidade aumenta gradualmente até atingir valor igual a 1 no início da seção **IV**. Com isso tem-se um aumento da presença da estrutura sonora gerada pelo *3f synthesis* acarretando em um adensamento. O caráter aleatório deste processo faz com que cada performance apresente particularidades no que diz respeito ao disparo do *3f synthesis*, assim, apenas o comportamento geral de adensamento é mantido.

Para o reconhecimento de ataques foi utilizado o objeto [bark~] (BRENT, 2011), também presente na biblioteca *timbreID*. Este objeto é um detector de *onset* que faz uso da

Figura 5.16: Emprego de estratégias de detecção de harmônicos e *pizzicato Bartok* no fim da seção III e no início da seção IV de *Do corpo ao pó*

escala *Bark*. O objeto usa um algoritmo muito semelhante ao [bonk~], no entanto o uso da escala *Bark* pode melhorar os resultados em algumas situações onde há ressonância entre os *onsets*. Como mostra a figura 5.17, nas subseções IIIA e IIIB o [bark~] é utilizado para reconhecer os ataques produzidos pelos acordes acentuados, destacados em vermelho. Para cada ataque detectado é disparado o *ataque palindromo*, responsável por gravar o acorde e reproduzi-lo no sentido inverso, produzindo assim, uma espécie de ressonância em *crescendo* para cada acorde acentuado.

## b) controle de parâmetros dos processos

As estratégias voltadas ao controle de parâmetros são empregadas nos processos *spectrum scale* e *delay spectral*. No *spectrum scale* foi utilizada a envoltória do *cepstrum* para controlar o fator escalar. A cada disparo do processo é extraído o *cepstrum* do violão e seus valores são armazenados em um *array*. Um [phasor~] é responsável por “varrer” esse *array* de tal forma que os valores são lidos como dados de controle que irão manipular o fator escalar. Na figura 5.18 é possível observar o *patch* responsável por extrair o *cepstrum*, e também o módulo que lê os valores armazenados no *array* “*envelope-spectrum*”. A implementação do *cepstrum* foi baseada no *patch* desenvolvido por Padovani (2009b) <sup>12</sup>.

<sup>12</sup>O *patch* está disponível em: <http://josehenriquepadovani.com/etc.html>

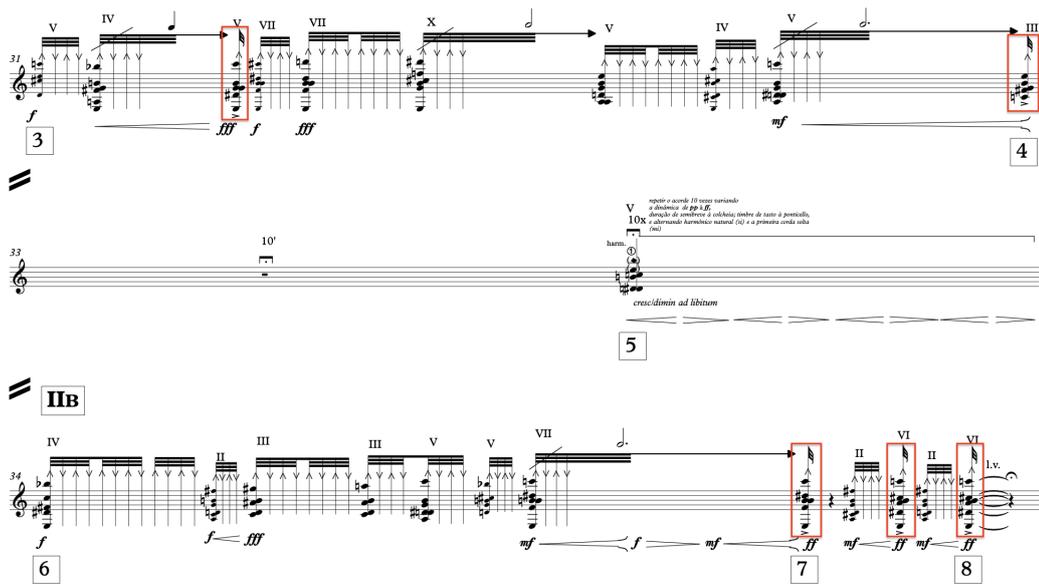


Figura 5.17: Reconhecimento dos acordes acentuados (vermelho) na seção II

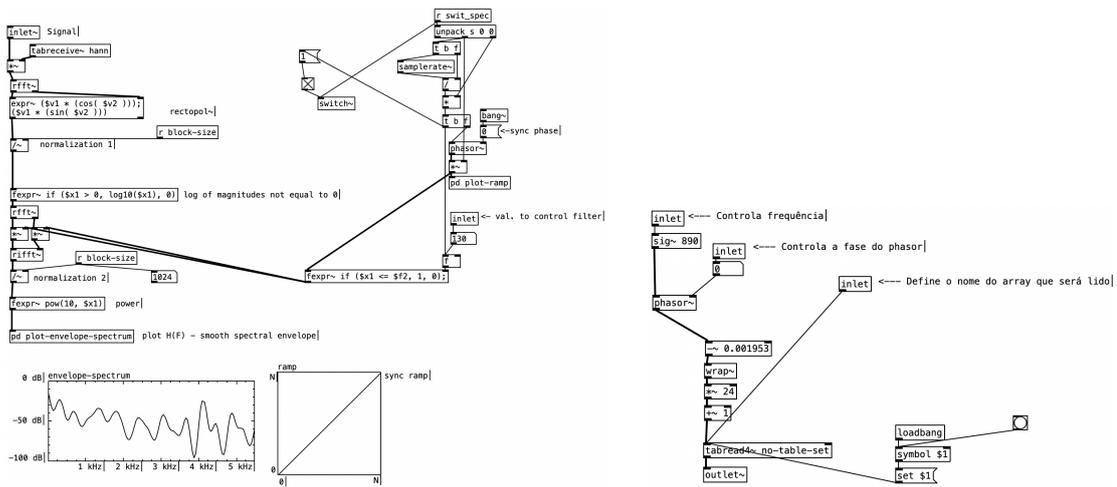


Figura 5.18: Implementação do *cepstrum* em *Pure Data* e do *patch* responsável por ler os dados extraídos

Para controlar as linhas de atraso de cada *bin* de frequência do *delay spectral* utilizamos a função de *auto-correlação especial normalizada* (*Special Normalized Autocorrelatoin – SNAC*) (MCLEOD, 2009) extraída de acordes específicos ao longo das seções I e II. O *patch* que calcula a SNAC foi implementado a partir de um exemplo da biblioteca [Helmholtz~]<sup>13</sup> e pode ser observado na figura 5.19.

Em alguns momentos específicos de *Do corpo ao pó* o sinal processado pelo *spectrum scale* passa pelo *delay spectral*. Essa combinação de processamentos visa expandir a paleta de possibilidades timbrísticas e morfológicas produzidas pela eletrônica. Como mostra a figura

<sup>13</sup>O objeto, assim como um texto demonstrando a sua implementação está disponível em: <https://www.katjaas.nl/helmholtz/helmholtz.html>

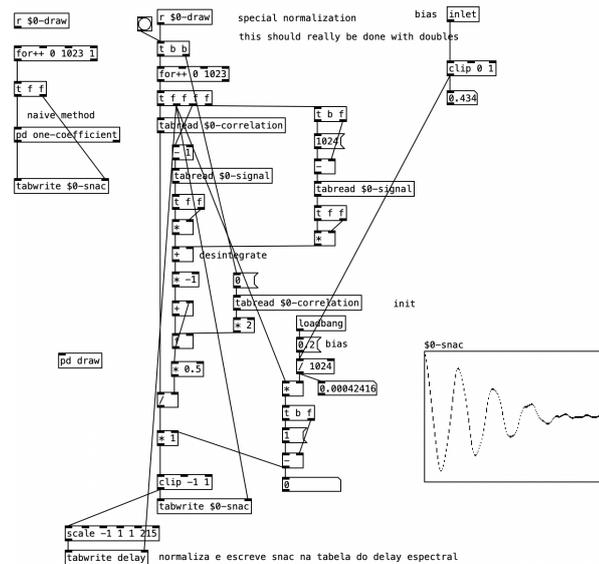


Figura 5.19: Implementação da SNAC em *Pure Data*

5.20, os eventos 5, 6, 8, 11 e 13 da seção I (destacados em vermelho) extraem o *cepstrum* em tempo-real a partir do sinal do violão e o mapeia para ser utilizado como controle do fator escalar, já para controlar as linhas de atraso do *delay espectral* é extraído o SNAC nos eventos 4, 7, 10 e 12 (destacados em azul). As setas indicam que a curva de auto-correlação extraída dos acordes destacados em azul é utilizada para manipular as linhas de atraso do *delay espectral* nos acordes destacados em vermelho.

Há uma relação entre a concepção desta estratégia e a noção de *territorialidade*. O próprio conceito de auto-correlação pode ser associado às relações que o corpo estabelece com ele mesmo no processo de constituição do território. Quando extraímos a curva de auto-correlação de um acorde estamos medindo de alguma forma sua relação consigo mesmo. Já a estratégia de utilizar os atributos extraídos do violão para controlar processos evoca a idéia de trocas materiais e imateriais entre os corpos e o terreno. Há neste processo uma idéia de troca que se acentua nos momentos em que ocorre a utilização de atributos extraídos de um gesto para manipular outro gesto. De acordo com (MORAIS, 2015), essas trocas fazem com que o indivíduo e o território se fundam até o ponto em que se tornam indissociáveis.

### c) Algoritmos e procedimentos generativos

A modelização de algoritmos e procedimentos foi utilizada em três aplicações distintas. A primeira trata-se da extração das três frequências mais proeminentes do sinal para realizar o cálculo do espectro produzido pelo *3f synthesis*. Nessa implementação

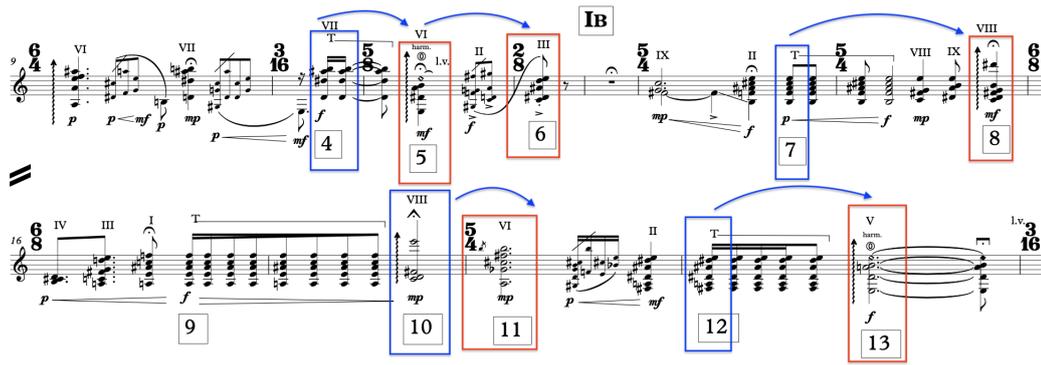


Figura 5.20: Exemplo de utilização do *cepstrum* e da *Special Normalized Autocorrelatoin* para controlar processos na seção I

utilizamos o [sigmund~] para extrair as frequências. O final da seção IV ilustra a aplicação desta estratégia. Neste trecho o cálculo do *3f synthesis* é acionado sempre que há a presença de harmônicos, *pizzicato Bartok* e acordes (indicados em vermelho na figura 5.21). Desse processo emerge uma estrutura sonora produzida pelo acionamento da síntese que adquire um perfil rítmico próximo ao representado na figura 5.21 (abaixo).

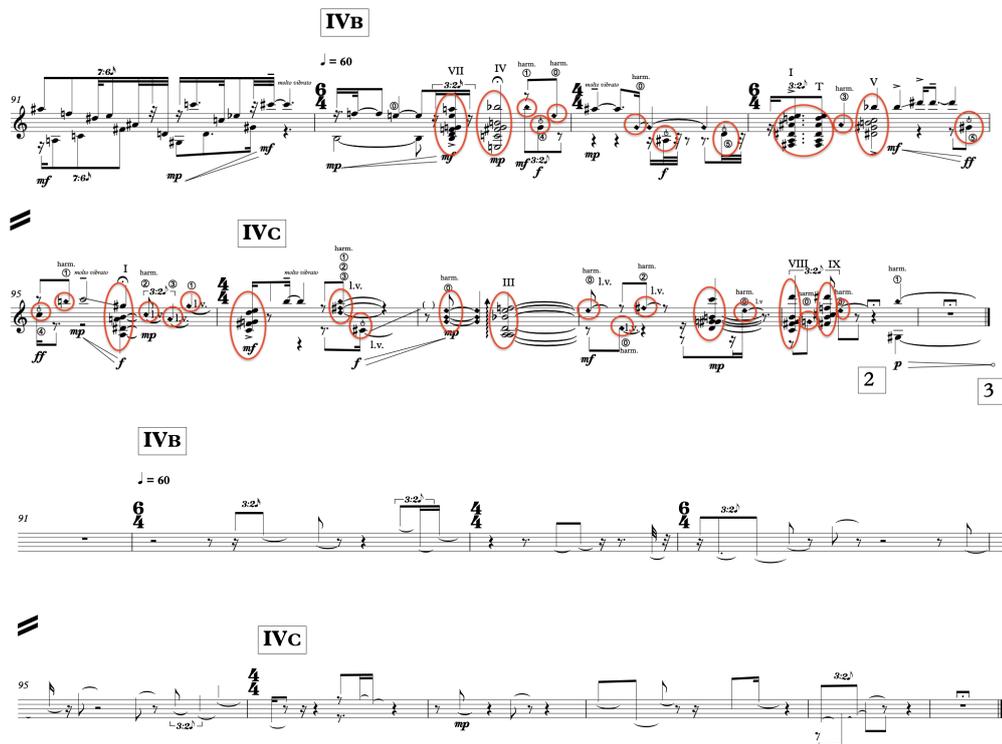


Figura 5.21: Utilização do *3f synthesis* para gerar estruturas sonoras/musicais na seção IV

A segunda aplicação foi voltada a concepção da *síntese concatenativa*. Para tanto, utilizamos o objeto [timbreID], uma vez que ele apresenta uma função específica para implementação deste tipo de processo. Em *Do corpo ao pó*, usamos como corpus sonoro sons

de diversos instrumentos de cordas executando diferentes técnicas, e como “som-alvo” o próprio sinal de entrada do violão. Para analisar estes sons foi utilizado o descritor *chroma*. Em alguns momentos específicos da seção III é acrescentada uma linha de atraso na *síntese concatenativa* com o objetivo de criar um maior movimento a partir da defasem, assim evitando a emergência de morfologias estáticas em relação aos gestos do violão. Também foi utilizado o *RMS* extraído do instrumento em tempo-real para controlar a amplitude de saída da síntese, mantendo assim uma correlação entre instrumento e eletrônica não só no nível das alturas, a partir do emprego do *chroma*, mas também no nível da amplitude, com a extração da envoltória de energia. O *patch* utilizado na peça foi baseado no exemplo de *síntese concatenativa* desenvolvido por Brent<sup>14</sup>.



Figura 5.22: *Patch* de síntese concatenativa retirado dos exemplos do timbreID

Por fim, a terceira aplicação consiste na extração da curva de *roughness* para derivar intervalos musicais. Esta estratégia foi empregada com o intuito de gerar uma sequência de alturas que irão modular os sons reproduzidos *gravador e reproduzidor de amostras*. Dada a curva de *roughness* de um certo espectro, é possível derivá-la em intervalos musicais medidos em *cents*. Pontos mínimos representam alinhamentos de parciais, e podem ser correlacionados com a sensação de intervalo consonante. Por outro lado, os pontos máximos podem ser correlacionados com a percepção de intervalos dissonantes (PORRES, A.; PIRES, 2009, p. 192). A implementação deste procedimento utiliza o objeto [*roughness*] da biblioteca *PSYCHO*, desenvolvida por Alexandre Torres Porres (2019) para *Pure Data*.

O emprego dessa estratégia pode ser ilustrado por um trecho da seção II reproduzido na figura 5.24. No início da subseção IIc o evento 10 aciona o *gravador e reproduzidor de*

<sup>14</sup>O *patch* está disponível em: <https://github.com/mxa/timbreID-examples>

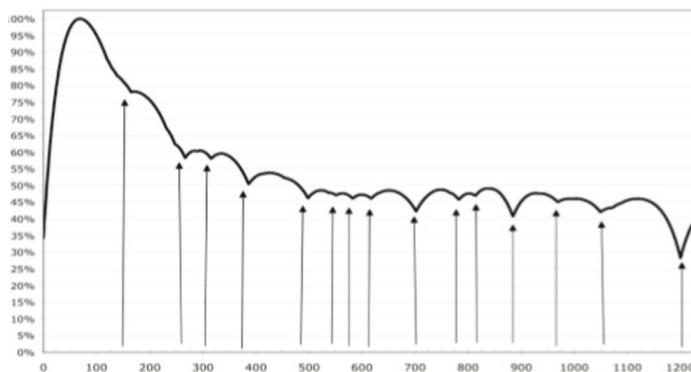


Figura 5.23: Curvas de *roughness* no eixo vertical, e intervalos em *cents* no eixo horizontal (PORRES, A.; PIRES, 2009, p. 193)

*amostras* que grava e reproduz o primeiro acorde tocado e ao mesmo tempo extrai a sua curva de *roughness*. Os intervalos escalares derivados da análise do *roughness* são escolhidos aleatoriamente com três espaços de tempo distintivos e mapeados como fator de transposição dados em *cents*. Com isso tem-se uma textura que se constitui pela sobreposição de duas morfologias sonoras/musicais: os acordes tocados reiteradamente pelo violonista com variações de dinâmica, duração e timbre; e as estruturas polirrítmicas arpejadas derivadas da eletrônica em tempo-real.

Figura 5.24: Extração da curva de *roughness* de um acorde para derivar intervalos escalares na seção IIc

## Considerações Finais

Esta pesquisa realizou um estudo acerca de estratégias composicionais no emprego de técnicas de ML e MIR enquanto ferramentas voltadas à criação no contexto da *Música interativa/live-electronics*. Refletindo a abordagem metodológica adotada, o trabalho divide-se em três frentes de pesquisa, voltadas: (1) à fundamentação teórica e à revisão de técnicas e de trabalhos com temáticas análogas àquelas propostas na pesquisa; (2) à investigação de estratégias criativas ecléticas relacionadas ao emprego dos recursos de interesse da pesquisa; (3) à pesquisa criativa, que consistiu na realização de um projeto composicional.

A fundamentação teórica compreende duas partes, a primeira consiste na discussão acerca do recorte da pesquisa. Nesta etapa, realizamos uma reflexão sobre os termos *música mista*, *live-electronics* e *sistemas musicais interativos/música interativa* em perspectiva às transformações tecnológicas apontando as particularidades estéticas de cada uma destas práticas musicais. Contextualizamos e definimos os conceitos relacionados às áreas de ML e MIR enquanto campos de pesquisa interdisciplinares. E por fim, abordamos as implicações técnicas, poéticas e éticas do emprego dessas ferramentas de forma a aproximá-las de um contexto exploratório e criativo dentro do escopo da *live-electronics* na atualidade.

A segunda parte da fundamentação teórica envolve a revisão técnica dos descritores de áudio abordados ao longo do trabalho de tal modo que foram divididos em duas famílias: descritores de baixo-nível, e descritores de alto-nível. Também foram introduzidas algumas medidas, escalas, técnicas e conceitos relacionados aos métodos de processamento digital de sinais que são utilizados na implementação dos descritores de áudio. Com isso, além de organizar uma revisão que possibilitou uma compreensão mais global dos arcabouços técnicos envolvidos na concepção dessas ferramentas, também possibilitará um maior entendimento da estruturação de alguns processos de transformação sonora que são comumente utilizados na *live-electronics*, como filtros; janelamento; transformada de *Fourier*; *vocoder*, etc.

A investigação de estratégias criativas se deu a partir de observações gerais acerca das principais características da escrita musical na *música interativa/live-electronics* associadas ao emprego de ML/MIR. Ao apontar algumas aplicações práticas e abordagens criativas em um conjunto de peças do repertório, circunscrevemos de forma geral as estratégias observadas categorizando-as em quatro paradigmas composicionais: *mapeamento*; *gatilho*; *notação*, e *procedimento*. Buscamos ilustrar tais paradigmas a partir de um repertório heterogêneo no que diz respeito à estética, instrumentação e emprego criativo das técnicas estudadas, o que nos forneceu um panorama geral das soluções técnicas e das diferentes estratégias composicionais que vêm sendo empregadas desde as primeiras experimentações com escuta de máquina e recuperação de informações musicais.

Por fim, fizemos uso de um amplo conjunto de ferramentas e estratégias que foram estudadas ao longo desta pesquisa, resultando em uma peça para violão e *live-electronics*, *Do corpo ao pó*. A peça desenvolve seu discurso musical orientada por uma imagem poética que evoca a noção de territorialidade segundo o povo *Kaiowá*. Tanto a escrita instrumental, como a modelização dos processos interativos foram concebidos como metáforas musicais que remetessem ao conceito de corpo e territorialidade.

No que diz respeito ao uso das técnicas de ML/MIR, buscamos criar e desenvolver estratégias criativas e soluções técnicas individuais que estabeleceram um diálogo entre gestos instrumentais e morfologias sonoras explorando aspectos temporais, timbrísticos e espaciais nessa interação. Além da elaboração de estratégias específicas para o projeto composicional, também adaptamos alguns dos procedimentos observados nas peças abordadas na fase de investigação criativa. É o caso da síntese *3f*, utilizada por Manoury em *Tensio*, que foi adaptada e implementada como objeto externo para *Pure Data* por José Henrique Padovani.

No decorrer de todo o processo criativo o violonista Gabriel Angelo, a quem a peça é dedicada, colaborou com sugestões acerca da escrita instrumental, notações de gestos, e com o envio de gravações de trechos da peça. Essa dinâmica colaborativa contribuiu decisivamente na elaboração de soluções técnicas, na experimentação e exploração de estratégias criativas com os descritores de áudio, e também possibilitou uma escrita que fosse coerente com as particularidades técnicas, gestuais e sonoras do violão.

Durante todo o desenvolvimento do trabalho foi fundamental estabelecer um diálogo entre todas as frentes de pesquisa. Essa relação de reciprocidade entre as fases de realização

deste trabalho favoreceu o bom andamento da pesquisa e possibilitou que fossem atingidos os objetivos de forma satisfatória. No entanto houve alguns desafios que dificultaram a realização do trabalho, e que de certa forma acabaram influenciando de forma decisiva o seu desenvolvimento.

A escolha do repertório abordado no capítulo 4 foi um dos obstáculos enfrentados devido a uma série de restrições relacionadas ao acesso a materiais, como partituras; gravações; *pitches*, e bibliografia acerca das peças. Primeiramente, essa dificuldade se deve ao fato de que as áreas de nosso interesse — apesar de terem se expandido consideravelmente nas últimas duas décadas e já estarem bem consolidadas — ainda são relativamente pouca exploradas no contexto da música interativa/*live-electronics*. Portanto, houve uma dificuldade em encontrar peças que abordassem de maneira mais sistematizada o uso de técnicas de ML/MIR, e que pudessem contribuir de alguma forma com este estudo. Ao mesmo tempo, a busca por um repertório abrangente e que se distanciassem, em certa medida, dos grandes centros e escolas/estéticas europeias e estadunidenses aumentou consideravelmente o desafio da escolha das peças.

Outro fator preponderante foi a dificuldade em entrar em contato com alguns compositores/as, o que impediu o acesso a materiais que não são disponibilizados por eles, que não estão disponíveis em base de dados de livre acesso, ou que somente podem ser acessados por meio da comercialização. É notório que este tem sido um dos grandes problemas enfrentados no desenvolvimento de pesquisas que necessitam de acesso a materiais, principalmente relacionados a produções musicais atuais. Enquanto no mundo acadêmico cada vez mais a circulação de *papers* migra pro modelo *open access*, no mundo da música de concerto ainda é predominante o acesso pago a partituras. No caso de peças que contam com códigos/*live-electronics*, soma-se a isso a ainda incipiente prática de publicar este tipo de material de maneira aberta em repositórios de livre acesso.

Também é importante ressaltar que desde o início da pesquisa houve a intenção de abordar peças em sua maioria compostas por mulheres. Essa escolha foi motivada por uma preocupação em mostrar a relevância das produções artísticas de compositoras na atualidade e, com isso, contribuir de alguma forma com a inserção de seus pensamentos composicionais no debate. Apesar de enfrentarmos as dificuldades relacionadas a esse modelo ainda bastante fechado de distribuição de partituras, *pitches*, aplicativos e códigos, conseguimos incluir no

trabalho um número significativo de peças compostas por mulheres. Das onze obras abordadas no capítulo 4, cinco são de compositoras.

Já no que diz respeito aos aspectos técnicos e criativos ligados ao projeto composicional, o grande desafio foi a elaboração de estratégias de aplicação das técnicas estudadas de forma conceitual e que pudesse estabelecer relações com o próprio processo criativo em todas as suas instâncias. Conforme discutido no capítulo 1 ( subseção 1.3.1), as técnicas de ML/MIR estão bem consolidadas e desenvolvidas, seus métodos são utilizados em uma variedade de aplicações. Contudo, ainda há a necessidade de pensar seu emprego, sobretudo, na criação musical.

Em um contexto onde há uma certa carência de trabalhos que se concentrem na exploração criativa das técnicas de escuta de máquina e recuperação de informações musicais, o presente trabalho buscou contribuir ao propor uma abordagem centrada nos aspectos criativos da aplicação destas técnicas no contexto específico da *música interativa* e da *live-electronics*. Acreditamos que nossa contribuição, juntamente com um número crescente de trabalhos que buscam investigar técnicas atuais de escuta de máquina a partir de uma perspectiva direcionada à criação e aos processos criativos, pode vir a ser de utilidade a compositores/as que tenham interesse em explorar tais técnicas em seus projetos composicionais.

A revisão técnica dos descritores de áudio e dos conceitos; técnicas; modelos, e medidas envolvidos na sua implementação foi realizada de forma bastante sistemática e organizada, o que pode facilmente se tornar um interessante material didático para consultas em disciplinas voltadas à composição com suportes tecnológicos e computação musical. Além disso, as questões acerca do emprego de ML/MIR no contexto da exploração criativa e musical que foram levantadas na subseção 1.3.1 podem suscitar discussões e reflexões teóricas prolíferas e com grande potencial de desdobramento em outras pesquisas e trabalhos.

Ao abordar tais técnicas a partir de uma perspectiva concentrada nos processos criativos na *live-electronics*, esperamos que o presente trabalho seja útil a outros/as pesquisadores/as e compositores/as, suscitando reflexões teóricas pertinentes a tais práticas, colaborando na compreensão das características de diferentes abordagens voltadas à composição para sistemas interativos/*live-electronics* na atualidade, e delimitando e

explicitando abordagens poéticas, ferramentas técnicas e problemáticas da exploração da ML/MIR nesses contextos.

De fato, a aplicação destas técnicas no contexto da *música interativa* e da *live-electronics* traz consigo uma série de questões pertinentes. Como manipular, interpretar, e tratar os dados extraídos para que seja possível elaborar procedimentos e estratégias composicionais mais contextuais. E de que modo os dados extraídos podem dialogar de alguma forma com o próprio procedimento de escrita musical. Estas são questões que foram levantadas durante todo o desenvolvimento deste trabalho, mas que não puderam ser exploradas satisfatoriamente.

Visando problematizar e abordar estas questões de forma estruturada, este trabalho aponta para uma pesquisa de doutorado que buscará explorar o emprego criativo de técnicas específicas de ML/MIR e aquelas voltadas à *Auditory scene analysis* (ASA) e *Computational auditory scene analysis* (CASA) associadas a métodos de aprendizado de máquina visando a concepção de procedimentos autônomos. Acreditamos que um estudo que considere o uso conjunto dessas ferramentas poderá viabilizar tanto uma reflexão teórica pertinente, quanto a criação de subsídios que possam levar a possíveis soluções para os problemas técnicos, estéticos e poéticos que o uso criativo dessas ferramentas no contexto da criação musical contemporânea podem trazer.

## Referências bibliográficas

AMATRIAIN. **An Object-Oriented Metamodel for Digital Signal Processing with a Focus on Audio and Music**. 2005. Tese (Doutorado) – Universitat Pompeu Fabra. DOI: <http://hdl.handle.net/10803/667051>.

ANTONIOU, A. **Digital Signal Processing: Signals, Systems and Filters**. New York: McGraw-Hill, 2005. ISBN 978-0-07-145424-7.

ARZT, A.; FROSTEL, H. et al. Artificial intelligence in the concertgebouw. In: PROCEEDINGS of the 24th International Conference on Artificial Intelligence. Buenos Aires, Argentina: AAAI Press, jul. 2015. (IJCAI'15), p. 2424–2430. ISBN 978-1-57735-738-4. Acesso em: 16 fev. 2021.

ARZT, A.; WIDMER, G.; DIXON, S. Automatic Page Turning for Musicians via Real-Time Machine Listening. In: ECAI 2008 - 18th European Conference on Artificial Intelligence. Patras, Grécia: [s.n.], jan. 2008. DOI: [10.3233/978-1-58603-891-5-241](https://doi.org/10.3233/978-1-58603-891-5-241).

AUDI, R.; AUDI, P. R. **The Cambridge Dictionary of Philosophy**. [S.l.: s.n.], 2017. ISBN 978-1-78684-333-3. Disponível em: <http://0-search.credoreference.com/emu.londonmet.ac.uk/content/title/cupdphil?institutionId=5061>.

BAKER, C.; ERBE, T. PD spectral Toolkit. In: ICMC. [S.l.: s.n.], 2013.

BELLO, J. P. et al. A tutorial on onset detection in music signals. English (US). **IEEE Transactions on Speech and Audio Processing**, v. 13, n. 5, p. 1035–1046, set. 2005. Publisher: Institute of Electrical and Electronics Engineers Inc. ISSN 1063-6676. DOI: [10.1109/TSA.2005.851998](https://doi.org/10.1109/TSA.2005.851998). Disponível em: <https://nyuscholars.nyu.edu/en/publications/a-tutorial-on-onset-detection-in-music-signals>. Acesso em: 27 mai. 2021.

BIGAND, E.; PARNCUTT, R.; LERDAHL, F. Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. en. **Perception & Psychophysics**, v. 58, n. 1, p. 125–141, jan. 1996. ISSN 1532-5962. DOI: [10.3758/BF03205482](https://doi.org/10.3758/BF03205482). Disponível em: <https://doi.org/10.3758/BF03205482>. Acesso em: 16 nov. 2021.

BIGAND, E.; PARNCUTT, R.; LERDAHL, F. Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. **Perception & psychophysics**, v. 58, p. 124–41, jan. 1996. DOI: [10.3758/BF03205482](https://doi.org/10.3758/BF03205482).

BÖCK, S.; WIDMER, G. Local Group Delay Based Vibrato and Tremolo Suppression for Onset Detection. en, nov. 2013. DOI: [10.5072/zenodo.244150](https://doi.org/10.5072/zenodo.244150). Disponível em: <https://sandbox.zenodo.org/record/244150>. Acesso em: 7 mar. 2022.

BÖCK, S.; WIDMER, G. Maximum filter vibrato suppression for onset detection. In: IN Proc. of the 16th International Conference on Digital Audio Effects (DAFx-13. [S.l.: s.n.], 2013.

BOGDANOV, D. et al. **ESSENTIA: an Audio Analysis Library for Music Information Retrieval**. [S.l.: s.n.], nov. 2013. Journal Abbreviation: Proceedings - 14th International Society for Music Information Retrieval Conference Publication Title: Proceedings - 14th International Society for Music Information Retrieval Conference.

BOGERT, B. P.; HEALT, M. J. R.; TUKEY, J. W. The Quefreny Analysis of Time Series for Echoes : Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum and Saphe Cracking. **Proc. Symposium on Time Series Analysis, 1963**, John Wiley & Sons, 1963.

BORN, G. Diversifying MIR: Knowledge and Real-World Challenges, and New Interdisciplinary Futures. en. **Transactions of the International Society for Music Information Retrieval**, v. 3, n. 1, p. 193–204, out. 2020. Number: 1 Publisher: Ubiquity Press. ISSN 2514-3298. DOI: 10 . 5334 / tismir . 58. Disponível em: <<http://transactions.ismir.net/articles/10.5334/tismir.58/>>. Acesso em: 11 abr. 2021.

BOULEZ, B. **Penser la musique aujourd'hui**. [S.l.]: Gallimard, 1964. Disponível em: <<https://www.librairie-gallimard.com/livre/9782070709014-penser-la-musique-aujourd-hui-pierre-boulez/>>. Acesso em: 27 abr. 2021.

BOWLER, I.; PURVIS, A.; MANNING, P. D.; BAILEY, N. New Techniques for a Real-time Phase Vocoder. In: PROCEEDINGS of the 1990 International Computer Music Conference, ICMC, 1990, Glasgow, Scotland, September 10-15, 1990. [S.l.]: Michigan Publishing, 1990. Disponível em: <<http://hdl.handle.net/2027/spo.bbp2372.1990.048>>.

BRENT, W. A perceptually based onset detector for real-time and offline audio parsing. In: ICMC. [S.l.: s.n.], 2011.

BRENT, W. A Timbre Analysis and Classification Toolkit for Pure Data. en, p. 6, 2009.

BRENT, W. Cepstral Analysis Tools for Percussive Timbre Identification. en, p. 7, 2009.

BRENT, W. **Physical and perceptual aspects of percussive timbre**. 2010. Tese (Doutorado) – UC San Diego. Disponível em: <<https://escholarship.org/uc/item/5bx4j1fj>>. Acesso em: 9 mar. 2022.

BROWN, J. Determination of the meter of musical scores by autocorrelation. **Journal of the Acoustical Society of America (JASA)**, v. 94, p. 1953, out. 1993. DOI: 10 . 1121/1 . 407518.

BULLOCK, J. **Implementing Audio Feature Extraction in Live Electronic Music**. 2008. Tese (Doutorado) – Birmingham City University, Birmingham. Acesso em: 1 jul. 2021.

BULLOCK, J. LIBXTRACT: A LIGHTWEIGHT LIBRARY FOR AUDIO FEATURE EXTRACTION. **Proceedings of the International Computer Music Conference, 2007. 11th International Society for Music Information Retrieval Conference**, jan. 2007.

BURG, J.; ROMNEY, J.; SCHWARTZ, E. **Digital Sound & Music: Concepts, Applications, and Science**. Portland, Oregon: Franklin, Beedle & Associates Inc, out. 2016. ISBN 978-1-59028-274-8.

BYRD, D.; FINGERHUT, M. The History of ISMIR - A Short Happy Tale. **D-lib Magazine**, v. 8, n. 11, 2002.

CANNAM, C. **The vamp audio analysis plugin api: A programmer's guide**. [S.l.: s.n.], 2009.

CANNAM, C.; LANDONE, C.; SANDLER, M.; BELLO, J. [S.l.: s.n.], out. 2006.

CELLA, C.-E. Music Information Retrieval and Contemporary Classical Music: A Successful Failure. en. **Transactions of the International Society for Music Information Retrieval**, v. 3, n. 1, p. 126–136, set. 2020. Number: 1 Publisher: Ubiquity Press. ISSN 2514-3298. DOI: 10.5334/tismir.55. Disponível em: <<http://transactions.ismir.net/articles/10.5334/tismir.55/>>. Acesso em: 6 abr. 2021.

CELLA, C.-E.; BURRED, J. J. **Advanced sound hybridizations by means of the theory of sound-types**. [S.l.: s.n.], ago. 2013.

CHADABE, J. Interactive Composing: An Overview. **Computer Music Journal**, The MIT Press, v. 8, n. 1, p. 22–27, 1984. ISSN 0148-9267. DOI: 10.2307/3679894.

CHENG, O.; ABDULLA, W.; SALCIC, Z. Performance Evaluation of Front-End Processing for Speech Recognition Systems. en, p. 43, 2005.

CHOMSKY, N. **Aspects of the Theory of Syntax**. Cambridge, Massachusetts: The MIT Press, 2015. OCLC: 908127961. ISBN 978-0-262-52740-8.

CHOWDHURY, G. G. **Introduction to Modern Information Retrieval**. [S.l.]: Facet Publishing, 2010. Google-Books-ID: cN4qDgAAQBAJ. ISBN 978-1-85604-694-7.

CLARKE, M.; DUFEU, F.; MANNING, P. **Inside Computer Music**. 1st edition. New York: Oxford University Press, set. 2020. ISBN 978-0-19-065965-3.

COLLINS, N. A Comparison of Sound Onset Detection Algorithms with Emphasis on Psychoacoustically Motivated Detection Functions. English. In: disponível em: <<https://www.aes.org/e-lib/browse.cfm?elib=13079>>. Acesso em: 28 mai. 2021.

COLLINS, N. Automatic Composition of Electroacoustic Art Music Utilizing Machine Listening. en. **Computer Music Journal**, v. 36, n. 3, p. 8–23, set. 2012. ISSN 0148-9267, 1531-5169. DOI: 10.1162/COMJ\_a\_00135. Disponível em: <[http://www.mitpressjournals.org/doi/10.1162/COMJ\\_a\\_00135](http://www.mitpressjournals.org/doi/10.1162/COMJ_a_00135)>. Acesso em: 14 mai. 2019.

COLLINS, N. Machine Listening in SuperCollider. In: THE SuperCollider Book. Cambridge, Mass: MIT Press, 2011. ISBN 978-0-262-23269-2.

COLLINS, N. SCMIR: A SuperCollider Music Information Retrieval Library, jan. 2011.

COLLINS, N. **Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems**. Jan. 2006. Tese (Doutorado) – University of Cambridge.

COLLINS, N. Using a Pitch Detector for Onset Detection. In: ISMIR 2005, 6th International Conference on Music Information Retrieval, London, UK, 11-15 September 2005, Proceedings.

[S.l.: s.n.], 2005. p. 100–106. Disponível em: <<http://ismir2005.ismir.net/proceedings/1008.pdf>>.

COLLINS, N. Imaginary Landscape – Electronics and Live Performance, 1989 and 1939. In: p. 191–194.

CONT, A. A Coupled Duration-Focused Architecture for Real-Time Music-to-Score Alignment. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 32, p. 974–987, jun. 2010. Publisher: Institute of Electrical and Electronics Engineers. DOI: 10.1109/TPAMI.2009.106. Disponível em: <<https://hal.archives-ouvertes.fr/hal-00479737>>. Acesso em: 16 fev. 2021.

CONT, A. Antescofo: Anticipatory Synchronization and control of Interactive parameters in Computer Music. In: PROCEEDINGS of International Computer Music Conference. Belfast: [s.n.], 2008. Acesso em: 1 mar. 2021.

CONT, A. On the Creative Use of Score Following and Its Impact on Research. **Proceedings of the 8th Sound and Music Computing Conference, SMC 2011**, jul. 2011.

COOK, P. R. **Real Sound Synthesis for Interactive Applications**. Zeroth. [S.l.]: A K Peters/CRC Press, jul. 2002. ISBN 978-0-429-09152-0. DOI: 10.1201/b19597.

CORNELIS, O.; SIX, J.; HOLZAPFEL, A.; LEMAN, M. Evaluation and Recommendation of Pulse and Tempo Annotation in Ethnic Music. **Journal of New Music Research**, v. 42, n. 2, p. 131–149, jun. 2013. Publisher: Routledge. eprint: <https://doi.org/10.1080/09298215.2013.812123>. ISSN 0929-8215. DOI: 10.1080/09298215.2013.812123. Disponível em: <<https://doi.org/10.1080/09298215.2013.812123>>. Acesso em: 11 abr. 2021.

COSTÈRE, E. **Lois et styles des harmonies musicales**. [S.l.]: Presses Universitaires de France, 1954.

COSTÈRE, E. **Mort ou transfiguration de l'harmonie**. Paris: Presses Universitaires France, 1962. ISBN 978-2-13-040382-1.

DALTER, V. Q. **Algoritmos para estimação da frequência fundamental**. 2016. Trabalho de conclusão de curso – Universidade Estadual de Londrina, Londrina. Disponível em: <<https://docplayer.com.br/46985110-Vinicius-quintal-dalter-algoritmos-para-a-estimacao-da-frequencia-fundamental.html>>. Acesso em: 17 nov. 2021.

DANNENBERG, B. R. An on-line algorithm for real-time accompaniment. In: 1984 International Computer Music Conference. [S.l.: s.n.], 1984. p. 193–198. Disponível em: <<https://www.cs.cmu.edu/~rbd/papers/icmc84accomp.pdf>>. Acesso em: 15 fev. 2021.

DANNENBERG, R. **Music Information Retrieval as Music Understanding**. [S.l.: s.n.], jan. 2001.

DANNENBERG, R. B. An Intelligent Multi-Track audio Editor. In: PROCEEDINGS of the 2007 International Computer Music Conference, ICMC 2007, Copenhagen, Denmark, August 27-31, 2007. [S.l.]: Michigan Publishing, 2007. Disponível em: <<http://hdl.handle.net/2027/spo.bbp2372.2007.129>>.

DAVIES, M. et al. Moving in time: Computational analysis of microtiming in Maracatu de baque solto. In: PROCEEDINGS of the 21st International Society for Music Information Retrieval Conference. Montreal, Canada: [s.n.], out. 2020. p. 795–802. Conference Name: International Society for Music Information Retrieval Conference (ISMIR 2020) Pages: 795-802 Publication Title: Proceedings of the 21st International Society for Music Information Retrieval Conference Publisher: ISMIR. DOI: 10 . 5281/zenodo . 4245554. Disponível em: <<https://zenodo.org/record/4245554#.YG9YURRkG-Q>>. Acesso em: 8 abr. 2021.

DE CHEVEIGNÉ, A.; KAWAHARA, H. YIN, a Fundamental Frequency Estimator for Speech and Music. en. **The Journal of the Acoustical Society of America**, v. 111, n. 4, p. 1917–1930, abr. 2002. ISSN 0001-4966. DOI: 10 . 1121/1 . 1458024.

DESAIN, P.; HONING, H. The Quantization of Musical Time: A Connectionist Approach. v. 13, p. 56–66, set. 1989. DOI: 10 . 2307/3680012.

DOLSON, M. The Phase Vocoder: A Tutorial. en. **Computer Music Journal**, v. 10, n. 4, p. 14, 1986. ISSN 01489267. DOI: 10 . 2307/3680093.

DOUGLAS, R. L. Formalizing an African-American aesthetic. English. **New art examiner.**, p. 18–24, 1991. OCLC: 39965335.

DOWNIE, J. S. Music Information Retrieval. en. **Annual Review of Information Science and Technology**, v. 37, n. 1, p. 295–340, 2003. ISSN 1550-8382. DOI: 10 . 1002/aris . 1440370108.

EINARSSON, A. **Singing the body electric - Understanding the role of embodiment in performing and composing interactive music**. Out. 2017. Tese (Doutorado) – Lund University Royal College of Music Stockholm.

EINARSSON, A. We Can Work It Out - Calibration as Artistic Method. **RUUKKU - Studies in Artistic Research**, n. 4, jul. 2015. Disponível em: <<https://www.researchcatalogue.net/view/142373/142374>>. Acesso em: 25 mai. 2021.

EINARSSON, A.; FRIBERG, A. Using Singing Voice Vibrato as a Control Parameter in a Chamber Opera. In: LOOKING Back, Looking Forward: Proceedings of the 41st International Computer Music Conference, ICMC 2015, Denton, TX, USA, September 25 - October 1, 2015. [S.l.]: Michigan Publishing, 2015. Disponível em: <<http://hdl.handle.net/2027/spo.bbp2372.2015.003>>.

EINARSSON, A.; FRIBERG, A. Using Singing Voice Vibrato as a Control Parameter in a Chamber Opera. In: 41ST International Computer Music Conference. University of North Texas: [s.n.], set. 2015. p. 18–21.

EMMERSON, P. S. **Living Electronic Music**. [S.l.]: Ashgate Publishing, Ltd., jan. 2013. Google-Books-ID: zNShAgAAQBAJ. ISBN 978-1-4094-9371-6.

EMMERSON, S. Ring Modulation and Structure. **Contact: A Journal for Contemporary Music**, v. 17, p. 14–20, 1977. ISSN 0308-5066. Disponível em: <<https://journals.gold.ac.uk/index.php/contactjournal/issue/view/contactjournal20/contactjournal20>>. Acesso em: 18 mai. 2021.

ESLING, P.; CHEMLA-ROMEU-SANTOS, A.; BITTON, A. Bridging Audio Analysis, Perception and Synthesis with Perceptually-regularized Variational Timbre Spaces. In: G: International Society for Music Information Retrieval Conference (ISMIR 2018). [S.l.: s.n.], 2018.

ESLING, P.; AGON, C. Composition Of Sound Mixtures With Spectral Maquettes, jun. 2010.

FARIA, R. R. A. **Auralização em ambientes audiovisuais imersivos**. Jun. 2005. text – Universidade de São Paulo. DOI: 10 . 11606 / T . 3 . 2005 . tde - 29092013 - 105249. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3142/tde-29092013-105249/>>. Acesso em: 4 dez. 2021.

FAVREAU, E. et al. Software Developments for the 4X Real-time System. In: PROCEEDINGS of the 1986 International Computer Music Conference, ICMC 1986, Den Haag, The Netherlands, October 20-24, 1986. [S.l.]: Michigan Publishing, 1986. Disponível em: <<http://hdl.handle.net/2027/spo.bbp2372.1986.074>>.

FELS, S.; HINTON, G. E. Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks. In: TESAURO, G.; TOURETZKY, D.; LEEN, T. (Ed.). **Advances in Neural Information Processing Systems**. [S.l.]: MIT Press, 1995. v. 7. Disponível em: <<https://proceedings.neurips.cc/paper/1994/file/c2aee86157b4a40b78132f1e71a9e6f1-Paper.pdf>>.

FERNANDEZ, J. M.; GIAVITTO, J.-L.; DONAT-BOUILLUD, P. AntesCollider: Control and Signal Processing in the Same Score. en. Disponível em: <<https://hal.inria.fr/hal-02159629>>. Acesso em: 27 abr. 2021.

FERRAZ, S. **Composição com suporte tecnológico**. Salvador, 1999. Disponível em: <<http://sferraz.mus.br/anp99.htm>>. Acesso em: 22 mai. 2021.

FERRAZ, S. **Livro da sonoridades [notas dispersas sobre composição]**. 2. ed. [S.l.]: 7 letras, 2013. ISBN 978-85-421-0140-9. Disponível em: <[https://www.academia.edu/45536335/LIVRO\\_DAS\\_SONORIDADES\\_notas\\_dispersas\\_sobre\\_composi%C3%A7%C3%A3o](https://www.academia.edu/45536335/LIVRO_DAS_SONORIDADES_notas_dispersas_sobre_composi%C3%A7%C3%A3o)>. Acesso em: 29 nov. 2021.

FLANAGAN, J. L.; GOLDEN, R. M. Phase Vocoder. en. **Bell System Technical Journal**, v. 45, n. 9, p. 1493–1509, 1966. ISSN 1538-7305. DOI: 10 . 1002 / j . 1538 - 7305 . 1966 . tb01706 . x.

FLETCHER, H. Auditory Patterns. **Reviews of Modern Physics**, American Physical Society, v. 12, n. 1, p. 47–65, jan. 1940. DOI: 10 . 1103 / RevModPhys . 12 . 47.

FLETCHER, H.; MUNSON, W. A. Loudness, Its Definition, Measurement and Calculation\*. en. **Bell System Technical Journal**, v. 12, n. 4, p. 377–430, 1933. ISSN 1538-7305. DOI: 10 . 1002 / j . 1538 - 7305 . 1933 . tb00403 . x.

FREIRE, S.; PADOVANI, J.; CAMPOS, C. **Estimation of Perceptual Qualities of Percussive Sounds Inspired by Schaefferian Criteria: Attack Profile, Mass, and Harmonic Timbre**. [S.l.: s.n.], nov. 2021.

FREIRE, S.; PADOVANI, J.; CAMPOS, C. **Real-time Qualification of Percussive Sounds Based on Correspondences Between Schaeffer's Solfège and Low-level Audio Descriptors**. [S.l.: s.n.], out. 2021.

FRIBERG, A.; SCHOONDERWALDT, E.; JUSLIN, P. CUEX: An Algorithm for Automatic Extraction of Expressive Tone Parameters in Music Performance from Acoustic Signals. **Acta Acustica united with Acustica**, v. 93, p. 411–420, mai. 2007.

FUJISHIMA, T. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music. In: ICMC. [S.l.: s.n.], 1999.

FUNARTE, M. **Brazilian International Songbook Online - Concert Music**. [S.l.]: Marcos Souza. v. 8. Disponível em: <<https://www.funarte.gov.br/partituras-brasileiras-online/>>. Acesso em: 3 jul. 2021.

FUTRELLE, J.; DOWNIE, J. Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000?2002. **Journal of New Music Research**, v. 32, p. 121–131, jun. 2003. DOI: 10.1076/jnmr.32.2.121.16740.

GABOR, D. Theory of Communication. Part 1: The Analysis of Information. en. **Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering**, v. 93, n. 26, p. 429–441, nov. 1946. ISSN 2054-0604. DOI: 10.1049/ji-3-2.1946.0074.

GARNETT, G. E. The Aesthetics of Interactive Computer Music. **Computer Music Journal**, The MIT Press, v. 25, n. 1, p. 21–33, 2001. ISSN 0148-9267.

GERZON, M. J. Periphony: With-Height Sound Reproduction. en. **undefined**, 1972. Disponível em: <<https://www.semanticscholar.org/paper/Periphony%3A-With-Height-Sound-Reproduction-Gerzon/d5e9c8ec308b025ea7616f9a9e83f763119c5fdf>>. Acesso em: 4 dez. 2021.

GÓMEZ, E. **Tonal Description of Music Audio Signals**. 2006. Tese (Doutorado) – Universitat Pompeu Fabra. DOI: <http://hdl.handle.net/10803/7537>.

GRUBB, L.; DANNENBERG, R. A Stochastic Method of Tracking a Vocal Performer. In: PROCEEDINGS 1997 International Computer Music Conference. [S.l.: s.n.], set. 1997. p. 301–308.

GRÜNWARD, J. **Theory, implementation and evaluation od digital phase vocoder in the context of audio effects**. 2010. Tese (Doutorado) – Graz University of Technology, Graz.

GUATIMOSIM, J.; PADOVANI, J.; GUATIMOSIM, C. Concatenative Sound Synthesis as a Technomorphic Model in Computer-Aided Composition. In: SIMPÓSIO Brasileiro de Computação Musical 2021. [S.l.: s.n.], out. 2021.

HELLER, E. J. **Why You Hear What You Hear: An Experiential Approach to Sound, Music, and Psychoacoustics**. [S.l.: s.n.], 2013. ISBN 978-0-691-14859-5.

HELMHOLTZ, H. **On the Sensations of Tone**. 2nd ed. edição. New York: Dover Publications, jun. 1954. ISBN 978-0-486-60753-5.

HENRIQUE, L. L. **Acústica musical**. [S.l.]: Fundação Calouste Gulbenkian, 2002. ISBN 978-972-31-0987-0.

HOLMES, T. **Electronic and experimental music: technology, music, and culture**. 3rd ed. New York: Routledge, 2008. ISBN 978-0-415-95781-6 978-0-415-95782-3 978-0-203-92959-9.

HOLZAPFEL, A.; STURM, B. L.; COECKELBERGH, M. Ethical Dimensions of Music Information Retrieval Technology. en. **Transactions of the International Society for Music Information Retrieval**, v. 1, n. 1, p. 44–55, set. 2018. Number: 1 Publisher: Ubiquity Press. ISSN 2514-3298. DOI: 10 . 5334 / tismir . 13. Disponível em: <<http://transactions.ismir.net/articles/10.5334/tismir.13/>>. Acesso em: 11 abr. 2021.

HORII, Y. Acoustic analysis of vocal vibrato: A theoretical interpretation of data. en. **Journal of Voice**, v. 3, n. 1, p. 36–43, mar. 1989. ISSN 0892-1997. DOI: 10 . 1016/S0892-1997(89)80120-1. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0892199789801201>>. Acesso em: 27 mai. 2021.

HOWARD, D. M.; ANGUS, J. **Acoustics and Psychoacoustics**. 5th Edition. New York ; London: Routledge, 14 jul. 2017. 518 p. ISBN 978-1-138-85987-6.

HUGUES, T. P. Technological momentum. In: DOES technology drive history? : the dilemma of technological determinism. Cambridge Mass.: MIT Press, 1994. p. 101–113. ISBN 978-0-262-19347-4.

HUNT, A.; DD, Y.; WANDERLEY, M.; PARADIS, M. The Importance of Parameter Mapping in Electronic Instrument Design. **Journal of New Music Research**, v. 32, mai. 2002. DOI: 10 . 1076/jnmr . 32 . 4 . 429 . 18853.

HUNT, A.; KIRK, R. Mapping Strategies for Musical Performance. **Trends in Gestural Control of Music**, jan. 2000.

HUYGENS, C. **Oeuvres complètes de Christiaan Huygens. Correspondance [de Christiaan Huygens], 1664-1665 / publ. par la Société hollandaise des sciences**. [S.l.: s.n.], 1888. Disponível em: <<https://gallica.bnf.fr/ark:/12148/bpt6k77853w>>. Acesso em: 2 mai. 2021.

IAZZETTA, F. Meaning in Musical Gesture. In: [s.l.: s.n.], jan. 2000. p. 259–268.

III SMITH, J. O. **Introduction to Digital Filters: With Audio Applications**. W3K Publishing: Createspace Independent Pub, ago. 2006. ISBN 978-0-9745607-1-7.

JEAN-LOUIS GIAVITTO; CONT, A.; ECHEVESTE, J. **Antescofo Reference**. [S.l.: s.n.]. <https://antescofo-doc.ircam.fr/Reference/1-intro/>.

JOHN ELLIS, A. On the Musical Scales of Various Nations | Sound & Science: Digital Histories. English. **Journal of the Society of Arts**, v. 33, n. 1688, p. 485–527, 1885.

JOHNSON, D. G. Computer ethics. In: THE Blackwell Guide to the Philosophy of Computing and Information. [S.l.]: Luciano Floridi, abr. 2008. Google-Books-ID: a37OrM9IUagC. ISBN 978-0-470-75676-8.

JOHNSON, M. **The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason**. 1ª edição. [S.l.]: University of Chicago Press, jan. 2013.

JU, Y. et al. Automatic figured bass annotation using the new bach chorales figured bass dataset. In: PROCEEDINGS of the 21st International Society for Music Information Retrieval Conference, Montreal, Canada: [s.n.], out. 2020. p. 640–646. Conference Name: International

Society for Music Information Retrieval Conference (ISMIR 2020) Pages: 640-646 Publication Title: Proceedings of the 21st International Society for Music Information Retrieval Conference Publisher: ISMIR. DOI: 10 . 5281 / zenodo . 4245512. Disponível em: <<https://zenodo.org/record/4245512#.YG9YDxRKg-Q>>. Acesso em: 8 abr. 2021.

KASSLER, M. Toward Musical Information Retrieval. **Perspectives of New Music**, Perspectives of New Music, v. 4, n. 2, p. 59-67, 1966. ISSN 0031-6016. DOI: 10.2307/832213.

KEHTARNAVAZ, N. **Digital Signal Processing System Design**. [S.l.]: Elsevier, 2008. ISBN 978-0-12-374490-6. DOI: 10 . 1016/B978-0-12-374490-6.X0001-3.

KOPPARAPU, S. K.; LAXMINARAYANA, M. Choice of Mel Filter Bank in Computing MFCC of a Resampled Speech. en. In: 10TH International Conference on Information Science, Signal Processing and Their Applications (ISSPA 2010). Kuala Lumpur, Malaysia: IEEE, mai. 2010. p. 121-124. ISBN 978-1-4244-7165-2. DOI: 10 . 1109/ISSPA . 2010 . 5605491.

KOSINSKI, M. et al. Manifestations of user personality in website choice and behaviour on online social networks. en. **Machine Learning**, v. 95, n. 3, p. 357-380, jun. 2014. ISSN 1573-0565. DOI: 10 . 1007/s10994-013-5415-y. Disponível em: <<https://doi.org/10.1007/s10994-013-5415-y>>. Acesso em: 12 abr. 2021.

KOZU, F. A complexidade, a figura eo ritmo no pensamento composicional de Brian Ferneyhough. en. **uel.br**. Disponível em: <[https://www.academia.edu/984403/A\\_complexidade\\_a\\_figura\\_eo\\_ritmo\\_no\\_pensamento\\_composicional\\_de\\_Brian\\_Ferneyhough](https://www.academia.edu/984403/A_complexidade_a_figura_eo_ritmo_no_pensamento_composicional_de_Brian_Ferneyhough)>. Acesso em: 26 nov. 2021.

KRAUSE, M.; ZALKOW, F.; WEISS, C.; MÜLLER, M. Classifying leitmotifs in recordings of operas by Richard Wagner. In: PROCEEDINGS of the 21st International Society for Music Information Retrieval Conference. Montreal, Canada: [s.n.], out. 2020. Conference Name: International Society for Music Information Retrieval Conference (ISMIR 2020) Pages: 473-480 Publication Title: Proceedings of the 21st International Society for Music Information Retrieval Conference Publisher: ISMIR. DOI: 10 . 5281 / zenodo . 4245472. Disponível em: <<https://zenodo.org/record/4245472#.YG9X8xRKg-Q>>. Acesso em: 8 abr. 2021.

KRIMPHOFF, J.; MCADAMS, S.; WINSBERG, S. Caractérisation du timbre des sons complexes.II. Analyses acoustiques et quantification psychophysique. **Journal de Physique IV Proceedings**, v. 04, n. C5, p. c5-625-c5-628, 1994. Publisher: EDP Sciences. DOI: 10 . 1051 / jp4 : 19945134. Disponível em: <<https://hal.archives-ouvertes.fr/jpa-00252811>>. Acesso em: 3 mar. 2022.

LARGE, E. W. Periodicity, Pattern Formation, and Metric Structure. **Journal of New Music Research**, v. 30, n. 2, p. 173-185, jun. 2001. ISSN 0929-8215. DOI: 10 . 1076 / jnmr . 30 . 2 . 173 . 7113. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1076/jnmr.30.2.173.7113>>. Acesso em: 1 mar. 2021.

LARTILLOT, O.; TOIVAINEN, P. **MIR in Matlab (II): A Toolbox for Musical Feature Extraction from Audio**. [S.l.: s.n.], jan. 2007. Journal Abbreviation: Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France Pages: 130 Publication

Title: Proceedings of the 10th International Conference on Digital Audio Effects, Bordeaux, France.

LEE, M.; WESSEL, D. Connectionist models for real-time control of synthesis and compositional algorithms. In: PROCEEDINGS of the 1992 International Computer Music Conference. San Jose, CA: [s.n.], 1992. p. 277–280.

LEMOUTON, S.; MANOURY, P. Suivi de partition, reconnaissance d'événements musicaux. **Journées d'Informatique musicale**, v. 98, p. 1–9, mai. 1998.

LEWIS, G. E. Improvising Tomorrow's Bodies: The Politics of Transduction. **Improvising Tomorrow's Bodies: The Politics of Transduction**, v. 4, n. 2, 2007.

LEWIS, G. E. Interacting with Latter-Day Musical Automata. **Contemporary Music Review**, Routledge, v. 18, n. 3, p. 99–112, jan. 1999. ISSN 0749-4467. DOI: 10.1080/07494469900640381.

LEWIS, G. E. Too Many Notes: Computers, Complexity and Culture in Voyager. **Leonardo Music Journal**, v. 10, p. 33–39, dez. 2000. ISSN 0961-1215. DOI: 10.1162/096112100570585.

LIANG, R. et al. Real-Time Speech Enhancement Algorithm Based on Attention LSTM. **IEEE Access**, v. 8, p. 48464–48476, 2020. Conference Name: IEEE Access. ISSN 2169-3536. DOI: 10.1109/ACCESS.2020.2979554.

LINDEMANN, E.; DECHELLE, F.; SMITH, B.; STARKIER, M. The architecture of the IRCAM musical workstation. **Computer Music Journal**, v. 15, n. 3, p. 41–49, jan. 1991.

LINDENMAYER, A. Mathematical models for cellular interaction in development: Parts I and II. **Journal of Theoretical Biology**, v. 18, 1968.

LOW, T. et al. The ISMIR Explorer - A Visual Interface for Exploring 20 Years of ISMIR Publications. In: Delft, The Netherlands. PROCEEDINGS OF THE 20TH INTERNATIONAL SOCIETY FOR MUSIC INFORMATION RETRIEVAL CONFERENCE. Delft, The Netherlands: ISMIR, nov. 2019. p. 754–760. DOI: 10.5281/zenodo.3527920. Disponível em: <<https://doi.org/10.5281/zenodo.3527920>>.

LOY, D. G. **Musimathics: The Mathematical Foundations of Music: Volume 1**. Cambridge, Mass. ; London: MIT Press, 2006. ISBN 978-0-262-12285-6.

LOY, G. **Musimathics: The Mathematical Foundations of Music: Volume 2**. Edição: Reprint. Cambridge, Mass.: Mit Press, 2007. ISBN 978-0-262-51655-6.

MAGALHAES, T.; BARROS, F.; LOUREIRO, M. Iracema: a Python library for audio content analysis. **Revista de Informática Teórica e Aplicada**, v. 27, p. 127–138, dez. 2020. DOI: 10.22456/2175-2745.107202.

MALT, M. Khorwa: A musical experience with « autonomous agents ». In: ICMC. [S.l.: s.n.], 2004.

MALT, M.; GENTILUCCI, M. **Real Time Vowel Tremolo Detection Using Low Level Audio Descriptors**. [S.l.: s.n.], 2015. arXiv: 1511.07008 [cs.SD].

MALT, M.; JOURDAN, E. Real-Time Uses of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa.Descriptors Library. en, p. 12, 2009.

MALT, M.; JOURDAN, E.; MALT, M.; JOURDAN, E. Zsa.Descriptors: A Library for Real-Time Descriptors Analysis. en, p. 4, 2008.

MANNING, P. **Electronic and Computer Music**. [S.l.]: Oxford University Press, 2004. Google-Books-ID: P2dCIS4LdPQC. ISBN 978-0-19-514484-0.

MANOURY, P. **BRAHMS - IRCAM: Tensio - Note de Programme**. [S.l.: s.n.], 2010. <http://brahms.ircam.fr/works/work/28236/>. Acesso em:

MANOURY, P. Compositional Procedures in Tensio. **Contemporary Music Review**, v. 32, n. 1, p. 61–97, fev. 2013. ISSN 0749-4467. DOI: 10.1080/07494467.2013.774514.

MANOURY, P. **La note et le son**. en-US. [S.l.: s.n.], 1990. Disponível em: <<http://www.philippemanoury.com/?p=330>>. Acesso em: 22 fev. 2021.

MANOURY, P. **Les Partitions Virtuelles**. en-US. [S.l.: s.n.].

MANOURY, P. **Philippe Manoury» Blog Archive » Considérations (toujours actuelles) sur l'état de la musique en temps réel**. en-US. [S.l.: s.n.], 2008. Disponível em: <<http://www.philippemanoury.com/?p=319>>. Acesso em: 8 fev. 2021.

MANOURY, P. **Philippe Manoury» Blog Archive » Les Grammaires Musicales Génératives**. en-US. [S.l.: s.n.], 2012. Disponível em: <<http://www.philippemanoury.com/?p=5025>>. Acesso em: 8 jul. 2019.

MANOURY, P. **Sur la synthèse 3F**. en-US. [S.l.: s.n.], 2020. Disponível em: <<http://www.philippemanoury.com/?p=7084>>. Acesso em: 11 jun. 2021.

MARCURIO, G. Do corpo ao pó. **Cadernos de Campo (São Paulo 1991)**, v. 27, p. 415–419, dez. 2018. DOI: 10.11606/issn.2316-9133.v27i1p415-419.

MAY, A. Philippe's Manoury Jupiter. In: **ANALYTICAL Methods of Electroacoustic Music**. [S.l.]: Routledge, dez. 2005. p. 145–185. ISBN 978-1-135-50335-2.

MCFEE, B. et al. **librosa: Audio and Music Signal Analysis in Python**. en, p. 7, 2015.

MCLEOD, P. **Fast, Accurate Pitch Detection Tools for Music Analysis**. [S.l.]: University of Otago, 2009.

MCNAB, R. J. et al. Towards the digital music library: tune retrieval from acoustic input. In: **PROCEEDINGS of the first ACM international conference on Digital libraries**. New York, NY, USA: Association for Computing Machinery, abr. 1996. (DL '96), p. 11–18. ISBN 978-0-89791-830-5. DOI: 10.1145/226931.226934. Disponível em: <<https://doi.org/10.1145/226931.226934>>.

MIRANDA, E.; WANDERLEY, M. **New digital musical instruments: Control and interaction beyond the keyboard**. [S.l.]: A-R Editions, 2006. (Computer music and digital audio series). tex.lccn: 2005035245. ISBN 978-0-89579-585-4. Disponível em: <<https://books.google.com.br/books?id=CGEwXZ7hcIoC>>.

MODLER, P. Neural Networks for Mapping Hand Gestures to Sound Synthesis Parameters, mai. 2021.

MONTEIRO, A. C. Criação e performance musical no contexto dos instrumentos musicais digitais. Português. [s.n.], 2012.

MORAIS, B. M. "Do que é feito um corpo?": uma crítica substantiva kaiowá e guarani ao agronegócio. en. **Revista Latinoamericana de Estudios en Cultura y Sociedad**. Disponível em: <[https://www.academia.edu/40504823/\\_Do\\_que\\_%C3%A9\\_feito\\_um\\_corpo\\_uma\\_cr%C3%ADtica\\_substantiva\\_kaiow%C3%A1\\_e\\_guarani\\_ao\\_agroneg%C3%B3cio](https://www.academia.edu/40504823/_Do_que_%C3%A9_feito_um_corpo_uma_cr%C3%ADtica_substantiva_kaiow%C3%A1_e_guarani_ao_agroneg%C3%B3cio)>. Acesso em: 29 set. 2021.

MORAIS, B. M. **Do corpo ao pó: crônicas da territorialidade Kaiowá e Guarani nas adjacências da morte**. Dez. 2015. text – Universidade de São Paulo. DOI: 10.11606/D.8.2016.tde-23032016-134741. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8134/tde-23032016-134741/>>. Acesso em: 27 set. 2021.

MORCIANO, L. **Ecriture du son, du temps et de l'espace dans l'interaction entre instruments et dispositifs numériques synchrones**. Dez. 2018. Tese de Doutorado – Paris Sciences et Lettres (ComUE). Disponível em: <<http://www.theses.fr/2018PSLET038>>. Acesso em: 1 mar. 2021.

MORCIANO, L. **Raggi Di Stringhe**. [S.l.: s.n.]. <http://brahms.ircam.fr/works/work/29525/>.

MORREALE, F.; DE ANGELI, A. Collaborating with an Autonomous Agent to Generate Affective Music. **Computers in Entertainment**, v. 14, p. 1–21, jan. 2017. DOI: 10.1145/2967508.

NOBLE, S. U. **Algorithms of Oppression: How Search Engines Reinforce Racism**. [S.l.]: NYU Press, 2018. ISBN 1479849944, 9781479849949. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=a4ecf6f20c1ea83f63b8ae5ea73de5b3>>.

NOLL, A. M. Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and maximum likelihood estimate. en. **Proceedings of the Symposium on Computer Processing in Communications**, v. 19, p. 779–797, 1969. Acesso em: 16 nov. 2021.

NUTTALL, T. et al. Contributing to New Musicological Theories with Computational Methods: The Case of Centonization in Arab-Andalusian Music. In: NG: International Society for Music Information Retrieval Conference (ISMIR 2019). Delft, The Netherlands: [s.n.], nov. 2019. p. 4–8. Conference Name: International Society for Music Information Retrieval Conference (ISMIR 2019) Pages: 223-228 Publication Title: Proceedings of the 20th International Society for Music Information Retrieval Conference Publisher: ISMIR. DOI: 10.5281/zenodo.3527784. Disponível em: <<https://zenodo.org/record/3527784#.YG91shRKg-Q>>. Acesso em: 8 abr. 2021.

OFNER, A.; STOBBER, S. Modeling perception with hierarchical prediction: Auditory segmentation with deep predictive coding locates candidate evoked potentials in EEG. In: INTERNATIONAL Society for Music Information Retrieval Conference (ISMIR 2020).

Montreal, Canada: [s.n.], out. 2020. Conference Name: International Society for Music Information Retrieval Conference (ISMIR 2020) Pages: 566-573 Publication Title: Proceedings of the 21st International Society for Music Information Retrieval Conference Publisher: ISMIR. DOI: 10 . 5281 / zenodo . 4245496. Disponível em: <<https://zenodo.org/record/4245496#.YG9ywhRKg-Q>>. Acesso em: 8 abr. 2021.

PADOVANI, J. **Representação, intuição e contato na composição com algoritmos**. Fev. 2009. Tese (Doutorado).

PADOVANI, J. **Spectral envelope extraction by means of cepstrum analysis and filtering in Pure Data**. [S.l.: s.n.], jan. 2009.

PADOVANI, J.; FERRAZ, S. Proto-história, Evolução e Situação Atual das Técnicas Estendidas na Criação Musical e na Performance. pt. **Revista Música Hodie**, v. 11, n. 2, 2011. Number: 2. ISSN 1676-3939. DOI: 10 . 5216 / mh . v11i2 . 21752. Disponível em: <<https://www.revistas.ufg.br/musica/article/view/21752>>. Acesso em: 10 mar. 2022.

PAMPALK, E. **Computational Models of Music Similarity and their Application in Music Information Retrieval**. Jan. 2006. Tese (Doutorado).

PASSOS, M. G. et al. Um Ambiente Para Processamento Digital de Sinais Aplicado à Comunicação Vocal Homem-Máquina. en. **Revista Principia**, Editora IFPB, v. 1, n. 14, p. 25-31, set. 2006. ISSN 1517-0306, 2447-9187. DOI: 10 . 18265 / 1517 - 0306 2015 v1 n14 p25 - 31.

PEETERS, G. A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project, p. 25, 2004.

PEETERS, G. et al. The Timbre Toolbox: Extracting Audio Descriptors from Musical Signals. en. **The Journal of the Acoustical Society of America**, v. 130, n. 5, p. 2902-2916, nov. 2011. ISSN 0001-4966. DOI: 10 . 1121 / 1 . 3642604.

PEREIRA, E. M. **Estudos sobre uma ferramenta de classificação musical**. 2009. Dissertação de Mestrado – Universidade Estadual de Campinas, Campinas.

PESTOVA, X. **Models of interaction in works for piano and live electronics**. 2008. Tese (Doutorado) – McGill University. Publisher: McGill University. Acesso em: 17 mai. 2021.

PLOMP, R.; LEVELT, W. J. M. Tonal Consonance and Critical Bandwidth. **The Journal of the Acoustical Society of America**, v. 38, n. 4, p. 548-560, out. 1965. Publisher: Acoustical Society of America. ISSN 0001-4966. DOI: 10 . 1121 / 1 . 1909741. Acesso em: 29 jun. 2021.

POLLARD, H. F.; JANSSON, E. V. A Tristimulus Method for the Specification of Musical Timbre. **Acta Acustica united with Acustica**, v. 51, n. 3, p. 162-171, ago. 1982.

PORRES, A.; PIRES, A. Um External de Aspereza para Puredata & MAX/MSP. In: SIMPÓSIO Brasileiro de Computação Musical 2009. Recife: [s.n.], 2009. p. 191-194.

PORRES, A. T. **Modelos psicoacústicos de dissonância para eletrônica ao vivo**. Abr. 2012. text – Universidade de São Paulo. DOI: 10 . 11606 / T . 27 . 2012 . tde - 22022013 - 141622. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27157/tde-22022013-141622/>>. Acesso em: 16 nov. 2021.

PORRES, A. T. PSYCHO library for Pure Data. **Anais do Simpósio Brasileiro de Computação Musical (SBCM 2019)**, 2019. DOI: 10.5753/sbcm.2019.10432.

POUSSEUR, H. **Apoteose de Rameau e outros ensaios**. 1. ed. [S.l.: s.n.], 2009. ISBN 978-85-7139-919-8. Disponível em: <<http://editoraunesp.com.br/catalogo/9788571399198>>, apoteose-de-rameau>. Acesso em: 29 jun. 2021.

POUSSEUR, H. Applications Analytiques de la 'technique des réseaux'. **Revue belge de Musicologie / Belgisch Tijdschrift voor Muziekwetenschap**, v. 52, p. 247–298, 1998. Publisher: Societe Belge de Musicologie. ISSN 0771-6788. DOI: 10.2307/3686929. Disponível em: <<https://www.jstor.org/stable/3686929>>. Acesso em: 29 jun. 2021.

PRAME, E. Vibrato extent and intonation in professional Western lyric singing. en. **The Journal of the Acoustical Society of America**, v. 102, n. 1, p. 616, ago. 1998. Publisher: Acoustical Society of AmericaASA. ISSN 0001-4966. DOI: 10.1121/1.419735. Disponível em: <<https://asa.scitation.org/doi/abs/10.1121/1.419735>>. Acesso em: 27 mai. 2021.

PROCKUP, M.; GRUNBERG, D.; HRYBYK, A.; KIM, Y. E. Orchestral Performance Companion: Using Real-Time Audio to Score Alignment. **IEEE MultiMedia**, v. 20, n. 2, p. 52–60, abr. 2013. Conference Name: IEEE MultiMedia. ISSN 1941-0166. DOI: 10.1109/MMUL.2013.26.

PUCKETTE, M. Max at Seventeen. **Computer Music Journal**, MIT Press, v. 26, n. 4, p. 31–43, dez. 2002. ISSN 0148-9267. DOI: 10.1162/014892602320991356.

PUCKETTE, M. **The Theory and Techniques of Electronic Music**. Hackensack, NJ: World Scientific Publishing Company, mai. 2007. ISBN 978-981-270-077-3.

PUCKETTE, M.; APEL, T.; ZICARELLI, D. Real-time audio analysis tools for Pd and MSP. In: PROCEEDINGS of International Computer Music Conference. São Francisco: [s.n.], out. 1998.

RABINER, L. R.; SCHAFER, R. W. **Digital Processing of Speech Signals**. Edição: 1. Englewood Cliffs, N.J: Prentice Hall, set. 1978. ISBN 978-0-13-213603-7.

RAPHAEL, C. Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 21, n. 4, p. 360–370, abr. 1999. ISSN 0162-8828. DOI: 10.1109/34.761266. Disponível em: <<https://doi.org/10.1109/34.761266>>. Acesso em: 17 fev. 2021.

RAPHAEL, C. Music Plus One and Machine Learning. In: PROCEEDINGS of the 27th International Conference on Machine Learning (ICML-10). Haifa, Israel: [s.n.], dez. 2009.

REGIUS, H.; LANGER, J. **Kung Kalle av Kurrekurreduttön: En resa i Efraim Långstrumps fotspår**. [S.l.: s.n.], 2002. Disponível em: <<https://biblioteket.stockholm.se/titel/833558>>. Acesso em: 25 mai. 2021.

REYNOLDS, R. Compositional Strategies in The Angel of Death for Piano, Chamber Orchestra, and Computer-Processed Sound. **Music Perception: An Interdisciplinary Journal**, v. 22, n. 2, p. 173–205, 2004. Publisher: University of California Press. ISSN 0730-7829. DOI: 10.1525/mp.2004.22.2.173. Disponível em: <<https://www.jstor.org/stable/10.1525/mp.2004.22.2.173>>. Acesso em: 7 abr. 2021.

RIMOLDI, G. Metaflauta : uma abordagem ecológica aplicada ao design e processos criativos a partir de um instrumento expandido. Português, 2018. Accepted: 2018-09-04T18:43:32Z  
 Publisher: [s.n.] Disponível em:  
 <<http://repositorio.unicamp.br/jspui/handle/REPOSIP/332105>>. Acesso em: 7 abr. 2021.

ROADS, C. **The Computer Music Tutorial**. Cambridge, Mass: Mit Press, fev. 1996. ISBN 978-0-262-68082-0.

ROSSETTI, D.; TEIXEIRA, W.; MANZOLLI, J. Emergent Timbre and Extended Techniques in Live-Electronic Music: An Analysis of Desdobramentos do Contínuo Performed by Audio Descriptors. en. **Revista Música Hodie**, v. 18, n. 1, p. 16, jun. 2018. ISSN 1676-3939, 2317-6776. DOI: 10 . 5216 / mh . v18i1 . 53568. Disponível em:  
 <<https://www.revistas.ufg.br/musica/article/view/53568>>. Acesso em: 2 jun. 2019.

ROSSING, T. D. (Ed.). **Springer Handbook of Acoustics**. New York, N.Y: Springer, 2007. ISBN 978-0-387-30446-5.

ROVAN, J.; WANDERLEY, M.; DUBNOV, S. Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance, dez. 1997.

ROWE, R. **Interactive Music Systems: Machine Listening and Composing**. Cambridge: MIT press, 1993.

ROWE, R. **Machine Musicianship**. Revised ed. Edição. Cambridge, Mass.: MIT Press, mar. 2004. ISBN 978-0-262-68149-0.

SCHAEFFER, P. **Traité des objets musicaux**. Nouv. éd edição. Paris: Le Seuil, 1966.

SCHEDL, M.; GÓMEZ, E.; URBANO, J. **Music Information Retrieval: Recent Developments and Applications**. Boston, Mass.: Now Publ, 2014. (Foundations and Trends in Information Retrieval, 8,2/3). ISBN 978-1-60198-806-5.

SCHEIRER, E. D.; SLANEY, M. Construction and evaluation of a robust multifeature speech/music discriminator. **1997 IEEE International Conference on Acoustics, Speech, and Signal Processing**, 1997. DOI: 10 . 1109/ICASSP . 1997 . 596192.

SCHLEUSING, O.; ZHANG, B.; WANG, Y. Onset detection in pitched non-percussive music using warping-compensated correlation. In: SCOPUS. [S.l.: s.n.], 2008. Accepted: 2013-07-04T08:04:37Z ISSN: 1520-6149. ISBN 978-1-4244-1484-0. Disponível em:  
 <<https://scholarbank.nus.edu.sg/handle/10635/40452>>.

SERRA, X. Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project. In: SEMANTIC Audio. [S.l.: s.n.], 2014.

SHEPARD, R. N. Geometrical approximations to the structure of musical pitch. eng. **Psychological Review**, v. 89, n. 4, p. 305–333, jul. 1982. ISSN 0033-295X.

SIMURRA, I. E. Y. **Contribuição Ao Problema Da Orquestração Assistida Por Computador Com Suporte de Descritores de Áudio**. 2016. Tese (Doutorado) – [s.n.]

STEVENS, S. S.; VOLKMANN, J. The Relation of Pitch to Frequency: A Revised Scale. **The American Journal of Psychology**, University of Illinois Press, v. 53, n. 3, p. 329–353, 1940. ISSN 0002-9556. DOI: 10.2307/1417526.

STOWELL, D. **Making music through real-time voice timbre analysis: machine learning and timbral control**. [S.l.: s.n.], 2010. Acesso em: 2 jun. 2019.

STOWELL, D.; MCLEAN, A. **(2) Live Music-Making: a Rich Open Task Requires a Rich Open Interface | Request PDF**. [S.l.: s.n.], 2013. Disponível em: <[https://www.researchgate.net/publication/318494429\\_Live\\_Music-Making\\_a\\_Rich\\_Open\\_Task\\_Requires\\_a\\_Rich\\_Open\\_Interface](https://www.researchgate.net/publication/318494429_Live_Music-Making_a_Rich_Open_Task_Requires_a_Rich_Open_Interface)>. Acesso em: 2 jun. 2019.

STURM, B. Adaptive Concatenative Sound Synthesis and Its Application to Micromontage Composition. **Computer Music Journal**, v. 30, p. 46–66, dez. 2006. DOI: 10.1162/comj.2006.30.4.46.

SYNFUL. **Synful Orchestra - Products**. [S.l.: s.n.], 2019. <https://www.synful.com/SynfulOrchestra.htm>. Acesso em:

TEMPELAARS, S. **Signal Processing, Speech and Music: 1**. Edição: Stan Tempelaars. 1ª Edição. [S.l.]: Routledge, jan. 2016. ISBN 978-1-138-98192-8.

THOMAZ, L. F. **Aplicação à música de um sistema de especialização sonora baseado em Ambisonics**. Fev. 2007. text – Universidade de São Paulo. DOI: 10.11606/D.3.2007.tde-12072007-155603. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/3/3142/tde-12072007-155603/>>. Acesso em: 4 dez. 2021.

TOFFOLO, R. B. G. [ **As redes harmônicas de Pousseur aplicadas à composição musical em tempo real**. Mar. 2014. Tese (Doutorado) – Universidade Estadual Paulista. Accepted: 2014-08-13T14:50:58Z Publisher: Universidade Estadual Paulista (UNESP). Disponível em: <<https://repositorio.unesp.br/handle/11449/108808>>. Acesso em: 29 jun. 2021.

TZANETAKIS, G.; COOK, P. MARSYAS: a framework for audio analysis. **Organised Sound**, v. 4, set. 2002. DOI: 10.1017/S1355771800003071.

VALIATI, J. F. **Reconhecimento de voz para comandos de direcionamento por meio de redes neurais**. 2000. Tese (Doutorado).

VERCOE, B. The synthetic performer in the context of live performance. In: v. 1984, p. 199–200. Disponível em: <<https://quod.lib.umich.edu/cache//b/b/p/bbp2372.1984.026/bbp2372.1984.026.pdf#page=1;zoom=75>>. Acesso em: 15 fev. 2021.

VOLKMANN, J.; STEVENS, S. S.; NEWMAN, E. B. A Scale for the Measurement of the Psychological Magnitude Pitch. en, p. 2, 1937.

WANG, A. An Industrial Strength Audio Search Algorithm. en. In: 4TH International Conference on Music Information Retrieval, Baltimore. Baltimore, Maryland, USA: [s.n.], 2003. p. 7.

WANG, S.; WANDERLEY, M.; SCAVONE, G. The Study of Mapping Strategies Between the Excitators of the Single-Reed Woodwind and the Bowed String. In: [s.l.: s.n.], jan. 2020. p. 107–119. ISBN 9789811527555. DOI: 10.1007/978-981-15-2756-2\_9.

WANG, W.; YU, X.; WANG, Y. H.; SWAMINATHAN, R. Audio fingerprint based on Spectral Flux for audio retrieval. In: 2012 International Conference on Audio, Language and Image Processing. [S.l.: s.n.], jul. 2012. p. 1104–1107. DOI: 10.1109/ICALIP.2012.6376781.

WIENER, J.; BRONSON, N. **Facebook's Top Open Data Problems**. en-US. [S.l.: s.n.], out. 2014. Disponível em: <<https://research.fb.com/blog/2014/10/facebook-s-top-open-data-problems/>>. Acesso em: 12 abr. 2021.

WILSON, S.; COTTLE, D.; COLLINS, N. (Ed.). **The SuperCollider Book**. Cambridge, MA, USA: MIT Press, abr. 2011. ISBN 978-0-262-23269-2.

WISHNICK, A. Time-Varying Filters for Musical Applications. en, p. 8, 2014.

WRIGHT, M.; FREED, A. Open SoundControl: A New Protocol for Communicating with Sound Synthesizers. **International Computer Music Conference Proceedings**, v. 1997, 1997. ISSN 2223-3881. Disponível em: <<http://hdl.handle.net/2027/spo.bbp2372.1997.033>>.

XENAKIS, I. **Formalized Music: Thought and Mathematics in Composition**. [S.l.]: Pendragon Press, 1992. Google-Books-ID: y6lL3I0vmMwC. ISBN 978-1-57647-079-4.

XU, H.; ZHANG, X.; JIA, L. The Extraction and Simulation of Mel Frequency Cepstrum Speech Parameters. In: 2012 International Conference on Systems and Informatics (ICSAI2012). Yantai, China: IEEE, mai. 2012. p. 1765–1768. ISBN 978-1-4673-0199-2. DOI: 10.1109/ICSAI.2012.6223385.

YSTAD, S. Sound modeling using a combination of physical and signal models, jan. 1998.

ZÖLZER, U.; AMATRIAIN, X. (Ed.). **DAFX: Digital Audio Effects**. Chichester ; New York, N.Y.: Wiley, 2002. ISBN 978-0-471-49078-4.

ZWICKER, E. Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). **The Journal of the Acoustical Society of America**, Acoustical Society of America, v. 33, n. 2, p. 248–248, fev. 1961. ISSN 0001-4966. DOI: 10.1121/1.1908630.

ZWICKER, E. Procedure for calculating loudness of temporally variable sounds. **The Journal of the Acoustical Society of America**, v. 62, n. 3, p. 675–682, set. 1977. Publisher: Acoustical Society of America. ISSN 0001-4966. DOI: 10.1121/1.381580. Disponível em: <<https://asa.scitation.org/doi/abs/10.1121/1.381580>>. Acesso em: 27 set. 2021.

ZWITTER, A. Big Data ethics. en. **Big Data & Society**, v. 1, n. 2, p. 2053951714559253, jul. 2014. Publisher: SAGE Publications Ltd. ISSN 2053-9517. DOI: 10.1177/2053951714559253. Disponível em: <<https://doi.org/10.1177/2053951714559253>>. Acesso em: 12 abr. 2021.

# Apêndice A

## Apêndice

A seguir consta a partitura de *Do corpo ao pó*, incluindo as informações técnicas necessárias para a performance, tais como: equipamentos e *software*, configuração dos alto-falantes. Os arquivos referentes ao *patch*, a gravação da peça e os códigos em *python* utilizados nas análises dos exemplos no capítulo 3 estão disponíveis no link <https://github.com/oviniusc>.

Vinicius Cesar

# Do corpo ao pó

(2021)

*Dedicado ao violonista Gabriel Angelo*

Para violão e *live-electronics*

## Do corpo ao pó

O título da peça faz referência à dissertação e ao livro *Do corpo ao pó*: crônicas da territorialidade Kaiowá e Guarani nas adjacências da morte, do antropólogo Bruno Moraes. Este texto traz reflexões acerca do genocídio sofrido pelos povos Kaiowá e Guarani, no Mato Grosso do Sul no contexto dos conflitos fundiários, a partir de suas relações com o mundo, sendo a terra e o corpo, vivo ou morto, indissociáveis nesse debate. A partir de reflexões sobre a violência, e a morte, Moraes busca percorrer dois modelos de territorialidade; (1) a partir de concepções próprias dos indígenas das categorias territoriais; (2) a partir da expressão *tekooha* como contingências políticas do Estado que historicamente vincularam os Kaiowá e Guarani a um território específico. A imagem poética que permeia a peça surge a partir deste debate em torno da territorialidade que se direciona à uma crítica kaiowá e guarani à colonização de suas terras. Esta crítica é traçada a partir do delineamento de uma genealogia própria do corpo, e de uma escatologia. A vida e sobretudo a morte, são responsáveis por registrar as pessoas na terra até o ponto em que acabam se confundindo a ela. Por fim, o sujeito está definitivamente inscrito no espaço — o corpo passa a integrar o terreno de fato. Em outras palavras, a terra enquanto espaço é composta pelos corpos e pelas relações dos corpos com ela e dos corpos entre si, na vida e na morte — trocas materiais e imateriais entre os corpos e a terra.

Trazendo essa imagem poética para o contexto da criação, buscou-se evocar metáforas que remetessem à essas relações, bem como à noção de territorialidade descrita acima. Seja nas relações estabelecidas entre os processos interativos e os gestos instrumentais, ou na estruturação e desenvolvimento dos materiais musicais no nível micro e macro, diferentes desdobramentos dessa metáfora foram utilizados na concepção da peça.

## Eletrônica

### Requisitos técnicos

#### *Hardware e software*

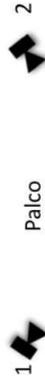
1 computador rodando *Pure Data* versão 0.51-4 ou superior

1 interface de áudio com suporte para 2 entradas e até 4 saídas

1 *fader* MIDI para ajustar volume dos processamentos durante a performance (opcional)

### Alto-falantes

4 alto-falantes posicionados ao redor do público



Palco

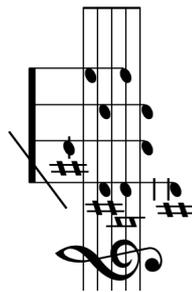


Público

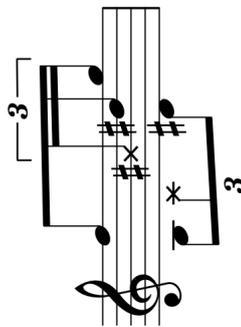


4

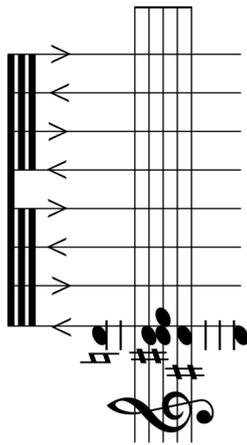
# Notas para performance



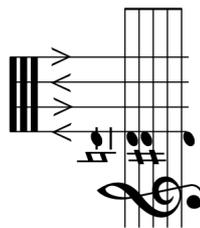
(a) Toçar o mais rápido possível.



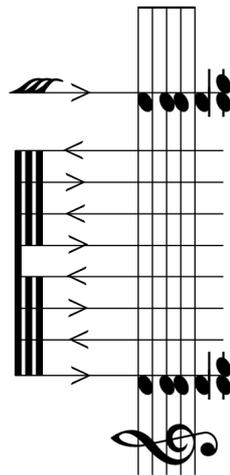
(b) Cabeça de nota em "x": martelar a corda com o dedo da mão esquerda sem atacar a nota com a mão direita.



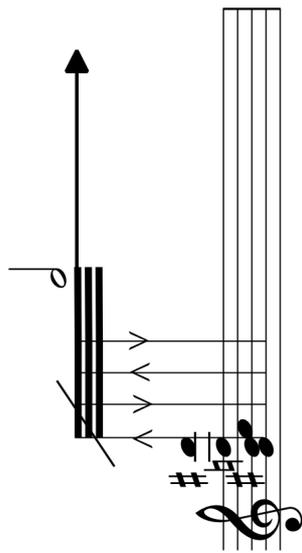
(c) Rasgueado: as setas indicam a direção da mão direita em relação à notas (seta para cima = das notas graves para as agudas, seta para baixo = das notas agudas para as graves).



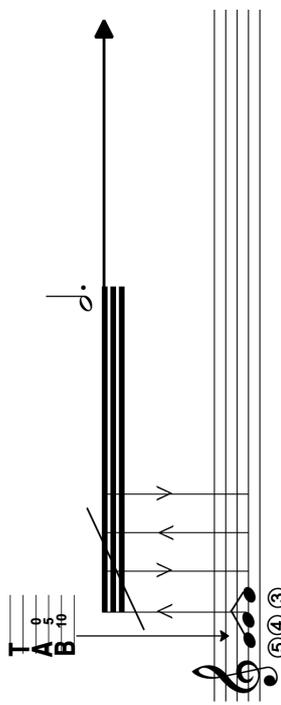
(d) *Apoggiatura*



(e) Rasgueado seguido de acorde acentuado.



(f) Rasgueado contínuo o mais rápido possível: a duração do evento deve respeitar a indicação de duração acima.



(g) Rasgueado em unísono: a tablatura acima indica em quais cordas as notas deverão ser tocadas.

# Do corpo ao pó

para Gabriel Angelo

Vinicius Cesar

**IA**

**IB**

**IIA**

**IIB**















