

UNIVERSIDADE ESTADUAL DE CAMPINAS

ADRIANO CLARO MONTEIRO

**CRIAÇÃO E PERFORMANCE MUSICAL NO CONTEXTO  
DOS INSTRUMENTOS MÚSICAIS DIGITAIS**

CAMPINAS, 2012



UNIVERSIDADE ESTADUAL DE CAMPINAS

ADRIANO CLARO MONTEIRO

**CRIAÇÃO E PERFORMANCE MUSICAL NO CONTEXTO  
DOS INSTRUMENTOS MUSICAIS DIGITAIS**

Dissertação apresentada ao curso de Mestrado em Música do Instituto de Artes da Universidade Estadual de Campinas como requisito parcial para obtenção do Título de Mestre em Música.

Área de Concentração: Processos Criativos

Orientador: Prof. Dr. Jônatas Manzolli

CAMPINAS, 2012

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO INSTITUTO DE ARTES DA UNICAMP**

M764c Monteiro, Adriano Claro.  
Criação e Performance Musical no Contexto dos Instrumentos  
Musicais Digitais /Adriano Claro Monteiro. – Campinas, SP:  
[s.n.], 2012.

Orientador: Jônatas Manzolli.  
Dissertação (mestrado) - Universidade Estadual de  
Campinas, Instituto de Artes.

1. Composição musical por computador 2. Música por  
computador. 3. Música eletrônica 4. Sistemas de recuperação da  
informação – Música. 5. Dispositivo em tempo real (Música)  
6. Meios Eletrônicos (Música) I. Manzolli, Jônatas.  
II. Universidade Estadual de Campinas. Instituto de Artes.  
III. Título.

(em/ia)

Informações para Biblioteca Digital

Título em inglês: Music Creation and Performance in the Context of Digital  
Musical Instruments.

Palavras-chave em inglês (Keywords):

Music composition with computers

Computer composition

Electronic Music

Systems for information retrieval – Music

Real-time devices (Music)

Electronic devices (Music)

Área de Concentração: Processos Criativos

Titulação: Mestre em Música

Banca examinadora:

Jônatas Manzolli [Orientador]

Mikhail Malt

Artemis Francelin Sanchez Moroni

Stéphan Oliver Schaub

Rafael Santos Mendes

Data da Defesa: 25-01-2012

Programa de Pós-Graduação: Música

**Instituto de Artes**  
**Comissão de Pós-Graduação**

Defesa de Dissertação de Mestrado em Música, apresentada pelo Mestrando Adriano Claro Monteiro - RA 030803 como parte dos requisitos para a obtenção do título de Mestre, perante a Banca Examinadora:

Prof. Dr. Jônatas Manzolli  
Presidente

Prof. Dr. Mikhail Malt  
Titular

Profa. Dra. Artemis Maria Francelin Sanchez Moroni  
Titular



À Aline,  
pelo carinho de  
todas as horas.



## **AGRADECIMENTOS**

Agradeço ao meu orientador Jônatas Manzolli por sua amizade e pela dedicação na orientação que foi imprescindível para realização dessa pesquisa.

Aos meus pais e meu irmão pelo apoio, ajuda e companheirismo nos caminhos da vida.

À Aline da Silva Alves pelo companheirismo, carinho e paciência antes, durante e depois das muitas horas de trabalho.

Aos Professores Mikhail Malt, Artemis Moroni, Stéphan Schaub e Rafael Santos Mendes por comporem a banca do exame de defesa dessa dissertação e por seus inestimáveis conselhos.

Aos Professores Denise Garcia, Adolfo Maia, Marcelo Gimenes pelos conselhos e contribuições para esse trabalho no exame de qualificação.

Aos amigos Lucas Araújo e Pedro Faria que colaboraram para o desenvolvimento dessa pesquisa nos trabalhos em parceria.

A todo o pessoal do NICS, alunos, professores, pesquisadores e funcionários que contribuíram com seu trabalho e com nossas trocas de idéias.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo auxílio financeiro para realização dessa pesquisa.



## **RESUMO**

Este trabalho apresenta um estudo sobre o computador como instrumento de performance musical. A partir deste estudo desenvolvemos processos de criação musical com base na representação algorítmica e no código computacional. Os dois principais recursos técnicos que utilizamos em nossos trabalhos foram: uma técnica de síntese sonora baseada na geração de forma de ondas por equações não-lineares; e métodos de recuperação de informação musical para prover análise e memória em tempo real dos parâmetros musicais de uma performance. O corpo da dissertação possui a seguinte estrutura: primeiramente, apresentamos as características dos instrumentos musicais digitais e as implicações de seu emprego na prática musical. Em seguida fazemos uma revisão do estado da arte da área de recuperação da informação musical e detalhamos os métodos específicos que utilizamos nesse trabalho, que incluem: a) a extração de parâmetros musicais de sinais de áudio monofônicos capturados de instrumentos acústicos; b) análises do comportamento dinâmico de medidas extraídas do sinal de áudio, através de mapas de Poincaré. Posteriormente, expomos o método de síntese por equações não-lineares. E por fim, apresentamos os processos de criação musical que desenvolvemos com base nos estudos realizados.



## **ABSTRACT**

This work presents a study about the computer as an instrument for musical performance. Based on this study we developed processes of music creation using the algorithmic representation and the computer code. The two main technical resources we applied in our works were: a sound synthesis method based on the generation of waveforms by non-linear equations; and methods for music information retrieval to provide a computer memory of the music parameters in real time. The dissertation has the following structure: first, we show the characteristics of the digital music instruments and the consequences of its using in the musical practice. Second, we review the state of the art of music information retrieval area and we detail the specific methods we used in our works, which include: a) the extraction of music parameters from monophonic audio signals captured from acoustic instruments; b) analysis of the dynamic behavior of measurements extracted from audio signal, by means of Poincaré maps. Next, we present the sound synthesis method by non-linear equations. Finally, we show the musical creative process we developed based on our studies.



## LISTAS DE FIGURAS

<b>Fig 1</b>	Representação dos músicos robôs do livro de al-Jazari - copiado por Farrukh ibn Abd al-Latif (1315 - tinta e ouro sobre papel)	<b>2</b>
<b>Fig 2</b>	Esquema que relaciona a estrutura interna dos Estudos Musicais apresentados no <b>Capítulo 5</b> com a organização deste trabalho	<b>6</b>
<b>Fig 3</b>	Representação em hélice das dimensões da percepção de notas musicais: Altura (sentido vertical) e Chroma (rotação da hélice). (FONTE: ONG, 2006)	<b>30</b>
<b>Fig 4</b>	Gravações de fragmentos melódicos com as funções extraídas, sendo respectivamente, de cima para baixo em cada quadrante: fluxo espectral, fluxo espectral positivo, fluxo espectral de diferença	<b>38</b>
<b>Fig 5</b>	O gráfico inferior apresenta a função de detecção obtida através do descritor ‘fluxo espectral de diferença’ que foi extraída da forma de onda no gráfico superior, que, por sua vez, contém a gravação de um fragmento melódico tocado por um saxofone tenor em articulação legato.	<b>40</b>
<b>Fig 6</b>	Função de detecção apresentada na Fig. 5 após o processamento do filtro passa-baixas	<b>41</b>
<b>Fig 7</b>	Função limiar extraída pelo algoritmo de limiarização dinâmica (equação 3.3) da função apresentada na Fig. 6.	<b>41</b>
<b>Fig 8</b>	Função resultante da subtração entre função de detecção função limiar. Gráfico superior apresenta a função resultante na mesma escala que os exemplos anteriores e gráfico inferior a mesma função normalizada	<b>42</b>
<b>Fig 9</b>	Exemplo final com a seleção dos picos apresentada no gráfico central. A fim de facilitar a comparação visual o gráfico superior retoma a função de detecção da Fig. 5 e o gráfico inferior sobrepõe os pontos de início selecionados com a forma de onda do som analisado.	<b>43</b>
<b>Fig 10</b>	Fluxograma do algoritmo de classificação harmônica	<b>46</b>
<b>Fig 11</b>	Extração de conteúdo harmônico de quatro acordes do segundo movimento da sonata opus K330 de Mozart. Na partitura acima os quatro acordes analisados e o gráfico central ilustra o espectrograma destes acordes sequenciais. Os gráficos inferiores contém a distribuição energética do chromograma de cada acorde nos <i>sliders</i> superiores e o vetor binário representado pelas filas de 12 quadrados	<b>47</b>
<b>Fig 12</b>	Extração do Chromograma dos dois primeiros compassos da peça número 3 das “Dez Peças para Quinteto de Sopros” de György Ligeti.	<b>49</b>
<b>Fig 13</b>	Exemplos de Mapas de Poincaré	<b>51</b>
<b>Fig 14</b>	Espectrogramas de duas seções da uma mesma gravação de <i>As Duas Criaturas que Estavam a Mesa de Chá não Tiveram esta Conversa</i>	<b>53</b>
<b>Fig 15</b>	Extração do Fluxo espectral das Seções A e B presentes na Fig. 14.	<b>54</b>

<b>Fig 16</b>	Mapas de Poincaré obtidos através das funções da Fig. 15. À esquerda mapa da seção A, e à direita mapa da seção B.	<b>54</b>
<b>Fig 17</b>	Tipologias de formas de onda geradas no móbile atrator. Acima, da esquerda para direita: ruído, onda periódica simples e onda periódica complexa. Abaixo, da esquerda para direita: som percussivo, silêncio e sequência de pulsos.	<b>62</b>
<b>Fig 18</b>	Exemplos gráficos comparando as funções: à esquerda antes do processo de interpolação e à direita após o processo de interpolação.	<b>63</b>
<b>Fig 19</b>	Resultado da aplicação do móbile de distorção. Na parte superior da figura uma onda periódica complexa a esquerda e a direita a mesma onda após o processo de waveshaping. Na parte inferior da figura mesma relação pré e pós waveshaping para uma amostra de ruído branco.	<b>69</b>
<b>Fig 20</b>	diagrama da relação entre entrada do contrabaixo e processamento computacional. Setas tracejadas indicam o fluxo do sinal de áudio e setas contínuas a relação gesto e resultado sonoro	<b>70</b>
<b>Fig 21</b>	Digrama do Móbile de síntese aditiva modulada por Phaser. Setas tracejadas indicam o fluxo do sinal de áudio e setas contínuas indicam os controles paramétricos em cada etapa.	<b>71</b>
<b>Fig 22</b>	GUI de <i>As duas criaturas que estavam à mesa de chá não tiveram esta conversa</i>	<b>72</b>
<b>Fig 23</b>	Esquema do algoritmo de automação da permutação dos valores de frequência de leitura da forma de onda	<b>73</b>
<b>Fig 24</b>	Esquema do novo sistema de <i>performance</i> com inclusão dos algoritmos para recuperação de informação musical e memória da performance	<b>74</b>
<b>Fig 25</b>	Interface de <i>As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram Esta Conversa</i> com adição do sistema de análise, memória e recuperação das informações da performance.	<b>76</b>
<b>Fig 26</b>	Interface de <i>As Duas Criaturas Baforaram seus Óculos</i>	<b>77</b>
<b>Fig 27</b>	À esquerda o fluxo de macroestruturas que definem a sequência de elementos da peça. À direita a construção interna da macroestrutura “Gesto soproII” e suas subestruturas que descrevem sequências de alterações paramétricas nos algoritmos de síntese e processamento sonoro.	<b>78</b>
<b>Fig 28</b>	Análise e segmentação de uma performance de <i>As Duas Criaturas Baforaram Seus Óculos</i> de acordo com os tipos de estruturas algorítmicas que a compõe. Acima as formas de onda dos dois canais resultante da gravação e abaixo os respectivos espectrogramas.	<b>79</b>
<b>Fig 29</b>	Análises espectrais dos elementos Sopro 1 e Sopro 2 referentes a segmentação da Fig. 28. Acima o espectrograma com seleção dos picos de energia e abaixo um gráfico que representa o perfil da frequência centro da distribuição de energia espectral e do desvio padrão da distribuição energética no espectro.	<b>81</b>

<b>Fig 30</b>	Interface de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i>	<b>87</b>
<b>Fig 31</b>	Formas de ondas geradas através da equação <i>Latoocarfian</i> para a primeira seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i> .	<b>88</b>
<b>Fig 32</b>	Acima as formas de onda dos quatro canais resultantes da gravação da primeira seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i> e abaixo os respectivos espectrogramas.	<b>89</b>
<b>Fig 33</b>	Formas de ondas geradas através da equação <i>Latoocarfian</i> para a segunda seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i> .	<b>90</b>
<b>Fig 34</b>	Espectrogramas dos quatro canais extraídos da gravação de uma performance da segunda seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i> .	<b>91</b>
<b>Fig 35</b>	Formas de ondas geradas através da equação <i>Latoocarfian</i> para a terceira seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i>	<b>92</b>
<b>Fig 36</b>	Acima as formas de onda dos quatro canais geradas pela gravação de uma performance da terceira seção de <i>Não Diziam-nos que iam Contar sobre nosso Passado</i> e abaixo os respectivos espectrogramas	<b>93</b>
<b>Fig 37</b>	Exemplos de tipos de envelope ADSR em diferentes instrumentos musicais	<b>119</b>
<b>Fig 38</b>	Centróide Temporal extraído das amostras sonoras de um evento tocado em uma caixa-clara (gráfico superior), contrabaixo (gráfico central) e violino (gráfico inferior).	<b>121</b>
<b>Fig 39</b>	A tabela direita apresenta uma função resultante da autocorrelação da onda dente de serra presente na tabela esquerda.	<b>123</b>
<b>Fig 40</b>	Fluxograma do algoritmo para extração de MFCC e BFCC	<b>135</b>

## LISTA DE TABELAS

<b>Tabela 1</b>	Descrição das Siglas utilizadas nas Tabelas 2, 3, 5 e 6	<b>94</b>
<b>Tabela 2</b>	Esquema de organização dos parâmetros e respectivos algoritmos do sistema de <i>As Duas Criaturas que estavam à Mesa de Chá Não Tiveram Esta Conversa</i>	<b>96</b>
<b>Tabela 3</b>	Esquema de organização dos parâmetros e respectivos algoritmos do sistema de <i>As Duas Criaturas Baforaram seus óculos.</i>	<b>98</b>
<b>Tabela 4</b>	Parâmetros estabelecidos para equação <i>Latoocarfian</i> no sistema de <i>As Duas Criaturas Baforaram seus Óculos.</i>	<b>99</b>
<b>Tabela 5</b>	Esquema de organização dos parâmetros e algoritmos relacionados a etapa de composição do sistema de <i>Não Diziam-nos que Iam Contar sobre Nosso Passado</i>	<b>100</b>
<b>Tabela 6</b>	Esquema de organização dos parâmetros e algoritmos relacionados a etapa de composição do sistema de <i>Não Diziam-nos que Iam Contar sobre Nosso Passado.</i>	<b>101</b>
<b>Tabela A.1</b>	Taxonomia empregada para os Descritores.	<b>117</b>

## SUMÁRIO

<b>INTRODUÇÃO</b>	<b>1</b>
<b>CAPÍTULO 1 – Conceitos Ligados À Luteria Digital</b>	<b>9</b>
1.1 Arquitetura dos Instrumentos Musicais Digitais	9
1.2 Principais Pontos da Pesquisa em Instrumentos Musicais Digitais	12
1.3 Aspectos Relevantes para Fundamentação Teórica da Pesquisa	15
1.3.1 Instrumentos Compostos	16
1.3.2 Performance com Instrumentos Musicais Digitais	18
<b>CAPÍTULO 2 - Extração de Conteúdo Musical do Sinal De Áudio</b>	<b>23</b>
2.1 Descritores Acústicos	24
2.1.1 Pré-Processamento	26
2.1.2 Fluxo Espectral	27
2.1.3 Chroma.	29
2.1.4 Média Quadrática (RMS).	31
<b>CAPITULO 3 - Metodologia de Análise</b>	<b>33</b>
3.1. Segmentação	36
3.1.1 Função para Detecção de Início (Onset)	37
3.1.2 Identificação de Picos	39
3.1.3 Função para Detecção de Fim (Offset)	44
3.2 Extração de Conteúdo Harmônico	44
3.3 Análise de Dinâmica Espectral via Mapas de Poincaré	50
3.3.1 Mapas de Poincaré.	50
3.3.2 Análise de Funções Extraídas por Descritores	52
<b>CAPÍTULO 4 - Síntese Sonora Digital via Equação <i>Latoocarfian</i></b>	<b>57</b>
4.1 Sistemas Dinâmicos Não-lineares em Composição e Síntese	57
4.1.1 Características dos Sistemas Dinâmicos	57
4.1.2 Dinâmicas Não-lineares, Fractais e Composição	59
4.2 Síntese Sonora via Equação <i>Latoocarfian</i>	61
<b>CAPÍTULO 5 - Processo Criativo</b>	<b>65</b>
5.1 As Duas Criaturas	65
5.2 <i>As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram esta Conversa</i>	67
5.2.1 <i>Waveshapper</i>	68
5.2.2 Transpositor	70
5.2.3 Sintetizador Aditivo e Phaser.	71
5.2.4 Sintetizador Não-linear via Equação <i>Latoocarfian</i>	72
5.2.5 Recuperação e Memória Computacional da Performance	74
5.3 <i>As Duas Criaturas Baforaram seus Óculos</i>	76

5.4 <i>Não Diziam-nos que Iam Contar sobre nosso Passado</i>	83
5.4.1 Primeira Seção	87
5.4.2 Segunda Seção	89
5.4.3 Terceira Seção.	91
5.5 Organização Paramétrica dos Sistemas	94
<b>CONCLUSÃO</b>	<b>103</b>
<b>REFERÊNCIAS</b>	<b>107</b>
<b>APÊNDICE A - CD Contendo os Exemplos de Áudio Referenciados no corpo do Texto e a Biblioteca de algoritmos Descritores PDescriptors</b>	
<b>ANEXO A- Descritores em PDescriptors: Taxonomia e Definições</b>	<b>117</b>
<b>A.1 Características Temporais</b>	<b>118</b>
<b>A.1.1 Características Temporais Globais</b>	<b>118</b>
<b>A.1.1.1 Centróide Temporal</b>	<b>119</b>
<b>A.1.2 Características Temporais Instantâneas</b>	<b>121</b>
<b>A.1.2.1 Média Quadrática (RMS)</b>	<b>121</b>
<b>A.1.2.2 Taxa de Cruzamento por Zero</b>	<b>122</b>
<b>A.1.2.3 Autocorrelação</b>	<b>122</b>
<b>A.2 Características Espectrais</b>	<b>124</b>
<b>A.2.1 Características do Envelope Espectral</b>	<b>125</b>
<b>A.2.1.1 Centróide Espectral</b>	<b>125</b>
<b>A.2.1.2 Desvio Padrão Espectral.</b>	<b>126</b>
<b>A.2.1.3 Obliquidade Espectral</b>	<b>126</b>
<b>A.2.1.4 Curtose Espectral</b>	<b>127</b>
<b>A.2.1.5 Nivelamento Espectral</b>	<b>127</b>
<b>A.2.1.6 Medida da Crista Espectral</b>	<b>128</b>
<b>A.2.1.7 Irregularidade Espectral</b>	<b>128</b>
<b>A.2.1.8 Inclinação Espectral</b>	<b>129</b>
<b>A.2.1.9 Decrescimento Espectral</b>	<b>129</b>
<b>A.2.1.10 <i>Roll-off</i></b>	<b>130</b>
<b>A.2.1.11 Conteúdo de Frequências Altas</b>	<b>130</b>
<b>A.2.2 Características da Dinâmica Espectral</b>	<b>131</b>
<b>A.2.2.1 Fluxo Espectral</b>	<b>131</b>
<b>A.2.2.2 Desvio de Fase</b>	<b>134</b>
<b>A.2.3 Descrição Global do Envelope Espectral Segundo Preceitos Psicoacústicos: MFCC e BFCC</b>	<b>135</b>
<b>A.2.4 Característica Espectral Segundo Preceitos Musicais: Chroma.</b>	<b>136</b>
<b>A.3 Características Psicoacústicas</b>	<b>137</b>
<b>A.3.1 Loudness</b>	<b>138</b>
<b>A.3.2 Energia Espectral por Banda Mel</b>	<b>138</b>
<b>A.3.3 Energia Espectral por Banda Bark</b>	<b>139</b>

# INTRODUÇÃO

O inventor árabe al-Jazari (1136–1206) descreveu diversas máquinas no “Livro de Conhecimentos sobre Dispositivos Mecânicos Engenhosos”<sup>1</sup>. Dentre relógios, calendários e outros objetos funcionais, vê-se também descrita uma “Banda musical Robótica” movida a energia hidráulica (Fig. 1). Antes de al-Jazari, também na chamada “Idade de Ouro Islâmica”, cerca de cem máquinas foram descritas no “Livro dos Mecanismos Engenhosos”<sup>2</sup>, escrito pelos irmãos Banū Mūsā na cidade que atualmente é Bagdá, Iraque, no ano de 850. Esse livro descreve um órgão movido a energia hidráulica que acionava cilindros intercambiáveis e um tocador de flauta automático (FOWLER, 1967).

Há indícios que, ainda anteriormente, Heron de Alexandria (cerca de 10-70 D.C.) foi o inventor de diversas máquinas automáticas, as quais usavam sistemas hidráulicos, eólicos e a vapor como fonte energética. Dentre elas destacamos um órgão movido a energia eólica (DRACHMANN, 1961).

Segundo Paine (2009) a automação em instrumentos musicais também possui uma longa história na Europa moderna, que inclui instrumentos mecânicos como as “caixas-de-música”, realejos, o órgão Wulitzer, a pianola, e a orquestra mecânica de Johann Maelzel (detentor da patente do metrônomo) chamada “Panharmonicon” composta de 42 “músicos robôs” e para a qual Beethoven compôs a primeira parte da marcha “Vitória de Wellington” Opus 91.

---

<sup>1</sup> *The Book of Knowledge of Ingenious Mechanical Devices: Kitāb fī ma'rifat al-hiyal al-handasiyya,*

<sup>2</sup> *The book of ingenious devices (Kitaab al-ohiyal)* (IBN SHAAKIR, MUHAMMAD IBN MUSAIBN et al., 1973)



**Fig. 1** - Representação dos músicos robôs do livro de al-Jazari - copiado por Farrukh ibn Abd al-Latif (1315 - tinta e ouro sobre papel)<sup>3</sup>

Esses são exemplos que ilustram ideários de máquinas musicais, desde a antiguidade até a modernidade. Tais inventos compartilham alguns princípios com a utilização musical de computadores. Pois os computadores podem ser descritos também como autômatos musicais. São máquinas que tentam modelar alguns dos “mecanismos físicos” da atividade musical. Talvez tenha sido também esta possibilidade que motivou os inventores de épocas anteriores a conceber os seus inventos. Não apenas descrever e re-inventar o mundo através de composições, mas poder somar a isso a possibilidade de descrever e re-inventar os mecanismos que regem o mundo dos sons musicais. Se este conceito fez realmente parte do ideário de todos esses inventores não nos é mais possível saber, mas certamente foi essa a motivação da pesquisa de mestrado aqui reportada.

<sup>3</sup> FONTE: [http://en.wikipedia.org/wiki/File:Al-Jazari\\_-\\_A\\_Musical\\_Toy.jpg#filehistory](http://en.wikipedia.org/wiki/File:Al-Jazari_-_A_Musical_Toy.jpg#filehistory), apud The Smithsonian's Museum of asian art.: <http://www.asia.si.edu/collections/zoomObject.cfm?ObjectId=9919> (consultado em 16/03/2012).

Na atualidade o estudo sobre Instrumentos Musicais Digitais ou, simplesmente, Instrumentos Digitais, concentra grande esforço de pesquisa interdisciplinar em Música e Tecnologia. Esses instrumentos vem alterando paradigmas do fazer musical que se consolidaram durante séculos de desenvolvimento e prática da Música Ocidental. Vários pesquisadores estudam de que maneira a concepção e utilização dessas ferramentas de produção sonora influenciam a prática musical e as relações sociais construídas em torno delas. Nos últimos anos, periódicos científicos destacados dedicaram volumes inteiros ao entendimento de vários aspectos deste tema. Dentre esses periódicos destacamos: *Organized Sound*, volume 14(2) de 2009 e volume 16(2) de 2011; *Contemporary Music Review*, volume 28(1) de 2009 e volume 22(4) de 2003; *Computer Music Journal*, volume 34(4) de 2010, *Journal of New Music Reserach* volume 38(3) de 2009, outros livros e fontes encontram-se nas referências bibliográficas.

O início da utilização de computadores em música data do final da década de 1950 quando Martin Klein e Douglas Bolitho criam nos EUA o programa “Tin Pan Alley” que gerava melodias a partir de um computador DATATRON, em 1956.

A primeira peça musical totalmente composta por intermédio de um computador foi criada por Lejaren Hiller e Leonard Issacson na Universidade de Illinois. Hiller e Issacson utilizaram um computador chamado Illiac no qual programaram um conjunto de algoritmos baseados em regras de contraponto e controle estocástico e com o qual criaram uma obra para quarteto de cordas chamada *Illiad Suite* (1956-57). Em 1960, em Paris, Pierre Bardaud criou as primeiras obras compostas por computadores no continente Europeu. Em 1962 Iannis Xenakis, um dos principais expoentes da computação musical no século XX, desenvolveu o programa *ST* para um computador IBM 7090 no qual implementou suas idéias de composicionais utilizando-se de processos estocásticos. Ele então criou uma série de peças geradas a partir desse mesmo algoritmo<sup>4</sup> (AMES, 1987) .

Contudo, a aplicação de computadores para performance musical ao vivo só foi possível a partir da década de 1980 quando os computadores tornaram-se potentes o suficiente para realizar processamento em tempo real. Mesmo assim, nesse período o

---

<sup>4</sup> As peças composta por Xenakis utilizando o programa ST são: ST/48-1,240162 ; ST/10 - 1,080262 ; ST/4 - 1,080262 ( edição para Quarteto de cordas do ST/10) ; ST/4 -1,030762 (Morsima - Amorisma) ; ST/10-1,030762 (Amorisma - Morsima) ; ST/10 -3,060962 (Atrées) ; ST/CosGauss (Polytope de Cluny) ; Parte de St'ratégies (1962) ; Parte de Eonta (1964).

alto custo das máquinas as restringia ao uso em grandes instituições de pesquisa onde foram aplicadas para processamento de áudio e em sistemas que acompanhavam a performance com instrumentos acústicos, visando proporcionar uma parte eletroacústica temporalmente maleável.

Algumas obras expoentes desse período são *Répons* (1981) de Pierre Boulez (uma das primeiras a utilizar esse recurso) e *Jupiter* de Philippe Manoury (1987). Os projetos de processamento computacional em tempo real dessas obras foram desenvolvidos no IRCAM durante a década de 1980 e início da década de 1990.

O ambiente computacional utilizado nestas duas composições foi o sistema de processamento digital 4X desenvolvido pelo pesquisador Di Giugno (DI GUIGNO, GERZSO, 1998) (LIPE, 1998) e o ambiente Max, desenvolvido por Miller Puckette, que era utilizado para controle de dados e comunicação via protocolo MIDI entre as máquinas hospedeiras e os processadores de sinal (PUCKETTE, 1988). No início da década de 1990 o sistema 4X foi substituído pela “IRCAM Signal Processing Station” (IPSW) planejada por Lindemann (LINDEMANN, et al. 1991) e composta de processadores Intel i860 conectados a uma máquina NeXT. Neste período Puckette incorporou uma biblioteca de processamento de sinais no sistema Max (PUCKETTE 1991).

Em meados da década de 1990 o computador popularizou-se como instrumento de criação e performance musical graças a expansão do mercado de computadores pessoais e o rápido desenvolvimento tecnológico dessas máquinas. Nesse período foram criados ambientes de programação especialmente voltados à música (e atualmente utilizados para artes digitais em geral) que encapsulavam procedimentos básicos da área de processamento de sinais para áudio e vídeo como unidades mínimas de linguagens de programação próprias.

Tais ambientes são extensivamente utilizados na atualidade para criação e pesquisa musical (os principais exemplos são MAX/MSP<sup>5</sup>, Pure Data<sup>6</sup>, Supercollider<sup>7</sup>, Kyma<sup>8</sup>, Chuck<sup>9</sup>). Surge com isso a prática de criação tanto de instrumentos digitais como de composições musicais experimentais que nascem da exploração e pesquisa

---

<sup>5</sup> <http://cycling74.com/>

<sup>6</sup> <http://puredata.info/>

<sup>7</sup> <http://supercollider.sourceforge.net/>

<sup>8</sup> <http://www.symbolicsound.com/cgi-bin/bin/view/Company/WebHome>

<sup>9</sup> <http://chuck.cs.princeton.edu/>

sobre as capacidades da tecnologia computacional como interatividade, automação e integração de recursos audiovisuais.

Na década de 2000, com a massiva expansão da rede mundial e com computadores cada vez mais portáteis, surgiram novas formas de Arte Musical Digital, que são próprias desses meios: as *webarts* (aplicativos ou páginas na *web* com os quais usuários interagem com processos sonoros) ou os aplicativos para telefones celulares e computadores de bolso (*smartphones*). A Arte Digital nos últimos anos saiu do domínio da pesquisa e da música experimental e tornou-se um gênero popular e de consumo em massa. Um exemplo é o popular aplicativo RjDj<sup>10</sup>, uma plataforma construída com o software Pure Data e criada para o desenvolvimento de aplicativos musicais para os sistemas operacionais *Apple* de computadores portáteis como iPhone ou iPad.

A pesquisa relatada nesta Dissertação insere-se dentro desse panorama atual da tecnologia digital aplicada à música. Os recursos e os conhecimentos para uso da computação em música são cada vez mais acessíveis e compartilhados mundialmente, o que tem estimulando o crescimento de novas práticas e linguagens musicais assim como a intensificação da pesquisa na área.

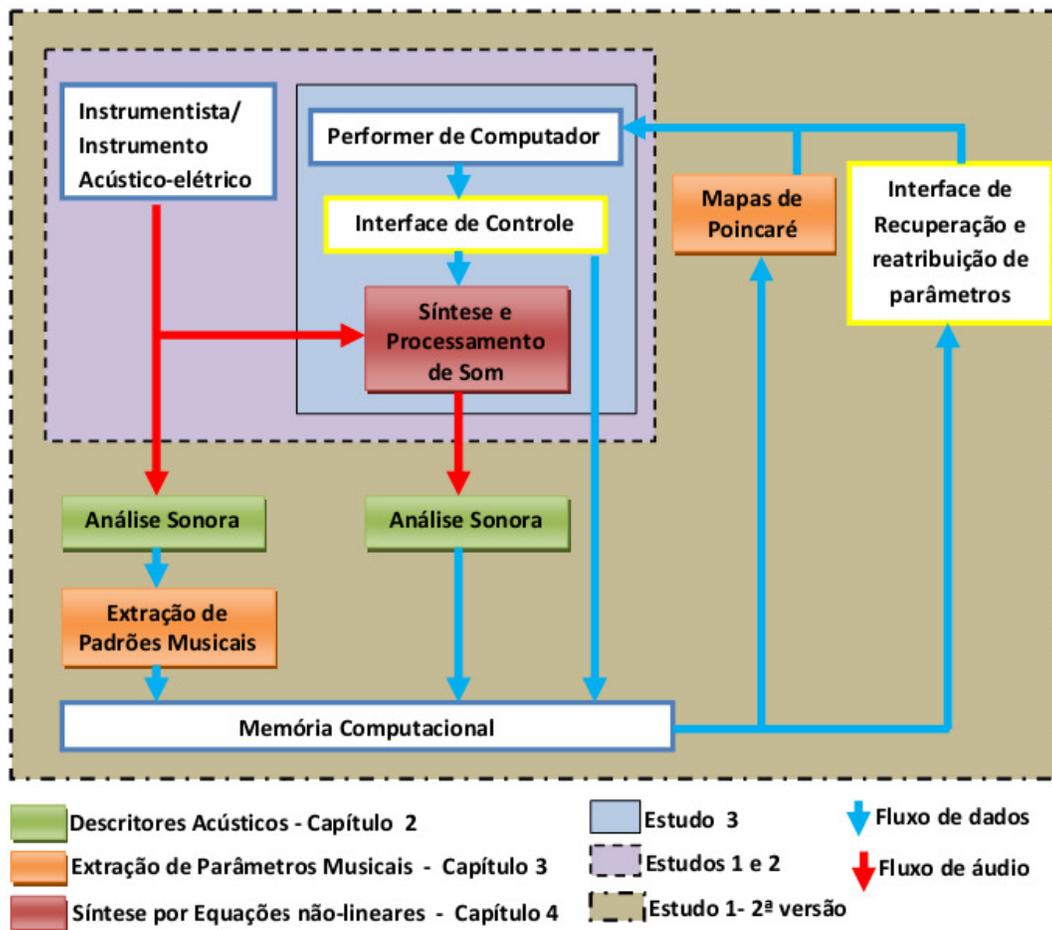
Especificamente, essa Dissertação apresenta um estudo sobre Instrumentos Musicais Digitais aplicados ao desenvolvimento de processos criativos musicais. As obras aqui apresentadas ao final dessa Dissertação estruturaram-se a partir de conceitos e métodos estudados durante nossa pesquisa que é reportada em cinco capítulos.

No **Capítulo 1** discutimos as noções de *luteria digital* que embasaram conceitualmente nossa pesquisa. Destacamos o conceito de *instrumento composto* e aspectos da performance com instrumentos digitais que são a principal fundamentação teórica para as composições apresentadas no **Capítulo 5** onde descrevemos três Estudos musicais algorítmicos.

A Fig. 2 apresenta um diagrama dos sistemas computacionais desenvolvidos para esses Estudos e relaciona-o também com os outros Capítulos dessa Dissertação nos quais descrevemos os procedimentos técnicos para implementação dos algoritmos que fizeram parte dos sistemas computacionais desenvolvidos.

---

<sup>10</sup> <http://rjdj.me/>



**Fig. 2** - Esquema que relaciona a estrutura interna dos Estudos Musicais apresentados no **Capítulo 5** com a organização deste trabalho.

Conforme a Fig. 2 indica, nos **Capítulos 2 e 3** apresentamos duas etapas dos métodos de análise de áudio que foram aplicados para a segunda versão do primeiro Estudo musical apresentado no **Capítulo 5**. Empregamos esses métodos em um sistema de análise computacional para extração de informações musicais paramétricas a partir do áudio capturado durante a *performance*. Os dados extraídos do áudio são armazenados na memória do computador e posteriormente utilizados pelo *performer* computacional que pode reatribuí-los ao algoritmo de geração musical como valores paramétricos que controlam o sistema de síntese sonora.

No **Capítulo 4** descrevemos um método de *síntese sonora por equações não-lineares*. Esse método é destacado em um Capítulo da dissertação, pois apresentou resultados sonoros que tiveram grande importância na criação musical dos três Estudos aqui reportados.

Finalmente, no **Capítulo 5** apresentamos os três Estudos algorítmicos musicais que fazem parte de uma série chamada *As Duas Criaturas* e são o resultado criativo de nossa pesquisa. Este capítulo é dedicado à discussão do processo criativo dos Estudos relacionando-os aos **Capítulos 2, 3 e 4** e aos conceitos sobre Instrumentos Digitais que foram elencados **Capítulo 1**.

A pesquisa foi voltada ao estudo e desenvolvimento de uma gama variada de processos computacionais que foram então concentrados na criação musical. Concebemos obras musicais algorítmicas que apresentam tanto propriedades de instrumento musical quanto propriedades de representação musical. Ou seja, os algoritmos computacionais desenvolvidos são ao mesmo tempo instrumentos para performance musical, pois possibilitam o controle de um performer humano e também definem estruturas que representam as idéias musicais do autor.

Essas duas características manifestam-se em diferentes circunstâncias e graus em cada um dos Estudos, que são os seguintes: *As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram essa Conversa*, para computador e contrabaixo elétrico, *As Duas Criaturas Baforaram seus Óculos*, para computador e trombone, *Não Diziam-nos que iam Contar Nosso Passado*, para computador solo.

Os resultados obtidos durante o mestrado também foram publicados em um conjunto de 7 publicações (MANZOLLI E MONTEIRO 2011; MONTEIRO E MANZOLLI 2010; MONTEIRO E MANZOLLI 2011a, 2011b, 2011c, 2011d, 2011e) sendo 3 artigos internacionais e 4 nacionais, assim como 4 apresentações de trabalhos, 1 em evento nacional e 3 em eventos internacionais. Junto ao corpo da dissertação apresentamos exemplos sonoros que auxiliam o entendimento do texto (vide o CD Apêndice A deste trabalho).

Como subproduto da pesquisa produzimos uma biblioteca de análise de áudio para o ambiente de programação Pure Data chamada *PDescriptors* e que está disponibilizada livremente em uma página da internet<sup>11</sup> e também está contida no CD Apêndice A desta Dissertação. Essa biblioteca contém as implementações dos métodos algorítmicos de análise de áudio estudados durante a pesquisa e apresentados no **Capítulo 1 e 2** e suas definições formais são apresentadas no **ANEXO A**.

---

<sup>11</sup> <https://sites.google.com/site/pdescriptors/>

Em suma, a pesquisa aqui reportada apresenta possibilidades de criação musical algorítmica e demonstra o potencial que essa abordagem tem de encapsular num mesmo sistema computacional tanto o projeto do Instrumento Musical Digital quanto a representação ou escritura dos eventos que constituem o discurso musical.

# CAPÍTULO 1

## CONCEITOS LIGADOS À LUTERIA DIGITAL

Neste primeiro capítulo é feito um embasamento conceitual sobre os instrumentos musicais digitais. Primeiramente, expomos os princípios de construção desses instrumentos, em seguida apontamos os principais aspectos técnicos e conceituais da área e os descrevemos resumidamente. Por fim, detalhamos os conceitos que foram importantes para concepção dos sistemas computacionais musicais detalhados ao final dessa dissertação.

### 1.1 Arquitetura dos Instrumentos Musicais Digitais

Nos Instrumentos Digitais, o processo de geração sonora é descrito em código computacional que é um esquema simbólico para representação de processos formais e não emite som. O corpo material de um instrumento digital é composto pelo *hardware* que executa os procedimentos algorítmicos (que é composto ao menos por uma CPU computacional, um aparelho conversor de sinal digital para analógico, e caixas acústicas para amplificação e emissão do sinal sonoro). Os sistemas computacionais podem funcionar automaticamente e não há a necessidade, *a priori*, de controle humano ou excitação por um fenômeno externo ao próprio computador (exceto a energia elétrica para manter o aparelho funcionando). Quando é necessário algum controle externo sobre o sistema, interfaces construídas com dispositivos eletro-eletrônicos (e.g. sensores, câmaras e botões) propiciam a comunicação com o ambiente ou com agentes humanos.

O pesquisador Magnusson (2009) compara os instrumentos musicais digitais os instrumentos musicais acústicos. Primeiramente, observa que tanto a construção, quanto a aprendizagem e o uso dos instrumentos acústicos podem ser entendidas como ações cognitivas corporificadas [de acordo com a teoria da cognição corporificada (*Embodied Cognition*) de Varela et al. (1991)].

A construção dos instrumentos acústicos é feita através de engajamento físico com o material. Quando pensamos no trabalho de um lutiê que utiliza métodos artesanais (e ainda hoje em dia boa parte da luteria é artesanal), podemos imaginar que esse artesão determina alterações no formato do material com o qual constrói o instrumento, realiza testes e, eventualmente, chega a acertos por tentativa e erro. Esse é

um processo predominantemente corporificado e pouco teórico no qual o método de construção é de-baixo-para-cima, ou seja, muitas vezes a construção instrumental é realizada sem o conhecimento das propriedades físicas do sistema de produção acústica do instrumento. O método se dá pelo conhecimento empírico, através de relações de causa e efeito produzidas pela contínua exploração dos materiais.

O instrumento musical resultante deste tipo de trabalho não nasce como uma invenção, mas como resultado de um processo evolutivo. Esse é constituído de um corpo material modelado para gerar uma resposta acústica em reação a um estímulo físico de energia. Interface e geração sonora são duas propriedades indissociáveis de um sistema de produção acústica.

Em contrapartida, nos instrumentos digitais o núcleo gerador de som é detreminado por instruções formais e lógicas que são descritas por uma notação simbólica. Trata-se de um código que pode ser desenvolvido e planejado independentemente das características materiais do equipamento computacional que irá executá-lo. A criação de um instrumento digital requer conhecimentos prévios sobre processamento digital de sinais, programação de computadores, acústica e interação homem-máquina.

Há um aumento na complexidade e quantidade de informações envolvidas na construção de instrumentos digitais que difere da construção de-baixo-para-cima dos instrumentos acústicos. Inversamente, instrumentos digitais são construído de-cima-para-baixo, isto é, são construídos através de algoritmos e conhecimentos sobre procedimentos de geração sonora que serão posteriormente executados por uma máquina.

Na visão de Magnusson (2009), instrumentos digitais podem ser entendidos sob a teoria da mente expandida [*Extended Mind* elaborada por Clark e Chalmers, (1998)]. O instrumento representa modelos mentais de alto nível simbólico que carregam um alto grau de conhecimento e podem compartilhar a execução de tarefas cognitivas com o usuário ou liberá-lo do tratamento de certas informações. Embora instrumentos acústicos também possuam um nível de encapsulamento de conhecimento, segundo Magnusson (2009), os instrumentos digitais potencializam essa relação e são registros de contextos culturais extremamente localizados, de uma pequena comunidade ou até mesmo de práticas pessoais.

Em resumo, instrumentos musicais digitais podem ser divididos nos elementos: a) *interface* e b) *algoritmo*, c) *equipamento de hardware para execução dos algoritmos e emissão acústica*.

Os computadores são uma tecnologia desenvolvida, *a priori*, fora do contexto da prática musical. Eles foram introduzidos na música assim como em quase todas as áreas do fazer humano devido à intrínseca generalidade de aplicações que um sistema de processamento de lógica formal apresenta. Eles são funcionais para tudo o que possa ser representado numericamente ou digitalmente. Por esse motivo, consideramos que os elementos que caracterizam os instrumentos musicais digitais, ou seja, os elementos que particularizam a generalidade do computador para aplicações como instrumentos musicais, na conjuntura atual, são:

- 1) os algoritmos voltados para fim musical;
- 2) as interfaces físicas específicas à música — que propiciam tipos de controle humano de maior interesse para a prática musical.

Quanto ao hardware para processamento, tipicamente um instrumento musical digital é desenvolvido sobre plataformas computacionais comerciais e com funcionalidades genéricas (equipamentos de *hardware* e sistemas operacionais).

Em relação à divisão da arquitetura dos instrumentos musicais digitais o enfoque do nosso trabalho recaiu sobre o elemento (1) descrito acima. Durante o mestrado pesquisamos algoritmos para o uso do sistema computacional como instrumento musical. A pesquisa sobre interfaces físicas (*hardware*) para performance musical esteve fora do escopo desse trabalho, porém, vale ressaltar que usamos as interfaces padrões do computador com o propósito de controle dos sistemas desenvolvidos (relatados no **Capítulo 5**), e também usamos interfaces musicais comerciais contendo *knobs* e *sliders*, e baseadas no protocolo de comunicação MIDI.

Nossa pesquisa sobre interface humano-computador em performance musical foi direcionada para utilização de métodos algorítmicos, baseados em procedimentos computacionais de análise de áudio para extração de padrões musicais, que foram aplicados como ferramentas para controle e interação musical por parte do performer computacional em um dos Estudos musicais criados. Os métodos de análise de áudio são relatados nos **Capítulos 2 e 3** desse trabalho e sua aplicação em um dos sistemas musicais desenvolvidos é apresentada no **Capítulo 5**.

Nas próximas subseções fazemos um levantamento da pesquisa sobre Instrumentos Musicais Digitais, bem como questões relativas à sua utilização na prática musical. Por fim, dissertamos sobre os pontos especificamente mais relevantes para o desenvolvimento dos três Estudos.

## **1.2 Principais Pontos da Pesquisa em Instrumentos Musicais Digitais**

A criação de instrumentos musicais digitais é uma atividade que envolve conhecimentos provenientes de diversas áreas. Por essa circunstância, a pesquisa em instrumentos musicais digitais divide-se em diversos focos de interesse. Encontram-se na literatura trabalhos voltados à pesquisa tecnológica para desenvolvimento dos instrumentos digitais, bem como estudos conceituais interessados no debate sobre as implicações trazidas à prática musical que envolve esses instrumentos. Duas referências abrangentes podem ser vista no trabalho de Miranda e Wanderley (2006) e Jordà (2005).

Dentre os assuntos técnicos e conceituais, destacamos a seguir alguns dos principais pontos de pesquisa sobre os instrumentos musicais digitais.

### **a) Interfaces**

A pesquisa relacionada com interfaces humano-computador aplicadas à música pode enfocar: I) questões sobre o desenvolvimento tecnológico de novas interfaces (GOTO, 1997; ORIO, 2002; WANDERLEY, ORIO, 2002); II) Instrumentos musicais aumentados (*Hyperinstruments*) (MACKOVER, 1992; YANG, 2002); III) a classificação do gesto físico humano (CADOZ, WANDERLEY, 2000; VAN NORT 2009) e a classificação dos tipos interface de captura do gesto (WANDERLEY, DEPALLE, 2004); IV) questões acerca do mapeamento<sup>12</sup> entre interface e parâmetros do algoritmo (CARAMIAUX, 2008; LEVITIN et al 2002; OVERHOLT, 2009), dentre outros.

Em relação a demais referências no assunto, evidenciamos um encontro científico internacional chamado NIME<sup>13</sup> (New Interface for Music Expression) voltado aos desenvolvimentos na área de interfaces musicais que centraliza grande parte das referências dessa área.

---

<sup>12</sup> Mapeamento é o termo utilizado

<sup>13</sup> <http://www.nime.org/> é o endereço da página do NIME que pode ser consultada na internet.

## **b) Síntese e processamento de sinais de áudio.**

A manipulação do sinal de áudio é um elemento condicional para criação de instrumentos digitais, pois é essa a classe de algoritmos responsável pela sonorização do instrumento. A literatura em síntese sonora digital e processamento de áudio digitalizado é extensa e contém uma diversidade de métodos que podem ser aplicados, inclusive combinadamente. Apontamos algumas referências importantes, pois reúnem grande parte das técnicas de síntese e processamento sonoro digital aplicado à música: os trabalhos de Curtis Roads (1996, 2001), de Miller Puckette (2006), e de Richard Boulanger (2000).

## **c) Generatividade**

Generatividade é o termo empregado para nomear a capacidade do instrumento musical digital de gerar material sonoro com certo grau de novidade e imprevisibilidade coerentemente contextualizado a uma performance musical específica. Tal possibilidade deriva da autonomia operacional dos sistemas computacionais e da sua capacidade de processar informação. A combinação desses recursos pode ser empregada para gerar automaticamente material sonoro-musical em tempo real. Existem diversos trabalhos que versam sobre as propriedades da generatividade musical e sobre tipos de algoritmo aplicados para esse fim (BERG, 2009; COLLINS, 2003, 2008; WHALLEY, 2009). Exemplos de questões relacionadas ao tema podem ser vistos no periódico *Contemporary Music Review* volume 28, edição 1, do ano de 2009.

## **d) Sistemas Musicais Interativos**

Apesar dos Sistemas Musicais Interativos possuírem um destaque particular na literatura da área de música computacional, nós os entendemos nesse trabalho como um tipo específico de Instrumento Musical Digital. A principal característica desses sistemas é explorar a interatividade propiciada pelo computador para estabelecer um diálogo musical com os músicos humanos durante uma performance. Esses sistemas interpretam os dados capturados do ambiente através de interfaces e geram uma resposta musical contextualizada às informações extraídas do sinal musical capturado.

Esta abordagem geralmente envolve técnicas das áreas de ciência cognitivas (GIMENES, 2008) inteligência artificial e aprendizagem de máquina (ASSAYAG et al. 2006; PACHET, 2003) Em suma, procura-se através nesses sistemas

modelar os comportamentos de um agente humano durante uma *performance* musical. Alguns exemplos de algoritmos que desempenham a função de agentes musicais autônomos em *performance* são Voyager (LEWIS, 2000), Ime (GIMENES, 2008), OMax (ASSAYAG et al. , 2006) Continuator (PACHET, 2003). Demais questões sobre interatividade computacional aplicada a música podem ser vistas em (DRUMMOND , 2009; GIMENES 2008; PAINE 2002). Vale destacar o trabalho de Rowe (1993) como uma das principais referências sobre a classificação de Sistemas Musicais Interativos.

#### **e) Criação de Instrumentos Digitais**

Enquadram-se nesse tópico estudos que visam entender as conseqüências dos Instrumentos Musicais Digitais nas etapas de produção musical anteriores a *performance*. Devido à natureza simbólica e esquemática de parte da anatomia dos Instrumentos Digitais — isto é, devido ao código computacional — um Instrumento Digital pode conter instruções para execução autônoma de estruturas musicais altamente elaboradas. Essa propriedade faz que certas características dos Instrumentos Digitais sejam próximas das características da notação musical, mesmo que mantenham suas propriedades instrumentais voltadas à *performance*. Instrumentos Digitais que encapsulam estruturas musicais específicas, altamente elaboradas e direcionados à composições musical particulares são denominados na literatura de *Instrumentos Compostos* (SCHNELL, BATTIER, 2002).

#### **f) Performance**

Considerando que certas propriedades dos Instrumentos Musicais Digitais são muito distintas ou inexistentes nos instrumentos acústicos tradicionais — como, por exemplo, os métodos generativos, os métodos de interatividade, a desconexão e múltiplas possibilidades de mapeamento entre interface e sistema de sonorização — o instrumentista de computador pode ter uma função muito distinta se comparada à função que habitualmente lhe é designada na música com instrumentos tradicionais. Dentre os tópicos que envolvem o estudo desse fenômeno estão: o papel das interfaces no controle dos algoritmos (MAGNUSSON, 2010); o papel do instrumentista na *performance* de sistemas generativos (GUREVICH, FYANS, 2011; WESSEL, WRIGHT, 1992); a atuação de um performer em sistemas musicais interativos (GUREVICH, FYANS, 2011; MANIATAKOS et al. 2010); novas formas de *performance* musical (COLLINS, 2003); o conceito virtuosismo instrumental em relação aos instrumentos digitais (BOWN, et al. 2009; GUREVICH , FYANS, 2011;

MAGNUSSON 2010); questões envolvendo a recepção do público de performances com instrumentos digitais (GUREVICH, FYANS, 2011), entre outros.

### **g) Manutenção e recuperação**

A conservação e recuperação do repertório dedicado a instrumentos eletrônicos e computacionais e a manutenção dos instrumentos são motivos de pesquisa atual (BATTIER, 2004; BULLOCK, COCCIOLI, 2006; POLFREMAN et al., 2006; TERUGGI, 2004; WETZEL, 2006; ZATTRA, 2004). Há um alto grau de obsolescência e desaparecimento dos instrumentos eletro-eletrônicos e computacional que ocorre rapidamente e é motivado pelo ideal de aperfeiçoamento e substituição tecnológica, por opções estéticas dos artistas entre a sonoridade dos diferentes equipamentos, e por questões comerciais e influência do mercado. Por exemplo, Roads (2001, p.45-49) apresenta uma listagem 99 instrumentos de tecnologia elétrico-eletrônicos criados entre 1899 a 1950, dentre os quais poucos são conhecidos popularmente hoje em dia. Como consequência direta, o desaparecimento dos instrumentos implica na impossibilidade de execução do repertório a eles dedicado.

Em relação aos instrumentos digitais o fator crítico de sua obsolescência não está no desaparecimento do *hardware*, que quase sempre cumpre as funções de memória e processamento de dados para execução de instruções determinadas nos programas. Mas está no desuso do código de representação desses algoritmos: a incompatibilidade de sistemas operacionais, softwares e linguagens de programação. Neste caso, os trabalhos de 'arqueologia' e recuperação requerem a compreensão do próprio software ou linguagem para readaptação dos procedimentos de escritura algorítmica do compositor e, em relação aos processos de manipulação do sinal de áudio neste suporte, a tradução e entendimento do código fonte dos softwares utilizados, quando acessíveis.

## **1.3 Aspectos Relevantes para Fundamentação Teórica da Pesquisa**

O desenvolvimento dos sistemas musicais relatado no **Capítulo 5** foi fundamentado em alguns dos tópicos de interesse da pesquisa em Instrumentos Musicais Digitais listados acima.

Sob o ponto de vista conceitual, dois tópicos fundamentaram o desenvolvimento dos sistemas musicais:

- I) o conceito de instrumentos compostos;

II) as questões pertinentes a performance com instrumentos musicais. Nas subseções seguintes detalhamos esses conceitos.

### **1.3.1 Instrumentos Compostos**

Como afirmamos anteriormente, os instrumentos musicais acústicos possuem propriedades inerentes ao seu comportamento material que antecipam os conhecimentos científicos sobre eles. Por exemplo, a manipulação dos parâmetros de altura e timbre do som existe nos instrumentos acústicos “... muito antes de Fourier e Helmholtz introduzirem suas teorias sobre funções sinusoidais e timbre” (MAGNUSSON, 2009, p. 171). Ao contrário, instrumentos digitais são construídos a partir de esquemas simbólicos, pois sua natureza (i.e. os algoritmos computacionais) é a representação abstrata. A construção de um instrumento digital é facilmente modificada e adaptada, pois uma simples alteração do esquema simbólico pode implicar em mudanças na resposta do instrumento sem que para isso se tenha a dificuldade e a quantidade de trabalho necessária na manipulação de algum material físico.

Por conta de sua natureza abstrata, os instrumentos digitais também incluem propriedades da notação musical. Muitas características do código são similares às da partitura, dentre elas podemos elencar que: são estruturas de representação simbólicas e sintáticas; possuem alta capacidade para reprodutibilidade e facilidade de compartilhamento graças ao formato textual; são maleáveis à intertextualidade, ou seja, à re-contextualização de suas partes ou estruturas típicas; são suscetíveis ao encapsulamento de estruturas musicais particulares e altamente elaboradas. Ou seja, programas computacionais aplicados à música podem apresentar um caráter ambíguo, pois incorporam duas funções: I) *são o processo de geração sonora e como tal instrumentos musicais* e II) *são algoritmos de representação do código musical e como tal incorporam as funções relacionadas à composição e à partitura*.

Devido a esses fatores, e também devido à autonomia energética e à capacidade de processamento de dados que permite que o instrumento realize as instruções musicais automaticamente, parte dos músicos que optaram pelo uso de instrumentos digitais começaram a programar seus próprios instrumentos com intuito de inserir nestes os elementos das suas idéias sobre estruturação musical, e, ao mesmo tempo, explorar as idiosincrasias que as propriedades que esse novo meio de criação introduzem à linguagem musical. Os produtos deste tipo de luteria que funde-se com a

composição musical são denominados na literatura de *instrumentos compostos* (SCHNELL, BATTIER, 2002).

Dummond (2009), assim como Bown et al. (2009), observam que este conceito não é novo, mas é utilizado desde o início da música eletrônica. Gordon Mumma, desenvolvendo seus trabalhos para trompa e eletrônica ao vivo, considerava ambos, composição e criação instrumental como parte do mesmo processo criativo. “Eu considero que o projeto e a construção de circuitos é realmente composição (...) meus instrumentos são inseparáveis das minhas composições” (MUMMA, 1967).<sup>14</sup>

A tecnologia digital potencializou essa prática graças à facilidade de aprendizagem, facilidade de operação, e o maior compartilhamento social dos procedimentos técnicos para programação de computadores, se comparada com o *design* de circuitos eletrônicos feitos por Mumma.

Também observando essa convergência, Bown et al. (2009) apresentam o conceito *objetos comportamentais* para definir o papel do *software* na prática musical. Segundo os autores, esse conceito enfatiza a natureza ativa de programas de computadores que exibem comportamentos complexos como máquinas ou estruturas orgânicas ao mesmo tempo em que desempenham o papel de “unidades tangíveis de troca social”, ou ferramentas para criação.

“Objetos comportamentais podem agir como mediadores entre pessoas no desenvolvimento de estilos e idéias musicais, movendo-se pelas redes sociais e desenvolvendo-se no ritmo das idéias, ilimitados pela materialidade dos objetos físicos. Objetos comportamentais podem ser extensivamente reconfigurados pelas pessoas, permitindo a adaptação flexível dos sistemas para os contextos de performance. Objetos comportamentais podem interagir entre eles com conseqüências generativas. Em cada uma dessas formas de interação, objetos comportamentais podem ser ativos na condução do processo global.” (Bown et al., 2009, p. 194)<sup>15</sup>

---

<sup>14</sup> I consider that my designing and building circuits is really ‘composing’ (...) my ‘instruments’ are inseparable from the compositions themselves.” (Mumma, 1967)

<sup>15</sup> Behavioural objects can act as mediators between people in the development of musical styles and ideas, moving through social networks and developing at the pace of ideas, unbounded by the materiality of physical objects. Behavioural objects can be extensively reconfigured by people, allowing the flexible adaptation of systems to performance contexts. Behavioural objects can interact with each other with generative consequences. In each of these forms of interaction, behavioural objects may be active in driving the overall process. (Bown et al., 2009, p. 194).

Bown et al. argumentam que o conceito de objeto comportamental está mais próximo dos papéis que de fato o código computacional desempenha na prática musical, se comparado com a abordagem que eles nomeiam de *paradigma do instrumento acústico*. Essa última é relacionada às abordagens dos modelos tradicionais de prática musical que distinguem as atividades como composição, interpretação e luteria.

Resumidamente, constata-se que o código computacional inserido na prática musical tem papel de pivô e é adaptável para as funções que essa prática abrange. Ele pode ser um instrumento para execução ou para auxílio da concepção de uma obra, pode ser o objeto de apreciação artística ele mesmo, o agente responsável pela criação musical, e um meio de representação e transmissão de conhecimento.

### **1.3.2 Performance com Instrumentos Musicais Digitais**

A consequência mais direta da separação entre interface e núcleo gerador de som é a desvinculação entre o *gesto instrumental* e a *resposta sonora*. Não existe, necessariamente, relação de causa e efeito entre o processo de produção sonora e as propriedades físicas de uma interface computacional, pois qualquer interface pode ser conectada a qualquer tipo de algoritmo gerador de som. Com isso, gestos idênticos, captados por uma mesma interface, ou por alguma outra, podem apresentar respostas sonoras completamente diferentes. Destacamos algumas características dos Instrumentos Digitais que contribuem para essa dissociação:

- As propriedades generativas e/ou mapeamentos indiretos tendem a criar uma dissociação entre a percepção visual do gesto físico do intérprete e a audição da resposta sonora.
- O desenvolvimento tecnológico separado das interfaces e dos algoritmos de síntese e tratamento de áudio colabora para a multiplicidade de associações entre diversos tipos desses elementos.
- A escolha pessoal da interface pode ser motivada pela sua capacidade de suprir ou neutralizar as limitações motoras do performer. Esse critério pode induzir a seleção de gestos de controle pequenos, inexpressivos visualmente e descorrelacionados com o resultado sonoro.

A prática do *live coding*<sup>16</sup> (MAGNUSSON, 2011; NILSON, 2007; SORENSEN, BROWN, 2007), por exemplo, está intrinsecamente associada ao problema de expressividade visual da performance, porque neste gênero a performance é o ato de programar os *scripts* em tempo real. Uma solução encontrada pelos praticantes de *live coding* foi a projeção em vídeo da tela do computador do *performer* com intuito que o público possa acompanhar as decisões do performer, lendo a interface gráfica do computador (COLLINS, 2003). Existem casos em que além do *script* (que requer um conhecimento prévio sobre a estrutura da linguagem de programação para que público o entenda) e das interfaces gráficas, o *performer* também digita na tela comentários sobre o que ele está fazendo, como uma guia ao público (GUREVICH, FYANS, 2011). E em alguns casos há projeção de vídeos que também são manipulados pela programação em tempo-real e compõe uma camada artística visual da *performance* eletrônica.

De modo geral, as possibilidades de controle que as interfaces dos Instrumentos Digitais oferecem são limitadas e diretamente estabelecidas no *design* e mapeamento do sistema, pois não há resposta sonora do corpo material do Instrumento Digital. A interface é dissociada do núcleo gerador de som, todos os controles são prescritos e, portanto, as possibilidades são finitas. Como exemplo inverso, nos instrumentos acústicos o controle sonoro é associado ao comportamento acústico do material que não é inteiramente previsível e abre campo para a exploração de sons e modos de utilização do instrumento não prescritos em seu *design*.

Em busca de solução para o problema da expressividade gestual alguns autores advogam por um *design* de Instrumentos Digitais que permita o máximo possível de controle sonoro e expressividade musical de acordo as acepções tradicionais do conceito. Normalmente esses autores optam por interfaces fundamentadas em modelos extraídos de instrumentos musicais acústicos e pelo mapeamento direto entre gesto instrumental e resposta sonora (PAINE, 2009). Essa abordagem busca estabelecer condições para o refinamento motor do performer e, por consequência, para o controle preciso das nuances da sonorização do sistema.

Em contraposição, Gurevich e Fyans (2011) colocam que a generalidade, a falta de especialização das interfaces colaboram com o deslocamento do virtuosismo do domínio da habilidade motora para o domínio da habilidade cognitiva. Segundo os

---

<sup>16</sup> Live coding é uma prática contemporânea na qual a performance musical se dá através do ato de programar os códigos para geração sonora-musical pelo computador em tempo real.

autores, os instrumentos digitais invertem a relação do músico virtuoso que passa por anos de treino para adaptar-se à interface de seu instrumento em prol da aquisição de habilidades motoras que lhe permitam um controle refinado da resposta sonora do instrumento. Pois, um músico programador em um curto período de tempo pode adaptar as características do *software* e da interface para que eles atendam aos limites de suas habilidades motoras, sem perder com isso a possibilidade de exploração de características e refinamentos sonoros do instrumento (BOWN et al. , 2009). Segundo esses autores, a relação de virtuosismo no instrumento digital está no modo como o instrumentista lida intelectualmente com as propriedades algorítmicas do programa que se configuram como limites da expressão musical, e como ele explora-as em performance através da interface de sua preferência.

Gurevich e Fyans (2011) propõem o termo *interações musicais digitais* (*Digital Music Interactions* - DMI) para definir todos os tipos de interações entre *performer* e sistemas digitais de performance musical, sob os mais diversos contextos. Essa acepção genérica do conceito, segundo os autores, contrapõe-se ao termo *instrumento digital* que segundo eles é muitas vezes utilizado de forma imprópria. Os autores adotam uma noção de instrumentalidade baseada em Cadoz (2009) que afirma que o critério de importância para definir a interação instrumental é a “relação mecânica através de processos que respeitam uma consistente troca energética” (GUREVICH, FYANS, 2011). As demais interações em DMIs os autores chamam de interações não-instrumentais. Gurevich e Fyans exemplificam classificações de interações não-instrumentais baseados em outros autores:

- Segundo Pressing (1990) que define dois tipos: a) ações que são tradicionalmente consideradas involuntárias ou não envolvem diretamente movimento no espaço; b) ações que envolvem modelar algum processo em desenvolvimento ou seus efeitos, em que modelar pode ser apenas ativar ou desativar, filtrar ou vários outros tipos de parâmetros de controle.
- No trabalho de Wessel e Wright (1992) que discute tipos de controle de parâmetros em processos generativos.
- Segundo Johnston et al. (2008) que, no contexto de interfaces visuais de instrumentos virtuais, detectam dois modos de operação além do

instrumental: a) o modo ornamental, no qual o controle primário do resultado sonoro é determinado pelo sistema e o *performer* age ornamentando o resultado, gerando certa variabilidade; b) modo conversacional, os *performers* encaram o sistema como um agente responsivo, permitindo que ele tenha uma participação na determinação da direção musical.

Outro aspecto interessante apontado por Gurevich e Fyans é relacionado à temporalidade dos sistemas de interação ‘não-instrumental’. Eles notam que em sistemas interativos autônomos e modelados para mimetizar o comportamento humano (e.g. ASSAYAG et al. 2006 ; GIMENES 2008; LEWIS 2000; PACHET 2003; WEINBERG, DRISCOLL 2006), a frequência do comportamento do sistema é parecida com a do agente humano. Neste caso eles assumem que uma baixa frequência da dinâmica do comportamento do sistema é propícia para interações ‘não-instrumentais’ que muitas vezes compartilham de processos generativos parecidos com os sistemas autônomos.

Por manipular processos com alterações lentas, muitas DMIs permitem aos usuários especificar respostas temporalmente distantes de modos que não são possíveis com instrumentos acústicos. Diferença temporal entre ação e resposta é um diferenciador primário das habilidades intelectuais em relação às habilidades motoras. (ROSENBAUM et al. 2001 apud. GUREVICH, FYANS 2011); Os objetivos simbólicos especificados por ações envolvendo habilidades intelectuais são menos imediatos e deficientes na precisão temporal requerida por ações envolvendo habilidades motoras e perceptuais. (GUREVICH, FYANS, 2011, p.169).<sup>17</sup>

Os autores apontam o *live coding* como exemplo deste conceito.

Resumindo, nota-se que existe um contínuo de possibilidades para o controle de instrumentos digitais, que varia entre: a) o modelo baseado em instrumentos acústicos tradicionais nos quais cada gesto é responsável por uma emissão acústica, e

---

<sup>17</sup> By manipulating slowly varying processes, many DMIs allow users to specify temporally remote outcomes in ways that are not possible with acoustic instruments. Temporal remoteness between action and outcome is a primary differentiator of intellectual from perceptual-motor skill ( ROSENBAUM et al. 2001 apud. GUREVICH, FYANS 2011 ) ; the symbolic goals specified by actions involving intellectual skills are less immediate and lacking in the temporal precision required by actions involving perceptual-motor skills. (GUREVICH, FYANS, 2011, p.169)

b) o controle de sistemas generativos em que procura-se modos de interface e gestos apropriados para o controle e gerenciamento de dados globais. Controles baseados nos modelo (a) tendem a reforçar a prática musical tradicional em que o instrumento serve para a realização física de uma idéia musical e tendem a eliminar aspectos de generatividade do sistema em prol do controle absoluto do performer. O modelo (b) privilegia a generatividade dos sistemas o que aproxima-o do conceito de instrumento composto e de uma maior estruturação musical intrínseca ao instrumento. Em nosso trabalho estamos interessados primordialmente no modelo de controle generativo, pois é mais propício a construção dos instrumentos compostos. Com isso, os tipos de interações e relações performer-computador listados acima relacionados a esse tipo abordagem são mais explorados.

Concluimos esse capítulo reforçando que os principais pontos conceituais que fundamentaram a criação dos sistemas musicais relatados no **Capítulo 5** foram o conceito de instrumentos compostos e a busca por um equilíbrio entre as características composicional do sistema — mantendo certas recorrências de estruturas musicais e processos de autonomia, geração e condução musical dando à performance do algoritmo recorrência e unidade suficiente para caracterizar uma obra musical fechada — e entre as características instrumentais dos sistemas — oferecendo parâmetros de controle a um performer humano que pode interferir em maior ou menor grau na autonomia do algoritmo. Em suma, buscamos o equilíbrio entre a propriedade de representação musical e a *instrumentalidade* do código computacional.

Nos capítulos seguintes abordaremos os principais aspectos técnicos relacionados às criações algorítmico-musicais apresentadas no **Capítulo 5**.

# CAPÍTULO 2

## EXTRAÇÃO DE CONTEÚDO MUSICAL DO SINAL DE ÁUDIO

O processo de representação do sinal elétrico para o sinal digital é denominado, na área de *Processamento de Sinais*, de *conversão analógica digital*. Essa transformação é realizada por um circuito eletrônico que converte o sinal analógico contínuo em uma sequência numérica discreta através da amostragem dos valores de amplitude do sinal elétrico. A representação computacional para o áudio gravado e digitalizado é um vetor unidimensional de valores numéricos inteiros e caracterizado por dois parâmetros: a) *taxa de amostragem (TA) dada em hertz* e b) *resolução em bits da amostragem (RB) dada em número de bits*.

A *taxa de amostragem (TA)* corresponde ao número de amostras por segundo. A *resolução em bits (RB)* representa o número de bits associado à amplitude do sinal elétrico de cada amostra. Quando um valor numérico é escrito na forma binária, cada *bit* corresponde a um espaço de memória do computador associado a dois estados: 0 ou 1. Um conjunto de  $n$  *bits* representa em notação binária um valor decimal máximo de  $2^n$ . Por exemplo, 3 *bits* representam o valor máximo  $2 \times 2 \times 2 = 2^3 = 8$ .

No processo de *conversão analógica digital*, esse número é associado à capacidade de amostragem do conversor sobre o sinal elétrico, por exemplo, uma resolução de 16 *bits* é capaz de amostrar 65536 valores diferentes de amplitude. Quanto maior o número de *bits* mais o âmbito de variação do sinal é dividido em pequenas partes. O aumento da resolução em *bits* propicia uma melhora na qualidade do sinal digital, pois através deste procedimento a representação digital é capaz de capturar cada vez mais nuances das variações de amplitude do sinal.

Neste capítulo denominaremos de  $x[n]$ , onde  $n = 0, 1, 2, 3, 4, 5, \dots, N$  o vetor numérico com as características apresentadas nos parágrafos anteriores, sendo  $N$  o número total de amostras de um som digitalizado e  $x[n]$  o valor de cada amostra. No caso de um sinal amostrado com a taxa de 44.1 kHz (44.100 amostras por segundo) serão 44.100 amostras de números inteiros para se obter um segundo de som digitalizado, ou seja,  $N = 44.100$ . Para a conversão da posição da amostra  $n$  para seu respectivo valor temporal no tempo  $t$ , nesta taxa de amostragem, calcula-se  $t = n/44.1$ .

Na maioria dos casos em que o som foi amostrado de um ambiente acústico (i.e. ondas mecânica propagadas pelo ar), este vetor  $x[n]$  está associado às interações de

diversas fontes acústicas (i.e. corpos que emitem ondas sonoras). Daí então nasce o interesse no estudo sobre a representação digital da informação acústica, pois através de procedimentos analíticos existe um potencial de recuperar estas diversas interações entre materiais, taxas de reverberação de uma sala, maneiras de articulação sonora de sons de instrumentos musicais, qualidade sonora e timbre dentre outras tantas possibilidades.

Esses procedimentos analíticos são descritos por um conjunto de funções matemáticas extraídas do vetor  $x[n]$ . O objetivo de tal procedimento é descrever as características do sinal digitalizado e, desta forma, recuperar algum aspecto específico da informação presente no sinal de áudio. Neste capítulo estes algoritmos são denominados de *Descritores Acústicos* e vamos tratar de uma de suas aplicações no campo dos Instrumentos Musicais Digitais de acordo com o que foi apontado na Fig. 2 da **Introdução**.

## 2.1 Descritores Acústicos

No domínio da psicoacústica (FASTL, ZWICKER, 2007) a percepção auditiva humana é entendida como um *fenômeno multidimensional* composto por diferentes características que chamaremos aqui de *dimensões da percepção auditiva*. Alguns exemplos destas dimensões são: a amplitude perceptiva chamada de *loudness*, a sensação de altura, e o timbre, que, ele próprio é entendido atualmente como um fenômeno perceptivo multidimensional, vide (LOUREIRO, DE PAULA, 2006; MCADAMS, 1999).

A pesquisa da área de Recuperação da Informação Musical (MIR - Music Information Retrieval)<sup>18</sup> utiliza a informação unidimensional da gravação sonora digitalizada (aqui representada pelo vetor  $x[n]$ ) para extrair informações associadas às diferentes dimensões da percepção auditiva. A hipótese de estudo é que uma vez identificada a relação entre os padrões do comportamento acústico com componentes perceptivos, é possível mesurar/parametrizar no som gravado e digitalizado características relacionadas com as múltiplas dimensões da percepção auditiva humana.

Um considerável número de métodos de análise de sinais com conteúdo musical baseia-se em estudos de psicoacústica, como os trabalhos de Jehan (2005) e

---

<sup>18</sup> Definições mais aprofundadas, exemplos dos temas e desenvolvimentos da área podem ser encontrados no principal evento científico ligado a de recuperação da informação musical: o ISMIR. Os anais desta conferência podem ser acessados na página da rede mundial: <http://www.ismir.net/>

Collins (2005). Outras abordagens referem-se diretamente ao comportamento físico do som, sem fazer associação direta com a percepção humana, como por exemplo, o estudo apresentado por Bello et al. (2005).

Em ambas as abordagens a primeira tarefa na recuperação de informação musical de sinais de áudio é a extração de medidas objetivas do sinal. Estes procedimentos são denominados como *extração de características (features)* ou *descritores*. O primeiro termo é predominante na literatura relacionada à Engenharia Elétrica e Computacional (BARBEDO, LOPES, 2007), pois descreve as medidas extraídas do sinal. Já o termo *descritores* é encontrado predominantemente na literatura de Música e Tecnologia, principalmente na área de Music Information Retrieval (MIR) (PEETERS, 2004).

Os dois termos se referem à análise do sinal digitalizado e o termo utilizado nesta dissertação será *Descritores Acústicos*. Através dele estamos nos referindo a algoritmos extratores que englobam desde medidas resultantes de cálculos estatísticos sobre o sinal de áudio no domínio do tempo e no domínio da frequência, até medições relacionadas à psicoacústica. Esse primeiro grupo é denominado de *Descritores de Baixo-nível*. Os algoritmos de um segundo grupo denominado de *Descritores de Alto-nível* extraem padrões musicais como ritmo, harmonia, melodia, entre outros, e para isso utilizam-se das medidas fornecidas pelo primeiro grupo. No **Anexo A** desta dissertação há uma compilação de vários destes algoritmos para extração de características do sinal de áudio. Os descritores que se encontram no **Anexo A** foram organizados com base na taxonomia apresentada no trabalho de Geoffroy Peeters (PEETERS, 2004). Este conjunto de descritores compõe uma biblioteca denominada de PDescriptors<sup>19</sup> que foi desenvolvida durante a pesquisa e foram construídas como abstrações programadas na linguagem Pure Data.

Nas próximas subseções o texto concentra-se em apontar os procedimentos específicos adotados durante a pesquisa no que concerne à aplicação de Descritores Acústicos para o sistema de análise e recuperação da informação musical implementado em uma das composições algorítmicas que são apresentadas no **Capítulo 5** e conforme o esquema mostrado na Fig. 2 da **Introdução** deste trabalho. Tendo em vista a grande variedade de descritores (vide **Anexo A**) apresentamos a seguir apenas os

---

<sup>19</sup> A biblioteca PDescriptors está contida no CD Apêndice A desta Dissertação e também pode ser obtida através da página na internet <https://sites.google.com/site/pdescriptors/>

procedimentos algorítmicos essenciais para o entendimento do processo desenvolvido durante a pesquisa e que foram divididos nas seguintes subseções: “pré-processamento” onde descrevemos uma etapa de processamento de áudio anterior à extração de medidas, porém necessária para dois dos descritores adotados. E nas três últimas subseções apresentamos os descritores denominados de *Fluxo Espectral*, *Chroma* e *Média Quadrática (RMS)*.

### 2.1.1 Pré-Processamento

De acordo com a taxonomia que adotamos, utilizamos descritores de dois tipos para os procedimentos de recuperação da informação musical: a) *Descritores Temporais Instantâneos* e b) *Descritores Espectrais*. A primeira classe de descritores é computada a partir do sinal de áudio digitalizado no domínio do tempo e em janelas<sup>20</sup> de tamanho invariável e que são deslocadas ao longo de toda a sequência de amostras  $x[n]$  de som gravado. Para cada janela o descritor retorna um único valor numérico correspondente à medida extraída. Este tipo de descritor não requer qualquer pré-processamento do sinal a não ser a divisão do sinal em janelas sucessivas.

Os algoritmos do segundo tipo, denominados *Descritores Espectrais* são computados no domínio da frequência, ou seja, após o sinal de áudio digitalizado passar por uma Transformada Discreta de Fourier (DFT - *Discrete Fourier Transform*) (OPENHEIM, SCHAFER, 2010). Tipicamente em procedimentos de análise aplica-se a Transformada de Fourier em janelas sucessivas de tamanho invariável. Este procedimento, chamado de *Short-Time Fourier Transform* (STFT) é bem documentado na literatura da área (KLAPURI, DAVY, 2006; OPPENHEIM, SCHAFER, 2010; ROADS 1996, 2001)

A STFT é definida na Eq. 2.1.

$$X_i[k] = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x[hn + m]w[m]e^{-\frac{2j\pi mk}{N}} \quad (2.1)$$

---

<sup>20</sup> Nesse contexto o termo *janela* significa um segmento do sinal de áudio que contém um número específico de amostras.

Onde  $X_i[k]$  representa a  $k^{\text{ésima}}$  componente espectral (*bin*) do  $n^{\text{ésimo}}$  quadro,  $w[m]$  é uma janela de  $N$  amostras, e  $h$  é o número de amostras de interpolação entre quadros sucessivos (*hop size*). Estas  $N$  amostras de cada quadro de uma STFT são números complexos tendo a parte real  $X_r$  e a parte imaginária  $X_{im}$ . A magnitude  $|X_i[k]|$  e a fase  $\varphi_i[k]$  do espectro são computadas da seguinte forma:

$$|X_i[k]| = \sqrt{X_r[k]^2 + X_{im}[k]^2} \quad (2.2)$$

$$\varphi_i[k] = \arctan \frac{X_{im}[k]}{X_r[k]} \quad (2.3)$$

Para implementação dos algoritmos em PDescriptors realizamos as Transformadas de Fourier através do objeto (algoritmo) interno da linguagem *Pure Data*: **FFT~**, que opera uma versão otimizada da STFT chamada Transformada Rápida de Fourier (*Fast Fourier Transform - FFT*) (OPPENHEIM, SCHAFER, 2010)

### 2.1.2 Fluxo Espectral

O descritor Fluxo Espectral mede a diferença das magnitudes do espectro entre duas janelas sucessivas de análise. Este descritor retorna valores baixos quando a variação do espectro é quase constante, ou seja, há pouca variação da energia entre um quadro e outro. Em contrapartida, a medida do Fluxo Espectral terá valores altos quando há grande variação de uma janela para a outra. Algumas das implementações desse descritor encontradas na literatura foram implementadas na biblioteca PDescriptors no **Anexo A** desta dissertação. Duas delas, utilizadas no sistema computacional que desenvolvemos durante a pesquisa, são apresentadas a seguir.

A primeira e mais simples medida do Fluxo Espectral é definida pela somatória da diferença quadrática entre as magnitudes de duas janelas consecutivas e é apresentada na Eq. 2.4.

$$SF_i = \sqrt{\sum_{k=1}^{K-1} \{|X_i[k]| - |X_{i-1}[k]|\}^2} \quad (2.4)$$

Onde  $|X_i[k]|$  é a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela com  $N$  amostras e  $K = N / 2$ .

A segunda medida, chamada de *Fluxo Espectral de Diferença*, é obtida combinando-se as duas outras medidas de fluxo espectral apresentadas nas equações Eq. 2.5 e Eq. 2.6. A implementação deste descritor é referente à documentação sobre algoritmos para detecção de *onset* encontrada na página do *Mazurka Project*<sup>21</sup>

Na primeira etapa realiza-se o cálculo do *Fluxo Espectral Positivo* (BELLO et al., 2005; DIXON, 2006). Esta medida é obtida pelo acréscimo de uma função de retificação de meia onda positiva na Eq. 2.4 definida acima. Ao anular os valores negativos resultantes da diferença entre as magnitudes de duas janelas consecutivas, esta medida computa apenas valores positivos os quais representam somente o aumento de energia.

Esta medida é principalmente útil na detecção do início de eventos musicais produzidos por instrumentos acústicos (i.e. notas e acordes), pois nas regiões dos picos do Fluxo Espectral detecta-se o início (*onset*) e o fim (*offset*) de cada evento. Em suma, a Eq. 2.5 descreve uma medida que indica os inícios de eventos onde há rápido aumento da energia.

$$SF_i^+ = \sqrt{\sum_{k=1}^{K-1} \{H^+(|X_i[k]| - |X_{i-1}[k]|)\}^2} \quad (2.5)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela com  $N$  amostras,  $K = N / 2$  e  $H^+(x)$  é a função de retificação de meia onda positiva definida na Eq. 2.6.

$$H^+(x) = \frac{x + |x|}{2} \quad (2.6)$$

A próxima etapa segue o mesmo raciocínio do anterior, mas com propósito inverso. Chamado de *Fluxo Espectral Negativo*, este algoritmo inclui uma função de retificação de meia onda negativa como descrita na Equação 2.7.

$$SF_i^- = \sqrt{\sum_{k=1}^{K-1} \{H^-(|X_i[k]| - |X_{i-1}[k]|)\}^2} \quad (2.7)$$

---

<sup>21</sup> Mazurka Project: <http://www.mazurka.org.uk/software/sv/plugin/MzSpectralFlux/> (consultado em 03/01/2012)

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela com N amostras,  $K = N / 2$  e  $H(x)$  é a função de retificação de meia onda negativa definida na Eq. 2.8.

$$H^-(x) = \frac{x - |x|}{2} \quad (2.8)$$

A medida do Fluxo Espectral Negativo não tem uma aplicação direta para a detecção de início de eventos, mas é subtraída do valor do Fluxo Espectral Positivo para finalmente obter-se a medida do Fluxo Espectral de Diferença, apresentada na Eq. 2.9.

$$SF_i^{\Delta} = H^+ (SF_i^+ - SF_i^-) \quad (2.9)$$

Onde  $SF^+$  representa o fluxo espectral positivo (Eq. 2.5),  $SF^-$  representa o fluxo espectral negativo (Eq. 2.7) e  $H^+(x)$  a função de retificação de onda positiva (Eq. 2.6)

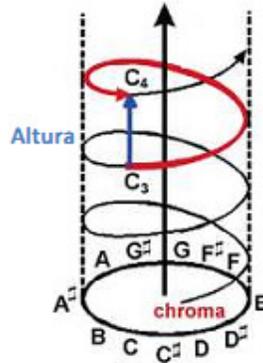
### 2.1.3 Chroma

O descritor denominado de *Chroma* também é encontrado na literatura com o nome de *Perfil de Classe de Notas*<sup>22</sup> (*Pitch Class Profile* - PCP) como pode ser visto no trabalho de Fujishima (1999). Como o próprio termo sugere, este descritor resulta em um vetor numérico que cada elemento indica um fator de presença de cada classe de nota em um trecho de áudio analisado. Essa análise é realizada através da comparação da energia espectral (Magnitude, vide Eq. 2.2) concentrada em diferentes subdivisões do espectro sonoro que são relativas às classes de notas. Desta forma, o Chroma pertence a uma classe de medidas espectrais que são construídas sobre conhecimento prévio de aspectos da estruturação musical. Sua medida baseia-se no fenômeno de equivalência de oitavas da percepção humana que é descrita por dois valores: 1) o valor do *Chroma* que corresponde à relação de equivalência da percepção de uma mesma nota em oitavas diferentes; 2) a *Altura* (*Height*) que é relativo a percepção de diferença entre duas ou mais regiões espectrais e que corresponde a diferenciação perceptiva entre grave e agudo, ou por exemplo, a diferença entre duas ou mais oitavas para uma mesma classe

---

<sup>22</sup> Em português, o termo altura pode ser equivalente ao termo nota. Contudo, em nosso trabalho designamos o termo nota como equivalente ao termo *pitch* em inglês, que corresponde às notas musicais: Dó, Dó sustenido, Ré, etc; e o termo altura para designar a percepção de registro, oitava, ou relação grave e agudo.

de notas. Este modelo perceptivo é ilustrado Fig. 3 por uma espiral onde a representação do parâmetro altura está associada ao eixo vertical, e a rotação da hélice sob a divisão de Escala Cromática representa a medida do Chroma.



**Fig. 3** - Representação em hélice das dimensões da percepção de notas musicais: Altura (sentido vertical) e Chroma (rotação da hélice). (FONTE: ONG, 2006)

A medida do Chroma é associada a um vetor de 12 valores, denotado a seguir por *Chroma* [n] para  $n = 0, 1, \dots, 11$ , que corresponde à energia do sinal para cada nota da escala cromática. A subdivisão em 12 valores normalmente é utilizada em aplicações que envolvem reconhecimento de estruturas harmônicas como acordes, *pitch-class* e tonalidade. Para se obter os valores de *Chroma* [n], o espectro é primeiramente dividido em bandas com larguras correspondentes a um semitom e com frequência central igual ao valor das notas da escala cromática, em seguida, para cada nota são somados os valores de energia das respectivas bandas *presentes* em todas as oitavas<sup>23</sup>. A Eq. 2.10 descreve o cálculo para obtenção do vetor Chroma para cada janela de análise segundo a descrição de Sheh e Ellis (2003).

$$Chroma_p = \sum_{K:p(k)=p} |X_i[k]|^2 \quad 2.10$$

Onde ,  $p = 0, 1, \dots, 11$ ;

---

<sup>23</sup> A percepção de notas é de ordem logarítmica, ou seja, em termos frequenciais absolutos, um intervalo de semitom é maior que seu antecessor descendente. Desse modo, em um espectro dividido em semitons, para uma nota (*pitch-class*) a largura das bandas e intervalos entre as frequências centro aumentam conforme se aumenta o parâmetro de altura, i.e oitava.

$$e \quad p[k] = \left[ 12 \log_2 \left( \frac{k f_{sr}}{N f_{ref}} \right) \right] \text{mod} 12 \quad 2.11$$

Onde  $f_{sr}$  é a frequência de amostragem e  $f_{ref}$  é a frequência relativa ao *Chroma*[0], ou seja, a frequência da primeira nota do vetor.

É possível e usual realizar o cálculo do Chroma com divisões menores que semitons, como por exemplo em quartos-de-tom ( $p = 0, \dots, 24$ ) ou sextos-de-tom ( $p = 0, \dots, 36$ ), (BELLO, PICKENS 2005; SHEH, ELLIS, 2003)

#### 2.1.4 Média Quadrática (RMS)

Este descritor calcula a média quadrática da variação da intensidade do sinal amostrado. (RMS é a sigla para *Root Mean Square* que corresponde ao termo ‘média quadrática’ em português). Este é um descritor calculado com os valores amostrados no domínio do tempo e em janelas de mesmo número de amostras e consecutivas sobre o sinal  $x[n]$ . Os valores de RMS extraídos de um evento sonoro descrevem o contorno de seu envelope de energia. O cálculo do RMS é definido na Eq. 2.12.

$$RMS_i = \sqrt{\frac{\sum_{n=0}^{N-1} x_i[n]^2}{N}} \quad (2.12)$$

Onde  $x_i[n]$  são os valores de amplitude na  $n^{\text{ésima}}$  amostra na  $i^{\text{ésima}}$  janela do áudio digitalizado, e  $N$  é o número de amostras das janelas.

Os três descritores definidos acima formam a base do sistema de análise implementado durante a pesquisa. O *Fluxo Espectral*, *Chroma* e *Média Quadrática (RMS)* foram utilizados em um dos algoritmos musicais apresentado no **Capítulo 5** fazendo parte de um sistema de recuperação de informação musical aplicado ao áudio capturado da performance. No próximo capítulo expomos como os algoritmos descritores são utilizados nas implementações dos métodos de extração de parâmetros musicais em tempo real via computador. Estes métodos incluem algoritmos para segmentação de eventos musicais do sinal de áudio baseados na detecção dos inícios dos eventos (onset) através da extração do Fluxo Espectral e na detecção dos finais de eventos (*offsets*) através da extração do RMS, e um algoritmo de classificação harmônica que emprega o descritor Chroma.



# CAPITULO 3

## METODOLOGIA DE ANÁLISE

Este capítulo apresenta a implementação de ferramentas de análise em tempo real para o áudio produzido por instrumentos acústicos que se utilizam dos três descritores introduzidos no **Capítulo 2**. Em nossa abordagem, os procedimentos de análise do áudio digitalizado abrem um canal de comunicação entre a emissão sonora dos instrumentos acústicos e o algoritmo para performance musical programado no computador. Durante a performance do sistema o performer computacional atua como mediador entre os algoritmos de análise para recuperação de dados musicais e os algoritmos para síntese sonora.

Sob esse ponto de vista, apresentamos nos próximos parágrafos a metodologia que adotamos para extrair a informação musical do áudio e construir este canal de interação entre o sistema computacional e os músicos atuantes na performance. Apresentamos procedimentos para extração de parâmetros musicais relacionados com a segmentação e com a medida de duração temporal de eventos musicais, bem como para extração de conteúdo harmônico de cada evento. Tais procedimentos foram embasados na literatura da área denominada de Music Information Retrieval (MIR), (BROSSIER, 2006; JEHAN 2005; KLAPURI, DAVY, 2006). Essas análises inserem-se num caso mais geral denominado na literatura de “*Transcrição Automática de Sinais Musicais*”, e trata-se de utilizar descritores de baixo-nível como apresentado no **Capítulo 2** para produzir medidas de alto-nível relacionadas com elementos estruturantes do discurso musical.

Expomos também nesse Capítulo a aplicação de *Mapas de Poincaré* para a análise de dinâmica espectral, apresentada durante a pesquisa no artigo (MONTEIRO, MANZOLLI, 2011c), aplicada ao som produzido por instrumentistas. Esta metodologia resulta da aplicação de experimentos que realizamos para visualização de dinâmicas de características espectrais vi Mapas de Poincaré (MONTEIRO, MANZOLLI, 2011a) e, portanto, não foi extraída da literatura sobre Transcrição Automática de Sinais Musicais como os demais procedimentos.

Segundo Scheirer (1995), o objetivo de um processo de transcrição musical é extrair do sinal de áudio informações simbólicas similares as estruturas musicais que podem ser vistas em uma partitura. Considera-se que uma transcrição completa implica

em determinar o conteúdo de alturas, temporalidade, e instrumentação de todos os eventos relacionados a música notada ocidental (KLAPURI, DAVY, 2006). A pesquisa em transcrição desenvolvida por Scheirer e Klapuri é contextualizada à tradição musical ocidental em que a tipologia sonora predominante é o som produzido por instrumentos acústicos que, majoritariamente, apresentam um envelope dinâmico de intensidade que pode ser descrito pelo perfil ADSR<sup>24</sup> e conteúdo espectral com características quasi-harmônicas (em instrumentos com altura definida), ou ruidosas (maioria dos instrumentos de percussão sem altura definida). Em síntese, a detecção e avaliação de um evento sonoro-musical produzido por um instrumento acústico e relacionado à notação musical tradicional do ocidente está diretamente relacionada a quatro parâmetros: *altura, duração, dinâmica e timbre*.

Desta forma, a primeira tarefa no processo de transcrição musical automática via modelo computacional é a segmentação do sinal de áudio. Nela determinam-se o início e a duração de cada evento. Em seguida, aplicam-se processos para extrair de cada evento medidas que detectem os três parâmetros principais: alturas, dinâmica e timbre. Finalmente, a tarefa de mais alto nível simbólico é localizar relações entre os eventos a partir de um ponto de vista de organização musical pré-estabelecida, como por exemplo, *perfis melódicos, padrões rítmicos, padrões harmônicos e texturas*. Alguns exemplos de trabalhos relacionados à transcrição automática de sinais são: (BELLO, 2003; BELLO, et al. 2000; BROSSIER, 2006; JEHAN, 2005; KLAPURI, DAVY, 2006; MÜLLER, 2007; MÜLLER et al. 2011). Outros estudos focam apenas procedimentos específicos envolvidos na tarefa de transcrição automática, como a detecção de início de evento musical (*onset*) e extração de padrões rítmicos (BELLO et al. 2005; COLLINS, 2005; DIXON, 2001, 2006; GOUYON 2005); extração de conteúdo harmônico (GÓMEZ 2006; LEE, SLANEY, 2008; PEETERS, 2006a; STARK, PLUMBLEY, 2009); classificação de timbre de instrumentos musicais (ERONEN, KLAPURI 2000; HERRERA et al. 2003; HERRERA-BOYER et al. In: KLAPURI, DAVY 2006; PARK, 2004; PEETERS 2003; PEETERS, RODET, 2003); extração de frequência(s) fundamental(ais) (CHEVEIGNÉ, KAWAHARA, 2002; KLAPURI, 2004; PEETERS, 2006b; YEH, 2008); análise e segmentação de estruturas musicais (ONG, 2006; PAULUS et al. 2010; PIRES, 2011; WU e BELLO 2010)

---

<sup>24</sup> Sobre o Modelo ADSR ver a definição no ANEXO A

Vê-se claramente que a gama de possibilidades é muito grande e a maioria das pesquisas concentra esforços para extrair parâmetros relacionados à estruturação do repertório tradicional da música erudita de concerto (música com estrutura métrica regular e harmonia tonal, característica da prática europeia dos séculos XVIII e XIX) assim como na maioria dos gêneros de música popular. Este também é o principal enfoque da área de *Music Information Retrieval* (MIR), pois há grande interesse no potencial comercial-econômico que esta tecnologia representa, principalmente para os setores ligados a rede mundial de computadores e mídia eletrônica. Um exemplo de pesquisa nesse campo é o desenvolvimento de sistemas dedicados à catalogação automática e à busca de arquivos musicais em base de dados, como pode ser visto na de pesquisa de (BARBEDO, LOPES 2007; PEREIRA, 2009).

Há também pesquisas que focam o desenvolvimento de ferramentas para descrição/identificação de eventos sonoros incomuns ao repertório descrito no parágrafo anterior, mas especificamente importantes à música de concerto contemporânea. Alguns exemplos podem ser vistos em aplicações para análise e re-síntese sonora (CAETANO, RODET, 2010; JEHAN, 2001; MASRI, 1996; SCHWARZ, 2004); detecção de gesto instrumental (ROSAMIMANANA et al. , 2009); classificação de morfologias sonoras, (BLOIT et al. 2010; JANER et al, 2009; PEETERS, DERUTY 2008); detecção de eventos relacionados a técnicas estendidas instrumentais como em (MALT, JOURDAN 2009).

A aplicação de sistemas de transcrição musical é muito diversificada e de acordo com Brossier (2006 p. 28-29) e Kapluri E Davy (2006, p. 5) alguns exemplos são: a) recuperação de informação musical; b) processamento musical — parâmetros musicais e de efeitos de áudio controlados adaptativamente; c) equipamento relacionado com música — por exemplo, controle de áudio ou vídeo em tempo real. d) análises musicológicas; e) ferramentas de transcrição para músicos amadores; f) interação entre homem e máquina — como sistemas computacionais para co-improvisação, acompanhamento, *score-following*, composição generativa.

A nossa pesquisa concentrou-se na última aplicação apontada no parágrafo anterior. Ou seja, utilizamos procedimentos da área de transcrição musical automática como método de captura de parâmetros musicais de uma performance em tempo real. No sistema desenvolvido, os dados extraídos durante a performance são correlacionados aos parâmetros do algoritmo gerador de som (vide **Capítulo 4**) e armazenados em memória computacional. O objetivo é prover ao *performer* computacional a

possibilidade de recuperar e analisar os parâmetros da performance utilizando-se de uma representação visual associada às características sonoras de instantes passados.

Nas subseções 3.1 e 3.2 deste capítulo, ao apresentarmos a metodologia que adotamos, por questão de simplicidade e de embasamento bibliográfico mais consolidado, implementamos o nosso modelo segundo a concepção de evento musical como discutido por Klapuri e Davy (2006 p.5) e predominantes nas pesquisas da área de Music Information Retrieval (MIR).

Esta metodologia divide-se em dois procedimentos principais: 1) segmentação do sinal de áudio de acordo com os trabalhos de Brossier (2006) e Dixon (2006); 2) classificação do conteúdo harmônico dos segmentos através da extração de vetores de Chroma baseado nos trabalhos de Jehan (2005) e Gómez (2006). As implementações destes métodos foram incluídas na biblioteca PDescriptors que foi apresentada nos anais do primeiro Encontro Internacional de Arte Sonora (MONTEIRO, MANZOLLI, 2010)

Na subseção 3.3 apresentamos o procedimento de análise via Mapas de Poincaré que aplicamos no sistema como um método de visualização das características espectrais de um segmento de áudio analisado.

### 3.1 Segmentação

A segmentação do sinal musical consiste em delimitar o início (*onset*) e final (*offset*) de cada evento musical. Este procedimento é fundamental para descrever as características rítmico-temporais de um sinal musical. Moelants e Rampazzo (1997, apud Brossier, 2006, p. 34) definem o início perceptual de um evento sonoro como o “*início apreendido de um evento discreto, determinado por um aumento pronunciado da intensidade ou por uma mudança súbita na altura ou no timbre*”<sup>25</sup>.

O método que implementamos nesse trabalho parte deste preceito e é dividido em duas partes: a) a construção de uma função de detecção que caracteriza momentos de alteração do sinal; b) a captura dos picos da função calculada em (a). A metodologia para detecção de *fnais* de eventos é composta de outra função de detecção que é submetida ao processamento de um limiar fixo para determinar o momento onde cada evento termina.

---

<sup>25</sup> “perceived beginning of a discrete event, determined by a noticeable increase in intensity, or by a sudden change in pitch or in timbre”.

### 3.1.1 Função para Detecção de Início (Onset)

Uma função ideal para detectar o início de eventos musicais (*onset*) deve apresentar um nível alto de redução da informação do sinal de áudio enquanto preserva a informação temporal necessária para determinar os inícios dos eventos musicais (Dixon, 2006, p.133). Ou seja, o procedimento adotado é alinhar os picos dessa função com o início de cada evento como descrito por Dixon (2006). Optamos nesse trabalho pela aplicação de um Descritor de conteúdo espectral (vide **Capítulo 2**), nossa escolha foi fundamentada nos resultados reportados nos anais do MIREX<sup>26</sup>. Destacamos que pela simplicidade de implementação e eficiência, este método permite a sua aplicação em tempo real sem a necessidade de processamento computacional anterior.

Escolhemos o descritor *Fluxo Espectral de Diferença* (Eq. 2.26), pois segundo Dixon (2006), o *Fluxo Espectral Positivo* possui uma performance equilibrada com bom desempenho, maior precisão temporal e baixo custo computacional. Nos testes desenvolvidos durante nossa pesquisa observamos que o *Fluxo Espectral de Diferença* otimiza os resultados, pois elimina picos indesejados e aumenta a diferença entre os picos que correspondem mais precisamente ao início dos eventos. Os segmentos da função com valores baixos estão associados mais ao corpo de sustentação do evento sonoro ou estado permanente do som.

A Fig. 4 exemplifica as medidas que realizamos na pesquisa para verificar a eficiência do método implementado: cada conjunto de quatro gráficos sucessivos no sentido vertical representa a forma de onda gravada de um instrumento musical tocando um fragmento melódico seguido das respectivas três funções extraídas pelas três variações do algoritmo de fluxo espectral apresentadas no **Capítulo 2** (Fluxo espectral, Fluxo espectral Positivo, Fluxo Espectral de Diferença).

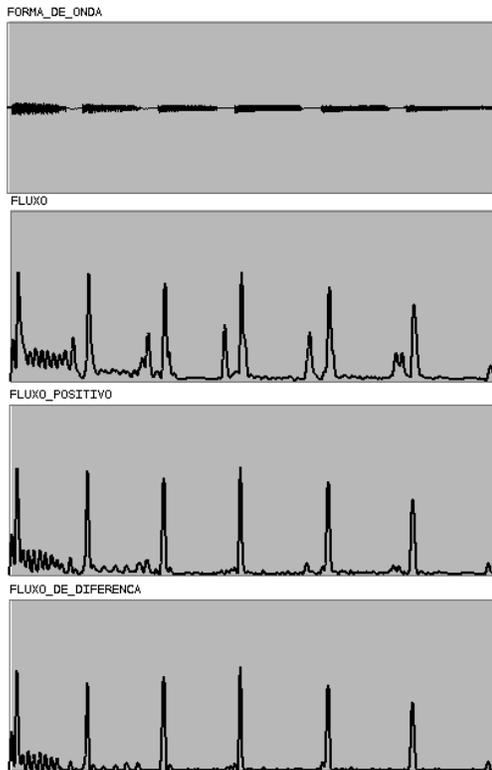
O quadrante superior esquerdo apresenta funções extraídas do sinal de um contrabaixo-baixo elétrico tocando um fragmento melódico em articulação *portato* (para ouvir: Faixa 1 do CD Apêndice A); o quadrante superior direito apresenta funções extraídas do sinal de um saxofone tenor tocando um fragmento melódico em articulação *staccato* (para ouvir: Faixa 2 - CD Apêndice A); o quadrante inferior esquerdo

---

<sup>26</sup> Uma importante fonte de referências para metodologias e de extração de parâmetros musicais é o MIREX ([http://www.music-ir.org/mirex/wiki/MIREX\\_HOME](http://www.music-ir.org/mirex/wiki/MIREX_HOME)), uma competição ligada a MIR onde são comparadas as performances de diferentes algoritmos para classificação e descrição de parâmetros musicais.

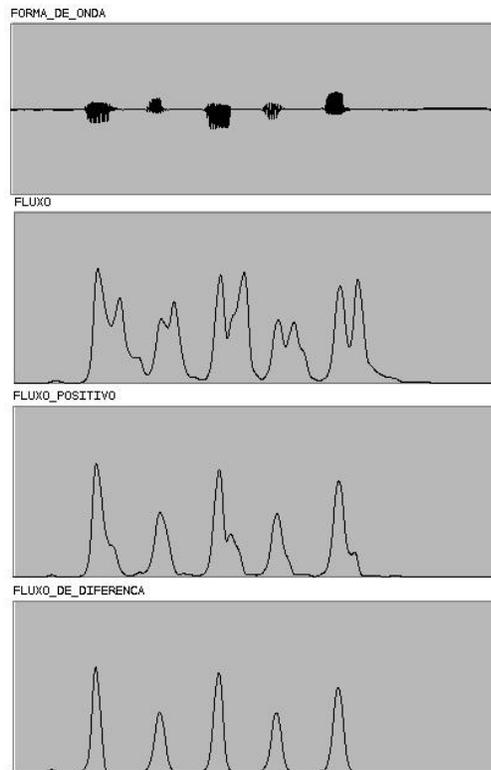
apresenta funções extraídas do sinal de um saxofone soprano tocando um fragmento melódico em articulação *legato* (para ouvir: Faixa 3 - CD Apêndice A); o quadrante inferior direito apresenta funções extraídas do sinal de um trombone tenor tocando um fragmento melódico em articulação *staccato* seguida de articulação *legato* (para ouvir: Faixa 4 - CD Apêndice A).

### CONTRABAIXO - Portato

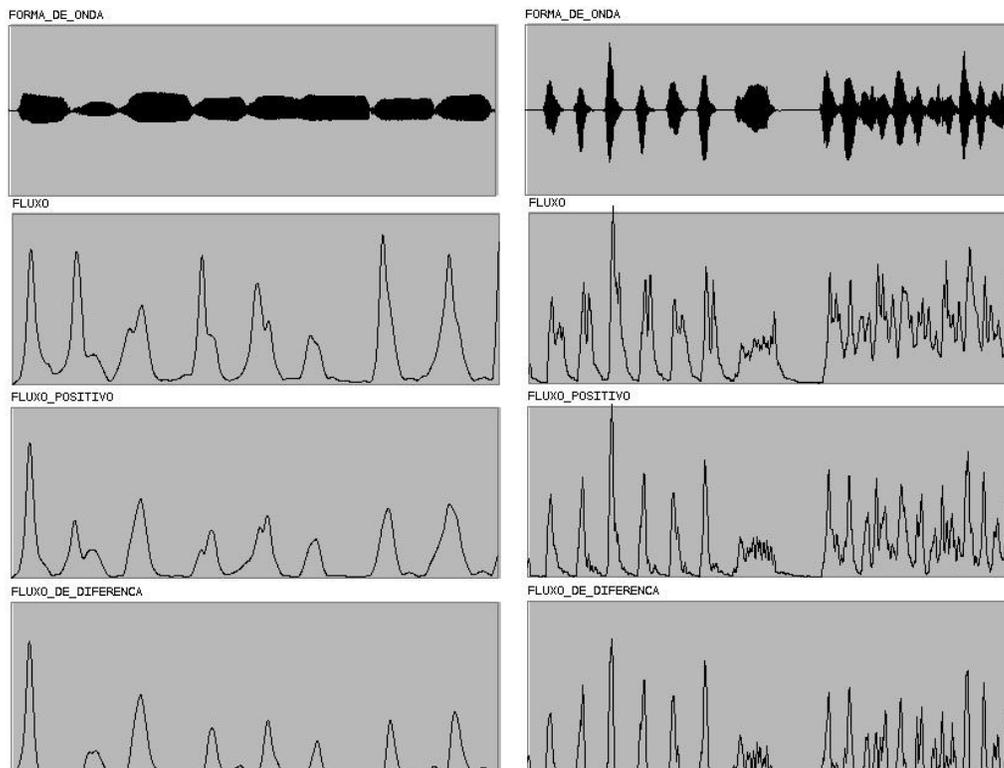


### SAXOFONE SOPRANO - Legato

### SAXOFONE TENOR - Staccato



### TROMBONE - Staccato / Legato



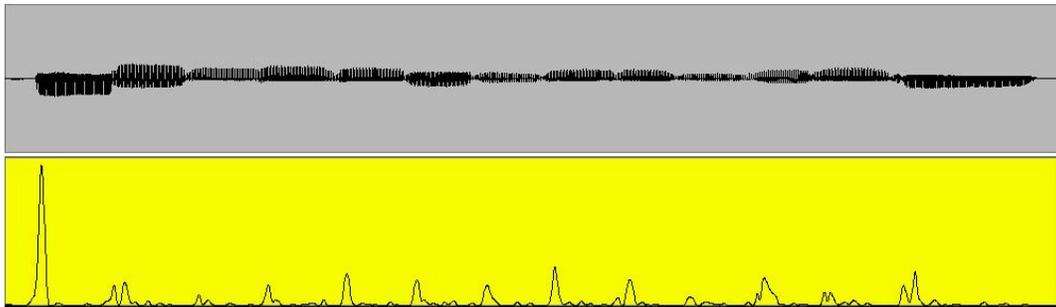
**Fig. 4** - Gravações de fragmentos melódicos com as funções extraídas, sendo respectivamente, de cima para baixo em cada quadrante: fluxo espectral, fluxo espectral positivo, fluxo espectral de diferença.

Nota-se nos segundos gráficos de cada quadrante que a extração do fluxo espectral apresenta picos sucessivos e relativamente próximos que indicam a alteração da energia do espectro sonoro tanto nos finais dos eventos quanto nos inícios. Nos terceiros gráficos (funções extraídas pelo descritor Fluxo Espectral Positivo) os picos relacionados aos finais de eventos são eliminados ou atenuados pela retificação de meia onda (vide no **Capítulo 2** a definição da Eq. 2.6). Por fim, nos últimos gráficos de cada quadrante o Fluxo Espectral de Diferença (Eq. 2.9, **Capítulo 2**) resulta em funções com contornos mais suaves e maior proeminência dos picos em relação ao restante dos pontos do gráfico da função, sendo, portanto, mais refinado para descrição dos inícios de eventos sonoros.

### 3.1.2 Identificação de Picos

Para se identificar os picos é desejável que se opere sobre um perfil normalizado da função de detecção, objetivando com isso que pequenas variações de dinâmica não prejudiquem este processo. Dessa forma é necessário determinar parâmetros fixos para definir um limiar de detecção dos picos. A obtenção deste perfil normalizado requer

pós-processamento sobre os valores da função de detecção, que aplicados em tempo diferido são tipicamente: a filtragem de altas frequências; a remoção do *DC-offset*; e a normalização do sinal. Todavia, não é possível realizar as duas últimas etapas em tempo real, pois devem ser aplicadas sobre toda a duração do sinal analisado. Para solucionar este impasse utilizamos o trabalho de Brossier (2006) que apresenta uma solução otimizada para o pós-processamento em tempo-real, dividindo-o nas etapas: I) filtro passa-baixa, II) limiar (*threshold*) dinâmico; III) seleção do valor máximo local. A Fig. 5 apresenta no gráfico superior a forma de onda gravada de um saxofone tenor tocando um fragmento melódico em articulação legato e no gráfico inferior uma função de detecção extraída desse fragmento pela aplicação do descritor Fluxo Espectral de Diferença. Esses dados serão utilizados a seguir como bases das ilustrações para os procedimentos de pós-processamento da função de detecção. O Fragmento de áudio contendo este exemplo pode ser escutado na Faixa 5 do CD, Apêndice A deste trabalho.



**Fig. 5** - O gráfico inferior apresenta a função de detecção obtida através do descritor ‘fluxo espectral de diferença’ que foi extraída da forma de onda no gráfico superior, que, por sua vez, contém a gravação de um fragmento melódico tocado por um saxofone tenor em articulação legato.

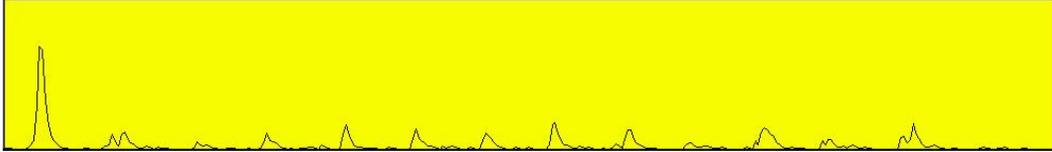
I) A implementação do filtro passa-baixa visa suavizar o contorno da curva da função de detecção e prevenir que haja identificações de falsos positivos, ou seja, picos que não indiquem os inícios de notas. Contudo, filtragens com frequências de corte baixas adicionam atraso na detecção do pico, por isso seu uso deve ser ponderado em vista do perfil da função de detecção. Caso a função já apresente um perfil geral suave o filtro pode não ser necessário. Todavia, se o perfil da função é muito irregular o filtro pode colaborar para prevenção de detecções de falsos positivos. Implementamos um filtro passa-baixas IIR (Infite Impulse Responde) que é definido na Eq. 3.1:

$$Lop_i = (SF_i^+ \cdot \alpha) + (Lop_{i-1} \cdot (1 - \alpha)) \quad (3.1)$$

Onde 
$$\alpha = \frac{f_c \cdot 2\pi \cdot \text{hopsize}}{\text{taxa de amostragem}} \quad (3.2)$$

e  $SF_i^+$  é o valor resultante do *Fluxo Espectral Positivo* para análise da janela  $i$ ,  $f_c$  é a frequência de corte em Hertz, e *hopsize* o número de amostras entre cada janela de análise espectral.

A Fig. 6 apresenta a função da Fig. 5 após a filtragem pelo filtro passa baixa. A frequência de corte  $f_c = 7Hz$ .

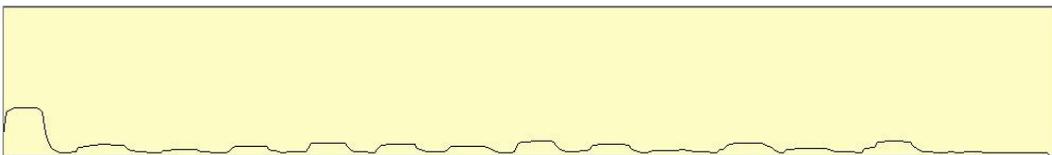


**Fig. 6** - Função de detecção apresentada na Fig. 5 após o processamento do filtro passa-baixas.

**II) Em seguida, a função de limiar dinâmico** é construída através da aplicação da mediana móvel adicionada à média móvel em pequenas janelas da função  $Lop_i$  (Eq. 3.1) computadas sobre algumas amostras anteriores e posteriores à amostra atual. O cálculo da mediana suaviza o perfil da função, reduz os picos indesejados e ressalta os picos mais proeminentes. A média substitui a remoção de DC e da normalização presentes na versão em tempo diferido do algoritmo, compensando a variação de amplitude e suavizando o perfil da função. O limiar dinâmico é definido formalmente como:

$$D_i = \gamma \cdot \text{mediana}(Lop_{i-a}, \dots, Lop_i, \dots, Lop_{i+b}) + \beta \cdot \text{média}(Lop_{i-a}, \dots, Lop_i, \dots, Lop_{i+b}) + \delta \quad (3.3)$$

Onde,  $\gamma$  e  $\beta$  são fatores de ponderação da mediana e da média respectivamente,  $\delta$  é um fator de ajuste fino da função de limite,  $a$  e  $b$  são números inteiros positivos correspondentes à meia janela anterior a amostra atual  $i$ , e meia janela posterior a amostra atual  $i$ , respectivamente. O exemplo na Fig. 7 mostra a função limiar  $D_i$  extraída da função  $Lop_i$  correspondente a Fig. 6, com  $\gamma = 0.3$ ,  $\beta = 0.3$ ,  $\delta = 0.35$ ,  $a = 8$  e  $b = 1$ .



**Fig. 7** - Função limiar extraída pelo algoritmo de limiarização dinâmica (equação 3.3) da função apresentada na Fig. 6.

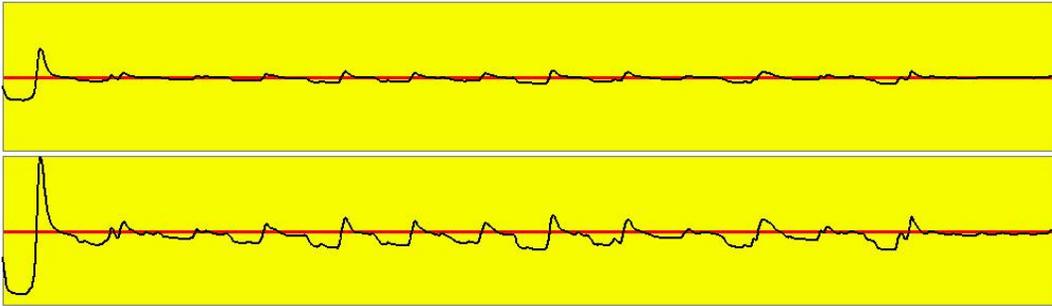
III) Finalmente, a função de limite é subtraída da função  $Lop_i$  (Eq. 3.4) e calcula-se o máximo local (Eq. 3.5) para identificar o pico.

$$\check{D}_i = Lop_i - D_i \quad (3.4)$$

$$Onset_i = \begin{cases} \check{D}_i > 0 \text{ e } \check{D}_i = \max(\check{D}_{i-c}, \dots, \check{D}_i, \dots, \check{D}_{i+c}) & 1 \\ else & 0 \end{cases} \quad (3.5)$$

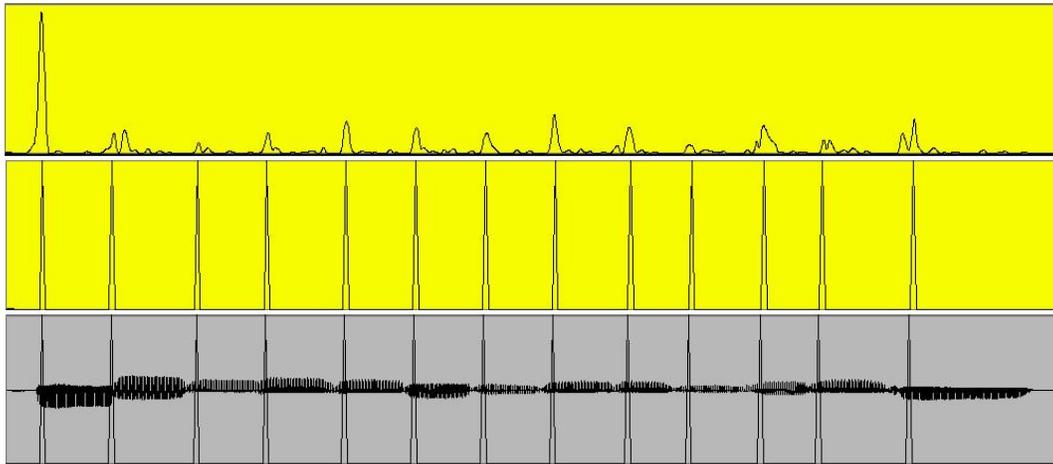
Onde  $c$  é um número inteiro positivo correspondente ao valor de meia janela anterior e posterior em relação à amostra  $i$ . Finalmente, adicionamos um algoritmo que filtra uma detecção de início de evento caso tenha já havido uma detecção até algumas amostras anteriores, sendo o número máximo de amostras parâmetro especificado pelo o usuário.

A Fig. 8 apresenta a Função  $\check{D}_i$  do exemplo mostrado nas etapas anteriores. O gráfico superior mostra a subtração entre as funções na mesma resolução e o gráfico inferior mostra a mesma função normalizada para que se possa ver os detalhes.



**Fig. 8** - Função resultante da subtração entre função de detecção função limiar. Gráfico superior apresenta a função resultante na mesma escala que os exemplos anteriores e gráfico inferior a mesma função normalizada.

A Fig. 9 mostra os picos capturados no gráfico central indicando o início dos eventos em comparação com a função de fluxo espectral de diferença, sem passar por pós-processamentos no gráfico superior, e sua sobreposição com a forma de onda do som analisado no gráfico inferior.



**Fig. 9** - Exemplo final com a seleção dos picos apresentada no gráfico central. A fim de facilitar a comparação visual o gráfico superior retoma a função de detecção da Fig. 5 e o gráfico inferior sobrepõe os pontos de início selecionados com a forma de onda do som analisado.

Contudo, o valor  $i$  corresponde a amostras da análise do *Fluxo Espectral de Diferença* que são obtidas de quadros de análise espectral com  $N$  amostras do sinal  $x[n]$ , e por isso, o ponto de segmentação do evento musical é impreciso se baseado apenas em  $i$ . Para garantir que a segmentação do sinal gravado seja feita sobre um valor de amplitude próximo de ‘zero’ (para não haver *clicks* quando o segmento de áudio for tocado), programamos um algoritmo para busca de uma amostra que seja o primeiro valor após um cruzamento por zero e anterior ao ponto dado por  $i$ . Este algoritmo, apresentado no Pseudo-código 1, é calculado de forma iterativa e decremental sobre o sinal de áudio a partir da amostra dada por:  $i \cdot N$ , se  $Onset_i = 1$ .

```

n = i * N
h = 0
Enquanto h=0, faça {
  Se (sig(x[n]) ≠ sig(x[n-1]))
    h = 1
    retorne n
  Senão
    n = n-1
}

```

**Pseudo-código 1** - algoritmo que descreve a função de busca de ponto de cruzamento por “zero” no sinal  $x[n]$

### 3.1.3 Função para Detecção de Fim (Offset)

Para detecção dos finais (*offsets*) dos eventos musicais utilizamos um algoritmo que compara o valor RMS extraído do sinal de áudio (ver Eq. 2.12) com um limiar mínimo estipulado pelo usuário. Caso o valor de RMS cruze o limiar no sentido decrescente, ou seja, sendo a amostra anterior maior que o limiar mínimo e a atual menor, o algoritmo detecta um offset (ver Eq. 3.6).

$$Offset_i = \begin{cases} RMS_i < 0 \text{ e } RMS_{i-1} \geq 0 & 1 \\ else & 0 \end{cases} \quad (3.6)$$

Em seguida, o algoritmo descrito no Pseudo-código 1 também é aplicado para a detecção de um ponto de cruzamento por zero anterior ao do ponto de detecção do *offset*. E eventos são segmentados entre dois *onsets* quando o  $RMS_i$  não decresce o suficiente para caracterizar um offset segundo a Eq. 3.6.

A posição de um evento no tempo em relação a um tempo inicial 0 é descrita pelo valor  $n$  retornado no algoritmo do Pseudo-código 1. A duração de um evento é dada pela amostra  $n$  de *onset* do segmento subtraída da amostra  $n$  seguinte, seja respectiva a um *onset* ou a um *offset*.

Por fim, os intervalos temporais entre a detecção de *offset* e de um novo *onset* são classificados como silêncio.

## 3.2 Extração de Conteúdo Harmônico

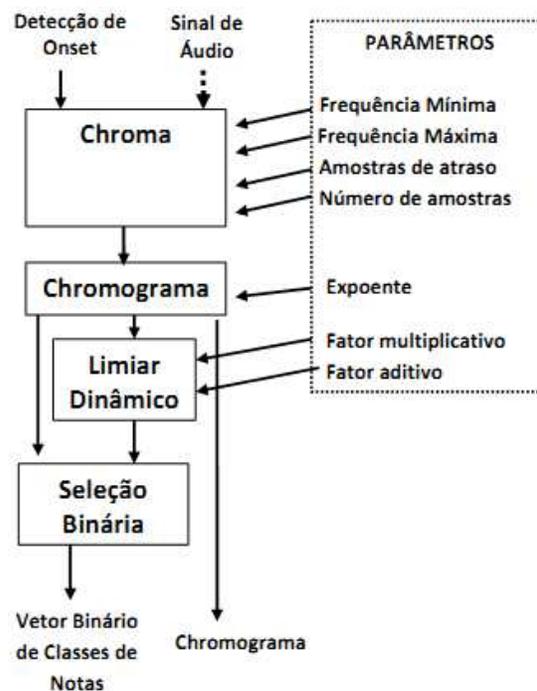
O vetor Chroma  $Chroma[n]$  (vide **Capítulo 2**) representa a distribuição de energia espectral de uma janela de análise e dada pela divisão em semitons do espectro sonoro. Tipicamente o descritor Chroma é computado em janelas de 8192 amostras que corresponde a um intervalo temporal de 185 milisegundos a uma taxa de amostragem de 44100Hz. Esse intervalo pode ser impreciso para representar o conteúdo harmônico de um segmento sonoro que evolui no tempo, ou seja, um intervalo temporal constante pode não ser capaz de identificar variações súbitas. Por isso implementamos um segundo procedimento denominado de Chromograma que é um histograma normalizado de sucessivos vetores de Chroma computado durante um número específico de janelas de análise do sinal.

Deixamos na nossa implementação deste algoritmo que algumas das determinações paramétricas sejam especificadas pelo usuário. Com isto pretendemos

que a extração do Chroma seja otimizada para o tipo de evento sonoro que estiver sendo analisado. Os parâmetros disponíveis ao usuário são:

- Segmento espectral de análise: Dois valores de frequência em Hertz definem um segmento do espectro para extração do chroma. Essa medida propicia que a extração da energia por classes-de-notas seja efetuada na região espectral onde se encontram as primeiras parciais harmônicas, as quais são mais importantes para definir a percepção de nota. Isso também previne a inserção de dados indesejados nos resultados pela análise de transientes de alta frequência. Jehan (2005) propõe aproximadamente seis oitavas de âmbito entre Dó 1 e Si 7 (de 65Hz a 7902Hz) enquanto Bello e Pickens (2005) propõem a análise entre 98Hz e 5250Hz. No caso específico da aplicação para qual propomos para essa ferramenta (para análise de instrumento acústico) esses parâmetros variam de acordo com a distribuição espectral e tessitura de cada instrumento em específico, por isso sua determinação foi configurada como parâmetro ao usuário.
- Janelas de análise consideradas: inclui dois parâmetros que são o número de vetores de chroma (ou janelas de análise) a desconsiderar após a detecção de onset para a computação do Chromograma; e o número de vetores de Chroma consecutivos a serem computados no Chromograma. Essa medida visa direcionar o segmento de análise para o corpo de sustentação do envelope sonoro, prevenindo com isso a inserção do ruído gerado pelos transientes de ataque e pelo aumento da relação sinal-ruído do sistema no decaimento do envelope.
- Exponenciação: aplicamos uma exponenciação após a somatória dos vetores de chroma e anteriormente a sua normalização com objetivo de ressaltar os picos energéticos e facilitar sua seleção. O valor do expoente é parâmetro do usuário.
- Limiar dinâmico: após a obtenção do chromograma normalizado, operamos a seleção dos picos através de uma limiarização dinâmica. O limiar é definido pela média dos valores do chromograma que em seguida é ponderada por um fator multiplicativo e adicionada a um fator de ajuste fino, ambos os fatores também são parâmetros do usuário.

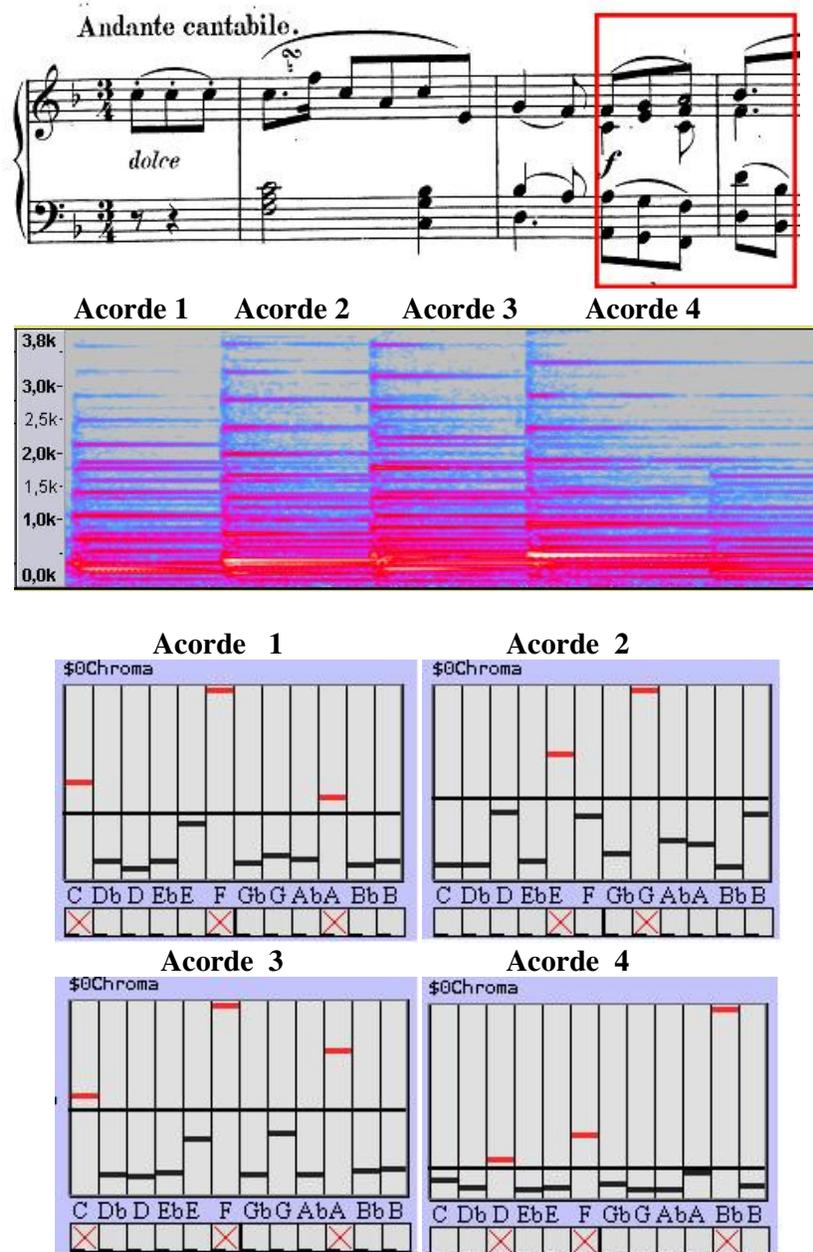
Após a limiarização e seleção dos picos o algoritmo retorna um vetor binário resultante da seleção dos picos maiores que a medida de limiar dinâmico. Estes picos representam a presença de determinadas classes de notas no segmento sonoro analisado. Por fim, o algoritmo pode retornar tanto o vetor cromático binário quanto o chromograma. O esquema geral do processo algorítmico para extração de conteúdo harmônico está representado na Fig. 10.



**Fig. 10** - Fluxograma do algoritmo de classificação harmônica

As Fig. 11 mostra um exemplo dos resultados obtidos pelo algoritmo para extração de conteúdo harmônico. Na parte superior da Fig. 11 está a partitura e o respectivo espectrograma do tempo final do 3º compasso e tempo inicial do 4º compasso do segundo movimento (contando o compasso de anacruse) da sonata para piano em Dó maior opus K330 de Mozart. O áudio relativo à partitura da Fig. 11 está contido na Faixa 6 do CD Apêndice A. Na parte inferior da Fig. 11 estão as representações dos resultados do algoritmo de extração de conteúdo harmônico para os quatro acordes, na mesma seqüência em que estão na partitura. Os retângulos superiores e segmentados nas representações de resultados do algoritmo são a representação gráfica do vetor do Chromograma. A linha contínua representa o limiar adaptativo e os

segmentos vermelhos os picos selecionados acima do limiar. A fileira de quadrados inferior representa o vetor binário cromático. As letras acima dos quadrados e abaixo do gráfico do Chromograma são as cifras das 12 notas musicais da escala cromática temperada e o símbolo X nos quadrados indica que aquela respectiva nota está contida no vetor binário.



**Fig. 11** - Extração de conteúdo harmônico de quatro acordes do segundo movimento da sonata opus K330 de Mozart. Na partitura acima os quatro acordes analisados e o gráfico central ilustra o espectrograma destes acordes sequenciais. Os gráficos inferiores contém a distribuição energética do chromograma de cada acorde nos *sliders* superiores e o vetor binário representado pelas filas de 12 quadrados.

Nota-se que o acorde 4 na partitura é fragmentado em dois ataques (conjunto de notas simultâneas em duas colcheias), apenas a análise do último ataque foi considerada. Por isso observa-se a predominância da nota Si bemol, pois a energia da ressonância das demais notas já havia decaído um pouco quando a nota Si bemol (dobrada em oitava) foi tocada.

Outro exemplo interessante é apresentado na Fig. 12 que mostra a extração do conteúdo harmônico dos 4 primeiros acordes da peça número 3 das “Dez Peças para Quinteto de Sopros” de György Ligeti. Ligeti inicia os dois primeiros compassos desta peça com o mesmo acorde formado pelas notas Si, Ré bemol e Mi bemol e cria uma estrutura permutativa para a associação entre notas e timbres instrumentais, como pode ser observado na partitura da parte superior da Fig. 12. O chromograma extraído da seção da peça mostra a variação do equilíbrio de energia espectral entre as notas causada pela modificação do timbre associado para cada nota e que é percebido pela escuta humana (durante a audição da peça) como uma alteração de colorido do acorde, vide parte inferior da Fig. 12. Uma gravação de áudio relativo à partitura da Fig. 12 pode ser escutada na Faixa 7 do CD Apêndice A.

# 3

Lento

(♩ = 40)

stets sehr weich einsetzen  
enter very softly throughout

1 2 3

Flauto contralto in Sol  
*pp* *dolciss., sempre cantabile*

Oboe d'amore  
*pp* *dolciss., sempre cantabile*  
Oboe d'amore:  
*sempre poco in rilievo*

stets sehr weich einsetzen  
enter very softly throughout

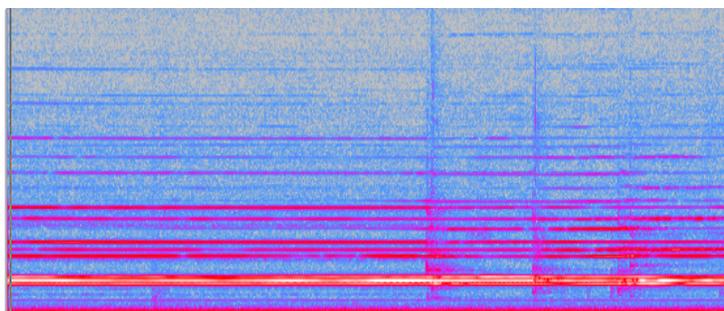
Clarinetto in Sib  
*pp* *dolciss., sempre cantabile*

stets sehr weich einsetzen  
enter very softly throughout

Corno in Fa  
*con sord.*  
*pp* *dolciss., sempre cantabile*

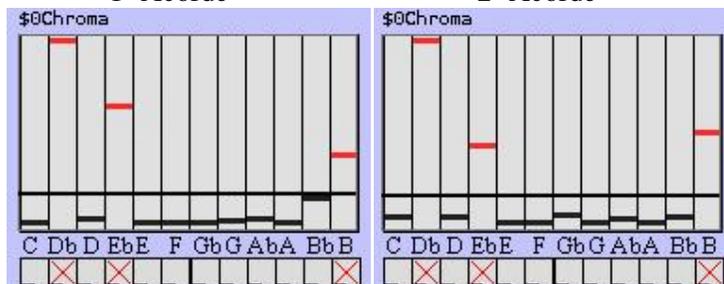
stets sehr weich einsetzen  
enter very softly throughout

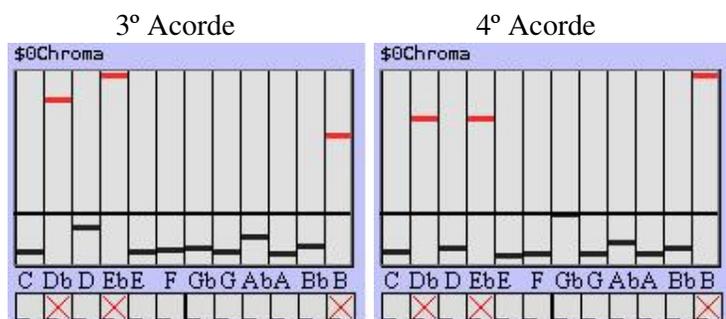
Fagotto  
*pp* *dolciss., sempre cantabile*



1° Acorde

2° Acorde





**Fig. 12** - Extração do Chromograma dos dois primeiros compassos da peça número 3 das “Dez Peças para Quinteto de Sopros” de György Ligeti.

### 3.3 Análise de Dinâmica Espectral via Mapas de Poincaré

Descrevemos nessa seção uma metodologia para análise do comportamento dinâmico de características de um sinal sonoro. O método constitui-se na aplicação de Mapas de Poincaré em séries temporais unidimensionais resultantes da extração de descritores acústicos. Primeiramente realizamos experimentos pilotos em trabalhos de análise musical (MONTEIRO e MANZOLLI, 2011a) em seguida incorporamos o método na construção dos sistemas musicais computacionais com propósito de recuperar informações musicais relacionadas a características de evolução sonoro-espectral da performance.

#### 3.3.1 Mapas de Poincaré

O matemático francês Henri Poincaré desenvolveu no começo do século XX um estudo sobre dinâmicas complexas e uma ferramenta útil para analisar este tipo de fenômeno: os “mapas de Poincaré” (POINCARÉ, 1952). A teoria de Poincaré tem várias aplicações importantes, como Jirsa e Kelso (2005) que utilizam sistemas dinâmicos e o teorema de Poincaré–Bendixson para estudar a geração de movimentos rítmicos. Os mapas de Poincaré são utilizados também na análise da variabilidade dos batimentos cardíacos (ACHARYA et al. , 2006) e do comportamento coletivo em interações sociais (CAMURRI et al. , 2010), entre muitos outros estudos em ciências naturais e aplicadas.

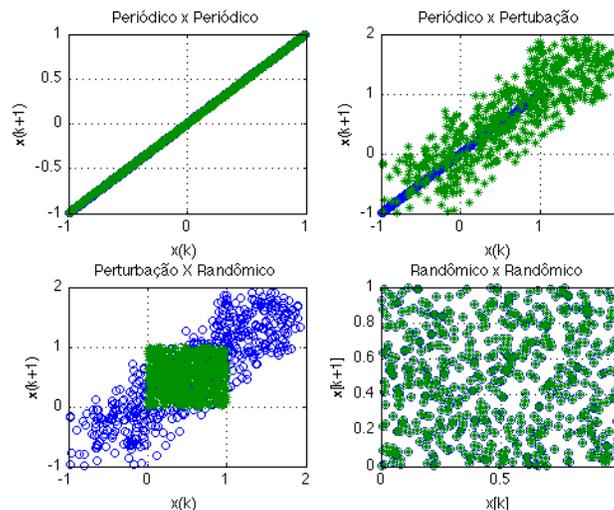
Definimos formalmente um mapa de Poincaré como:

*Dada uma série temporal unidimensional  $S(t) : [a...b] \subset \mathfrak{R} \rightarrow [c...d] \subset \mathfrak{R}$  com valores tomados em  $t = t_1, t_2, t_3 \dots t_n$  descritos como  $A = \{x_1, x_2, x_3, \dots, x_n\}$ , o mapa de*

Poincaré é o gráfico associado ao conjunto de pares ordenados  $P = \{(x_1, x_2), (x_2, x_3) \dots (x_{n-1}, x_n)\}$ .

O gráfico resultante é um plano bidimensional em que em um dos eixos descreve os valores da própria função e o outro descreve os valores da função com atraso de um elemento. Com isso, as coordenadas dos pontos no gráfico são determinadas por dois elementos consecutivos da função, cada um associado a um eixo do gráfico.

Os mapas de Poincaré descrevem “*recorrência de padrões no tempo*” que é uma propriedade fundamental de uma grande variedade de sistemas dinâmicos e caracteriza a regularidade ou não-regularidade temporal dos estados de sistemas complexos ou caóticos (BRADLEY, MANTILLA, 2002). Eles sintetizam o comportamento complexo de uma série temporal num gráfico bi-dimensional descrevendo recorrências quando uma mesma região do gráfico é reiterada formando aglomerados de pontos ou *clusters*. Exemplificando, quando o mapa descreve uma sequência randômica, o gráfico é totalmente preenchido com pontos. Em oposição quando a informação é periódica, apenas a faixa da diagonal principal é preenchida. Quanto mais estreita for a faixa da diagonal mais recorrente é o sinal. Os mapas da Fig. 13 são: (acima, esquerda) mapa de uma sequência periódica; (acima, direita) sequência periódica vs. uma perturbada com números aleatórios; (abaixo, esquerda) sequência perturbada vs. uma aleatória; (abaixo, direita) superposição de duas sequências aleatórias.



**Fig.13:** Exemplos de Mapas de Poincaré

### 3.3.2 Análise de Funções Extraídas por Descritores

As funções extraídas pelos algoritmos descritores são séries temporais que podem ser analisadas em um mapa de Poincaré para que possam ser observadas as características de seu comportamento dinâmico. Nossa motivação para este tipo de aplicação parte da hipótese de que a análise visual por mapas de Poincaré traga alguns pontos desejáveis no contexto da performance musical:

- Os mapas de Poncaré privilegiam a observação de recorrências dos dados analisados o que pode colocobarar para a difrenciação entre os principais elementos de articulação musical e os elementos pouco recorrentes, tendo em vista o parâmetro analisado;
- A possibilidade de se observar características do comportamento de séries temporais com uma grande quantidade de dados sem limitações geradas pela resolução visual da interface gráfica. Por exemplo, uma função com milhares de pontos tem sua definição visual limitada pelo número de pixels horizontais da interface (tela) se estiver em função do tempo, sendo os dados ‘plotados’ em seqüência no eixo horizontal. Nos mapas de Poincaré, sendo o âmbito de valores da função menor ou igual ao número de pixels do sistema gráfico, não importa a quantidade de pontos da função a sua análise não será prejudicada por falta de definição, sendo que os pares consecutivos da função são associados como coordenadas no plano.

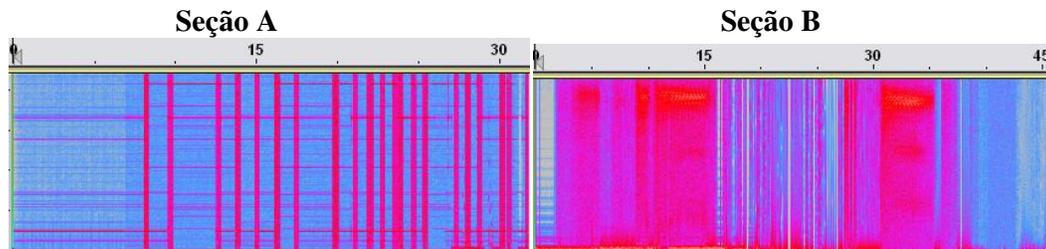
Um modo efeciente de observação dos dados extraídos pela análise e que ressalte as informações mais relevantes para o fim requerido são fatores importantes para o contexto deste trabalho, pois é desejável a prontidão de compreensão e resposta do usuário considerando que esses métodos foram aplicados em um sistema de performance musical em tempo real.

Tendo em vista o propósito de recuperação de informação musical, a escolha do descritor que será associado com o mapa de Poincaré depende da decisão sobre qual característica sonora é importante para organização musical do caso analisado.

No caso do sistema computacional que desenvolvemos na obra “As Duas Criaturas que Estavam à Mesa de Chá não Tiveram esta Conversa” a relação entre ruído (com conteúdo espectral com alta taxa de variação na distribuição energética) e sinal constante (com conteúdo espectral com baixa taxa de variação na distribuição energética) é importante para articulação musical na performance do sistema e é interessante que esta relação seja capturada. Por isso, escolhemos o Fluxo Espectral (vide equação 2.4) como descritor apropriado para captura das características desejadas e posterior análise nos mapas de Poincaré.

Mais detalhes sobre as causas dos padrões sonoros citados acima, assim como sobre a implementação e propriedades do sistema computacional podem ser vistos no **Capítulo 5**. Nesse capítulo apenas aludimos a essa característica da resposta sonora para apresentarmos um exemplo da aplicação de Mapas de Poincaré como método de recuperação da informação musical, que segue abaixo.

A Fig. 14 contém o espectrograma de duas seções da gravação de uma performance de “As Duas Criaturas que Estavam à Mesa de Chá não Tiveram esta Conversa”. Os respectivos áudios das seções A e B podem ser escutados nas Faixas 8 e 9 do CD Apêndice A.

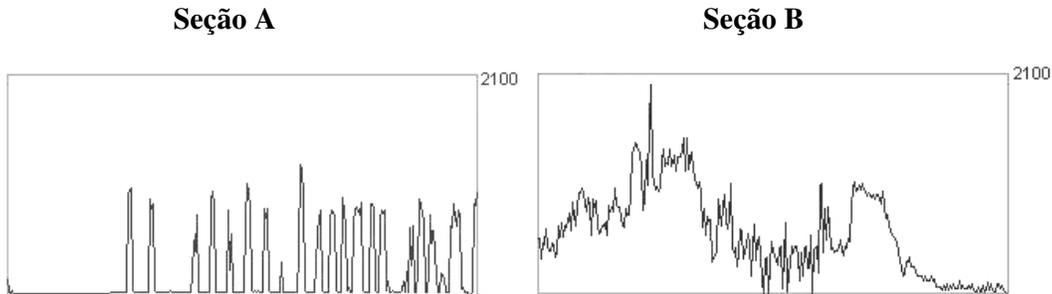


**Fig. 14** - Espectrogramas de duas seções de uma mesma gravação de *As Duas Criaturas que Estavam a Mesa de Chá não Tiveram esta Conversa*

As regiões predominantemente azuladas nas seções A e B representam sonoridades com baixa energia (com conteúdo espectral com baixa taxa de variação na distribuição energética) e as linhas avermelhadas horizontais representam a permanência de energia espectral em bandas frequências estreitas. As regiões com tonalidade avermelhada em todo o eixo vertical representam sonoridade ruidosa (sinais com conteúdo espectral com alta taxa de variação na distribuição energética). Na seção A nota-se uma estrutura de permutação entre essas duas características formando padrões visuais com certa simplicidade geométrica e na seção B nota-se a transição

entre os dois estados com característica interpolada gerando um aspecto visual de *dégradé* de tons avermelhados.

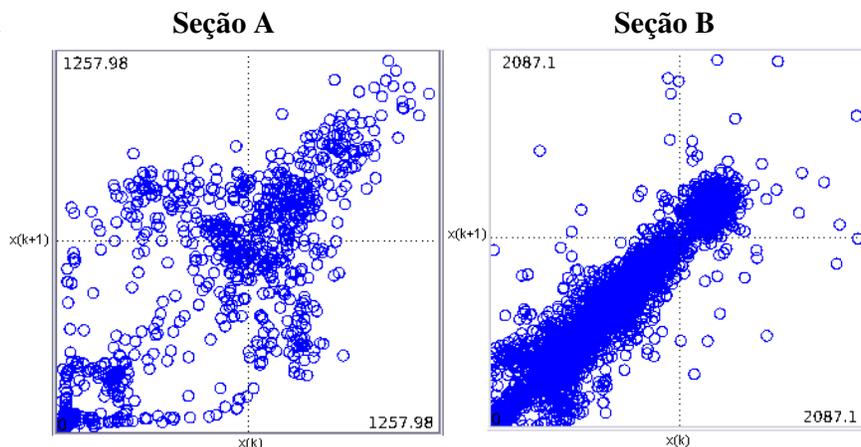
A Fig. 15 apresenta os gráficos gerados da extração do fluxo espectral das seções apresentadas nos espectrogramas da Fig. 14.



**Fig. 15** - Extração do Fluxo espectral das Seções A e B presentes na Fig. 14.

Comparando as funções na Fig. 15 com os respectivos espectrogramas na Fig. 14, nota-se que as regiões de valores altos nos gráficos da Fig. 15 ocupam a mesma posição no eixo horizontal que as regiões com predominância de cores escuras em todo sentido vertical nos espectrogramas da Fig. 14. Ou seja, os picos das análises são encontrados nas mesmas regiões que a representação de ruído (envelope espectral com alta variabilidade), confirmando a utilidade da extração de Fluxo Espectral para detecção de sonoridades com envelope espectral de alta variabilidade

Na Fig. 16 apresentamos os mapas de Poincaré gerados a partir das funções da Fig. 15.



**Fig. 16** - Mapas de Poincaré obtidos através das funções da Fig. 15. À esquerda mapa da seção A, e à direita mapa da seção B.

A análise dos dados espectrais nos mapas de Poincaré possibilita a visualização da complexidade da dinâmica das medidas extraídas. A densidade de

pontos no quadrante superior indica a presença de conteúdo espectral com alta taxa de variação (sonoridade ruidosa) enquanto pontos acumulados próximos da origem do gráfico indicam conteúdo espectral estático. A concentração de pontos na faixa diagonal do mapa é índice da recorrência dos valores de Fluxo Espectral e a dispersão fora da diagonal representa alterações dos valores. A seção A possui maior dispersão de pontos no gráfico (à esquerda na Fig. 16), evidenciando mais permutações e mudanças súbitas entre valores altos e baixos de Fluxo Espectral que correspondem à permutação entre as características de alta variação e estabilidade do envelope espectral. A concentração de pontos verticais na seção B (à direita na Fig. 16) evidencia estabilidade/recorrência das componentes espectrais e mudanças mais graduais.

Com os **Capítulos 2 e 3** sintetizamos os procedimentos de análise automática em tempo real via processo computacional que implementamos durante a pesquisa. Como já mencionamos estes procedimentos tiveram a função de estabelecer um canal de comunicação entre o processo de performance musical com instrumentos acústicos e o intérprete computacional. Estes dois capítulos encerram um bloco da pesquisa no qual estudamos e desenvolvemos uma busca intensiva na literatura recente sobre Descritores Acústicos e suas aplicações em análise automática. Nos próximos capítulos focaremos um mecanismo generativo de síntese digital via funções não-lineares e, finalmente, no **Capítulo 5** apresentamos o sistema para performance interativa que implementamos e as obras que foram criadas em conjunto com o desenvolvimento gradual e aprimoramento da programação.



# CAPÍTULO 4

## SÍNTESE SONORA DIGITAL VIA EQUACÃO DE LATOOCARFIAN

Este capítulo apresenta o método de síntese digital que utilizamos para gerar a resposta dos Instrumentos Digitais. Trata-se de um método que produz amostras sonoras digitais no domínio do tempo a partir da amostragem de órbitas numéricas geradas por equações não-lineares. Este foi um dos principais recursos empregados nas criações musicais algorítmicas descritas no **Capítulo 5**. Primeiramente fazemos uma breve revisão de aplicações de sistemas dinâmicos não-lineares e fractais em composição e síntese sonora, em seguida expomos o método desenvolvido e alguns exemplos das formas de ondas geradas através da amostragem das iterações numéricas do sistema não-linear descrito pela equação de Latoocarfián (PICKOVER 1994).

### 4.1 Sistemas Dinâmicos Não-lineares em Composição e Síntese

A área de Sistemas Dinâmicos é bem conhecida e determinada dentro da Matemática Aplicada e há vários livros de referência que podem ser pesquisados. Este capítulo foca uma aplicação específica em síntese sonora digital, pois o estudo de sistemas dinâmicos foge ao escopo desta dissertação.

É bem conhecido na literatura que sistemas dinâmicos não-lineares podem gerar comportamentos numéricos variados e complexos (HASSELBLATT, KATOK 1997; MOREIRA 1993). Um dos aspectos mais interessantes deste tipo de processo matemático é que órbitas numéricas complexas são produzidas por iterações de regras simples, como acontece, por exemplo, em autômatos celulares, fractais e em sistemas físicos tais como formação de galáxias e estruturas cosmológicas entre outros. Esta seção descreve alguns aspectos importantes da teoria de sistemas dinâmicos com o objetivo de introduzir a natureza do comportamento numérico aplicado à síntese sonora digital estudada durante a pesquisa.

#### 4.1.1 Características dos Sistemas Dinâmicos

As origens da teoria dos sistemas dinâmicos estão relacionadas com o estudo da mecânica celeste desenvolvido por Johannes Kepler e a mecânica clássica Newtoniana no final do século XVII. O estudo contemporâneo de Sistemas Dinâmicos reporta-se ao começo do século XX com os trabalhos de Birkhoff (1927) e de Poincaré (1952, 1993).

Sob o ponto de vista físico os sistemas dinâmicos são descritos como “sistemas fora do equilíbrio”, pois os seus estados mudam com o tempo. Em outras palavras, caracterizam fenômenos que têm estados que mudam em função do tempo, por exemplo, mudanças climáticas, variações de espécies biológicas, comportamentos financeiros, entre outros.

Estes sistemas podem ser definidos por equações discretas ou contínuas, no caso da aplicação estudada durante a pesquisa, vamos utilizar sistemas discretos nos quais os estados são determinados por passos, ou iterações.

Se chamarmos os estados do sistema de  $X_n$ , onde “n” é o passo que varia de  $n = 0, 1, 2, 3, \dots$ , a forma geral que descreve o comportamento temporal é dada por  $X_{n+1} = F(X_n)$ . Em outras palavras, o estado ou passo futuro “n+1” do sistema depende do passo atual “n” que é transformado por uma regra ou função “F”.

A sequência numérica gerada por esta regra, ou seja, os estados nos passos  $n = 0, 1, 2, 3, \dots$ , é denominada de trajetória ou órbita que é escrita como  $\{X_0, X_1, X_2, X_3, X_4, \dots\}$ . É essa órbita numérica que foi utilizada durante a pesquisa como meio de produzir amostras sonoras digitais a partir da Equação Latoocarfian (Eq. 4.1).

O segundo aspecto a ser descrito para se entender a aplicação estudada relaciona-se com a classificação entre sistemas *lineares* e *não-lineares*. Os chamados sistemas lineares são geralmente formulados por uma regra ou equação simples na qual grandezas relacionam-se por uma proporção, por exemplo, se utilizarmos uma equação do primeiro grau temos que  $X_{n+1} = K \cdot X_n$ . Em outras palavras os valores da órbita são proporcionais a constante “K”. Em contrapartida, os sistemas não-lineares caracterizam-se por equações que estabelecem relações de outra ordem, por exemplo, uma equação do segundo grau onde  $X_{n+1} = K \cdot (X_n)^2$ .

A partir destas duas propriedades brevemente descritas acima, ou seja, a *geração de órbitas* e a *não-linearidade*, destacamos que há sistemas dinâmicos que têm o seu comportamento temporal associado a uma regra ou função não-linear. Estes sistemas são chamados de *deterministas*, pois a regra ou função que define os seus estados futuros é bem determinada analiticamente. Em outras palavras, um sistema dinâmico não-linear implica que a geração dos seus estados *individualmente* não são aleatórias, e sim, previsíveis.

Todavia, quando observa-se a órbita numérica produzida por um número grande de iterações de um sistema não-linear, vê-se que elas apresentam comportamento altamente complexo impossibilitando a predição de estados futuros distantes. Estes sistemas evoluem com comportamento temporal aperiódico, onde o seu estado futuro é

extremamente dependente de seu estado atual, e transforma-se radicalmente com pequenas mudanças nos seus parâmetros de controle. Ou seja, apesar de apresentar características caóticas, o comportamento desses sistemas é determinístico e altamente dependente de seu estado inicial e dos parâmetros. Toda vez que o sistema apresentar os mesmos valores paramétricos e o mesmo estado inicial o seu comportamento será idêntico.

A propriedade de imprevisibilidade dos sistemas não-lineares é conhecida na área pela denominação de *caos determinístico* como descrito por Gleick (1989). Se analisarmos as possibilidades de comportamento numérico produzido pelas órbitas dos sistemas dinâmicos podemos observar as seguintes características gerais:

- a) convergência e estabilidade para um valor fixo;
- b) convergência para uma órbita numérica que oscila em valores específicos;
- c) comportamento aparentemente caótico também denominado de “atrator estranho”;
- d) ‘expansão’ ou divergência numérica para valores infinitamente grandes ou pequenos.

Nos casos a, b e c onde a órbita numérica ocupa um subespaço dos estados do sistema, é dado a denominação de “atrator” ao estado ou conjunto de estados para o quais o sistema converge.

Estas possibilidades de gerar comportamentos numéricos variados ou quase-periódicos foi uma das principais propriedades dos sistemas dinâmicos não-lineares que fomentaram a nossa pesquisa. Nosso objetivo foi explorar as peculiaridades sonoras do som digital a partir de comportamentos numéricos complexos.

#### **4.1.2 Dinâmicas Não-lineares, Fractais e Composição**

Em meados da década de 1970 o matemático Benoit Mandelbrot que criou a Geometria Fractal (MANDELBROT, 1982) descreveu propriedades fractais na órbita numérica ou atratores produzidos por funções não-lineares. Através do mapeamento das funções não-lineares como imagens bidimensionais num plano cartesiano, Mandelbrot obteve órbitas numéricas com propriedades geométricas fractais que apresentaram resultados estéticos interessantes e que então cativaram interesses no campo das artes visuais e na música.

Por exemplo, o compositor Gyorgy Ligeti inspirado pela complexidade autossemelhante dos fractais escreveu seu *Estudo para Piano nº 1 - Désordre* (1985). A partir do final da década de 1980 alguns músicos propuseram a utilização de equações não-lineares em sistemas computacionais para geração ou manipulação de material musical. Em alguns trabalhos vemos o mapeamento de sistemas não-lineares para parâmetros musicais como altura, duração, dinâmica como descrito em (BIDLACK, 1992; CHAPEL, 2003; DIAZ-JEREZ, 2000; DODGE, 1988; GOGINS, 1991; PRESSING, 1988; PETERS, 2010), ou para procedimentos de variação de material pré-existente (DABBY, 1995).

Em outras abordagens as órbitas numéricas produzidas por equações não-lineares foram utilizados na síntese digital de áudio para controlar os parâmetros de Síntese Granular (DI SCIPIO, 1990; TRUAX, 1990), ou as órbitas produzidas por várias iterações foram diretamente mapeadas como amostras de áudio digital (DI SCIPIO, PRIGNANO, 1996; MANZOLLI, 1993a, 1993b; MONRO, 1995), ou ainda a síntese por equações não-lineares foi obtida através de um procedimento de retro-alimentação cruzada em síntese FM e síntese AM (VALSAMAKIS, MIRANDA, 2005), há também um caso em que o sistema não-linear foi utilizado como princípio para construção de um instrumento sintetizador analógico (SLATER, 1998).

Ao utilizar-se o resultado das órbitas numéricas estes procedimentos partem de modelos matemáticos abstratos para gerar sinal sonoro no domínio do tempo e no nível ‘microscópico’ de amostras digitais. Outro exemplo dessa classe de método de síntese que foi concebida na mesma época é a *Síntese Estocástica* (HOFFMANN, 2009; SERRA, 1993) elaborada por Iannis Xenakis e que emprega modelos estocásticos para gerar formas de onda que são obtidas por linhas poligonais que ligam pontos gerados por distribuições de probabilidade.

No método de síntese aqui reportado utilizamos funções obtidas de iterações de equações não-lineares como método de síntese sonora digital no domínio do tempo. Como descrito no trabalho de Manzolli (1993a) os sistemas dinâmicos não-lineares estão relacionados com a Geometria Fractal e podem ser mecanismos para síntese de som e modelo para interação musical em tempo real. No método *FracWave* (DAMIANI et al. 1999; HOFFMANN, MANZOLLI, 2005; MANZOLLI, 1993a, 1993b), órbitas numéricas produzidas por equações não-lineares são amostradas e armazenadas num *buffer* circular e utilizadas como *forma de onda* denominada *dynamic wavetable*. Segundo os autores:

(...) podemos, com isso, através de um controle algorítmico das formas de onda, emular sons com características que, possivelmente, venham refletir a dinâmica de processos complexos, assim como os NLD (sistemas dinâmicos não-lineares) são utilizados para descrever turbulência encontrada em outros fenômenos naturais (DAMIANI et al., 1999).

Um aspecto importante produzido pelo uso da síntese digital via equações não-lineares é a alta variabilidade no parâmetro da altura perceptiva do som. Como mencionado a seguir:

Vale considerar, porém, que a definição perceptual de altura depende profundamente das características do sinal proveniente da forma de onda gerada pela amostragem da órbita numérica (...) as não-linearidades dificultam o controle preciso do fator frequencial, variando de quasi-periódico a caótico ..., muitas vezes, a altura “perceptiva” do som pode variar, já que a forma de onda torna-se mais complexa e irregular, com variações muito bruscas (HOFFMANN, MANZOLLI, 2005): .

#### 4.2 Síntese via Equação de Latoocarfian

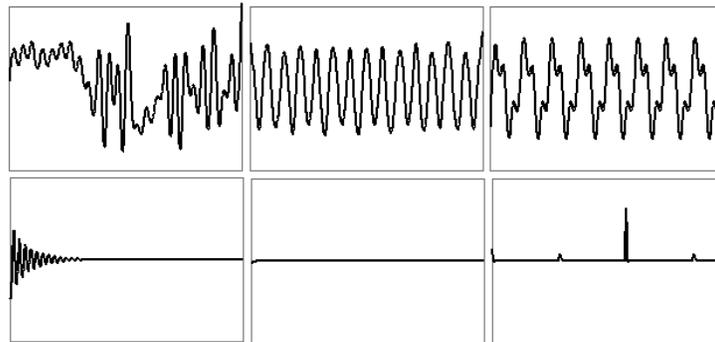
O método empregado nesse trabalho é similar ao visto no método de síntese digital FracWave de Manzolli (1993a, 1993b). Utilizamos uma equação não-linear nomeada por Pickover (1994) de *atrator Latoocarfian* (Equação 4.1) para gerar formas de onda. Para isso, associamos as órbitas numéricas geradas por iterações da equação com indexes de uma *look-up table* que posteriormente são amostrados como forma de onda para síntese sonora. A equação Latoocarfian é definida como:

$$\begin{cases} X_{n+1} = \sin(bY_n) + c \sin(bX_n) \\ Y_{n+1} = \sin(aX_n) + d \sin(aY_n) \end{cases} \quad (4.1)$$

Em que  $-3 < a < 3$ ,  $-3 < b < 3$ ,  $0.5 < c < 1.5$ ,  $0.5 < d < 1.5$

As sequências  $X_n$  e  $Y_n$  geram órbitas numéricas que são armazenadas em *arrays* numéricos ou *look-up tables*. Dentro da diversidade de comportamentos numéricos das órbitas geradas pela Eq. 4.1, identificamos a recorrência de 6 tipologias principais: *aleatoriedade*, *órbita periódica simples*, *órbita periódica complexa*, *rápida transição entre um comportamento anterior para um valor fixo*, *estabilidade em um valor fixo*, *sequência de pulsos*. É importante destacar que estes seis comportamentos estão diretamente associados à natureza do atrator como apresentado na seção 4.1.1. Desta forma, quando estas órbitas numéricas são amostradas produzem comportamentos

sonoros que são respectivamente: *ruído*, *sons periódicos simples*, *sons periódicos complexos*, *som percussivo*, *silêncio* e *sequência de pulsos* (vide Fig. 17). Vale ressaltar também que as formas de onda geradas não são apenas seis, mas sim, elas sempre apresentam características que são variações dessas seis tipologias mais gerais que listamos acima. Exemplos sonoros das tipologias podem ser escutados na Faixa 10 do CD Apêndice A desse trabalho. No áudio gravado, a sequência de tipologias tem a seguinte ordem: pulsos, som periódico simples, som periódico complexo, ruído e som percussivo.



**Fig. 17-** Tipologias de formas de onda geradas no móbile atrator. Acima, da esquerda para direita: ruído, onda periódica simples e onda periódica complexa. Abaixo, da esquerda para direita: som percussivo, silêncio e sequência de pulsos.

Anteriormente ao processo de leitura das formas de onda realizamos um procedimento de interpolação utilizando uma curva senoidal entre os pontos da função para suavizar seu contorno. Isso acarreta uma alteração do conteúdo espectral do som resultante da forma-de-onda, pois diminui a energia nos parciais altos e torna o timbre perceptivamente menos ruidoso e estridente.

O algoritmo de interpolação é descrito pelo Pseudocódigo 2 a seguir. Sendo que,  $P_{max}$  é o parâmetro que determina o número pontos de interpolação;  $N$  é parâmetro que determina o número total de iterações da equação Latoocarfian;  $Y_n$  é o valor da amostra gerada pela equação Latoocarfian na iteração  $n$ ;  $Y_{n-1}$  é uma variável interna que armazena o valor obtido na iteração passada ( $n-1$ ) da equação Latoocarfian,  $p$  e  $n$  são variáveis internas que correspondem aos contadores respectivamente do índice de interpolação e o índice de iteração da equação de *Latoocarfian*.  $Z$  é a variável interna que armazena o cálculo para obtenção do valor de interpolação no índice  $p$ ; e por fim *Tabela* é um vetor de memória que armazena os valores da função interpolada a cada iteração do algoritmo.

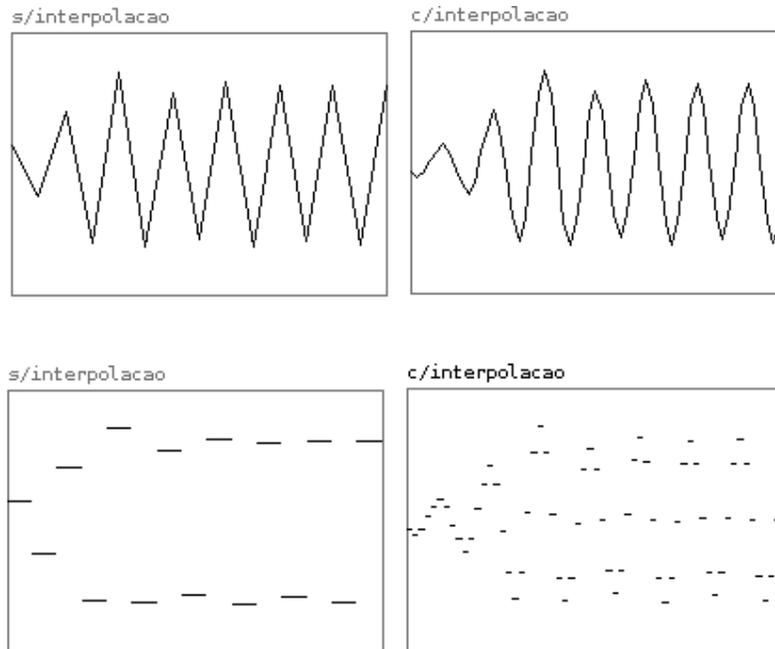
```

var p
var Z
var N = número de pontos iterações da função de Latoocarfian
var Pmax = número de pontos de interpolação
vetor Tabela = redimensiona para N * Pmax
var n = 0
var Yn-1 = 0
Enquanto (n < N) faça {
    p = 0
    Enquanto (p < Pmax) faça {
        Z = sin((p - 1)*(π / Pmax) - 1,57) / 2
        Z = Z * (Yn - Yn-1)
        Z = Z + Yn-1
        Tabela [n + p] = Z
        Yn-1 = Yn
        p = p + 1
    }
    n = n + 1
}

```

**Pseudocódigo 2** - Algoritmo para interpolação aplicado nas funções geradas pelo atrator Lattocarfian

A Fig. 18 compara uma mesma forma de onda gerada após 15 iterações da equação Latoocarfian (Eq. 4.1) sem interpolação (à esquerda) e após a interpolação de 4 pontos (à direita).



**Fig. 18** Exemplos gráficos comparando as funções: à esquerda antes do processo de interpolação e à direita após o processo de interpolação.

Os gráficos superiores da Fig. 18 ilustram através de segmentos de retas as formas de ondas e privilegiam a visualização da curvatura gerada pela interpolação senoidal. Os gráficos inferiores da Fig. 18 ilustram através de pontos (tracejados) as formas de ondas e privilegiam a visualização dos valores exatos dos pontos de interpolação.

Após a indexação de cada iteração da Eq. 4.1 (e os respectivos pontos de interpolação entre valores) na tabela, operamos a síntese sonora através da leitura da tabela como uma forma de onda fixa. A frequência de leitura é um parâmetro do sistema que é definido pelo usuário em tempo real. Tipicamente as funções resultantes da Eq. 4.1 apresentam âmbito maior que -1 e 1 (padrão de amplitude do sinal de áudio). Através de observações constatamos que operando com variáveis com 4 casas decimais (relação que mantemos fixa no algoritmo), os resultados da função não ultrapassam valores entre -3 e 3. Esses valores excedem a máxima de amplitude possível para o sinal de áudio o que acarreta em distorção do sinal. Para que isso não ocorra multiplicamos as saídas de áudio do sistema de síntese pelo fator 0.3. Dessa forma impedimos que o sinal distorça por exceder o limite de amplitude e conservamos as variações de amplitudes inerentes do funcionamento do algoritmo.

O próximo capítulo apresenta como o método de síntese sonora digital aqui descrito foi utilizado como meio de produção sonora do computador em tempo real. As características dinâmicas e grande variedade de sonoridades produzidas pelas formas de onda geradas por este método foram fundamentais para a caracterização timbrística das obras criadas durante a pesquisa.

# CAPÍTULO 5

## PROCESSO CRIATIVO

Neste capítulo apresentamos três Estudos Composicionais. Estes Estudos fazem parte da série denominada de *As Duas Criaturas* e representam o processo criativo resultante dos métodos estudados, pois neles aplicamos as técnicas descritas nos capítulos anteriores desta dissertação.

A primeira parte deste capítulo é dedicada à apresentação das idéias gerais com as quais a série de Estudos musicais sobre um mesmo princípio algorítmico foi composta. Em seguida cada uma das três peças é apresentada nas suas peculiaridades, a saber: “*As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram essa Conversa*” duo para computador e contrabaixo elétrico; “*As Duas Criaturas Baforaram seus Óculos*” duo para computador e trombone; “*Não Diziam-nos que iam Contar Nosso Passado*” para computador solo. Durante a pesquisa foi também gerada uma nova versão da primeira obra e nela foram aplicados os métodos de recuperação de informação musical, conforme será descrito nas próximas seções. Este capítulo apresenta também análises gráficas de segmentos das obras criadas com intuito de demonstrar a organização musical resultante do processo algorítmico e da interação entre o(s) interprete(s) em tempo real.

O método de análise adotado neste Capítulo está relacionado com o processo de análise sonora apresentado no **Capítulo 2**. Utilizamos espectrogramas que representam graficamente as magnitudes de cada janela de análise espectral (vide seção 2.1.1) em função do tempo, e gráficos construídos a partir de dados extraídos pelos Descritores Acústicos: Centróide espectral (vide **ANEXO A** - seção A.2.1.1), Desvio Padrão Espectral (vide **ANEXO A** - seção A.2.1.2) e RMS (vide seção 2.1.4) . Em suma, no **Capítulo 5** há uma convergência de todos os elementos estudados que foram aplicados ao processo criativo dando origem às três obras da série.

### 5.1 *As Duas Criaturas*

Trata-se de uma série de estudos musicais algorítmicos que foram criados como um conjunto de variações do primeiro Estudo da série. O projeto de composição de “*As Duas Criaturas*” não foi planejado antecipadamente, mas se concretizou durante a criação do segundo estudo “*As Duas Criaturas Baforaram seus Óculos*”. A idéia de

uma série de composições nasceu da necessidade de se alterar as configurações do sistema computacional da primeira peça (*As Duas Criaturas que Estavam à Mesa de Chá não Tiveram Esta Conversa*) com o intuito de compor uma nova peça em parceria com o trombonista Pedro Faria.

Esta composição foi encomendada para ser estreada no “Concerto Multimodal” de abertura do “XIII Simpósio Brasileiro de Computação Musical” (SBCM) em Vitória, Setembro de 2011. O procedimento adotado foi adaptar os parâmetros e configurações da obra anterior como uma releitura do seu processo algorítmico. Com isso surgiu a idéia de desenvolvermos *uma série de estudos musicais sobre um mesmo princípio algorítmico* o qual seria remodelado de acordo com as necessidades de cada nova obra. Com este intuito foram estabelecidos alguns princípios como base para criação de novas composições. Os objetivos foram os seguintes:

- Explorar novas técnicas para interação entre instrumentos acústicos e/ou eletroacústicos e computador.
- Explorar o conceito de *instrumento composto* (como definido na seção 1.3.2 do **Capítulo 1**), ou seja, a dualidade entre instrumento e representação do algoritmo musical com as seguintes características:
  - do ponto de vista instrumental, verificar quais os parâmetros de controle dos algoritmos são propícios para obter-se controle e expressividade musical dos algoritmos (sobre performance com instrumentos digitais vide **Capítulo 1** seções 1.3.4 e 1.3.5)
  - do ponto de vista composicional, verificar como a noção de interpretação do algoritmo interfere na relação entre organizações musicais realizadas em tempo diferido e a condução desejada para o resultado musical em tempo real (sobre dualidade entre instrumento e representação musical vide **Capítulo 1** seções 1.3.2 e 1.3.3).
- Explorar um conjunto específico de processos de síntese sonora digital que serão utilizados de forma variada nos diversos contextos de organização musical e estéticos que for estabelecido para cada novo Estudo.

A partir destes objetivos gerais, optamos por não estabelecer um planejamento antecipado para organização dos Estudos em seções ou movimentos, por duas razões:

a) estamos interessados na experimentação empírica com os métodos que estudamos. Desta forma, o objetivo é fomentar condições para descoberta de novas relações e funcionalidades tanto no campo técnico quanto no campo poético e estéticos das obras compostas;

b) no caso da criação dos dois duos, eles foram compostos em parceria com os dois intérpretes. A imposição de uma organização prévia poderia impor barreiras ao fluxo de idéias advindo do processo de criação em parceria.

A adoção dessas duas posturas não significa que não houve planejamento algum na criação dos Estudos, mas sim que o planejamento global foi substituído por um processo emergente. Nele a organização final das obras foi sendo construída pouco a pouco durante o processo criativo, dialogando com a experimentação técnica e com relações no campo da estética e poética que surgiram durante o próprio processo criativo. Todavia, há dois pontos de similaridades entre as obras. O primeiro, de ordem técnica, é o conceito de *variação de um mesmo algoritmo*. O segundo, de ordem poética, está relacionado *com os títulos das peças*. Os títulos estabelecem uma lógica sucessória que descreve uma narrativa. Dois dos títulos, do primeiro e do terceiro Estudo, foram adaptados de frases extraídas de poemas de Fernando Pessoa, respectivamente: do fragmento 370 do ‘Livro do Desassossego’ do heterônimo Bernardo Soares e do poema dramático O Marinheiro. O título do segundo Estudo foi criado livremente.

## **5.2 As Duas Criaturas que Estavam à Mesa de Chá não Tiveram esta Conversa**

O primeiro Estudo, desenvolvido em parceria com o músico contrabaixista Lucas Araújo, constitui-se num ambiente algorítmico computacional para improvisação musical e surgiu de uma encomenda realizada pelo núcleo de música experimental IBRASOTOPE, local onde foi estreiado em 2009. Uma performance dessa peça integra uma faixa do CD “CCJ Experimental” de música experimental gravado na cidade de São Paulo. Sua principal apresentação ao vivo ocorreu no PDcon09 (Pure Data Convention) também na cidade de São Paulo. A instrumentação utilizada para esse

Estudo é *computador em sistema de áudio estéreo 2.0 e contrabaixo elétrico*. O sinal do contrabaixo foi conectado apenas à entrada de áudio do computador e não foi amplificado diretamente pelo sistema de difusão sonora da peça, ou seja, o contrabaixo foi utilizado somente como fonte de processamento sonoro. Para criação desse estudo foram preestabelecidos dois critérios:

- a) a obra não teria organização temporal predeterminada, deveria apenas caracterizar um ambiente para improvisação com recorrência de sonoridades;
- b) a sonoridade geral deveria ser predominantemente ruidosa.

Esta composição, quando analisada através do conceito de *instrumento composto* (vide **Capítulo 1**, seção 1.3.2), apresenta uma forte característica instrumental porque delimita um campo sonoro para improvisação no qual os gestos dos instrumentistas, acústico e computacional, entrelaçam-se nos mecanismos de controle sonoro. Esta obra foi estruturada em quatro módulos algorítmicos de processamento e/ou síntese sonora. Denominamos cada um dos módulos desenvolvidos para a primeira composição de *móviles*. Justamente porque cada um deles funciona como peça de encaixe livre que é inserida no discurso musical de acordo com a interação entre o contrabaixista e o *performer* computacional. Os quatro móveis apresentados a seguir são os seguintes: *Waveshapper*, *Transpositor*, *Sintetizador Aditivo processado por Phaser* e *Sintetizador Não-linear utilizando a equação Latoocarfian*.

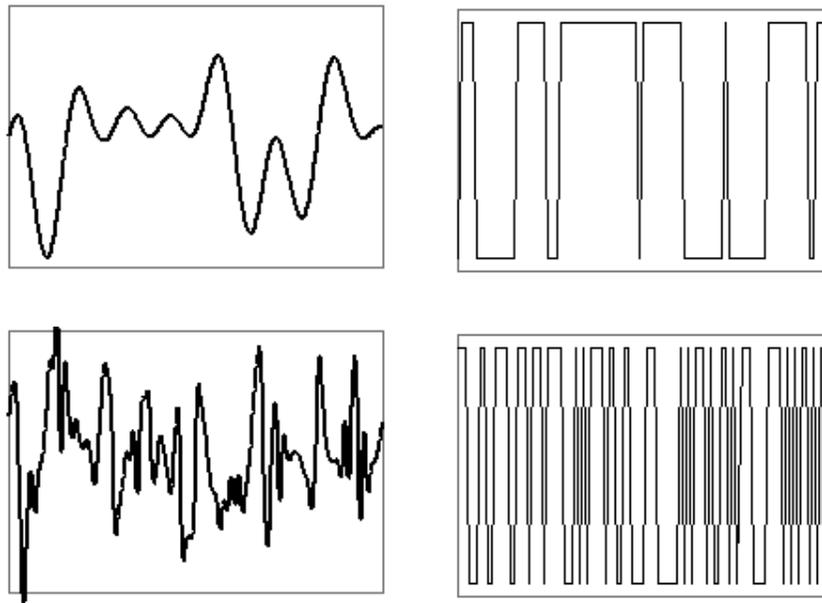
### 5.2.1 *Waveshapper*

O primeiro móbil é uma distorção do sinal do contrabaixo na qual utilizamos o método de *waveshaping* (ROADS, 1979). A distorção é obtida a partir da seguinte *função de transferência*.

$$y[n] = \begin{cases} 1, & x[n] > 0 \\ 0, & x[n] = 0 \\ -1, & x[n] < 0 \end{cases} \quad (5.1)$$

Onde  $x[n]$  é uma amostra do sinal de entrada e  $y[n]$  a saída.

Verifica-se facilmente que qualquer valor positivo ou negativo do sinal  $x[n]$  é associado ao seu valor máximo. Este procedimento transforma o sinal de entrada  $x[n]$  gerado pelo contrabaixo numa onda quadrada  $y[n]$  (vide Fig. 19) e iguala a intensidade do sinal do contrabaixo com o ruído do sistema elétrico-eletrônico, pois maximiza o valor de qualquer amostra  $x[n]$  com o máximo de intensidade da saída. O sinal de saída  $y[n]$  perde seu perfil dinâmico de intensidade e permite ao contrabaixista alternar, através do gesto instrumental, entre o ruído branco e o som distorcido com altura definida.

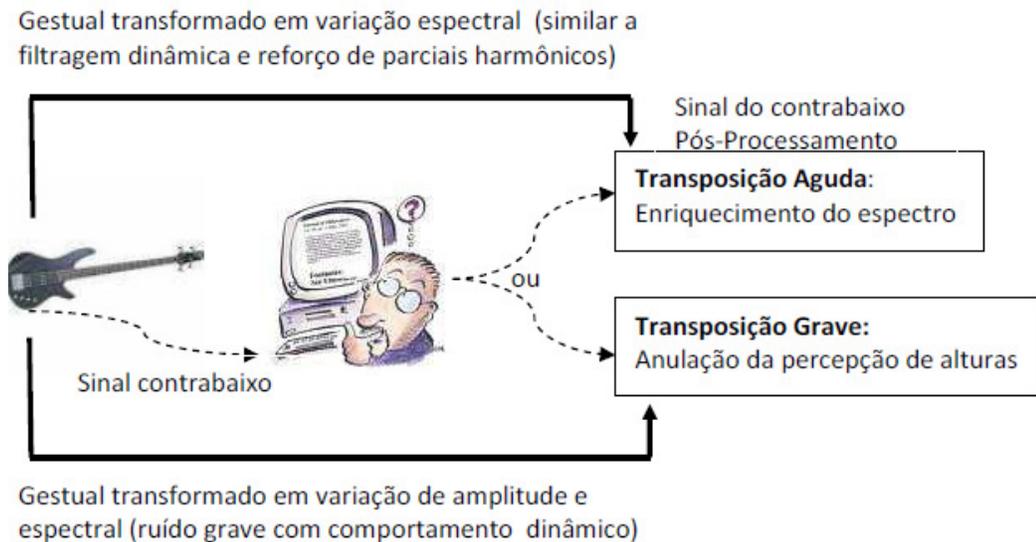


**Fig. 19** - Resultado da aplicação do móbile de distorção. Na parte superior da figura uma onda periódica complexa a esquerda e a direita a mesma onda após o processo de waveshaping. Na parte inferior da figura mesma relação pré e pós waveshaping para uma amostra de ruído branco.

A diminuição da amplitude do contrabaixo, quando realizada gradualmente através dos controles de ganho do sinal (como pedal de volume ou o próprio potenciômetro do contrabaixo), gera uma interpolação gradual do som do contrabaixo distorcido com altura definida para o ruído branco originado pelo *offset* do sistema elétrico-eletrônico e amplificado pelo algoritmo. A interpolação oposta também ocorre com o aumento gradual da amplitude do sinal do contrabaixo.

## 5.2.2 Transpositor

O segundo móbil foi construído sobre um processo denominado na literatura de *pitch-shifter* (PUCKETTE, 2006, p.202-208). Desenvolvemos dois modos de transposição. No primeiro o sinal é transposto para frequências muito graves o que causa a anulação da percepção de alturas do som do contrabaixo. Este efeito gera um ruído grave com dinâmica interna controlada pelo gesto do contrabaixista. O segundo modo transpõe qualquer altura do contrabaixo para um mesmo som de frequência fundamental alta, constante e de espectro rico. Nesse modo, o sinal de controle do contrabaixista gera um som similar a um filtro dinâmico com direção e velocidade dependentes do gesto físico do intérprete (Fig. 20). Nestes dois modos de uso do *pitch-shifter* mapeiam-se os gestos instrumentais do contrabaixista (que na execução tradicional do instrumento geravam uma diversidade sonora com variações de altura e timbre) para modulações internas de um som agudo com frequência constante, ou para um som muito grave numa faixa de frequências subsônicas.

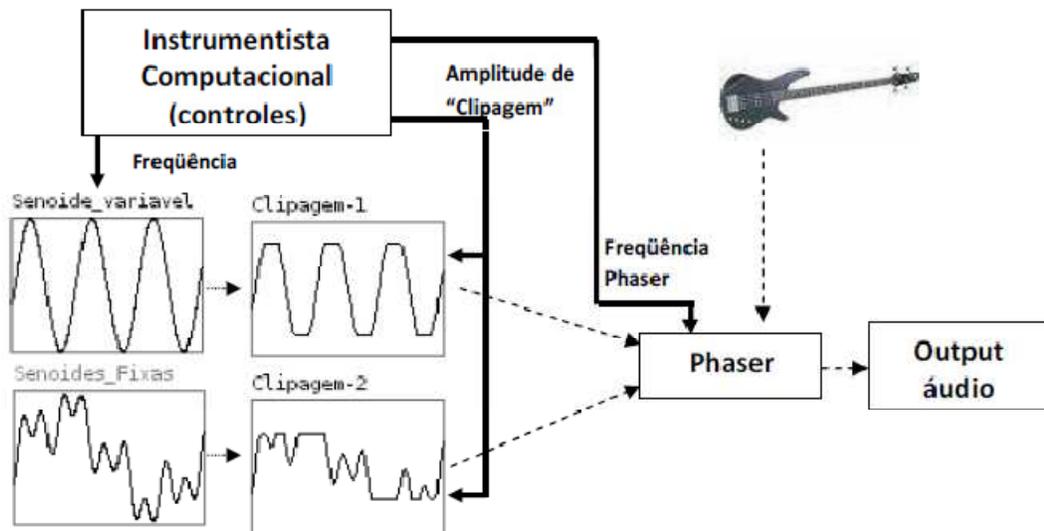


**Fig. 20** - diagrama da relação entre entrada do contrabaixo e processamento computacional. Setas tracejadas indicam o fluxo do sinal de áudio e setas contínuas a relação gesto e resultado sonoro

### 5.2.3 Sintesador Aditivo e *Phaser*

O terceiro móbile é baseado na aplicação do processamento de filtragem de áudio denominado *Phaser* (PUCKETTE 2006, p.261). O algoritmo implementado gera inicialmente um sinal composto por quatro ondas senoidais: três com frequências fixas e uma quarta com frequência variável através da interface do computador ou algum periférico (e.g teclado MIDI). As ondas senoidais podem ser ‘clipadas’ (distorcidas) simetricamente com o valor da amplitude de *clipping* variando nos intervalos [0...1] e [-1...0]. Estes valores limites são também determinados pelo performer computacional através da interface de controle do sistema.

Na segunda etapa do processo, os sons sintetizados são adicionados ao sinal do contrabaixo e o sinal resultante passa pela modulação de um *phaser*. A frequência de modulação do *phaser* também é um parâmetro de controle que varia no intervalo [0...10] Hz. O esquema geral deste móbile é apresentado na Fig. 21.



**Fig. 21** - Digrama do Móbile de síntese aditiva modulada por Phaser. Setas tracejadas indicam o fluxo do sinal de áudio e setas contínuas indicam os controles paramétricos em cada etapa.

O sinal de saída produzido por esse móbile cria um contraste com os demais móbiles, pois os sons são poucos ruidosos e o som do contrabaixo não é distorcido. Em outra palavras, cria-se um sinal com características senoidais que contrasta grandemente com os sons predominantemente ruidosos gerados pelos móbiles anteriores.

## 5.2.4 Sintetizador Não-linear via Equação *Latoocarfian*

O quarto módulo engloba o processo de síntese por equação não-linear descrita no **Capítulo 4**. Como descrito anteriormente, há uma tabela que contém as formas de onda geradas pela equação *Latoocarfian* que serve como *look-up table* para quatro algoritmos de síntese sonora, distribuídos dois para cada canal do sistema estéreo. O botão *single-change* na interface gráfica (ver Fig. 22) dispara uma única alteração na forma de onda, ou seja, uma nova sequência de iterações numéricas da equação gera uma nova órbita que é amostrada como forma de onda. Antes das iterações da equação serem disparadas seus parâmetros são alterados aleatoriamente dentro do âmbito estipulado por Pickover (1994) (ver **Capítulo 4**). A Fig. 22 apresenta a interface de controle do sintetizador não-linear.

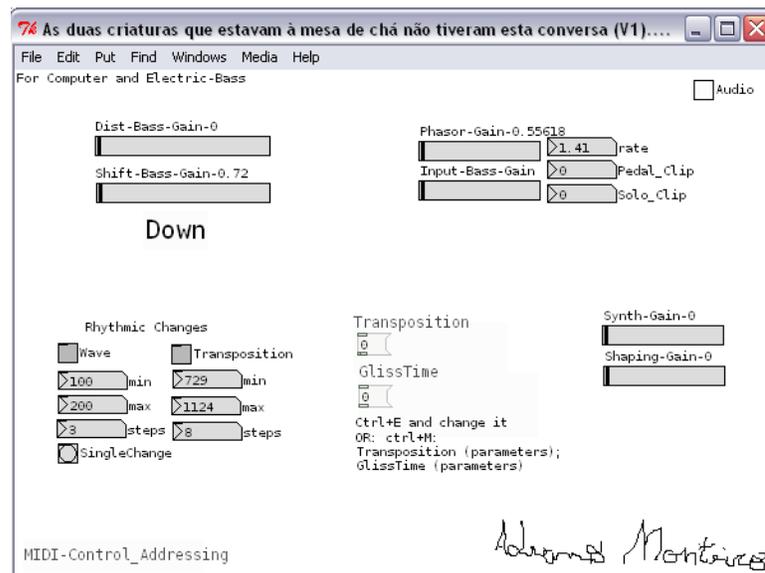
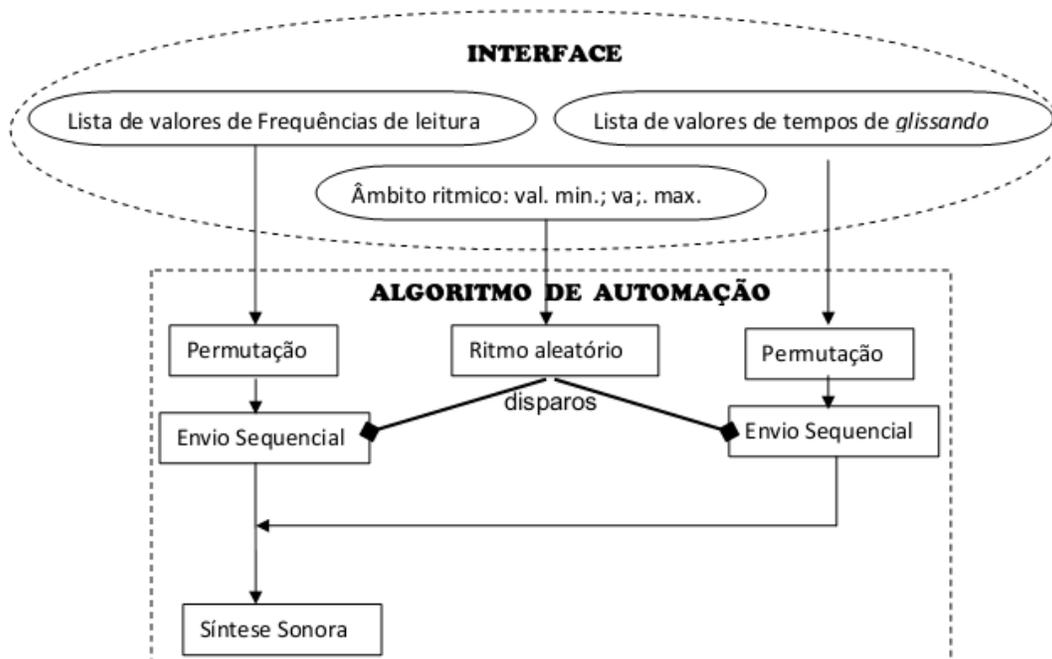


Fig. 22 - GUI de *As duas criaturas que estavam à mesa de chá não tiveram esta conversa*

As listas com valores paramétricos de frequência de leitura da *lookup table* e de tempo de *glissando* (i.e. interpolação entre os valores da lista anterior) são inseridos através da interface gráfica (vide parâmetros *Transposition* e *GlissTime* na Fig. 22). Em seguida, as listas são enviadas para os algoritmos que as permutam e enviam sequencialmente os valores contidos nelas aos parâmetros de síntese. Esse algoritmo é replicado para cada um dos quatro sintetizadores. A transmissão dos parâmetros de síntese é controlada por algoritmos que geram disparos independentemente e em intervalos com diferentes durações, que são determinados por

geradores de números aleatórios. O esquema de automação dos parâmetros de síntese é mostrado na Fig. 23.



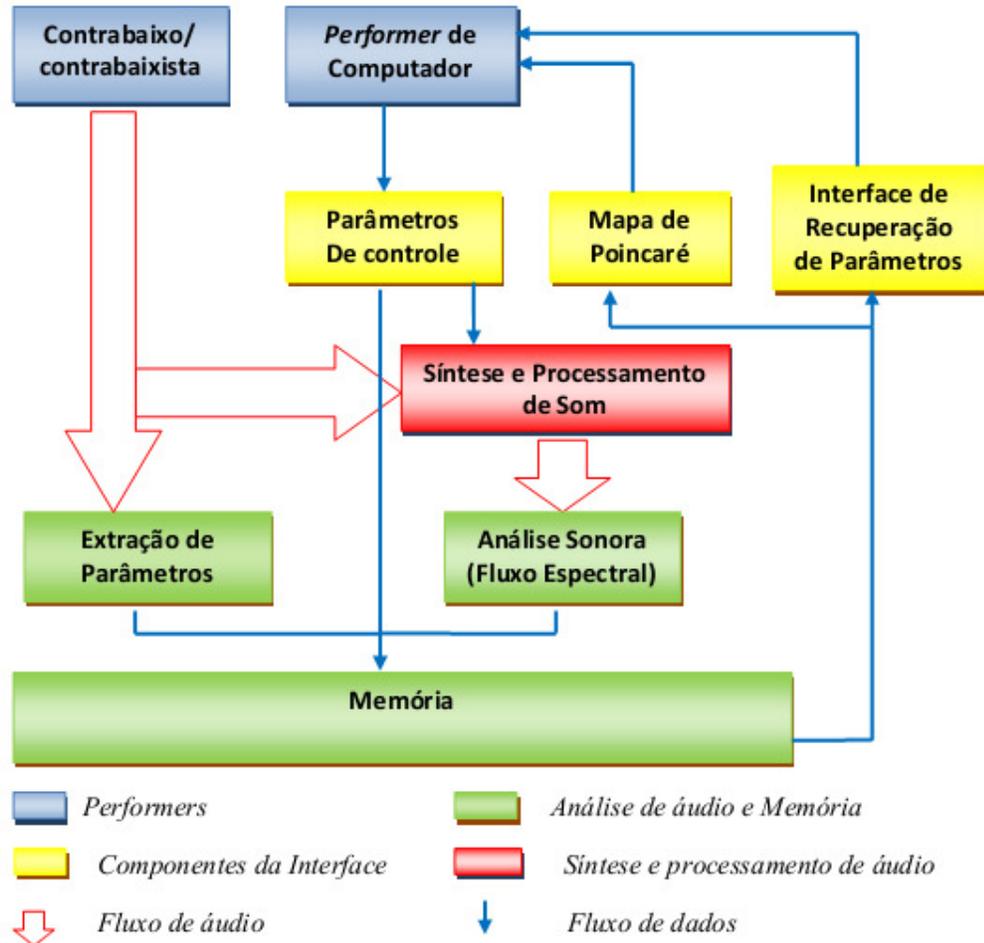
**Fig. 23** - Esquema do algoritmo de automação da permutação dos valores de frequência de leitura da forma de onda.

Os valores que definem o âmbito de geração de números aleatórios também são parâmetros controlados via interface. Esse algoritmo que gera disparos ritmados também é aplicado para automação da alteração de forma de onda. Na interface foram disponibilizadas as possibilidades de ativar ou desativar ambos os geradores de disparos ritmados. Por fim, a forma de onda gerada pela equação Latoocarfian também é utilizada como função de transferência em um processo de *waveshaping* utilizado para distorcer o som do contrabaixo.

Três performances de *As duas criaturas que estavam à mesa de chá não Tiveram esta conversa* podem ser escutadas nas Faixas 11, 12 e 13 do CD Apêndice A deste trabalho. As Faixas 11 e 12 foram gravadas em estúdio sem edição e a Faixa 13 tem edição de Mário Del Nunzio e é a Faixa que integra o CD “CCJ Experimental”.

### 5.2.5 Recuperação e Memória Computacional da Performance

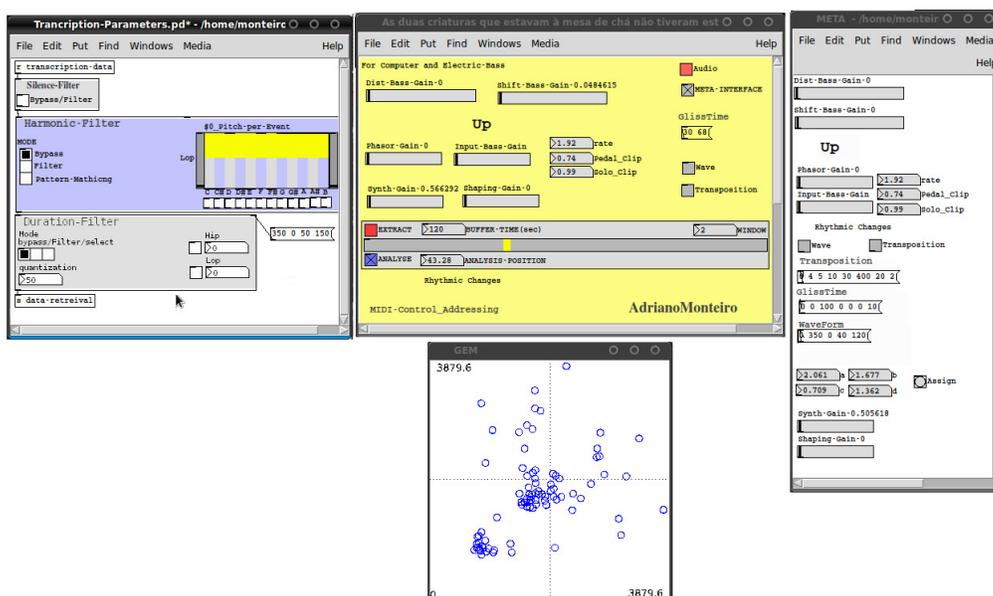
Nesta seção descrevemos a aplicação dos métodos de recuperação da informação musical descritos no **Capítulo 3** em uma segunda versão de *As Duas Criaturas que Estavam a Mesa de Chá não Tiveram esta Conversa*. Utilizamos esses métodos para construção de um sistema que provê uma memória computacional de alguns parâmetros musicais extraídos durante a *performance* da peça. Este desenvolvimento foi apresentado no 13º Simpósio Brasileiro de Computação Musical (MONTEIRO, MANZOLLI 2011c) e Convenção sobre Pure Data (PDcon) 2011 (MONTEIRO E MANZOLLI 2011b). O diagrama na Fig. 24 mostra a arquitetura do sistema com os métodos implementados.



**Fig. 24** - Esquema do novo sistema de *performance* com inclusão dos algoritmos para recuperação de informação musical e memória da performance

A análise do áudio digital gerado pelo método de síntese, assim como pelo gesto do contrabaixo/contrabaixista e as informações paramétricas de controle da interface são gravados durante a *performance* na mesma taxa de amostragem. O instrumentista computacional pode acessar segmentos da memória gravada que são definidos por dois parâmetros: a) um intervalo de tempo dado em segundos e b) um ponto na memória da gravação que é o centro do segmento. Em seguida, um gráfico com um mapa de Poincaré é associado ao segmento de memória selecionado, de acordo ao procedimento especificado no **Capítulo 3**, seção 3.3 (e respectivas subseções) que associa a função extraída pelo descritor Fluxo Espectral com o mapa. Este gráfico é plotado numa janela gráfica na interface computacional para representar as características do segmento selecionado. Simultaneamente, uma terceira janela exibe os parâmetros de controle associados ao mesmo segmento. Esta nova janela compartilha da mesma configuração da janela principal, porém seus ícones de interface (e.g. *sliders*, botões, caixas de números, etc) não transmitem informações paramétricas para os algoritmos geradores de som, apenas exibem os valores dos parâmetros selecionados anteriormente pelo músico — as únicas exceções são os parâmetros da equação *Latoocarfican*, que não possuem interface na versão inicial e podem ser reatribuídos pela nova interface. Uma quarta janela apresenta a informação de parâmetros musicais recuperada do som do contrabaixo de acordo aos procedimentos de recuperação da informação musical especificados no **Capítulo 3**, seções 3.1 e 3.2 (e suas respectivas subseções). Todos esses dados podem ser manipulados pelo *performer* computacional de acordo com os algoritmos presentes nesta interface e enviados como parâmetros de controle do algoritmo de síntese sonora.

A Fig. 25 apresenta a interface gráfica do novo sistema na qual a janela centro é a interface principal herdada da versão anterior. As janelas esquerda e direita são respectivamente: a interface de manipulação de dados extraídos do som do contrabaixo e uma interface apenas para exibição dos valores paramétricos recuperados da memória do sistema. A janela inferior é o mapa de Poincaré que mostra os dados de análise espectral de um segmento da função de Fluxo Espectral gravada em tempo real. Este segmento é relacionado ao trecho de memória selecionado.



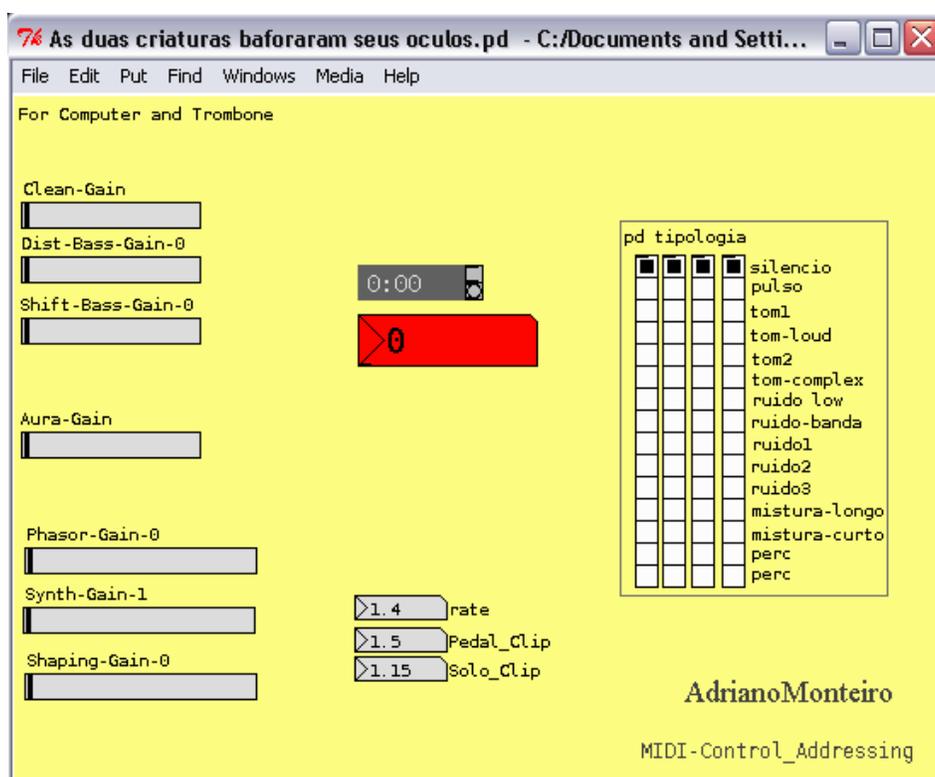
**Fig. 25** - Interface de *As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram Esta Conversa* com adição do sistema de análise, memória e recuperação das informações da performance.

### 5.3 *As Duas Criaturas Baforaram seus Óculos*

Esta primeira variação algorítmica, como já foi apontado acima, foi criada em parceria com o trombonista Pedro Faria. Ela foi apresentada no 13º Simpósio Brasileiro de Computação Musical, em setembro de 2011 na cidade de Vitória. A sua instrumentação é composta por computador e trombone com surdina contendo um microfone interno. A surdina tem a função de bloquear a propagação acústica do som do trombone e capturar o sinal de áudio através deste microfone interno que é ligado ao computador. Se necessário, pede-se um microfone dinâmico ligado diretamente à mesa de som (sem passar pelo computador) para balancear a sonoridade do trombone com a síntese digital do computador em alguns momentos da obra. Ou seja, o conjunto trombone e surdina foi utilizado na segunda obra com a função de prover material sonoro para processamento computacional como já utilizado anteriormente com o contrabaixo na primeira obra.

Ao contrário do primeiro Estudo, empregamos em *As Duas Criaturas Baforaram seus Óculos* um método de composição determinista. Nesta segunda obra os móveis de processamento sonoro serviram como bases sobre as quais organizamos estruturas algorítmicas para manipulação de seus parâmetros de controle. Essas estruturas foram encadeadas sequencialmente. O processo de criação iniciou com a

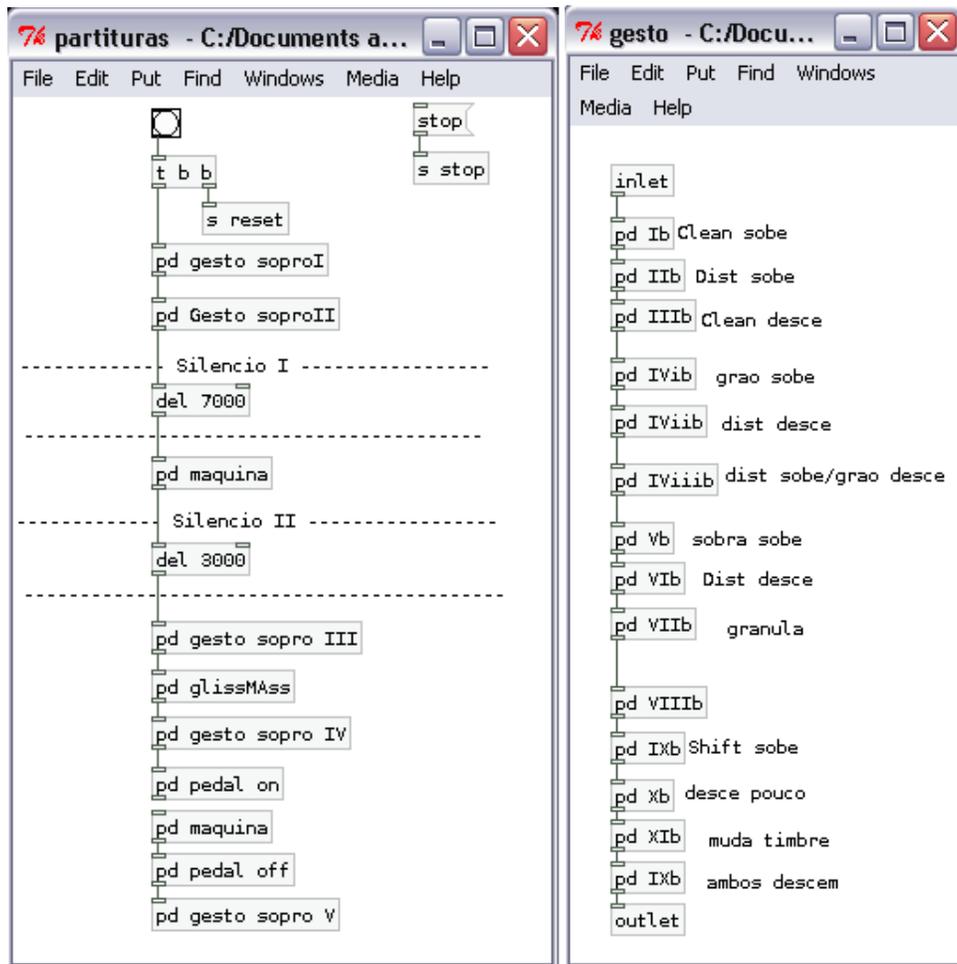
realização de um *ateliê* de improvisação em que escolhemos combinações entre sons digitais produzidos pela síntese não-linear e sonoridades do trombone. Primeiro escolhemos algumas formas de onda geradas pela equação *Latoocarfian* (vide seção 4.2.). A escolha foi feita a partir da Escuta das sonoridades resultantes da síntese inicializada randomicamente. Selecionamos 15 sonoridades e gravamos os parâmetros da equação para que as correspondentes formas de onda pudessem ser retomadas durante a performance. Adicionamos um ícone na interface (vide Fig. 26) que contém botões independentes para cada módulo de síntese (mantivemos os quatro algoritmos de síntese do primeiro Estudo) e que disparam a transformação das formas de onda de acordo com os parâmetros pré-gravados.



**Fig. 26** - Interface de *As Duas Criaturas Baforaram seus Óculos*

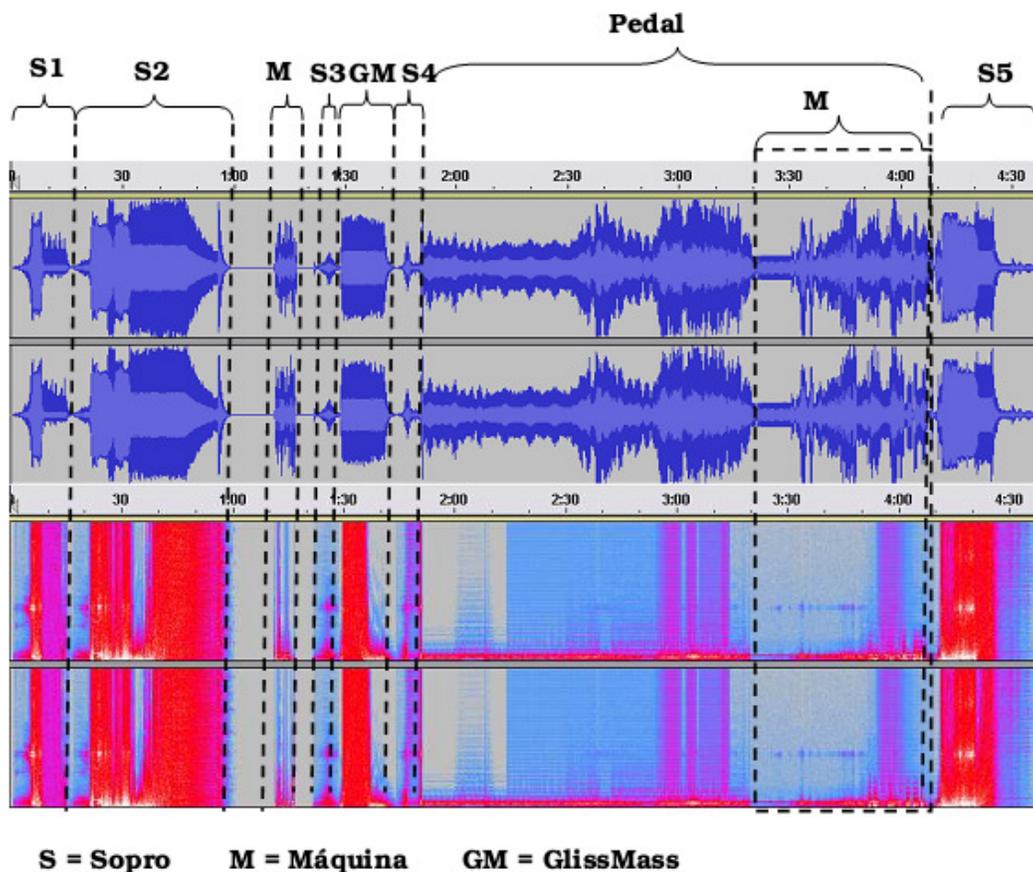
Após fixar o conjunto de parâmetros de síntese e as sonoridades do trombone, pudemos então organizar a composição em macroestruturas que concatenam as alterações paramétricas para resultar no que denominaremos de *elementos sonoros*. Desta forma, construímos uma espécie de partitura sequencial que pode ser vista à esquerda na Fig. 27. Cada macroestrutura é uma *patch* em Pure Data que contém uma sequência de subestruturas ou *sub-patches* que correspondem a etapas sequenciais (vide

*sub-patches* à direita da Fig. 27). Cada uma delas dispara alterações paramétricas nos módulos de síntese e processamento sonoro simultaneamente.



**Fig. 27-** À esquerda o fluxo de macroestruturas que definem a sequência de elementos da peça. À direita a construção interna da macroestrutura “Gesto sopraII” e suas subestruturas que descrevem sequências de alterações paramétricas nos algoritmos de síntese e processamento sonoro.

A Fig. 28 apresenta a análise espectral e segmentação de uma gravação da performance de *As Duas Criaturas Baforaram seus Óculos* que pode ser escutada na Faixa 14 do CD Apêndice A. Os segmentos correspondem aos sons gerados pelas respectivas macroestruturas ou seções da Fig. 27. Nesta ilustração as seções “gesto sopra I”, “gesto sopra II” etc. são designados apenas como ‘*sopra 1*’, ‘*sopra 2*’ etc. respectivamente.



**Fig. 28** - Análise e segmentação de uma performance de *As Duas Criaturas Baforaram Seus Óculos* de acordo com os tipos de estruturas algorítmicas que a compõe. Acima as formas de onda dos dois canais resultante da gravação e abaixo os respectivos espectrogramas.

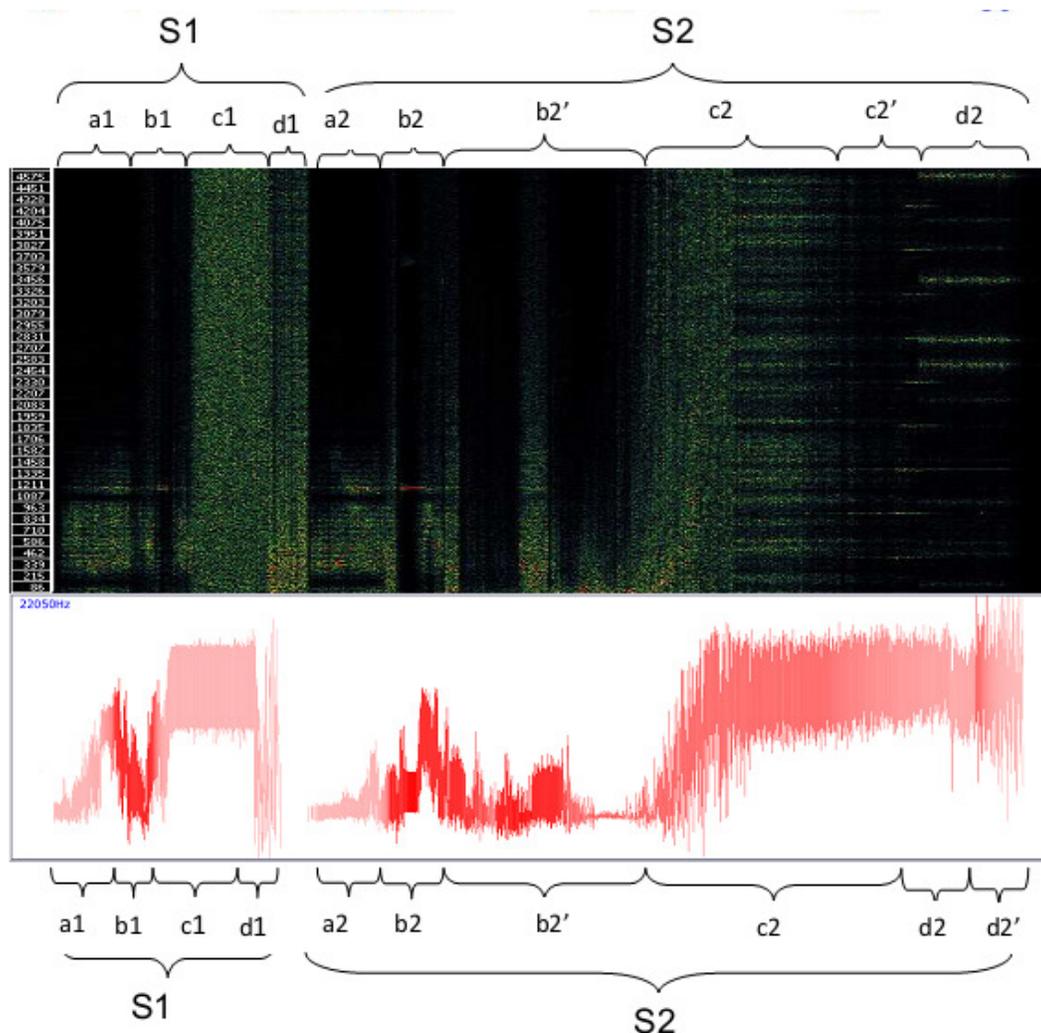
Os elementos nomeados de **Sopro (S)** resultam em estruturas sonoras em que o trombone (com ou sem processamento) e os algoritmos de síntese se articulam como partes de mesmo elemento sonoro. A construção básica é o som resultante do ato de soprar no trombone sem emitir nota, que é gradualmente saturado pela distorção e em seguida altera-se o timbre da massa sonora resultante, justapondo timbres semelhantes resultante de outros processos de síntese sonora ou aplicando novas transformações ao som de sopro. Em geral esse gesto descreve sonoramente movimentos em arco ou semi-arco. As trajetórias das massas sonoras descrevem contornos dos quais os parâmetros de articulação musical são: a frequência centro, a largura de banda (ou saturação do espetro), e a distribuição espectral — e consequentemente a característica timbrística.

A Fig. 29 ilustra através de dois tipos de análise espectral o trecho que abrange os elementos **S1** e **S2** e que pode ser escutado na Faixa 15 do CD Apêndice A.

O gráfico superior é um espectrograma que mostra apenas os picos de energia do espectro e que o *dégradé* verde-amarelo-vermelho representa, nessa ordem, aumento de energia. O gráfico inferior apresenta a representação de uma análise de áudio em que a posição centro no eixo horizontal das regiões avermelhadas representa o centro de energia do espectro (centróide espectral) e a largura vertical representa a média do desvio de energia em relação à posição central (desvio padrão espectral) que é correspondente à média da largura de banda da distribuição de energia no espectro<sup>27</sup>. As regiões de maior intensidade do tom avermelhado indicam uma média de energia (RMS) mais alta, enquanto regiões mais claras (mais próximas da cor branca) representam regiões com menos energia. Esse gráfico é composto de consecutivos traços avermelhados horizontais e foi gerado pela combinação das funções extraídas pelos Descritores Acústicos: Centróide Espectral (**Anexo A**, seção A.2.1.1), associado à posição central dos traços vermelhos; Desvio Padrão Espectral (**Anexo A**, seção A.2.1.2) associado ao tamanho vertical de cada traço; e RMS (**Capítulo 2**, seção 2.1.4) associado à saturação da cor vermelha.

---

<sup>27</sup> Esse gráfico foi baseado no exemplo mostrado pelo pesquisador Mikhail Malt durante uma palestra no Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música (ANPPOM) em agosto de 2011.



**Fig. 29** - Análises espectrais dos elementos Sopro 1 e Sopro 2 referentes a segmentação da Fig. 28. Acima o espectrograma com seleção dos picos de energia e abaixo um gráfico que representa o perfil da frequência centro da distribuição de energia espectral e do desvio padrão da distribuição energética no espectro.

Nota-se que os trechos **S1** e **S2** possuem organizações similares. Em ambos, os inícios **a1** e **a2** representam o som de sopro no trombone que gradualmente é saturado até chegar em **b1** e **b2** com a distorção do módulo descrito na seção 5.2.1. Em **S2** há uma expansão desse trecho, **b2'**, que foi construída pela permutação entre a distorção e uma textura grave e ritmada gerada pelo módulo de síntese. Em seguida, nos trechos **c1** e **c2** a energia decai, a largura de banda aumenta e a distribuição espectral muda (mudança do timbre da distorção para o som gerado pelo algoritmo de síntese), no elemento **S2** a massa é gradualmente conduzida para região aguda. Em **d1** e **d2** há uma segunda mudança na distribuição espectral. **d2** é subdividido em dois outros trechos: o

primeiro é caracterizado pela entrada do *pitch-shifter* no agudo que é mais visível no espectrograma superior (vide seção 5.2.3); e a segunda, **d2'**, caracteriza-se pela mudança de timbre do algoritmo de síntese, assim como em **d1**.

A estrutura que descrevemos para **S1** e **S2** são também as bases para a organização dos demais elementos **S**. É importante notar que a divisão analítica feita na Fig. 29 não é a mesma da parte direita da Fig. 27. A primeira é uma análise dos componentes de organização do resultado sonoro e a outra representa o mecanismo algorítmico que gera este resultado, e não há necessariamente uma correspondência direta (de um para um) entre as divisões.

O elemento **Máquina** da Fig. 28 é uma massa sonora formada por rápidas permutações de frequências e timbres que são entrecortadas de silêncios curtos. Essa sonoridade resulta do funcionamento do algoritmo de síntese. No primeiro momento esse elemento aparece isolado e, mais ao fim da peça, aparece sobreposto ao elemento **Pedal**.

O elemento **GlissMass** compartilha da mesma morfologia que os elementos **Sopro** e foi concebido como uma variação dele, derivado da seção **c2** na Fig. 29 em que a massa sonora é deslocada continuamente para a região aguda. Esse elemento difere-se dos elementos **Sopro** por ser gerado apenas pelo algoritmo de síntese. Em suma, ele é uma massa de sons sintetizados que realiza um glissando do grave para o agudo e retorna ao grave.

Finalmente, **Pedal** é o resultado sonoro gerado pelo módulo *phaser* (vide seção 5.2.3) com momentos de sobreposição do som do trombone. Nesse segmento foi estabelecida uma improvisação. O performer computacional improvisa com os parâmetros do módulo *phaser*: “rate”, Pedal\_Clip” e “Solo\_Clip” (vide interface na Fig. 26) e o trombonista improvisa criando efeitos de modulação oscilatórias de amplitude e frequência sobre uma nota sustentada. Ou seja, ambos improvisam com frequências de modulações e saturação espectral.

Concluindo, *As Duas Criaturas Baforaram Seus Óculos* é o segundo estudo de *As Duas Criaturas*. Nele foi adotada uma composição determinista que impõe uma forma fixa e uma condução temporal linear. A propriedade instrumental do algoritmo foi muito reduzida, com exceção no elemento **Pedal**, não há necessidade de controle ou interpretação do algoritmo por um performer computacional. Quanto à interação com o instrumento acústico (trombone), existem dois momentos distintos: o primeiro caracteriza os elementos denominados **Sopro** na Fig. 28 (e “gesto sopra” na Fig. 27),

neles, trombone e computador operam numa relação de dependência para emissão sonora, o que significa que o som de sopro emitido pelo trombone é processado pelo computador para gerar uma nova sonoridade. As ações do computador e do trombonista resultam em um mesmo elemento sonoro. O segundo momento caracteriza o elemento **Pedal** na Fig. 27 e Fig. 28, nele, trombone e computador são fontes de emissão sonora independentes. A interação se dá pela percepção e manipulação paramétrica por parte dos instrumentistas que agem sob uma pré-convenção a respeito das morfologias sonoras em jogo na *performance* improvisada desse trecho.

O método que adotamos para variação do algoritmo de *As Duas Criaturas que Estavam a Mesa de Chá Não Tiveram esta Conversa* foi o encapsulamento dos algoritmos. Circunscrevemos os parâmetros do algoritmo original em macroestruturas ou seções (*Patches*) que organizam os parâmetros de forma muito específica e detalhada, criando relações que não seriam factíveis pelas possibilidades da estrutura algorítmica e da interface originais, e com isso propiciamos uma nova camada de estruturação e controle composicional gerando uma nova organização sonora musical.

#### **5.4 Não Diziam-nos que iam Contar sobre nosso Passado**

Este é o terceiro Estudo da série *As Duas Criaturas* e é uma peça composta para computador solista conectado em um sistema de saída de áudio quadrifônico. Este estudo teve sua primeira apresentação pública em novembro de 2011, no Departamento de Música da Universidade Estadual de Campinas, cidade de Campinas, como programa da série de concertos Nova Música Eletroacústica (NME). Assim como em *As Duas Criaturas Baforaram seus Óculos* ele possui uma estrutura temporal sequencial. Dois preceitos guiaram sua criação: 1) conceber e explorar um tipo específico de interpretação musical para esta música composta e tocada através de um computador; 2) aprofundar o uso do método de síntese por equações não-lineares apresentado na subseção 5.2.4 e no **Capítulo 4**.

Primeiramente, elaboramos novas abordagens para gerar material sonoro através da equação *Latoocarfian*. Observamos os resultados de três diferentes manipulações da equação, com intuito de detectar comportamentos gerais que fossem proveitosos aos nossos interesses de aplicação musical:

- a) Interpolações lineares entre valores paramétricos durante as iterações da equação. Realizamos interpolação de 1 parâmetro apenas e a de 2

parâmetros simultaneamente. Definimos interpolação nesse contexto como a soma iterada de um fator numérico ao valor paramétrico inicial e que gera como resultados números entre o parâmetro inicial e o final em uma sucessão linear.

Observações: As interpolações resultaram em transições graduais entre uma tipologia e outra das formas de onda geradas pelo sistema (listadas na seção 4.2). Também foram observadas mudanças graduais e direcionais apenas no que seria relativo à amplitude da onda.

- b) Observamos a relação entre as duas variáveis da equação durante uma mesma série de iterações (foram observados casos com os parâmetros fixos ou com interpolação linear entre valores paramétricos).

Observações: Os resultados gerados pelas duas variáveis da equação *Latoocarfian* são similares quanto aos tipos de orbitas que descrevem. Todavia, detectamos algumas diferenças, como por exemplo, atrasos entre as funções; diferenças nos valores das órbitas, mas que mantém uma mesma tipologia (segundo as descritas na seção 4.2).

- c) Associação das duas variáveis da equação para geração de uma mesma forma de onda. Utilizamos o algoritmo apresentado no Pseudocódigo 2 (seção 4.2) para mapear os valores resultantes da equação como forma de onda, sendo que uma das variáveis é mapeada para os valores da função de onda e a outra como parâmetro de número de pontos de interpolação. Este procedimento se aproxima do método utilizado por Xenakis em GENDY, no qual utiliza duas variáveis uma para controlar a distribuição dos pontos e outra para controlar o número de pontos interpolados (SERRA, 1993).

Observações: O comportamento das funções se mantém, mas a variação de interpolação resulta em sonoridades menos artificiais quando as formas de onda geradas são sonorizadas pelo algoritmo de síntese.

Em seguida partimos das observações feitas para guiar o processo de composição musical. Estabelecemos quatro tabelas nas quais os resultados de iterações consecutivas das equações não-lineares, sob os diferentes modos de manipulação paramétrica (conforme os listados acima), podem ser gravados sequencialmente.

Utilizamos sempre o último estado gerado em cada tabela para inicializar as variáveis da nova seção de iterações obtendo com isso uma transição gradual entre os estados consecutivos. As tabelas contendo as formas de onda geradas foram utilizadas como *look-up tables* para a síntese sonora na qual empregamos o método descrito como *Dynamic Wavetable* (DAMIANI et al, 1999; MANZOLLI 1993a, 1993b). Esse método é caracterizado por utilizar para a síntese de som, um trecho da tabela que contém a forma de onda (e que chamaremos de *janela*), e que se desloca ao longo da tabela. Em outras palavras, os pontos de início e fim de leitura da forma de onda estabelecem uma relação diferencial fixa e menor que o tamanho da tabela e se movem em paralelo ao longo dela. Esse método visa propiciar uma alteração constante da forma de onda utilizada para síntese, e por conseqüência, uma constante alteração da sonoridade, timbre e conteúdo espectral do som gerado. Por fim, para cada tabela foram associados dois algoritmos de síntese (igualmente aos outros estudos da série *As Duas Criaturas*) com parâmetros de frequências de leitura independentes.

Esse processo foi aplicado para cada um dos quatro canais de áudio. Apesar da configuração em quatro canais, não há nesta peça a programação de algoritmos para espacializações sonoras. Cada caixa de som foi usada como um instrumento de emissão sonora local e independente, relacionado a uma das tabelas com as formas de onda.

Quanto ao sistema de interpretação musical, elaboramos um algoritmo para controle de parâmetros do processo de *dynamic wavetable* pelo *performer* computacional. As janelas de leitura da forma de onda de cada tabela se deslocam paralelamente e com a mesma velocidade (que é um parâmetro de controle propiciado ao *performer*). Chamaremos esse ponto de leitura que progride linearmente de *ponto guia*. O *performer* computacional pode interferir na linearidade do ponto de leitura independentemente para cada canal e através de dois modos:

- Variações locais da velocidade de leitura que podem ser controladas gestualmente gerando efeitos de *scratch* sobre a forma de onda,
- Saltos randômicos e ritmados do ponto de leitura, em que o ritmo dos intervalos temporais entre saltos é controlado por uma tabela que contém órbitas geradas por uma equação *Latoocarfian*.

Ambos os modos de interferência sobre a progressão linear do ponto de leitura da tabela são delimitados por uma região que tem o *ponto guia* como centro, e por isso a região de interferência também se desloca linearmente, avançando para o final da tabela. A interferência operada pelo *performer* se sobrepõe ao movimento constante e linear do ponto guia. Elaboramos esse método para que haja uma condução linear e determinística da peça e habilitamos a interferência na leitura da forma de onda como parâmetro aberto para a interpretação.

Por fim, os parâmetros de frequência de leitura nos algoritmos de síntese também foram organizados durante a etapa de composição. Mapeamos os valores resultantes de iterações da equação *Latoocarfian* em valores de frequências dentro de âmbitos frequenciais determinado por nós a cada caso, conforme a necessidade composicional (intuitiva) requeresse. Usamos os conjuntos de valores de frequência obtidos como parâmetros dos algoritmos de síntese, mas não fomos ortodoxos na aplicação deste sistema, pois nos demos liberdade para alterar os dados gerados conforme as necessidades musicais determinadas por critérios subjetivos e embasado na percepção auditiva do resultado.

A Fig. 30 apresenta a interface de *Não Diziam-nos que iam Contar sobre nosso Passado*. As tabelas superiores armazenam as formas de ondas para síntese. Os elementos inferiores são controles paramétricos para interpretação do algoritmo. Os quatro *sliders* verticais à esquerda são controles de intensidade dos quatro canais de áudio (i.e. resultado da síntese por leitura de cada tabela). Os quadrados ao centro são os módulos de controle para gerar interferência na linearidade de leitura das tabelas, conforme descrito acima. Em nossas interpretações do algoritmo, associamos diversos parâmetros de controle com interfaces MIDI que contém *sliders* e *knobs* que propiciaram o controle de diversos parâmetros simultaneamente e com maior agilidade gestual.

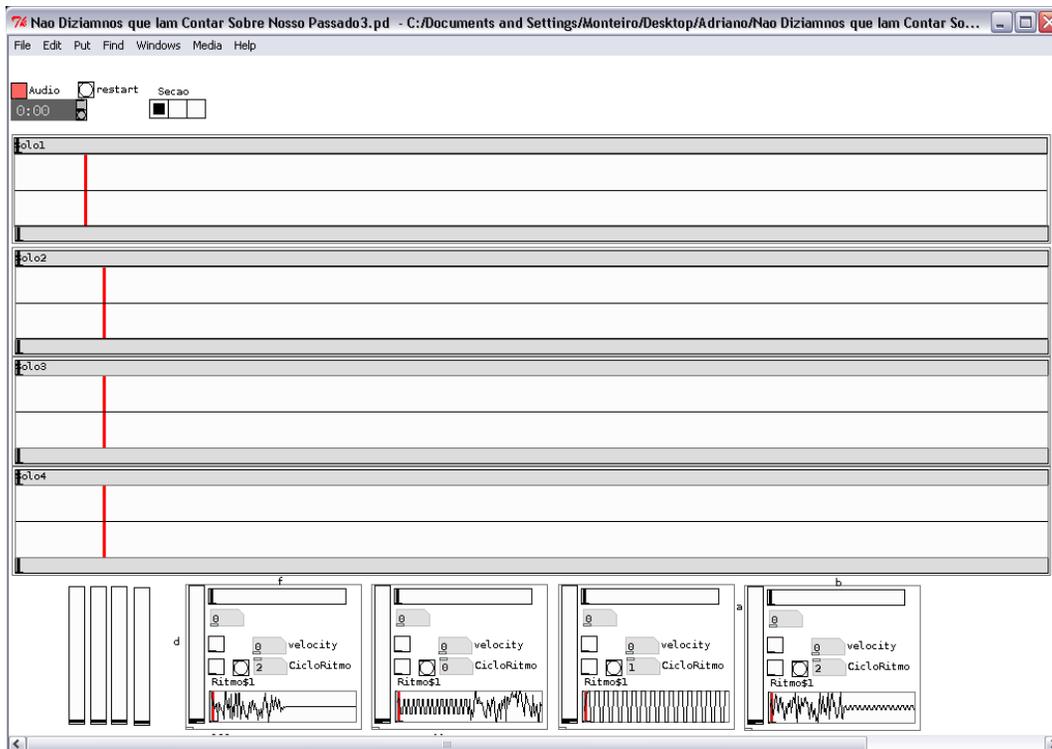
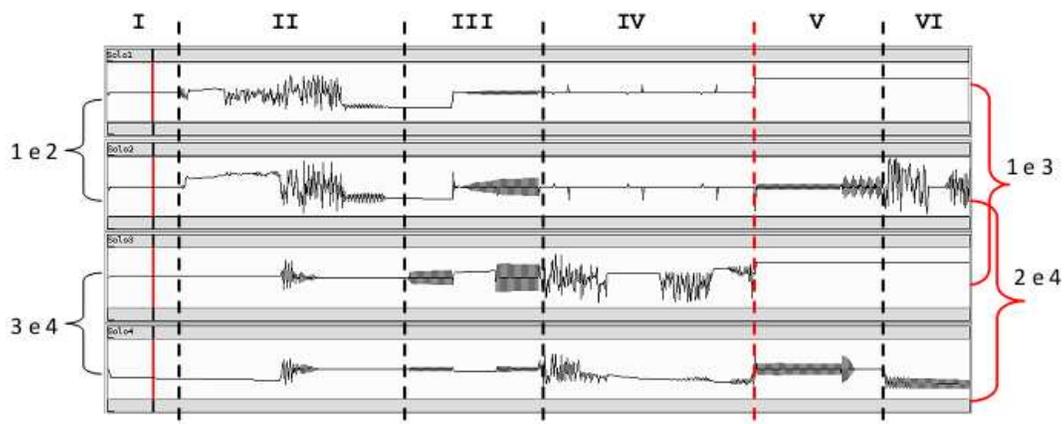


Fig. 30 - Interface de *Não Diziam-nos que iam Contar sobre nosso Passado*.

Após estabelecido o ‘esqueleto’ do algoritmo, a próxima etapa da composição foi o processo de geração e seleção de sequências de formas de onda para cada canal. A peça é dividida em três seções em que abordamos os métodos de síntese de modos distintos.

#### 5.4.1 Primeira Seção

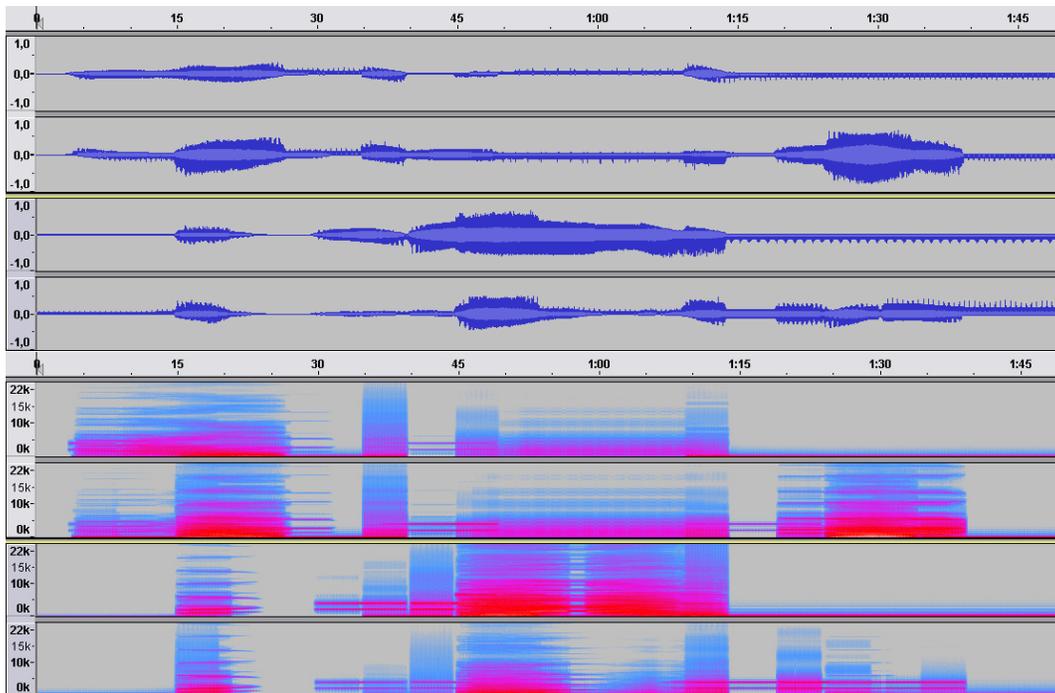
A Fig. 31 apresenta a primeira seção. Nela associamos pares de tabelas com as duas variáveis de uma mesma equação e utilizamos a interpolação de parâmetros da equação para gerar uma dinamicidade da forma de onda com transições entre diferentes tipologias.



**Fig. 31** - Formas de ondas geradas através da equação *Latoocarfian* para a primeira seção de *Não Diziamos que iam Contar sobre nosso Passado*.

As linhas tracejadas na Fig. 31 representam divisões de segmentos onde novos parâmetros da equação foram aplicados para gerar continuidade da forma de onda. A linha tracejada vermelha marca a mudança de associação entre pares de canais (e conseqüentemente pares de caixas de som). O segmento anterior à linha vermelha possui pares formados entre os canais 1-2 e 3-4; e o segmento posterior entre 1-3 e 2-4. A sonoridade geral oscila entre massas sonoras complexas em contínua transformação (e.g todos os canais no segmento II; canais 3 e 4 segmento IV; canal 2 no segmento VI) e sons tonais mais estáticos mas também com alguma variação interna principalmente quanto o envelope de amplitude de onda (e.g. todos os canais segmento III; canais 2 e 4 segmento V; canal 4, segmento 4). Há a presença da sonoridade de pulsos iterados (pela leitura da janela sobre a forma de onda ser iterada) que pode ser visto no segmento III canal 2 e segmento IV canais 1 e 2.

É importante explicitar que as formas de onda inseridas nas tabelas não são a representação dos resultados sonoros. O som de cada canal é resultante do processo de *dynamic wavetable* sobre cada tabela da Fig. 31, e por tanto, varia com a velocidade da janela de leitura, com as frequências de leitura em cada algoritmo de síntese (dois por tabela), e com a interferência no ponto de leitura operada pelo intérprete computacional. A somatória desses parâmetros gera o som resultante para cada tabela/canal de áudio. A Fig. 32 apresenta a forma de onda da gravação da seção e seu respectivo espectro sonoro. O áudio relativo a essa gravação pode ser escutado nas Faixas 16, 17 e 18 do CD Apêndice A. As Faixas 16 e 17 contêm um arquivo estéreo com os canais 1-2 e 3-4 respectivamente separando os canais do original quadrifônico. A Faixa 18 contém uma mixagem em estéreo do resultado dos quatro canais.



**Fig. 32** - Acima as formas de onda dos quatro canais resultantes da gravação da primeira seção de *Não Diziam-nos que iam Contar sobre nosso Passado* e abaixo os respectivos espectrogramas.

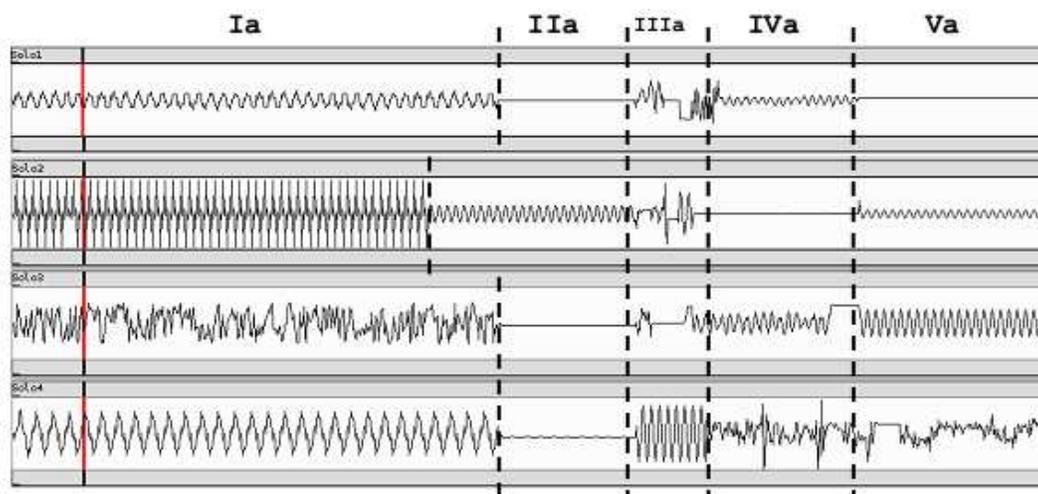
Como pode ser observado na Fig. 32 acima e escutado nas gravações, o resultado musical da seção 1 pode ser descrito como uma sucessão de massas sonoras com alta densidade espectral que alguns momentos convergem para sons tonais (com altura reconhecível).

Quanto às propriedades de interpretação do algoritmo, detectamos que a alta variação das formas de onda obtidas pela interpolação dos parâmetros da equação *Latoocarfian* (Fig. 31) resultaram em pouca abertura para interpretações realizadas pelo performer computacional. Concluímos isso pela observação e prática da interpretação dessa seção através das quais constatamos que as interferências na linearidade da leitura de onda (parâmetros para execução/interpretação do algoritmo) causaram um excesso de complexidade da dinâmica sonora. Para essa seção, apenas alterações na velocidade de leitura comum entre as tabelas (*ponto guia*) e manipulações sobre a intensidade sonora dos canais de áudio foram os parâmetros para sua interpretação.

#### 5.4.2 Segunda Seção

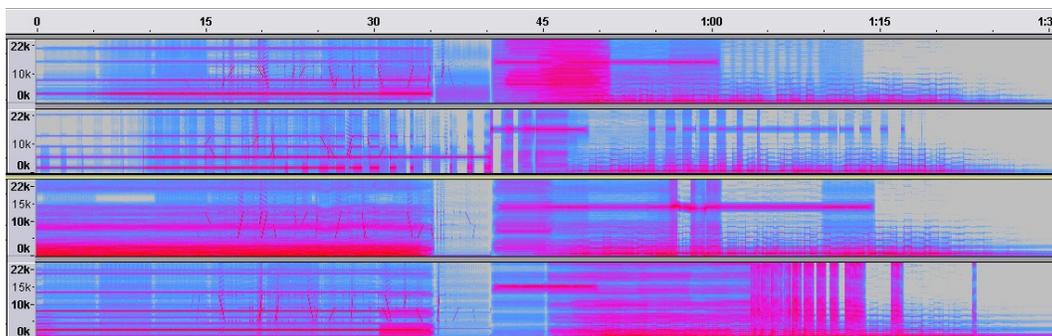
Na segunda seção (Fig. 33) as tabelas foram tratadas separadamente, ou seja, a forma de onda em cada tabela foi gerada independentemente utilizando diferentes instancias do algoritmo que implementa a equação *Latoocarfian*. Por isso cada tabela contém formas

de onda diferentes umas das outras. Suas alterações foram operadas sem interpolações e sincronicamente, com exceção de uma antecipação no canal 2 (de cima para baixo) do segmento IIa.



**Fig. 33-** Formas de ondas geradas através da equação *Latoocarfian* para a segunda seção de *Não Diziam-nos que iam Contar sobre nosso Passado*.

Optamos por não utilizar interpolações entre os parâmetros da equação *Latoocarfian* para otimizar a possibilidade de interpretação (interferência na linearidade da leitura da forma de onda). Os resultados de interpretação foram mais satisfatórios do que na primeira seção, pois há mais invariância do resultado sonoro resultante da leitura linear da tabela pelo algoritmo de síntese, o que abre espaço para ação do intérprete sem que resulte em um excesso de complexidade sonora. A Fig. 34 apresenta o espectrograma resultante da gravação de uma performance da seção 2. O áudio relativo a essa gravação pode ser escutado nas Faixas 19, 20 e 21 do CD Apêndice A. As Faixas 19 e 20 contêm um arquivo estéreo com os canais 1-2 e 3-4 respectivamente separando os canais do original quadrifônico. A Faixa 21 contém uma mixagem em estéreo do resultado dos quatro canais.

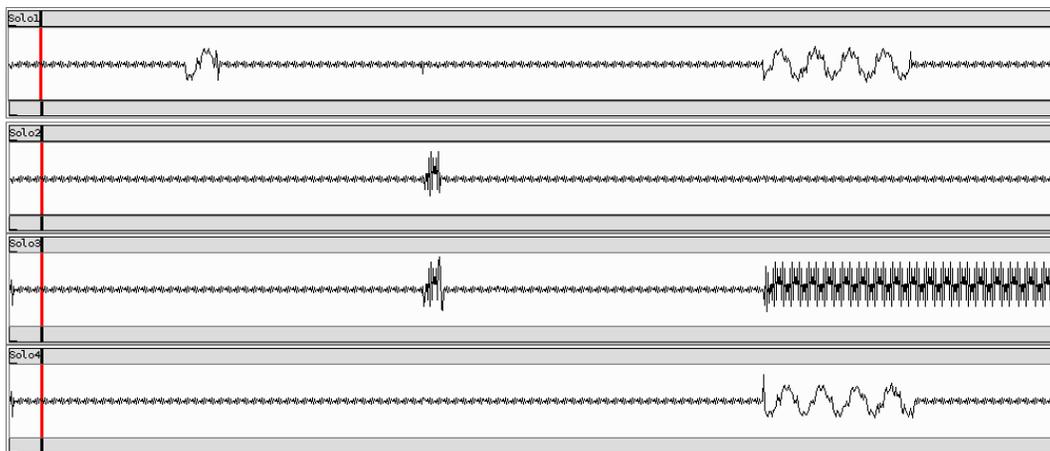


**Fig. 34-** Espectrogramas dos quatro canais extraídos da gravação de uma performance da segunda seção de *Não Diziam-nos que iam Contar sobre nosso Passado*.

Nota-se no espectrograma que as massas sonoras geradas são mais dispersas no espectro (menos densas), pois apresentam picos de energia espaçados. Nessa seção, sobrepomos ao resultado sonoro gerado pela leitura das formas de onda com texturas sonoras geradas pelo algoritmo descrito na seção 5.2.4 e que podem ser observadas no espectrograma nas curtas linhas diagonais na primeira metade do espectrograma. Nota-se que a interpretação gerou variação em relação à estrutura apresentada na Fig. 33. Por exemplo, observa-se no segundo canal da Fig. 34 uma repetição constante de um padrão curto que foi causada por um modo de tocar em *loop* gerado pela aceleração da velocidade de leitura que estava restrita a um pequeno âmbito entorno do *ponto guia* (responsável por manter a leitura linear da forma de onda). Esse mesmo padrão anulou a antecipação da mudança de forma de onda do segundo canal no segmento IIa, e causou a extensão do segmento Va nos segundo e quarto canais, ao fim da peça.

### 5.4.3 Terceira Seção

Na terceira seção, apresentada na Fig. 35, realizamos uma construção bem mais simples que nas seções anteriores. As formas de ondas são geradas sem interpolações paramétricas e há menos mudanças, ou seja, os padrões permanecem por mais tempo. Outra característica é a utilização das mesmas formas de onda para todas as tabelas, o que cria uma estrutura musical de permutação de sonoridades entre os canais de áudio.



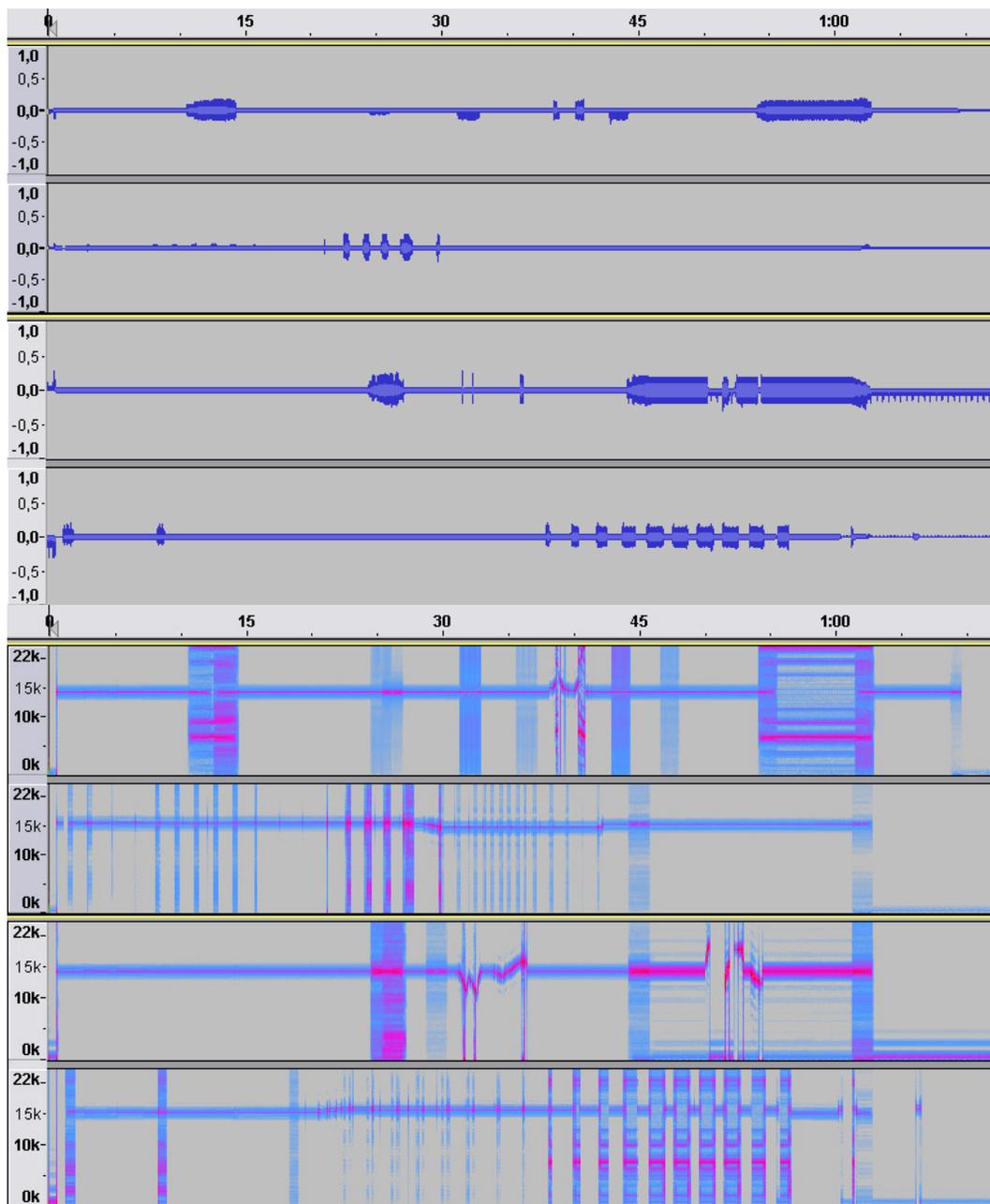
**Fig. 35** - Formas de ondas geradas através da equação *Latoocarfián* para a terceira seção de *Não Diziamos que iam Contar sobre nosso Passado*

Objetivamos com essa abordagem de organização mais simples enfatizar a instrumentalidade do algoritmo e a possibilidade de interpretação da estrutura musical composta sem descaracterizá-la em relação a sua sonoridade e a sua organização temporal. Os resultados foram satisfatórios nesse sentido como pode ser observado na Fig. 36 que contem as formas de ondas gravadas de uma interpretação da seção 3, assim como seus respectivos espectrogramas. O áudio relativo a essa gravação pode ser escutado nas Faixas 22, 23 e 24 do CD Apêndice A. As Faixas 22 e 23 contêm um arquivo estéreo com os canais 1-2 e 3-4 respectivamente separando os canais do original quadrifônico. A Faixa 24 contém uma mixagem em estéreo do resultado dos quatro canais.

Nela é possível identificar as estruturas musicais correspondentes entre as formas de ondas das Fig. 35 e o resultado de sua interpretação nas formas de onda e espectrograma da Fig. 36, assim como os padrões de variação resultantes da interpretação do *performer* computacional. Nota-se que um tipo de variação obtida pelo *performer* foi a repetição de elementos, como por exemplo, a sequência de picos de amplitude do segundo canal da Fig. 36 (entre os segundos 15 e 30) que são derivados do segmento de forma de onda com maior amplitude, na tabela 2 da Fig. 35. Esta repetição foi causada pelos movimentos em *loop* (ou em vai-e-vem) operados sobre o segmento de maior amplitude e através da variação de velocidade de leitura da tabela.

Outra consequência da manipulação gestual da velocidade de leitura foi a transposição de frequência e sua consequente geração de *glissandos*, que podem ser

vistos no terceiro canal do espectrograma, logo após a marca de 30 segundos, e no mesmo canal entre as marcas de 45 segundos e 1 minuto de gravação.



**Fig. 36** - Acima as formas de onda dos quatro canais geradas pela gravação de uma performance da terceira seção de *Não Diziam-nos que iam Contar sobre nosso Passado* e abaixo os respectivos espectrogramas

Foi nesta seção que obtivemos maior êxito quanto ao equilíbrio entre propriedades de composição e instrumentalidade do algoritmo. Porém essa maleabilidade instrumental e interpretativa veio associada a uma maior dificuldade

técnica para operação do algoritmo. Observamos que para obtenção de um resultado musical satisfatório é necessário o estudo e engajamento intelectual sobre as propriedades do algoritmo, assim como habilidade motora mais sofisticada para controle dos parâmetros simultaneamente. Acreditamos que a potencialidade das relações musicais que estão estruturadas nesse estudo ainda não foram completamente exploradas pelas interpretações que realizamos.

## 5.5 Organização Paramétrica dos Sistemas

Nesta subseção ilustramos uma visão geral da organização dos três sistemas desenvolvidos na série de Estudos musicais. Nas **Tabelas 2, 3, 5, e 6**, relacionamos os parâmetros dos sistemas com as respectivas etapas do processamento algorítmico, seja para síntese sonora ou automação de controle. Apontamos quais parâmetros são disponibilizados para o performer como controle do sistema via interfaces gráfica e/ou dispositivos eletrônicos e quais parâmetros foram utilizados para composição de estruturas musicais.

A **Tabela 1** contém a descrição das siglas utilizadas nas **Tabelas 2, 3, 5, e 6** que exemplificam a organização dos sistemas.

**Tabela 1-** Descrição das Siglas utilizadas nas **Tabelas 2, 3, 5 e 6**

<b>Lista de Siglas e suas respectivas descrições organizadas em ordem alfabética.</b>	
<b>Sigla</b>	<b>Descrição</b>
<b>Acr/dcr. Un.Rand.</b>	Acréscimo ou Decréscimo unitário randômico - adiciona um valor randômico positivo ou negativo ao ponto de leitura da tabela, dentro do âmbito de interferência. Resulta em saltos randômicos dentro do âmbito e entorno do ponto de leitura
<b>Acr/dcr.Itr.</b>	Acréscimo ou Decréscimo iterado - itera um valor (positivo ou negativo) sobre o ponto de leitura. Resulta em movimentos lineares dentro do âmbito e entorno do ponto de leitura.
<b>Amb.Interf.</b>	Âmbito de interferência no ponto e leitura da tabela tendo o ponto guia como centro.
<b>Amb.Map</b>	Âmbito de mapeamento para os valores gerados
<b>Amp.</b>	Controle de amplitude do sinal de áudio
<b>Assoc. Tab.</b>	Tipo de associação dos valores resultantes da Equação com as tabelas
<b>Aut.Rit</b>	Automação de disparos rítmicos
<b>Aut.Perm</b>	Automação de permutação de listas
<b>Clip.Ped.</b>	Valor de “clipping” para os osciladores com frequência estática
<b>Clip.Solo</b>	Valor de “clipping” para o oscilador com frequência variável
<b>Comb.P.</b>	Combinações paramétricas
<b>D.man.</b>	Disparo Manual (realizado via interface)
<b>F.Trans.</b>	Fator de transposição
<b>Freq.</b>	Valor de Frequência
<b>Freq. Mod.</b>	Frequência de Modulação
<b>Freq. Osc.</b>	Frequência dos Osciladores

<b>Freq.L.T.</b>	Frequência de Leitura da Tabela (wavetable)
<b>Func.Offset</b>	<i>Offset</i> para atribuição de valores de frequência extraídos da função gerada e enviados aos algoritmos de síntese. Extrai-se uma lista de oito valores (dois para cada par de algoritmos de leitura da forma de onda) a cada disparo da atribuição
<b>Inc.Dyn.Wvt</b>	Valor de incremento que é iterado ao ponto de leitura da tabela para estabelecimento do algoritmo de <i>Dynamic Wavetable</i>
<b>Ini. Var</b>	Inicialização das Variáveis
<b>L/D</b>	Liga ou Desliga
<b>L.Freq</b>	Lista de valores de frequência
<b>L.T.interp</b>	Lista de valores de tempo de interpolação
<b>N.itr.</b>	Número de iterações
<b>N.p.Interp.</b>	Número de pontos de interpolação
<b>N.val.</b>	Número (quantidade) de valores tomados no âmbito escolhido
<b>Param.Eq.</b>	Parâmetros da Equação
<b>Rand.</b>	Randômico (gerador de números aleatórios)
<b>T.interp.</b>	Tempo de interpolação
<b>Tam.Tab</b>	Tamanho da tabela
<b>Val.Ad.Itr.</b>	Valores adicionados, um a cada parâmetro da equação, e iterados simultaneamente a cada iteração da equação. Esse procedimento resulta em interpolações entre estados do sistema, e conseqüentemente entre diferentes formas de onda.
<b>Val.itr.</b>	Valor iterado
<b>Var</b>	Variável

As células amarelas nas **Tabelas 2, 3, 5 e 6** indicam que os parâmetros inclusos (nessas células) são atribuídos para a interface de controle disponibilizada ao *performer*. Âmbitos de valores paramétricos são designados por pares de colchetes que incluem o valor inicial e o final do âmbito separados por reticências, e.g. [0 ... 1 ], quando não há o valor final significa que apenas um valor mínimo para o parâmetro é determinado e não há delimitação para o valor máximo, e.g. [ 0... ].

A **Tabela 2** apresenta a organização do sistema de *As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram esta Conversa*. Ela é dividida horizontalmente em quatro seções relacionadas aos quatro móveis de geração/organização sonora.

A primeira coluna da **Tabela 2** apresenta os móveis; a segunda coluna apresenta os parâmetros de controle de cada procedimento de síntese; a terceira coluna apresenta, para cada parâmetro, uma delimitação paramétrica, ou uma atribuição paramétrica fixa, ou uma alteração do parâmetro segundo os procedimentos algorítmicos de automação de controle, estes últimos, por sua vez, são expostos na quarta coluna. Por fim, a quinta coluna apresenta a segunda camada de parâmetros referentes aos algoritmos de automação.

Conforme é visível na **Tabela 2**, existem dois níveis de controle no sistema de *As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram esta Conversa*: a) o controle ligado diretamente aos parâmetros de síntese relacionados na terceira coluna; e b) o controle ligado aos sistemas de automação do controle dos parâmetros de síntese relacionados na quinta coluna. Ambos são voltados a performance em tempo real e não há o uso de parâmetros do sistema para elaboração previa de estruturas musicais conforme existe nos próximos sistemas que serão mostrados a seguir.

**Tabela 2** - Esquema de organização dos parâmetros e respectivos algoritmos do sistema de *As Duas Criaturas que estavam à Mesa de Chá Não Tiveram Esta Conversa*

Tipo de Processamento	Parâmetros Primários	Primeira Delimitação Paramétrica	Algoritmos de Automação		Segunda Delimitação Paramétrica	
<b>Distorção</b>	Amp.	[ 0...1 ]				
<b>4 Osciladores -&gt; clipagem-&gt; Phaser</b>	Freq. Osc.	204.707				
		204.855				
		250.399				
		[ 0... ]				
	Freq. Mod.	[ 0 ... 10 ]Hz				
	Clip.Ped.	[ 0...1 ]				
	Clip.Solo	[ 0...1 ]				
Amp	[ 0...1 ]					
<b>Pitch-shifter</b>	F.Trans.	0.178764 ou 23.9283				
<b>Síntese via Equações não Lineares</b>	Param.Eq.	Rand.	Aut.Rit.		L/D	
		(D.Man.)			Amb.	
					N.val.	
	Ini. Var.	Rand.				
	N.p.Interp.	4				
	N.itr.	1000				
	Freq.L.T.	Freq.	Aut.Rit .-> Aut.Per.	Aut.Rit.		L/D
		T.interp.		Aut.Per.		Amb.
Amp.	[ 0...1 ]				N.val.	
					L.Freq.	
					L.T.interp	

A **Tabela 3** mostra a arquitetura do sistema de *As Duas Criaturas Baforaram seus Óculos*. Sua estrutura é muito similar com a da **Tabela 2**. Conforme pode ser observado na **Tabela 3**, os móveis de geração sonora são mantidos entre esta peça e a peça relacionada à **Tabela 2**, com a exceção do algoritmo para automação rítmica da mudança de forma de onda, que foi substituído pela possibilidade de escolha entre 15 combinações paramétricas fixas.

Na **Tabela 3**, as células verdes indicam os parâmetros usados na organização dos *elementos sonoros* (que são macroestruturas de concatenação de parâmetros, conforme descritas na subseção 5.3). Seguindo uma contagem temporal linear e progressiva, os algoritmos que geram os elementos sonoros concatenam em um mesmo instante no tempo determinações ou alterações de valores para os parâmetros que estão ressaltados nas células verdes da **Tabela 3**. Ou seja, as macroestruturas (vide subseção 5.3) são formadas por um conjunto de simples determinações paramétricas para os móveis sonoros sincronizadas em um mesmo instante e que caracterizam uma sonoridade típica.

**Tabela 3** - Esquema de organização dos parâmetros e respectivos algoritmos do sistema de *As Duas Criaturas Baforaram seus olhos*.

Tipo de Processamento	Parâmetros Primários	Primeira Delimitação Paramétrica	Algoritmos de Automação		Segunda Delimitação Paramétrica
Distorção	Amp.	[ 0...1]			
4 Osciladores -> clipagem-> Phaser	Freq. Osc.	204.707			
		204.855			
		250.399			
		256			
	Freq. Mod.	[ 0 ... 10 ] Hz			
	Clip.Ped.	[ 0...1]			
	Clip.Solo	[ 0...1]			
	Amp.	[ 0...1]			
Pitch-shifter	F.Trans.	0.178764 ou 23.9283			
Síntese via Equações não Lineares	Param.Eq.	[ 15 * Combinações Paramétricas ]			
	Ini. Var.	Rand.			
	N.p.Interp.	4			
	N.itr.	1000			
	Freq.L.T.	Freq.	Aut.Rit .-> Aut.Per.	Aut.Rit.	L/D
		T.interp.		Aut.Per.	Amb.
					N.val.
				L.Freq.	
				L.T.interp	
	Amp.	[ 0...1 ]			

Também na **Tabela 5**, as células amarelas são parâmetros da interface de controle utilizados durante a performance do sistema.

As quinze configurações paramétricas assinaladas pelo símbolo \* na **Tabela 3** estão disponibilizadas na **Tabela 4**, seguindo a mesma sequência estabelecida na interface do sistema (vide Fig.26) e com os mesmos nomes.

**Tabela 4** - Parâmetros estabelecidos para equação Latoocarfian no sistema de *As Duas Criaturas Baforaram seus Óculos*

Nomes Atribuídos na Interface	Parâmetros da Equação Latoocarfian			
	a	b	c	d
silencio	- 0.1894	1.4622	0.7053	0.9849
pulso	- 0.6455	0.6688	1.4996	1.4967
tom1	-2.6628	1.063	0.5926	0.8758
tom-loud	1.8411	1.8142	1.1609	0.8675
tom2	-2.3504	1.4664	1.4491	0.9579
tom-complex	-2.7895	2.0079	0.8677	0.623
ruido-grave	-2.9397	0.1025	1.3028	0.9251
ruido-banda	2.4305	-2.1506	1.4495	0.6474
ruido1	1.3588	2.0255	1.1911	1.0876
ruido2	2.0617	1.6777	0.7095	1.3622
ruido3	-2.5773	-0.4408	0.803	1.1928
mistura-longo	-1.26	2.3675	0.5484	1.4314
mistura-curto	-1.2133	2.8976	0.7718	0.956
Perc.	0.4123	-1.3226	0.9055	1.2319
Perc.	1.3983	-2.1585	0.8363	0.786

O sistema do Estudo *Não Diziam-nos que Iam Contar Sobre nosso Passado* apresenta duas etapas de manipulações algorítmicas pelo usuário: a primeira relativa ao momento da composição de estruturas musicais, que neste caso são representadas por forma de ondas (vide Fig. 31, 33 e 35, seção 5.4); e a segunda relativa à performance da peça através da manipulação dos parâmetros de leitura das tabelas que contêm as formas de onda compostas. Por esse motivo representamos nossa organização do sistema em duas tabelas: **Tabela 5** e **Tabela 6**. A primeira correspondente aos procedimentos e parâmetros envolvidos na etapa de composição (determinação do perfil das formas de onda) e a segunda aos parâmetros e procedimentos envolvidos na etapa de performance (síntese sonora pela leitura das tabelas).

**Tabela 5** - Esquema de organização dos parâmetros e algoritmos relacionados a etapa de composição do sistema de *Não Diziam-nos que Iam Contar sobre Nosso Passado*.

Etapas do Processo de manipulação do sistema	Parâmetros Primários	Primeira Delimitação Paramétrica	Parâmetros secundários	Segunda Delimitação Paramétrica
Composição de Formas de Ondas e parâmetros de síntese via Equação Latoocarfian	Param.Eq.	[ Rand. ]		
	Val.Ad.Itr	[0...]		
	Var.Ini.	última gerada		
	N.p.Interp.	[ 2 Tipos: - 4 - Var X ]		
	Assoc.Tab.	[ 3 Tipos: - 1 Tab. -> Var.Y - 2 Tab. -> Var. X e Y - 1 Tab. -> Var. Y e Var X = N.p.Interp. ]		
	N.itr.	[ 0... ]		
	Freq.L.T	Mapeamento = Equação Latoocarfian	Param.Eq.	[ Rand.]
		Var.Ini.	[ Rand.]	
		N.itr.	250	
		Func.Offset	[ 0...252 ]	
		Amb.Map. -> Freq	[ 0 ... ]	

As células verdes da **Tabela 5** indicam os parâmetros utilizados pelo usuário do sistema para composição das formas de onda através da equação Latoocarfian. Apenas o parâmetro de *frequência de leitura da tabela* (designado pela sigla Freq.L.T), é mapeado para uma função numérica gerada por uma segunda implementação da equação Latoocarfian. Os parâmetros dessa equação, bem como dois valores designando os limites de um âmbito, usado para o mapeamento dos valores gerados pela equação para um âmbito de valores em frequência, são parâmetros de controle do usuário do sistema.

A **Tabela 6** apresenta a organização dos algoritmos de performance de *Não Diziam-nos que Iam Contar sobre Nosso Passado*. Esse algoritmos foram construídos sobre manipulações do procedimento de *Dynamic WaveTable* utilizado para leitura linear das tabelas que contêm as formas de onda compostas. Neste procedimento há um constante incremento no ponto de leitura que faz com que a janela de leitura da forma de onda (menor que o tamanho total da tabela) percorra linearmente toda a tabela. Esse valor adicionado iteradamente ao ponto de leitura determina uma progressão linear para

a peça, pois faz com que formas de ondas distintas no decorrer da tabela sejam sonorizadas pelo algoritmo de síntese consecutivamente.

Qualquer interferência no ponto de leitura implica em alterações na linearidade da execução da peça. Por esse motivo foram determinados modos de interferência ao ponto leitura como sendo parâmetros de controle da performance. Com isso propiciamos variabilidade ao resultado sonoro a cada execução/interpretação, mas mantendo uma mesma progressão linear e semelhança entre as diversas execuções da peça.

Na **Tabela 6** as células amarelas representam os parâmetros de controle do sistema. Os parâmetros primários que estão em amarelo na primeira coluna da tabela foram disponibilizados para controle do performer e são relativos à leitura linear da tabela no processo de *Dynamic Wavetable* e à amplitude do sinal resultante. Os parâmetros contidos nas demais células amarelas são relativos aos algoritmos de mapeamento e automação de interferência na linearidade de leitura da tabela.

**Tabela 6** - Esquema de organização dos parâmetros e algoritmos relacionados a etapa de composição do sistema de *Não Diziam-nos que Iam Contar sobre Nosso Passado*.

Etapas do Processo de manipulação do sistema	Parâmetros Primários	Primeiro Mapeamento Paramétrico	Segundo mapeamento Paramétrico	Terceiro Mapeamento paramétricos
Performance . Leitura de Formas de onda para síntese sonora.	inc.Dyn.Wvt( ponto Guia) [0...]			
	Interferência no ponto de leitura	Acr/dcr.Un.Rand . [amb.interf]	Mapeamento Rítmico = Equação Latoocarfian	Param.Eq. [ rand. ]
				Amb.Map.-> Dur [0 ...]
				N.itr. [1 ...]
			L/D	
		Acr/dcr.Itr. [amb.interf.]	Val.itr. [0...]	
	Amb.Interf. [0 ... +/- Tam.Tab.]	L/D		
	Amp. [0...1]			

## **CONCLUSÃO**

Apresentamos nesse trabalho um estudo sobre os Instrumentos Musicais Digitais que abordou tanto aspectos conceituais e debates a respeito da construção e uso destes instrumentos, quanto técnicas específicas de análise e síntese de áudio como fundamentos para sua criação. Aplicamos os conhecimentos aprendidos na composição de 3 Estudos musicais que compreenderam abordagens originais para construção de instrumentos musicais digitais. Além desta Dissertação, nossa pesquisa resultou na produção bibliográfica de um conjunto de 7 publicações em anais de eventos científicos (MANZOLLI, MONTEIRO 2011; MONTEIRO, MANZOLLI 2010; MONTEIRO, MANZOLLI 2011a, 2011b, 2011c, 2011d, 2011e) sendo 5 internacionais e 2 nacionais, assim como 3 apresentações de trabalhos, 1 em evento nacional e 2 em eventos internacionais. Também implementamos uma biblioteca de algoritmos computacionais descritores de áudio denominada de PDescriptors.

No **Capítulo 1** fizemos um levantamento teórico sobre as características dos instrumentos musicais digitais. Decorremos sobre suas propriedades, como a automação e a interatividade, que são propiciadas pelos sistemas computacionais; decorremos também sobre os aspectos de sua construção — como a divisão em interface física e abstração algorítmica computacional —; e sobre as conseqüências de seu uso para a prática musical — como a dissociação entre gesto e resposta sonora e o conceito de *instrumentos compostos*.

Nos **Capítulos 2 e 3** descrevemos procedimentos de análise de áudio baseados na metodologia da área de Recuperação da Informação Musical (MIR) que foram empregados no sistema desenvolvido para um dos Estudos musicais.

Especificamente no **Capítulo 2** foram apresentamos algoritmos para análise de áudio que são denominados na literatura da área de Descritores Acústicos. Esses algoritmos retornam funções numéricas que são relacionadas com medidas do comportamento do sinal. Estas funções são as bases para implementação dos algoritmos de extração de parâmetros musicais que apresentamos no **Capítulo 3**. No terceiro Capítulo os algoritmos implementados foram: 1) procedimentos para segmentação do sinal de áudio capturado de instrumentos musicais, localizando o início e fim de eventos musicais (e.g notas); 2) procedimentos para classificação dos eventos segmentados através da extração do perfil de classes de notas (ou Chroma), definido no **Capítulo 2**; 3) uma metodologia de análise embasada em mapas de Poincaré para visualização do comportamento dinâmico de funções relacionadas às características espectrais do som, que são obtidas através de Descritores Acústicos.

No **Capítulo 4** descrevemos a síntese sonora por equações não-lineares que foi utilizada com processo de sonorização nos três Estudos. Trata-se de um método que produz amostras sonoras digitais no domínio do tempo a partir da amostragem de órbitas numéricas geradas por equações não-lineares. Primeiramente fizemos uma breve revisão de aplicações de sistemas dinâmicos não-lineares e fractais em composição e síntese sonora, em seguida expusemos o método desenvolvido e alguns exemplos das formas de onda geradas através da amostragem das iterações numéricas do sistema não-linear descrito pela equação *Latoocarfian* (Pickover 1994).

Finalmente no **Capítulo 5** apresentamos três Estudos algorítmicos musicais que fazem parte de uma série chamada *As Duas Criaturas* e são os produtos de nossa pesquisa. Descrevemos o processo criativo da concepção dos Estudos e como se inserem nesse contexto as aplicações das técnicas descritas nos **Capítulos 2, 3 e 4** e os conceitos envolvendo a prática musical com instrumentos digitais que foram elencados no **Capítulo 1**. Primeiramente apresentamos o contexto e os preceitos para criação da série *As Duas Criaturas*. Em seguida descrevemos as três peças que compõe esta série: *As Duas Criaturas que Estavam à Mesa de Chá Não Tiveram essa Conversa*, para computador e contrabaixo elétrico, *As Duas Criaturas Baforaram seus Óculos*, para computador e trombone, *Não Diziam-nos que iam Contar Nosso Passado*, para computador solo.

No primeiro Estudo criamos um sistema algorítmico voltado para a improvisação musical baseado em quatro módulos de síntese e processamento sonoro. O algoritmo elaborado nesse Estudo serviu de base para os demais, pois a série de Estudos *As Duas Criaturas* consiste em variações deste primeiro algoritmo. Implementamos uma segunda versão da peça em que empregamos os algoritmos para recuperação da informação musical apresentados nos **Capítulos 2 e 3** com intuito de prover um sistema de extração e memória dos parâmetros sonoros-musicais em jogo na performance para que possam ser recuperados e manipulados pelo *performer* durante a improvisação. A nova versão implementada é um protótipo no qual estamos realizando melhorias e o próximo passo dessa pesquisa será a realização de teste para constatarmos as influências desse sistema na performance da peça.

No segundo Estudo, *As Duas Criaturas Baforaram Seus Óculos*, exploramos uma variação determinista do algoritmo da peça anterior que impõe uma forma fixa e uma condução temporal linear. Criamos uma nova camada de estruturação algorítmica que organiza os parâmetros do algoritmo original em macroestruturas

(*Patches*) de forma específica e detalhada, criando relações que não seriam factíveis pelas possibilidades da estrutura algorítmica e da interface originais, e com isso propiciamos uma nova camada de estruturação e controle composicional, bem como uma nova organização sonora.

Experimentalmente os resultados de *As Duas Criaturas Baforaram Seus Óculos* foram satisfatórios, mas os resultados musicais não satisfizeram os autores principalmente quanto à estruturação e desenvolvimento do material sonoro-musical. Por isso, mesmo já tendo apresentado publicamente uma versão da peça, objetivamos retomar sua construção e modificações deverão ocorrer em relação à versão documentada nesse trabalho.

Por fim, no terceiro Estudo optamos por explorar apenas a síntese por equação não-linear apresentada no **Capítulo 4** através de uma abordagem mais controlada. Utilizamos formas de onda geradas sob diferentes manipulações paramétricas da equação *Latoocarfian* e sua leitura e sonorização pelo processo de *Dynamic Wavetable*. Concluímos que experimentalmente obtivemos resultados satisfatórios que evidenciaram o potencial dos métodos empregados tanto para estruturação musical quanto para síntese sonora via equações não-lineares. Houve uma notória progressão durante a composição das seções em direção a uma abordagem equilibrada entre as propriedades instrumentais do algoritmo e sua unidade enquanto composição musical.

Em termos de composição musical, as abordagens adotadas para cada seção apresentaram resultados bastante diversos em relação às sonoridades obtidas, o que demonstra a riqueza do método e a possibilidade de sua aplicação em diversos contextos. Reaplicá-las em prol do aprofundamento da composição e de uma maior elaboração do discurso é um dos nossos próximos objetivos. Acreditamos que este estudo serviu como uma primeira exploração da metodologia, mas ainda não representa o potencial de elaboração musical que se pode alcançar com sua aplicação para a organização de um discurso musical, por isso a composição deste estudo será retomada e espera-se a obtenção de resultados ainda mais satisfatórios.

A pesquisa aqui reportada deu-nos subsídios e incentivo para prosseguir na direção de uma exploração mais específica das técnicas e processos estudados. Para o Doutorado, propomos o desenvolvimento de um modelo para criação de sistemas musicais interativos em que o computador atuará como: instrumento musical para geração sonora; gerenciador de processos de representação e memória musical; e como dispositivo de auxílio/complementaridade às decisões e tarefas cognitivas de um

*performer* computacional durante a performance musical. O sistema computacional cumprirá funções como sugerir e auxiliar na continuidade e interpolação do fluxo musical, prover *informação estruturada de memória* da performance, *reforçar a interação* multi-sensorial da informação sonora extraída durante a performance, entre outras. Para isso o sistema criará em tempo real modelos dos padrões de organização de parâmetros musicais realizados pelos músicos humanos e, a partir deles, inferir as soluções das tarefas listadas acima.

Em síntese, como produtos da pesquisa para o Doutorado iremos construir um modelo de interação homem-máquina que utilize o sistema computacional para auxiliar os processos cognitivos do *performer* computacional, mantendo nossa abordagem relacionada aos *instrumentos compostos* apresentada nesse trabalho. Iremos implementar um *framework* computacional para que os métodos desenvolvidos possam ser reutilizados em outros trabalhos do próprio autor e compartilhado com a comunidade de pesquisa e produção artística na área. Nesse sentido, a biblioteca PDescriptors, desenvolvida durante a pesquisa de mestrado, representa os primeiros passos. Os algoritmos para extração dos parâmetros musicais e análise de áudio serão as bases para construção da etapa de captura paramétrica dos sistemas, e por isso compõe parte do *framework* que desenvolveremos.

## REFERÊNCIAS

- ACHARYA, U.R.; JOSEPH, K. P.; KANNATHAL, N.; LIM, C. M.; SURI, J. S. Heart rate variability: a review. *Medical and Biological Engineering and Computing*, [S.l.] v. 44 n.12, p. 1031–1051, 2006.
- AL-JAZARÍ. *The Book of Knowledge of Ingenious Mechanical Devices: Kitáb fíma'rifat al-hiyal al-handasiyya*, [S.l.]: Springer, 1973.
- AMES, C. Automated Composition in Retrospect: 1956-1986. *Leonardo*, [S.l.] v. 20, n. 2, Special Issue: Visual Art, Sound, Music and Technology, 1987.
- ASSAYAG, G.; BLOCH, G.; CHEMILLIER, M.; CONT, A.; DUBNOV, S. OMax Brothers: a Dynamic Topology of Agents for Improvization Learning. In: AUDIO AND MUSIC COMPUTING FOR MULTIMEDIA CONFERENCE, 1, 2006, Santa Barbara, California, USA. *Anais...* ACM, 2006.
- BATTIER, M. Electroacoustic music studies and the danger of loss. *Organised Sound*, v.9, n.1, p. 47–53, 2004.
- BARBEDO, J. G.; LOPES, A. Automatic Genre Classification of Musical Signals. *EURASIP Journal on Advanced Signal Processing*, v.2007, n.1, 2007.
- BELLO, J.P. *Towards the automated analysis of simple polyphonic music: A knowledge-based approach..* Tese (Doutorado) Queen Mary University of London, London, 2003.
- BELLO, J.P.; DAUDET, L.; ABDALLAH, S.; DUXBURY, C.; DAVIES, M.; SANDLER, M.A. Tutorial on Onset Detection in Music Signals. *IEEE Transactions on Speech and Audio Processing*, v.13, p. 1035–1047, 2005.
- BELLO, J. P.; MONTI, G.; SANDLER, M. Techniques for automatic music transcription. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 2000, Plymouth, Massachusetts, USA. *Anais...* Plymouth: International Society for Music Information Retrieval, 2000, p.23–25.
- BELLO J.P. ; PICKENS J. A robust mid-level representation for harmonic content in music signals. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 2005, London, UK, *Anais:* London: International Society for Music Information Retrieval. 2005. p. 304–311,. Disponível em <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.6957&rep=rep1&type=pdf> , acessado em 24/06/2011.
- BERG, P. Composing Sound Structures with Rules. *Contemporary Music Review* vol. 28, n. 1, p. 75–87, 2009.
- BIDLACK, R. Chaotic Systems as Simple (But Complex) Compositional Algorithms. *Computer Music Journal*, v. 16, n. 3, p. 33-47 1992.
- BIRKHOFF, G. *Dynamical Systems*. New York, American Mathematical Society, 1927.
- BLOIT, J. ; RASAMIMANANA, N. ; BEVILACQUA, F. Modeling and segmentation of audio descriptor profiles with segmental models. *Pattern Recognition Letters*, v. 31, p. 1507-1513, 2010.
- BOULANGER, R. (ed). *The CSound Book: Perspectives in software synthesis, sound design, signal processing, and programming*. MIT Press, Cambridge, Massachusetts. 2000.

- BOWN O. ; ELDRIDGE, A. ; MCCORMACK, J. Understanding Interaction in Contemporary Digital Music: from instruments to behavioural objects. *Organised Sound*, v. 14, n.2, p. 188–196, 2009.
- BRADLEY, E. ; MANTILLA, R. Recurrence plots and unstable periodic orbits. *Chaos*, v. 12 , n. 3, p. 596–600, 2002.
- BROSSIER, P.M. Automatic Annotation of Musical Audio for Interactive Applications. Tese (Doutorado), Queen Mary University of London, London, 2006.
- BULLOCK, J. ; COCCIOLI, L. Modernising musical works involving Yamaha DX-based synthesis: a case study. *Organised Sound*, v. 11, n. 3, p. 221–227, 2006.
- CADOZ, C. Supra-Instrumental Interactions and Gestures. *Journal of New Music Research*, v.38, n.3, p. 215–230, 2009.
- CADOZ, C. ; WANDERLEY, M. Gesture - Music. In: WANDERLEY, M E BATTIER, M. (Ed.) *Trends in Gestural Control of Music*, Paris, France: Institut de Recherche et Coordination Acoustique Musique — Centre Pompidou, 2000. Disponível em [http://www.music.mcgill.ca/~mwanderley/Trends/Trends\\_in\\_Gestural\\_Control\\_of\\_Music/DOS/P.CadWan.pdf](http://www.music.mcgill.ca/~mwanderley/Trends/Trends_in_Gestural_Control_of_Music/DOS/P.CadWan.pdf). acessado em 19/12/2011.
- CAETANO, M. ; RODET, X Automatic Timbral Morphing of Musical Instrument Sounds By High-Level Descriptors. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, New York, 2010. *Anais...* New York: International Computer Music Association, 2010.
- CAMURRI, A. ; VARNI, G. ; VOLPE, G. Towards Analysis of Expressive Gesture in Groups of Users: Computational Models of Expressive Social Interaction. In: *Lecture Notes in Computer Science*, Springer: Berlin, 5934/2010, 122–133, 2010.
- CARAMIAUX, B. *Gestification du son: mapping adaptatif geste/son dans un contexte d'écoute et performance musical*. Tese (Doutoramento) Université Paris VI - Pierre et Marie Curie, Paris, 2008.
- CHAPEL, R. H. *Realtime Algorithmic Music Systems from Fractals and Chaotic Functions: Toward na Active Musical Instrument*. Dissertação (submetida ao exame parcial para aquisição de Doutorado) Universitat Pompeu Fabra, Barcelona 2003.
- CHEVEIGNÉ, A. DE ; KAWAHARA, H. YIN, A fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, v.111, 2002.
- CLARK, A. ; CHALMERS D. The Extended Mind. *Analysis*, v. 58, p. 10-23, 1998.
- COLLINS, N. Generative Music and Laptop Performance. *Contemporary Music review*, v. 22, n. 4, p. 67-79, 2003.
- COLLINS, N. A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions. In: 118th AUDIO ENGINEERING SOCIETY CONVENTION, 118, 2005, Barcelona. *Anais...* Barcelona: Audio Engineering Society, 2005.
- COLLINS, N. The Analysis of Generative Music Programs. *Organised Sound*, v. 13, n. 3, p. 237–248, 2008.
- DABBY, D. *Musical Variations from a Chaotic Mapping*. Tese (Doutorado), Massachusetts Institute of Technology, Cambridge, 1995.

- DAMIANI, F. ; MANZOLLI, J. ; TATSCH, P. J. A non-linear algorithm for the design and production of digitally synthesized sounds, In: INTERNATIONAL CONFERENCE ON MICROELECTRONICS AND PACKING, 1999, Campinas, Anais... ICMP '99, Sociedade Brasileira de Micro-eletrônica, 1999, pp. 196-199.
- DI SCIPIO, A. Composition by exploration of non-linear dynamics systems. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1990, Glasgow. Escócia. Anais ... Glasgow: International Computer Music Association, 1990, p. 324-328.
- DI SCIPIO, A. ; PRIGNANO, I. Synthesis by Functional Iterations. A Revitalization of Non-Standard Synthesis. *Journal of New Music Research*, v. 25, n. 1, p. 31-46, 1996.
- DIAZ-JEREZ, G. *Algorithmic Music: using mathematical models in music composition*. Dissertação (submetida ao exame parcial para requerimentos em Doutorado em Música). Manhattan School of Music, New York, 2000.
- DI GUIGNO, G., ; GERZSO, A. La Station de Travail Musical 4X. IRCAM. Technical Report. IRCAM, Paris. 1986.
- DIXON, S. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, vol. 30, p. 39-58, 2001.
- DIXON, S. Onset Detection Revisited. In: INTERNETAIONAL. CONFERENCE ON DIGITAL AUDIO EFFECTS, 9, 2006, Montreal. Anais... DAFx-06, Montreal, Canada, 2006 p. 133–137.
- DODGE, C. Profile: A musical fractal, *Computer Music Journal*, v.12, n. 3, p.10-14, 1988.
- DRACHMANN, A.G. Heron's Windmill, *Centaurus*, v. 7, p. 145-151, 1961.
- DRUMMOND, J. Understanding Interactive Systems. *Organised Sound*, v. 14, n. 2, p. 124-133, 2009.
- ERONEN, A. ; KLAPURI, A. Musical instrument recognition using cepstral coefficients and temporal features. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, 2000, Istanbul, Turkey, Anais... ICASSP'00, Istanbul, 2000, p. II753–II756 vol.2.
- FASTL,H.; ZWICKER. E. *Psychoacoustics : facts and models*. Berlin; New York, N.Y.: Springer, 2007.
- FOWLER, C.B. The Museum of Music: A History of Mechanical Instruments. *Music Educators Journal*, v. 54, n. 2, p. 45-49, 1967.
- FUJISHIMA T. Realtime chord recognition of musical sound: A system using common lisp music In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1999, Beijing. Anais... Beijing: International Computer Music Association, 1999 Disponível em <http://ci.nii.ac.jp/naid/10013545881/>, acessado 24/06/2011.
- GIMENES, M. *An Approach to Machine Development of Musical Ontogeny*. Tese (Doutorado) School of Computing, Communications and Electronics, University of Plymouth, Plymouth, Reino Unido, 2008.
- GLEICK, J. *Caos: a criação de uma nova ciência*. Campos. 1989.

- GÓMEZ, E. Tonal Description of Polyphonic Audio for Music Content Processing. *INFORMS Journal on Computing* , v. 18, p. 294-304, 2006.
- GOGINS, M. Iterated Functions Systems Music. *Computer Music Journal*, v. 15, n.1, p. 40-48, 1991.
- GOUYON, F. *A computational approach to rhythm description-Audio features for the computation of rhythm periodicity functions and their use in tempo induction and music*, Dissertação (submetida ao exame parcial para aquisição de Doutorado) Universitat Pompeu Fabra, Barcelona, 2005.
- GOTO, S. The aesthetics and technological aspects of virtual musical instruments: The case of the SuperPolm MIDI violin. *Leonardo Music Journal*, v.9, p. 115 –120, 1999.
- GUREVICH, M. ; FYANS A. C. Digital Musical Interactions: Performer–system relationships and their perception by spectators. *Organised Sound* , v. 16, n. 2, p. 166–175, 2011.
- HASSELBLATT, B. ; KATOK, A. *The modern theory of dynamical systems. Encyclopedia of mathematics and its applications*, 57. Cambridge University Press ,1997.
- HERRERA, P. ; DEHAMEL, A. ; GOUYON, F. Automatic Labeling of Unpitched Percussion Sounds. In: AUDIO ENGINEERING SOCIETY CONVENTION, Amsterdam , 114, Amsterdam, 2003. *Anais...* Amsterdam: Audio Engineering Society, 2003.
- HERRERA-BOYER, P. ; KLAPURI, A. ; DAVY, M. Automatic Classification of Pitched Musical Instrument. In: KLAPURI A. ; DAVY (Ed). *Signal Processing Methods for Music Transcription*. New York, NY, USA: Springer Science+Business Media LLC, 2006.
- HOFFMANN, P. *Music Out of Nothing? A Rigorous Approach to Algorithmic Composition by Iannis Xenakis*. Tese (Doutorado), Technischen Universität Berlin, 2009.
- HOFFMANN, M. A.; MANZOLLI, J. Sistemas dinâmicos não-lineares aplicados ao design sonoro. In: SEMINÁRIO MÚSICA, CIÊNCIA E TECNOLOGIA, 2, 2005, SCIELO Campinas, *Anais...* Campinas: Unicamp, p. 1-6, 2005.
- IBN SHAKIR, MUHAMMAD IBN MUSA; IBN SHAKIR, AHMAD IBN MUSA; IBN SHAKIR, HASAN IBN MUSA. *The book of ingenious devices (Kitaab al-ohiyal)*, 873. Tradução Donald Routledge Hill, Dordrecht, Boston: D. Reidel, c1979.
- JANER, J. ; HARO, M. ; ROMA, G. ; FUJISHIMA, T. ; KOJIMA, N. Sound Object Classification for Symbolic Audio Mosaicing: a Proof-of-Concept. In: SOUND AND MUSIC COMPUTING CONFERENCE, 6, 2009, Porto. *Anais...* Porto: Sound and Music Computing research community, 2009, p.297-302.
- JEHAN T. *Perceptual Synthesis Engine : An Audio-Driven Timbre Generator*. Dissertação (Mestrado) Massachusetts Institute of Technology , Cambridge, 2001.
- JEHAN, T. *Creating Music by Listening*. Tese (Doutoramento), Massachussets Institute of Technology, Cambridge, 2005.
- JIRSA, V. K. ; KELSO, J.A.S. The Excitator as a Minimal Model for the Coordination

- Dynamics of Discrete and Rhythmic Movement Generation”. *Journal of Motor Behavior*, v. 37, n. 1, p. 35–51, 2005.
- JOHNSTON, A. ; CANDY, L. ; EDMONDS, E. Designing and Evaluating Virtual Musical Instruments: Facilitating Conversational User Interaction. *Design Studies*, v. 29, n. 6, p. 556–71, 2008.
- JORDÀ, S.. *Digital Lutherie: Crafting Musical Computers for New Musics Performance and Improvisation*. Tese (Doutorado) Universitat Pompeu Fabra, Barcelona, 2005.
- KLAPURI, A. *Signal Processing Methods for the Automatic Transcription of Music*. Tese (doutorado), Tampere University of Technology, 2004.
- KLAPURI A. ; DAVY (Ed). *Signal Processing Methods for Music Transcription*. New York, NY, USA: Springer Science+Bussiness Media LLC, 2006.
- LEE, K. ; SLANEY, M. Acoustic Chord Transcription and Key Extraction From Audio Using Key-Dependent HMMs Trained on Synthesized Audio. *IEEE Transactions on Audio, Speech, and Language Processing* v.16, p. 291-301, 2008.
- LEVITIN, D. ; MCADAMS, S. ; ADAMS, R. Control parameters for musical instruments: a foundation for new mappings of gesture to sound. *Organised Sound*, v. 7, n.2, p. 171–189, 2002.
- LEWIS, G.E. Too Many Notes: Computers, Complexity and Culture in Voyager. *Leonardo Music Journal*, v. 10, p. 33– 39, 2000.
- LINDEMANN, E. ; DECHELLE, F. ; SMITH B. E. ; STARKIER M. The Architecture of the IRCAM Musical Workstation. *Computer Music Journal* v. 15, n. 3, p. 41-49, 1991.
- LIPE, C. Real-Time Interactive Digital Signal Processing: A View of Computer Music. *Computer Music Journal*, v. 20, n. 4, p. 21-24, 1996.
- LOUREIRO M.A. ; PAULA H.B. Timbre de um instrumento musical: caracterização e representação. *PER MUSI-Revista Acadêmica de Música*. 2006, 1457-81 Disponível em: [http://www.musica.ufmg.br/permusi/port/numeros/14/Num14\\_cap\\_05.pdf](http://www.musica.ufmg.br/permusi/port/numeros/14/Num14_cap_05.pdf), acessado em 24/06/2011.
- MACHOVER, T. Hyperinstruments - A Progress Report 1987 - 1991. Technical report, Massachusetts Institutof Technology, 1992.
- MAGNUSSON, T. Of Epistemic Tools: musical instruments as cognitive extensions. *Organised Sound*, v. 14, n. 2, p. 168–176, 2009.
- MAGNUSSON, T. Designing constraints: Composing and performing with digital musical systems. *Computer Music Journal* v. 34, n. 4, p. 62–73, 2010.
- MAGNUSSON, T. *Confessions of a Live Coder*. 2011 acessado em <http://www.ixi-software.net/thor/confessions.pdf>, consultado em 19/12/2011.
- MALT, M. ; JOURDAN, E. Real-Time Issues of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa.Descriptors Library. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO MUSICAL, 12, 2009, Anais Recife: Sociedade Brasileira de Computação Musical, 2009.
- MANDELBROT, B. *The Fractal Geometry of Nature*. rev. ed. New York: W.H.

- Freeman Company, 1982.
- MANIATAKOS, F. ; ASSAYAG G., BEVILACQUA, F.; AGON, C. On Architecture and Formalisms For Computer-Assisted Improvisation. In: SOUND AND MUSIC COMPUTING CONFERENCE, 7, Barcelona, 2010, *Anais ...* Barcelona: Sound and Music Computing reserach community, 2010.
- MANZOLLI, J. *Non-linear Dynamics and Fractals as a Model for Sound Synthesis and Real time Composition*. Tese (Doutorado), University of Nottingham, Inglaterra, 1993a.
- MANZOLLI, J. Musical application derived from FracWave sound synthesis method. In: AUDIO ENGINEERING SOCIETY CONVENTION, 94, Berlim, 1993, *Anais...* Berlim: Audio Engineering Society, 1993b.
- MANZOLLI, J. ; MONTEIRO, A.C., An interactive system for synthesing singing voices duets. In: P3S - INTERNATIONAL WORKSHOP ON PERFORMATIVE SPEECH AND SINGING SYNTHESIS, Vancouver, University of British Columbia Canadá, Março 14-15, 2011.
- MASRI P. *Computer Modelling of Sound for Transformation and Synthesis of Musical Signals*. Tese (Doutorado), University of Bristol, Department of Electrical and Electronic Engineering, Bristol, 1996.
- MCADAMS, S. Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal*, v. 23, n. 3, p. 85-10, 1999.
- MIRANDA, E.R. ; WANDERLEY M.M. *New Digital Musical Instruments: Control and Interaction beyond the Keyboard*, A-R Editions, Spring, 2006.
- MOELANTS, D. ; RAMPAZZO, C.. A computer system for the automatic detection of perceptual onsets in a musical signal. In CAMURRI, A. (Ed.), *KANSEI - The Technology of Emotion*, p 141–146, Genova: AIMI-DIST, 1997. Apud. BROSSIER, P.M. *Automatic Annotation of Musical Audio for Interactive Applications*. Tese (Doutorado), Queen Mary University of London, London, 2006, p34.
- MONTEIRO ,A.C.; MANZOLLI J. Aplicação de Descritores em Sistema Interativo para Música Mista, Revista EIMAS 2010, disponível em: <http://www.eimas.net/port/revista2010.html>.
- MONTEIRO, A.C.; MANZOLLI, J. Análise Computacional de Texturas Sonoras via Mapas de Poincaré. In: CONGRESSO DA ANPPOM, 21, 2011, Uberlândia. *Anais...* : Uberlândia: Associação Nacional de Pesquisa e Pós-Graduação em Música, 2011a. CD-room.
- MONTEIRO, A.C.; MANZOLLI, J., A Framework for Real-Time Instrumental Sound Segmentation and Labeling. In: PURE DATA CONVENTION, 4, 2011, Weimar. *Anais ...* : Weimar: PDcon'11, 2011b.
- MONTEIRO, A.C. ; MANZOLLI, J. Análise de Áudio e Recuperação da Informação Musical em um Ambiente Computacional Voltado a Improvisação, In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO MUSICAL, 13, 2011, Vitória. *Anais...* : Vitória: Sociedade Brasileira de Computação Musical, 2011c CD-room.
- MONTEIRO, A.C. MANZOLLI J., Estudo de Performance e Interação Utilizando Processamento em Tempo Real. IN: Encontros de Investigação em Pefromance, 2011, Aveiro, *Anais...* Aveiro: Performa'11, 2011d.

- MONTEIRO, A.C, MANZOLLI J., “DUETO: System for Synthesises Singing Duets in Real Time”. CONGRESSO DE ENGENHARIA DE ÁUDIO, 9, 2011, São Paulo, *Anais...* São Paulo: Audio Engineering Society, 2011e.
- MONRO, G. Fractal Interpolation WaveForms. *Computer Music Journal*, v. 19, n. 1, p. 88-98, 1995.
- MOREIRA, I.C. Sistemas caóticos em física: uma introdução. *Revista Brasileira de Ensino de Física*. v. 15, ns. 1-4, p.163-181, 1993.
- MÜLLER, M. *Information retrieval for music and motion*. Springer-Verlag, New York Inc: Bonn, Germany, 2007.
- MÜLLER, M. ; ELLIS, D.P.W. ; KLAPURI, A. ; RICHARD, G. Signal Processing for Music Analysis. *IEEE Journal of Selected Topics in Signal Processing*, v. 5, n. 6, p.1088 - 1110, 2011.
- MUMMA, G. Creative Aspects of Live Electronic Music Technology. In: NATIONAL CONFERENCE OF AUDIO ENGINEERING SOCIETY, 33, 1967, New York. Acessado em <http://brainwashed.com/mumma/creative.htm>, consultado em 19/12/2011.
- NILSON, C. Live Coding Practice. In: NEW INTERFACE FOR MUSIC EXPRESSION, 7, 2007, New York. *Anais. NIME'07 2007*. Acessado em <http://www.sussex.ac.uk/Users/nc81/research/livecodingpractice.pdf> consultado em 19/12/2011.
- ONG, B. S. *Structural analysis and Segmentation of Music Signals*. Tese (Doutorado) Universitat Pompeu Fabra, Barcelona, 2006.
- OPENHEIM, A. ; SHAFER, R. *Discrete-Time Signal Processing*, 3. ed. Upper Saddle River: Prentice Hall, 2010.
- ORIO, N. A Gesture Interface Controlled by the Oral Cavity. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1997, Thessaloniki, Grécia. *Anais...* International Computer Music Association, 1997, p. 141–144.
- OVERHOLT D. The Musical Interface Technology Design Space. *Organised Sound*. v. 14, n. 2, p. 217-226, 2009.
- PACHET, F. The Continuator: Musical Interaction with Style. *Journal of New Music Research* v. 32, n. 3, p. 333–41, 2003.
- PAINE, G. Interactivity, where to from here? *Organised Sound*, v. 7, n.3, p. 295-304, 2002.
- PAINE, G. Towards Unified Design Guidelines for New Interfaces for Musical Expression. *Organised Sound*, v. 14, n. 2, p. 142–155, 2009.
- PAULUS, J. ; MÜLLER M. ; KLAPURI A. Audio-based music structure analysis. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 11, 2010, Utrecht. *Anais...* Utrecht: International Society for Music Information Retrieval, 2010. Acessado em: <http://ismir2010.ismir.net/proceedings/ismir2010-107.pdf> (consultado em 19/12/2011)
- PARK, T.H. *Towards Automatic Musical Instrument Timbre Recognition*. Tese (Doutorado) Princenton University, Princeton, 2004.

- PEREIRA, E. M. *Estudos Sobre uma Ferramenta de Classificação Musical*. Dissertação (Mestrado), Universidade Estadual de Campinas, Campinas, 2009.
- PEETERS G. Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization. In: AUDIO ENGINEERING SOCIETY CONVENTION, 115, 2003, New York. *Anais...* New York: Audio Engineering Society, 2003. p. 1-14.
- PEETERS, G., A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project, CUIDADO I.S.T. Project Report. 2004.
- PEETERS, G. Chroma-based estimation of musical key from audio-signal analysis. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 7, 2006, Victoria, Canadá. *Anais...* Victoria: International Society for Music Information Retrieval, 2006, p. 115–120.
- PEETERS, G. Music Pitch Representation by Periodicity Measures Based on Combined Temporal and Spectral Representations. In: IEEE INTERNATIONAL CONFERENCE OF ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2006, Toulouse, *Anais...* Toulouse: ICASSP v. 5, 2006b
- PEETERS, G.; DERUTY, E. Sound Indexing Using Morphological Description. *Audio, Speech, and Language Processing, IEEE Transactions on Speech and Language Processing*, pp. 675–687, 2010.
- PEETERS, G. ; RODET, X. Hierarchical Gaussian Tree with Inertia ratio Maximization for the Classification of Large Musical Instrument Databases. In: CONFERENCE ON DIGITAL AUDIO EFFECTS,6, London, 2003, *Anais...* London: Queen Mary University, DAFx-03,1-6, 2003.
- PETERS, M. From Strange to Impossible Interactive Attractor Music. *Contemporary Music Review*, v. 29, n. 4, p- 395-404, 2010.
- PICKOVER, C.A. *Chaos in wonderland: Visual Adventures in Fractal World*. New York, USA: St. Martin's Press, 1994.
- PIRES, A. S.. *Métodos de Segmentação Musical Baseados em Descritores Sonoros*. Dissertação (Mestrado), Universidade de São Paulo, São Paulo, 2011.
- POINCARÉ, H. *Science and method*. New York: Dover Publications. Paper bound students' ed. 1952.
- POINCARÉ, H. *Les méthodes nouvelles de la mécanique céleste*. Gauthier-Villars, 1893. Vol. 1-3. Republicado por Blanchard, Paris, 1993.
- POLFREMAN, R.; SHEPPARD, D.; DEARDEN, I. Time to re-wire? Problems and strategies for the maintenance of live electronics. *Organised Sound*, v.11 n. 3, p. 229–242, 2006.
- PRESSING, J. Nonlinear maps as generators of musical design. *Computer Music Journal*, v. 12, n. 2, p. 35-46, 1988.
- PRESSING, J. Cybernetic Issues in Interactive Performance Systems. *Computer Music Journal* v. 14, n. 1, p. 12–25, 1990.
- PUCKETTE, M. The Patcher. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1988, Kologne. *Anais...* Kologne: International Computer Music Association, 1988. pp. 420-429

- PUCKETTE, M. Combining Event and Signal Processing in the MAX Graphical Programming Environment. *Computer Music Journal*, v. 15, n. 3, p. 68-77, 1991.
- PUCKETTE, M. The Theory and Technique of Electronic Music. Draft: December 30, 2006. Disponível em: <http://crca.ucsd.edu/~msp/techniques/latest/book.pdf>, acessado em 03/01/2012
- ROADS, C. A Tutorial on Non-Linear Distortion or Waveshaping Synthesis. *Computer Music Journal*, v. 3, n. 2, p. 29-34, 1979.
- ROADS, C. *The Computer Music Tutorial*. MIT Press, Cambridge, Massachusetts. 1996.
- ROADS, C. *Microsound*. MIT Press, Cambridge, Massachusetts. 2001.
- ROSAMIMANANA, N. ; KAISE, F. ; BEVILACQUA F. Perspectives on Gesture–Sound Relationships Informed from Acoustic Instrument Studies. *Organised Sound*, v. 14, n. 2, p. 208–216, 2009.
- ROWE, R. *Interactive Music Systems: machine listening and composing*, Cambridge, MIT Press, 1993.
- SCHEIRER, E. *Extracting expressive performance information from recorded music*. Tese (Mestrado) Massachusetts Institute of Technology, Cambridge, 1995.
- SCHNELL, N. ; BATIER, M. Introducing Composed Instruments, Technical and Musicological Implications. CONFERENCE ON NEW INSTRUMENTS FOR MUSICAL EXPRESSION, 2002, Dublin . *Anais...* Dublin: NIME-02 , 2002, p.1-5.
- SCHWARZ, D. *Data-Driven Concatenative Sound Synthesis*. Tese (Doutorado), Université Paris VI - Pierre et Marie Curie, Paris, 2006.
- SERRA, M. Stochastic Composition and Stochastic Timbre: GENDY3 by Iannis Xenakis. *Perspectives of New Music*, v.31, n. 1, p. 236-257, 1993.
- SHEH A, ELLIS DPW. Chord segmentation and recognition using EM-trained hidden Markov models. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIEVAL, 2003, Baltimore, USA. *Anais...* Baltimore: International Society for Music Information Retrieval, 2003. p. 185–191.
- SLATER, D. Chaotic Sound Synthesis. *Computer Music Journal*, v. 22, n. 2, p. 12-19, 1998.
- SORENSEN, A.; BROWN A. aa-cell In Practice: an approach to musical live coding. In: PROCEEDINGS OF THE INTERNATIONAL COMPUTER MUSIC CONFERENCE, 2007, Copenhagen. *Anais...* Copenhagen: International Computer Music Association, 2007 pp. 292-299
- STARK, A.M. ; PLUMBLEY, M.D. Real-time Chord Recognition for Live Performance. In PROCEEDINGS OF INTERNATIONAL COMPUTER MUSIC CONFERENCE, 2009, Montreal, Canadá. *Anais...* International Computer Music Association, 2009.
- STEVENS, S. ; VOLKMAN, J. ; NEWMAN, E. A scale for the pitch measurement of the psychological magnitude of pitch. *Journal of Acoustical Society for America*, v. 8, n. 3, p.185-190, 1937.

- TERUGGI, D. Electroacoustic preservation projects: how to move forward. *Organised Sound*, v. 9, n. 1, p. 55–62, 2004.
- TRUAX, B. Chaotic non-linear systems and digital synthesis: na exploratory study. In: INTERNATIONAL COMPUTER MUSIC CONFERENCE, 1990. Glasgow Anais... Glasgow: International Computer Music Association, 1990, p. 100-103.
- VALSAMAKIS, N.; MIRANDA, E. R. Interactive Sound Synthesis by means of cross-coupled digital oscillators. *Digital Creativity*, v. 16, n. 2, p. 79-92, 2005.
- VAN NORT, D. Instrumental Listening: sonic gesture as design principle. *Organised Sound*. v, 14 n. 2, p. 177-187, 2009.
- VARELA, F.; THOMPSON, E. ; ROSCH, E. *The Embodied Mind*. Cambridge, Massachusetts: MIT Press, 1991.
- YANG, D. The Hyperbow Controller: Real-Time Dynamics Measurement of Violin Performance. In: CONFERENCE ON NEW INSTRUMENTS FOR MUSICAL EXPRESSION , 2002. Dublin, Anais... Dublin: NIME-02, 2002.
- YEH, C. Multiple *Fundamental Frequency Estimation of Polyphonic Recordings*. Tese (Doutorado) Université Paris VI - Pierre et Marie Curie, Paris, 2008.
- WHALLEY, I. Software Agents in Music and Sound Art Research/Creative Work: current state and a possible direction. *Organised Sound*, v. 14, n. 2, p. 156–167, 2009.
- WANDERLEY, M. DEPALLE, P. Gestural Control of Sound Synthesis. *Proceedings of the ieee*, v. 92, n. 4, 2004, p.632-644.
- WANDERLEY M.M. ; ORIO N. Evaluation of Input Devices for Musical Expression: Borrowing Tools from HCI. *Computer Music Journal*, v. 26, n.3, p. 62-76, 2002.
- WEINBERG, G. AND DRISCOLL, S.. Toward Robotic Musicianship. *Computer Music Journal* v. 30, n. 4, p. 28–45, 2006.
- WESSEL, D. e WRIGHT, M. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal*, v. 26, n. 3, p. 11–22, 2002.
- WETZEL, D. B. A model for the conservation of interactive electroacoustic repertoire: analysis, reconstruction, and performance in the face of technological obsolescence. *Organised Sound*, v. 11, n. 3, p. 273–284, 2006.
- WU H., BELLO J. P. *Audio-Based Music Visualization for Music Structure Analysis*. In: PROCEEDINGS OF SOUND AND MUSIC COMPUTING CONFERENCE 2010. Barcelona, Anais..., Barcelona: Sound and Music Computing research community, 2010.
- ZATTRA, L. Searching for lost data: outlines of aesthetic–poietic analysis. *Organised Sound*, v. 9, n. 1, p.35–46, 2004.
- ZWICKER, E. Subdivision of the audible frequency range into critical bands. *The Journal of the Acoustical Society of America*, v. 33, 1961.

## **ANEXO A**

### **DESCRITORES EM PDESCRIPTORS: TAXONOMIA E DEFINIÇÕES**

Apresentamos neste anexo as formalizações matemáticas dos descritores de sinal de áudio inclusos em PDescriptors que pode ser encontrada no CD Apêndice A desta Dissertação. A taxonomia para organização dos descritores aqui apresentados é descrita na **Tabela A.1**.

**Tabela A.1** - Taxonomia empregada para os Descritores.

Características Temporais	Globais	Centróide Temporal
	Instantâneos	RMS
		Auto-correlação
		Taxa de cruzamento por zero
Características Espectrais	Características do envelope espectral	Centróide Espectral
		Desvio padrão Espectral
		Obliquidade Espectral
		Curtose Espectral
		Nivelamento Espectral
		Medida de Crista Espectral
		Irregularidade Espectral
		Inclinação Espectral
		Decréscimo Espectral
		Roll-off
		Conteúdo de Frequências Altas
	Características da dinâmica espectral	Fluxo Espectral
		Desvio de Fase
	Descrição Global segundo preceitos psicoacústicos	MFCC
BFCC		
Características segundo preceitos musicais	Chroma	
Características Psicoacústicas	-----	Loudness
		Energia por banda MEL
		Energia por banda Bark

Descritores de características temporais são computados no domínio do tempo, descritores de características espectrais são computados no domínio das frequências (após a transformada de Fourier do sinal analisado) e descritores psicoacústicos são baseados em modelos da fisiologia da audição humana. As especificidades de cada classe de descritores serão descritas nas correspondentes subseções.

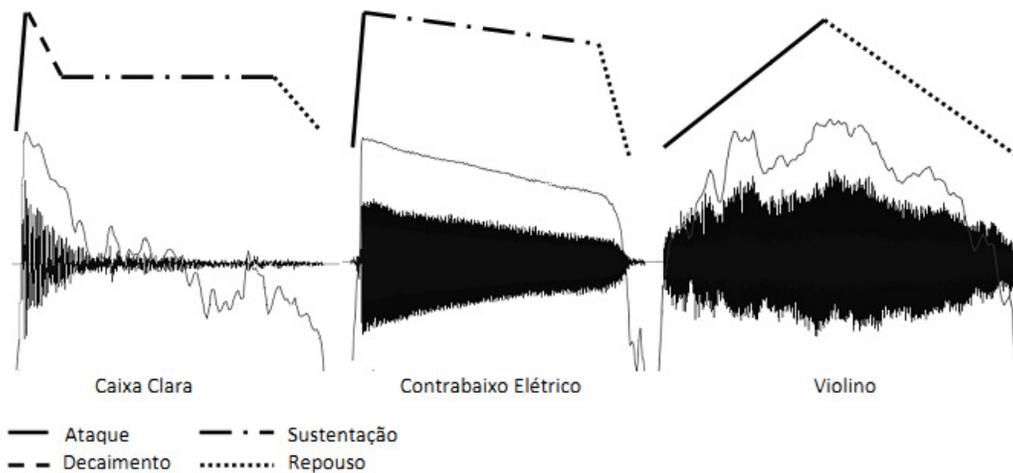
### A.1 Características Temporais

Descritores temporais são computados no domínio do tempo, ou seja, diretamente sobre o sinal de áudio  $x[n]$ .

### A.1.1 Características Temporais Globais

*Descritores Temporais Globais* extraem medidas do sinal de áudio que descrevem aspectos de seu envelope de energia no domínio do tempo. O envelope energético de instrumentos musicais é comumente descrito por quatro componentes: 1) ataque, 2) decaimento, 3) sustentação e 4) repouso (*Attack, Decay, Sustain e Release - ADSR*). O ataque é o início do som, o momento em que a energia cresce do silêncio, do valor inicial “zero” até o seu pico de energia máximo. O decaimento é um decréscimo na energia antes da sustentação, este é o momento caracterizado pela dissipação da energia de excitação logo após a excitação cessar. A sustentação é o momento de estabilidade da energia e, na maioria das vezes, associada à ressonância de ondas estacionárias. Por fim, o repouso é o decréscimo de energia após a sustentação e até o fim do evento sonoro.

A Fig. 37 ilustra alguns envelopes analisados de três sons de instrumentos musicais: caixa clara percutada, contrabaixo elétrico com apenas uma corda tangida, violino com apenas uma corda friccionada pelo arco. As curvas finas e sobrepostas às representações temporais dos sons gravados (formas de ondas escuras na parte inferior da figura) indicam a energia do envelope de amplitude calculado pelo quadrado da amplitude. As curvas superiores relativas à legenda da figura representam uma análise dos componentes ADSR dos envelopes dos sons amostrados.



**Fig. 37** - Exemplos de tipos de envelope ADSR em diferentes instrumentos musicais.

Através da comparação entre as curvas sobrepostas (envelope energético do sinal) e as superiores (modelo ADSR extraído) em relação à forma de onda dos sons amostrados, nota-se que o modelo ADSR é uma simplificação da complexidade contida nas pequenas nuances do envelope energético dos sons. Este modelo é extraído de um padrão macro detectado como recorrente para boa parte dos sons gerados por instrumentos musicais. Nota-se também que em alguns casos, componentes do envelope podem ser muito reduzidos tornando sua identificação pouco clara ou inexistente, por exemplo, no caso do contrabaixo em que o decaimento parece omitido e a ressonância da sustentação possui um decréscimo gradual de energia, ou no caso da amostra de violino que possui um envelope crescendo-decrescendo omitindo os componentes decaimento e sustentação. Esse tipo de envelope é factível no violino pela constante e controlável reiteração da energia de excitação propiciada pelo arco.

Recapitulando, os Descritores de características temporais globais fornecem dados para classificação de alguns aspectos do envelope energético de um som e de suas características em relação ao modelo ADSR descrito acima. Para isso, é necessário que o cálculo de um descritor global seja operado sobre a totalidade do som a ser analisado.

#### **A.1.1.1 Centróide Temporal**

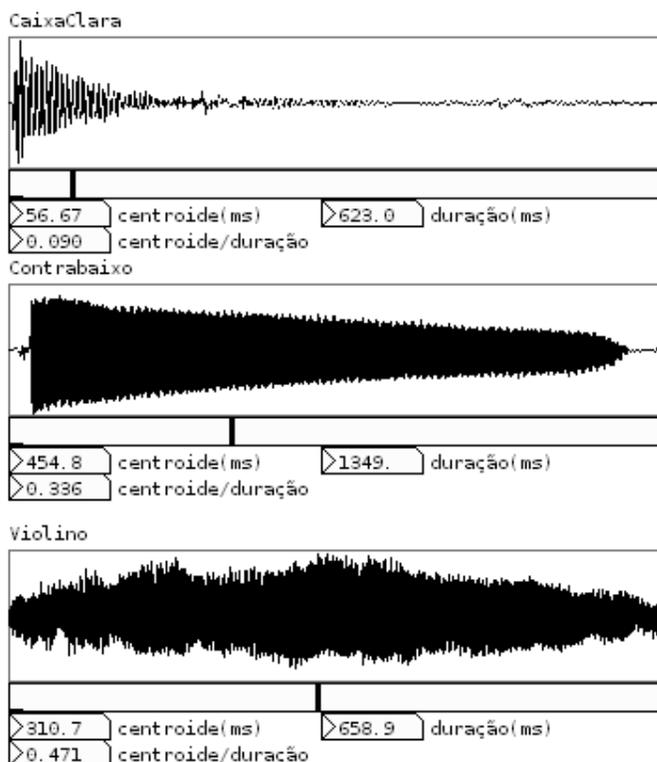
O *Centróide Temporal* é o baricentro (ou o centro de gravidade) do envelope de energia de um som. Este descritor retorna o valor no eixo temporal que divide o som analisado em duas partes que possuem a mesma quantidade de energia (somatória da energia no trecho). O cálculo é dado pela soma da energia de cada amostra temporal, multiplicada pelo valor temporal e dividida pela somatória da energia de todas as amostras. A Equação A.1 apresenta a descrição formal deste descritor para um sinal de áudio digitalizado  $x[n]$ .

$$CT = \frac{\sum_{n=0}^{N-1} e[n]n}{\sum_{n=0}^{N-1} e[n]} \quad (\text{A. 1})$$

Onde  $N$  é o número total de amostras do som analisado;  $e[n]$  é a energia da amostra definida como o quadrado do valor da amplitude na  $n^{\text{ésima}}$  amostra:  $e[n] = x[n]^2$ .

A Fig. 38 apresenta o cálculo do Centróide Temporal sobre as mesmas amostras sonoras em que foram extraídos os envelopes na figura anterior.

O Centróide Temporal é utilizado, por exemplo, para distinção entre sons percussivos e sustentados. Envelopes percussivos tendem ao acúmulo de energia próximo ao início do som, resultando em valores de Centróide Temporal baixos, enquanto sons sustentados têm a distribuição da energia mais simétrica no tempo, resultando em valores de Centróides Temporais mais centralizados em relação à duração do som.



**Fig. 38-** Centróide Temporal extraído das amostras sonoras de um evento tocado em uma caixa-clara (gráfico superior), contrabaixo (gráfico central) e violino (gráfico inferior).

### A.1.2 Características Temporais Instantâneas

*Descritores Temporais Instantâneos* são computados do sinal de áudio no domínio do tempo e em uma janela de amostras que é deslocada ao longo de todo o som gravado. A cada janela analisada o descritor retorna um único valor numérico correspondente à medida da característica extraída.

#### A.1.2.1 Média Quadrática (RMS)

O *RMS* (ou *Média Quadrática - Root Mean Square*) é o cálculo da média quadrática dos valores de amplitude em uma janela de amostras. Os valores de RMS extraídos de um evento sonoro descrevem o contorno de seu envelope de energia. A extração do RMS é definida formalmente na Equação A.2.

$$RMS_i = \sqrt{\frac{\sum_{n=0}^{N-1} x_i[n]}{N}} \quad (\text{A.2})$$

Onde  $x_i[n]$  são os valores de amplitude na  $n^{\text{ésima}}$  amostra da  $i^{\text{ésima}}$  janela do áudio digitalizado, e  $N$  é o número de amostras da  $i^{\text{ésima}}$  janela.

### A.1.2.2 Taxa de Cruzamento por Zero

A *Taxa de Cruzamento por Zero (Zero-Crossing Rate - ZCR)* mede no domínio do tempo o número de vezes que um sinal cruza o valor de amplitude zero para em cada janela de amostras. Sinais periódicos tendem a ter poucos cruzamentos por zero enquanto sinais ruidosos (estocásticos) tendem a apresentar altas taxas de cruzamentos por zero.

Para sinais muito simples, como ondas sinusoidais, essa medida é útil para identificação da altura, mas falha para sinais mais complexos. Aplicada em sinais complexos a Taxa de Cruzamento por Zeros fornece indicações sobre a ruídosidade do sinal. Por exemplo, quando aplicado à voz resulta em valores baixos para vogais e altos para consoantes. Este descritor é definido formalmente na Equação A.3.

$$ZCR_i = \frac{1}{2N} \sum_{n=0}^{N-1} |\text{sgn}(x_i[n]) - \text{sgn}(x_i[n-1])| \quad (\text{A.3})$$

Onde,

$$\text{sgn}(n) = \begin{cases} x > 0, & 1 \\ x < 0, & -1 \\ x = 0, & 0 \end{cases} \quad (\text{A.4})$$

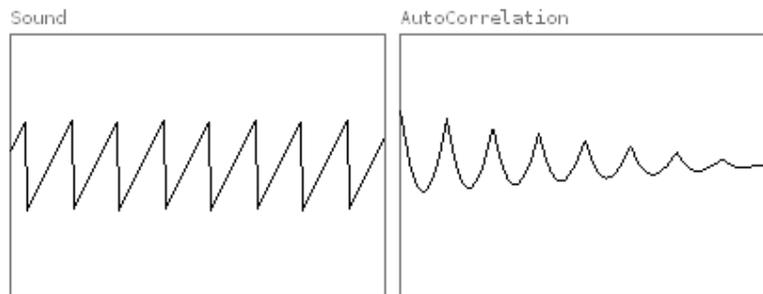
e  $x_i[n]$  são os valores de amplitude na  $n^{\text{ésima}}$  amostra do áudio digitalizado, e  $N$  é o número de amostras da  $i^{\text{ésima}}$  janela.

### A.1.2.3 Autocorrelação

A *Autocorrelação* (*Autocorrelation Function* - ACF) é a medida de correlação de um sinal com ele mesmo. Por sua vez, correlação é a medida de similaridade entre dois sinais (ou duas formas de onda) calculada amostra por amostra, sendo uma das formas de onda estacionária e a outra deslocando no tempo (com atraso). O resultado da autocorrelação é um vetor que representa no domínio do tempo a distribuição espectral do sinal. Aplicada em um sinal periódico a função de autocorrelação retorna valores altos para múltiplos inteiros do período (Fig. 39). É um algoritmo vastamente utilizado em métodos para detecção de altura (*pitch*) e é definido formalmente na Equação A.5.

$$ACF_{\tau} = \frac{1}{N} \sum_{n=0}^{N-\tau-1} x[n]x[n + \tau] \quad (A.5)$$

Onde  $x[n]$  é o sinal de áudio digitalizado,  $N$  o número de amostras e  $\tau$  é a defasagem em número de amostras.



**Fig. 39** - A tabela direita apresenta uma função resultante da autocorrelação da onda dente de serra presente na tabela esquerda.

No processo de detecção de altura seleciona-se o valor de Autocorrelação mais alto e que seja correspondente a um atraso diferente de zero (eixo horizontal do gráfico).

O algoritmo de Autocorrelação definido na Equação A.5 é bastante intensivo computacionalmente e pode ser otimizado quando calculado no domínio das frequências (ver seção A.2). Considerando que a operação de multiplicação no domínio do tempo corresponde a operação de convolução no domínio das frequências (e vice-e-

versa) pode-se computar a Autocorrelação elevando a potência de dois o sinal no domínio das frequências. Esse método é chamado de *Autocorrelação Rápida* e é expresso formalmente na Equação A.6.

$$AFC_{\tau} = IFT \{|X_i[k]|^2\} \quad (A.6)$$

Onde,  $|X_i(k)|$  representa a magnitude da  $k^{ésima}$  componente espectral (*bin*) do  $n^{ésimo}$  quadro é o sinal de áudio, e IFT é a transformada inversa de Fourier.

Na implementação desse algoritmo em PDescriptors utilizamos a transformada rápida de Fourier, que é uma versão otimizada da transformada de Fourier (ver seção A.2).

## A.2 Características Espectrais

*Descritores Espectrais* são computados no domínio da frequência, ou seja, após o sinal de áudio digitalizado passar por uma Transformada Discreta de Fourier (DFT - *Discrete Fourier Transform*). Tipicamente em procedimentos de análise de sinal são aplicadas a Transformada de Fourier sobre janelas de amostras sucessivas e de tamanho invariável. Este procedimento, chamado de *Short-Time Fourier Transform* (STFT), causa uma melhora na resolução temporal, mas diminui a resolução frequencial da transformada. Janelas curtas acentuam este efeito enquanto janelas longas aumentam a resolução frequencial em consequência da perda de resolução temporal. A STFT é definida na Equação A.7.

$$X_i[k] = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x[hn + m]w[m]e^{-\frac{2j\pi mk}{N}} \quad (A.7)$$

Onde  $X_i[k]$  representa a  $k^{ésima}$  componente espectral (*bin*) do  $n^{ésima}$  janela,  $w(m)$  é uma janela de  $N$  amostras, e  $h$  é o número de amostras de interpolação (*hop size*). Estas  $N$  amostras de cada janela de uma STFT são números complexos tendo a

parte real  $X_r$  e a parte imaginária  $X_{im}$ . A magnitude  $|X_i[k]|$  e a fase  $\varphi_i[k]$  do espectro são computadas da seguinte forma:

$$|X_i[k]| = \sqrt{X_r[k]^2 + X_{im}[k]^2} \quad (A.8)$$

$$\varphi_i[k] = \arctan \frac{X_{im}[k]}{X_r[k]} \quad (A.9)$$

Para implementação dos algoritmos em PDescriptors realizamos as Transformadas de Fourier através do objeto interno da linguagem *Pure Data*: **fft~**, que opera uma versão otimizada da STFT chamada Transformada Rápida de Fourier (*Fast Fourier Transform* - FFT) possibilitando que os cálculos sejam realizados em tempo real.

## A.2.1 Características do Envelope Espectral

*Descritores de características do envelope espectral* extraem medidas do sinal de áudio no domínio das frequências que indicam propriedades da distribuição de energia no espectro frequencial.

### A.2.1.1 Centróide Espectral

O *Centróide Espectral* (*spectral centroid*) é definido como o centro de gravidade (a média ou primeiro momento central) da distribuição energética no envelope espectral de um som e é extraído a cada janela de análise. Esta medida é relacionada com a percepção de brilho do sinal. Valores altos indicam predominância energética em frequências agudas caracterizando sonoridades mais brilhantes e valores baixos indicam predominância energética nas frequências graves e menor brilho da sonoridade.

A definição formal deste algoritmo é apresentada na Equação A.10.

$$SC_i = \frac{\sum_{k=1}^{K-1} f[k] |X_i[k]|^2}{\sum_{k=1}^{K-1} |X_i[k]|^2} \quad (A.10)$$

Onde, 
$$f[k] = \frac{\text{taxa de amostragem}}{2K} k \quad (A.11)$$

e  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral (*bin*) do  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*) e  $f[k]$  é a frequência centro do componente  $k$  definida na Equação A.11 .

### A.2.1.2 Desvio Padrão Espectral

Também encontrado na literatura como *spectral spread* ou *largura de banda*, o *Desvio Padrão Espectral* é definido pela raiz quadrada da medida de variância (segundo momento central) da distribuição energética do envelope espectral. Ele mede a dispersão estatística da distribuição de energia espectral, ou seja, a média de dispersão em relação ao centróide e que neste caso pode ser entendido como uma estimação média da largura de banda espectral. A Equação A.12 mostra sua descrição formal.

$$SP_i = \sqrt{\frac{\sum_{k=1}^{K-1} \{f[k] - SC_i\}^2 |X_i[k]|^2}{\sum_{k=1}^K |X_i[k]|^2}} \quad (A.12)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral (*bin*) do  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*),  $f[k]$  é a frequência em Hz centro do componente  $k$ , definida na Equação A.11, e  $SC$  é o Centróide Espectral em Hz definido na Equação A.10.

### A.2.1.3 Obliquidade Espectral

A *Obliquidade Espectral* (*spectral skewness*) fornece uma medida para a assimetria da distribuição em torno da média (centróide). É computada sobre o terceiro momento central da distribuição energética do envelope espectral. Se o valor da obliquidade é igual a zero significa que a distribuição é simétrica. Valores positivos indicam maior energia nas frequências abaixo do centróide e valores negativos indicam maior energia nas frequências acima do centróide. A Equação A.13 apresenta sua descrição formal

$$SSk_i = \frac{\sum_{k=1}^{K-1} \{f[k] - SC_i\}^3 |X_i[k]|^2}{SP_i^3 \sum_{k=1}^{K-1} |X_i[k]|^2} \quad (A.13)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral (*bin*) do  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*),  $f[k]$  é a frequência em Hz centro do componente  $k$ , definida na Equação A.11,  $SC_i$  é o Centróide Espectral em Hz, definido na Equação A.10, e  $SP_i$  é o Desvio Padrão Espectral definido na equação A.12

#### A.2.1.4 Curtose Espectral

Calculado sobre o quarto momento central, a *Curtose Espectral* (*spectral kurtosis*) é uma medida de dispersão que caracteriza o “achatamento” da distribuição energética do envelope espectral. Valores de curtose igual a 3 significam uma distribuição gaussiana da energia no envelope espectral; a curtose maior que 3 significa uma distribuição com picos, ou seja, caracteriza a presença de valores que se afastam da média a vários múltiplos do desvio padrão; valores de curtose menores que 3 indicam uma distribuição da energia mais “achatada ou plana” do que a gaussiana. A descrição formal deste algoritmo apresentada na Equação A.14.

$$SK_i = \frac{\sum_{k=1}^{K-1} \{f[k] - SC_i\}^4 |X_i[k]|^2}{SP_i^4 \sum_{k=1}^{K-1} |X_i[k]|^2} \quad (A.14)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral do  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*),  $f[k]$  é a frequência em Hz centro do componente  $k$  definida na Equação A.11,  $SC_i$  é o centróide espectral em Hz definido na Equação A.10 e  $SP_i$  é o desvio padrão espectral definido na equação A.12

#### A.2.1.5 Nivelamento Espectral

A *Medida de Nivelamento Espectral* (*Spectral Flatness Measure*) indica o quanto homogênea é a distribuição de energia no espectro, ou seja, o quanto o envelope espectral é próximo de ‘plano’. Pode ser entendida também como uma medida da relação ruído/som-tonal. Uma distribuição de energia homogênea no espectro corresponde a um sinal contendo ruído branco e uma distribuição com picos é relativa à presença de sons tonais. O algoritmo retorna valores baixos para o primeiro caso e altos

para o segundo. Essa medida é computada pela razão entre a média geométrica e a média aritmética da energia em uma banda espectral (Equação A.15).

$$SFM_i = \frac{(\prod_{q=1}^Q |X_i[q]|)^{\frac{1}{Q}}}{\frac{1}{Q} \sum_{q=1}^Q |X_i[q]|} \quad (A.15)$$

Onde  $|X_i[q]|$  designa a magnitude da banda espectral  $q$  na janela  $i$  que possui  $Q$  bandas no total. A Medida de Nivelamento Espectral é usualmente extraída em função de bandas do espectro. Jehan (2001) propõe a extração desta medida em função das bandas da escala Bark, enquanto Peeters (2004) propõe uma extração para 4 bandas de oitavas com início em 250Hz e fim em 4000Hz.

A Medida de Nivelamento Espectral pode ser usada para se obter o *Coefficiente de Tonalidade (tonality coefficient)* [ou *Ruidosidade (noisiness)*].

$$SFM_{db} = 10 \log_{10}(SFM[n]) \quad (A.16)$$

$$Tonality = \min\left(\frac{SFM_{db}}{-60}, 1\right) \quad (A.17)$$

O Coeficiente de Tonalidade retorna valores próximos de 1 para sinais tonais e próximos de 0 para sinais ruidosos.

#### A.2.1.6 Medida da Crista Espectral

Outro descritor que extrai características da forma do envelope espectral é a *Medida da Crista de Espectral (Spectral Crest Measure - SCM)* que calcula a razão entre a componente de máxima energia em uma banda espectral e a média de energia na banda. Ou seja, esse descritor calcula o quão grande é o pico de energia em uma banda espectral. A formalização deste algoritmo é apresentada na Equação A.18

$$SCM_i = \frac{\max |X_i[q]|}{\frac{1}{Q} \sum_{q=1}^Q |X_i[q]|} \quad (A.18)$$

Onde  $|X_i[q]|$  designa a magnitude da banda espectral  $q$  na janela  $i$  que possui  $Q$  bandas no total.

### A.2.1.7 Irregularidade Espectral

Também encontrado na literatura com o nome de *Spectral Smoothness*, o algoritmo para extração da *Irregularidade Espectral* (Spectral Irregularity - SI) mede o grau da diferença entre magnitudes de componentes espectrais adjacentes. Essa medida descreve se o contorno da envoltória espectral é suave ou se apresenta picos, ou seja, com característica ‘endentada’. Equação A.19 apresenta a descrição formal do algoritmo em que a média da energia dos componentes espectrais atual, anterior e posterior é comparada com a energia do quadro atual.

$$SI_i = \sum_{k=2}^{K-1} \left| 20 \log(|X_i[k]|) - \frac{20 \log(|X_i[k-1]|) + 20 \log(|X_i[k]|) + 20 \log(|X_i[k+1]|)}{3} \right| \quad (A.19)$$

Jensen (1999) propõe outra versão para algoritmo descrita na equação A.20

$$SI'[n] = \frac{\sum_{k=1}^{K-1} \{|X_i[k]| - |X_i[k-1]|\}^2}{\sum_{k=1}^{K-1} |X_i[k]|^2} \quad (A.20)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*)

### A.2.1.8 Inclinação Espectral

A *Inclinação Espectral* (*Spectral Slope*) é uma estimação do decrescimento da energia espectral computada por regressão linear no espectro de magnitudes. Sua definição formal é apresentada na Equação A.21

$$SSl_i = \frac{1}{\sum_{k=1}^{K-1} |X_i[k]|} \frac{N \sum_{k=1}^K f[k] |X_i[k]| - \sum_{k=1}^K f[k] \sum_{k=1}^K |X_i[k]|}{N \sum_{k=1}^{K-1} f[k]^2 - (\sum_{k=1}^{K-1} f[k])^2} \quad (A.21)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*),  $f[k]$  é a frequência centro do componente  $k$  e medida em Hz

### A.2.1.9 Decrescimento Espectral

O algoritmo para extração do *Decrescimento Espectral (Spectral Decrease)* (Equação A.22) tem característica similar com a *Inclinação Espectral*, pois também representa o decrescimento de energia no espectro de magnitude. De acordo com Peeters (2004) este descritor é derivado de estudos psicoacústicos e é mais correlato com a percepção humana que o algoritmo da subseção anterior.

$$SD_i = \frac{1}{\sum_{k=2}^{K-1} |X_i[k]|} \sum_{k=2}^{K-1} \frac{|X_i[k]| - |X_i[1]|}{k - 1} \quad (A.22)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral da  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*).

#### A.2.1.10 Roll-off

O ponto de *Roll-off* é definido como um valor de frequência que assinala o ponto em que a somatória da energia dos componentes espectrais abaixo desse ponto contém  $X\%$  do total da energia do espectro. É usual para diferir sons com grande quantidade de energia nos transientes de ataque de alta frequência dos sons mais constantes e com predominância de energia em componentes espectrais de baixa frequência, como por exemplo, para diferir vogais e consoantes da voz humana. Sua definição formal é apresentada na Equação A.23

$$\sum_{k=1}^{R_i} |X_i[k]|^2 f[x] = b \sum_{k=1}^{K-1} |X_i[k]|^2 f[x] \quad (A.23)$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral da  $i^{ésima}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*),  $f[k]$  é a frequência em Hz centro do componente  $k$  definida na Equação A.11,  $R_i$  é a frequência do ponto de *Roll-off*, e  $b$  é a porcentagem do espectro de magnitudes que define o ponto de *Roll-off*. Tipicamente  $b$  é definido entre 0.85 e 0.95.

#### A.2.1.11 Conteúdo de Frequências Altas

O Descritor *Conteúdo de Altas Frequências (High Frequency Content - HFC)* pondera linearmente a contribuição de cada componente (*bin*) de energia espectral

de acordo com seu valor em frequência. Ou seja, as energias dos componentes de frequências altas são realçadas em relação às de frequências baixas. A descrição formal do algoritmo é apresentada na Equação A.24.

$$HFC_i = \sum_{k=1}^{K-1} k \cdot |X_i[k]|^2 \quad (\text{A.24})$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*).

Este descritor é funcional para localização de *onsets* de eventos musicais, pois ressalta a energia dos transientes de alta frequência em ataques, principalmente de instrumentos percutidos.

## A.2.2 Características da Dinâmica Espectral

Nesta classe de descritores são medidas as relações entre a distribuição de energia e dinâmica de fases entre sucessivas janelas de análise espectral. Este tipo de descritor é utilizado principalmente em métodos para detecção de início de eventos sonoros gerados por instrumentos musicais (detecção de *onset*) que é uma característica base para transcrição e análise rítmica de um sinal musical.

### A.2.2.1 Fluxo Espectral

O *Fluxo Espectral* mede a diferença de magnitude entre sucessivos quadros de análise espectral. Retorna valores baixos quando o sinal é estático, ou tem poucas alterações de energia entre uma janela de análise e outra, e valores altos quando o envelope espectral sofre grandes alterações. São encontradas diversas variações desse algoritmo, algumas delas foram implementadas em PDescriptors.

A primeira é definida pela subtração entre janelas de análise consecutivas. Sua definição formal é apresentada na Equação A.25.

$$SF_i = \sqrt{\sum_{k=1}^{K-1} \{|X_i[k]| - |X_{i-1}[k]|\}^2} \quad (\text{A.25})$$

Onde  $|X_i(k)|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela e  $K$  é a metade do número de amostras da janela de análise (*FFT size*).

A segunda variação deste algoritmo é chamada de *Fluxo Espectral Positivo*. O cálculo deste algoritmo é apresentado na Equação A.26 e também pode ser encontrado nos trabalhos de Bello et al. (2005) e Dixon (2006). O Fluxo Espectral Positivo é definido pelo acréscimo de uma função de retificação de meia onda positiva na Equação A.26. Este procedimento implica a anulação de valores negativos resultantes da subtração e como consequência a seleção dos valores positivos que representam aumentos de energia entre janelas de análise. Na detecção de eventos musicais (notas e acordes) emitidos por instrumentos tradicionais, as regiões de picos do Fluxo Espectral estão no início (*onset*) e fim (*offset*) de cada evento. O Fluxo Espectral Positivo seleciona apenas os inícios de eventos onde há rápido aumento da energia.

$$SF_i^+ = \sqrt{\sum_{k=1}^{K-1} \{H^+(|X_i[k]| - |X_{i-1}[k]|)\}^2} \quad (\text{A.26})$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela,  $K$  é metade do número de amostras da janela de análise (*FFT size*) e  $H^+(x)$  é a função de retificação de meia onda positiva:

$$H^+(x) = \frac{x + |x|}{2} \quad (\text{A.27})$$

A terceira, quarta e quinta variação, têm como referência a documentação sobre extração do Fluxo Espectral para detecção de *onset* na página do *Mazurka Project*<sup>28</sup>. A terceira variação do algoritmo de Fluxo Espectral segue o mesmo raciocínio do anterior, mas com propósito inverso. Chamado de *Fluxo Espectral Negativo* este algoritmo implementa a função de retificação de meia onda negativa. Sua formalização é apresentada na Equação A.28.

$$SF_i^- = \sqrt{\sum_{k=1}^{K-1} \{H^-(|X_i[k]| - |X_{i-1}[k]|)\}^2} \quad (\text{A.28})$$

---

<sup>28</sup> Mazurka Project: <http://www.mazurka.org.uk/software/sv/plugin/MzSpectralFlux/> (consultado em 15/06/2011)

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela,  $K$  é a metade de número de amostras da janela de análise (*FFT size*) e  $H(x)$  é a função de retificação de meia onda negativa:

$$H^-(x) = \frac{x - |x|}{2} \quad (\text{A.29})$$

O Fluxo Espectral Negativo não é interessante por ele mesmo para detecção de início de evento, mas sim combinado com o Fluxo Espectral Positivo para a quarta variação do algoritmo chamada de *Diferença de Fluxo Espectral*, apresentada na Equação A.30.

$$SF_i^\Delta = H^+(SF_i^+ - SF_i^-) \quad (\text{A.30})$$

Onde  $SF^+$  representa o Fluxo Espectral Positivo (Equação A.26),  $SF^-$  representa o Fluxo Espectral Negativo (Equação A.28) e  $H^+$  a função de retificação de onda positiva (Equação A.27)

Finalmente, a Diferença de Fluxo Espectral é aplicada na construção do *Fluxo Espectral Composto* que é apresentado na Equação A.31

$$SF_i^C = \frac{SF_i^\Delta}{|SF_i^+ - SF_i^-|} \quad (\text{A.31})$$

Onde  $SF^+$  representa o Fluxo Espectral Positivo (Equação A.26),  $SF^-$  representa o Fluxo Espectral Negativo (Equação A.28) e  $SF^\Delta$  a Diferença de Fluxo Espectral (Equação A.30)

A sexta variação do Fluxo Espectral é obtida pela diferença quadrática entre o logaritmo das magnitudes de janelas sucessivas. Esse procedimento minimiza influências na variação de amplitude do sinal. O algoritmo pode ser visto no trabalho de Barbedo e Lopes (2006) e é definido pela Equação A.32

$$SF_i^{lg} = \sum_{k=1}^{K-1} \{\log_{10}(|X_i[k]|) - \log_{10}(|X_{i-1}[k]|)\}^2 \quad (\text{A.32})$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela,  $K$  é a metade do número de amostras da janela de análise (*FFT size*).

Por fim, a sétima variação do Fluxo Espectral utiliza a correlação normalizada entre quadros de análise sucessivos, ao invés da sua diferença. Esse algoritmo, chamado de *Fluxo de Correlação Espectral*, é definido na Equação A.33 e é encontrado também no trabalho de Peeters (2004).

$$SF_i^{cr} = 1 - \frac{\sum_{n=1}^{N-1} |X_i[k]| |X_{i-1}[k]|}{\sqrt{\sum_{k=1}^{K-1} |X_i[k]|^2} \sqrt{\sum_{k=1}^{K-1} |X_{i-1}[k]|^2}} \quad (\text{A.33})$$

Ond  $|X_i[k]|$  representa a magnitude da  $k^{ésima}$  componente espectral do  $i^{ésima}$  janela,  $K$  é a metade do número de amostras da janela de análise (*FFT size*).

O Fluxo de Correlação Espectral resulta o valor 0 quando as janelas espectrais sucessivos são iguais, e resulta o valor 1 quando são muito distintas.

#### A.2.2.2 Desvio de Fase

Este Descritor calcula a diferença de fase entre sucessivos quadros de análise espectral. O cálculo do *Desvio de Fase* é descrito pelas Equações A.34, A.35 e A.36

A taxa de alteração na fase de um componente de uma STFT é relativa à estimação de sua frequência instantânea. Considerando  $\varphi_i[k]$  a fase correspondente a  $|X_i[k]|$  (Equação 2.9.9), e  $-\pi < \varphi_i[k] \leq \pi$ , a frequência instantânea é dada por  $\varphi'_i[k]$

$$\varphi'_i[k] = \varphi_i[k] - \varphi_{i-1}[k] \quad (\text{A.34})$$

Em que  $-\pi < \varphi'_i[k] \leq \pi$ . Então, a alteração na frequência instantânea é dada pela segunda diferença de fase  $\varphi''_i[k]$ :

$$\varphi''_i[k] = \varphi'_i[k] - \varphi'_{i-1}[k] \quad (\text{A.35})$$

A equação A.36 descreve média dos Desvios de Fases resultantes na Equação A.35 e é um algoritmo eficiente para detecção de *onsets* de eventos musicais que não possuem um ataque suave sem um grande aumento de energia ou presença de transientes. A diferença de fase neste caso é um bom indicativo para detecção de mudança de altura.

$$PD_i = \frac{\sum_{k=1}^K |\varphi_i''[k]|}{N} \quad (\text{A.36})$$

Dixon (2006) propõe uma versão da Equação A.37 ponderada pelas magnitudes do espectro e normalizada:

$$WPD_i = \frac{\sum_{k=1}^K |\varphi_i''[k]| |X_i[k]|}{\sum_{k=1}^K |X_i[k]|} \quad (\text{A.37})$$

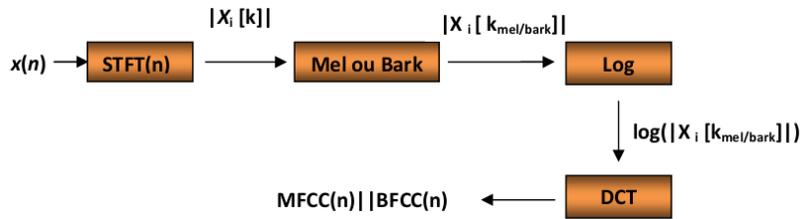
Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral do  $i^{\text{ésima}}$  janela,  $K$  é a metade do número de amostras da janela de análise (*FFT size*)

### **A.2.3 Descrição Global do Envelope Espectral Segundo Preceitos Psicoacústicos: MFCC e BFCC**

Algoritmos para descrição global do envelope espectral retornam um vetor de dados numéricos que representa as magnitudes dos componentes espectrais de um sinal. É usual a prática de alteração da representação do contorno espectral para que os dados sejam facilmente manipulados, em determinados casos, ou destaquem certas características do contorno espectral, como por exemplo, é o caso da representação cepstral. Para alguns casos é interessante a aplicação de modelos psicoacústicos a descrição do contorno espectral, pois aproxima sua descrição com características da resposta em frequência do sistema auditivo humano e diminui a quantidade de dados necessários para descrição do envelope espectral.

Os *Coeficientes Cepstrais por Frequencias Mel (Mel Frequency Cepstral Coeficients - MFCC)* e os *Coeficientes Cepstrais por Frequencias Bark (Bark Frequency Cepstral Coeficients - BFCC)* são representações do espectro sonoro filtrados por escalas que modelam propriedades da percepção auditiva humana e são obtidas através de experimentos psicoacústicos. A escala Mel divide o espectro auditivo humano em alturas perceptivamente equidistantes. A escala Bark possui 24 valores relacionados às frequências limites das bandas críticas da percepção sonora. A filtragem dos coeficientes espectrais por essas escalas fornece uma redução de dados para a descrição global do envelope espectral, de acordo com esses respectivos modelos da percepção sonora. O cepstro é definido como a transformada de Fourier do logaritmo do

espectro de magnitudes de um sinal. O processo de extração destes dois descritores possui várias etapas e é descrito na Fig. 40



**Fig. 40** - Fluxograma do algoritmo para extração de MFCC e BFCC.

Os algoritmos para o cálculo de BFCC e MFCC são similares, diferem-se apenas no tipo de escala usado para filtragem do espectro de magnitudes do sinal (Mel ou Bark). O detalhamento do algoritmo de filtragem do espectro de magnitudes por tais escalas derivadas de modelos psicoacústicos é apresentado nas seções A.3.2 e A.3.3, respectivamente para o MFCC e para o BFCC.

No cálculo do MFCC ou BFCC substitui-se a segunda transformada de Fourier (para obtenção dos coeficientes cepstrais) pela Transforma Discreta do Cosseno (*Discrete Cosine Transform*) definida na Equação A.38

$$DCT_i(j) = \sum_{k_{band}=1}^{T_{band}} |X_i[k_{band}]| \cos \left[ \frac{\pi}{T_{band}} j \left( k_{band} - \frac{1}{2} \right) \right] \quad (A.38)$$

Onde  $|X_i[k_{band}]|$  representa a  $k_{band}$  <sup>ésima</sup> banda mel ou bark da  $i$  <sup>ésima</sup> janela;  $T_{band}$  é o número total de bandas mel ou bark ; e  $i = 1, 2, \dots T_{band}$

#### A.2.4 Característica Espectral Segundo Preceitos Musicais: Chroma

O *Chroma* pertence a uma classe de algoritmos de análise e representação espectral que é construída sobre conhecimento prévio de aspectos de organização musical (especificamente sob um aspecto da percepção sonora humana que influencia neste tipo de estruturação), enquanto os demais algoritmos aqui apresentados são baseados em propriedades do sinal acústico digitalizado ou em modelos da percepção auditiva humana.

A extração do Chroma, também encontrada com o nome originalmente dado por Fujishima (1999) *Pitch Class Profile* (PCP), é baseada no fenômeno de equivalência em oitavas da percepção humana das notas musicais. A percepção de notas é entendida com tendo duas dimensões: 1) o *chroma* que é a relação de equivalência da percepção de mesma nota para sons distanciados por oitavas; 2) a *altura* (*Height*) que está diretamente associado ao aumento e diminuição absolutos da frequência no âmbito do espectro freqüencial percebido por humanos. Este modelo é ilustrado pela figura Fig. 3 (subseção 2.1.3) como uma espiral onde a representação do parâmetro altura está associada ao eixo vertical, e a rotação da hélice sob a divisão de escala cromática representa o Chroma.

O algoritmo de extração do Chroma retorna um vetor de 12 valores correspondendo à energia do sinal para cada nota da escala cromática. Para obtenção deste vetor o espectro é dividido em bandas com larguras correspondentes a 1 semitom e frequência centro no valor das notas do temperamento (sendo que a percepção de notas é de ordem logarítmica, a largura das bandas e os intervalos entre as frequências centro aumentam conforme se aumenta o parâmetro de altura, i.e oitava), em seguida, para cada nota são somados os valores de energia das respectivas bandas presentes em todas as oitavas. A formalização do algoritmo segundo descrição de Sheh e Ellis (2003) é apresentada na Equação A.39.

$$Chroma_p = \sum_{K:p(k)=p} |X_i[k]|^2 \quad A.39$$

Onde ,  $p = 0, 1, \dots, 11$ ;

$$e \quad p[k] = 1 \left[ 2 \log_2 \left( \frac{k f_{sr}}{N f_{ref}} \right) \right] \text{mod} 12 \quad A.40$$

Onde  $f_{sr}$  é a frequência de amostragem e  $f_{ref}$  é a frequência relativa ao *Chroma* [0], ou seja, a frequência correspondente a primeira nota do vetor.

É usual realizar os cálculos do Chroma com divisões menores que semitons, como por exemplo, em quartos-de-tom ( $p = 0, \dots, 24$ ) ou sextos-de-tom ( $p = 0, \dots, 36$ ). O

vetor de Chroma é utilizado em trabalhos que envolvem reconhecimentos de estruturas harmônicas como acordes, *pitch-class*, e tonalidade.

### **A.3 Características Psicoacústicas**

Descritores de características psicoacústicas são baseados em estudos de modelos matemáticos para representação de características da percepção auditiva humana. Os algoritmos dessa classe de Descritores convertem o sinal de áudio em funções que representam a resposta auditiva de certas características do sistema auditivo humano.

### A.3.1 Loudness

O *loudness* é uma medida psicoacústica relativa à percepção de intensidade sonora. Existem diversos modelos matemáticos para a percepção da intensidade levando em conta estudos sobre as propriedades físicas do sistema auditivo. Um deles, visto no trabalho de Pereira (2009), é baseado na modelagem da resposta em frequência dos ouvidos médio e externo e atua como uma função de ponderação  $W(k)$  que atenua ou enfatiza os componentes espectrais de acordo com o modelo de resposta em frequência do sistema auditivo (Equação A.41).

$$W[k] = -0,6 \cdot 3.64 \cdot f[k]^{-0.8} + 6.5 \cdot e^{-0.6 \cdot (f[k]-3.3)^2} - 10^{-3} f[k]^4 \quad (\text{A.41})$$

Onde  $f[k]$  (definida na Equação A.11) é a frequência centro em Hz do componente espectral  $k$ .

Desta forma o *loudness*  $L_i$  de uma janela de análise espectral é dado por:

$$L_i = \sum_{k=1}^{K-1} |X_i[k]|^2 10^{W[k]/20} \quad (\text{A.42})$$

Onde  $|X_i[k]|$  representa a magnitude da  $k^{\text{ésima}}$  componente espectral da  $i^{\text{ésima}}$  janela,  $K$  é metade de número de amostras da janela de análise (*FFT size*)

### A.3.2 Energia Espectral por Banda da Escala Mel

A *Escala Mel* foi elaborada através de estudos de psicoacústica realizados por Stevens, Volkman e Newman (1937) e divide o espectro auditivo em uma escala de alturas equidistantes segundo critérios perceptivos. O nome ‘Mel’ vem da palavra ‘melodia’ e indica que a escala é baseada em comparações entre alturas. A Equação A.43 é uma função que converte valores de frequências em Hertz para a escala mel.

$$MEL(f) = 2595 \log_{10} \left( 1 + \frac{f}{700} \right) = 1127 \log_e \left( 1 + \frac{f}{700} \right) \quad (\text{A.43})$$

Onde  $f$  representa valores de frequência em Hz.

Para medir a energia espectral sob a Escala Mel, usualmente filtram-se as magnitudes dos componentes espectrais em  $k_{mel}$  filtros Mel. Filtros Mel têm formato triangular e são equidistantes em termos de frequências Mel. Cada componente espectral com frequência centro em Hertz e inclusa numa das bandas Mel tem sua energia multiplicada pela resposta de magnitude do filtro. Por fim, soma-se o quadrado da energia de todos componentes daquela banda.

### A.3.3 Energia Espectral por Banda da Escala Bark

A *Escala Bark* é uma escala psicoacústica proposta por Zwicker (1961) que divide o espectro auditivo humano em 24 valores que correspondem as frequências limites das bandas críticas da percepção sonora. A equação A.44 apresenta a função que mapeia valores de frequência em Hertz para valores da escala Bark.

$$BARK(f) = 13 \arctan\left(\frac{0.76f}{1000}\right) + 3.5 \arctan\left(\left(\frac{f}{7500}\right)^2\right) \quad (A.44)$$

Onde  $f$  representa valores de frequência em Hz.

Para extração da energia espectral por banda da escala Bark somam-se as magnitudes dos componentes espectrais com frequência centro (em Hz) com mesmo valor segundo a escala Bark, de acordo com a função A.44 para conversão entre as escalas.