



Universidade Federal de Pernambuco
Centro de Informática

Pós-graduação em Ciência da Computação

**DESCOBERTA AUTOMÁTICA DE
CONHECIMENTO EM INTERPRETAÇÕES
MUSICAIS: O CASO DO
ACOMPANHAMENTO RÍTMICO AO
VIOLÃO**

Ernesto Trajano de Lima Neto

TESE DE DOUTORADO

Recife
08 de março de 2007

Universidade Federal de Pernambuco
Centro de Informática

Ernesto Trajano de Lima Neto

**DESCOBERTA AUTOMÁTICA DE CONHECIMENTO EM
INTERPRETAÇÕES MUSICAIS: O CASO DO
ACOMPANHAMENTO RÍTMICO AO VIOLÃO**

*Trabalho apresentado ao Programa de Pós-graduação em
Ciência da Computação do Centro de Informática da Uni-
versidade Federal de Pernambuco como requisito parcial
para obtenção do grau de Doutor em Ciência da Com-
putação.*

Orientador: *Prof. Dr. Geber Lisboa Ramalho*

Recife
08 de março de 2007

AGRADECIMENTOS

Em primeiro lugar, agradeço a Geber, meu orientador, pelos conselhos e, principalmente, pela paciência, especialmente neste último ano.

A todos os colegas e amigos no CIn, pela ajuda. Em especial a Marcio, que me ajudou muito e em vários momentos. Mas o agradecimento vai mesmo é pelas nossas longas conversas, que começavam sérias e centradas em problemas de nossas teses, mas que logo descambavam para assuntos aleatórios, todavia, extremamente interessantes. Isto quando não íamos parar lá na cantina de Dona Deda conversando sobre sei-lá-o-quê! Ou melhor, no coreano, comendo carne de cachorro!

Ao Centro de Informática por todo o apoio e infraestrutura.

Ao Departamento de Música da Universidade Federal de Pernambuco, em especial aos professores Carlos Sandroni e Osman Gioia, que possibilitaram a gravação das obras utilizadas no decorrer da pesquisa, sem as quais muito pouco poderia ter sido feito.

Ao CNPq pelo apoio financeiro, incluindo a bolsa de doutorado sanduíche. À CAPES pela instauração do portal periódicos, que é uma verdadeira “mão na roda”. Que ele sempre esteja disponível para a comunidade científica nacional.

Ich möchte mich bei Gerhard bedanken, der mir sehr gute wissenschaftliche Hinweise gegeben hat. Auch die Hinweise über Leberkäse haben mir sehr viel geholfen! Liebe Grüße an die Kollegen vom ÖFAI, sowie an die Ex-ÖFAIers, bzw. Arthur, Asmir, Elias, Paolo, Simon, Søren, Tim, und Werner. Bei euch habe ich ja viel gelernt. Ich bedanke mich auch bei Isabella, die mir sehr, sehr viel geholfen hat.

Este projeto de pesquisa foi premiado pelo Instituto Itaú Cultural através de seu Projeto Rumos Pesquisa no ano de 2003. Sou muito grato pelo auxílio advindo deste prêmio.

RESUMO

Qualquer músico, com um mínimo de experiência, sabe que tocar uma música da forma exata como ela está grafada resulta em algo mecânico e artificial. De fato, a representação simbólica da música ou *notação* serve apenas como um guia para o músico, que a modifica de acordo com suas intenções musicais, com um conjunto de regras, aprendidas durante o decorrer de sua vida musical e, em sua grande maioria, não explícitas, e de acordo com uma análise, ainda que superficial e intuitiva, da obra a ser interpretada. À forma como a música é variada pelo músico dá-se o nome de *interpretação*.

Diversas pesquisas estudam variados aspectos deste fenômeno. Resultados mostram que, apesar de tratar-se de uma forma de arte, existem aspectos em comum nos diferentes intérpretes e em diferentes interpretações. Estas pesquisas concentram-se, todavia, na Música Clássica composta para piano, o que exclui partes significativas do repertório musical, em especial, a Música Popular Brasileira, em estilos como a Bossa Nova e o Samba, e em outros instrumentos como, por exemplo, o violão. Por outro lado, na Computação Musical, diversos pesquisadores têm desenvolvido métodos para a análise automática de interpretações. Dentre eles, alguns utilizam a Aprendizagem de Máquina e a Mineração de Dados, mas ainda restringem-se à análise da Música Clássica composta para piano.

A despeito dessas pesquisas, existe um grande número de questões sobre a interpretação que permanecem sem resposta. No presente trabalho, explorou-se um subconjunto particular dessas questões, a saber: questões que tratam de como o acompanhamento rítmico ao violão na Bossa Nova é construído pelos intérpretes. Neste novo escopo, o trabalho de descoberta automática de conhecimento musical partiu de um elemento estrutural unificador, o padrão rítmico, e estabeleceu um conjunto de processos computacionais e algoritmos para a extração automática que forneceram elementos para uma melhor compreensão do fenômeno.

Resultaram da pesquisa as seguintes contribuições: um algoritmo para a derivação do dedilhado da mão direita do violonista; uma representação textual desse dedilhado, assim como um processo para a redução da dimensionalidade dessa representação; adaptações de algoritmos de extração de padrões; além de um conjunto de dados tratados e de ferramentas de análise automática.

Palavras-chave: Inteligência Artificial, Computação Musical, Análise Inteligente de Dados, Aprendizagem de Máquina, interpretação musical, ritmo, padrão rítmico, Bossa Nova

ABSTRACT

It is common sense that playing music in the exact way it is written in the score results in a mechanical and uninteresting succession of sounds. In fact, the score is just a guide to the musician, who changes it according to his musical intentions, to some sort of empirical analysis of the musical structure and performance rules he learned during his musical life. These modifications are called *expressive performance*.

Several researchers study different aspects of expressive performance. Results show that, although expressive performance is an art form, it still has common features across performers and performances. However, these results are restricted to Classical Music composed for the piano. Musical styles such as Bossa Nova and Samba and instruments like the acoustic guitar are rarely studied. In Computer Music, other researchers are concerned with the automatic analysis of expression. Some of them use techniques from fields such as Machine Learning and Data Mining in order to build models of musical expression. Nevertheless, they still analyze Classical Music composed for the piano.

Despite all efforts, there still is a great number of unanswered questions about expressive performance. In this work, we explored a particular subset of these questions: Actually, how Bossa Nova's rhythmic accompaniment is built by the performers. Given this new scope, we used a single unifying structural element, the rhythmic pattern, and established computational processes and algorithms whose results provided a better understanding of the phenomenon.

The work resulted in the following contributions: an algorithm for the induction of the guitarist's right hand fingering; a textual representation of the fingering, as well as a process for dimensionality reduction of this fingering; pattern extraction algorithm's adaptations; and a prepared data set, as well as computational tools for the automatic rhythmic analysis.

Keywords: Artificial Intelligence, Computer Music, Intelligent Data Analysis, Machine Learning, expressive performance, rhythm, rhythmic pattern, Bossa nova

SUMÁRIO

Capítulo 1—Introdução	1
1.1 O Problema	4
1.2 Objetivos	6
1.3 Abordagem	8
1.4 Requisitos	9
1.5 Contribuições	11
1.6 Estrutura do Documento	12
Capítulo 2—Interpretação Musical	13
2.1 Caracterização do Problema	14
2.2 A Pesquisa em Interpretação Musical	18
2.3 Da Necessidade do Computador	21
2.4 Análise Automática de Interpretações Musicais	25
2.4.1 O Projeto de Interpretação Musical Vienense	25
2.4.2 A Pesquisa na Suécia: Comunicação Musical	30
2.4.3 Outras Pesquisas	32
2.5 Discussão	34
Capítulo 3—Construção de Modelos da Interpretação Musical	37
3.1 Extração de padrões	39
3.2 Extração de Padrões Musicais	40
3.2.1 GCTMS: <i>General Computational Theory of Musical Structure</i>	46
3.2.2 FLEXPAT	48
3.2.3 SimilaritySegmenter	53
3.2.4 Outros	54
3.2.5 Discussão	56
3.3 Descoberta Automática de Conhecimento Musical	58
3.3.1 Aprendizagem de Máquina e Análise de Interpretações	62
Capítulo 4—Análise Automática do Acompanhamento Musical	65
4.1 Um Processo para a Análise do Acompanhamento	67
4.2 Aquisição de Dados	70
4.3 Preparação dos Dados	71
4.4 Representação dos Dados	73

4.4.1	Determinação da Pulsação e Quantização	74
4.4.2	Dedilhado da Mão Direita e Abstrações Derivadas	75
4.4.3	Abstração e Representação dos Dados	80
Capítulo 5—Experimentos e Protótipo		85
5.1	Extração de Padrões	85
5.1.1	Avaliação dos Algoritmos de Extração	85
5.1.2	Experimento 1: Casamento Exato	87
5.1.3	Experimento 2: Casamento Inexato	88
5.1.4	SFIEXPath: uma versão simplificada de FIEXPath	93
5.1.5	Etapa de Extração de Padrões: Discussão dos Resultados	96
5.2	Ferramenta de Auxílio à Análise	97
5.3	Considerações Finais	102
Capítulo 6—Conclusões		105
6.1	Contribuições	106
6.2	Limitações e Trabalhos Futuros	108
Apêndice A—Catálogo de Padrões		125
	Padrão Principal	125
	Padrões Cíclicos	125
	Padrões de Início de Música	126
	Padrões Especiais	126
	Padrões de Virada	127
Apêndice B—SFIEXPath: Padrões Induzidos		129
B.1	Padrões do Intérprete 1	129
	A Felicidade	129
	Desafinado	130
	Eu Sei Que Vou Te Amar	130
	Garota De Ipanema	131
	Insensatez	131
	Samba De Uma Nota Só	131
	Tarde Em Itapuã	132
	Wave	132
B.2	Padrões do Intérprete 2	132
	Barquinho	132
	Bim Bom	133
	Garota De Ipanema	133
	Insensatez	133
	Só Danço Samba	133
	Wave	133

LISTA DE FIGURAS

2.1	O percurso da música: do compositor ao ouvinte	13
2.2	Exemplos de cifra	15
2.3	Exemplo de partitura	16
2.4	Influências na interpretação	19
2.5	Visão geral do projeto austríaco	26
2.6	O Verme Interpretativo em ação	29
2.7	A análise-por-síntese	31
3.1	Duas representações para o início de <i>Desafinado</i>	42
3.2	Repetição inexata de um padrão na <i>Appassionata</i>	45
3.3	As operações de fragmentação e consolidação	46
3.4	A GCTMS	47
3.5	Etapas do algoritmo FLEXPAT	49
3.6	Exemplo de um subgrafo estrela	51
3.7	Exemplo de parte de um grafo utilizado por SIMILARITYSEGMENTER	54
3.8	Exemplo de padrões encontrados pelo algoritmo SIA	56
3.9	O ciclo de descoberta de conhecimento	60
4.1	Processo de descoberta de conhecimento sobre o acompanhamento rítmico	69
4.2	Exemplo de regra de filtragem	72
4.3	Exemplo de correção manual	73
4.4	Exemplo de quantização	75
4.5	Acorde de dó maior	77
4.6	Conjunto \mathcal{P} utilizado na avaliação do algoritmo.	79
4.7	Dedilhado obtido para o início da canção <i>Insensatez</i>	80
4.8	Dedilhado obtido para o início da canção <i>Eu Sei Que Vou Te Amar</i>	80
4.9	Representação completa da canção	81
4.10	Dedilhado e representação monofônica da canção <i>Insensatez</i>	82
4.11	Representação dos dados	83
5.1	Exemplos de padrões rítmicos do catálogo	86
5.2	Pequenos desvios nos padrões	88
5.3	Padrão encontrado por FLEXPAT	92
5.4	Componentes dos sistema	98
5.5	Dedilhado em sua versão não-segmentada	99
5.6	Dedilhado em sua versão segmentada	99
5.7	Entrada de parâmetros para FLEXPAT (A) e SFLEXPAT (B)	100

5.8	Grafo de equipolência	100
5.9	Exemplo de experimento utilizando a ferramenta <i>KnowledgeFlow</i>	102

CAPÍTULO 1

INTRODUÇÃO

As tentativas de estudo da Música através de máquinas vêm de longa data. Quando Charles Babage, nos idos de 1842, concebeu a sua Máquina Analítica, *Lady Ada Lovelace* fez a seguinte observação acerca de suas possibilidades (*apud* [Rich e Knight, 1993, p. 7–8]):

“O mecanismo operacional pode até entrar em ação independentemente de ter um objeto no qual operar (é claro que neste caso nenhum resultado poderá ser desenvolvido). Novamente, ele pode operar sobre outras coisas que não números, caso haja objetos cujas relações mútuas fundamentais possam ser expressas pela ciência abstrata das operações e que devam também ser suscetíveis a adaptações, à ação da notação operacional e ao mecanismo da máquina. Supondo-se, por exemplo, que as relações fundamentais dos sons na ciência da harmonia e da composição musical sejam suscetíveis a tais expressões e adaptações, a máquina poderia então compor peças de música elaboradas e científicas em qualquer grau de complexidade ou extensão”.

—ADA LOVELACE

Foi, no entanto, apenas com o advento dos computadores, após a Segunda Guerra Mundial, que idéias como essa puderam, de fato, ser postas em prática.

Um dos motivos deste crescente interesse é que a Música oferece aos pesquisadores um conjunto de problemas cujas formalização e investigação via computador são bastante complexas [Balaban et al., 1992]. Na Música, estão em jogo, ao mesmo tempo, elementos subjetivos, tais como criatividade, emoção e intuição, e elementos formais e estruturais, que permitem, por exemplo, classificar uma obra como pertencente a um determinado estilo ou compositor. Por que é possível diferenciar músicas compostas por Bach e por João Gilberto, ou ainda diferenciar músicas compostas pelo mesmo compositor em diferentes fases composicionais (Beethoven em sua primeira fase *vs.* Beethoven em sua terceira e última fase)? Será que isto se deve apenas a um elemento cultural ou intuitivo, ou será que existe algum elemento estrutural, intrínseco à própria música, que possibilita que esta distinção se torne possível?

Um dos problemas que tem despertado a atenção de pesquisadores é o da *interpretação musical*, ou melhor posto, o da compreensão deste fenômeno. Qualquer músico, com um mínimo de experiência, sabe que tocar uma música da forma exata como ela está grafada resulta em algo mecânico e artificial. De fato, a representação simbólica da música ou *notação*, quando existente, serve apenas como um guia para o músico, que a modifica de acordo com suas intenções musicais, com um conjunto de regras, aprendidas durante o decorrer de sua vida musical e, em sua grande maioria, não explícitas, e de acordo com uma análise, ainda que superficial e intuitiva, da obra a ser interpretada. Este conjunto de ações realizadas pelo músico resulta, na prática, na variação de diversos parâmetros musicais: aceleração ou desaceleração do ritmo, aumento ou diminuição da intensidade (dinâmica), modificações na articulação dos eventos (colocação de *staccati*, ligaduras¹, etc.), entre outros [Widmer, 1998]. A este conjunto de variações dá-se o nome de *interpretação musical*.

De uma maneira geral, como notado por Palmer [Palmer, 1997], a grande maioria dos estudos sobre a interpretação se concentra em obras para as quais existe uma partitura disponível, em especial na Música Clássica composta para piano. Além disso, muito se avançou na análise harmônica e melódica, hoje consolidadas e estudadas nos conservatórios como disciplinas práticas. No entanto, apesar de seu importante papel, diversas questões interessantes sobre o *ritmo* e sua interpretação permanecem em aberto. Alguns exemplos de questões são: dada a infinidade de ritmos possíveis, que fatores levam os músicos, compositores e intérpretes a considerarem certas combinações rítmicas apropriadas e outras não? Estes fatores são os mesmos em todos os estilos musicais (Música Clássica e Bossa Nova, por exemplo), isto é, estes fatores são universais? Será que existe alguma influência cultural na determinação destes fatores? Além destas, outras tantas perguntas semelhantes poderiam ser, ainda, formuladas.

Na Musicologia, área da Música que está preocupada com o pensamento, a pesquisa e o conhecimento de variados aspectos da música [Kerman, 1987], algumas questões sobre o ritmo, sua estruturação, interpretação e percepção, têm sido consideradas, na medida do possível [Clarke, 1982, Cooper e Meyer, 1960, Large e Palmer, 2002, Shaffer et al., 1985, Lerdahl e Jackendorff, 1983, Shmulevich e Povel, 2000].

¹Notas marcadas com um *staccato*, um ponto acima ou abaixo do símbolo que representa a nota, são executadas com metade de sua duração. Já notas com ligaduras são executadas sem qualquer interrupção entre elas. Em instrumentos de sopro, por exemplo, não existe interrupção no fluxo de ar durante a execução das notas ligadas.

Uma das limitações neste tipo de pesquisa é que o *corpus* musical estudado, isto é, a quantidade de dados investigados, é, via de regra, muito pequeno. Além disso, o uso de recursos computacionais ainda é bastante limitado. Em geral, o computador é utilizado para atividades como o controle de execuções em experimentos cognitivos, o estabelecimento de estatísticas sobre os dados coletados, na implementação e validação de alguns modelos da música, etc. Existe ainda uma grande parte do trabalho analítico que é feita manualmente, praticamente sem auxílio do computador². Isto pode demandar muito tempo e, por, muitas vezes, se tratar de um trabalho repetitivo, pode levar a erros, principalmente quando o *corpus* a ser analisado é extenso. Daí a necessidade de ferramentas computacionais que auxiliem o músico ou musicólogo, acelerando o processo de descoberta do conhecimento e evitando que possíveis erros invalidem os resultados obtidos. Além disso, o computador permite que uma maior quantidade de dados seja utilizada nas investigações.

No âmbito da Computação Musical [Roads et al., 1996], área da Ciência da Computação que tem como domínio de aplicação a música, as pesquisas e estudos da interpretação musical concentram-se, via de regra, na análise das dimensões melódica e harmônica [Cambouropoulos e Widmer, 2000, Pickens e Crawford, 2002, Widmer, 2003, Pachet et al., 1996, Pardo e Birmingham, 1999, Cambouropoulos et al., 1999]. Apenas recentemente, alguns aspectos da dimensão rítmica, a despeito de sua evidente importância estrutural, têm recebido alguma atenção [Cemgil et al., 2000a, Dixon, 2000, Dixon et al., 2003, Honing, 2001, Dixon et al., 2004, Desain e Honing, 2003]. Uma possibilidade para a existência deste viés é que existem certas “leis”, tanto de harmonização, quanto de condução melódica, que são aceitas pelos músicos e que, por conta disto, podem ser transcritas para o computador³. No caso do ritmo, entretanto, não foi possível encontrar na literatura científica descrições de sistematizações, mesmo que parciais, que indiquem quais combinações rítmicas são desejáveis ou não⁴.

²A segmentação (delimitação de trechos musicais) é um exemplo de tarefa laboriosa e ainda feita manualmente.

³Isto não significa, entretanto, que a transcrição e a implementação destas regras são tarefas fáceis. Muito pelo contrário, a elicitación de tais regras é uma tarefa complexa, uma vez que se trata de conhecimento relativamente informal que é, quase sempre, descrito pelo especialista de maneira bastante subjetiva. Em virtude disto, alguns pesquisadores têm optado pela indução automática de tais regras, utilizando, para tanto, técnicas de Aprendizagem de Máquina (*Machine Learning*) [Widmer, 1998, Dannenberg, 2000].

⁴Em um trecho de seu *Harmonielehre*, Schönberg chega a classificar tentativas neste sentido como “supérfluas” (no original, *überflüssig*) [Schönberg, 1922, p. 242].

1.1 O PROBLEMA: DESCOBERTA AUTOMÁTICA DE CONHECIMENTO EM INTERPRETAÇÕES MUSICAIS

Como visto anteriormente, a interpretação é um fenômeno complexo e que está sujeito a elementos subjetivos, como, por exemplo, as intenções dos intérpretes. Apesar de sua inerente subjetividade, várias pesquisas mostraram a existência de aspectos comuns em diferentes interpretações. As análises do fenômeno, entretanto, são feitas, em sua grande maioria, com o uso de um número reduzido de dados, ou seja, a partir da interpretação de algumas obras ou de trechos de obras, tocadas por poucos intérpretes.

Uma das partes do repertório musical cuja interpretação é relativamente pouco estudada, principalmente se comparada à pesquisa com foco na Música Clássica, é a Música Popular Brasileira, em estilos como a Bossa Nova. Neste âmbito, sabe-se, por um lado, que o violão é o instrumento acompanhador por excelência, *vide*, por exemplo, a grande quantidade de intérpretes que o utilizam como instrumento de acompanhamento, principalmente do canto (João Gilberto, João Bosco, Djavan, Gilberto Gil, Baden Powell, entre outros)⁵.

Por outro lado, sabe-se que o ritmo na Bossa Nova é uma dimensão de análise extremamente importante [Garcia, 1999, Gava, 2002, Sandroni, 1988, Sandroni, 1996]. Importante porque o ritmo, em si, é uma dimensão musical relevante, capaz, inclusive, de definir estilos—basta ver, por exemplo, o grande número de danças que são caracterizadas por um ritmo específico. Além desta importância quase que natural da dimensão rítmica, é interessante notar que, dada a ausência de notação, fato comum no tipo de música aqui estudada, o ritmo só passa a existir a partir do momento em que a música é executada. Na Música Clássica, ao contrário da Bossa Nova, o ritmo já se encontra prescrito na partitura, e pode, portanto, ser analisado de antemão, sem a necessidade obrigatória da execução. Existe, desta forma, uma intrínseca relação entre o ritmo e o intérprete, já que é esse que de fato cria o ritmo, que poderá, apenas então, ser estudado.

No que tange a *construção* do acompanhamento rítmico, existe forte evidência teórica de que esta construção se dá a partir da concatenação de grupos ou células rítmicas recorrentes ou *padrões rítmicos* [Garcia, 1999, Gava, 2002, Sandroni, 1988, Sandroni, 1996]⁶. Os músicos, inclusive, criaram um jargão próprio, denominando um padrão rítmico de

⁵A própria expressão “um banquinho e um violão” denota, de certa maneira, o que é suficiente para a execução na Bossa Nova.

⁶Especificamente na Bossa Nova, Gava chega a classificar o padrão rítmico como um dos *traços estruturais* do estilo [Gava, 2002, p. 44].

batida. Assim, tomando como base a idéia de que o padrão parece ser a de estrutura ou *bloco fundamental* na construção do ritmo, a análise do ritmo deve, portanto, ser realizada justamente a partir dos padrões, ou seja, deve ser uma *análise orientada a padrões*⁷.

Ainda que se tenha construído um certo nível de conhecimento sobre o ritmo na Bossa Nova, existem, todavia, diversas questões interessantes sobre esta dimensão da música que se encontram em aberto ou, no mínimo, necessitam de mais elementos para que sejam melhor respondidas. Alguns exemplos de tais perguntas são as seguintes:

1. existe algum padrão rítmico ou batida preferida por um determinado executante? Quais são estes padrões? Em que situações e com que frequência estes padrões aparecem na música? Existem variações destes padrões? Que variações são estas? Como os padrões podem ser agrupados?
2. Como se caracteriza um padrão rítmico de violão? O andamento possui algum tipo de influência nos padrões rítmicos encontrados? E a intensidade? E a melodia? E a harmonia? Será que existem outras dimensões que possuem influência direta na construção do ritmo? Quais são estas dimensões e que influência é esta?
3. Que variações (temporais, de dinâmica, etc.) são “aceitáveis” dentro de um padrão? Isto é, a partir de que ponto variações realizadas em parâmetros musicais como andamento e dinâmica, entre outros, passam a descaracterizar um certo padrão?
4. Será que, de fato, tudo é padrão? Ou seja, todos os trechos são realmente recorrentes?
5. Existe alguma relação entre o bordão (corda mais grave do violão, tangida com o polegar) e os outros dedos da mão direita?
6. É possível construir um vocabulário de padrões rítmicos para um certo executante?
7. Caso seja possível a construção de tal vocabulário, será que ele muda de estilo para estilo em um mesmo executante (entre Bossa Nova e Samba, por exemplo)? E de executante para executante? Existem mudanças? Como são essas mudanças?

⁷Vale salientar que existe uma grande quantidade de musicólogos que estudam a música através da análise de padrões, principalmente os que estudam o Jazz e seus intérpretes [Kernfeld, 1983, Owens, 1974, Smith, 1983, *apud* [Ramalho, 1997]].

8. Existe uma gramática gerativa capaz de sintetizar formalmente as transformações rítmicas realizadas por um executante? Se for possível construir essa gramática, que relações existem entre as gramáticas de dois executantes distintos?

Vale salientar que perguntas de natureza semelhante podem ser encontradas na pesquisa em Musicologia e Etnomusicologia atual [Nettl, 1983, de Ulhôa, 2003, Pereira, 2003, Trotta, 2003, Napolitano, 2003].

1.2 OBJETIVOS

Levando-se em consideração o que foi previamente exposto, a saber:

- que a interpretação musical é um fenômeno com vários aspectos a serem elucidados;
- que existem limitações na pesquisa em interpretação, seja no reduzido tamanho dos *corpora* estudados ou seja em um certo viés dado à Música Clássica composta para piano;
- que o ritmo, em particular a sua interpretação, é uma das dimensões da música que recebe menos atenção por parte dos pesquisadores em Computação Musical; e
- que o computador é uma ferramenta essencial de auxílio para o musicólogo, possibilitando, inclusive, a análise de grandes *corpora* de maneira eficaz;

pode-se dizer que o objetivo geral desta pesquisa foi a investigação da automação da análise da interpretação musical, ou seja, a *descoberta automática de conhecimento em interpretações musicais*. Especificamente, de maneira a delimitar o escopo de pesquisa, a descoberta automática de conhecimento se deu a partir do acompanhamento rítmico ao violão na Bossa Nova. Em particular, o trabalho tratou da *extração automática de padrões rítmicos de Bossa Nova*. Isto significa que apenas algumas das perguntas anteriormente formuladas foram respondidas ou foram fornecidos elementos fundamentais para a sua resposta, a saber: perguntas 1, 4, 5 e 6.

É importante notar que a delimitação de escopo aqui estabelecida não é gratuita. A escolha da Bossa Nova, e não de um outro estilo, justifica-se, em primeiro lugar, por ser uma manifestação da cultura nacional e cujo estudo poderá se beneficiar muito com a existência de ferramentas do tipo das aqui construídas. Em segundo lugar, é interessante que a interpretação musical seja estudada a partir de outros gêneros, e não apenas da

Música Clássica, como é usual. Já a escolha pelo violão decorre quase que automaticamente da escolha do estilo, já que este é o instrumento, digamos, icônico daquele. Por fim, tratou-se do acompanhamento pois este é, via de regra, o papel fundamental do violonista na Bossa Nova.

Além disso, como encontra-se descrito em maiores detalhes no Capítulo 2, existem muitos elementos que influenciam o fenômeno da interpretação musical, entre eles, o estilo, o instrumento musical e a função do intérprete. Tais elementos, a depender de sua configuração, podem modificar significativamente a natureza do problema, ainda que este seja, em sua essência, a descoberta de conhecimento sobre a interpretação do ritmo. Desta forma, o escopo foi delimitado de maneira tal que o sistema tratasse da análise automática de uma determinada configuração do problema.

Cabem aqui, ainda, algumas considerações sobre as conseqüências desta delimitação de escopo. Em primeiro lugar, da delimitação de estilo (Bossa Nova), decorre também, quase que automaticamente, uma delimitação na forma de notação, já que a maioria das obras nestes estilos encontra-se disponível, no máximo, na forma de cifra (uma notação parcial⁸, portanto), não sendo exagero dizer que, no caso de músicos profissionais, a execução se dê quase que exclusivamente através da imitação e da adaptação, sem o uso de qualquer tipo de notação.

Deve-se ter em mente que, devido à ausência de notação na prática, o intérprete possui um alto grau de liberdade na execução. No que concerne a harmonia, por exemplo, o intérprete tem a liberdade de montar os acordes da forma que lhe for mais apropriada, isto é, da forma que mais saliente suas intenções interpretativas, respeitando, obviamente, certas convenções estilísticas, como, por exemplo, as identificadas por Gava [Gava, 2002]. No caso específico da Bossa Nova, esta liberdade pode ser ainda maior, já que não é incomum que músicos experientes façam suas próprias mudanças na harmonia. No que tange o ritmo, ainda que certas restrições estilísticas se apliquem [Garcia, 1999, Sandroni, 1988], o intérprete possui liberdades ainda maiores, já que, mesmo que exista uma cifra, nenhum de seus aspectos está especificado.

Considerando especificamente o aspecto rítmico do problema, é importante notar que o padrão rítmico encontra-se em um nível estrutural acima do nível da nota. Isto significa que ele não pode ser obtido diretamente a partir da representação da música utilizada, sendo necessário um procedimento que estabeleça se uma determinada seqüência de

⁸Definições e maiores detalhes sobre as formas de notação podem ser obtidas no Capítulo 2.

durações de notas deve ser considerada um padrão rítmico ou não. Neste procedimento, deve-se também determinar um grau de recorrência mínimo para que tal seqüência passe a ser, de fato, um padrão.

Por fim, cabe salientar que a escolha do violão faz com que alguns parâmetros utilizados na análise da música para piano não sejam mais válidos ou aplicáveis. Por exemplo, as medições dos pedais feitas na análise de música para piano e a conseqüente mudança de sonoridade que resulta de sua utilização, não mais poderão ser analisadas no caso da música para violão, já que este instrumento não possui tais recursos. Outros parâmetros, como, por exemplo, KOT (*key overlap time*—tempo de sobreposição de teclas) ou KDT (*key detach time*—tempo de separação de teclas), que são utilizadas na pesquisa sobre articulação na música para piano [Bresin, 2000], também não são válidos neste trabalho. No entanto, outros parâmetros e atributos como, por exemplo, a corda ou a posição na qual uma determinada nota é tocada, se e que o nível de vibrato foi executado, etc., podem e devem ser utilizados.

1.3 ABORDAGEM

Para a determinação da abordagem aqui seguida, resultados de variadas pesquisas anteriores foram considerados. De início, resultados advindos da pesquisa em Música conseguiram isolar regularidades sistemáticas na interpretação musical [Palmer, 1997]. Isto significa que, nestas pesquisas, foi possível, em um primeiro momento, estabelecer padrões de interpretação em execuções variadas e, posteriormente, com o estudo de como estas regularidades acontecem, também foi possível modelar a interpretação, ou ao menos parte dela.

Note, entretanto, que, para o estabelecimento de tais padrões, faz-se necessário que dados que descrevam uma interpretação musical sejam de alguma maneira analisados em busca de regularidades. No âmbito da Ciência da Computação, inclusive na Computação Musical [Arcos et al., 1997, de Mántaras e Arcos, 2002, Widmer, 1998, Widmer, 2003], a análise de dados em busca de regularidades é feita com o auxílio de técnicas variadas das áreas de Aprendizagem de Máquina, Mineração de Dados e Reconhecimento de Padrões, técnicas estas que, além de apontarem regularidades, são também capazes de gerar automaticamente modelos que descrevem, em um determinado nível de precisão, o comportamento dos dados advindos da interpretação.

Na Aprendizagem de Máquina e na Mineração de Dados existe um processo de ex-

tração de conhecimento já consagrado, conhecido como o *ciclo de extração de conhecimento*, que, até certo ponto, descreve os passos necessários para se chegar a um modelo dos dados. Pretende-se, aqui, seguir este processo tanto quanto possível, adaptando-o à natureza do problema e aos propósitos aqui pretendidos.

Ainda como parte da abordagem seguida, estabeleceu-se o material que foi utilizado como entrada para a construção dos modelos. Na Musicologia, existe um certo consenso de que a música, além de possuir várias dimensões, como já foi descrito, é organizada em diferentes níveis estruturais, níveis estes que mantêm entre si uma relação hierárquica [Cooper e Meyer, 1960, Lerdahl e Jackendorff, 1983, Cook, 1994, Guigue, 1997]. Assim, deve ser delimitado a partir de que nível da estrutura musical a construção dos modelos acontece, isto é, se o modelo deve descrever relacionamentos entre notas individuais ou se ele deve versar sobre conexões entre elementos de maior nível estrutural, como a frases ou seções, por exemplo.

No caso específico do problema aqui estudado, o nível estrutural utilizado na construção de modelos pôde ser determinado através de um outro resultado importante: existe evidência, empírica e teórica, de que o acompanhamento rítmico na Bossa Nova é feito a partir da concatenação de seqüências de durações particulares, chamadas de padrões rítmicos ou ainda, no jargão dos músicos, de *batida* [Garcia, 1999, Sandroni, 1988]. Tomando este resultado como ponto de partida inicial, assumiu-se neste trabalho que a modelagem deveria acontecer a partir do nível dos padrões, e que faria, *a priori*, pouco sentido descrever relações abaixo deste nível, como relações entre notas, por exemplo.

Ainda como parte da abordagem foi construída uma ferramenta de apoio à análise automática, ferramenta esta que possibilitou a extração de padrões rítmicos dos dados coletados. Além da ferramenta, foi estabelecida uma metodologia de análise associada ou um processo de análise que, mesmo que em linhas gerais, fundamentou o procedimento analítico decorrente do uso da ferramenta.

1.4 REQUISITOS

Feita a explanação do problema e da abordagem seguida é possível apontar alguns requisitos e funcionalidades que um musicólogo esperaria ou gostaria que um sistema de análise da interpretação possuísse.

O primeiro requisito vem da constatação de que a Música possui aspectos subjetivos e imprecisos. Isto significa que não existe *uma única* maneira de se analisar uma obra

musical e sua interpretação, mas sim, *uma possível* maneira de se analisar uma obra. O sistema deve, portanto, permitir que o analista construa, através de experimentações, a versão de análise que lhe pareça mais coerente, isto é, o sistema deve ser *flexível*. É importante notar que este requisito cria um conflito com a idéia de que grandes quantidade de dados devem ser utilizados na análise de interpretações musicais. Como a análise sucessiva de grandes massas de dados pode ser custosa computacionalmente, estes experimentos podem ser proibitivos. Deve-se, portanto, levar em consideração que o nível de experimentação permitido deve ser limitado de acordo com a quantidade de dados e recursos computacionais disponíveis.

Uma outra questão importante em um sistema como o que se propõe construir aqui, é que ele deve ser utilizado por músicos ou musicólogos e, como tal, deve fornecer meios para que estes indivíduos possam interagir com o sistema de forma intuitiva. Isto significa que não deve ser necessário que o usuário conheça profundamente conceitos da Ciência da Computação para que possa formular questões essencialmente musicais. Deve existir, portanto, uma interface capaz de fazer a tradução necessária de uma pergunta feita nos termos do especialista para uma pergunta que o sistema “compreenda”. Outros requisitos, mais comuns em sistemas deste tipo, mas que devem ser mencionados, são a capacidade de se ouvir, visualizar e manipular os dados musicais.

Por fim, como o objetivo é servir de apoio às atividades de análise musical, um outro requisito é que o resultado obtido pelo computador deve consistir de um *modelo da interpretação explícito e geral* (ou mais geral possível), permitindo, assim, que o analista compare os modelos encontrados (*sínteses*) e tire suas conclusões analíticas sobre como a interpretação acontece. Levando-se em conta o primeiro requisito descrito (o da flexibilidade), seria importante também que os diferentes modelos de uma mesma obra pudessem ser comparados e que esta comparação fosse de alguma maneira possível intermediada pelo computador.

Em suma, os requisitos de um sistema para a análise de interpretações aqui enumerados são os seguintes:

- possibilidade de construção de diferentes e variados experimentos (flexibilidade);
- existência de uma interface que permita que o analista trabalhe nos experimentos sem a necessidade de conhecimentos profundos em Ciência da Computação;
- capacidade de manipulação dos dados musicais por parte do usuário;

- possibilidade de construção de modelos explícitos e gerais da interpretação; e
- possibilidade de comparação destes modelos entre si.

1.5 CONTRIBUIÇÕES

Uma das contribuições fundamentais da presente pesquisa foi a derivação de uma representação original dos dados. Reduzindo a dimensionalidade dos dados (de multidimensionais para unidimensionais), essa representação foi crucial para uma eficiente extração de padrões. Tal representação, inclusive, se utilizou de conceitos utilizados pelos próprios músicos e poderia, em tese, com algum treino, ser lida por eles. Isto significa que, além de eficiente para a extração de padrões, a representação também é potencialmente efetiva para os músicos.

Uma outra contribuição fundamental foi a criação de um algoritmo original para a indução do dedilhado da mão direita do violonista [Trajano et al., 2004]. Esse foi, inclusive, o resultado que serviu de base para a construção da representação dos dados previamente mencionada.

Foram, ainda, criadas adaptações no algoritmo FLEXPAT, o que constitui uma terceira contribuição fundamental.

Além dessas, pode-se ainda elencar outras contribuições secundárias. A primeira delas foi a aquisição de dados realizada. Tais dados, após tratamento adequado, formam o maior, e possivelmente único, conjunto de dados de Bossa Nova conhecido. Implementou-se também um protótipo de ferramenta de auxílio à análise do acompanhamento e adaptou-se o ciclo de descoberta de conhecimento para as especificidades do problema estudado, o que funcionou como metodologia de análise automática, orientando, também, a utilização da ferramenta.

Resultados experimentais obtidos durante a realização da pesquisa foram publicados em dois artigos científicos [Trajano et al., 2005, Trajano et al., 2006].

Vale salientar, ainda, que este projeto foi premiado no *Rumos Itaú Cultural Pesquisa 2003*, que é um programa criado pelo Instituto Itaú Cultural e que tem o objetivo de incentivar projetos de pesquisa acadêmica sobre arte e mídia, especificamente os que explorem as várias interseções entre tecnologia, arte, ciência e mídia. Das mais de 330 submissões recebidas pela instituição, apenas oito foram aceitas, sendo este projeto o único selecionado no Norte e Nordeste do país.

Além deste prêmio, ao autor foi concedida uma bolsa de doutorado sanduíche no exte-

rior (SWE) pelo CNPq (processo número 201232/2004-9), que foi usufruída no Instituto Austríaco de Pesquisas em Inteligência Artificial (ÖFAI) em Viena, Áustria. A pesquisa neste instituto foi realizada sob a supervisão do Univ.-Prof. Dr. Gerhard Widmer, no período de fevereiro a agosto de 2005.

1.6 ESTRUTURA DO DOCUMENTO

O restante do texto conta com outros cinco capítulos, organizados como descrito a seguir. No Capítulo 2, descreve-se o fenômeno da interpretação musical: o que é, quais as suas características, o que o influencia e como ele é analisado pelos pesquisadores em Música e Musicologia. Mostra-se ainda que o computador pode ser um elemento fundamental neste tipo de pesquisa, suprimindo lacunas e deficiências existentes. Neste Capítulo descreve-se, ainda, como a pesquisa em Computação Musical trata da análise de interpretações, isto é, como se dá a análise automática de interpretações musical. Para tanto, os principais grupos de pesquisas que tratam do problema são identificados, assim como as abordagens utilizadas por estes e alguns dos resultados que foram obtidos. Discute-se ainda que abordagem será seguida nesta proposta.

No Capítulo 3, mostra-se que o problema da análise do acompanhamento rítmico ao violão pode ser dividido em duas partes: a identificação dos padrões rítmicos utilizados pelo intérprete e a construção de modelos que descrevam como estes padrões são tocados e utilizados. Em seguida, descrevem-se algumas técnicas e algoritmos de extração de padrões musicais e como modelos da interpretação podem ser construídos. Discutem-se ainda algumas questões acerca da metodologia de construção de modelos.

Nos dois capítulos seguintes (Capítulos 4 e 5), descreve-se em detalhes o trabalho elaborado nesta pesquisa, a saber: a adaptação do ciclo de descoberta do conhecimento, as tarefas realizadas, os problemas resolvidos e os resultados obtidos em cada uma das etapas do ciclo adaptado. Apresenta-se também uma série de experimentos conduzidos com os dados coletados e com a ferramenta de análise desenvolvida.

Já no último Capítulo deste trabalho (Capítulo 6), apresentam-se algumas conclusões e as contribuições do trabalho, assim como os trabalhos futuros.

CAPÍTULO 2

INTERPRETAÇÃO MUSICAL

A interpretação de uma música (*expressive music performance*) é um fenômeno muito estudado, tanto pela Musicologia, quanto pelas Ciências Cognitivas e pela Psicologia. Como ilustrado na Figura 2.1, a interpretação é, nesses estudos, usualmente vista como parte de um sistema de comunicação no qual compositores codificam suas idéias musicais de alguma maneira (na forma de notação musical ou mesmo por transmissão oral, por exemplo) e os ouvintes recodificam tais idéias a partir do sinal acústico que é produzido por músicos, chamados intérpretes [Kendall e Carterette, 1990, *apud* [Palmer, 1997]].

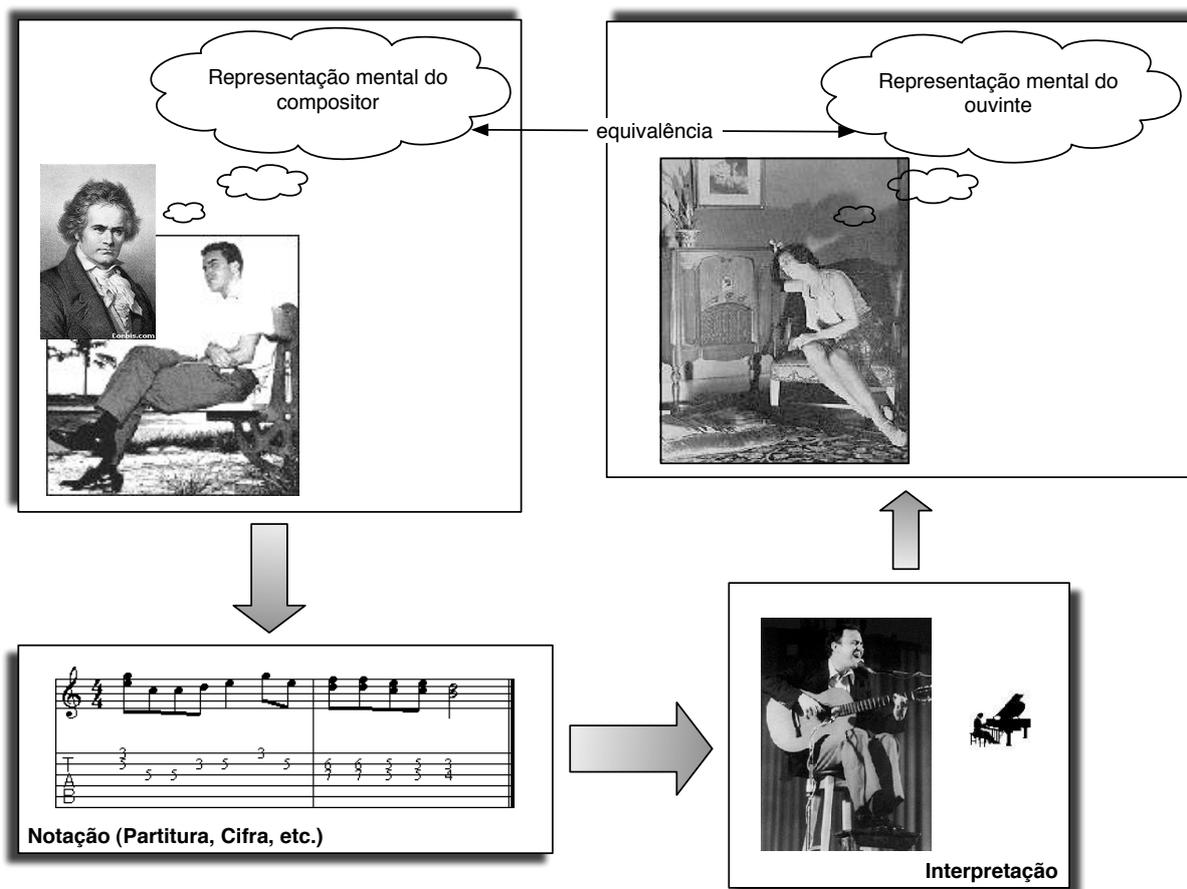


Figura 2.1 O percurso da música: do compositor ao ouvinte

Palmer identifica três componentes principais deste fenômeno [Palmer, 1997]: a interpretação, o planejamento e o movimento. O primeiro componente diz respeito à modelagem individual da obra musical, feita de acordo com as próprias idéias ou intenções musicais do intérprete. O segundo refere-se a aos mecanismos de planejamento e de recuperação da memória realizados durante a interpretação, isto é, como os músicos preparam as seqüências musicais para a interpretação. Já o terceiro e último componente trata da produção e do controle da interpretação propriamente dita.

Segundo Sloboda [Sloboda, 2000], dois fatores compõem a interpretação: um componente técnico e outro expressivo. O primeiro diz respeito à mecânica necessária para a produção fluente e coordenada de som. O segundo, a *expressão*, deriva das variações intencionais dos parâmetros da interpretação, variações estas escolhidas pelo músico de forma a influenciar os resultados cognitivos e estéticos do ouvinte.

Este capítulo trata especificamente da interpretação: que fenômeno é este, o que o influencia e como ele é analisado pelos pesquisadores. O objetivo é caracterizar a tarefa “interpretar uma música”, identificar lacunas e limitações na pesquisa em interpretação musical e mostrar que o computador pode ser uma peça chave para o preenchimento destas lacunas, assim como pode fornecer novos elementos para a sua melhor compreensão.

O restante do capítulo está organizado da seguinte maneira: na Seção 2.1, apresentam-se a definição e a caracterização do ato de interpretar uma música. Na Seção 2.2, apresenta-se, em linhas gerais, como se dá a pesquisa em interpretação musical no âmbito da Música. Na Seção 2.3, apresentam-se argumentos para que o computador funcione como um elemento fundamental neste tipo de pesquisa, assim como as dificuldades existentes quando se usa o computador com este intuito. Já na Seção 2.4 apresentam-se pesquisas que tratam da automação da análise de interpretações, isto é, como se dá a pesquisa em interpretação musical no âmbito da comunidade de Computação Musical. Por fim, na Seção 2.5, discutem-se as abordagens apresentadas na Seção 2.4 levando-se em consideração se e como elas podem ser aplicadas especificamente ao problema do acompanhamento rítmico.

2.1 CARACTERIZAÇÃO DO PROBLEMA

Qualquer músico com um mínimo de experiência sabe que tocar uma música da forma exata como ela está grafada resulta em algo mecânico e artificial. Na verdade, a representação simbólica da música ou *notação*, quando existente, serve apenas como um guia

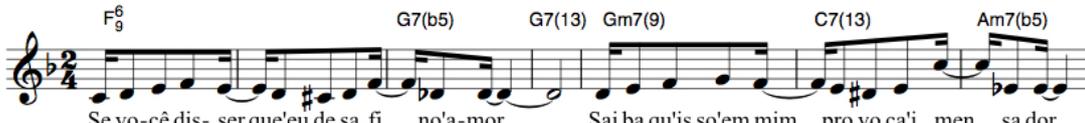
para o músico, que a modifica de acordo com suas intenções musicais, com um conjunto de regras e de acordo com uma análise da obra a ser executada. Assim, a interpretação musical ou expressão pode ser definida da seguinte maneira.

Definição 2.1 (Interpretação Musical). *Interpretação musical é a arte de moldar uma peça musical de acordo com intenções, regras e análises do intérprete.*

Desta definição depreende-se, portanto, que a interpretação de uma música é um fenômeno complexo, podendo ser influenciado de diversas maneiras.

Uma particularidade que exerce uma forte influência na interpretação é a forma como a música encontra-se representada, isto é, que tipo de *notação* é disponibilizada para o músico. De uma maneira geral, a notação musical pode ser classificada em dois tipos: completa ou parcial. Um exemplo típico de notação completa é a partitura¹. Exemplos de notações parciais são a cifra, a tablatura e o baixo cifrado. A influência destes tipos de representação se reflete em certas liberdades que são concedidas ou não ao músico. Enquanto que na partitura cabe ao músico, via de regra, a escolha do andamento e o estabelecimento dos níveis de intensidades, não cabendo a ele a escolha de que notas tocar, por exemplo, na música cifrada ele deve escolher, além do andamento e dinâmica, as notas e o ritmo, já que estes não se encontram explicitamente representados. As Figuras 2.3 e 2.2 ilustram duas maneiras de apresentação de cifras e um exemplo de partitura, respectivamente. Note que, no caso da cifra, apenas a harmonia está especificada (em casos mais raros, como na letra B, a linha melódica encontra-se descrita também).

A F₉⁶ / / / G7(b5) / G7(13) / Gm7(9) / C7(13) / Am7(b5) /
Se você disser que eu desafino amor Saiba que isso em mim provoca imensa dor [...]

B 
Se vo-cê dis- ser.que'eu de sa fi__no'a-mor Sai ba qu'is so'em mim__pro vo ca'i men__sa dor

continua

Figura 2.2 Exemplos de cifra

¹É importante frisar, entretanto, que a noção de completude aqui não deve ser vista com rigor matemático e, muito menos, confundida com a idéia de precisão. Existe um grande número de informações bastante importantes para a interpretação que não estão objetivamente representadas na partitura (informações sobre a estrutura, por exemplo) e outros tipos de informações que, apesar de representadas, não são estabelecidas através de medidas precisas (andamento e dinâmica são exemplos típicos).

Clair de lune

(Suite bergamasque, Paris 1905)
(L. 75 - 1890)

Claude DEBUSSY
(1862-1918)

Andante très expressif

Piano *pp* con sordina

Figura 2.3 Exemplo de partitura

Ainda sobre a influência da representação da música, existe o caso de sua inexistência, isto é, a música a ser tocada não se encontra registrada na forma de notação². A ausência de notação é prática bastante comum na música popular de maneira geral e na música folclórica [Napolitano, 2003].

A *modalidade de interpretação* é outra forte influência. Palmer enumera tais modalidades da seguinte maneira: leitura à primeira vista (tocar uma obra desconhecida a partir de alguma notação); tocar, de memória ou a partir de notação, obras já estudadas e preparadas de antemão; improvisação e tocar “de ouvido” [Palmer, 1997]. No caso da leitura à primeira vista, a interpretação é construída dinamicamente, a partir da análise feita no momento em que o símbolo da notação é lido. Este tipo de construção dinâmica, feita com conhecimento parcial da estrutura, pode resultar em diversos erros como estruturação incorreta, incapacidade de compreensão de grandes estruturas (seções, frases, etc.) [Clarke, 1988]. Em obras previamente estudadas, a interpretação é um processo mais laborioso, indo desde a análise da estrutura da música, até o estabelecimento de intenções interpretativas. O músico pode, inclusive, testar diversas interpretações diferentes, verificando que versão é a mais adequada. É, portanto, um processo muito mais profundo que o primeiro caso.

²Note que, nos dias atuais, o fonograma (CD, disco, fita, etc.) é uma forma de registro extremamente importante, mas não se constitui em uma forma de notação.

No caso da improvisação, como também é o caso da leitura à primeira vista, a interpretação é criada dinamicamente. Existe, no entanto, algumas diferenças fundamentais: na improvisação, o músico, em geral, já possui algum tipo de intenção interpretativa *a priori*. Além disso, cabe ao próprio músico a delimitação e, principalmente, a criação da estrutura musical durante a improvisação, se aproximando este, portanto, do compositor. Por fim, no caso de se tocar uma música “de ouvido”, a interpretação pode ser baseada em uma imitação do que foi ouvido anteriormente. Esse caso está intrinsecamente ligado ao caso da ausência de notação musical previamente citado.

Outras características que ainda exercem influência sobre a interpretação são: o grau de familiaridade com a obra a ser executada e a própria experiência do intérprete. Esta última, aliás, é apontada como um dos principais fatores que pode influenciar a interpretação [Ericsson et al., 1993, Sloboda et al., 1996, *apud* [Palmer, 1997]]. Todas estas influências são, todavia, muito difíceis de se medir com exatidão, dada a sua alta subjetividade.

O *papel do intérprete* na música é outra característica importante. Os dois principais papéis que um músico pode exercer na música são solista e acompanhador. No primeiro caso, o músico, como em uma peça teatral, tem o papel principal, sendo responsável pelas principais escolhas interpretativas. Já o acompanhador deve sustentar tais decisões, mantendo-se em segundo plano. Isto não significa, entretanto, que o acompanhador não possui qualquer tipo de liberdade. Dentro de suas limitações, ele pode escolher se esta ou aquela nota deve ser destacada, se uma ou outra duração é mais adequada, etc.

Outra influência na interpretação vem das *dimensões da música*. Melodia, harmonia e ritmo, entre outras, fazem com que as escolhas interpretativas se adaptem às suas estruturas (tamanho das frases, estrutura métrica, estrutura harmônica, etc.). O problema aqui é que estas dimensões, apesar de sua aparente independência, são, na verdade, interdependentes: decisões de como a estrutura harmônica deve ser interpretada podem se basear nas escolhas feitas para a estrutura melódica e vice-versa; a acentuação resultante da estrutura rítmica da obra pode influenciar na escolha da interpretação de uma certa linha melódica; e assim por diante.

O *instrumento musical* também influencia o fenômeno, já que cada um deles possui recursos próprios que podem ser utilizados pelo intérprete, afetando, assim, a interpretação. No violão, por exemplo, uma mesma nota pode ser tocada em diferentes posições, cada uma destas possuindo uma qualidade de som ou timbre resultante diferente. O intérprete pode se utilizar desta particularidade do instrumento para favorecer esta ou aquela in-

tenção interpretativa. Já em outros instrumentos, como o piano, só existe uma única posição (neste caso, uma tecla) para a produção de uma nota musical, não sendo possível neste caso, portanto, basear a interpretação em diferentes qualidades tímbricas.

Além das particularidades de cada instrumento, o conjunto ou grupo de instrumentos para o qual a música foi composta também influencia a interpretação. Para manter o equilíbrio sonoro entre as partes, por exemplo, grupos compostos por instrumentos de diferentes famílias (cordas e metais, cordas e madeiras, etc.) podem impor certas restrições nos níveis de volume sonoro de cada integrante (um trompete possui um volume sonoro muito maior que o violino ou qualquer outro instrumento de corda, por exemplo). Mesmo em grupos compostos por instrumentos elétricos (guitarras, baixos elétricos e afins), este tipo de preocupação existe: o engenheiro de som, via de regra, é quem faz esta tarefa, equalizando os sinais dos diversos instrumentos.

Outra característica importante para a interpretação é o *estilo* da obra a ser tocada. Segundo Palmer [Palmer, 1997], a interpretação é limitada por propriedades sintáticas específicas do estilo e da época de composição de cada obra. Além disso, é bastante plausível imaginar que obras de períodos diferentes não sejam interpretadas da mesma maneira. Tais propriedades estilísticas transcendem, portanto, o individualismo de cada interpretação.

Feita esta análise, pode-se dizer que a interpretação é um fenômeno multicritério, com algumas de suas dimensões possuindo caráter subjetivo e individual, de cada intérprete, portanto. Outras destas dimensões, porém, transcendem características individuais, e são comuns a todos os intérpretes. O mais importante é fazer notar que cada confluência destas características, isto é, cada ponto neste espaço multicritério requer o uso de um conhecimento musical específico, ainda que exista um subconjunto de conhecimento musical básico, comum a todo músico. A Figura 2.4 ilustra o passo entre a notação e a interpretação, indicando os fatores que influenciam esta última.

2.2 A PESQUISA EM INTERPRETAÇÃO MUSICAL

Quando se fala em pesquisa sobre a interpretação musical, em particular pesquisas que pretendem analisá-la e compreendê-la, uma questão precisa ser respondida de antemão: será que a interpretação musical, sendo uma manifestação artística, não é um fenômeno intangível, que não pode, de maneira alguma, ser generalizado? Apesar da interpretação musical nos parecer uma atividade artística extremamente individual e subjetiva, como

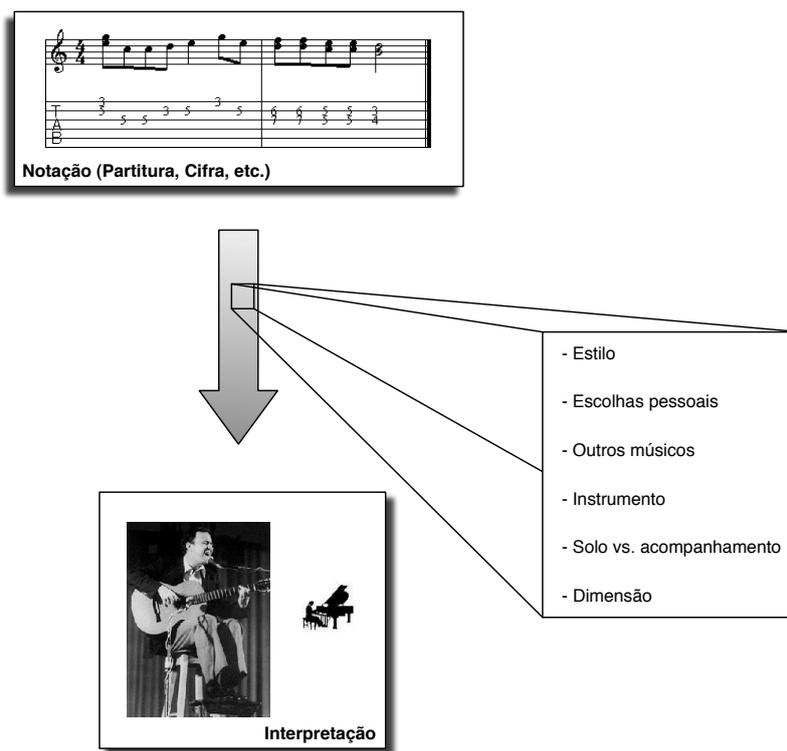


Figura 2.4 Influências na interpretação

foi indicado na seção anterior, existem resultados que indicam a existência de diversos aspectos da expressão que são comuns a todos os intérpretes [Clarke, 1988, Palmer, 1997]. De acordo com Palmer, “a interpretação é baseada tanto em elementos individuais, que diferenciam os intérpretes, quanto em aspectos normativos que são compartilhados por estes mesmos intérpretes” [Palmer, 1997]³.

Feita a constatação de que a interpretação é um fenômeno que possui aspectos passíveis de análise, resta saber que aspectos são estes. Uma das primeiras dimensões da música em que padrões sistemáticos de expressão foram detectados foi o ritmo [Palmer, 1997]. Dois exemplos de padrões de expressividade nesta dimensão vem de pesquisas sobre as interpretações de valsas vienenses e de alguns estudos que tiveram como sujeitos pianistas e percussionistas. No primeiro caso, as valsas eram sistematicamente interpretadas com o primeiro tempo do compasso sempre mais curto, enquanto que o segundo tempo era sempre mais longo [Askenfelt, 1986, Bengtsson e Gabrielsson, 1977, apud [Palmer, 1997]]. Já

³É interessante notar que estes aspectos gerais também parecem se refletir na percepção musical. Sloboda relata que ouvintes não-treinados foram capazes de julgar certas seqüências musicas como bem formadas (gramaticais) ou mal formadas (não gramaticais) [Sloboda, 2000].

os estudos com pianistas e percussionistas mostraram que, na interpretação de padrões rítmicos, estes músicos realizam desvios sistemáticos na duração das notas (se comparados a um tempo metronômico) e nas intensidades [Gabrielsson, 1974, apud [Palmer, 1997]].

Além do ritmo e suas variadas facetas, também foram identificados padrões de expressividade em outras dimensões da música, a saber: dinâmica [Nakamura, 1987, Repp, 1999b, Sloboda e Lehmann, 2001]; andamento, em particular de que maneira ele é variado pelos músicos (*timing*) [Clarke, 1982, Palmer, 1988, Repp, 1999a, Repp, 1998a]; articulação [Repp, 1998b]; melodia [Palmer, 1989, Rasch, 1979]; entre outros. A título de exemplo, pesquisas na âmbito da dimensão melódica encontraram padrões sistemáticos de desvios na condução melódica: em geral, a nota da melodia é tocada entre 20 e 30 milissegundos antes das notas do acompanhamento. Isto foi percebido tanto na música para piano, quanto em música tocada por outros instrumentos [Sundberg et al., 1991]⁴. Também é importante notar que existem padrões de expressão em duas ou mais dimensões, simultaneamente. Um exemplo é o fato de que marcas estruturais importantes (finais de frase, por exemplo) são muitas vezes indicadas pelo intérprete através de uma diminuição do andamento e da intensidade (dinâmica) [Henderson, 1936, apud [Palmer, 1997]].

Outras pesquisas têm caráter mais geral, tentando, por exemplo, determinar quais as funções da interpretação. Estas, de um modo geral, apontam para duas funções principais: salientar determinados conteúdos estruturais [Clarke, 1988] e destacar um conteúdo emotivo em particular [Palmer, 1997]. Existe ainda outro grande conjunto de pesquisas que estuda a interpretação e sua influência na percepção da música [Clarke, 1989, Large e Palmer, 2002, Repp, 1992, Sadakata et al., 2002].

De maneira geral, como notado por Palmer, a grande maioria dos estudos se concentra em obras para as quais existe uma partitura disponível e na música para piano [Palmer, 1997]. No primeiro caso, escolhe-se a partitura pois ela, comparada a outras notações como a cifra ou tablatura, fornece metas de interpretação com um menor nível de ambigüidade. Já o piano, devido à forma como o som é produzido, é um instrumento que facilita sobremaneira as medições necessárias para estas pesquisas⁵. Ainda devido a esta combinação (partitura e piano), a grande maioria das pesquisas se concentra na

⁴No caso do piano, dada a natureza de seu mecanismo de produção sonora, este é um resultado de certa forma esperado. Mas a identificação do mesmo fenômeno em outros instrumentos é bastante inusitada.

⁵Existem, inclusive, pianos de concerto, como o Boesendorfer SE290, que apesar de serem instrumentos acústicos, possuem sistemas computadorizados que medem com grande precisão o abaixar e levantar de teclas, assim como o uso de pedais.

Música Clássica.

No que concerne a metodologia, a pesquisa em interpretação se dá, em geral, de maneira empírica: um analista humano cria uma teoria ou um modelo matemático de aspectos da interpretação e, em seguida, tenta validar seu modelo empiricamente, utilizando para isto dados coletados de execuções reais [Widmer, 2002b].

Por fim, toda pesquisa em interpretação deve levar em consideração que a validação dos resultados obtidos pode ser problemática, afinal é muito complicado determinar objetivamente se uma interpretação está correta ou não, já que se trata de algo extremamente subjetivo. No entanto, pode-se assumir, com alguma propriedade, que intérpretes como o pianista Vladimir Horowitz ou o violonista João Gilberto produzem interpretações corretas, isto é, que as interpretações de músicos reconhecidamente competentes podem ser utilizadas para efeito de comparação.

2.3 DA NECESSIDADE DO COMPUTADOR

Apesar das pesquisas sobre a interpretação musical apresentarem resultados interessantes, é possível apontar algumas de suas deficiências e limitações. Uma primeira limitação, bastante evidente, é observada no tamanho do *corpus* analítico utilizado pelos pesquisadores: via de regra, utilizam-se interpretações de uma ou duas obras, como é o caso, por exemplo, de Clarke [Clarke, 1982, Clarke, 1985]; ou ainda a interpretação de trechos de uma ou algumas obras feitas por alguns intérpretes, como é o caso, por exemplo, de Repp [Repp et al., 2002], que utilizou as interpretações de uma melodia de oito compassos, em quatro andamentos diferentes, feitas por 12 pianistas diferentes. Em suma, utiliza-se uma quantidade relativamente pequena de dados.

A consequência imediata da utilização de um número limitado de obras e intérpretes reflete-se nas limitações de resultados obtidos: caracterização de alguns aspectos da interpretação [Repp et al., 2002, Shaffer et al., 1985], exame da interpretação de uma única obra [Beran e Mazzola, 2000, Clarke, 1982, Clarke, 1985], etc. Como a idéia, de uma maneira geral, é a de obter resultados gerais que caracterizem a interpretação da melhor maneira possível, fica a pergunta se resultados como estes são representativos, capazes de realmente modelar o fenômeno. Para evitar este tipo de questionamento, a pesquisa em interpretação musical deve, portanto, dispor de uma quantidade razoável de dados, de forma que os modelos obtidos sejam, de fato, representativos.

Por outro lado, o exame de uma grande quantidade de dados, como dito anteriormente,

é praticamente inviável sem o auxílio do computador. Um exemplo que confirma este fato é a pesquisa sobre Charlie Parker conduzida por Owens [Owens, 1974]: o objetivo, nesta pesquisa, era verificar a existência de eventuais padrões melódicos nas improvisações feitas por Charlie Parker⁶. Para tanto, Owens teve que analisar um *corpus* musical bastante extenso⁷, de forma a poder encontrar ou não tais padrões. Por conta disto, sua tese de doutorado só pôde ser terminada após 16 anos de pesquisa, já que, por falta de ferramentas adequadas, muito tempo foi gasto com transcrições, revisões e correções, além das comparações necessárias para a realização da análise em si.

O tempo excessivamente longo leva ainda a algumas considerações que podem, eventualmente, invalidar alguns aspectos da pesquisa: será que o pesquisador consegue ser metodologicamente consistente durante um período de tempo tão extenso? Dado que, neste tipo de pesquisa, o analista deve realizar um trabalho comparativo substancial, envolvendo uma grande quantidade de dados, quantos erros podem ocorrer durante o processo como um todo? Será que estes eventuais erros não invalidam de alguma forma a análise realizada? Como se vê, este tipo de pesquisa estará quase sempre limitada a um repertório pequeno, caso não se esteja disposto a enfrentar todas as dificuldades inerentes ao processo de análise de grandes quantidades de dados.

É justamente por isto que o computador é uma ferramenta essencial em pesquisas como estas. De fato, a pesquisa em Mineração de Dados [Han e Kamber, 2001, Witten e Frank, 2000] e em Aprendizagem de Máquina [Langley, 1996, Mitchell, 1997] já mostrou diversos resultados interessantes em pesquisas que utilizavam grandes quantidades de dados e nos mais variados domínios [Langley e Simon, 1995, Mitchell, 1997], inclusive no domínio da Música [Widmer, 2000, Widmer, 2001a]. Além disso, como posto por Widmer, “o computador, por estar livre de preconceitos e expectativas, pode, mais facilmente, gerar hipóteses novas e possivelmente surpreendentes” [Widmer, 2002b].

Esta argumentação pode dar a idéia de que o computador não é utilizado de maneira alguma pelos pesquisadores e pesquisas anteriormente citadas, o que não é verdade. Ele é, sim, utilizado por estes pesquisadores. Todavia, o uso que é feito da máquina não corresponde às possibilidades que ela poderia proporcionar, principalmente no que concerne a descoberta de regularidades nos dados musicais. Uma hipótese plausível para esta limitação talvez seja a virtual inexistência de ferramentas computacionais que sirvam de apoio às atividades de análise musical necessárias para a identificação e posterior

⁶Por padrão melódico, entende-se certos trechos de melodias que são utilizados de maneira recorrente.

⁷Na verdade, mais extenso impossível, já que ele analisou todas as obras de Parker.

compreensão do que significam estas regularidades nas interpretações⁸.

Uma outra limitação vem do fato de que toda a pesquisa se concentra, quase que exclusivamente, na Música Clássica composta para piano. Ainda que este seja um instrumento bastante representativo, existem épocas, anteriores ao surgimento deste instrumento, cujas interpretações não podem ser analisadas, como é o caso, por exemplo, do período Barroco e dos períodos anteriores a este. Isto, mais uma vez, pode se refletir nos tipos de perguntas que podem ser formuladas e respondidas: não é possível saber, por exemplo, se existem regularidades interpretativas comuns a todas as épocas da Música Clássica e, muito menos, caracterizá-las⁹. Além disso, alguns compositores importantes da própria Música Clássica praticamente não compuseram para o piano (Berlioz, por exemplo, compôs suas obras mais importantes para orquestra). E, acima de tudo, existem outros estilos musicais nos dias de hoje, como o Jazz ou mesmo a Música Popular Brasileira, cuja interpretação é pouco analisada, mas que são representantes importantes da vida musical e cultural.

Se, por um lado, o uso do computador pode facilitar a vida do músico, a Música, em especial a análise de interpretações musicais, é um domínio que traz diversas dificuldades para o cientista da computação.

O primeiro grande problema é o da representação. Este, por sua vez, possui dois aspectos: a granularidade e a multidimensionalidade. No que concerne a granularidade, existem os prós e os contras de se escolher entre uma representação contínua ou discreta do som [Roads et al., 1996]. Em uma representação contínua, como os dados de arquivos áudio por exemplo, existe uma grande riqueza na representação de algumas dimensões da música (o timbre, por exemplo), o que, muitas vezes, não existe em uma representação discreta, como no caso de um arquivo MIDI. Todavia, identificação, a partir de arquivo áudio, das notas executadas ainda é um problema em aberto, ainda que com diversos avanços recentes [Abdallah e Plumbley, 2004, Cabral et al., 2005, Yoshioka et al., 2004, Wen e Sandler, 2005]. Já na representação discreta, é trivial saber qual nota está sendo

⁸De fato, isto pode ser generalizado para qualquer tipo de análise musical. Apesar de existir uma quantidade significativa de aplicativos de auxílio à composição musical, não existe nenhum que tenha sido desenvolvido exclusivamente para o auxílio à análise musical. Quando muito, existem algumas extensões de sistemas de composição que são utilizadas para analisar certas características da estrutura musical [Guigue, 2003, Meudic, 2000], o que não é, nem de longe, o mais adequado, já que as duas disciplinas possuem naturezas e objetivos completamente diferentes.

⁹Isto é quase que uma tentativa de encontrar elementos da interpretação que sejam gerais, partes elementares do tal sistema de comunicação, citado no início deste capítulo e que é assumido por grande parte dos pesquisadores.

executada, já que trata-se de uma informação diretamente acessível. A depender do problema, uma ou outra representação será mais ou menos adequada.

Quanto à multidimensionalidade, o que se tem é a possibilidade de representação de um elemento da música através de muitos atributos. Pode-se, por exemplo, pensar com facilidade em diversos atributos para se descrever uma simples nota, o nível mais baixo da música: altura ou frequência, duração, intensidade, instrumento, se é uma nota da melodia ou do acompanhamento, se é nota da harmonia ou uma nota de passagem, articulação utilizada (*staccato*, legato, etc.), grau da escala, função harmônica do acorde subjacente, posição métrica, relacionamentos (intervalos) com notas anteriores e posteriores, intervalo com a nota do baixo, etc. Obviamente, existe uma quantidade mínima de atributos para a representação de cada elemento da música (no caso específico da nota, este atributo seria a frequência), mas, a depender da tarefa a ser automatizada, mais atributos podem ser essenciais. A dificuldade reside, portanto, em definir, para cada tarefa, que conjunto de atributos é este. Vale salientar, ainda, que alguns dos atributos *supra* citados nem sempre podem ser obtidos automaticamente com facilidade, já que não se encontram representados diretamente na música (função harmônica de um acorde, por exemplo). Desta forma, além do problema da definição do número ideal de atributos, a derivação automática de alguns destes atributos também pode ser um problema.

Outras dificuldades vêm do fato de a Música, em particular a interpretação musical, possuir um conjunto de conhecimento que é parcialmente subjetivo. Um exemplo disto é que a própria determinação da estrutura musical, passo importante para o estabelecimento da interpretação, é um processo que não é consenso entre os musicólogos. Um reflexo disto é a grande quantidade de teorias existentes que tratam das formas como a música pode ou deve ser estruturada (*e. g.* [Cooper e Meyer, 1960, Cook, 1994, Lerdahl e Jackendorff, 1983, Narmour, 1990]). Elas, ainda que possuam alguns elementos em comum (a visão de que a música é constituída de uma hierarquia de estruturas, mais notadamente), divergem em muitos aspectos.

Além de possuir aspectos subjetivos, parte do conhecimento musical também é impreciso. Por exemplo, como já foi dito, a partitura, ainda que represente a forma mais completa de notação musical, utiliza um grande conjunto de símbolos cuja interpretação é completamente imprecisa, entre eles: símbolos que denotam a dinâmica e a sua variação (*forte*, *piano*, *mezzo-forte*, *crescendo*, etc.), símbolos que denotam o andamento e sua variação (*allegro*, *accelerando*, etc.), símbolos que denotam a articulação (*staccato*, *portamento*, etc.).

2.4 ANÁLISE AUTOMÁTICA DE INTERPRETAÇÕES MUSICAIS

No âmbito da Ciência da Computação, em especial na Computação Musical, existem alguns grupos que se destacam na pesquisa em interpretação musical, a saber: o grupo de Processamento Musical Inteligente e Aprendizagem de Máquina do Instituto Austríaco de Pesquisas em Inteligência Artificial (ÖFAI)¹⁰ e o Grupo de Pesquisas em Acústica Musical do Instituto Real de Tecnologia (KTH)¹¹ em Estocolmo, Suécia.

Apesar de possuírem abordagens baseadas no uso do computador como ferramenta de análise fundamental, cada um dos grupos possui uma ênfase e, em virtude disto, uma abordagem diferente para o problema. A seguir são descritas as pesquisas realizadas por esses grupos. Nesta descrição, salientam-se os objetivos dessas pesquisas, assim como as técnicas e os resultados já alcançados. O objetivo aqui é mostrar um panorama geral da pesquisa em interpretação musical feita por pesquisadores de Computação Musical, fornecendo subsídios para a identificação de possíveis similaridades entre estas pesquisas e os propósitos da pesquisa apresentada nesta tese.

2.4.1 O Projeto de Interpretação Musical Vienense

Este projeto, coordenado pelo pesquisador austríaco Gerhard Widmer, situa-se na interseção entre Musicologia e Inteligência Artificial e pode ser descrito, em linhas gerais, como um projeto que procura “desenvolver e utilizar métodos de Aprendizagem de Máquina e Mineração de Dados para estudar o fenômeno complexo da expressão musical” [Widmer, 2001b], gerando modelos formais e quantitativos de interpretações, o que é um grande problema em aberto na Musicologia contemporânea.

O projeto busca, a partir de gravações de pianistas profissionais e aplicando técnicas de Aprendizagem de Máquina, construir modelos gerais e robustos dos vários aspectos da interpretação, como p. e., um conjunto de regras capaz de estipular os desvios de tempo que um pianista deverá fazer ao tocar uma determinada peça. Uma vez construídos, os modelos devem ser validados, o que pode ser feito de diversas maneiras: comparando-os com resultados teóricos da Musicologia; aplicando-os a novas peças e analisando a qualidade musical das interpretações geradas pelo computador; e, obviamente, medindo-se a acurácia em um conjunto de testes. A Figura 2.5 ilustra todas as fases do projeto.

¹⁰<http://www.ofai.at/research/impml/index.html>.

¹¹<http://www.speech.kth.se/music/>.

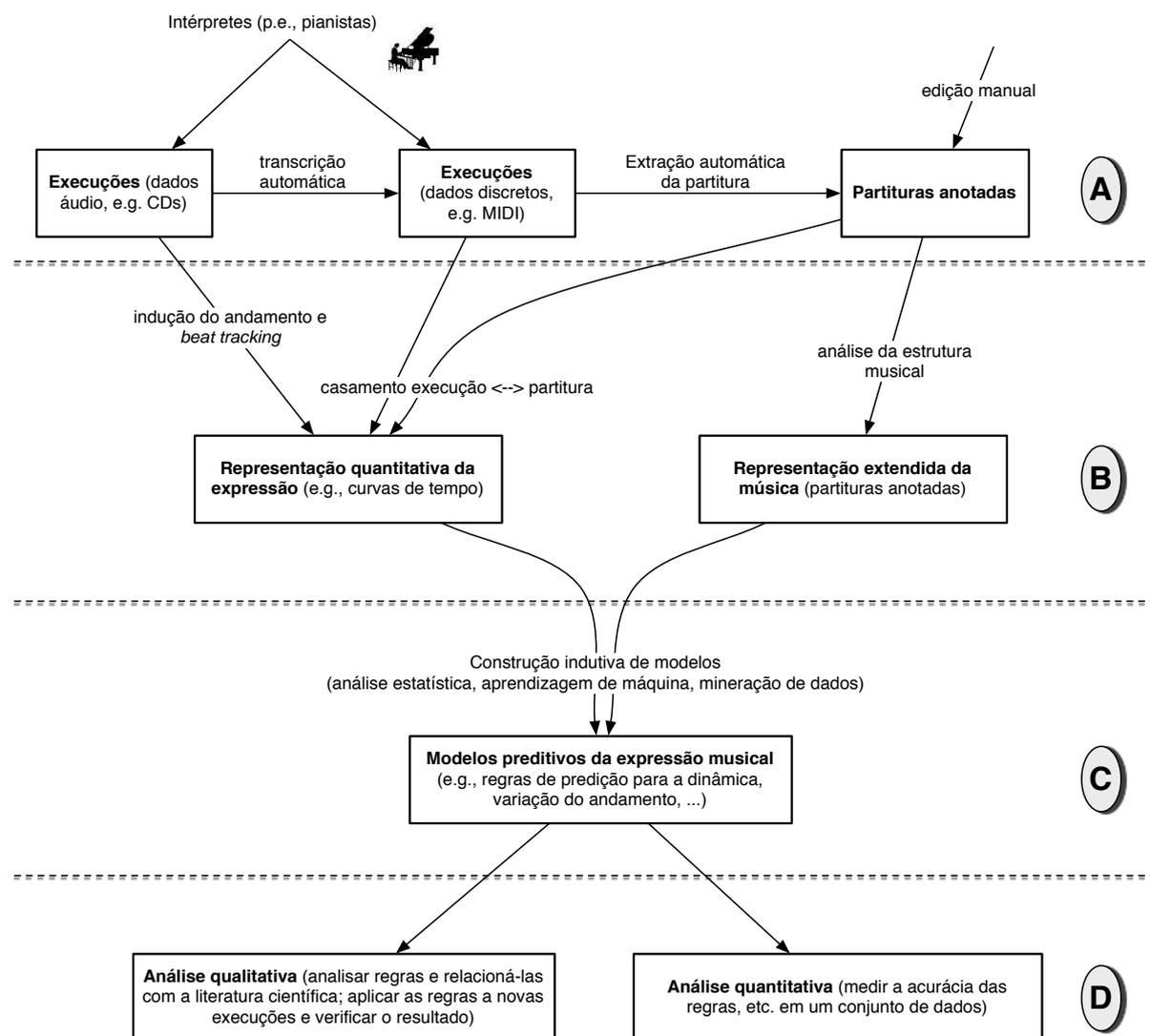


Figura 2.5 Visão geral do projeto austríaco

Especificamente, a metodologia utilizada no projeto passa por diversas fases, sendo a primeira delas a aquisição de dados (Figura 2.5, letra A). Nela, utilizou-se um piano Boesendorfer modelo SE290, que é um piano de concerto—acústico, portanto—que mede com precisão cada movimento de tecla, martelo e pedal e grava estas medições em formato simbólico, similar ao MIDI, mas com uma precisão muito maior. O conjunto de dados adquiridos constitui-se de: 17 sonatas completas de Wolfgang Amadeus Mozart, correspondendo a cerca de cinco horas e meia de música e 150.000 notas; e praticamente toda a obra para piano de Frédéric Chopin, correspondendo a mais de nove horas de música, 300.000 notas e dois milhões de medições de pedal. Esta é, por sinal, a maior coleção de dados de interpretações já coletada e estudada. Além destes dados, eles também utili-

zam gravações em áudio digital com o intuito de estudar, ainda que de maneira limitada, alguns aspectos da interpretação de diversos artistas famosos [Widmer, 2002a].

Outra fase do projeto trata da preparação e do pré-processamento dos dados (Figura 2.5, letra B). É o resultado desta fase que é utilizado como entrada para os algoritmos de Aprendizagem de Máquina. O principal problema, aqui, é a reconstrução da partitura original, já que ela é necessária para estabelecer o nível de desvios de um executante. Realizá-la manualmente seria inviável, dado o grande número de dados. O que se conseguiu foi um processo semi-automático, onde algoritmos geram, a partir das execuções, grande parte das partituras. Este resultado, entretanto, precisa ser corrigido e completado à mão. Em seguida, estabelece-se uma correspondência nota a nota, entre a interpretação e a partitura gerada.

Um problema, neste momento, é representar a música de maneira adequada para que os algoritmos de Aprendizagem de Máquina possam gerar modelos interessantes. É necessário estabelecer relacionamentos entre estruturas da música (estruturas harmônica, métrica e frasal, por exemplo) e padrões na interpretações (um *crescendo*, por exemplo). Obviamente, a análise estrutural é necessária para que estes relacionamentos sejam construídos. Fazer a análise manualmente, é, mais uma vez, impraticável devido à grande quantidade de dados existentes. Foram desenvolvidos alguns algoritmos para realizar a análise: algoritmos para a segmentação [Cambouropoulos, 2001b], para a criação de categorias de motivos melódicos e suas recorrências [Cambouropoulos e Widmer, 2000] e para encontrar os mais diversos padrões harmônicos e rítmicos, como postulado pela teoria musical. A análise computada por estes algoritmos pode, então, ser utilizada como descritores adicionais na representação da peças.

Uma vez representada a música, inicia-se a fase de investigação utilizando técnicas de Aprendizagem de Máquina (Figura 2.5, letra C). A pesquisa, neste momento, segue duas linhas distintas: uma primeira onde se tenta caracterizar e prever as coincidências entre as interpretações, e uma segunda onde se tenta caracterizar as diferenças entre os intérpretes. No primeiro caso, espera-se construir descrições dos padrões que são comuns na maioria das execuções e tratar as situações onde as interpretações diferem como ruído. Já no segundo caso, o problema é encontrar classes de situações onde existam diferenças explícitas e sistemáticas no comportamento dos músicos, levando à sua possível caracterização.

Uma vez construídos os modelos, eles podem ser testados de duas maneiras: uma análise qualitativa ou uma análise quantitativa (Figura 2.5, letra D). No primeiro caso, a

análise qualitativa, verifica-se se as regras resultantes da modelagem possuem correlatos na literatura científica (textos musicológicos, análise prévias, etc.) ou se uma interpretação gerada a partir delas passa pelo crivo de um músico competente. Já na análise quantitativa, as regras resultantes são verificadas em um conjunto de testes de forma a medir sua acurácia, isto é, nos moldes da própria pesquisa em Aprendizagem de Máquina e Mineração de Dados [Mitchell, 1997, Witten e Frank, 2000].

Esta pesquisa vem sendo conduzida há alguns anos e já apresentou diversos resultados relevantes. Um dos primeiros resultados desta pesquisa foi um meta algoritmo original de indução de regras, chamado PLCG—*Partition+Learn+Cluster+Generalize*— [Widmer, 2001a, Widmer, 2003]. Ele foi capaz de descobrir princípios de expressividade bastante simples, mas extremamente gerais e robustos, como, por exemplo, a seguinte regra:

```
louder IF
  dir_prev = up &
  int_prev > p4 &
  metr_strength > 2.
```

Segundo ela, uma nota deve ser enfatizada (tocada mais forte) caso ela seja precedida de um salto melódico ascendente com intervalo maior que uma quarta justa.

Também foram criados diversos programas para diminuir o tempo gasto na fase de preparação dos dados, entre eles, um programa para a inferência da estrutura métrica das obras (*beat induction* e *tempo tracking*), chamado *BeatRoot*, que é capaz de realizar esta inferência mesmo em músicas onde existem grandes e bruscas mudanças no andamento [Dixon, 2000, Dixon, 2001]. Também foram desenvolvidos algoritmos para a quantização [Cambouropoulos, 2000] e para a indução da correta enarmonia das notas (sol♯ *vs.* lá♭, por exemplo) [Cambouropoulos, 2001a].

Além destes, resultados teóricos também foram obtidos. Um exemplo de resultado teórico é que os modelos gerados para um determinado pianista tendem a se deteriorar (diminuir a acurácia) se aplicados para a predição da interpretação de outros pianistas. Outro resultado, que levou à criação do algoritmo PLCG, é que a construção de modelos completos das execuções, pelo menos modelos completos construídos no nível da nota, não leva a resultados satisfatórios. Uma árvore de decisão que discrimina entre *accelerando* e *ritardando*, por exemplo, teve acurácia de 58,09% e 3.037 folhas, o que corresponde a 3.037 regras de classificação [Widmer, 2002b]. Isto se dá, muito possivelmente, porque

algumas das escolhas interpretativas do executante são relacionadas com outros níveis da estrutura musical. Com o algoritmo PLCG, o resultado é um modelo parcial que “explica, no nível da nota, apenas o que pode ser explicável” [Widmer, 2002b].

Entre as ferramentas desenvolvidas pelo grupo uma que merece ser mencionada é o chamado “Verme Interpretativo” (*Performace Worm*) [Dixon et al., 2002a, Dixon et al., 2002b]. Baseado em uma representação desenvolvida pelo musicólogo Jörg Langner [Langner e Goebel, 2002], o sistema possibilita a visualização, em tempo real, de duas dimensões da interpretação: o tempo, medido em batidas por minuto e representado na abscissa do plano, e o volume (*loudness*), medido em decibéis e representado pela ordenada. A Figura 2.6 ilustra o Verme em ação. Nela, encontram-se duas trajetórias que representam a interpretação dos últimos compassos (24–28) da sonata para piano KV.279 de W. A. Mozart feita por dois pianistas renomados (Daniel Barenboim e Andrés Schiff). Os pontos mais escuros representam o instante atual, enquanto que instantes anteriores da trajetória são representados por pontos mais esmaecidos. Além da representação bastante intuitiva, o sistema também possui um algoritmo de indução de pulsação em tempo real capaz de se adaptar a modificações feitas pelo intérprete, o que, de fato, é uma contribuição importante para o problema da indução da pulsação.

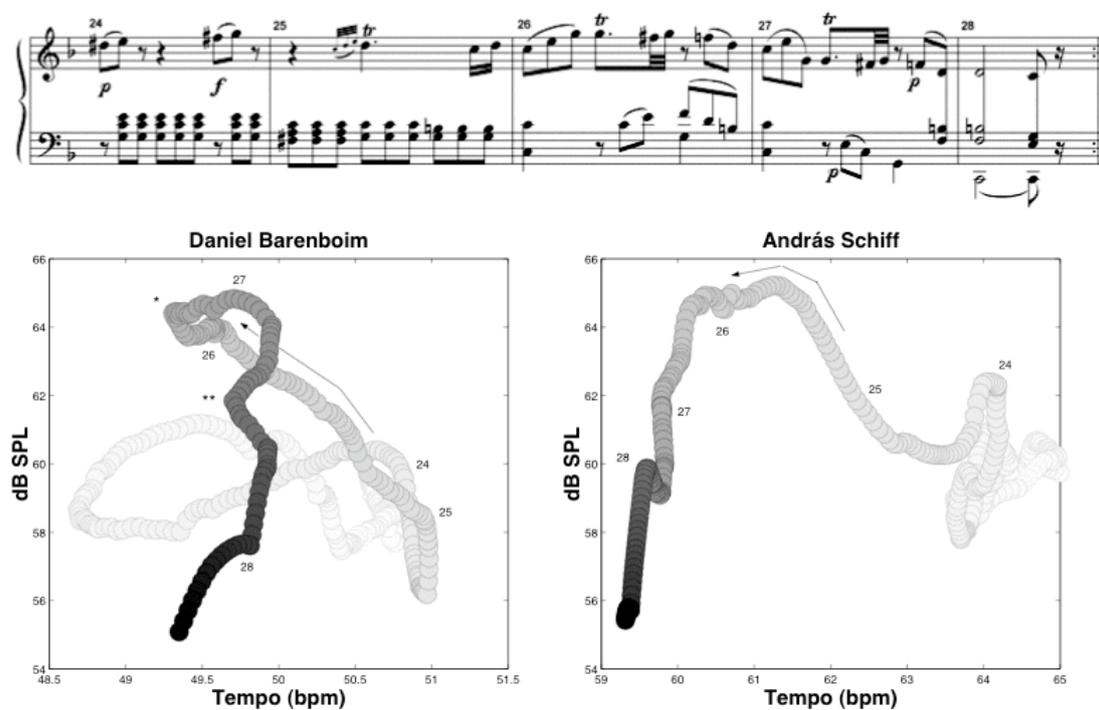


Figura 2.6 O Verme Interpretativo em ação

Mais recentemente, as pesquisas conduzidas por este grupo concentraram-se na tentativa de construção de modelos em um nível hierárquico acima do nível da nota [Widmer e Tobudic, 2002] e na investigação da possibilidade de se caracterizar e quantificar formalmente aspectos do estilo individual de um artista. Analisando gravações digitais de vários pianistas famosos, o grupo chegou a resultados que apontam para a possibilidade de se chegar a certos padrões de interpretação de pianistas famosos [Widmer, 2002a, Widmer et al., 2003]. Em outros trabalhos, tratou-se da separação de vozes e identificação de melodias, ambos os casos em arquivos MIDI [Madsen e Widmer, 2006b, Madsen e Widmer, 2007a, Madsen e Widmer, 2007b] e do alinhamento automático¹² de diferentes interpretações da mesma obra musical [Dixon e Widmer, 2005, Dixon, 2005].

Outros temas de pesquisa, estes mais alinhados com a área de Recuperação de Informação Musical (*Music Information Retrieval*), tratam do desenvolvimento de algoritmos para a extração de informação de alto nível e musicalmente relevante, assim como de metadados, a partir de arquivos áudio [Dixon et al., 2004], de métodos para a automação da estruturação e visualização de grandes coleções de músicas e da construção de ferramentas para a navegação destas coleções [Pampalk et al., 2004], da criação de algoritmos para a classificação automática de gêneros e estilos musicais a partir de gravações de áudio [Dixon et al., 2003]

2.4.2 A Pesquisa na Suécia: Comunicação Musical

O grupo de pesquisa liderado pelo pesquisador Johan Sundberg é, certamente, o mais antigo a ainda realizar pesquisas em interpretação musical, iniciando suas pesquisa nos anos 1970. O grupo possui uma linha de pesquisa bastante clara [Sundberg et al., 1991, Sundberg e Lindblom, 1991]: partindo-se da hipótese de que os desvios realizados pelos executantes são, na verdade, maneiras de comunicar a estrutura musical para o ouvinte, as pesquisas visam a maior compreensão de como esta comunicação musical acontece.

Àquela época (anos 1970), acreditava-se que não era possível gerar uma interpretação minimamente decente de uma obra musical utilizando regras, já que se considerava a interpretação como um fenômeno único. Tais crenças não estavam de acordo com ob-

¹²Por alinhamento entende-se a identificação das mesmas estruturas musicais em interpretações diferentes de uma mesma obra, a despeito das variações e dos desvios de andamento que normalmente acontecem em diferentes interpretações.

servações empíricas feitas pelo músico profissional e integrante do grupo Lars Frydén, que havia identificado uma série de regularidades na interpretação que mereciam testes [Sundberg et al., 2001].

Para testar as observações de Frydén, o grupo utilizou uma abordagem chamada análise-por-síntese (*analysis-by-synthesis*) [Sundberg et al., 1991]. Nela, os modelos de interpretação são construídos a partir das hipóteses iniciais sobre a música e sua estrutura coletadas por Frydén, hipóteses estas codificadas na forma de regras, que são, em seguida, utilizadas na síntese de uma execução. O resultado da síntese é avaliado (através de sua audição) e, se necessário, a hipótese inicial é modificada ou refinada e o processo é repetido. O resultado é, portanto, uma regra ou conjunto de regras que modelam uma execução (Figura 2.7).

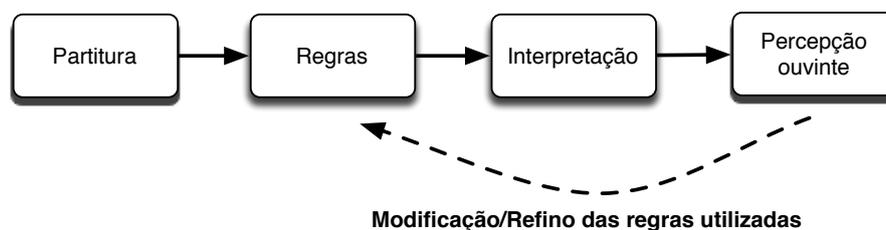


Figura 2.7 A análise-por-síntese

Um dos resultados mais significativos já obtidos pelo grupo foi a criação de uma gramática gerativa com cerca de 30 regras quantitativas, que prescrevem modificações que devem ser aplicadas a cada uma das notas de forma a gerar uma interpretação [Sundberg et al., 1991]. As regras da gramática são de dois tipos: regras de agrupamento (*grouping rules*) e regras de diferenciação (*differentiation rules*). As regras de agrupamento ajudam os ouvintes a agrupar notas, formando estruturas musicais de nível maior que o da nota (motivos, frases, etc.). Já as regras de diferenciação ajudam o ouvinte a identificar categorias de alturas e durações, como, por exemplo, notas acentuadas, *staccati*, etc. Esta gramática encontra-se incorporada ao sistema *Director Musices* [Friberg, 1995]. Neste sistema, cada nota pode ser processada por diversas regras da gramática e o efeito, em geral, é resultante da adição das modificações realizadas pelas regras utilizadas. O grupo também estudou a correlação entre a música, a fala e o movimento [Carlson et al., 1989, Friberg e Sundberg, 1999], identificando, por exemplo, analogias entre padrões de movimentos do corpo e certos *ritardandi*.

Mais recentemente, o grupo investiga, entre outras coisas, a possibilidade de se reproduzir de forma automática o estilo interpretativo de um pianista [Sundberg et al., 2003].

Outra linha de pesquisa é a simulação da emoção. Em uma das vertentes desta linha de pesquisa, varia-se a quantidade de parâmetros das regras e o número de regras em si, conseguindo-se com isto um modelo que corresponde a algo semelhante a uma “paleta de emoções” (suave, solene, alegre, triste, etc.) [Bresin, 2000]. Em outra vertente, utiliza-se Lógica Nebulosa para a análise da emoção na execução [Friberg, 2005]. Os resultados foram aplicados à visualização das emoções [Friberg et al., 2002] e ao jogo colaborativo *Ghost in the Cave* [Rinman et al., 2004]¹³.

2.4.3 Outras Pesquisas

Além destes, outros grupos de pesquisa e pesquisadores também conduzem, em maior ou menor número, pesquisas sobre a interpretação. Elas estão descritas aqui de maneira bastante resumida, já que, ainda que tratem da interpretação, possuem objetivos muito diferentes dos pretendidos nesta proposta e não serão consideradas na discussão feita na Seção 2.5.

O grupo de pesquisas holandês MMM (*Music, Mind, Machine*), coordenado pelos pesquisadores Henkjan Honing e Peter Desain, tem como propósito central a melhor compreensão da cognição musical, em particular o entendimento de como a execução musical acontece [Desain et al., 2001]. Para tanto, o grupo constrói modelos cognitivos da execução, utilizando uma metodologia chamada modelagem computacional [Desain et al., 1998]. A pesquisa, iniciada no final dos anos 80, já apresentou diversos resultados, entre eles pode-se destacar os seguintes: um estudo sobre a indução de pulsação (*beat induction*), estudo este que resultou em um modelo baseado em regras simbólicas [Desain e Honing, 1999]; estudos sobre a relevância do contexto musical em decisões interpretativas tomadas por um executante, como, por exemplo, as decisões sobre o uso do *rubato* [Timmers et al., 2000]; e um estudo sobre apojaturas (*grace notes*) que mostrou que questões estilísticas, e não a estrutura musical, parecem ser os fatores mais influentes na forma como o ornamento é executado [Windsor et al., 2000].

Na Espanha, no Instituto de Pesquisas em Inteligência Artificial (IIIA), pesquisadores, liderados por Ramon López de Mantaras, utilizam técnicas de Raciocínio Baseado em Casos (RBC) [Kolodner, 1993] para gerar a interpretação de uma música. Eles possuem um sistema, chamado SaxEx [Arcos et al., 1997, de Mántaras e Arcos, 2002], capaz de, dada como entrada uma melodia sem qualquer interpretação, aplicar modificações nota a nota

¹³<http://www.speech.kth.se/music/projects/Ghostgame/>.

gerando uma melodia interpretada. É importante notar ainda que o sistema utiliza tanto uma representação MIDI, quanto uma representação áudio da música: a primeira é utilizada pelo subsistema RBC, enquanto que a representação áudio é processada via técnicas de modelagem espectral (*Spectral Modelling Synthesis*—SMS). Os valores extraídos a partir da aplicação destas técnicas também são utilizados pelo subsistema RBC. Uma vez recuperado um caso, ele é sintetizado utilizando-se técnicas de modelagem espectral e o resultado é um arquivo áudio com a interpretação da música.

Atualmente, o grupo constrói um outro sistema, também baseado em RBC e semelhante ao SaxEx, chamado *Tempo-Express* [Grachten et al., 2006], capaz de interpretar uma peça adequadamente, mesmo quando a entrada encontra-se em um andamento diferente do caso armazenado na base. Além disso, alguns avanços foram feitos na tentativa de se descrever automaticamente se notas extras devem ser tocadas, se algumas notas devem ter sua duração modificada ou ainda se determinadas notas não devem ser tocadas. Isto é chamado pelo grupo de “anotar um caso” e é feito utilizando-se programação dinâmica e a distância de edição [Arcos et al., 2003].

Também na Espanha, na Universidade Pompeu Fabra, o Grupo de Tecnologia Musical (*Music Technology Group*) possui uma linha de pesquisa em interpretação musical. Esta linha de pesquisa possui uma orientação bastante semelhante à do grupo austríaco (Seção 2.4.1). Nas palavras do próprio grupo: “ao invés de se modelar a interpretação manualmente, testando em seguida o modelo obtido em um conjunto de dados, deixa-se que o computador, através do uso de técnicas de aprendizagem de máquina, descubra regularidades e princípios de execução a partir de dados interpretativos reais”¹⁴. Os trabalhos do grupo se concentram na modelagem da interpretação de linhas melódicas de Jazz (o instrumento é o saxofone), utilizando para tanto técnicas de aprendizagem de máquina indutiva [Ramirez e Hazan, 2004, Ramirez et al., 2004] e algoritmos genéticos [Hazan et al., 2006, Hazan e Ramirez, 2006].

Na Universidade de Pádua, no Departamento de Eletrônica e Informática (DEI), Itália, existe um grupo, liderado pelo pesquisador Giovanni De Poli, que também trabalha com a interpretação, mais precisamente com o controle do som. A idéia é desenvolver modelos que mapeiem intenções expressivas como “brilhante”, “suave”, etc. em parâmetros acústicos de baixo nível que sirvam, então, para controlar uma execução com a intenção pretendida [Canazza et al., 1999]. Recentemente, o grupo realizou pesquisas, em cola-

¹⁴Trecho retirado da página internet do grupo que trata da linha de pesquisa em interpretação musical (<http://www.iaa.upf.edu/mtg/pages/topics/single?id=9>).

oração com o grupo de pesquisa da Áustria (Seção 2.4.1), no intuito de gerar modelos capazes de reconhecer pianistas famosos a partir de suas execuções [Zanon e Widmer, 2003a, Zanon e Widmer, 2003b]. O grupo começou, também, a trabalhar de maneira oposta, isto é, analisar execuções para extrair as intenções expressivas dos músicos, ao invés de construir execuções baseadas em intenções expressivas pré-definidas [Canazza et al., 2003]. Esta pesquisa está mais alinhada com a idéia de que uma das funções da música é salientar determinado conteúdo emotivo (Seção 2.2).

2.5 DISCUSSÃO

De início, é importante deixar claro que as abordagens apresentadas têm objetivos diferentes: enquanto que no grupo sueco a intenção é entender como se dá a comunicação entre músico e ouvinte (que pode ser um músico também), construindo, para isto, modelos que produzam resultados que “soem bem”, o grupo austríaco está preocupado com o desenvolvimento de modelos que se assemelhem ao máximo com as execuções analisadas, isto é, eles estão interessados em saber *como* uma determinada interpretação acontece. É necessário, portanto, que estas diferenças sejam levadas em consideração durante qualquer discussão sobre estas abordagens.

Posto isto, uma primeira diferença importante entre as pesquisas descritas encontra-se na validação dos resultados: no caso de grupo sueco, a avaliação é feita através da audição do resultado, o que requer a presença de ao menos um especialista, algo nem sempre possível. Além disso, para que esta validação seja estatisticamente significativa, é necessário um número razoável de avaliadores, já que pode haver discrepância entre eles, o que torna o processo mais complicado ainda. O sucesso das regras, no caso da pesquisa sueca, depende inteiramente da formulação de sua seqüência de aplicação e da existência de ouvintes competentes. Além disso, pode ser difícil identificar a regra ou as regras que estão causando problemas: “um desvio específico em uma nota pode ser originado a partir de diferentes princípios. Assim, a ‘verdadeira’ origem pode ser impossível de ser identificada” [Friberg, 1995]. O interessante, entretanto, é que é possível realizar diversos experimentos com um mesmo conjunto de regras, modificando apenas alguns parâmetros e/ou a ordem das regras, o que pode levar a alguns *insights* musicológicos.

No grupo austríaco, por sua vez, a validação é feita tanto de maneira quantitativa quanto qualitativa. Para a validação quantitativa, verifica-se, em primeiro lugar, o grau de adequação (*degree of fit*) das regras: ele mede a cobertura e a acurácia das regras. Em

seguida, verifica-se a generalidade do modelo obtido, isto é, verifica-se se o modelo pode ser utilizado com sucesso na predição dos desvios expressivos em obras diferentes das utilizadas na sua construção. Neste teste, pode-se avaliar ainda se o modelo também se aplica a outros artistas. Na validação qualitativa, procura-se na literatura musicológica teorias que modelem a interpretação de maneira semelhante às regras obtidas ou ainda, se interpretações geradas a partir da aplicação destas regras são satisfatórias.

Um segundo ponto onde as abordagens diferem é na forma como os dados são adquiridos: na Áustria, a aquisição é feita com o objetivo de se obter grandes quantidades de dados, isto é, um número razoável de obras completas, tocadas por pianistas profissionais em um piano de concerto. Já na Suécia não é preciso adquirir dados no sentido de que um músico deve tocar uma obra ou trecho de obra, como no caso da pesquisa austríaca, isto porque a pesquisa funciona modificando uma execução neutra com o intuito de torná-la expressiva. A única necessidade do grupo é codificar a obra em um formato legível pelo sistema que realiza as modificações expressivas. No caso da pesquisa austríaca, cabe uma pergunta sobre os dados e o processo de aquisição: será que os dados são de fato “do mundo real” ou o equipamento utilizado introduz algum tipo de distorção nos dados coletados? E mais, se tais distorções existirem, elas são levadas em consideração no momento da construção dos modelos ou em sua avaliação?

Por fim, vale salientar um ponto comum em todas as pesquisas: o uso do computador como ferramenta ou instrumento básico de pesquisa. Não se admitem mais, portanto, pesquisas nos moldes da já citada pesquisa realizada por Owens¹⁵.

A decisão de que abordagem deveria ser seguida na presente pesquisa levou em consideração os seus objetivos (cf. Seção 1.2). Desta forma, pela semelhança nos objetivos (descrever *como* uma determinada interpretação acontece), optou-se por uma abordagem nos moldes da utilizada pelo grupo austríaco, isto é, utilizar Aprendizagem de Máquina e Mineração de Dados como ferramentas computacionais básicas para a construção de modelos do acompanhamento rítmico na Música Popular Brasileira. É importante notar, entretanto, que a escolha pela uma abordagem nos moldes da austríaca não implica em ignorar os resultados obtidos nas outras pesquisas. Eles certamente foram considerados no decorrer deste trabalho, já que fornecem informações e pontos de vistas diferentes que podem, eventualmente, ajudar a esclarecer questões importantes do domínio do problema.

Existem, no entanto, diferenças entre o que foi aqui realizado e o que é pesquisado em

¹⁵ Vide página 22 para maiores informações.

Viena. No Quadro 2.1 apontam-se algumas destas diferenças. Tais diferenças e suas conseqüências serão, entretanto, tratadas apropriadamente no próximo capítulo (Capítulo 3, Seção 3.3).

	Pesq. Tradicional	ÖFAI	CIn
Estilo	Mús. Clássica	Mús. Clássica	MPB
Notação	partitura	partitura	cifras
Instrumento	piano	piano	violão
Dimensão	outras	outras	ritmo
Função	solista	solista	acompanhador
Quantidade de Dados	poucos	muitos	muitos

Quadro 2.1: Comparação entre pesquisas em tradicional em Música (sem computador), a austríaca e a aqui apresentada

CAPÍTULO 3

CONSTRUÇÃO DE MODELOS DA INTERPRETAÇÃO MUSICAL

No Capítulo 2, foram apresentadas uma caracterização da interpretação musical e diversas abordagens, no âmbito da Computação Musical, para a investigação de como este fenômeno acontece. De maneira geral, pode-se dizer que todas essas abordagens, de uma forma ou de outra, tentam formalizar, em maior ou menor grau, diversos aspectos da “interpretação musical”. Isto acontece tanto para os aspectos comuns a todas as interpretações, quanto para os aspectos que diferem e são particulares a cada músico. O resultado destas tentativas de formalização é, desta forma, um *modelo* do fenômeno.

No que tange o ritmo, existem ao menos três diferentes níveis de abstração da estrutura musical a partir dos quais um modelo de interpretação pode ser construído: a nota, a estrutura métrica e o grupo rítmico. O nível da nota corresponde ao das durações individuais — de cada nota — tocadas pelo intérprete. É, portanto, o nível mais baixo de análise. A estrutura métrica corresponde à pulsação, isto é, àquela sensação de recorrência que faz com que os ouvintes “batam o pé” em intervalos aparentemente regulares de tempo. Este nível, por organizar as durações em períodos de tamanho equivalente, encontra-se acima do nível das notas¹. Já o terceiro nível corresponde à definição de grupo proposta por Lerdahl e Jackendorff, isto é, à “organização natural do sinal sonoro em unidades” [Lerdahl e Jackendorff, 1983, p. 12], limitada aqui, obviamente, ao aspecto rítmico. Trata-se, desta forma, de um terceiro nível estrutural, que se encontra acima do nível da estrutura métrica.

Quanto aos modelos de interpretação atualmente construídos, pode-se afirmar que, em grande parte, eles refletem a interpretação no nível da nota, isto é, descrevem apenas como se dão as relações entre uma nota e suas vizinhas imediatas. Sabe-se, entretanto, que o acompanhamento rítmico na Bossa Nova é construído através da concatenação de

¹É importante frisar que os intervalos temporais entre pulsações consecutivas podem não ser exatamente os mesmos, se medidos com precisão, mas que, a despeito deste fato, seu *significado* é o mesmo, isto é, eles correspondem, na teoria, ao mesmo lapso temporal.

diversos grupos rítmicos específicos. Garcia e Sandroni chegam, inclusive, a catalogar, a partir de transcrições de acompanhamentos feitos por João Gilberto, alguns destes grupos rítmicos [Garcia, 1999, Sandroni, 1988]. Desta maneira, na Bossa Nova, a construção de modelos da interpretação feita no nível individual da nota pode não dizer muito sobre como estes grupos rítmicos são escolhidos, já que modelos neste nível sequer são capazes de identificar os grupos utilizados. Posto isto, conclui-se que, para que sejam de alguma utilidade analítica, os modelos de interpretação devem refletir o nível dos grupos rítmicos².

Existe, ainda, uma particularidade que deve ser observada. Como foi exposto no Capítulo 1, o grupo rítmico assume, na Bossa Nova, um papel estrutural de grande importância. Neste estilo, ao conceito de grupo rítmico adiciona-se a noção de batida ou padrão rítmico, ou seja, a idéia de que um grupo rítmico passa a possuir um certo grau de recorrência, podendo, desta maneira, indicar ou delimitar algum tipo de estrutura musical. Assim, não é a partir de qualquer grupo rítmico que os modelos devem ser construídos, mas sim a partir dos grupos rítmicos que se apresentam de maneira recorrente, ou seja, a partir dos padrões musicais³.

Isso posto, define-se que a modelagem do acompanhamento rítmico da Bossa Nova deve ser realizada em dois estágios, a saber: a determinação dos padrões rítmicos utilizados pelos intérpretes ou a *extração de padrões rítmicos*; e a formalização de como tais padrões são utilizados ou a *construção de um modelo* que descreva a interpretação dos padrões previamente encontrados de alguma maneira.

No restante do presente Capítulo, são apresentados os problemas específicos de cada um destes estágios, assim como diversas técnicas que tratam destes problemas específicos. Na Seção 3.1, discorre-se sobre a extração de padrões de maneira geral, ou seja, fora do domínio particular da Música. Na Seção 3.2, detalha-se o problema da descoberta ou extração de padrões musicais, desde a definição do que é um padrão musical, até algumas técnicas para sua extração, passando pela especificação das dificuldades inerentes deste processo. Na Seção 3.3, discute-se a construção de modelos da interpretação musical, suas dificuldades e algumas técnicas de modelagem. Discute-se, ainda, que abordagem

²De fato, repetições de determinados grupos, sejam eles rítmicos, melódicos ou harmônicos, não é uma exclusividade da Bossa Nova: tais repetições são tidas por muitos pesquisadores como um fenômeno perceptivo e estrutural da música de grande importância [Cambouropoulos, 1998, Lerdahl e Jackendorff, 1983, Meredith et al., 2002]. Schenker, por exemplo, diz que a repetição é “a base da música como uma arte” [Schenker, 1954, p. 5, apud [Meredith et al., 2002]].

³Os padrões podem, desta maneira, ser vistos como uma espécie de especialização do conceito de grupo rítmico anteriormente citado.

ou abordagens são mais adequadas para o problema do acompanhamento rítmico.

3.1 EXTRAÇÃO DE PADRÕES

A descoberta, indução ou extração de padrões a partir de dados brutos encerra um conjunto de técnicas e algoritmos bastante estudado na Ciência da Computação. Tais algoritmos servem aos mais variados propósitos, entre eles a própria identificação dos padrões, a correção de erros (*error-correction*) e o estabelecimento de relações entre padrões [Sankoff e Kruskal, 1999]. Os domínios onde estas técnicas e algoritmos podem ser aplicados são os mais diversos possíveis: análise de dados financeiros, análise de estruturas moleculares, reconhecimento de impressões digitais, compressão de dados, etc. Em geral, os dados são representados como seqüências ordenadas que são posteriormente decompostas e analisadas de maneira tal que determinadas recorrências — os padrões — possam ser identificadas.

No âmbito desse conjunto de técnicas e algoritmos, dois conceitos são de fundamental importância, a saber: o conceito de similaridade e o de recorrência. O primeiro define uma maneira através da qual é possível comparar duas seqüências e dizer se elas são similares ou não. Exemplos de medidas que são utilizadas no cômputo da similaridade entre duas seqüências s e t são a distância euclidiana ($[\sum_{i=1}^n (s_i - t_i)^2]^{1/2}$), a distância de Manhattan ou *city block distance* ($\sum_{i=1}^n |s_i - t_i|$) e a distância de Hamming, que corresponde ao número de posições nas quais os elementos correspondentes são diferentes. Já a recorrência diz respeito ao grau de repetição de uma seqüência necessário para que ela seja considerada padrão.

De maneira geral, existem duas classes de algoritmos de extração de padrões: os algoritmos que tratam de padrões exatos e os que tratam de padrões inexatos ou aproximados. Via de regra, nos algoritmos de extração de padrões exatos, utiliza-se, na comparação das seqüências, um procedimento denominado casamento (*matching*). Nele, partes de uma seqüência, chamadas *subseqüências*, são comparadas e, através da medição da similaridade entre elas, consideradas como repetições ou não. Assim, duas subseqüências são consideradas padrões quando são idênticas, isto é, são do mesmo tamanho e formadas pelos mesmos elementos, organizados na mesma ordem. Para realizar esta tarefa, existem algoritmos extremamente eficientes, como, por exemplo, o algoritmo Knuth-Morris-Pratt [Knuth et al., 1977] e o Boyer-Moore [Boyer e Moore, 1977].

Já a segunda classe de algoritmos, os que tratam dos padrões inexatos ou aproxi-

mados, utilizam medidas de similaridade que permitem que algumas diferenças entre seqüências sejam desprezadas, fazendo com que seqüências antes tidas como diferentes sejam consideradas, agora, de alguma forma similares, isto é, sejam consideradas padrões. Uma destas medidas de similaridade é a chamada distância de edição, ou *distância de Levenshtein* [Levenshtein, 1965]. Ela é definida da seguinte maneira.

Definição 3.1 (Distância de Levenshtein). *A distância de Levenshtein ou distância de edição $D_L(S_1, S_2)$ entre duas seqüências S_1 e S_2 é o número mínimo de operações que inserem, apagam ou substituem um símbolo necessário para derivar S_2 a partir de S_1 .*

Em linhas gerais, ela calcula o número mínimo de operações necessárias para a transformação de uma seqüência em outra. Tais operações são a inserção (*insertion*), a deleção (*deletion*) e a substituição (*substitution*). Na inserção, calcula-se o número de inserções necessárias para se transformar uma seqüência na outra. Já na deleção, calcula-se o número de elementos que precisam ser retirados para se transformar uma seqüência na outra. Por fim, na substituição, calcula-se a quantidade de substituições de elementos necessária para que uma seqüência seja transformada em outra. Por exemplo, para se transformar a seqüência CASA em CASAL é necessário *inserir* um elemento, a letra L. No caso de se transformar CAUSA em CASA precisa-se *retirar* a letra U. Já para transformar a seqüência CAÇA na seqüência CASA é necessário que a letra Ç seja *substituída* pela letra S. É possível, também, que combinações de operações sejam necessárias como, por exemplo, na transformação da seqüência CESAR em CASA, onde são necessárias uma deleção (a letra R) e uma substituição (a letra E pela letra A). Neste caso, a seqüência CESAR estaria mais distante (mais dissimilar, portanto) de CASA do que, por exemplo, a seqüência CAUSA, já que seriam necessárias duas operações para realizar a primeira transformação, ao passo que na segunda, uma única operação seria necessária.

Na implementação do cômputo da distância de edição, os algoritmos que realizam a indução de padrões aproximados utilizam uma técnica denominada programação dinâmica. Gusfield [Gusfield, 1997] descreve, além da própria técnica de programação dinâmica, vários algoritmos de indução de padrões inexatos, entre eles o algoritmo de Hirschberg e o método BYP (Baeza-Yates e Perleberg).

3.2 EXTRAÇÃO DE PADRÕES MUSICAIS

Além dos domínios citados na 3.1, a Música também já se mostrou um domínio de aplicação interessante para algoritmos de extração de padrões [Crawford et al., 1998].

Como nos outros casos, ela também pode ser pensada como sendo uma seqüência de eventos (notas, durações, intensidades, etc.) ordenados no tempo. Um padrão musical, desta forma, pode ser considerado como uma determinada subsequência de eventos que se repete acima de um determinado limiar. A extração de um padrão musical consiste, portanto, em identificar automaticamente estes padrões recorrentes.

Existem, todavia, particularidades que fazem com que a extração de padrões a partir de dados musicais seja bastante complexa. A primeira particularidade é que os objetos musicais são multi-atributo, isto é, para se descrever adequadamente um objeto musical qualquer são necessários, via de regra, vários atributos. Para o objeto *nota*, por exemplo, atributos como frequência, nome da nota (dó, ré, mi, etc.), instrumento que a toca, intensidade etc. podem ser utilizados para sua descrição. Uma segunda peculiaridade é que a música é multi-objeto, isto é, vários objetos musicais podem acontecer simultaneamente, como é caso, por exemplo, quando vários instrumentos tocam diferentes notas concomitantemente. Por fim, ainda existe o componente temporal que faz com que os vários objetos multi-atributo possuam uma ordenação no tempo.

Além destas particularidades e das dificuldades que elas impõem, pode-se ainda dizer que um outro problema central na identificação de padrões, de uma maneira geral e na música em particular, é a definição da medida de similaridade, já que é ela que determinará se uma subsequência deve ou não ser considerada similar à outra, isto é, é ela que define de fato o que pode ou não ser considerado recorrência.

A extração de padrões musicais possui ainda outros dois subproblemas básicos: a representação da música e a técnica ou algoritmo de extração de padrões em si. No que concerne a representação, via de regra, as abordagens atuais utilizam cadeias de símbolos (*strings*) como forma de representação da música e, ainda que exista uma grande diversidade na forma como estas cadeias são montadas, é possível reduzi-las a duas categorias básicas: cadeias de eventos (*event strings*) e cadeias de intervalos (*interval strings*) [Meredith et al., 2002].

Nas cadeias de eventos, cada símbolo x_i em uma cadeia $x = x_1 \dots x_n$ representa um evento musical. Se x_i e x_j são símbolos em x , então $j > i$ se, e somente se, o evento representado por x_j ocorrer na música após o evento x_i . Exemplos típicos deste tipo de cadeia são as formadas por símbolos que representam a altura de uma nota. A Figura 3.1 (A) ilustra esta representação. Este tipo de representação tem sido empregada com sucesso na área de Recuperação de Informação Musical, mas, por ser uma representação bastante simples, pode oferecer obstáculos quando se pretende representar música ou ainda quando

há a necessidade de representação de níveis estruturais situados acima do nível da nota. Ainda assim, feitas certas simplificações, é possível utilizar cadeias de eventos para a representação de outros níveis estruturais [Conklin e Anagnostopoulou, 2001].

Já nas cadeias de intervalos, cada símbolo x_k pertencente à cadeia representa a transformação necessária para criar, a partir do evento $e_1(x_k)$, o evento imediatamente consecutivo $e_2(x_k)$. Se x_i e x_j são dois símbolos quaisquer na cadeia x , então $j > i$ se, e somente se, $e_1(x_i) \prec e_2(x_i) \preceq e_1(x_j) \prec e_2(x_j)$, onde \prec significa “precede” e \preceq “precede ou é igual a”. Exemplos de cadeias deste tipo são as formadas por elementos que representam o intervalo musical entre duas notas consecutivas. A Figura 3.1 (B) ilustra este tipo de representação. Exceto por algumas situações, descritas em [Lemström e Ukkonen, 2000], este tipo de representação facilita que ocorrências de padrões invariantes quanto à transposições sejam encontrados. Com este esquema de representação também é mais fácil representar a polifonia, podendo ela, inclusive, ser reduzida a uma cadeia de *bitvectors*⁴ [Lemström e Tarhio, 2000].



Se vo - cê dis - ser que'eu de sa fi - no'a - mor

(A) dó ré mi fá mi ré dó# ré fá réb réb

(B) +2 +2 +1 -1 -2 -1 +1 +3 -4 0

Figura 3.1 Duas representações para o início de *Desafinado*

Na representação da música deve-se, ainda, lidar também com as especificidades do domínio previamente citadas, a saber: objetos multi-atributo, ocorrências de vários objetos simultaneamente e ordenação temporal. Nas representações ilustradas na Figura 3.1, por exemplo, optou-se por uma grande simplificação: no exemplo A apenas os nomes das notas encontram-se representados, enquanto que no exemplo B representa-se apenas o perfil intervalar, isto é, a diferença em semitons entre duas notas consecutivas. Em ambos os casos o aspecto temporal da música não é representado explicitamente: assume-se,

⁴*Bitvectors* são cadeias binárias, isto é, cadeias formadas unicamente pelos numerais zero e um.

na representação, que os eventos acontecem em seqüência, uns após os outros. E ainda, por se tratar, nesse caso específico, de uma linha melódica, não existe polifonia, isto é, não existem notas simultâneas.

É importante notar, ainda, que a representação utilizada pode favorecer ou dificultar a extração de certos tipos de padrões. Ela deve, portanto, ser criteriosamente escolhida de acordo com as características dos padrões que se pretende extrair.

Considerando as duas classes de algoritmos descritas da Seção 3.1, no caso particular da Música, existem algumas particularidades que fazem com que o uso de técnicas que encontram padrões não sejam as mais adequadas. A razão mais importante é que as repetições na música raramente são exatas, principalmente no âmbito da interpretação musical. Mesmo quando exatas, como é algumas vezes o caso na recapitulação na forma sonata [Rosen, 1972], é praticamente impossível que um músico as toque da mesma maneira. Durações e intensidades, por exemplo, certamente serão diferentes em diferentes execuções⁵.

Além desta, existe outra razão que corrobora para a necessidade de técnicas mais elaboradas no domínio da música: a maioria das repetições na música não é significativa do ponto de vista perceptivo, mesmo as repetições exatas. “A tarefa de desenvolver um algoritmo que isole repetições significativas na música envolve a caracterização formal do que de fato é interessante em certas repetições e do que as distinguem de outras tantas que o ouvinte não percebe e o analista não considera importante” [Meredith et al., 2002]. Isto certamente não é uma tarefa trivial, já que existe uma variedade muito grande na classe de repetições musicalmente significativas, e os algoritmos de casamento exato certamente falhariam em lidar com esta variedade.

Meredith *et al.* apontam, ainda, duas razões principais para a existência de uma grande variedade na classe de repetições musicalmente significativas: (1) características estruturais das repetições, que fazem com que elas possam variar desde pequenos motivos, a seções inteiras de uma música; e (2) as diferentes maneiras pelas quais um padrão pode ser modificado. Estas razões podem, ainda, ser postas de outra maneira. A primeira diz respeito ao nível da hierarquia na estrutura musical onde a extração de padrões é realizada, ou ainda, ao nível da estrutura musical onde os padrões mais importantes do

⁵Vale lembrar que a menção à interpretação aqui deve-se ao contexto no qual esta pesquisa está inserida. É claro que se a idéia for apenas a análise da partitura, isto é, se a análise for da notação da música, os algoritmos que realizam o casamento exato se constituirão em excelentes ferramentas, já que casamentos exatos, apesar de raros, existirão e, possivelmente, marcarão pontos importantes na estrutura da música, como é o caso da recapitulação anteriormente citado.

ponto de vista perceptivo se encontram. Motivos e seções, por exemplo, estão, claramente, em níveis estruturais diferentes. Cabe ao analista decidir em que nível da estrutura sua análise se dará. No caso do ouvinte, a percepção das repetições dependerá de sua experiência no idioma musical em questão [Lerdahl e Jackendorff, 1983].

Já a segunda razão diz respeito à grande variedade de técnicas composicionais possíveis. Na forma musical conhecida como “tema com variações”, por exemplo, os compositores utilizam um grande conjunto de técnicas que fazem com que o tema possa ser modificado, em maior ou menor grau, mas mantenha características que permitem que o resultado de tais modificações possa ainda ser relacionado ao tema original. Desta forma, é praticamente impossível determinar *a priori* um conjunto de restrições que sirva de guia durante a extração de possíveis padrões. Em suma, pode-se dizer que, para o caso da música, um algoritmo de extração de padrões deve ser flexível o suficiente para que estas particularidades do domínio possam ser contempladas, ou seja, deve-se favorecer técnicas e algoritmos relacionados à extração de padrões aproximados.

Uma dessas possíveis técnicas é permitir a existência de buracos (*gaps*) nas seqüências. Com a existência de buracos, é possível, por exemplo, que padrões que incorporem silêncios ou pausas em suas repetições possam ser encontrados. No exemplo ilustrado na Figura 3.2 — trechos do segundo movimento da Sonata para Piano Op. 57 (*Appassionata*) de Ludwig van Beethoven — um algoritmo que não leve em consideração a existência de buracos falharia em encontrar a ocorrência do padrão da mão direita dos compassos 1–3 que se encontra nos compassos 17–19. Note que as notas são exatamente as mesmas, residindo a diferença apenas em suas durações, que se encontram, na ocorrência dos compassos 17–19, reduzidas pela metade por meio de pausas.

No caso particular da interpretação, esta técnica é ainda mais essencial, uma vez que, como já foi exposto, o instrumentista, ao tocar, realiza diversas modificações que podem alterar as durações das notas, aumentando-as ou diminuindo-as, a depender das intenções do intérprete, permanecendo a notação, quando existente, exatamente a mesma. Isso, entretanto, não muda o fato de que existem repetições estruturais que, a despeito de modificações realizadas na sua interpretação, encontram-se definidas na própria notação (partitura, no exemplo) e são visíveis após a análise desta. As repetições da forma sonata anteriormente citada constituem um exemplo entre muitos. De fato, há pesquisas, como a descrita na Seção 2.4.2, que se baseiam na hipótese de que as modificações realizadas pelos músicos refletem justamente suas intenções em “comunicar” ao ouvinte a estrutura da música. Desta forma, é importante que estas repetições sejam identificadas na música

The image displays two musical excerpts from the piano part of Beethoven's 'Appassionata'. The top excerpt, labeled 'Compassos 1-3', shows measures 1-3 with dynamics 'p' e dolce' and 'sfp', and the word 'continua'. The bottom excerpt, labeled 'Compassos 17-19', shows measures 17-19 with the word 'continua'. A dashed arrow points from the end of the first excerpt to the beginning of the second, illustrating an inexact repetition of a pattern.

Figura 3.2 Repetição inexata de um padrão na *Appassionata*

interpretada, uma vez que podem dar pistas importantes de como a estrutura da música é trabalhada pelo intérprete.

Além de permitir a existência de buracos, outras técnicas podem ser utilizadas a fim de se obter casamentos aproximados. Uma das possibilidades é utilizar uma medida de similaridade que, em seu cômputo, leve em conta diferenças entre seqüências, como, por exemplo, a distância de edição previamente descrita.

Em um estudo sobre a aplicação da distância de edição no domínio da música [Mongeau e Sankoff, 1990], propôs-se, ainda, que outras duas operações fossem adicionadas: a fragmentação (*fragmentation*) e a consolidação (*consolidation*). A fragmentação consiste na divisão de um elemento da seqüência em vários, de maneira que a soma das durações individuais dos elementos resultantes não modifique a duração do elemento original. A consolidação é a operação inversa da fragmentação, isto é, ela é a união de elementos da seqüência feita de maneira que a soma das durações individuais destes elementos seja preservada (Figura 3.3).

A seguir (Seções 3.2.1 a 3.2.4), segue a descrição mais detalhada de alguns algoritmos e abordagens para o problema da extração de padrões musicais. Na Seção 3.2.5, é feita ainda uma discussão sobre a abordagem ou abordagens que seriam mais apropriadas para os objetivos desta pesquisa.

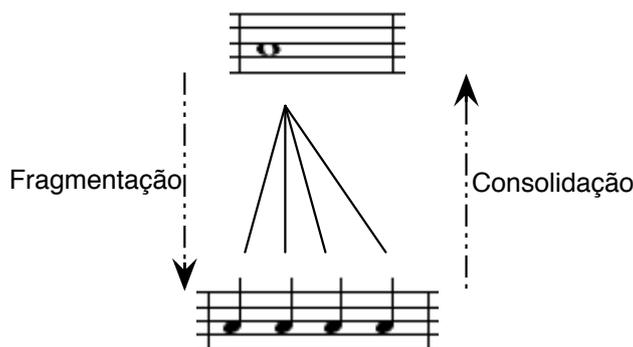


Figura 3.3 As operações de fragmentação e consolidação

3.2.1 GCTMS: General Computational Theory of Musical Structure

A GCTMS é uma “teoria que pode ser empregada para se obter uma descrição estrutural — ou conjunto de descrições — de uma superfície musical” [Cambouropoulos, 1998]. A teoria é independente de estilo ou idioma musical, uma vez que se baseia em princípios lógicos e cognitivos gerais, e pode ser aplicada a qualquer superfície musical⁶. Sua meta é alcançar uma descrição estrutural da obra musical que possa ser considerada “plausível” ou, ao menos, “possível” por parte do analista humano.

A análise nesta teoria é composta de dois estágios distintos, mas intimamente ligados: (1) o estabelecimento de uma série de componentes individuais que abordam tarefas analíticas especializadas; e (2) uma descrição elaborada e detalhada de como estes componentes se relacionam e interagem uns com os outros de forma que descrições estruturais plausíveis de uma superfície musical possam ser alcançadas.

Para a realização do primeiro estágio, foram desenvolvidos algoritmos que lidam com aspectos específicos da superfície musical, como por exemplo, um algoritmo que modela as estruturas métrica e de acentuação e outro para a indução de padrões. Já no segundo estágio, a descrição é obtida comparando-se os resultados dos algoritmos empregados no estágio anterior.

Este processo analítico pode ser acompanhado na Figura 3.4: uma superfície musical, composta por eventos discretos é convertida para uma outra superfície — *superfície musical (1)* — que é formada por perfis intervalares em vários níveis de abstração (para as alturas, por exemplo, intervalos exatos, intervalos conjuntos, intervalos disjuntos, contor-

⁶O autor implementou um protótipo do sistema e algumas partes encontram-se implementadas como parte da biblioteca *Kant* [Meudic, 2000] para o ambiente *OpenMusic* [Agon, 1998].

nos, etc.). Esta nova superfície é submetida a um algoritmo que localiza possíveis rupturas. O resultado, denominado proto-segmentação, é então submetido a outros algoritmos que tentam validar os segmentos encontrados ou mesmo identificar novos segmentos. Por fim, a segmentação, ou conjunto de segmentações, é organizada e rotulada baseada em conceitos de similaridade.

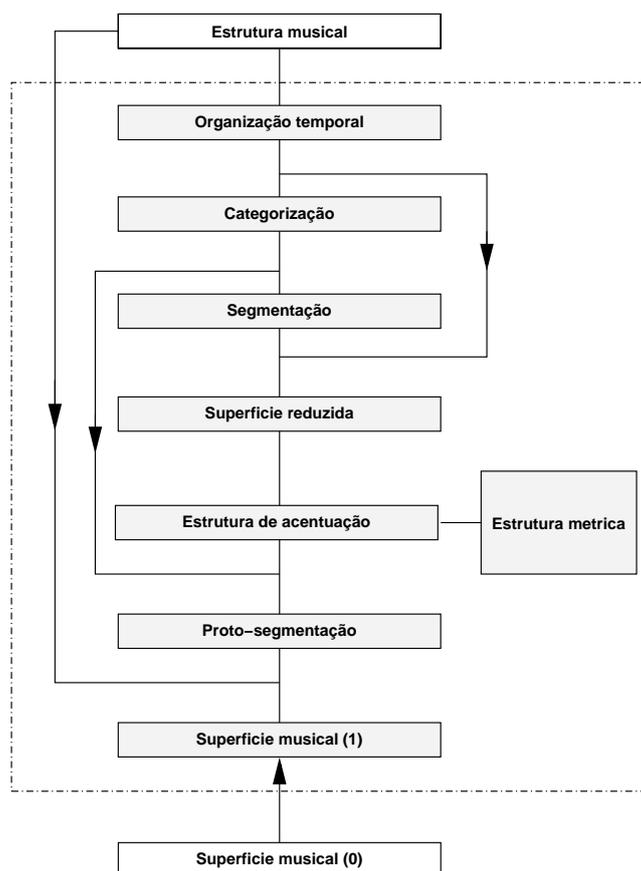


Figura 3.4 A GCTMS

Para realizar a análise da *superfície reduzida*, o autor utilizou um algoritmo desenvolvido por Crochemore [Crochemore, 1981]. Neste algoritmo, os padrões são fatores periódicos máximos (*maximal periodic factors*) de uma cadeia de símbolos. Se $x = x_1 \dots x_n$ é uma cadeia, então w é um fator se, e somente se, existe um i e um j tal que $w = x_i \dots x_j$. Posto de outra forma, um fator é um padrão que não contém buracos. Uma cadeia w é um fator periódico de outra cadeia x se, e somente se, w é um fator de x e w pode ser formada pela concatenação de duas ou mais cópias de alguma outra cadeia y . A complexidade temporal do algoritmo de Crochemore é de $O(n \log n)$.

Uma observação final sobre a GCTMS [Cambouropoulos, 1998, p. 46]:

“[ela] não é uma teoria linear onde a análise é feita de maneira unidirecional, seja ela bottom-up ou top-down. Ela também não é uma teoria onde agentes totalmente independentes interagem uns com os outros. Ela é uma teoria onde diferentes componentes, em vários níveis, interagem uns com os outros (resultados de níveis analíticos mais baixos facilitam o emprego de procedimentos de nível mais alto e resultados destes, por sua vez, informam e ajudam a esclarecer resultados de níveis analíticos mais baixos) e onde existe uma direcionalidade relativamente livre [...] (por exemplo, não é computacionalmente prático nem cognitivamente plausível começar com o modelo de categorização antes que alguma segmentação, mesmo que preliminar, tenha sido obtida)”.

—E. CAMBOUROPOULOS

3.2.2 FIEXPAT

FLEXPAT é um algoritmo de extração capaz de realizar casamento aproximado de padrões [Rolland, 1999, Rolland, 2001]. Ele baseia-se em algoritmos da bioinformática utilizados para a análise de cadeias de DNA, e tem como medida de similaridade a distância de edição em conjunto com as operações de fragmentação e consolidação descritas anteriormente. Sua entrada consiste de uma seqüência de dados simbólicos, representando uma linha monofônica⁷, e sua saída é formada por um conjunto de padrões. Cada um dos padrões extraído, denominado classe, é formado por dois elementos: um protótipo e um conjunto de ocorrências deste protótipo. O protótipo é a subsequência que melhor representa as subsequências de uma mesma classe. Já as ocorrências são as subsequências, similares ao protótipo, que foram encontrados na seqüência original.

Como diversos algoritmos de extração de padrões, FLEXPAT funciona em duas etapas consecutivas, como ilustrado na Figura 3.5: a etapa de comparação, onde a seqüência de entrada é dividida em subsequências e as similaridades entre todos os pares de subsequências são calculadas; e a etapa de classificação, onde as classes, isto é, os protótipos e suas ocorrências são identificadas⁸.

A primeira etapa, a etapa de comparação, tem como objetivo a construção de um grafo, chamado grafo de equípolência. Equípolência é definida da seguinte maneira: duas seqüências s_1 e s_2 são ditas equípolentes se, e somente se, sua similaridade é maior ou

⁷É possível, com algumas simplificações, também representar a polifonia.

⁸Para uma descrição formal do algoritmo *vide* [Rolland, 2001].

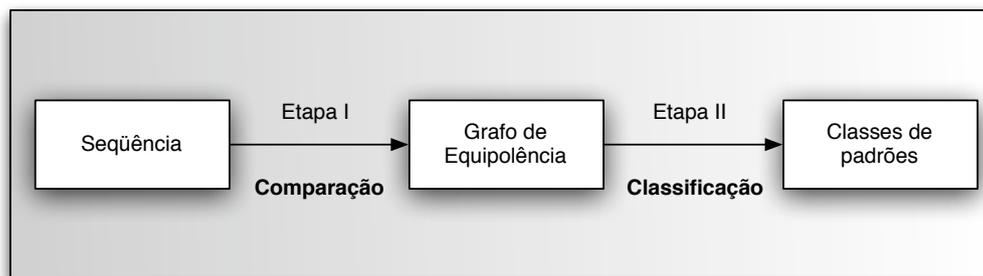


Figura 3.5 Etapas do algoritmo FLEXPAT

igual a um dado limiar⁹. O grafo de eqüipolência é um grafo não direcional onde cada vértice representa uma subseqüência, enquanto que as arestas indicam as similaridades entre estas subseqüências. Desta forma, o grafo é uma estrutura que representa todas as subseqüências e as similaridades entre elas.

O problema de encontrar este grafo pode ser formalizado da seguinte maneira: considere uma seqüência s de tamanho n . Chama-se de $s_{[i][t]}$ a subseqüência de s que começa na posição i e tem tamanho t , onde $1 \leq i \leq n$ e $t \leq (n - i)$. A construção do grafo se dará pela cálculo da função de similaridade \mathcal{S} a seguir.

$$\mathcal{S}(s_{[i_1][t_1]}, s_{[i_2][t_2]}), \forall i_1, i_2, t_1, t_2, i_1 \geq 1, i_2 \leq n, t_1 \leq (n - i_1), t_2 \leq (n - i_2) \quad (3.1)$$

Posto desta forma, entretanto, o problema é extremamente custoso do ponto de vista computacional: considerando o casamento exato, que, como dito anteriormente, é bastante eficiente, teríamos um limite superior de $O(n^5)$. Medidas de similaridade que computam o casamento aproximado levariam a ordens de complexidade ainda maiores. Todavia, é possível, utilizando técnicas de otimização como a programação dinâmica [Cormen et al., 2002], reduzir a ordem de complexidade do problema. Além disso, o próprio algoritmo possui alguns parâmetros de entrada que podem ser ajustados e que funcionam como restrições, ajudando na diminuição da ordem de complexidade do algoritmo. Estes parâmetros são os seguintes:

- o tamanho mínimo dos padrões (t_{min}), que faz com que apenas subseqüências de

⁹O termo equipolência não é utilizado porque implica em uma série de propriedades matemáticas (transitividade, por exemplo) que não são, necessariamente, válidas neste caso específico [Rolland, 2001].

tamanho maior que este valor sejam consideradas;

- o tamanho máximo dos padrões (t_{max}), de função similar à do parâmetro anterior;
- a máxima diferença de tamanho entre padrões (dt_{max}), que evita que subsequências de tamanhos muito diferentes sejam testadas;
- a máxima sobreposição entre os padrões (sob_{max}), que faz com que duas subsequências sejam comparadas apenas se a quantidade de elementos que se sobrepõem for menor que esta constante e
- o limiar de similaridade local (l), que define a equípolência entre duas subsequências.

Utilizando programação dinâmica, é possível armazenar os valores já calculados e reaproveitá-los em cálculos posteriores, reduzindo, assim, a complexidade temporal¹⁰. No caso de FLEXPAT, os valores das similaridades já calculadas para as subsequências de tamanho $t - 1$ e $t' - 1$ são aproveitados para calcular a similaridade das subsequências de tamanho t e t' , de maneira indutiva, de acordo com a seguinte expressão

$$\mathcal{S}(s_{[i][t]}, s_{[j][t']}) = \min \left\{ \begin{array}{l} \mathcal{S}(s_{[i][t-1]}, s_{[j][t'-1]}) + \mathit{SUB}(s_{[i+t]}, s_{[j+t']}) \\ \mathcal{S}(s_{[i][t-1]}, s_{[j][t']}) + \mathit{DEL}(s_{[i+t]}) \\ \mathcal{S}(s_{[i][t]}, s_{[j][t'-1]}) + \mathit{INS}(s_{[j+t']}) \end{array} \right\}, \quad (3.2)$$

onde as funções $\mathit{SUB}(x, y)$, $\mathit{DEL}(x)$ e $\mathit{INS}(x)$ calculam os custos de se substituir o elemento x pelo elemento y , apagar o elemento x e inserir o elemento x na seqüência s , respectivamente. Desta forma, com o uso dos parâmetros de entrada e da programação dinâmica, a complexidade, tanto temporal, quanto espacial, nesta etapa do algoritmo diminui para $O(n^2)$.

Uma vez construído o grafo de equípolência, passa-se para a etapa de classificação, a segunda etapa do algoritmo. Nela, as subsequências similares são agrupadas em classes. O algoritmo é mais simples que o algoritmo da etapa anterior e menos caro computacionalmente. Sua entrada consiste do próprio grafo de equípolência $G(V, A)$ e de uma constante q , chamada limiar de *quorum* mínimo, que indica quantas recorrências de uma determinada subsequência são necessárias para que ela seja considerada uma classe. O

¹⁰É importante notar, entretanto, que a complexidade espacial aumenta devido à necessidade de se armazenar os valores já calculados. No caso de FLEXPAT, segundo o autor, “foram desenvolvidas técnicas que permitem que, em qualquer momento, apenas t_{max}^2 valores de similaridades precisam estar armazenados na memória” [Rolland, 2001].

algoritmo em si consiste no cálculo, para cada vértice $v \in V$ do grafo de equípólencia, da similaridade vizinha total de v ($simVizTotal(v)$). Esta função retorna um conjunto de vértices similares a cada vértice v do grafo. Cada conjunto, então, é comparado ao *quorum* mínimo e, caso possua um número maior de subseqüências, é considerado uma classe. Cada classe é representada na forma de um subgrafo estrela, que é um subgrafo que satisfaz as seguintes propriedades: possui p vértices $\{v_1, v_2, \dots, v_p\}$ e $p - 1$ arestas; e existe um inteiro c , $1 \leq c \leq p$, tal que v_c é o vértice central deste subgrafo, isto é, existe uma aresta que liga v_c a todos os outros vértices. A Figura 3.6 ilustra um subgrafo deste tipo. O vértice central (em negrito) representa o protótipo da classe, enquanto que os demais vértices ligados a ele representam suas ocorrências. Nas arestas, encontra-se o grau de similaridade entre os vértices.

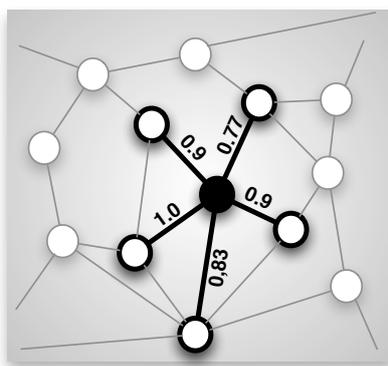


Figura 3.6 Exemplo de um subgrafo estrela

O algoritmo foi testado com o material musical resultante das pesquisas realizadas por Owens [Owens, 1974], isto é, todos os solos de Charlie Parker. Estes solos, devidamente convertidos, resultaram em uma base de seqüências com cerca de 60.000 eventos (notas ou pausas). Imitando a metodologia empregada por Owens, foi possível encontrar novamente grande parte dos padrões originais. Além disso, pelo menos uma classe padrões não identificada por Owens foi encontrada pelo algoritmo.

As duas etapas do algoritmo encontram-se descritas de maneira simplificada a seguir (Algoritmos 1 e 2)¹¹.

¹¹Para uma descrição completa e detalhada *vide* [Rolland, 2001].

Algoritmo 1 Etapa de comparação do FLEXPAT

Para todos os pares de subsequências de s , $s_{[i][t]}$ e $s_{[j][t']}$,
por ordem crescente de t e t' , faça

$$\begin{aligned}
 cDel &\leftarrow \mathcal{S}(s_{[i][t-1]}, s_{[j][t']}) + \mathcal{DEL}(s_{[i][t]}) \text{ \{cálculo do custo de deleção\}} \\
 cIns &\leftarrow \mathcal{S}(s_{[i][t]}, s_{[j][t'-1]}) + \mathcal{INS}(s_{[j][t']}) \text{ \{cálculo do custo de inserção\}} \\
 cSub &\leftarrow \mathcal{S}(s_{[i][t-1]}, s_{[j][t'-1]}) + \mathcal{SUB}(s_{[i][t]}, s_{[j][t']}) \text{ \{cálculo do custo de substituição\}} \\
 \mathcal{S}(s_{[i][t]}, s_{[j][t']}) &\leftarrow \min(cDel, cIns, cSub)
 \end{aligned}$$

Se $\mathcal{S}(s_{[i][t]}, s_{[j][t']}) \geq l$ $\{l$ é o limiar de similaridade $\}$ **Então**

Adicione uma aresta no grafo de equiipolência entre estas duas
subsequências, com o valor de \mathcal{S} calculado

fim Se

Algoritmo 2 Etapa de classificação do FLEXPAT

Criar uma lista de padrões lp , inicialmente vazia

Para cada vértice v no grafo de equiipolência

Compute a $simVizTotal(v)$

Se $simVizTotal(v) > q$ $\{q$ é o *quorum* mínimo $\}$ **Então**

adicione a lp o maior subgrafo estrela cujo vértice central é v

fim Se

3.2.3 SimilaritySegmenter

SIMILARITYSEGMENTER [Madsen e Widmer, 2005, Madsen e Widmer, 2006a] é um algoritmo genético capaz de encontrar padrões em uma música representada na forma de um grafo. A idéia central é a de que um padrão ou *objeto musical*, como denominado pelo autor, pode ser derivado de outro a partir de uma seqüência de operações ou *transformações* [West et al., 1991]. Desta forma, para se determinar se um objeto musical é similar a outro ou não — a ocorrência de um padrão, portanto — dois passos são necessários: determinar quais são as funções de transformação e qual deve ser a seqüência de aplicação destas funções. De maneira geral, tais transformações podem ser classificadas em dois tipos: simples e não-simples¹². No primeiro caso, o objeto resultante da aplicação da função de transformação mantém o mesmo tamanho. Já no segundo caso, a aplicação da função de transformação resulta em um objeto de tamanho diferente do original.

Como dito, a busca por padrões é feita a partir de um grafo que representa a música. Trata-se de um grafo dirigido, construído da seguinte forma: cada evento musical, pausa ou nota, é representado por um vértice e a aresta que une dois vértices do grafo representa um tipo determinado de relação entre eles, relação esta especificada pelo rótulo da aresta. Estes rótulos, por sua vez, podem ser de dois tipos: “*follows*”, indicando que os dois eventos representados pelos vértices ocorrem um após o outro, ou “*simultaneous*”, indicando que os eventos começam no mesmo ponto do tempo. Na Figura 3.7 mostra-se um exemplo deste grafo. Nela, três eventos musicais, *i*, *j* e *k*, encontram-se representados e as arestas indicam os relacionamentos entre eles: os vértices *i* e *k* acontecem simultaneamente, enquanto os vértices *i* e *j*, assim como *k* e *j*, acontecem sequencialmente.

Em linhas gerais, o algoritmo funciona da seguinte maneira: inicialmente, uma população de *Ocorrências Similares* — doravante OS — é inicializada aleatoriamente¹³. Cada OS representa uma estimativa de que dois subgrafos são similares. Utilizando as operações usuais dos algoritmos genéticos, isto é, cruzamento (*crossover*), mutação e seleção, e modificando o tamanho e a posição de cada OS, o algoritmo evolui sua população com o intuito de identificar subgrafos cada vez mais similares. Após um certo número de gerações, o algoritmo termina e o OS mais apto, determinado por uma função objetivo (*fitness function*), é comparado a um limiar a fim de verificar se os subgrafos

¹²Manteve-se aqui o termo usando no original — *non-simple*.

¹³No original, o termo utilizado é *Similarity Statement*.

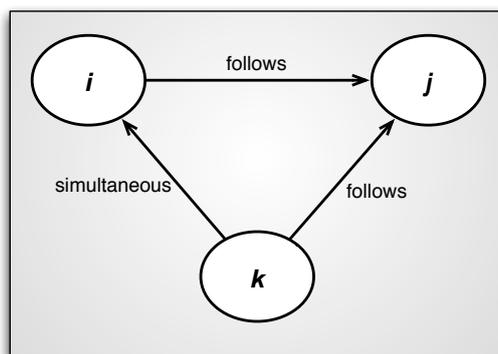


Figura 3.7 Exemplo de parte de um grafo utilizado por SIMILARITYSEGMENTER

são suficientemente similares para serem considerados padrões. Caso isto aconteça, uma busca é feita para encontrar outras possíveis ocorrências do padrão encontrado, antes que uma nova população seja gerada e o processo reinicie. Desta forma, a função objetivo pode ser utilizada para determinar os tipos de similaridades desejadas, enquanto que o limiar de similaridade define quanta dissimilaridade será permitida.

Este algoritmo foi testado na Fuga em Dó Menor (BWV 847) do Cravo Bem-Temperado (*das wohltemperierte Klavier*) de J. S. Bach, sendo capaz de encontrar diversos padrões e segmentando a peça coerentemente.

3.2.4 Outros

Além dos trabalhos anteriormente descritos, outros ainda merecem menção. Estes, todavia, não serão considerados na discussão já que se tratam de trabalhos que não são os mais adequados para os objetivos deste trabalho.

Um trabalho bastante citado na literatura de extração de padrões musicais é o realizado por Hsu *et al.* [Hsu et al., 1998]. Nele, para realizar a extração de padrões de linhas monofônicas, utiliza-se uma estrutura de dados chamada matriz de correlação (*correlative matrix*) e técnicas de programação dinâmica. Como no caso do algoritmo FLEXPAT, esta abordagem possui duas etapas: a construção da matriz e a identificação dos padrões propriamente ditos. A matriz de correlação é uma matriz $n \times n$, onde n é o tamanho da melodia a ser analisada, que armazena resultados intermediários. Estes resultados são construídos comparando-se a i -ésima à j -ésima nota da melodia. Caso elas sejam as mesmas, atribuí-se 1 à célula da matriz determinada pela i -ésima linha e j -ésima coluna. Além disso, se a $(i + 1)$ -ésima e $(j + 1)$ -ésima notas forem iguais, atribuí-se 2 à posição

da matriz determinada pela $(i + 1)$ -ésima linha e $(j + 1)$ -ésima coluna e assim por diante. A matriz representa, desta forma, o conjunto de candidatos a padrões e contém todos os elementos da melodia, mesmo os que só aparecem uma única vez, juntamente com a sua frequência. A complexidade temporal de construção desta matriz é de $O(n^2)$. Na segunda etapa da abordagem, a matriz de correlação é analisada e os candidatos que não formam um padrão são eliminados. De uma maneira geral, o pior caso do algoritmo é $O(n^4)$, mas, na prática, ele é mais eficiente.

HUMDRUM é um sistema computacional de apoio à análise musical que permite que analistas e pesquisadores façam “perguntas” sobre a estrutura de uma determinada obra musical [Huron, 1999]. Ele consiste de um conjunto de mais de 70 ferramentas capazes de manipular os dados ASCII descritos por uma gramática. Estas manipulações podem ser de vários tipos: busca, contagem, classificação e comparação são alguns exemplos de manipulações realizadas pela ferramenta. Entre essas ferramentas, HUMDRUM oferece alguns algoritmos de seleção e extração de informação, de reconhecimento de padrões e de classificação. Apesar do sistema facilitar investigações sobre a estrutura musical, caso o analista não possua um problema ou pergunta clara a ser colocada, o sistema será de pouca utilidade, já que ele é incapaz de construir um modelo da obra sozinho.

SIA e SIATEC [Meredith et al., 2001, Meredith et al., 2002] são algoritmos que diferem dos algoritmos até agora expostos: ao contrário dos algoritmos anteriores, que utilizam o processamento de cadeias, estes algoritmos utilizam uma abordagem geométrica para a extração de padrões. Nela, a música a ser analisada é representada como um conjunto de dados multidimensionais (*multidimensional datasets*), isto é, um conjunto de pontos em um espaço euclidiano multidimensional, e os padrões são encontrados através de um pequeno número de projeções ortogonais desta representação multidimensional da música. Segundo os autores, “certos tipos de repetições significativas do ponto de vista perceptivo, que não podem ser descobertas com algoritmos de processamento de cadeias, podem ser descobertas facilmente com algoritmos que processem conjuntos de dados multidimensionais”. Ainda segundo os autores, este método não é uma substituição dos métodos baseados em cadeias, sendo apenas uma alternativa para a descoberta de padrões que são difíceis ou praticamente impossíveis de serem descobertos com técnicas tradicionais.

Os algoritmos foram testados em 52 conjuntos de dados de dimensionalidade e tamanho variados (2 a 5 dimensões e tamanho de 6 a 3500 pontos), a maioria retirada do Cravo Bem-Temperado de J. S. Bach. De maneira geral, o número de padrões encon-

trados por SIA constitui apenas uma pequena fração de todos os padrões do conjunto de dados. Quanto a SIATEC, os experimentos realizados sugerem que, utilizado com heurísticas bem escolhidas, é possível isolar diversas repetições interessantes. Mesmo assim, SIATEC ainda encontra milhares de repetições, mesmo em peças relativamente curtas, e apenas algumas destas são, de fato, relevantes. Na Figura 3.8 pode-se observar um exemplo dos resultados obtidos com o algoritmo SIA. Nela duas classes de repetições foram encontradas: uma formada pelos padrões A e D e outra pelos padrões C e B.

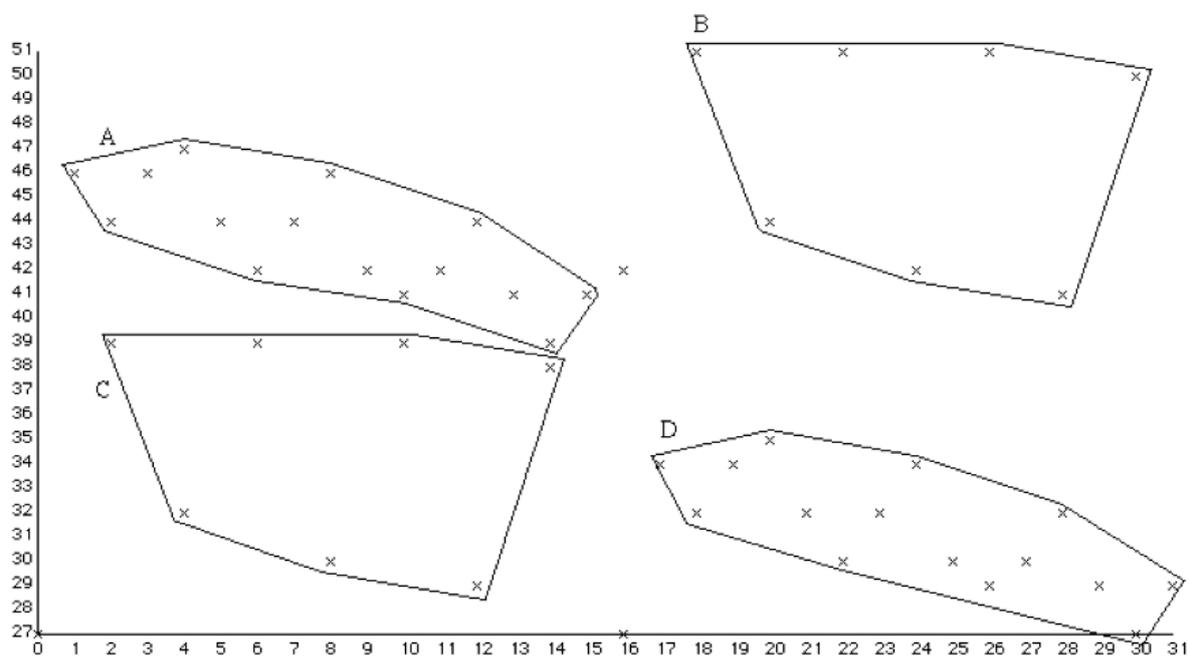


Figura 3.8 Exemplo de padrões encontrados pelo algoritmo SIA

Outros trabalhos importantes sobre a extração de padrões foram desenvolvidos na Finlândia [Perttu, 2000, Lemström et al., 1998, Lemström et al., 1999, Lemström, 2000, Lemström et al., 2001]. Em geral, estes trabalhos tratam, além da extração de padrões a partir de cadeias em si, da aplicação destas técnicas em uma área chamada Recuperação de Informação Musical (*Music Information Retrieval*—MIR).

3.2.5 Discussão

A presente discussão tem como objetivo identificar, entre os algoritmos e ferramentas anteriormente apresentadas, as que mais se adequam à presente pesquisa.

Na abordagem descrita na Seção 3.2.1, os tipos de padrões que o algoritmo é capaz

de identificar são muito restritos. Eles não podem conter buracos, o que diminui a possibilidade de se encontrar padrões onde existe variação na articulação, e estes padrões só podem ser identificados em música monofônica. Para diminuir estas limitações, uma alternativa seria a utilização de um outro algoritmo de extração de padrões. Isto, todavia, implicaria na avaliação da consistência dos resultados obtidos, pois não fica claro, a despeito das observações sobre a relativa independência entre os módulos, se a teoria é suficientemente modular para sofrer uma mudança deste tipo sem afetar a consistência dos resultados. Cada módulo parece estar limitado por resultados de etapas anteriores da teoria, em especial por resultados da etapa de proto-segmentação (*vide* Figura 3.4, página 47). Por fim, para utilizá-la, seria necessário uma adesão à teoria, o que, não necessariamente, é o ideal no caso deste trabalho.

As abordagens descritas nas Seções 3.2.2 e 3.2.3 são bem mais elaboradas. Tanto em FLEXPAT, quanto em SIMILARITYSEGMENTER, a flexibilidade, algo bastante importante para o problema aqui tratado, é uma característica relevante: a possibilidade de ajuste dos pesos e dos parâmetros de entrada no primeiro algoritmo, e a possibilidade de definir a função objetivo e o limiar de similaridade no segundo, claramente favorecem à experimentação por parte do usuário.

Assim, considerando principalmente a necessidade de flexibilidade na extração de padrões imposta pelo domínio, os algoritmos FLEXPAT e SIMILARITYSEGMENTER se constituem em bons candidatos para a identificação de padrões na Bossa Nova¹⁴.

Antevêem-se, porém, alguns problemas com estes algoritmos: o primeiro é que FLEXPAT é incapaz de extrair padrões em músicas polifônicas. O segundo problema vem de observações feitas por Mongeau e Sankoff em um artigo sobre a extração de padrões musicais em seqüências [Mongeau e Sankoff, 1990]. Nele, os autores afirmam que “a atribuição de pesos às transformações [i. e., às operações de edição] deve corresponder aos tipos de diferenças musicais que se pretende medir”. Isto é bastante lógico, mas é difícil estabelecer com clareza a relação entre estes pesos e os objetivos analíticos. O que significa, objetivamente, atribuir pesos 2, 2 e 3 às operações de inserção, deleção e substituição, respectivamente, e qual a relação destes valores com a idéia de encontrar recorrências na Bossa Nova? E se estes valores fossem 2, 3 e 3 os resultados seriam significativamente diferentes?

¹⁴Nesta pesquisa, a extração de padrões está restrita à dimensão rítmica da música. Estes algoritmos, todavia, podem ser empregados na extração de padrões em outras dimensões, como a melódica ou harmônica.

Por fim, o Quadro 3.1 resume características relevantes de alguns dos algoritmos aqui apresentados.

	Hsu	GCTMS	SIA	SIATEC	FLEXPAT	SIMSEG.
permite buracos	não	não	sim	sim	sim	sim
casamento aproximado	não	sim	restrito	restrito	sim	sim
invariância à transposição	sim	sim	sim	sim	sim	sim
considera polifonia	não	não	sim	sim	restrito	sim
complexidade temporal	$O(n^4)$	$O(n \log n)$	$O(n^2 \log_2 n)$	$O(n^3)$	$O(n^2)$?
complexidade espacial	?	?	$O(n^2)$	$O(n^3)$	$O(n^2)$?

Quadro 3.1: Características de alguns algoritmos de extração de padrões musicais

3.3 DESCOBERTA AUTOMÁTICA DE CONHECIMENTO MUSICAL

Como apontado no início deste Capítulo, uma vez identificados os padrões rítmicos utilizados pelo intérprete, é necessário estabelecer que elementos fazem com que este ou aquele padrão seja escolhido, isto é, um modelo que explique como o acompanhamento rítmico é feito. Isto significa que os dados musicais devem ser examinados de forma tal que regularidades consistentes e gerais sejam identificadas.

A descoberta automática de conhecimento (*Knowledge Discovery in Databases* ou KDD) é uma área da Ciência da Computação onde diversos métodos, de diversas disciplinas, são utilizados para que conhecimento possa ser obtido a partir de uma certa quantidade de dados armazenados em algum repositório¹⁵. Os exemplos de sucesso da aplicação destes métodos, nos mais diversos domínios, são vários: análise de dados financeiros (concessão de crédito, detecção de crimes financeiros, etc.), classificação automática de objetos celestes, prevenção de falhas em transformadores elétricos, predição de grandes tempestades, e em diversas aplicações em várias áreas do conhecimento humano como a Medicina, Genética, Telecomunicações e na Música [Dannenber, 2000, Han e Kamber, 2001, Langley e Simon, 1995, Widmer, 2000, Widmer, 2003, Widmer et al., 2003].

¹⁵Prefere-se o uso do termo repositório, já que banco de dados pode induzir o leitor a pensar em um banco de dados real, o que nem sempre é o caso. De fato, muitas vezes os dados estão disponíveis em arquivos texto simples.

A descoberta de conhecimento envolve um processo formado por um conjunto de estágios, conjunto esse que é bem conhecido e documentado na literatura científica da área [Han e Kamber, 2001, Witten e Frank, 2000]. Este processo é comumente chamado de *ciclo de descoberta do conhecimento* (Figura 3.9). Este ciclo, por sua vez, é formado pelas seguintes etapas:

- 1) aquisição de dados, onde os dados são obtidos;
- 2) pré-processamento, onde diversas técnicas são utilizadas no preparo dos dados para etapa seguinte;
- 3) construção do modelo, onde métodos inteligentes são empregados para que regularidades nos dados sejam efetivamente encontradas¹⁶; e
- 4) uma última etapa onde o conhecimento encontrado é validado e, eventualmente, visualizado, com a ajuda de ferramentas adequadas¹⁷.

Cada uma destas fases possui características, dificuldades e métodos próprios. Na aquisição de dados, a depender do domínio, a dificuldade pode ser a própria obtenção dos dados. A Música, por um infeliz acaso, é um destes domínios. No caso particular da análise de interpretações, as fontes de dados são, basicamente, duas: dados obtidos a partir de gravações em CDs ou dados obtidos com gravações em instrumentos computadorizados ou instrumentos adaptados e ligados a computadores. No primeiro caso, apesar da evidente profusão de gravações de qualidade, como já foi discutido (Seção 2.3), existem algumas limitações nos atributos que podem ser obtidos ou derivados a partir deste tipo de representação, o que pode limitar ou dificultar a aquisição. No segundo caso, os instrumentos computadorizados são caros e, em sua maioria, de difícil acesso. Já os instrumentos adaptados, que são ligados a computadores, padecem de limitações tecnológicas e podem não ser muito precisos.

Existe ainda uma terceira possibilidade: o uso de tecnologias como OCR (*Optical Character Recognition*) para a obtenção de dados a partir da notação musical (partituras, principalmente). Mas, existem ainda limitações tecnológicas que impossibilitam o uso

¹⁶Muitas vezes esta etapa, também conhecida como *mineração de dados*, é confundida com o próprio processo de descoberta de conhecimento. Mas, como dito por Han & Kamber [Han e Kamber, 2001], trata-se apenas de uma das fases do processo.

¹⁷É importante notar que este processo é realmente cíclico: pode-se, a partir de qualquer etapa, retornar para uma etapa anterior.

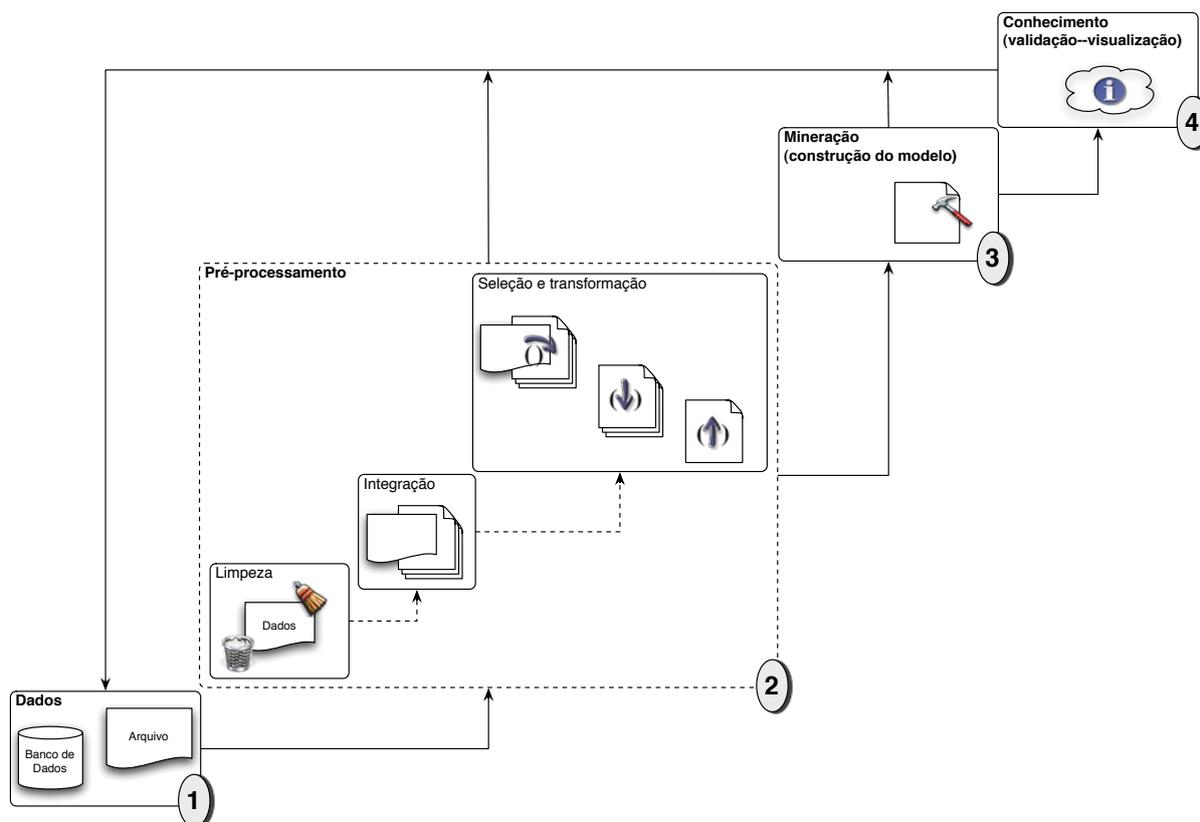


Figura 3.9 O ciclo de descoberta de conhecimento

efetivo desta abordagem. Além disso, existem outros dois problemas: em primeiro lugar, a notação não é a melhor fonte de dados para o caso da análise de interpretações, já que, como já foi apontado no Capítulo 2, o intérprete realiza diversas modificações ao tocar uma obra musical; o outro problema é que, dada a natureza da MPB, não faz muito sentido ter a partitura como fonte de dados, já que não é prática comum utilizar este tipo de notação para o registro de uma composição.

A fase de pré-processamento é tida por muitos como a mais laboriosa de todo o processo [Han e Kamber, 2001, Pyle, 1999, Witten e Frank, 2000]. O objetivo principal nesta etapa é melhorar a qualidade dos dados e, em consequência disto, melhorar também a qualidade dos resultados obtidos com a etapa de construção dos modelos. Para esta etapa, existe uma grande variedade de técnicas. Na *limpeza dos dados*, inconsistências são corrigidas e ruídos são removidos. Na *integração*, dados provenientes de eventuais fontes diferentes são integradas de forma coerente¹⁸. Na *transformação*, diversas operações,

¹⁸Uma tendência atual é armazenar os dados limpos e integrados em um repositório chamado *data warehouse* [Han e Kamber, 2001].

como a normalização de valores, são feitas. Por fim, na *seleção*, os dados resultantes são manipulados (redundâncias são removidas, reduz-se o volume de dados por meio de agregações, etc.) de maneira tal que formem uma entrada apropriada para a fase de mineração. No caso particular da Música, pode-se dizer que, de uma maneira geral, a idéia, nesta etapa, é a obtenção de uma representação que seja suficientemente abstrata para a etapa seguinte. Este processo de abstração passa, por exemplo, pela segmentação ou pelo agrupamento (que servem para determinar as estruturas musicas de mais alto nível como motivos, frases, etc.), análise harmônica, determinação da pulsação (*beat tracking*), etc.

Na etapa de mineração de dados, que tem como objetivo a construção de um modelo capaz de descrever os dados previamente preparados, utilizam-se técnicas e algoritmos desenvolvidos por pesquisadores da área conhecida como Aprendizagem de Máquina [Langley, 1996, Mitchell, 1997]. Um exemplo de modelo que pode ser construído é um que diga se, sob certas condições climáticas, vai chover ou não. Outro exemplo é determinar, a partir do quadro clínico de um paciente, que doença ele possui — as doenças poderiam ser *faringite*, *rinite*, *otite*, etc. Na Computação Musical, um exemplo de modelo que poderia ser construído seria um que indicasse se, dada uma certa seqüência de notas, a próxima nota a ser executada deveria ser tocada com maior, menor ou mesma intensidade que a imediatamente anterior. Nota-se, desta maneira, que os tipos de modelos que podem ser gerados ou aprendidos são os mais variados e dependem essencialmente do tipo de tarefa de aprendizagem que se pretende realizar. Predição, classificação, associação, entre outras, são exemplos de tais tarefas [Mitchell, 1997].

Cabe ainda salientar que, no caso da mineração de dados, onde a aprendizagem é feita a partir de grandes conjuntos de dados, existe também uma certa preocupação com aspectos como eficiência e escalabilidade dos algoritmos e implementações, uma vez que o uso de muitos dados não pode ser um impedimento para que o modelo seja construído em tempo aceitável¹⁹.

Na última etapa, os modelos resultantes são avaliados e, eventualmente, visualizados. Para avaliar a validade de um modelo existem diversas metodologias. Supondo como exemplo a tarefa de classificação, uma possível metodologia de avaliação seria medir a taxa de erro do classificador obtido, sendo tal classificador rejeitado caso a taxa de erro esteja acima de um determinado patamar. Nesta metodologia, utiliza-se, via de regra,

¹⁹A idéia do que é aceitável ou não, obviamente, varia de acordo com o domínio de aplicação.

cerca de dois terços dos dados disponíveis (o *conjunto de treinamento*) para a construção do classificador, enquanto que nos testes utiliza-se o restante dos dados (chamado de *conjunto de teste*). Uma outra metodologia de avaliação é a chamada validação cruzada (*cross-validation*). Nela, estabelece-se, de início, um número n de partições nos dados (*folds*), constrói-se o modelo a partir dos dados de $n - 1$ partições e verifica-se a taxa de erro utilizando-se a enésima parte restante como conjunto de testes. Isto é repetido sucessivamente até que cada partição tenha sido utilizada uma vez como conjunto de testes. A taxa de erro global do modelo é a média de todas as n taxas encontradas durante o processo. Não existe uma prova de que método de avaliação é o melhor, mas diversos experimentos levam a crer que a validação cruzada feita com 10 partições (*tenfold cross-validation*) possui resultados ótimos, o que fez com que esta metodologia se tornasse a padrão em termos práticos [Witten e Frank, 2000].

Apenas recentemente é que a visualização dos dados e resultados da mineração vem recebendo atenção por parte dos pesquisadores. A intenção, aqui, é fornecer ferramentas que possibilitem que o usuário, possivelmente um especialista no domínio do problema e não necessariamente em Aprendizagem de Máquina ou Mineração de Dados²⁰, possa visualizar os dados ou os resultados da etapa de mineração de diferentes maneiras e em diferentes níveis de abstração. No caso da visualização de dados, curvas, histogramas, superfícies, etc, podem ajudar no entendimento e na caracterização dos dados que se encontram em um repositório, podendo auxiliar, inclusive, na formulação dos problemas de aprendizagem. Já na visualização dos resultados, os modelos são ilustrados de maneiras diversas, como, por exemplo, através de gráficos que mostram árvores de decisão, que são as estruturas resultantes da classificação feita a partir do algoritmo ID3 [Quinlan, 1986] e seus derivados, como o C4.5 [Quinlan, 1993] ou o J48 [Witten e Frank, 2000].

3.3.1 Aprendizagem de Máquina e Análise de Interpretações

Quando se pensa em utilizar técnicas e algoritmos da área de Aprendizagem de Máquina para gerar modelos de interpretações musicais, é necessário fazer algumas perguntas sobre a real aplicabilidade deste tipo de abordagem. Tais questões estão muito bem formuladas por Widmer [Widmer, 2001b] e são as seguintes: de fato, existe algo a ser aprendido? Será que a interpretação é algo que não pode ser modelado, intangível, portanto? Como já foi mostrado anteriormente, resultados, tanto na Musicologia [Palmer, 1997, Sloboda, 2000],

²⁰Este é exatamente o caso na presente pesquisa.

quanto na Computação Musical [Widmer, 1994], mostram que, de fato, é possível modelar o fenômeno da expressão musical, isto é, existem tanto semelhanças quanto diferenças sistemáticas que possibilitam que o fenômeno seja aprendido.

Existindo esta evidência, é necessário saber quanto do fenômeno pode ser aprendido e formalizado. Será que apenas uma parte da interpretação pode ser modelada ou ela é completamente previsível? Por se tratar de uma forma de arte, é de se esperar que alguma parte da expressão não seja previsível. Há de levar em consideração, portanto, que apenas uma parte do fenômeno será modelada. Como salienta Widmer, “este requisito favorece algoritmos que, ao invés de tentar cobrir todo o espaço de instâncias, satisfaçam apenas aqueles subespaços onde algo possa ser realmente aprendido e produzam modelos que indiquem claramente quando algo está fora de sua área de conhecimento (*expertise*)” [Widmer, 2001b].

Outra questão importante é saber que conceitos devem ser aprendidos ou modelados (*target concepts*). Esta pergunta deve, ainda, levar em consideração resultados da pesquisa musicológica que indicam que a interpretação não é um algo puramente expressivo, mas também um código estrutural [Clarke, 1988, Palmer, 1997] e que a estrutura da música é hierárquica [Lerdahl e Jackendorff, 1983]. Desta forma, além de saber o que exatamente deve ser modelado ou aprendido, é necessário saber em que nível da estrutura essa modelagem deve acontecer. Widmer aponta ainda que o contexto musical (“janela” de observação do fenômeno) e representação das estruturas da música são diretamente afetadas por esta problemática.

Por fim, existe o problema da avaliação e validação dos resultados. Como validar os resultados quando não existe uma única interpretação correta e sim a idéia de que existem diversas interpretações aceitáveis? Será que, por exemplo, os métodos de avaliação usuais na área de Aprendizagem de Máquina (medir a taxa de erro, validação cruzada, etc.) são suficientes para o correto julgamento dos modelos obtidos? Ou será que considerações estilísticas e essencialmente musicais também devem ser levadas em consideração?

Como foi adiantado na Seção 2.5, a abordagem seguida nesta pesquisa foi semelhante à abordagem utilizada na Áustria. A escolha por abordagem nos moldes da abordagem austríaca não implica, entretanto, em uma reprodução dos métodos lá utilizados. Existem diferenças nos intuítos desta pesquisa que fazem com que o problema mude de maneira significativa. A primeira é que, enquanto lá os parâmetros analisados são o andamento, a articulação e a dinâmica na Música Clássica para piano, aqui, a análise será do acompanhamento rítmico ao violão na Bossa Nova. Outra diferença é que, aqui, se

pretende construir modelos no nível dos padrões rítmicos (lá, os modelos refletem o nível da nota). Esta mudança no nível estrutural implica em uma mudança na representação e os atributos utilizados nesta representação devem refletir esta mudança, o que pode não ser trivial.

Além disso, algumas considerações devem ser feitas. A pesquisa conduzida na Áustria indica que “os modelos gerados para um determinado pianista tendem a se deteriorar se aplicados para a predição da interpretação de outros pianistas”. Será que o mesmo é válido para o caso do ritmo? Isto é, será que existem dimensões da música onde se pode construir uma teoria geral, válida para todos os casos, de como a interpretação é feita?

Chegou-se a conclusão, ainda, que não é interessante construir modelos completos da interpretação musical no nível da nota, sendo muito mais frutífero a construção de modelos parciais [Widmer, 2002b]. Em um dos modelos, que diferenciava entre *accelerando* e *ritardando*, obteve-se um total de 3.037 regras, com 58,09% de acurácia. Um modelo longe do ideal, principalmente quando se pensa que o objetivo é a análise. Como este resultado vem de uma pesquisa que constrói modelos para dimensões da dinâmica, do andamento e da articulação, será que, no caso da dimensão rítmica, os modelos que a descrevem devem também ser parciais ou será que para o ritmo é possível construir modelos completos no nível da nota e, ainda assim, concisos? E no caso de modelos no nível dos padrões: será que eles podem ser completos? No fundo, a pergunta é a seguinte: o que faz com que modelos completos da interpretação não possam ser construídos? A dimensão ou dimensões utilizadas, ou o nível da estrutura musical?

ANÁLISE AUTOMÁTICA DO ACOMPANHAMENTO MUSICAL

No decorrer do documento, mostrou-se, de início, que a interpretação musical é um fenômeno complexo, sujeito à influência de diversos elementos, incluindo elementos de caráter subjetivo, mas que, apesar disso, apresenta aspectos comuns, mesmo quando diferentes interpretações são examinadas (Seção 2.1). Salientou-se, ainda, que a interpretação, na pesquisa tradicional realizada por musicólogos, é, em geral, estudada a partir de alguma dimensão da música e que as pesquisas sobre o assunto tendem a se concentrar na Música Clássica composta para piano, o que exclui uma grande parte do repertório musical (Seção 2.2). Além disso, mostrou-se que este tipo de análise se beneficiaria bastante caso fosse automatizada, ao menos em parte, principalmente por possibilitar o uso de uma maior quantidade de dados (Seção 2.3).

Mostrou-se, ainda, que existem pesquisadores no âmbito da Computação Musical que estão preocupados com a automação da análise de interpretações (Seção 2.4). De maneira geral, estes pesquisadores têm como idéia norteadora tentar descrever diferentes aspectos da interpretação através da construção de modelos computacionais. Modelos para a indução da pulsação, para a quantização, modelos capazes de distinguir diferentes intérpretes, entre outros, são exemplos de resultados já obtidos por estes pesquisadores. Ainda como no caso da pesquisa tradicional, a Música Clássica composta para piano se constitui, via de regra, no foco principal de atenção.

Mostrou-se também, no capítulo introdutório, que, no âmbito da Bossa Nova, a despeito do conhecimento já obtido em diversas pesquisas, existe um grande número de perguntas e problemas em aberto, em particular, um conjunto de perguntas sobre como é construído o acompanhamento rítmico na Bossa Nova (Seção 1.1, página 5). Estabeleceu-se, a partir daí, o escopo ptou-se aqui por uma mudança, que resultou nas seguintes características para o escopo: o estilo musical é a Bossa Nova, o instrumento é o violão, a dimensão de análise é o ritmo, a notação, caso exista, é a cifra e a função do músico é a de acompanhador.

Discutiu-se, ainda, a partir de que nível da estrutura musical (nota, frase, etc.) as investigações deveriam ser feitas (Seções 1.1 e 1.3, e início do Capítulo 3). Desta discussão, que utilizou como base as perguntas previamente formuladas e resultados de pesquisa sobre a construção do acompanhamento, determinou-se que o padrão rítmico deveria ser o ponto a partir do qual a automação da análise deveria ser implementada. A análise, desta forma, deve ser orientada a padrões, isto é, assume-se que o padrão musical, ou *batida* no jargão dos músicos, funciona como bloco fundamental na construção do acompanhamento e que não faz sentido, por exemplo, analisar relações entre notas individuais.

Automatizar a análise do acompanhamento tendo como nível estrutural mínimo o padrão rítmico implica, todavia, em encontrar tais padrões, já que, independentemente do tipo de representação adotada (partitura, cifra, arquivos MIDI ou áudio), eles não se encontram explicitamente indicados na música. Para tanto é necessário utilizar ou desenvolver técnicas e algoritmos capazes de encontrar os padrões musicais, tendo em mente, ainda, que a Música se constitui em um domínio bastante peculiar e que requer que certas características, como, por exemplo, a capacidade de encontrar recorrências inexatas, sejam satisfeitas para que padrões musicais interessantes possam ser extraídos. Várias destas técnicas e algoritmos para a extração de padrões, assim como diferentes aspectos relativos à apropriada representação de dados musicais para tal tarefa, foram apresentados e discutidos (Seção 3.2). Desta discussão e levando-se em conta um dos requisitos necessários para a extração na presente pesquisa (a flexibilidade), foram apontados algoritmos candidatos à extração de padrões rítmicos (Seção 3.2.5).

Apresentaram-se e discutiram-se, também, alguns aspectos das disciplinas relacionadas à descoberta automática de conhecimento diretamente ligados aos problemas abordados na presente pesquisa (Seção 3.3). Neste capítulo, destacou-se, ainda, um processo para a descoberta do conhecimento, amplamente adotado pela comunidade, que descreve, em linhas gerais, um conjunto de passos e procedimentos que devem ser adotados.

Considerando o exposto, no presente Capítulo descreve-se como o problema central desta pesquisa, a descoberta automática de conhecimento a partir de acompanhamento rítmico ao violão na Bossa Nova, foi abordado. Tal abordagem passou, em primeiro lugar, pela adaptação do ciclo de descoberta do conhecimento para as especificidades do problema aqui tratado. Em seguida, os problemas referentes a cada um dos passos resultantes desta adaptação foram estudados, o que resultou no estabelecimento de um conjunto de técnicas, novas e adaptadas, e na criação de uma representação da música adequada para a extração de padrões. Esta representação foi, em seguida, utilizada em

uma série de experimentos que tinham como foco a extração de padrões, o que serviu tanto para validar a representação criada quanto como estudo de adequação dos algoritmos anteriormente selecionados para a tarefa de extração de padrões rítmicos. Além disso, foi implementado um protótipo para apoio às atividades aqui desenvolvidas, protótipo este que conta, inclusive, com um módulo para a visualização dos resultados.

O restante do capítulo está organizado da seguinte maneira: na Seção 4.1, descrevem-se as adaptações realizadas no ciclo de descoberta do conhecimento. Em seguida, nas Seções 4.2, 4.3 e 4.4, são apresentadas as etapas iniciais do processo adaptado. Nesta apresentação, detalham-se o processo de aquisição dos dados, as técnicas empregadas em sua preparação e a representação da música adotada. A apresentação dos experimentos, assim como a descrição da ferramenta e aspectos de sua implementação são descritos no Capítulo 5.

4.1 UM PROCESSO PARA A ANÁLISE DO ACOMPANHAMENTO

Tanto na pesquisa em Música quanto na pesquisa em Ciência da Computação, os processos metodológicos que levam à descoberta de conhecimento são elementos fundamentais. Como exposto, no contexto da presente pesquisa, o *ciclo de descoberta do conhecimento* (Figura 3.9, p. 60) é a metodologia que, ainda que em linhas bastante genéricas, descreve os passos necessários para que uma efetiva descoberta de conhecimento seja realizada. É natural, portanto, utilizá-lo como base para a definição de um processo para a análise do acompanhamento.

De início, o processo de análise poderia ser formado por quatro fases distintas, a saber: (1) a aquisição dos dados, que poderia ser feita, por exemplo, através do uso de gravações de obras musicais; (2) o pré-processamento, que consistiria na preparação dos dados e na derivação de parâmetros relevantes para a análise; (3) a descoberta de conhecimento em si, onde, para tanto, algoritmos apropriados seriam utilizados; e, por fim, (4) a validação e visualização dos resultados. Cabe aqui, entretanto, a seguinte pergunta: seria esta configuração inicial suficiente para a análise do acompanhamento aqui discutida? Ou seriam necessárias adaptações específicas ao problema para que conhecimento seja efetivamente extraído dos dados musicais?

De imediato, é possível apontar que a extração de padrões não consta nesta adaptação inicial do ciclo. Decerto ela deve se “encaixar” em algum ponto do processo, mas onde este “encaixe” deve acontecer? Ela deve integrar a segunda etapa do processo (etapa de

pré-processamento), ser parte da terceira ou, ainda, formar uma etapa própria?

Para definir exatamente em que ponto a adaptação do ciclo deve ocorrer, identificaram-se o papel da extração de padrões na análise do acompanhamento e o tipo de resultado que a ela produz. Dada a natureza do problema aqui estudado, é patente que a fase de extração de padrões deve acontecer antes da etapa de mineração ou construção de modelos (etapa 3, no ciclo adaptado), afinal é a extração que identificará o material básico a partir do qual conhecimento sobre o acompanhamento rítmico será construído. Neste caso, o papel da extração é o de alimentar com dados a etapa de mineração.

Mas, por outro lado, o resultado de um algoritmo de extração de padrões não deixa de ser uma espécie de conhecimento sobre a obra analisada, já que seu resultado identifica um tipo de *estrutura musical*, ou seja, um determinado subconjunto de notas que possui algum tipo de semântica associada. Além disso, na extração estabelece-se uma *relação de similaridade* entre padrões, isto é, além da identificação de estruturas musicais, os algoritmos são capazes de determinar graus de semelhança entre tais estruturas. Alguns algoritmos, inclusive, traduzem de maneira estruturada tais relações de similaridades entre padrões¹, o que pode servir como elemento para uma melhor compreensão da forma de uma obra musical. Pode-se, ainda, utilizar o resultado obtido com a execução dos algoritmos de extração como substrato para uma nova etapa de extração, e este segundo resultado como entrada para uma terceira fase de extração, e assim, recursivamente, até que se chegue a uma espécie de estrutura fundamental, dando, talvez, um quê de análise schenkeriana [Cook, 1994] à investigação da obra²!

Fica claro, então, que a extração de padrões ocupa um lugar fundamental no processo de descoberta de conhecimento, ora integrando a etapa mineração, ora servindo como provedor de dados para esta etapa, sendo parte, portanto, do pré-processamento de dados. Considerando o exposto, a adaptação do ciclo para o problema desta pesquisa passa a possuir os seguintes passos:

- 1) aquisição dos dados;
- 2) pré-processamento ou preparação dos dados;
- 3) extração de padrões;

¹Os grafos de equípólicia gerados pelo algoritmo FLEXPAT são bons exemplos disto.

²Na análise schenkeriana, procedimento analítico idealizado por Heinrich Schenker [Cook, 1994], tem-se por princípio a idéia de que sucessivas reduções da estrutura musical podem levar a uma mesma estrutura fundamental ou *Ursatz*, como denominada pelo autor.

- 4) construção de modelos; e
- 5) validação e visualização dos resultados.

É importante ressaltar, por fim, a importância da representação simbólica da música. Como previamente mencionado, a etapa de pré-processamento é reconhecidamente uma das mais trabalhosas do ciclo³. No caso específico da análise de interpretações, a adequada preparação dos dados passa, muitas vezes, pela solução de problemas bastante complexos, como por exemplo a indução automática da pulsação ou a quantização. Mesmo nestes casos, onde existem soluções automáticas, algumas vezes é necessária a intervenção humana para corrigir pequenas imperfeições⁴. Fica claro, desta forma, que o resultado de todo este processo não deve ser perdido, sendo necessário a existência de uma representação da música que preserve os resultados do pré-processamento.

O ciclo adaptado, ressaltando, também, a representação simbólica como um resultado importante no processo, encontra-se ilustrado na Figura 4.1.

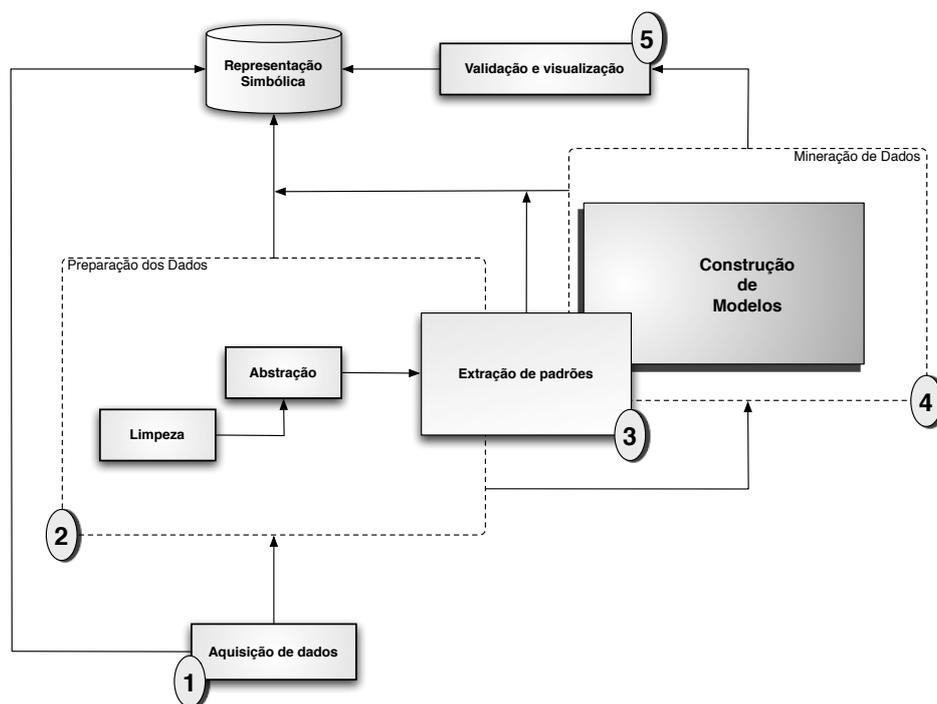


Figura 4.1 Processo de descoberta de conhecimento sobre o acompanhamento rítmico

³ Vide p. 60 para maiores informações.

⁴Para se ter um exemplo prático, durante a indução da pulsação realizada no *corpus* utilizado nesta pesquisa (16 gravações) foram necessárias cerca de 3 horas por gravação na correção de pequenos problemas, como, por exemplo, pulsações mal posicionadas.

No presente documento, os passos 1, 2 e 3 são discutidos na Seção 4.2, enquanto que os passos subseqüentes, em especial a extração de padrões, são discutidos no Capítulo 5.

4.2 AQUISIÇÃO DE DADOS

Na etapa de aquisição de dados foi utilizado um violão acústico tradicional com um captador adaptado⁵. À saída deste captador foi conectado o conversor MIDI GR-33, fabricado pela Roland⁶, que é responsável por transformar o sinal do captador em eventos MIDI. Estes eventos MIDI foram, então, transmitidos a um computador para armazenamento. Tal arranjo permitiu que o violonista utilizasse um instrumento acústico ao qual está acostumado, ao invés de um controlador MIDI em formato de violão, como, por exemplo, o controlador ZTAR⁷, o que teoricamente evita distorções na interpretação devido a diferenças no instrumento. Uma outra vantagem é que foi possível obter, para uma mesma interpretação, tanto dados discretos (arquivos MIDI), quanto dados contínuos (arquivos áudio).

O processo de gravação foi realizado da seguinte maneira: dois violonistas com vários anos de experiência, um deles, inclusive, estudioso da Bossa Nova, foram convidados a um estúdio, onde gravaram diversas canções pertencentes ao seu repertório corrente. Durante a gravação, a única restrição exigida foi que os músicos tocassem as canções seguindo cifras fornecidas previamente⁸. Permitiu-se apenas mudanças nas superestruturas harmônicas dos acordes (nonas, décimas primeiras, etc.). Isso foi feito para garantir que as mesmas canções tocadas por diferentes intérpretes possuísem a mesma harmonização, o que, em tese, faz com que eventuais escolhas por padrões diferentes motivadas pela dimensão harmônica sofram exatamente a mesma influência⁹. No total, foram realizadas 16 gravações, o que corresponde a cerca de meia hora de música. A distribuição das gravações encontra-se descrita no Quadro 4.1.

⁵O captador adaptado no violão foi o *RMC Poly-Drive II* da RMC Pickup Co. — <http://www.rmcpickup.com/polydriveii.html>.

⁶Trata-se, na verdade, de um sintetizador MIDI. Suas funções de síntese, entretanto, não foram utilizadas na pesquisa, sendo necessário apenas o uso de suas capacidades de conversão do sinal de áudio para MIDI (*pitch to MIDI conversion*). Mais informações sobre o sintetizador podem ser obtidas em <http://www.roland.com/products/en/GR-33/index.html>.

⁷<http://www.starrlabs.com/>

⁸As cifras foram retiradas dos *songbooks* sobre a Bossa Nova de Almir Chediak [Chediak, 1990].

⁹No fundo, a intenção é manter sob controle o número de variáveis da interpretação.

	Intérprete 1	Intérprete 2
Bim Bom		×
Eu Sei Que Vou Te Amar	×	
O Barquinho		×
Samba de uma Nota Só	×	
Desafinado	×	
Garota de Ipanema	×	×
Só Danço Samba		×
A Felicidade	×	
Insensatez	×	×
Wave	×	×
Corcovado	×	
Tarde em Itapoã	×	
Chega de Saudade	×	

Quadro 4.1: Lista de obras gravadas

4.3 PREPARAÇÃO DOS DADOS

Uma vez realizada a aquisição de dados, passou-se à segunda etapa do processo: a preparação dos dados. Nesta etapa, duas atividades distintas foram realizadas: a limpeza dos dados e a identificação e derivação de abstrações da estrutura musical pertinentes ao problema em questão.

Antes de descrever como a limpeza dos dados foi realizada, é necessária uma pequena explicação sobre a sua necessidade. Ao se trabalhar com os dados coletados mais de perto, observou-se o que parecia ser um certo nível de ruído, isto é, notou-se a presença de eventos que não aconteceram durante a execução ou ainda algum tipo de modificação indevida (modificação na altura da nota, ou redução de sua duração, por exemplo)¹⁰.

De maneira geral, tais ruídos puderam ser classificados em três tipos básicos: notas excessivamente curtas, notas de intensidade muito baixa e notas deslocadas, isto é, notas cuja altura (frequência) foi, por algum motivo, modificada pelo equipamento¹¹.

Com base nesta classificação, optou-se, para a realização da limpeza dos dados, pela

¹⁰O leitor pode se perguntar se tais ruídos, na verdade, poderiam ser erros dos intérpretes. Tal possibilidade foi cogitada de início, mas logo eliminada após a audição dos arquivos áudio coletados.

¹¹Não foi possível, infelizmente, encontrar literatura científica que descrevesse estes problemas em detalhes ou mesmo que realizasse uma comparação entre os diferentes violões e captadores existentes no mercado (algo semelhante a um *benchmark*). Todavia, a busca por informações sobre o problema, principalmente em fóruns de usuários de violões MIDI na *internet*, como o *MIDI Guitar* (<http://launch.groups.yahoo.com/group/midiguitar/>), revelou que este tipo de problema é bastante comum em equipamentos como os aqui utilizados.

construção de um filtro, capaz de processar o arquivo MIDI de maneira tal a eliminar estas classes de ruídos identificadas. Para tanto, utilizou-se, como ferramenta de auxílio, o seqüenciador MIDI *Cubase SX*¹². Nele, foi possível construir um conjunto de regras que, de acordo com parâmetros de entrada fornecidos pelo usuário, pode eliminar os ruídos dos dois primeiros tipos identificados anteriormente (notas curtas e notas de baixa intensidade). Um exemplo de regra utilizada encontra-se ilustrada na Figura 4.2. Nela, notas com intensidade, ou velocidade (na nomenclatura MIDI), menor que 20 unidades são apagadas.



Figura 4.2 Exemplo de regra de filtragem

Já para o terceiro tipo de ruído, o caso onde as notas são transpostas, não foi possível construir um conjunto de regras com o uso da ferramenta e a correção foi realizada manualmente. A Figura 4.3 exemplifica o procedimento realizado.

A limpeza dos dados, portanto, foi realizada pelo autor através de um processo semi-automático, onde parte do ruído é extraída com o uso de regras e outra parte precisa ser consertada à mão. Mesmo no caso das regras, os valores para cada parâmetro utilizado (duração e intensidade das notas) deve ser identificado pelo usuário e estabelecido de maneira específica para cada obra a ser analisada. Além disso, algumas notas que foram de fato tocadas pelos intérpretes foram capturadas com suas durações reduzidas, podendo, desta forma, ser confundidas com as notas curtas não tocadas, o que também requer

¹²Cubase SX é um produto da Steinberg — <http://www.steinberg.net>.

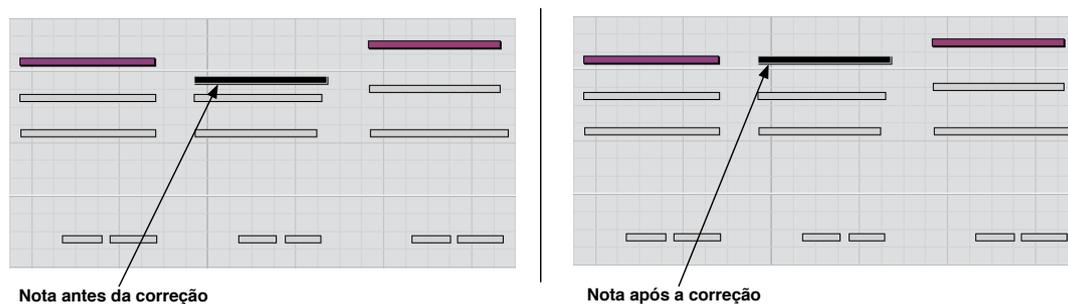


Figura 4.3 Exemplo de correção manual

um exame individual. Vale salientar, ainda, que um quarto tipo de ruído também foi observado: nele, notas executadas pelo intérprete não são captadas pelos equipamentos e, conseqüentemente, não são “gravadas” no arquivo MIDI resultante. Tal ruído, todavia, não foi tratado nesta pesquisa.

4.4 REPRESENTAÇÃO DOS DADOS

Uma vez realizada a limpeza dos dados, passou-se ao processo de derivação de abstrações da música. Neste processo, levou-se em consideração, em primeiro lugar, que a dimensão da música a ser analisada é o ritmo. Isto implica, *grosso modo*, em analisar os *ataques* e as *durações* das notas. Em segundo lugar, considerou-se também que as abstrações resultantes devem favorecer a derivação do nível da estrutura musical onde a análise será feita, isto é, o nível dos padrões. Isto significa que a representação adotada deve permitir que os padrões sejam encontrados com os algoritmos selecionados a partir da análise feita no Capítulo 3.

Isso posto, o processo de abstração se deu em três passos: de início, encontrou-se a pulsação de cada uma das obras gravadas, assim como estabeleceu-se um procedimento de quantização deste resultado, e, em seguida, tratou-se da derivação do dedilhado da mão direita. Tudo isto, junto com conhecimento sobre o acompanhamento na Bossa Nova, foi utilizado como elemento para a derivação de abstrações e da representação a ser utilizada nas análises. Explicações sobre cada um dos passos, assim como a justificativa para a escolha exata do dedilhado da mão direita como forma de abstração, serão explicitadas nas seções que se seguem. Na Seção 4.4.1, descreve-se como foi realizada a determinação da pulsação, assim como a quantização das obras. Já na Seção 4.4.2, descreve-se o algoritmo de indução do dedilhado. Melhoras e adequações nas abstrações da música obtidas e a representação estendida da música original são, por fim, descritas na Seção 4.4.3.

4.4.1 Determinação da Pulsação e Quantização

Como já foi frisado anteriormente, durante a interpretação de uma obra musical, diversos parâmetros são variados pelo intérprete. Em particular, para o caso do ritmo, pode-se citar o *andamento* como um dos parâmetros passíveis de variação. A consequência imediata de flutuações neste parâmetro é que torna-se praticamente impossível estabelecer *a priori* uma associação entre a duração efetiva de uma nota — em segundos ou milissegundos, por exemplo — e sua duração teórica — uma colcheia, uma semínima. Supondo-se, a título de exemplo, que, em um determinado momento da música, uma nota com duração teórica de colcheia possua duração efetiva de 1 segundo, dado o exposto, nada garante que em outro momento esta duração teórica terá a mesma duração em segundos. O intérprete pode perfeitamente variar sua duração, executando a colcheia em 0,9 ou 1,1 segundos, por exemplo, ou dependendo de outros fatores, como a articulação ou marcas expressivas, variar ainda mais a duração efetiva da nota.

Como, então, traçar uma correspondência entre duração efetiva e duração teórica de uma nota? A base para a construção deste mapeamento é fornecida pela própria música, ou melhor posto, a construção se dá com base em um fenômeno da percepção humana sobre a música, a saber: a *pulsação*. Sua importância para o problema aqui tratado se dá porque à ela é possível associar uma duração teórica. Assim, se a pulsação corresponder a uma semínima, por exemplo, isto significa dizer que o espaço entre dois pulsos quaisquer, independentemente de diferenças em sua duração efetiva, corresponde a uma semínima. Os pulsos funcionam, desta maneira, como marcos que delimitam espaços que podem ser mapeados em durações teóricas de mesmo valor.

Para a determinação da pulsação, foi utilizado o aplicativo *BeatRoot*, já citado no Capítulo 2, página 28. Trata-se de uma ferramenta de indução da pulsação que funciona de maneira interativa: o sistema, em primeiro lugar, induz as pulsações da obra em análise e o usuário, em seguida, pode ouvir os resultados, corrigindo eventuais imprecisões, como, por exemplo, pulsos que foram induzidos erroneamente.

Uma vez determinadas estas referências temporais, passou-se ao processo de quantização, isto é, à determinação de subdivisões ou frações da pulsação, que, de maneira semelhante a ela, representam durações teóricas de mesmo valor. O algoritmo idealizado é extremamente simples, *naïve* até: para cada par de pulsos consecutivos, P_i e P_{i+1} , aplica-se uma razão numérica que divide o espaço entre eles em partes de igual duração. Assim, considerando uma pulsação no nível da semínima, caso um fator igual a 2 seja

aplicado, as sub-partes resultantes corresponderão à duração de uma colcheia. Caso o fator seja igual a 4, o resultado será formado por pedaços que correspondem à semicolcheia, e assim por diante. No caso da música ser anacrúsica, isto é, se alguma nota for tocada antes do primeiro pulso, a referência para a quantização entre o início da música e o primeiro pulso é o intervalo entre o primeiro e o segundo pulso¹³. A Figura 4.4 ilustra o processo.

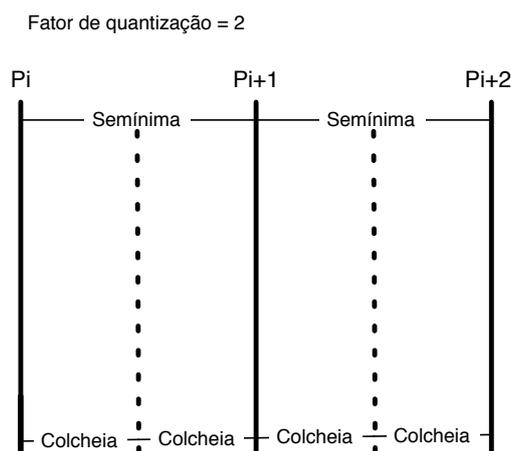


Figura 4.4 Exemplo de quantização

O resultado desta primeira etapa na preparação dos dados é um conjunto de arquivos MIDI, um para cada canção, com as pulsações estabelecidas no nível da colcheia e sobre os quais se pode aplicar o algoritmo de quantização descrito.

4.4.2 Dedilhado da Mão Direita e Abstrações Derivadas

O ato de tocar violão pode ser grosseiramente reduzido a duas tarefas básicas: escolher em que locais—cordas e casas—as notas serão tocadas e tanger as cordas adequadamente, executando as notas desejadas. Com a mão esquerda¹⁴, o violonista escolhe as cordas e as casas de acordo com as notas que devem ser tocadas, pressionando as cordas contra o braço do instrumento. Já com a mão direita, o músico utiliza uma determinada combinação de dedos de maneira tal a pinçar as cordas onde as notas deverão ser produzidas. À combinação de dedos utilizada em cada uma destas operações dá-se o nome de *dedilhado*.

¹³Esta estratégia foi adotada assumindo como hipótese que o intervalo entre o pulso zero—inexistente e apenas imaginado pelo intérprete—e o primeiro pulso tende ser o mesmo que o existente entre o primeiro e o segundo pulso.

¹⁴Assume-se, para todos os efeitos, que o violonista é destro.

Dado este mecanismo de produção, pode-se dizer que a mão esquerda determina as notas que serão executadas, enquanto que a mão direita determina *quando* e como estas notas serão executadas. É, portanto, a mão direita a responsável pela determinação do ritmo. Cabe aqui, porém, o seguinte questionamento: que tipo de adição a determinação do dedilhado traz para o estudo do ritmo? Não será o entendimento das relações entre as durações da nota, resultante do processo anteriormente descrito de derivação da pulsação e quantização, muito mais relevante para este estudo? De fato, a duração e, principalmente, o momento do ataque da nota são elementos essenciais para o estudo do ritmo. Todavia, o dedilhado serve a dois propósitos: em primeiro lugar, o dedo serve como um descritor extra para a nota, como um elemento adicional, correlato ao ritmo, mas que poderá ser utilizado como parâmetro na construção de modelos do ritmo. Mas o motivo principal é que é possível descrever os padrões como seqüências de bordões ou *baixos* (notas tocadas com o polegar) e acordes ou *puxadas* (notas tocadas conjuntamente pelos outros dedos), como, por exemplo, faz Garcia [Garcia, 1999]. E, como será exposto na Seção 4.4.3, é justamente a partir da determinação do dedilhado que estas abstrações podem ser construídas.

Apesar de algoritmos para a derivação do dedilhado da mão esquerda já terem sido criados [Sayegh, 1991, Cabral et al., 2001], desconhecem-se algoritmos para o estabelecimento do dedilhado da mão direita. O algoritmo de indução do dedilhado da mão direita aqui proposto segue uma abordagem semelhante à seguida por Sayeg [Sayegh, 1991], a saber: pensa-se no dedilhado como uma seqüência de estados e que a cada mudança de estado corresponde uma transição. Além disso, a cada transição está associado um custo, que representa o esforço necessário para realizar a mudança de estado, isto é, para mudar o dedilhado. O custo, por sua vez, é construído baseado na combinação da dificuldade e da uniformidade necessárias para realizar a mudança de estado. Desta forma, cada dedilhado possível¹⁵ é representado como um nó em um grafo, enquanto que os custos estão associados às arestas que ligam os nós. O problema de encontrar o dedilhado pode então ser reduzido à descoberta de um caminho que minimize o custo global das mudanças de estado.

Antes da descrição do algoritmo, denominado RHFA, é necessário definir alguns con-

¹⁵Teoricamente o número de dedilhados possíveis para a mão esquerda, no pior caso que é a execução de 5 notas simultâneas, é de $120n$, onde n representa o número de notas da música. Todavia, raramente o dedo mínimo da mão direita é utilizado, o que reduz significativamente a quantidade de dedilhados possíveis para $24n$.

ceitos. A representação adotada para os dedos é a seguinte: T para o polegar, F para o indicador, M para o dedo médio, R para o anular, e L para o dedo mínimo¹⁶. Outros conceitos são definidos a seguir.

Definição 4.1. *Sejam N e \mathcal{S} duas n -uplas que representam uma seqüência de n notas (n_1, \dots, n_n) e o conjunto de cordas do violão utilizadas para tocar cada nota em N , respectivamente.*

Definição 4.2. *Posição (p) é um arranjo dos dedos da mão direita (posição da mão direita) com relação à corda ou cordas que eles irão tocar.*

Cada posição p é representada como uma sêxtupla, $\langle x_1, x_2, x_3, x_4, x_5, x_6 \rangle$, onde cada posição da tupla identifica uma corda, sendo a primeira posição relativa à corda $E2$ (a corda mais grave do instrumento), a segunda à corda $A2$, e assim por diante. Cada x_i , por sua vez, indica o dedo utilizado para tangir a corda. Caso a corda não seja utilizada o símbolo $-$ é utilizado. Assim, para, por exemplo, tocar o acorde de dó maior mostrado na Figura 4.5, utiliza-se a posição $p = \langle -, T, F, M, R, - \rangle$ (na ilustração, os dedos da mão direita encontram-se ao lado das cordas pinçadas).

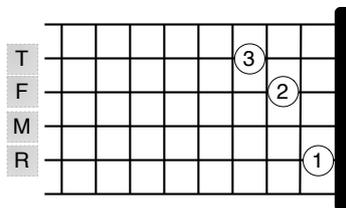


Figura 4.5 Acorde de dó maior

Definição 4.3. *Conjunto de Posições (\mathcal{P}) é um conjunto onde cada elemento é uma posição p .*

Definição 4.4. *Custo de Transformação ($\tau_{(p_i, p_{i+1})}$) é uma função que estabelece o custo de mudança da posição p_i para a posição p_{i+1} , onde $p_i, p_{i+1} \in \mathcal{P}$.*

Os custos de transformação foram definidos levando-se em consideração os seguintes critérios: facilidade de execução, flexibilidade e agilidade. Por facilidade de execução entende-se que as transições devem tentar preservar ao máximo a posição dos dedos

¹⁶Ainda que o dedo mínimo não seja utilizado na prática, ele será mantido na formalização do algoritmo por uma questão de completude.

em relação às cordas (mudanças na posição dos dedos são, desta maneira, penalizadas com custos mais altos). Por flexibilidade, entende-se que transições que levem em consideração posições futuras, preparando-as de alguma maneira, são favorecidas, possuindo custos menores. Por agilidade, entende-se que transições que são mais adequadas para passagens solo (seqüências de notas individuais) possuem custos menores. É importante frisar que estes custos somados ao conjunto de posições \mathcal{P} servem como mecanismos para modelar tanto restrições anatômicas da mão, quanto restrições culturais (que conjunto de dedilhados é o mais adequado para este ou aquele instrumentista ou estilo musical, por exemplo).

Definição 4.5. *Custo de Aplicação* ($\alpha(p, s)$) é uma função que computa o custo de aplicação da posição $p \in \mathcal{P}$ ao elemento $s \in S$.

Definição 4.6. *Seqüência das Melhores Posições* ($Best_p$) é uma seqüência formada por n melhores posições, $(best_{p_1}, \dots, best_{p_n})$. Este é justamente o resultado do algoritmo RHFA.

O algoritmo que computa $Best_p$ é descrito pelo seguinte pseudocódigo:

Algoritmo RhFA

- 1 **para cada** $s_i \in S$ e $best_{p_i} \in Best_p$
- 2 **para cada** $p_j \in \mathcal{P}$
- 3 **faça** $\alpha(p_j, s_i) = \min_{p \in \mathcal{P}} [\alpha(p_j, s_{i-1}) + \tau(p_j, p)]$
- 4 $best_{p_i} = \operatorname{argmin}_{p \in \mathcal{P}} [\alpha(p, s_i)]$

Em linhas gerais, o algoritmo calcula o menor custo de aplicação possível de cada elemento de \mathcal{P} a todos os elementos de S , levando em conta todos os custos de aplicações anteriores e o custo de transição entre posições (linha 3). Em seguida, é computada a melhor seqüência de transições (linha 4 no pseudocódigo). Utilizou-se, para tanto, o algoritmo de Viterbi [Viterbi, 1967], que, implementado com o uso de técnicas de programação dinâmica, possui complexidade de $O(nm)$, onde n é o número de elementos em \mathcal{P} e m é o número de notas na música.

Para avaliar o algoritmo, utilizaram-se os dados coletados anteriormente. Antes de relatar os resultados obtidos, porém, é necessário descrever as outras entradas do algoritmo, a saber: o conjunto \mathcal{P} de posições da mão, e as funções τ e α .

No estabelecimento de \mathcal{P} , o objetivo principal era dividir um conjunto de posições que modelasse adequadamente o acompanhamento na Bossa Nova. Para tanto, observou-se

como os dedilhados são “montados” pelos músicos. Desta observação e com o devido auxílio de violonistas, constatou-se, em primeiro lugar, que o baixo é sempre executado com o polegar (T). Além disso, verificou-se também que a puxada é executada com o restante dos dedos, usualmente na seguinte ordem: F para a nota imediatamente acima do baixo, M para a nota imediatamente acima da nota tocada com o dedo indicador e R para nota acima da nota executada com o dedo médio. De maneira a aumentar a flexibilidade, satisfazendo eventuais casos particulares, introduziram-se dois outros símbolos: 1, que representa a possibilidade de execução da corda em questão com o dedo indicador ou com o dedo médio (F ou M); e 2, que representa a possibilidade de pinçar a corda com os dedos médio ou anelar (M ou R). Assim, na posição $p = \langle -, -, T, -, F, 2 \rangle$, tanto o dedo médio, quanto o dedo anelar podem ser utilizados para tocar a corda mais aguda no instrumento. O conjunto resultante destas considerações é formado por 20 posições diferentes, e pode ser visto na Figura 4.6.

$$\begin{aligned} \mathcal{P} = \{ & \langle T, -, F, M, R, - \rangle, \langle T, -, -, F, M, R \rangle, \langle -, T, F, M, R, - \rangle, \langle -, T, -, F, M, R \rangle, \\ & \langle -, -, T, F, M, R \rangle, \langle T, -, -, 1, R, - \rangle, \langle -, T, -, 1, R, - \rangle, \langle -, -, T, 1, R, - \rangle, \\ & \langle T, -, -, -, 1, R \rangle, \langle -, T, -, -, 1, R \rangle, \langle -, -, T, -, 1, R \rangle, \langle T, -, -, F, 2, - \rangle, \\ & \langle -, T, -, F, 2, - \rangle, \langle -, -, T, F, 2, - \rangle, \langle T, -, -, -, -, 1 \rangle, \langle -, T, -, -, -, 1 \rangle, \\ & \langle -, -, T, -, -, 1 \rangle, \langle T, -, -, -, F, 2 \rangle, \langle -, T, -, -, F, 2 \rangle, \langle -, -, T, -, F, 2 \rangle \} \end{aligned}$$

Figura 4.6 Conjunto \mathcal{P} utilizado na avaliação do algoritmo.

No que diz respeito às funções τ e α , as outras entradas do algoritmo, pode-se dizer que: a função que calcula o custo de transformação (τ) foi pensada de maneira a evitar mudanças de posição, punindo com um custo cada vez maior quaisquer mudanças de posição entre estados. Assim, quanto maior o número de mudanças entre estados consecutivos, maior o custo da transição. Já para a função que calcula o custo de aplicação (α), adotou-se uma estratégia bem simples: ela pune com o custo mais alto possível (custo infinito, ∞) posições que não utilizam a corda na qual a nota é produzida, e com o menor custo possível (custo zero) o caso contrário.

Na Figura 4.7, é possível visualizar parte dos resultados obtidos com o algoritmo. Ela ilustra o dedilhado obtido para o início da canção *Insensatez*, tocada pelo intérprete 2 (a corda mais aguda do instrumento é representada pela linha superior).

Já a Figura 4.8 ilustra o resultado do dedilhado da canção *Eu Sei Que Vou Te Amar*,

```

E -----
B R-----R-----R---
G M-----M-----M---
D F-----F-----F--- [...]
A -----T-----
E T-----T-----

```

Figura 4.7 Dedilhado obtido para o início da canção *Insensatez*.

tocada pelo intérprete 1.

```

E -----
B R-----R-----
G M-----M-----
D F-----F----- [...]
A ----T---T-----T---T
E -----

```

Figura 4.8 Dedilhado obtido para o início da canção *Eu Sei Que Vou Te Amar*.

Uma vez encontrados os dedilhados das obras coletadas, passou-se à validação dos resultados. Nela, 3 violonistas competentes analisaram os resultados em busca de dedilhados errôneos ou inaceitáveis, como, por exemplo, dedilhados anatomicamente incorretos ou mesmo impossíveis. Com a configuração descrita anteriormente, não se encontrou nenhum dedilhado errado ou impossível, sendo todos classificados como aceitáveis, mesmo que houvesse, ainda, alguma discordância em certos trechos¹⁷.

4.4.3 Abstração e Representação dos Dados

De posse da canção com o dedilhado devidamente identificado, pode-se aplicar a ela o que resulta dos procedimentos descritos na Seção 4.4.1, isto é, a pulsação e a quantização. Introduzindo novos símbolos, + para representar a ocorrência de um pulso, e | para designar a barra de compasso, obtém-se a representação de acordo com o ilustrado na Figura 4.9. Como nos casos anteriores, cada linha da matriz representa uma corda do violão, com exceção da última, que representa os momentos onde acontecem pulsações.

¹⁷Este algoritmo e os resultados aqui descritos foram publicados na *International Computer Music Conference*, edição de 2004 [Trajano et al., 2004].

O resultado de todo o processo de pré-processamento dos dados (a determinação da pulsação, a quantização, o dedilhado da mão direita e a representação monofônica derivada) é capturado por uma representação orientada a objetos, cujo diagrama de classes é apresentado na Figura 4.11¹⁹.

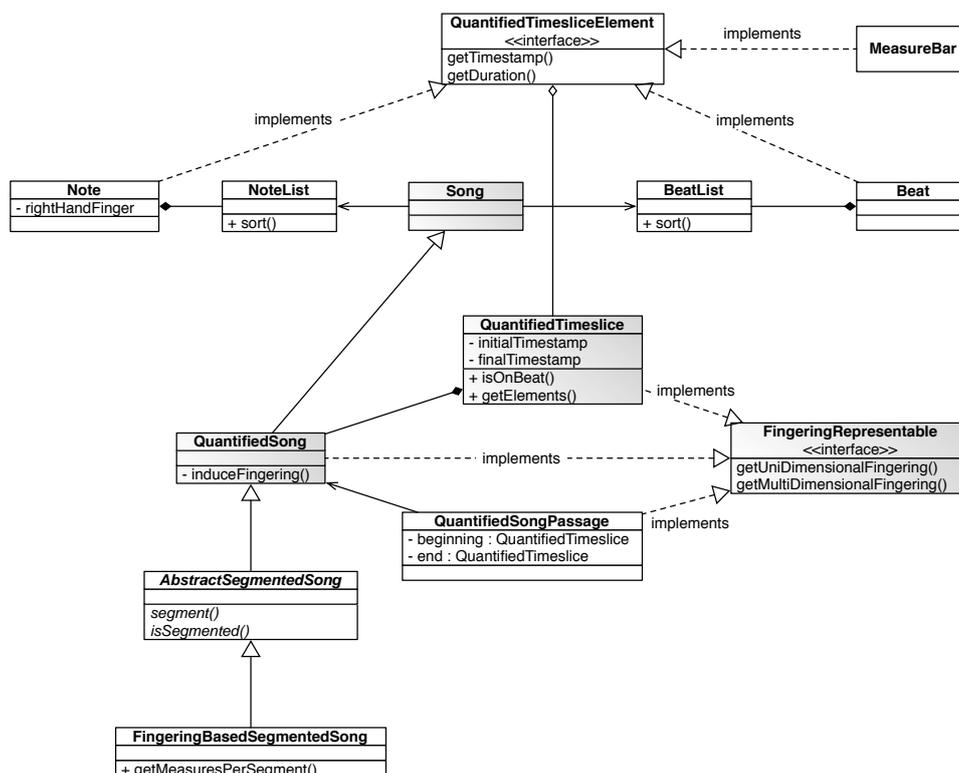


Figura 4.11 Representação dos dados

Cabe aqui salientar que essa representação é suficiente para tratar o problema aqui proposto, isto é, a extração de padrões rítmicos.

As classes mais importantes são SONG, QUANTIFIEDSONG, QUANTIFIEDTIMESLICE e a interface FINGERINGREPRESENTABLE (todas encontram-se salientadas em cinza no diagrama). A classe SONG encapsula um arquivo MIDI que, assume-se, já se encontra com a pulsação previamente determinada. Além de encapsular o arquivo, cabe a esta classe construir listas ordenadas contendo as notas (NOTE) e os pulsos (BEAT), otimizando o acesso às informações relativas a estas entidades²⁰.

¹⁹Apenas alguns atributos e métodos são mostrados no diagrama.

²⁰Em um arquivo MIDI, a duração de uma nota não é diretamente acessível, sendo necessário varrer o arquivo em busca do início e do fim da nota desejada e só então calcular a duração. Estas listas otimizam o acesso a esta informação.

A classe `QUANTIFIEDSONG`, como seu nome já antecipa, representa uma canção devidamente quantizada. Para tanto, esta classe constrói uma lista de objetos do tipo `QUANTIFIEDTIMESLICE`, que são objetos que representam porções do tempo de mesma duração teórica. Para que estes objetos possam ser construídos, o fator de quantização, isto é, o número que especifica em quantas vezes o espaço entre dois pulsos consecutivos será dividido, deve ser conhecido de antemão (em tempo de construção, precisamente). Além da quantização, à classe `QUANTIFIEDSONG` cabe também a indução o dedilhado da mão direita, utilizando para tanto o algoritmo `RHFA`.

Já a interface `FINGERINGREPRESENTABLE` é importante principalmente para o processo de derivação da representação monofônica da música. Por contrato, classes que a implementam devem ser capazes de se descrever tanto em termos do dedilhado em sua forma polifônica (como resultante do algoritmo `RHFA`), quanto em termos do dedilhado em sua forma monofônica. Isto significa que as classes que implementam esta interface devem encapsular o cálculo de como esta transformação é feita.

É importante notar, por fim, que, a despeito da seqüência de transformações necessárias para se chegar a certas abstrações, com a forma de representação aqui adotada, não existe perda de informação. Tomando as transformações que levam à representação monofônica da música como exemplo, pode-se dizer que, se tais transformações fossem operadas diretamente na representação *string* da música, e não sobre objetos que podem, eventualmente, se descrever como cadeias de caracteres, haveria perda de informação. Neste caso, não seria possível, por exemplo, dizer exatamente em que cordas do instrumento uma puxada ou um baixo foi tocado. Já com uma representação orientada a objetos apresentada, esta informação não é perdida.

EXPERIMENTOS E PROTÓTIPO

Neste Capítulo, descreve-se como foi feita a extração de padrões no conjunto de dados de Bossa Nova previamente preparado (Seção 5.1), assim como o desenvolvimento de um protótipo que serviu como ferramenta de apoio às atividades analíticas (Seção 5.2). Na Seção 5.3 são expostas, ainda, algumas considerações finais sobre o trabalho realizado.

5.1 EXTRAÇÃO DE PADRÕES

Esta seção, que trata da extração de padrões, está organizada da seguinte maneira: em primeiro lugar, na Seção 5.1.1 discute-se um dos problemas no contexto da presente pesquisa que é a avaliação dos resultados obtidos após a execução dos algoritmos de extração. Nela, apresenta-se também a solução aqui adotada. Nas duas seções seguintes (Seções 5.1.2 e 5.1.3) apresentam-se dois experimentos realizados com os dados preparados. Já na Seção 5.1.4, apresenta-se uma modificação realizada no algoritmo FLEXPAT, assim como os resultados obtidos. Por fim, na Seção 5.1.5 discutem-se os resultados obtidos nos experimentos.

5.1.1 Avaliação dos Algoritmos de Extração

Na Seção 3.2.5 foram expostos e discutidos uma série de fatores que indicavam que os algoritmos FLEXPAT e SIMILARITYSEGMENTER, seriam, dos algoritmos estudados, os mais adequados para a extração de padrões no contexto desta pesquisa. Mas, mesmo assumindo tais premissas, como avaliar os resultados da extração de padrões? Como dizer se os padrões encontrados são válidos ou não? Como decidir que algoritmo produz os melhores resultados?

Uma das possíveis maneiras de se validar os resultados obtidos com os algoritmos seria consultar um especialista no domínio capaz de validá-los adequadamente. Tal abordagem significaria que, a cada experimento realizado, em tese, o especialista deveria ser consultado. Todavia, tal procedimento é muito custoso, pois, em geral, gasta-se muito

tempo na validação dos resultados¹. Além disso, um especialista nem sempre está disponível, ou mesmo disposto a realizar esta tarefa, que pode ser bastante tediosa. Assim, para minimizar este problema, na avaliação dos resultados seguiu-se uma abordagem diferente.

Em um trabalho realizado pelo pesquisador Marcio Dahia, integrante do grupo de pesquisas em computação musical do Centro de Informática/UFPE, foi criado um aplicativo capaz de gerar automaticamente o acompanhamento de músicas no estilo Bossa Nova [Dahia et al., 2004]. A base para a geração do acompanhamento é um catálogo composto de 21 padrões rítmicos, que fora adquirido manualmente por um especialista a partir de gravações de João Gilberto, e complementado com consultas à literatura pertinente. Exemplos de padrões deste catálogo encontram-se ilustrados na Figura 5.1².

The figure displays three examples of rhythmic patterns, each consisting of two staves: 'Puxada' (top) and 'Baixo' (bottom). All patterns are in 2/4 time. The first pattern, 'Padrão P1', shows a sequence of notes in the 'Puxada' part and corresponding notes in the 'Baixo' part. The second pattern, 'Padrão P15', shows a more complex sequence of notes in the 'Puxada' part. The third pattern, 'Padrão P17', shows a sequence of notes in the 'Puxada' part and corresponding notes in the 'Baixo' part.

Figura 5.1 Exemplos de padrões rítmicos do catálogo

Considerando que o catálogo foi formado a partir de padrões executados por João Gilberto, o criador da Bossa Nova, é razoável assumir que tais padrões estão corretos e são relevantes para o estilo em questão. Devido a esta relevância estilística, também é razoável assumir que alguns destes padrões devem ser encontrados em gravações de outros intérpretes. Desta forma, utilizou-se o catálogo como referência para a avaliação dos resultados obtidos com os algoritmos de extração, isto é, os padrões do catálogo foram tomadas como o universos de padrões e os encontrados pelos algoritmos foram comparados a este universo.

¹Este tempo tende a aumentar de acordo com o tamanho do conjunto de dados utilizado.

²Uma reprodução do catálogo pode ser encontrada no Apêndice A.

5.1.2 Experimento 1: Casamento Exato

Como exposto na Seção 3.2, algoritmos que realizam o casamento exato de padrões não são as ferramentas mais adequadas para a extração a partir de dados musicais, em especial a partir de dados que representam interpretações musicais. Apesar disso, resolveu-se realizar um experimento com um algoritmo deste tipo. Neste experimento verificou-se se os padrões do catálogo previamente apresentado poderiam ser encontrados nos dados coletados. O intuito principal era explorar os dados da maneira mais completa possível.

Para esta tarefa, foi escolhido o algoritmo Boyer–More [Boyer e Moore, 1977], que é um algoritmo de casamento exato bastante conhecido. Além disso, por possuir complexidade temporal sub-linear, uma vez que utiliza apenas uma parte da cadeia de entrada em seu processamento, é considerado o algoritmo padrão para o casamento exato de cadeias de caracteres [Gusfield, 1997].

O experimento foi conduzido da seguinte maneira: cada um dos padrões que constituem o catálogo foi transcrito para a representação monofônica aqui adotada e, em seguida, utilizado como uma das entradas do algoritmo. A segunda entrada era a própria cadeia que representava uma canção. Este procedimento foi repetido em cada uma das 16 gravações.

Os resultados obtidos foram os seguintes: dos 21 padrões que integram o catálogo, apenas um único padrão pode ser identificado pelo algoritmo no conjunto de dados. O padrão encontrado foi o padrão P1 (|A---P---B-p+----|A---P---B-p+----), que é tido como o padrão principal da Bossa Nova. Considerando as 16 gravações realizadas, este padrão só foi observado em apenas três delas. A Tabela 5.1 apresenta os resultados obtidos com este algoritmo. Cada valor na tabela representa o número de ocorrências de P1 para um dado intérprete.

Canção	Intérprete 1	Intérprete 2
Desafinado	1	0
Garota de Ipanema	6	0
Insensatez	29	4

Tabela 5.1 Número de ocorrências do padrão P1

Como se verá nas Seções 5.1.3 e 5.1.4, entretanto, vários outros padrões que compõem

o catálogo, e não apenas o padrão P1, foram utilizados pelos intérpretes³. Desta maneira, com estes resultados, confirma-se, então, que um algoritmo de casamento exato, como o Boyer–Moore, é incapaz de lidar adequadamente com os dados de interpretações musicais.

Uma análise mais aprofundada dos dados e dos resultados mostra que o problema se encontra principalmente nos pequenos desvios usualmente realizados pelos intérpretes. A título de exemplo, considere uma parte do padrão P1 transcrito como A---P---B-p+----. Tal padrão pode perfeitamente ser tocado com pequenos desvios como A---P---Bp---+---- (a antecipação da última puxada) ou A---P---B--p+---- (um atraso da puxada). Outra possível modificação é adicionar uma nota de passagem no final do padrão, como em A---P---B-p+-s- (Figura 5.2). Tais modificações não são capturadas pelo algoritmo, que considera, na verdade, as ocorrências como quatro padrões distintos, o que, neste caso, não é o mais adequado.

Figura 5.2 Pequenos desvios nos padrões

5.1.3 Experimento 2: Casamento Inexato

Uma vez verificada a inadequação de algoritmos de casamento exato para a extração de padrões no contexto da presente pesquisa, passou-se a experimentos com os algoritmos de casamento inexato previamente identificados: FLEXPAT e SIMILARITYSEGMENTER.

Antes da descrição dos experimentos, é importante salientar as diferenças entre estes algoritmos e o algoritmo utilizado no experimento descrito na Seção 5.1.2: o algoritmo utilizado no primeiro experimento procurava pela ocorrência de cada um dos padrões pertencentes ao catálogo na canção gravada, isto é, o algoritmo apontava ocorrências de um padrão previamente conhecido. Já neste experimento, os algoritmos *induziram* um conjunto de padrões que satisfaziam a certas restrições, que foram, apenas então, compa-

³Após uma simples audição das canções, é possível, inclusive, identificar sem maiores dificuldades outros padrões do catálogo.

rados aos padrões do catálogo. Estudou-se neste experimento, portanto, a adequação dos algoritmos FLEXPAT e SIMILARITYSEGMENTER, verificando, para isto, até que ponto padrões existentes no catálogo poderiam ser encontrados automaticamente no conjunto de dados.

Isso posto, os experimentos com os algoritmos de casamento inexato foram realizados da seguinte maneira: em primeiro lugar, transcreveram-se as canções para a representação monofônica anteriormente apresentada. Em seguida, os algoritmos foram configurados, e estas configurações juntamente com as canções representadas formaram suas entradas. Estabelecidas as entradas, os dois algoritmos foram executados e os resultados obtidos, vários padrões induzidos pelos algoritmos, foram então analisados. Nesta análise, verificou-se a ocorrência dos padrões do catálogo nos resultados, como já descrito na Seção 5.1.1.

Nas configurações utilizadas para o algoritmo SIMILARITYSEGMENTER, permitiu-se até duas diferenças de símbolos por compasso. Além disso, foram especificadas duas restrições estruturais, a saber: que os padrões deveriam começar no início do compasso (seu primeiro símbolo deveria ser igual a $|$) ou que eles poderiam iniciar a partir do símbolo l . Isto é, permitiu-se padrões com uma das seguintes estruturas: $|B$ ou $l * B$, onde B representa o restante do padrão e $*$ representa uma seqüência, eventualmente vazia, de sinais de subtração ($-$). Assim, $|A---P---B-p-+---$ e $l-|B---P---Bp---+l$ seriam padrões, enquanto que $P---B-p-+---|$ não seria considerado um padrão já que não possui a estrutura especificada.

Para o algoritmo FLEXPAT, as configurações utilizadas foram as seguintes:

- $m_{min} = 17$ (tamanho mínimo do padrão, em número de caracteres);
- $m_{max} = 34$ (tamanho máximo do padrão, em número de caracteres);
- limiar de similaridade local: 0,75 (valor normalizado); e
- limiar de similaridade total: 0,85 (valor normalizado).

Os parâmetros m_{min} e m_{max} foram especificados de forma a corresponder a padrões variando entre um e dois compassos. No caso de FLEXPAT, entretanto, isto não quer dizer que os padrões induzidos necessariamente começam no início do compasso. Na verdade, este tipo de restrição estrutural não pode ser especificada para este algoritmo.

Os resultados, no caso do algoritmo FLEXPAT, são representados na forma de classes de padrões, isto é, na forma de um grafo estrela, onde o nó central representa o protótipo

da classe ou padrão mais representativo, enquanto que os outros nós representam as ocorrências deste protótipo⁴. O Quadro 5.1 ilustra um exemplo de classe de padrões encontrada por FLEXPAT na canção Garota de Ipanema, executada pelo Intérprete 1 (grafo estrela n° 54). Salienta-se, ainda, que na avaliação dos resultados no caso de FLEXPAT, para simplificar o processo e unificar as comparações, apenas os protótipos de cada classe foram comparados a padrões do catálogo.

Protótipo:	A---P---B-p+---- P---P---Bp---+1--
	--P---B-p+---- P---P---Bp---+
	--P---B-p+---- P---P---Bp---+1
	--P---B-p+---- P---P---Bp---+1-- B
	--P---B-p+---- P---P---Bp---+1--
	-P---B-p+---- P---P---Bp---+1-
Ocorrências:	-P---B-p+---- P---P---Bp---+1-- B-
	P---B-p+---- P---P---Bp---+1--
	P---B-p+---- P---P---Bp---+1-- B
	---B-p+---- P---P---Bp---+1-- B-
	--B-p+---- P---P---Bp---+1-- B---
	--B-p+---- P---P---Bp---+1-- B---A
	-B-p+---- P---P---Bp---+1-- B---A-

Quadro 5.1: Exemplo de classe de padrão encontrada por FLEXPAT

Na média, foram encontradas 48 classes de padrões para cada canção (o menor número de classes foi 40 e o maior 78). Considerando apenas os protótipos de cada classe, isto significa que, por canção, na média, existem 48 padrões diferentes. No conjunto de padrões induzido, apenas 6 padrões puderam ser correlacionados a padrões do catálogo, ou seja, apenas cerca de 28,5% do total de padrões do catálogo, o que representa um número muito pequeno. Isto significa que, na média, apenas 12,5% dos resultados (6 padrões em um total de 48) estavam corretos⁵. O Quadro 5.2 indica quais dos 21 padrões do catálogo foram induzidos por FLEXPAT, assim como quais dos intérpretes executou o padrão em questão.

Já o algoritmo SIMILARITYSEGMENTER induziu, na média, 13 padrões por canção (número máximo de padrões por canção foi 24 e mínimo 4). Destes padrões, apenas 5

⁴Cf. Seção 3.2.2, em especial p. 50, para maiores informações.

⁵Vale lembrar aqui que o conceito de correção, no contexto deste experimento, significa apenas que o padrão devia ser previamente conhecido — i. e., deve constar no catálogo.

Padrão	Intérop. 1	Intérop. 2	Padrão	Intérop. 1	Intérop. 2
P1	×	×	P12		
P2	×		P13		
P3			P14		
P4	×		P15	×	
P5			P16		
P6			P17		
P7			P18		
P8			P19		
P9	×	×	P20		
P10	×	×	P21		
P11			/	/	/

Quadro 5.2: Padrões do catálogo que foram induzidos por FLEXPAT

pertenciam ao catálogo, isto é, o algoritmo foi capaz de induzir ca. de 23,8% do total de padrões do catálogo. Apesar de ter induzido um número menor de padrões do catálogo, o índice de acerto de SIMILARITYSEGMENTER foi maior: ca. de 38,4% (5 padrões em um total de 13) estavam corretos. O Quadro 5.3 indica quais foram os 5 padrões do catálogo que SIMILARITYSEGMENTER foi capaz de induzir e quem os tocou.

Padrão	Intérop. 1	Intérop. 2	Padrão	Intérop. 1	Intérop. 2
P1	×	×	P12		
P2			P13		
P3			P14	×	
P4			P15		×
P5			P16		
P6			P17		
P7			P18		
P8			P19		
P9	×	×	P20		
P10		×	P21		
P11			/	/	/

Quadro 5.3: Ocorrências dos padrões segundo SIMILARITYSEGMENTER

No que diz respeito a este experimento, pode-se concluir o seguinte: foi possível induzir, a partir do conjunto de dados coletado, alguns padrões que constavam de um catálogo previamente construído. A quantidade dos padrões induzidos pode não parecer

grande de início, mas como não foi realizado nenhum levantamento prévio dos padrões que foram utilizados em cada uma das canções, nada, de fato, pode ser objetivamente dito sobre esta quantidade (se ela é pequena ou não, por exemplo). Pode-se dizer, todavia, que os algoritmos se mostraram minimamente adequados para a extração de padrões rítmicos na Bossa Nova.

Cabem aqui, ainda, alguns comentários mais específicos sobre os resultados obtidos. Na canção *Garota de Ipanema*, tocada pelo intérprete 1, por exemplo, FLEXPAT identificou o padrão $A---P---B-p+----|P---P---Bp--+1--$, que corresponde ao padrão $|A---P---B-p+----|A---P---B-p++-1-$ do catálogo. Observando mais atentamente as modificações realizadas pelo músico (*vide* Tabela 5.1), é possível notar que ele recorrentemente substituiu o evento *A* por uma puxada (*P*) no início do segundo compasso do padrão e que ele antecipa as duas últimas puxadas deste mesmo compasso. A Figura 5.3 ilustra estas variações.

The figure displays two musical staves, 'Puxada' (top) and 'Baixo' (bottom), in 2/4 time. The 'Padrão original' section shows a sequence of notes: Puxada (quarter, quarter, quarter, quarter) and Baixo (quarter, quarter, quarter, quarter). The 'Padrão encontrado por FLEXPAT' section shows a variation: Puxada (quarter, quarter, quarter, quarter) and Baixo (quarter, quarter, quarter, quarter). The notation includes stems, beams, and dots, with a double bar line separating the two sections.

Figura 5.3 Padrão encontrado por FLEXPAT

Já na canção *Wave*, tocada pelo intérprete 2, FLEXPAT identificou, como protótipo da classe 18, o padrão $|A---P---B-p++1--|B---P---Bp--+1-$, que corresponde ao padrão $|A---P---B-p++-1|B---P---B-p++-1-$ no catálogo. Como no caso anterior, as modificações realizadas pelo intérprete são principalmente antecipações.

Ainda que, na análise dos resultados obtidos com FLEXPAT, tenha sido possível encontrar regularidades como as apontadas, alguns problemas surgiram. O principal deles foi um grande número de padrões induzidos que não possuíam uma estrutura como a dos padrões do catálogo, isto é, não começavam com o símbolo $|$ nem com o símbolo l , como, por exemplo, no caso do padrão $--B---B---+----|P---B---B---+----|P$ ⁶. Esta espécie de má formação estrutural trouxe problemas principalmente para a avaliação dos resultados, já que foi difícil validar os padrões encontrados. Em alguns casos, inclusive,

⁶Neste caso específico, FLEXPAT também foi capaz de encontrar o padrão esperado, ou seja, $|P---B---B---+----|P---B---B---+--$, mas isto não era sempre o caso.

grande parte da classe de padrões era composta por padrões deste tipo, o que tornava a validação ainda mais difícil. Já no caso de SIMILARITYSEGMENTER, este tipo de problema não aconteceu, muito possivelmente devido à possibilidade de se especificar restrições estruturais sobre os padrões.

Quanto aos resultados obtidos com SIMILARITYSEGMENTER, algo bastante interessante é que o algoritmo encontrou padrões de grande extensão. Na canção *Insensatez*, tocada pelo intérprete 1, por exemplo, o algoritmo encontrou três ocorrências do seguinte padrão (136 símbolos, isto é, 8 compassos):

```
|A---P---B-p-+---|A---P---B-p-+---|A---P---B-p-+---
|A---P---B-p-+---|A---P---B-p-+---|A---P---B-p-+---
|A---P---B-p-+---|A---P---B-p-+---
```

Ainda que formado pela mesma célula rítmica ($|A---P---B-p-+---$), é significativo que o algoritmo tenha encontrado um padrão tão longo. Analisando com atenção os dados, observa-se que, neste caso específico, o intérprete 1 utiliza este padrão e suas ocorrências como acompanhamento da maior parte do tema. Ele possui, desta forma, implicações estruturais. Outro exemplo de padrão longo é o seguinte:

```
|B---P---B-p-+-s-|A---P---B-s-+-l-|B-p-+-p-B-p-+-l-
|B-p-+-p-B---P---
```

Este padrão, com 68 símbolos, ocorreu duas vezes na canção *Tarde em Itapoã*, também tocada pelo intérprete 1, não sendo possível, entretanto, correlacioná-lo a nenhuma estrutura formal da obra. O algoritmo, entretanto, também encontrou lixo, como, por exemplo, $|B-$ e $|A----$.

5.1.4 SFIEXPAT: uma versão simplificada de FIEXPAT

No experimento anterior mostrou-se que os algoritmos previamente selecionados apresentaram resultados apenas razoáveis. Além de terem induzido uma grande quantidade de padrões, vários deles possuíam uma estrutura muito diferente da estrutura dos padrões conhecidos, principalmente no caso dos padrões induzidos por FLEXPAT, o que dificultou sobremaneira o processo de validação dos resultados. Além disso, ainda no caso de FLEXPAT, quando as cadeias de símbolos eram muito longas, aconteciam problemas com

o espaço em memória, devido à necessidade de armazenar uma matriz de dissimilaridade grande demais. Na tentativa de minorar ou, até mesmo, eliminar estes problemas com o algoritmo FLEXPAT, foram introduzidos na representação mais elementos do conhecimento sobre o domínio, o que levou a uma variação simplificada de FLEXPAT, denominada SFLEXPAT.

Mas exatamente que conhecimento é este? Observando o catálogo de padrões, nota-se que todo padrão a ele pertencente possui uma extensão fixa, contando com exatos dois compassos. Assim, foi implementado um algoritmo de segmentação que incorpora este conhecimento sobre a estrutura do padrão, ou seja, o algoritmo percorre a cadeia de símbolos criando segmentos a cada dois compassos. Esta implementação foi incorporada à classe FINGERINGBASEDSEGMENTEDSONG, que estende ABSTRACTSEGMENTEDSONG e integra a representação da música adotada (*vide* Seção 4.4.3).

Utilizar uma instância da classe FINGERINGBASEDSEGMENTEDSONG como entrada para indução de padrões fez com que fosse necessária uma mudança na etapa de comparação do algoritmo. Agora, ao invés de percorrer a representação monofônica símbolo por símbolo, o algoritmo percorre-a segmento por segmento. Assim, a incorporação deste conhecimento do domínio otimiza tanto o tempo de execução do algoritmo quanto a utilização de memória, já que reduz significativamente o tamanho da entrada e a quantidade de comparações necessárias.

Para verificar se SFLEXPAT levaria a melhoras efetivas, quando comparado a FLEXPAT, realizou-se um experimento semelhante ao descrito na Seção 5.1.3. Neste experimento foram mantidos os dados do experimento anterior, assim como a metodologia de validação. No que tange as configurações do algoritmo SFLEXPAT, os tamanhos máximos e mínimos dos padrões não fazem parte de sua configuração, uma vez que este tamanho já está pré-determinado (é o tamanho do segmento). Assim, das configurações originais, apenas os limiares de similaridade local e total foram utilizados (os valores foram mantidos).

Em média, foram encontradas 14 classes de padrões por canção, sendo o menor número de classes igual a 6 e o maior igual a 34. Estes números iniciais já representavam uma redução significativa no número classes de padrões encontrados, o que levou a crer que os resultados obtidos por SFLEXPAT poderiam realmente ser melhores. E isto foi, de fato, confirmado com a avaliação dos resultados. Agora, ao contrário do experimento anterior, foram encontrados, de maneira sistemática em cada uma das canções, padrões pertencentes ao catálogo. No total, 12 dos padrões induzidos pertenciam ao catálogo, ou

seja, foi possível encontrar nos dados 57,1% dos padrões do catálogo, um aumento expressivo se comparado ao experimento anterior. Além disso, houve uma melhora significativa no percentual de acerto, que passou a ser de 85,7%. O Quadro 5.4 mostra que padrões pertencentes ao catálogo foram encontrados pelo algoritmo nos dados coletados.

Padrão	Intérop. 1	Intérop. 2	Padrão	Intérop. 1	Intérop. 2
P1	×	×	P12		
P2	×	×	P13		
P3	×	×	P14		
P4	×		P15		×
P5	×		P16	×	×
P6	×		P17		
P7			P18		×
P8			P19	×	
P9	×	×	P20		
P10	×	×	P21		
P11			/	/	/

Quadro 5.4: Ocorrências dos padrões segundo SFLEXPAT

Vale salientar ainda que, além de ter encontrado uma quantidade muito maior de padrões do catálogo, as classes de padrões induzidas por SFLEXPAT eram formadas por uma quantidade bem menor deles. Além disso, estes padrões, por possuírem uma estrutura semelhante à dos padrões do catálogo, puderam ser correlacionados a estes logo à primeira vista. O Quadro 5.5 consta um exemplo de classe de padrão identificada por SFLEXPAT na canção *Wave*, executada pelo intérprete 2 (classe de padrões nº 1).

Protótipo:	A-p+-p-B-p+-l- B-p+-p-B-p+-l-
Ocorrências:	A---P---B-p+-l- B-p+-p-B-p-B-s- A-p+-p-B-p+-l- B-p+-p-B-p+-l- l- B--p+++a+p+-l- B-p+-p-B-p+-s- B-a+p--A-p+-l- B-p+-p-B-p+-s- A---+p-B-p+-s- A--pS-p-B-p+-s- l- A---P---B--p+-l- B---P---B-p+-l- l- B---P---B-p+-l- B-p+-p-B-p+-l- l- B---P---B-p+-s- A-p+-p-B-p-S-s-

Quadro 5.5: Exemplo de classe de padrão encontrada por SFLEXPAT

5.1.5 Etapa de Extração de Padrões: Discussão dos Resultados

No que diz respeito aos experimentos realizados, pode-se dizer, a partir dos resultados obtidos, que, para uma efetiva extração de padrões, no contexto desta pesquisa, é essencial que os algoritmos possuam, no mínimo, duas características básicas: em primeiro lugar, conforme os resultados obtidos com o primeiro experimento, eles devem ser algoritmos de *casamento inexato*. Além disso, a análise dos resultados mostrou a *flexibilidade* como uma segunda característica essencial. Tal flexibilidade serve tanto para lidar com as pequenas variações interpretativas dos músicos, quanto para restringir os padrões que devem ser induzidos.

Quanto à avaliação dos resultados, pode-se dizer que mostrou-se adequada a abordagem de avaliação adotada, isto é, verificar os resultados dos experimentos tendo como base o catálogo de padrões. Seria bastante complicado e demorado submeter os resultados de variados experimentos à inspeção de um especialista. Além disso, nos experimentos comparou-se o desempenho de algoritmos que utilizam um conjunto diferente de restrições. Sem o uso do catálogo, que funcionou como uma espécie de *benchmark*, poderia ser bastante complicado estabelecer critérios de comparação entre algoritmos tão diversos.

Ainda que o primeiro experimento não tenha produzido resultados interessantes, é relevante notar que, de todos os padrões do catálogo, o algoritmo de casamento exato tenha encontrado justamente o padrão P1 (e apenas ele), que é considerado o padrão principal da Bossa Nova. É difícil explicar este resultado, mas uma hipótese plausível, ao menos *a priori*, seria que, por este ser o padrão característico do estilo, os intérpretes tendam a executá-lo com poucas ou mesmo nenhuma variação rítmica — talvez em uma tentativa de “afirmar” musicalmente o estilo que estão interpretando. Todavia, trata-se apenas de uma hipótese: seriam necessários mais dados e experimentos na tentativa confirmá-la ou rejeitá-la.

Os algoritmos de casamento inexato também suscitaram outras dúvidas. No caso de SIMILARITYSEGMENTER, dada à característica não-determinística própria do algoritmo, restou a seguinte dúvida: se um padrão do catálogo não foi encontrado isto se deu por que ele de fato não foi tocado ou por que SIMILARITYSEGMENTER não executou por tempo suficiente? Este pode ter sido o caso, por exemplo, do padrão P10 para o intérprete 1: ele foi encontrado diversas vezes por FLEXPAT, mas não foi induzido por SIMILARITYSEGMENTER (*vide* Tabelas 5.2 e 5.3).

É interessante notar ainda que o contrário também aconteceu: dois dos padrões encon-

trados por SIMILARITYSEGMENTER, P14 e P15 (para o intérprete 2), não foram encontrados por FLEXPAT. Como FLEXPAT não possui características não-determinísticas, os padrões deveriam ter sido induzidos. Com uma análise mais aprofundada dos resultados, entretanto, foi possível identificar que, ao satisfazer o limiar de similaridade total especificado nos experimentos, FLEXPAT acabava descartando algumas classes de padrões, classes estas que continham os padrões em questão.

No que concerne ao problema da má formação estrutural dos resultados obtidos com FLEXPAT, deve-se tentar examinar o problema por outro ângulo. Observando o catálogo de padrões, é possível verificar que vários dos padrões possuem partes de sua estrutura em comum, isto é, existem “pedaços” dos padrões que são iguais. Assim, considere \mathcal{S} como sendo o conjunto de padrões do catálogo que possuem o mesmo segundo compasso (i. e., o mesmo ritmo) e \mathcal{P} como sendo o conjunto de padrões cujo ritmo no primeiro compasso é o mesmo. Os resultados de FLEXPAT podem, na verdade, estar revelando que existe uma quantidade tão significativa de subsequências formados por elementos de \mathcal{S} seguidos de elementos de \mathcal{P} , que tais conexões acabam sendo “promovidas” a padrões. Assim, ao invés de um problema, isto pode significar algo completamente diferente: que o interessante na Bossa Nova não é tanto os padrões em si, mas sim como eles são concatenados! Obviamente, esta hipótese precisa ser melhor investigada.

Por fim, a partir dos resultados obtidos com SFLEXPAT, foi possível identificar diferenças no tipo de padrão rítmico utilizado pelos dois intérpretes. No catálogo⁷, os padrões estão agrupados de acordo com características relacionadas à estrutura da música. Há, por exemplo, um grupo de padrões que são tocados em “viradas”, isto é, trechos na estrutura da música que marcam mudanças de frases ou seções; outros que são tocados no início da música; existe o padrão principal, que marca o estilo em questão; etc. Relacionando os padrões identificados pelo algoritmo a estes grupos, foi possível observar uma clara preferência do intérprete 1 por padrões do tipo cíclico (padrões que cuja última nota está ligada à primeira do padrão seguinte).

5.2 FERRAMENTA DE AUXÍLIO À ANÁLISE

Todo o processo analítico descrito até o momento culminou na modelagem e implementação de uma ferramenta de auxílio à análise do acompanhamento. Em linhas gerais, esta ferramenta possui quatro componentes principais, ilustrados na Figura 5.4.

⁷Os padrões que formam o catálogo encontram-se listados no Apêndice A.

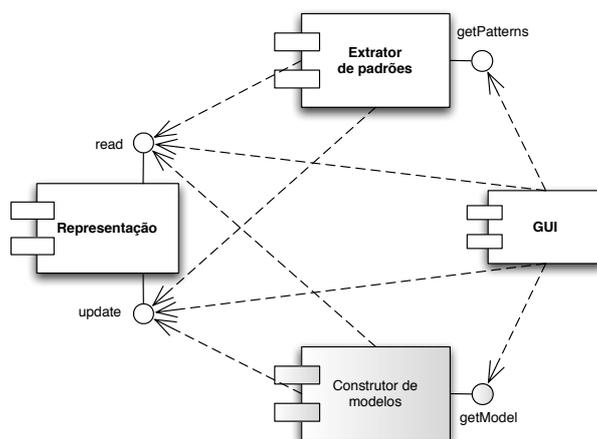


Figura 5.4 Componentes dos sistema

O componente **Representação** corresponde à representação estendida da música, já descrita na Seção 4.4.3, e acrescida das classes `FINGERINGBASEDSEGMENTEDSONG` e `PATTERN`, conforme explicado anteriormente. O componente **Extrator de Padrões** representa os algoritmos de extração de padrões `FLEXPAT` e `SFLEXPAT`. O componente **GUI** representa a interface gráfica construída, enquanto que o componente **Construtor de Modelos** encapsula os algoritmos de aprendizagem adequados para a construção de modelos do acompanhamento. É importante notar que, mesmo que a extração de padrões possa ser vista, no contexto da presente pesquisa, como uma parte da etapa de mineração de dados, o que poderia fazer com que o componente **Construtor de Modelos** encapsulasse também tais algoritmos, preferiu-se separar explicitamente estas duas etapas em componentes diferentes. Tomou-se esta decisão principalmente porque, além de integrar a etapa de mineração de dados, a extração de padrões também pode ser parte da etapa de pré-processamento. O componente **Construtor de Modelos** é o único que não foi implementado, justamente porque em nenhum dos experimentos foram utilizados algoritmos de aprendizagem.

O protótipo foi implementado em Java⁸ e é composto por 46 classes e cerca de 9000 linhas de código. Para a implementação da interface gráfica, seguiu-se o padrão de projeto MVC (*Model-View-Controller*) [Gamma et al., 2000]. Aqui, o modelo corresponde às classes da representação estendida que possuem ou necessitam de algum tipo de visualização. No caso dos algoritmos de extração, pensou-se na adoção do padrão *Strategy*, mas dado o baixo número de algoritmos (dois apenas), optou-se por uma solução mais

⁸<http://java.sun.com>

simples e direta: o controlador que dispara a extração de padrões escolhe o algoritmo de acordo com a entrada fornecida (objetos do tipo `ABSTRACTSEGMENTEDSONG` são processados por `SFLEXPAT`, enquanto objetos do tipo `QUANTIFIEDSONG` são processados por `FLEXPAT`). Todavia, no caso da incorporação de outros algoritmos ao sistema, o padrão *Strategy* pode ser utilizado sem muito esforço. A seguir, seguem algumas capturas de tela apresentando funcionalidades do sistema.

A primeira captura de tela (Figura 5.5) mostra a interface criada para a visualização do dedilhado. Nela, é possível, além desta visualização, alternar o fator de quantização e escutar o arquivo MIDI em análise. É possível corrigir eventuais falhas editando cada um dos quadros contendo os símbolos que representam o dedilhado. Esta é, portanto, visão fornecida pelo sistema da classe de representação `QUANTIFIEDSONG`.

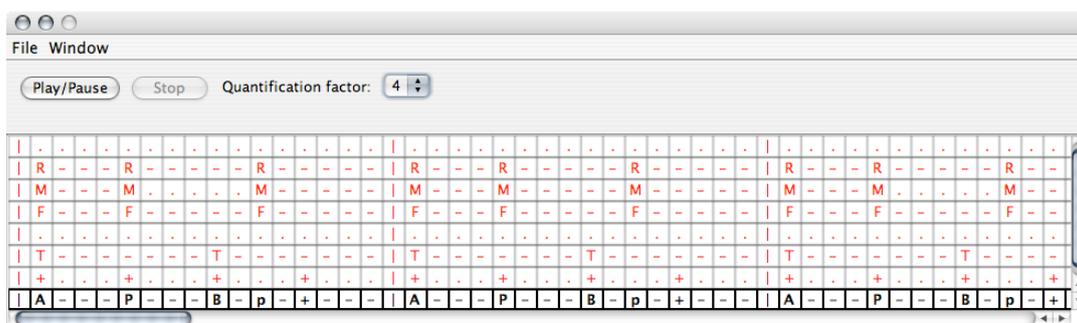


Figura 5.5 Dedilhado em sua versão não-segmentada

Já a segunda captura de tela (Figura 5.6) mostra como um objeto do tipo `ABSTRACTSEGMENTEDSONG` é visualizado pelo usuário. Aqui, cada segmento da obra é identificado visualmente, através de trechos em cores alternadas (branco e cinza). As mesmas funcionalidades da tela anterior (edição dos símbolos, alternância do fator de quantização, etc.) também são fornecidos para nesta tela.



Figura 5.6 Dedilhado em sua versão segmentada

No que concerne os algoritmos FLEXPAT e SFLEXPAT, a ferramenta oferece duas interfaces: uma para que o usuário forneça os parâmetros de entrada e outra para visualizar os resultados (o grafo de eqüipolência). Na Figura 5.7, mostram-se as interfaces de entrada dos parâmetros, enquanto que, na Figura 5.8, ilustra-se a representação do grafo de eqüipolência.

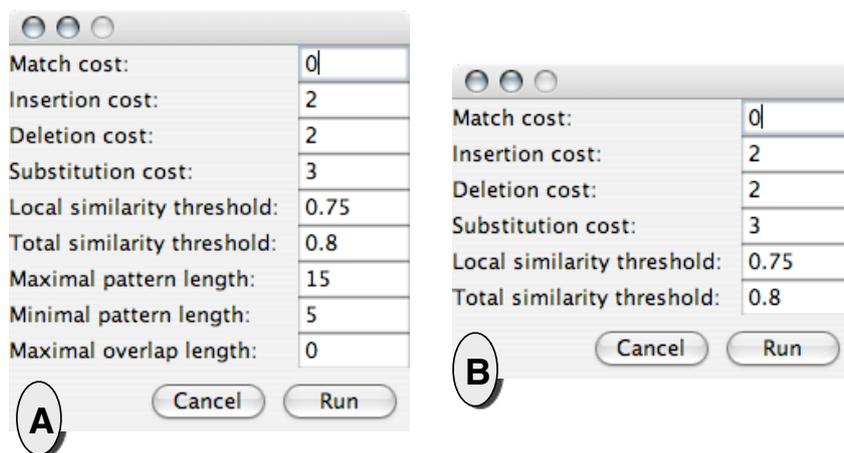


Figura 5.7 Entrada de parâmetros para FLEXPAT (A) e SFLEXPAT (B)

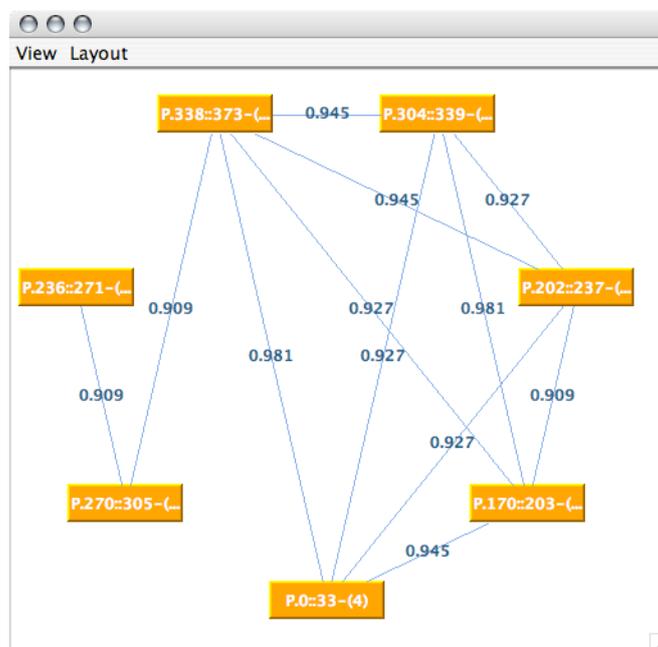


Figura 5.8 Grafo de eqüipolência

No exemplo da Figura 5.8, cada nó do grafo de eqüipolência representa um padrão. O rótulo de cada nó deve ser compreendido da seguinte maneira: $P.i :: j - (n)$ significa

“padrão que se inicia no símbolo de índice i da representação monofônica e vai até o índice j , ocorrendo n vezes na obra”. Os valores ao lado das arestas entre os nós representam a similaridade entre os padrões.

É importante ressaltar que na implementação da ferramenta levou-se em conta, na medida do possível, os requisitos previamente apresentados (cf. Seção 1.4). Deles, observou-se, em especial, a flexibilidade, a capacidade de manipulação dos dados musicais e a interface, que abstraiu conceitos específicos da Ciência da Computação. Quanto à possibilidade de construção de modelos explícitos e gerais da interpretação, a ferramenta disponibiliza unicamente a criação de experimentos com os algoritmos FLEXPAT e SFLEXPAT, cujos resultados podem ser estudados através da análise dos grafos resultantes. Já no que diz respeito ao auxílio na comparação entre modelos, ainda que seja possível comparar os modelos obtidos através da análise visual dos grafos resultantes, não foi implementada uma funcionalidade específica que satisfaça este requisito.

Como dito anteriormente, o componente **Construtor de Modelos** não foi implementado. Apesar disso, é possível descrever uma das dificuldades com as quais a implementação deste componente terá que lidar, a saber: a montagem de experimentos que envolvam algoritmos de aprendizagem de máquina. Existem, fundamentalmente, dois problemas que devem ser resolvidos, ambos ligados a requisitos elicitados anteriormente. O primeiro deles está ligado à necessidade de minimizar o conhecimento de aspectos computacionais necessários para que o usuário possa utilizar a ferramenta. O segundo problema é que, ao minimizar o conhecimento de aspectos essencialmente computacionais, não se pode abrir mão da flexibilidade na montagem dos experimentos. Uma possível abordagem é construir algo semelhante à ferramenta *KnowledgeFlow*, uma das ferramentas que integram o ambiente WEKA [Witten e Frank, 2000]. Nesta ferramenta, cada funcionalidade do ambiente possui uma representação icônica, sendo os experimentos montados através da configuração dos parâmetros de cada funcionalidade e da conexão entre ícones. A Figura 5.9 ilustra a montagem de um experimento na ferramenta.

Cabe salientar, por fim, que a ferramenta não foi testada por um musicólogo. Imagina-se que possam surgir problemas não identificados durante a implementação e os testes realizados, mas nada que torne seu uso inviável. O maior entrave para um efetivo uso da ferramenta é o tipo de dado utilizado: gravações MIDI, especialmente gravações como as aqui utilizadas, não são tão fáceis de se obter.

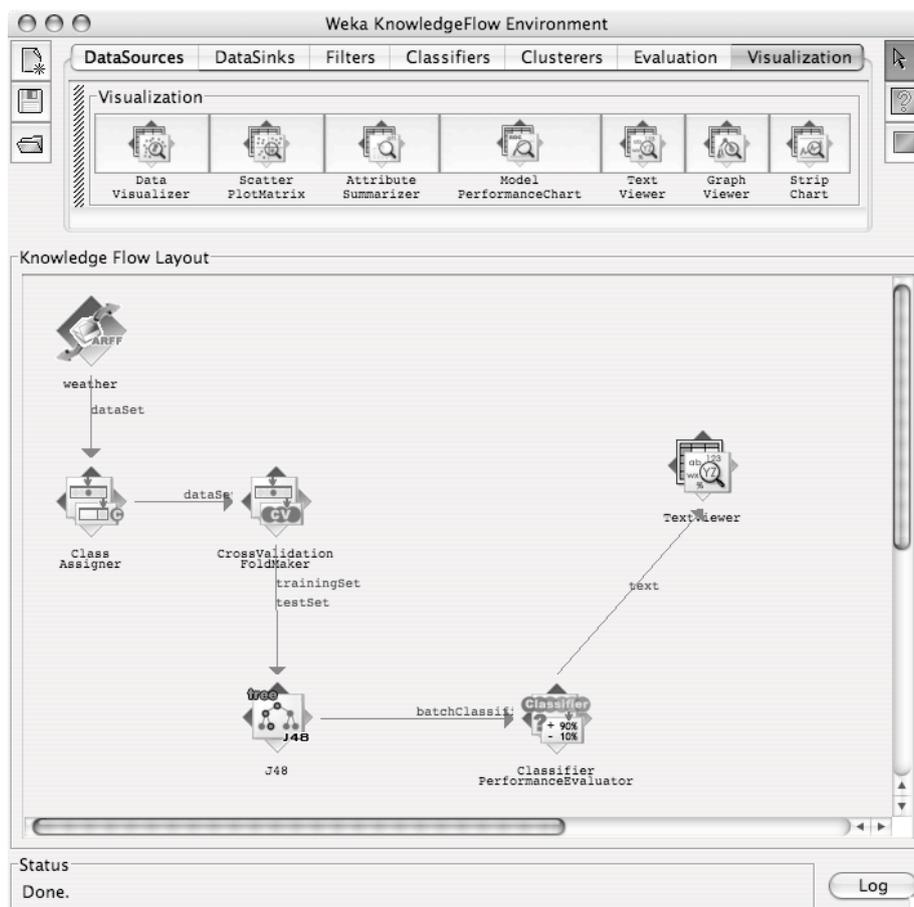


Figura 5.9 Exemplo de experimento utilizando a ferramenta *KnowledgeFlow*

5.3 CONSIDERAÇÕES FINAIS

Considerando o exposto, cabem ainda algumas considerações finais sobre trabalho realizado. Em primeiro lugar, durante a descrição do trabalho realizado, foi dada uma certa ênfase à possibilidade de investigação a partir de uma grande quantidade de dados. A quantidade de dados utilizada no trabalho, entretanto, não corresponde à ênfase dada, principalmente comparando os volumes de dados aqui utilizados ao volume disponível para análise na pesquisa em interpretação austríaca. Isto se deveu a dificuldades inerentes à aquisição de dados musicais, principalmente dados simbólicos, que necessitam de intérpretes dispostos a gravá-los. Apesar disto, cabe salientar que as canções gravadas representam o maior conjunto de dados de Bossa Nova. Além disso, é o único que se conhece onde os dados simbólicos (MIDI) estão sincronizados com os dados contínuos (áudio).

No que concerne os experimentos aqui conduzidos, ainda que com um caráter inicial

e uso de poucos dados, os resultados obtidos mostraram as possibilidades e o potencial das contribuições que a análise automática pode proporcionar. Destes resultados é muito importante ressaltar a elaboração de novas perguntas sobre a formação do acompanhamento rítmico, como, por exemplo, identificar a concatenação entre padrões como uma informação possivelmente tão relevante quanto os padrões em si.

Por fim, várias das perguntas formuladas anteriormente acabaram ficando sem resposta (cf. p. 5). Entretanto, apenas com experimentos realizados com quantidade mínimas de dados, foi possível apontar possíveis respostas para algumas delas. Verificou-se, por exemplo, que existem, de fato, relações entre os dedos da mão. Tais relações foram, inclusive, exploradas na criação da representação da música. Além disso, os resultados obtidos com SFLEXPAT forneceram subsídios que fortalecem a possibilidade da existência de preferências na escolha dos padrões por parte dos intérpretes. Se existem preferências, talvez seja possível construir um vocabulário de padrões específico de um intérprete. Quanto à dúvida de tudo ser padrão ou não, com os dados examinados, pode-se dizer que sim.

CONCLUSÕES

Neste documento, foram examinados problemas relacionados à descoberta automática de conhecimento em interpretações musicais. Em primeiro lugar, mostrou-se que a interpretação musical é um fenômeno complexo, sujeito a elementos subjetivos, mas que apresenta aspectos comuns, mesmo em diferentes interpretações. Mostrou-se ainda que tanto a pesquisa em Música, quanto a pesquisa em Computação Musical analisam o fenômeno no âmbito da Música Clássica composta para piano, excluindo uma boa parte do repertório musical. Posto isto, estabeleceu-se um escopo de pesquisas diferente do usual: a descoberta automática de conhecimento a partir do acompanhamento rítmico ao violão no estilo Bossa Nova.

Sistemas e metodologias para este fim são relevantes em dois aspectos. O primeiro deles é que analistas e musicólogos passam a ter uma série de novas possibilidades investigativas, pois tais sistemas e metodologias possibilitam a investigação dos mais diferentes aspectos da interpretação do ritmo. Exemplos de tais possibilidades vão desde variadas comparações entre padrões rítmicos, até o estabelecimento de relações entre diferentes dimensões da música, todas podendo se basear na análise de uma quantidade de dados muito maior que o usual e podendo ser reformuladas sem maiores dificuldades e a qualquer tempo com a investigação de novos dados. O segundo aspecto é que estes sistemas, metodologias e, principalmente, o conhecimento resultante das análises podem ser empregados na construção de outros tipos de sistemas musicais, como, por exemplo, sistemas para ensino de violão, assistentes de arranjo, assistentes de ensaio e sistemas de acompanhamento automático.

Estabelecido este escopo e identificada sua relevância, buscou-se, por um lado, por conceitos teóricos-musicais que fundamentassem o processo de automação da análise, e por outro, por elementos da pesquisa em Ciência da Computação que pudessem servir de inspiração aos propósitos desta pesquisa. Identificou-se, em primeiro lugar, a batida ou padrão rítmico, que funciona como bloco fundamental na construção do acompanhamento. Identificou-se também um processo utilizado na automação da descoberta de

conhecimento, processo este que foi adaptado e moldado para que uma análise do acompanhamento orientada a padrões pudesse ser realizada. Além deste processo, também foram identificadas diversas técnicas aplicáveis à análise da interpretação, em especial à análise aqui realizada.

Com base nos elementos identificados, definiu-se um processo analítico que passa por cinco fases distintas: a aquisição dos dados, a preparação dos dados, a extração de padrões, a construção de modelos e a validação e visualização dos resultados. No estabelecimento deste processo, verificou-se, ainda, que a representação simbólica dos dados era um resultado muito importante e que a etapa de extração de padrões possuía natureza dupla, podendo tanto ser vista como um passo na preparação dos dados, quanto um elemento na etapa de construção de modelos. Cada uma destas fases foi discutida e diversos métodos e técnicas foram estabelecidos, exceção feita à fase de construção de modelos. Ao final do processo, o protótipo de uma ferramenta de apoio a partes deste processo analítico foi desenvolvida. Foram realizados, ainda, dois experimentos que testaram e puseram em prática elementos do processo aqui definido.

6.1 CONTRIBUIÇÕES

Uma das contribuições obtidas com a presente pesquisa se deu na construção do conjunto de dados utilizado nos experimentos. Trata-se do maior, talvez único, conjunto de dados de Bossa Nova, e é o único conhecido onde os dados simbólicos (MIDI) estão corretamente sincronizados com os dados contínuos (áudio).

Outra contribuição foi a criação de uma representação original da massa de dados coletada. Tal representação, além de ter reduzido a dimensionalidade dos dados sem perdas de expressividade, simplificando sobremaneira a extração de padrões, baseou-se em conceitos criados pelos próprios músicos. Isto significa que, *a priori*, os resultados obtidos na fase de extração podem ser compreendidos pelos músicos sem maiores dificuldades. Especula-se ainda que, com algum treino, os intérpretes, ou mesmo musicólogos, possam utilizar a representação como forma de anotar os padrões executados, o que representa também uma contribuição para a transcrição e notação do acompanhamento rítmico, pois viabiliza a superação de uma barreira que é inexistência de uma notação para o acompanhamento.

Criou-se, ainda, um algoritmo para a indução do dedilhado da mão direita, RHFA, algoritmo este que produziu os resultados que serviram de base para a construção de

várias abstrações, inclusive, as que constituem a representação utilizada na transcrição do ritmo.

Dos experimentos realizados também resultaram contribuições. Uma delas foram as adaptações que levaram à criação de uma versão simplificada de FLEXPAT, o SFLEXPAT. Outra contribuição foi o estabelecimento, mesmo a partir de um conjunto de dados reduzido, de novas hipóteses analíticas que podem e devem ser exploradas em outros experimentos, como, por exemplo, a possibilidade das concatenações entre padrões serem tão relevantes quando os padrões em si.

O próprio conjunto de perguntas que serviu de motivação inicial e guia para a condução da pesquisa também se constituiu em uma contribuição, pois, mesmo que encontradas na literatura científica, tais perguntas estavam dispersas em variados trabalhos e com objetivos diferenciados. Aqui, as perguntas foram reunidas e unificadas em torno de um elemento único, o padrão rítmico. Além disso, as perguntas auxiliaram na identificação e especificação de um escopo onde novas pesquisas sobre a análise da interpretação podem ser conduzidas, o que também não deixa de ser uma das contribuições do trabalho.

Por fim, tem-se como contribuições a ferramenta de auxílio à análise do acompanhamento e a adaptação do ciclo de descoberta de conhecimento para as especificidades do problema. Ainda que apenas um protótipo, a ferramenta possibilitou que fossem estabelecidas novas hipóteses analíticas e que subsídios para responder a algumas das perguntas formuladas fossem encontrados. Já a adaptação do ciclo, além de orientar a utilização da ferramenta, funcionou como metodologia de análise automática.

As contribuições do presente trabalho podem, então, ser enumeradas da seguinte maneira:

1. um conjunto de perguntas reunidas e unificadas em torno de um elemento estrutural único;
2. o conjunto de dados de Bossa Nova;
3. uma representação original para os dados coletados;
4. um algoritmo original para a indução do dedilhado da mão direita;
5. o algoritmo SFLEXPAT, gerado a partir de modificações e adaptações do algoritmo FLEXPAT;
6. uma ferramenta de auxílio à análise;

7. um processo que serve como metodologia de análise automática; e
8. um conjunto de novas hipóteses analíticas.

6.2 LIMITAÇÕES E TRABALHOS FUTUROS

Dado o caráter pioneiro deste trabalho, limitações, e os decorrentes trabalhos futuros, não faltam. Deve-se citar, de imediato, a necessidade de completar o ciclo de descoberta adaptado com a etapa de construção de modelos. Ela é vital para a completa validação do processo. Deve ser considerada, no âmbito destas atividades, a implementação de uma solução para a montagem de experimentos, como, por exemplo, algo nos moldes do que foi realizado no ambiente WEKA (cf. Seção 5.2, p. 101), ou mesmo uma solução alternativa.

Intrinsecamente ligada à construção de modelos está a ampliação da representação aqui adotada. Por hora, apenas elementos diretamente ligados ao ritmo são representados. Todavia, mesmo para o estudo do ritmo, abstrações ligadas a outras dimensões da música necessitarão ser representadas. Antevê-se, por exemplo, a estrutura harmônica como um elemento a integrar a representação. Além da harmonia, a melodia é outra dimensão fundamental que precisa ser representada.

No que concerne etapa de aquisição, é necessário compreender melhor que aspectos do equipamento utilizado influenciam a geração dos ruídos anteriormente relatados. Caso não seja possível eliminar o ruído diretamente na fonte, é necessário fazer com que a limpeza aqui apresentada seja feita da maneira mais automática possível.

Ainda no âmbito da aquisição dos dados, citou-se a captação de arquivos áudio em conjunto com os dados MIDI. Seria interessante verificar que tipo de informação adicional pode ser extraída do áudio e incorporada à representação. É relevante salientar que a retirada de informações de arquivos áudio encerra uma série de novos desafios e problemas computacionais ainda não resolvidos. Mas, um conjunto de dados onde informações contidas em arquivos áudio encontram-se devidamente sincronizadas com informações simbólicas pode ser extremamente útil na solução destes desafios e problemas.

O algoritmo de quantização aqui utilizado é ingênuo demais. Ele, por exemplo, assume que não existem desvios locais de tempo, dividindo o espaço entre pulsos consecutivos igualmente, o que nem sempre é verdade. Como a pulsação foi determinada em um nível estrutural relativamente baixo, as limitações do algoritmo foram de certo modo camufladas, mas, ainda assim, ocorreram algumas distorções que necessitaram de correção ma-

nual, que foi realizada com o auxílio da ferramenta implementada. Além disso, figurações como quiáleras não são quantizadas corretamente. Assim, é necessário melhorar esta etapa da preparação dos dados.

O algoritmo RHFA, por sua vez, se mostrou adequado para a indução do dedilhado para o acompanhamento acórdico. Ainda que seqüências de notas isoladas (solos) tenham sido consideradas à época da construção do algoritmo, é preciso verificar com cuidado sua eficácia para estes casos.

É muito importante trabalhar na automação da análise dos resultados. Ainda existe uma grande parte do trabalho de validação que é feita manualmente. Com a quantidade de dados aqui analisada, ainda foi possível realizar esta tarefa manualmente, mesmo que isto tenha sido pouco produtivo e tedioso. Mas com o eventual aumento da massa de dados, realizar esta tarefa manualmente se torna impraticável. Seria valioso também a existência de um módulo que calculasse estatísticas variadas dos dados em análise (número de notas, número de acordes, acordes mais tocadas, durações mais freqüentes, etc.). Isto ajudaria o analista a ter uma melhor visão dos dados e até a identificar novos experimentos.

Verificou-se nos experimentos que os algoritmos de extração de padrões possuíam vantagens diferentes. Com FLEXPAT os resultados eram apresentados de maneira estruturada (o grafo estrela). Já no caso de SIMILARITYSEGMENTER, como não existiam limitações nos tamanhos máximos dos padrões, foi possível induzir padrões longos que puderam, em seguida, ser associados a estruturas formais da música. Seria interessante, portanto, investigar a possibilidade de combinação dos diferentes algoritmos (estes dois ou mesmo outros) em uma espécie de meta-extrator, de maneira a reunir suas vantagens e diminuir as desvantagens.

No que concerne à ferramenta, além de melhorias nas funcionalidades implementadas, existe uma ampla gama de possibilidades no que tange a visualização dos dados e dos modelos. Imagina-se que boas ferramentas de visualização possam ser extremamente valiosas na compreensão dos resultados, principalmente fornecendo alternativas à inexistência de notação musical inerente ao domínio.

Durante a pesquisa foi coletado um outro conjunto de dados. Este conjunto consta de onze gravações de diversos sambas realizadas por três intérpretes diferentes. Seria extremamente interessante aplicar o que foi aqui idealizado a este conjunto de dados. Um novo *corpus* analítico certamente melhoraria a metodologia, as técnicas e as ferramentas aqui desenvolvidas, assim como ajudaria a validar tudo o que foi exposto. Pode-se pensar em,

além de ampliar as pesquisas para outros estilos, ampliá-las para outros instrumentos ou conjuntos de instrumentos. A extração de conhecimento em vários estilos e instrumentos, desta forma, pode se constituir em um grande projeto de descoberta científica sobre a música brasileira.

Como se pode ver, ainda existe muito trabalho a ser realizado. Espera-se, entretanto, que este trabalho sirva de matéria-prima para a compreensão de que o computador deve funcionar como uma máquina de raios X e que o resultado de suas computações, o *raio X musical*, revele ao olho atento os interiores da música!

REFERÊNCIAS BIBLIOGRÁFICAS

- [Abdallah e Plumbley, 2004] Abdallah, S. A. e Plumbley, M. D. (2004). Polyphonic music transcription by non-negative sparse coding of power spectra. In *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR'04)*, Barcelona, Spain.
- [Agon, 1998] Agon, C. (1998). *OpenMusic: Un langage visuel pour la composition musicale assistée par ordinateur*. PhD thesis, IRCAM/Paris VI, Paris.
- [Arcos et al., 1997] Arcos, J. L., de Mántaras, R. L., e Serra, X. (1997). SaxEx: a case-based reasoning system for generating expressive musical performances. In Cook, P. R., editor, *ICMC*, pages 329–336. International Computer Music Association.
- [Arcos et al., 2003] Arcos, J. L., Grachten, M., e de Mántaras, R. L. (2003). Extracting performers' behaviours to annotate cases in a CBR system for musical tempo transformations. In Ashley, K. D. e Bridge, D. G., editors, *LNAI*, volume 2689, pages 20–34. Springer Verlag, Berlin.
- [Askenfelt, 1986] Askenfelt, A. (1986). Measurement of bow motion and bow force in violin playing. *Journal of the Acoustical Society of America*, 80:1007–1015.
- [Balaban et al., 1992] Balaban, M. et al., editors (1992). *Understanding Music with AI: Perspectives on Music Cognition*. AAAI Press e MIT Press, Cambridge (MA).
- [Bengtsson e Gabrielsson, 1977] Bengtsson, I. e Gabrielsson, A. (1977). Rhythm research in Uppsala. In *Music, Room and Acoustics*. Royal Swedish Academy of Music.
- [Beran e Mazzola, 2000] Beran, J. e Mazzola, G. (2000). Timing microstructure in Schumann's "Träumerei" as an expression of harmony, rhythm, and motivic structure in music performance. *Computers & Mathematics with Applications*, 39(5–6):99–130.
- [Boyer e Moore, 1977] Boyer, R. S. e Moore, J. S. (1977). A fast string searching algorithm. *Communications of the ACM*, 20(10):762–772.
- [Bresin, 2000] Bresin, R. (2000). *Vitual Virtuosity: Studies in Automatic Music Performance*. Doctoral dissertation, Kungliga Tekniska Högskolan (KTH), Stockholm, Sweden.
- [Cabral et al., 2005] Cabral, G., Pachet, F., e Briot, J.-P. (2005). Automatic x traditional descriptor extraction: The case of chord recognition. In *Proc. of the 6th International Conference on Music Information Retrieval (ISMIR'05)*.

- [Cabral et al., 2001] Cabral, G., Zanforlin, I., Lima, R., Santana, H., e Ramalho, G. L. (2001). Playing along with D'Accord Guitar. In *Anais do VIII Simpósio Brasileiro de Computação Musical (SBCM'01)*, pages 858–866, Fortaleza. Sociedade Brasileira de Computação (SBC).
- [Cambouropoulos, 1998] Cambouropoulos, E. (1998). *Towards a General Computational Theory of Musical Structure*. PhD thesis, The University of Edinburgh.
- [Cambouropoulos, 2000] Cambouropoulos, E. (2000). From MIDI to traditional musical notation. In *Proceedings of the AAAI'2000 Workshop on Artificial Intelligence and Music*, Menlo Park (CA). 17th National Conference on Artificial Intelligence (AAAI'2000), Austin, TX, AAAI Press.
- [Cambouropoulos, 2001a] Cambouropoulos, E. (2001a). Automatic pitch spelling: From numbers to sharps and flats. In *Proceedings of the 8th Brazilian Symposium on Computer Music*, pages 795–803, Fortaleza. Sociedade Brasileira de Computação (SBC).
- [Cambouropoulos, 2001b] Cambouropoulos, E. (2001b). The Local Boundary Detection Model (LBDM) and its application in the study of expressive timing. In *ICMC*, San Francisco (CA). International Computer Music Association.
- [Cambouropoulos et al., 1999] Cambouropoulos, E., Crawford, T., e Iliopoulos, C. S. (1999). Pattern processing in melodic sequences: Challenges, caveats & prospects. In *Proceedings of the AISB'99 Convention (Artificial Intelligence and Simulation of Behaviour)*, Edinburgh.
- [Cambouropoulos e Widmer, 2000] Cambouropoulos, E. e Widmer, G. (2000). Automatic motivic analysis via melodic clustering. *Journal of New Music Research*, 29(4).
- [Canazza et al., 1999] Canazza, S., De Poli, G., Di Federico, R., Drioli, C., e Rodà, A. (1999). Symbolic and audio processing to change the expressive intention of a recorded music performance. In *Proceedings of the second COST G-6 Workshop on Digital Audio Effects (DAFx'99)*, pages 1–4, Trondheim, Norway.
- [Canazza et al., 2003] Canazza, S., De Poli, G., Mion, L., Rodà, A., Vidolin, A., e Zanon, P. (2003). Expressive classifiers at CSC: an overview of the main research streams. In *Proceedings of the XIV Colloquium on Musical Informatics*, pages 64–68, Florença.
- [Carlson et al., 1989] Carlson, R., Friberg, A., Frydén, L., Granström, B., e Sundberg, J. (1989). Speech and music performance: Parallels and contrasts. *Contemporary Music Review*, 4:391–404.
- [Cemgil et al., 2000a] Cemgil, A. T., Desain, P., e Kappen, B. (2000a). Rhythm quantization for transcription. *Computer Music Journal*, 24(2):60–76.
- [Cemgil et al., 2000b] Cemgil, A. T., Kappen, B., Desain, P., e Honing, H. (2000b). On tempo tracking: Tempogram representation and Kalman filtering. In *ICMC*, pages 352–355, San Francisco (CA). ICMA.

- [Chediak, 1990] Chediak, A., editor (1990). *Songbook: Bossa Nova*, volume 1–5. Lumiar Editora, Rio de Janeiro.
- [Clarke, 1982] Clarke, E. F. (1982). Timing in the performance of Erik Satie’s ‘Vexations’. *Acta Psychologica*, 50(1):1–19.
- [Clarke, 1985] Clarke, E. F. (1985). Some aspects of rhythm and expression in performances of Erik Satie’s “Gnossienne no. 5”. *Music Perception*, 2(3):299–328.
- [Clarke, 1988] Clarke, E. F. (1988). Generative principles in music performance. In Sloboda, J. A., editor, *Generative Processes in Music: The psychology of performance, improvisation, and composition*, chapter 1, pages 1–26. The Clarendon Press, Oxford.
- [Clarke, 1989] Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research*, 51(1):2–9.
- [Conklin e Anagnostopoulou, 2001] Conklin, D. e Anagnostopoulou, C. (2001). Representation and discovery of multiple viewpoint patterns. In *ICMC*, pages 479–485, Havana, Cuba. International Computer Music Association.
- [Cook, 1994] Cook, N. (1994). *A Guide to Musical Analysis*. Oxford University Press, Oxford.
- [Cooper e Meyer, 1960] Cooper, G. e Meyer, L. B. (1960). *The Rhythmic Structure of Music*. University of Chicago Press, Chicago.
- [Cormen et al., 2002] Cormen, T. H., Leiserson, C. E., Rivest, R. L., e Stein, C. (2002). *Algoritmos: Teoria e Prática*. Campus, Rio de Janeiro.
- [Crawford et al., 1998] Crawford, T., Iliopoulos, C. S., e Raman, R. (1998). String-matching techniques for musical similarity and melodic recognition. In Hewlett, W. e Selfridge-Field, E., editors, *Melodic similarity: Concepts, procedures, and applications*, volume 11 of *Computing in Musicology*, pages 73–100. MIT Press, Cambridge, MA.
- [Crochemore, 1981] Crochemore, M. (1981). An optimal algorithm for computing the repetitions in a word. *IPL*, 12(5):244–250.
- [Dahia, 2004] Dahia, M. (2004). Gerando acompanhamento rítmico automático para violão: o estudo de caso Cyber-João. Master’s thesis, Centro de Informática, Universidade Federal de Pernambuco, Recife. 102 páginas.
- [Dahia et al., 2004] Dahia, M., Santana, H., Trajano, E., Sandroni, C., Ramalho, G. L., e Cabral, G. (2004). Using patterns to generate rhythmic accompaniment for guitar. In *Proc. of Sound and Music Computing (SMC’04)*, pages 111–115.
- [Dannenberg, 2000] Dannenberg, R. B. (2000). Artificial intelligence, machine learning and music understanding. In *Anais do VII Simpósio Brasileiro de Computação Musical, XX Congresso Nacional da Sociedade Brasileira de Computação—SBC*, volume 1. Champagnat.

- [de Mántaras e Arcos, 2002] de Mántaras, R. L. e Arcos, J. L. (2002). AI and Music: From composition to expressive performance. *AI Magazine*, pages 43–57.
- [de Ulhôa, 2003] de Ulhôa, M. T. (2003). *Isto é bom! ou Yayá, você quer morrer? – a tradição oral e a tradição escrita no lundu*. In *Anais do XIV Encontro Nacional da ANPPOM*.
- [Desain e Honing, 1999] Desain, P. e Honing, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, 28(1):29–42.
- [Desain e Honing, 2003] Desain, P. e Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3):341–365.
- [Desain et al., 2001] Desain, P., Honing, H., e Timmers, R. (2001). Music performance panel: NICI/MMM position statement. In *MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, Spain.
- [Desain et al., 1998] Desain, P., Honing, H., van Thienen, H., e Windsor, L. (1998). Computational modeling of music cognition: Problem or solution? *Music Perception*, pages 151–166.
- [Dixon, 2000] Dixon, S. (2000). A lightweight multi-agent musical beat tracking system. In *Pacific Rim International Conference on Artificial Intelligence*, pages 778–788.
- [Dixon, 2001] Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58.
- [Dixon, 2005] Dixon, S. (2005). Live tracking of musical performances using on-line time warping. In *Proc. of the 8th International Conference on Digital Audio Effects (DAFx'05)*, Madrid.
- [Dixon et al., 2002a] Dixon, S., Goebel, W., e Widmer, G. (2002a). The performance worm: Real time visualisation of expression based on Langner's tempo-loudness animation. In *ICMC*, pages 361–364, Göteborg, Sweden. International Computer Music Association (ICMA).
- [Dixon et al., 2002b] Dixon, S., Goebel, W., e Widmer, G. (2002b). Real time tracking and visualization of musical expression. In *Proceedings of the 2nd International Conference on Music and Artificial Intelligence (ICMAI'2002)*, pages 58–68, Edinburgh, Scotland.
- [Dixon et al., 2004] Dixon, S., Gouyon, F., e Widmer, G. (2004). Towards characterization of music via rhythmic patterns. In *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR'04)*, pages 509–516, Barcelona, Spain.
- [Dixon et al., 2003] Dixon, S., Pampalk, E., e Widmer, G. (2003). Classification of dance music by periodicity patterns. In *Proc. of the 4th International Conference on Music Information Retrieval (ISMIR'03)*, Washington, D.C.

- [Dixon e Widmer, 2005] Dixon, S. e Widmer, G. (2005). MATCH: A music alignment tool chest. In *Proc. of the 6th International Conference on Music Information Retrieval (ISMIR'05)*, London.
- [Ericsson et al., 1993] Ericsson, K. A., Krampe, R. T., e Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363–406.
- [Friberg, 1995] Friberg, A. (1995). *A Quantitative Rule System for Musical Performance*. PhD thesis, Royal Institute of Technology (KTH), Estocolmo, Suécia.
- [Friberg, 2005] Friberg, A. (2005). A fuzzy analyzer of emotional expression in music performance and body motion. In Sundberg, J. e Brunson, B., editors, *Proceedings of Music and Music Science (MMS'04)*.
- [Friberg et al., 2002] Friberg, A., Schoonderwaldt, E., Juslin, P. N., e Bresin, R. (2002). Automatic real-time extraction of musical expression. In *Proc. of the International Computer Music Conference (ICMC)*, pages 365–367, San Francisco. International Computer Music Association (ICMA).
- [Friberg e Sundberg, 1999] Friberg, A. e Sundberg, J. (1999). Does music performance allude to locomotion? a model of final *Ritardandi* derived from measurements of stopping runners. *Journal of the Acoustical Society of America*, 105(3):1469–1484.
- [Gabrielsson, 1974] Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology*, 15:63–72.
- [Gamma et al., 2000] Gamma, E., Helm, R., Johnson, R., e Vlissides, J. (2000). *Padrões de Projeto: Soluções Reutilizáveis de Software Orientado a Objetos*. Bookman, Porto Alegre.
- [Garcia, 1999] Garcia, W. (1999). *Bim Bom: A Contradição sem Conflitos de João Gilberto*. Editora Guerra e Paz.
- [Gava, 2002] Gava, J. E. (2002). *A Linguagem Harmônica da Bossa Nova*. Editora Unesp, São Paulo.
- [Goto e Muraoka, 1995] Goto, M. e Muraoka, Y. (1995). Beat tracking based on multiple-agent architecture: A real-time beat tracking system for audio signals. In Lesser, V., editor, *Proceedings of the First International Conference on Multi-Agent Systems—ICMAS'96*, pages 103–110. MIT Press.
- [Grachten et al., 2006] Grachten, M., Arcos, J. L., e de Mántaras, R. L. (2006). A case based approach to expressivity-aware tempo transformation. *Machine Learning*, 65(2–3):411–437.
- [Guigue, 1997] Guigue, D. (1997). *Une étude “pour les Sonorités Opposées”: Pour une analyse orientée objets de l’œuvre pour piano de Debussy et de la musique du XXe siècle*. Presses Universitaires du Septentrion, Villeneuve d’Ascq.

- [Guigue, 2003] Guigue, D. (2003). *Sonic Object Analysis Library—User’s Manual*. GMT—Grupo de Música, Musicologia e Tecnologia Aplicada, João Pessoa.
- [Gusfield, 1997] Gusfield, D. (1997). *Algorithms on Strings, Trees, and Sequences*. Cambridge University Press.
- [Han e Kamber, 2001] Han, J. e Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufman, San Francisco.
- [Hazan et al., 2006] Hazan, A., Grachten, M., e Ramirez, R. (2006). Evolving performance models by performance similarity: Beyond note-to-note transformations. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR’06)*, Victoria, Canada.
- [Hazan e Ramirez, 2006] Hazan, A. e Ramirez, R. (2006). Modelling expressive performance using consistent evolutionary regression trees. In *Proc. of ECAI Workshop on Evolutionary Computation*, Siena, Italy.
- [Henderson, 1936] Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In Seashore, C. E., editor, *Psychology of Music*, pages 281–305. McGraw-Hill, New York.
- [Honing, 2001] Honing, H. (2001). From time to time: The representation of timing and tempo. *Computer Music Journal*, 25(3).
- [Howell et al., 1991] Howell, P., West, R., e Cross, I., editors (1991). *Representing Musical Structure*. Academic Press, London.
- [Hsu et al., 1998] Hsu, J.-L., Liu, C.-C., e Chen, A. L. P. (1998). Efficient repeating pattern finding in music databases. In *Proceedings of the seventh International Conference on Information and Knowledge Management*, pages 281–288. ACM Press.
- [Huron, 1999] Huron, D. (1999). *Music Research Using Humdrum: A User’s Guide*. CCARH, Menlo Park. Disponível em <http://www.music-cog.ohio-state.edu/Humdrum/guide.toc.html>. Último acesso 12 dez. 2003.
- [Kendall e Carterette, 1990] Kendall, R. A. e Carterette, E. C. (1990). The communication of musical expression. *Music Perception*, 8:129–164.
- [Kerman, 1987] Kerman, J. (1987). *Musicologia*. Coleção Opus-86. Martins Fontes, São Paulo.
- [Kernfeld, 1983] Kernfeld, B. (1983). Two Coltranes. *Annual Review of Jazz Studies*, 20:7–66.
- [Knuth et al., 1977] Knuth, D. E., Morris Jr, J. H., e Pratt, V. R. (1977). Fast pattern matching in strings. *SICOMP*, 6(2):323–350.

- [Kolodner, 1993] Kolodner, J. B. (1993). *Case-based Learning*. Kluwer Academic Publishers, Boston.
- [Langley, 1996] Langley, P. (1996). *Elements of Machine Learning*. Morgan Kaufmann, San Francisco.
- [Langley e Simon, 1995] Langley, P. e Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11):55–64.
- [Langner e Goebel, 2002] Langner, J. e Goebel, W. (2002). Representing expressive performance in tempo-loudness space. In *ESCOM 10th anniversary conference on Musical Creativity*, Liège, Belgium.
- [Large e Palmer, 2002] Large, E. W. e Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26(1):1–37.
- [Lemström, 2000] Lemström, K. (2000). *String Matching Techniques for Music Retrieval*. PhD thesis, University of Helsinki, Department of Computer Science.
- [Lemström et al., 1998] Lemström, K., Haapaniemi, A., e Ukkonen, E. (1998). Retrieving music—to index or not to index. In *Proc. Art Demos - Technical Demos - Poster Papers - The Sixth ACM International Multimedia Conference (MM '98)*, pages 64–65, Bristol.
- [Lemström et al., 1999] Lemström, K., Laine, P., e Perttu, S. (1999). Using relative interval slope in music information retrieval. In *ICMC*, pages 317–320, Beijing, China.
- [Lemström e Tarhio, 2000] Lemström, K. e Tarhio, J. (2000). Searching monophonic patterns within polyphonic sources. In *Proceedings of Content-Based Multimedia Information Access (RIAO'2000)*, volume 2, pages 1261–1279, Paris, France.
- [Lemström e Ukkonen, 2000] Lemström, K. e Ukkonen, E. (2000). Including interval encoding into edit distance based music comparison and retrieval. In *Proceedings of the Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science*, pages 53–60, Birmingham, United Kingdom.
- [Lemström et al., 2001] Lemström, K., Wiggins, G. A., e Meredith, D. (2001). A three-layer approach for music retrieval in large databases. In *Proceedings ISMIR 2001 2nd Annual International Symposium on Music Information Retrieval*, pages 13–14, Bloomington, Indiana.
- [Lerdahl e Jackendoff, 1983] Lerdahl, F. e Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, MA.
- [Levenshtein, 1965] Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions and reversals. *Cybernetics and Control Theory*, 10(8):707–710.
- [Madsen e Widmer, 2005] Madsen, S. T. e Widmer, G. (2005). Evolutionary search for musical parallelism. In *Proc. of the EvoWorkshops2005*, Lecture Notes in Computer Science 3449, pages 488–497. Springer Verlag.

- [Madsen e Widmer, 2006a] Madsen, S. T. e Widmer, G. (2006a). Exploring pianist performance styles with evolutionary string matching. *International Journal on Artificial Intelligence Tools, (Special Issue on Artificial Intelligence in Music and Art)*, 15(4):495–513.
- [Madsen e Widmer, 2006b] Madsen, S. T. e Widmer, G. (2006b). Separating voices in MIDI. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR'06)*.
- [Madsen e Widmer, 2007a] Madsen, S. T. e Widmer, G. (2007a). A complexity-based approach to melody track identification in MIDI files. In *Proc. of the Intern. Workshop on Artificial Intelligence and Music (MUSIC-AI'07) held in conjunction with the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, page to appear, Hyderabad, India.
- [Madsen e Widmer, 2007b] Madsen, S. T. e Widmer, G. (2007b). Towards a computational model of melody identification in polyphonic music. In *Proc. of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07)*, page to appear, Hyderabad, India.
- [Meredith et al., 2001] Meredith, D., Lemström, K., e Wiggins, G. A. (2001). SIA and SIATEC: Efficient algorithms for translation-invariant pattern discovery in multi-dimensional datasets. Document submitted to UK Patent Office, City University, School of Informatics, Department of Computing. Available online at <http://www.titanmusic.com/papers/public/patent.pdf>.
- [Meredith et al., 2002] Meredith, D., Lemström, K., e Wiggins, G. A. (2002). Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, 31(4):321–345.
- [Meudic, 2000] Meudic, B. (2000). *OpenMusic: Bibliothèque Kant*. Paris.
- [Mitchell, 1997] Mitchell, T. (1997). *Machine Learning*. McGraw Hill.
- [Mongeau e Sankoff, 1990] Mongeau, M. e Sankoff, D. (1990). Comparison of musical sequences. *Computer and the Humanities*, 24:161–175.
- [Nakamura, 1987] Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics*, 41(6):525–533.
- [Napolitano, 2003] Napolitano, M. (2003). O fonograma como fonte para a pesquisa histórica sobre música popular: Problemas e perspectivas. In *Anais do XIV Encontro Nacional da ANPPOM*.
- [Narmour, 1990] Narmour, E. (1990). *The Analysis of Cognition of Basic Melodic Structures: The Implication-Realization Model*. The Chicago University Press, Chicago.

- [Nettl, 1983] Nettl, B. (1983). *The Study of Ethnomusicology: Twenty-nine Issues and Concepts*. University of Illinois Press, Urbana and Chicago.
- [Owens, 1974] Owens, T. (1974). *Charlie Parker: Techniques of Improvisation, 2 vols.* PhD thesis, Univeristy of Los Angeles California.
- [Pachet et al., 1996] Pachet, F., Ramalho, G. L., e Carrive, J. (1996). Representing temporal musical objects and reasoning in the MusES system. *Journal of New Music Research*, 5(3):252–275.
- [Palmer, 1988] Palmer, C. (1988). *Timing in Skilled Piano Performance*. PhD thesis, Cornell University.
- [Palmer, 1989] Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):331–346.
- [Palmer, 1997] Palmer, C. (1997). Music performance. *Annual Review of Psychology*, 48:115–138.
- [Pampalk et al., 2004] Pampalk, E., Dixon, S., e Widmer, G. (2004). Exploring music collections by browsing different views. *Computer Music Journal*, 28(2):49–62.
- [Pardo e Birmingham, 1999] Pardo, B. e Birmingham, W. (1999). Automated partitioning of tonal music. Technical report, Electrical Engineering and Computer Science Department, University of Michigan.
- [Pereira, 2003] Pereira, S. L. (2003). Música e memória: o trabalho com as tramas memorialísticas dos ouvintes da Bossa Nova. In *Anais do XIV Encontro Nacional da ANPPOM*.
- [Perttu, 2000] Perttu, S. (2000). Combinatorial pattern matching in musical sequences. Master’s thesis, Department of Computer Science, Univeristy of Helsinki.
- [Pickens e Crawford, 2002] Pickens, J. e Crawford, T. (2002). Harmonic models for polyphonic music retrieval. In *Proceedings of the eleventh International Conference on Information and Knowledge Management*, pages 430–437. ACM Press.
- [Pyle, 1999] Pyle, D. (1999). *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, San Francisco.
- [Quinlan, 1986] Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1:81–106.
- [Quinlan, 1993] Quinlan, J. R. (1993). *Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo.
- [Ramalho, 1997] Ramalho, G. L. (1997). *Construction d’un Agent Rationnel Jouant du Jazz*. PhD thesis, Universidade Paris VI.

- [Ramirez e Hazan, 2004] Ramirez, R. e Hazan, A. (2004). Rule induction for expressive music performance modeling. In *Proc. of Advances in Inductive Rule Learning Workshop*, Pisa, Italy.
- [Ramirez et al., 2004] Ramirez, R., Hazan, A., Gómez, E., e Maestre, E. (2004). Understanding expressive transformations in saxophone jazz performances using inductive machine learning. In *Proc. of Sound and Music Computing (SMC'04)*, Paris, France.
- [Rasch, 1979] Rasch, R. (1979). Synchronization in performed ensemble music. *Acustica*, 43:121–131.
- [Repp, 1992] Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, 44(3):241–281.
- [Repp, 1998a] Repp, B. H. (1998a). A microcosm of musical expression: I. quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 104:1085–110.
- [Repp, 1998b] Repp, B. H. (1998b). Perception and production of staccato articulation on the piano (unpublished manuscript).
- [Repp, 1999a] Repp, B. H. (1999a). Control of expressive and metronomic timing in pianists. *Journal of Motor Behavior*, 31(2):145–164.
- [Repp, 1999b] Repp, B. H. (1999b). A microcosm of musical expression: I. quantitative analysis of pianists' dynamics in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, 105:1972–1988.
- [Repp et al., 2002] Repp, B. H., Windsor, W. L., e Desain, P. (2002). Effects of tempo on the timing of simple musical rhythms. *Music Perception*, 19(4):565–593.
- [Rich e Knight, 1993] Rich, E. e Knight, K. (1993). *Inteligência Artificial*. Makron Books, São Paulo.
- [Rinman et al., 2004] Rinman, M.-L., Friberg, A., Bendiksen, B., Cirotteau, D., Dahl, S., Kjellmo, I., Mazzarino, B., e Camurri, A. (2004). Ghost in the cave—an interactive collaborative game using non-verbal communication. In Camurri, A. e Volpe, G., editors, *Gesture-based Communication in Human-Computer Interaction*, volume 2915 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 549–556, Berlin. Springer Verlag.
- [Roads et al., 1996] Roads, C. et al. (1996). *The Computer Music Tutorial*. The MIT Press, Cambridge (MA).
- [Rolland, 1999] Rolland, P.-Y. (1999). Discovering patterns in musical sequences. *Journal of New Music Research*, 28(4):334–350.

- [Rolland, 2001] Rolland, P.-Y. (2001). FExPat: Flexible extraction of sequential patterns. In Cercone, N., Lin, T. Y., e Wu, X., editors, *Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01)*, pages 481–488. IEEE Computer Society.
- [Rosen, 1972] Rosen, C. (1972). *The Classical Style: Haydn, Mozart, Beethoven*. The Norton Library. W. W. Norton & Company, New York.
- [Sadakata et al., 2002] Sadakata, M., Desain, P., e Honing, H. (2002). The relation between rhythm perception and production: towards a bayesian model. *Transaction of Technical Committee of Psychological and Physiological Acoustics*, 32(10).
- [Sandroni, 1988] Sandroni, C. (1988). O olhar do aprendiz: Observações sobre a prática do violão popular no brasil, fartamente documentada por exemplos em partitura. Unpublished manuscript.
- [Sandroni, 1996] Sandroni, C. (1996). Mudanças de padrão rítmico no samba carioca, 1917-1937. *Revista Transcultural de Música*, 2. Available online at <http://www.sibetrans.com/trans/trans2/sandroni.htm>.
- [Sankoff e Kruskal, 1999] Sankoff, D. e Kruskal, J., editors (1999). *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. CSLI Publications.
- [Sayegh, 1991] Sayegh, S. L. (1991). Fingering for string instruments with the optimum path paradigm. In Todd, P. M. e Loy, D. G., editors, *Music and Connectionism*, pages 243–255, Cambridge (MA). The MIT Press.
- [Schenker, 1954] Schenker, H. (1954). *Harmony*. University of Chicago Press, London.
- [Schönberg, 1922] Schönberg, A. (1922). *Harmonielehre*. Universal Edition, Wien.
- [Shaffer et al., 1985] Shaffer, L. H., Clarke, E. F., e Todd, N. P. (1985). Metre and rhythm in piano playing. *Cognition*, 20(1):61–77.
- [Shmulevich e Povel, 2000] Shmulevich, I. e Povel, D.-J. (Feb. 2000). Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1):61–69.
- [Sloboda, 2000] Sloboda, J. A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences*, 4(10):397–403.
- [Sloboda et al., 1996] Sloboda, J. A., Davidson, J. W., Howe, M. J. A., e Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, 87(2):287–309.
- [Sloboda e Lehmann, 2001] Sloboda, J. A. e Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception*, 19(1):87–120.

- [Smith, 1983] Smith, G. (1983). *Homer, Gregory, and Bill Evans? The Theory of Formulaic Composition in the Context of Jazz Piano Improvisation*. PhD thesis, Harvard University.
- [Sundberg et al., 2001] Sundberg, J., Friberg, A., e Bresin, R. (2001). Music performance panel: Position statement. In *MOSART Workshop on Current Research Directions in Computer Music*, Barcelona, Spain.
- [Sundberg et al., 2003] Sundberg, J., Friberg, A., e Bresin, R. (2003). Attempts to reproduce a pianist's expressive timing with Director Musices performance rules. *Journal of New Music Research*, 32(3):317–326.
- [Sundberg et al., 1991] Sundberg, J., Friberg, A., e Frydén, L. (1991). Common secrets of musicians and listeners: an analysis-by-synthesis study of musical performance. In Howell, P., West, R., e Cross, I., editors, *Representing Musical Structure*, pages 161–197. Academic Press.
- [Sundberg e Lindblom, 1991] Sundberg, J. e Lindblom, B. (1991). Generative theories for describing musical structure. In Howell, P., West, R., e Cross, I., editors, *Representing Musical Structure*, Cognitive Science Series, pages 245–272. Academic Press, London.
- [Timmers et al., 2000] Timmers, R., Ashley, R., Desain, P., e Heijink, H. (2000). The influence of musical context on tempo rubato. *Journal of New Music Research*, 29(2):131–158.
- [Trajano et al., 2004] Trajano, E., Dahia, M., Santana, H., e Ramalho, G. L. (2004). Automatic discovery of right hand fingering in guitar accompaniment. In *Proc. of the International Computer Music Conference (ICMC)*, pages 722–725.
- [Trajano et al., 2005] Trajano, E., Madsen, S. T., Dahia, M., Widmer, G., e Ramalho, G. L. (2005). Extracting patterns from guitar accompaniment data: Some experimental results. In de Paula, H. B. e Yehia, H. C., editors, *Proc. of the X Brazilian Symposium on Computer Music (SBCM'05)*, pages 154–163, Belo Horizonte.
- [Trajano et al., 2006] Trajano, E., Madsen, S. T., Dahia, M., Widmer, G., e Ramalho, G. L. (2006). Extracting patterns from brazilian guitar accompaniment data. In *Proceedings of the 1st European Workshop on Intelligent Technologies for Cultural Heritage Exploitation at the 17th European Conference on Artificial Intelligence (ECAI'06)*, pages 50–54, Riva del Garda, Italy.
- [Trotta, 2003] Trotta, F. (2003). O acorde-tempestade: Considerações sobre a utilização do IV⁷ no repertório do samba. In *Anais do XIV Encontro Nacional da ANPPOM*.
- [Viterbi, 1967] Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *TIT*, 13(2):260–269.

- [Wen e Sandler, 2005] Wen, W. e Sandler, M. (2005). A partial searching algorithm and its application for polyphonic music transcription. In *Proc. of the 6th International Conference on Music Information Retrieval (ISMIR'05)*, pages 690–695, London, UK.
- [West et al., 1991] West, R., Howell, P., e Cross, I. (1991). Musical structure and knowledge representation. In [Howell et al., 1991], pages 1–30.
- [Widmer, 1994] Widmer, G. (1994). The synergy of music theory and AI: Learning multi-level expressive interpretation. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), Seattle (WA)*, pages 114–119, Menlo Park (CA). AAAI Press.
- [Widmer, 1998] Widmer, G. (1998). Applications of machine learning to music research: Empirical investigations into the phenomenon of musical expression. In Michalski, R. S., Bratko, I., e Kubat, M., editors, *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*. Wiley & Sons, Chichester (UK).
- [Widmer, 2000] Widmer, G. (2000). Large-scale induction of expressive performance rules: First quantitative results. In *ICMC, San Francisco (CA)*. International Computer Music Association.
- [Widmer, 2001a] Widmer, G. (2001a). Discovering strong principles of expressive music performance with the PLCG rule learning strategy. In *Proceedings of the 11th European Conference on Machine Learning (ECML'01)*, Berlin. Springer Verlag.
- [Widmer, 2001b] Widmer, G. (2001b). The musical expression project: A challenge for machine learning and knowledge discovery (invited talk). In *Proceedings of the 12th European Conference on Machine Learning (ECML'2001)*, Berlin. Springer Verlag.
- [Widmer, 2002a] Widmer, G. (2002a). In search of the Horowitz factor: Interim report on a musical discovery project. invited paper. In *Proceedings of the 5th International Conference on Discovery Science (DS'02), Lübeck, Germany*, Berlin. Springer Verlag.
- [Widmer, 2002b] Widmer, G. (2002b). Machine discoveries: A few simple, robust local expression principles. *Journal of New Music Research*, 31(1):27–50.
- [Widmer, 2003] Widmer, G. (2003). Discovering simple rule in complex data: A meta-learning algorithm and some surprising musical discoveries. *Artificial Intelligence*, 146(2):129–148.
- [Widmer et al., 2003] Widmer, G., Dixon, S., Goebel, W., Pampalk, E., e Tobudic, A. (2003). In search of the Horowitz factor. *AI Magazine*, 24(3):111–130.
- [Widmer e Tobudic, 2002] Widmer, G. e Tobudic, A. (2002). Playing mozart by analogy: Learning multi-level timing and dynamics strategies. Technical Report TR-2002-23, Österreichisches Forschungsinstitut für Artificial Intelligence, Viena.

- [Windsor et al., 2000] Windsor, W. L., Desain, P., Honing, H., Aarts, R., Heijink, H., e Timmers, R. (2000). On time: The influence of tempo, structure and style on the timing of grace notes in skilled musical performance. In Desain, P. e Windsor, W. L., editors, *Rhythm Perception and Production*, pages 217–223. Swets & Zeitlinger, Lisse.
- [Witten e Frank, 2000] Witten, I. H. e Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco.
- [Yoshioka et al., 2004] Yoshioka, T., Kitahara, T., Komatani, K., Ogata, T., e Okuno, H. G. (2004). Automatic chord transcription with concurrent recognition of chord symbols and boundaries. In *Proc. of the 5th International Conference on Music Information Retrieval (ISMIR'04)*, Barcelona, Spain.
- [Zanon e Widmer, 2003a] Zanon, P. e Widmer, G. (2003a). Learning to recognize famous pianists with machine learning techniques. In *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03)*, pages 581–584, Stockholm, Sweden.
- [Zanon e Widmer, 2003b] Zanon, P. e Widmer, G. (2003b). Recognition of famous pianists using machine learning algorithms: First experimental results. In *Proceedings of the XIV Colloquium on Musical Informatics*, pages 84–89, Florence.

APÊNDICE A

CATÁLOGO DE PADRÕES

Neste Apêndice são apresentados os padrões que integram o catálogo utilizado na avaliação dos experimentos realizados. Nesta apresentação, os padrões seguem a mesma organização do catálogo original [Dahia, 2004], ou seja, reunidos de acordo com sua função estrutural na canção, o que resultou em 5 grupos diferentes. Apresenta-se também a transcrição do padrão para a representação utilizada durante a pesquisa.

Padrão Principal

P1	A---P---B-p-+--- A---P---B-p-+---	
-----------	-----------------------------------	--

Padrões Cíclicos

P2	1- B---P---B-p-+-1- B---P---B-p-+-1-	
P3	1- B---P---B-p-+--- A---P---B-p-+-1-	
P4	1- B-p-+-p-B---P--- A---P---B-p-+-1-	
P5	1- B-p-+-p-B-p-+--- A---P---B-p-+-1-	

P6	1- B-p+----A---P--- A---P---B-p+-1-	
P7	1- B----+p-B---P--- A---P---B-p+-1-	
P8	1- B----+p-B-p+---- A---P---B-p+-1-	

Padrões de Início de Música

P9	A---P---B-p+-1- B---P---B-p+-1-	
P10	A---P---B-p+---- A---P---B-p+-1-	
P11	A----+p-B---P--- A---P---B-p+-1-	
P12	A----+p-B-p+---- A---P---B-p+-1-	

Padrões Especiais

P13	1- B----+p-B----+1- B----+p-B----+1-	
------------	--------------------------------------	--

<p>P14</p>	<p>1- B-p-+-p-B-p-+-l- B-p-+-p-B-p-+-l-</p>	
-------------------	---	--

Padrões de Virada

<p>P15</p>	<p> A---P---B-p-+--- A-p-+-p-B-p-+-l-</p>	
<p>P16</p>	<p>1- B---P---B-p-+--- A-p-+-p-B-p-+-l-</p>	
<p>P17</p>	<p>1- B-p-+-p-B---P--- A-p-+-p-B-p-+-l-</p>	
<p>P18</p>	<p>1- B-p-+-p-B-p-+--- A-p-+-p-B-p-+-l-</p>	
<p>P19</p>	<p>1- B-p-+---A---P--- A-p-+-p-B-p-+-l-</p>	
<p>P20</p>	<p>1- B---+-p-B---P--- A-p-+-p-B-p-+-l-</p>	
<p>P21</p>	<p>1- B---+-p-B-p-+--- A-p-+-p-B-p-+-l-</p>	

APÊNDICE B

SFLEXPAT: PADRÕES INDUZIDOS

Neste Apêndice são apresentados os padrões induzidos pelo algoritmo SFLEXPAT. Como nos experimentos foram utilizados apenas os protótipos e também por uma questão de espaço, apenas os protótipos de cada grafo estrela são apresentados. Os padrões estão agrupados em duas seções (Seções B.1 e B.2) correspondendo aos padrões encontrados para os intérpretes 1 e 2, respectivamente.

B.1 PADRÕES DO INTÉRPRETE 1

A Felicidade

1- B-s-+-p-B-s-+-1- B-s-+-p--B-p-+-1-	1- B-p-+----A---S-1- B---P---B-p-+-1-
1- B---P---B-p-+---- A---P---B-p-+-1-	1-- Bp-+----A---P--- A---+-p-B-p-+-1-
1- B-p-+----A---P--- A---P---B--p	1- B-p-+----A---+-1- B---P---B-p-+-1-
1-- B-p-+----A---P--- A---S-p-Bp-+-1--	P-s-+-p-B-s-+-1- B-s-+-p-B-s-+-1-
1- B-p-+----A---P--- A---+p--B-p-+-1-	1-- +-a-+----A---P--- A---P---B-p-+-1-
1- B-s-+-p-B-s-+-1- B-s-+-p-B-s-+-1-	1- B-p-+----A---P--- A---P---+-a-+-1-
1- B-p-+----A--p+---- A---P---B-p-+-1--	1-- Bp-+----A---S-p- A---P---B-p-+-1--
1- Bp-+----A---P-- A---P---B--p-+-1-	1-- B-p-+----A---P--- A---P---+-a-+-1--
1- B-p-+----A---P--- +a---P---B-p-+-1-	1- +bp-+----A---P--- A---P---B-p-+-1-
1-- Bp-+----A---P--- A---P---B-p-+p--	1-- Bp-+----A---P--- A---P---Bp-+-1-
1- Bp-+----A---P--- A---P---Bp-+-1-	1- B-p-+----A---P--- A---P---Bp-+-1--
1- B-p-+----A---P--- A---P---B-p-+-1--	1- B-p-+----A---P--- A---P---B-p-+-1--
1- B-p-+----A---P--- A---P---B-p-+-1-	1-- B-p-+----A---P--- A---P---B-p-+-1-
1-- B-p-+----A---P--- A---P---B-p-+-1-	1-- B-p-+----A---P--- A---P---B-p-+-1--
1- B-p-+----A---P--- A---P---B-p-+-1--	1- Bp-+----A---P--- A---P---B-p-+-1--
1-- B-p-+----A---P--- A---P---B-p-+-1-	1- B-p-+----A---P--- A---P---B-p-+-1-
1- B-p-+----A---P--- A---P---B-p-+-1-	1- B-p-+----A---P--- A---P---B-p-+-1--

Garota De Ipanema

1-- B---P---+bp+---- A---P---B-p+p--	A---P---Bp---+---- A---P---B-p++1--
A---P---Bp---+---- A---P---Bp---+1--	1-- B---P---Bp---+---- A---P---Bp---+1--
A---P---B-p+---- P---P---Bp---+1--	A---P---B-p+---- A---P---B-p++1-
1-- B---P---B-p+---- A---P---Bp---+1--	A---P---B-p+---- A---P---B-p+----
1-- B---A---B-p+---- A---P---B-p+1--	1-- B---+p-B-p+---- A---P---B-p+----
1- B---P---B-p+---- A---P---B-p+----	A---P---B-p+---- A---P---B-p++1-
A---P---B-p+---- A---P---B-p++1-	A---P---B-p+---- A---P---B-p+----
1-- B---P---B-p+---- A---P---B-p++1-	1- B---P---B-p+---- A---P---B-p+----
1- B---P---B-p+---- A---P---B-p++1-	1- B---P---B-p+---- A---P---B-p+----
1- B---P---B-p+---- A---P---B-p+----	/

Insensatez

1- B---+p-B-p+---- A---P---B-p++1-	1-- B---P---Bp---+---- A---P---+p+----
1- B---+p-B-p+---- A---P---B-p++1-	1- B---+p-B-p+---- A---P---B-p+----
1-- B---P---B-p+---- A---P---+p+1--	A---P---B-p+---- A---P---B-p+----
A---P---B-p+---- A---A---B-p++1-	A---P---B-p+---- A---P---B-p++1--
A---P---B-p+---- A---P---B-p+----	1- B---P---B-p+---- A---P---B-p++1-
1- B---P---B-p+---- A---P---Bp---+----	1-- B---P---B-p+---- A---P---B-p+----
A---P---Bp---+---- A---P---B-p+----	1- B---P---B-p+---- A---P---B-p+----
1- B---P---B-p+---- A---P---B-p+----	A---P---B-p+---- A---P---B-p++1--
A---P---B-p+---- A---P---B-p++1-	A---P---B-p+---- A---P---B-p++1-
A---P---B-p+---- A---P---B-p+----	A---P---B-p+---- A---P---B-p+----
A---P---B-p+---- A---P---B-p+----	/

Samba De Uma Nota Só

1- B---P---B-p+---- A---P---B-p++a-	A---P---Bp---+---- A---P---Bp---+1--
1- B---P---B-p+---- A-p++a+----P---	1- B---P---B-p+---- A---P---B-p++1-
1- B---P---B-p+---- A---P---B-p++1-	A---P---B-p+---- A---P---B-p++1-

Tarde Em Itapuã

1- B-p+-p-B---P--- A---P---B-p+-l-	1-- B---P---+p+--- A---P---Bp---+l-
1- B-p+---A---P--- P---S-s-B-p+-l-	1- B-p+-p-B-p+-l- B-p+-p-B---P---
1- B-p+---A---P--- A---P---Bp---+a-	1- B-p+---A---P--- A---+p-B-p+-l-
1-- B-p+---A---P--- A---P---B-p+-l-	1-- B-p+---A---P--- A---P---Bp---+l--
1- B-p+-p-B-p+-l- B-p+-p-B---P---	A---P---B-p+-l- B---P---B-p+-l-
B---P---B-p+-s- A---P---B-s+-l-	1- B-p+-p-B-p+-l- B-p+-p-B---Pb--
A---P---B-p+--- A---P---B-p+-l-	1- B-p+---A---P--- A---P---B-p+-l-
1- B---P---B-p+--- A-p+-p-B-p+-l-	1- B---P---B-p+--- B---P---B-p+-l-
+---P---B-p+--- A---P---B-p+-l-	1- B---P---B-p+--- A---P---Bp---+l-
1- B-p+---B-p+--- A---P---B-p+-l-	1-- B---P---B-p+--- A---P---B-p+-l-
1- B-p+---A---P--- A---P---B-p+-a-	1- B-p+---A---P--- A---P---B-p+-l-
+---P---B-p+--- A---P---B-p+-l-	1- B---P---B-p+--- A---P---B-p+a--
1- B---P---B-p+--- A---P---Bp---+l-	1-- B---P---B-p+--- A---P---B-p+-l-
1- B---P---B-p+--- A---P---B-p+-l-	1- B---P---B-p+--- A---P---B-p+-l-
1- B---P---B-p+--- A---P---B-p+-l-	/

Wave

1- B-p+-p-B-p+-l- B-p+-p-B---P---	A---P---B-p+--- A---P---Bp---+l--
1- B-p+---A---P--- A---P---B-p+l--	1-- B-p+---A---P--- A---P---B-p+l--
A---P---Bp---+--- A---P---Bp---+l--	1- B---P---Bp---+--- A---P---B-p+---
A---P---A---+l- B-p+-p-B-p+-l-	A---P---A---+l- B-p+-a+---P---
A---P---B-p+--- A---P---B-p+l--	A---P---A---+l- B-p+-p-B---P---
A---P---A---+l- B-p+-p-B---P---	1-- B---P---B-p+--- A---P---B-p+l--
1- B---P---B-p+--- A---P---B-p+-l-	1-- B---P---B-p+--- A---P---B-p+p--

B.2 PADRÕES DO INTÉRPRETE 2

Barquinho

1-- B---P---B-p+-l- B---P---B-p+-l-	A---P---B-p+--- A---P---B-p+l--
A---P---B-p+-l- B---P---B-p+-l-	A---P---B-p+-l- B---P---B-p+-b-
A---P---B-p+l-- B---P---B-p+-s-	A---P---B-p+-l- B---P---B-p+-b-
A---P---B-p+-l- B---P---B-p+-l-	1- B---P---B-p+-l- B---P---B-p+-s-

Bim Bom

A---P---A---S-l- B-p-+-p-B-p-+-l-	l- B-p-+-p-B-p-+-l-- B-p-+-p-B-p-+-l-
A---P-b-A---+-l- B-p-S-p-B-p-B-s-	A---P-b-A-p-S-l- B-p-S-p-Bp-+-l--
A---A---A---S-l- B-p-+-p-B-p-B-s-	A---P-b-A---S-l- B-p-+-p-B-p-+-l-
A---P-b-A---S-l- B-p-S-p-B-p-B-s-	A---P-b-A-p-S-l- B-p-+-p-B-p-B-s-
A---P-b-A---s+-l- B-p-+-p-B-p-+-l-	/

Garota De Ipanema

B-p-+p-bA-p-+-l- B---P---B-p-+-s-	l- B---P---bB-p-+-s- A---P---B-p-+---
l- B---P---B-p-+--- A---P---B-p-+-l-	l-- B---P---bB-p-+-s- A---P---B-p-+-l-
l- B---P---B-p-+-s- A---P---B-p-+-l-	A---P---B-p-+-s- A---P---B-p-+-l-
A---+---+---+--- A---+---a+---+---	A---+---+---+--- A---+---a+---+---

Insensatez

A---P---B-p-+-s- A-p-+-p-B-p-B---	A---A---B-p-+-l- B---P---B-p-+-s-
A---P---B-p-+b-- A-p-+-p-B-p-+-s-	A---P---B-p-+-s- A-a-+-p-B-p-+-s-
P--p+-a-A---a+---a- B---P---B--p-B-s-	A---P---B-p-+-s- A-p-+-p-B-p-+-s-
A---P---B-p-+-s-- A---P---B-p-B-l-	A---P---B--p-+-s- A---P---B-p-+-s-
A---P---B-p-+--- A---P---B-p-+-s--	A---A---A---S-a- +a-+-a-+-a-+-s-
A---P---B-p-+--- A---P---B-p-+---	A---P---B-p-+-s- A---+---p-B-p-+-s-
A---P---B-p-+-s- A---P---B-p-+-s--	A---P---B-p-+--- A---P---B-p-+---
A---P---B-p-+-s- A---P---B-p-+-s-	A---P---B-p-+-s- A---P---B-p-+-s-
A---A---A---S-a- +a-+-a-+-a-+-s-	A---P---B-p-+-s- A---P---B-p-+---
A---P---B-p-+-s- A---P---B-p-+-s-	A---A---A---S-a- +a-+-a-+-a-+---
A---A---A---S-a- +a-+-a-+-a-+-s-	A---A---A---S-a- +a-+-a-+-a-+-s-

Só Danço Samba

l- B-p-+-s-A---P-b- A-p-P-p-Bp-+-l--	l- B-p-+-s-A---P-b- A-p-+-p-B-p-B--l
l-- B-p-+-s-A---Pb-- A-p-S-p-B-p-+-l-	A---P---A---+l-- B-p-+-p-Bp-+-p--
l- B-p-+-b-A---P--- A-p-P-p-B-p-+-l-	B-p-+-s-A---Psb- A-p-P-p-B-p-+-l-
A---P---A---+-l- B-p-+-p-B-p-B---	l- B-p-+-p-B-p-+-l- B-p-+-p-B--pB-s-
l- B-p-+-p-B-p-+-l- B--p+---p-B-p-B-s-	/

Wave

A-p-+-p-B-p-+-l- B-p-+-p-B-p-+-l-	l- B---P---B-p-+-l- B-p-+-p-B-p-+-l-
l- B---P---B-p-+-s- A-p-+-p-B-p-S-s-	A---P---B-p-Bl-- B---P---B-p-+-l--
A---A---A---S-a- +a-+-a-+-a-+---	A---A---A---S-a- +a-+-a-+-a-+---