

Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

**Representação e Análise de Interpretações Expressivas —
Um Estudo da Dinâmica e da Agógica**

Autor:

Lucas Simões Maia

Orientador:

Prof. Luiz Wagner Pereira Biscainho, D. Sc.

Examinador:

Filipe Castello da Costa Beltrão Diniz, D. Sc.

Examinador:

Prof. Wallace Alves Martins, D. Sc.

DEL

Agosto de 2013

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Escola Politécnica – Departamento de Eletrônica e de Computação

Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária

Rio de Janeiro – RJ CEP 21949-900

Este exemplar é de propriedade da Universidade Federal do Rio de Janeiro, que poderá incluí-lo em base de dados, armazenar em computador, microfilmear ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

*“Quem crê, nunca está sozinho
Nem na vida nem na morte.”*

Bento XVI

*“A música é uma arte e é, ao mesmo tempo, uma ciência.
Como arte, não é senão a manifestação do belo por meio
dos sons. Esta manifestação repousa sobre uma ciência
exata, formada pelo conjunto das leis que regem a produção
dos sons e suas relações de altura e duração.”*

Hugo Riemann

DEDICATÓRIA

Para meu irmão.

AGRADECIMENTO

Agradeço, primeiramente, a Deus, pela vida que me foi dada.

Agradeço a meus pais, por todo incentivo e apoio nessa longa caminhada, e também a minhas irmãs, pelos sorrisos e gritos.

Agradeço a meus avós, pelos exemplos que são e sempre foram.

Agradeço a meu orientador, pela oportunidade que me ofereceu de trabalhar com um tema que muito me interessa, e a meus colegas de faculdade, sem os quais minhas manhãs teriam sido muito mais difíceis.

Agradeço a minha namorada, sem a qual este trabalho não seria terminado, cujos sorriso e palavras de carinho foram verdadeiros motores para meu ânimo.

RESUMO

Este trabalho apresenta um conjunto de ferramentas que permitem visualizar o andamento e a intensidade que um músico confere a uma peça musical.

Entende-se que existe uma “assinatura” do intérprete, que o caracteriza na interpretação de diversas obras. Assim, as ferramentas aqui desenvolvidas auxiliam na compreensão das estratégias de expressividade empregadas.

Alguns conceitos de teoria musical são introduzidos, para dar a base necessária ao leitor. Além disso, é apresentado um modelo psicoacústico para a audição humana.

Como objeto de análise, foram escolhidos os 24 Prelúdios, Opus 28, de Frédéric Chopin, na execução de diversos artistas.

Os parâmetros de expressividade andamento e dinâmica são estudados em uma peça executada por diferentes músicos, e em um grupo de peças executadas por um dado músico.

São extraídas estatísticas para os andamentos dos diversos prelúdios e os valores médios encontrados apresentam leves desvios com relação às indicações metronômicas convencionais. Nesta análise estatística, observa-se a redução da intensidade média executada e percebida no decorrer dos anos.

As características dos intérpretes são comparadas e os resultados obtidos pela técnica dos mapas de correlação, que medem a similaridade nas execuções, estão de acordo com o que se sabe das relações interpessoais e profissionais dos mesmos. Esta técnica permite também avaliar a abrangência do modelo de Todd, que supõe uma relação linear entre andamento e intensidade (“*the faster the louder, the slower the softer*” — quanto mais rápido a peça é executada, maior a intensidade colocada em cada nota e, ao tocar mais lentamente, o intérprete também suaviza seu toque).

Por fim, mostra-se que a análise da trajetória de uma interpretação no espaço bidimensional da dinâmica e da agógica pode ajudar na assimilação dos recursos de interpretação expressiva.

Palavras-chave: dinâmica, agógica, psicoacústica.

ABSTRACT

This work presents a set of tools that allow the visualization of tempo and intensity adopted by a performer while playing a piece of music.

It is believed that there exists a “signature” of each performer, which characterizes him/her on the interpretation of music pieces. Thus, the tools developed here aid the understanding of expressive performance strategies.

Basic concepts of music theory are introduced and a psychoacoustic model of human hearing is described.

The Frédéric Chopin’s 24 Préludes, Opus 28, were chosen as object of analysis.

Agogics and dynamics are studied on one piece played by different performers and on a group of pieces performed by a given musician.

Tempo statistics are extracted and the mean values for each prelude show a small deviation from the conventional metronomic markings. The values of mean dynamics are found to decrease over time.

The similarities discovered by scape plotting the performers’ characteristics are in agreement with what is known about their interpersonal and professional relationships. This technique also allows evaluating the range of Todd’s model of musical expression — “the faster the louder, the slower the softer”.

Finally, it is shown that studying the trajectory of a performance in the two-dimensional space of tempo and loudness can help in inferring expressive performances techniques.

Key-words: dynamics, agogics, psychoacoustics.

SIGLAS

CD – *Compact Disc*

EBU – European Broadcasting Union

FFT – *Fast Fourier Transform*

ITU – International Telecommunication Union

MIDI – *Musical Instrument Digital Interface*

MIR – *Musical Information Retrieval*

SI – Sistema Internacional

SPL – *Sound Pressure Level*

STFT – *Short-Time Fourier Transform*

UFRJ – Universidade Federal do Rio de Janeiro

TL – *Track Level*

XML – *Extensible Markup Language*

Sumário

1	Introdução	1
1.1	Objetivo e Organização	1
1.2	Representação Musical	2
1.2.1	Notação Musical	2
1.2.2	A Pianola e o MIDI	6
1.2.3	Outras Formas de Representação	10
1.3	Psicoacústica	10
1.3.1	Pressão Sonora	10
1.3.2	Audibilidade	13
1.3.3	Faixas de uma Oitava e de um Terço de Oitava	15
2	Ferramentas de Expressividade	18
2.1	Agógica	20
2.2	Dinâmica	22
3	Extração dos Parâmetros Expressivos	23
3.1	<i>Beat Tracking</i>	23
3.1.1	Determinação de <i>Onsets</i>	25
3.1.2	Determinação do Andamento	28
3.2	Intensidade Percebida	30
3.2.1	Medição do Nível de Intensidade Percebida	31
3.2.2	Determinação da Dinâmica	32
4	Análise dos Dados	38
4.1	Descrição dos Arquivos de Teste	38
4.2	Curvas de Andamento e de Intensidade	40

4.3	Similaridade	41
4.3.1	Coeficiente de Correlação de Pearson	41
4.3.2	<i>Scape Plots</i>	42
4.4	Espaço de Dinâmica e Agógica	44
4.5	Análise 1: Estatísticas das Interpretações	45
4.6	Análise 2: Mapas de Correlação	49
4.7	Análise 3: Modelo de Todd	56
4.8	Análise 4: Trajetórias do Prelúdio 6	61
5	Conclusão	64
5.1	Considerações Finais	64
5.2	Projetos Futuros	65
	Bibliografia	67
A	Códigos da Implementação	71
A.1	Códigos Principais	72
A.2	Códigos Auxiliares	79
A.3	Códigos Externos	83

Lista de Figuras

1.1	Notação neumática quadrada e quironômica/sangaliana	4
1.2	Divisão binária dos valores das figuras	4
1.3	Outra representação para a divisão binária dos valores das figuras	5
1.4	Exemplo de partitura na notação moderna	6
1.5	Exemplo de rolo de papel perfurado	9
1.6	Exemplo de <i>software</i> sequenciador/editor de MIDI	9
1.7	Forma de onda de um sinal periódico	11
1.8	Faixa de níveis sonoros para a audição humana	13
1.9	Curvas de igual audibilidade	15
2.1	Experimento da reconstrução do “movimento interno” de Truslit	19
3.1	Divisão temporal	24
3.2	Exemplo de STFT de um sinal discreto	27
3.3	Exemplo de função de fluxo espectral para um sinal de áudio	29
3.4	Diagrama do algoritmo de extração da curva de dinâmica	37
4.1	Domínio da técnica de <i>scape plotting</i>	43
4.2	Exemplo de trajetória de uma execução no espaço bidimensional da dinâmica e da agógica	44
4.3	“Interpretação média” para o prelúdio 1	46
4.4	“Interpretação média” para o prelúdio 7	47
4.5	Curvas de andamento para as interpretações “média” e de M. PE- RAHIA para o prelúdio 7	48
4.6	Curvas de andamento das interpretações de S. D. TIEMPO e de G. OHLSSON (1995) para o prelúdio 7	49
4.7	Média das curvas de dinâmica por ano da execução	50

4.8	Média das curvas de dinâmica por ano de nascimento do intérprete	50
4.9	Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT	51
4.10	Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT sem considerar outros anos de execução	53
4.11	Mapas de correlação para o andamento das interpretações do prelúdio 7 por N. MAGALOFF e M. ARGERICH	55
4.12	Mapas de correlação para o andamento das interpretações do prelúdio 1 por A. CORTOT sem considerar outros anos de execução	57
4.13	Mapas de correlação para a dinâmica das interpretações do prelúdio 7 por A. CORTOT sem considerar outros anos de execução	58
4.14	Mapas de correlação entre andamento e dinâmica das interpretações do prelúdio 1 para cada intérprete	59
4.15	Curvas de andamento e dinâmica normalizadas para a interpretação do prelúdio 1 por B. DAVIDOVICH	61
4.16	Trajetória das execuções no espaço bidimensional da dinâmica e da agógica para os dois primeiros compassos do prelúdio 6	62
4.17	Trajetória das execuções no espaço bidimensional da dinâmica e da agógica para os oito primeiros compassos do prelúdio 6	63

Lista de Tabelas

1.1	Frequências centrais das faixas de uma oitava e de um terço de oitava	17
2.1	Tipos de andamento e respectivas indicações metronômicas	21
4.1	Interpretações expressivas com <i>onsets</i> anotados	39
4.2	Informações sobre os prelúdios anotados	40
4.3	Valores médio, máximo e mínimo entre intérpretes dos andamentos médios para os prelúdios anotados	45

Capítulo 1

Introdução

1.1 Objetivo e Organização

O objetivo deste projeto é a elaboração de um conjunto de ferramentas que permita destrinchar visualmente a expressividade inculcada por um intérprete em uma peça musical.

A análise dos parâmetros expressivos durante a execução de uma obra “lança luz” sobre as intenções de seu intérprete. Concentrando-se em poucos, mas essenciais, parâmetros da expressão musical, pode-se inferir uma “assinatura” associada a este intérprete, que tenta caracterizá-lo nas interpretações de diferentes obras.

Neste capítulo, são introduzidos alguns conceitos musicais, acústicos e psicoacústicos fundamentais para a compreensão do presente trabalho. Diferentes formas de notação musical, bem como sua evolução através dos séculos, são apresentadas. Além disso, o instrumento mecânico popularmente conhecido como pianola é evidenciado como motivação para o estudo dos parâmetros expressivos.

No Capítulo 2, duas importantes componentes da intenção musical — agógica e dinâmica — são elucidadas. Estas serão as principais ferramentas para a análise das características expressivas dos intérpretes.

No Capítulo 3, estuda-se a extração destas características, *i.e.*, como serão obtidas a partir do conjunto de dados. Um método de determinação automática de *onsets* é analisado e proposto como forma de validação dos pulsos da música (*beats*) marcados de forma manual. Posteriormente, é sugerido um modelo psicoacústico para determinação da curva de dinâmica.

No Capítulo 4, o conjunto de dados é apresentado — um subconjunto de uma base de dados formada pelo ciclo dos 24 Prelúdios, Opus 28, de Frédéric Chopin, na execução de 27 artistas. Diferentes análises são propostas sobre os parâmetros extraídos. Destaca-se a representação no espaço bidimensional da dinâmica-agógica.

Por fim, no Capítulo 5, é realizada uma discussão dos resultados e das possibilidades de extensão do presente projeto.

1.2 Representação Musical

Nesta seção, é feita uma breve revisão histórica da representação musical. Parte-se da notação efonética e apresentam-se os desenvolvimentos graduais que nos levaram da notação neumática até a notação atual, na qual cada nota escrita na pauta informa altura e duração de um som [1].

Alguns conceitos da teoria de compassos são visitados. A pianola e os rolos de papel perfurado são apresentados como forma de representação de *pitch*¹, andamento e dinâmica, no domínio do tempo, e tratados como motivação para o estudo das intenções dos artistas. Por fim, são descritas formas de representação musical que podem ser “lidas” por um computador.

1.2.1 Notação Musical

O sistema moderno de notação musical, que indica, em uma partitura, ao menos duas características — altura e duração — de uma sequência de sons, data do início do século XVII [2].

Esta representação tem suas raízes nas notações efonéticas como as de Aristófanos de Bizâncio (para pontuação e acentuação gramatical do grego) e da tradição de canto judaica [2]. Originalmente, estes sistemas de notação consistiam em um número limitado de sinais dispostos sobre as palavras ou sílabas que deveriam ser enfatizadas durante uma leitura (*e.g.*, de um texto litúrgico). Embora a execução exata de alguns destes sinais hoje seja desconhecida, entende-se que os

¹*Pitch* refere-se à altura com que um som é percebido auditivamente pelo ser humano. Quanto maior a frequência fundamental de uma nota musical, mais agudo é o som percebido. Nesse sentido, o *pitch* está diretamente relacionado à frequência fundamental, mas é um conceito psicoacústico.

sinais de pontuação serviam para clarificar as estruturas das sentenças enquanto os sinais de acentuação eram associados às variações ascendentes e descendentes do *pitch* [2].

Bebendo da fonte das notações ecfonéticas, desenvolveu-se, ao longo dos séculos V a VII, a notação neumática [2, 1]. Esta representação mnemônica “primitiva” dos neumas² também não definia a altura exata das notas nem suas durações, indicando apenas, de forma imprecisa, o contorno melódico. Por volta do século IX, esta deficiência dos neumas *in campo aperto*, *i.e.*, sem moldura ou pauta, começou a ser contornada com a adição de uma linha, que representava a nota fá, sob e sobre a qual se dispunham as figuras, indicando sons mais graves e agudos, respectivamente, que a dada referência [1].

Considera-se o monge italiano Guido d’Arezzo (992–1050) como quem sugeriu o emprego do tetragrama para esta notação diastemática, isto é, da busca pela precisão na notação dos intervalos, utilizado até hoje na notação gregoriana. No tetragrama, aproveitam-se, ordenadamente, as linhas e os espaços entre estas para representar a gradação dos sons — mais agudo quanto mais alta a posição na pauta³.

O século XIII se iniciou com a preocupação acerca da duração das notas. Surgiram, então, a notação quadrada dos neumas (que viria a substituir os símbolos quironômicos mais vagos) e as figuras *longa*, *brevis* e *semibrevis*, em ordem decrescente de duração. As figuras *minima* e *seminima* apareceram, pela primeira vez, no século XIV. Contudo, a proporção entre as durações destas figuras não foi fixada e podia ser diferente (*i.e.*, imperfeita ou “duas para cada uma”, perfeita ou “três para cada uma”) nas diversas mensurações — *modus* (relação entre *brevis* e *longa*, em geral, imperfeita), *tempus* (*semibrevis*–*brevis*) e *prolatio* (*minima*–*semibrevis*) [2].

Um exemplo da evolução histórica da notação neumática pode ser observado na Figura 1.1, que mostra a notação quadrada dos neumas no tetragrama e o contorno melódico equivalente da tradição manuscrita para as primeiras notas do canto “Ave Maria” no Ofertório do 4^o domingo do Advento.

²Como as figuras da notação moderna, os neumas são os elementos básicos deste tipo de notação.

³As quatro linhas do tetragrama se mostraram suficientes para as melodias gregorianas, nas quais o *ambitus*, *i.e.*, a distância entre a nota mais baixa e a mais alta de uma melodia, é pouco desenvolvido. Embora o emprego de linhas suplementares seja raro, ao contrário do que se observa na música moderna, a notação gregoriana permite a mudança de clave no decorrer da peça [3].



Figura 1.1: Notação neumática quadrada (no tetragrama) e quironômica/sangaliana (abaixo). Adaptada de [3].

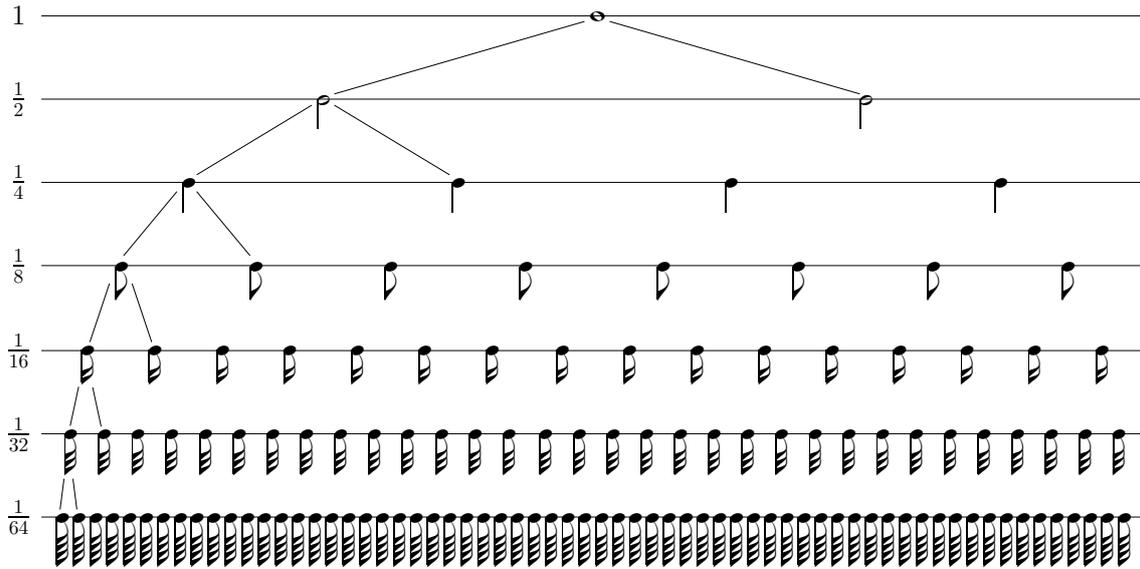


Figura 1.2: Divisão binária dos valores das figuras. A duração relativa de cada figura está indicada à esquerda. Adaptada de [1].

No século XV, surgiu a notação mensural branca, substituindo as notas originalmente pretas e, no século XVI, a mensuração tripla (*tempus perfectum*, *prolatio perfecta*) foi amplamente abandonada, assim como as ligaduras características dos neumas. Neste ponto, os princípios de notação eram, praticamente, os da notação moderna, excetuando-se o pentagrama, sistema com cinco linhas paralelas, e a notação redonda, que foram adotados somente no século XVII [1].

A duração relativa das figuras mais utilizadas na notação moderna pode ser observada nas Figuras 1.2 e 1.3. Na Figura 1.2, as figuras estão dispostas, de cima para baixo, em ordem decrescente de duração: semibreve, mínima, semínima, colcheia, semicolcheia, fusa e semifusa. Utiliza-se a mensuração dupla (divisão binária dos valores), *e.g.*, duas colcheias equivalem a uma semínima. Uma duração absoluta só pode ser determinada com as indicações de andamento (*cf.* Tabela 2.1).

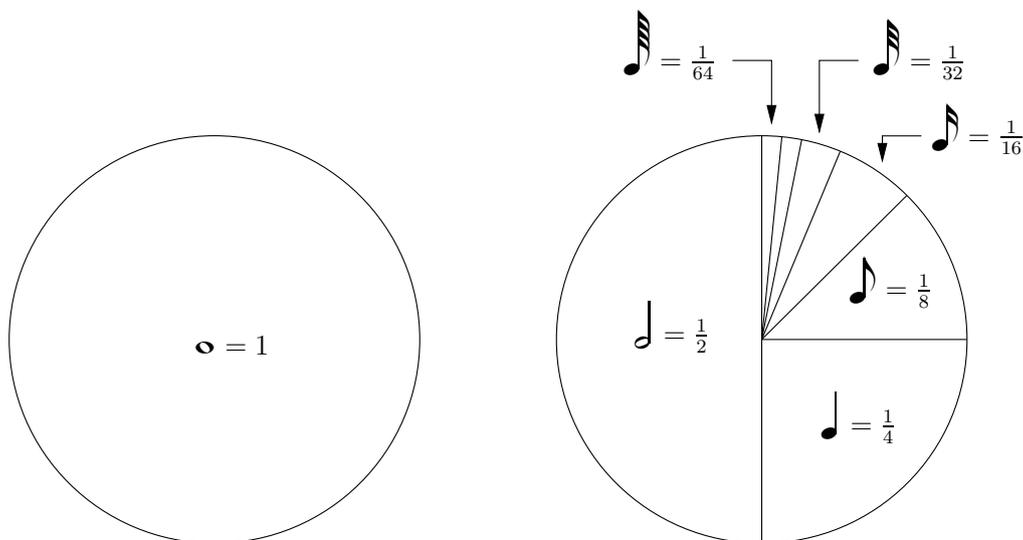


Figura 1.3: Outra representação para a divisão binária dos valores das figuras. Adaptada de [1].

Na notação moderna, destaca-se ainda o uso da fórmula e das barras de compasso. As teorias de compasso foram desenvolvidas a partir das proporções da notação mensural vistas anteriormente. O conceito de compasso enquanto divisão temporal regular da notação musical surgiu por volta do século XVII.

A fórmula de compasso é colocada no início de uma peça musical e informa, através de dois números dispostos na forma de fração, o tamanho dos compassos. O numerador indica a quantidade de figuras que cabem no compasso e o denominador assinala sua espécie, *e.g.*, a fórmula⁴ $\frac{3}{4}$ (lê-se “três por quatro”) indica que cada compasso pode conter 3 semínimas (uma semínima corresponde a $\frac{1}{4}$ da semi-breve, como pode ser verificado nas Figuras 1.2 e 1.3) ou uma combinação de figuras com duração total equivalente. No ritmo normal⁵, cada tempo do compasso⁶ deve obedecer à métrica⁷. Em geral, o primeiro tempo é “forte” e os subsequentes são

⁴Ao longo deste trabalho, as fórmulas de compasso serão apresentadas em negrito para serem diferenciadas dos valores das figuras.

⁵Ritmo é a organização do tempo ou, segundo Platão, “ritmo é a ordem do movimento” [1].

⁶Dentro de um compasso, uma unidade de tempo ou, simplesmente, tempo (*beat*) é o valor que se utiliza por “unidade de movimento” [1]. No exemplo do compasso $\frac{3}{4} = 3 \times \frac{1}{4}$, contam-se três tempos (*beats*) em cada compasso, uma vez que a figura de referência é a semínima.

⁷Existem acentos que modificam a métrica regular do compasso. Estes acentos são exemplificados no Capítulo 2.

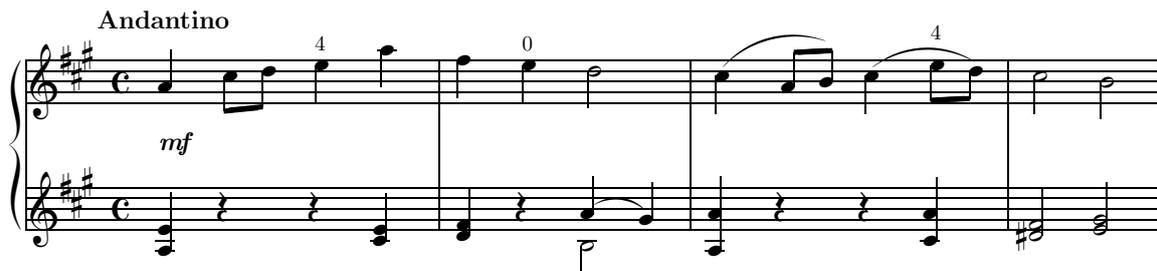


Figura 1.4: Exemplo de partitura na notação moderna: quatro primeiros compassos de um estudo para violino em *andantino*, de Nicolas Laoureux. Adaptada de [4].

ditos “fracos”⁸. Os compassos são separados por uma linha vertical, chamada barra de compasso. A Figura 1.4 traz um exemplo de partitura na notação moderna. Neste caso, **C** indica o compasso quaternário $\frac{4}{4}$ (lê-se “quatro por quatro”) e, portanto, cada compasso contém quatro semínimas ou uma combinação de figuras com duração equivalente⁹. O leitor interessado é convidado a buscar mais informações a respeito das teorias de compasso em [1].

1.2.2 A Pianola e o MIDI

Do século XIV ao XIX, os instrumentos mecânicos, *i.e.*, que atuam “automaticamente” sem a necessidade de um intérprete, utilizavam, em sua maior parte, o sistema de cilindro com pinos para executar sequências musicais predeterminadas [2]. Esta representação foi utilizada em carrilhões, relógios, realejos, caixas de música, cravos e até mesmo em órgãos de tubos. Nestes instrumentos, a rotação do cilindro fazia com que os pinos acionassem martelos, tocassem sinos, abrissem válvulas para a passagem do ar etc.

No final do século XIX, a tecnologia dos cartões perfurados, aplicada inicialmente na indústria têxtil francesa, substituiu os cilindros e pinos dos órgãos mecânicos. A vantagem dos rolos de papel perfurado, além de sua inquestionável

⁸Em compassos quaternários, o terceiro tempo é “meio forte”.

⁹No exemplo da Figura 1.4, o primeiro compasso da primeira voz contém três semínimas e um par de colcheias. Vemos que $3 \times \text{semínima} + 2 \times \text{colcheias} = 3 \times \frac{1}{4} + 2 \times \frac{1}{8} = \frac{4}{4}$, o que completa o compasso. Similarmente, o segundo compasso da mesma voz é composto de duas semínimas e uma mínima. Neste caso, $2 \times \text{semínimas} + \text{mínima} = 2 \times \frac{1}{4} + \frac{1}{2} = \frac{4}{4}$. Alguns compassos da segunda voz apresentam a figura ξ , que corresponde à pausa com duração relativa equivalente à da semínima.

“leveza”, era a capacidade de produção em massa desta representação, ainda que os rolos-base precisassem ser perfurados manualmente. A posição dos furos e seu comprimento correspondiam à altura e à duração dos tons a serem reproduzidos [2]. A ação dos rolos era pneumática — o ar somente era liberado para os tubos do instrumento quando passava através dos furos no papel.

As primeiras pianolas¹⁰ foram construídas e patenteadas na metade da década de 1890, inicialmente como um móvel externo ao piano que acionava suas teclas sob o comando de um rolo de piano [5]. Estes instrumentos cobriam 65 teclas (perdendo quase uma oitava em cada uma das extremidades do piano)¹¹, possuíam um conjunto de pedais para a sucção do ar e alavancas para o controle de andamento, dinâmica, sustentação, avanço e rebobinação do rolo de papel perfurado. Neste sentido, as pianolas não podem ser caracterizadas como instrumentos simplesmente “mecânicos”, uma vez que a expressividade da música era “controlada” pelo pianolista¹².

Em 1903, a Aeolian Company, uma das mais famosas fabricantes de pianolas, introduziu o *Metrostyle*. Inventado por Francis L. Young, entusiasta da pianola, o *Metrostyle* consistia de uma ponteira conectada à alavanca de andamento e que permitia acompanhar o fraseamento da música de uma linha vermelha traçada sobre o rolo de piano, alterando o andamento conforme indicado por essa linha. Na mesma

¹⁰A literatura faz a diferenciação entre os *piano players* e os *player pianos*. Os *piano players* eram equipamentos colocados na frente dos pianos convencionais e acionavam as teclas através pequenos “dedos” de madeira e feltro. Os *player pianos* eram pianos especiais, contendo um mecanismo embutido em seu corpo para o acionamento automático dos martelos. Ambos utilizavam o mesmo sistema pneumático com rolos de papel perfurado. A incerteza da novidade dos rolos de piano fez com que os compactos *piano players* fossem explorados comercialmente antes de seus “irmãos”; afinal, se os clientes possuísem um piano em casa, não iam querer um outro, de uma marca desconhecida. A *Pianola* original era um *piano player* da Aeolian Company, que começou a ser comercializado em 1898. Em 1903, a Aeolian comprou a renomada Weber Piano Company e, no ano seguinte, começou a desenvolver uma série de *player pianos* com o nome *Pianola*. Atualmente, o uso do termo enquanto marca desapareceu e serve para indicar, de maneira genérica, os *player pianos* acionados por pedais de sucção [5].

¹¹A transição para 88 notas só aconteceu em 1909.

¹²O piano reproduzidor, ao contrário das pianolas, era completamente independente de interação humana, reproduzindo música diretamente do rolo de papel perfurado. A informação de dinâmica era anotada no rolo e podia ser lida pelo piano. O primeiro piano reproduzidor foi, provavelmente, o *Welte Mignon*. Em 1914, a Aeolian lançou seu próprio piano reproduzidor, o *Duo-Art* [5].

época, o mecanismo da pianola foi dividido em dois, permitindo o controle independente dos graves e agudos. Em 1904, a Aeolian lançou a série de pianos especiais (*Pianola Pianos*) com o mecanismo pneumático embutido¹³. Em 1909, o *Themodist* é adicionado comercialmente aos dois carros-chefes da empresa — trata-se de uma ferramenta de dinâmica, aproveitando a separação das seções de graves e de agudos, para intensificar a linha melódica e reduzir o volume do acompanhamento [5]. A Figura 1.5 apresenta um rolo de papel perfurado para pianola com algumas fendas e indicações de andamento e dinâmica.

Muitos rolos de pianola eram reproduções baseadas em execuções de pianistas *virtuosos* ou preparadas com suas supervisões, *i.e.*, com suas indicações de interpretação. Claramente, isto foi de grande valor comercial para os fabricantes: seus clientes podiam ter acesso a músicos famosos em suas próprias casas! Existiram até peças compostas especificamente para as pianolas, e que contavam com movimentos impossíveis de serem executados por um pianista humano, *e.g.*, acordes com dezenas de dedos e nos baixos, médios e altos simultaneamente [2]. Contudo, ao longo da década de 20, os fabricantes de pianolas começaram a enfrentar dificuldades financeiras, que se agravam com a crise da bolsa, em 1929. Além disso, a expansão da transmissão de rádio e o advento do sistema elétrico de gravação, em 1925, terminaram de remover o caráter único desses instrumentos. No início dos anos 30, a produção de pianolas cessa¹⁴ [5].

Hoje em dia, ainda é possível acompanhar a herança que a pianola deixou para a música. Substituiu-se o mecanismo pneumático por um sistema eletromecânico-magnético com funcionamento equivalente e os tradicionais rolos de pianola pelo interfaceamento MIDI¹⁵. Mídias digitais regraváveis permitem guardar e reproduzir milhares de músicas e, portanto, o armazenamento digital apresenta uma enorme flexibilidade frente aos rolos de papel perfurado. Alguns exemplos de sistemas que

¹³A Aeolian já havia lançado a série *Aeriol Pianos* no final da década de 1890, sem grande sucesso comercial.

¹⁴As pianolas só retornariam nos anos 60 e 70, com novos estilos musicais gravados nos rolos.

¹⁵MIDI é um protocolo de comunicação padrão entre instrumentos musicais e equipamentos eletrônicos. Consiste em uma série de instruções que informam ao equipamento (placa de som, instrumento musical, telefone celular etc.) como reproduzir cada nota. As instruções indicam, por exemplo, em que instantes a nota começa e termina, e qual a intensidade sonora.

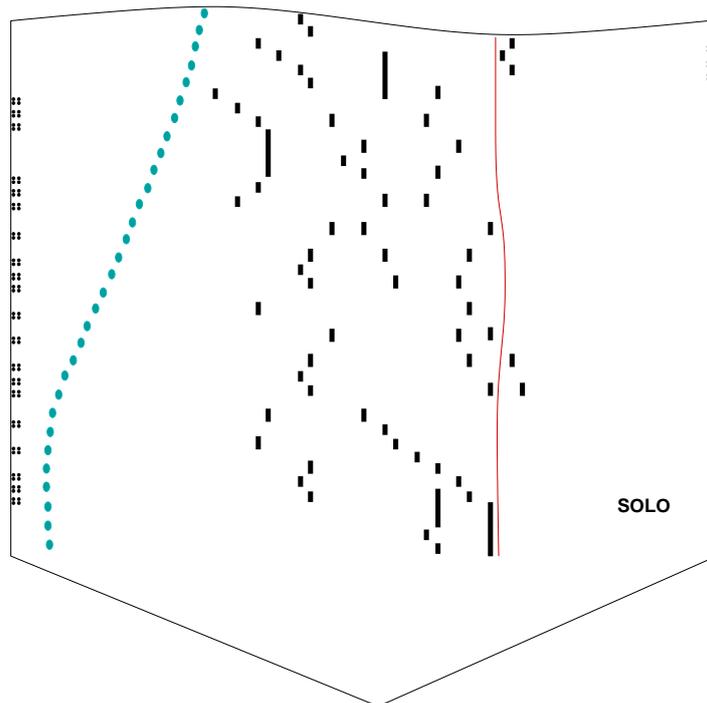


Figura 1.5: Exemplo de rolo de papel perfurado para pianola. Notar a linha vermelha para acompanhamento do *Metrostyle* e as perfurações laterais do *Thomodist*, para acentuação dos graves e dos agudos. Adaptada de [5].

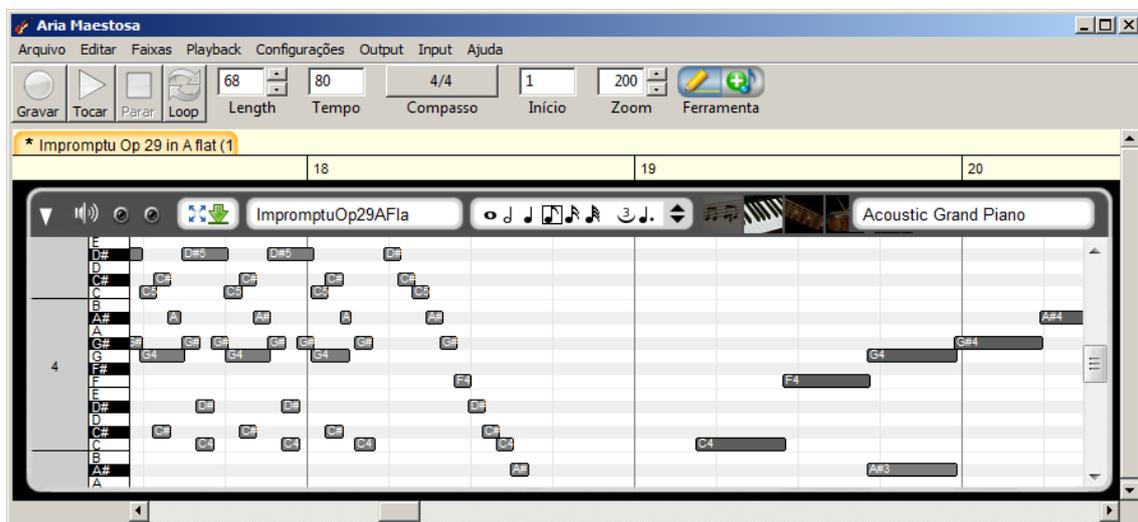


Figura 1.6: Exemplo da tela de um *software* sequenciador/editor de MIDI exibindo alguns compassos do Impromptu No. 1 em Lá bemol maior, Opus 29, de Chopin. A posição e comprimento das barras indica, respectivamente, o *pitch* e a duração das notas, assim como em um rolo de piano convencional. A intensidade fica por conta da coloração — quanto mais “forte”, mais escura é a barra.

operam desta forma são: *Disklavier* (Yamaha), *PianoDisc*, *CEUS* (Bösendorfer) e *LX* (Stahnke)¹⁶. Por fim, é notável a similaridade entre o rolo para pianola e as telas dos sequenciadores MIDI modernos, exemplificada na Figura 1.6.

1.2.3 Outras Formas de Representação

Destacam-se ainda outras formas de notação musical: MusicXML (baseada na linguagem XML) e as notações GUIDO e ABC.

Estas notações buscam representar a estrutura musical sem estarem “presas” às linguagens específicas dos diversos aplicativos de edição de partituras. Buscam, portanto, retratar a música de uma maneira portátil e compartilhável. Por serem representações em arquivos-texto das partituras convencionais, também podem carregar informações de dinâmica e andamento como os acentos que serão vistos no Capítulo 2.

1.3 Psicoacústica

A Psicoacústica é a área da Psicofísica que estuda a percepção do som pelo sistema auditivo humano. Baseia-se na representação dos sinais de áudio para modelar a resposta subjetiva deste sistema.

Segundo [6], “a Psicoacústica, que descreve a relação entre o mais baixo nível (o estímulo) e o maior nível possível (a sensação), é uma maneira atraente de avaliar nosso sistema auditivo”.

Nesta seção, a atenção é voltada para as ondas sonoras, que são, essencialmente, ondas mecânicas longitudinais. Estuda-se brevemente a sua geração e a sua propagação, através da análise da pressão sonora. Em seguida, são apresentados os princípios psicoacústicos mais relevantes para a compreensão deste trabalho.

1.3.1 Pressão Sonora

Do ponto de vista físico, ondas sonoras são geradas através da vibração de uma fonte, *e.g.*, a corda e a caixa acústica de um violino ou a membrana de um

¹⁶Os sistemas citados são melhor classificados como pianos reprodutores, e não pianolas, uma vez que funcionam com um nível ínfimo de interação com o ser humano.

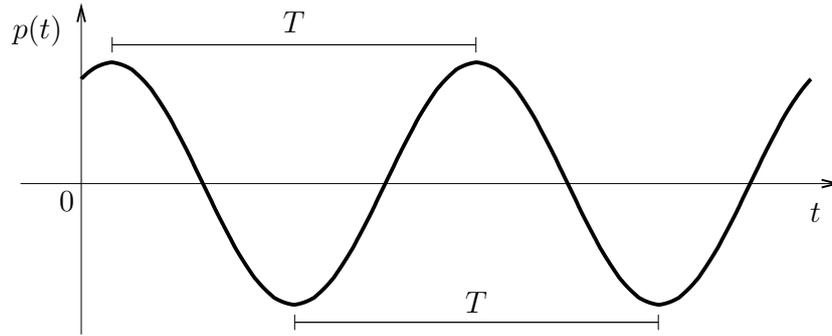


Figura 1.7: Forma de onda de um sinal periódico com período T .

tambor [7], e se propagam devido a forças mecânicas que atuam sobre as partículas do ar na forma de compressões e rarefações locais. De forma geral, as ondas sonoras são longitudinais¹⁷ e podem ser representadas através de sua forma de onda $p(x, t)$, que corresponde ao gráfico da pressão local em função do tempo e da posição na direção de propagação. A unidade do SI para medição de pressão é pascal (Pa). A forma de onda pode ser entendida como a superposição de uma pressão média, p_m , não afetada pelo sinal de perturbação e uma variação na pressão $\Delta p(x, t)$, *i.e.*,

$$p(x, t) = p_m + \Delta p(x, t). \quad (1.1)$$

Analisando-se a forma de onda no receptor, *e.g.*, o ouvido humano ou o diafragma de um microfone, e atentando-se apenas à variação de pressão com relação à média, o sinal $p(t)$ pode ser descrito como função somente do tempo. Quando a forma de onda $p(t)$ apresenta um padrão de repetição, o tempo necessário para esta repetição é chamado período e indicado por T . A frequência, medida em hertz (Hz), é dada pelo inverso do período¹⁸. Um exemplo de sinal periódico proveniente de uma fonte senoidal pura¹⁹ é mostrado na Figura 1.7.

Em psicoacústica, estudam-se valores de pressão entre 10^{-5} Pa (limiar de audição) e 10^2 Pa (limiar de dor) [6]. Para que seja possível trabalhar com uma

¹⁷O movimento de compressão e rarefação das partículas se dá na direção de propagação.

¹⁸Se um sinal apresenta uma regularidade, *i.e.*, um período de repetição, a este sinal pode ser associado um *pitch*, conforme definiu-se anteriormente. Embora, fisicamente, a frequência de um sinal possa variar de 0 Hz até infinito, só faz sentido pensar em *pitch* nos limites da audição humana (aproximadamente entre 20 Hz e 20 kHz).

¹⁹Sinais reais são, geralmente, compostos por uma combinação de frequências como, por exemplo, uma fundamental e seus harmônicos.

faixa dinâmica tão larga, costuma-se utilizar a escala logarítmica do nível de pressão sonora (em inglês, *Sound Pressure Level* — SPL). A relação entre a pressão sonora, p , e o nível de pressão sonora, L_p , é dada por

$$L_p = 20 \log \frac{p}{p_0} \text{ dB}, \quad (1.2)$$

na qual o valor de referência p_0 é, por padrão, igual a $20 \mu\text{Pa}$, que corresponde à pressão sonora de um tom de 2 kHz no limiar de audição [8]. O nível de pressão sonora se relaciona com o nível de intensidade sonora²⁰ através de

$$L_p = 20 \log \frac{p}{p_0} \text{ dB} = 10 \log \frac{I}{I_0} \text{ dB}, \quad (1.3)$$

onde a intensidade de referência I_0 é adotada como 10^{-12} W/m^2 , que equivale à onda de referência com pressão igual a p_0 . O valor L_p também pode ser chamado de nível sonoro.

É importante destacar que as ondas de pressão aqui descritas são sinais analógicos no tempo contínuo. A representação destes sinais em um computador, *e.g.*, em arquivos WAV, é obtida após os processos de amostragem (discretização no tempo), quantização (discretização na amplitude) e codificação. O sinal de áudio resultante é, portanto, a versão digital da forma de onda após uma transformação com perdas. Admitindo-se que a frequência de amostragem f_s satisfaz o Teorema de Nyquist²¹, o erro introduzido pelo processo de quantização, também chamado de ruído de quantização, pode ser feito imperceptível ao ouvido humano. Em [7, 8], estuda-se o exemplo da representação digital em um CD de áudio, no qual se armazenam dados a uma taxa de 44 100 Hz, *i.e.*, o sinal amostrado apresenta 44 100 amostras por segundo²². Por fim, cada amostra é representada (codificada) com 16 *bits*, correspondendo a 65 536 níveis de quantização.

²⁰A intensidade sonora é dada pela potência média da onda por unidade de área, e é proporcional a p^2 .

²¹Segundo o Teorema de Nyquist, para que a reconstrução de um sinal passa-baixas amostrado para o tempo discreto seja possível, a frequência de amostragem precisa ser maior que o dobro da maior frequência do sinal original. Assim, evita-se o efeito de *aliasing*.

²²Neste caso, considerando o limite superior da faixa audível, tem-se que 44 100 Hz é maior que $(2 \times 20\,000)$ Hz. Assim, esta frequência de amostragem satisfaz o Teorema de Nyquist e o sinal de áudio pode ser reproduzido com fidelidade.

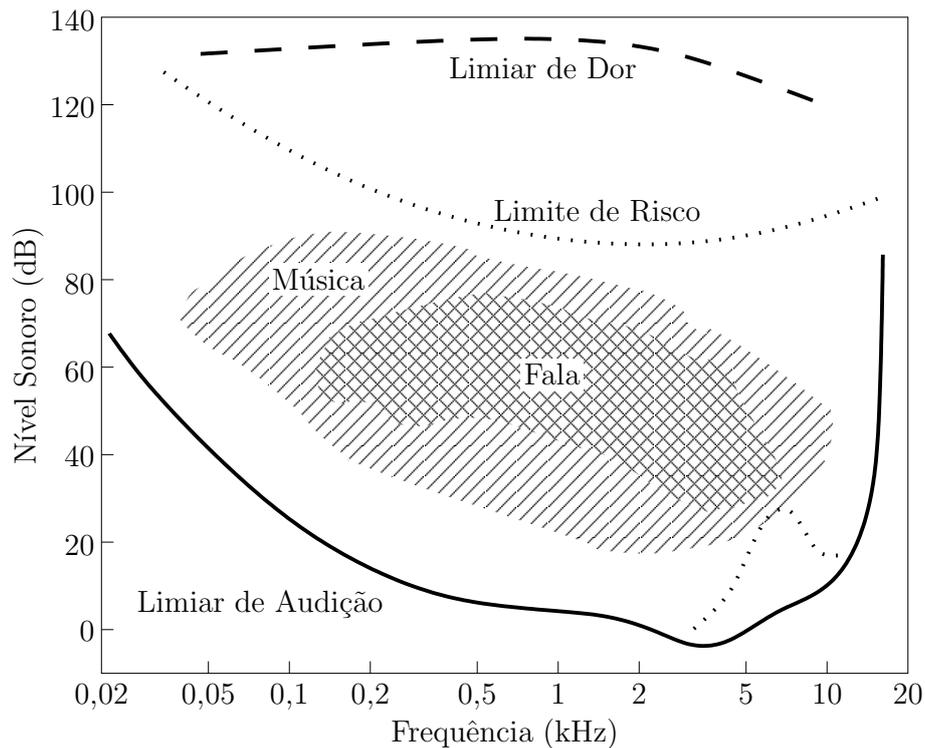


Figura 1.8: Faixa de níveis sonoros para a audição humana. Adaptada de [6].

1.3.2 Audibilidade

Como visto na Seção 1.3.1, o ouvido humano trabalha com uma larga faixa dinâmica de pressão sonora. A Figura 1.8 apresenta esta faixa, também chamada de região audível (em inglês, *hearing area*), que equivale à área do gráfico nível sonoro *versus* frequência onde se localizam os sons audíveis. Portanto, a região está inserida entre 20 Hz e 20 kHz, limitada inferiormente pelo limiar de audição (na figura, representado através de uma linha cheia) e superiormente pelo limiar de dor (linha tracejada). Destacam-se ainda sub-regiões que compreendem a maior parte dos sinais de música e de fala, e o limite de risco de danos à audição (linha pontilhada). A exposição a sons de grande intensidade produz variações temporárias no limiar de audição. Se esta exposição for prolongada, acaba causando alterações permanentes nos limites da região audível. Na mesma figura, uma linha pontilhada junto ao limiar de audição exemplifica estas perdas auditivas. Algum nível de perdas em altas frequências também é esperado com o avanço da idade [6, 8].

É importante interpretar a forma do limiar de audição apresentada na Figura 1.8. O limiar de audição indica o nível sonoro (de pressão ou de intensidade) a

partir do qual um som composto por uma frequência pura é audível. Desta maneira, a intensidade sonora percebida, *i.e.*, a sensibilidade do ouvido humano, varia com a frequência. Para baixas frequências, é necessária uma pressão sonora maior para tornar o tom audível. O mesmo ocorre para altas frequências. Na faixa de 2 kHz a 5 kHz, em geral, nota-se uma maior sensibilidade do ouvido humano, *i.e.*, pode ser aplicada uma pressão sonora menor para que o som seja ouvido.

Na realidade, a percepção da intensidade sonora é também função do nível de pressão sonora do som emitido. A grandeza que expressa esta sensação da intensidade sonora é denominada nível de audibilidade ou nível de intensidade percebida (em inglês, *loudness level*) [6]. O nível de audibilidade L_N de um som é definido como o nível de pressão sonora de um tom de 1 kHz em onda plana e de incidência frontal que soa tão audível quanto o som [6]. A unidade de medida²³ do nível de audibilidade é o phon.

A Figura 1.9 apresenta as curvas de igual audibilidade e seus respectivos valores em phons para tons puros, que podem ser medidas experimentalmente. Da definição da unidade, percebe-se que todas as curvas assumem o mesmo valor em phons e em dB_{SPL} (nível de pressão sonora) em 1 kHz, *e.g.*, a curva de igual audibilidade para 20 phons admite o valor de 20 dB_{SPL} em 1 kHz. A curva do limiar de audição também é uma curva de igual audibilidade e está indicada na Figura 1.9 com a linha mais espessa. Como o nível de pressão sonora para um tom de 1 kHz minimamente audível é de 3 dB, a curva do limiar de audição é designada com 3 phons [6]. A partir da mesma figura, é possível notar que a variação da sensibilidade é tanto menor quanto maior o ponto de operação (audibilidade). Para todas as curvas, a faixa de frequências de 2 kHz a 5 kHz é sempre a região de maior sensibilidade.

A Figura 1.9 apresenta também os valores em sones das curvas de mesma audibilidade. Esta unidade de intensidade percebida utiliza o nível de intensidade sonora de 40 dB_{SPL} em 1 kHz como referência (1 sone) e se relaciona com o phon através da equação:

$$N = 2^{\frac{L_N - 40}{10}}, \text{ para } L_N > 40 \text{ phons}, \quad (1.4)$$

²³Não faz parte do SI de unidades.

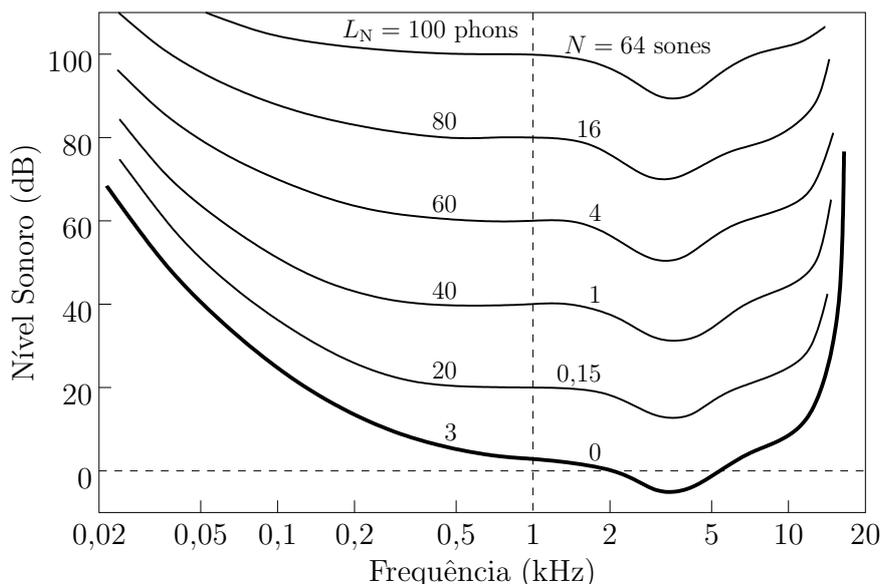


Figura 1.9: Curvas de igual audibilidade. O limiar de audição está indicado como a curva de maior espessura. Adaptada de [6].

de forma que um aumento de $10 \text{ dB}_{\text{SPL}}$ corresponde a dobrar a intensidade percebida do tom de referência [6]. Para níveis de pressão sonora inferiores a 40 phons, a lei exponencial descrita pela equação (1.4) não pode mais ser utilizada como aproximação, uma vez que a intensidade percebida decresce cada vez mais rápido com a diminuição do nível sonoro, atingindo 0 sone para um nível de $3 \text{ dB}_{\text{SPL}}$ (limiar da audição).

Discrepâncias encontradas entre as curvas de audibilidade medidas experimentalmente por Fletcher e Munson (1933) [9], por Churcher e King (1937) [10], por Zwicker e Feldtkeller (1955) [11], por Robinson e Dadson (1956) [12], e em estudos mais recentes [13] deram origem à revisão da norma ISO 226:1987 para o novo padrão ISO 226:2003 [14]. Embora a nova norma seja considerada por muitos a leitura definitiva sobre o assunto, alguns pesquisadores apontam inconsistências em cálculos e definições do padrão [15].

1.3.3 Faixas de uma Oitava e de um Terço de Oitava

Uma vez que a faixa audível é extensa, em diversas aplicações, mostra-se interessante avaliá-la em subfaixas de frequências. De maneira geral, há dois tipos de divisão em subfaixas que costumam ser utilizados na literatura:

- Subfaixas com largura constante — a diferença entre as frequências inferior e superior de cada subfaixa é constante ao longo da faixa audível;
- Subfaixas com largura proporcional à frequência central — a razão entre as frequências inferior e superior de cada subfaixa é constante ao longo da faixa audível.

As faixas de uma oitava e de um terço de oitava estão inseridas neste último tipo.

Um intervalo de uma oitava é equivalente a tomar o dobro da frequência do tom original [2]. As escalas musicais nomeiam como a “mesma nota” frequências cuja razão seja uma potência (inteira) de 2, pois assim são percebidas [2].

A divisão em faixas de uma oitava baseia-se neste mesmo princípio e, portanto, está aliada à percepção humana. Dada a frequência central de uma subfaixa, f_{central} , as frequências inferior e superior podem ser obtidas através de

$$\begin{cases} f_{\text{superior}} = 2^{\frac{1}{2k}} f_{\text{central}}; \\ f_{\text{inferior}} = \frac{f_{\text{central}}}{2^{\frac{1}{2k}}}, \end{cases} \quad (1.5)$$

onde $k = 1$ para faixas de uma oitava. Dessa forma, o intervalo entre os limites inferior e superior em uma subfaixa é de exatamente uma oitava, visto que

$$\frac{f_{\text{superior}}}{f_{\text{inferior}}} = 2, \quad (1.6)$$

e a frequência central pode ser obtida a partir da média geométrica destes limites,

$$f_{\text{central}} = \sqrt{f_{\text{superior}} f_{\text{inferior}}}. \quad (1.7)$$

Por fim, a largura de cada subfaixa é proporcional à frequência central,

$$\Delta f = f_{\text{superior}} - f_{\text{inferior}} = \frac{\sqrt{2}}{2} f_{\text{central}}, \quad (1.8)$$

o que decorre imediatamente da sua definição. Isto significa dizer que, quanto maior a frequência central da subfaixa que se está avaliando, maior é a banda e, consequentemente, o número de *bins* de uma representação discreta na frequência.

Cada faixa de uma oitava pode ser subdividida em três faixas de um terço de oitava, para uma maior resolução no domínio da frequência. Neste caso, utiliza-se $k = 3$ na equação (1.5).

Tabela 1.1: Frequências centrais das faixas de uma oitava e de um terço de oitava, em concordância com o padrão ISO 226 (valores aproximados).

# da faixa	Faixas de uma oitava		Faixas de um terço de oitava	
	f_{central} (Hz)		f_{central} (Hz)	
11, 12, 13	16	12,5	16	20
14, 15, 16	31,5	25	31,5	40
17, 18, 19	63	50	63	80
20, 21, 22	125	100	125	160
23, 24, 25	250	200	250	315
26, 27, 28	500	400	500	630
29, 30, 31	1000	800	1000	1250
32, 33, 34	2000	1600	2000	2500
35, 36, 37	4000	3150	4000	5000
38, 39, 40	8000	6300	8000	10 000
41, 42, 43	16 000	12 500	16 000	20 000

A Tabela 1.1 apresenta os valores aproximados²⁴ para as frequências centrais em cada uma das subfaixas de uma oitava e de um terço de oitava.

²⁴A sequência de valores exatos pode ser obtida tomando-se a frequência de 1 kHz como referência e efetuando divisões e multiplicações por $2^{\frac{1}{k}}$, com $k = 1$ para faixas de uma oitava e $k = 3$ para faixas de um terço de oitava.

Capítulo 2

Ferramentas de Expressividade

Retirar notas da partitura para dar-lhes algum sentido é parte integrante da rotina de um estudante de música. Assim também é o trabalho de um intérprete: perscrutar a intenção do compositor, compreender seu estado de espírito ao escrever uma dada peça. Muitas vezes, o próprio título da obra o auxilia neste desafio. Temos vários exemplos disto na música erudita [1].

Em [16, 17], entende-se que a sincera execução de uma peça musical advém, antes, da compreensão de seu “movimento interno” (*innere Bewegtheit* [16]) tanto pelo intérprete quanto pelo compositor. São a experiência interna do artista e sua qualidade criativa que se unem para, conscientemente, dar forma e conteúdo à peça. Da mesma maneira, o ouvinte, que, em geral, ignora os processos técnicos de transcrição e interpretação, deve reviver este movimento internamente (*inneres Mitgestalten* [16]) para apreciar a música em sua plenitude. Como o “movimento interno” gera a música, então, similarmente, a música pode provocar a sensação de “movimento natural” no ouvinte. Na música, este conceito pode ser observado, em algum grau, nas partituras, em termos como *andante* e *lento*, *accelerando* e *ritardando* ou *crescendo* e *diminuendo*, que carregam a noção de movimento.

Em 1938, o musicólogo Alexander Truslit [16] apresentou um experimento, reproduzido aqui na Figura 2.1, no qual movimentos intencionados por um indivíduo foram expressados em trechos musicais e, posteriormente, recuperados por outro indivíduo. Neste experimento, a comunicação foi feita apenas através de partituras e ambos estavam familiarizados com a representação do pesquisador para as curvas de movimento, a ferramenta de comparação para as intenções e interpretações.

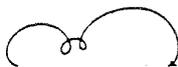
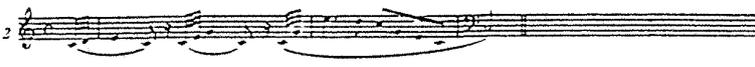
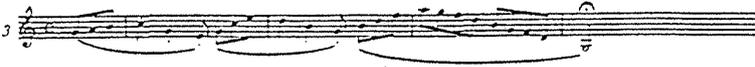
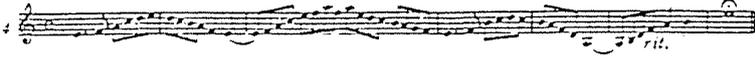
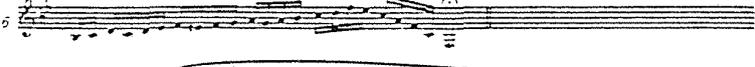
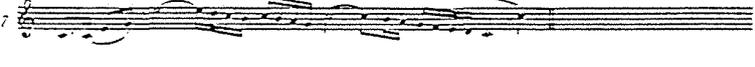
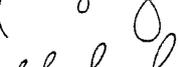
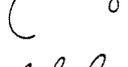
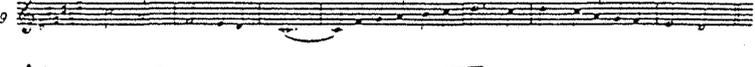
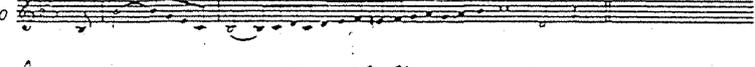
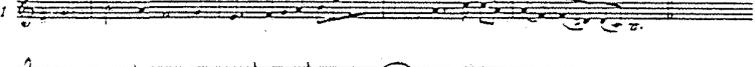
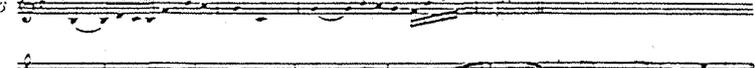
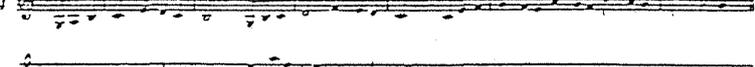
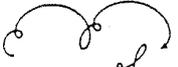
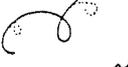
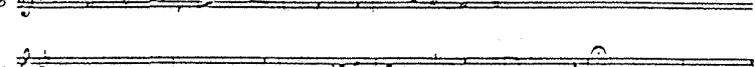
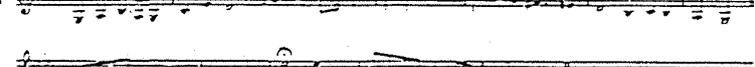
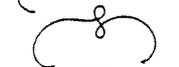
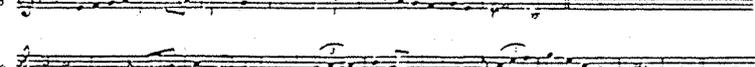
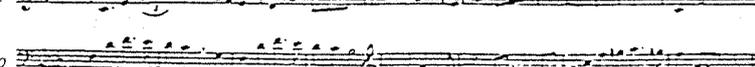
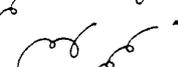
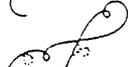
1 Musikalische Form der Urbewegung	2 Urbewegung	3 Festgestellte Beweg. (Reproduzierte Urbeweg.)
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		
		

Figura 2.1: Experimento da reconstrução do “movimento interno” de Truslit [16, 17].
 (1) Transcrição dos movimentos originais pretendidos pelo indivíduo A; (2) *Urbewegung* – movimentos originais (intencionados pelo indivíduo A); (3) *Festgestellte Bewegung* – movimentos determinados (recuperados pelo indivíduo B).

De forma geral, a expressão (ou expressividade) compreende tudo aquilo que não pode ser representado somente pelas notas ou, em seu extremo, por qualquer símbolo ou sinal na partitura. São vários os parâmetros que atuam para dar forma à música — altura, timbre, intensidade, duração, fraseamento, arcada (*e.g.*, violino), pedalização (*e.g.*, piano) etc.—, alguns com maior importância para o intérprete, outros para o compositor. São os chamados parâmetros expressivos.

A seguir, definem-se dois destes parâmetros: a agógica e a dinâmica, relacionados à duração e à intensidade dos sons, respectivamente.

2.1 Agógica

Definição 2.1 (Agógica). Do substantivo masculino grego $\alpha\gamma\omega\gamma\acute{o}\varsigma$ — aquele que guia, condutor —, derivado, por sua vez, do verbo $\acute{\alpha}\gamma\omega$ — guiar, conduzir [18]. Utilizado pela primeira vez, em contexto musical, em alemão (*Agogik*).

O andamento é a indicação da velocidade que um músico deve imprimir em uma peça. É uma indicação precisa e absoluta, que figura no início da partitura, da duração das notas e de suas pausas. O andamento pode ser medido em batidas por minuto (bpm). A Tabela 2.1 mostra alguns andamentos e as respectivas indicações metronômicas [1].

Na língua escrita, os acentos são sinais indicativos da ênfase ou da tonicidade que deve ser dada a certas sílabas. Assim também, na música, acentos são marcações que informam a alteração que deve ser executada sobre notas em uma frase musical. Dentre os diversos tipos de acento, são de interesse deste trabalho, os acentos agógicos e os acentos dinâmicos. Os acentos agógicos são aqueles que alteram não a intensidade ou a altura de uma nota (ou de um conjunto de notas), mas a sua duração, isto é, são acentos que atuam sobre o andamento da música [2].

Como a origem grega da palavra, a agógica representa a articulação dos sons, a alteração do andamento. Portanto, o músico (compositor ou intérprete) se utiliza das modificações rítmicas para “conduzir” a peça. O termo foi utilizado pela primeira vez neste sentido pelo musicólogo e compositor Hugo Riemann, em 1884, para apresentar estes desvios do andamento formal como parte essencial do ritmo e da métrica, necessários à compreensão das frases musicais [19].

Tabela 2.1: Tipos de andamento e as respectivas indicações metronômicas.

Tipo	Andamento	Indicação
		Metronômica (bpm, ref. ♩)
Lento	Grave	40
	Largo	44–48
	Lento	50–54
	Adagio	54–58
	Larghetto	60–63
Médio	Andante	63–72
	Andantino	69–80
	Sostenuto	76–84
	Moderato	88–92
	Allegretto	104–108
	Animato	120
Rápido	Allegro	132
	Vivace	160
	Presto	184

Alguns exemplos de indicação de alteração no andamento, como em [1], são:

1. Para apressar o andamento: *accelerando*, *affretando*, *stringendo*, *stretto*;
2. Para retardar o andamento: *ritardando*, *ritenuto*, *allargando*, *rallentando*;
3. Para retornar ao andamento: *a tempo*, *tempo primo*, *lo stesso tempo*;
4. Para suspender o andamento (interpretação livre): *ad libitum*, *senza tempo*, *a piacere*, coroa (sinal de prolongamento);
5. Para flexibilizar a divisão de tempo dentro do compasso: *tempo rubato* ou, simplesmente, *rubato*.

Algumas destas indicações são anotadas sob a pauta na forma abreviada (*e.g.*, *accel.* no lugar de *accelerando*).

2.2 Dinâmica

Definição 2.2 (Dinâmica). Do grego antigo, $\delta\upsilon\nu\alpha\mu\iota\kappa\acute{o}\varsigma$, adjetivo para “poderoso”/“forte” derivado de $\delta\upsilon\nu\alpha\mu\iota\varsigma$ — poder, força, potência. As duas palavras têm sua raiz em $\delta\upsilon\nu\alpha\mu\alpha\iota$ (grego), que significa “ser capaz”, “ter autoridade”, “ser poderoso” e, curiosamente, “ser elevado ao quadrado” [18]. Em alemão, *Dinamik*.

Segundo [1], a dinâmica é a “graduação da intensidade de som”, “é o grau de intensidade com que o som é emitido ou articulado”. Bohumil divide a dinâmica em natural, a alternância de tempos fortes e fracos própria da métrica, e artificial, que é a ferramenta de expressão a ser indicada na partitura pelo compositor.

Novamente, as indicações de dinâmica são expressas em termos italianos e grafadas sob a pauta de maneira abreviada — *fortississimo* **fff**, *fortissimo* **ff**, *forte* **f**, *mezzo forte* **mf**, *mezzo piano* **mp**, *piano* **p**, *pianissimo* **pp**, *pianississimo* **ppp** — aqui apresentadas na ordem decrescente de intensidade, isto é, do extremamente forte para o extremamente suave. As variações gradativas de dinâmica ficam por conta das marcações *crescendo* (<) e *decrescendo* (ou *diminuendo* >). Há ainda acentos dinâmicos pontuais, como o *marcato*, o *tenuto*, o *sforzato* (**sfz**) e o *forte piano* (**fp**). Todos estes sinais são ferramentas do compositor para guiar o intérprete a uma execução “mais viva”.

Se o intérprete executa as notas com a mesma intensidade e com durações de acordo com o que está determinado na partitura, sem considerar os acentos, a peça soa “pobre” [19], “sem vida e mecânica” [17]. A expressividade é, portanto, o que dá forma à música, através da variação destes e de outros parâmetros. Contudo, não deve ser enxergada como infidelidade à partitura e ao compositor [17]. É, pelo contrário, uma tentativa de se aproximar de suas intenções originais e humanas, adicionando uma contribuição criativa do próprio intérprete.

Neste trabalho, serão analisadas apenas as variações de andamento e intensidade em obras selecionadas. Isto não é feito em uma tentativa de reduzir a música a alguns elementos básicos, mas pela compreensão de que estas duas dimensões descrevem suficientemente bem o “movimento interno” e, portanto, lançam luz sobre as estratégias de expressividade empregadas pelos intérpretes.

Capítulo 3

Extração dos Parâmetros Expressivos

No Capítulo 2, foram definidas componentes expressivas que serão utilizadas como as principais ferramentas de análise no presente trabalho. Neste capítulo, estuda-se como cada uma dessas características será “capturada” a partir das diferentes execuções dos intérpretes.

3.1 *Beat Tracking*

Todos nós somos capazes, com maior ou menor aptidão, de seguir a “batida” de uma música através das palmas, dos pés ou até mesmo balançando a cabeça. No Brasil, é lugar-comum acompanhar o ritmo da música “Parabéns para você” com palmas, marcando-se, justamente, os tempos (*beats*) da música¹. Essa definição informal é a essência do *beat tracking*².

Acompanhar o ritmo de uma peça, *i.e.*, localizar o instante em que cada pulso ocorre, nos dá a noção de como o andamento da música está variando no tempo e, portanto, permite-nos descrever como um intérprete faz uso dos acentos agógicos. Quando há uma partitura, espera-se que os instantes percebidos como

¹Como a Figura 3.1 apresenta, é possível marcar também múltiplos e subdivisões do tempo.

²O termo em inglês será utilizado ao longo deste capítulo uma vez que um equivalente em português — “determinação”, “rastreamento” ou “acompanhamento” de tempo, de andamento ou de ritmo — soa estranho para o autor.

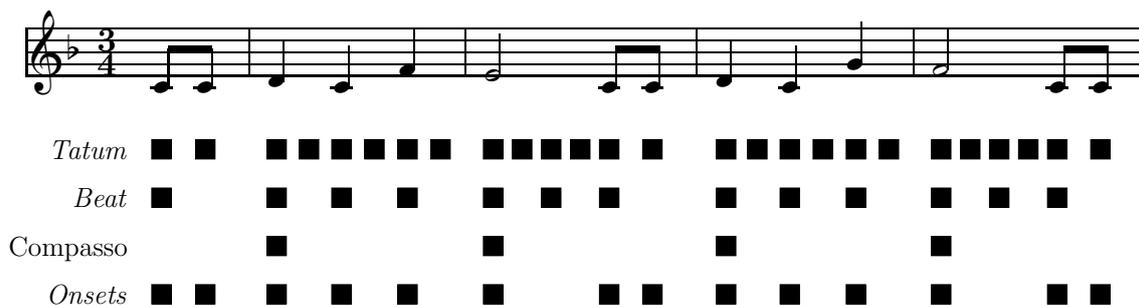


Figura 3.1: Divisão temporal em níveis *sub-* (*tatum*), *super-* (compasso) e ao nível do *beat* para a música “Parabéns para você”. Também são mostrados os ataques (*onsets*). Adaptada de [20].

tempos estejam de acordo com o que está notado na mesma, *i.e.*, recaiam sobre as notas que ocupam estas posições na estrutura temporal da música.

Na literatura, os algoritmos que se propõem a estimar a posição de cada tempo assim como um ser humano treinado faria são chamados de algoritmos de *beat-tracking* [20]. Em geral, os pulsos da música podem ser encontrados junto a notas³ e, assim, determinar o andamento da música se traduz em um problema de distinguir onde se encontram os ataques das notas (*note onsets*) e quais deles compõem a estrutura rítmica. Algoritmos de *beat-tracking* recebem, como entrada, uma representação simbólica (*e.g.*, instantes em que *onsets* ocorrem) ou direta (*e.g.*, um arquivo MIDI ou WAV) da música que se está estudando. Contudo, vale a pena destacar que, no caso em que a entrada é um arquivo de áudio WAV ou similar, alguns algoritmos transformam o fluxo de *bits* convencional em uma representação tabular, através da detecção dos *onsets*, sobre os quais um processamento posterior atua para determinar o andamento da música [20].

A Figura 3.1 exhibe alguns níveis temporais com os quais um trecho da música “Parabéns para você” pode ser analisado. O *tatum*, por vezes chamado de “átomo temporal” [21], representa o nível mais baixo da métrica, ou seja, o menor intervalo de tempo⁴ a ocorrer regularmente na música [20]. *Beat* e compasso são indicações

³Os tempos também são encontrados em pausas ou durante notas longas. Nesses casos, contudo, a determinação exata de onde o “tempo” intencionado pelo intérprete se encontra não é possível e, dessa forma, serão eventos ignorados, por enquanto.

⁴Deve-se evitar a confusão entre os intervalos de tempo, que são marcados em segundos, e os tempos da música, também chamados de pulsos ou de *beats*.

determinadas pelas unidades de tempo e de compasso⁵, respectivamente. Outro nível de representação exemplificado na figura é o conjunto de *onsets*, os ataques das notas. Note que os *onsets* não correspondem, necessariamente, aos *beats* e, ao mesmo tempo, nem todo *beat* pode ser associado a um *onset*.

Alguns algoritmos de *beat-tracking* foram testados para este trabalho. Nenhum foi capaz de acompanhar completamente as nuances dos diferentes músicos em suas interpretações expressivas. Como entende-se que, para modelar bem a expressividade de cada intérprete, é necessário um certo rigor na determinação dos tempos em cada peça, a marcação dos *beats* foi feita manual e cuidadosamente com o auxílio do *software* Sonic Visualiser [22]. Os *beats* marcados foram então “validados” pela proximidade com os *onsets* destacados por um algoritmo à parte. Na Seção 3.1.1, será apresentada esta função para determinação de *onsets*, conhecida como Fluxo Espectral (*Spectral Flux*). A Seção 3.1.2 trata da determinação do andamento.

3.1.1 Determinação de *Onsets*

Embora o andamento da música seja entendido como uma função contínua no tempo, os elementos musicais disponíveis para descrevê-lo são essencialmente discretos. Como foi visto no início deste capítulo, o problema de acompanhamento do andamento da música pode ser separado em dois problemas de igual peso e um passo de pós-processamento como a seguir:

1. Encontrar os *onsets* das notas;
2. Determinar, dentro dos *onsets* encontrados, quais correspondem à estrutura temporal, *i.e.*, *beat*, *super-beat* ou *sub-beat*, que se está avaliando;
3. Estimar, a partir dos instantes de tempo obtidos como resultado, uma versão discreta do andamento conferido à peça em análise.

A função de fluxo espectral [23] é uma medida de quanto o espectro de potência de um sinal no tempo varia de quadro para quadro. Assume-se que variações “bruscas” estão intimamente relacionadas à localização dos *onsets* e, é claro,

⁵A unidade de compasso é a figura ou figura pontuada que, sozinha, preenche completamente o compasso. No exemplo da Figura 1.4, a unidade de compasso é a semibreve. No exemplo da Figura 3.1, $\mathcal{C} = \mathcal{C} + \mathcal{C}$ corresponde à unidade de compasso.

também ocorrem nos tempos (*beats*), *i.e.*, obedecem ao andamento da música [24]. Dessa forma, o fluxo espectral pode ser utilizado na caracterização dos *onsets*.

Basicamente, o fluxo espectral é calculado tomando-se a derivada temporal do módulo da transformada de Fourier de curta duração⁶ (do inglês, *Short-Time Fourier Transform* — STFT) e somando-se o resultado para todas as frequências no domínio da transformada [23].

A Figura 3.2 apresenta um exemplo do cálculo da STFT de um sinal discreto. Pela definição da STFT, se $X(n, k)$ corresponde à k -ésima componente de frequência no n -ésimo quadro e o sinal no tempo $x(m)$ é janelado por $w(m)$, de comprimento N par, então [25, 23]

$$X(n, k) = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(hn + m)w(m)e^{-j\frac{2\pi}{N}km}, \quad (3.1)$$

para $k = -\frac{N}{2}, \dots, \frac{N}{2} - 1$ e onde h representa o passo com o qual a janela avança (*hop size*), em amostras⁷. Por exemplo, podemos ter uma janela de comprimento $N = 2048$ e com passo de $h = 441$ amostras, que equivale a 10 ms a uma taxa de amostragem $f_s = 44\,100$ kHz.

Assim, a função de fluxo espectral pode ser obtida através de [23]

$$\text{SF}(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n, k)| - |X(n-1, k)|), \quad (3.2)$$

onde $H(x) = \frac{x+|x|}{2}$ é o retificador de meia onda. Esta operação de retificação de meia onda é necessária para contabilizar no somatório apenas as contribuições de componentes de frequências em que há aumento de energia [25].

Foram propostas diversas modificações para esta função. Em algumas dessas modificações, o sinal é subamostrado para que se obtenha uma representação tempo-frequencial mais precisa, ou então é primeiro dividido em subfaixas e o cálculo do fluxo espectral é efetuado para cada subfaixa. Pode-se, inclusive, calcular uma função de fluxo para cada componente de frequência da representação no domínio da transformada (sem efetuar o somatório) ou ainda utilizar o logaritmo do módulo

⁶A transformada de Fourier de curta duração é uma generalização da transformada de Fourier que permite analisar, localmente, *i.e.*, em uma janela temporal, o conteúdo frequencial de um sinal.

⁷Se $h = N$, então não há sobreposição (*overlap*) entre os quadros adjacentes.

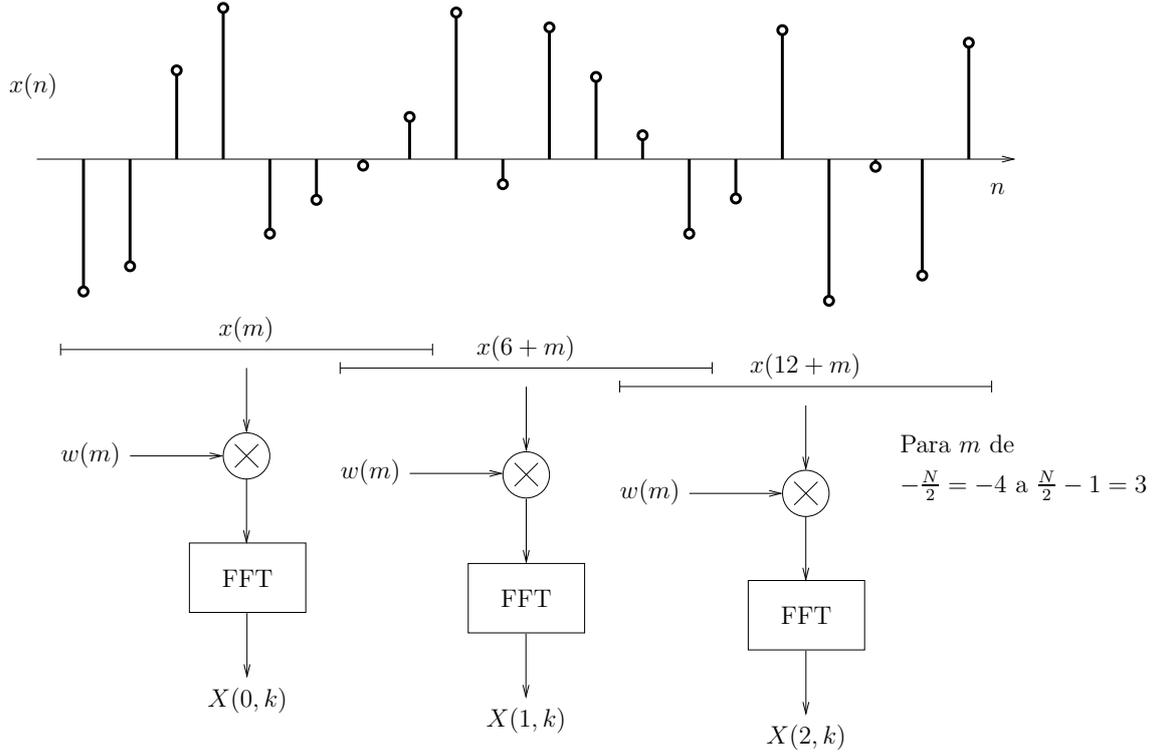


Figura 3.2: Exemplo de STFT de um sinal discreto, com $N = 8$ e $h = 6$.

da STFT para o cálculo da diferença, no método da distância de Kullback–Leibler modificada [20, 24].

Em vez da norma L_1 , pode-se tomar qualquer outra norma L_p , com $p \neq 1^8$, para o cálculo do fluxo espectral, *i.e.*,

$$\text{SF}_p(n) = \left\{ \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} [H(|X(n, k)| - |X(n-1, k)|)]^p \right\}^{\frac{1}{p}}. \quad (3.3)$$

Observa-se que valores de p menores levam, geralmente, a picos mais agudos na função de detecção e, portanto, a uma melhor resolução da posição dos ataques das notas⁹ [27].

Com a função de fluxo espectral montada, resta ainda identificar quais picos correspondem a verdadeiros *onsets*. Em [23, 27], a seguinte estratégia é traçada:

1. Primeiro, a função de fluxo espectral é normalizada para ter média zero e variância unitária;

⁸Estritamente falando, para $0 < p < 1$, a equação (3.3) não define uma norma.

⁹Em [26], utiliza-se o quadrado da norma L_2 . Entende-se que a aplicação ou não da raiz p -ésima (no caso de [26], a raiz quadrada) não altera a posição dos picos espectrais na função de detecção. Aplicar esta raiz se faz importante na comparação das funções $\text{SF}_p(n)$ para diferentes valores de p .

2. Um pico no instante $t_p = \frac{n_p h}{f_s}$ (em segundos) identifica um *onset* se satisfaz, simultaneamente, os critérios:

- Máximo local:

$$\text{SF}(n_p) \geq \text{SF}(n), \text{ para } n_p - M < n < n_p + M. \quad (3.4)$$

O pico no fluxo espectral precisa ser igual ao máximo local em um intervalo de $\pm M$ quadros, *e.g.*, $M = 3$.

- Limiar da média móvel:

$$\text{SF}(n_p) \geq \delta + \frac{1}{(\beta + 1)M + 1} \sum_{n=n_p-\beta M}^{n_p+M} \text{SF}(n). \quad (3.5)$$

O pico no fluxo espectral precisa ser maior que a média local acrescida de um fator δ . Aqui, β é um fator multiplicativo para garantir que a média seja calculada em um grande intervalo antes do pico, *e.g.*, $\beta = 3$.

- Limiar do decaimento exponencial:

$$\text{SF}(n_p) \geq g_\alpha(n). \quad (3.6)$$

O pico no fluxo espectral não pode ser menor que uma curva de decaimento exponencial determinada por

$$g_\alpha(n) = \max\{\text{SF}(n), \alpha g_\alpha(n - 1) + (\alpha - 1)\text{SF}(n)\}. \quad (3.7)$$

A Figura 3.3 apresenta um esboço da função de fluxo espectral e dos critérios de determinação dos ataques das notas para um sinal de áudio.

3.1.2 Determinação do Andamento

Conforme já foi mencionado, para este trabalho, os *beats* foram marcados manualmente para cada execução. O *plugin* MzSpectralFlux [27], desenvolvido pela equipe do Mazurka Project para o Sonic Visualiser, implementa a detecção de *onsets* segundo a formulação proposta em [23]. Para cada faixa do conjunto de dados, o grupo de *onsets* determinado por esse *plugin* foi utilizado na validação dos tempos marcados manualmente, segundo o critério do vizinho mais próximo. Dessa forma,

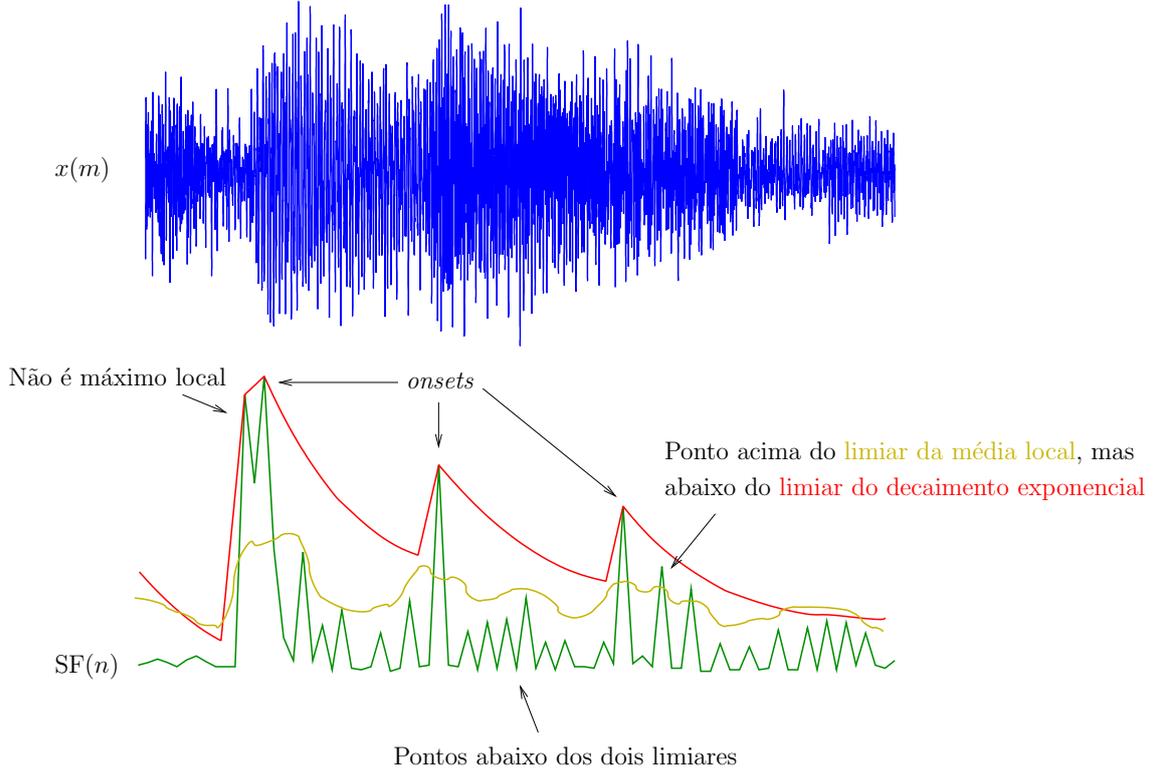


Figura 3.3: Exemplo de função de fluxo espectral para um sinal de áudio. Os *onsets* são calculados para o sinal $x(m)$ (em azul) a partir da função de fluxo espectral $SF(n)$ (em verde). Os limiares da média móvel (em amarelo) e do decaimento exponencial (em vermelho) são mostrados para auxiliar na compreensão do papel de cada um dos critérios de decisão. Adaptada de [27].

seja a sequência $\mathcal{B}_i = \{b_0, b_1, \dots, b_{B-1}\}$ de *beats* determinados manualmente após uma rigorosa audição da faixa i e a sequência de *onsets* $\mathcal{O}_i = \{o_0, o_1, \dots, o_{O-1}\}$ encontrados através do fluxo espectral da mesma faixa, onde b_j e o_k são marcados em segundos. Cada *beat* b_j é aproximado por um *onset* o_k , seu vizinho mais próximo dentro do grupo de *onsets*, através de

$$b'_j = \arg \min_{o_k \in \mathcal{O}_i} |b_j - o_k|. \quad (3.8)$$

A partir da sequência \mathcal{B}'_i de *beats* validados, é possível determinar o andamento da faixa i , em bpm, no instante em que cada *beat* b'_j ocorre,

$$A_i^j = \begin{cases} \frac{60}{b'_j - b'_{j-1}}, & \text{para } 1 \leq j \leq B-1, \\ A_i^1, & \text{para } j = 0. \end{cases} \quad (3.9)$$

Em alguns casos, contudo, o nível de marcação (em inglês, *Track Level* — TL) não é de *beats*, *i.e.*, os *onsets* selecionados não estão relacionados à unidade de tempo da partitura, mas correspondem a algum nível *super-beat* ou *sub-beat*. Nesse caso, seja a faixa i cujo compasso indicado na partitura é da forma $\frac{P}{Q}$, onde $\frac{1}{Q}$ representa a figura da unidade de tempo, se a figura de referência para marcação (TL) é $\frac{1}{Q'}$, com duração maior ou menor que a unidade de tempo, a partir da sequência \mathcal{B}_i de *super-beats* ou *sub-beats*, podemos determinar o andamento através do escalamento do resultado da equação (3.9),

$$A_{i,UT}^j = \frac{1/Q'}{1/Q} A_i^j = \frac{Q}{Q'} A_i^j, \quad (3.10)$$

e $A_{i,UT}^j$ é novamente dado em bpm.

Agora, $A_{i,UT}^j$ representa o andamento medido na unidade de tempo da partitura. Como vimos na Tabela 2.1, o andamento e as indicações metronômicas são comumente referenciados à semínima. Novamente, um simples fator multiplicativo nos leva para essa referência:

$$A_{i,J}^j = \frac{1/Q}{1/4} A_{i,UT}^j = \frac{4}{Q} A_{i,UT}^j = \frac{4}{Q} \frac{Q}{Q'} A_i^j = \frac{4}{Q'} A_i^j. \quad (3.11)$$

A função `extrairAgogica`, escrita em MATLAB, permite a extração da curva de andamento a partir do conjunto de *onsets* marcados e se encontra no Apêndice A.

3.2 Intensidade Percebida

Pode-se obter uma medida objetiva da intensidade sonora, em dB_{SPL} , a partir da STFT e do espectrograma de um sinal, e, com isso, descrever de que forma um intérprete faz uso da dinâmica expressiva. Contudo, essa medida ainda não incorpora as especificidades subjetivas aliadas à percepção do som.

Como foi destacado no Capítulo 1, a sensibilidade do ouvido humano e, portanto, o nível de audibilidade de um dado tom são funções da frequência e do nível objetivo de pressão sonora. Dessa forma, é interessante que a curva de dinâmica extraída carregue, ao menos, as características expressas pelas curvas de audibilidade [14], buscando representar, mais fielmente, o processo auditivo pelo qual passam o intérprete e os demais ouvintes.

A Seção 3.2.1 apresenta, rapidamente, a especificação EBU Tech Doc 3341, baseada na recomendação ITU-R BS.1770, para medição do nível de intensidade sonora percebida em programas de televisão. Na Seção 3.2.2 será descrito o algoritmo de determinação da dinâmica desenvolvido para o presente trabalho.

3.2.1 Medição do Nível de Intensidade Percebida

A recomendação ITU-R BS.1770 [28] descreve um algoritmo para medição do nível de intensidade sonora percebida ao longo de uma transmissão televisiva. O ITU-R, setor de radiocomunicação da ITU (International Telecommunication Union), entende que o áudio das transmissões modernas admite uma faixa dinâmica muito grande e que os telespectadores desejam, em geral, que o volume da programação seja uniforme tanto entre programas e intervalos comerciais, quanto entre diferentes canais de televisão.

A solução publicada pela EBU [29] (European Broadcasting Union) baseia-se na recomendação ITU-R BS.1770 e revisões seguintes. A implementação do comitê da EBU atua em três diferentes escalas temporais:

- Intensidade percebida “instantânea” — utiliza uma janela retangular deslizante com 400 ms de comprimento;
- Intensidade percebida de “curta duração” — utiliza uma janela retangular deslizante com 3 s de comprimento;
- Intensidade percebida “integrada” — avaliação do programa ou segmento como um todo, considerando a seleção de quadros (*gating*) de acordo com os limiares definidos pela ITU-R BS.1770-2.

Para um programa com vários canais de áudio, as medidas “instantânea” e de “curta duração” utilizam o algoritmo descrito pela recomendação ITU-R BS.1770-2, que consiste de três estágios:

1. Ponderação das componentes do sinal janelado em cada canal pela curva K — uma compensação das curvas de audibilidade implementada através da cascata de dois filtros de segunda ordem;

2. Cálculo do valor quadrático médio (“perceptivo”) do intervalo janelado em cada canal;
3. Soma ponderada da contribuição de cada canal para o nível de intensidade percebida no intervalo, *e.g.*, canais *surround* têm maior peso.

Para a medida “integrada”, cujo objetivo é qualificar o sinal completo com um único valor do nível de intensidade percebida, utiliza-se o método de *gating*. Inicialmente, o sinal é dividido em quadros com comprimento de 400 ms e 75% de sobreposição e o nível de intensidade percebida é avaliado para cada quadro, assim como na medida “instantânea”. Deve-se, então:

1. Excluir os quadros 70 dB abaixo da referência em fundo de escala (limiar absoluto);
2. Avaliar o nível de intensidade percebida para o programa inteiro utilizando-se a média dos quadros que restaram acima do limiar absoluto;
3. Excluir os quadros 10 dB abaixo desse valor (limiar relativo);
4. Recalcular o nível de intensidade percebida, utilizando-se os quadros que restaram;

O procedimento de *gating* visa a eliminar, do cálculo “integrado”, quadros de silêncio ou que contêm apenas ruído de fundo.

O “medidor de volume” descrito pela especificação EBU Tech Doc 3341 tem aplicação na área de produção e distribuição de conteúdo para a televisão digital. Contudo, os conceitos ali destacados são importantes para qualquer proposta de avaliação do nível de intensidade percebida. O algoritmo descrito na Seção 3.2.2 e utilizado neste trabalho para determinação da curva de dinâmica baseia-se nesses conceitos.

3.2.2 Determinação da Dinâmica

Em [30], desaprova-se a pouca importância dada, na literatura, à avaliação da dinâmica. Nesse artigo, o musicólogo Jörg Langner discute que os acentos dinâmicos são vistos como secundários frente às informações temporais quanto a sua capacidade

de qualificar, expressivamente, peças e seus intérpretes. Cita, como exemplo, o modelo de Todd, que simplifica a relação agógica–dinâmica, acoplando a evolução das duas em uma peça musical (“*the faster the louder, the slower the softer*”, *i.e.*, quanto maior a velocidade, maior a intensidade e, quando o ritmo é desacelerado, também diminui-se pressão sobre cada tecla, corda etc.) [31].

Segundo Langner, a pouca atenção dada à dinâmica se deve à falta de métodos de análise qualificados, que impossibilita uma maior compreensão de sua verdadeira natureza e de sua complexa relação com o andamento. O pesquisador propõe uma análise da variação da intensidade sonora percebida em diversos níveis, *i.e.*, avaliada através do janelamento de trechos da execução, para diferentes larguras de janela. Esta abordagem permite detectar arcos dinâmicos em suas diferentes camadas, uma vez que, quanto maior a largura da janela, maior a suavização que a curva de dinâmica sofre, e, portanto, analisa-se a variação da intensidade percebida em um contexto maior da peça [30]. Já pequenas janelas proporcionam uma análise local da dinâmica.

Este trabalho propõe uma maneira simples de se avaliar localmente o nível de intensidade sonora percebida e extrair a curva de dinâmica.

Assim como no caso da curva de andamento, a curva de intensidade é uma função contínua do tempo que será calculada em alguns pontos. Como não faz sentido comparar as interpretações expressivas no tempo absoluto, o nível de intensidade sonora percebida será calculado sobre o conjunto de *beats* ou de nível superior/inferior ao do *beat*, *i.e.*, em uma vizinhança limitada dos instantes em que cada evento de interesse ocorre — uma janela que contém cada *onset*¹⁰.

Conforme o proposto na recomendação ITU-R BS.1770, o trecho ao redor do *onset* será tratado no domínio perceptivo, aplicando-se uma curva de ponderação que “equaliza” a curva de audibilidade referente ao nível médio de intensidade percebida do trecho. Como o arquivo de áudio representa uma medida proporcional à pressão sonora, a intensidade de cada componente de frequência precisa ser “compensada” de forma a cancelar as diferenças apontadas pela curva de audibilidade. Por exemplo, seja uma senoide de 200 Hz cujo nível de pressão sonora é de 70 dB_{SPL}. Sabe-se

¹⁰O termo *onset* será utilizado ao longo deste texto para indicar o nível da métrica que se está avaliando. Contudo, destaca-se que nem sempre um *onset* corresponde a um pulso na música.

que este tom será percebido com 60 phons, *i.e.*, tão audível quanto uma senoide de 1000 Hz com nível de pressão sonora de 60 dB_{SPL}. Assim, a partir da medida de 70 dB_{SPL} para o tom de 200 Hz, é necessário que a curva de ponderação, que leva para o domínio perceptivo operando em 60 phons, confira uma atenuação de 10 dB_{SPL} nesta frequência. Considerando-se todas as frequências, vemos que a curva de ponderação utilizada para 60 phons corresponde à curva de audibilidade para este mesmo nível invertida ao redor de 60 dB_{SPL}.

Do proposto, surgem dois questionamentos:

- A norma ISO 226:2003 define os valores das curvas de audibilidade apenas para as frequências da série de um terço de oitava entre 20 Hz e 12 500 Hz, inclusive. Essas frequências não estarão necessariamente representadas ao tratar-se o sinal no domínio da frequência através da FFT (transformada de Fourier rápida, em inglês, *Fast Fourier Transform*) e existem outras tantas componentes de frequência. Como proceder com a ponderação das frequências não representadas pela ISO 226:2003?
- O algoritmo busca determinar o nível de intensidade percebida do trecho em questão, mas, para isso, utiliza a curva de ponderação que compensa o efeito da percepção humana e depende do conhecimento prévio do nível de audibilidade sobre o qual está operando. Como descobrir, então, o nível de intensidade percebida?

Para o primeiro ponto, uma solução simples seria utilizar o valor de ponderação da frequência central para todas as frequências nas subfaixas de um terço de oitava, evitando, assim, a necessidade de interpolar a curva de audibilidade. Quanto ao cálculo do ponto de operação da curva de audibilidade (e da curva de ponderação), propõe-se que seja obtido de forma iterativa, como em um método de aproximações sucessivas, valendo-se das iterações até que o nível médio de intensidade sonora percebida (*i.e.*, após a ponderação) seja equivalente ao valor da curva de ponderação.

Finalmente, os níveis de intensidade percebida na vizinhança de cada *onset* compõem a curva de dinâmica desejada.

Assim, sejam um sinal de áudio $x(m)$ amostrado a uma taxa f_s e a sequência de pulsos $\mathcal{B}'_i = \{b'_0, b'_1, \dots, b'_{B-1}\}$ determinados manualmente e validados. O algoritmo desenvolvido para a extração da curva de dinâmica pode ser resumido como:

1. Selecionar a região no entorno do j -ésimo *onset*, utilizando uma janela retangular com $\text{floor}(3f_s)$ amostras (3 s de comprimento);
2. Calcular a STFT $X(n, k)$ do trecho selecionado, utilizando janelas de Hann com $N = \text{floor}(0,4f_s)$ amostras (400 ms de comprimento) e 75% de *overlap*. O tamanho L da FFT de cada quadro deve ser grande o suficiente para que cada subfaixa de um terço de oitava seja representada por, ao menos, um *bin* na frequência, *e.g.*, $L = 2^{15}$ amostras;
3. Para cada quadro n da representação na STFT, calcular a potência em cada subfaixa de um terço de oitava:

$$P(n, \tilde{k}) = \sum_{F_{\tilde{k}}} |X(n, k')|^2, \quad (3.12)$$

onde \tilde{k} representa o número da subfaixa (considerando as subfaixas centradas em 20 Hz e 12 500 Hz como subfaixas 13 e 41, respectivamente, conforme descrito na Tabela 1.1) e $F_{\tilde{k}} = \{k' | f_{\text{inferior}} \leq \frac{f_s}{L}k' < f_{\text{superior}}\}$ é o conjunto de *bins* limitados entre as frequências inferior e superior da subfaixa;

4. Para cada subfaixa, calcular a média da potência ao longo dos quadros:

$$\bar{P}(\tilde{k}) = \frac{1}{M} \sum_{n=0}^{M-1} P(n, \tilde{k}); \quad (3.13)$$

5. Obter a potência média do trecho:

$$P_{\text{média}} = \sum_{\tilde{k}=13}^{41} \bar{P}(\tilde{k}); \quad (3.14)$$

6. Calcular o nível de pressão sonora L_p do trecho de 3 s através de:

$$L_p = 10 \log \frac{P_{\text{média}}}{P_{\text{ref}}}, \quad (3.15)$$

onde P_{ref} é uma constante de proporcionalidade que incorpora o valor de I_0 e os efeitos da janela de Hann e do comprimento L da FFT para que a senoide de referência de 1000 Hz em fundo de escala (amplitude igual a 1, no MATLAB) tenha um nível de audibilidade de 80 phons;

7. Enquanto a diferença entre duas aproximações sucessivas para o valor do nível de audibilidade L_N for maior que ε :

- i) Ponderar a potência média obtida no item 4 pela curva de ponderação correspondente a um nível de audibilidade L_N , calculado na iteração anterior¹¹;
 - ii) Calcular a potência média percebida do trecho, assim como é feito no item 5;
 - iii) Calcular a nova aproximação do nível médio de audibilidade L_N do trecho, similarmente ao indicado no item 6;
8. Retornar ao item 1 com $j = j + 1$, para calcular o nível de audibilidade na vizinhança do próximo *onset*.

Assim, D_i^j corresponde ao nível de audibilidade L_N para o j -ésimo *onset* da faixa i e é medido em “phons”. A sequência $\{D_i^j\}$ pode ser transformada para uma escala linear (em “sones”) através da equação (1.4).

A Figura 3.4 apresenta um diagrama em blocos simplificado para o algoritmo de extração de dinâmica acima descrito.

A função `extrairDinamica`, escrita em MATLAB, permite a extração da curva de dinâmica a partir do conjunto de *onsets* marcados e pode ser vista no Apêndice A.

¹¹Para a primeira iteração, utilizar L_p como estimativa do nível de audibilidade.

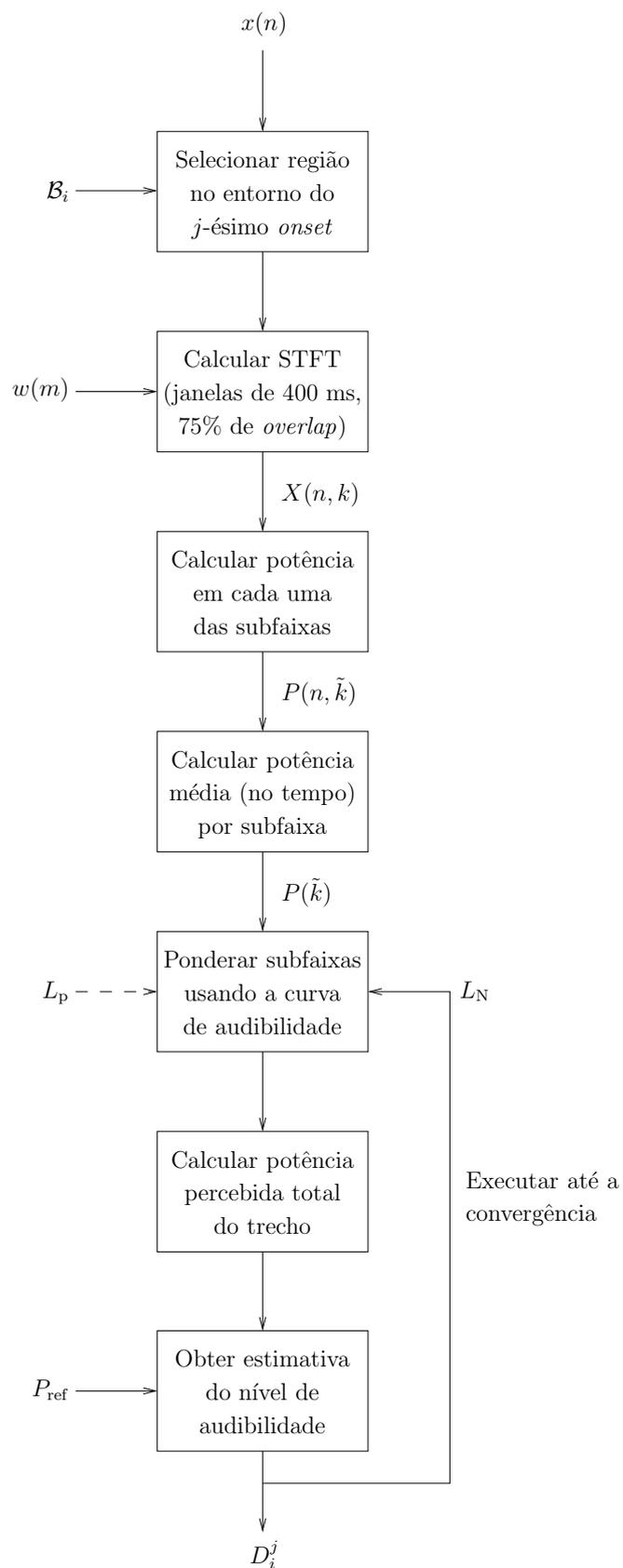


Figura 3.4: Diagrama do algoritmo de extração da curva de dinâmica.

Capítulo 4

Análise dos Dados

Neste capítulo, apresenta-se o conjunto de gravações expressivas estudado. Também são apresentadas diferentes técnicas de análise sobre os parâmetros extraídos — andamento e dinâmica — e os principais resultados obtidos.

A Seção 4.1 descreve a base de gravações utilizada neste trabalho.

Na Seção 4.2, as curvas de andamento e de intensidade são revisitadas.

A Seção 4.3 introduz a técnica de *scape plots*, que permite comparar a similaridade dos intérpretes quanto aos acentos agógicos e dinâmicos em uma dada peça musical através da correlação de Pearson.

Na Seção 4.4, a ferramenta de estudo das interpretações no espaço bidimensional da dinâmica e da agógica é apresentada.

Por fim, nas Seções 4.5–4.8, são realizadas algumas análises sobre os dados anotados tendo por base os métodos descritos nas Seções 4.2–4.4.

4.1 Descrição dos Arquivos de Teste

O conjunto de dados analisado no presente trabalho é um subconjunto de uma base formada pelo ciclo dos 24 Prelúdios, Opus 28, de Frédéric Chopin, na execução de 27 artistas.

A Tabela 4.1 apresenta o conjunto de intérpretes e execuções cujos *onsets* foram anotados e devidamente verificados conforme o descrito no Capítulo 3.

A Tabela 4.2 informa o andamento, o compasso e o nível de marcação (TL, de *Track Level*) para cada um dos prelúdios anotados.

Tabela 4.1: Interpretações expressivas com *onsets* anotados.

Intérprete	Prelúdio											
	1	2	3	4	5	6	7	9	10	11	19	
Alfred Cortot (1925)	●					●	●					●
Alfred Cortot (1934)	●					●	●					●
Alfred Cortot (1942)	●						●					
Alfred Cortot (1955)	●						●					
Alfred Cortot (1957)	●						●					
Alicia de Larrocha	●						●					
Adam Harasiewicz	●						●					
Arthur Moreira Lima	●						●					
Abdel Rahman El Bacha	●						●					
Arthur Rubinstein	●						●					
Bella Davidovich	●						●					
Claudio Arrau	●					●	●					●
Cyprien Katsaris	●						●					
François-René Duchâble	●	●	●	●	●	●	●	●	●	●		
Guiomar Novaes	●						●					
Garrick Ohlsson (1974)	●						●					
Garrick Ohlsson (1995)	●						●					
Ivo Pogorelich	●						●					
Martha Argerich	●						●					
Maria João Pires	●						●					
Murray Perahia	●						●					
Maurizio Pollini	●						●					
Miguel Proença	●						●					
Nelson Freire	●						●					
Nikita Magaloff	●						●					
Rafał Blechacz	●						●					
Shura Cherkassky	●						●					
Sergio Daniel Tiempo	●						●					
Samson François	●						●					
Vladimir Ashkenazy	●						●					
Wojciech Świątała	●						●					
Yevgeniy Kissin	●						●					

Tabela 4.2: Informações sobre os prelúdios anotados.

Prelúdio	Andamento	Compasso	TL
1	Agitato	$\frac{2}{8}$	$\frac{1}{8}$
2	Lento	$\frac{2}{2}$	$\frac{1}{8}$
3	Vivace	$\frac{2}{2}$	$\frac{1}{16}$
4	Largo	$\frac{2}{2}$	$\frac{1}{8}$
5	All ^o molto	$\frac{3}{8}$	$\frac{1}{16}$
6	Lento assai	$\frac{3}{4}$	$\frac{1}{8}$
7	Andantino	$\frac{3}{4}$	$\frac{1}{4}$
9	Largo	$\frac{4}{4}$	$\frac{1}{4}$
10	All ^o molto	$\frac{3}{4}$	$\frac{1}{4}$
11	Vivace	$\frac{6}{8}$	$\frac{1}{8}$
19	Vivace	$\frac{3}{4}$	$\frac{1}{4}$

Todas as gravações foram gentilmente fornecidas pelo professor Luiz Wagner, a partir de sua coleção pessoal, e estão amostradas a $f_s = 44\,100$ Hz.

4.2 Curvas de Andamento e de Intensidade

Como foi visto no Capítulo 2, é interessante caracterizar uma interpretação expressiva através dos acentos agógicos e dinâmicos na sua execução. O Capítulo 3 descreve o ferramental necessário para, a partir dos *onsets* anotados, extraírem-se as curvas de andamento e de intensidade percebida.

Dependendo do nível de marcação, as curvas de andamento podem ser mais ou menos “nervosas”. Já as curvas de dinâmica, que avaliam o nível de audibilidade de uma região grande¹, são mais suaves. Assim sendo, é possível suavizar as curvas de andamento através da equação a diferenças

$$y(n) = \gamma x(n) + (1 - \gamma)y(n - 1), \quad (4.1)$$

¹O intervalo de 3 s ao redor do *onset* é considerado suficiente para que um ser humano “perceba” o volume de um sinal de áudio.

utilizando $\gamma = 0,5$ e o comando `filtfilt` do MATLAB para garantir uma filtragem sem distorção de fase.

Uma curva de andamento suavizada simula a percepção do andamento, uma vez que o ser humano tende a subestimar longas pausas e a superestimar sequências rápidas de notas, de acordo com o pesquisador Simon Dixon em [32].

Para as análises realizadas neste trabalho, este tipo de suavização da curva de andamento somente será utilizado com a técnica da Seção 4.4.

4.3 Similaridade

Em [33], entende-se o processo de interpretação como uma leitura simultânea de dois conjuntos de instruções: as instruções escritas, *i.e.*, notadas na partitura, e as “regras implícitas”. Este último conjunto, que é transmitido “de ouvido” entre os intérpretes e pode também ser ensinado pelos professores a seus alunos, é composto pelas ferramentas de expressão, como foi visto no Capítulo 2. Segundo [33], estas sugestões não-escritas são aplicáveis a apenas uma peça, ao trabalho de um compositor, a um gênero musical ou, até mesmo, a uma era musical inteira, e podem ser uma combinação de várias fontes (*e.g.*, estilos de diferentes pianistas famosos) ou uma “novidade” na contramão das interpretações convencionais.

Assim, é interessante estudar a similaridade das execuções de uma mesma peça por diferentes intérpretes, utilizando como parâmetro de comparação as curvas de andamento e dinâmica extraídas. A técnica de *scape plots* permite executar essa análise através do coeficiente de correlação de Pearson.

4.3.1 Coeficiente de Correlação de Pearson

O coeficiente de correlação entre duas sequências pode ser usado para apontar quão similares essas sequências são. O coeficiente de correlação de Pearson, definido como

$$r(x, y) = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_n (x_n - \bar{x})^2 \sum_n (y_n - \bar{y})^2}}, \quad (4.2)$$

assume valores entre -1 e 1 , e pode ser interpretado como a seguir:

- $r(x, y) = 1$: existe uma relação linear exata entre x e y , e quando x cresce, y cresce com essa relação;
- $r(x, y) = -1$: existe uma relação linear exata entre x e y , e quando x cresce, y decresce com essa relação;
- $r(x, y) = 0$: não existe qualquer relação linear entre x e y ;
- Para outros valores de $r(x, y)$ positivos, o coeficiente de correlação descreve a tendência que as sequências têm de estar acima ou abaixo de suas respectivas médias, conjuntamente. Os valores negativos de $r(x, y)$ dão conta do contrário, *i.e.*, $(x - \bar{x})$ e $(y - \bar{y})$ apresentam mais frequentemente sinais opostos.

O coeficiente de correlação pode, portanto ser utilizado para verificar a similaridade de interpretações de uma mesma peça. Foi de grande uso nos episódios de fraude em coletâneas dos pianistas Sergio Fiorentino e Joyce Hatto, auxiliando a identificação dos verdadeiros pianistas por trás das execuções [34, 35]. Nos dois casos, as faixas copiadas de outros artistas foram minimamente tratadas e tiveram o andamento alterado (acelerado ou reduzido). Como o coeficiente de correlação “normaliza” a sequência, é imune a este tipo de manipulação e pôde apontar com segurança as origens de cada faixa.

4.3.2 *Scape Plots*

Associar a similaridade das sequências de andamento (ou de intensidade) S_k e $S_{k'}$ a um único número significa ignorar os detalhes de interpretação de cada frase musical que compõe a peça. Para avaliar a similaridade localmente, podemos separar estas sequências em pedaços e medir a correlação de cada pedaço com seu correspondente na outra interpretação.

Por exemplo, sejam $S_k = \{a, b, c, d, e, f\}$ e $S_{k'} = \{a', b', c', d', e', f'\}$, um par de sequências com os valores do andamento em cada *onset* de duas interpretações de uma mesma peça. Podemos calcular a correlação entre as subsequências $\{a, b, c\}$ e $\{a', b', c'\}$, bem como para $\{d, e\}$ e $\{d', e'\}$ ou ainda para $\{c, d, e, f\}$ e $\{c', d', e', f'\}$. De fato, avaliando a correlação de todas as subsequências pareadas, podemos estudar

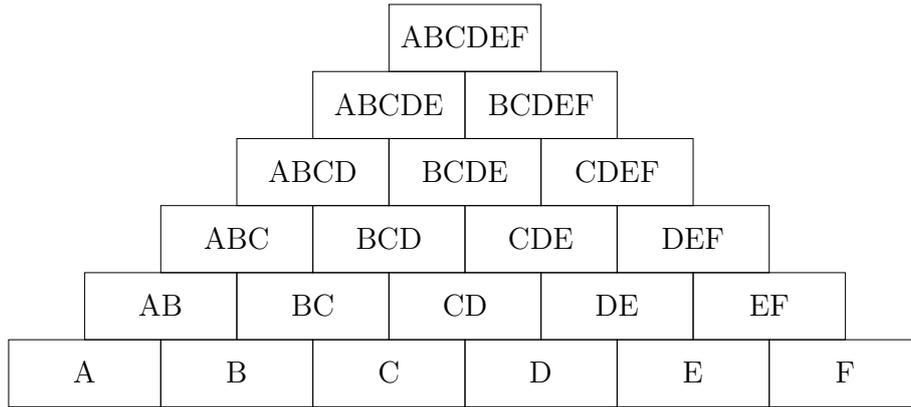


Figura 4.1: Domínio da técnica de *scape plotting*.

a similaridade entre duas interpretações em cada trecho de uma peça musical e com a abrangência que quisermos.

O método proposto em [33] avalia o coeficiente de correlação de Pearson para todas as combinações possíveis dos elementos das sequências, mantendo sua ordenação. O resultado é um gráfico hierárquico “triangular”, no qual quanto mais alto, maior o número de elementos das sequências de andamento ou de intensidade utilizados para o cálculo.

A Figura 4.1 apresenta um esboço desta técnica, chamada de *scape plotting*, para as sequências S_k e $S_{k'}$. A célula A representa a avaliação da correlação entre os elementos a e a'^2 . Já a célula AB corresponde ao valor do coeficiente de correlação para as subsequências $\{a, b\}$ e $\{a', b'\}$, e assim por diante até o topo da pilha, onde as duas sequências são comparadas como um todo.

Este mapa de correlação pode ser montado a partir da comparação de duas sequências e, então, utiliza-se uma graduação de cor para descrever os valores dos coeficientes de correlação em cada célula, *e.g.*, do preto ($r(S_k, S_{k'}) = -1$) ao branco ($r(S_k, S_{k'}) = 1$). O mapa também pode ser traçado na comparação de um intérprete contra todos os outros. Neste caso, a cor de cada célula é associada ao intérprete que apresenta maior coeficiente de correlação com o intérprete em análise³.

²Avaliar o coeficiente de correlação entre dois elementos resulta em indeterminação. A análise da correlação de dois elementos, sem a extração da média e do desvio-padrão da sequência, também não é interessante. Assim, a base da pilha na Figura 4.1 não foi implementada para este método.

³Alternativamente, podem ser utilizados os critérios de menor coeficiente de correlação ou de coeficiente com o menor valor absoluto (descorrelação).

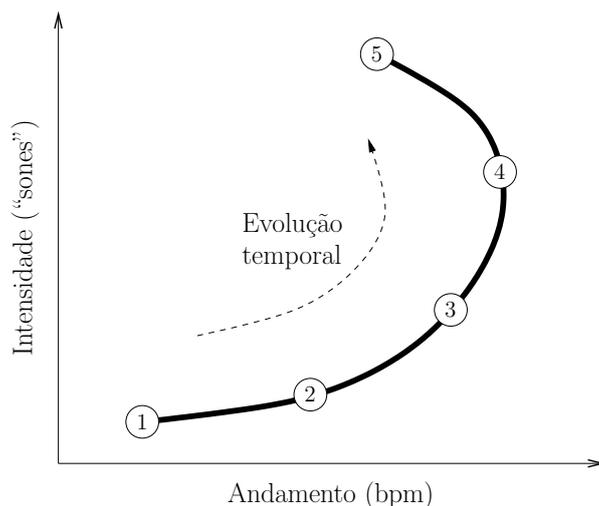


Figura 4.2: Exemplo de trajetória de uma execução no espaço bidimensional da dinâmica e da agógica.

Técnicas de visualização diferentes podem ser utilizadas. No caso dos mapas multi-intérpretes, é possível realizar uma gradação da cor de cada célula para informar quão forte é a correlação. Pode-se fazer um mapa para um único intérprete medindo a correlação entre as sequências de andamento e de intensidade ao longo da peça. Também para um único intérprete, utilizando apenas a sequência de andamento (ou de intensidade), pode-se calcular a média dos valores da sequência para avaliar o andamento médio (ou a intensidade média) nos diversos trechos da música.

4.4 Espaço de Dinâmica e Agógica

Em [36, 37], desenvolve-se uma ferramenta de visualização de uma interpretação expressiva no espaço bidimensional da dinâmica e da agógica. Assim, as curvas de andamento e de intensidade para uma execução são suavizadas e dispostas graficamente como na Figura 4.2: o andamento no eixo horizontal e a intensidade no eixo vertical. A visualização é, na verdade, uma animação exibida em sincronia com a música.

Esta é uma representação em alto nível das estratégias de interpretação em uma dada execução e permite encarar os dois parâmetros — agógica e dinâmica — simultaneamente para analisar a complexa relação entre os dois. A ferramenta nos remete às figuras representativas do movimento interno de Truslit (*cf.* Capítulo 2),

mas aqui são as informações de andamento e de intensidade extraídas da interpretação expressiva que atuam, e não a sensação de movimento.

4.5 Análise 1: Estatísticas das Interpretações

O objetivo desta primeira análise é obter algumas informações estatísticas sobre as interpretações expressivas anotadas.

A partir das curvas de agógica de cada execução, pode-se obter os valores médio, máximo e mínimo entre intérpretes do andamento médio de cada prelúdio. Para tanto, primeiro calcula-se o andamento médio (ou a intensidade média) deste prelúdio na execução de cada intérprete e, então, analisam-se média, máximo e mínimo. A Tabela 4.3 apresenta os resultados para os prelúdios anotados.

Tabela 4.3: Valores médio, máximo e mínimo entre intérpretes dos andamentos médios para os prelúdios anotados.

Prelúdio	Execuções	Andamento (bpm, ref. ♩)		
		Médio	Máximo	Mínimo
1	32	70,0	85,3	57,2
2	1	46,0	46,0	46,0
3	1	181,3	181,3	181,3
4	1	48,5	48,5	48,5
5	1	145,9	145,9	145,9
6	4	43,9	47,8	35,5
7	32	72,4	98,6	42,0
9	1	39,2	39,2	39,2
10	1	137,4	137,4	137,4
11	1	149,3	149,3	149,3
19	3	197,6	213,0	184,7

Comparando as Tabelas 2.1, 4.2 e 4.3 para os prelúdios 2–4, 7, 9, 11 e 19, que têm indicações de andamento simples, pode-se notar que as maiores discrepâncias relativas entre os valores médios e as referências tabeladas são as dos prelúdios 3 e 19,

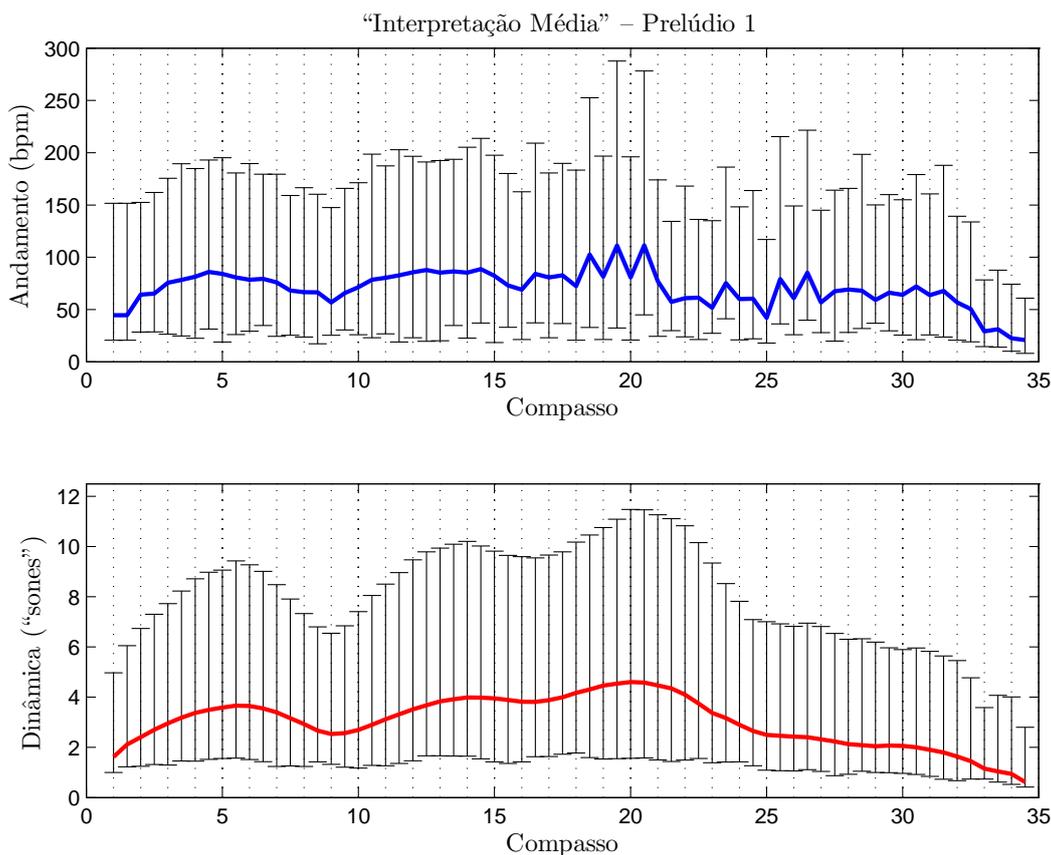


Figura 4.3: “Interpretação média” para o prelúdio 1. A barra em cada ponto da curva dá conta dos valores máximo e mínimo para aquele *onset* em todas as interpretações.

13,3% e 23,5%, respectivamente. O prelúdio 9 apresenta 10,9% de discrepância em relação ao limite inferior da indicação metronômica para andamentos *largo*. Deve-se notar que, para esses prelúdios, são poucos os dados disponíveis para a estatística.

É interessante, também, inferir uma “interpretação média” calculando-se os valores médio, máximo e mínimo para os parâmetros em cada *onset*. As Figuras 4.3 e 4.4 apresentam estes resultados para os prelúdios 1 e 7, respectivamente.

Para o prelúdio 1, observa-se que as maiores dispersões sobre a média, em termos tanto do valor máximo quanto do desvio padrão (que não é exibido na figura), ocorrem no segundo *onset* anotado dos compassos 18–20 e 25–26. Nestes compassos são encontradas, justamente, variações nas subdivisões rítmicas que tendem a ser exageradas pelos intérpretes⁴.

⁴Nestes compassos e também no compasso 23, há a ocorrência de quáteras na razão 5:4. O resto da peça apresenta apenas grupos de quáteras em 3:2.

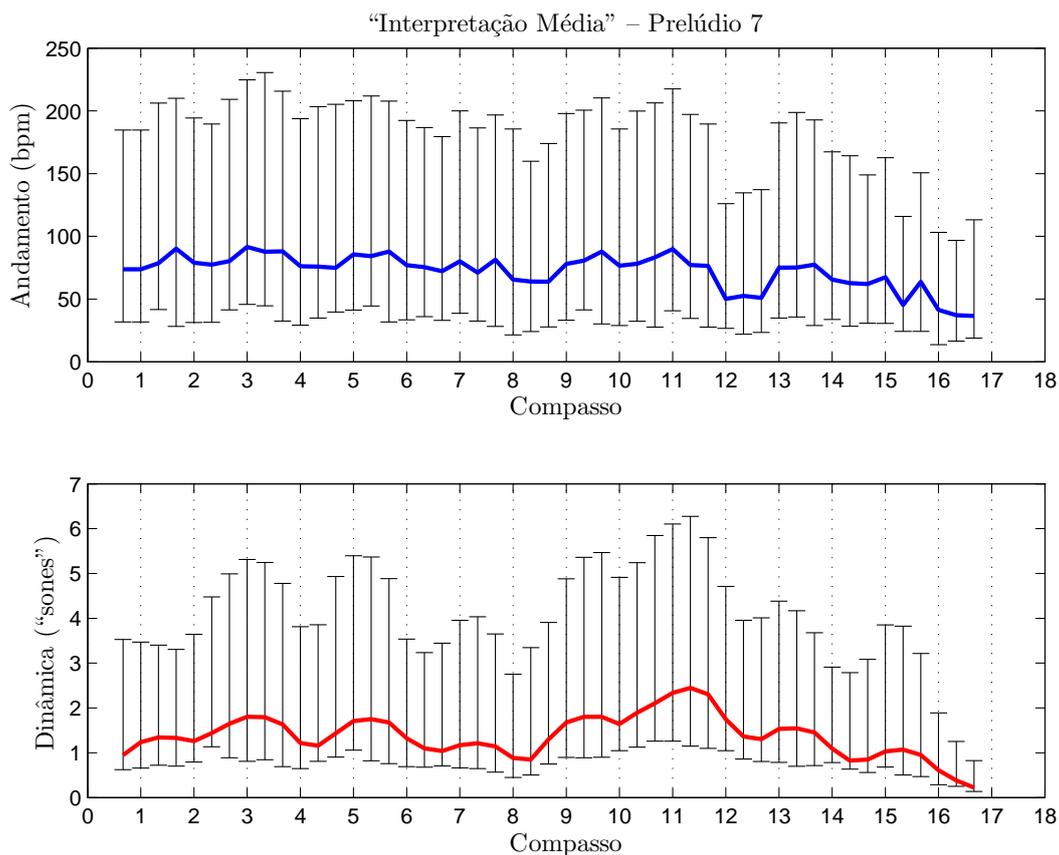


Figura 4.4: “Interpretação média” para o prelúdio 7. A barra em cada ponto da curva dá conta dos valores máximo e mínimo para aquela *onset* em todas as interpretações.

De maneira geral, das Figuras 4.3 e 4.4 percebe-se que a razão entre os valores máximo e mínimo de andamento e de dinâmica é, constantemente, maior do que 4. Isso significa que, em cada *onset*, a execução mais “forte” soa entre 4 e 11 vezes mais “forte” que a mais “suave”. Praticamente o mesmo pode ser dito da velocidade dos intérpretes ao tocarem aquela nota, com relação às execuções mais “rápida” e mais “lenta”, para os dois prelúdios.

Para o prelúdio 7, avaliando-se o coeficiente de correlação entre os andamentos da “interpretação média” e de cada uma das interpretações, considerando-se a peça inteira, chega-se à conclusão que o melhor correlacionado é M. PERAHIA, conforme é demonstrado na Figura 4.5, que exhibe as duas curvas de andamento.

Também para o prelúdio 7, a Figura 4.6 apresenta as interpretações de S. D. TIEMPO e G. OHLSSON (1995), exemplificando a variação nos acentos agógicos e no andamento médio para esta peça musical.

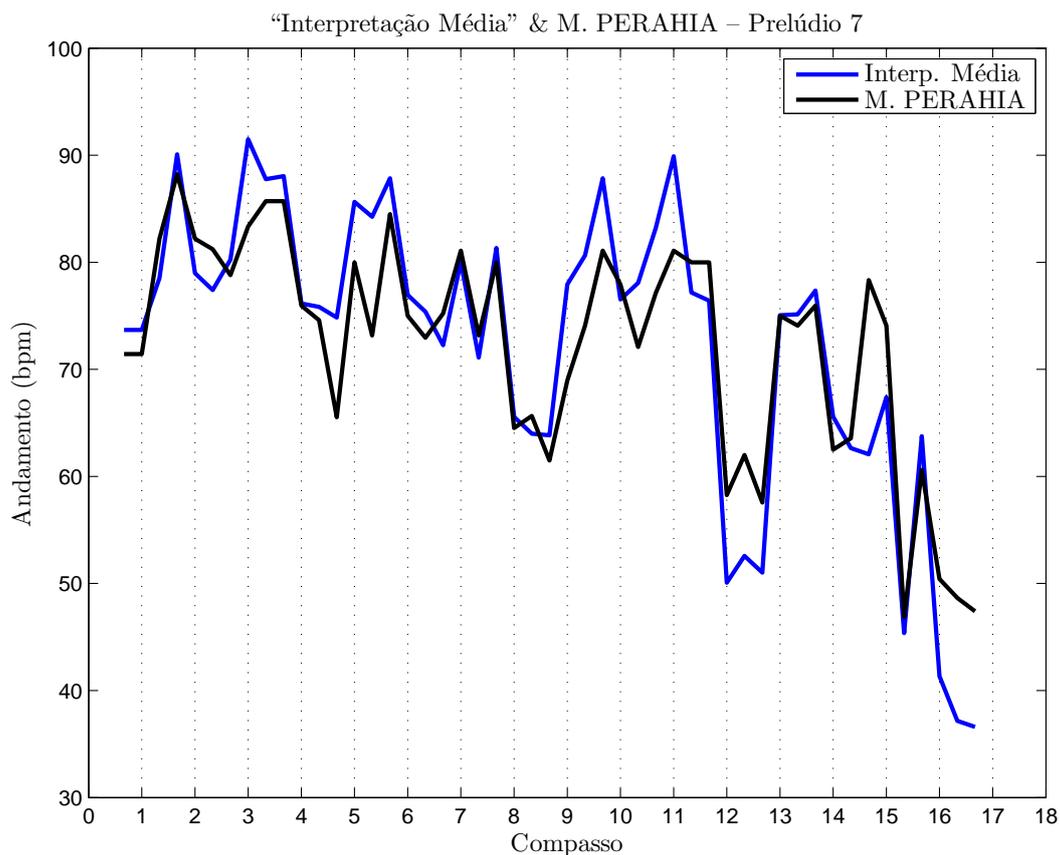


Figura 4.5: Curvas de andamento para as interpretações “média” e de M. PERAHIA para o prelúdio 7. Considerando-se a peça inteira, este foi o valor melhor correlacionado ($r = 0,92$).

Por fim, conhecendo-se o ano de cada execução, pode-se analisar como as estratégias de expressividade variaram ao longo do tempo. A Figura 4.7 apresenta a variação temporal da média da curva de dinâmica (intensidade média) para os prelúdios 1, 6 e 7. Os dados apresentam uma tendência decrescente, *i.e.*, a intensidade média das interpretações tem diminuído com os anos. Nenhuma tendência foi observada para a média do andamento.

A Figura 4.8 apresenta análise e resultados similares, mas, neste caso, a média da curva de dinâmica é exibida contra o ano de nascimento do intérprete. Novamente, os valores médios para as curvas de andamento não indicaram nenhuma tendência.

Como foi visto nesta seção, a partir das curvas de andamento e de intensidade, pode-se obter informações importantes que auxiliam na compreensão das estratégias de execução empregadas pelos intérpretes.

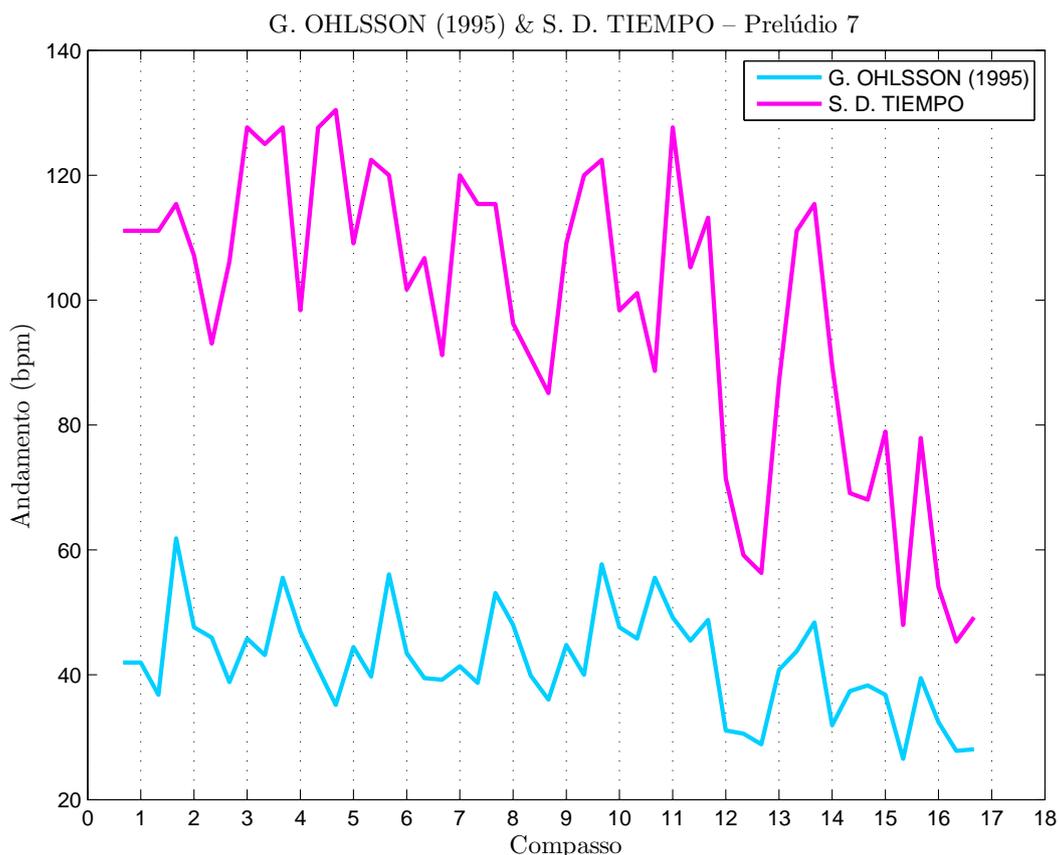


Figura 4.6: Curvas de andamento das interpretações de S. D. TIEMPO e de G. OHLSSON (1995) para o prelúdio 7.

4.6 Análise 2: Mapas de Correlação

O objetivo desta seção é utilizar os mapas de correlação (*scape plots*) para avaliar a similaridade do pianista A. CORTOT com ele mesmo (em outras execuções) e outros intérpretes nos prelúdios 1 e 7, em termos do andamento.

Primeiramente, geram-se os mapas de correlação para o prelúdio 7 conforme visto na Seção 4.3. A Figura 4.9 apresenta o conjunto de mapas para as execuções do prelúdio 7 por A. CORTOT nos anos 1925–1957. Pode-se perceber que A. CORTOT apresenta uma razoável similaridade com N. MAGALOFF, que aparece nas Figuras 4.9(a)–(c), sobretudo no início da peça. Na Figura 4.9(c), nota-se uma forte consonância entre A. CORTOT e a pianista M. ARGERICH. Os pianistas M. PERAHIA, M. J. PIRES e G. NOVAES também apresentam correlação com A. CORTOT, em trechos da peça. Como era de se esperar, as maiores similaridades são de A. CORTOT com ele mesmo, *e.g.*, entre as execuções de 1955 e 1957.

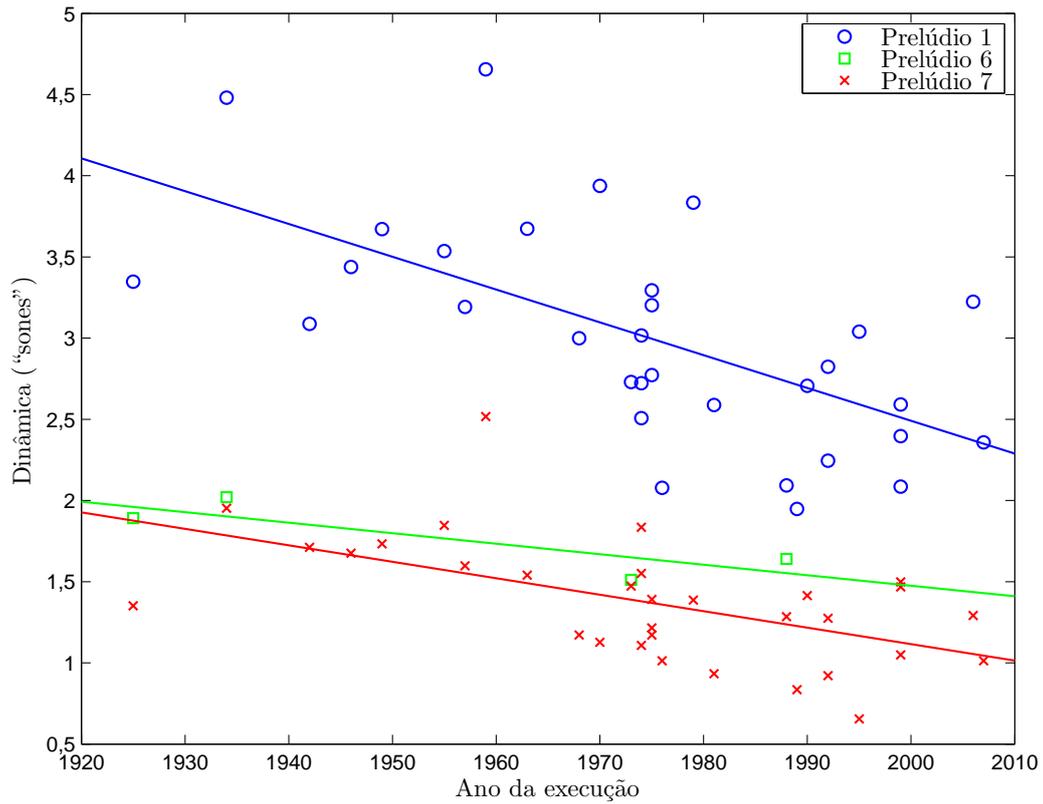


Figura 4.7: Média das curvas de dinâmica por ano da execução.

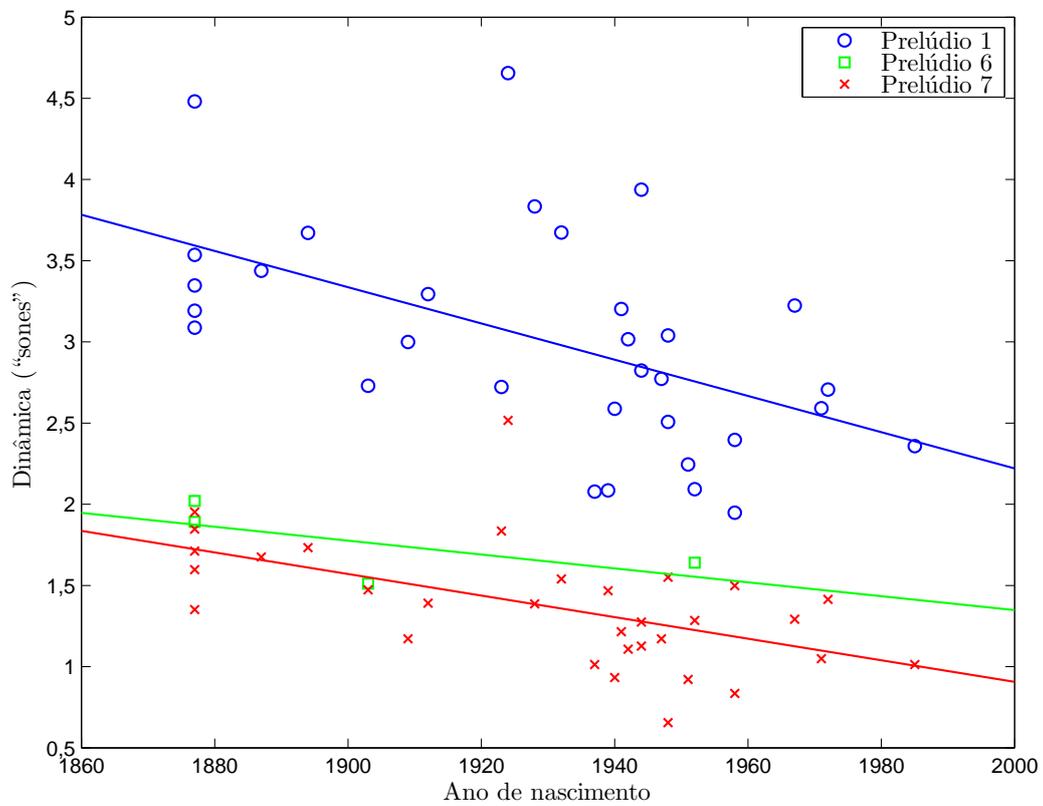
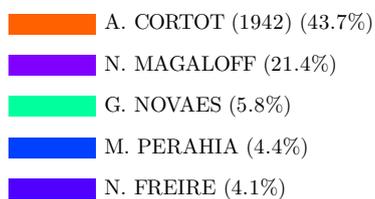
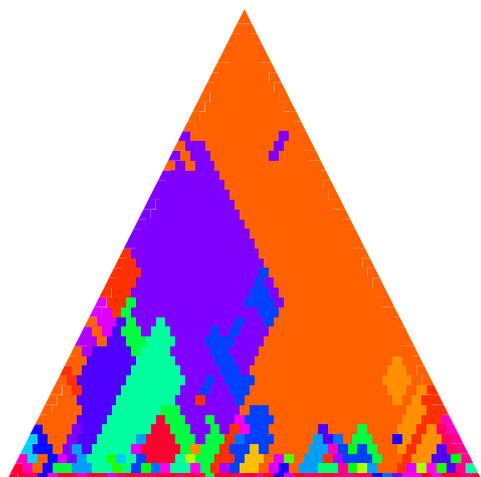


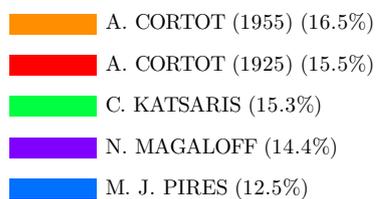
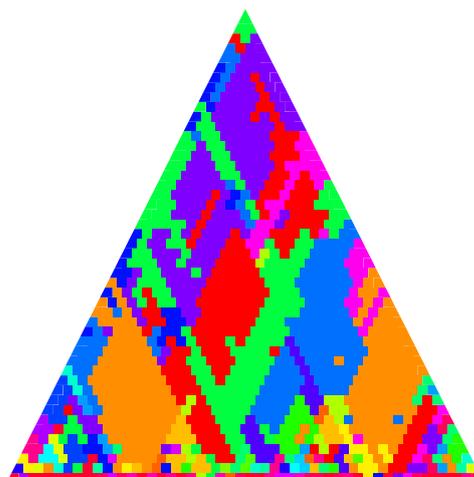
Figura 4.8: Média das curvas de dinâmica por ano de nascimento do intérprete.

A. CORTOT (1925) – Prelúdio No.7



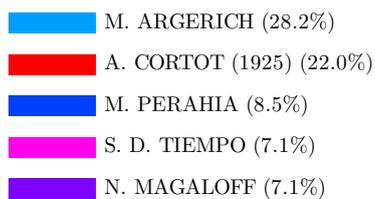
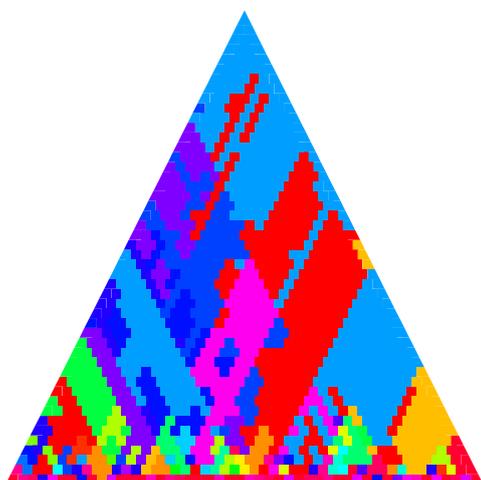
(a)

A. CORTOT (1934) – Prelúdio No.7



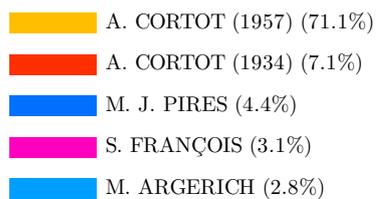
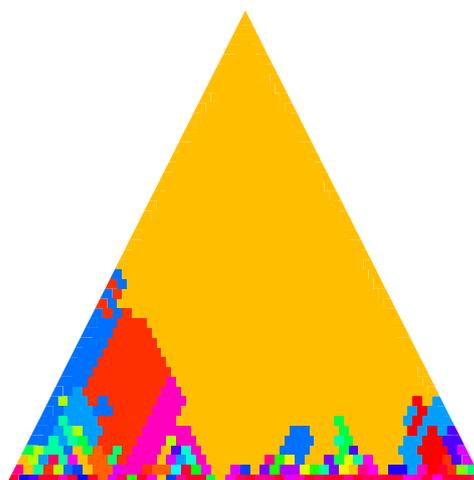
(b)

A. CORTOT (1942) – Prelúdio No.7



(c)

A. CORTOT (1955) – Prelúdio No.7



(d)

Figura 4.9: Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT em: (a) 1925; (b) 1934; (c) 1942; (d) 1955.

A. CORTOT (1957) – Prelúdio No.7

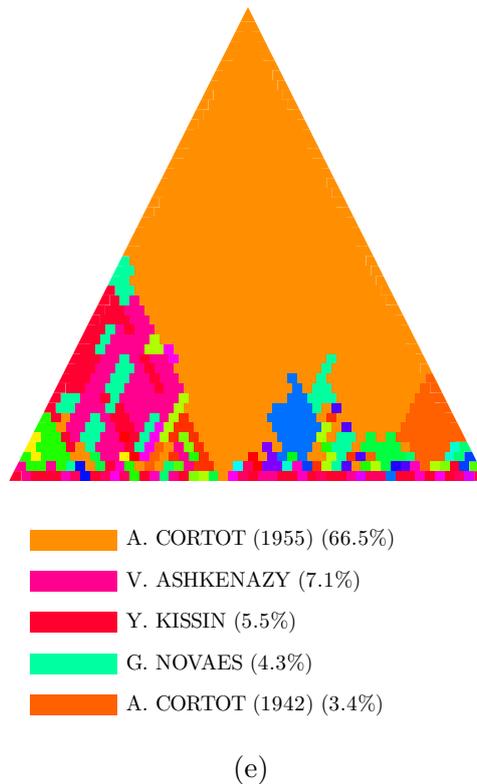
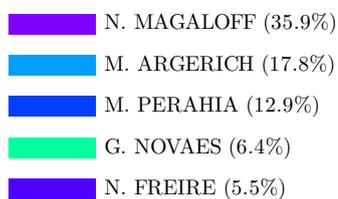
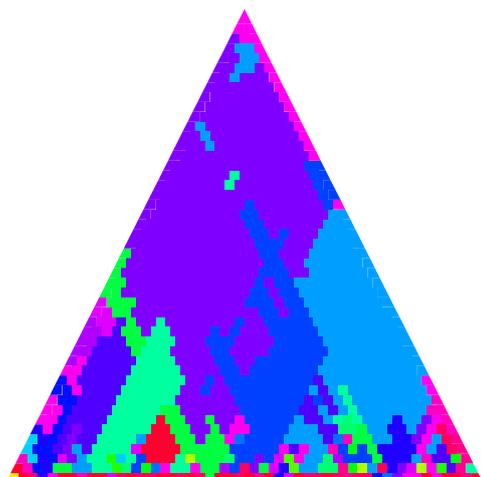


Figura 4.9: (*cont.*) Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT em: (e) 1957.

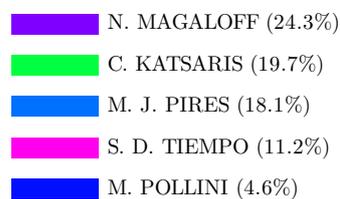
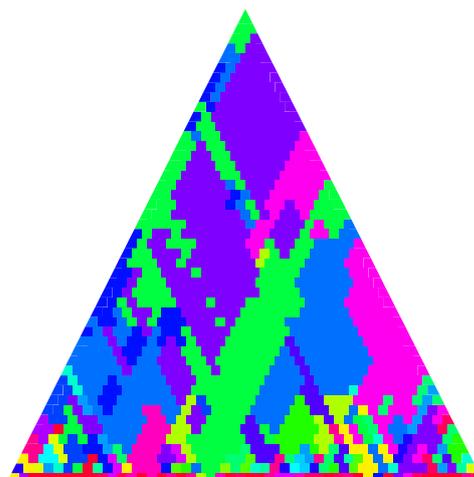
Ao analisar-se A. CORTOT contra ele mesmo, perdemos as informações de quão similares são suas interpretações com as dos demais intérpretes. Assim, é possível “descamar” os mapas de correlação gerados através da rotina `scapePlot` (*cf.* Apêndice A), ignorando a correlação de suas interpretações com outros anos de execução do mesmo prelúdio. A Figura 4.10 apresenta o conjunto de mapas para as mesmas execuções do prelúdio 7, agora sem considerar outras execuções de A. CORTOT em anos diferentes. De novo, é notável a similaridade com N. MAGALOFF em níveis mais gerais da peça. Observando as Figuras 4.10(a), (c) e (d), nota-se também alta correlação com M. ARGERICH, em especial nos momentos finais da execução. Provavelmente, essas correlações podem ser explicadas através dos relacionamentos pessoais e profissionais dos intérpretes. Em 1922, Nikita Magaloff iniciou seus estudos com Isidor Philipp, amigo pessoal de Alfred Cortot, no Conservatório de Paris. O próprio Cortot foi professor no Conservatório até 1923. Anos mais tarde, Magaloff seria tutor de Martha Argerich.

A. CORTOT (1925) – Prelúdio No.7



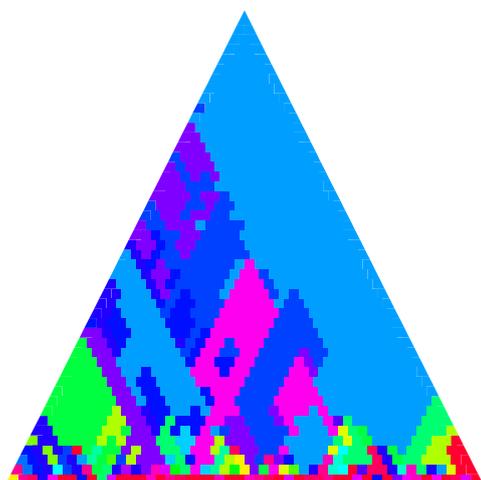
(a)

A. CORTOT (1934) – Prelúdio No.7



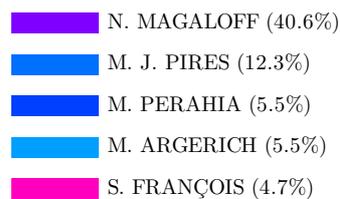
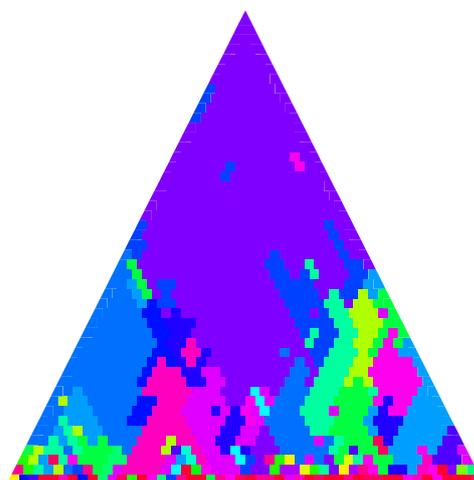
(b)

A. CORTOT (1942) – Prelúdio No.7



(c)

A. CORTOT (1955) – Prelúdio No.7



(d)

Figura 4.10: Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT sem considerar outros anos de execução.

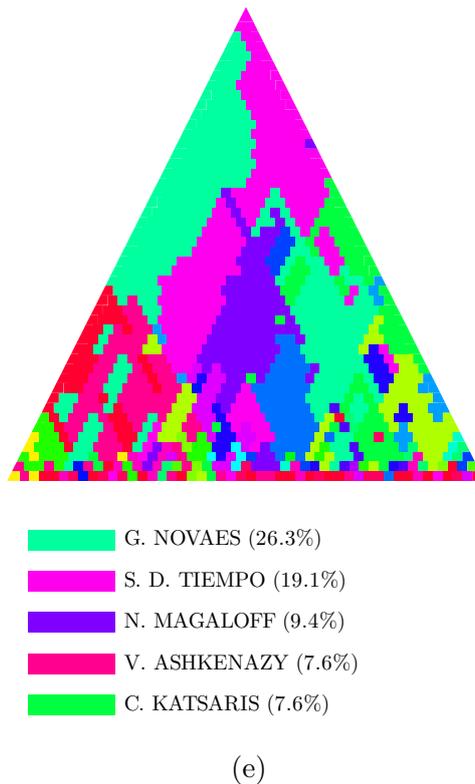
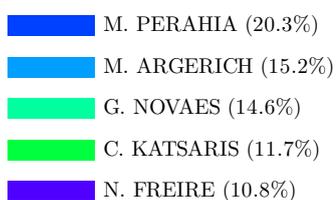
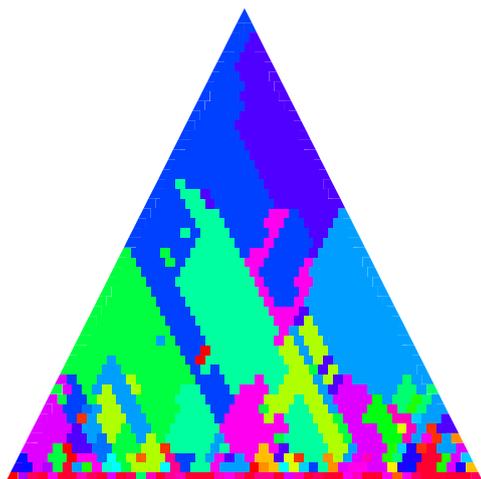


Figura 4.10: (*cont.*) Mapas de correlação para o andamento das interpretações do prelúdio 7 por A. CORTOT sem considerar outros anos de execução.

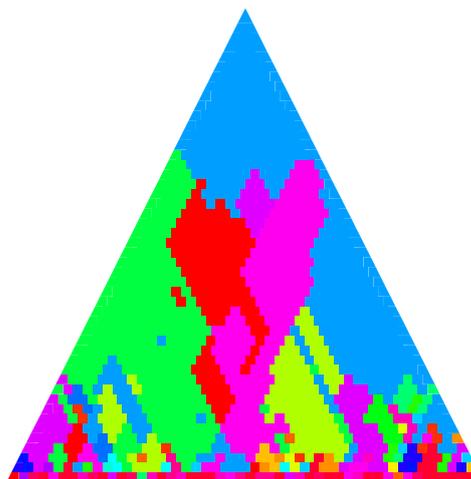
Calculando-se os mapas de correlação para N. MAGALOFF e M. ARGERICH, contudo, não percebemos uma simetria nesta correlação. Nos dois casos, é necessário desconsiderar os intérpretes brasileiros N. FREIRE⁵ e G. NOVAES. Para N. MAGALOFF, desconsidera-se também o americano M. PERAHIA e, para M. ARGERICH, o brasileiro A. M. LIMA. A Figura 4.11 exhibe os mapas para N. MAGALOFF e M. ARGERICH considerando-se todos os pianistas (em (a) e (c)) e após “descamar” alguns intérpretes (em (b) e (d)).

Neste ponto, as correlações com A. CORTOT começam a ficar mais evidentes, mas ainda não são comparáveis às exibidas na Figura 4.10. A interpretação deste fenômeno é que, para o prelúdio 7, embora a interpretação de A. CORTOT seja muito similar às de N. MAGALOFF e M. ARGERICH, as interpretações destes

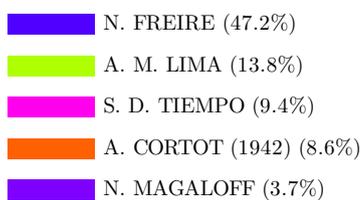
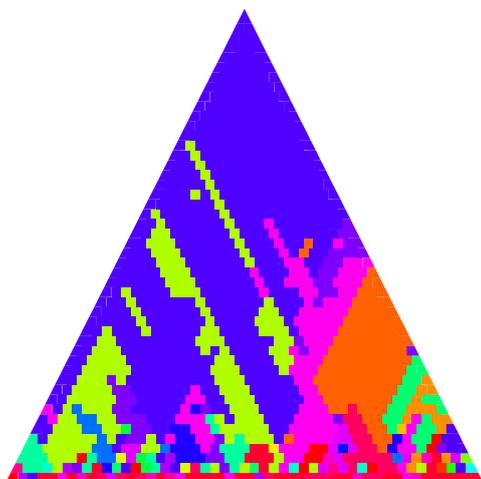
⁵Martha Argerich tem como um de seus grandes amigos o pianista brasileiro Nelson Freire. Os dois se conheceram na década de 60 e, desde então, trabalharam juntos inúmeras vezes. Esta amizade explica a similaridade das execuções dos dois intérpretes, ambas gravadas nos anos 70.



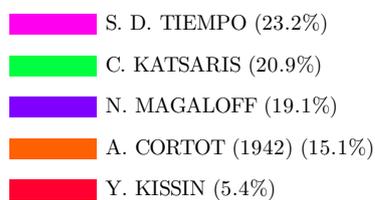
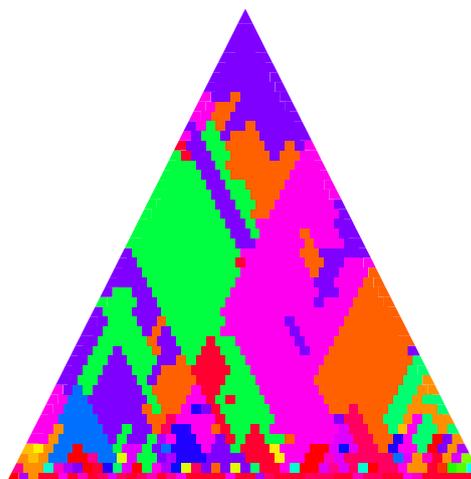
(a)



(b)



(c)



(d)

Figura 4.11: Mapas de correlação para o andamento das interpretações do prelúdio 7 por N. MAGALOFF e M. ARGERICH para: (a) e (c) todos os intérpretes; (b) desconsiderando M. PERAHIA, N. FREIRE e G. NOVAES; (d) desconsiderando N. FREIRE, A. M. LIMA e G. NOVAES.

dois apresentam maior correlação com as execuções de outros intérpretes, como S. D. TIEMPO⁶ e C. KATSARIS.

A Figura 4.12 apresenta os *scape plots* multi-intérpretes para o andamento das interpretações de A. CORTOT do prelúdio 1, nos anos 1925–1955. Em cada mapa de correlação, foram descartadas as outras execuções de A. CORTOT em anos diferentes. Para este prelúdio, percebe-se como a evolução do andamento imposto por A. CORTOT ao longo da peça se manteve com o passar dos anos. A semelhança entre os mapas é maior do que para o prelúdio 7, como visto na Figura 4.10.

Como mencionado na Seção 4.3, pode-se também calcular o mapa de correlação para as sequências de dinâmica utilizando a mesma função `scapePlot`. A Figura 4.13 traz esses resultados para A. CORTOT, desconsiderando-se outras execuções do mesmo. Aqui, nota-se uma forte correlação com as estratégias para a dinâmica empregadas por M. POLLINI e, novamente, M. ARGERICH.

Esta seção abordou a técnica de *scape plots*, que elucida as semelhanças entre as execuções dos intérpretes. Esta técnica permite que todas as subsequências de *onsets* sejam analisadas e auxilia na compreensão das relações interpessoais e profissionais dos pianistas.

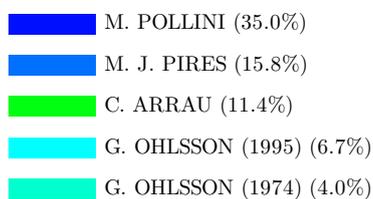
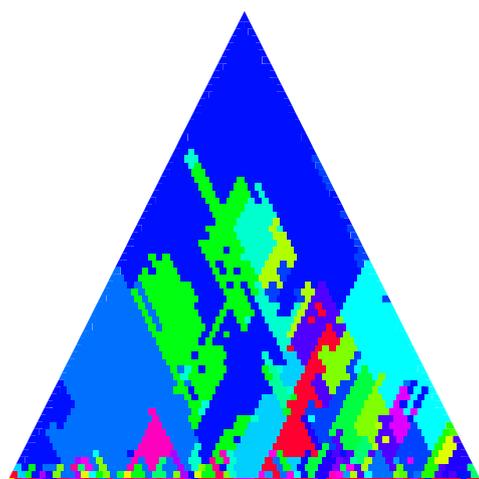
4.7 Análise 3: Modelo de Todd

A técnica de *scape plots* também permite avaliar a validade do modelo de Todd [31]. Segundo este modelo, a relação agógica–dinâmica é tal que, quando o ritmo é acelerado, aumenta-se a intensidade das notas e, quando se toca mais lentamente, toca-se com maior suavidade, *i.e.*, Todd infere que o coeficiente de correlação entre os dois parâmetros é sempre positivo.

Para tanto, adapta-se a função `scapePlot` para realizar o cálculo da correlação apenas entre as sequências de andamento e de intensidade de um mesmo intérprete, e exibir o resultado de forma que os limites -1 e 1 da correlação sejam mapeados em tons de azul e vermelho, respectivamente. Assim, -1 (azul) significa que as sequências apresentam coeficiente de correlação negativo e “caminham” em sentidos opostos. Já 1 (vermelho) significa máxima correlação entre andamento e

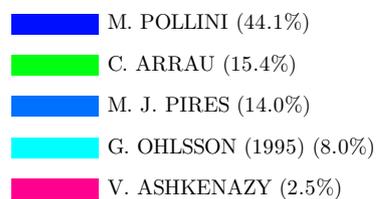
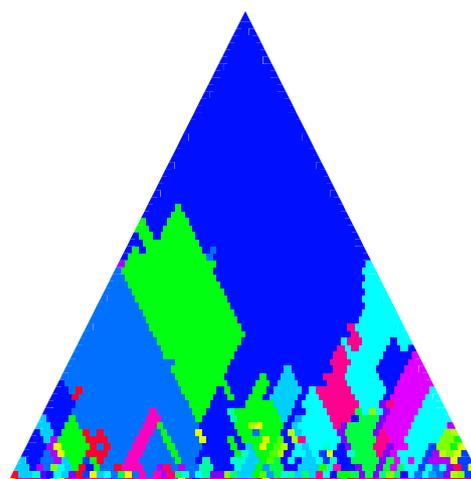
⁶Martha Argerich é mentora do pianista venezuelano Sergio Daniel Tiempo.

A. CORTOT (1925) – Prelúdio No.1



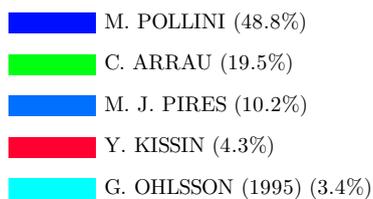
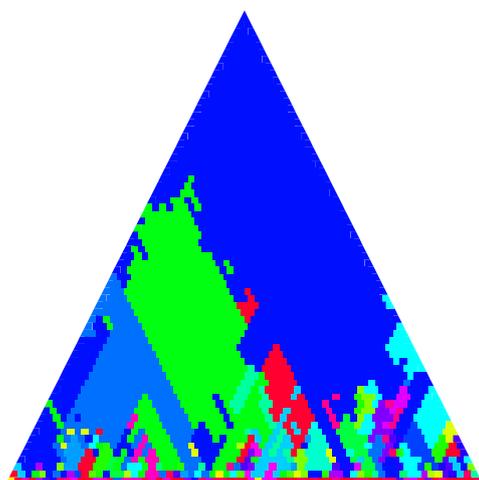
(a)

A. CORTOT (1934) – Prelúdio No.1



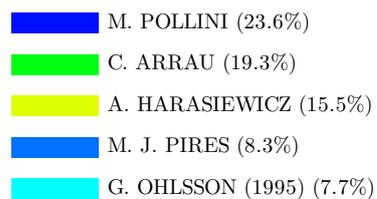
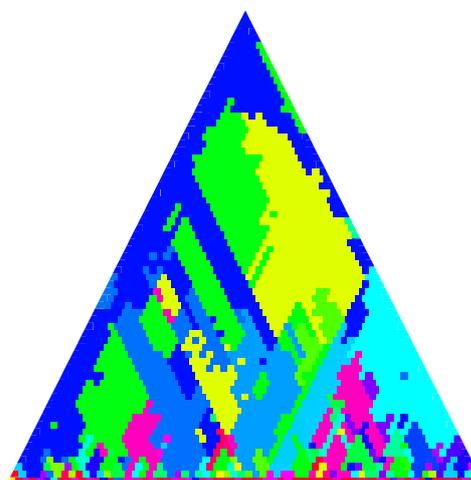
(b)

A. CORTOT (1942) – Prelúdio No.1



(c)

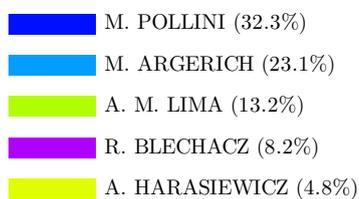
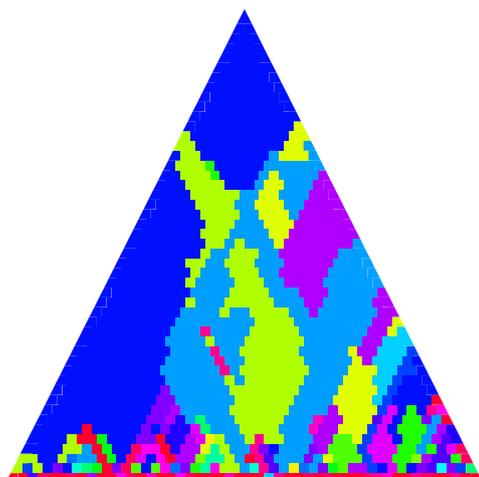
A. CORTOT (1955) – Prelúdio No.1



(d)

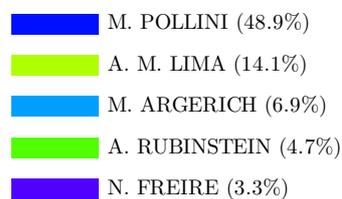
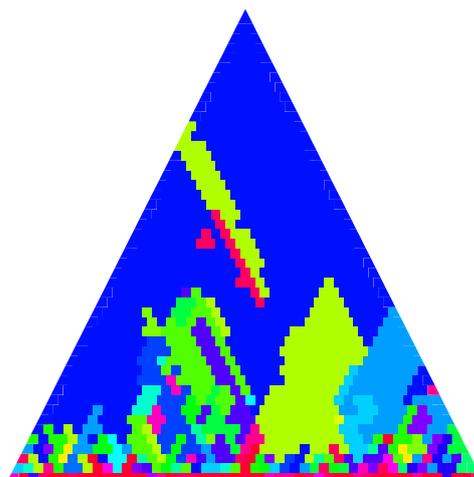
Figura 4.12: Mapas de correlação para o andamento das interpretações do prelúdio 1 por A. CORTOT sem considerar outros anos de execução.

A. CORTOT (1934) – Prelúdio No.7



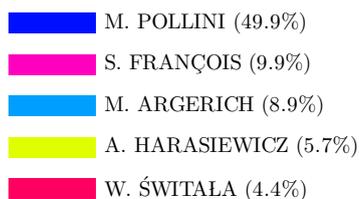
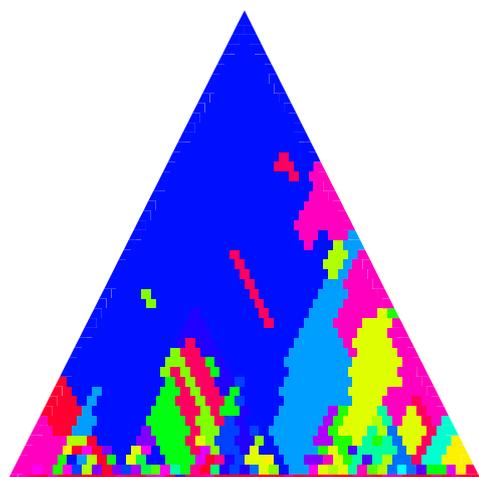
(a)

A. CORTOT (1942) – Prelúdio No.7



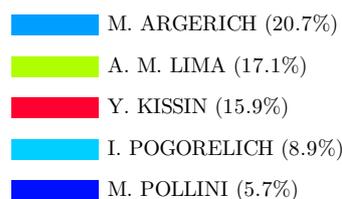
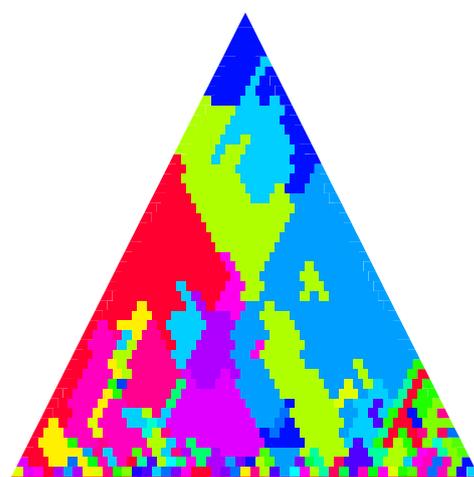
(b)

A. CORTOT (1955) – Prelúdio No.7



(c)

A. CORTOT (1957) – Prelúdio No.7



(d)

Figura 4.13: Mapas de correlação para a dinâmica das interpretações do prelúdio 7 por A. CORTOT sem considerar outros anos de execução.

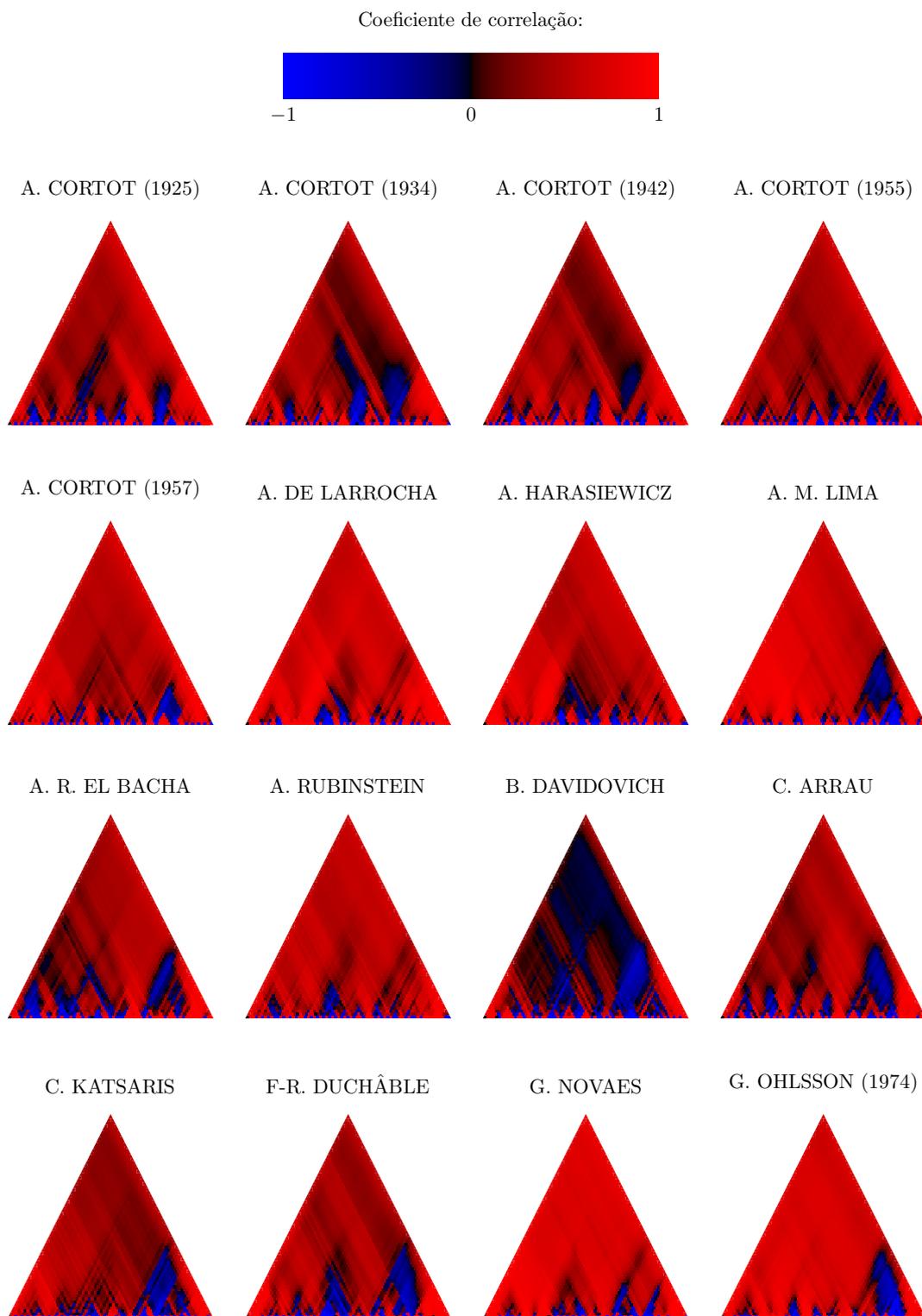


Figura 4.14: Mapas de correlação entre andamento e dinâmica das interpretações do prelúdio 1 para cada intérprete.

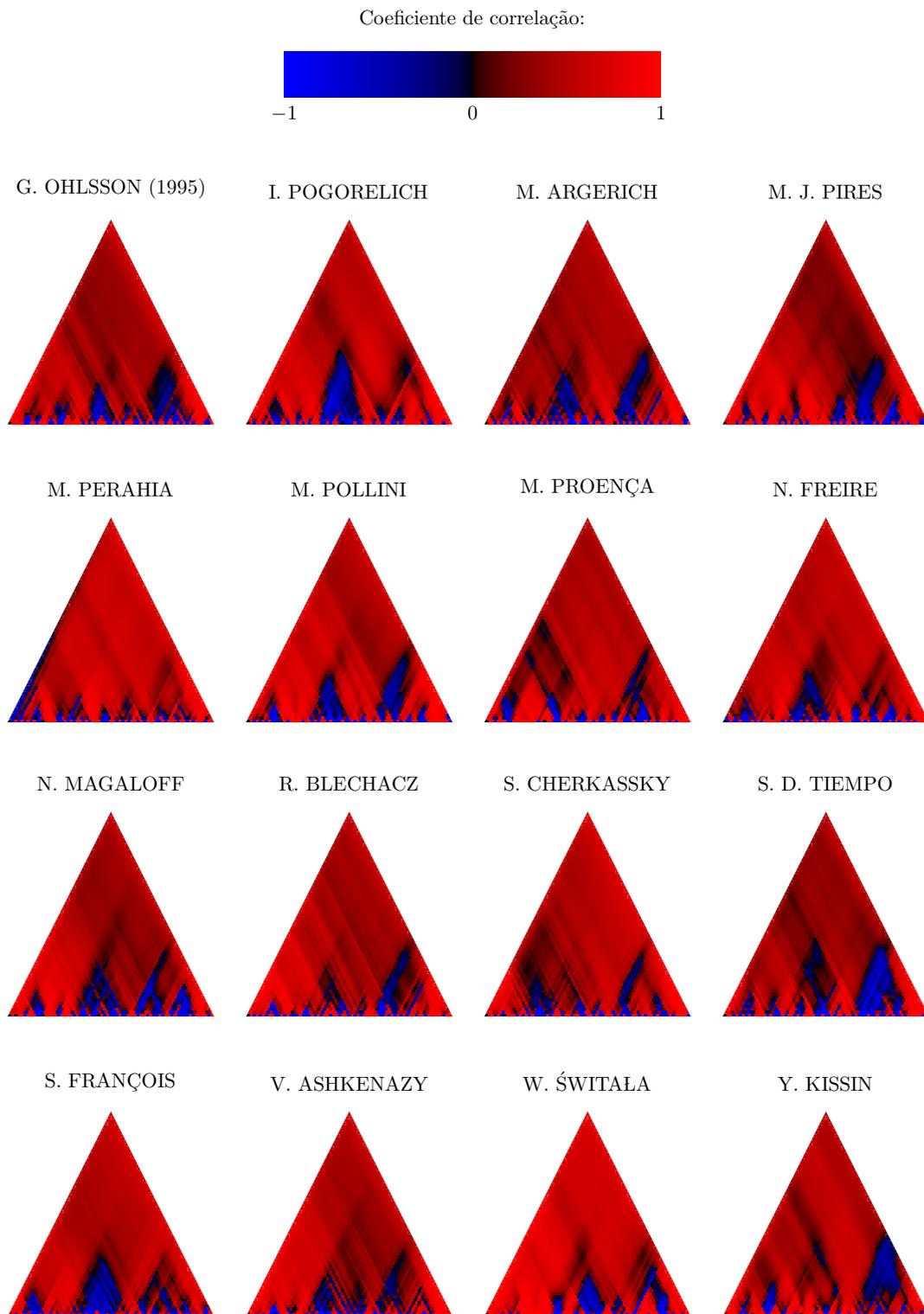


Figura 4.14: (*cont.*) Mapas de correlação entre andamento e dinâmica das interpretações do prelúdio 1 para cada intérprete.

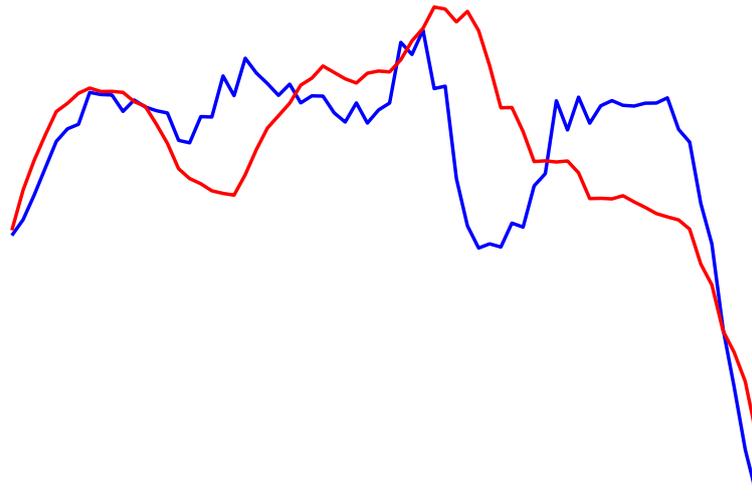


Figura 4.15: Curvas de andamento (suavizada, em azul) e dinâmica (em vermelho) normalizadas para a interpretação do prelúdio 1 por B. DAVIDOVICH.

intensidade. Valores entre -1 e 1 seguem lógicas similares, considerando-se 0 (preto) como desconexão entre as duas sequências.

Avaliando-se a correlação das duas sequências inteiras, *i.e.*, o coeficiente no topo do “triângulo”, encontra-se, em todos os casos, um valor positivo. Contudo, observando-se a Figura 4.14, identificam-se, para os vários intérpretes, diversos trechos (subsequências de *onsets*) para os quais a correlação é próxima de zero ou até mesmo negativa. Isso mostra que o modelo de Todd para a relação agógica–dinâmica é muito simples e pode até ser válido, numericamente, para a peça como um todo, mas ao refinar-se a análise dos trechos, são várias as regiões que vão contra este paradigma. A Figura 4.15 apresenta as curvas de andamento e intensidade na interpretação de B. DAVIDOVICH. É possível notar diferentes tendências das duas sequências, o que explica seu mapa de correlação “escuro” na Figura 4.14.

4.8 Análise 4: Trajetórias do Prelúdio 6

A representação da trajetória de uma execução no espaço bidimensional da dinâmica e da agógica, proposta por Langner e Goebel em [36], permite a visualização simultânea dos parâmetros expressivos considerados essenciais. Assim sendo, essa forma de representação pode auxiliar a compreensão da complexa relação existente entre essas duas ferramentas de expressão.

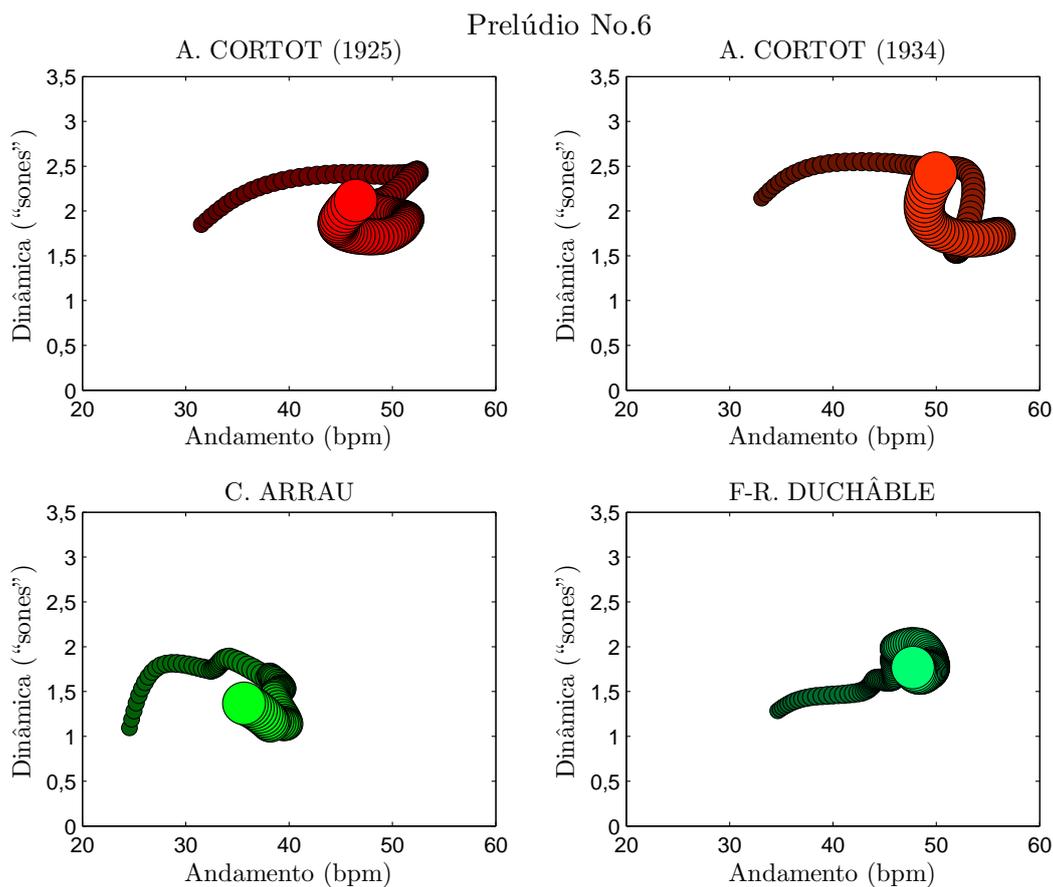


Figura 4.16: Trajetória das execuções no espaço bidimensional da dinâmica e da agógica para os dois primeiros compassos do prelúdio 6. Intérpretes: A. CORTOT (1925), A. CORTOT (1934), C. ARRAU e F-R. DUCHÂBLE.

O objetivo desta seção é apenas ilustrar as capacidades da ferramenta. Para tanto, estuda-se o prelúdio 6 na execução de A. CORTOT (1925 e 1934), C. ARRAU e F-R. DUCHÂBLE.

A Figura 4.16 apresenta a trajetória dos dois primeiros compassos para cada execução. Todos os intérpretes começam acelerando o andamento, com alguma variação dinâmica. Observa-se que, na metade da trajetória (primeiro *onset* do segundo compasso), os intérpretes começam a suavizar as notas, exceto por F-R. DUCHÂBLE, que coloca um pouco mais de força sobre a primeira nota do segundo compasso, em relação ao que vinha executando até aquele ponto. Ao final do segundo compasso, F-R. DUCHÂBLE mantém estáveis o andamento e a intensidade das notas, enquanto C. ARRAU desacelera. Em suas duas execuções, o pianista A. CORTOT acelera e imprime maior força, “antecipando” o terceiro compasso desta peça. Vê-se que, para os primeiros dois compassos, todas as interpretações descrevem

Prelúdio No.6

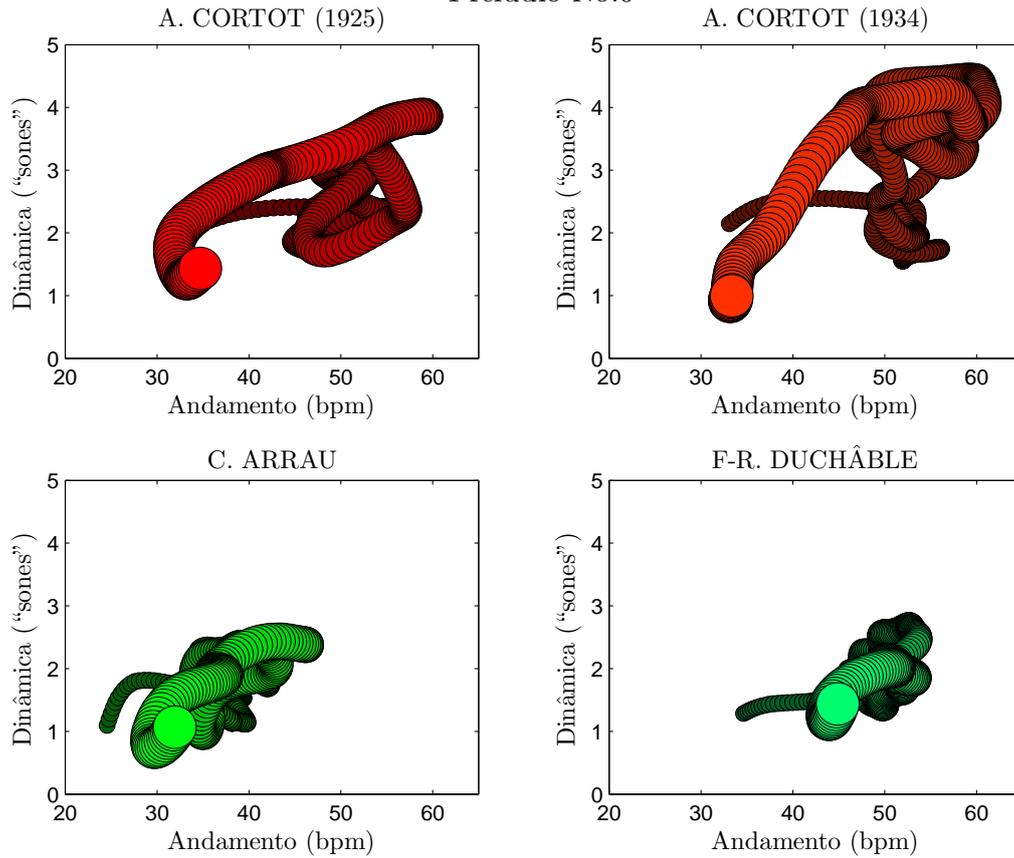


Figura 4.17: Trajetória das execuções no espaço bidimensional da dinâmica e da agógica para os oito primeiros compassos do prelúdio 6. Intérpretes: A. CORTOT (1925), A. CORTOT (1934), C. ARRAU e F-R. DUCHÂBLE.

um movimento anti-horário e, novamente, a relação entre os parâmetros parece ser mais complexa do que o proposto pelo modelo de Todd [31].

Lançando um olhar mais geral sobre as interpretações, a Figura 4.17 apresenta a trajetória para as mesmas execuções, mas, desta vez, para os oito primeiros compassos. Aqui, percebe-se como A. CORTOT faz uso, em larga escala, dos acentos agógicos e dinâmicos. Em termos do andamento, C. ARRAU também mostra grande variação, dobrando e reduzindo à metade seu valor ao longo deste trecho. Embora F-R. DUCHÂBLE pareça apresentar a mesma distribuição de agógica, por estar tocando em um andamento médio maior, suas variações são proporcionalmente menores, *i.e.*, “compactas”, menos expressivas.

Capítulo 5

Conclusão

5.1 Considerações Finais

O presente trabalho teve como objetivo o desenvolvimento de um conjunto de ferramentas que permitisse analisar as técnicas de expressão empregadas por um intérprete em uma peça musical.

Como motivação para o estudo destes parâmetros, o conceito de “movimento” na música foi evidenciado tanto na história da notação musical quanto nas palavras utilizadas hoje em dia para descrever as atitudes de interpretação. Foi visto que nem todas as indicações interpretativas se encontram na partitura, *i.e.*, a experiência da expressão depende de um “movimento interno” que não destrói as intenções do compositor, mas constrói sobre estas.

Outra motivação importante foi o instrumento mecânico popularmente conhecido como pianola. O pianolista tem grande liberdade para variar a interpretação da peça que está executando e, além disso, rolos especiais com interpretações “ao estilo” de pianistas famosos foram comercializadas, indicando a importância de se descobrir a “assinatura” de um intérprete.

Na tentativa de caracterizar as interpretações expressivas, os parâmetros andamento e dinâmica foram extraídos de um banco de dados contendo gravações dos 24 Prelúdios de Chopin por diversos artistas. A partir dos *onsets* anotados “manualmente”, foi possível calcular a curva de andamento. Para a curva de dinâmica, foi considerada a percepção humana e um algoritmo foi desenvolvido utilizando as curvas de igual audibilidade fornecidas pela norma ISO 226:2003.

Foram realizadas 4 análises sobre os parâmetros extraídos.

Levantando-se as estatísticas das curvas de andamento, uma “interpretação média” foi obtida. Além disso, uma análise temporal dos valores médios da curva de dinâmica de cada prelúdio demonstrou uma tendência à redução da intensidade, tanto em função do ano de nascimento do intérprete, quanto do ano de execução da peça musical. Para as informações de dinâmica, contudo, faz-se apenas uma ressalva: os níveis de intensidade das interpretações dependem fortemente de como suas gravações foram tratadas e comprimidas.

Os mapas de correlação são importantes ferramentas para se medir a similaridade de interpretação dos pianistas. A partir desta técnica, foi possível inferir sobre as relações pessoais dos intérpretes, *i.e.*, se são contemporâneos ou têm uma relação professor–aluno ou fazem parte de uma mesma “escola” de pianistas, que foram claramente retratadas em alguns dos mapas. Certamente, uma alta correlação para uma ou mais interpretações não significa, necessariamente, que os intérpretes conviveram ou mesmo trabalharam juntos. Ao mesmo tempo, o fato de terem uma relação pessoal/profissional não obriga que suas interpretações sejam similares.

Verificou-se que o modelo de Todd se restringe a uma visão global da peça. Quando investiga-se os detalhes nas frases melódicas este modelo não é válido e a correlação entre andamento e intensidade pode ser, inclusive, negativa.

Finalmente, demonstrou-se a capacidade da análise das trajetórias de interpretações no espaço bidimensional da dinâmica e da agógica. Esta poderosa ferramenta permite uma visualização em “alto nível” da complexa relação entre estes dois parâmetros e auxilia na compreensão tanto dos detalhes quanto dos aspectos mais gerais de cada interpretação, como foi feito para o prelúdio 6.

Assim, os métodos aqui descritos são capazes de elucidar as questões interpretativas, podem auxiliar estudantes e amantes da música a entenderem seus intérpretes favoritos e a melhor vivenciar este “movimento interno”.

5.2 Projetos Futuros

Em termos práticos, para que esse trabalho possa contribuir mais com a área de extração de informações musicais (em inglês, *Musical Information Retrieval* —

MIR) é necessário marcar todo o conjunto dos prelúdios na execução dos 27 pianistas e completar a estatística, para então tirar mais conclusões.

De fato, as ferramentas aqui desenvolvidas podem e devem ser utilizadas para a análise de outros pianistas, outras interpretações, estilos musicais e, até mesmo, outros instrumentos expressivos, *e.g.*, o violino.

Com relação à marcação dos *onsets*, ainda é necessário um método que melhor avalie estas interpretações expressivas, uma vez que os *beat trackers* disponíveis não parecem conseguir acompanhar as variações abruptas de ritmo e intensidade características destas execuções, e assim possibilite uma redução no tempo de preparação dos arquivos de áudio. O modelo psicoacústico desenvolvido também pode ser aperfeiçoado para considerar o efeito de mascaramento.

Além disso, há a necessidade de melhor entender a “assinatura” do intérprete a partir das curvas de andamento e intensidade. Talvez seja interessante utilizar uma rede neural para capturar estas características para um dado intérprete e, possivelmente, ser capaz de classificar uma gravação como sendo ou não executada por ele tendo, novamente, as curvas expressivas como parâmetros. Se a “assinatura” do intérprete for capturada, pode-se “sintetizar” novas interpretações para aquele intérprete de peças musicais que ele nunca tenha executado.

É interessante também, a partir dos mapas de correlação, montar uma “rede” das relações interpianistas, na qual os ramos que ligam um pianista a outro têm um peso proporcional aos valores dos coeficientes de correlação.

Por fim, sugere-se a “manipulação” das interpretações. Por exemplo, pode-se trocar as estratégias de expressividade dos pianistas ou alterar suas curvas de andamento e dinâmica através do espaço bidimensional de representação, sem que ocorra alteração do *pitch*.

Referências Bibliográficas

- [1] MED, B., *Teoria da Música*. 4 ed. Brasília, Brasil, MusiMed, 1996.
- [2] APEL, W., *Harvard Dictionary of Music*. 2 ed. Cambridge, USA, Belknap Press, 1969.
- [3] CARDINE, E., *Primeiro Ano de Canto Gregoriano e Semiologia Gregoriana*. 1 ed. São Paulo, Brasil, Attar Editorial/Palas Athena, 1989.
- [4] LAOUREUX, N., *Metodo Practico de Violin — Principios Fundamentales de la Técnica*, vol. 1. Buenos Aires, Argentina, Ricordi, 1916.
- [5] “The Pianola Institute — History of the Pianola”, Disponível em: <http://www.pianola.org/history/history.cfm>, Acesso em: 16 de março de 2013.
- [6] FASTL, H., ZWICKER, E., *Psychoacoustics: Facts and Models*. 3 ed. New York, USA, Springer, 2007.
- [7] MÜLLER, M., *Information Retrieval for Music and Motion*. New York, USA, Springer, 2007.
- [8] BOSI, M., GOLDBERG, R., *Introduction to Digital Audio Coding and Standards*, The Kluwer International Series in Engineering and Computer Science. Dordrecht, Nederland, Kluwer Academic Publishers, 2003.
- [9] FLETCHER, H., MUNSON, W. A., “Loudness, its Definition, Measurement and Calculation”, *The Journal of the Acoustical Society of America*, vol. 5, n. 2, pp. 82–108, Oct. 1933.
- [10] CHURCHER, B. G., KING, A. J., “The Performance of Noise Meters in Terms of the Primary Standard”, *Journal of the Institution of Eletrical Engineers*, vol. 81, n. 487, pp. 57–81, Jul. 1937.

- [11] ZWICKER, E., FELDTKELLER, R., “Über die Lautstärke von Gleichförmigen Geräuschen”, *Acustica*, vol. 5, pp. 303–316, 1955.
- [12] ROBINSON, D. W., DADSON, R. S., “A Re-Determination of the Equal-Loudness Relations for Pure Tones”, *British Journal of Applied Physics*, vol. 7, n. 5, pp. 166–181, May 1956.
- [13] SUZUKI, Y., TAKESHIMA, H., “Equal-Loudness-Level Contours for Pure Tones”, *The Journal of the Acoustical Society of America*, vol. 116, n. 2, pp. 918–933, Aug. 2004.
- [14] ISO/TC 43, *Acoustics — Normal Equal-Loudness-Level Contours*, ISO 226:2003(E), International Standardization Organization, Geneva, Switzerland, Oct. 2003.
- [15] PARMANEN, J., “Some Reasons to Revise the International Standard ISO 226:2003: Acoustics — Normal Equal-Loudness-Level Contours”, *Open Journal of Acoustics*, vol. 2, n. 4, pp. 143–149, Dec. 2012.
- [16] TRUSLIT, A., *Gestaltung und Bewegung in der Musik*. Berlin, Deutschland, Chr. Friedrich Vieweg, 1938.
- [17] REPP, B., “Music as Motion: A Synopsis of Alexander Truslit’s (1938) Gestaltung und Bewegung in der Musik”, *Psychology of Music*, vol. 21, n. 1, pp. 48–72, Jan. 1993.
- [18] PEREIRA, I., *Dicionário Grego–Português e Português–Grego*. 8 ed. Braga, Portugal, Apostolado da Imprensa, 1998.
- [19] RIEMANN, H., *Musikalische Dynamik und Agogik*. Hamburg, Deutschland, D. Rahter, 1884.
- [20] HAINSWORTH, S., “Beat Tracking and Musical Metre Analysis”. In: Klapuri, A., Davy, M. (eds.), *Signal Processing Methods for Music Transcription*, capítulo 4, New York, USA, Springer, 2006.
- [21] BILMES, J. A., *Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in*

- Percussive Rhythm*. Tese de mestrado, Program in Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, USA, Sep. 1993.
- [22] CANNAM, C., LANDONE, C., SANDLER, M., “Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files”. In: *Proceedings of the ACM Multimedia 2010 International Conference*, pp. 1467–1468, Firenze, Italy, Oct. 2010.
- [23] DIXON, S., “Onset Detection Revisited”. In: *Proceedings of the International Conference on Digital Audio Effects (DAFx-06)*, pp. 133–137, Montreal, Canada, 2006.
- [24] NUNES, L. O., BISCAINHO, L. W. P., “Tempo Estimation: Evaluation of Different Spectral Flux Computation Methods”. In: *Anais do 10^o Congresso de Engenharia de Áudio da AES-Brasil*, pp. 125–132, São Paulo, Brasil, 2012.
- [25] BELLO, J. P., DAUDET, L., ABDALLAH, S., *et al.*, “A Tutorial on Onset Detection in Music Signals”, *IEEE Transactions on Speech and Audio Processing*, vol. 13, n. 5, pp. 1035–1047, Sep. 2005.
- [26] DUXBURY, C., SANDLER, M., DAVIES, M., “A Hybrid Approach to Musical Note Onset Detection”. In: *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*, pp. 33–38, Hamburg, Germany, Sep. 2002.
- [27] Mazurka Project (CHARM), “Manpage for SV Mazurka Plugin: MzSpectralFlux”, Disponível em: <http://www.mazurka.org.uk/software/sv/plugin/MzSpectralFlux>, Acesso em: 22 de maio de 2013.
- [28] ITU-R, *Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level*, Recommendation ITU-R BS.1770, International Telecommunication Union, Geneva, Switzerland, Jul. 2006.
- [29] EBU/PLAUD, *Loudness Metering: ‘EBU Mode’ Metering to Supplement Loudness Normalisation in Accordance with EBU R 128*, EBU – Tech Doc 3341, European Broadcast Union, Geneva, Switzerland, Aug. 2011.

- [30] LANGNER, J., “Multidimensional Dynamic Shaping”. In: Gabrielsson, A. (ed.), *Proceedings of the Third Triennial ESCOM Conference*, pp. 713–718, Uppsala, Sweden, Jun. 1997.
- [31] TODD, N. P. M., “The Dynamics of Dynamics: A Model of Musical Expression”, *The Journal of the Acoustical Society of America*, vol. 91, n. 6, pp. 3540–3550, Jun. 1992.
- [32] DIXON, S., GOEBL, W., “Pinpointing the Beat: Tapping to Expressive Performances”. In: *7th International Conference on Music Perception and Cognition (ICMPC7)*, pp. 617–620, Sydney, Australia, 2002.
- [33] SAPP, C. S., “Comparative Analysis of Multiple Musical Performances”. In: *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pp. 497–500, Vienna, Austria, 2007.
- [34] Mazurka Project (CHARM), “Fiorentino Fakes: Continuing Saga of Concert Artist Recordings”, Disponível em: <http://mazurka.org.uk/fiorentino>, Acesso em: 22 de maio de 2013.
- [35] COOK, N., SAPP, C. S., “Purely Coincidental? Joyce Hatto and Chopin’s Mazurkas”, Disponível em: http://www.charm.rhul.ac.uk/projects/p2_3_2.html, Acesso em: 22 de maio de 2013.
- [36] LANGNER, J., GOEBL, W., “Visualizing Expressive Performance in Tempo-Loudness Space”, *Computer Music Journal*, vol. 27, n. 4, pp. 69–83, Dec. 2003.
- [37] DIXON, S., GOEBL, W., WIDMER, G., “The Performance Worm: Real Time Visualisation of Expression Based on Langner’s Tempo-Loudness Animation”. In: *Proceedings of the International Computer Music Conference (ICMC)*, pp. 361–364, Göteborg, Sweden, 2002.

Apêndice A

Códigos da Implementação

Neste apêndice encontram-se os códigos MATLAB utilizados para a extração e análise dos parâmetros expressivos.

A Seção A.1 contém os principais códigos desenvolvidos, as funções

- `extrairAgogica`;
- `extrairDinamica`;
- `scapePlot`.

Na Seção A.2, são apresentadas algumas rotinas auxiliares necessárias à implementação deste trabalho. São elas:

- `aproximarOnsets`, que dá conta de aproximar os valores dos *onsets* marcados (após correção manual) para os valores mais próximos dos pulsos encontrados pela rotina de fluxo espectral, `MzSpectralFlux`;
- `lerColuna`;
- `separarString`;
- `calibracao`, que gera o valor de referência de potência para o cálculo da intensidade sonora percebida ao redor de cada *onset*;
- `lerInterpretes`.

Finalmente, o código da função `iso226`, de autoria de Jeff Tackett, é encontrado na Seção A.3. Esta função permite calcular os valores das curvas de igual

audibilidade para qualquer valor entre 0 e 90 phons, de acordo com as equações da norma ISO 226:2003.

A.1 Códigos Principais

```
1 function andamento = extrairAgogica(interprete,prelude,trackLevel)
2
3 % EXTRAIRAGOGICA Retorna o andamento em bpm (beats por minuto).
4 %   Seja uma execucao musical com onsets devidamente marcados,
5 %   a funcao ANDAMENTO = EXTRAIRAGOGICA(I,P,TL) retorna o
6 %   andamento estimado, dado o trackLevel TL no qual a musica
7 %   foi anotada.
8 %
9 %   Exemplo:
10 %       % Compasso: 3/4
11 %       % Track level: 1/4
12 %       TL = 1/4;
13 %       andamento = extrairAgogica('A. M. LIMA',7,TL);
14 %
15 %   See also LERCOLUNA.
16
17 if (nargin == 1)
18     warning('Necessario informar o prelude desejado.');
```

```
19 else
20     if (nargin == 3)
21         Ql = 1/trackLevel;
22     else
23         Ql = 4;
24         if (nargin == 1)
25             warning('Assumindo nivel de marcacao igual a 1/4.');
```

```
26         end
27     end
28
29     dados = str2double( ...
30         lerColuna([interprete '\Track ' num2str(prelude) ...
31             '.txt'],1));
```

```

32     IOI = diff(dados);           % Inter-onset intervals
33     IOI = [IOI(1); IOI];       % Repetindo o primeiro tempo
34     andamento = (4/Q1)*(60./IOI); % Andamento em bpm
35 end
36 end

```

```

1  function intensidade = extrairDinamica(interprete,prelude)
2
3  % EXTRAIRDINAMICA Retorna a curva de dinamica (em phons).
4  %   Seja uma execucao musical com onsets devidamente marcados,
5  %   a funcao DINAMICA = EXTRAIRDINAMICA(I,P) retorna a curva de
6  %   dinamica estimada, utilizando o espectrograma (com janelas
7  %   de 400 ms e 75% de overlap na vizinhanca de 3 s de cada
8  %   onset) para o calculo do nivel de audibilidade.
9  %
10 %   Exemplo:
11 %       dinamica = extrairDinamica('A. M. LIMA',7);
12 %
13 %   See also LERCOLUNA, CALIBRACAO.
14
15 if (nargin == 1)
16     warning('Necessario informar o prelude desejado.');
```

```

17 else
18     % Leitura do arquivo de audio e do arquivo com os onsets
19     dados = str2double( ...
20         lerColuna([interprete '\Track ' num2str(prelude) ...
21             '.txt'],1));
22     [data,fs] = wavread([interprete '\Track ' num2str(prelude) ...
23         '.wav']);
24     x = mean(data,2);
25
26     % Amostra do sinal em que cada onset esta localizado
27     pulso = floor(dados*fs);
28
29     % Definicoes basicas
30     Nfft = 2^15;           % Bins no dominio da frequencia
31     Lm = floor(0.4*fs); % Comprimento da janela momentanea: 400 ms
32     Ls = floor(3*fs);    % Comprimento da janela curta duracao: 3 s

```

```

33     noverlap = floor(0.75*Lm);      % Overlap das janelas momentaneas
34     win = hann(Lm);
35
36     % Valor de referencia de potencia
37     ref_db = 80;
38     P_ref = calibracao(ref_db,win,Nfft,fs);
39
40     % Bandas de 1/3 de oitava
41     centro_faixa_nom = [20 25 31.5 40 50 63 80 100 125 160 200 ...
42                        250 315 400 500 630 800 1000 1250 1600 ...
43                        2000 2500 3150 4000 5000 6300 8000 ...
44                        10000 12500];
45     B = length(centro_faixa_nom);
46
47     centro_faixa_real = zeros(1,B);
48     lim_superior_faixa = zeros(1,B);
49     lim_inferior_faixa = zeros(1,B);
50     for i = 1:B,
51         centro_faixa_real(i) = 1000*(2^(1/3))^(i-18);
52         lim_inferior_faixa(i) = centro_faixa_real(i)/2^(1/6);
53         lim_superior_faixa(i) = centro_faixa_real(i)*2^(1/6);
54     end
55
56     kinf = ceil(Nfft*lim_inferior_faixa/fs);
57     ksuf = floor(Nfft*lim_superior_faixa/fs);
58
59     % Estendendo o sinal x (espelhado)
60     Ls_esq = ceil(Ls/2);
61     Ls_dir = Ls - Ls_esq;
62     pulso_estendido = pulso + Ls_esq;
63     inicio = flipdim(x(2:Ls_esq+1),1);
64     final = flipdim(x(end-Ls_dir:end-1),1);
65     x_estendido = [inicio; x; final];
66
67     intensidade = zeros(1,length(pulso_estendido));
68     % Calculo da intensidade percebida na regio de cada pulso
69     for i = 1:length(pulso_estendido)
70
71         % Separando trecho ao redor do onset

```

```

72     trecho = x_estendido(pulso_estendido(i)-Ls_esq: ...
73                       pulso_estendido(i)+Ls_dir-1);
74
75     X = abs(spectrogram(trecho, win, noverlap, Nfft));
76     P = zeros(B, size(X, 2));
77
78     % Calculando potencia por faixa em cada quadro
79     for j = 1:B
80         P(j, :) = sum(X(kinf(j):ksup(j), :).^2, 1);
81     end
82
83     % Calculando potencia/energia em cada faixa de 1/3 de oitava
84     PSDm = mean(P, 2);
85     P_passo = sum(PSDm);
86
87     Ln = 10*log10(P_passo/P_ref);
88
89     if (Ln < 3)
90         Ln = 3;
91     elseif (Ln > 90)
92         Ln = 90;
93     end
94
95     diferenca = 100;
96     while (abs(diferenca) > 0.01)
97         % Assumindo um loudness equivalente ao SPL calculado
98         [Lp, freq] = iso226(Ln);
99         curva_L = Lp(freq==1000)-Lp;
100        ponderacao = (10.^(curva_L/10))';
101
102        % Levando para o dominio perceptivo
103        PSDp = PSDm.*ponderacao;
104        P_passo = sum(PSDp);
105
106        L_old = Ln;
107        Ln = 10*log10(P_passo/P_ref);
108
109        if (Ln < 3)
110            Ln = 3;

```

```

111         elseif (Ln > 90)
112             Ln = 90;
113         end
114
115         diferenca = Ln - Lold;
116     end
117
118     intensidade(i) = Ln;
119 end
120 intensidade = intensidade(:);
121 end
122 end

```

```

1 function ...
    scapePlot(interprete, listaInterpretes, prelude, sequencias, cor)
2
3 % SCAPEPLOT Gera o scape plot para uma execucao.
4 %   SCAPEPLOT (INTERPRETE, LISTAINTERPRETES, PRELUDE, SEQUENCIAS, COR)
5 %   gera o mapa de correlacao para o INTERPRETE tendo como base
6 %   todos os interpretes em LISTAINTERPRETES e suas SEQUENCIAS de
7 %   andamento ou de intensidade. Os resultados sao exibidos nas
8 %   em COR, para cada interprete. LISTAINTERPRETES e SEQUENCIAS
9 %   sao do tipo cell.
10 %
11 %   See also EXTRAIRAGOGICA, EXTRAIRDINAMICA.
12
13 interpreteBase = strmatch(interprete, listaInterpretes, 'exact');
14
15 N = length(sequencias{1});
16
17 % Selecionando quais interpretes serao comparados
18 comparar = 1:length(listaInterpretes);
19 comparar = comparar(comparar ≠ interpreteBase);
20
21 % Preenchendo a variavel scape
22 scape = cell(N-1,1);
23 for i = 1:N-1
24     for inicio = 1:N-i

```

```

25     maiorCorrelacao = comparar(1);    % Caso Base
26     fim = inicio+i;
27     maximo = -Inf;
28     for interprete = comparar
29         coef = corr(sequencias{interpreteBase}(inicio:fim), ...
30                     sequencias{interprete}(inicio:fim));
31         maximo = max(maximo,coef);
32         if maximo == coef
33             maiorCorrelacao = interprete;
34         end
35     end
36     scape{i}(inicio) = maiorCorrelacao;
37 end
38 end
39
40 % Plotando a variavel scape
41 figure;
42 subplot(3,1,[1 2], 'Position', [0.13 0.3696 0.775 0.5154]);
43 hold on;
44 L = 1;
45 H = 1;
46 for i = 1:N-1
47     inicio = (i-1)/2;
48     totalCampos = (N-1)-i+1;
49     if i < N-1
50         % Primeiro campo da linha - Trapezio
51         x = inicio-L/4;
52         y = i;
53         fill([x;x+5*L/4;x+5*L/4;x+L/2], [y;y;y+H;y+H], ...
54             cor(scape{i}(1),:), 'LineStyle', 'none');
55         for j = 2:totalCampos-1
56             % Campos intermediarios - Retangulos
57             x = inicio+(j-1)*L;
58             rectangle('Position', [x,y,L,H], ...
59                     'FaceColor', cor(scape{i}(j),:), ...
60                     'LineStyle', 'none');
61         end
62         % Ultimo campo da linha - Trapezio
63         x = inicio+(totalCampos-1)*L;

```

```

64         fill([x;x+5*L/4;x+3*L/4;x],[y;y;y+H;y+H], ...
65             cor(scape{i}(totalCampos),:),'LineStyle','none');
66     else
67         % Campo do topo - Triangulo
68         x = inicio-L/4;
69         y = i;
70         fill([x;x+3*L/2;x+3*L/4],[y;y;y+3*H/2], ...
71             cor(scape{i}(1),:),'LineStyle','none');
72     end
73 end
74 axis equal;
75 axis off;
76
77 % Caculando a contagem de cada interprete
78 linha = zeros(1,sum(1:N-2));
79 inicio = 1;
80 for i = 1:N-2
81     totalCampos = length(scape{i});
82     linha(inicio:inicio+totalCampos-1) = scape{i};
83     inicio = inicio+totalCampos;
84 end
85
86 [n,x] = hist(linha,1:length(listaInterpretes));
87 dados = flipdim(sortrows([n;x]',1),1);
88
89 % Legenda dos 5 melhores correlacionados
90 subplot(3,1,3);
91 for j = 1:5
92     elemento = dados(j,2);
93     porcentagem = 100*dados(j,1)/sum(dados(:,1));
94     rectangle('Position',[0.4,5-j,0.2,0.5], ...
95             'FaceColor',cor(elemento,:),'LineStyle','none');
96     text(0.62,5-j+0.25,[listaInterpretes{elemento} ' (' ...
97         sprintf('%.1f',porcentagem) '\%)'],'Interpreter','Latex');
98 end
99 axis([-0.2 2 -0.1 4.7]);
100 axis off;
101
102 % Colocando titulo

```

```

103 set(gcf, 'NextPlot', 'add');
104 axes;
105 h = title(strcat(listaInterpretes(interpreteBase), ...
106     ' — Prel\''udio No. ', num2str(prelude)), ...
107     'FontSize', 12, 'Interpreter', 'Latex');
108 set(gca, 'Visible', 'off');
109 set(h, 'Visible', 'on');
110 end

```

A.2 Códigos Auxiliares

```

1 function aproximarOnsets(arquivoTicks, arquivoOnsets)
2
3 % APROXIMARONSETS Aproximar as marcacoes de dois arquivos.
4 % Para um arquivo contendo os pulsos marcados manualmente,
5 % a funcao escreve um arquivo temp.txt contendo a melhor
6 % aproximacao destes pulsos aos onsets determinados
7 % automaticamente.
8 %
9 % Exemplo:
10 %     arquivoBeats = 'Track 7 - beats.txt';
11 %     arquivoOnsets = 'Track 7 - onsets.txt';
12 %     aproximarOnsets(arquivoBeats, arquivoOnsets);
13 %
14 % See also FOPEN, LERCOLUNA.
15
16 temp = fopen('temp.txt', 'w');
17
18 ticks = str2double(lerColuna(arquivoTicks, 1));
19 labels = lerColuna(arquivoTicks, 2);
20 onsets = str2double(lerColuna(arquivoOnsets, 1));
21
22 ordenado = sort([ticks; onsets]);
23
24 N = length(ticks);
25 M = length(onsets);

```

```

26 for i = 1:N
27     indice = find(ordenado == ticks(i));
28     indice = indice(1);
29     elemento = ordenado(indice);
30     if (indice > 1) && (indice < N+M)
31         superior = ordenado(indice+1);
32         inferior = ordenado(indice-1);
33         if (superior-elemento) < (elemento-inferior)
34             aproximacao = superior;
35         else
36             aproximacao = inferior;
37         end
38     elseif indice == 1
39         aproximacao = ordenado(2);
40     else
41         aproximacao = ordenado(N+M-1);
42     end
43     fprintf(temp, '%s %s\r\n', num2str(aproximacao, '%.9f'), labels{i});
44 end
45
46 fclose(temp);
47 end

```

```

1 function coluna = lerColuna(arquivo, idColuna)
2
3 % LERCOLUNA Le uma coluna de dados de um arquivo.
4 % Para um arquivo contendo informacoes dispostas em colunas,
5 % a chamada COLUNA = LERCOLUNA(ARQUIVO,N) retorna os dados da
6 % N-esima coluna do arquivo em ARQUIVO.
7 %
8 % See also FOPEN, SEPARARSTRING.
9
10 id = fopen(arquivo, 'r');
11
12 nlinhas = 0;
13 while (fgetl(id) ≠ -1)
14     nlinhas = nlinhas + 1;
15 end

```

```

16 fseek(id,0,'bof');
17
18 coluna = cell(nlinhas,1);
19 for i = 1:nlinhas
20     linha = fgetl(id);
21     coluna{i} = separarString(linha,idColuna);
22 end
23
24 fclose(id);
25 end

```

```

1 function string = separarString(linha,elemento)
2
3 % SEPARARSTRING Separa os dados de uma linha.
4 %     STRING = SEPARARSTRING(LINHA,N) retorna a N-esima substring
5 %     em LINHA. Admite-se que o delimitador eh o caractere ' '.
6 %
7 %     See also TEXTSCAN.
8
9 aux = textscan(linha, '%s');
10 if elemento <= length(aux{1})
11     string = cell2mat(aux{1}(elemento));
12 else
13     string = '';
14 end

```

```

1 function P_ref = calibracao(ref_db,win,Nfft,fs)
2
3 % CALIBRACAO Gera um fator de escalamento de potencia calibrado.
4 %     P_REF = CALIBRACAO(REF_DB,WIN,NFFT,FS) retorna o fator de
5 %     calibracao para que uma senoide de referencia de 1 kHz com
6 %     Vp = 1 tenha o nivel de pressao sonora de REF_DB. O fator
7 %     de calibracao aqui considera tanto o efeito da janela (WIN)
8 %     quanto o comprimento da FFT (NFFT) utilizada para a
9 %     representacao do espectrograma.
10 %

```

```

11 % See also SPECTROGRAM.
12
13 L = length(win);
14 f_ref = 1000;
15 n = (0:L-1)';
16
17 sinal_ref = cos(2*pi*f_ref*n/fs);
18 S_ref = abs(fft(sinal_ref.*win,Nfft));
19
20 P_ref = sum(S_ref(1:Nfft/2+1).^2)/10^(ref_db/10);

```

```

1 function interprete = lerInterpretes(arquivo)
2
3 % LERINTERPRETES Le os nomes dos interpretes.
4 % Para um arquivo contendo os nomes dos interpretes dispostos
5 % em uma coluna, INTERPRETE = LERINTERPRETE(ARQUIVO) retorna
6 % uma lista com estes nomes.
7 %
8 % See also FOPEN.
9
10 id = fopen(arquivo,'r');
11
12 nlinhas = 0;
13 while (fgetl(id) ≠ -1)
14     nlinhas = nlinhas + 1;
15 end
16 fseek(id,0,'bof');
17
18 interprete = cell(nlinhas,1);
19 for i = 1:nlinhas
20     interprete{i} = fgetl(id);
21 end
22
23 fclose(id);
24 end

```

A.3 Códigos Externos

```
1 function [spl, freq] = iso226(phon);
2 %
3 % Generates an Equal Loudness Contour as described in ISO 226
4 %
5 % Usage: [SPL FREQ] = ISO226(PHON);
6 %
7 % PHON is the phon value in dB SPL that you want the equal
8 % loudness curve to represent. (lphon = 1dB @ 1kHz)
9 % SPL is the Sound Pressure Level amplitude returned for
10 % each of the 29 frequencies evaluated by ISO226.
11 % FREQ is the returned vector of frequencies that ISO226
12 % evaluates to generate the contour.
13 %
14 % Desc: This function will return the equal loudness contour for
15 % your desired phon level. The frequencies evaluated in
16 % this function only span from 20Hz - 12.5kHz, and only 29
17 % selective frequencies are covered. This is the limitation
18 % of the ISO standard.
19 %
20 % In addition the valid phon range should be 0 - 90 dB SPL.
21 % Values outside this range do not have experimental values
22 % and their contours should be treated as inaccurate.
23 %
24 % If more samples are required you should be able to easily
25 % interpolate these values using spline().
26 %
27 % Author: Jeff Tackett 03/01/05
28
29
30
31 % /-----\
32 %%%%%%%%%% TABLES FROM ISO226 %%%%%%%%%%
33 % \-----/
34 f = [20 25 31.5 40 50 63 80 100 125 160 200 250 315 400 500 630 ...
35      800 1000 1250 1600 2000 2500 3150 4000 5000 6300 8000 10000 ...
```

```

36     12500];
37
38 af = [0.532 0.506 0.480 0.455 0.432 0.409 0.387 0.367 0.349 ...
39     0.330 0.315 0.301 0.288 0.276 0.267 0.259 0.253 0.250 ...
40     0.246 0.244 0.243 0.243 0.243 0.242 0.242 0.245 0.254 ...
41     0.271 0.301];
42
43 Lu = [-31.6 -27.2 -23.0 -19.1 -15.9 -13.0 -10.3 -8.1 -6.2 -4.5 ...
44     -3.1 -2.0 -1.1 -0.4 0.0 0.3 0.5 0.0 -2.7 -4.1 ...
45     -1.0 1.7 2.5 1.2 -2.1 -7.1 -11.2 -10.7 -3.1];
46
47 Tf = [ 78.5 68.7 59.5 51.1 44.0 37.5 31.5 26.5 22.1 ...
48     17.9 14.4 11.4 8.6 6.2 4.4 3.0 2.2 2.4 ...
49     3.5 1.7 -1.3 -4.2 -6.0 -5.4 -1.5 6.0 12.6 ...
50     13.9 12.3];
51 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
52
53 %Error Trapping
54 if((phon < 0) | (phon > 90))
55     disp('Phon value out of bounds!')
56     spl = 0;
57     freq = 0;
58 else
59     %Setup user-defined values for equation
60     Ln = phon;
61
62     %Deriving sound pressure level from loudness level (iso226
63     %sect 4.1)
64     Af = 4.47E-3 * (10.^(0.025*Ln) - 1.15) + ...
65         (0.4*10.^(((Tf+Lu)/10)-9)).^af;
66     Lp = ((10./af).*log10(Af)) - Lu + 94;
67
68     %Return user data
69     spl = Lp;
70     freq = f;
71 end

```