ALGORITMOS PARA ANÁLISE RÍTMICA COMPUTACIONAL

Leonardo de Oliveira Nunes

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Luiz Wagner Pereira Biscainho

Rio de Janeiro
Setembro de 2014

ALGORITMOS PARA ANÁLISE RÍTMICA COMPUTACIONAL

Leonardo de Oliveira Nunes

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Eduardo Antonio Barros da Silva, Ph.D.

Prof. Luiz Pereira Calôba, Dr.Ing.

Prof. Marcio Nogueira de Souza, D.Sc.

Prof. Vítor Heloiz Nascimento, Ph.D.

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2014

Nunes, Leonardo de Oliveira

Algoritmos para Análise Rítmica Computacional/Leonardo de Oliveira Nunes. – Rio de Janeiro: UFRJ/COPPE, 2014.

XVII, 184 p.: il.; 29,7cm.

Orientador: Luiz Wagner Pereira Biscainho

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 173 – 184.

1. processamento de sinais. 2. processamento de sinais acústicos. 3. análise rítmica computacional. I. Biscainho, Luiz Wagner Pereira. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

Para:
Malu e
Domingos José de Oliveira

Agradecimentos

Inicialmente, devo agradecer a minha família pelo apoio dado durante todos esses anos de educação e, acima de tudo, pelo amor. Meus pais, minha irmã e meus avós criaram o ambiente no qual este trabalho se tornou realidade. E minha esposa que me apoiou incondicionalmente durante este período e é a pessoa mais importante do meu universo.

Ao Prof. Luiz Wagner devo agradecer por ter acreditado em mim 10 anos atrás, e ter se tornado um grande mentor e amigo. É impossível quantificar a influência e a ajuda que recebi nesses anos todos.

Não posso deixar de lado todos os amigos que fiz no Grupo de Processamento de Áudio e no Laboratório de Sinais, Multimídia e Telecomunicações e cujos trabalhos estão de alguma forma associados a este.

Agradeço a todos os examinadores por terem aceitado o convite para participar da banca desta dissertação.

Por fim, agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ) pelo apoio financeiro.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

ALGORITMOS PARA ANÁLISE RÍTMICA COMPUTACIONAL

Leonardo de Oliveira Nunes

Setembro/2014

Orientador: Luiz Wagner Pereira Biscainho

Programa: Engenharia Elétrica

Sinais de música usualmente possuem uma estrutura temporal bem definida, em que eventos acontecem em determinados instantes de tempo. Perceptivamente, esta estrutura temporal define um pulso dominante, cuja frequência determina o andamento da música. Também se podem definir uma divisão e um múltiplo desse pulso e com esses três níveis métricos descrever a estrutura rítmica global de uma peça musical. Este trabalho tem como objetivo o estudo e desenvolvimento de algoritmos que extraem esta informação rítmica de sinal de áudio. Tal informação é de extrema valia para diversas aplicações, entre elas reconhecimento automático de gênero musical, transcrição musical automática e compressão de sinais.

Esta tese é dividida em duas partes. Na primeira, é abordado o problema de estimação do andamento. Para isto, é feito um estudo sobre os atributos de sinais de áudio usualmente adotados para esta tarefa e como eles podem ser melhor obtidos. Também são propostas modificações sobre algoritmos para estimação de andamento encontrados na literatura. Por fim, é feita a comparação do desempenho dos algoritmos originais e de suas modificações.

Na segunda parte da tese, é estudado o problema de rastreamento métrico e são propostos para atacá-lo modelos probabilísticos, baseados em modelos ocultos de Markov. Estes modelos procuram encontrar, dentro do sinal de áudio, as ocorrências de cada nível métrico. Também são feitas simplificações sobre este modelo, para rastrear apenas o pulso do sinal. Acoplados aos modelos de rastreamento, também são obtidos diversos modelos de observação de cada um dos níveis métricos que procuram quantificar a chance de o valor de um determinado atributo estar associado às transições em cada nível métrico. Por fim, também é proposto um modelo baseado em padrões rítmicos apropriado para o rastreamento de estruturas rítmicas mais complexas.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ALGORITHMS FOR COMPUTATIONAL RHYTHMICAL ANALYSIS

Leonardo de Oliveira Nunes

September/2014

Advisor: Luiz Wagner Pereira Biscainho

Department: Electrical Engineering

Music signals usually exhibit a well-defined temporal structure, in which events can occur at determined time instants. Perceptually, this temporal structure defines a dominant pulse (beat), whose frequency determines the tempo of the music. One can also define a multiple and a submultiple of this pulse, and use these three metrical levels to describe the overall rhythmical structure of a musical piece. This dissertation has as an objective the development of algorithms for the extraction of this rhythmic information from music signals. Such information can be used in several applications, such as automatic music genre recognition, automatic music transcription, and signal compression.

This work is split into two parts. In the first one, the problem of tempo estimation is approached. In particular, a study on how signal features can be extracted so as to improve tempo estimates is presented. Modifications to standard algorithms intended for this task are also proposed. At last, performances of original and modified algorithms are compared, highlighting the contributions at each processing step.

In the second part of this dissertation, the metrics tracking problem is tackled and new probabilistic models, based on Hidden Markov Models, are proposed to this end. These models track, in a pre-recorded audio signal, the occurrence of each metrical level. Simplifications that restrict this model to estimate only the beats, are also described. Associated with the tracking models, observation models that quantify the probability that a given feature value is observed at each metrical level transition are developed. At last a pattern-based model, suited to track more complex rhythmical structures is also proposed.

Sumário

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Nesta tese, são estudados algoritmos para análise rítmica computacional, que procuram estimar o ritmo de uma gravação musical ou notação simbólica de música. Considerando a importância do ritmo para a percepção musical, estes algoritmos podem servir como base para diversos outros que assumem conhecimento desta informação. Desta forma, o problema abordado nesta tese é uma peça importante dentro de sistemas de extração de informação musical.

Neste capítulo, são apresentados os temas principais abordados nesta tese. Inicialmente é feita na Seção 1.1 uma breve descrição de conceitos associados ao ritmo e como estes serão abordados. Na Seção 1.2, são apresentados aspectos sobre percepção rítmica que serviram de guias no desenvolvimento dos algoritmos descritos neste trabalho. Na Seção 1.3 o problema da análise rítmica computacional propriamente dita é introduzido, sendo feitas ali uma breve revisão bibliográfica e uma descrição de suas principais aplicações. Por fim, na Seção 1.4, os objetivos da tese são apresentados em conjunto com a estrutura do restante deste documento.

1.1 Ritmo

Nesta seção é feita uma breve introdução em que será definida a visão de ritmo adotada neste trabalho. Deve-se ressaltar que o tratamento dado aqui não pretende esgotar o tema; procura-se apenas definir alguns conceitos que serão empregados ao longo do texto. Além disso, a discussão a seguir e o trabalho como um todo assume que as músicas sendo analisadas induzem todo o tempo a sensação de ritmo. Com isso, excluem-se estilos musicais contemporâneos em que o ritmo pode não estar evidente durante uma parcela da música.

Dentre as partes em que tradicionalmente se decompõe a Música, o ritmo [3] é responsável pela estrutura temporal [4], que induz a sensação de pulsação que se tem ao escutá-la [5]. É comum traçar um paralelo entre a regularidade temporal observada na música e fenômenos biológicos também regulares, como a respiração e

os batimentos cardíacos [6]. O ritmo e estes fenômenos biológicos estão associados à percepção de uma sucessão de eventos regularmente espaçados ao longo do tempo, e a maneira como, internamente, seres humanos tendem a se "sincronizar" a eles.

Independentemente de sua origem perceptiva, o ritmo é um aspecto importante tanto da forma como percebemos música como da forma como a música é produzida. Por isso, faz-se necessário pensar em maneiras de capturar esse aspecto da música—inicialmente através da notação simbólica. Nas suas primeiras formas, as indicações de altura e duração das notas eram imprecisas. Os primeiros esforços de padronização da notação de ritmo na forma que conhecemos datam do séc. XIII [7, 8]. Adotaram-se símbolos para diferentes durações relativas precisamente definidas, na forma de subdivisões de uma unidade temporal pré-estabelecida. Essa unidade temporal definiria o pulso anotado da música, ou o intervalo temporal mais frequente. A notação musical também define a quantidade de vezes que esse pulso se repete dentro de uma célula de maior duração, o compasso. Essa divisão hierárquica também é chamada de métrica, e é usualmente escolhida de forma a tanto facilitar a notação quanto capturar simbolicamente a sensação rítmica induzida pela música.

A notação musical tradicional, no entanto, deixa de capturar alguns aspectos importantes relacionados ao ritmo, como, por exemplo, informação sobre interpretação—pequenos ajustes que um músico realiza ao executar uma determinada música[1]. Mais restritiva é a incapacidade da partitura de representar características importantes rítmicas que são compartilhadas entre diferentes peças de determinado gênero. Por exemplo, o chamado *swing*, que é uma parte importante na execução de determinados estilos de Jazz, não é capturado em partituras, embora em geral fique claro para um músico se a execução de uma música o apresenta ou não. Por fim, pode-se mencionar a incapacidade de a notação tradicional representar grande parte da música contemporânea, como música eletroacústica, apesar de esta poder possuir ritmo claramente definido pelo autor (ainda que localmente).

Levando em consideração esses problemas, fica claro que a notação em partitura não é suficiente para capturar completamente a percepção do ritmo. Uma forma alternativa de conceitualizar o ritmo procura caracterizá-lo essencialmente como um fenômeno perceptivo [1]. Assim, pode-se representar o ritmo utilizando termos adotados da notação musical, como compasso; porém, redefinindo-os sob uma visão perceptiva: o compasso, o pulso e suas subdivisões, seriam definidos através da marcação por um ouvinte dos momentos em que ele percebe os eventos rítmicos. Desta forma, pode-se quantificar aspectos como *swing* (procurando-se desvios entre os intervalos esperados e os marcados) e até mesmo identificar trechos em que não há uma estrutura rítmica bem definida. Neste trabalho será considerada apenas essa

[1]Deve-se notar que muitas vezes não se deseja anotar essas informações, permitindo que fiquem a cargo do intérprete.

visão perceptiva do ritmo, a qual será mais detalhada adiante. Deve-se notar que a partitura e essa notação "perceptiva" procuram quantificar o mesmo fenômeno; logo, o que é anotado na partitura e as marcações feitas por um ouvinte usualmente carregam informações similares. Pode-se pensar, inclusive, que se a marcação fosse feita puramente por músicos treinados, ambas as notações seriam idênticas para músicas com ritmo bem definido. Afinal, pode-se assumir que, internamente, o compositor realiza a marcação perceptiva e dela deriva a partitura [1].

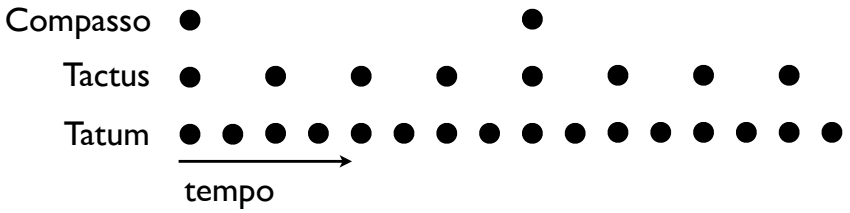
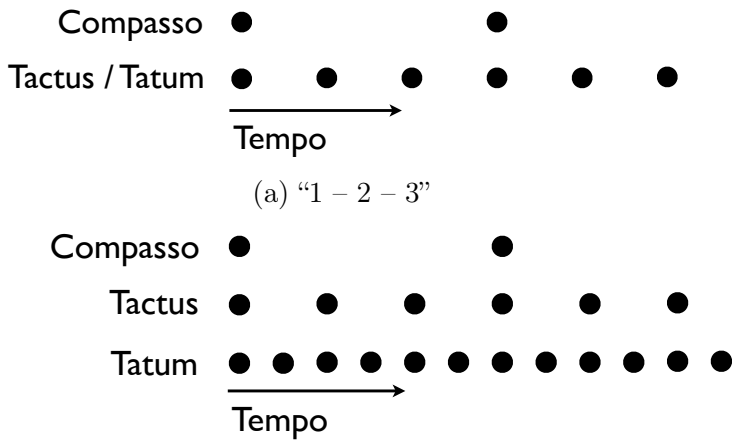


Figura 1.1: Níveis perceptivos que compõem uma possível estrutura rítmica. No exemplo, a estrutura estaria associada a uma contagem do tipo "1 – e – 2 – e – 3 – e – 4 – e".

Perceptivamente, pode-se propor uma estrutura rítmica da música [1] em dois ou três níveis hierárquicos. A Figura 1.1 ilustra uma possível estrutura para uma música com três níveis hierarquicamente estruturados: o tatum (com menor distância temporal entre os eventos), o tactus e o compasso. No caso do exemplo, um período do tactus ocorreria a cada dois períodos do tatum e, por sua vez, o compasso a cada dois períodos do tactus. Informalmente, pode-se associar esses níveis hierárquicos com a forma com que músicos normalmente "contam" durante a execução de uma música; no caso do exemplo, a contagem seria "1 – e – 2 – e – 3 – e – 4 – e", sendo o menor intervalo (determinado pela presença dos "e"s) o tatum. Nota-se nesta contagem que o que é efetivamente "contado" é o tactus. Tal escolha não é arbitrária: em geral o tactus é que ancora a percepção rítmica, servindo como referência para os demais níveis. Neste caso, o tactus também é chamado de pulso da peça, e será discutido mais detalhadamente na próxima seção.

Deve-se notar que existe uma correspondência entre a hierarquia obtida perceptivamente e a métrica anotada numa partitura. No entanto, duas músicas com partituras de mesma métrica podem induzir uma representação perceptiva diferente. Por exemplo, uma música com métrica anotada em $\frac{3}{4}$ em que o compasso é dividido em três pulsos pode levar a contagens como "1 – 2 – 3" ou como "1 – e – 2 – e – 3 – e". No primeiro caso, apenas dois níveis hierárquicos estão presentes (o tactus e o tatum se confundem); já no segundo, há três níveis hierárquicos. A Figura 1.2 exibe as duas estruturas correspondentes. Na próximo seção, serão discutidos aspectos da percepção que guiam, por exemplo, a escolha entre estes dois níveis hierárquicos.



(a) "1 – 2 – 3"

(b) "1 – e – 2 – e – 3 – e"

Figura 1.2: Dois diferentes padrões rítmicos que podem estar associados à mesma métrica anotada numa partitura.

Além disso, deve ser notado que o padrão rítmico fornecido por um ouvinte não precisa necessariamente corresponder ao padrão fornecido por um outro ouvinte, sendo a percepção rítmica influenciada por diversos fatores, tais como educação musical e familiaridade com o gênero da música sob análise.

1.2 Percepção Rítmica

Nesta seção, será feita uma breve introdução à percepção rítmica e serão sucintamente ilustrados alguns de seus aspectos. Deve-se ressaltar que este é um tema de pesquisa ativo, com diversos tópicos ainda sob investigação [1, 9].

Conforme mencionado na seção anterior, o ritmo é associado à regularidade (repetição) de eventos ao longo do tempo. No entanto, a taxa de repetição dos eventos é fundamental para determinar como essa repetição será percebida [10, 11]. Para ilustrar esse fenômeno, será utilizado o trem de impulsos com período T mostrado na Figura 1.3. Para valores de T acima de aproximadamente 50 ms, os impulsos se confundem e não é percebida uma série de eventos, mas sim apenas um evento com altura definida (*pitch*). Com T igual a 100 ms os eventos passam a ser percebidos de forma individualizada, percebendo-se a sua regularidade. Ao mesmo tempo, para valores de T acima de aproximadamente 8 s, os eventos são percebidos de forma individualizada e desconexa. Assim sendo, o trem de impulsos é percebido ritmicamente quando T possui valores entre 100 ms e 8 s. Pode-se dizer que é essa a escala temporal em que o ritmo ocorre [11].

Dentro desta escala, diversos modelos procuram explicar a sensação rítmica provocada por uma série de eventos sucessivos. Por exemplo, o modelo descrito em [12] procura explicar o ritmo como um princípio de organização. Em [10], é descrito um

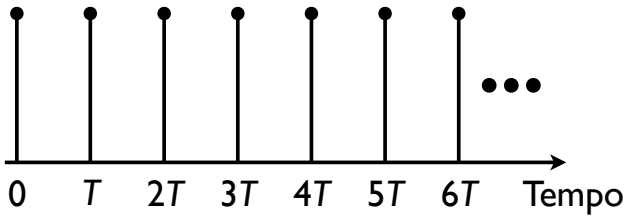


Figura 1.3: Trem de impulsos com período T usado para ilustrar aspectos da percepção rítmica.

modelo em que o ouvinte sincroniza um relógio interno com a estrutura rítmica do sinal. Neste último caso, a percepção do ritmo surgiria da relação entre este hipotético relógio interno e a música. Uma visão mais abrangente insere o ritmo numa camada mais geral da percepção [13] em que o ritmo estaria associado a processos básicos da mente, como percepção, expectativa e atenção [11]. A associação entre ritmo e atenção também é explorada em [1], onde se procura explicar o ritmo como um mecanismo de foco e uma forma de filtrar informações indesejadas.

Independentemente do modelo adotado, diversos fatos já são conhecidos sobre a maneira como o ritmo é percebido. O aspecto mais estudado é a percepção do tactus, que está associado ao pulso da música e é considerado o nível rítmico primário [3]. O pulso também define através de seu período o andamento da música, que está ligado à percepção de se uma música é rápida ou lenta. Em geral, o período do pulso se situa num intervalo delimitado inferiormente numa região aproximada entre 200 e 250 ms [1, 11] e superiormente, em 2 s. Dentro dessa faixa, períodos entre 600 ms e 700 ms são mais favorecidos [1, 14].

Com relação ao tatum, já foi estabelecido um limite inferior para o seu período em 100 ms. Superiormente, o período do tatum é limitado pelo limite inferior do tactus [1]. Na realidade, já foi encontrado experimentalmente que sinais de teste exibindo repetições com período abaixo de 200 ms induzem um tactus que é um múltiplo deste período [15].

O limite inferior para o tatum é de particular interesse, pois é determinado por limites psicoacústicos [1]. Devido a isso e outros achados experimentais, é considerado que os níveis hierárquicos são construídos do mais veloz para o mais lento [1]: primeiro o tatum, depois o tactus e, por fim, o compasso. Sendo assim, o limite inferior do tactus deve variar dependendo da quantidade de períodos do tatum que o compõem. Essa visão permite a explicação da existência da hierarquia de níveis rítmicos e também mostra os limites perceptivos para a sua existência. A Tabela 1.1, adaptada de [1], ilustra os limiares para diferentes períodos do tactus quando são consideradas divisões simples (2 períodos do tatum para um período do tactus) e compostas (3 períodos do tatum para um período do tactus). Em última análise

pode-se pensar que a existência do tactus é ditada pela possibilidade de o tatum ser também percebido [1].

Tabela 1.1: Limites absolutos para a percepção do tactus em função do andamento e da divisão. Adaptado de [1].

Período do tactus	Divisão simples	Divisão composta
$< 200\,\text{ms}$	—	—
$[200,\ 300\,]\text{ms}$	$[100,\ 150\,]\text{ms}$	—
$[300,\ 500\,]\text{ms}$	$< 250\,\text{ms}$	$< 250\,\text{ms}$
$[500,\ 750\,]\text{ms}$	$> 250\,\text{ms}$	$< 250\,\text{ms}$
$> 750\,\text{ms}$	$> 250\,\text{ms}$	$> 250\,\text{ms}$

Até o momento ainda não foi discutido quais variações num sinal podem induzir a percepção do ritmo. Numa análise superficial, pode-se pensar que apenas variações na intensidade da música, como a provocada pela execução de uma nota musical, definiriam um padrão temporal que induz o ritmo. No entanto, qualquer mudança percebida no sinal, como, por exemplo, mudança no *pitch* sem alteração na intensidade sonora, pode fazê-lo [11].

1.3 Análise Rítmica Computacional

A análise rítmica computacional se preocupa em estimar informações rítmicas a partir de alguma representação da música. Eis uma breve explicação das tarefas mais usuais encontradas em algoritmos de análise rítmica computacional:

- Estimação de andamento – procura estimar o período preferencial do tactus;
- Rastreamento de pulso – procura estimar os tempos que delimitam cada tactus;
- Estimação da estrutura rítmica – procura estimar os tempos que delimitam cada tatum, tactus e, possivelmente, compasso.

Os algoritmos que tentam estimar essas informações podem ser divididos em duas grandes famílias, de acordo com a sua entrada: os que utilizam informação simbólica e os que utilizam um sinal gravado. A seguir, é feita uma breve descrição de cada uma dessas famílias, sendo dada maior ênfase a soluções que empregam o sinal de áudio.

A estimação de informações rítmicas a partir da representação simbólica da música [16–19] (um arquivo MIDI [20], por exemplo) utiliza, em geral, apenas a informação sobre o ataque (*onset*) de cada nota musical. De particular interesse é o intervalo entre *onsets* (IOI, do inglês *inter-onset interval*), que armazena a diferença

entre os tempos de ocorrência de dois ataques consecutivos. Podem ser encontrados na literatura algoritmos que, a partir do IOI, estimam o pulso [21], o andamento [21] e todos os níveis hierárquicos [16, 19].

A estimação de informações rítmicas a partir do sinal tem sido um tópico ativo de pesquisa [22]. Ao se utilizar o sinal de áudio no lugar de uma representação simbólica, a quantidade de informações presentes no sinal aumenta, juntamente com a dificuldade de se extrair essas informações e a consequente complexidade dos sistemas de análise. Em geral, os algoritmos descritos na literatura procuram estimar informações rítmicas específicas, as mais comuns sendo o andamento e o pulso. A estimação do pulso foi abordada inicialmente em [23], sendo que o problema foi atacado em diversos outros trabalhos [24–27]. O problema de se estimar o andamento será abordado em mais detalhes no próximo capítulo, onde é feita uma revisão mais detalhada da literatura. Além do andamento e do pulso, há trabalhos que procuram estimar a partir do sinal de áudio a sua divisão rítmica [28], ou estimar conjuntamente o tactus e o tatum [29]. Considerando a estrutura de mais longa duração de músicas, alguns trabalhos descrevem algoritmos [30–32] para estimar a forma da música (versos e refrões, em músicas populares, por exemplo). Tais trabalhos, em geral, empregam informação rítmica de mais curta duração (como o pulso) para estimar a forma do sinal.

1.3.1 Aplicações

Tendo-se a estrutura musical de um sinal, é possível pensar em diversas utilizações potenciais dessas informações [11]. A primeira consiste na elaboração de algoritmos de edição musical que utilizam a informação do pulso para selecionar as posições temporais mais adequadas para se cortar e/ou inserir trechos de áudio.

Também se pode utilizar a informação rítmica em algoritmos de processamento de sinais. Por exemplo, algoritmos que operam em blocos poderiam dividir o sinal de modo síncrono com o ritmo. Esse esquema poderia ser especialmente benéfico em algoritmos de compressão de áudio, já que os blocos selecionados de forma síncrona com o pulso, por exemplo, possuiriam maior similaridade interna do que blocos cujos tamanho e início são escolhidos de forma arbitrária. Por exemplo, o algoritmo de compressão descrito em [33] utiliza uma heurística para modificar o tamanho dos blocos de análise que pode se beneficiar do conhecimento da estrutura rítmica do sinal.

Uma aplicação direta da informação rítmica se dá em algoritmos para transcrição musical automática [34, 35]. A informação rítmica, neste caso, é necessária para a correta escrita da partitura. Além disso, o conhecimento prévio da estrutura rítmica do sinal pode também auxiliar na detecção da ocorrência de notas musicais,

melhorando o desempenho dos algoritmos já existentes.

Em extração de informação musical (MIR, do inglês *Music Information Retrieval*), a informação sobre a estrutura rítmica pode auxiliar na detecção do gênero de músicas gravadas [36]. Além disso, a informação de se uma peça possui um andamento elevado ou lento é importante para sistemas de recomendação de músicas. Por último, cabe citar a utilização da informação rítmica para busca de sinais similares e do problema de encontrar uma música apenas a partir de uma sequência de pulsos fornecida por uma pessoa [37].

Outras possíveis aplicações da informação rítmica incluem transformações rítmicas de sinais de áudio [38], sincronização entre uma execução e uma partitura [39] e efeitos digitais de áudio [11].

1.4 Objetivos da Tese

Esta tese se ocupa do desenvolvimento de algoritmos para estimação de informações rítmicas a partir do sinal de áudio. De particular interesse é a estimação da estrutura rítmica como um todo: a estimação dos tempos de ocorrência de tatum, tactus e compasso.

Para chegar ao seu objetivo, inicialmente será abordado o problema de estimação de andamento (a frequência do tactus). O objetivo desta escolha se deve ao fato de, conforme visto na Seção 1.2, o período do tactus carregar informações importantes sobre os demais níveis hierárquicos, servindo como ponto de partida na solução do problema. Além disso, a informação sobre o andamento é útil por si só, definindo se a execução da música é percebida como rápida ou lenta. O Capítulo 2 contém a descrição dos algoritmos de estimação de andamento encontrados na literatura. O Capítulo 3 compara o desempenho de diversos destes algoritmos, ao mesmo tempo em que apresenta um algoritmo protótipo. O Capítulo 4 descreve melhorias sobre este algoritmo e analisa o ganho em desempenho resultante das melhorias propostas.

Em seguida, a tese aborda o problema de rastreamento métrico, ou seja, a detecção de quando ocorreu um novo tatum, tactus ou compasso. Inicialmente, são propostos modelos probabilísticos para a estrutura rítmica de um sinal de música no Capítulo 5. Em particular, serão descritos modelos ocultos de Markov capazes de estimar esses eventos considerando diferentes estruturas métricas e também pequenas variações no andamento. Já no Capítulo 6, é descrito um banco de sinais com os três níveis métricos anotados. Este banco vai servir de base para o desenvolvimento dos dois capítulos seguintes. No Capítulo 7, são propostos modelos probabilísticos que associam um atributo observado com a ocorrência de um determinado nível métrico (ou ausência de informação rítmica). Já no Capítulo 8 é avaliado o desempenho dos modelos propostos. As conclusões desta tese são apresentadas no Capítulo 9.

Parte I

Estimação de Andamento

Capítulo 2

Estimação de Andamento/Tempo

Neste capítulo é feita uma revisão de algoritmos para estimação do andamento de uma música. Serão apresentados algoritmos bem estabelecidos na literatura e, quando possível, serão discutidas modificações propostas na literatura sobre esses algoritmos.

Na Seção 2.1 é feita uma breve introdução à estimação de andamento. Na Seção 2.2, são apresentadas as três etapas usualmente encontradas em algoritmos de estimação de andamento. A seguir, nas Seções 2.3, 2.4 e 2.5 são descritas soluções para cada uma dessas etapas. Por fim, na Seção 2.6 são feitos os comentários finais do capítulo.

2.1 Visão Geral

A estimação do andamento de uma peça procura descobrir o período preferencial dos pulsos de uma música. Conforme discutido no capítulo anterior, o conceito de andamento é encontrado em teoria musical [3] e cognição [1], e está associado à percepção de velocidade de uma peça, isto é, se ela é rápida ou lenta. Uma peça com andamento rápido tenderá a possuir notas com ocorrências mais próximas entre si no tempo. Além disso, as repetições ditadas pela estrutura rítmica da peça tenderão a ocorrer com um período menor. Em última análise, até mesmo a duração de uma música pode ser determinada pelo seu andamento.

Em geral, pode-se estimar o andamento de uma peça musical a partir de uma representação simbólica da execução [21] ou a partir do sinal capturado, conforme mencionado no Seção 1.3. A estimação do andamento a partir de um registro de uma peça musical (isto é, de um sinal de áudio) é uma tarefa desafiadora. Sem a informação de onde as notas ocorrem, precisa-se descobrir diretamente a partir da forma de onda quais regiões são similares a outras e com que frequência ocorrem suas repetições. Uma das primeiras e mais populares soluções para esse problema pode ser encontrada em [23]; ela inspirou diversas outras (como as descritas em [24, 28, 40])

e também foi a primeira a utilizar as três etapas descritas na próxima seção. Esse trabalho, além de propor as etapas de processamento, também descreveu o uso das variações na energia do sinal como um atributo relacionado ao ritmo e a busca por picos numa função de periodicidade como uma forma de se estimar o andamento de sinais de música.

A estimação do andamento a partir do sinal de áudio tem sido um tópico ativo de pesquisa, principalmente após a inclusão, em 2004, da tarefa de estimação de andamento na competição anual de algoritmos de extração de informação musical (o MIREX). Esta competição estabeleceu métricas e bancos de sinais (utilizados nos próximos capítulos) que são adotados pelos pesquisadores do tema. Contudo, a tarefa de estimação de andamento continua apresentando desafios como, por exemplo, o fato de o andamento poder variar ao longo da execução da peça. Outra característica que precisa ser abordada é a variação do andamento percebido de pessoa para pessoa [14], que dificulta a avaliação do desempenho dos métodos.

2.2 Diagrama de Blocos

Nesta seção, é apresentada uma visão geral de algoritmos de estimação de andamento a partir de um sinal de áudio. Em particular, serão descritas as tarefas usualmente empregadas nos algoritmos da literatura. O objetivo desta seção é motivar cada uma das etapas; os algoritmos que realizam estas tarefas serão descritos nas seções seguintes.

De forma geral, os algoritmos de estimação de andamento podem ser descritos [41] como na Figura 2.1. O andamento do sinal de entrada, $x[n]$, é estimado em três etapas: a extração de atributos, a computação da periodicidade e a determinação do andamento propriamente dita. Para ilustrar cada um destes estágios, será utilizado um sinal de exemplo. Este sinal contém um *cowbell* sendo tocado a uma taxa de 120 batidas por minuto (BPM). A taxa de amostragem do sinal é $f_s = 44{,}1\,\text{kHz}$. O sinal no domínio do tempo pode ser visto na Figura 2.2, onde se identifica claramente cada um dos ataques presentes no sinal. A seguir, serão detalhadas cada uma das etapas.

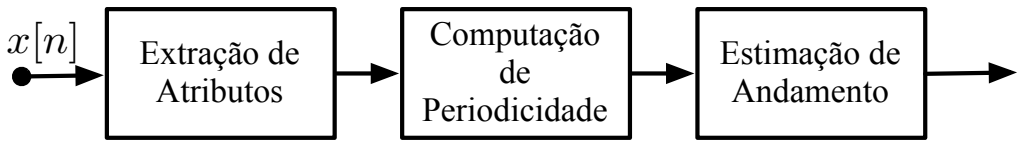


Figura 2.1: Três etapas de processamento usualmente utilizadas em algoritmos de estimação de andamento.

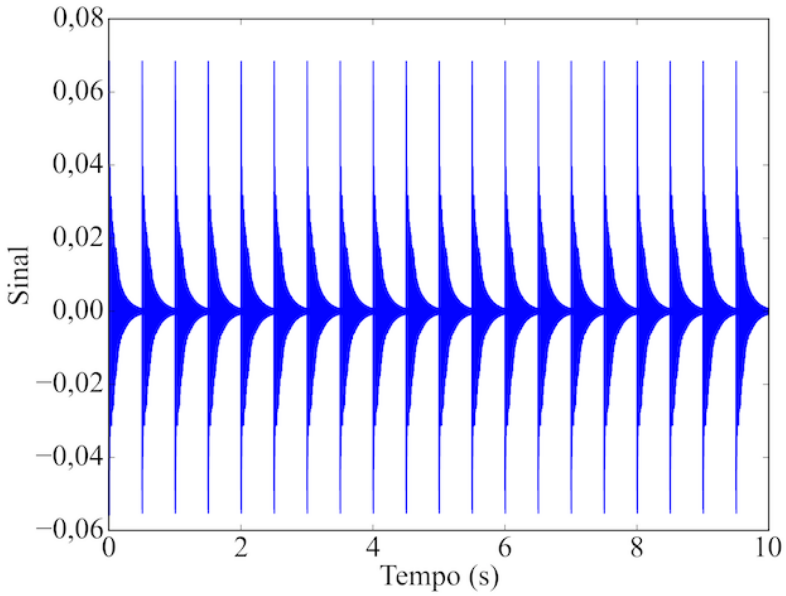


Figura 2.2: Sinal de exemplo no domínio do tempo.

A estimação de atributos do sinal de áudio tem como objetivo reduzir a quantidade de dados a serem analisados pelos próximos estágios, facilitando sua tarefa, já que idealmente deixa apenas informações pertinentes à estimação do andamento. Espera-se, portanto, que os atributos extraídos sejam altamente correlacionados com eventos temporais que ocorrem no sinal de áudio, como, por exemplo, *onsets*. Na Figura 2.3, é exibido um atributo extraído a partir do sinal de exemplo. Pode ser observado que o atributo captura a ocorrência de cada ataque do *cowbell*, enquanto remove informação irrelevante para estimação do andamento, como, por exemplo, a duração de cada "nota" executada. Além disso, a taxa de amostragem neste exemplo pôde ser reduzida de 44,1 kHz para 100 Hz. Deve ser ressaltado ainda que a saída deste bloco pode ser um único atributo variante no tempo (como no exemplo) ou um conjunto de atributos variantes no tempo (cada um calculado para uma diferente faixa de frequências do sinal, por exemplo).

O próximo estágio consiste na estimação do período dos atributos extraídos. A saída deste estágio é uma função de periodicidade, que representa quão prováveis diferentes andamentos (usualmente expressos na forma de BPM) são. A Figura 2.4 exibe uma função de periodicidade para o sinal de exemplo. Como pode ser visto, ocorrem picos em 60 e 120 BPM, indicando que estes dois valores são bons candidatos a representarem o andamento do sinal. A função de periodicidade pode integrar informações de diferentes atributos ou ser computada para cada atributo, resultando nesse caso num conjunto de funções.

De posse da função de periodicidade, o próximo estágio consiste em estimar o andamento ou os candidatos a andamento do sinal de entrada. As estratégias

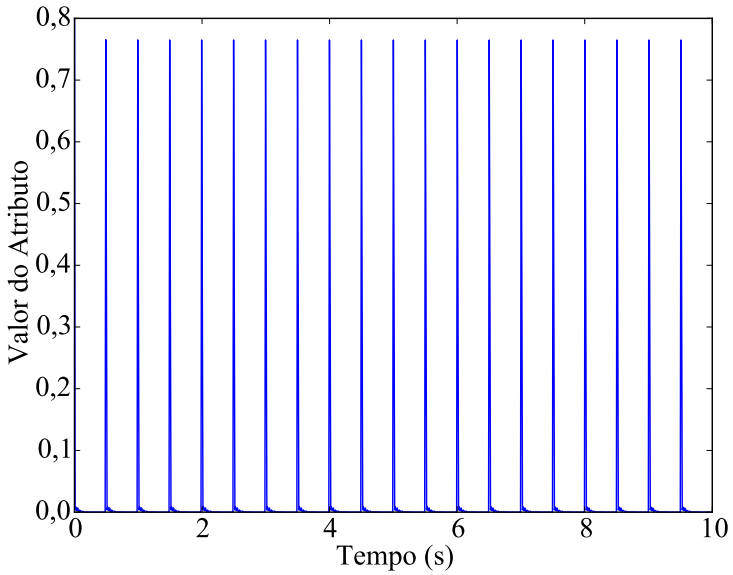


Figura 2.3: Atributo extraídos do sinal de exemplo da Figura 2.2.

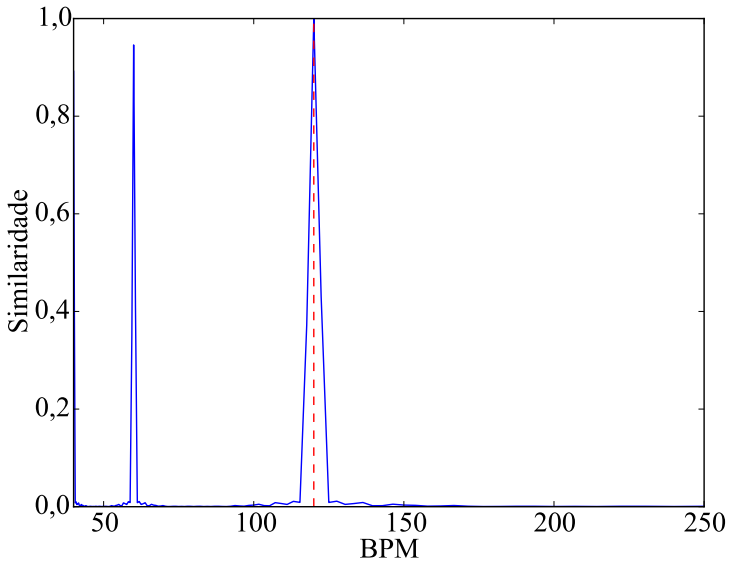


Figura 2.4: Função de periodicidade calculada a partir dos atributos exibidos na Figura 2.3. A linha tracejada vertical marca o andamento do sinal.

utilizadas nesta tarefa são diversas: variam de um simples algoritmo de seleção de picos até algoritmos mais sofisticados que incorporam informações de teoria musical e cognição [42]. Para o sinal de exemplo, um simples algoritmo de seleção de picos seria suficiente—o maior pico da função de periodicidade corresponderia ao andamento do sinal.

2.3 Extração de Atributos

Os atributos utilizados para estimação de andamento são usualmente projetados para capturar mudanças abruptas no sinal, tais como o aumento da energia ou uma mudança na linha melódica. Para que os atributos carreguem informação sobre o andamento, supõe-se que tais mudanças ocorrem seguindo alguma estruturação temporal, ou seja, respeitam a estrutura rítmica da música.

Por exemplo, atributos que capturam mudanças harmônicas/melódicas são utilizados em [43]. Nesse trabalho, atributos tonais são extraídos utilizando-se a Transformada de Q Constante [44]. Em [45], uma idéia similar é empregada, porém utiliza-se um método diferente para obtenção dos atributos tonais. Outra abordagem consiste em detectar mudanças no timbre do sinal [46] através do uso de Coeficientes Cepstrais na Escala Mel [47].

É possível notar, no entanto, que atributos que capturam mudanças de energia são mais populares. Estes atributos usualmente estimam a evolução temporal da envoltória de energia do sinal (ou de uma de suas sub-bandas). Normalmente, estas variações de energia estão associadas aos ataques de notas (*onsets*) que, por sua vez, estão associados à percepção da estrutura temporal de uma música [1, 19, 22]. Na literatura, atributos desta natureza são nomeados de diversas formas (dependendo da forma com que são calculados): função de novidade [48], função de detecção de *onset* [49] ou função de acentuação [50]. Uma das primeiras publicações a usar tais atributos [21] primeiro divide o sinal em 6 sub-bandas, estima a envoltória de energia para cada sub-banda, calcula a derivada temporal das envoltórias e obtém os atributos pela retificação de meia onda das derivadas. Essa sequência de passos (divisão em sub-bandas, diferenciação ao longo do tempo e retificação de meia-onda) é encontrada em diversos artigos que utilizam atributos baseados em energia. Em [29] e [51] são descritas modificações sobre o algoritmo de [23], onde são utilizados diferentes números de sub-bandas e pré-processamentos para obtenção dos atributos.

Dentre os atributos baseados em energia, o fluxo espectral [52] tem sido usado com frequência para a estimação de andamento [2, 25, 42, 48, 53–58]; por esse motivo, será detalhado separadamente na próxima seção.

2.3.1 Fluxo Espectral

Inicialmente apresentado em [40], o fluxo espectral possui mais de uma definição. Nesta seção, será descrita a formulação de [52], por ser a mais utilizada e por servir como base para diversas modificações encontradas na literatura.

Para o cálculo do fluxo espectral do sinal $x[n]$, inicialmente é calculada a sua Transformada de Fourier de Curta Duração (STFT, do inglês *Short-Time Fourier*

Transform), definida como

$$X[m,k] = \frac{1}{N}\sum_{n=0}^{N-1} w[n]x[n+mH]e^{-jk\frac{2\pi}{N}n}, \tag{2.1}$$

onde N é o número de raias de frequência k calculadas para um quadro de índice m, $w[n]$ é uma janela de suavização [59] tal que $w[n] = 0$ fora do intervalo $0 \leq n < N$, e o salto (do inglês, *hop*) entre janelas adjacentes é de H amostras.

De posse da STFT do sinal, o fluxo espectral é obtido através de

$$F^{\mathrm{SF}}[m] = \sum_{k=0}^{\lfloor \frac{N}{2} \rfloor} \mathrm{HWR}\left(|X[m,k]| - |X[m-1,k]|\right), \tag{2.2}$$

onde $\lfloor a \rfloor$ denota o arredondamento para o inteiro mais próximo não maior que a e $\mathrm{HWR}(\cdot)$ é a função de retificação de meia onda:

$$\mathrm{HWR}(a) = \begin{cases} a, & \text{se } a > 0 \\ 0, & \text{em caso contrário.} \end{cases} \tag{2.3}$$

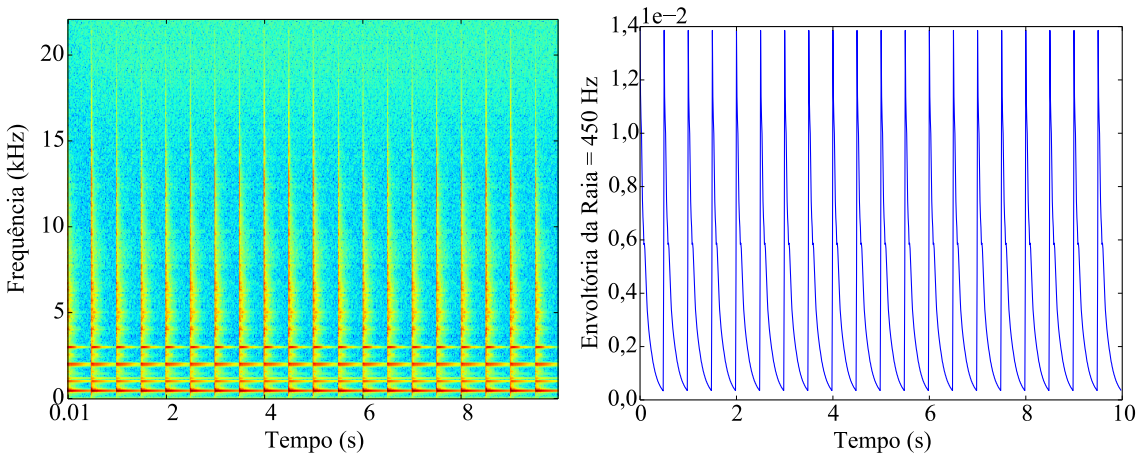
Com isso, pode-se ver que o fluxo espectral é obtido através da soma das raias[1] após a retificação de meia onda da primeira diferença do módulo da STFT.

A seguir, será exibido para o sinal de exemplo o resultado de cada um dos passos do cálculo do fluxo espectral, com o objetivo de motivar cada um pela ilustração de seu efeito sobre um sinal de fácil interpretação.

Na Figura 2.5, podem ser vistos: o módulo da STFT do sinal para todas as raias na Figura 2.5a, e apenas a evolução do módulo da raia de maior energia na Figura 2.5b. Pode ser observada a presença de linhas verticais no módulo da STFT no momento em que ocorre o ataque de cada nota do *cowbell*, e o mesmo pode ser observado na raia de maior energia. Idealmente, deseja-se que apenas essas linhas verticais estejam presentes no sinal, já que elas contêm a informação relativa ao surgimento de energia no sinal. Como será visto, os passos subsequentes no cálculo do fluxo espectral procuram isolar e refinar essa informação sobre os ataques.

Na Figura 2.6, pode ser vista a diferença entre quadros adjacentes do módulo da STFT. Essa diferença permite considerar apenas regiões de grande variação de energia, tornando regiões mais "bem comportadas" da STFT (onde linhas verticais não são observadas) aproximadamente zero. Isso fica mais claro ao se comparar as Figuras 2.5b e 2.6—os picos, que antes possuíam um decaimento lento, são mais bem localizados temporalmente após o cálculo da diferença.

[1] Como o sinal de entrada é real, essa soma só precisa considerar as raias correspondentes às frequências positivas.



(a) Todas as raias. (b) Apenas raia de frequência 450 Hz.

Figura 2.5: STFT do sinal de exemplo.

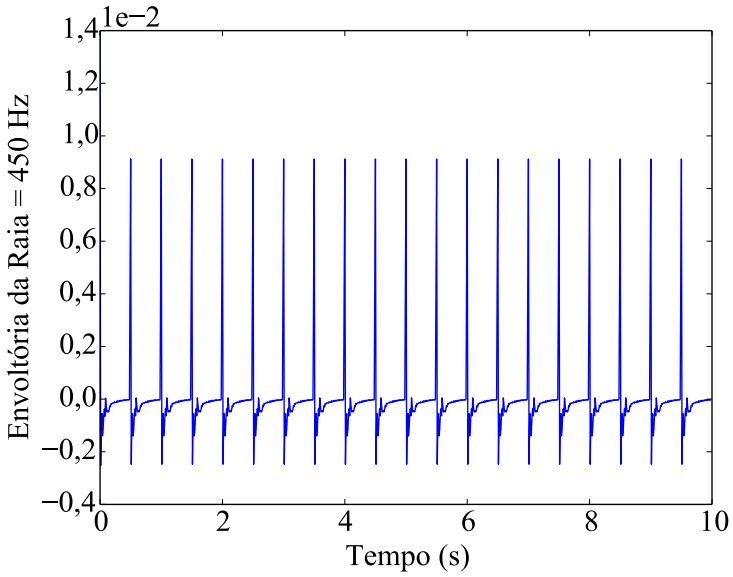


Figura 2.6: Raia de frequência 450 Hz da diferença entre os módulos da STFT do sinal de exemplo em quadros adjacentes.

Por fim, os valores negativos da diferença são descartados e os resultados obtidos para as diferentes raias são somados. Para o sinal de exemplo, o fluxo espectral obtido é exibido na Figura 2.7. Os valores negativos podem ser descartados porque carregam a informação de perdas abruptas de energia (por exemplo, durante o *offset* de uma nota). Para o sinal de exemplo, somar as raias permite a redução da quantidade de dados sem grande perda de informação, já que suas variações de energia são coerentes ao longo das raias. Para sinais mais complexos, as variações de energia podem ser bem distintas em torno de diferentes frequências. Nestes casos, a combinação das diferentes raias pode ser prejudicial à estimação do andamento.

Os exemplos exibidos anteriormente foram calculados utilizando-se uma janela

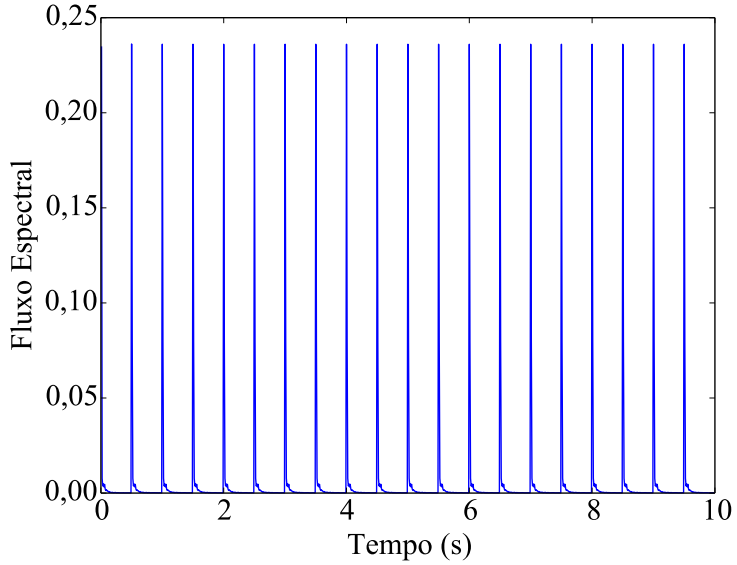


Figura 2.7: Fluxo espectral obtido para o sinal de exemplo.

de Hann [59], um comprimento de janela correspondente a 20 ms (882 amostras @ 44,1 kHz) e um salto de 10 ms (441 amostras @ 44,1 kHz). Uma escolha adequada destes dois últimos parâmetros é importante para a obtenção de resultados coerentes na estimação do andamento. Em particular, o salto entre janelas vai definir a taxa de amostragem do Fluxo Espectral. Quanto menor o salto, mais precisa temporalmente é a informação obtida sobre a evolução de energia do sinal. Em contrapartida, isto aumenta a complexidade computacional do cálculo do atributo e das etapas de processamento posteriores. O comprimento da janela, por sua vez, está associado à escala em que se deseja observar a variação de energia do sinal. Idealmente a janela deve ser longa o suficiente para gerar uma estimativa suave da variação de energia, mas também não pode ser tão longa que acabe integrando a energia de múltiplos *onsets*. Logo, o comprimento da janela N limita o maior valor em BPM que poderá ser detectado, sendo este igual a $\frac{60 f_s}{N}$ (N em amostras, f_s em Hertz). Como será descrito na próxima seção, existe uma grande variabilidade nos valores escolhidos para estes parâmetros na literatura.

2.3.2 Modificações Sobre o Fluxo Espectral

Podem ser encontradas na literatura muitas modificações sobre o algoritmo apresentado na seção anterior. Sob este ponto de vista, o fluxo espectral serve como uma plataforma sobre a qual são propostos ajustes e modificações. Nesta seção, são descritas algumas dessas modificações.

Uma primeira etapa de processamento do fluxo espectral pode incluir a subamostragem do sinal de entrada. Considerando a taxa de amostragem do fluxo

espectral (normalmente em torno de 100 Hz), pode-se pensar em reduzir já de início a taxa do sinal de entrada (usualmente 44,1 kHz ou mais) de forma a reduzir a complexidade computacional. Por exemplo, nos trabalhos [2, 57, 60–63] a taxa de amostragem é reduzida de 44,1 kHz para 11,05 kHz, enquanto que em [25] a taxa final é de 8 kHz. Deve-se ressaltar que com essa modificação, descarta-se conteúdo de alta frequência no sinal que pode conter informações relevantes sobre sua estrutura rítmica.

Considerando que diferentes regiões de frequência de um sinal podem conter informações diferentes sobre o andamento, alguns trabalhos procuram calcular o fluxo espectral em diferentes faixas do espectro. Por exemplo, em música popular, o fluxo espectral calculado em faixas do espectro de baixa frequência poderia conter os ataques do baixo, enquanto regiões de alta frequência conteriam os ataques dos instrumentos percussivos de acompanhamento (como um prato da bateria). Em [48], o fluxo espectral é calculado para 5 sub-bandas linearmente espaçadas, enquanto que em [53] são utilizadas 8 sub-bandas linearmente espaçadas.

Outra forma alternativa de se considerar diferentes regiões de frequência é o mapeamento das raias da STFT numa escala inspirada na percepção humana. A escala perceptiva mais utilizada, neste caso, é a Mel [64, 65] encontrada em [25, 58, 61, 66–70]. Ela é obtida através de filtros passa-faixa espaçados geometricamente que podem ser aplicados diretamente à STFT do sinal. Usualmente, basta considerar as saídas de um número pequeno de filtros Mel (em torno de 20), o que reduz a quantidade de raias de frequência em duas ordens de magnitude, em alguns casos.

Uma modificação frequentemente utilizada [2, 25, 48, 53, 57, 60, 61, 63, 66–71] é a compressão dos valores de magnitude da STFT, normalmente através da função logaritmo. A compressão é aplicada antes da diferenciação e é motivada pela percepção humana, comprimida de forma aproximadamente logarítmica, das variações de intensidade. Deve-se observar que se interpretarmos a diferença no cálculo do fluxo espectral como uma aproximação de derivada, o uso do logaritmo equivale a dividir a derivada pelo valor da magnitude da STFT; assim, estaria sendo calculada a variação relativa, e não absoluta, de energia.

Outras modificações que podem ser encontradas na literatura incluem o uso de filtros de suavização antes [57, 61] e após a diferenciação [25, 42, 57]. Em outros casos, a soma da equação (2.2) não é feita [61, 67, 70], sendo calculado um fluxo espectral para cada raia da STFT.

Além das modificações sobre a forma de cálculo do fluxo espectral, também ocorre grande variabilidade nos valores escolhidos para os parâmetros. Por exemplo, o tamanho da janela utilizada na STFT pode variar de 20 ms [67] até 92,8 ms [2] e o salto entre janelas de 4 [25] até 28 ms [58]. Nos trabalhos que mapeiam a STFT para escala Mel, o número de filtros Mel utilizados varia de 9 [61] até 40 [25].

2.4 Cálculo da Periodicidade

Nesta seção serão apresentados métodos para o cálculo da periodicidade. Conforme descrito na Seção 2.2, o objetivo destes métodos é obter uma medida de quão provável um determinado andamento (ou período) é a partir dos atributos extraídos do sinal. De forma geral, existem duas formas de se obter essa medida de periodicidade: procurar regiões temporais do atributo que são consistentemente similares a regiões do mesmo sinal atrasado por um valor constante; ou medir quão similares são os atributos em relação a um sinal de teste com período conhecido. A função de autocorrelação [24, 40], que será detalhada na Seção 2.4.1, é um exemplo do primeiro tipo de método para cálculo da função de periodicidade. A transformada de Fourier, detalhada na Seção 2.4.2, e o banco de filtros-pente [23, 50] são exemplos do segundo tipo. Existem ainda métodos híbridos que combinam as duas abordagens, um dos quais será descrito na Seção 2.4.3.

A descrição feita a seguir se refere ao cálculo da periodicidade de apenas um atributo F. No caso de existir um conjunto de atributos, pode-se estimar uma função de periodicidade para cada um e obter uma função de periodicidade média posteriormente. A periodicidade também pode ser computada para uma determinada janela temporal do sinal ou para todo o sinal. No primeiro caso, é obtida uma função de periodicidade para cada janela, o que pode ser útil caso se deseje estimar a variação do andamento ao longo do tempo. Em todo caso, a duração do trecho sob análise deve ser tal que o maior período a ser estimado possa ser observado[2].

2.4.1 Autocorrelação

A função de periodicidade definida pela autocorrelação para um atributo $F[m]$ é obtida através de

$$P^{\text{corr}}[l_\tau] = \sum_{m=0}^{M-1} F^*[m]F[m + l_\tau], \tag{2.4}$$

onde $l_\tau = \text{round}\left(\frac{\tau f_s}{H}\right)$ é o período em amostras, H o salto utilizado para se obter os atributos, $\tau \in [\tau_{\min}, \tau_{\max}]$ é o período em segundos e M é o comprimento (em amostras) da região em que a função de periodicidade será calculada. Os limites $\tau_{\min}$ e $\tau_{\max}$ limitam a região de suporte da função de periodicidade e normalmente são escolhidos dentro da faixa tolerável de andamentos (digamos entre 40 BPM e 250 BPM). Observando-se a equação (2.4), é possível notar que a autocorrelação para um período l_τ é estimada computando-se o produto ponto a ponto entre o atributo original e o atributo atrasado de l_τ amostras. Quanto mais similar for

[2]Normalmente, é desejável que haja pelo menos 5 ciclos do maior período dentro de uma janela. Com isso, para se obter uma estimativa do andamento de um sinal de 40 BPM, o trecho sob de análise deveria possuir pelo menos 7,5 s de duração.

$F[m]$ a $F[m + l_\tau]$, maior resultará o somatório para o período l_τ.

A principal desvantagem desse método é o fato de $P^{\text{corr}}[l_\tau]$ resultar periódico com período igual a um hipotético período τ_0 de $F[m]$. Com isso, no lugar de um impulso no atraso correspondente ao período do sinal, é obtido um trem de impulsos. Isso ocorre mesmo que o sinal avaliado não seja exatamente periódico, mas exiba uma forte tendência a periodicidade para um determinado andamento.

A Figura 2.8 mostra o resultado da aplicação da função de autocorrelação ao fluxo espectral obtido para o sinal de exemplo. No caso, $P[l_\tau]$ é exibida como função da frequência em BPM em vez de do período, através do mapeamento $f(\text{BPM}) = \frac{60}{\tau}$, para τ em segundos. Com isso, como o sinal possui uma forte componente periódica em 120 BPM, também são observadas componentes periódicas em 60 BPM ($2\tau_0$), 40 BPM ($3\tau_0$), 30 BPM ($4\tau_0$) etc. Deve-se notar que o maior pico ocorre no andamento correto e que os demais picos possuem uma amplitude decrescente em τ. Isso pode acontecer porque o sinal não é exatamente periódico e/ou possui duração finita[3].

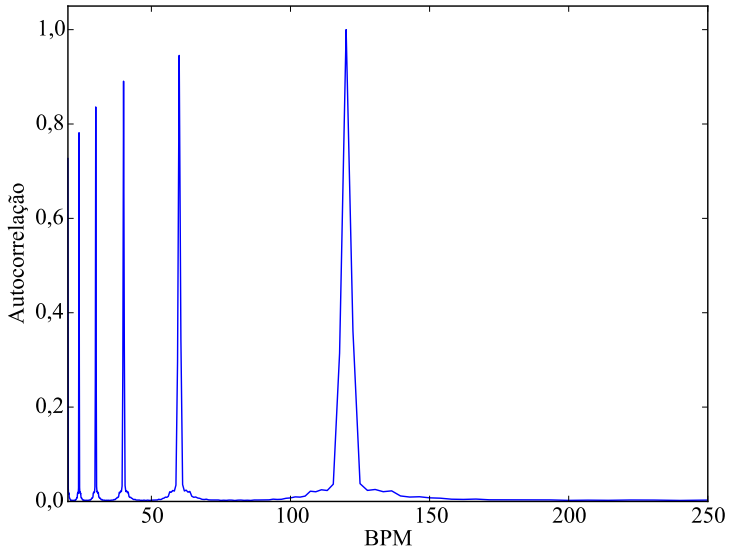


Figura 2.8: Autocorrelação calculada a partir do fluxo espectral do sinal de exemplo.

2.4.2 Módulo da Transformada de Fourier Discreta

Uma alternativa à função de autocorrelação como função de periodicidade é o uso do módulo da Transformada de Fourier Discreta (DFT, do inglês *Discrete Fourier*

[3] A duração finita faz com que os números de ciclos observados para os períodos maiores (menores valores de BPM) sejam menores, reduzindo, assim, a quantidade de termos que aparecem no somatório em m.

Transform):

$$P^{\mathrm{DFT}}[\gamma_f] = \left| \sum_{m=0}^{M-1} F[m] e^{-j\gamma_f \frac{2\pi}{M} m} \right|, \tag{2.5}$$

onde M é o comprimento da janela de observação, f é a frequência em Hz (numericamente igual ao inverso do período $\tau = \frac{1}{f}$) e $\gamma_f = \mathrm{round}\left(\frac{fMH}{fs}\right)$ é a sua versão discretizada. De forma similar ao que se fez para τ, escolhe-se f entre um valor mínimo (f_{min}) e valor máximo (f_{max}). A periodicidade do sinal é estimada, então, comparando-se os atributos $F[m]$ com cada exponencial complexa de frequência γ_f; quanto mais parecido for o atributo com a exponencial de período conhecido, maior o valor da função de periodicidade.

De forma similar à função de autocorrelação, $P^{\mathrm{DFT}}[\gamma_f]$ herda qualquer hipotética periodicidade do vetor de atributos $F[m]$ na frequência f_0. Diferentemente da obtida para a autocorrelação, a curva de $\overline{P}^{\mathrm{DFT}}[l_\tau]$ como função da frequência em BPM agora apresenta picos em múltiplos da frequência fundamental $(2f_0, 3f_0, ...)$.

A Figura 2.9 mostra a função de periodicidade obtida após o cálculo do módulo da DFT do fluxo espectral extraído do sinal de exemplo. A função obtida exibe um pico em 120 BPM (o andamento do sinal de exemplo) e em seus múltiplos. De forma similar à autocorrelação, os picos decaem conforme f aumenta (novamente, devido à janela de observação finita).

Deve-se notar que o vetor de atributos usualmente é similar a um trem de impulsos (ver Figura 2.7), e não a um sinal senoidal (empregado pela DFT). Essa discrepância faz com que a DFT não seja muito utilizada para a estimação de periodicidade. Contudo, como será visto na próxima seção, a informação obtida através da DFT pode ser utilizada para melhorar a função de periodicidade.

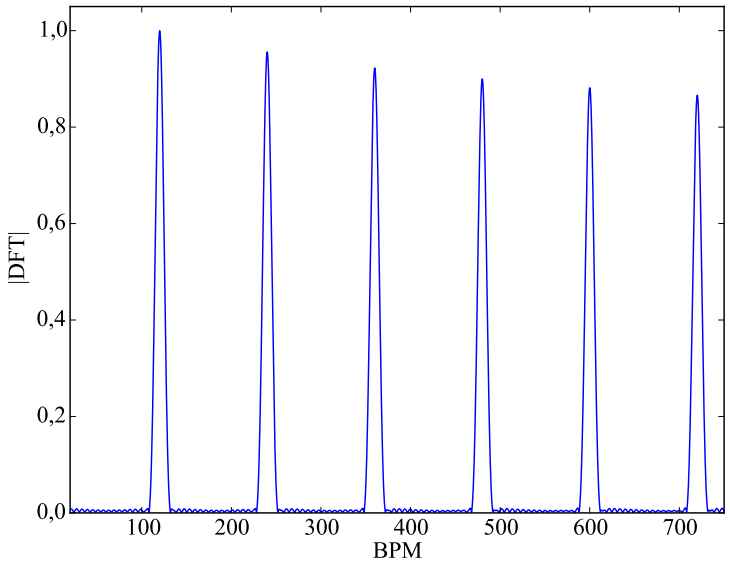


Figura 2.9: Módulo da DFT do fluxo espectral do sinal de exemplo.

2.4.3 Produto Autocorrelação × Módulo da DFT

Observando-se as funções de similaridade obtidas utilizando a autocorrelação e a DFT, é possível notar que elas são calculadas em domínios recíprocos: período e frequência, respectivamente. Com isso, os picos indesejados aparecem em posições diferentes: múltiplos do período e múltiplos da frequência, respectivamente. Levando isso em consideração, em [2] é proposta uma função de periodicidade obtida através do produto entre o módulo da DFT e a autocorrelação, chamada aqui de $P^{\text{prod}}[\gamma_f]$

Para se obter o produto, inicialmente é necessário mapear as funções no mesmo domínio. Nesta seção, será apresentada a solução chamada de função de autocorrelação mapeada na frequência [2], que mapeia período em frequência por interpolação da função de autocorrelação. São detalhadas em [57] outras possíveis formas de cálculo desta função de periodicidade.

O primeiro passo para a obtenção de $P^{\text{prod}}[l_\tau]$ é a obtenção de $P^{\text{DFT}}[\gamma_f]$. Para melhorar a resolução do método para frequências baixas (períodos longos), usualmente é realizada uma DFT com comprimento mais longo que o do atributo, estendendo-o com zeros antes do cálculo. Resultados adequados foram obtidos utilizando-se uma DFT com 4 vezes o comprimento do sinal.

Em seguida, é obtido $P^{\text{corr}}[l_\tau]$. De posse dos períodos τ e das frequências f, obtém-se uma versão $\overline{P}^{\text{corr}}[\gamma_f]$ desta periodicidade, mapeada na frequência através da interpolação da função $P^{\text{corr}}[l_\tau]$ para os valores de frequência associados a γ_f. Observou-se experimentalmente que uma simples interpolação linear é suficiente para que se obtenham resultados adequados. Finalmente, calcula-se a função de periodicidade através de

$$P^{\text{prod}}[\gamma_f] = \overline{P}^{\text{corr}}[\gamma_f] P^{\text{DFT}}[\gamma_f]. \tag{2.6}$$

Para ilustrar o procedimento, mostra-se na Figura 2.10 a função de autocorrelação da Figura 2.8 mapeada para as frequências correspondentes à DFT cujo módulo é mostrado na Figura 2.9. Realizando-se o produto entre a autocorrelação mapeada e o módulo da DFT do fluxo espectral obtido no sinal de exemplo, obtém-se a função de periodicidade mostrada na Figura 2.11. Como se vê, a função de periodicidade possui apenas um pico proeminente, localizado no andamento do sinal.

2.5 Estimação do Tempo

Uma vez obtida a função de periodicidade, é necessário escolher uma estratégia para selecionar o andamento do sinal, ou um conjunto de candidatos ao andamento. Essa etapa, por ser a final, é muito dependente da escolha feita nas etapas anteriores,

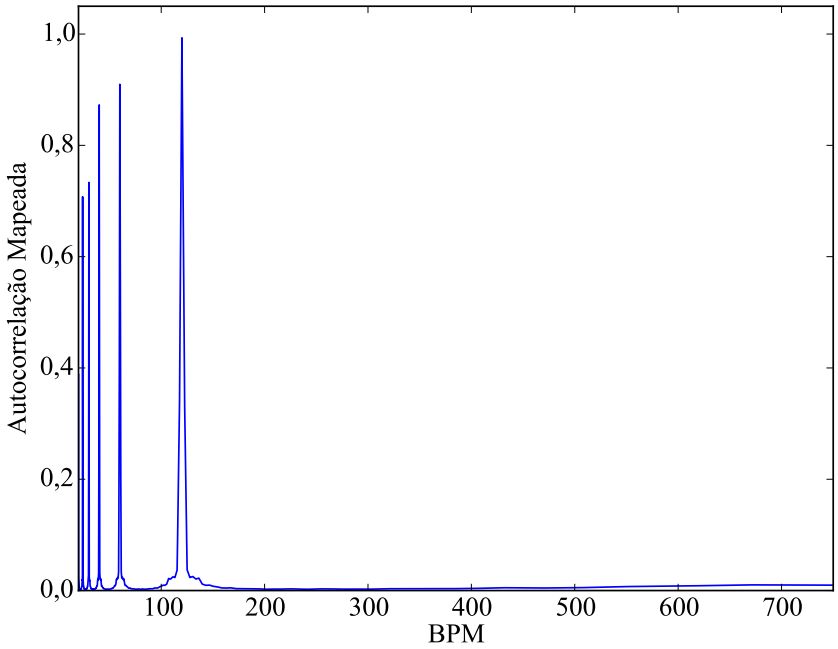


Figura 2.10: Autocorrelação da Figura 2.8 mapeada para as frequências da Figura 2.9.

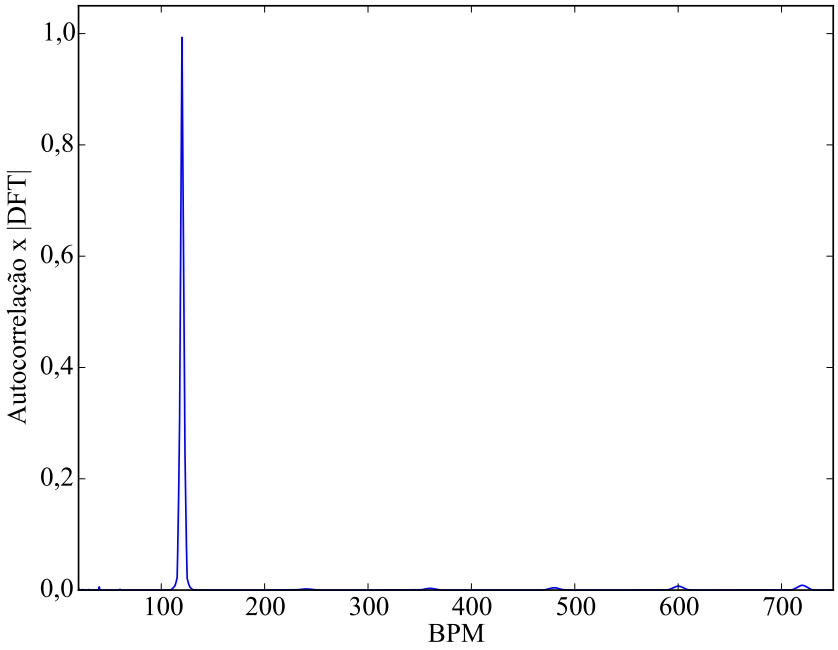


Figura 2.11: Produto autocorrelação e DFT obtido a partir do fluxo espectral do sinal de exemplo.

sendo, quase sempre, projetada para um conjunto específico de atributos e funções de periodicidade. De forma geral, a detecção é feita assumindo-se que regiões em que a função de periodicidade é mais elevada são boas candidatas ao andamento do sinal. Sendo assim, diversos métodos empregam uma simples seleção do pico

mais proeminente da função de periodicidade [45, 55, 56, 61, 67, 68]. Dentre os demais métodos, devem-se destacar os que utilizam informação do comportamento rítmico de sinais de música e os que utilizam informação cognitiva. A seguir, serão brevemente revisados métodos que empregam essas duas estratégias.

2.5.1 Inclusão de Informação de Alto Nível

Idealmente, o pico de maior intensidade na função de periodicidade corresponderia ao andamento. No entanto, como visto na seção anterior, o próprio cálculo da função de periodicidade pode levar ao surgimento de picos com alta intensidade indesejados. Além disso, a estrutura rítmica de uma peça pode levar ao surgimento de picos em múltiplos e/ou submúltiplos do andamento. Por exemplo, se uma música é ternária (um compasso é formado por três tempos—cada um correspondendo a um tactus), é possível que ocorra um pico representando esse período reduzido (na posição igual ao triplo do andamento da música).

Foram propostas na literatura diversas estratégias que procuram contornar o problema mencionado no parágrafo anterior. Por exemplo, em [24], são escolhidos como candidatos ao andamento os três picos de maior intensidade da função de autocorrelação; o andamento é selecionado a partir de heurísticas aplicadas sobre os candidatos. Em [43], foi proposta uma função que mede as chances de dois picos da função de periodicidade estarem associados ao andamento da música. Em [46] e [56], são utilizados histogramas que medem a chance de um determinado pico na função de periodicidade corresponder ao andamento do sinal. De forma geral, estes métodos aproveitam o fato de que um sinal com andamento f_0 exibe picos em sua função de periodicidade em múltiplos e submúltiplos de f_0.

É mostrada em [2] uma forma de levar em consideração a estrutura rítmica do sinal ao se estimar o andamento. Esta estaria associada a padrões rítmicos, cada um derivado para um tipo de compasso (por exemplo, binário, ternário e composto). Cada um desses padrões rítmicos informa quais múltiplos e sub-múltiplos são esperados, permitindo estimar as combinações mais prováveis de andamento e divisão rítmica, dada uma função de periodicidade observada [2]. Será feita no Capítulo 4 uma descrição mais detalhada desses padrões rítmicos. Deve-se ressaltar que esses padrões também podem ser "aprendidos", levando-se em conta, por exemplo, o gênero da música sob análise [72].

2.5.2 Inclusão de Dicas Cognitivas

Dicas cognitivas podem ser incluídas nos algoritmos de detecção de picos através do uso de curvas de ponderação que modelam a preferência observada por andamentos próximos a 120 BPM [1]. Diferentes curvas de ponderação foram propostas em [25,

49, 73–75] e utilizadas para detecção de tempo em [27, 76, 77]. A seguir, será descrita uma dessas curvas.

Em [73, 74] foram realizados diversos testes subjetivos onde os participantes marcaram o pulso de músicas populares escolhidas cuidadosamente. A partir da análise do intervalo entre os pulsos anotados, os autores geraram um modelo para a preferência por cada andamento. O modelo gerado, chamado de modelo de ressonância, é

$$W[\gamma_f] = \frac{1}{\sqrt{(f_0^2 - f^2)^2 + \beta f^2}} - \frac{1}{\sqrt{f_0^4 + f^4}}, \tag{2.7}$$

onde f_0 é a frequência de ressonância do modelo, escolhida como $f_0 = 138/60$ e $\beta = 5$ é o fator de amortecimento. É mostrada na Figura 2.12 a função de ponderação W para diferentes f. Pode-se ver que W exibe um pico em aproximadamente 120 BPM e decai rapidamente para 0 fora da faixa de andamentos representada. De posse

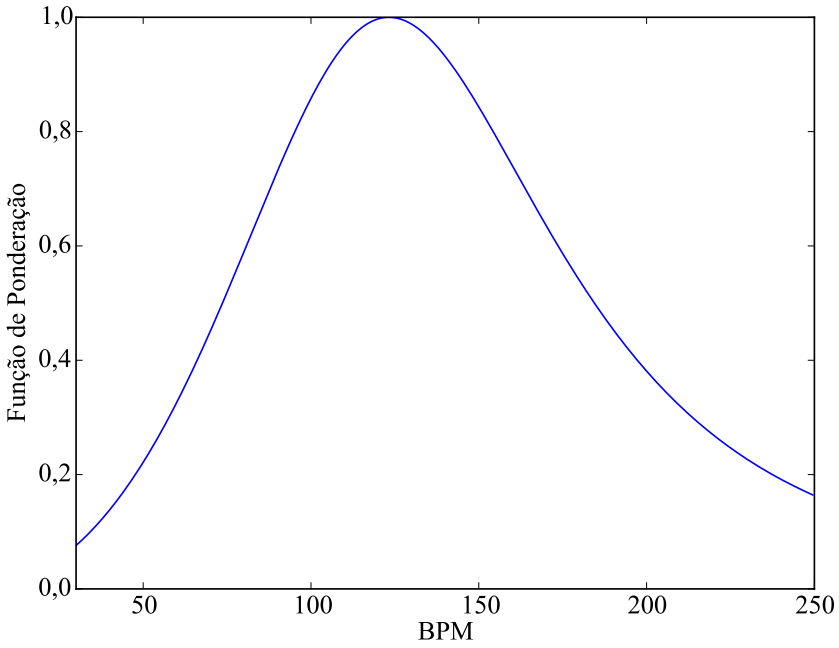


Figura 2.12: Função de ponderação utilizando o modelo de ressonância.

de W, pode-se multiplicar cada andamento da função de periodicidade pela sua "preferência", levando a uma função de periodicidade ponderada por informação cognitiva. Neste caso, ao se selecionar o máximo desta nova função de periodicidade, será levada em consideração a preferência por andamentos próximos de 120 BPM. Este modelo foi validado utilizando-se um outro conjunto de sinais. Observou-se que o modelo é mais bem respeitado para sinais com andamentos próximos de 120 BPM do que para sinais com andamentos abaixo de 75 BPM e acima de 175 BPM [73]. Isso ocorre porque há uma variação maior entre as opiniões de diferentes participantes do teste para sinais cujos andamentos estão nas regiões de menor amplitude da curva.

Deve ser ressaltado, no entanto, que, apesar de modelarem adequadamente a distribuição de opiniões para um conjunto de sinais, as funções de ponderação não modelam a distribuição dos andamentos atribuídos a uma mesma música, que se mostra bimodal [14] com o andamento correspondendo a uma moda e um múltiplo ou submúltiplo, à outra moda. Isso caracteriza uma confusão perceptiva sobre qual nível hierárquico conteria o verdadeiro andamento da peça [1].

2.6 Conclusão

Neste capítulo foi feita uma breve revisão dos métodos de estimação do andamento de uma música. Estes foram organizados numa estrutura modular em três etapas, que permitiu uma breve descrição de uma grande quantidade de algoritmos da literatura, ressaltando-se as contribuições individuais de cada trabalho.

Conforme mencionado anteriormente, foi considerado apenas o caso em que se deseja estimar um único andamento para todo o sinal de áudio. Isso é válido se o andamento variar pouco dentro da mesma música. Caso isso não aconteça, pode-se aplicar os métodos descritos neste capítulo a blocos do sinal, gerando para cada um uma estimativa ou um conjunto de estimativas para o andamento da música. No caso de mais de uma estimativa de andamento, pode-se ordená-las por suas probabilidades, conforme feito em [53, 58, 72].

No próximo capítulo, será realizada uma avaliação do desempenho de alguns dos métodos descritos neste capítulo. Também será feita uma análise de como estes métodos podem ser combinados de forma a gerar um algoritmo de estimação do andamento. Em particular, serão estudados os pontos fracos de cada algoritmo, de forma a mapear o que pode ser melhorado.

Capítulo 3

Comparação de Métodos para Estimação do Andamento

Neste capítulo, serão feitas avaliações de desempenho dos métodos envolvidos na estimação de andamento descritos no capítulo anterior. O objetivo destas avaliações é escolher dentre eles os de melhor desempenho, a partir dos quais se criará um método-protótipo de estimação de andamento para servir de plataforma para teste de novos algoritmos. Uma vez montando este método, ele próprio será submetido a uma série de análises de desempenho com a finalidade de mapear suas principais deficiências.

Para realizar a avaliação de desempenho, inicialmente serão descritos os bancos de sinais utilizados e as figuras de mérito empregadas. Em seguida, o método de estimação será definido através de uma etapa de cada vez: primeiro os atributos, depois a função de similaridade e, por fim, a estratégia de seleção do andamento. Em cada um desses testes, será adotada a mesma metodologia: apenas o bloco referente à etapa sob teste será variado, sendo fixados os demais algoritmos.

Este capítulo é organizado da seguinte forma. Inicialmente, são descritos na Seção 3.1 os bancos de sinais. As figuras de mérito utilizadas são apresentadas na Seção 3.2. Diferentes formas do cálculo do fluxo espectral são avaliadas na Seção 3.3. Na Seção 3.4, são comparados os desempenhos de funções de periodicidade. O impacto do uso de informação cognitiva é estudado na Seção 3.5. O método construído a partir dos resultados das seções anteriores é avaliado na Seção 3.6. Por fim, são apresentadas as conclusões na Seção 3.7.

3.1 Banco de Sinais

Nesta seção, são apresentados os três bancos de sinais utilizados neste capítulo, escolhidos por possuírem sinais com o andamento anotado manualmente. Além disso, estes bancos foram empregados em outros trabalhos da literatura, o que facilita a comparação dos resultados aqui obtidos com os naqueles relatados.

Os três bancos de sinais são chamados *Ballroom Dancer*, *Song Excerpts* e *Hainsworth*. Os dois primeiros foram desenvolvidos originalmente para o MIREX de 2004 e são detalhados em [22]. O terceiro foi desenvolvido para a tese de doutorado de Stephen Hainsworth [78], onde é descrito. A Tabela 3.1 contém informação relevante sobre cada banco de sinais. Os três bancos de sinais foram gentilmente cedidos pelo Dr. Fabien Gouyon, do INESC. Dois bancos (*Ballroom* e *Hainsworth*) possuem também anotação do gênero de cada sinal. Os sinais presentes nos três bancos foram amostrados a 44,1 kHz com uma precisão de 16 bits.

Tabela 3.1: Informações gerais sobre os bancos de sinais empregados.

	Ballroom	*Excerpts*	*Hainsworth*
Número de sinais	698	465	221
Duração de cada sinal (s)	30	20	45
Duração total (s)	20940	9300	11906
Faixa de andamento (BPM)	[60, 224]	[24, 242]	[52, 198]

3.2 Figuras de Mérito

A fim de quantificar o desempenho de algoritmos de estimação de andamento, foram propostas na literatura diversas figuras de mérito [22, 74], dentre as quais duas se destacam por serem amplamente utilizadas em publicações: a Acurácia 1 e a Acurácia 2, descritas abaixo para um conjunto arbitrário de sinais.

- Acurácia 1 = percentual de andamentos estimados com uma diferença menor que 4 % do valor anotado;
- Acurácia 2 = percentual de andamentos estimados com uma diferença menor que 4 % do valor anotado, ou da metade, ou de um terço, ou do dobro, ou do triplo do seu valor.

A Acurácia 1 é uma figura de mérito mais restringente, considerando como acertos apenas os casos em que o andamento estimado e anotado são iguais dentro de uma

margem estreita de erro. Por outro lado, a Acurácia 2 é mais leniente, considerando como corretos, dentro de uma mesma faixa, andamentos que são múltiplos e divisores do andamento anotado. Se o objetivo do método for estimar o andamento de um sinal, claramente a Acurácia 1 é a figura de mérito mais adequada. Se o objetivo for utilizar o andamento como entrada de algum outro método (rastreamento de pulsos, por exemplo), então errar o andamento para um divisor ou múltiplo não é tão problemático, especialmente se for considerado que até mesmo seres humanos cometem esse tipo de confusão [14]. Deve-se ressaltar que o valor 4 % é o tipicamente utilizado na literatura (e também nesta tese), mas valores diferentes podem ser adotados dependendo da aplicação sendo avaliada.

O MIREX, a partir de sua edição em 2006, adotou uma nova figura de mérito chamada de *pscore* [73, 74]. Esta figura de mérito leva em consideração que a distribuição de opiniões acerca de um mesmo sinal é bimodal, e utiliza esse fato para quantificar o desempenho de um algoritmo. Infelizmente, essa figura de mérito demanda que um mesmo sinal seja anotado por uma grande quantidade de pessoas, para que as duas modas possam ser corretamente estimadas. Como os bancos disponíveis publicamente não possuem esse tipo de anotação, tal figura de mérito acaba não sendo utilizada em publicações, tendo seu uso restrito ao MIREX.

3.3 Comparação de Formas de Cálculo do Fluxo Espectral

Conforme mencionado na Seção 2.3.1, o fluxo espectral foi escolhido como plataforma para extração de atributos. É de interesse, portanto, estudar como as diferentes formas de obtê-lo influenciam o desempenho de um método de estimação de andamento, objetivo desta seção. Seu texto é fortemente baseado em [79], onde foram publicados os resultados nela apresentados.

3.3.1 Metodologia

Nesta seção, será apresentada a metodologia empregada para a comparação entre diferentes formas de cálculo do fluxo espectral. Para medir o impacto das diferentes escolhas de algoritmos para análise foi criado um algoritmo-protótipo que consiste em três etapas: cálculo do fluxo espectral (extração de atributo), autocorrelação (cálculo da função de similaridade) e seleção do pico de maior amplitude (detecção do andamento). A seguir, cada uma destas etapas será detalhada.

A etapa de cálculo do fluxo espectral pode ser dividida nos estágios mostrados na Figura 3.1. Na figura, são opcionais os estágios denotados por caixas cujas bordas

são tracejadas. Os demais compõem etapas básicas para obtenção do fluxo espectral (descritas na Seção 2.3.1).

Resumidamente, são utilizados os seguintes passos no cálculo do atributo: O sinal pode ser sub-amostrado por D antes do cálculo da STFT. A STFT é calculada utilizando-se uma janela de Hann com comprimento N e salto H e o número de raias na frequência é determinado pelo comprimento da janela (não é feita inserção de zeros no sinal). A STFT pode ser mapeada para a escala Mel utilizando-se um número arbitrário de filtros E; se $E = 0$ o mapeamento não é realizado. Em seguida, é calculada a diferença entre quadros adjacentes utilizando-se o valor absoluto da STFT ou o seu logaritmo na base 10 (se L=Verdadeiro, é utilizado o logaritmo). O sinal resultante é então retificado. Finalmente, os sinais de cada raia da frequência podem ser somados (S=Verdadeiro), obtendo-se apenas um fluxo espectral $F^{\mathrm{SF}}[m]$ para cada quadro m analisado (de forma similar à equação (2.2)). Se S=Falso, a soma não é feita e um fluxo espectral $F^{\mathrm{SF}}[m, k]$ é obtido para cada raia de frequência k e quadro m (equivalente a omitir o somatório da equação (2.2)).

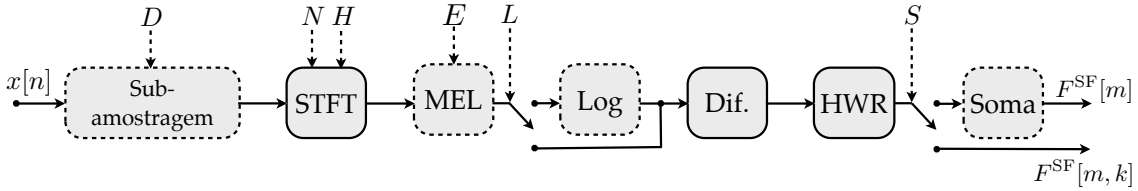


Figura 3.1: Estágios empregados no cálculo do fluxo espectral. São opcionais os estágios denotados por caixas com borda tracejada. As linhas tracejadas verticais denotam os parâmetros que podem ser escolhidos em cada estágio.

A função de periodicidade empregada foi a autocorrelação (descrita na Seção 2.4.1), calculada para atrasos (períodos) máximo e mínimo equivalentes a 250 BPM e 40 BPM, respectivamente. No caso de mais de um fluxo espectral (quando S=Falso), é calculada uma função de periodicidade para cada fluxo espectral, normalizada de forma que seu maior valor seja igual a 1. A função de periodicidade final é obtida como a média das funções para cada raia. O andamento é estimado como o pico de maior amplitude da função de similaridade.

O algoritmo utilizado neste experimento enfatiza a etapa de extração de atributos, permitindo que esta incorpore uma grande quantidade das modificações mais comumente empregadas sobre o fluxo espectral (ver Seção 2.3.2). Foram escolhidos para as demais etapas algoritmos conhecidos da literatura que permitem uma fácil interpretação dos resultados. Além disso, a simplicidade dos algoritmos empregados garante que eles não mascaram nenhuma deficiência do fluxo espectral propriamente dito, o que poderia acontecer se algoritmos mais elaborados fossem escolhidos.

3.3.2 Avaliação

O objetivo desta avaliação é verificar o impacto de cada parâmetro opcional sobre o desempenho de um algoritmo de estimação de andamento. Além disso, serão testados dois diferentes valores para o comprimento da janela e o tamanho do salto. Os valores testados para cada parâmetro podem ser vistos na Tabela 3.2.

Tabela 3.2: Parâmetros escolhidos para a avaliação de desempenho do fluxo espectral.

Parâmetro (Estágio)	Valor
D (Sub-amostragem)	$\{1, 4\}$
N (STFT)	$\{20, 40\}$ ms
H (STFT)	$\{5, 10\}$ ms
E (Mel)	$\{0, 20\}$
L (Log)	{Verdadeiro, Falso}
S (Soma)	{Verdadeiro, Falso}

Utilizando esses parâmetros, foram executadas 64 avaliações de desempenho para os bancos de sinais *Ballroom* e *Excerpts*, cada avaliação consistindo de uma possível combinação dos parâmetros exibidos na Tabela 3.2. Ao fim de cada avaliação foram calculadas as figuras de mérito Acurácia 1 e Acurácia 2 para cada banco de sinais e para um terceiro, composto pela união dos sinais nos dois bancos avaliados, que será chamado de banco *Conjunto*.

3.3.3 Resultados

Os resultados serão apresentados separadamente para cada figura de mérito. Como o objetivo é avaliar o impacto de cada atributo individualmente sobre o desempenho de um método de estimação de andamento, procurou-se uma figura de mérito que medisse quão consistentemente melhor é uma determinada escolha de um parâmetro em relação à outra. Para isso, os resultados da Acurácia 1 e da Acurácia 2 foram calculados para cada um dos 64 conjuntos distintos de parâmetros. Para cada figura de mérito, estes conjuntos foram ordenados de acordo com seu desempenho de forma decrescente: o conjunto de parâmetros com melhor resultado na primeira posição e com pior resultado na última. Em seguida, para cada figura de mérito, computou-se a média entre as 32 posições em que aparecia cada valor de parâmetro nos 64 testes possíveis. Dessa forma, uma posição média em torno de 32 indicaria uma pequena influência da escolha sobre os resultados, uma posição menor que 32 uma influência positiva nos resultados, e uma posição acima uma influência negativa. Também foi calculada a melhor posição obtida por cada parâmetro: um valor igual

Tabela 3.3: Posição média para a Acurácia 1 para cada banco de sinais e valor de parâmetro. É mostrada entre parênteses a melhor posição obtida.

Banco	D		N		H	
	1	4	20 ms	40 ms	5 ms	10 ms
Excerpts	**25** (2)	41 (2)	36 (5)	**30** (1)	34 (1)	**31** (2)
Ballroom	40 (3)	**26** (1)	35 (5)	**30** (1)	**32** (1)	34 (5)
Conjunto	32 (5)	**30** (1)	35 (2)	**30** (1)	**32** (1)	33 (4)

Banco	E		L		S	
	0	20	Falso	Verdadeiro	Falso	Verdadeiro
Excerpts	45 (22)	**20** (1)	**30** (1)	35 (2)	**27** (1)	39 (7)
Ballroom	38 (3)	**28** (1)	**22** (1)	34 (5)	36 (2)	**29** (1)
Conjunto	44 (5)	**22** (1)	**23** (1)	43 (10)	**32** (1)	33 (3)

a 1 indica que este parâmetro foi escolhido no teste que obteve o melhor resultado numa determinada métrica.

Acurácia 1

São mostradas na Tabela 3.3 as posições médias e a melhor posição para cada valor de parâmetro para os três bancos de sinais. A partir da tabela, podem ser feitas as seguintes observações:

- O fator de sub-amostragem D parece gerar uma pequena piora nos resultados;
- Janelas mais longas geram resultados melhores;
- Saltos menores parecem levar a melhores resultados, apesar de o efeito parecer pequeno;
- O uso da escala Mel leva a melhores resultados na Acurácia 1;
- O uso do logaritmo leva a piores resultados;
- Foram observados resultados inconclusivos para o parâmetro S.

Dentre as 64 avaliações executadas, a seguinte configuração obteve o maior valor para Acurácia 1:

- *Excerpts*: 32 % para $\{D = 4, N = 40\,\text{ms}, H = 5\,\text{ms}, E = 20, L = \text{Falso}, S = \text{Falso}\}$
- *Ballroom*: 41 % para $\{D = 4, N = 40\,\text{ms}, H = 5\,\text{ms}, E = 20, L = \text{Verdadeiro}, S = \text{Falso}\}$;

- *Conjunto*: 37 % para $\{D = 1, N = 40\,\text{ms}, H = 10\,\text{ms}, E = 20, L = \text{Verdadeiro}, S = \text{Falso}\}$.

Como pode ser observado, os resultados para o *Ballroom* foram melhores que os obtidos para o *Excerpts*; este mesmo fato foi observado em outros estudos [80]. Além disso, a configuração com melhor desempenho médio sempre utilizou $D = 4$, mapeamento para escala Mel e $L = \text{Falso}$.

Quando são selecionados os parâmetros que obtiveram a melhor posição média (os valores em negrito na Tabela 3.3), são encontrados os seguintes resultados:

- *Excerpts*: 31 % para $\{D = 1, N = 40\,\text{ms}, H = 10\,\text{ms}, E = 20, L = \text{Falso}, S = \text{Falso}\}$;
- *Ballroom*: 41 % para $\{D = 4, N = 40\,\text{ms}, H = 5\,\text{ms}, E = 20, L = \text{Falso}, S = \text{Verdadeiro}\}$;
- *Conjunto*: 28 % para $\{D = 1, N = 40\,\text{ms}, H = 5\,\text{ms}, E = 20, L = \text{Falso}, S = \text{Falso}\}$.

Para os bancos *Ballroom* e *Excerpts*, os resultados obtidos com os parâmetros com menor posição média são similares aos que obtiveram maiores valores de Acurácia 1. Para o caso do *Excerpts*, os únicos valores de parâmetros que diferem são o fator de sub-amostragem e o salto entre janelas, produzindo uma diferença de 1 ponto percentual em relação ao melhor resultado. Isso corrobora a observação anterior de que estes parâmetros têm pequeno impacto sobre o desempenho do método. Para o caso do *Ballroom*, os parâmetros com menor posição média foram os que levaram aos maiores valores de Acurácia 1, exceto L e S. Já considerando-se a união dos dois bancos de sinais, houve uma piora significativa do desempenho, apesar de apenas um parâmetro ter sido modificado (D). Isso mostra que os resultados para um banco de sinais, pelo menos para D, não são necessariamente consistentes com os de outro conjunto. O caso da sub-amostragem ficará mais claro na próxima seção, quando serão analisados os resultados da Acurácia 2.

Acurácia 2

São mostradas na Tabela 3.4 a posição média e melhor posição para os três bancos de sinais. As seguintes observações podem ser feitas a partir do resultado:

- Diferentemente do que foi observado para a Acurácia 1, o uso de sub-amostragem teve um grande impacto no desempenho;
- Novamente, resultados melhores foram obtidos para comprimentos maiores de janela;

Tabela 3.4: Posição média para a Acurácia 2 para cada banco de sinais e valor de parâmetro. É mostrada entre parênteses a melhor posição obtida.

Banco	*D*		*N*		*H*	
	1	4	20 ms	40 ms	5 ms	10 ms
Excerpts	**20** (1)	46 (19)	36 (6)	**30** (1)	34 (3)	**32** (1)
Ballroom	**30** (1)	36 (17)	37 (7)	**29** (1)	34 (1)	**31** (1)
Conjunto	**23** (1)	42 (21)	36 (7)	**29** (1)	35 (5)	**31** (1)

Banco	*E*		*L*		*S*	
	0	20	Falso	Verdadeiro	Falso	Verdadeiro
Excerpts	42 (12)	**23** (1)	**32** (6)	34 (1)	**30** (1)	35 (2)
Ballroom	34 (4)	**31** (2)	**17** (1)	48 (27)	**30** (2)	35 (1)
Conjunto	39 (11)	**26** (1)	**22** (1)	44 (17)	**29** (2)	36 (1)

- Saltos maiores foram favorecidos, apesar de a diferença ser muito pequena;
- O uso de filtros Mel melhorou o desempenho, como observado para Acurácia 1;
- O uso do logaritmo piorou os resultados;
- S=Falso levou a melhores resultados, indicando que a combinação de funções de periodicidade pode facilitar a detecção de múltiplos ou submúltiplos do andamento verdadeiro do sinal.

Com isso, pode-se dizer que os resultados obtidos com a Acurácia 2 tenderam a corroborar os resultados observados para a Acurácia 1. O fato encontrado mais interessante é a influência do fator de sub-amostragem, que pode indicar que a região de frequências altas (descartadas quando $D = 4$) carrega informação relevante sobre múltiplos e submúltiplos do andamento de referência. Há um forte indicativo que janelas maiores levam a melhores desempenhos. Além disso, apesar de os resultados indicarem que o uso da escala Mel é vantajoso, ainda é necessário avaliar o impacto do número de filtros sobre o desempenho.

3.3.4 Testes Adicionais

Nesta seção, serão descritos dois testes adicionais: um para avaliar a influência do comprimento da janela e do tamanho do salto, e outro para medir a influência do número de filtros Mel. Cada teste será descrito separadamente.

Comprimento da Janela e Tamanho do Salto

O objetivo deste teste é verificar como diferentes valores para o comprimento da janela e o tamanho do salto afetam o desempenho da estimação de andamento.

Serão considerados para esses parâmetros valores numa faixa mais ampla do que no teste anterior. Tal como observado anteriormente, janelas mais longas parecem levar a melhores resultados, mas os resultados foram pouco conclusivos.

Os seguintes valores foram escolhidos para o comprimento da janela e o tamanho do salto: $N \in \{40, 80, 120\}$ ms e $H \in \{10, 20, 40\}$ ms. Com exceção de N e H, os outros parâmetros foram fixados com valores que produziram bom desempenho no teste anterior: $D = 1$, $E = 20$, $L =$ Falso. Como o impacto de S no desempenho não ficou claro no teste anterior, seus valores também foram variados neste teste: $S \in \{\text{Verdadeiro, Falso}\}$.

A Tabela 3.5 contém os resultados para todas as combinações de comprimento de janela e tamanho do salto quando é calculada a média para os dois valores de S.

Para Acurácia 1, os resultados indicam que o tamanho do salto igual a 20 ms pode levar a melhores resultados para o *Ballroom*, o que se reflete no *Conjunto*. Os resultados para janelas mais longas é menos claro, com o melhor resultado variando em função do salto escolhido e do banco de sinais.

Quando se considera a Acurácia 2, saltos menores tendem a produzir melhores resultados, e os ganhos obtidos com saltos menores são maiores que os obtidos com janelas mais longas. Em particular, o desempenho cresce significativamente quando o salto é reduzido de 40 ms para 20 ms. Em relação ao comprimento da janela, a melhora é mais clara para o *Excerpts* do que para o *Ballroom*, indicando que o comprimento ideal de janela pode variar de acordo com as características do sinal.

Tabela 3.5: Resultados médios para maior número de valores para o comprimento da janela e para o valor do salto. Todos os valores em %.

	Acurácia 1								
	Excerpts			*Ballroom*			*Conjunto*		
$N \setminus H$ (ms)	10	20	40	10	20	40	10	20	40
40	30	30	30	30	35	28	30	33	29
80	32	32	32	28	34	28	30	33	30
120	34	32	33	30	31	28	32	32	30

	Acurácia 2								
	Excerpts			*Ballroom*			*Conjunto*		
$N \setminus H$ (ms)	10	20	40	10	20	40	10	20	40
40	73	72	67	84	83	74	79	79	71
80	76	75	70	84	83	76	80	80	74
120	78	76	71	83	82	76	81	80	74

Número de Filtros Mel

No teste descrito na Seção 3.3.3, os resultados indicaram que utilizar o mapeamento para escala Mel tende a melhorar o desempenho. Nesta seção, é estudado o efeito do número de filtros Mel empregados. Para isso, o número de filtros foi variado dentro do conjunto {10, 20, 40, 80}, enquanto os outros parâmetros foram fixados como $D = 1$, $N = 80\,\text{ms}$, $H = 10\,\text{ms}$, $L =$ Falso, e $S \in \{\text{Verdadeiro, Falso}\}$, seguindo os mesmos critérios adotados para o teste descrito na seção anterior.

Podem ser vistos na Tabela 3.6 os resultados para ambas as figuras de mérito, dos quais se pode concluir que o número de filtros tem um impacto pequeno no desempenho. Similarmente ao que foi reportado para a divisão do sinal em sub-bandas [22, 23], o número de filtros Mel parece não ser um parâmetro significativo, desde que o mapeamento seja empregado. Quanto ao valor de S, ele apenas teve influência na Acurácia 1 para o banco *Excerpts*; precisam ser realizadas mais investigações para entender a influência deste parâmetro sobre o desempenho.

Tabela 3.6: Resultados para o teste variando o número de filtros Mel. Todos os valores em %.

	Acurácia 1					
	Excerpts		*Ballroom*		*Conjunto*	
$E \setminus S$	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
10	33	26	32	32	32	29
20	32	26	31	31	31	28
40	33	25	31	31	32	27
80	32	24	29	29	32	26

	Acurácia 2					
	Excerpts		*Ballroom*		*Conjunto*	
$E \setminus S$	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
10	76	77	85	83	81	81
20	77	76	84	85	81	81
40	77	75	83	84	81	81
80	77	75	82	83	80	80

3.3.5 Normalização da Função de Similaridade

Dados os resultados inconclusivos obtidos ao se calcular funções de periodicidade para cada sub-banda, considerou-se necessário investigar se a normalização aplicada em cada função de periodicidade tem influência no desempenho global, quando a soma não é utilizada.

Para isso, foram investigadas duas formas de normalizar a função de periodicidade. A primeira forma é a já utilizada nos testes anteriores, onde a função de periodicidade para a raia k, $P_k[m]$, é normalizada por $\max(|P_k[l]|)$. A forma alternativa investigada considera a normalização pela energia do sinal na raia k, o que equivale a normalizá-la por $\sum_m |F[m,k]|^2$.

Para os resultados exibidos a seguir, foram utilizados os seguintes parâmetros: $D = 1$, $N = 80\,\text{ms}$, $H = 10\,\text{ms}$, $E \in \{0, 20\}$, $L =$ Falso, e $S =$ Falso. Estes valores foram escolhidos por terem produzido melhor desempenho nas diferentes avaliações realizadas anteriormente. Como o efeito da normalização pode variar se for utilizado ou não o mapeamento para escala Mel (já que a distribuição de energia de cada raia é modificada pelo mapeamento), decidiu-se testar as normalizações com e sem o mapeamento. O parâmetro S foi escolhido como Falso porque se deseja comparar entre si apenas os efeitos das diferentes normalizações nessa etapa.

A Tabela 3.7 contém os resultados da avaliação. Como se pode observar, não houve grande diferença entre os resultados produzidos pelas duas formas de normalização, o que permite supor que o emprego de uma normalização diferente não deve alterar os resultados obtidos nas avaliações anteriores.

Então, aparentemente a escolha de se somar ou não os atributos precisa levar em consideração outros fatores como os algoritmos empregados e qual informação das diferentes raias é integrada.

Tabela 3.7: Resultados para o teste variando a normalização da função de periodicidade. Todos os valores em %.

	Acurácia 1					
	Excerpts		*Ballroom*		*Conjunto*	
E \ Norm.	Máximo	Energia	Máximo	Energia	Máximo	Energia
0	27	27	25	25	26	26
20	31	31	26	26	28	28

	Acurácia 2					
	Excerpts		*Ballroom*		*Conjunto*	
E \ Norm.	Máximo	Energia	Máximo	Energia	Máximo	Energia
0	66	66	81	81	81	81
20	76	76	85	85	85	85

3.3.6 Discussão

De forma geral, constatou-se que o uso da escala Mel melhora o desempenho do algoritmo de estimação de andamento, comprimir o valor da STFT utilizando o

logaritmo piora o desempenho, e a sub-amotragem melhora o desempenho da Acurácia 1, mas piora o da Acurácia 2.

Também foi estudado o comprimento da janela de observação e do salto entre janelas, com resultados previsíveis: janelas mais longas e saltos mais curtos melhoram o resultado (especialmente se a Acurácia 2 for considerada), mas o ganho no desempenho é pequeno se levado em consideração o custo computacional. O efeito de somar as raias da STFT no cálculo do fluxo espectral obteve resultados inconclusivos. Aparentemente, o benefício ou prejuízo de se somar as raias varia de sinal para sinal. O efeito do número de filtros Mel escolhido também foi investigado, sem que se constatasse influência sua sobre o desempenho dos métodos.

Em termos geral, o objetivo destes testes foi servirem de guias para escolha dos parâmetros do fluxo espectral. Em particular, utilizando estes resultados foi possível escolher com segurança os atributos empregados nas análises de desempenho que serão realizadas a seguir.

3.4 Comparação dos Métodos de Seleção de Periodicidade

Nesta seção, será comparado o desempenho dos três métodos de cálculo de função de periodicidade descritos na Seção 2.4. O objetivo é verificar qual dos três métodos obtém o melhor resultado e estudar experimentalmente o comportamento deles.

3.4.1 Metodologia

Seguindo a metodologia empregada no estudo do fluxo espectral, um algoritmo-protótipo será utilizado na análise de desempenho de diferentes funções de cálculo de periodicidade. O algoritmo empregado consiste, como antes, em realizar, nessa ordem: cálculo do fluxo espectral, estimação da periodicidade e seleção do pico de maior intensidade.

Para a obtenção do fluxo espectral, o sinal não foi sub-amostrado. A STFT utilizou uma janela e um salto com durações equivalentes a 40 ms e 5 ms, respectivamente. Foram utilizados 20 filtros para o mapeamento do espectro para a escala Mel e não foi utilizado o logaritmo. Estes parâmetros foram escolhidos observando-se os resultados encontrados na seção anterior. Como os testes anteriores não indicaram se vale a pena somar ou não as diferentes raias do sinal, escolheu-se realizar o teste com os fluxos espectrais calculados para cada raia e também para a sua soma. No caso de múltiplos fluxos espectrais, foi utilizado o expediente de calcular a periodicidade para cada fluxo e posteriormente a periodicidade média (com cada função de periodicidade normalizada pelo seu valor máximo).

Como função de periodicidade, foi escolhido cada um dos três métodos descritos na Seção 2.4: a autocorrelação, o módulo da DFT e o produto do módulo da DFT pela autocorrelação. Para os três métodos, a periodicidade foi calculada entre 40 e 250 BPM. No caso do produto, foi utilizada a solução que mapeia a autocorrelação na frequência através de uma interpolação linear, e foi calculada uma DFT com comprimento igual a 4 vezes o do atributo.

Foram calculadas, então, a Acurácia 1 e a Acurácia 2 para as 6 combinações possíveis: os três métodos de estimação de periodicidade, com e sem a soma dos atributos, para os bancos *Ballroom*, *Excerpts* e *Conjunto*. Na próxima seção, serão apresentados os resultados encontrados.

3.4.2 Resultados

Os resultados para os 6 diferentes métodos avaliados podem ser vistos na Tabela 3.8. Como se pode observar, exceto no *Excerpts* avaliado pela Acurácia 1, em que a Autocorrelação se mostrou superior, o produto entre a autocorrelação e o módulo da DFT produziu melhores resultados, independentemente da escolha feita em relação à soma dos atributos. Em particular, o resultado obtido para a Acurácia 2 usando o produto foi extremamente elevado, sendo os andamentos de aproximadamente 90 % dos sinais corretamente estimados segundo esta figura de mérito. No entanto, mesmo os melhores resultados para a Acurácia 1 ainda são insatisfatórios.

Tabela 3.8: Resultados para o teste variando a função de periodicidade. Todos os valores em %.

	Acurácia 1					
	Excerpts		*Ballroom*		*Conjunto*	
Método \ S	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
Autocorrelação	27	30	35	29	32	29
DFT	14	16	42	41	31	31
Produto	24	27	46	42	36	36

	Acurácia 2					
	Excerpts		*Ballroom*		*Conjunto*	
Método \ S	Verdadeiro	Falso	Verdadeiro	Falso	Verdadeiro	Falso
Autocorrelação	72	71	83	84	78	79
DFT	69	69	89	89	81	81
Produto	83	84	94	94	89	90

Para compreender melhor os erros cometidos por cada um dos métodos sob

estudo, foi calculado para cada sinal i o seguinte erro

$$\mathrm{err}_i^{\log} = \log_2 \left(\frac{\gamma_{f,i}^{\dagger}}{\gamma_{f,i}^{\mathrm{ref}}} \right), \tag{3.1}$$

onde $\gamma_{f,i}^{\dagger}$ e γ_i^{ref} são o andamento estimado e o anotado para o sinal i, respectivamente. Esse erro deixa claro quais múltiplos e submúltiplos do andamento de referência estão sendo selecionados no lugar do valor desejado. Por exemplo, quando $\mathrm{err}^{\log} = 0$, o andamento foi corretamente estimado; já quando $\mathrm{err}^{\log} = 1$ ou $\mathrm{err}^{\log} = -1$, o andamento estimado foi o dobro ou a metade do andamento anotado, respectivamente.

As Figuras 3.2, 3.3 e 3.4 exibem o histograma desta figura de mérito para os sinais de ambas as bases de dados quando são utilizados a autocorrelação, o módulo da DFT e seu produto, respectivamente. Serão analisados apenas os resultados quando a soma é realizada; os resultados quando os atributos não foram somados são similares. Como pode ser observado, o histograma relativo à autocorrelação possui 3 modas bem definidas: uma no andamento desejado, uma em -1 e outra em 1, indicando que os andamentos de uma grande quantidade dos sinais analisados foram confundidos com o andamento cujo valor é a metade ou o dobro do valor anotado. Como é sabido que a autocorrelação não provoca picos espúrios em múltiplos do andamento (apenas submúltiplos), o pico em 1 deve ter sido causado por características inerentes aos sinais (este fato vai ficar mais claro na análise realizada na próxima seção). A confusão para metade do valor de referência já não acontece para a DFT, já que ela só provoca picos espúrios nos múltiplos do andamento do sinal, mas a confusão com o dobro do andamento é consideravelmente maior que na autocorrelação. O produto, por sua vez, melhora o resultado mas ainda possui uma moda proeminente em 1, indicando mais sinais do que a observada no mesmo valor para a autocorrelação. Aparentemente, o uso do produto reduz apenas os erros diferentes de 1.

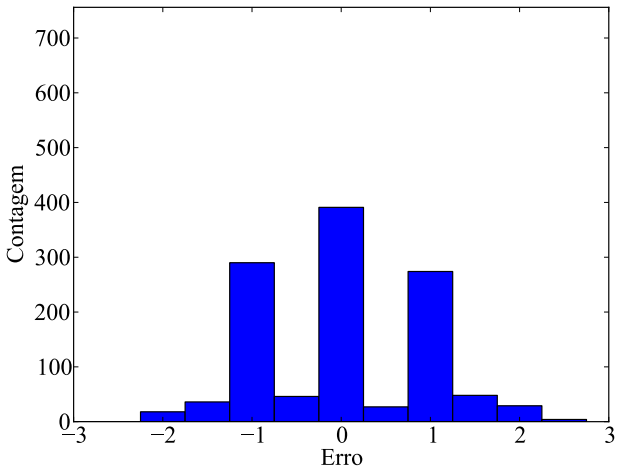


Figura 3.2: Histograma do erro quando a autocorrelação é utilizada.

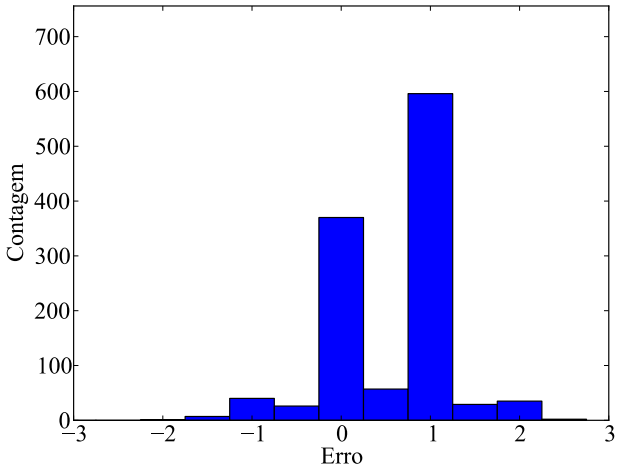


Figura 3.3: Histograma do erro quando o módulo da DFT é utilizado.

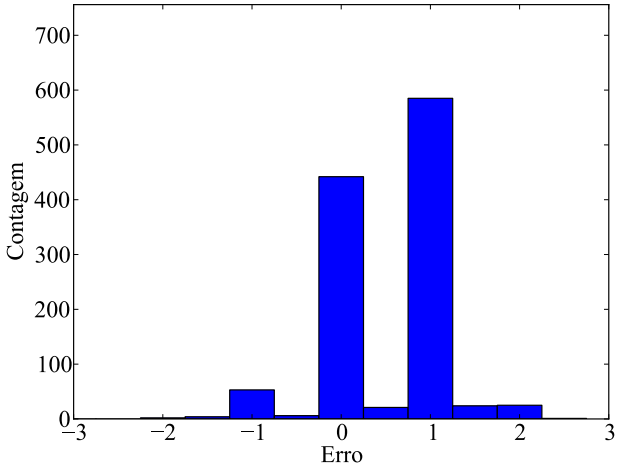


Figura 3.4: Histograma do erro quando o produto autocorrelação e módulo da DFT é utilizado.

3.4.3 Discussão

Nesta seção, três métodos para o cálculo da função de periodicidade foram comparados. Foi observado que o resultado quando se utiliza o produto entre a autocorrelação e o módulo da DFT é quase sempre superior ao obtido quando é utilizado uma dessas funções individualmente. Também foi observada a distribuição dos erros (destacando-se os mais comuns) cometidos por cada método. A conclusão obtida desta análise é que mesmo o produto confunde o andamento da maioria dos sinais com o dobro do seu valor.

Deve-se destacar que o resultado obtido para Acurácia 2 foi bastante elevado quando o produto é utilizado, sendo corretamente estimados cerca de 90 % dos sinais analisados. Isso demonstra que estimar algum múltiplo ou submúltiplo do andamento relacionado ao andamento do sinal é uma tarefa consideravelmente mais fácil que acertar exatamente o andamento, como era esperado. Nesse último caso, a taxa de acerto foi reduzida para 31 %. Na próxima seção, será investigado o uso de informações cognitivas para se tentar melhorar a Acurácia 1. Outro fator importante que ainda não está sendo levado em consideração é a estrutura rítmica dos sinais. Ao se escolher o dobro do andamento anotado, pode-se estar selecionando o andamento do nível hierárquico errado. Para corrigir esse problema, deverão ser consideradas informações sobre a estrutura rítmica de sinais musicais. Será apresentado no próximo capítulo um algoritmo que utiliza essa informação.

3.5 Uso de Informação Cognitiva

Nesta seção, será avaliado o impacto do uso da curva de ponderação descrita na Seção 2.5.2. É de particular interesse o seu efeito sobre a Acurácia 1, tendo em vista que, ao se incluir informação cognitiva, espera-se que o andamento de referência seja detectado um número maior de vezes. A seguir serão descritos a metodologia empregada e os resultados encontrados.

3.5.1 Metodologia

Da mesma forma que nas avaliações anteriores, será utilizado um algoritmo-protótipo em que é variado apenas o uso da função de ponderação. O algoritmo escolhido, com base nos resultados anteriores, consiste no cálculo do fluxo espectral seguido da obtenção da sua periodicidade através do produto entre a autocorrelação e o módulo da DFT. Em seguida, a periodicidade pode ser multiplicada pela curva descrita pela equação (2.7) ou não. Por fim, é escolhido o andamento que possui o maior pico na função de periodicidade. O fluxo espectral foi obtido utilizando-se os mesmos parâmetros utilizados na Seção 3.4. Neste teste, contudo, optou-se por

Tabela 3.9: Resultados quando a função de periodicidade é ponderada. Todos os valores em %.

	Acurácia 1		
	Excerpts	*Ballroom*	*Conjunto*
Sem ponderação	24	46	37
Com ponderação	33	64	51

	Acurácia 2		
	Excerpts	*Ballroom*	*Conjunto*
Sem ponderação	83	94	89
Com ponderação	84	91	88

sempre somar os atributos de cada raia da frequência antes do cálculo da função de similaridade.

3.5.2 Resultados e Discussão

Os resultados desta avaliação são exibidos na Tabela 3.9. Pode-se observar que o uso da informação cognitiva realmente melhorou a Acurácia 1, tendo-a elevado em torno de 14 pontos percentuais quando todos os sinais são considerados. A Acurácia 2, no entanto, é reduzida em torno de 1 ponto percentual quando a ponderação é utilizada. De forma geral, pode-se considerar que a função de ponderação tem um impacto positivo no desempenho, sendo a informação cognitiva relevante ao se estimar o andamento de cada sinal. Apesar disso, o desempenho em relação à Acurácia 1 ainda é insatisfatório. Na próxima seção, o algoritmo com ponderação será investigado mais profundamente, para se descobrir as suas principais deficiências.

3.6 Validação

Até este momento, neste capítulo, as diversas etapas de um método de estimação de tempo foram testadas individualmente. Nesta seção, o método final, escolhido a partir dos resultados mais promissores da seção anterior, será validado para um novo banco de sinais. O objetivo é verificar se o desempenho observado nas seções anteriores se mantém, para sinais desconhecidos. Após essa validação, o desempenho do método será avaliado para diferentes agrupamentos de sinais, de acordo com o andamento anotado e de acordo com o seu gênero. Por fim, os resultados obtidos pelo método serão comparados com resultados relatados na literatura.

O método empregado nas seções seguintes é o mesmo utilizado na Seção 3.5, quando a curva de ponderação é empregada.

3.6.1 Desempenho para o Banco *Hainsworth*

Até esta seção, foram utilizados apenas 2 dos 3 bancos de sinais descritos na Seção 3.1 para o desenvolvimento do método de estimação de andamento. Nesta seção, são apresentados resultados para a Acurácia 1 e Acurácia 2 quando é utilizado o banco *Hainsworh*.

Os resultados para o banco de sinais *Hainsworth* foram 66 % e 87 % para a Acurácia 1 e Acurácia 2, respectivamente. Estes resultados são consistentes com os resultados obtidos com o banco *Ballroom* avaliado pela Acurácia 1 (64 %) e os 3 bancos anteriores avaliados pela Acurácia 2 (84 %, 95 % e 88 %). Este fato sugere que o método de estimação de andamento escolhido não foi especializado para os bancos utilizados no seu desenvolvimento.

3.6.2 Resultado por Faixa de Andamento

Nesta seção, será verificado o desempenho do método (utilizando-se a Acurácia 1 e 2) para diferentes faixas de andamento. Para isso, os sinais dos três bancos foram combinados e agrupados de acordo com seu andamento. Em seguida, foram calculadas a Acurácia 1 e a Acurácia 2 para cada grupo de sinais.

Os resultados, para cada grupo de andamentos, são exibidos na Figura 3.5. A figura apresenta um gráfico de barras, cada uma mostrando o valor da Acurácia 1 (barra escura) ou da Acurácia 2 (barra clara) calculados para os sinais de cada grupo.

Como pode ser observado, a Acurácia 2 se comporta de forma aproximadamente uniforme para as diferentes faixas de andamento. Já a Acurácia 1 indica uma clara polarização para andamentos mais rápidos, sendo seu valor abaixo de 30 % para sinais com andamento abaixo de 120 BPM. Esta discrepância pode ser explicada pelo fato de a região de busca do andamento ficar entre 40 e 250 BPM. Se considerarmos que a função de periodicidade tende a errar para o dobro do andamento correto (como observado na Seção 3.4), só seria possível cometer esse erro para sinais com andamentos abaixo de 125 BPM. Logo, o salto no desempenho da Acurácia 1 observado pode se dever puramente à escolha da região permitida de andamentos. De qualquer forma, os resultados indicam uma deficiência do método: se houver essa possibilidade, o andamento estimado corresponde ao dobro do andamento anotado.

3.6.3 Resultado por Gênero

O desempenho do método também pode ser analisado para diferentes gêneros musicais. Dentre os bancos de sinais utilizados, dois possuem o gênero de cada sinal

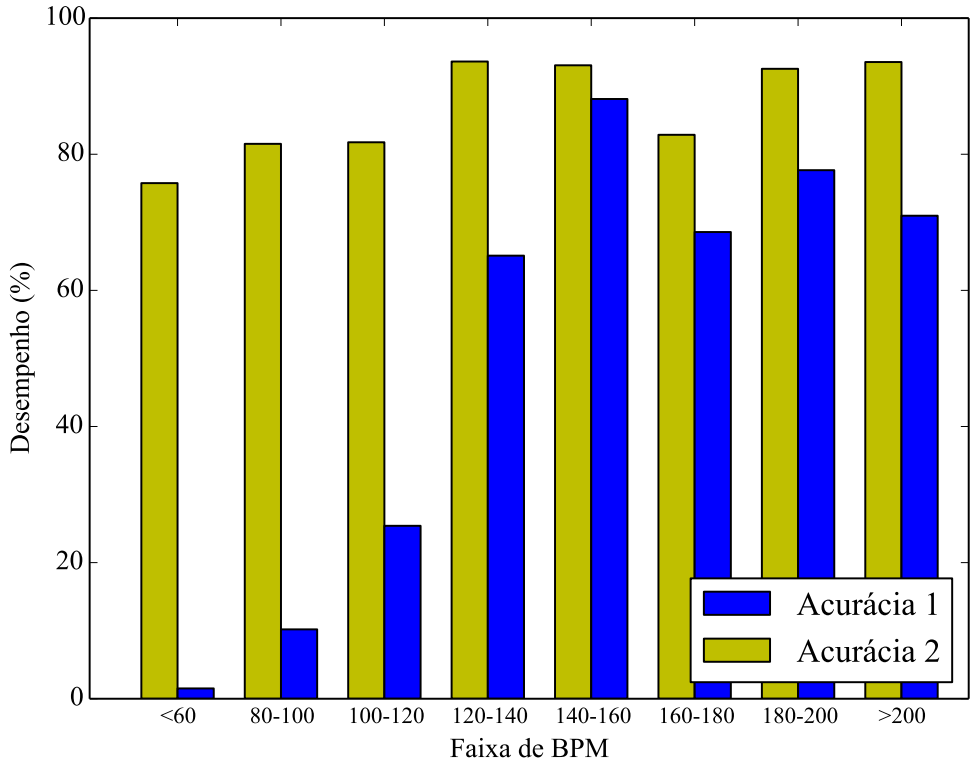


Figura 3.5: Acurácia 1 e Acurácia 2 para diferentes faixas de andamento.

anotado: o *Ballroom* e o *Hainsworth*. Para o cálculo das Acurácias 1 e 2 os sinais de cada banco foram agrupados em 6 gêneros. O mapeamento entre os gêneros anotados e os gêneros exibidos está descrito no Anexo A. Podem ser vistos na Figura 3.6 os valores obtidos para as Acurácias 1 e 2 para cada gênero mapeado. A figura exibe um gráfico de barras similar ao utilizado na seção anterior.

O pior resultado, para as duas figuras de mérito, foi observado para música clássica. Em geral, os sinais nesta categoria possuem uma série de características que dificultam a estimação do seu andamento:

- Ataques de notas mais sutis, com possível utilização de *glissandi*;
- Variação do andamento ao longo de uma mesma peça;
- Andamento, em geral, mais lento que os dos sinais de outros gêneros.

Se for considerada apenas a Acurácia 2, mais de 20 % dos sinais deste gênero tiveram o seu andamento incorretamente estimado. Esse tipo de erro indica que, possivelmente, o atributo utilizado não conseguiu capturar o andamento de uma quantidade significativa de sinais nesta categoria.

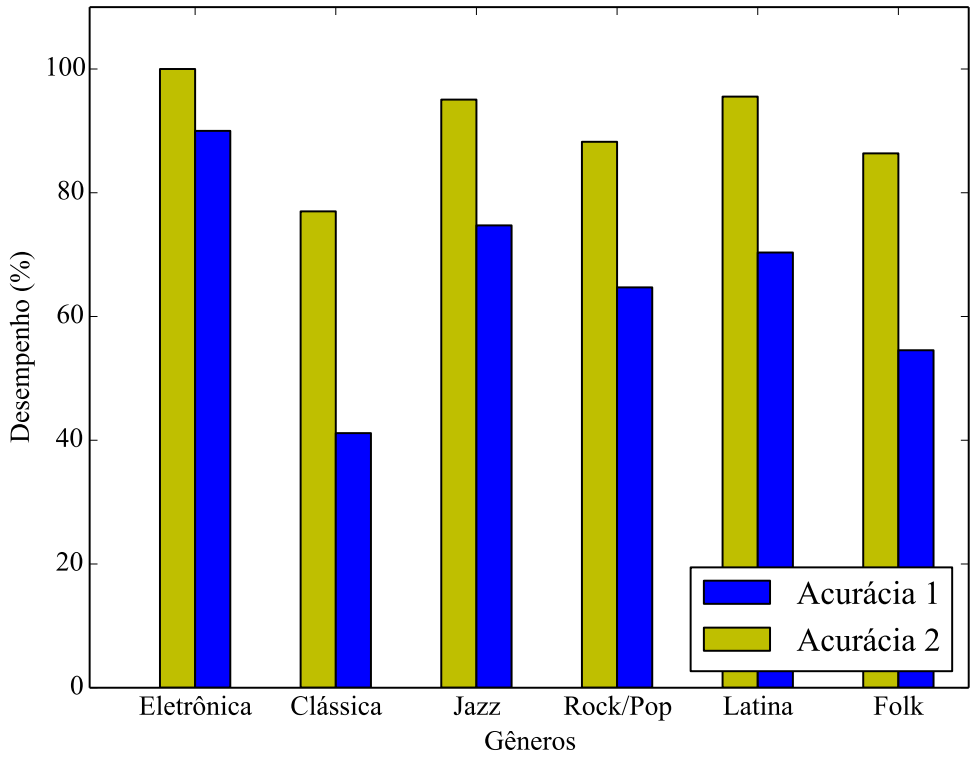


Figura 3.6: Acurácia 1 e Acurácia 2 para sinais de diferentes gêneros.

3.6.4 Comparação com Resultados da Literatura

As figuras de mérito utilizadas e os bancos de sinais *Ballroom* e *Excerpts* foram utilizados em diversos trabalhos na literatura. Com isso, é possível comparar o resultado relatado em diferentes publicações com o obtido pelo método desenvolvido.

Na Tabela 3.10 o desempenho do método derivado neste capítulo é comparado com o de outros 4 trabalhos publicados nos últimos 5 anos. Pode-se perceber que quase sempre os resultados obtidos já são comparáveis aos atingidos por outros métodos. Isto demonstra que o caminho seguido gerou um método de estimação adequado, sendo o seu desempenho pior que, porém próximo ao do estado da arte. Deve-se levar em consideração que, diferentemente dos trabalhos utilizados na comparação, o método desenvolvido não utiliza nenhuma informação da estrutura rítmica da peça.

Tabela 3.10: Comparação entre o método final deste capítulo e resultados reportados na literatura. Todos valores em %.

	Ballroom		*Excerpts*	
Método	Acurácia 1	Acurácia 2	Acurácia 1	Acurácia 2
Protótipo	64	94	33	84
[2]	65.2	93.1	49.5	83.7
[42]	48	83	30	73
[77]	69.2	94.1	50.4	91.8
[75]	57.3	80.8	51.8	69.1

3.7 Conclusão

Neste capítulo os algoritmos descritos no capítulo anterior foram utilizados para montar um método de estimação de andamento. O método foi construído progressivamente, escolhendo-se os algoritmos de cada etapa através de seus resultados experimentais. Ao final, o andamento é estimado utilizando-se o fluxo espectral como atributo, o produto da autocorrelação pelo módulo da DFT como função de periodicidade e a seleção do pico mais proeminente após a aplicação de uma curva de ponderação que utiliza informações cognitivas como a estimativa do andamento propriamente dita. O objetivo das avaliações foi gerar o método que servirá como ponto de partida para o desenvolvimento de novos algoritmos para extração de atributos, cálculo da função de periodicidade e seleção do andamento.

Considerando que o método gerado servirá como uma plataforma onde novos algoritmos serão testados, o seu desempenho foi avaliado sob diferentes aspectos. Observou-se que a maior parte das falhas do algoritmo acontecem para sinais com andamentos lentos (menores que 100 BPM). Deste resultado, pode-se concluir que a inclusão de informação rítmica e melhorias na função de periodicidade poderiam corrigir algumas das deficiências encontradas no método, já que há uma clara polarização para andamentos mais elevados nos resultados obtidos. O desempenho do método para sinais de diferentes gêneros musicais também foi analisado. Neste caso, observou-se um desempenho insatisfatório para músicas clássicas. Com isso, conclui-se que podem ser necessários atributos mais adequados para estes tipos de sinais (que muitas vezes não exibem mudanças abruptas de energia). No próximo capítulo, serão descritos algoritmos desenvolvidos que tentam remediar estes dois problemas.

Capítulo 4

Propostas de Algoritmos para Estimação de Tempo

Neste capítulo serão apresentados novos algoritmos para estimação de andamento com o intuito de melhorar o desempenho obtido com o método descrito no capítulo anterior. Em particular, será dado foco na melhoria da figura de mérito Acurácia 1, buscando-se melhorar o desempenho para sinais com andamento abaixo de 120 BPM. Serão propostas soluções para as três etapas da estimação: extração de atributo, cálculo da função de periodicidade e seleção do andamento.

Ao longo do capítulo, será utilizado para ilustrar o funcionamento do algoritmo o sinal de exemplo do *cowbell* que havia sido empregado ao longo do Capítulo 2 com o mesmo fim.

4.1 Separação Transitório/Permanente

Nesta seção será descrita uma técnica que procura melhorar a etapa de extração de atributos através de uma modificação sobre a STFT do sinal que enfatiza a sua parcela transitória. Observando-se o módulo da STFT (que foi definida na Seção 2.3.1) do sinal de exemplo na Figura 4.1, pode-se notar que a maior parte da energia se concentra em linhas horizontais ou verticais [81]. As linhas verticais representam a parcela transitória da uma batida no *cowbell*. Já as linhas horizontais representam a parcela em regime permanente do sinal, associada às ressonâncias consecutivas à percussão do instrumento.

Em [82] é descrito um algoritmo que procura decompor o módulo da STFT do sinal nessas duas partes. Para isto, foi proposta a aplicação de um filtro mediana-móvel nas linhas e nas colunas, obtendo-se, assim, representações que contêm apenas a parcela em regime permanente ou a parcela transitória, respectivamente, as quais foram então utilizadas para gerar dois novos sinais de áudio contendo predominan-

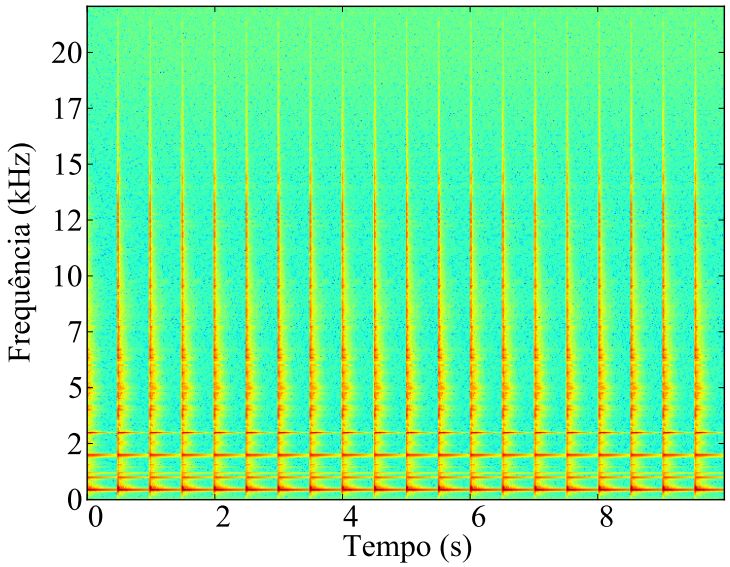


Figura 4.1: Módulo da STFT do sinal de exemplo.

temente uma dessas parcelas.

Neste trabalho, será utilizada a ideia de separar a parcela transitória da permanente do módulo da STFT para melhorar a etapa de extração de atributos usados na estimação do andamento. Para isso, é empregado um método denominado SSE (do inglês, *Stochastic Spectrum Estimation*), originalmente proposto em [83] e também descrito em [84, 85]. Esse método procura calcular a média de um sinal sem levar em consideração picos de grande intensidade. Esta característica é extremamente desejável, já que a estimação da parcela transitória consiste em estimar o "chão" do sinal sem a presença dos picos induzidos pela parcela permanente (e, analogamente, para a estimação da parcela ressonante). O filtro mediana móvel utilizado em [82] não é tão robusto a influência picos quanto o SSE [84]. O método SSE pode ser descrito, de forma geral, como a aplicação dos seguintes três passos [84] sobre um sinal positivo $S[n]$:

1. Filtrar $S[n]$ por um filtro média-móvel com três coeficientes, numa tentativa de remover valores nulos, obtendo-se $\overline{S}[n]$;

2. Calcular o sinal $R[n] = \frac{1}{\overline{S}[n]}$;

3. Obter uma versão suavizada de $R[n]$ através da aplicação de um filtro média-móvel de comprimento N^{sse}, resultando no sinal $\overline{R}[n]$;

4. Computar a curva suavizada $E[n] = \frac{1}{\overline{R}[n]}$.

Como se pode observar, o método procura remover picos da curva $S[n]$ através da suavização de $R[n]$, onde os picos de $S[n]$ se tornam valores pequenos e são

removidos. Este método pode ser utilizado para obtenção da parcela não ressonante da STFT de um sinal através de sua aplicação ao espectro obtido em cada quadro: $S[n]$ seriam os espectros em cada quadro e se desejaria remover os picos induzidos sobre a parcela "residual" deles. Já para a obtenção da parcela transitória, estes três passos seriam aplicados sobre os sinais variando ao longo dos quadros de cada raia, sendo os picos removidos associados a eventos transitórios.

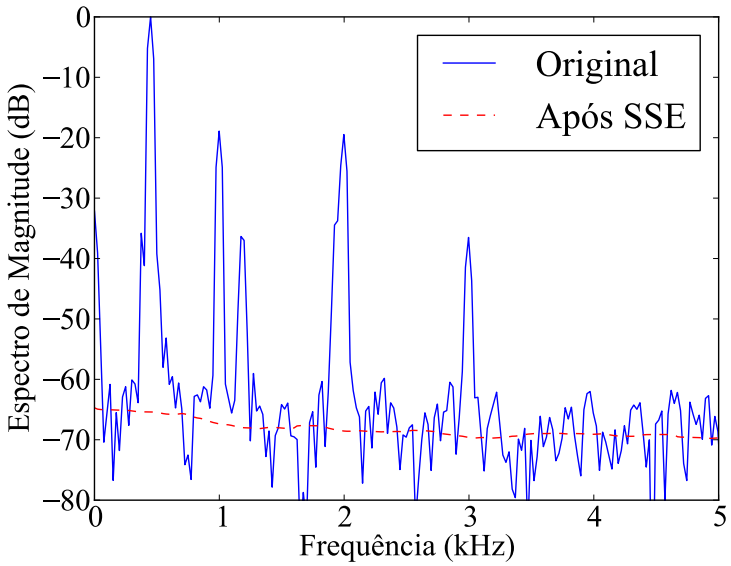
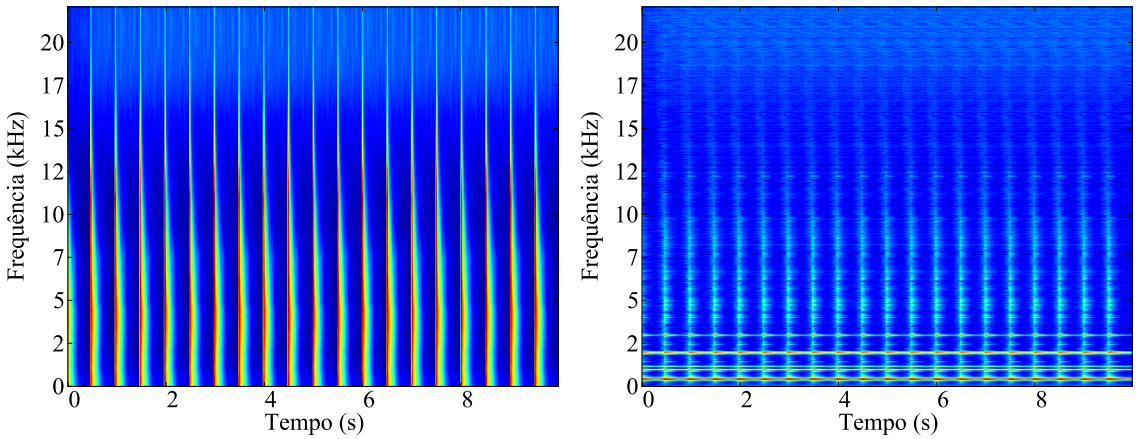


Figura 4.2: Curva obtida quando o SSE é aplicado numa das colunas do módulo da STFT.

A Figura 4.2 mostra o resultado da aplicação do SSE a uma das colunas do módulo da STFT do sinal de exemplo. Como se pode observar, a saída consiste de uma estimativa do "chão do sinal" que não sofre influências dos picos. O SSE possui apenas um parâmetro a ser escolhido, N^{sse}; nas figuras, adotou-se o valor de 31 amostras para este parâmetro. Deve-se ressaltar que não foi observada grande influência desse parâmetro sobre os resultados obtidos.

Quando o método SSE é aplicado ao longo das linhas e ao longo das colunas do módulo da STFT do sinal de exemplo, é obtido o resultado mostrado nas Figuras 4.3a e 4.3b, respectivamente. Pode-se ver que na Figura 4.3a restaram apenas as linhas verticais. Já na Figura 4.3b, as linhas horizontais estão mais evidentes, porém ainda é possível ver as linhas verticais.

Considerando o problema de estimação rítmica, é intuitivo assumir que a parcela transitória seja muito mais informativa do que a parcela permanente, pois carrega consigo a informação dos *onsets* das notas musicais. Por exemplo, se o fluxo espectral for calculado apenas a partir da parcela transitória, é obtido o sinal mostrado na Figura 4.5. Se comparado ao fluxo espectral encontrado sobre o módulo da STFT não modificada, exibido na Figura 4.4, a energia do atributo extraído a partir da



(a) Parcela transitória.

(b) Parcela permanente.

Figura 4.3: Resultado da aplicação do SSE ao longo das linhas e das colunas do módulo da STFT mostrada na Figura 4.1.

parcela transitória é mais concentrada em torno das batidas do cowbell, com os pulsos também sendo mais estreitos.

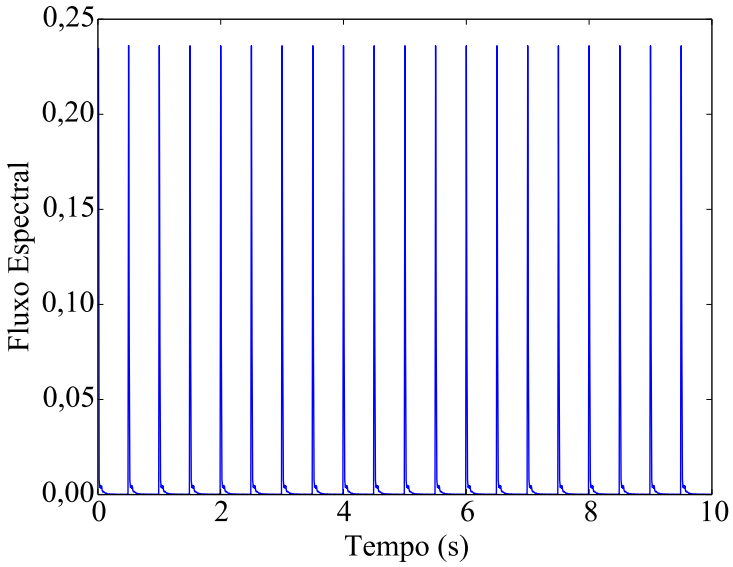


Figura 4.4: Fluxo espectral obtido para o sinal de exemplo. Reprodução da Figura 2.7.

4.2 Modificações sobre o Produto Autocorrelação × Módulo da DFT

O método sendo utilizado para obtenção da função de periodicidade consiste no produto entre a autocorrelação e o módulo da DFT. Conforme mencionado na Seção 2.4.3, que descreve este método, devido ao fato de a autocorrelação e a DFT

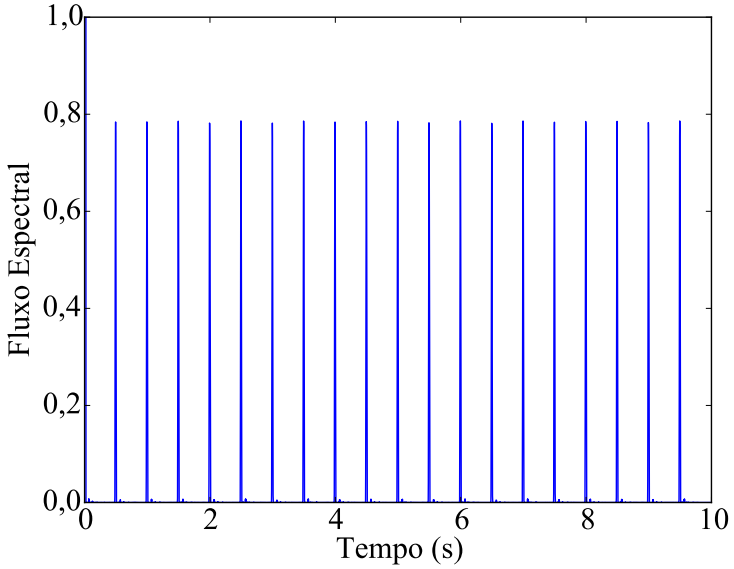


Figura 4.5: Fluxo espectral obtido a partir da parcela transitória do sinal de exemplo.

serem obtidas em domínios recíprocos (atraso e frequência), faz-se necessário um mapeamento do domínio de uma função no da outra. No método que vem sendo utilizado, os atrasos da autocorrelação são mapeados nos valores de frequência associados à DFT.

Nesta seção, é descrita uma solução que dispensa esse mapeamento. Para isso, no lugar do módulo da DFT, será empregado o módulo da Transformada de Fourier de Tempo Discreto (DTFT, do inglês *Discrete-Time Fourier Transform*), definida para uma frequência $\Omega \in [-\pi, \pi)$ e para um sinal $x[m]$ de comprimento M:

$$P^{\mathrm{DTFT}}(\Omega) = \left| \sum_{m=0}^{M-1} x[m] e^{-j\Omega n} \right|. \tag{4.1}$$

A DTFT é definida, então, em qualquer frequência, sendo possível avaliá-la apenas nos valores de frequência f que correspondem ao inverso dos atrasos contínuos τ associados aos atrasos discretos l_τ da função de autocorrelação. Assim, seria obtida a DTFT avaliada apenas nos atrasos:

$$\overline{P}^{\mathrm{DTFT}}[l_\tau] = \left| \sum_{m=0}^{M-1} x[m] e^{-j\frac{2\pi}{l_\tau} n} \right|, \tag{4.2}$$

e a função de periodicidade produto modificada ficaria:

$$\overline{P}^{\mathrm{prod}}[l_\tau] = P^{\mathrm{corr}}[l_\tau] \overline{P}^{\mathrm{DTFT}}[l_\tau]. \tag{4.3}$$

Para fins de ilustração, é mostrada na Figura 4.6 a função de periodicidade

$\overline{P}^{\mathrm{DTFT}}[l_\tau]$ análoga à vista na Figura 2.9, porém calculada apenas para os atrasos l_τ associados à função de autocorrelação exibida na Figura 2.8. Neste caso, a função de similaridade obtida pode ser observada na Figura 4.7. Novamente, a principal vantagem deste método é o fato de ele não necessitar de um mapeamento artificial sobre a autocorrelação que, em geral, acaba privilegiando regiões de andamento lento já que a função de periodicidade calculada pela autocorrelação possui mais pontos nessa região do que a obtida pela DFT. Uma discussão mais detalhada sobre estes mapeamentos, e suas desvantagens, pode ser encontrada em [57].

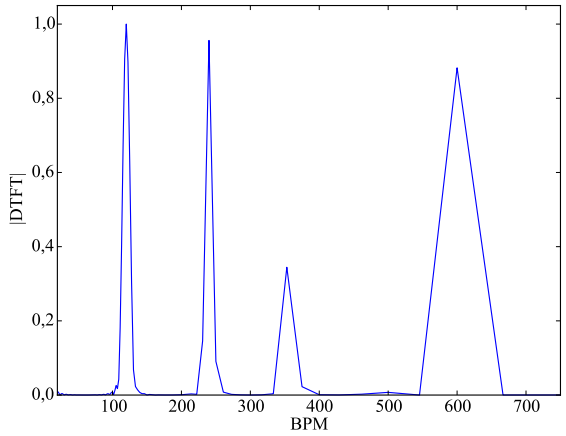


Figura 4.6: Módulo da DTFT calculada apenas para os atrasos correspondentes aos da Figura 2.8, para o fluxo espectral do sinal de exemplo.

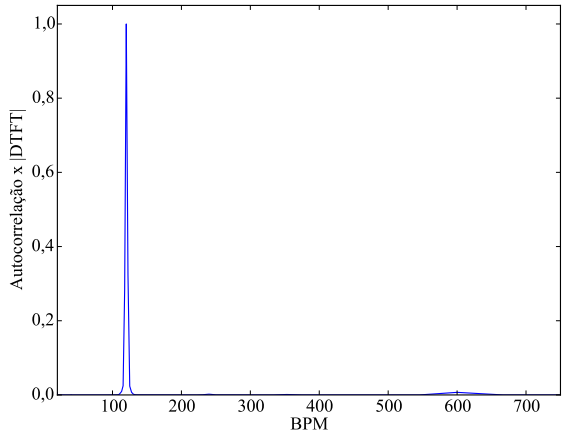


Figura 4.7: Função de periodicidade obtida como o produto do módulo da DTFT e da autocorrelação para o fluxo espectral do sinal de exemplo.

4.3 Estimação de Tempo usando Padrões Rítmicos

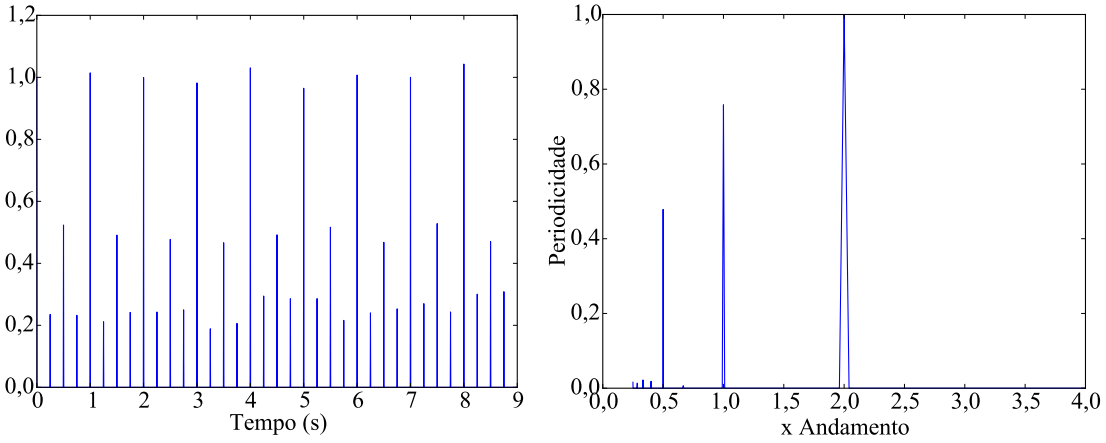
A seleção do melhor candidato ao andamento do sinal de áudio, até o momento, consistiu de uma simples detecção do pico mais proeminente da função de periodicidade após a sua multiplicação por uma curva de ponderação cognitiva. Nenhuma informação sobre a estrutura rítmica de sinais de música foi incorporada, não sendo considerada, portanto, a hierarquia entre os diferentes níveis rítmicos (ver Seção 1.1).

Em [2], foi proposta uma série de padrões rítmicos que podem ser utilizados para melhorar a etapa de estimação do andamento, os quais podem ser motivados por vetores de atributos que simulam sequências rítmicas idealizadas. As sequências utilizadas podem ser vistas na Figura 4.8, juntamente com a função de periodicidade estimada a partir de cada uma. Como se pode observar, cada sequência provoca o aparecimento de picos em posições diferentes da função de similaridade. A informação da posição desses picos é que define os chamados padrões rítmicos.

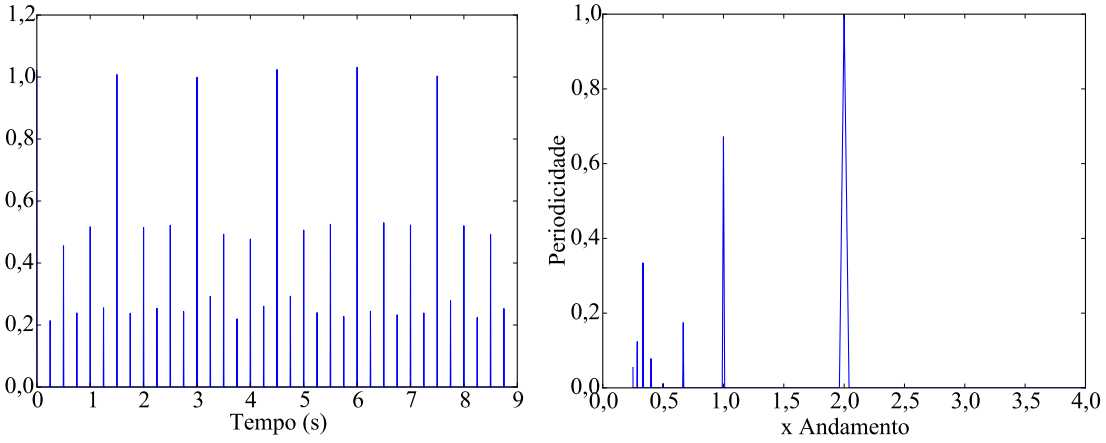
Os padrões rítmicos propostos em [2] consistem, para um andamento f, dos valores exibidos na Tabela 4.1. O objetivo dos padrões, então, é indicar quais múltiplos e submúltiplos do andamento de interesse deverão ser agregados ou ignorados, assumindo-se uma determinada estrutura rítmica para o sinal. Valores negativos são utilizados para punir num determinado padrão rítmico periodicidades que indiciam outro padrão rítmico. Deve-se notar também que é igual a zero o peso efetivo para periodicidades que não são múltiplas ou submúltiplas do andamento. Deve-se notar que essa escolha de valores (-1, 0, 1) foi realizada em [2] através da observação de quais múltiplos e sub-múltiplos deveriam ser considerados numa função de periodicidade. Métodos mais sofisticados, como de aprendizado estatístico, poderiam ser utilizados para refinar essa proposta inicial.

Tabela 4.1: Pesos $\alpha_{r,v}$ associados aos padrões rítmicos propostos em [2] (Eq. (4.4)).

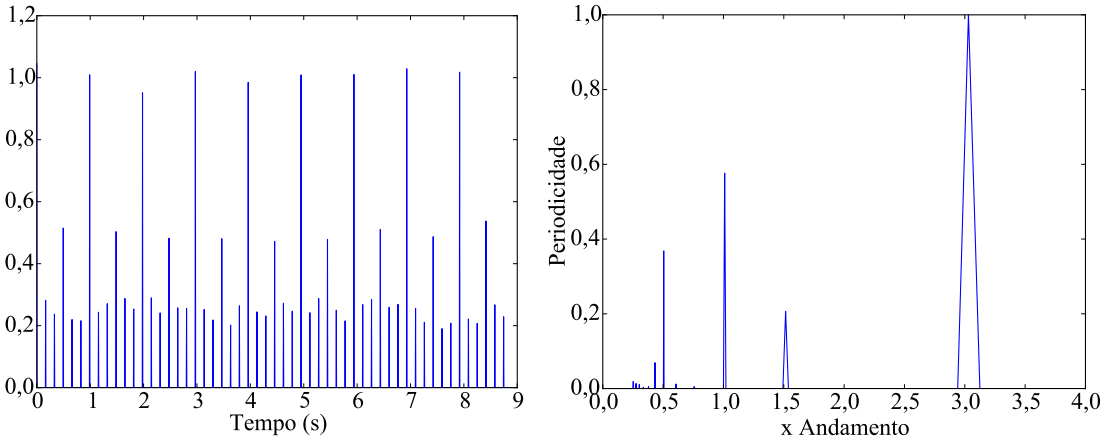
v	Estrutura	$\frac{f}{3}$	$\frac{f}{2}$	f	$1{,}5f$	$2f$	$3f$
1	Duplo/Simples	-1	1	1	-1	1	-1
2	Triplo/Simples	-1	1	1	-1	-1	1
3	Duplo/Composto	-1	1	1	-1	1	1



(a) Duplo/Simples.



(b) Triplo/Simples.



(c) Duplo/Composto.

Figura 4.8: Sequências idealizadas e suas periodicidades para os três padrões rítmicos utilizados neste trabalho. As três sequências estão associadas a um andamento percebido com valor em 120 BPM. Para facilitar a visualização, o eixo x da função de periodicidade foi normalizado pelo valor do andamento dos sinais; assim, o valor 2,0 indica um pico no dobro do andamento do sinal (nesse caso, 240 BPM).

Os padrões rítmicos podem ser utilizados de diversas formas para melhorar a estimação do andamento. Neste trabalho, eles serão aplicados da mesma forma que foi feito em [2], sendo a principal diferença entre a abordagem apresentada aqui e a do artigo a forma como o andamento é selecionado. Na referência, a periodicidade é calculada para trechos de 6 s e para diferentes padrões. Em seguida, um modelo probabilístico é utilizado para se obter uma sequência de andamentos e padrões mais prováveis ao longo do tempo. Neste trabalho, é obtido apenas o valor que melhor aproxima o andamento de toda a peça, logo são propostas novas formas de se escolher o andamento e padrão mais adequado. Além disso, em [2] não é utilizada a curva de ponderação da equação (2.7).

Os padrões rítmicos são utilizados para gerar uma função de periodicidade modificada definida, para um padrão rítmico v e uma função de periodicidade $P[\gamma_f]$, através de

$$P_v^{\text{mod}}[\gamma_f] = \sum_{r=1}^{5} \alpha_{r,v} P\left[\overline{\gamma_f \beta_r}\right], \tag{4.4}$$

onde β_r é o r-ésimo elemento de $\beta = \left[\frac{1}{3}\ \frac{1}{2}\ 1\ 1{,}5\ 2\ 3\right]$, $\overline{\gamma_f \beta_r}$ é o valor $\gamma_f \beta_r$ arredondado para o inteiro mais próximo e $\alpha_{r,v}$ é o peso associado ao v-ésimo padrão rítmico (ver Tabela 4.1). Com isso, a função de periodicidade modificada por um padrão rítmico num determinado andamento não considera apenas a sua periodicidade, mas também a de seus possíveis múltiplos e submúltiplos, além de descontar a periodicidade de múltiplos e submúltiplos que não podem ocorrer no padrão em questão.

Com base nessas funções de periodicidade, o seguinte algoritmo é utilizado para estimar o andamento:

1. Calcular as funções de periodicidade modificadas P_1^{mod}, P_2^{mod} e P_3^{mod};
2. Ponderar cada uma das funções de periodicidade pela equação (2.7);
3. Selecionar como andamento o pico mais proeminente entre os das funções P_1, P_2 e P_3.

Para ilustrar o funcionamento do método, será utilizado um trecho de um sinal contendo o registro de uma música popular que exibe um andamento bem definido com valor 129 BPM. A Figura 4.9 mostra o fluxo espectral deste sinal, onde se pode perceber claramente um padrão rítmico similar ao "Duplo/Simples" da Figura 4.8a. A Figura 4.10 exibe a função de periodicidade obtida, com picos indesejados aparecendo em múltiplos e submúltiplos do andamento do sinal. Já a Figura 4.11 exibe as funções de periodicidade modificadas por cada padrão rítmico antes e após a ponderação pela função cognitiva. Observando as diferentes funções de periodicidade, é possível notar claramente que os picos com maior intensidade ocorrem para o padrão

rítmico "duplo/simples", como esperado. Além disso, nota-se que utilizando apenas o padrão rítmico, dois valores se tornam mais proeminentes: um no andamento correto (129 BPM) e outro na metade de seu valor. Após a aplicação da função de ponderação cognitiva, no entanto, o pico no andamento desejado fica mais evidente e o seu valor correspondente em BPM deve ser selecionado como andamento do sinal pelo algoritmo proposto.

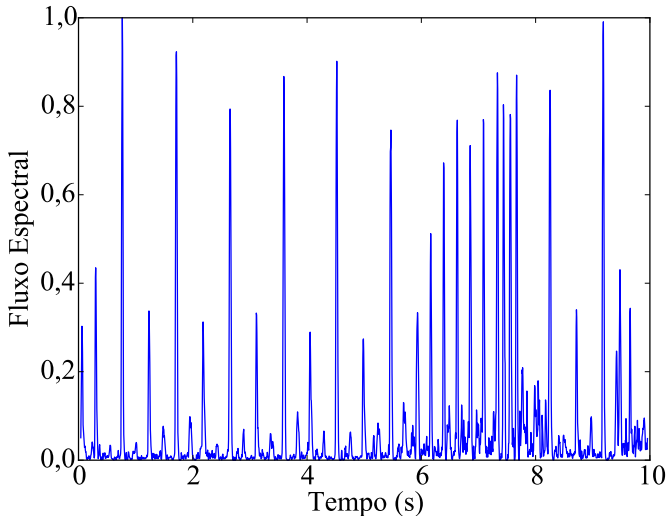


Figura 4.9: Fluxo spectral obtido do sinal usado para ilustrar o algoritmo de seleção de andamento.

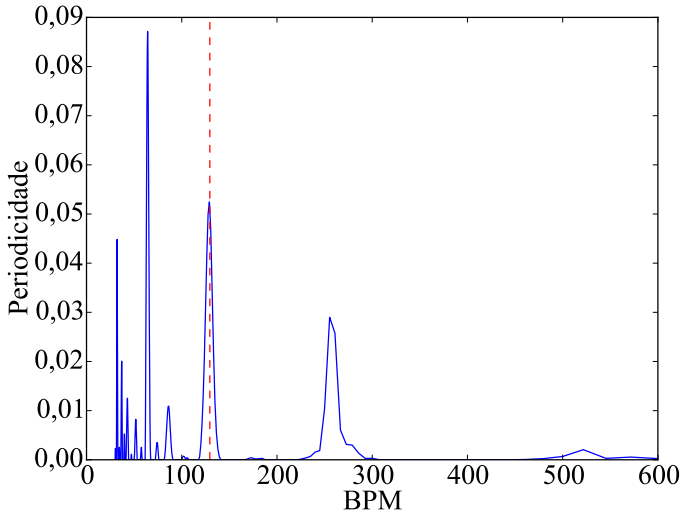
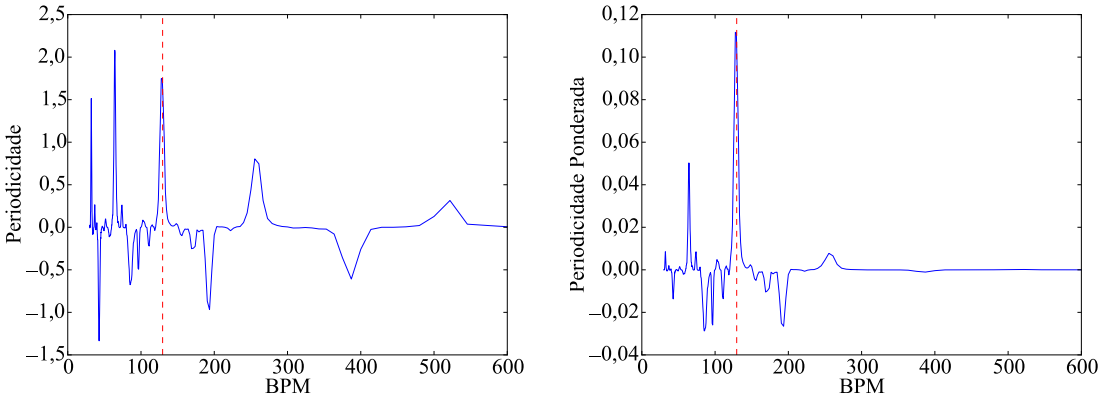
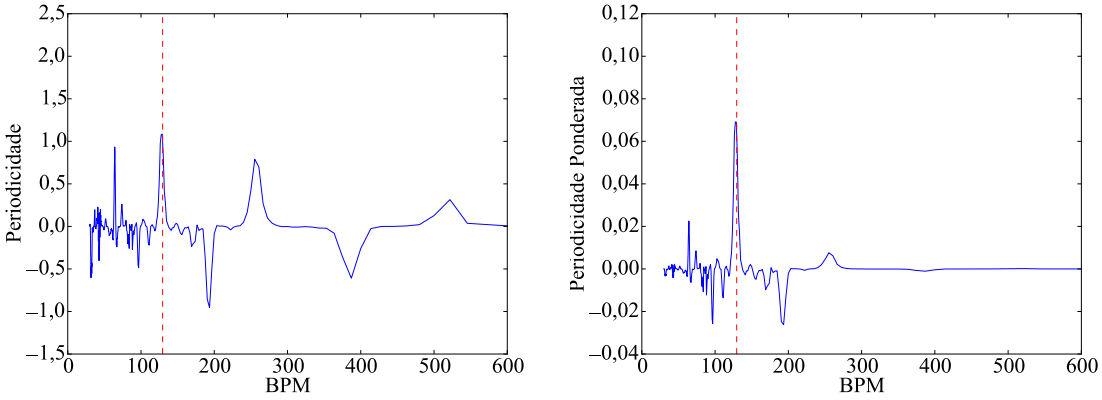


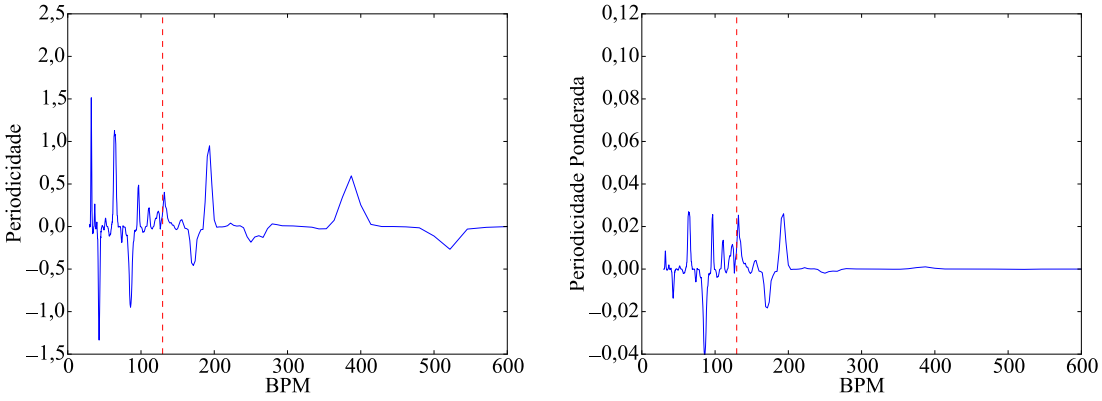
Figura 4.10: Periodicidade obtida a partir do fluxo spectral da Figura 4.9. A linha tracejada vertical denota o andamento do sinal.



(a) Duplo/Simples: esquerda (periodicidade modificada), direita (após ponderação).



(b) Triplo/Simples: esquerda (periodicidade modificada), direita (após ponderação).



(c) Duplo/Composto: esquerda (periodicidade modificada), direita (após ponderação).

Figura 4.11: Periodicidades modificadas por cada padrão rítmico antes e após a aplicação da função de ponderação cognitiva. A linha tracejada vertical denota o andamento do sinal.

4.4 Avaliação de Desempenho

Nesta seção, serão descritos os resultados obtidos quando são utilizados os algoritmos apresentados nas seções anteriores. A seguir, serão descritos o método utilizado

juntamente com os parâmetros selecionados, os bancos de sinais e as figuras de mérito empregados na avaliação. Para a escolha dos parâmetros dos métodos, serão utilizados os resultados obtidos no capítulo anterior.

4.4.1 Método Utilizado

O método utilizado consiste nas seguintes etapas:

1. Cálculo da STFT do sinal;
2. Separação das partes transitória/permanente do módulo da STFT utilizando o SSE;
3. Obtenção do fluxo espectral a partir apenas da parcela transitória da STFT;
4. Cálculo da periodicidade utilizando o produto entre a autocorrelação e módulo da DTFT;
5. Estimação do andamento usando padrões rítmicos.

A STFT é obtida utilizando-se uma janela de 40 ms de duração e com um salto de 5 ms que foram escolhidos por demonstrarem um bom desempenho em testes preliminares com os algoritmos propostos. O fluxo espectral mapeia a parcela transitória da STFT na escala Mel utilizando 20 filtros, não aplica o logaritmo e realiza a soma ao longo das raias. A função de periodicidade é calculada para atrasos entre 0,25 ms e 1,5 s. O SSE utiliza um filtro média-móvel com 31 amostras de comprimento.

4.4.2 Metodologia

O método descrito na seção anterior foi utilizado para estimar o andamento dos sinais dos três bancos descritos na Seção 3.1. Em seguida, foram obtidas a Acurácia 1 e a Acurácia 2 (ver Seção 3.2) para cada banco de sinais. Também foram calculadas as Acurácias 1 e 2 para os sinais agrupados em faixas de diferentes andamento e agrupados segundo o seu gênero anotado (quando disponível).

4.4.3 Resultados

Podem ser vistos na Tabela 4.2 os resultados obtidos pelo método descrito neste capítulo, pelo método protótipo obtido ao final do capítulo anterior e também resultados reportados na literatura. Como pode ser observado, o resultado obtido pelo método modificado melhorou significativamente o resultado da Acurácia 1 para todos os bancos de sinais quando comparado ao algoritmo protótipo. Em relação à Acurácia 2, foi observada queda de 1 ponto percentual no desempenho para os bancos

Ballroom e *Hainsworth*. Deve-se ressaltar o resultado para Acurácia 1 do *Excerpts*, que passou de 33 % para 52 %, igual ao melhor dos outros resultados reportados.

Tabela 4.2: Comparação entre o método descrito neste capítulo e resultados reportados na literatura e no capítulo anterior. Todos os valores em %.

	Ballroom		*Excerpts*		*Hainsworth*	
Método	Acur. 1	Acur. 2	Acur. 1	Acur. 2	Acur. 1	Acur. 2
Proposto	67	93	52	87	73	86
Protótipo	64	94	33	84	66	87
[2]	65	93	50	84	–	–
[42]	48	83	30	73	–	–
[77]	69	94	50	92	–	–
[75]	57	81	52	69	–	–

A Figura 4.12 mostra o resultado para diferentes faixas de andamento. Pode-se perceber que o método descrito apenas estima corretamente andamentos entre 100 e 180 BPM. Como a maior parte dos sinais dos bancos sob análise está concentrada nesta faixa, o desempenho geral é satisfatório. Fica claro, no entanto, que é necessário investigar mais profundamente como melhorar o desempenho para os sinais fora desta região. De qualquer forma, o resultado para a Acurácia 2 é uniforme ao longo das faixas de andamento, indicando que pelo menos um andamento que é múltiplo ou submúltiplo do andamento correto costuma ser selecionado.

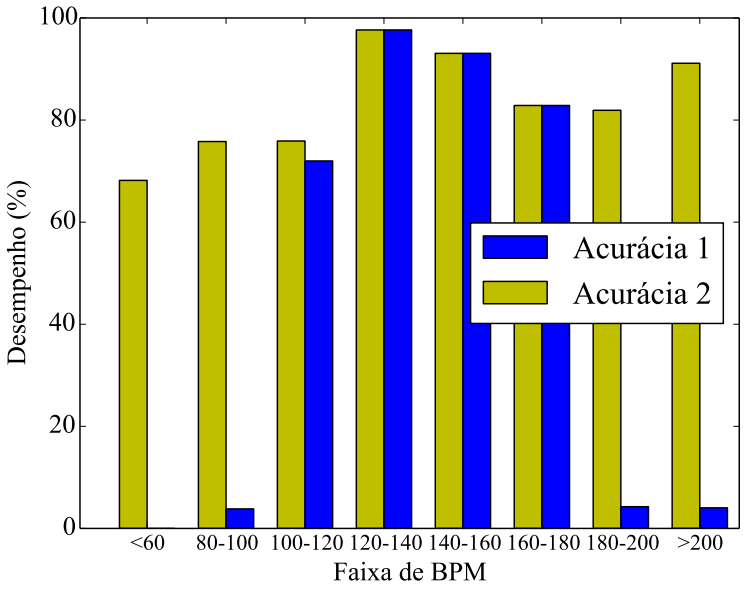


Figura 4.12: Acurácia 1 e Acurácia 2 para diferentes faixas de andamento para o método proposto.

Na Figura 4.13 pode ser vista a distribuição das Acurácias 1 e 2 de acordo com o gênero do sinal. O desempenho para sinais de música clássica melhorou consideravelmente em relação ao desempenho obtido no capítulo passado. O resultado para sinais de Jazz, no entanto, piorou bastante. Considerando que sinais de Jazz possuem uma estrutura rítmica que poderia ser associada ao "triplo/simples" ou ao "duplo/composto", pode ser necessário ajustar os padrões rítmicos de forma a melhorar o desempenho para este gênero.

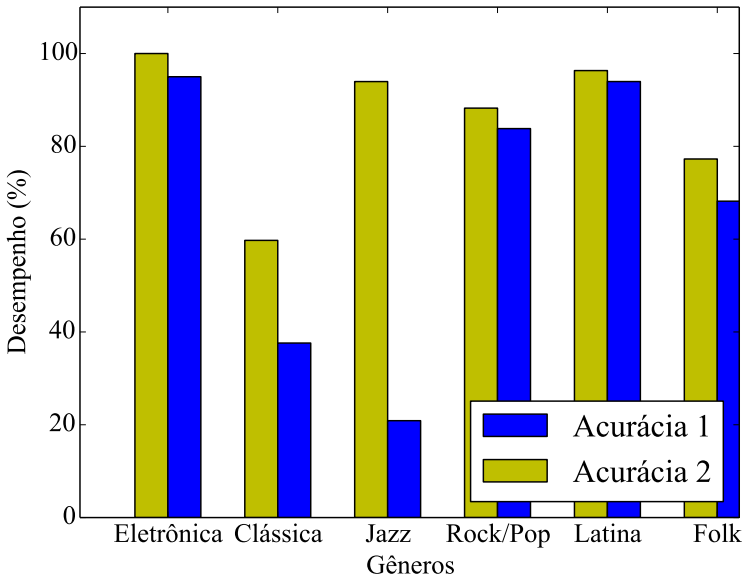


Figura 4.13: Acurácia 1 e Acurácia 2 para sinais de diferentes gêneros.

Deve-se ressaltar que os mesmos resultados foram obtidos quando se omitiu a etapa de separação transitória/permanente. Isso indica que, pelo menos para detecção de andamento, a separação transitório/permanente não traz grandes benefícios. Devido a isso, nos resultados apresentados nos capítulos seguintes, não foi utilizada a separação transitório/permanente.

4.5 Conclusão

Neste capítulo foram apresentadas novas soluções para estimação de andamento que demonstraram uma melhora no desempenho quando comparado aos resultados obtidos no capítulo anterior e com métodos encontrados na literatura. No entanto, pode-se perceber que são pequenos os ganhos sobre os resultados anteriores, indicando que alguma informação extra precisa ser utilizada para se dar um salto significativo no desempenho, principalmente no da Acurácia 1.

Nos próximos capítulos, serão apresentados métodos que procuram estimar

quando os níveis métricos foram percebidos. Estes métodos serão desenvolvidos inspirados nos resultados obtidos para o andamento e utilizarão os resultados como uma estimativa inicial do andamento. Além dos métodos, também será descrita uma nova anotação para um dos bancos de sinais utilizados. Esta marcação foi feita para todos os níveis métricos por apenas um músico profissional, garantindo a consistência dos dados anotados. Por fim, espera-se que ao se procurar modelar quando cada um dos níveis métricos ocorreu, o andamento possa ser reestimado em uma maior precisão, principalmente na Acurácia 1, já que as relações entre os diferentes níveis métricos serão modeladas diretamente.

Parte II

Análise Métrica

Capítulo 5

Modelos para Análise Rítmica Computacional

Neste capítulo, serão apresentados modelos para o rastreamento da métrica de um sinal de áudio. Para isto, os algoritmos propostos procurarão identificar, a partir de um atributo extraído do sinal de áudio, quais quadros estão associados a cada nível métrico—tatum, tactus ou compasso. Desta forma, é obtida a informação sobre a métrica de um sinal de áudio e como ela evolui ao longo do tempo.

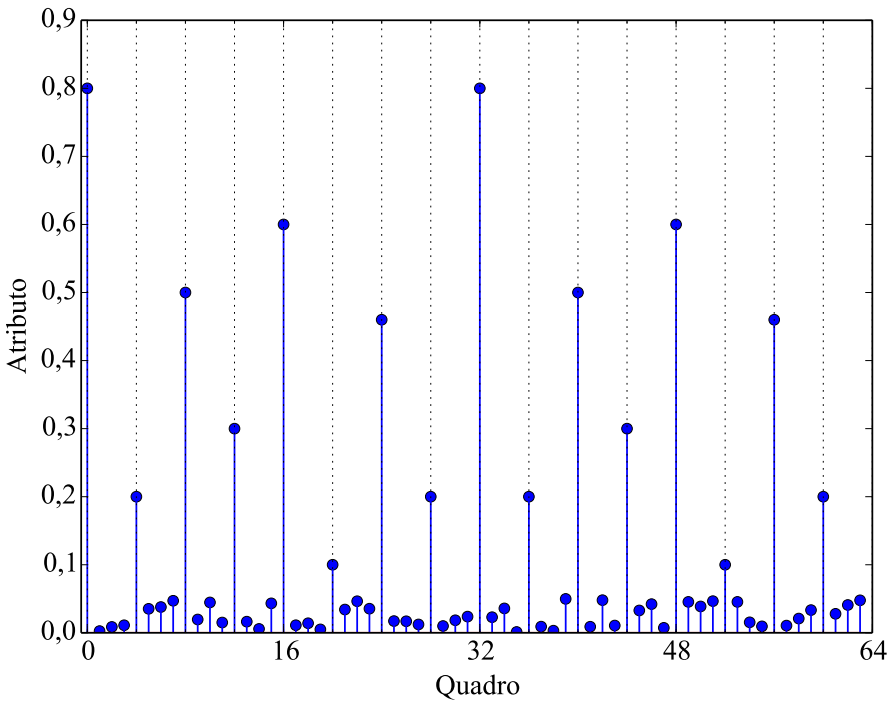


Figura 5.1: Exemplo de um atributo que exibe informação métrica. Neste caso, pode-se observar que o tactus possui um período discreto de 16 quadros, com o tatum ocorrendo a cada 4 quadros e o compasso a cada 32.

Para exemplificar a tarefa a ser realizada pelos algoritmos a serem apresentados, será utilizado o atributo visto na Figura 5.1, que foi gerado artificialmente. Observando-se o atributo, pode-se perceber que determinados valores se repetem com uma certa periodicidade, sendo uma observação com menor intensidade acontecendo a cada 4 quadros, uma com média intensidade acontecendo a cada 8 quadros e uma mais intensa ocorrendo a cada 32 quadros. Neste caso, considerando as divisões observadas, pode-se dizer que os valores de menor intensidade estão associados ao tatum, os de média ao tactus e os de alta ao compasso. No caso do exemplo, os atributos estariam associados a uma estrutura hierárquica como a definida na Figura 5.2 (reproduzida do Capítulo 1).

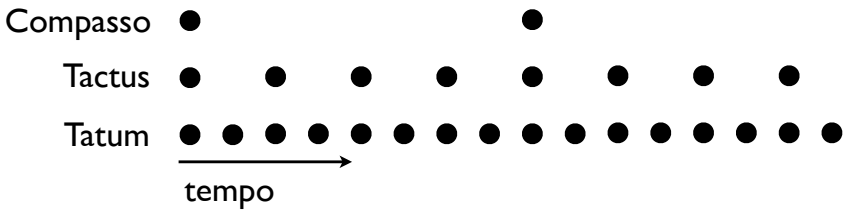


Figura 5.2: Estrutura rítmica associada ao Exemplo da Figura 5.1. Neste caso, a estrutura estaria associada a uma contagem do tipo "1 – e – 2 – e – 3 – e – 4 – e".

Observando novamente a Figura 5.1, pode-se definir a principal tarefa de um algoritmo de rastreamento métrico, então, como a identificação de quais atributos em quais quadros estão associados a um dos níveis métricos, e quais atributos não estão associados a nenhum nível métrico (não possuem informação rítmica). O principal desafio desta tarefa consiste na variação temporal das ocorrências dos níveis métricos, que, apesar de apresentarem uma periodicidade induzida, não precisam ocorrer em intervalos idênticos no tempo. Outro desafio é a divisão entre os eventos poder variar numa mesma música, por exemplo, com o período do tatum se tornando um terço do período do tactus. Em casos como esse, o período do tactus (que define a nossa percepção do andamento) é preservado. Em outros casos, o período do tactus também pode variar ao longo de uma mesma música. Por fim, algumas músicas exibem trechos de silêncio ou regiões onde não está presente uma estrutura métrica.

Na literatura, podem ser encontrados diversos algoritmos para rastreamento do tactus, porém poucos que tratam da análise dos demais níveis métricos. De forma geral, os algoritmos podem ser divididos entre os que utilizam uma anotação simbólica e os que utilizam o sinal de áudio. Neste trabalho, o foco será na extração de informação métrica a partir de sinais de áudio.

A seguir faz-se uma breve revisão dos principais trabalhos sobre rastreamento do tactus. Uma abordagem para esse problema consiste na formulação de um problema de programação dinâmica, onde procura-se encontrar a sequência de inícios de tactus

que minimiza uma função-custo associada ao período do tactus (que foi previamente estimado ou é conhecido). Tal abordagem foi proposta em [25] e versões modificadas deste algoritmo podem ser encontradas em [86–88]. Outros métodos procuram definir diversos agentes, cada um responsável por manter a hipótese de uma possível sequência de tactus. As hipóteses armazenadas em cada agente são atualizadas a cada quadro e, dependendo de um conjunto de heurísticas, um agente pode dar origem a novos agentes ou pode ser extinguido. Ao final, uma figura de mérito é utilizada para encontrar o agente com a hipótese de tactus mais promissora. Essa abordagem foi proposta inicialmente em [89] e diversas modificações foram propostas em [90, 91]. Em [27, 92], modelos ocultos de Markov (HMM, do inglês *Hidden Markov Model*) são utilizados para realizar o rastreamento do tactus. Diferentes modelos são propostos mas, de forma geral, todos os trabalhos procuram encontrar a sequência de tactus mais provável utilizando um modelo em que a probabilidade de se iniciar um tactus num quadro é definida pela probabilidade de se iniciar um tactus no quadro anterior. Será dada na próxima seção uma explicação detalhada sobre modelos ocultos de Markov. Além dos algoritmos já citados, cabe também mencionar as soluções propostas em [23, 24, 70], onde, uma vez tendo-se o conhecimento do período do tactus, procura-se gerar um sequência de tactus respeitando esse período e, em seguida, ajustar temporalmente cada ocorrência de acordo com o sinal observado.

A literatura sobre rastreamento da métrica é menos extensa que a de rastreamento do tactus. Uma primeira tentativa de análise métrica pode ser encontrada em [93], onde é proposto um HMM para busca de padrões rítmicos em sinais de áudio. Neste trabalho, o padrão é definido em função de uma variável aleatória arbitrária e permite a identificação da ocorrência do tactus e do compasso, mas não do tatum. Este modelo foi refinado através do uso de padrões aprendidos a partir de um banco de sinais em [94, 95]. Outra família de algoritmos para rastreamento de padrões rítmicos pode ser encontrada nos trabalhos [96–99], onde a principal modificação ocorre na forma como os padrões são definidos, utilizando aprendizagem estatística, e na formulação de um algoritmo de inferência específico para o modelo proposto. Uma outra versão modificada do modelo de rastreamento da métrica através de padrão rítmico ainda foi proposta em [100]. Em [101], é descrito um algoritmo de rastreamento de compassos que utiliza informação do gênero da música sob análise. Já em [102], é descrito um algoritmo próprio para analisar música indiana que procura utilizar o padrão rítmico cíclico deste gênero durante a análise. Em [50] também é utilizado um modelo oculto de Markov para análise métrica; porém, o modelo proposto apenas estima os períodos de cada nível hierárquico, sendo o instante de ocorrência estimado heuristicamente. Serão feitas ao longo da apresentação dos modelos propostos comparações com os trabalhos citados.

Serão apresentadas nas próximas seções duas soluções que são potencialmente capazes de extrair, a partir de um sinal de áudio, toda a informação métrica sobre seu conteúdo musical. Os algoritmos propostos se baseiam em modelos ocultos de Markov, que foram escolhidos por fornecerem um bom compromisso entre flexibilidade na modelagem dos fenômenos de interesse e custo computacional, além de já terem sido empregados com sucesso em problemas de análise métrica. Nas soluções propostas, dois modelos serão apresentados: um que rastreia diretamente os três níveis métricos e outro que utiliza padrões rítmicos para a identificação da métrica. Também serão propostas simplificações sobre o primeiro modelo: quando apenas o tactus é rastreado e quando o rastreamento do tatum é feito isoladamente do rastreamento dos demais níveis métricos.

O restante desse capítulo será organizado da seguinte forma. Inicialmente, na Seção 5.1, será feita uma breve apresentação sobre modelos ocultos de Markov. Em seguida, é descrito na Seção 5.2 o modelo hierárquico, que modela diretamente os três níveis métricos. Nas Seções 5.3 e 5.4, são propostas duas simplificações do modelo hierárquico. É apresentado, na Seção 5.5, o modelo de rastreamento por padrão rítmico. Por fim, as conclusões do capítulo são apresentadas na Seção 5.6.

5.1 Modelos Ocultos de Markov

Nesta seção, será feita uma breve introdução a modelos ocultos de Markov com o objetivo de introduzir os seus principais conceitos e familiarizar o leitor com a notação matemática que será empregada no restante do capítulo. HMMs possuem uma vasta literatura, já tendo sido utilizados em diversas aplicações que variam de reconhecimento e síntese de fala [65] até sequenciamento de DNA [103]. Deve-se notar que não será feita uma apresentação detalhada de HMMs; tutoriais com esse fim podem ser encontrados em [104] e [34].

De forma geral, modelos ocultos de Markov são utilizados para modelar fenômenos cujas probabilidades de ocorrência variam ao longo do tempo. Sendo assim, pode-se classificá-los como modelos dinâmicos, que procuram capturar a variação de uma certa quantidade ao longo do tempo. Usualmente, cada modelo pode conter uma ou mais variáveis aleatórias (VAs) cujo valor mais provável será estimado a partir da observação de um conjunto de atributos. Nesse caso, o conjunto de atributos seria a variável observada, e as VAs a serem estimadas (não observadas) seriam as variáveis ocultas. As VAs ocultas seguem por hipótese um modelo de Markov de primeira ordem, isto é, a probabilidade de cada uma delas assumir um determinado valor num determinado ponto da sequência de dados observados depende apenas de seus valores no ponto imediatamente anterior da sequência.

Nas próximas seções, serão apresentadas as partes que compõem um modelo de

Markov e também algoritmos de inferência.

5.1.1 Modelagem

Nesta seção será descrita um HMM com apenas duas VAs ocultas; a extensão para um número maior de VAs pode ser feita sem grandes dificuldades.

Num HMM, são consideradas aleatórias tanto as variáveis a serem estimadas (ocultas) quanto as variáveis que serão observadas [105]. As variáveis a serem estimadas em cada quadro m precisam ser discretas e são descritas matematicamente como

$$\mathbf{x}_m^{\mathrm{I}} \in \mathcal{N}^{\mathrm{I}} = \left\{x_0^{\mathrm{I}}, x_1^{\mathrm{I}}, \ldots, x_{N^{\mathrm{I}}-1}^{\mathrm{I}}\right\} \tag{5.1}$$

$$\mathbf{x}_m^{\mathrm{II}} \in \mathcal{N}^{\mathrm{II}} = \left\{x_0^{\mathrm{II}}, x_1^{\mathrm{II}}, \ldots, x_{N^{\mathrm{II}}-1}^{\mathrm{II}}\right\}. \tag{5.2}$$

Note que as VAs estão sendo denotadas em negrito e o índice m indica o quadro ao qual elas estão associadas. Observe também que os conjuntos $\mathcal{N}^{\mathrm{I}}$ e $\mathcal{N}^{\mathrm{II}}$, que são, respectivamente, os conjuntos amostrais das VAs $\mathbf{x}_m^{\mathrm{I}}$ e $\mathbf{x}_m^{\mathrm{II}}$, não variam ao longo do tempo.

Uma forma simples de descrever os possíveis valores assumidos por cada VA em cada quadro é obtida se for feita a seguinte enumeração:

$$e_{m,0} \triangleq (\mathbf{x}_m^{\mathrm{I}} = x_0^{\mathrm{I}}, \mathbf{x}_m^{\mathrm{II}} = x_0^{\mathrm{II}}) \tag{5.3}$$

$$e_{m,1} \triangleq (\mathbf{x}_m^{\mathrm{I}} = x_1^{\mathrm{I}}, \mathbf{x}_m^{\mathrm{II}} = x_0^{\mathrm{II}}) \tag{5.4}$$

$$\ldots$$

$$e_{m,|\mathcal{N}|-1} \triangleq (\mathbf{x}_m^{\mathrm{I}} = x_{N^{\mathrm{I}}-1}^{\mathrm{I}}, \mathbf{x}_m^{\mathrm{II}} = x_{N^{\mathrm{II}}-1}^{\mathrm{II}}), \tag{5.5}$$

onde $e_{m,0}$ é o estado no quadro m associado ao primeiro par ordenado contido no conjunto $\mathcal{N} = \mathcal{N}^{\mathrm{I}} \times \mathcal{N}^{\mathrm{II}}$, $e_{m,1}$ está associado ao segundo par ordenado de $\mathcal{N}$ e assim por diante. Esta notação alternativa, que enumera todas as possíveis combinações de valores assumidos pelas VAs ocultas do HMM, simplificará a descrição do HMM e facilita a sua extensão para um número maior de VAs ocultas.

Quando comparada com a VA oculta, a variável que é observada possui menos restrições, podendo ser contínua. Considerando o caso mais geral, ela pode ser descrita como

$$\vec{\mathbf{y}}_m \in \mathbb{R}^{K\times 1}, \tag{5.6}$$

onde o símbolo $\vec{\cdot}$ denota um vetor, o índice m indica o quadro associado à VA e K é o comprimento do vetor.

Uma vez definidas as VAs ocultas e as observadas, é necessário descrever a densidade de probabilidade conjunta das VAs ocultas em cada quadro. É nesta etapa

que são feitas as principais hipóteses do modelo de Markov. Inicialmente, será feita a hipótese de que a densidade de probabilidade conjunta das variáveis num quadro depende apenas do quadro anterior (elas formam uma cadeia de Markov de ordem 1). Esta hipótese pode ser descrita matematicamente como

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{I}}, x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}, x_{m-2}^{\mathrm{I}}, x_{m-2}^{\mathrm{II}}, \ldots) = p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{I}},x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}). \quad (5.7)$$

Com isso, é considerado que a influência de todo o passado na probabilidade para uma VA no quadro m está resumida no que aconteceu no quadro imediatamente anterior. Esta é a hipótese que mais limita a aplicação do HMM a um determinado problema, mas também é a hipótese que permite a existência dos algoritmos de inferência computacionalmente eficientes que serão vistos a seguir. Dá-se o nome de modelo de transição à probabilidade $p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{I}},x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}})$, uma vez que ele representa o efeito da transição de um quadro ao outro sobre a densidade de probabilidade de cada VA oculta. Note que no caso das cadeias homogêneas, com as quais lidaremos, esta probabilidade não varia de quadro para quadro, sendo que o efeito de uma variável ter assumido um determinado valor sobre as probabilidades no quadro seguinte é sempre o mesmo.

A probabilidade de transição descreve como ocorre a evolução da probabilidade de cada VA ao longo do tempo, mas ainda é necessário definir qual é a densidade de probabilidade conjunta das VAs no primeiro quadro. Esta probabilidade será denominada a prior[1] do modelo, e é escrita como $p_{\mathbf{x}_0^{\mathrm{I}},\mathbf{x}_0^{\mathrm{II}}}(x_0^{\mathrm{I}},x_0^{\mathrm{II}})$ e define a chance de ocorrer um determinado estado no primeiro quadro, quando não há informação de quadros anteriores. Esta probabilidade pode ser escolhida com base em conhecimentos prévios, por exemplo, a frequência de ocorrência de um determinado valor em um conjunto de dados. Neste caso, como será visto, a sequência de valores estimados será polarizada de acordo com a probabilidade prior. Essa polarização poderá melhorar o resultado da inferência, se for uma boa escolha, mas também pode levar a resultados piores. Caso não se tenha nenhuma informação prévia ou seja preferível uma estimativa não polarizada, pode-se adotar uma $p_{\mathbf{x}_0^{\mathrm{I}},\mathbf{x}_0^{\mathrm{II}}}(x_0^{\mathrm{I}},x_0^{\mathrm{II}})$ uniforme.

Uma outra hipótese assumida no HMM é que é possível obter um modelo para a densidade de probabilidade da VA observada no quadro m que depende apenas da VA oculta nesse mesmo quadro. Matematicamente, deve ser possível obter a densidade $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|x_m^{\mathrm{I}}, x_m^{\mathrm{II}})$ e, idealmente, esta densidade deve ser capaz de explicar as variações na observação. De forma alternativa, para cada estado espera-se que a densidade da VA observada seja concentrada em torno de algum valor. Esta propri-

[1]Neste trabalho, será usado o termo "probabilidade prior" (singular) e "probabilidades priores" (plural) para estas probabilidades. Outros termos usualmente encontrados na literatura para essas probabilidades seria "probabilidade *a priori*". Analogamente, será utilizado o termo "probabilidade posterior" no lugar de "probabilidade *a posteriori*".

edade será fundamental para a estimação das variáveis ocultas e dependerá de como foi definida a densidade $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|x_m^{\mathrm{I}}, x_m^{\mathrm{II}})$. Usualmente, a densidade $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|x_m^{\mathrm{I}}, x_m^{\mathrm{II}})$ é chamada de modelo de observação, já que descreve qual a probabilidade de um conjunto de valores da VA oculta ter gerado a informação observada.

As interdependências entre as variáveis ocultas e as variáveis observadas podem ser resumidas graficamente, como na Figura 5.3, onde é exibida uma das possíveis relações entre as duas variáveis em dois quadros: $m-1$ e m. Note que devido à hipótese markoviana, apenas estes dois quadros são suficientes para descrever todo o modelo. As setas no gráfico indicam a dependência na forma de condicionamento, a variável de onde a seta parte aparece como condicionante na densidade da variável de chegada da seta. Além disso, as variáveis ocultas são escritas dentro de círculos, enquanto a variável observada é escrita dentro de um quadrado. No caso da

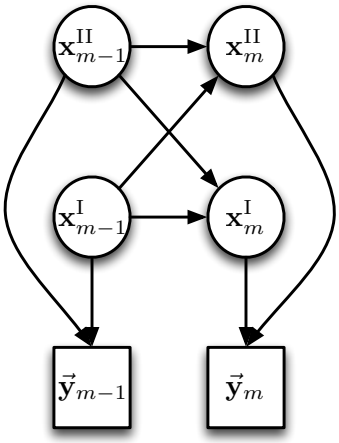


Figura 5.3: Exemplo de uma representação gráfica de um HMM com duas variáveis ocultas.

Figura 5.3, são representadas as seguintes dependências:

$$p_{\mathbf{x}_m^{\mathrm{I}}}(x_m^{\mathrm{I}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}) \tag{5.8}$$

$$p_{\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}) \tag{5.9}$$

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|x_m^{\mathrm{I}}, x_m^{\mathrm{II}}). \tag{5.10}$$

Nesse exemplo, a expressão para a probabilidade condicional conjunta das variáveis ocultas no quadro m pode ser obtida através das expressões anteriores da seguinte maneira:

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{I}}, x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}) = p_{\mathbf{x}_m^{\mathrm{I}}}(x_m^{\mathrm{I}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}) p_{\mathbf{x}_m^{\mathrm{II}}}(x_m^{\mathrm{II}}|x_{m-1}^{\mathrm{I}}, x_{m-1}^{\mathrm{II}}), \tag{5.11}$$

que expressa a independência condicional das variáveis. Note que é possível ter um HMM com outras decomposições da densidade conjunta. A representação gráfica

também permite a verificação da validade do modelo gerado. Se as setas ligando as variáveis formarem ciclos, então o modelo não é válido [105]. A evolução da cadeia é globalmente descrita por

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}},\mathbf{x}_{m-1}^{\mathrm{I}},\mathbf{x}_{m-1}^{\mathrm{II}}}(e_{m,i},e_{m-1,j}) = p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{m,i}|e_{m-1,j})p_{\mathbf{x}_{m-1}^{\mathrm{I}},\mathbf{x}_{m-1}^{\mathrm{II}}}(e_{m-1,j}). \tag{5.12}$$

5.1.2 Representação Matricial

É possível, também, realizar uma formulação matricial para HMMs que será útil para a descrição dos algoritmos de inferência a serem vistos em seguida. Esta formulação também facilitará a implementação do HMM de forma eficiente.

Inicialmente, o modelo de transição será descrito em função dos estados da seguinte forma:

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{m,i}|e_{m-1,j}) = k_{i,j}, \tag{5.13}$$

onde o valor $k_{i,j} \in \mathbb{R}$ armazena a probabilidade de se sair do j-ésimo estado no quadro $m-1$ para o i-ésimo estado no quadro m. De posse dos valores $k_{i,j}$ para todos os possíveis estados, pode-se organizá-los numa matriz[2] $\vec{A} \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ onde o elemento na i-ésima linha e na j-ésima coluna equivale a $k_{i,j}$.

A matriz $\vec{A}$ é chamada de matriz de transição do HMM: sua i-ésima linha armazena as probabilidades de se chegar ao i-ésimo estado no quadro m tendo vindo de qualquer estado j no quadro $m-1$; sua j-ésima coluna armazena as probabilidades de se chegar a qualquer estado i no quadro m tendo vindo do estado j no quadro $m-1$. Note que a matriz de transição não depende do quadro atual m, nem das variáveis observadas; ela apenas depende das probabilidades de transição que são inerentes ao modelo criado. Com isso, ela eficientemente captura toda a informação da dependência temporal entre as variáveis do modelo.

Deve-se ressaltar que, normalmente, a construção de um modelo gera estados que só permitem transições para poucos outros estados (com probabilidade não-nula). Quando isto acontece para um grande número de estados, a matriz $\vec{A}$ se torna esparsa, possuindo a maior parte de seus elementos iguais a zero. Como será visto mais adiante, pode-se aproveitar essa propriedade de matriz para reduzir o custo computacional dos algoritmos de inferência.

A observação no quadro m também pode ser escrita em função dos estados, neste caso

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|e_{m,i}) = \phi_{m,i}, \tag{5.14}$$

e os valores $\phi_{m,i}$ podem ser organizados numa matriz diagonal $\vec{O}_m \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ de forma que o i-ésimo elemento de sua diagonal principal equivale a $\phi_{m,i}$. Com isso,

[2]Tomamos a liberdade de denotar matrizes simplesmente modificando a variável minúscula do vetor por uma variável maiúscula.

o i-ésimo elemento da diagonal desta matriz, chamada de matriz de observação, armazena a densidade da probabilidade do valor $\vec{y}_m$ observado no quadro m se o estado corrente é $e_{m,i}$. Observe que esta matriz depende da observação no quadro m, logo precisa ser atualizada a cada quadro.

Por fim, a probabilidade prior de cada estado pode ser armazenada num vetor chamado $\vec{\pi} \in \mathbb{R}^{|\mathcal{N}|\times 1}$, cujo i-ésimo elemento guarda a probabilidade $p_{\mathbf{x}_0^{\mathrm{I}},\mathbf{x}_0^{\mathrm{II}}}(e_{0,i})$, ou seja, a probabilidade de um determinado estado iniciar a sequência de estados percorridos.

5.1.3 Algoritmos de Inferência

Até o momento, foi realizada apenas a descrição de um HMM. Ainda não foi discutido como se pode obter uma sequência de valores mais prováveis para cada VA oculta a partir da observação de um sinal; tal tarefa é realizada pelos algoritmos de inferência. Para que se possa descrever os algoritmos de inferência, inicialmente é necessário definir exatamente o que se deseja inferir do HMM. Diferentes algoritmos existirão para se estimar diferentes quantidades [106].

A primeira quantidade que pode ser obtida do HMM é a probabilidade de se ter passado por cada estado em cada quadro conhecendo-se todos os dados disponíveis. Essa quantidade pode ser descrita matematicamente da seguinte forma

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{i,m}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_{M-1}), \tag{5.15}$$

sendo M o número total de observações. A probabilidade acima pode ser interpretada como a probabilidade posterior dos estados, ou seja, após a observação de todos os dados. Note que os valores observados em todos os quadros são utilizados para se obter a probabilidade de cada possível valor das VAs ocultas no quadro m. Uma vez obtida a quantidade $p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{i,m}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_{M-1})$, pode-se obter uma sequência de estados mais prováveis através de

$$\hat{i} = \arg\max_i p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{i,m}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_{M-1}). \tag{5.16}$$

A estimativa $e_{\hat{i},m}$ será a estimativa *maximum a posteriori* (MAP) para o quadro m.

Note que uma estimativa MAP no quadro m não depende da estimativa MAP em nenhum outro quadro. Isso pode levar a escolhas de sequências de estimativas que, apesar de individualmente serem ótimas, não o são conjuntamente. Para se obter a sequência de estimativas ótima, deverá ser obtida a probabilidade *conjunta* após a observação de todos os dados. De forma matemática, ter-se-ia

$$p_{\{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}},\forall m\}}(e_{\hat{i}_0,0},e_{\hat{i}_1,1},\ldots,e_{\hat{i}_{M-1},M-1}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_{M-1}), \tag{5.17}$$

onde agora ao se maximizar a quantidade $p_{\{\mathbf{x}_m^{\mathrm{I}}, \mathbf{x}_m^{\mathrm{II}}, \forall m\}}$ é obtida uma sequência de estados $e_{\hat{i}_0,0}, e_{\hat{i}_1,1}, \ldots, e_{\hat{i}_{M-1},M-1}$ conjuntamente ótimos. Esta sequência de estados também é chamada de percurso, já que define uma sequência válida de valores para cada VA oculta a cada quadro dentro do HMM.

Nas próximas seções serão descritos dois algoritmos de inferência, um que maximiza a equação (5.15) e outro que calcula o percurso mais provável maximizando a equação (5.17).

Algoritmo *Forward-Backward*

O algoritmo *Forward-Backward* [105, 106] calcula a probabilidade posterior após a observação de todos os dados (equação (5.15)) de forma recursiva. Para isso, ele divide o cálculo em duas etapas. Na primeira (chamada de *Forward*), a probabilidade de cada estado no quadro m é calculada utilizando a probabilidade cada estado no quadro $m-1$. Na segunda (chamada *Backward*), a probabilidade de cada estado no quadro m é obtida a partir da probabilidade de cada estado no quadro $m+1$. A seguir, serão descritas essas duas etapas. Note que não será feita nesta seção uma derivação formal do algoritmo; será feita, no lugar, a apresentação de cada etapa seguida de sua interpretação.

A etapa *forward* é definida através da seguinte recursão:

$$\vec{\alpha}_m = K_m^{\alpha} \vec{O}_{m-1} \vec{A} \vec{\alpha}_{m-1}, \tag{5.18}$$

onde $K_m^{\alpha} = \left(\sum_i \vec{\alpha}_m[i]\right)^{-1}$ é uma constante de normalização e a recursão é inicializada através de $\vec{\alpha}_0 = \vec{\pi}$. As quantidades $\vec{\alpha}_m$ calculadas para cada quadro m são a probabilidade de cada estado tendo-se observado os dados do quadro 0 até o quadro m. A recursão descrita na equação (5.18) pode ser interpretada da seguinte forma: a probabilidade no quadro $m+1$ de um determinado estado é igual à soma da probabilidade de se chegar nesse estado multiplicada pela probabilidade do estado de origem no quadro m ($\vec{A}\vec{\alpha}_m$). O resultado, então, é ponderado por um fator proporcional à chance de ter sido gerado o dado observado pelo estado (multiplicação por $\vec{O}_m$). A multiplicação por K_{m+1}^{α} garante que o resultado é uma probabilidade e também evita problemas numéricos.

A etapa *backward* pode ser expressa da seguinte forma:

$$\vec{\beta}_m = K_m^{\beta} \vec{O}_{m+1} \vec{A}^T \vec{\beta}_{m+1}, \tag{5.19}$$

onde $K_m^{\beta} = \left(\sum_i \vec{\beta}_m[i]\right)^{-1}$ é uma constante de normalização e a recursão é inicializada com $\vec{\beta}_M = \vec{1}$ (um vetor cujos elementos são todos iguais a 1). A interpretação da recursão *backward* não é tão direta quanto a *forward*. De forma simples, o vetor

$\vec{\beta}_m$ pode ser entendido como a probabilidade de um determinado estado ter gerado a sequência de dados observada a partir do quadro m. O vetor utilizado para inicialização da recursão indica que a sequência de dados pode terminar em qualquer um dos estados com igual chance.

Uma vez calculados os vetores $\vec{\alpha}_m$ e $\vec{\beta}_m$ para todos os quadros, é possível obter a probabilidade de cada estado condicionada aos dados observados através de

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{i,m}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_{M-1}) = \vec{\alpha}_m[i]\vec{\beta}_m[i]. \tag{5.20}$$

Uma vez de posse desta quantidade, os estados mais prováveis podem ser obtidos através de uma busca em linha pelo estado mais provável em cada quadro.

O algoritmo *forward-backward* possui também a vantagem de facilmente explorar a possível esparsidade da matriz $\vec{A}$. Como sua implementação envolve apenas multiplicações entre matrizes e vetores, o número de operações realizadas no caso de modelos esparsos pode ser facilmente reduzido através da utilização de um pacote numérico para matrizes esparsas.

Deve-se notar que uma versão causal do algoritmo pode ser obtida se apenas a recursão *forward* for utilizada. Neste caso, as probabilidades encontradas são

$$p_{\mathbf{x}_m^{\mathrm{I}},\mathbf{x}_m^{\mathrm{II}}}(e_{i,m}|\vec{y}_0,\vec{y}_1,\ldots,\vec{y}_m) = \vec{\alpha}_m[i], \tag{5.21}$$

devendo-se ter em mente que dessa forma apenas os dados observados até o quadro m são utilizados para se obter a sua probabilidade.

Algoritmo de Viterbi

O algoritmo de Viterbi [107] procura calcular diretamente a sequência de estados válidos (percurso) conjuntamente mais prováveis, sem calcular explicitamente a probabilidade conjunta descrita na equação (5.17). Para isso, o algoritmo usa o fato de que num quadro apenas o percurso mais provável terminando num determinado estado precisa ser armazenado (em vez de todos os percursos que podem chegar nesse estado) [106]. Isto limita a quantidade de percursos a serem buscados ao número total de estados, reduzindo a complexidade de busca. Novamente, não será feita uma derivação formal do algoritmo nesta seção, sendo o algoritmo apresentado e suas principais propriedades discutidas.

A probabilidade de um percurso mais provável terminar no estado i pode ser calculada (a menos de uma constante de proporcionalidade) através da seguinte recursão:

$$\vec{\mu}_m[i] = \vec{O}_m[i,i] \max_j \left(\vec{A}[j,i]\vec{\mu}_{m-1}[j]\right), \tag{5.22}$$

onde é feita a inicialização $\vec{\mu}_0 = \vec{O}_0\vec{\pi}$. Note que a maximização no lado direito da

equação (5.22) garante que é selecionado apenas o melhor percurso passando pelo i-ésimo estado, com a quantidade $\vec{A}[j,i]\vec{\mu}_{m-1}[j]$ sendo proporcional à probabilidade de a sequência passando pelo estado j no quadro $m-1$ chegar no estado i no quadro m. Desta forma, é necessário apenas atualizar as quantidades $\vec{\mu}_m$ para cada estado para se obter os percursos mais prováveis até o quadro m.

A equação (5.22) informa apenas qual é o percurso mais provável que passa por um determinado estado no quadro m. Para se obter o percurso propriamente dito, é necessário armazenar os estados visitados por cada percurso até o quadro m. Para isso é necessário realizar a seguinte atualização em cada quadro:

$$\hat{j} = \arg\max_j \left(\vec{A}[j,i]\vec{\mu}_{m-1}[j]\right), \tag{5.23}$$

$$\mathcal{P}_{m,i} = \mathcal{P}_{m-1,\hat{j}} \cup \{i\}. \tag{5.24}$$

Cada conjunto $\mathcal{P}_{m,i}$ contém a sequência de estados visitados terminando no estado i no quadro m.

Deve-se notar que se calculadas como descrito na equação (5.22), as quantidades $\vec{\mu}_m[i]$ ficarão progressivamente menores a cada quadro, devido à sequência de multiplicações por números positivos menores que 1. Para evitar problemas numéricos, usualmente trabalha-se com o logaritmo desta quantidade:

$$\log(\vec{\mu}_m[i]) = \log(\vec{O}_m[i,i]) + \max_j \left(\log(\vec{A}[j,i]) + \log(\vec{\mu}_{m-1}[j])\right). \tag{5.25}$$

Desta forma, evitam-se problemas numéricos sem a perda da otimalidade do algoritmo.

Algoritmo de Viterbi Esparso

A formulação anterior do algoritmo de Viterbi não explora a possível esparsidade da matriz de estados para reduzir o número de operações realizadas em cada quadro. Além disso, o fato de se utilizar o logaritmo, transformando multiplicações em somas, dificulta o uso de soluções esparsas, já que a esparsidade de dois vetores não é preservada após a sua soma. Nesta seção, é apresentada uma implementação do algoritmo de Viterbi eficiente para matrizes de transição esparsas.

Para poder definir a versão esparsa do algoritmo de Viterbi, inicialmente será definido o conjunto de vizinhança de um dado estado i como

$$\mathcal{V}_s = \{i | \vec{A}[i,s] \neq 0\}. \tag{5.26}$$

Assim, o conjunto de vizinhança do estado s contém apenas os estados que podem ser alcançados a partir do estado s. Observe que o conjunto de vizinhança de

cada estado pode ser pré-calculado e armazenado, dependendo apenas da matriz de transição.

Utilizando os conjuntos de vizinhanças, é possível reescrever a recursão do algoritmo de Viterbi de forma que a maximização considere apenas os estados alcançáveis. Isto pode ser alcançado modificando-se a equação (5.22) de forma que

$$\vec{\mu}_m[i] = \vec{O}_m[i,i] \max_{j \in \mathcal{V}_i} \left(\vec{A}[j,i] \vec{\mu}_{m-1}[j] \right). \tag{5.27}$$

Note que utilizando esta nova formulação, também evitam-se multiplicações por zero, já que são utilizados apenas os elementos não-nulos da matriz $\vec{A}$.

Uma outra possível melhoria no desempenho consiste em evitar que estados que não possam ser alcançados sejam atualizados. Para isto, é calculada a quantidade

$$\vec{\zeta}_m = \vec{A} \vec{\mu}_{m-1}, \tag{5.28}$$

que é proporcional à soma das probabilidades de transição para um estado ponderadas pelas probabilidades dos percursos que levam a ele. Note que essa multiplicação pode ser feita de forma eficiente caso a matriz $\vec{A}$ seja esparsa. Se o $\vec{\zeta}_m$ for zero para um determinado estado, então ele não é alcançável por nenhum percurso que possui probabilidade não-nula. Assim sendo, a probabilidade de um percurso terminar no estado em questão será zero e a recursão da equação (5.27) não precisa ser calculada. De forma matemática, deve-se criar o conjunto,

$$\mathcal{U}_m = \{ i | \vec{\zeta}_m[i] \neq 0 \} \tag{5.29}$$

e atualizar apenas os percursos terminando em $i \in \mathcal{U}_m$. Deve-se notar que a cardinalidade do conjunto $\mathcal{U}_m$ tende a aumentar ao longo dos quadros, pois conforme os percursos evoluem ao longo do tempo, a chance de um estado não ser atingível diminui. Com isso, uma boa estratégia é monitorar a cardinalidade de $\mathcal{U}_m$, e caso ela passe de um determinado limiar, parar de usar $\mathcal{U}_m$.

Algoritmo de Viterbi Posterior

Será descrito nesta seção um algoritmo que combina o baixo custo computacional do algoritmo *forward-backward* em encontrar a probabilidade posterior para matrizes de transição esparsas com a capacidade do algoritmo de Viterbi de obter um percurso ótimo. O algoritmo consiste na busca pelo melhor percurso válido sobre a probabilidade posterior [108] através da aplicação do algoritmo de Viterbi utilizando a probabilidade posterior obtida a partir do algoritmo *forward-backward*.

Para se calcular a sequência mais provável será utilizada a seguinte recursão

$$\vec{v}_m[i] = \vec{\alpha}_m[i]\vec{\beta}_m[i] \max_{j \in \mathcal{V}_i} \left(\vec{v}_{m-1}[j]\right), \tag{5.30}$$

onde o vetor $\vec{v}_m$ possui uma interpretação similar à do vetor $\vec{\mu}_m$ no algoritmo de Viterbi convencional. Note, que diferentemente do algoritmo de Viterbi, é utilizada no lugar da observação a probabilidade posterior de se ter observado o i-ésimo estado. Além disso, é escolhido, de um quadro para outro, apenas o percurso que está na vizinhança de i que possui a maior probabilidade. A ponderação pela probabilidade de transição não se faz necessária, pois já foi incorporada no vetor $\vec{v}_{m-1}[j]$ ao ser multiplicado por $\vec{\alpha}_{m-1}[j]\vec{\beta}_{m-1}[j]$.

Da mesma forma que é feito no algoritmo de Viterbi, a sequência de estados visitadas em cada percurso deve ser armazenada, fazendo-se

$$\hat{j} = \arg\max_j \left(\vec{v}_{m-1}[j]\right) \tag{5.31}$$

$$\mathcal{P}_{m,i} = \mathcal{P}_{m-1,\hat{j}} \cup \{i\}. \tag{5.32}$$

Dessa forma, é obtida a sequência de estados mais prováveis segundo a probabilidade posterior. Observe que, neste caso, a probabilidade posterior de cada estado da sequência mais provável pode ser armazenada também. Esta informação fornece uma medida de quão confiável é a escolha, permitindo avaliar quão bem o modelo escolhido descreve os dados observados.

5.2 Modelo Hierárquico

É apresentado nesta seção o primeiro modelo proposto para rastreamento da métrica. Considerando que a percepção métrica acontece em três níveis distintos (o tatum, o tactus e o compasso), o modelo proposto procura modelar diretamente cada nível e a forma como eles estão interligados. Em particular, serão explorados os fatos de que o compasso só pode iniciar junto com um tactus, e que um tactus só pode iniciar junto com um tatum, para se criar um modelo oculto de Markov hierárquico, onde a ocorrência de um nível está sujeita à ocorrência do nível cujo período é menor.

Além de rastrear a ocorrência dos três níveis, o modelo também considerará mudanças no período relativo entre os níveis hierárquicos e também mudanças no andamento. Em particular, o período do tactus (associado ao andamento) servirá de âncora, sendo o período do tatum um submúltiplo seu, e o período do compasso um múltiplo. O modelo também permitirá mudanças lentas no período do tactus, não sendo modeladas mudanças bruscas no andamento.

O modelo terá variáveis aleatórias associadas a cada nível hierárquico. Ideal-

mente, três VAs binárias modelariam se um determinado quadro inicia ou não um tatum, tactus ou compasso. Restrições sobre as probabilidades de transição destas variáveis impediriam a ocorrência de novo tactus sem a ocorrência de novo tatum, e da ocorrência de novo compasso sem a ocorrência de novo tactus. Infelizmente, um número maior de VAs deve ser utilizado, pois é necessário controlar o período com que ocorrem as transições para novo tatum, novo tactus e novo compasso. Para isto, duas outras VAs serão utilizadas para cada nível hierárquico: três contadores que controlam respectivamente há quantos quadros não se observa um novo tatum, um novo tactus ou um novo compasso, e três períodos: um que no caso do tactus armazena o andamento atual do sinal, outro que para o tatum armazena o inverso do divisor do período do tactus e outro que para o compasso armazena o multiplicador do período do tactus. A seguir, serão formalmente definidas todas as variáveis.

5.2.1 Variáveis Aleatórias

Nesta seção, serão apresentadas as VAs para cada nível hierárquico. Novamente, as VAs serão notadas em negrito, sendo o sub-índice relacionado ao quadro em que a VA ocorre. Os nomes dados às VAs também seguirão uma convenção. Um superíndice será utilizado para indicar a qual nível hierárquico a VA pertence: t para tatum, b para tactus e c para compasso. Além disso, o nome da VA estará associado à sua função: **I** para os indicadores de nova ocorrência de um dos níveis hierárquicos num quadro, **c** para os contadores e $\boldsymbol{\ell}$ para os períodos.

Note que a escolha feita para alguns valores (notadamente sobre a razão entre os períodos do tactus e do compasso) escolhidos para as variáveis a serem definidas considera as métricas mais encontradas na prática. Com isso, a escolha restringe o modelo, já que ele será apenas capaz de rastrear sinais que exibem as métricas geradas pelas combinações dos períodos do tactus e do compasso. Contudo, novos valores podem ser adicionados de forma a adicionar novas métricas ao modelo através da inclusão de novos possíveis valores para os períodos e consequente ajuste nas demais VAs.

A variável $\mathbf{I}_m^{\mathrm{t}} \in \{0, 1\}$ modela se o quadro m marca a ocorrência de novo tatum ou não. O período $\boldsymbol{\ell}_m^{\mathrm{t}} \in \left\{\frac{1}{2}, \frac{1}{3}\right\}$ indica se a divisão do período do tactus por tatum é binária ou ternária. O contador $\mathbf{c}_m^{\mathrm{t}} \in \{0, \ldots, \ell^{\max}/2 + \sigma_t - 1\}$, que conta a quantidade de quadros desde o último início de tatum, pode variar de 0 até um valor associado ao máximo período do tactus $\ell^{\max}$ e a uma tolerância σ_t que será detalhada futuramente.

O indicador $\mathbf{I}_m^{\mathrm{b}} \in \{0, 1\}$ modela se o quadro m marca a ocorrência de novo tactus ou não. O período do tactus $\boldsymbol{\ell}_m^{\mathrm{b}} \in \{\ell^{\min}, \ldots, \ell^{\max}\}$; no entanto, está diretamente associado ao andamento e pode variar desde o menor andamento aceitável, de período

$\ell^{\max}$, ao maior, de período $\ell^{\min}$, ambos os valores em número de quadros. O contador $\mathbf{c}_m^{\mathrm{b}} \in \{0, 1, 2\}$, no caso do tactus, conta a quantidade de tatums que iniciaram após o último tactus. Observe que a presença do tatum permite a simplificação desse contador, não sendo necessário medir os tactus em função dos quadros diretamente.

Por fim, as variáveis associadas ao compasso são definidas de forma muito similar às do tatum. O indicador de novo compasso é definido como $\mathbf{I}_m^{\mathrm{c}} \in \{0, 1\}$. O seu período $\boldsymbol{\ell}_m^{\mathrm{c}} \in \{2, 3, 4\}$, indica a quantidade de tactus que podem existir num compasso. O contador $\mathbf{c}_m^{\mathrm{c}} \in \{0, 1, 2, 3\}$, conta o número de tactus que iniciaram após o último compasso.

5.2.2 Modelo de Transição

Nesta seção, será descrito o modelo de transição do modelo hierárquico, sendo definidas as probabilidades de transição de cada uma das VAs apresentadas na seção anterior. O modelo de transição apresentado será responsável por garantir a coerência temporal entre as VAs de um mesmo nível e as VAs dos outros níveis. Serão explicadas separadamente as transições das VAs de cada nível hierárquico.

Tatum

A probabilidade de o contador do tatum assumir um valor num determinado quadro m pode ser descrita da forma abaixo[3]

$$p_{\mathbf{c}_m^{\mathrm{t}}}(c_m^{\mathrm{t}} | c_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{t}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{t}} = c_{m-1}^{\mathrm{t}} + 1) \wedge (I_{m-1}^{t} = 0) \\ 1, & \text{se } (c_m^{\mathrm{t}} = 0) \wedge (I_{m-1}^{t} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.33}$$

dependendo do valor do contador e do indicador do tatum no quadro anterior. Caso o indicador seja 0 (não houve um novo tatum), o valor do contador é incrementado. Caso o indicador seja 1, o contador é zerado. Observe que, dados o contador e o indicador no quadro $m - 1$, o valor do contador no quadro m é uma quantidade determinística. Esse fato faz com que a maior parte das transições possíveis desta variável tenham probabilidade nula, gerando um modelo esparso. Com esta formulação, a probabilidade de um determinado valor do contador estará sempre intrinsecamente associada ao número de quadros que se passaram desde a última ocorrência de tatum.

A probabilidade do período do tatum no quadro m pode ser definida através da

[3]Durante a descrição das probabilidades, o símbolo $\wedge$ denotará um "E" lógico.

seguinte expressão

$$p_{\ell_m^{\mathrm{t}}}(\ell_m^{\mathrm{t}}|\ell_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{c}}) = \begin{cases} \sigma_{\mathrm{s}}, & \text{se } (\ell_m^{\mathrm{t}} \neq \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{c}} = 1) \\ 1 - \sigma_{\mathrm{s}}, & \text{se } (\ell_m^{\mathrm{t}} = \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{c}} = 1) \\ 1, & \text{se } (\ell_m^{\mathrm{t}} = \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{c}} = 0) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.34}$$

onde o valor σ_s define a probabilidade de o número de tatums por tactus mudar entre dois compassos. Observe que a probabilidade depende do indicador do compasso, que impede o período do tatum de se alterar no meio de um compasso (o que é usualmente respeitado para as músicas de interesse neste trabalho). O valor de σ_s deve ser encontrado através de treinamento ou empiricamente, e é um parâmetro livre do modelo. Novamente, a maior parte das transições do período possuem probabilidade nula, contribuindo para a esparsidade do modelo.

A probabilidade do indicador do tatum é uma das mais importantes do modelo hierárquico, já que os indicadores dos outros níveis hierárquicos dependem desta quantidade. Além disso, a sua transição precisa ser definida de tal maneira que o período do tatum seja respeitado, mas também permitindo uma margem de segurança para imprecisões que podem ocorrer. Considerando estes fatores, a seguinte probabilidade de transição foi escolhida para o indicador do tatum no quadro m:

$$p_{\mathbf{I}_m^{\mathrm{t}}}(I_m^{\mathrm{t}}|c_m^{\mathrm{t}}, \ell_m^{\mathrm{t}}, \ell_m^{\mathrm{b}}) = \begin{cases} \overline{w}[c_m^{\mathrm{t}} - \mathrm{round}(\ell_m^{\mathrm{b}}\ell_m^{\mathrm{t}}) + 1], & \text{se } I_m^{\mathrm{t}} = 1 \\ 1 - \overline{w}[c_m^{\mathrm{t}} - \mathrm{round}(\ell_m^{\mathrm{b}}\ell_m^{\mathrm{t}}) + 1], & \text{se } I_m^{\mathrm{t}} = 0, \end{cases} \tag{5.35}$$

onde $\overline{w}[.]$ é uma janela simétrica em relação à origem, normalizada para $\sum_i \overline{w}[i] = 1$ com $2\sigma_t + 1$ valores não nulos. Esta expressão pode ser compreendida da seguinte forma: a probabilidade de ocorrer novo tatum aumenta conforme o contador de tatum incrementado de 1 se aproxima do período do tatum convertido para número de quadros (obtido multiplicando-se o período do tactus pelo período do tatum); a probabilidade de não ocorrer novo tatum é seu complemento. Neste trabalho, foi escolhida uma janela de Hann para parametrizar esta probabilidade. A Figura 5.4 exibe o formato da janela juntamente com os valores de probabilidade para $\sigma_t = 4$. Observe que é nula a chance de novo tatum ocorrer para contadores cujos valores são muito diferentes do período atual. Conforme o contador se aproxima do período, a chance aumenta, sendo máxima quando os dois são iguais. Com isso, a quantidade de valores não-nulos da janela pode ser interpretada como a tolerância do modelo a imprecisões de execução: quanto maior σ_t, maior a tolerância. Deseja-se escolher valores para σ_t que permitam que imprecisões de execução sejam toleradas, mas também que não sejam tão grandes a ponto de impedirem o modelo de seguir cor-

retamente o tatum. Na literatura, são encontrados valores para $2\sigma_t + 1$ próximos de 50 ms, quando estes são convertidos para segundos [21, 109].

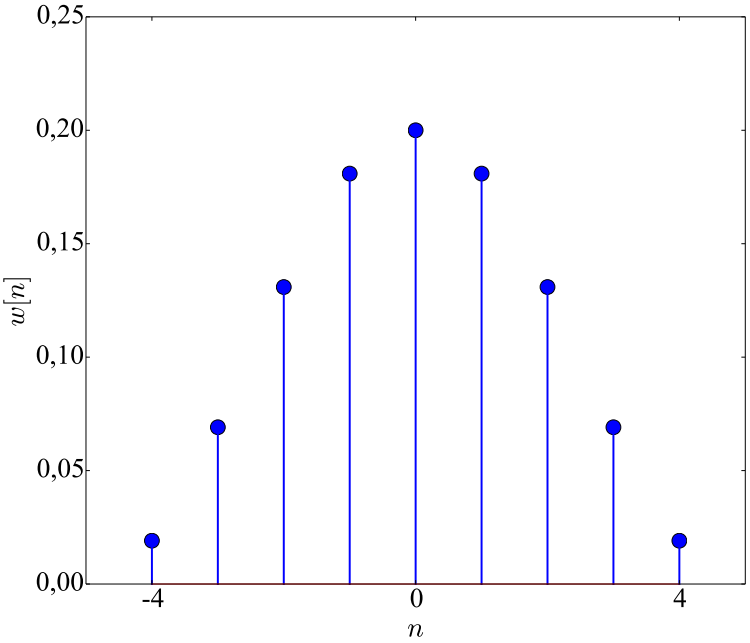


Figura 5.4: Exemplo da parametrização utilizada para a probabilidade de se observar novo tatum no quadro m para um determinado valor de $c_m - \ell_m + 1$, com $\sigma_t = 2$.

Tactus

O contador do tactus precisa apenas armazenar quantos tatums iniciaram após o início do tactus atual e deve ser zerado após o início de novo tactus. Com isso, matematicamente, tem-se

$$p_{\mathbf{c}_m^{\mathrm{b}}}(c_m^{\mathrm{b}} | c_{m-1}^{\mathrm{b}}, I_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{b}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{b}} = c_{m-1}^{\mathrm{b}} + 1) \wedge (I_{m-1}^{\mathrm{t}} = 1) \wedge (I_{m-1}^{\mathrm{b}} = 0) \\ 1, & \text{se } (c_m^{\mathrm{b}} = c_{m-1}^{\mathrm{b}}) \wedge (I_{m-1}^{\mathrm{t}} = 0) \\ 1, & \text{se } (c_m^{\mathrm{b}} = 0) \wedge (I_{m-1}^{\mathrm{b}} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.36}$$

onde os três casos possíveis para o contador são cobertos: incrementar por ter iniciado novo tatum sem ter iniciado novo tactus, permanecer com o mesmo valor quando não iniciou novo tatum e assumir um valor nulo por ter iniciado novo tactus. Repare, novamente, que a maior parte das transições possuem probabilidades nulas, contribuindo para que seja criado um modelo esparso.

O período do tactus determina o período dos demais níveis métricos e tem que

ser atualizado cuidadosamente, já que precisa ser definido de modo a acompanhar variações no andamento. Por isso, foi escolhida a seguinte probabilidade de transição

$$p_{\boldsymbol{\ell}_m^{\mathrm{b}}}(\ell_m^{\mathrm{b}}|\ell_{m-1}^{\mathrm{t}},\ell_{m-1}^{\mathrm{b}},I_{m-1}^{\mathrm{b}}) = \begin{cases} 1, & \text{se } \left(\ell_m^{\mathrm{b}} = \frac{c_{m-1}^{\mathrm{t}}}{\ell_{m-1}^{\mathrm{t}}}\right) \wedge (I_{m-1}^{\mathrm{b}} = 1) \\ 1, & \text{se } \left(\ell_m^{\mathrm{b}} = \ell_{m-1}^{\mathrm{b}}\right) \wedge (I_{m-1}^{\mathrm{b}} = 0) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.37}$$

onde se pode notar que o período pode ser alterado apenas após a observação de novo tactus e que seu valor sempre é atualizado de acordo com o valor do contador do tatum convertido pela divisão pelo período do tatum. Como o valor do contador do tatum sempre precisa estar próximo do período, o valor atualizado do período do tactus sempre estará próximo do valor passado, permitindo, assim, apenas o acompanhamento de mudanças lentas no andamento.

A transição do indicador do tactus pode ser expressa através de

$$p_{\mathbf{I}_m^{\mathrm{b}}}(I_m^{\mathrm{b}}|c_m^{\mathrm{b}},\ell_m^{\mathrm{t}},I_m^{\mathrm{t}}) = \begin{cases} 1, & \text{se } (I_m^{\mathrm{b}} = 1) \wedge \left(c_m^{\mathrm{b}} = \frac{1}{\ell_m^{\mathrm{t}}} - 1\right) \wedge (I_m^{\mathrm{t}} = 1) \\ 1, & \text{se } (I_m^{\mathrm{b}} = 0) \wedge \left(c_m^{\mathrm{b}} \neq \frac{1}{\ell_m^{\mathrm{t}}} - 1\right) \wedge (I_m^{\mathrm{t}} = 1) \\ 1, & \text{se } (I_m^{\mathrm{b}} = 0) \wedge (I_m^{\mathrm{t}} = 0) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.38}$$

onde a transição para início de tactus ocorre apenas se um tatum iniciar no mesmo quadro e se o contador de tactus incrementado de 1 for igual ao inverso do período do tatum. Desta forma, deixa-se para o indicador do tatum a modelagem das imprecisões temporais, sendo a indicação de novo tactus responsável apenas por garantir que um tactus não pode iniciar ocorrer sem um tatum iniciar, e que é respeitado o número de tatums dentro de um tactus.

Compasso

No caso do compasso, a probabilidade de seu contador é definida de forma similar à do contador do tactus, sendo

$$p_{\mathbf{c}_m^{\mathrm{c}}}(c_m^{\mathrm{c}}|c_{m-1}^{\mathrm{c}},I_{m-1}^{\mathrm{b}},I_{m-1}^{\mathrm{c}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{c}} = c_{m-1}^{\mathrm{c}} + 1) \wedge (I_{m-1}^{\mathrm{b}} = 1) \wedge (I_{m-1}^{\mathrm{c}} = 0) \\ 1, & \text{se } (c_m^{\mathrm{c}} = c_{m-1}^{\mathrm{c}}) \wedge (I_{m-1}^{\mathrm{b}} = 0) \\ 1, & \text{se } (c_m^{\mathrm{c}} = 0) \wedge (I_{m-1}^{\mathrm{c}} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.39}$$

onde os três casos permitidos para atualização do contador são: se um tactus iniciou um compasso, o contador é incrementado; caso não tenha ocorrido um tactus, o

contador permanece com o mesmo valor; e se um compasso iniciou, o contador deve ser zerado.

O período do compasso só pode ter o seu valor alterado durante uma mudança de compasso, sendo que as transições ocorrem com probabilidades fixas. Esta restrição segue a utilizada no período do tatum e pode ser escrita matematicamente como

$$p_{\boldsymbol{\ell}^{\mathrm{c}}_m}(\ell^{\mathrm{c}}_m|\ell^{\mathrm{c}}_{m-1}, I^{\mathrm{c}}_{m-1}) = \begin{cases} \frac{\sigma_c}{N^{\mathrm{c}}-1}, & \text{se } (\ell^{\mathrm{c}}_m \neq \ell^{\mathrm{c}}_{m-1}) \wedge (I^{\mathrm{c}}_{m-1} = 1) \\ 1-\sigma_c, & \text{se } (\ell^{\mathrm{c}}_m = \ell^{\mathrm{c}}_{m-1}) \wedge (I^{\mathrm{c}}_{m-1} = 1) \\ 1, & \text{se } (\ell^{\mathrm{c}}_m = \ell^{\mathrm{c}}_{m-1}) \wedge (I^{\mathrm{c}}_{m-1} = 0) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.40}$$

onde σ_c é a chance de o período do compasso mudar após uma mudança de compasso e N^{c} é o número de possíveis compassos. Com isso, a chance de se mudar de um período para qualquer outro é a mesma e pode ser obtida através de uma etapa de treinamento ou escolhida arbitrariamente.

O indicador do compasso é formulado de forma similar ao indicador do tactus, isto é, se o número de tactus iniciados após o último início de compasso for igual ao período do compasso decrementado de 1 e o quadro atual for de início de tactus, então um novo compasso terá iniciado. Matematicamente, isto pode ser descrito da seguinte forma

$$p_{\mathbf{I}^{\mathrm{c}}_m}(I^{\mathrm{c}}_m|c^{\mathrm{c}}_m, \ell^{\mathrm{c}}_m, I^{\mathrm{b}}_m) = \begin{cases} 1, & \text{se } (c^{\mathrm{c}}_m = \ell^{\mathrm{c}}_m - 1) \wedge (I^{\mathrm{c}}_m = 1) \wedge (I^{\mathrm{b}}_m = 1) \\ 1, & \text{se } (c^{\mathrm{c}}_m \neq \ell^{\mathrm{c}}_m - 1) \wedge (I^{\mathrm{c}}_m = 0) \wedge (I^{\mathrm{b}}_m = 1) \\ 1, & \text{se } (I^{\mathrm{c}}_m = 0) \wedge (I^{\mathrm{b}}_m = 0) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.41}$$

Desta forma, é preservada a hierarquia dos níveis métricos, garantindo que são modeladas apenas métricas válidas.

Exemplo

Para ilustrar como as diferentes variáveis estão interligadas, é mostrada na Figura 5.5 uma sequência válida (segundo o modelo adotado) para as variáveis propostas. A sequência escolhida, em particular, está associada ao exemplo de métrica exibido na Figura 5.2.

Deve ser ressaltado que a sequência gerada é apenas uma entre muitas sequências válidas segundo o modelo. O modelo de observação, em conjunto com os algoritmos de referência, será responsável por selecionar, dentre todas as possíveis sequências válidas, qual gerou com maior probabilidade o sinal sob análise.

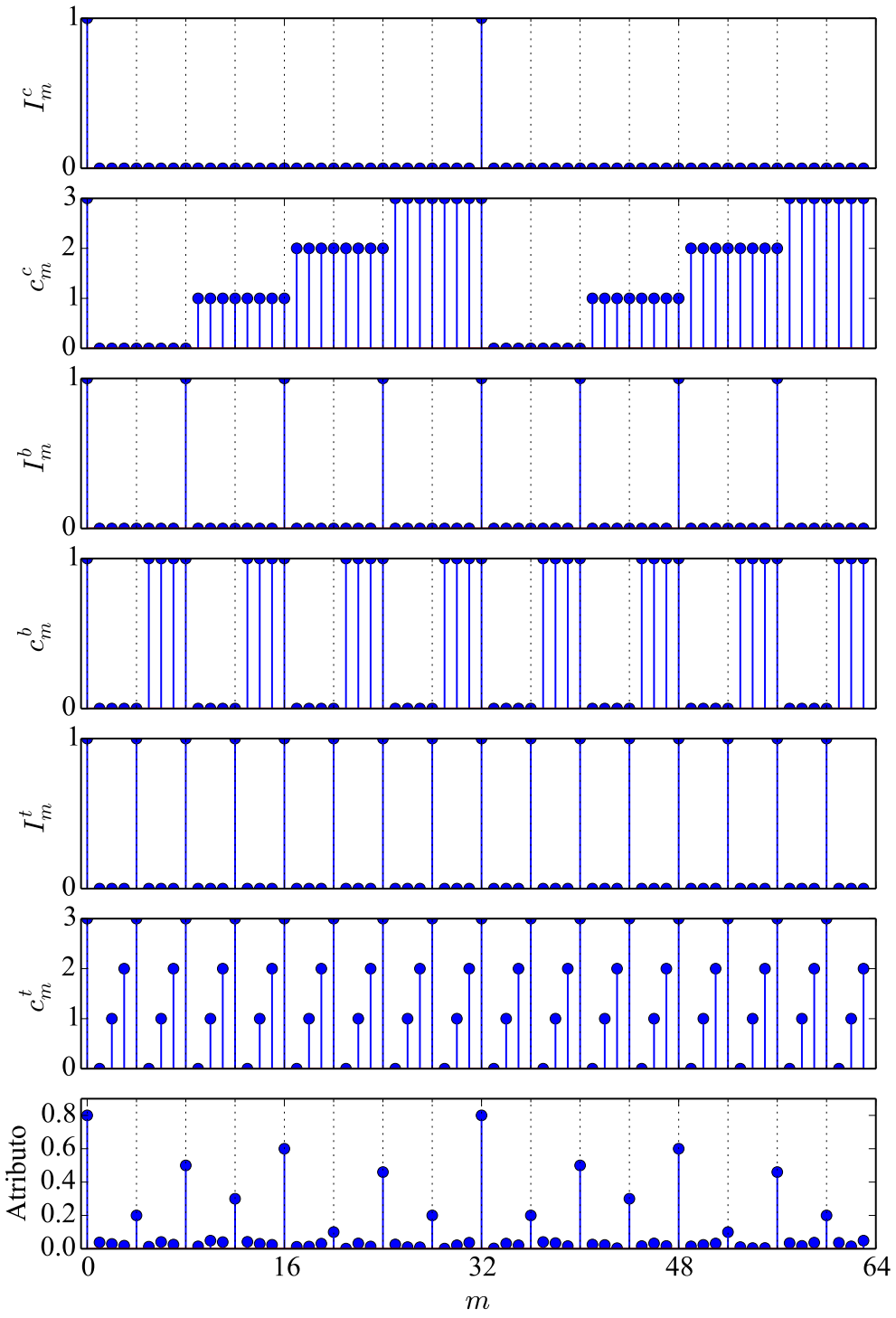


Figura 5.5: Esta figura exibe uma sequência válida de valores (segundo o modelo proposto) para as variáveis do modelo hierárquico. Neste exemplo, o período do tactus é igual a 8 quadros, o do tatum igual a $\frac{1}{2}$ tactus e o do compasso igual a 4 tactus; e os valores dos períodos não variam ao longo dos quadros.

5.2.3 Observação

Nesta seção, será brevemente apresentado o modelo de observação. Em particular, será definido o vetor de observação e com quais VAs do modelo ele está associado. Não será feita uma apresentação detalhada de como se podem obter as distribuições do modelo de observação, pois isso será feito no Capítulo 7.

O vetor de observações no quadro m será composto por uma versão normalizada do fluxo espectral calculado para diferentes sub-bandas da escala MEL, sendo o vetor de observações definido como

$$\vec{\mathbf{y}}_m = \left[\overline{F}^{\mathrm{SF}}[m,0], \ldots, \overline{F}^{\mathrm{SF}}[m,K-1]\right], \tag{5.42}$$

onde K é o número de sub-bandas. Será feita uma descrição no Capítulo 7 da normalização utilizada.

A densidade de probabilidade de observação do vetor de atributos será definida como

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m | I_m^{\mathrm{t}}, I_m^{\mathrm{b}}, I_m^{\mathrm{c}}), \tag{5.43}$$

onde apenas a informação de se o vetor observado está associado a um novo tatum, um novo tactus ou um novo compasso é utilizada para se gerar o modelo de observação. Com isso, assume-se que pode ser obtido um modelo em que os valores para o fluxo espectral em cada sub-banda podem ser determinados apenas através do conhecimento de se o quadro em questão está associado apenas a um novo tatum, a um novo tatum e um novo tactus, ou a um novo tatum, um novo tactus e um novo compasso. Espera-se encontrar um modelo que consiga informar a chance de o fluxo espectral extraído do sinal num quadro estar associado a um início de nível métrico ou se ele denota a ausência de informação métrica. No Capítulo 7, são discutidas diferentes formas de se obter este modelo e diferentes maneiras de parametrizar a sua distribuição de probabilidade. Note que ao se escolher utilizar o fluxo espectral calculado para diferentes sub-bandas, deixa-se a critério do modelo a melhor forma de se combinar a informação nas sub-bandas.

5.2.4 Prior

A probabilidade prior de cada uma das VAs será descrita nesta seção.

Para os contadores, uma distribuição inicial uniforme será considerada. Esta escolha assume que um determinado sinal pode iniciar com os contadores diferentes de zero (no caso de uma música exibindo anacruse ou um arquivo de áudio em que o início não coincide com o início da música). Distribuições uniformes também serão escolhidas para os indicadores; novamente, o primeiro quadro pode ou não coincidir com o primeiro tactus, tatum ou cabeça de compasso.

A probabilidade prior para cada valor dos períodos do compasso e do tatum será escolhida através da distribuição dos valores médios observados no banco de sinais sob análise[4]. Esta escolha é feita porque, em geral, divisões binárias são muito mais comuns que ternárias, assim como ritmos simples são mais comuns que compostos [19]. A melhor escolha, então, é selecionar valores iniciais que sejam representativos deste conhecimento prévio sobre essas variáveis.

Por fim, o período do tactus, que está associado ao andamento, pode ser inicializado de diversas maneiras. Caso ele seja inicializado com uma distribuição uniforme, o algoritmo de inferência ficará responsável por procurar o período que melhor modela os dados observados. No entanto, conforme visto na Parte I da tese, é possível obter-se estimativas confiáveis do andamento. Por isso, uma estratégia mais adequada seria inicializar o algoritmo com estas estimativas. Desta forma, a probabilidade prior para o período do tactus seria concentrada em poucos possíveis períodos iniciais, por exemplo: nos múltiplos e submúltiplos do andamento estimado pelo algoritmo descrito no Capítulo 4. Note que é possível adotar outras estratégias para a escolha dessas probabilidades e para a escolha dos candidatos.

5.2.5 Resumo do Modelo

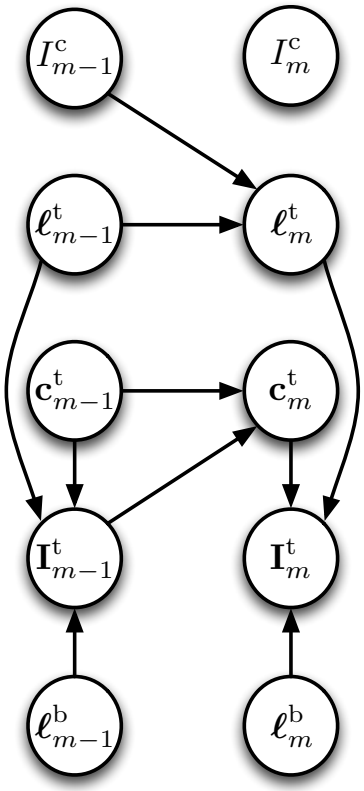
O modelo hierárquico proposto pode ser resumido através dos diagramas mostrados na Figura 5.6, onde se pode observar como as diferentes variáveis estão interligadas e quais variáveis de um determinado nível influenciam as variáveis de outro nível.

O modelo proposto consegue capturar diversos ritmos e rastrear a ocorrência dos três níveis hierárquicos. O custo desta característica, entretanto, é a grande quantidade de estados gerados pelo modelo, o que deixa a sua inferência muito complexa. Por exemplo, o número total de estados do modelo é:

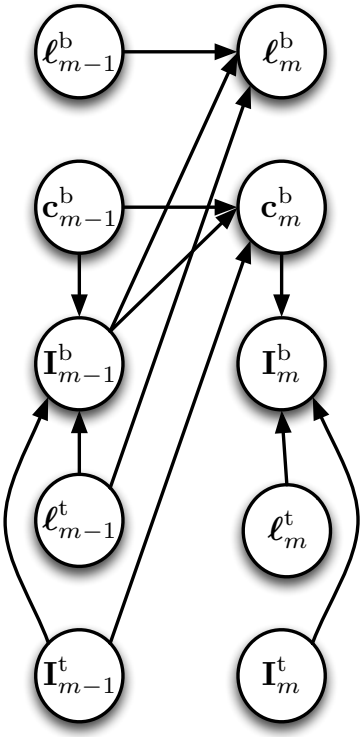
- Indicadores $\rightarrow$ 8 estados
- Contadores $\rightarrow 9(0{,}5\ell^{\max} + \sigma_t - 1)$ estados
- Períodos $\rightarrow 4(\ell^{\max} - \ell^{\min} + 1)$ estados
- Total $\rightarrow 288(0{,}5\ell^{\max} + \sigma_t - 1)(\ell^{\max} - \ell^{\min} + 1)$ estados.

Considerando valores típicos para os parâmetros $\ell^{\max}$ e $\ell^{\min}$, assumindo que um sinal pode possuir um andamento entre 40 BPM e 250 BPM, e assumindo-se que o fluxo espectral é calculado a cada 40 ms, o número total de estados seria 180576. Neste caso, a matriz de transição possuiria dimensão 180576×180576. Mesmo sendo a

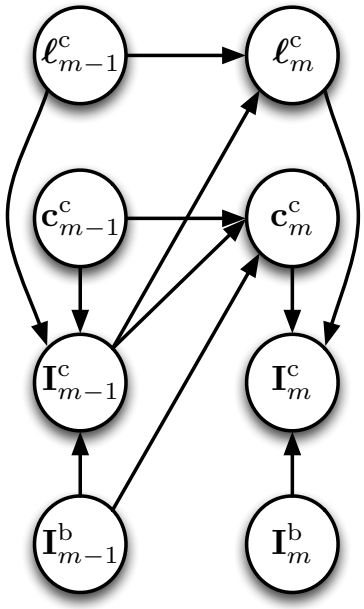
[4]Idealmente, esses valores são computados para uma partição do conjunto de sinais utilizada apenas para treinar o modelo.



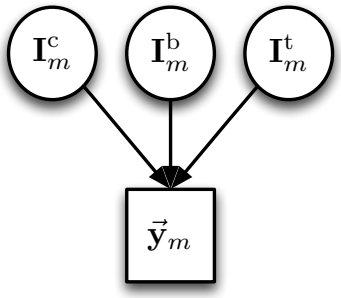
(a) Tatum



(b) Tactus



(c) Compasso



(d) Observação

Figura 5.6: Representação gráfica para cada nível do modelo hierárquico.

maior parte de seus elementos nulos, é alto o custo computacional para realizar inferências sobre o algoritmo.

Uma das soluções para reduzir a complexidade seria limitar os períodos máximo e mínimo a serem utilizados dependendo do andamento estimado do sinal. Considerando que há uma dependência quadrática do total de estados com $\ell^{\max}$, dessa forma o número total de estados pode ser consideravelmente reduzido. Em todo caso, nas próximas seções serão apresentados modelos simplificados que possuem um número reduzido de estados.

5.3 Modelo para Rastreamento do Tactus

Nesta seção, é apresentado um modelo para rastreamento apenas do tactus. Este modelo pode ser visto com uma simplificação do modelo anterior onde são mantidas apenas as VAs associadas ao nível hierárquico do tactus. Com isso, este modelo não utiliza nenhuma informação dos outros níveis hierárquicos para encontrar os quadros em que o tactus ocorreu.

O modelo apresentado nessa seção se assemelha ao descrito em [27], com ambos possuindo uma VA que conta o número de quadros que se passaram desde o último início de tactus. Aqui, no entanto, modela-se explicitamente o indicador do tactus (em [27] o indicador é modelado implicitamente, através de um estado especial) e inclui-se o rastreamento do período do tactus ao longo do sinal. A forma como a probabilidade do indicador do tatum é parametrizada também é diferente nos dois modelos.

As três variáveis relacionadas ao tactus do modelo hierárquico são mantidas neste modelo, não havendo mudanças na definição do indicador $\mathbf{I}_m^{\mathrm{b}} \in \{0, 1\}$ e do período $\boldsymbol{\ell}_m^{\mathrm{b}} \in \{\ell^{\min}, \ell^{\min}+1, \ldots \ell^{\max}\}$. Já a definição do contador precisa ser alterada, uma vez que o início de novo tactus será rastreado diretamente a partir do sinal e não indiretamente através dos tatums. Com isso, a nova definição do contador passa a ser $\mathbf{c}_m^{\mathrm{b}} \in \{0, 1, \ldots, \ell^{\max} + \sigma_{\mathrm{b}}\}$, semelhante à definição do contador de tatum do modelo hierárquico.

A probabilidade do contador no quadro m é definida da seguinte maneira neste modelo:

$$p_{\mathbf{c}_m^{\mathrm{b}}}(c_m^{\mathrm{b}} | c_{m-1}^{\mathrm{b}}, I_{m-1}^{\mathrm{b}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{b}} = c_{m-1}^{\mathrm{b}} + 1) \wedge (b_{m-1} = 0) \\ 1, & \text{se } (c_m^{\mathrm{b}} = 0) \wedge (I_{m-1}^{\mathrm{b}} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.44}$$

onde o contador é incrementado caso não tenha iniciado um tactus no quadro anterior e é zerado em caso contrário.

Para o indicador, será adotada uma probabilidade de transição similar à do indicador do tatum do modelo hierárquico. Com isso, a seguinte probabilidade foi adotada

$$p_{\mathbf{I}_m^{\mathrm{b}}}(I_m^{\mathrm{b}}|c_m^{\mathrm{b}},\ell_m^{\mathrm{b}}) = \begin{cases} \overline{w}[c_m^{\mathrm{b}} - \ell_m^{\mathrm{b}} + 1], & \text{se } I_m^{\mathrm{b}} = 1 \\ 1 - \overline{w}[c_m^{\mathrm{b}} - \ell_m^{\mathrm{b}} + 1], & \text{se } I_m^{\mathrm{b}} = 0, \end{cases} \tag{5.45}$$

onde a janela $\overline{w}[.]$ é análoga à janela utilizada no modelo hierárquico, porém possui comprimento igual a $2\sigma_{\mathrm{b}} + 1$, sendo σ_{b} uma variável que controla a tolerância a imprecisões no tactus. Novamente, a chance de iniciar um novo tactus no quadro m dependerá dos valores atuais do contador e do período: quanto mais próximo o contador incrementado de 1 estiver do período, maior a chance de novo tactus. Lembrando que o contador armazena o número de quadros após o último tactus, a expressão acima garante que o indicador permanece no estado 0 uma quantidade de quadros igual ao período. A transição do valor 1 para 0 em um quadro é garantida pelo contador, que é zerado no quadro seguinte à validação do indicador.

O período do tactus é atualizado a cada quadro para armazenar o valor passado do contador. Isso garante a modelagem de variações lentas no andamento, em que o período não pode aumentar (ou diminuir) de um valor maior que σ_b. Assim, a probabilidade do período no quadro m é

$$p_{\boldsymbol{\ell}_m^{\mathrm{b}}}(\ell_m^{\mathrm{b}}|\ell_{m-1}^{\mathrm{b}}, c_{m-1}^{\mathrm{b}}, I_{m-1}^{\mathrm{b}}) = \begin{cases} 1, & \text{se } \left(\ell_m^{\mathrm{b}} = c_{m-1}^{\mathrm{b}} + 1\right) \wedge \left(I_{m-1}^{\mathrm{b}} = 1\right) \\ 1, & \text{se } \left(\ell_m^{\mathrm{b}} = \ell_{m-1}^{\mathrm{b}}\right) \wedge \left(I_{m-1}^{\mathrm{b}} = 0\right) \\ 0, & \text{caso contrário.} \end{cases} \tag{5.46}$$

O modelo de observação do modelo de tactus será discutido em detalhes no Capítulo 7. Nele serão considerados os mesmos atributos do modelo hierárquico, porém dependendo apenas do indicador do tactus, podendo ser escrito como $P_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{b}})$. Note que contar apenas com a informação do tactus o faz necessariamente menos completo que o modelo de observação hierárquico, que utiliza informação de todos os níveis métricos.

A Figura 5.7 exibe a representação gráfica do modelo apresentado. O número total de estados do modelo para rastreamento do tactus pode ser obtido através de:

- Indicador $\rightarrow$ 2 estados
- Período $\rightarrow \ell^{\max} - \ell^{\min} + 1$
- Contador $\rightarrow \ell^{\max} + 1 + \sigma_{\mathrm{b}}$
- Total $\rightarrow 2(\ell^{\max} - \ell^{\min} + 1)(\ell^{\max} + 1 + \sigma_{\mathrm{b}})$

Considerando períodos entre 40 BPM e 250 BPM e o cálculo de atributos a cada 40 ms, têm-se 2640 estados no total. Deve-se notar que a maior parte destes estados

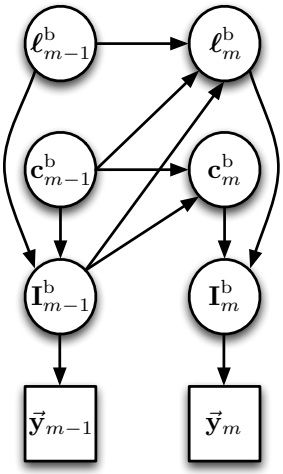


Figura 5.7: Representação gráfica para o modelo de rastreamento do tactus.

geram poucas transições, sendo o modelo esparso. Em todo caso, pode-se observar a redução drástica do número de estados em relação ao modelo hierárquico, levando a uma redução significativa na complexidade para se realizar inferências utilizando este modelo.

5.4 Modelo Hierárquico por Camadas

O modelo anterior reduz a complexidade do modelo hierárquico abrindo mão da busca simultânea dos três níveis métricos e concentrando-se apenas no rastreamento do tactus. O modelo proposto nesta seção procura rastrear os três níveis métricos de forma menos complexa que o modelo hierárquico através da sua divisão em dois submodelos. No primeiro, apenas o tatum é rastreado. No segundo, que utiliza a sequência de estados mais prováveis do primeiro submodelo, são modeladas apenas as informações do tactus e do compasso. Como os dois sub-modelos são desacoplados, o número de estados em cada um é reduzido. A seguir, cada submodelo será apresentado em separado.

5.4.1 Modelo de Rastreamento do Tatum

Neste modelo, apenas o tatum será rastreado. Para isso, a parcela do tatum do modelo hierárquico será despida da dependência dos outros níveis métricos. Com isso, as VAs utilizadas por esse modelo são as mesmas do tatum do modelo hierárquico. Considerando que o período do tatum não é separável do período do tactus, o período do tactus também será rastreado neste modelo. Isso permite a modelagem de variações lentas do andamento e de mudanças de divisões binárias para divisões ternárias do período do tactus. A seguir, será apresentado o modelo de transição

simplificado adotado para as variáveis (todas definidas como no modelo hierárquico.)

A transição do contador do tatum é idêntica à adotada no modelo hierárquico, sendo

$$p_{\mathbf{c}_m^{\mathrm{t}}}(c_m^{\mathrm{t}}|c_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{t}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{t}} = c_{m-1}^{\mathrm{t}} + 1) \wedge (I_{m-1}^{t} = 0) \\ 1, & \text{se } (c_m^{\mathrm{t}} = 0) \wedge (I_{m-1}^{t} = 1) \\ 0, & \text{caso contrário.} \end{cases} \tag{5.47}$$

A probabilidade do período do tatum é atualizada da forma abaixo, podendo mudar de uma divisão binária para ternária (ou vice-versa) com probabilidade σ_c:

$$p_{\boldsymbol{\ell}_m^{\mathrm{t}}}(\ell_m^{\mathrm{t}}|\ell_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{t}}) = \begin{cases} \sigma_{\mathrm{c}}, & \text{se } (\ell_m^{\mathrm{t}} \neq \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{t}} = 1) \\ 1 - \sigma_{\mathrm{c}}, & \text{se } (\ell_m^{\mathrm{t}} = \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{t}} = 1) \\ 1, & \text{se } (\ell_m^{\mathrm{t}} = \ell_{m-1}^{\mathrm{t}}) \wedge (I_{m-1}^{\mathrm{t}} = 0) \\ 0, & \text{caso contrário.} \end{cases} \tag{5.48}$$

Note que, como a informação de compasso foi removida do modelo, não há mais restrições para quando uma mudança de período do tatum pode ocorrer.

A probabilidade do período do tatum é atualizada da mesma forma que no período hierárquico, sendo

$$p_{\boldsymbol{\ell}_m^{\mathrm{b}}}(\ell_m^{\mathrm{b}}|\ell_{m-1}^{\mathrm{t}}, I_{m-1}^{\mathrm{t}}) = \begin{cases} 1, & \text{se } \left(\ell_m^{\mathrm{b}} = \frac{c_{m-1}^{\mathrm{t}}}{\ell_{m-1}^{\mathrm{t}}}\right) \wedge (I_{m-1}^{\mathrm{t}} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.49}$$

onde a única diferença é que o período agora pode ser atualizado a cada tatum, no lugar de a cada tactus, já que apenas o indicador do tatum está disponível.

A probabilidade de se observar um tactus também é atualizada da mesma forma que no modelo hierárquico:

$$p_{\mathbf{I}_m^{\mathrm{t}}}(I_m^{\mathrm{t}}|c_m^{\mathrm{t}}, \ell_m^{\mathrm{t}}, \ell_m^{\mathrm{b}}) = \begin{cases} \overline{w}[c_m^{\mathrm{t}} - \mathrm{round}(\ell_m^{\mathrm{b}}\ell_m^{\mathrm{t}}) + 1], & \text{se } I_m^{\mathrm{t}} = 1 \\ 1 - \overline{w}[c_m^{\mathrm{t}} - \mathrm{round}(\ell_m^{\mathrm{b}}\ell_m^{\mathrm{t}}) + 1], & \text{se } I_m^{\mathrm{t}} = 0, \end{cases} \tag{5.50}$$

onde a janela $\overline{w}[.]$ é idêntica à utilizada no modelo hierárquico.

O modelo de observação para o rastreamento do tactus é descrito através de $P_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}})$, onde se considera apenas se o quadro em questão possui informação rítmica ou não. Serão descritas no Capítulo 7 estratégias para a obtenção destas probabilidades.

Pode ser vista na Figura 5.8, uma representação gráfica do modelo para rastreamento do tatum. O número total de estados deste modelo pode ser calculado da

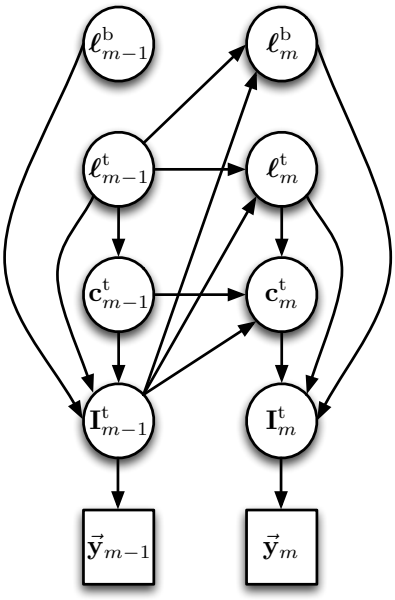


Figura 5.8: Representação gráfica do modelo para rastreamento de tatum utilizado no modelo por camadas.

seguinte forma:

- Indicador: 2 estados;
- Contador: $(0{,}5\ell^{\max} + \sigma_t - 1)$ estados;
- Períodos: $2\left(\ell^{\max} - \ell^{\min} + 1\right)$;
- Total: $4(0{,}5\ell^{\max} + \sigma_t - 1)\left(\ell^{\max} - \ell^{\min} + 1\right)$.

Considerando períodos para o tactus entre 40 BPM e 250 BPM e o cálculo de atributos a cada 40 ms, tem-se 1254 estados no total. Note que, ao se rastrear o tatum, o número total de estados é menor que no modelo de rastreamento do tactus. Isso acontece devido ao número reduzido de estados do contador do tatum, que é, aproximadamente, metade do número necessário para se contar o tactus.

Será aplicado sobre o modelo descrito o algoritmo de Viterbi para se obter a sequência de estados mais provável para um determinado sinal. Essa sequência irá determinar um novo índice temporal $\overline{m}$, que está associado apenas aos quadros em que iniciou novo tatum na sequência de estados estimadas. Tal vetor pode ser criado através de $\overline{m} \in \{m \mid \hat{I}^{\mathrm{t}}_m = 1\}$, onde $\hat{I}^{\mathrm{t}}_m$ é o valor mais provável estimado pelo algoritmo de Viterbi para I^{t}_m.

5.4.2 Modelo de Rastreamento Métrico

O segundo modelo da solução por camadas consiste em identificar, dentre os quadros $\overline{m}$ que foram associados a novo tatum estimado, quais também são de novo tactus

e, dentre estes, quais são de novo compasso. Por isso, será chamado de modelo de rastreamento métrico, pois é responsável em identificar apenas as informações métricas. Com isso, a saída da aplicação do modelo por camadas seria a informação métrica completa, porém estimada em duas etapas, cada etapa com complexidade computacional reduzida.

As VAs incluídas neste modelo são as VAs do modelo hierárquico que não foram utilizadas no modelo de rastreamento por tatum. Dessa forma, este modelo possui cinco VAs: os indicadores do tactus e de compasso, os contadores do tactus e do compasso e o período do compasso. Todas as variáveis são definidas da mesma forma que no modelo hierárquico. Será também utilizado o andamento estimado para o tatum $\hat{\mathbf{l}}^{\mathrm{t}}_{\overline{m}}$, porém aparecendo apenas como variável observada.

As probabilidades de transição para as variáveis do tactus neste modelo são obtidas modificando-se as variáveis do modelo hierárquico de forma que restem apenas os casos em que $\mathbf{I}^{\mathrm{t}}_{\overline{m}} = 1$. Como as variáveis do compasso não dependem de $\mathbf{I}^{\mathrm{t}}_{\overline{m}} = 1$, não é necessário modificá-las. Já a probabilidade de transição do indicador do tactus fica igual a

$$p_{\mathbf{I}^{\mathrm{b}}_{\overline{m}}}(I^{\mathrm{b}}_{\overline{m}}|c^{\mathrm{b}}_{\overline{m}}, \ell^{\mathrm{t}}_{\overline{m}}) = \begin{cases} 1, & \text{se } (I^{\mathrm{b}}_{\overline{m}} = 1) \wedge \left(c^{\mathrm{b}}_{\overline{m}} = \frac{1}{\ell^{\mathrm{t}}_{\overline{m}}} - 1\right) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.51}$$

e o contador do tactus igual a

$$p_{\mathbf{c}^{\mathrm{b}}_{\overline{m}}}(c^{\mathrm{b}}_{\overline{m}}|c^{\mathrm{b}}_{\overline{m}-1}, I^{\mathrm{b}}_{\overline{m}-1}) = \begin{cases} 1, & \text{se } (c^{\mathrm{b}}_{\overline{m}} = c^{\mathrm{b}}_{\overline{m}-1} + 1) \wedge (I^{\mathrm{b}}_{\overline{m}-1} = 0) \\ 1, & \text{se } (c^{\mathrm{b}}_{\overline{m}} = 0) \wedge (I^{\mathrm{b}}_{\overline{m}-1} = 1) \\ 0, & \text{caso contrário.} \end{cases} \tag{5.52}$$

O modelo de observação neste caso depende apenas dos indicadores do tactus e do compasso, sendo escrito como $P_{\vec{\mathbf{y}}_{\overline{m}}}(\vec{y}_{\overline{m}}|I^{\mathrm{b}}_{\overline{m}}, I^{\mathrm{c}}_{\overline{m}})$. Deve-se notar que este modelo, que será melhor descrito no Capítulo 7, precisa apenas considerar como distinguir um vetor de observações que é de um novo tatum de um que é de novo tatum e novo tactus, e de outro que é de novo tatum, novo tactus e novo compasso.

A Figura 5.9 exibe uma representação gráfica do modelo métrico. Considerando que as VAs deste modelo podem assumir poucos valores, este modelo apresenta um baixo número de estados, sendo 72 (4 dos Indicadores, 9 dos contadores e 2 do período) no total. Note que este modelo não é esparso, porém opera sobre um subconjunto dos quadros do sinal. Com isso, esse modelo é de baixa complexidade computacional quando comparado aos modelos vistos até o momento. O preço desta baixa complexidade, no entanto, é a sua alta dependência da qualidade das estimativas providas pelo rastreador do tatum.

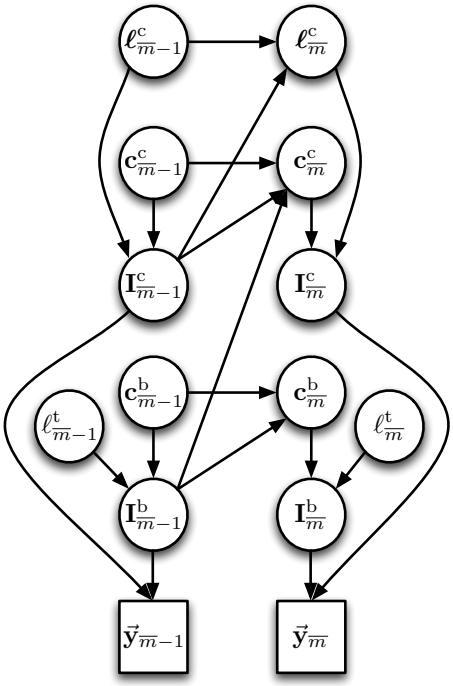


Figura 5.9: Representação do modelo métrico utilizado no modelo por camadas.

5.5 Modelo por Padrão Rítmico

O modelo hierárquico e suas simplificações modelam a estrutura métrica de uma música de forma construtiva, estimando cada nível hierárquico sem assumir nenhum padrão pré-determinado para eles a não ser os temporais. O modelo apresentado nessa seção, no entanto, irá assumir que a estrutura métrica da peça é definida por um conjunto de padrões. Essa metodologia pode ser entendida como dedutiva, em que uma sequência de acentuações, definidas por um padrão rítmico, é buscada dentro da peça. Esse tipo de modelo é especialmente adequado para sinais que apresentam métricas não cobertas pelo modelo hierárquico (por exemplo, métricas ímpares) ou para músicas exibindo ritmos compostos, cujos compassos podem conter um grande número de tatums.

O modelo apresentado nesta seção pode ser considerado uma adaptação do modelo originalmente descrito em [93] e posteriormente utilizado em [94, 95]. No modelo descrito nesta seção, no entanto, os padrões serão definidos a partir do tatum, sendo cada unidade do padrão associada à percepção da métrica, ao passo que em [93] os padrões são definidos em função de um contador de tempo arbitrário, que não está necessariamente associado a um nível métrico. Como será visto nos próximos capítulos, essa diferença facilita a definição dos padrões rítmicos, que se tornam mais facilmente interpretáveis. Isso também permite que o padrão rítmico utilizado no modelo proposto contenha toda a informação da métrica e dos relacionamentos

entre os períodos do tatum, do tactus e do compasso, enquanto em [93] uma VA extra precisa ser adicionada para a modelagem desses relacionamentos. Além disso, o rastreamento do andamento é feito diretamente sobre o tatum, e não de forma indireta como feito em [93].

A seguir, será apresentado o modelo por padrão rítmico, que, como o modelo hierárquico, também extrai a informação dos três níveis métricos do sinal de áudio. Esta informação estará contida nos padrões rítmicos utilizados para buscar o sinal, como ficará mais claro nas próximas seções.

5.5.1 Padrões Rítmicos

Nesta seção, os padrões rítmicos serão matematicamente definidos. Será dada uma breve explicação do tipo da informação contida em cada padrão, porém não será discutida a forma com que os padrões podem ser obtidos, tema que será abordado no Capítulo 8.

O i-ésimo padrão rítmico será definido pelas seguintes informações

- O conjunto $A_i = \{\mathcal{A}_{i,0}, \mathcal{A}_{i,1}, \ldots, \mathcal{A}_{i,L_i-1}\}$ onde $\mathcal{A}_{i,j}$ é um parâmetro que define a observação e L_i é o comprimento do padrão em tatums;
- Um multiplicador para o andamento $\Theta_i \in \{1/2, 1/3, 1/4, \ldots\}$.

O padrão, portanto, é definido em função dos tatums, sendo que o j-ésimo elemento do conjunto A_i define como o j-ésimo início de tatum deve ser observado. Por exemplo, $\mathcal{A}_{i,j}$ pode determinar se um determinado início de tatum dentro do padrão é acentuado ou não. Neste caso, o parâmetro armazenado estaria associado ao valor médio que determina uma acentuação no vetor de atributos. Parâmetros mais complexos, que incluem informação espectral (além de dinâmica) também podem ser criados. Em todo caso, o tipo da informação armazenada no padrão é transparente ao modelo: ele irá influenciar o resultado de uma inferência, mas não como é realizada a inferência. O comprimento do padrão L_i, usualmente, determina um compasso, e dentro de um mesmo padrão, mais de um elemento pode estar associado à ocorrência de um tactus. Por isso, para permitir a transição entre diferentes padrões, mantendo-se o período do tactus constante, é necessário saber a taxa de tactus em função do número de tatums de cada período. Esse valor é armazenado em Θ_i.

5.5.2 Variáveis Aleatórias

São definidas a seguir, as variáveis utilizadas no modelo por padrão rítmico. Como o padrão rítmico é definido em função do tatum, será utilizada uma parcela do

modelo de rastreamento de tatum. Já para rastrear o período, será usada uma variável similar ao período do tactus do modelo hierárquico. As VAs do modelo por padrão rítmico serão denotadas por um superíndice r.

Considerando, então, que é necessário rastrear o tatum dentro do modelo por padrão rítmico, serão utilizados um indicador para o tatum definido como $\mathbf{I}_m^{\mathrm{r}} \in \{0, 1\}$ e um contador definido por $\mathbf{c}_m^{\mathrm{r}} \in \{0, \ldots, \ell^{\max} \max_i(\Theta_i) + \sigma_t - 1\}$. Essas variáveis modelarão o mesmo que suas análogas no modelo de rastreamento do tatum. Note que o limite do contador agora é determinado pelo maior multiplicador dentre todos os padrões rítmicos sendo utilizados. Isso garante que o valor máximo do contador é grande o suficiente para rastrear o pior caso em termos de período e padrão rítmico. Como será visto a seguir, σ_t é uma constante com interpretação similar à constante de mesmo nome do modelo de rastreamento do tatum.

O período do padrão rítmico (associado ao andamento do sinal sob análise) é definido como $\boldsymbol{\ell}_m^{\mathrm{r}} \in \{\ell^{\min}, \ldots, \ell^{\max}\}$, onde $\ell^{\min}$ e $\ell^{\max}$ correspondem ao maior e ao menor andamentos a serem rastreados em amostras, respectivamente.

Uma variável será responsável por modelar qual ponto do padrão rítmico está associado a um determinado quadro m. Para isso, a variável é definida como $\mathbf{j}_m^{\mathrm{r}} \in \{0, \ldots, \max_i L_i\}$, cujo valor máximo é definido pelo maior comprimento dentre todos os padrões. Novamente, esta variável associa um quadro m a uma posição dentro do padrão, lembrando que o padrão é definido em função dos tatums. Com isso, essa variável acaba por contar quantos tatums ocorreram após o início do padrão ter sido detectado.

Por fim, a última variável a ser escolhida é o seletor do padrão rítmico, definido como $\mathbf{a}_m^{\mathrm{r}} \in \{0, \ldots, N_A - 1\}$, onde N_A é o número de padrões. Esta variável é responsável por informar qual padrão rítmico está ativo em um determinado quadro.

5.5.3 Modelo de Transição

Nesta seção, será apresentado o modelo de transição do modelo por padrão rítmico.

A probabilidade do contador do tatum do padrão pode ser escrita da seguinte forma

$$p_{\mathbf{c}_m^{\mathrm{r}}}(c_m^{\mathrm{r}} | c_{m-1}^{\mathrm{r}}, I_{m-1}^{\mathrm{r}}) = \begin{cases} 1, & \text{se } (c_m^{\mathrm{r}} = c_{m-1}^{\mathrm{r}} + 1) \wedge (I_{m-1}^{\mathrm{r}} = 0) \\ 1, & \text{se } (c_m^{\mathrm{r}} = 0) \wedge (I_{m-1}^{\mathrm{r}} = 1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.53}$$

onde fica claro que, dadas as informações do indicador e do contador no quadro $m-1$, o valor do contador no quadro m é determinístico. Note que a probabilidade do contador se concentra em três casos. No primeiro caso, quando não ocorre início de tatum no quadro anterior, o valor do contador é incrementado. No segundo,

quando ocorre início de tatum no quadro anterior, o valor do contador é zerado. Em qualquer outra situação, a probabilidade de o contador assumir qualquer valor é igual a zero.

A probabilidade do período no quadro m foi definida como

$$p_{\boldsymbol{\ell}^{\mathrm{r}}_m}(\ell^{\mathrm{r}}_m|c^{\mathrm{r}}_{m-1},\ell^{\mathrm{r}}_{m-1},I^{\mathrm{r}}_{m-1},a^{\mathrm{r}}_{m-1}) = \begin{cases} 1, & \text{se } (\ell^{\mathrm{r}}_m = \ell^{\mathrm{r}}_{m-1}) \wedge (I^{\mathrm{r}}_{m-1}=0) \\ 1, & \text{se } \left(\ell^{\mathrm{r}}_m = \frac{c_{m-1}}{\Theta_{a^{\mathrm{r}}_{m-1}}}\right) \wedge (I^{\mathrm{r}}_{m-1}=1) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.54}$$

onde foi utilizada a mesma abordagem utilizada no modelo de rastreamento do tatum, em que se utiliza o valor do contador ao se detectar novo tatum para se atualizar o período do tatum. Desta forma, é possível rastrear variações graduais do andamento.

No caso do indicador, foi adotada a mesma forma de descrever a probabilidade dos modelos previamente apresentados:

$$p_{\mathbf{I}^{\mathrm{r}}_m}(I^{\mathrm{r}}_m|c^{\mathrm{r}}_m,\ell^{\mathrm{r}}_m,a^{\mathrm{r}}_m) = \begin{cases} \overline{w}[c^{\mathrm{r}}_m - \mathrm{round}(\Theta_{a^{\mathrm{r}}_m}\ell^{\mathrm{r}}_m)+1], & \text{se } I^{\mathrm{r}}_m = 1 \\ 1-\overline{w}[c^{\mathrm{r}}_m - \mathrm{round}(\Theta_{a^{\mathrm{r}}_m}\ell^{\mathrm{r}}_m)+1], & \text{se } I^{\mathrm{r}}_m = 0, \end{cases} \tag{5.55}$$

onde $w[.]$ é uma janela normalizada de comprimento igual a $2\sigma_t + 1$ utilizada para parametrizar imprecisões sobre o tempo de ocorrência do início de tatum.

A probabilidade de se estar numa determinada posição dentro de um dado padrão rítmico no quadro m pode ser escrita como

$$p_{\mathbf{j}^{\mathrm{r}}_m}(j^{\mathrm{r}}_m|j^{\mathrm{r}}_{m-1},a^{\mathrm{r}}_{m-1},I^{\mathrm{r}}_{m-1}) = \begin{cases} 1, & \text{se } (j^{\mathrm{r}}_m = j^{\mathrm{r}}_{m-1}+1) \wedge (j^{\mathrm{r}}_{m-1}+1 \neq L_{a^{\mathrm{r}}_{m-1}}) \wedge (I^{\mathrm{r}}_{m-1}=1) \\ 1, & \text{se } (j^{\mathrm{r}}_m = 0) \wedge (j^{\mathrm{r}}_{m-1}+1 = L_{a^{\mathrm{r}}_{m-1}}) \wedge (I^{\mathrm{r}}_{m-1}=1) \\ 1, & \text{se } (j^{\mathrm{r}}_m = j^{\mathrm{r}}_{m-1}) \wedge (I^{\mathrm{r}}_{m-1}=0) \\ 0, & \text{caso contrário.} \end{cases} \tag{5.56}$$

Será discutido separadamente cada um dos casos acima. No primeiro caso, se um novo tatum foi iniciado no quadro $m-1$ e se a posição não é a última dentro do padrão atual, então a posição dentro do padrão é incrementada. No segundo caso, iniciou no quadro $m-1$ um novo tatum e o padrão estava em sua última posição; então, a posição deve retornar para o início do padrão. No terceiro caso, iniciou novo tatum; logo, a posição não deve ser alterada. Por fim, é nula a probabilidade de o valor não ser nenhum dos especificados anteriormente. Similarmente ao contador, a posição do padrão no quadro m é conhecida de forma determinística, se forem conhecidos os valores da posição no quadro, do índice do padrão e o do indicador no quadro $m-1$.

Resta definir o índice que determina o padrão ativo no quadro m. Para isto, será considerado que só pode haver mudança de padrão quando o padrão anterior tiver terminado (similarmente à escolha realizada em [93]). Isso evita mudanças abruptas, que interromperiam um padrão no meio e que são pouco comuns nos sinais sendo analisados. Com isso, a probabilidade de se encontrar um determinado índice no quadro m pode ser escrita como

$$p_{\mathbf{a}_m^{\mathrm{r}}}(a_m^{\mathrm{r}}|a_{m-1}^{\mathrm{r}}, j_{m-1}^{\mathrm{r}}) = \begin{cases} \frac{\sigma_r}{N_A}, & \text{se } (a_m^{\mathrm{r}} \neq a_{m-1}^{\mathrm{r}}) \wedge (j_{m-1}^{\mathrm{r}} - 1 = L_{a_{m-1}^{\mathrm{r}}}) \\ 1 - \sigma_r, & \text{se } (a_m^{\mathrm{r}} = a_{m-1}^{\mathrm{r}}) \wedge (j_{m-1}^{\mathrm{r}} - 1 = L_{a_{m-1}^{\mathrm{r}}}) \\ 1, & \text{se } (a_m^{\mathrm{r}} = a_{m-1}^{\mathrm{r}}) \wedge (j_{m-1}^{\mathrm{r}} - 1 \neq L_{a_{m-1}^{\mathrm{r}}}) \\ 0, & \text{caso contrário,} \end{cases} \tag{5.57}$$

onde σ_r é uma constante que determina a probabilidade de se permanecer com o mesmo padrão e N_A é o número de padrões rítmicos do modelo. Essa descrição assume que a probabilidade de cada padrão é a mesma. No entanto, seria simples modificar a equação acima de forma a considerar probabilidades individualizadas para cada padrão, com a única restrição de que a soma destas probabilidades fosse σ_r. Dessa forma, as transições para alguns padrões poderiam ser mais prováveis que para outros.

5.5.4 Observação

O modelo de observação do modelo por padrão rítmico pode ser descrito em dois casos distintos. No primeiro, deve ser obtido um modelo para quando não há informação rítmica presente no sinal. Este modelo seria similar ao caso em que o início de um tatum não é observado nos modelos anteriores. Já quando o início de um novo tatum é observado, a distribuição dos atributos deve ser fornecida por cada padrão rítmico.

Matematicamente, pode-se descrever o modelo de observação como

$$P_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{r}}, a_m^{\mathrm{r}}, j_m^{\mathrm{r}}), \tag{5.58}$$

onde a densidade de probabilidade dos atributos é determinada de duas formas:

- Quando $I_m^{\mathrm{r}} = 0$, é utilizado um modelo para ausência de informação rítmica para se calcular a probabilidade de se ter observado os dados;
- Quando $I_m^{\mathrm{r}} = 1$, é utilizado um modelo parametrizado pelo conteúdo do conjunto $\mathcal{A}_{a_m,j_m}$ para se calcular a probabilidade de se ter observado os dados.

Note que, diferentemente do modelo hierárquico, onde os modelos utilizam apenas a informação de se o quadro marca o início de um tatum, um início de tactus ou

um início de compasso, o modelo por padrão rítmico utiliza também a informação da posição dentro de um compasso. Isso permite uma modelagem mais complexa, que pode descrever um padrão de acentuação pouco comum. Por exemplo, imagine um ritmo em que nem todos os inícios de tactus são acentuados (ao contrário do que ocorre na maior parte da música popular). Tal ritmo seria de difícil análise usando o modelo hierárquico, pois este não tem como diferenciar um tactus do outro. Já o modelo por padrão rítmico poderia contemplar este ritmo através de um padrão específico. O custo desta flexibilidade é o crescimento do número de padrões utilizados, o que pode deixar o modelo muito complexo. Por isso, este modelo é considerado mais adequado para gêneros específicos, em que o número de padrões rítmicos é limitado. No Capítulo 7, será feita uma descrição de como os modelos para um padrão rítmico podem ser obtidos.

5.5.5 Prior

As probabilidades priores para as variáveis associadas ao rastreamento do tatum podem seguir escolhas similares às realizadas para as suas análogas do modelo hierárquico.

A probabilidade prior do padrão rítmico considera o fato de que, usualmente, pode-se inicializar uma música apenas em determinados pontos do ritmo. Em particular, é muito comum se começar no início do padrão ou no meio do padrão (em anacruse). Com isso, a prior da posição dentro do padrão pode ser inicializada de forma a refletir isso.

No caso da escolha da prior do índice, pode-se considerar a taxa de ocorrência de cada um dos padrões dentro de um conjunto de dados. Desta forma, é selecionada uma prior que reflita a chance média de se encontrar um determinado padrão. No caso de gêneros específicos, em que há informação prévia de quais padrões podem começar uma música, a prior pode ser escolhida de forma a refletir essa informação.

5.5.6 Resumo

Pode ser vista na Figura 5.10 uma representação gráfica do modelo por padrão rítmico. Considerando que o modelo rítmico é altamente dependente dos padrões escolhidos, uma análise mais detalhada do modelo será feita após a apresentação de alguns modelos extraídos a partir de um banco de sinais. Além disso, para demonstrar a flexibilidade do modelo por padrão rítmico de se adaptar a ritmos complexos, é realizado um estudo de caso utilizando o Candombe uruguaio (um polirritmo em que cada compasso é composto por 16 tatums, sendo que dos 4 tactus, apenas 1 é acentuado).

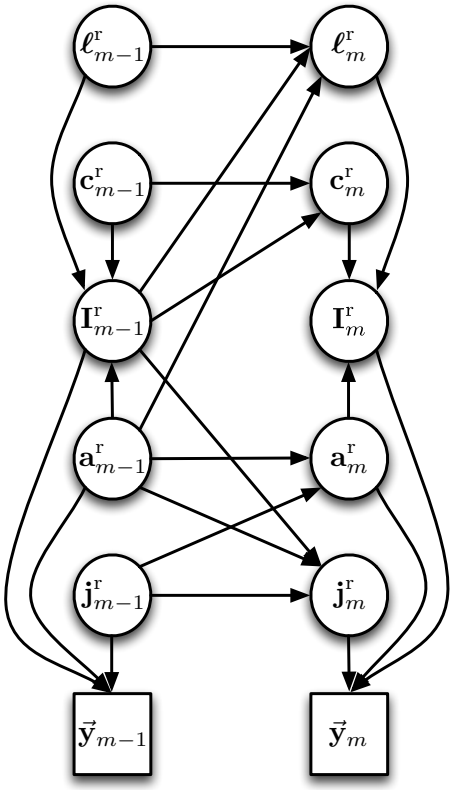


Figura 5.10: Representação do modelo por padrão rítmico.

5.6 Conclusão

Neste capítulo foram apresentados modelos ocultos de Markov para o rastreamento da métrica de sinais de música. Após uma breve descrição de modelos ocultos de Markov e de algoritmos de inferência, foram apresentados dois modelos principais, o modelo hierárquico e o modelo por padrão rítmico. Também foram descritas duas simplificações do modelo hierárquico que reduzem a sua complexidade.

Nos próximos capítulos, serão descritos o banco de sinais utilizados e modelos de observação, que associam as variáveis aleatórias utilizadas nos modelos descritos com os atributos extraídos do sinal de áudio. Uma vez descritos os modelos de observação, será avaliado o desempenho dos algoritmos propostos para um conjunto de sinais, considerando os diferentes modelos e possíveis escolhas dos seus parâmetros livres.

Capítulo 6

Banco Métrico

Neste capítulo, é descrita uma nova anotação para um dos bancos de sinais utilizados na Parte I desta tese. Esta anotação tem como objetivo permitir o treinamento e a avaliação dos modelos para análise rítmica apresentados no capítulo anterior, fornecendo um conjunto de dados em que são conhecidos os três níveis métricos. Na literatura são encontradas descrições de bancos de sinais que possuem o andamento anotado, apenas o tactus ou o tactus e o compasso, porém não é do conhecimento do autor o relato de um banco de sinais com a anotação de tatum, tactus e compasso. Sendo assim, esse banco consiste num passo importante para a criação de modelos que procuram extrair a informação rítmica dos três níveis métricos de um sinal de áudio. Outra vantagem do banco é que ele foi anotado por um mesmo músico profissional, garantindo a consistência das escolhas realizadas durante o processo de anotação.

Neste capítulo, também são descritas análises realizadas sobre os sinais anotados que procuram corroborar escolhas feitas nos modelos propostos. Tal análise tanto facilitará a interpretação de resultados obtidos utilizando-se esse banco como pode guiar a escolha de alguns parâmetros dos modelos para análise rítmica.

Este capítulo está organizado da seguinte forma. Na Seção 6.1 é feita uma descrição do banco escolhido e sua escolha é justificada. Já na Seção 6.2, é detalhado o processo de anotação. O andamento e o tactus anotados são analisados na Seção 6.3, o tatum na Seção 6.4, o compasso na Seção 6.5. Por fim, o capítulo é concluído na Seção 6.6.

6.1 Visão Geral

Nesta seção, será realizada uma visão geral do banco de sinais com foco na composição do banco escolhido. Uma versão resumida das informações desta seção foi apresentada na Seção 3.1.

O banco de sinais escolhido foi originalmente criado por Stephen Hainsworth [78]

para sua tese de doutorado. Este banco consiste de sinais de diferentes durações e gêneros, e na sua versão original possui a anotação do andamento dos sinais realizada pelo próprio autor. Os sinais desse banco foram escolhidos por possuírem uma duração maior que as dos sinais dos outros bancos utilizados. Além disso, o banco tem sinais de diferentes gêneros musicais e graus de dificuldade de análise. Todos os sinais do banco foram amostrados a 44,1 kHz com uma precisão de 16 bits.

Pode ser vista na Figura 6.1 a distribuição da duração dos sinais do banco escolhido. Como pode ser observado, a maior parte dos sinais possui duração entre 40 e 60 s, sendo a duração média 53 s. O sinal com menor duração possui 12 s e o de maior duração, 96 s. Conforme mencionado, o fato de a maior parte dos sinais desse banco possuírem duração acima de 30 s foi um forte motivador para a escolha desse banco para a anotação.

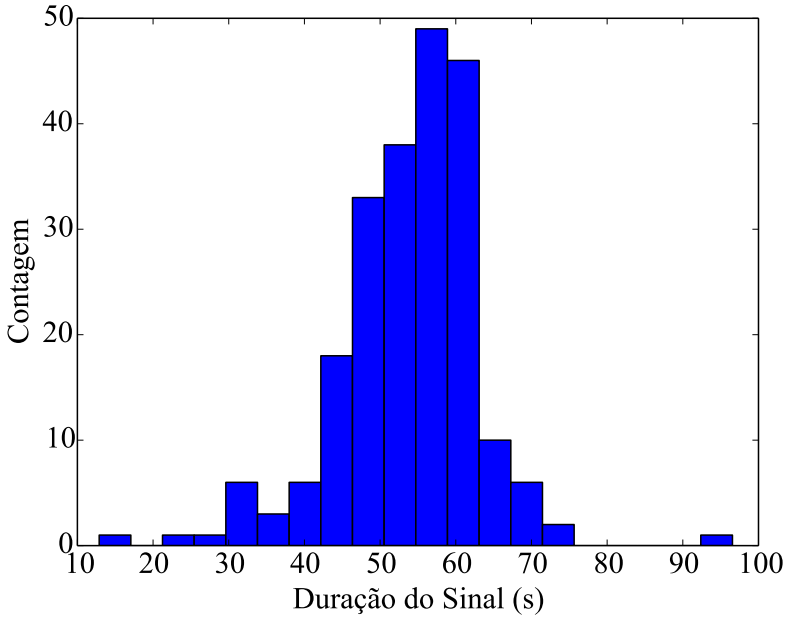


Figura 6.1: Distribuição da duração dos sinais no banco métrico.

Os sinais possuem o seu gênero anotado, sendo que a classificação original foi alterada de acordo com o esquema descrito no Anexo A. Na Tabela 6.1, é exibido o número de sinais em cada gênero. Deve ser ressaltado que no caso de música clássica, alguns sinais de música coral não exibem uma estrutura métrica regular, não se enquadrando completamente na classe de sinais de interesse desta tese. De qualquer forma, a presença desses sinais permite a análise do desempenho dos algoritmos propostos para sinais que não obedecem as hipóteses feitas. Por fim, esses sinais são minoria no banco, sendo a maior parte do banco composta por sinais de gêneros que exibem boa regularidade rítmica.

Tabela 6.1: Número de sinais de cada gênero no banco métrico.

Gênero	Número de sinais
Rock/Pop	68
Jazz	40
Eletrônica	40
Folk	22
Clássica	51

6.2 Anotação

A anotação dos níveis métricos foi feita por um músico profissional que possui vasta experiência em transcrição musical. Os três níveis métricos foram marcados para cada sinal, sendo que houve cuidado de se manter a coerência entre os critérios adotados para cada nível métrico e para cada sinal. De forma geral, foi pedido que fossem marcados os inícios dos tatums, tactus e compassos percebidos nos sinais.

A marcação do tatum foi realizada com precisão de 100 ms e cada início de tatum foi marcado como sendo, possivelmente, também de tactus e de compasso. As anotações foram salvas em tabelas que, posteriormente, foram importadas para o código gerado nesta tese.

As marcações de cada sinal foram validadas pelo autor da tese através da audição das marcações juntamente com o sinal de áudio e de inspeção visual de figuras do sinal com as marcações sobrepostas. Para tal, foi gerado um *script* que permite a audição de qualquer sinal somado ao sinal de um *cowbell* sendo executado na marcação do tatum, tactus ou compasso.

6.3 Análise do Tactus e do Andamento

Nesta seção, serão analisados o tactus anotado e o seu período. O objetivo é verificar os valores anotados, de forma a se obter um conhecimento maior sobre esses parâmetros. Um foco maior será posto na variação temporal do andamento num mesmo sinal.

O andamento de cada sinal foi obtido como a mediana do intervalo entre tactus consecutivos. Na Figura 6.2, podem ser vistos os andamentos anotados para os sinais. O menor andamento observado foi de 50 BPM, o maior de 260 BPM e o médio de 120 BPM. Deve-se notar que a distribuição dos andamentos segue a das curvas de ponderação adotadas na estimação do andamento, com uma concentração de sinais cujo andamento fica entre 100 e 150 BPM.

O andamento do banco escolhido já havia sido marcado anteriormente. Assim

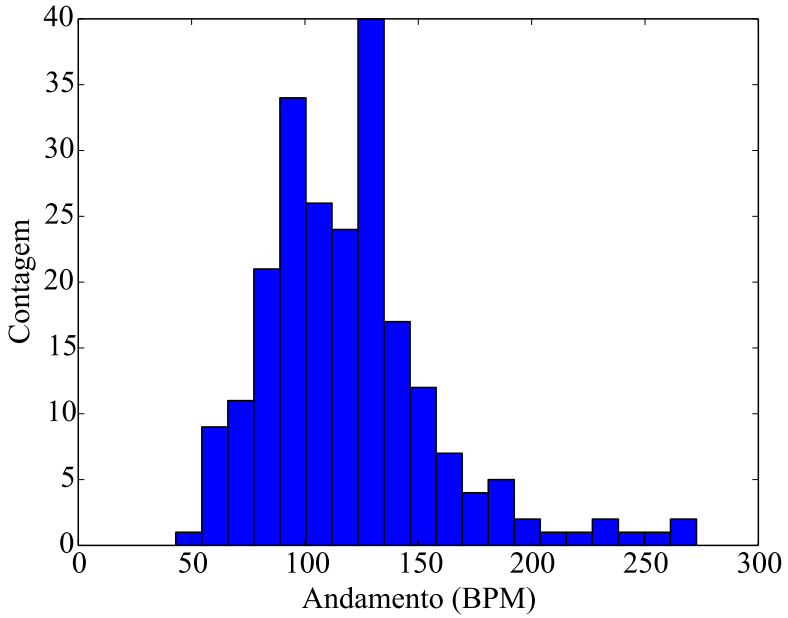


Figura 6.2: Histograma dos andamentos obtidos a partir do tactus anotado.

sendo, é pertinente uma comparação entre a marcação anterior e a atual. Na Figura 6.3 pode-se observar a diferença relativa entre o andamento atual e o antigo. No histograma, são exibidos apenas os andamentos cuja diferença relativa é menor que 20 %, cenário que abrange 197 dos 221 sinais presentes no banco. Dentre os sinais com discrepâncias de andamento superiores a 20 %, a diferença consiste em estimativas no dobro (6 sinais), na metade (14 sinais) e em um terço (1 sinal) do andamento antigo. Essa diferença mostra que o tactus foi percebido de forma diferente pelas duas pessoas que marcaram os bancos, o que não é raro. Deve-se sempre lembrar que a marcação realizada reflete apenas a opinião de uma pessoa, mesmo se tratando de um músico profissional com vasta experiência em transcrição de sinais de música.

Outra informação importante é a variação do período do tactus ao longo de um mesmo sinal. Para isso, foi calculada a diferença entre o maior e o menor períodos de tactus para um mesmo sinal, apresentada na Figura 6.4. Como pode ser visto, a maior diferença observada é de 200 ms para a maior parte dos sinais, indicando que o andamento fica em torno de um valor médio para a maior parte deles. Deve-se notar, no entanto, que para 25 sinais é observada uma grande diferença; dentre eles, 19 são músicas clássicas cujo andamento varia consideravelmente ao longo do tempo. Nos demais casos, são observadas mudanças abruptas de andamento, que contrariam a hipótese realizada no capítulo anterior sobre mudanças suaves no andamento. Esses sinais podem, então, ser utilizados para analisar como os métodos propostos se comportam sob tais condições.

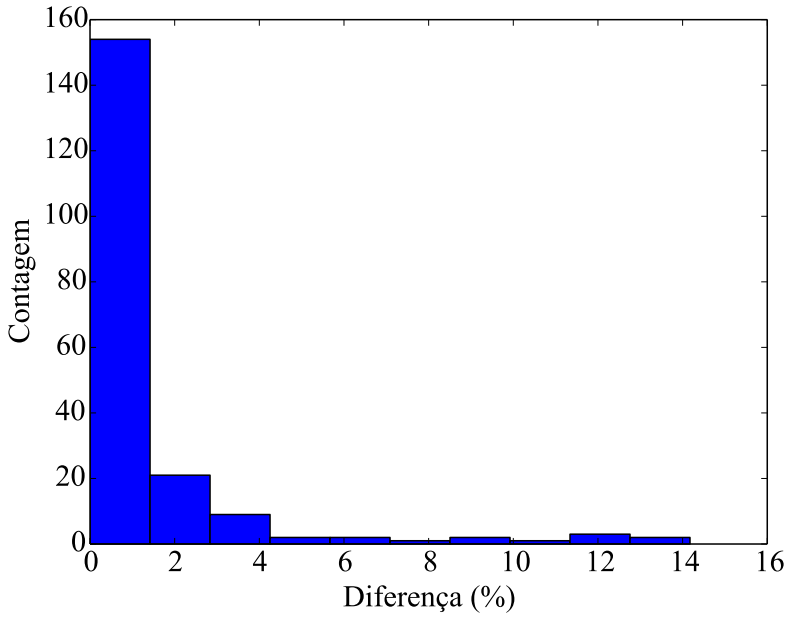


Figura 6.3: Diferença percentual entre os andamentos extraídos a partir do tactus e os andamentos anotados anteriormente no banco.

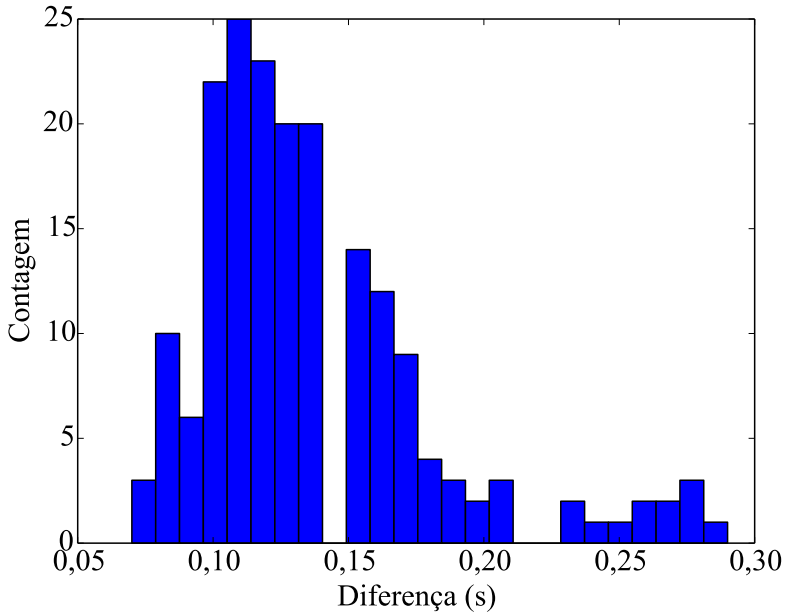


Figura 6.4: Histograma para a diferença absoluta entre o maior e o menor período de tactus de cada sinal do banco métrico.

Uma outra avaliação sobre o tactus consiste na verificação de como a diferença entre os períodos de tactus consecutivos se comporta, que auxilia o entendimento de como o andamento varia: grandes diferenças indicam mudanças bruscas no andamento. Histogramas das diferenças consecutivas máximas e das diferenças medianas para cada sinal podem ser vistos nas Figuras 6.5 e 6.6, respectivamente. Através deles, pode-se perceber que em geral o período do tactus não se altera muito ao longo de um sinal. Em particular, para a maior parte dos sinais, a diferença mediana fica abaixo de 50 ms, indicando que ocorrem mudanças suaves do andamento na maior parte dos sinais. Novamente, grandes variações são observadas para os sinais com mudanças abruptas no andamento discutidos no parágrafo anterior.

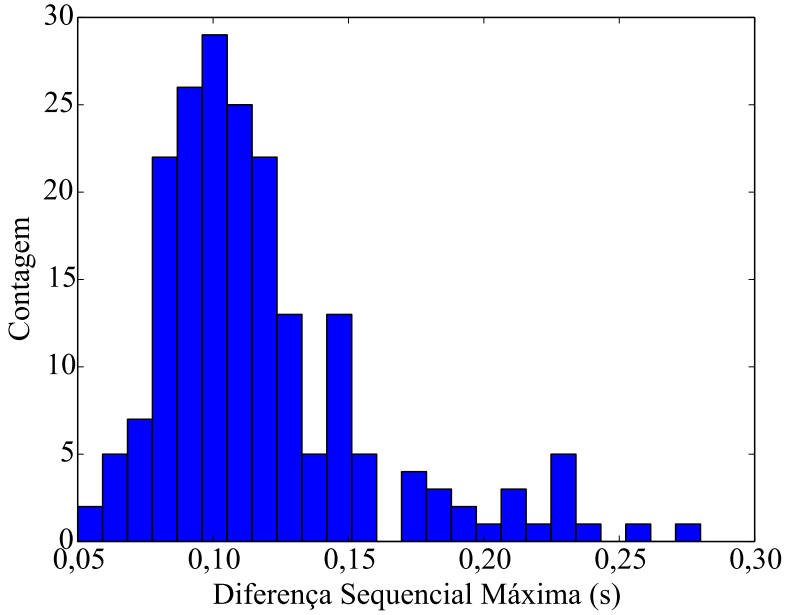


Figura 6.5: Histograma para todos os sinais da maior diferença entre os períodos de tactus consecutivos de um mesmo sinal.

6.4 Análise do Tatum

O período do tatum anotado para cada sinal foi computado em termos do período do tactus. No total, 216 sinais apresentaram períodos do tatum que são metade do período do tactus e 5 sinais exibiram períodos que são um terço do período do tactus. Não foi observada nenhuma mudança do período do tatum dentro de um mesmo sinal. Esses valores indicam que, apesar de o modelo considerar mudanças no período do tatum, tal atributo não poderá ser avaliado, já que nenhum sinal apresentou essa característica. Considerando também a baixa incidência de divisões ternárias (apenas 2 % dos sinais a exibem), pode-se abrir mão desta possibilidade caso seja necessária uma redução na complexidade do modelo hierárquico.

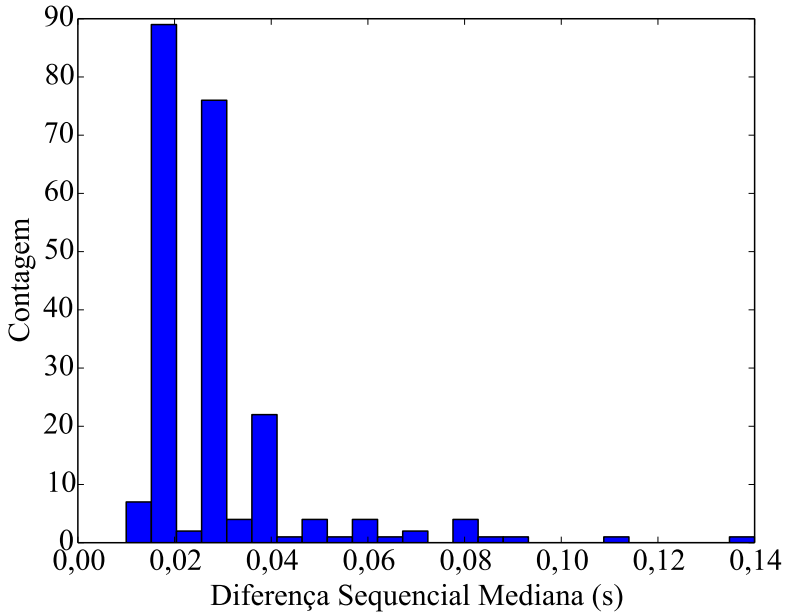


Figura 6.6: Histograma para todos os sinais da diferença mediana entre períodos de tactus consecutivos de um mesmo sinal.

Uma informação importante para o banco de sinais é quanto o período de tatum varia dentro de um mesmo sinal. Na Figura 6.7, é mostrado o desvio padrão dos períodos de tatum. Na figura, pode-se perceber que para a maior parte dos sinais, essa quantidade varia menos que 20 ms. No total, apenas 30 sinais (13 % dos sinais) possuem desvio padrão do período do tatum maior que 30 ms. Isso indica uma relativa estabilidade, conforme se assumiu no modelo hierárquico.

Outra informação importante que pode ser extraída do tatum anotado é a distribuição da diferença entre períodos de tatum consecutivos. Para isso, todas as diferenças entre tatums consecutivos foram computadas para todos os sinais são exibidas no histograma da Figura 6.8. Vale lembrar que nos modelos apresentados no capítulo anterior, essa diferença foi modelada como uma janela de comprimento finito. Observando os dados obtidos, pode-se notar que, realmente, a maior parte dos tatums possui um intervalo absoluto de aproximadamente 50 ms. Apenas 0,9 % dos intervalos ficam fora desta faixa. Deve-se notar que a janela a ser utilizada nos modelos pode ser computada a partir de um subconjuto dos dados, possivelmente levando a um melhora no desempenho dos modelos.

6.5 Análise do Compasso

Nesta seção, será avaliada a marcação dos compassos. A principal informação a ser analisada do compasso anotado é o seu período, e poderá auxiliar na validação das hipóteses feitas sobre este e a forma como evolui dentro de uma mesma música.

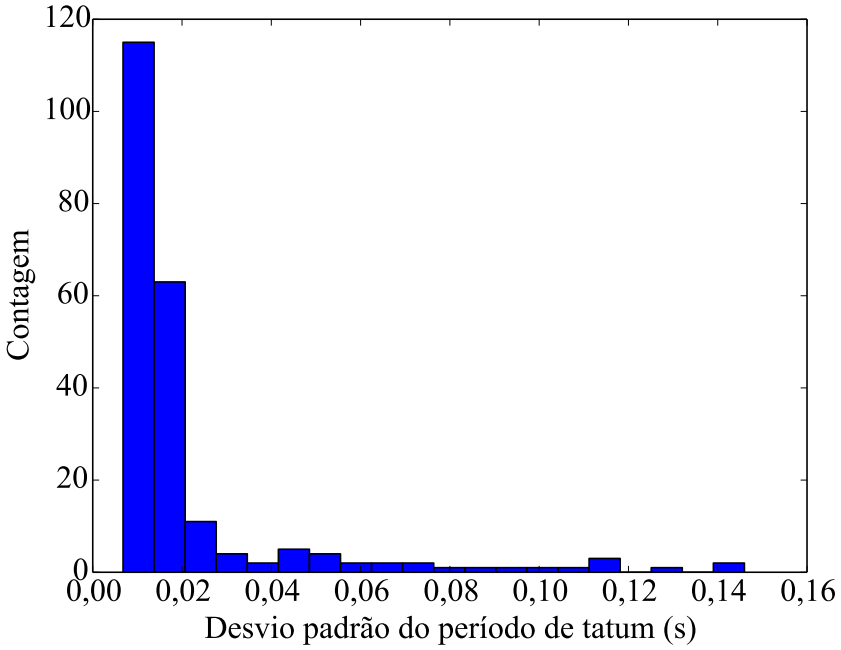


Figura 6.7: Desvio padrão do período de tatum calculado para os sinais do banco métrico.

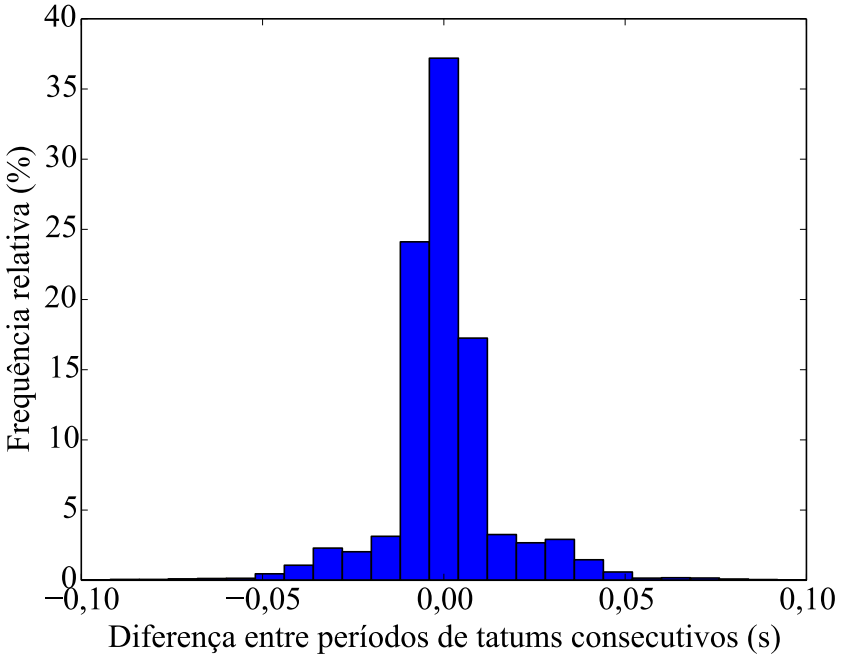


Figura 6.8: Distribuição da diferença entre períodos de tatums consecutivos. Os dados cobrem todos os tatums anotados no banco métrico.

O período do compasso em função do número de tactus que ocorreram dentro dele foi calculado para todos os sinais do banco. Para cada sinal, o período mediano foi computado e é exibido na Tabela 6.2. Como pode ser visto, a maior parte dos sinais possui períodos binários ou quartenários. Apenas 5,4 % dos sinais exibem compassos ternários e 1,8 % dos sinais possuem um valor distinto de 2, 3 ou 4. Novamente, considerando a baixa incidência de períodos diferentes de 2 ou 4, o modelo hierárquico poderia ser simplificado, desconsiderando-se valores diferentes destes para o período do compasso.

Tabela 6.2: Período mediano anotado para cada sinal do banco métrico.

Período	Número de Sinais
2	51
3	12
4	149
Outros	4

Outro fator a ser avaliado do compasso é se o valor de seu período varia ao longo de um mesmo sinal. Para o banco anotado, apenas 15 sinais (6,8 % dos sinais) possuem variação de andamento. Destes 15 sinais, 10 são de música coral que não exibe um compasso regular. Considerando isto, uma simplificação do modelo hierárquico poderia considerar a probabilidade de mudança de compasso como nula.

6.6 Conclusão

Neste capítulo foi apresentado o banco de sinais que será utilizado para o treinamento e análise dos modelos rítmicos propostos. O banco contém 221 sinais com os seus três níveis métricos anotados por um músico profissional, formando uma base sólida sobre a qual os algoritmos poderão ser adaptados e seu desempenho, analisado.

Através da análise das anotações, foram sugeridas algumas simplificações ou modificações sobre os modelos rítmicos propostos. A primeira consiste em não modelar a transição de períodos do tatum binário para ternário, já que estas transições não foram observadas no banco. Outra simplificação que poderia levar a uma redução na complexidade de análise seria não considerar períodos ternários para o tatum, já que apenas 2 % o possuem. Outra simplificação que pode ser adotada é abrir mão de permitir a transição entre períodos de compasso, que foram anotadas em aproximadamente 7 % dos sinais, sendo 70 % dessas ocorrências em sinais que não exibem compasso regular. Por fim, observou-se que a maior parte dos sinais que não são bem modelados pelos algoritmos propostos são músicas corais ou clássicas. Isso

indica uma deficiência do modelo, que poderia ser tornado mais abrangente para melhor modelar alguns desses sinais; alternativamente, um novo modelo mais adequado para estes gêneros poderia ser proposto. De qualquer forma, essa deficiência deve ser levada em consideração ao se realizar a análise do desempenho dos modelos propostos.

Nos próximos capítulos, o banco de sinais métrico será usado para treinar o modelo de observação dos modelos rítmicos propostos e para a sua análise de desempenho.

Capítulo 7

Modelos de Observação para Análise Rítmica

Neste capítulo, serão descritos a metodologia e os algoritmos utilizados para a obtenção dos modelos de observação necessários aos algoritmos para análise rítmica descritos no Capítulo 5. Em particular, serão discutidos modelos para a chance de se observar um determinado valor para o fluxo espectral assumindo que este esteja associado à percepção de novo tatum, tactus ou compasso ou à ausência de informação rítmica no quadro em que foi calculado. Neste sentido, enquanto os modelos descritos no Capítulo 5 procuram capturar como os três níveis métricos evoluem ao longo do tempo, os modelos deste capítulo consideram a informação de apenas um quadro.

Na literatura, podem ser encontrados diferentes modelos de observação para análise do tactus. Em [27], é descrito um modelo simples em que a probabilidade de um quadro associado ao tactus ter gerado um determinado valor aumenta com o valor observado. Logo, assume-se que valores elevados de fluxo espectral estão diretamente associados a ocorrência de um tactus. Uma estratégia similar também é aplicada em [25]. Em [92], uma distribuição gama é utilizada para modelar o valor de uma função de detecção de onset para quadros onde ocorreu um tactus. No caso de algoritmos que procuram modelar mais de um nível métrico, em [93] e [94] foram utilizadas distribuições gama para modelar a distribuição de um atributo associado à intensidade para diferentes instantes dentro de um padrão rítmico. É descrita em [100] uma abordagem em que as probabilidades de observação são aprendidas diretamente de um conjunto de dados.

Para obtenção dos modelos a seguir, será utilizado o banco de sinais apresentando no Capítulo 6, dividido em dois conjuntos: um apenas para obtenção dos modelos e outro para sua avaliação. Diferentemente dos trabalhos descritos no parágrafo anterior, serão abordados modelos para quadros associados aos três níveis métricos e apresentando ausência de informação rítmica. Além disso, também serão estudados

algoritmos que utilizam a média do fluxo espectral ao longo de diversas sub-bandas Mel e também algoritmos que procuram explorar a informação espectral para melhor distinguir entre os níveis métricos.

Os algoritmos a serem explorados podem ser classificados em duas formas: geradores ou discriminativos [105, 110]. Modelos geradores procuram modelar a probabilidade dos dados diretamente para cada um dos níveis rítmicos. Eles possuem esse nome porque, uma vez obtidos, podem ser utilizados para criar novos dados. Modelos discriminativos procuram apenas aprender a discriminar entre dados associados a diferentes classes.

No contexto do trabalho atual, serão estudados modelos geradores para o fluxo espectral calculado em diversas sub-bandas Mel para o modelo hierárquico. Dois modelos geradores serão estudados: o modelo de misturas gaussiana (GMM, do inglês *Gaussian Mixture Models*) e a análise de fatores (FA, do inglês *Factor Analysis*).

Será explorada também uma estratégia utilizando modelos discriminativos. Nesta estratégia, cada quadro de um sinal seria classificado como sendo de um determinado nível rítmico e a probabilidade de acerto do classificador seria encontrada e utilizada na HMM. Os modelos discriminativos serão obtidos a partir dos seguintes classificadores: máquina de vetores de suporte (SVM, do inglês *Support Vector Machine*), regressão logística e florestas aleatórias (RF, do inglês *Random Forests*).

São apresentados na Seção 7.1 os objetivos deste capítulo, juntamente com uma revisão do que precisa ser modelado para o modelo hierárquico. Os dados utilizados na modelagem são descritos na Seção 7.2. Já na Seção 7.3, é apresentada a figura de mérito que será utilizada na avaliação da qualidade dos modelos obtidos. Na Seção 7.4, é realizada uma modelagem preliminar considerando apenas o valor médio do fluxo espectral e na Seção 7.5 um estudo sobre quais regiões de frequência são mais relevantes ao problema. As soluções empregando GMMs e FA são apresentadas nas Seções 7.6 e 7.7, respectivamente. Os modelos discriminativos são propostos na Seção 7.8. A adaptação dos modelos para o rastreamento apenas do tactus e do rastreamento por camadas é apresentada na Seção 7.9. Por fim, são apresentadas as conclusões do capítulo na Seção 7.10.

7.1 Objetivo

No Capítulo 5 foram vistos modelos ocultos de Markov para modelagem rítmica de sinais de áudio. Uma parte integrante desses modelos é o chamado modelo de observação que informa a probabilidade de um determinado estado ter gerado os dados observados num quadro.

Neste capítulo, o foco será o modelo hierárquico descrito na Seção 5.2, por incluir

a modelagem dos três níveis métricos e servir como base para os demais modelos. A Figura 7.1 representa graficamente o modelo de observação que será estimado neste capítulo.

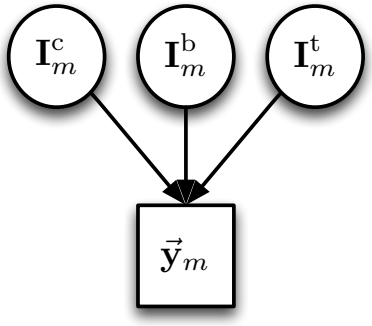


Figura 7.1: Representação gráfica do modelo de observação do modelo hierárquico.

Matematicamente, o modelo pode ser descrito através de

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}}, I_m^{\mathrm{b}}, I_m^{\mathrm{c}}), \tag{7.1}$$

onde I_m^{t}, I_m^{b} e I_m^{c} são variáveis binárias indicadoras da ocorrência de novo tatum, tactus e compasso, respectivamente. Considerando apenas as combinações válidas das variáveis indicadoras, quatro densidades de probabilidade precisariam ser obtidas:

1. $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}} = 0, I_m^{\mathrm{b}} = 0, I_m^{\mathrm{c}} = 0)$: modelando a probabilidade de os dados observados no quadro m não estarem associados a nenhum dos níveis métricos;

2. $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 0, I_m^{\mathrm{c}} = 0)$: modelando a probabilidade de os dados observados estarem associados à ocorrência de novo tatum no quadro m, porém sem a ocorrência de novo tactus;

3. $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 1, I_m^{\mathrm{c}} = 0)$: modelando a probabilidade de os dados observados estarem associados à ocorrência de novo tactus no quadro m, porém sem a ocorrência de novo compasso;

4. $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 1, I_m^{\mathrm{c}} = 1)$: modelando a probabilidade de os dados observados estarem associados à ocorrência de novo compasso no quadro m.

Na Figura 7.2, é mostrada uma ilustração das diferentes classificações que serão empregadas sobre um quadro e a nomenclatura adotada no restante deste capítulo. No topo à esquerda é ilustrado um conjunto denotando todos os quadros. Dentre todos os quadros, os que indicam novo tatum ($I_m^{\mathrm{t}} = 1$) são limitados pelo maior círculo. Dentre estes, são encontrados os que indicam novo tactus ($I_m^{\mathrm{b}} = 1$) e dentre estes, alguns também indicam novo compasso ($I_m^{\mathrm{c}} = 1$). Nas outras quatro figuras,

as regiões sombreadas denotam quais dados não contém informação rítmica ($I_m^{\mathrm{t}} = 0$, $I_m^{\mathrm{b}} = 0$ e $I_m^{\mathrm{c}} = 0$); quais são "apenas novo tatum" ($I_m^{\mathrm{t}} = 1$ e $I_m^{\mathrm{b}} = 0$); quais são "apenas novo tactus" ($I_m^{\mathrm{b}} = 1$, $I_m^{\mathrm{c}} = 0$); e quais são "novo compasso". No decorrer deste texto, cada uma destas possibilidades será, por vezes, chamada de classe dos dados.

O ponto de partida dos dados utilizados no modelo é o fluxo espectral obtido no quadro m para diferentes sub-bandas MEL k, denotado por $F^{\mathrm{SF}}[m, k]$. Conforme discutido nos capítulos anteriores, o fluxo espectral procura capturar mudanças abruptas de energia dentro de cada uma das sub-bandas e esperam-se valores mais elevados para quadros e sub-bandas associados a algum evento rítmico. No entanto, o fluxo espectral acaba por carregar informação de dinâmica de longo prazo: trechos da música que exibem dinâmica mais forte (mais energéticos) irão exibir valores mais elevados para o fluxo espectral que trechos de dinâmica mais fraca (menos energéticos). Para deixar os atributos utilizados menos suscetíveis a esse tipo de variação, é proposta a utilização de uma normalização sobre o fluxo espectral que deixa as variações de intensidade de curto prazo, porém remova as variações de longo prazo. Matematicamente, a normalização proposta é

$$\overline{F}^{\mathrm{SF}}[m, k] = \frac{F^{\mathrm{SF}}[m, k]}{\sqrt[p]{\sum_{i=-\sigma^{\mathrm{Norm}}}^{\sigma^{\mathrm{Norm}}} (F^{\mathrm{SF}}[m+i, k])^p}}, \tag{7.2}$$

onde $p, \sigma^{\mathrm{Norm}} \in \mathbb{N}$ são parâmetros que serão definidos a seguir. A normalização opera de forma independente para cada sub-banda, onde o valor do fluxo espectral para o quadro m é divido pela norma p de uma janela de comprimento $2\sigma^{\mathrm{Norm}} + 1$ centrada em m. O tamanho da janela deve ser escolhido de forma que uma quantidade razoável de eventos rítmicos tenham ocorrido.

Uma boa prática para escolha do valor de σ^{Norm}, adotada neste trabalho, é fazer o tamanho desta janela proporcional ao período (inverso do andamento) estimado do sinal. Matematicamente,

$$\sigma^{\mathrm{Norm}} = \nu l_{\hat{\tau}} \tag{7.3}$$

onde $l_{\hat{\tau}}$ é o período estimado em amostras e ν, um número inteiro positivo.

Já para o parâmetro p, é sugerido um valor elevado de forma que, se o fluxo espectral calculado no quadro m for próximo do valor máximo dentro da janela, ele seja normalizado para um valor próximo de 1. Desta forma, mantém-se a relevância local da intensidade do fluxo espectral, mas removem-se variações de longa duração. Além disso, podem-se interpretar valores normalizados próximos de 1 como indicadores de que um evento rítmico relevante (quando comparado com os seus vizinhos) ocorreu neste quadro, e valores próximos de 0 como a ausência de um evento rítmico. Por fim, note que ao realizar a normalização para cada sub-banda de forma

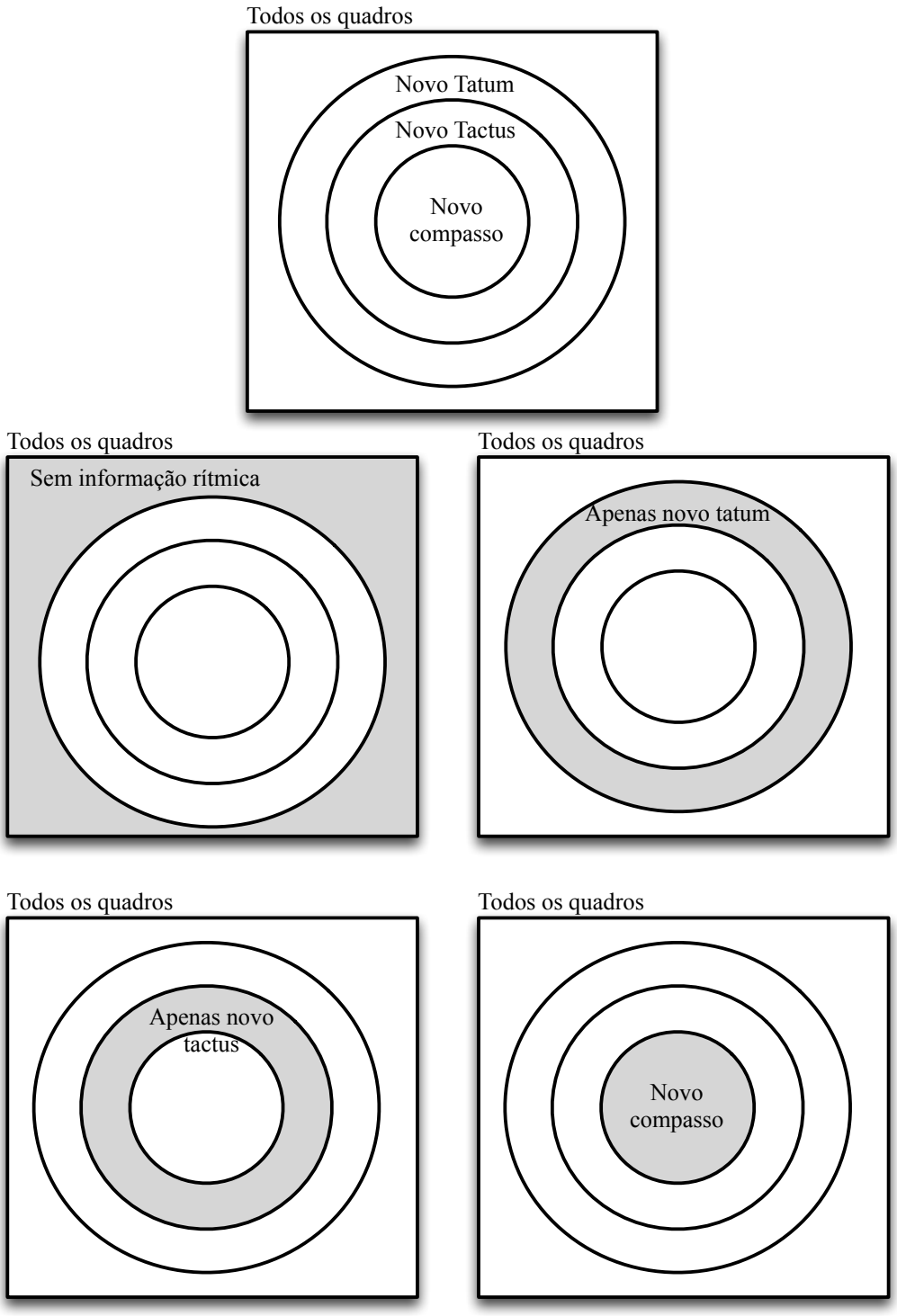
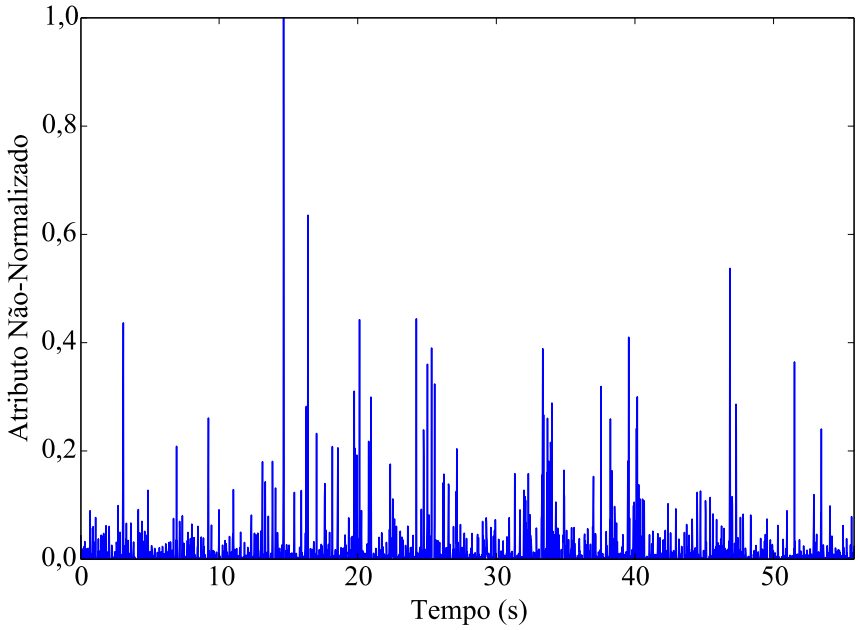
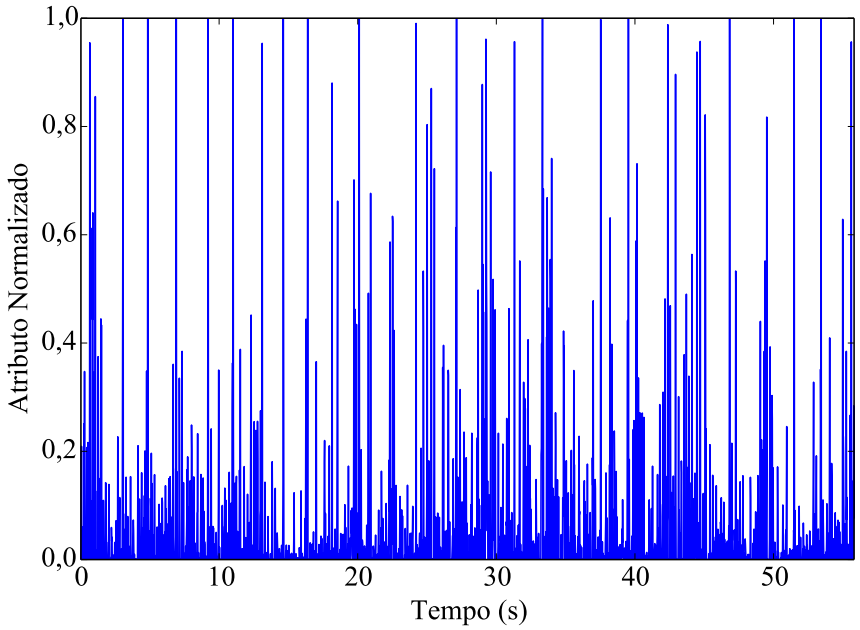


Figura 7.2: Visualização das diferentes classes que serão estudadas e sua classificação. As regiões em cinza denotam as classes que serão estudadas.

independente, a distribuição dos valores ao longo das sub-bandas pode ser alterada em relação à distribuição antes da normalização. De toda forma, como os eventos ritmicamente relevantes tendem a ocorrer simultaneamente em todas as sub-bandas, não é esperado que essa mudança seja significativa.



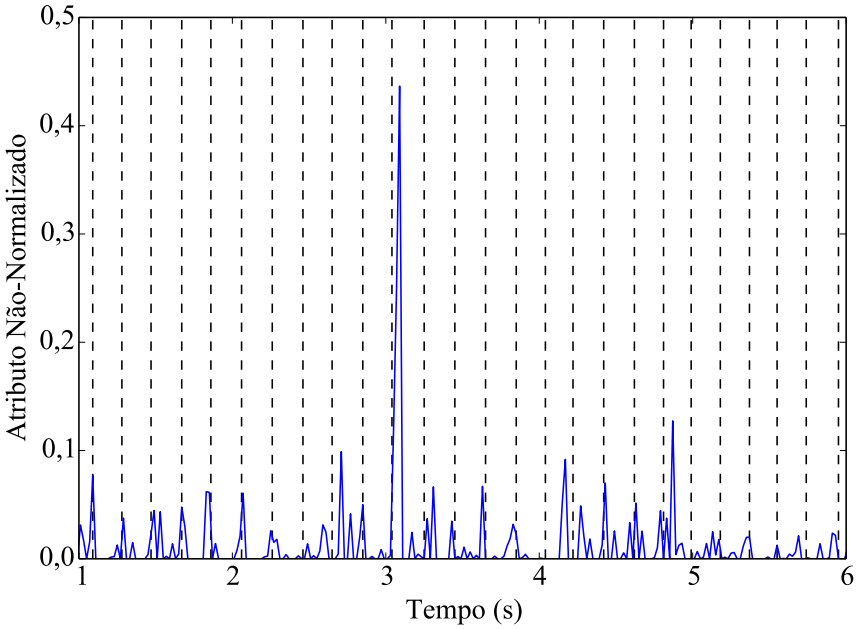
(a) Não normalizado.



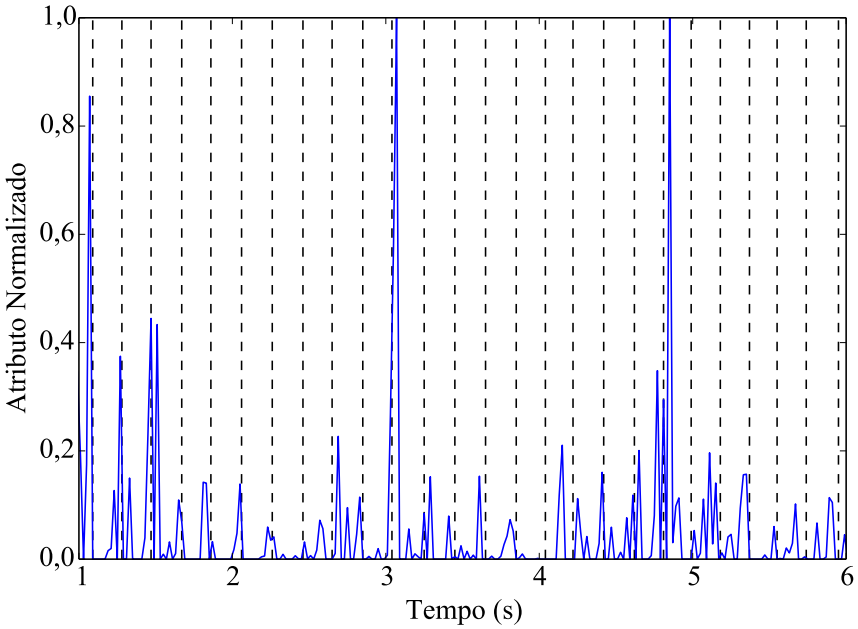
(b) Normalizado

Figura 7.3: Exemplo do sinal numa sub-banda Mel antes e após a normalização.

Na Figura 7.3 é exibido um exemplo do resultado da aplicação da normalização em que é exibido o sinal de uma única sub-banda Mel antes e após a normalização.



(a) Não normalizado.



(b) Normalizado

Figura 7.4: Detalhe do exemplo exibido na Figura 7.3. As linhas tracejadas verticais denotam os tactus anotados.

Como se pode perceber, os valores do atributo variam de intensidade ao longo do sinal antes da normalização. Após a normalização, os valores são mais bem distribuídos, sendo eliminada a variação de longo prazo enquanto é mantida a variação de curto prazo. Na Figura 7.4, pode ser visto um detalhe de apenas alguns segundos para o mesmo sinal com os tactus marcados como linhas verticais tracejadas. Nova-

mente, fica claro que as variações de curto prazo são preservadas pelo procedimento de normalização proposto.

Nos modelos a serem apresentados, os dados a serem modelados sempre serão chamados de $\mathbf{y}$ ou $\vec{\mathbf{y}}$ conforme forem unidimensionais ou não, respectivamente. Um superíndice irá ser usado para distinguir entre os diferentes modelos. Como no Capítulo 6, variáveis aleatórias são denotadas em negrito.

7.2 Geração dos Dados para Treinamento e Validação

Nesta seção, serão apresentados os conjuntos de dados a serem utilizados no treinamento e na validação dos modelos propostos. Em particular, será discutido como o banco métrico será dividido e como os atributos serão avaliados. Além disso, os parâmetros utilizados na extração dos atributos são discutidos. Por fim, é apresentada também a metodologia utilizada para seleção dos dados a fazerem parte dos conjuntos a serem utilizados.

O banco métrico, descrito no Capítulo 6, foi escolhido para se obter os modelos de observação necessários. Conforme mencionado anteriormente, esta base possui 221 sinais com seus tatums, tactus e compassos demarcados por um músico profissional. Além disso, o gênero de cada sinal também é conhecido.

O banco métrico será dividido em duas partes: uma para a obtenção dos modelos de observação (treinamento) e outra para a análise do desempenho dos modelos obtidos (validação). Como a maior parte dos métodos utilizados para se obter o modelo de observação procuram minimizar uma função-custo para um determinado conjunto de dados, utilizar um segundo conjunto que não fez parte da otimização realizada é considerada uma boa prática em aprendizado de máquina [105]. Considerando que os diferentes gêneros musicais apresentaram diferentes características rítmicas, a proporção de sinais de cada gênero foi preservada em cada uma das partições em relação à proporção encontrada na base toda. Ao final, 60 % dos sinais (133 sinais) foram associados à partição de treinamento e os restantes, à partição de validação.

Para cada sinal no banco, o fluxo espectral foi extraído para 40 sub-bandas MEL (E) e com dois conjuntos de parâmetros para o comprimento da janela (N) e salto entre janelas (H): $\{N = 40$ ms, $H = 20$ ms$\}$ e $\{N = 80$ ms, $H = 40$ ms$\}$. Cada conjunto de parâmetros irá gerar um diferente conjunto de dados que serão utilizados na obtenção do modelo de observação. Essa escolha foi feita para se investigar se esses parâmetros influenciam os resultados obtidos e se existe uma vantagem em se utilizar janelas e saltos menores que justifique o aumento de complexidade que esta

escolha pode causar na estimação do modelo rítmico (ver Seção 5.2). Uma vez que o fluxo espectral é obtido para cada sinal, ele é normalizado seguindo o procedimento descrito na seção anterior com $p = 8$ e $\nu = 8$.

Até este ponto tem-se um vetor de 40 elementos contendo o fluxo espectral normalizado para cada quadro de cada sinal do banco métrico. Agora, faz-se necessário identificar quais destes vetores estão associados a algum nível métrico e quais não estão associados a nenhum. O maior desafio desta etapa é alinhar o instante em que cada início de tatum foi marcado com um determinado quadro do atributo extraído. Para isto, é realizado o seguinte procedimento:

1. É calculado um sinal auxiliar composto da soma ao longo das sub-bandas do fluxo espectral normalizado em cada quadro, $F^{\mathrm{AUX}}[m] = \sum_k \overline{F}^{\mathrm{SF}}[m, k]$;
2. É centralizada em torno de cada início de tatum anotado uma janela de 80 ms;
3. São gerados, então, dois conjuntos:
   (a) O primeiro contendo os quadros associados ao maior valor de $F^{\mathrm{AUX}}[m]$ dentro da janela centrada no início do tatum;
   (b) O segundo contendo os quadros que não estão dentro de nenhuma janela.

Os fluxos espectrais associados aos elementos do primeiro conjunto irão formar o conjunto de observações que estão associadas a algum nível métrico; já o segundo conjunto conterá as observações que não estão associadas a nenhum nível métrico.

O procedimento descrito no parágrafo acima garante que os quadros que possuem o maior valor para o fluxo espectral são efetivamente associados ao início de tatum marcado, corrigindo pequenos desalinhamentos entre tatums marcados e o sinal. Cabe lembrar que os tatums foram anotados com uma precisão de 100 ms, logo a correção aplicada ainda fica dentro da incerteza da marcação. Além disso, esse procedimento remove dos conjuntos quadros que estão muito próximos dos quadros escolhidos para os inícios de tatum, que ainda poderiam conter uma parte da mudança de energia associada ao evento anotado. Isto remove do conjunto de dados uma série de observações que poderiam causar confusão, facilitando que os modelos aprendam a distinguir apenas o que claramente não é uma observação associada a um nível métrico de uma que é.

Na Figura 7.5, pode ser visto um exemplo do atributo auxiliar com as transições entre tatums anotadas e as janelas utilizadas sobrepostas. Pela figura, pode-se observar que apesar de valores mais elevados frequentemente estarem associados ao quadro mais próximo do início de tatum anotado, isso nem sempre ocorre; entretanto, a janela adotada é capaz de muitas vezes compensar esse desvio.

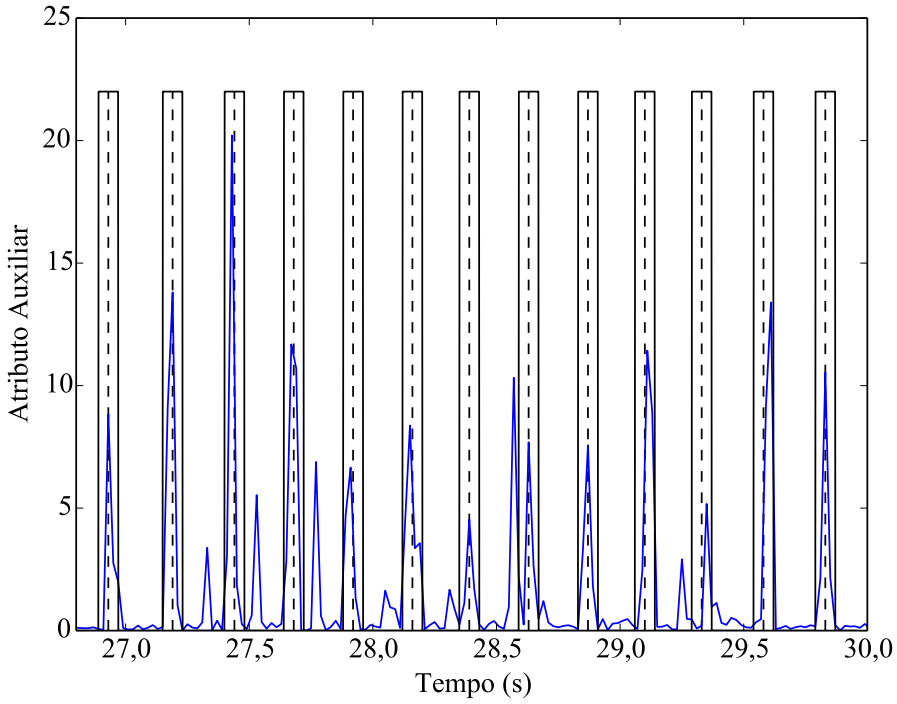


Figura 7.5: Exemplo das janelas utilizadas para obtenção das observações. As linhas tracejadas verticais denotam o início de tatum anotado, a linha sólida mais escura demarca as janelas em torno dos início de tatums anotados e a linha sólida azul o atributo auxiliar.

Uma vez tendo sido obtidas as observações associadas aos níveis métricos anotados, cabe agora agrupá-las nos três conjuntos que serão utilizados para o treinamento dos modelos:

1. Um contendo as observações associadas a apenas um início de tatum ($I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 0, I_m^{\mathrm{c}} = 0$);
2. Um contendo as observações associadas a apenas um início de tactus ($I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 1, I_m^{\mathrm{c}} = 0$);
3. Um contendo as observações associadas a um início de compasso ($I_m^{\mathrm{t}} = 1, I_m^{\mathrm{b}} = 0, I_m^{\mathrm{c}} = 1$).

A Figura 7.2 ilustra como estes dados são obtidos a partir dos dados gerais. Desta forma, são obtidos 4 conjuntos de dados que irão ser utilizados para a obtenção dos 4 modelos. Cabe lembrar que o procedimento acima é realizado para todos os sinais das partições de treinamento e de validação e que as observações para cada conjunto contêm as observações de todos os sinais da partição correspondente. Na Tabela 7.1, é mostrada a quantidade de elementos em cada conjunto para os atributos extraídos com diferentes parâmetros. Como era esperado, a quantidade de observações sem informação rítmica é muito maior que as das observações nos demais conjuntos. Além disso, a quantidade de elementos observados diminui conforme o nível métrico aumenta.

Tabela 7.1: Número de observações nas partições de treinamento e validação para cada conjunto de parâmetros para os conjuntos contendo as observações sem informação rítmica (SIR), apenas início de tatum (AIT), apenas início de tactus (AIB) e início de compasso (IC).

Treinamento				
Parâmetros	SIR	AIT	AIB	IC
$\{N = 40$ ms, $H = 20$ ms$\}$	224011	14218	9912	4309
$\{N = 80$ ms, $H = 40$ ms$\}$	98039	13910	9584	4250
Validação				
Parâmetros	SIR	AIT	AIB	IC
$\{N = 40$ ms, $H = 20$ ms$\}$	139391	9219	6262	2685
$\{N = 80$ ms, $H = 40$ ms$\}$	61751	9527	6573	2735

7.3 Avaliação

Antes de iniciar a modelagem dos dados, cabe uma discussão sobre como poderão ser comparados os desempenhos dos modelos obtidos por diferentes estratégias.

Idealmente, esses modelos de observação devem ser avaliados quando utilizados em conjunto com o HMM completo, comparando-se a sequência de estados obtida com a anotada. No entanto, tal análise teria uma complexidade computacional muito elevada, e é vantajoso poder excluir o mais cedo possível modelos que não apresentam um bom desempenho.

Uma figura de mérito para a comparação dos modelos de observação deveria levar em consideração se a probabilidade atribuída a uma observação que está associada a apenas um início de tatum, por exemplo, recebe uma alta probabilidade a partir do modelo obtido para $p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{t}}=1, I_m^{\mathrm{b}}=0, I_m^{\mathrm{c}}=0)$ e baixas probabilidades para os demais modelos. O principal desafio, no entanto, é como obter uma medida consistente considerando que as probabilidades estimadas não podem ser diretamente comparadas, já que os condicionantes de cada modelo são diferentes.

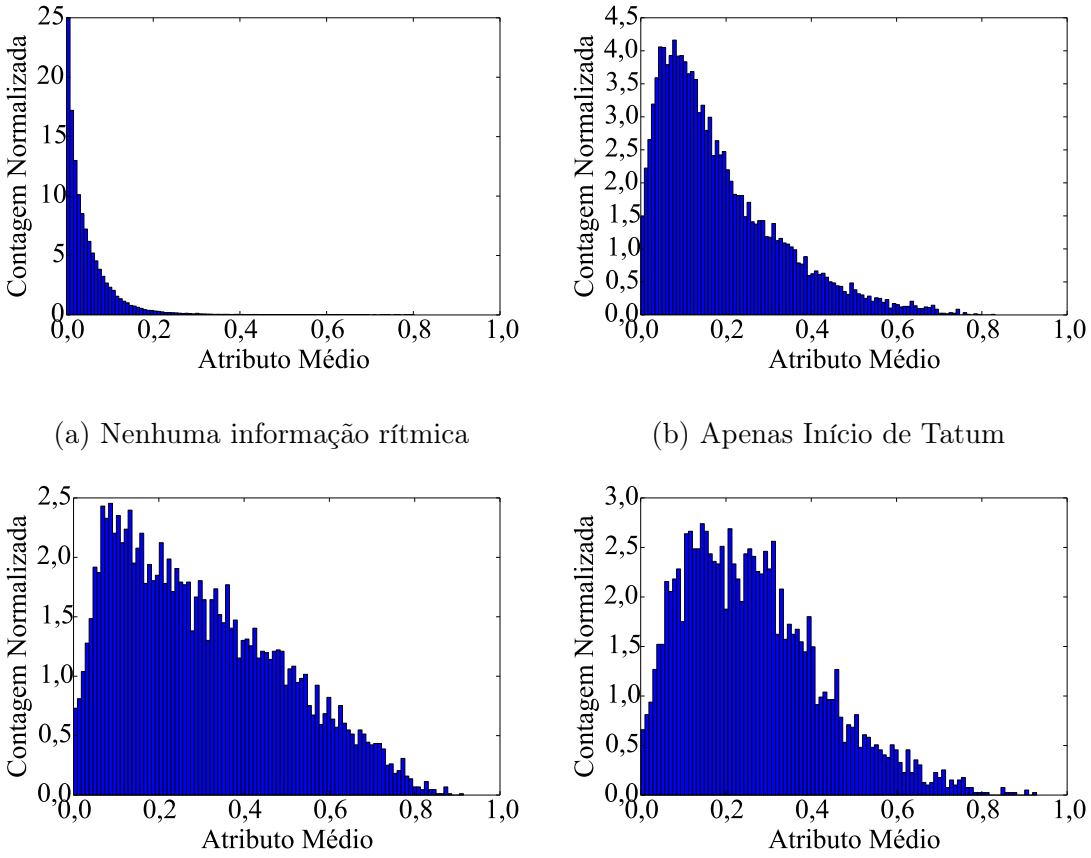
A solução utilizada acaba por apenas comparar as probabilidades, atribuindo um acerto para quando a probabilidade do modelo correspondente ao conjunto ao qual a observação pertence é maior que todas as demais. Ao se realizar essa comparação para todos os dados na partição de validação, pode-se chegar a uma idéia de quão consistentes são os modelos gerados, obtendo-se uma figura de mérito que informa para qual percentual dos dados em cada um dos conjuntos foi corretamente atribuída a maior probabilidade.

Em geral, os resultados obtidos não precisam ser necessariamente elevados, lembrando que uma parcela importante da informação necessária para a classificação dos quadros virá dos HMMs. Neste sentido, a característica mais importante desta métrica é apontar quais modelos são mais adequados e prover um indicativo da dificuldade que será enfrentada pelo HMM.

7.4 Modelagem Preliminar

Nesta seção, será realizada uma modelagem preliminar dos dados com o objetivo de adquirir mais informações que auxiliarão na escolha de modelos mais complexos. Nesta modelagem, o foco será em dados unidimensionais, que podem ser facilmente visualizados. Além disso, como será visto, tais dados podem ser adequadamente modelados por distribuições de probabilidade univariadas que podem ser facilmente ajustadas aos dados observados.

Nas próximas seções, serão exibidas as distribuições de probabilidade para o valor médio ou mediano do fluxo espectral normalizado para cada conjunto de dados. Uma vez que a distribuição dos valores é observada, será escolhida e ajustada uma função que aproxima esta distribuição. Por fim, a figura de mérito descrita na Seção 7.3 será utilizada para se avaliar o desempenho destes modelos.



(a) Nenhuma informação rítmica

(b) Apenas Início de Tatum

(c) Apenas Início de Tactus

(d) Início de Compasso

Figura 7.6: Histogramas dos valores médio do fluxo espectral normalizado para os dados em cada conjunto da partição de treinamento.

7.4.1 Atributos Médios

Nesta seção, é apresentado um modelo de observação para o valor médio do atributo espectral normalizado. Neste caso, uma amostra desta observação seria obtida como

$$y_m^{\mathrm{AVG}} = \frac{1}{K} \sum_{k=1}^{K} \overline{F}^{\mathrm{SF}}[m, k], \tag{7.4}$$

para um determinado quadro m de um sinal de áudio. Intuitivamente, pode-se pensar que esta variável modela o comportamento médio do fluxo espectral.

Os histogramas mostrados na Figura 7.6, exibem os valores assumidos pela VA $\mathbf{y}_m^{\mathrm{AVG}}$ para os quadros em cada um dos conjuntos de dados da partição de treinamento. A contagem de cada valor foi normalizada de forma que a área coberta pelas barras seja igual a 1. São exibidos os histogramas para o fluxo espectral calculado com $\{N = 80$ ms, $H = 40$ ms$\}$, resultados similares são obtidos para $\{N = 40$ ms, $H = 20$ ms$\}$.

Observando os histogramas, pode-se notar que, como esperado, a distribuição

da observação quando não há informação rítmica se concentra em zero, sendo que a chance de se encontrar um valor elevado cai de forma aparentemente exponencial com o valor médio do atributo. Já a distribuição de apenas início de tatum exibe uma média mais elevada, com uma moda em torno de 0,1 e uma distribuição menos concentrada. As distribuições de apenas início de tactus e início de compasso são similares, com valores médios novamente mais elevados que a distribuição de apenas tatum. No entanto, ambas as distribuições são similares entre si, apontando para uma dificuldade para se distinguir uma observação de início de compasso de uma de apenas início de tactus.

Considerando os histogramas observados, foram escolhidas uma distribuição exponencial [105] para os dados que não exibem informação rítmica e distribuições gama [105] para os demais. Matematicamente, pode-se escrever

$$p_{\mathbf{y}_m^{\mathrm{AVG}}}(y_m^{\mathrm{AVG}}|I_m^{\mathrm{t}}=0, I_m^{\mathrm{b}}=0, I_m^{\mathrm{c}}=0) = \lambda^{\mathrm{n}} e^{-\lambda^{\mathrm{n}} y_m^{\mathrm{AVG}}}, \tag{7.5}$$

$$p_{\mathbf{y}_m^{\mathrm{AVG}}}(y_m^{\mathrm{AVG}}|I_m^{\mathrm{t}}=1, I_m^{\mathrm{b}}=0, I_m^{\mathrm{c}}=0) = \frac{(y_m^{\mathrm{AVG}})^{\lambda^{\mathrm{t}}} e^{-\frac{y_m^{\mathrm{AVG}}}{\theta^{\mathrm{t}}}}}{(\theta^{\mathrm{t}})^{\lambda^{\mathrm{t}}}\Gamma(\lambda^{\mathrm{t}})} \tag{7.6}$$

$$p_{\mathbf{y}_m^{\mathrm{AVG}}}(y_m^{\mathrm{AVG}}|I_m^{\mathrm{t}}=1, I_m^{\mathrm{b}}=1, I_m^{\mathrm{c}}=0) = \frac{(y_m^{\mathrm{AVG}})^{\lambda^{\mathrm{b}}} e^{-\frac{y_m^{\mathrm{AVG}}}{\theta^{\mathrm{b}}}}}{(\theta^{\mathrm{b}})^{\lambda^{\mathrm{b}}}\Gamma(\lambda^{\mathrm{b}})} \tag{7.7}$$

$$p_{\mathbf{y}_m^{\mathrm{AVG}}}(y_m^{\mathrm{AVG}}|I_m^{\mathrm{t}}=1, I_m^{\mathrm{b}}=1, I_m^{\mathrm{c}}=1) = \frac{(y_m^{\mathrm{AVG}})^{\lambda^{\mathrm{c}}} e^{-\frac{y_m^{\mathrm{AVG}}}{\theta^{\mathrm{c}}}}}{(\theta^{\mathrm{c}})^{\lambda^{\mathrm{c}}}\Gamma(\lambda^{\mathrm{c}})}, \tag{7.8}$$

onde $\{\lambda^{\mathrm{n}},\ \lambda^{\mathrm{t}},\ \theta^{\mathrm{t}},\ \lambda^{\mathrm{b}},\ \theta^{\mathrm{b}},\ \lambda^{\mathrm{c}},\ \theta^{\mathrm{c}}\}$ são parâmetros livres a serem ajustados e $\Gamma(\cdot)$ é a função gama [105]. Os parâmetros livres foram obtidos através da maximização da verossimilhança [105] entre as distribuições ajustadas e os dados na partição de treinamento.

Na Figura 7.7, podem ser vistas as distribuições obtidas após o ajuste dos parâmetros. Como pode ser observado, as distribuições ajustadas se aproximam dos histogramas exibidos na Figura 7.6. Além disso, pode ser observado que as distribuições para apenas o início de tactus e início de compasso são muito próximas. Por fim, pode-se notar também que apesar de possuir valores próximos da origem, a distribuição obtida para quando há ausência de informação rítmica se sobrepõe às demais, principalmente na região de valores entre 0,05 e 0,2.

Na Tabela 7.2, é mostrada a figura de mérito adotada para a avaliação de desempenho calculada para a partição de validação dos dados. Vale lembrar que a figura de mérito denota o percentual dos dados cuja probabilidade atribuída pelo modelo correto foi maior que a probabilidade atribuída pelos demais modelos. Pode-se notar que o desempenho para os dados que não exibem informação rítmica foi razoável, com mais de 80 % dos dados desta classe tendo a maior probabilidade a ele atribuída. Já para as demais classes, o desempenho foi ruim, com desempenho abaixo de 50 %,

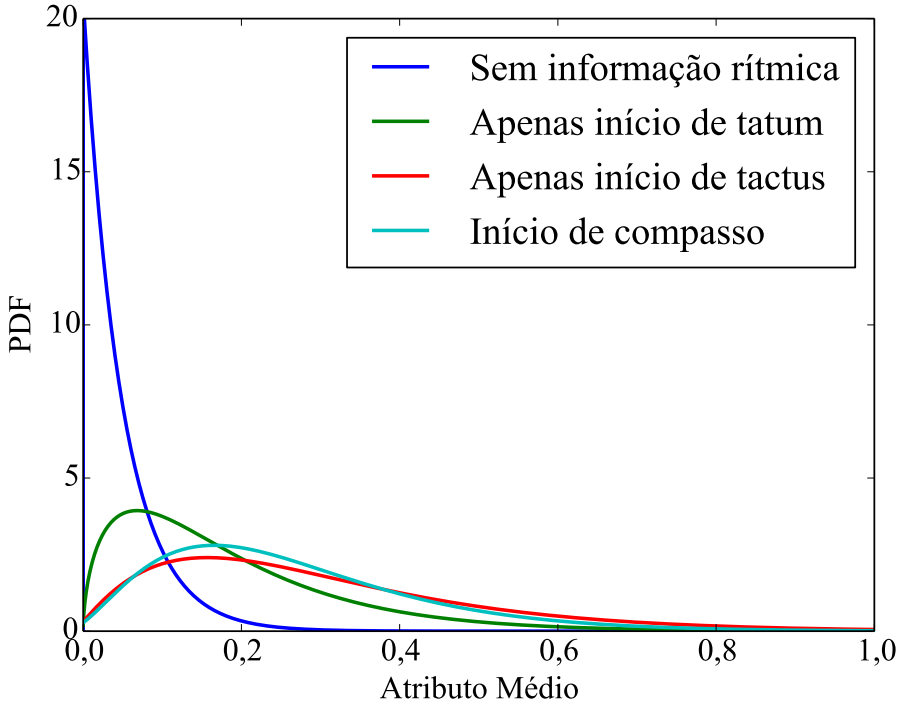


Figura 7.7: Distribuições ajustadas aos dados na partição de treinamento para os valores médios do fluxo espectral ao longo das raias.

especialmente para apenas início de tatum. Por fim, pode-se perceber que a escolha dos parâmetros do fluxo espectral alterou pouco o desempenho.

Tabela 7.2: Desempenho do modelo ajustado sobre a média ao longo das raias do fluxo espectral normalizado (SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso).

Parâmetros	SIR	AIT	AIB	IC
$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$	83%	35%	35%	45%
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$	82%	27%	34%	44%

Se o atributo mediano [88] for utilizado no lugar da média para a obtenção da observação, chega-se a um resultado melhor, como pode ser visto na Tabela 7.3. A melhoria de desempenho ocorre pelo fato de a mediana não ser tão influenciada por valores anômalos. No entanto, como ser observado, usar a mediana piora consideravelmente os resultados obtidos para o início de compasso. Na discussão realizada na próxima seção será vista uma possível razão para esta queda de desempenho.

Tabela 7.3: Desempenho do modelo ajustado sobre a mediana ao longo das raias do fluxo espectral normalizado (SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso).

Parâmetros	SIR	AIT	AIB	IC
$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$	91%	37%	48%	12%
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$	90%	32%	48%	12%

7.4.2 Variação ao Longo da Frequência

Observando-se os resultados da seção anterior, fica claro que utilizar apenas a média (ou mediana) do fluxo espectral ao longo das raias não fornece um bom desempenho, principalmente na caracterização de início de compasso e de apenas início de tatum. Para entender como o desempenho pode ser melhorado futuramente nos modelos descritos, será feita nesta seção uma análise sobre como os dados se comportam em cada sub-banda.

Na Figura 7.8 é mostrado o valor médio do fluxo espectral normalizado para cada uma das 40 sub-bandas Mel para cada conjunto de dados da partição de treinamento. O fluxo espectral normalizado exibido foi obtido com $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$. Como era esperado, o valor médio para os dados que não exibem informação rítmica está abaixo do valor médio para apenas início de tatum para todas as raias. O mesmo acontece entre os valores médios de apenas início de tatum e apenas início tactus, apesar de a diferença neste caso ser menor.

O resultado mais interessante obtido é entre apenas início tactus e início de compasso, em que pode-se ver que os dois conjuntos possuem valores médios próximos e que o segundo exibe um valor menor para a maior parte das sub-bandas. A exceção ocorre nas sub-bandas mais graves, em que o valor médio no início de compasso é superior a no início de tactus. O fato de que em música popular o início de compasso é muitas vezes marcado por uma batida de um tambor grave poderia explicar esse comportamento dos atributos. Esse comportamento anômalo em poucas sub-raias também poder explicar o fato da mediana ter um desempenho tão ruim para o início de compasso, já que ela reduziria a influência das raias que mais se desviam da média.

A breve análise realizada neste capítulo aponta para a necessidade de se utilizar informação espectral no modelo de observação. Essa informação é de extrema importância, principalmente para se distinguir as observações de apenas início de tactus das de início de compasso.

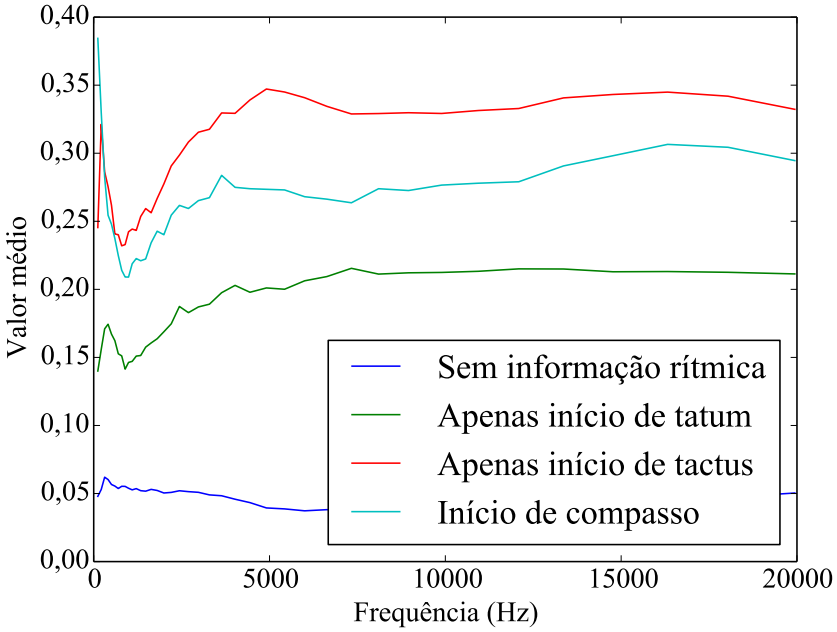


Figura 7.8: Valor médio para cada sub-banda Mel para a partição de treinamento.

7.5 Redução de Dimensionalidade

Nesta seção, serão avaliados métodos para redução da dimensão do vetor de observação a ser utilizado. Tal redução poderia facilitar a obtenção dos modelos que serão estudados a seguir, permitindo melhor modelagem dos dados. Além disso, o processo de redução de dimensionalidade também traz sobre o problema informações que serão úteis na escolha e interpretação dos resultados obtidos posteriormente.

Duas formas de redução de dimensionalidade serão estudadas: compressão das sub-bandas Mel num número menor de componentes relevantes e seleção apenas das sub-bandas Mel que apontam para a existência de informação rítmica. Ambas as estratégias são utilizadas na literatura [111] e são soluções distintas para o problema de redução de dimensionalidade.

Como será visto, ambas as soluções mostram que a dimensão dos dados observados não é facilmente reduzida, indicando que todas as sub-bandas Mel contêm informação relevante ao problema. Isso se deve à alta correlação entre os dados de uma sub-banda com os observados nas demais.

7.5.1 Análise de Componentes Principais

Nesta seção, é descrita a análise de componentes principais (PCA, do inglês *Principal Component Analysis*) [112] dos dados observados, procurando-se combinações dos dados em cada sub-banda que mantêm a informação relevante ao problema.

A PCA é um procedimento que procura extrair componentes que são, potencialmente, descorrelacionados. Cada componente, então, seria responsável por explicar uma parcela da variação observada nos dados. Para obter esses componentes principais, inicialmente é criada a matriz $\vec{X}$ com número de linhas igual ao número de dados e número de colunas igual ao número de sub-bandas Mel. Em seguida, é realizada a decomposição em valores singulares [112] dessa matriz, de forma que:

$$\vec{X} = \vec{U}\vec{\Sigma}\vec{W}^T, \tag{7.9}$$

onde $\vec{\Sigma}$ é uma matriz diagonal contendo os valores singulares, que podem ser interpretados como a parcela da variância de $\vec{X}$ explicada por cada componente principal, e $\vec{U}$ e $\vec{W}$ são matrizes contendo os vetores ortogonais a partir dos quais podem ser obtidas os componentes principais. Com isso, pode-se gerar uma versão reduzida dos dados selecionando-se apenas os vetores ortogonais associados aos z maiores valores principais, obtendo-se uma versão comprimida dos dados através de

$$\vec{X}_z = \vec{X}\vec{W}_z, \tag{7.10}$$

onde $\vec{X}_z$ é uma matriz com apenas p colunas e $\vec{W}_p$ é formado pelas linhas de $\vec{W}$ associadas aos p maiores valores singulares.

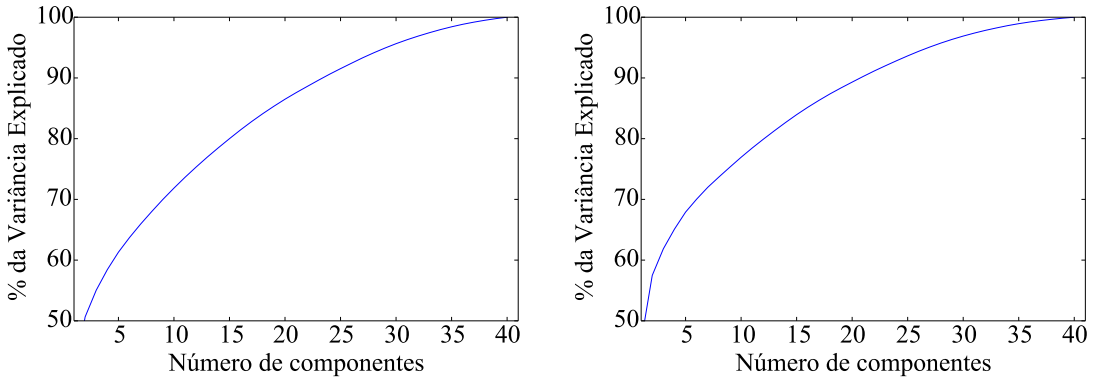
Observando-se os elementos na diagonal da matriz $\vec{\Sigma}$, pode-se determinar o número de elementos que concentram uma determinada parcela da variância dos dados observados. Na Figura 7.9 é exibida essa quantidade que informa a variação explicada quando se utiliza um determinado número de componentes principais para os dados obtidos com diferentes parâmetros para o fluxo espectral. Conforme pode ser notado, a maior parte da variância só é explicada quando se atinge o número máximo de sub-bandas, indicando que são necessários todos os componentes principais para se obter modelar os dados.

7.5.2 Seleção de Atributos

Nesta seção, será apresentada uma tentativa de se descobrir quais sub-bandas Mel são mais importantes na discriminação entre informação não-rítmica, apenas início de tatum, apenas início de tactus e início de compasso. Desta forma, sub-bandas que não contribuem significativamente para esta discriminação poderiam ser eliminadas.

O método adotado se baseia na regressão logística, que pode ser definida para um dado como

$$y_m^{\text{Log}} = \Phi\left(\sum_{k=1}^{K} w_k^{\text{Log}} \overline{F}^{\text{SF}}[m,k] + w_0\right), \tag{7.11}$$



(a) $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$ (b) $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$

Figura 7.9: Percentual da variância que é explicado de acordo com o número de componentes principais.

onde

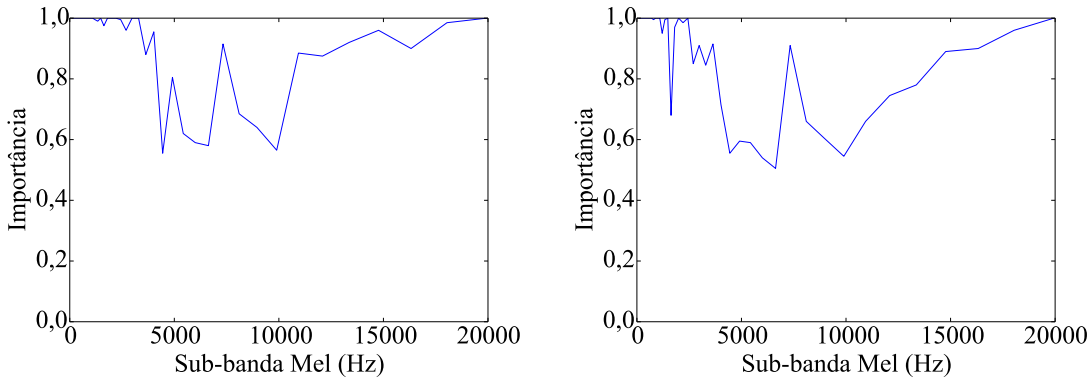
$$\Phi(x) = \frac{1}{1 + e^x} \tag{7.12}$$

é a função logística e w_k são pesos a serem determinados. A função logística possui valores apenas entre 0 e 1 e os pesos são obtidos de forma a maximizar a chance de o resultado ser um para quando o dado observado pertence a uma determinada classe e zero em caso contrário. Dessa forma, quatro pesos devem ser obtidos, cada um ajustado para detectar uma das quatro classes desejadas. A função-custo utilizada para a obtenção dos coeficientes também procura minimizar os seus valores, de forma que a coeficientes que pouco contribuem para o problema de classificação seja atribuído o valor zero.

O método de seleção de atributos utilizado, descrito em [113] e conhecido como seletor de estabilidade, consiste em treinar diversos classificadores, cada um com um sub-conjunto dos atributos e verificar quais deles são consistentemente selecionados (possuem coeficiente não-nulo). Ao final do procedimento, obtém-se uma medida que é proporcional ao número de vezes que cada atributo foi selecionado, sendo que um valor próximo de 0 indica um atributo que nunca foi selecionado (pouco importante); e um valor perto de 1, indica um atributo que foi sempre selecionado (muito importante).

Na Figura 7.10, podem ser vistos os resultados para os dois conjuntos de parâmetros utilizados. Como já foi observado na Seção 7.4.2, as regiões de baixa frequência e alta frequência são relativamente mais importantes que as de média frequência. No entanto, pode-se notar que mesmo as sub-bandas com menor valor de importância ainda foram selecionadas um número significativo de vezes. Isso indica que todas as sub-bandas Mel são importantes para a discriminação, sendo que a exclusão de uma

dada sub-banda levaria a uma piora na discriminação entre os dados das diferentes classes.



(a) $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$ (b) $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$

Figura 7.10: Importância de cada sub-banda Mel para a discriminação entre dados observados de diferentes classes.

7.6 GMMs

Nesta seção, será apresentada a modelagem dos dados observados utilizando-se o modelo de mistura de gaussianas. Este modelo procura descrever a probabilidade de se ter observado um determinado dado através da combinação de gaussianas multidimensionais, cada uma com média, matriz de covariância e peso próprios dentro do modelo.

Para GMMs, a variável observada é um vetor com dimensão igual ao número de sub-bandas Mel, sendo cada elemento do vetor definido como

$$\vec{y}_m^{\,\text{GMM}}[k] = \overline{F}^{\text{SF}}[m, k], \tag{7.13}$$

ou seja, o fluxo espectral normalizado calculado numa determinada sub-banda Mel.

A densidade de probabilidade de cada variável será definida como um mistura de gaussianas, onde os parâmetros serão ajustados de forma independente para cada

uma das quatro classes. Matematicamente, tem-se

$$p_{\vec{\mathbf{y}}_m^{\text{GMM}}}(\vec{y}_m^{\text{GMM}}|I_m^{\text{t}}=0, I_m^{\text{b}}=0, I_m^{\text{c}}=0) = \sum_{i=0}^{G^{\text{n}}} w_i^{\text{n}} \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{n}}, \vec{\Sigma}_i^{\text{n}}\right) \tag{7.14}$$

$$p_{\vec{\mathbf{y}}_m^{\text{GMM}}}(\vec{y}_m^{\text{GMM}}|I_m^{\text{t}}=1, I_m^{\text{b}}=0, I_m^{\text{c}}=0) = \sum_{i=0}^{G^{\text{t}}} w_i^{\text{t}} \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{t}}, \vec{\Sigma}_i^{\text{t}}\right) \tag{7.15}$$

$$p_{\vec{\mathbf{y}}_m^{\text{GMM}}}(\vec{y}_m^{\text{GMM}}|I_m^{\text{t}}=1, I_m^{\text{b}}=1, I_m^{\text{c}}=0) = \sum_{i=0}^{G^{\text{b}}} w_i^{\text{b}} \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{b}}, \vec{\Sigma}_i^{\text{b}}\right) \tag{7.16}$$

$$p_{\vec{\mathbf{y}}_m^{\text{GMM}}}(\vec{y}_m^{\text{GMM}}|I_m^{\text{t}}=1, I_m^{\text{b}}=1, I_m^{\text{c}}=1) = \sum_{i=0}^{G^{\text{c}}} w_i^{\text{c}} \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{c}}, \vec{\Sigma}_i^{\text{c}}\right), \tag{7.17}$$

onde $G^{\cdot}$ é o número de componentes gaussianas de cada modelo, w_i o peso de cada componente e

$$\mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}, \vec{\Sigma}\right) = \frac{1}{(2\pi)^E |\vec{\Sigma}|} e^{\left(-\frac{1}{2}(\vec{y}-\vec{\mu})^T \vec{\Sigma}^{-1}(\vec{y}-\vec{\mu})\right)} \tag{7.18}$$

é uma distribuição gaussiana para E variáveis com vetor de médias $\vec{\mu}$ e matriz de covariância $\vec{\Sigma}$.

O problema de obtenção da GMM para cada uma das classes, então, é obter os parâmetros $\{G,\ w_i,\ \mu_i,\ \vec{\Sigma}_i\}$ para cada componente de cada classe. Para isto, inicialmente são escolhidos um determinado número de componentes e uma possível estrutura para a matriz de covariância. O algoritmo *expectation-maximization* (EM) é, então, utilizado para se obter os parâmetros restantes a partir dos dados observados na partição de treinamento. Para cada classe, foram obtidos GMMs para G variando entre 1 e 12 e restringindo-se a matriz de covariância a ser diagonal ou cheia. Em seguida, utilizou-se a figura de mérito calculada sobre o conjunto de treinamento para se escolher o modelo final.

Ao final do procedimento descrito no parágrafo anterior, foram obtidos os parâmetros mostrados na Tabela 7.4 para cada um dos conjuntos de dados. Como se pode observar, sempre foi escolhida a matriz de covariância cheia, indicando, novamente, a alta correlação entre os dados das diversas cada sub-bandas. Além disso, um número elevado de componentes foram escolhidos, indicando que os dados possuem uma distribuição que difere muito de uma gaussiana.

Utilizando-se os modelos treinados, foi calculada a figura de mérito descrita na Seção 7.3 para a partição de validação, e os resultados obtidos podem ser vistos na Tabela 7.5. Como se pode notar, os resultados melhoraram para o modelo de início de compasso e do apenas início de tatum em relação ao obtido utilizando-se apenas a mediana dos dados, indicando que a informação presente nas sub-bandas auxiliou a melhor modelagem destes dados. Este aumento no desempenho, no entanto, resultou num pior desempenho para não informação rítmica e apenas início de tactus,

Tabela 7.4: Melhores parâmetros, segundo a figura de mérito, para o modelo GMM (SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso).

G				
Parâmetros	SIR	AIT	AIB	IC
$\{N = 40$ ms, $H = 20$ ms$\}$	9	9	6	4
$\{N = 80$ ms, $H = 40$ ms$\}$	9	8	6	4
Formato de $\vec{\Sigma}$				
Parâmetros	SIR	AIT	AIB	IC
$\{N = 40$ ms, $H = 20$ ms$\}$	cheia	cheia	cheia	cheia
$\{N = 80$ ms, $H = 40$ ms$\}$	cheia	cheia	cheia	cheia

possivelmente devido ao fato de as distribuições entre as classes ainda possuírem uma grande sobreposição.

Tabela 7.5: Desempenho do GMM (SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso).

Parâmetros	SIR	AIT	AIB	IC
$\{N = 40\,\text{ms}, H = 20\,\text{ms}\}$	80%	42%	37%	24%
$\{N = 80\,\text{ms}, H = 40\,\text{ms}\}$	80%	37%	38%	27%

7.7 Análise de Fatores

Uma metodologia alternativa para obtenção de modelos geradores é a análise de fatores [105]. Neste modelo, a variação observada nos dados é modelada através de combinações lineares de variáveis ocultas, chamadas de fatores. Uma vez obtido o modelo, é possível obter a probabilidade de um determinado dado ter sido gerado pelo modelo.

Matematicamente, a variável observada é definida similarmente ao que se fez na seção anterior:

$$\vec{y}_m^{\,\text{FA}}[k] = \overline{F}^{\text{SF}}[m, k], \tag{7.19}$$

e assume-se que uma dada observação pode ser descrita como

$$\vec{\mathbf{y}}_m^{\,\text{FA}} = \vec{W}\vec{h}_m + \vec{\epsilon}, \tag{7.20}$$

onde $\vec{h}_m$ é o vetor-coluna contendo os z fatores que descrevem a observação no quadro m, $\vec{W}$ é uma matriz que indica a contribuição de cada fator nos dados observados,

e $\vec{\epsilon}$ é um vetor de erros, que são modelados como gaussianos de média $\vec{\mu}$ e matriz de covariância $\vec{\Sigma}$ (que é assumida diagonal). Em geral, z é um parâmetro a ser escolhido na modelagem; devido à análise realizada na Seção 7.5, z foi escolhido igual ao número de sub-bandas.

Para cada classe dos dados observados, então, são obtidos os parâmetros $\{\vec{W}, \vec{\mu}, \vec{\Sigma}\}$ a partir dos dados de treinamento para os fatores $\vec{h}_m$. Uma vez obtidos esses parâmetros, a distribuição dos dados pode ser definida como

$$p_{\vec{\mathbf{y}}_m^{\text{FA}}}(\vec{y}_m^{\text{FA}}|I_m^{\text{t}}=0, I_m^{\text{b}}=0, I_m^{\text{c}}=0) = \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{n}}, \vec{W}^{\text{n}}(\vec{W}^{\text{n}})^T + \vec{\Sigma}_i^{\text{n}}\right) \tag{7.21}$$

$$p_{\vec{\mathbf{y}}_m^{\text{FA}}}(\vec{y}_m^{\text{FA}}|I_m^{\text{t}}=1, I_m^{\text{b}}=0, I_m^{\text{c}}=0) = \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{t}}, \vec{W}^{\text{t}}(\vec{W}^{\text{t}})^T + \vec{\Sigma}_i^{\text{t}}\right) \tag{7.22}$$

$$p_{\vec{\mathbf{y}}_m^{\text{FA}}}(\vec{y}_m^{\text{FA}}|I_m^{\text{t}}=1, I_m^{\text{b}}=1, I_m^{\text{c}}=0) = \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{b}}, \vec{W}^{\text{b}}(\vec{W}^{\text{b}})^T + \vec{\Sigma}_i^{\text{b}}\right) \tag{7.23}$$

$$p_{\vec{\mathbf{y}}_m^{\text{FA}}}(\vec{y}_m^{\text{FA}}|I_m^{\text{t}}=1, I_m^{\text{b}}=1, I_m^{\text{c}}=1) = \mathcal{N}\left(\vec{\mathbf{y}}|\vec{\mu}_i^{\text{c}}, \vec{W}^{\text{c}}(\vec{W}^{\text{c}})^T + \vec{\Sigma}_i^{\text{c}}\right). \tag{7.24}$$

Note que a distribuição final dos dados é gaussiana com apenas uma componente e matriz de covariância cheia. A diferença em relação ao modelo GMM consiste na forma como a matriz de covariânca é obtida, através da maximização da verossimilhança entre os dados observados e o modelo adotado.

Os resultados obtidos utilizando-se análise de fatores para os dados da partição de treinamento podem ser vistos na Tabela 7.6. Como pode ser observado, em relação à GMM os resultados para não informação rítmica melhoraram consideravelmente, enquanto o modelo para apenas início de tatum e apenas tactus piorou. Já o modelo para início de compasso não teve o seu desempenho alterado.

Tabela 7.6: Desempenho da Análise de Fatores (SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso).

Parâmetros	SIR	AIT	AIB	IC
$\{N = 40\,\text{ms}, H = 20\,\text{ms}\}$	92%	38%	22%	24%
$\{N = 80\,\text{ms}, H = 40\,\text{ms}\}$	92%	32%	26%	25%

7.8 Abordagem via Classificação

Nesta seção, será descrita uma abordagem alternativa à apresentada nas últimas duas seções. Agora, o atributo observado será a saída de um classificador multi-classe treinado para discriminar entre as quatro classes de interesse. Desta forma, a observação passa a ser unidimensional e discreta, permitindo o cômputo direto da distribuição a partir da partição de validação.

Matematicamente, a observação será obtida através

$$y_m^{\text{Class}} = \Xi\left(\overline{F}^{\text{SF}}[m,k]\right), \tag{7.25}$$

onde Ξ é a saída de um classificador com $y_m^{\text{Class}} \in \{\text{n}, \text{t}, \text{b}, \text{c}\}$. Com isso, é necessário modelar a probabilidade de o classificador predizer que o fluxo espectral medido pertence a uma classe quando ele na realidade pertence a outra classe. Ao final, serão obtidos 16 valores definindo a probabilidade de observação de cada uma das variáveis, dada a classe assumida.

O grande desafio desta abordagem é a definição da função de decisão Ξ. Idealmente, ela deve ser capaz de identificar perfeitamente cada uma das classes. Contudo, sabe-se que a informação temporal é muito importante para a distinção entre as diferentes classes, conforme discutido no Capítulo 5. Com isso, pode-se pensar que estes classificadores fornecem uma primeira estimativa da classe de cada quadro, e os HMMs apresentados no Capítulo 5 refinarão estas estimativas através da inclusão da informação de periodicidade.

Nas próximas seções, são descritas diferentes estratégias para a obtenção da função Ξ, cada uma explorando um classificador diferente.

7.8.1 Regressão Logística

O primeiro classificador a ser avaliado é a regressão logística, já brevemente apresentada na Seção 7.5.2. A regressão logística procura combinar a informação em cada sub-banda e depois aplica a função logística para comprimir o resultado da soma para valores entre 0 e 1, conforme descrito na equação (7.11), gerando assim uma função que permite distinguir entre duas classes. Durante o treinamento, um classificador que aprende a distinguir uma classe de todas as outras é obtido, totalizando 4 funções de classificação. Para se obter o valor predito, o resultado de cada classificador é comparado com os demais, e o que mais se aproxima de 1 é escolhido como vencedor. Esta estratégia de classificação é usualmente denominada um-contra-todos (OVA, do inglês *one-versus-all*).

Os quatro classificadores de regressão logística foram obtidos utilizando-se a partição de treinamento. Como as quantidades de dados em cada classe são muito diferentes, a função-custo utilizada leva essa informação em consideração: se uma classe possui dez vezes menos amostras que outra, um erro na classificação dessa classe durante o treinamento é multiplicado por dez.

O desempenho dos classificadores pode ser avaliado diretamente através da matriz de confusão, que mede a quantidade de dados da classe x que foram classificados como y. Através da normalização das linhas dessa matriz também é possível obter as probabilidades utilizadas no modelo de observação. Na Tabela 7.7, pode ser vista

a matriz de confusão para a regressão logística. Como pode ser notado, para os dois conjuntos, apenas início de tatum e início de compasso não foram bem modelados pela distribuição, sendo a maior parte dos dados considerados como sem informação rítmica. Apenas início de tactus obteve um resultado um pouco melhor, com metade dos dados sendo corretamente identificados, mas com na maior parte dos erros sendo confundido com sem informação rítmica. De forma geral, pode-se perceber que o classificador aprendeu corretamente a classificação para ausência de informação rítmica, mas não foi capaz de aprender as demais classes de forma satisfatória.

Tabela 7.7: Matriz de confusão para a regressão logística. (Todos valores em %. Verd. = Verdadeiro, Est. = Estimado, SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso)

$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	97	1	1	1
AIT	54	13	25	8
AIB	24	11	50	15
IC	23	12	39	26
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	94	2	2	2
AIT	48	17	25	10
AIB	24	13	47	16
IC	23	11	35	31

7.8.2 SVM

Nesta seção, é descrita a classificação utilizando máquina de vetores de suporte (SVM, do inglês *Support Vector Machines*) [114]. SVM já foi utilizada em diversos problemas de classificação e vem apresentando um desempenho muitas vezes superior quando comparado com outros esquemas de classificação [115]. O procedimento de treinamento da SVM consiste em procurar os dados, chamados vetores de suporte, que definem o hiperplano capaz de separar os dados de duas classes. Variantes da SVM podem incluir o mapeamento dos dados para uma dimensão superior utilizando-se funções não-lineares. Estes mapeamentos foram avaliados neste trabalho, porém não forneceram melhores resultados; todos os resultados aqui apresentados com SVM consideram apenas o caso linear (sem uso de mapeamento).

Similarmente à regressão logística, SVM separa o sinal em apenas duas classes,

sendo necessária a utilização de um esquema similar ao utilizado na regressão logística. No caso da SVM, é utilizado o esquema um contra um (OVO, do inglês *one versus one*) [116], onde uma SVM é treinada para cada par de classes, ou seja, cada SVM aprende a distinguir entre duas classes. Para se obter a predição dos dados, é realizado um esquema de votação: cada SVM treinada "vota" numa classe, e é escolhida como saída da SVM a classe que foi mais votada. A função custo de cada SVM também é ponderada pelo número de elementos em cada classe.

Na Tabela 7.8, podem ser vistos os resultados obtidos utilizando-se a SVM com o esquema OVO para a partição de validação. Em ambos os conjuntos, pode-se perceber um melhor desempenho da SVM em relação à regressão logística, com os maiores valores sempre aparecendo na diagonal da matriz de confusão. Além disso, os erros tendem a ocorrer em classes que são "próximas", indicando a consistência no classificador. Contudo, os índices de acerto são muito baixos, com menos da metade das amostras sendo corretamente classificadas para as classes que contêm informação rítmica. Deve-se lembrar que essas probabilidades ainda devem ser associadas ao modelo temporal para que, no fim, a decisão sobre os dados seja tomada.

Tabela 7.8: Matriz de confusão para SVM. (Todos valores em %. Verd. = Verdadeiro, Est. = Estimado, SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso)

$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	86	11	1	2
AIT	27	47	15	11
AIB	10	32	36	22
IC	9	33	24	34
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	86	11	1	2
AIT	28	42	17	14
AIB	13	28	35	25
IC	11	26	19	43

7.8.3 SVM Hierárquico

Uma estratégia alternativa seria, no lugar de uma classificação OVO, gerar um número menor de classificadores, porém numa estrutura em árvore, de forma similar à utilizada em [117]. A principal vantagem dessa estratégia de classificação é o fato de os classificadores serem treinados com conjuntos que são mais próximos da

forma como são organizados os dados. Além disso, são utilizados apenas os dados relevantes em cada nível, reduzindo o problema do desbalanceamento dos dados em cada classe.

A Figura 7.11 exibe como foram organizados os classificadores. No primeiro nível, a SVM1 é responsável por separar os dados que possuem informação rítmica dos que não a possuem. No segundo nível, a SVM2 identifica quais dados são apenas início de tatum dos que são apenas início de tactus ou início de compasso. Por fim, a SVM3 é responsável por separar os que são apenas início de tactus dos dados que são início de compasso. Os dados em cada nível são treinados com subconjuntos da partição de treinamento contendo as classes relevantes para aquele nível. Por exemplo, a SVM2 é treinada apenas com dados que possuem informação rítmica (os sem informação rítmica são removidos do conjunto de treinamento). A principal desvantagem deste esquema é o fato de erros serem propagados.

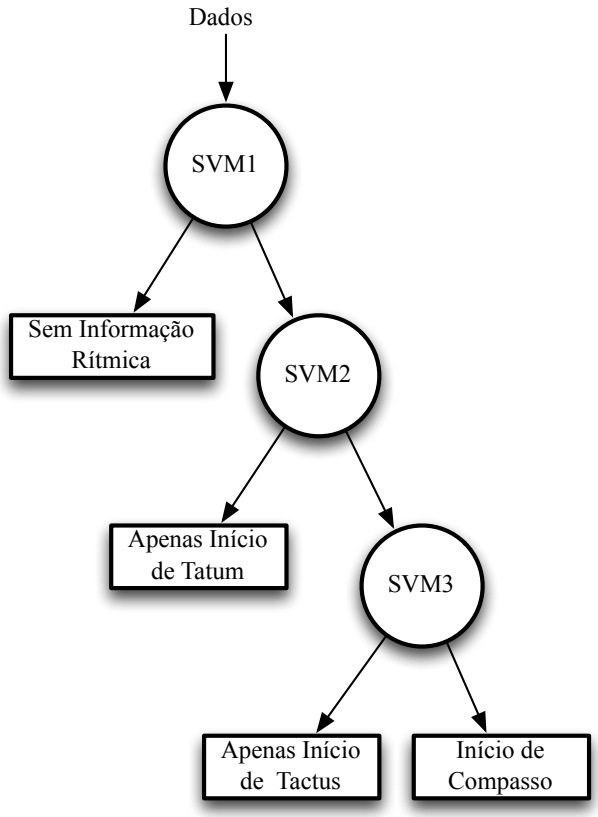


Figura 7.11: Diagrama ilustrando a estrutura de classificação para a SVM hierárquica.

As três SVMs, lineares, foram obtidas, então, utilizando-se a partição de treinamento. Os dados da partição de validação, então, foram classificados e a matriz de confusão obtida pode ser vista na Tabela 7.9. Os resultados obtidos são similares aos da SVM OVO para todas as classes, com exceção do início de compasso, para a qual, pode ser observada uma piora no desempenho. Isso pode ser explicado pelo o

fato de ser esta a última classe a ser classificada, sofrendo, portanto, a acumulação dos erros de classificação cometidos nos níveis anteriores.

Tabela 7.9: Matriz de confusão para SVM Hierárquica. (Todos valores em %. Verd. = Verdadeiro, Est. = Estimado, SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso)

$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	89	8	1	1
AIT	32	46	13	9
AIB	12	35	33	20
IC	11	35	22	32
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	88	10	1	1
AIT	30	49	11	10
AIB	11	34	34	21
IC	10	37	21	33

7.8.4 Floresta Aleatória

Os classificadores apresentados na seção anterior eram capazes apenas de identificar duas classes, demandando o uso de diferentes estratégias para classificação de mais de um par de classes. Nesta seção, será estudado o uso de um classificador multiclasse: a floresta aleatória (RF, do inglês *Random Forests*) [118]. A floresta aleatória é composta por diversos classificadores em árvore [105], cada um treinado por um sub-conjunto dos dados de treinamento. Durante a classificação, o dado é atribuído à classe que foi selecionada pelo maior número de árvores.

Os resultados utilizando a RF não foram positivos, como pode ser visto na Tabela 7.10 para 10 árvores. Resultados similares foram obtidos para diferentes números de árvores. Em geral, a RF não foi capaz de identificar corretamente as classes rítmicas, sendo os erros todos polarizados sem informação rítmica.

7.9 Modelos Parciais

Nesta seção, são apresentados modelos de observação para as simplificações do modelo hierárquico apresentadas no Capítulo 5. Os passos seguidos serão os mesmos que para o modelo hierárquico: inicialmente será apresentado o que precisa ser mo-

Tabela 7.10: Matriz de confusão para Floresta Aleatória com 10 árvores. (Todos valores em %. Verd. = Verdadeiro, Est. = Estimado, SIR = Sem informação rítmica, AIT = Apenas início de tatum, AIB = apenas início de tactus e IC = início de compasso)

$\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	93	4	2	1
AIT	62	19	13	6
AIB	42	20	28	10
IC	44	20	24	11
$\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$				
Verd. \ Est.	SIR	AIT	AIB	IC
SIR	88	7	4	1
AIT	51	25	17	8
AIB	34	23	32	12
IC	34	24	27	15

delado para, em seguida, serem apresentadas algumas formas alternativas de se obter o modelo a partir do conjunto de dados.

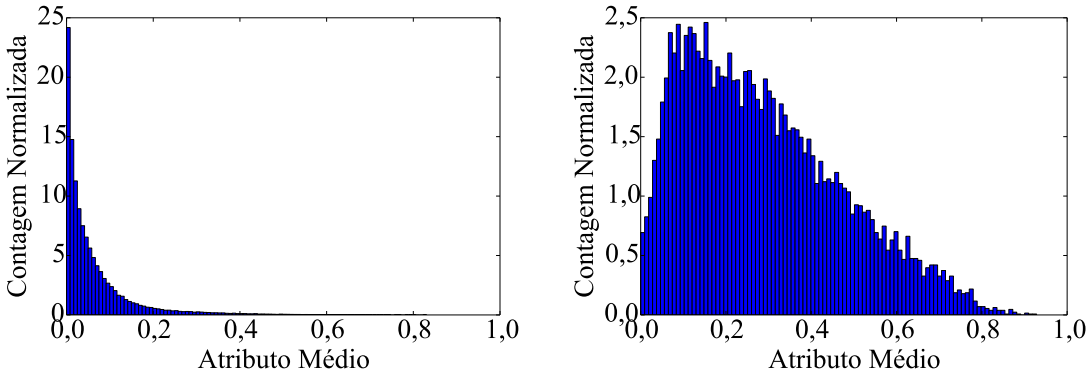
Para todos os casos, os dados a serem analisados serão os mesmos que os descritos na Seção 7.2 e utilizados nas seções anteriores. A única mudança será na forma como as anotações são agrupadas para ficarem coerentes com as necessidades de cada um dos modelos parciais.

7.9.1 Rastreamento do Tactus

No modelo de rastreamento do tactus, descrito originalmente na Seção 5.3, abre-se mão de se obter o instante de ocorrência de novo tatum e de novo compasso, procurando-se apenas identificar quando um novo tactus ocorreu. Com isso, é necessário modelar apenas dois tipos de observação: quando não há informação de início de tactus e quando há. Nesse caso, os dados precisam ser combinados em dois conjuntos:

1. Contendo os quadros anteriormente marcados como não possuindo informação e os que contêm apenas início de tatum, ou seja, $\mathbf{I}^{\text{b}} = 0$;

2. Contendo os quadros anteriormente marcados como apenas início de tactus e como início de compasso, ou seja, $\mathbf{I}^{\text{b}} = 1$.



(a) Sem Informação de Tactus

(b) Início de Tactus

Figura 7.12: Histogramas dos valores médio do fluxo espectral normalizado para o modelo de rastreamento do tactus para os dados da partição de treinamento.

Uma vez obtidos estes dois conjuntos, deve-se obter os seguintes modelos

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{b}}=0) \tag{7.26}$$

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\mathrm{b}}=1). \tag{7.27}$$

Considerando os resultados encontrados ao se obter o modelo de observação para todos os níveis métricos, nas próximas seções apenas três soluções serão abordadas: uma utilizando apenas o valor médio do atributo, uma utilizando GMMs e uma terceira utilizando uma SVM.

Modelo Univariado

Seguindo o trabalho realizado na Seção 7.4, serão obtidos nesta seção modelos para apenas início de tactus quando a observação é definida como na equação (7.4). São exibidas na Figura 7.12 a distribuição obtida, neste caso, para a partição de treinamento para as duas classes a serem abordadas quando o fluxo espectral é calculado com $\{N = 80\,\mathrm{ms},\ H = 40\,\mathrm{ms}\}$. Como pode ser visto, os dados sem informação de tactus se aproximam de uma distribuição exponencial, enquanto os dados de início de tactus podem ser modelados por uma distribuição gama. Na Figura 7.13, podem ser vistas estas distribuições ajustadas para os dados da partição de treinamento.

Se a figura de mérito descrita na Seção 7.3 for calculada para os dados da partição de validação, é obtido o seguinte resultado:

- Para $\{N = 40\,\mathrm{ms},\ H = 20\,\mathrm{ms}\}$: 88 % dos sem informação de tactus e 85 % dos inícios de tactus são corretamente classificados;
- Para $\{N = 80\,\mathrm{ms},\ H = 40\,\mathrm{ms}\}$: 84 % dos sem informação de tactus e 80 % dos

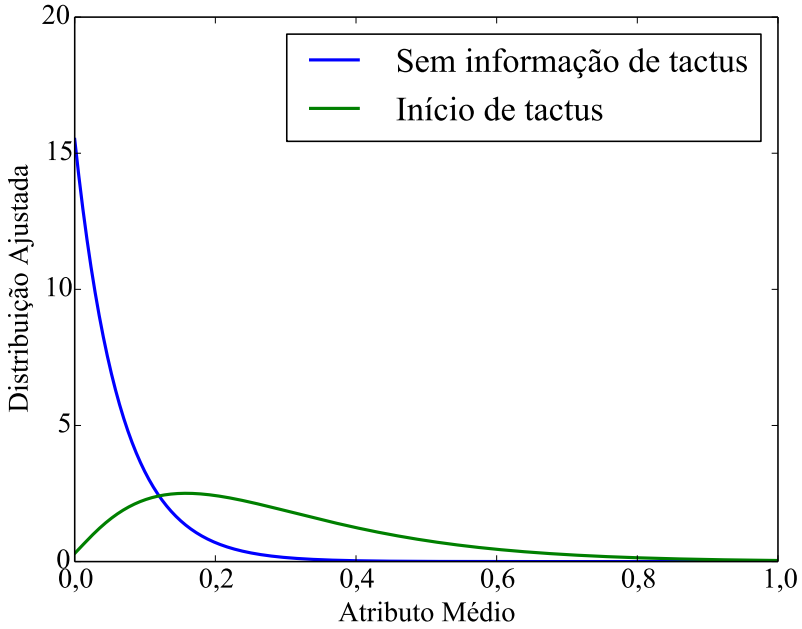


Figura 7.13: Distribuições ajustadas para o modelo considerando apenas o tactus para a partição de treinamento.

inícios de tactus são corretamente classificados.

Como pode-se notar, os resultados obtidos são positivos, indicando que usar apenas a média parece fornecer um bom modelo para rastrear o tactus.

GMM

Seguindo os mesmos passos utilizados na Seção 7.6 para os três níveis hierárquicos, um modelo utilizando misturas de gaussianas foi treinado para os dados. Novamente, foram testados GMMs com número de componentes variando entre 1 e 12 e diferentes formatos para a matriz de covariância, sendo escolhido o que obteve melhor resultado segundo a figura de mérito da Seção 7.3 para a partição de validação.

Para o conjunto de dados obtido com $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$, foram corretamente classificados, no melhor caso, 86 % dos sem informação rítmica e 80 % dos inícios de tatums, quando foram utilizados 6 gaussianas e um modelo de matriz de covariância cheio. Já para os dados do conjunto obtido com $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$, foram classificados, no melhor caso, 83 % dos sem informação de tactus e 74 % dos inícios de tactus, quando foram utilizadas 12 componentes e matriz de covariância cheia. Quando comparados com o resultado da seção anterior, pode-se perceber que a inclusão de informação das sub-bandas trouxe pouco benefício para a distinção entre início ou não de tactus. Conforme discutido para o modelo hierárquico, a informação espectral é mais importante para melhorar a identificação de apenas tactus e de compasso, trazendo pouco benefício para os modelos obtidos

para rastreamento apenas do tactus.

SVM

Uma máquina de vetores de suporte também foi treinada para discriminar entre quadros que são de início de tactus e os demais. Como é necessário discriminar apenas entre duas classes, neste caso, apenas uma SVM necessita ser treinada.

Para o conjunto obtido com parâmetros $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$, 90 % dos sem informação de tactus e 83 % dos inícios de tactus foram corretamente classificados pela SVM. Já para os dados do conjunto obtido com $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$, 87 % dos sem informação de tactus e 78 % dos inícios de tactus foram corretamente classificados. Como pode-se notar, os resultados foram similares aos obtidos anteriormente usando-se GMM e a distribuição univariada.

7.9.2 Modelo Hierárquico por Camadas

Nesta seção, serão descritos os modelos obtidos para o modelo hierárquico por camadas originalmente descrito na Seção 5.4. Este modelo utiliza duas HMMs independentes para modelar os três níveis hierárquicos: a primeira identifica quais quadros são início de tatum (modelo de rastreamento do tatum), enquanto a segunda identifica, dentre eles, quais também são início de tactus e início de compasso (modelo de rastreamento métrico).

Com isso, dois modelos de observação precisam ser obtidos. O primeiro funciona de forma similar ao apresentado anteriormente, sendo que o modelo precisa aprender apenas duas distribuições:

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\text{t}} = 0) \tag{7.28}$$

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\text{t}} = 1), \tag{7.29}$$

ou seja, uma para quadros associados ao início de tatum e outra para os demais. Serão estudadas nas seções seguintes diferentes formas de se obter essas distribuições, seguindo os mesmos passos utilizados na obtenção do modelo para rastreamento do tactus.

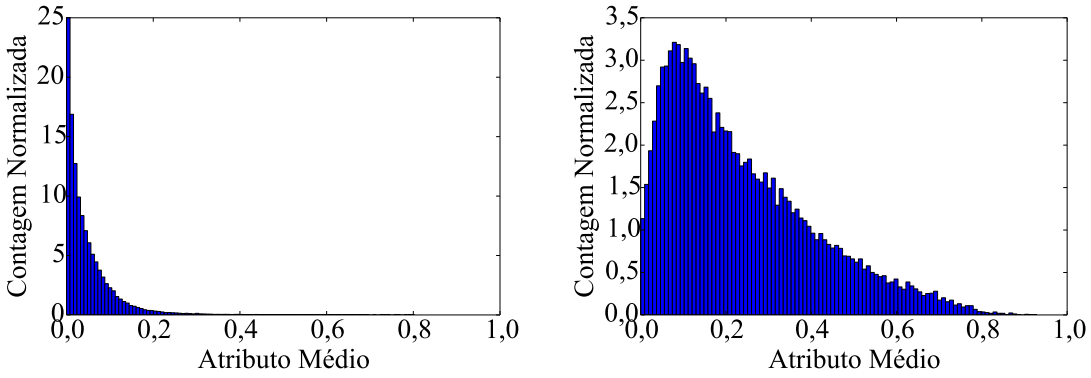
O segundo modelo, então, precisa obter as distribuições

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\text{t}} = 1, I_m^{\text{b}} = 0, I_m^{\text{c}} = 0) \tag{7.30}$$

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\text{t}} = 1, I_m^{\text{b}} = 1, I_m^{\text{c}} = 0) \tag{7.31}$$

$$p_{\vec{\mathbf{y}}_m}(\vec{y}_m|I_m^{\text{t}} = 1, I_m^{\text{b}} = 1, I_m^{\text{c}} = 1), \tag{7.32}$$

ou seja, os modelos para apenas início de tatum, apenas início de tactus e início de



(a) Sem Informação de Tatum

(b) Início de Tatum

Figura 7.14: Histogramas dos valores médio do fluxo espectral normalizado para o modelo de rastreamento do tatum para os dados da partição de treinamento.

compasso. Estas distribuições já foram estudadas para o modelo hierárquico, e os resultados obtidos com os modelos generativos (FA e GMM) podem ser utilizados sem adaptação. No caso dos modelos discriminativos, os dois últimos níveis da SVM hierárquica (SVM2 e SVM3) realizam a classificação necessária, identificando os inícios de tactus dentre os inícios de tatums, e os inícios de compasso dentre os inícios de tactus.

Nas próximas seções, serão apresentados os resultados obtidos para o rastreamento do tatum utilizando apenas a média do fluxo espectral, GMM e SVM.

Modelo Univariado

Seguindo os mesmos passos do modelo de rastreamento apenas do tactus, os dados para o conjunto de treinamento obtido com $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$ foram estudados apenas considerando duas classes: sem informação de tatum e início de tatum (que engloba as classes anteriores apenas início de tatum, apenas início de tactus e início de compasso). Para esse estudo inicial, a média do fluxo espectral normalizado foi obtida para cada dado. Pode ser visto na Figura 7.14 o histograma dos resultados obtidos para cada uma das classes. Os formatos das distribuições são similares aos obtidas anteriormente, com os sem informação de tatum seguindo uma distribuição aparentemente exponencial e os inícios de tatum uma distribuição próxima da gama.

Na Figura 7.15, são mostradas as distribuições ajustadas para os dados da partição de treinamento. Utilizando essas distribuições para se calcular a probabilidade dos dados da partição de validação, chega-se aos seguintes resultados:

- Para $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$: 87 % dos sem informação de tatum e 80 % dos inícios de tatum corretamente classificados;

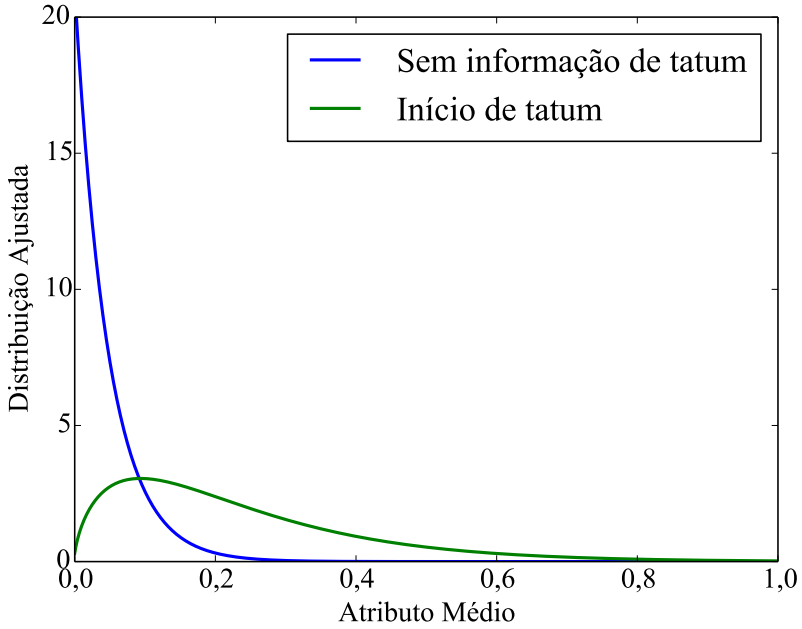


Figura 7.15: Distribuições ajustadas para o modelo considerando apenas o tactus para a partição de treinamento.

- Para $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$: 87 % dos sem informação de tatum e 78 % dos inícios de tatum corretamente classificados.

Com isso, pode ser visto que o modelo considerando apenas a média é capaz de rastrear adequadamente os tatums. Nas próximas seções, são estudados modelos alternativos utilizando a informação espectral completa.

GMM

Foram também realizados testes com GMM para modelar as distribuições dos quadros sem informação de tatum e de início de tatum. Para isso, foram testadas GMMs com 1 até 12 componentes e com matriz de covariância diagonal e cheia. Para cada possibilidade desses parâmetros, foi escolhido o que forneceu o melhor resultado para a partição de treinamento e foi calculado o desempenho para a partição de validação segundo a figura de mérito da Seção 7.3.

Os resultados para a partição de validação foram:

- Para $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$: 86 % sem informação de tatum e 72 % dos inícios de tatum foram corretamente classificados para 8 componentes e matriz de covariância cheia;

- Para $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$: 74 % dos sem informação de tatum e 78 % dos inícios de tatum foram corretamente classificados para 9 componentes e matriz de covariância cheia.

Com isso, pode-se perceber que os resultados são similares, porém um pouco piores, aos obtidos quando se utiliza apenas a média. De forma similar ao que foi observado para o tactus, a informação das sub-bandas Mel não parece melhorar o modelo para apenas início de tatum.

SVM

Foi treinada uma SVM para a detectar quais quadros são de início de tatum e quais não têm informação de tatum. Como este é um problema de classificação binário, apenas uma SVM precisou ser treinada utilizando todos os dados da partição de treinamento.

Usando a partição de validação, 90 % e 78 % dos sem informação de tatum e dos inícios de tatum foram corretamente classificados, respectivamente, para o conjunto com parâmetros $\{N = 40\,\text{ms},\ H = 20\,\text{ms}\}$. Já para o conjunto calculado com $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$, 88 % e 76 % dos sem informação de tatum e dos inícios de tatum foram corretamente classificados, respectivamente. Os resultados obtidos utilizando-se a SVM são similares aos resultados encontrados previamente, sendo o desempenho da SVM um pouco superior ao da GMM e ao do modelo univariado.

7.10 Conclusão

Neste capítulo foram apresentados os modelos de observação para estimação rítmica. Estes modelos têm como objetivo modelar a probabilidade de um determinado atributo observado ter sido gerado num quadro associado a um determinado nível métrico. Para a obtenção destes modelos, o banco métrico foi utilizado fornecendo tanto dados para obtenção do modelo quanto para a sua validação. Uma normalização dos atributos medidos também foi proposta: esta normalização permite a remoção de variações de longo prazo de energia do sinal enquanto preserva a estrutura de curto prazo essencial para a estimação rítmica.

Considerando o desempenho dos modelos obtidos, pode-se perceber a dificuldade de se distinguir entre os diferentes níveis métricos apenas com a informação de um quadro. Tal conclusão reforça a necessidade do uso da informação de periodicidade através das HMMs apresentadas no Capítulo 5. Considerando os modelos treinados, o uso de GMM como um modelo gerador forneceu resultados satisfatórios. Considerando os modelos discriminativos, a SVM foi a solução que forneceu o resultado mais satisfatório.

Os resultados obtidos para os modelos parciais apontam certa facilidade na sua obtenção. Esse resultado é importante principalmente para o modelo por camadas, pois aponta a viabilidade deste modelo como uma alternativa de baixo custo

computacional ao modelo hierárquico completo.

No próximo capítulo, os modelos obtidos serão combinados com os HMMs e o seu desempenho, analisado. Os resultados obtidos para cada teste irão auxiliar na escolha do modelo. Do ponto de vista dos atributos, os obtidos com parâmetros $\{N = 80\,\text{ms},\ H = 40\,\text{ms}\}$ parecem fornecer resultados similares a um custo computacional reduzido. Quando possível, ambos ainda serão analisados para HMMs com menor custo computacional.

Capítulo 8

Avaliação de Desempenho de Modelos para Análise Rítmica

Neste capítulo, será avaliado o desempenho da associação dos modelos de observação descritos no Capítulo 7 com os modelos de rastreamento métrico do Capítulo 5. O principal objetivo deste capítulo é mostrar quais algoritmos de inferência e modelos de observação são mais adequados para cada modelo de rastreamento. Além disso, será realizado um estudo de caso de como o modelo por padrão rítmico pode ser adaptado para um gênero específico que não seria bem modelado por nenhum dos outros modelos propostos.

Dentre os modelos de rastreamento métrico propostos, serão avaliados neste capítulo o modelo de rastreamento do início de tactus e o modelo de rastreamento por camadas. Estes modelos serão analisados utilizando o andamento anotado e também o andamento estimado pelo algoritmo proposto no Capítulo 4. Além disso, três modelos de observação serão estudados para cada algoritmo: o unidimensional, a GMM e a SVM. Por fim, também é estudado o desempenho dos três algoritmos de inferência descritos na Seção 5.1.3: o Viterbi, o *forward-backward*, e o Viterbi Posterior. Para todas essas avaliações será utilizada a partição de validação do banco métrico, descrita na Seção 7.2. Deve-se informar que o objetivo deste capítulo é ilustrar o desempenho dos modelos propostos, sem um ajuste muito fino dos seus parâmetros. Pode-se considerar os resultados apresentados como uma primeira estimativa do desempenho destes métodos, que podem ser refinados através de um ajuste de seus parâmetros ou dos modelos de observação propostos.

Este capítulo é organizado da seguinte forma. Na Seção 8.1 são descritas as figuras de mérito que serão utilizadas para medir o desempenho dos métodos. A Seção 8.2 é dedicada ao modelo de rastreamento do início de tactus. Já na Seção 8.3, é avaliado o modelo de rastreamento por camadas. Um estudo de caso do modelo de rastreamento por padrão rítmico é descrito na Seção 8.4. Por fim, são apresentadas as conclusões deste capítulo na Seção 8.5.

8.1 Figuras de Mérito

Diversas figuras de mérito existem para avaliação de algoritmos de rastreamento do início de tactus [119]. Neste trabalho, quatro figuras de mérito que são adotadas em diversos outros trabalhos [27, 91, 95, 98] serão utilizadas para estudar o desempenho dos algoritmos. Considerando a necessidade de medir o desempenho do rastreamento do tatum e do compasso, além do tactus, a definição dessas figuras de mérito será expandida de forma a incluir também estes níveis métricos. Para isto, na descrição abaixo um "evento rítmico" poderá se referir a um início de tatum, início de tactus ou início de compasso, dependendo do algoritmo sob análise.

Para se chegar às figuras de mérito, é definido inicialmente um critério de aceitação para um evento rítmico estimado. Para isto, o j-ésimo evento rítmico estimado e_j será considerando como correto se existir um evento rítmico anotado a_n tal que sejam respeitados os seguintes critérios [119]:

1. $|e_n - a_j| < \Delta_{a_j} \Upsilon$;
2. $|e_{n-1} - a_{j-1}| < \Delta_{a_{j-1}} \Upsilon$;
3. $\left| \frac{\Delta_{e_n}}{\Delta_{a_j}} - 1 \right| < \Upsilon$.

onde Δ_{a_j} é o intervalo entre a_j e a_{j-1}, Δ_{e_n} é o intervalo entre e_n e e_{n-1} e Υ é uma tolerância cujo valor corresponde a 17,5 %. A Figura 8.1 ilustra como as quantidades referenciadas no critério de aceitação estão interligadas. O primeiro critério garante que o valor estimado está próximo do valor estimado e que essa tolerância é proporcional ao intervalo entre eventos. O segundo critério exige que esta condição também seja verdadeira para o evento estimado precedendo o evento sendo avaliado. Por fim, a terceira condição obriga que o intervalo induzido pelo evento sendo avaliado fique próximo do intervalo entre os eventos anotados.

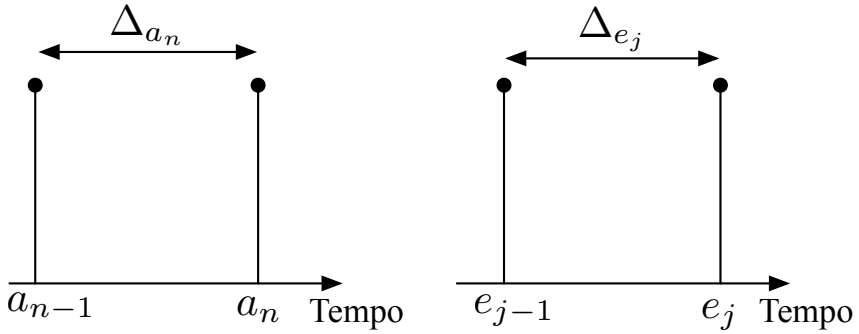


Figura 8.1: Ilustração das quantidade envolvidas na obtenção das figuras de mérito de continuidade.

Considerando os critérios acima, as figuras de mérito utilizadas são definidas como [119]

- CMLc (do inglês, *Correct Metrical Level, continuity required*): maior número de eventos consecutivos detectados corretamente, considerando apenas o nível métrico correto;
- CMLt (do inglês, *Correct Metrical Level, continuity not required*): número total de eventos corretamente detectados no nível métrico correto;
- AMLc (do inglês, *Allowed Metrical Levels, continuity required*): maior número de eventos consecutivos detectados corretamente, em qualquer um dos níveis métricos permitidos;
- AMLt (do inglês, *Allowed Metrical Levels, continuity not required*): número de eventos corretamente detectados, em qualquer um dos níveis métricos permitidos;

Os níveis métricos permitidos são anotações do evento no dobro ou na metade do período anotado e também anotações com o período correto mas marcado entre dois eventos corretos (marcações no "e")[1]. Esses novos níveis métricos são gerados manipulando-se diretamente a sequência de eventos anotados. Como exemplo, uma nova sequência de eventos anotados na metade do período original é gerada selecionando-se todos os elementos ímpares da sequência anotada original. Os mesmos critérios de aceitação descritos acima, então, são utilizados para se verificar o desempenho da sequência estimada considerando-se estas novas sequências anotadas. É escolhido para se calcular as métricas AMLt e AMLc o maior valor entre todos os possíveis níveis métricos.

Por fim, os valores destas figuras de mérito são normalizados pelo número de eventos detectados. Desta forma, um valor de 100 % para um sinal no CMLc indica que todos os eventos detectados correspondem a eventos anotados. Já um valor de 51 % para o CMLc, indica que a maior sequência de eventos corretamente estimados corresponde a 51 % do total de eventos estimados. Em geral, será apresentada a média destes valores para todos os sinais sob análise.

8.2 Rastreamento do Tactus

Nesta seção, será avaliado o algoritmo de rastreamento do início de tactus descrito na Seção 5.3. O motivo de se iniciar com este algoritmo é o fato de ele procurar rastrear apenas um nível métrico, permitindo uma avaliação inicial dos métodos que servirá como base para os algoritmos avaliados posteriormente. Além disso, os

[1]Similarmente ao que ocorre com algoritmos de estimação de andamento, algoritmos de rastreamento do início de tactus tendem a cometer erros no dobro e na metade do andamento correto. Com isso, as figuras de mérito começando por "AML" são similares à Acurácia 2 utilizada na avaliação dos algoritmos de estimação de andamento.

bons resultados preliminares dos modelos de observação indicam que os atributos utilizados permitem uma boa caracterização do tactus.

O atributo utilizado será o extraído com janelas de 80 ms e sobreposição temporal de 40 ms entre janelas adjacentes. Esta escolha foi feita por apresentar um bom resultado no modelo de observação e reduzir a complexidade computacional dos algoritmos de inferência. Três modelos de observação serão comparados nos testes a seguir: o univariado, a GMM e a SVM. O procedimento utilizado para se obter cada um desses modelos foi descrito na Seção 7.9.1. Quanto aos algoritmos de inferência, serão comparados os desempenhos do Viterbi, do *Forward-Backward* e do Viterbi Posterior.

8.2.1 Configuração do Algoritmo

O modelo de rastreamento do início do tactus é composto de três VAs, indicador do tactus, contador do tactus e período do tactus, e é necessário definir os valores assumidos por essas variáveis, assim como a distribuição prior escolhida.

O principal parâmetro livre do modelo são os períodos do tactus que serão explorados pelo algoritmo. Considerando que o número de estados a serem explorados crescem quadraticamente com o número de possíveis períodos (ver Seção 5.3), uma região de busca maior que a necessária pode levar a um aumento substancial na complexidade computacional dos modelos. Considerando esses fatores, foi escolhida uma janela de busca de 300 ms em torno do andamento estimado. Esta duração se mostrou adequada, uma vez que cobre a variação observada para os sinais do banco métrico, conforme pode ser visto na Figura 6.4 na Seção 6.3.

Outro parâmetro importante no algoritmo de inferência é a largura da janela que modela a imprecisão do tactus. Novamente, considerando a análise realizada na Seção 6.3, especialmente na Figura 6.6, foi escolhida uma janela de imprecisão de largura igual a 80 ms.

Por fim, foi utilizada uma prior que inicializa a busca sobre o andamento conhecido.

8.2.2 Desempenho com o Andamento Anotado

Nesta seção serão apresentados resultados em que o andamento utilizado para limitar a região de busca dos períodos e também utilizado na distribuição prior foi estimado como o valor mediano do intervalo entre tactus. Com isso, é isolado o desempenho do modelo de rastreamento do tactus do desempenho do estimador do andamento, permitindo a obtenção de um limite superior do desempenho do algoritmo de rastreamento. Na próxima seção, será feita a análise do desempenho do

rastreador quando este utiliza o andamento estimado pelo algoritmo proposto neste trabalho.

Na Tabela 8.1 podem ser vistas as quatro figuras de mérito obtidas pelo modelo para os três algoritmos de inferência e os três modelos de observação. Conforme pode ser observado, todas as combinações obtiveram resultados positivos, tendo acertado, na maior parte das vezes, mais de 70 % dos tactus se considerado o nível métrico correto. Também pode-se perceber que a GMM obteve os piores resultados e que o modelo univariado forneceu os melhores resultados. Além disso, o algoritmo *forward-backward* teve o pior desempenho, enquanto que o posterior, que impõe restrições de transição sobre a saída do *forward-backward*, obteve o melhor quando o nível métrico correto é avaliado. Já considerando qualquer nível métrico, o algoritmo de Viterbi obteve resultados melhores[2]. De forma geral, os resultados obtidos são satisfatórios exibindo a capacidade do modelo de encontrar os tactus corretos dentro dos sinais da base de validação.

Tabela 8.1: Resultados para o modelo de rastreamento do tactus utilizando o andamento anotado. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	55	74	60	90
FB	49	74	55	80
Posterior	54	77	59	83

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	53	73	57	79
FB	51	73	55	78
Posterior	51	75	56	80

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	53	70	63	81
FB	35	68	43	78
Posterior	51	73	60	83

É também interessante analisar os casos em que o modelo falhou, quando o valor obtido para o CMLt ficou abaixo de 20 %. No total, foram observados 8 sinais (9 % do total) para os quais este fato ocorreu considerando todas as combinações de algoritmos de inferência e modelos de observação testados. Dentre estes sinais, 3

[2]Considerando que o andamento indicado é o correto, os erros de nível métrico obtidos indicam uma tendência do método de escolher a sequência atrasada de meio período da sequência original.

exibem variações abruptas no andamento, que não foram contempladas no modelo. Outros 4 exibem variações constantes no andamento que são de difícil rastreamento. Por outro lado, o modelo de rastreamento do tactus acertou 100 % dos tactus em 25 sinais (28 % do total).

8.2.3 Resultados com o Andamento Estimado

Nesta seção, são apresentados os resultados para o rastreamento do tactus quando são utilizados os andamentos estimados pelo algoritmo proposto[3] no Capítulo 4. Considerando apenas os sinais na partição de validação, o algoritmo proposto para estimação de andamento obteve um desempenho de 76 % e 90 % nas Acurácias 1 e 2, respectivamente. Com isso, espera-se uma queda no desempenho do algoritmo, principalmente para as figuras de mérito CMLc e CMLt.

Na Tabela 8.2, podem ser vistos os resultados obtidos quando é utilizado o andamento estimado. O comportamento dos diferentes algoritmos é similar ao obtido com o andamento anotado, com uma queda de desempenho de aproximadamente 10 pontos percentuais para as figuras de mérito CMLc e CMLt. Para alguns casos, as figuras de mérito AMLc e AMLt mostraram uma melhora no desempenho. A principal razão para este fato é que os atributos extraídos (utilizados para a estimação do andamento e para o rastreamento) realmente apontavam para um tactus num nível métrico diferente do que foi anotado. Dessa forma, ao se inicializar o rastreador com o valor mais apropriado para os dados observados, ele foi capaz de encontrar os melhores candidatos ao tactus, apesar de estarem fora do nível métrico correto.

Os resultados obtidos nessa seção não podem ser diretamente comparados com outros encontrados na literatura devido aos diferentes sinais utilizados na avaliação de desempenho. Com isso, serão apenas descritos resultados de outros algoritmos encontrados na literatura e o banco de sinais utilizado para a sua avaliação. Em [91] é reportado um resultado de 50 % para um banco de 1238 sinais de curta duração e andamento bem comportado. Valores de 63 % são descritos em [98] para um banco de 474 sinais. Mais recentemente, em [95] são apresentados resultados de 83 % para o banco *Ballroom* (que possui sinais com o tactus bem marcado).

8.3 Avaliação do Modelo por Camadas

Nesta seção, é avaliado o desempenho do modelo por camadas apresentado no Capítulo 5.4. Inicialmente, será analisado o desempenho do rastreador do início de

[3]Deve-se notar que a etapa de normalização dos atributos (ver Seção 7.2) também utiliza a informação do andamento. Os atributos utilizados para a obtenção dos resultados nesta seção foram normalizados utilizando-se o andamento estimado.

Tabela 8.2: Resultados para o modelo de rastreamento do tactus utilizando o andamento estimado. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	49	63	64	80
FB	44	63	59	80
Posterior	48	65	61	81

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	47	62	59	77
FB	46	63	59	80
Posterior	46	63	58	77

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	47	59	67	81
FB	30	57	45	76
Posterior	48	65	61	82

tatum para, em seguida, ser feita a análise do desempenho do modelo de rastreamento métrico.

8.3.1 Modelo de Rastreamento do Tatum

O rastreador do tatum é muito similar ao rastreador do tactus cujo desempenho foi descrito na seção anterior. Este modelo possui quatro VAs: período do tatum, período do tactus, indicador do tatum e contador do tatum. Conforme discutido no Capítulo 6, a grande maioria dos sinais no banco métrico exibem uma divisão binária entre o período do tactus e do tatum, conforme discutido na Seção 6.4. Com isto, apenas um valor para essa VA foi avaliado (1/2). Já o período do tactus foi escolhido da mesma forma que o descrito na Seção 8.2. Os resultados apresentados a seguir foram gerados utilizando uma janela de imprecisão de 50 ms, que cobre boa parte da variância do tatum, conforme pode ser visto na Figura 6.8. Por fim, a prior deste modelo inicializa a busca no período do tactus conhecido (que pode ser o anotado ou o estimado).

Considerando os atributos a serem utilizados, testes preliminares mostraram que os atributos extraídos com janelas de 40 ms e saltos de 20 ms são mais adequados para este modelo.

As figuras de mérito que vêm sendo utilizadas até esse momento, foram desenvolvidas para a análise de desempenho de rastreadores do início de tactus e, por

considerarem o erro relativo ao intervalo entre eventos[4], podem não capturar o desempenho do rastreador do início de tatum. Considerando este fato, o critério para aceitação de um início de tatum foi modificado, sendo considerado um início de tatum corretamente detectado qualquer um que esteja dentro de uma janela de 60 ms em torno do anotado. De forma análoga aos CMLc, CMLt, AMLc e AMLt, as novas figuras de mérito serão denominadas $\overline{\text{CMLc}}$, $\overline{\text{CMLt}}$, $\overline{\text{AMLc}}$ e $\overline{\text{AMLt}}$ possuindo a mesma interpretação que as figuras de mérito de continuidade, porém utilizando um critério absoluto de aceitação.

Resultados com Andamento Anotado

Nesta seção, são analisados os resultados do modelo de rastreamento de tatum quando é utilizado o andamento anotado. Neste caso, o procedimento para obtenção do período do tactus foi o mesmo descrito na Seção 8.2.

Podem ser vistos na Tabela 8.3 os resultados obtidos para o modelo de rastreamento de tatum. De forma geral, os resultados foram positivos, mostrando a viabilidade do rastreamento desse nível métrico. No entanto, pode-se notar que o algoritmo comete erros em que foram indicados eventos que estavam entre os desejados. Além disso, a figura de mérito $\overline{\text{CMLc}}$ obteve valores baixos, indicando que mais de um início de tatum foi perdido na maioria dos sinais usados na avaliação. Por fim, o desempenho dos diversos algoritmos foi similar, com o modelo univariado obtendo resultados um pouco melhores que os demais.

Uma forma alternativa de analisar o desempenho do rastreador de tatum consiste em decimar os inícios de tatum detectados num sinal por dois, gerando uma sequência de possíveis inícios de tactus. Esta sequência, então, pode ser analisada usando-se as figuras de mérito AMLc e AMLt, que assim forneceriam um limite superior para o desempenho do modelo de rastreamento métrico. Na Tabela 8.4, podem ser vistos estes valores que indicam que o desempenho de um rastreador do início de tactus utilizando a informação do início de tatum estimado, apesar de obter resultados razoáveis, não será melhor que o do rastreador dedicado ao início de tactus que foi apresentado na Seção 8.2.

Resultados com Andamento Estimado

Nesta seção, o algoritmo de rastreamento do início de tatum é avaliado quando é utilizado o andamento estimado. Neste caso, o início de tatum pode ser inicializado com um valor que é um múltiplo ou submúltiplo do seu valor anotado.

Na Tabela 8.5, podem ser vistos os resultados que demonstram uma queda no desempenho nas figuras de mérito que requerem o nível métrico correto e um leve

[4]Lembrando que o período do tatum é, usualmente, metade do período do tactus.

Tabela 8.3: Resultados para o modelo de rastreamento do tatum utilizando o andamento anotado. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	36	80	59	90
FB	36	80	60	90
Posterior	40	81	61	91

GMM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	31	77	56	90
FB	31	77	56	90
Posterior	34	78	59	90

SVM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	35	78	59	90
FB	35	78	59	90
Posterior	38	79	61	91

Tabela 8.4: Resultados utilizando os candidatos a início de tactus obtidos a partir da estimativa do início de tatum. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	AMLc	AMLt
Viterbi	58	69
FB	45	58
Posterior	59	67

GMM

Algoritmo	AMLc	AMLt
Viterbi	51	62
FB	49	59
Posterior	51	62

SVM

Algoritmo	AMLc	AMLt
Viterbi	54	62
FB	24	44
Posterior	53	64

acréscimo nas demais figuras de mérito. De forma geral, os resultados se mostraram bastante robustos ao fato de ser utilizado o andamento estimado no lugar do andamento correto.

Tabela 8.5: Resultados para o modelo de rastreamento do início de tatum utilizando o andamento estimado.

Univariado

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	36	78	58	89
FB	35	77	60	90
Posterior	37	78	61	91

GMM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	33	77	58	89
FB	31	75	58	89
Posterior	35	77	59	90

SVM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	36	77	59	90
FB	35	76	60	90
Posterior	37	77	61	90

8.3.2 Modelo de Rastreamento Métrico

Nesta seção, é avaliado o modelo de rastreamento métrico (ver Seção 5.4.2) que procura, a partir das estimativas do modelo de rastreamento do início de tatum, encontrar os inícios de tactus e os inícios de compasso.

Considerando o modelo de rastreamento métrico, é necessário apenas definir um parâmetro: o valor para a probabilidade de mudança de período do compasso. Este parâmetro foi escolhido como 0,1 %, levando em conta que esta é a probabilidade de transição de período num dado início de compasso e também considerando como essas transições são raras no banco sob análise (ver Seção 6.5).

A prior do modelo métrico foi definida de duas formas: uma sendo inicializada no período correto para início de compasso e outra onde são atribuídas probabilidades para os três possíveis períodos para o compasso. No caso das probabilidades atribuídas, foram escolhidos os valores 25 %, 5 % e 70 % para períodos binários, ternários e quaternários, respectivamente, que é uma aproximação da taxa em que cada um desses períodos foi encontrado no banco métrico (ver Tabela 6.2) e também observadas apenas na partição de treinamento do banco.

Resultados com Tatum Anotado

Nesta seção, são apresentados os resultados do modelo de rastreamento métrico quando é utilizado o tatum anotado. Dessa forma, o desempenho do modelo de rastreamento métrico pode ser avaliado de forma independente à do modelo de rastreamento do início de tatum. Além disso, essa abordagem fornece um limite superior ao desempenho do modelo de rastreamento métrico.

Serão feitas duas diferentes inicializações do período do compasso: uma no valor correto e outra sem esta informação. Em ambos os casos, será avaliado o desempenho em relação à detecção do início de tactus e do início de compasso. No primeiro caso, será analisada apenas a capacidade do sistema de estimar a fase do tactus e do compasso. No segundo, também será avaliada a capacidade do sistema de escolher o período correto para o compasso.

Nas Tabelas 8.6 e 8.7, são mostrados o desempenho para a estimação do início de tactus e do início de compasso, respectivamente, para janelas de 40 ms e saltos de 20 ms (os mesmos usados na estimação do início de tatum). Nos resultados, pode-se claramente ver a importância do uso da informação espectral para distinguir os quadros que são apenas início de tatum dos que são apenas início de tactus e também início de compasso. O modelo obtido a partir do SVM, particularmente, obteve resultados bastante positivos. Considerando o desempenho para o início de tactus, vemos que o sistema é capaz de encontrar a fase correta do tactus para a maioria dos sinais. Isso indica que, a princípio, se a informação do tatum for confiável, o tactus pode ser bem rastreado. Já o compasso não é tão facilmente rastreado.

Considerando os resultados acima, especialmente para o início de compasso, o mesmo teste foi executado porém utilizando sinais estimados com janelas de 80 ms e saltos de 40 ms, porém mantendo janelas de 40 ms para o rastreamento do tatum. Os resultados para este teste podem ser vistos nas Tabelas 8.8 e 8.9 para o início de tactus e início de compasso, respectivamente. Como pode ser notado, o desempenho do sistema melhora consideravelmente quando é utilizado esse novo tamanho de janela, principalmente para o compasso. Além disso, o bom desempenho do SVM fica mais claro, mostrando que pelo menos metade dos inícios de compasso conseguem ser bem rastreados se o tatum for estimado perfeitamente. De forma geral, o ganho no desempenho mais que justifica o aumento na complexidade computacional de se utilizarem dois conjuntos de atributos: um para o rastreamento de tatum e outro para o de tactus e de compasso.

Os resultados quando o período do compasso não é informado podem ser vistos na Tabela 8.10 para janelas de 80 ms e saltos de 40 ms. Pode-se dizer que esses resultados apontam para uma piora considerável no desempenho do método quando

Tabela 8.6: Resultados para o início de tactus obtidos a partir modelo de rastreamento métrico utilizando o andamento e o período anotados para janelas de 40 ms e saltos de 20 ms. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	57	57	96	96
FB	57	57	74	74
Posterior	80	80	96	96

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	84	84	96	96
FB	84	84	96	96
Posterior	84	84	96	96

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	88	88	97	97
FB	88	88	97	97
Posterior	88	88	96	96

Tabela 8.7: Resultados para o início de compasso obtidos a partir modelo de rastreamento métrico utilizando o início de tatum e período do compasso anotados para janelas de 40 ms e saltos de 20 ms. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	24	25	50	52
FB	23	23	48	49
Posterior	25	25	60	61

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	27	29	44	46
FB	26	28	43	45
Posterior	26	28	43	45

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	43	43	67	69
FB	41	41	66	68
Posterior	41	41	66	69

Tabela 8.8: Resultados para o início de tactus obtidos a partir modelo de rastreamento métrico utilizando o andamento e o período anotados para janelas de 80 ms e saltos de 40 ms. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	80	81	96	97
FB	80	81	96	96
Posterior	80	81	96	97

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	81	82	96	96
FB	81	82	96	96
Posterior	82	82	96	97

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	80	81	96	97
FB	80	81	96	97
Posterior	80	81	96	97

Tabela 8.9: Resultados para o início de compasso obtidos a partir modelo de rastreamento métrico utilizando o início de tatum e período do compasso anotados para janelas de 80 ms e saltos de 40 ms. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	42	41	72	73
FB	42	34	62	63
Posterior	42	42	72	74

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	34	34	64	65
FB	33	34	62	63
Posterior	33	34	63	64

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	52	52	83	84
FB	51	51	80	81
Posterior	51	51	80	81

Tabela 8.10: Resultados para o início de tactus obtidos a partir modelo de rastreamento métrico utilizando o início de tatum anotado porém sem informação do período do compasso. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	30	30	70	71
FB	30	30	70	71
Posterior	30	30	71	71

GMM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	32	32	63	64
FB	31	32	62	63
Posterior	31	32	63	64

SVM

Algoritmo	CMLc	CMLt	AMLc	AMLt
Viterbi	33	33	81	81
FB	33	33	79	79
Posterior	33	33	80	80

o período do compasso não é informado. Isso aponta para a dificuldade de discernir um quadro que é apenas início de tactus de um quadro que é início de compasso, como foi discutido no capítulo anterior. De forma geral, pode-se notar que o uso da informação espectral melhora os resultados (os modelos GMM e SVM obtiveram resultados superiores ao univariado), mas fica claro que são necessários diferentes atributos para capturar o início de compasso.

Resultados com Tatum Estimado

Nesta seção, são apresentados os resultados do rastreamento métrico para quando é estimado o início de tatum. O início de tatum utilizado foi estimado usando o andamento estimado, o modelo de observação gerado pela SVM e o algoritmo de inferência Viterbi posterior. Os atributos utilizados para a estimação do início de tatum foram extraídos com saltos de 20 ms, enquanto que o utilizado para o rastreamento métrico foi extraído utilizando saltos de 40 ms. Neste caso, o quadro mais próximo do início de tatum anotado foi utilizado no rastreamento métrico. O período do compasso também não foi informado ao algoritmo.

Tabela 8.11: Resultados para o início de tactus obtidos a partir modelo de rastreamento métrico utilizando o andamento e o período anotados. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	36	45	58	69
FB	36	50	56	68
Posterior	36	45	58	69

GMM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	40	50	56	68
FB	40	50	56	68
Posterior	40	50	56	68

SVM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	41	51	56	69
FB	41	51	56	69
Posterior	41	51	56	69

Tabela 8.12: Resultados para o início do compasso obtidos a partir modelo de rastreamento métrico utilizando o andamento e o período anotados. Valores médios (em %) para os sinais na partição de validação.

Univariado

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	18	19	49	52
FB	18	19	49	52
Posterior	18	19	50	52

GMM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	25	26	50	52
FB	25	26	49	51
Posterior	25	26	49	51

SVM

Algoritmo	$\overline{\text{CMLc}}$	$\overline{\text{CMLt}}$	$\overline{\text{AMLc}}$	$\overline{\text{AMLt}}$
Viterbi	21	22	56	58
FB	21	22	56	58
Posterior	21	22	56	58

Nas Tabela 8.11 e 8.12, é mostrado o desempenho do algoritmo para os diferentes modelos de observação e algoritmos de inferência para o rastreamento do início do tactus e do início do compasso, respectivamente. Considerando os resultados para o início de tactus, vemos uma queda considerável quando comparado aos resultados obtidos pelo modelo de rastreamento de tactus. A maior razão para isto é da propagação dos erros na estimação do tatum. O mesmo problema é observado na estimação do início do compasso. De forma geral, pode-se dizer que um melhor desempenho do rastreamento do tatum é necessário para viabilizar o rastreamento métrico.

8.4 Estudo de Caso do Modelo por Padrões Rítmicos

Nesta seção[5], será realizado um estudo de caso que ilustra como o modelo por padrões rítmicos pode ser utilizado para o rastreamento de ritmos que não conseguem ser bem modelados pelo modelo hierárquico.

Em particular, será discutido um modelo para o Candombe uruguaio, que apresenta diversos desafios para um algoritmo de rastreamento métrico, como será discutido adiante. Nas próximos seções, será feita uma breve introdução ao Candombe. Em seguida, são propostos um padrão rítmico e um modelo de observação simples para o Candombe. Por fim, é realizada uma breve análise de desempenho utilizando sinais sintéticos e também gravações de campo.

8.4.1 Candombe Uruguaio e seu Padrão Rítmico

O Candombe uruguaio é um ritmo influenciado pela música africana que é usualmente tocado por três tambores de diferentes tamanhos: o *chico* (de menor tamanho), o *piano* (de maior tamanho) e o *repique* (de tamanho intermédiario). Um determinado conjunto de Candombe pode ser formado por 1 até diversos tambores de cada um dos tipos. Tradicionalmente, os músicos tocam Candombe enquanto se movem e os passos marcam o início de tactus, que nem sempre é acentuado [120].

O papel do *piano* e do *chico* são bem definidos, formando uma base rítmica, enquanto o *repique* é livre para improvisar. Na Figura 8.2 é possível ver o padrão destes dois tambores, que pode ser considerado constante durante uma dada execução. Como pode ser observado, um compasso de Candombe contém 16 tatums e 4 tactus. Este padrão também é executado com períodos muitos curtos, sendo um andamento usual para o início de tactus entre 110 e 150 BPM, o que leva a uma

[5]Os resultados exibidos nesta seção foram desenvolvidos em colaboração com Prof. Martín Rocamora, da UdelaR, Uruguai.

taxa de até 600 tatums por minuto. Além desses padrões, um padrão de *clave* (batida na lateral) pode ser executado pelos tambores. Este padrão por vezes ajuda a sincronizar os tambores e auxilia na definição do pulso da execução [120].

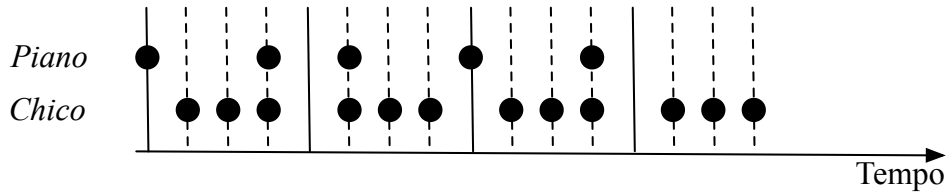


Figura 8.2: Padrões para os dois tambores de candombe que formam a base rítmica. As linhas tracejadas verticais marcam os tatum e as sólidas, o início de tactus.

Observando o padrão, pode-se perceber que dois dos quatro inícios de tactus não são acentuados, o que dificulta o uso de um algoritmo desenvolvido para música popular ocidental, que assume que todos os inícios de tactus são acentuados[6]. Desta forma, o modelo por padrão rítmico se torna essencial, pois permite a codificação dessa esperada falta de acentuação. Na próxima seção, será discutida a adaptação do modelo para o rastreamento de Candombe.

8.4.2 Adaptação do Algoritmo

Os resultados apresentados na próxima seção considerarão que há apenas um padrão rítmico que será definido como padrão de acentuação, informando quando é esperado um toque de tambor.

Com isso, o padrão para o Candombe é definido com um padrão de comprimento $L = 16$ e multiplicador do andamento $\Theta = 1/4$. O conjunto de parâmetros é definido como

$$A = [1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1], \tag{8.1}$$

onde os elementos irão definir quais são os tatums que são acentuados dentro do padrão. Note que este padrão é a combinação dos padrões do *piano* e do *chico* exibidos na Figura 8.2.

O atributo a ser utilizado é o fluxo espectral calculado com janelas de 40 ms e saltos de 20 ms e em 40 sub-bandas Mel. É utilizado como observação o valor médio ao longo das sub-bandas, após a normalização descrita na Seção 7.2. Nesta seção

[6]Na realidade, a marcação do início de tactus é difícil para pessoas que não são familiarizadas com o ritmo.

será empregado um modelo de observação simples[7], definido pela seguinte expressão:

$$p_{\mathbf{y}_m}(y_m|I_m^{\mathrm{r}}, a_m^{\mathrm{r}}) = \begin{cases} N_{\sigma_t}(y_m), \text{ se } I_m^{\mathrm{r}} = 1 \\ N_{\sigma_t}(y_m - A_{a_m}), \text{ se } I_m^{\mathrm{r}} = 1, \end{cases} \tag{8.2}$$

onde $N_{\sigma_t}(.)$ é uma gaussiana de média zero e variância σ_t. Dessa forma, espera-se um valor próximo de zero para ausência de informação rítmica. Para o modelo de observação do início de tatum, o valor médio esperado depende da posição do tatum atual dentro do padrão. Para os tatums acentuados, é esperado um valor médio 1; caso contrário, o valor médio é 0.

Nos resultados apresentados a seguir, será utilizada uma janela de 60 ms para a imprecisão do tatum. Além disso, $\sigma_t = 0.5$, de forma que o modelo de observação considera uma sobreposição considerável entre os valores observados quando há e não há acentuação. Além disso, o andamento é assumido como conhecido.

8.4.3 Resultados

Um conjunto de experimentos foi desenvolvido para avaliar o algoritmo proposto num ambiente controlado utilizando sinais sintéticos de padrões de Candombe e duas gravações de campo. O primeiro conjunto evita alguns aspectos problemáticos de gravações de campo, como reverberação e ruído, e simplifica o processo de criação de anotações. O segundo permite a validação do método em um cenário mais realista. O desempenho do algoritmo será avaliado apenas considerando o desempenho do rastreamento de tactus.

Análise de Padrões Simples

O primeiro passo neste estudo consiste em gerar padrões rítmicos básicos de Candombe e avaliar o desempenho do algoritmo para cada sinal. Os resultados são mostrados na Figura 8.3, para os primeiros 4 compassos de cada sinal (de um total de aproximadamente 14 compassos). Os padrões do *chico* e *clave* foram corretamente extraídos quando eles aparecem sozinhos ou combinados com outros padrões. O último exemplo considera a adição do padrão de piano, gerando a base rítmica mais usual do Candombe. Note que todos os sinais iniciam com a marcação da *clave*, como numa execução real.

O algoritmo proposto obtém nota máxima (100 %) para todos os sinais de exemplo com as figuras de mérito utilizadas, o que demonstra a capacidade do método de rastrear estes padrões simples, apesar de eles não serem exatamente idênticos ao

[7] Diferentemente dos outros modelos, a quantidade de sinais anotados disponíveis até o momento da execução deste trabalho não permitiu o desenvolvimento de um modelo de observação mais avançado.

padrão rastreado. É especialmente notável a capacidade do algoritmo de encontrar a fase correta no padrão de *chico*, onde nenhum dos inícios de tactus é acentuado. Um algoritmo convencional de rastreamento de início de tactus iria provavelmente acertar o período do tactus, mas certamente errar a sua fase.

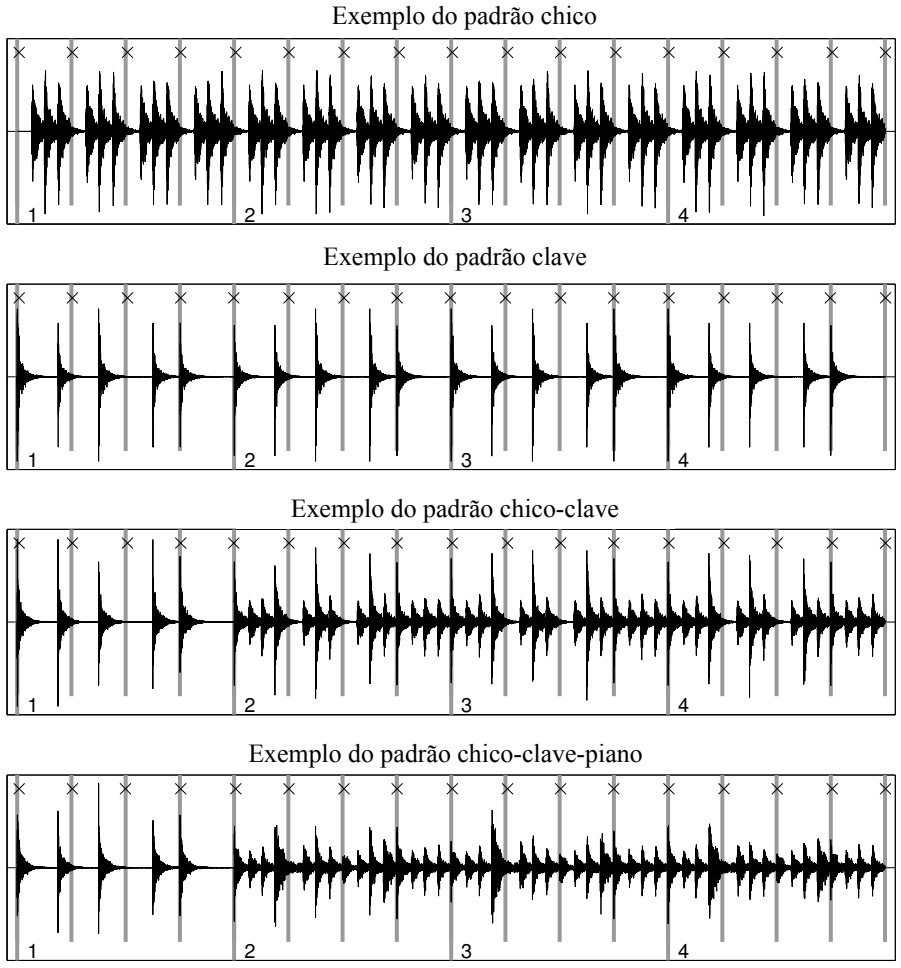


Figura 8.3: Exemplos do desempenho do algoritmo de rastreamento por padrões rítmicos para padrões simples de Candombe. Os inícios de tactus anotados são marcados pelas linhas verticais, enquanto os inícios de tactus estimados são marcados por 'x'.

Resultados Utilizando Sinais Sintéticos

Os sinais apresentados na seção anterior são simples por diversos aspectos

- O tambor *repique*, responsável pela improvisação, não está presente;
- Existem diversos padrões de *clave* que podem ser utilizados, mas apenas o mais comum foi avaliado;
- O padrão do *piano* pode sofrer alterações durante uma execução.

Considerando esses fatores, foram criados dois exemplos sintéticos que procuram simular uma execução real. As partes do *repique* foram criadas com base em transcrições de execuções realizadas por músicos renomados. Para o início de cada sinal, foram considerados dois casos entre os mais comuns: um em que todos os tambores tocam o padrão da *clave* (exemplo 1) e outro em que o *piano* começa em anacruse seguido por um padrão de *clave* (exemplo 2). A duração de ambos os sinais é de 65 s e o seu andamento é de 136 BPM, sem variações durante a execução (porém imprecisões nas batidas são modeladas pelo algoritmo de síntese). Tais escolhas geraram sinais mais simples que os encontrados, porém ainda apresentando um certo grau de desafio ao algoritmo. Além disso, na próxima seção serão consideradas gravações reais que incluem variações de andamento.

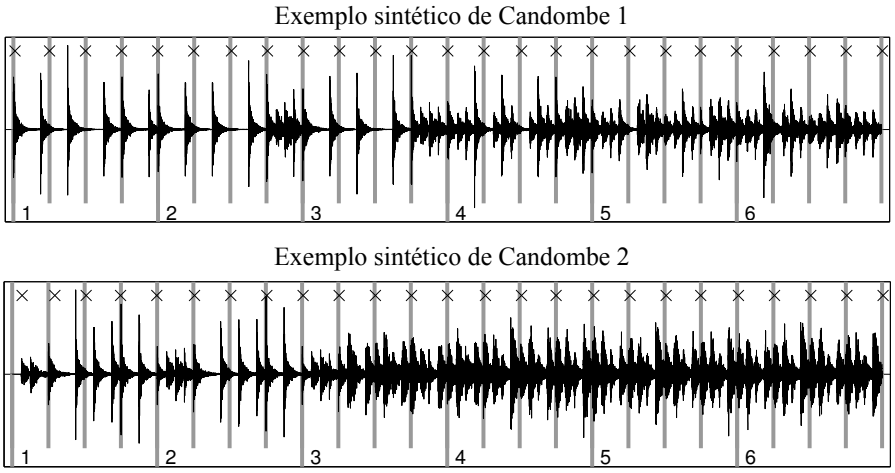


Figura 8.4: Seis primeiros compassos dos dois exemplos sintéticos de Candombe os inícios de tactus anotados (barras verticais) e estimados ('x').

São mostrados na Figura 8.4 os seis primeiros compassos de cada exemplo sintético com os inícios de tactus anotados e estimados. Apesar da complexidade maior destes exemplos, o algoritmo obteve um desempenho similar ao da seção anterior: 100 % e 98,7 % em todas as figuras de mérito para os exemplos 1 e 2, respectivamente. Note na Figura 8.4 que os dois primeiros tactus do exemplo 2 são deslocados devido à presença da anacruse, mas o algoritmo rapidamente se adapta ao tactus correto e permanece na fase correta até o fim do sinal.

Resultados Utilizando Sinais Gravados

Finalmente, duas gravações de campo são usadas para ilustrar o funcionamento do algoritmo proposto. Ambas as gravações possuem um andamento de 130 BPM, e foram executadas por diferentes grupos de músicos.

Como mostrado na Figura 8.5, o algoritmo é capaz de rastrear o tactus do primeiro exemplo corretamente, com um resultado igual a 100 % em todas as figuras de

mérito. Já o segundo exemplo apresenta maior dificuldade de análise, com todos os tambores sendo executados desde o início da gravação. Além disso, o padrão do *repique* é mais complexo do que nos sinais estudados até aqui, sendo menos repetitivo e contendo mais improvisações. Estas características atrapalham o algoritmo, que perde a fase do sinal em dois momentos, sendo um deles exibido na Figura 8.5. O resultado para este sinal é igual a 45,2 % para as métricas que exigem continuidade (CMLc e AMLc) e 60.6 % para as demais métricas (CMLt e AMLt).

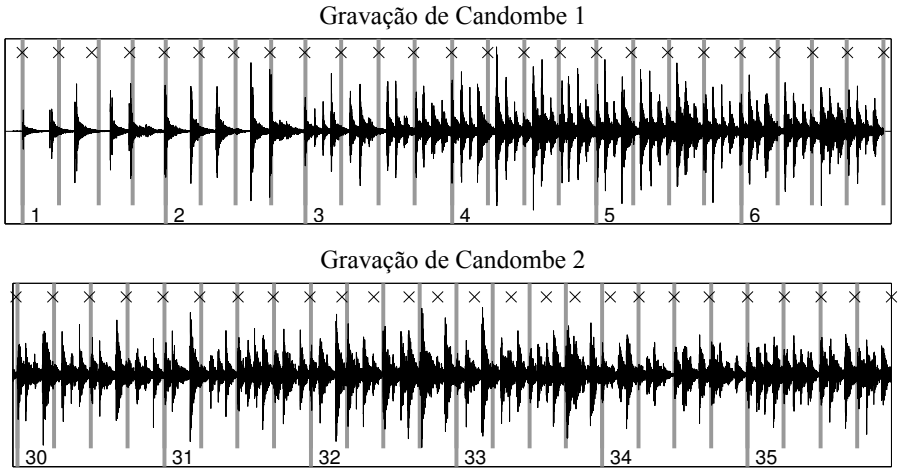


Figura 8.5: Seis compassos de duas gravações de Candombe com os inícios de tactus anotados (barras verticais) e estimados ('x').

8.5 Conclusão

Neste capítulo, foi realizado um estudo do desempenho dos modelos propostos para rastreamento métrico. Em particular, foram analisados os desempenhos de três modelos de observação e de três algoritmos de inferência. Considerando os resultados apresentados, os modelos de observação usando SVMs apresentaram bons resultados em todos os casos. Já nos algoritmos de inferência, vê-se uma pequena vantagem em se utilizar o Viterbi posterior, que no entanto deve ser ponderada pelo custo computacional adicional de executá-lo.

Considerando os modelos de rastreamento, a solução proposta é promissora quando combinada aos modelos de observação gerados e ao método de estimação do andamento. Em particular, o modelo de rastreamento do tactus mostrou resultados satisfatórios. Já o algoritmo de rastreamento do tatum precisa ser melhor ajustado, considerando os resultados obtidos. Em contrapartida, quando o tatum é corretamente detectado, o algoritmo de rastreamento métrico é uma opção de baixo custo computacional para a estimação do início de tactus e do início de compasso. Nos resultados obtidos, também ficou clara a necessidade de atributos complementares

aos utilizados que sejam mais adequados ao rastreamento do tatum e do compasso.

Por fim, a versatilidade do modelo por padrão rítmico foi demonstrada por um estudo de caso envolvendo o ritmo uruguaio Candombe. Em particular, foi discutido como padrões rítmicos complexos podem ser codificados neste modelo, permitindo que seja realizada uma busca por padrões de acentuação que seriam muito mal modelados por algoritmos que não levam em consideração a particularidade de um determinado ritmo. Notadamente, o ritmo estudado possui tactus que não são acentuados, que dificultariam a análise pela maioria dos algoritmos encontrados na literatura. No modelo por padrão rítmico, essa informação foi agregada ao modelo, auxiliando-o.

Capítulo 9

Conclusão

Esta tese apresentou soluções para algoritmos de análise rítmica computacional. Em particular, foram propostas soluções para estimação do andamento e rastreamento métrico de sinais de música. Dentre as principais contribuições pode-se citar

1. Estudo da melhor forma de obtenção do fluxo espectral;
2. Proposta de um método de estimação de andamento que inclui o uso de informação de padrões rítmicos e curvas de ponderação, além de um método eficiente para o cálculo da periodicidade de um sinal de áudio;
3. Criação de modelos para o rastreamento rítmico que permitem a estimação dos três níveis métricos, considerando variações lentas no andamento e mudanças de métrica;
4. Criação e estudo de uma base de sinais em que os níveis métricos foram anotados por um músico profissional;
5. Estudo de diferentes metodologias para obtenção de modelos probabilísticos para os atributos utilizados, incluindo modelos que exploram a informação espectral.

A análise sobre a forma do cálculo espectral foi publicada em [79], enquanto o algoritmo de rastreamento de tactus foi submetido para publicação [121]. A seguir, serão descritas em detalhes as principais contribuições do trabalho, juntamente com possíveis continuações da pesquisa.

9.1 Estimação do Andamento

A estimação de andamento foi utilizada para realizar um estudo inicial de técnicas para análise métrica. Em particular, o fluxo espectral foi estudado em detalhe e foram comparadas diferentes maneiras de computá-lo. Tal tarefa permitiu a avaliação

de diferentes formas de se estimar o andamento de um sinal e como elas contribuíam para a melhoria do desempenho de um sistema de análise rítmica.

Considerando os resultados da análise, então, foram propostas diferentes modificações e melhorias sobre um algoritmo de estimação de andamento. Em particular, foi descrita a separação transitório/permanente como um possível pré-processamento, uma forma mais eficiente e precisa de cálculo da função de periodicidade que utiliza a Transformada de Fourier de Tempo Discreto e padrões rítmicos. Em geral, o algoritmo proposto apresentou um bom resultado, com um desempenho similar ou superior ao de métodos descritos na literatura.

O conhecimento adquirido nesta etapa foi fundamental no desenvolvimento dos algoritmos de rastreamento métrico da segunda parte deste trabalho. Considerando a continuação desta pesquisa, pode-se pensar em integrar a estimação de andamento com os modelos de rastreamento. Neste caso, uma função de periodicidade computada para um trecho do sinal poderia ser utilizada para atualizar as estimativas do período num determinado quadro. Além disso, utilizando o formalismo introduzido na segunda etapa, modelos de observação para funções de periodicidade podem ser obtidos para diferentes métricas, vindo a substituir os padrões rítmicos empregados.

9.2 Rastreamento Métrico

A principal contribuição desta tese para rastreamento métrico foi o desenvolvimento do modelo hierárquico, utilizando modelos ocultos de Markov, que serviu de base conceitual para os demais modelos propostos. O uso de HMMs definidos a partir de diversas variáveis aleatórias permite uma melhor compreensão de como essas variáveis interagem na formação métrica, permitindo formalizar de diversas heurísticas utilizadas em outros trabalhos. Além disso, utilizando os algoritmos de inferência descritos, é possível realizar inferência de forma eficiente e com um entendimento claro das quantidades que estão sendo maximizadas. Por fim, o modelo proposto se mostrou flexível o suficiente para gerar sub-modelos com complexidade computacional reduzida.

Considerando os objetivos da tese, uma base descrita na literatura teve seus três níveis métricos anotados, permitindo o desenvolvimento dos algoritmos propostos. Uma análise das anotações foi realizada e utilizada para corroborar e auxiliar a escolha de parâmetros dos modelos propostos. Por fim, por ser anotado de forma consistente por um músico profissional, este banco pode servir como base para diversos futuros trabalhos de rastreamento métrico, especialmente considerando que este é um dos poucos bancos (se não o único) a conter todos os três níveis métricos anotados.

Usando os dados da base e as definições propostas nos modelos rítmicos, foram

propostos diversos modelos de observação para os três níveis métricos. Em particular, foi descrita uma nova normalização para os atributos que reduz problemas relacionados a variações de dinâmica de longo prazo. Dentre os modelos de observação propostos, devem-se destacar os modelos que procuram utilizar informação espectral para diferenciar ocorrências dos diferentes níveis métricos. Essa informação se mostrou importante principalmente para a distinguir inícios de compasso de inícios de tactus, e indica um importante caminho futuro no desenvolvimento de novos atributos específicos para a detecção do compasso. Além disso, foi proposta uma abordagem utilizando classificação que se mostrou bastante promissora.

Por fim, foi realizado um estudo do desempenho dos diferentes métodos propostos comparando-se diferentes algoritmos de inferência. Os resultados obtidos demonstram a viabilidade dos métodos propostos, em particular para o rastreamento do tactus. Dentro os modelos de observação, o univariado (que se baseia na média dos atributos) se mostrou particularmente adequado para os níveis métricos que só precisam se diferenciar da ausência de informação rítmica (inícios vs não-início de tatum ou inícios vs não-inícios de tactus). Já os modelos de observação que empregam a informação espectral obtiveram melhor desempenho na diferenciação entre diferentes níveis métricos, principalmente entre o início de tactus e o início de compasso. Por fim, o algoritmo de inferência Viterbi posterior se mostrou uma alternativa ao Viterbi, com a vantagem do conhecimento da probabilidade posterior de cada estado estimado mas a desvantagem de uma complexidade computacional mais elevada.

Como trabalho futuro, novos atributos que melhor caracterizem o início do compasso podem ser agregados ao modelo. Outra abordagem seria utilizar um método de aprendizado para extrair a informação relevante diretamente dos sinais de áudio ou da STFT do sinal, como descrito em [122]. Graças ao formalismo do modelo hierárquico descrito, estratégias como essas podem ser empregadas e comparadas de forma mais direta. Os modelos propostos podem incorporar informação de mudanças de andamento (como discutido na seção anterior) e também podem ser ampliados para incluir a estimação de outros parâmetros num sinal de áudio como, por exemplo, *onsets* e acordes. Neste caso, modelos específicos para estes eventos podem ser desenvolvidos e acoplados ao modelo hierárquico, permitindo a estimação conjunta destes fenômenos.

Apêndice A

Mapeamento de Gêneros Utilizado

Os sinais dos bancos *Ballroom* e *Hainsworth* foram classificados de acordo com o seu gênero pelos autores de cada um dos bancos. As classificações utilizadas para cada banco, no entanto, são muito diferentes, sendo que sinais que pertenceriam a grupos distintos do *Ballroom* apareceriam no mesmo grupo do *Hainsworth*, e vice-versa. Com isso, se faz necessário um reagrupamento dos gêneros de forma a compatibilizar os dois bancos de sinais.

Podem ser vistos na Tabela A.1, os novos gêneros e os gêneros anotados correspondentes aos bancos *Hainsworth* e *Ballroom*. Como pode ser observado, os gêneros finais são mais abrangentes, sempre correspondendo a um ou mais gêneros anotados nos bancos de sinais.

Tabela A.1: Mapeamento entre os gêneros utilizados e os gêneros anotados no *Ballroom* e *Hainsworth*.

Gênero Final	Gêneros do *Ballroom*	Gêneros do *Hainsworth*
Jazz/Blues	Jive, Quickstep	Jazz, BigBandJazz
Rock/Pop		Rock/Pop, Pop60s
Eletrônica		Dance
Clássica	Waltz, Viennesse-Waltz	Choral, Classical, ClassicalSolo
Folk		AcousticFolk, Folk
Latina	Rumba-American, Samba, Rumba-International, Rumba-Misc, Tango	

Referências Bibliográficas

[1] LONDON, J. *Hearing in Time: Psychological Aspects of Musical Meter*. London, UK, Oxford University Press, 2004.

[2] PEETERS, G. "Template-Based Estimation of Time-Varying Tempo", *EURASIP Journal on Advances in Signal Processing*, v. 2007, n. 67215, pp. 1–14, December 2007.

[3] COOPER, G., MEYER, L. B. *The Rhythmic Structure of Music*. Chicago, USA, The University of Chicago Press, 1960.

[4] SCHOLES, P. A. *Oxford Companion to Music*. London, UK, Oxford University Press, 1970.

[5] PARNCUTT, R. "A Perceptual Model of Pulse Salience and Metrica Accent in Musical Rhythms", *Music Perception*, v. 11, n. 4, pp. 409–464, November 1994.

[6] BETTERMANN, H., CYSARZ, D., LEEUWEN, P. V. "Detecting cardiorespiratory coordination by respiratory pattern analysis of heart period dynamics – the musical rhythm approach", *International Journal of Bifurcation & Chaos*, v. 10, n. 10, pp. 2349–2360, October 2000.

[7] MICHELS, U. *Guide Illustré de la Musique*. Paris, France, Fayard, 1998.

[8] ABROMONT, C., DE MONTALEMBERT, E. *Guide de La Theorie de la Musique*. Paris, France, Fayard, 2001.

[9] VINKE, L. N. *Factors Affecting The Perceived Rhythmic Complexity Of Auditory Rhythms*. Master Thesis, Graduate College of Bowling Green State University, Bowling Green, USA, 2010.

[10] POVEL, D.-J., ESSENS, P. "Perception of Temporal Patterns", *Music Perception*, v. 2, n. 4, pp. 411–441, Summer 1985.

[11] SETHARES, W. A. *Rhythm and Transforms*. London, UK, Springer-Verlag, 2007.

[12] POVEL, D.-J. "A Theoretical Framework for Rhythm Perception", *Psychological Review*, v. 45, pp. 315–337, March 1984.

[13] JONES, M. R. "Time, our lost dimension: Toward a new theory of perception, attention, and memory", *Psychological Review*, v. 83, n. 5, pp. 323–355, September 1976.

[14] LEVY, M. "Improving Perceptual Tempo Estimation With Crowd-Sourced Annotations". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pp. 317–322, Miami, USA, October 2011.

[15] DOWLING, J. W., HARWOOD, D. L. *Music Cognition*. Orlando, USA, Academic Press, 1986.

[16] BROWN, J. C. "Determination of the Meter of Musical Scores by Autocorrelation", *Journal of the Acoustical Society of America*, v. 94, n. 4, pp. 1953–1957, October 1993.

[17] HONING, H. "Issues on the Representation of Time and Structure in Music", *Contemporary Music Review*, v. 9, n. 1, pp. 221–238, January 1993.

[18] DESAIN, P., HONING, H. "Computational Models of Beat Induction: The Rule-Based Approach", *Journal of New Music Research*, v. 28, n. 1, pp. 29–42, January 1999.

[19] TEMPERLEY, D. *Music and Probability*. Cambridge, USA, The MIT Press, 2006.

[20] MIDI MANUFACTURERS ASSOCIATION. "The Standard MIDI Files (SMF) Specification". Standard, January 1998.

[21] DIXON, S. "Automatic Extraction of Tempo and Beat from Expressive Performances", *Journal of New Music Research*, v. 30, n. 1, pp. 39–59, January 2001.

[22] GOUYON, F. *A Computational Approach to Rhythm Description: Audio Features for the Computation of Rhythm Periodicity Functions and Their Use in Tempo Induction and Music Content Processing*. Phd thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2005.

[23] SCHEIRER, E. D. "Tempo and Beat Analysis of Acoustic Musical Signals", *Journal of the Acoustical Society of America*, v. 103, n. 1, pp. 588–601, January 1998.

[24] ALONSO, M., DAVID, B., RICHARD, G. "Tempo and Beat Estimation of Musical Signals". In: *Proceedings of the International Conference on Music Information Retrieval*, pp. 158–163, Barcelona, Spain, October 2004.

[25] ELLIS, D. P. W. "Beat Tracking by Dynamic Programming", *Journal of New Music Research*, v. 36, n. 1, pp. 51–60, March 2007.

[26] GOUYON, F., DIXON, S., WIDMER, G. "Evaluating Low-Level Features for Beat Classification and Tracking". In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 4, pp. 1309–1312, April 2007.

[27] DEGARA, N., RUA, E., PENA, A., et al. "Reliability-Informed Beat Tracking of Musical Signals", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 20, n. 1, pp. 290–301, January 2012.

[28] SEPPÄNEN, J. *Computational Models Of Musical Meter Recognition*. M.Sc. Thesis, Tampere University of Technology, Tampere, Finland, November 2001.

[29] SEPPÄNEN, J., ERONEN, A., HIIPAKKA, J. "Joint Beat & Tatum Tracking from Music Signals". In: *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, October 2006.

[30] PAULUS, J., KLAPURI, A. "Music Structure Analysis by Finding Repeated Parts". In: *Proceedings of the 1st Audio and Music Computing for Multimedia Workshop*, pp. 59–68, Santa Barbara, USA, October 2006.

[31] PEETERS, G. "Sequence Representation of Music Structure Using Higher-Order Similarity Matrix and Maximum-Likelihood Approach". In: *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, September 2007.

[32] PAULUS, J., KLAPURI, A. "Music Structure Analysis Using a Probabilistic Fitness Measure and a Greedy Search Algorithm", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 17, n. 6, pp. 1159–1170, August 2009.

[33] HEUSDENS, R., JENSES, J., KLEIJN, W. B., et al. "Bit-Rate Scalable Intra-frame Sinusoidal Audio Coding Based on Rate-Distortion Optimization", *Journal of the Audio Engineering Society*, v. 54, n. 3, pp. 167–188, March 2006.

[34] Klapuri, A., Davy, M. (Eds.). *Signal Processing Methods for Music Transcription*. New York, NY, USA, Springer, 2006.

[35] DINIZ, F. C. C. B. *Transcrição Musical Automática Usando Representação Frequencial Eficiente por Banco de Filtros de Alta Seletividade*. D.sc. Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, August 2009.

[36] GOUYON, F., DIXON, S., PAMPALK, E., et al. "Evaluating Rhythmic Descriptors For Musical Genre Classification". In: *Proceedings of the 25th International AES Conference*, pp. 1–9, London, UK, June 2004.

[37] PAULUS, J., KLAPURI, A. "Measuring the Similarity of Rhythmic Pattern". In: *Proceedings of the International Symposium on Music Information Retrieval*, Paris, France, October 2002.

[38] HOCKMAN, J. A., BELLO, J. P. "Automated Rhythmic Transformation Of Musical Audio". In: *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 2008.

[39] ORIO, N., LEMOUTON, S., SCHWARZ, D. "Score Following: State of the Art and New Developments". In: *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*, pp. 36–41, Montreal, Canada, May 2003.

[40] LAROCHE, J. "Efficient Tempo and Beat Tracking in Audio Recordings", *Journal of the Audio Engineering Society*, v. 51, n. 4, pp. 226–233, April 2003.

[41] ZAPATA, J. R., GÓMEZ, E. "Comparative Evaluation and Combination of Audio Tempo Estimation Approaches". In: *Proceedings of the 42nd International Conference of the Audio Engineering*, Ilmenau, Germany, July 2011.

[42] OLIVEIRA, J. L., GOUYON, F., MARTINS, L. G., et al. "IBT: A Real-Time Tempo And Beat Tracking System". In: *Proceedings of the 11th International Society for Music Information Retrieval*, Utrecht, Netherlands, October 2010.

[43] GKIOKAS, A., KATSOUROS, V., CARAYANNIS, G. "Ilsp Audio Tempo Estimation Algorithm For Mirex 2011". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Miami, USA, October 2011.

[44] BROWN, J. C., PUCKETTE, M. S. "A High Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform", *Journal of the Acoustical Society of America*, v. 94, n. 2, pp. 662 – 667, August 1993.

[45] GKIOKAS, A., KATSOUROS, V., CARAYANNIS, G. "Tempo Induction Using Filterbank Analysis And Tonal Features". In: *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pp. 555–558, Utrecht, Netherlands, October 2010.

[46] ANTONOPOULOS, I., PIKRAKIS, A., THEODORIDIS, S. "Self-Similarity Analysis Applied on Tempo Induction from Music Recordings", *Journal of New Music Research*, v. 36, n. 1, pp. 27–38, March 2007.

[47] LOGAN, B. "Mel Frequency Cepstral Coefficients for Music Modeling". In: *Proceedings of the International Symposium on Music Information Retrieval*, pp. 1–11, Plymouth, USA, October 2000.

[48] GROSCHE, P., MÜLLER, M. "A Mid-Level Representation for Capturing Dominant Tempo And Pulse Information In Music Recordings". In: *Proceedings of the 10th International Society for Music Information Retrieval Conference*, pp. 189–194, Kobe, Japan, October 2009.

[49] DAVIES, M. E. P., PLUMBLEY, M. D. "Context-Dependent Beat Tracking of Musical Audio", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 15, n. 3, pp. 1009–1020, March 2007.

[50] KLAPURI, A., ERONEN, A., ASTOLA, J. "Analysis of the meter of acoustic musical signals", *Audio, Speech, and Language Processing, IEEE Transactions on*, v. 14, n. 1, pp. 342–355, January 2006.

[51] LARTILLOT, O. "MIRTEMPO: Tempo Estimation Through Advanced Frame-By-Frame Peaks Tracking". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[52] DIXON, S. "Onset Detection Revisited". In: *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, pp. 133–137, Montreal, Canada, September 2006.

[53] ALONSO, M., RICHARD, G., DAVID, B. "Tempo Estimation for Audio Recordings", *Journal of New Music Research*, v. 36, n. 1, pp. 17–25, March 2007.

[54] DAVIES, M., ROBERTSON, A., PLUMBLEY, M. "Mirex 2009 Audio Beat Tracking Evaluation: Davies, Robertson and Plumbley". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Kobe, Japan, October 2009.

[55] TZANETAKIS, G. "MARSYAS Submission to MIREX 2010". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[56] AYLON, E., WACK, N. "Beat Detection Using PLP". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[57] PEETERS, G. "Spectral and Temporal Periodicity Representations of Rhythm for the Automatic Classification of Music Audio Signal", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 19, n. 5, pp. 1242–1252, July 2011.

[58] WU, F.-H. F. "A Statistic Learning Approach To Tempo Estimation For Audio Music". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Miami, USA, October 2011.

[59] HARRIS, F. J. "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform", *Proceedings of the IEEE*, v. 66, n. 1, pp. 51–83, January 1978.

[60] PEETERS, G. "Mirex-09 "Audio Beat Tracking" Task: ircambeat Submission". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Kobe, Japan, October 2009.

[61] EYBEN, F., SCHULLER, B. "Tempo Estimation from Tatum and Meter Vectors (MIREX 2010 Submission)". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[62] PEETERS, G. "Mirex-2010 "Audio Beat Tracking" Task: Ircambeat Submission". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, October 2010.

[63] PEETERS, G. "MIREX-2011 "Audio Beat Tracking" Task: Ircambeat Submission". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Miami, USA, October 2011.

[64] MOORE, B. C. J. *An Introduction to the Psychology of Hearing*. Fifth ed. New York, USA, Elsevier, 2004.

[65] RABINER, L., SCHAFER, R. W. *Theory and Application of Digital Speech Processing*. Upper Saddle River, USA, Prentice Hall, 2010.

[66] SCHULLER, B., EYBEN, F., RIGOLL, G. "Fast and Robust Meter and Tempo Recognition for the Automatic Discrimination of Ballroom Dance Styles". In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 217–220, Honolulu, USA, April 2007.

[67] GKIOKAS, A., KATSOUROS, V., CARAYANNIS, G. "Audio Tempo Extraction Algorithm for MIREX 2010". In: *Proceedings of Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[68] BÖCK, S., EYBEN, F., SCHULLER, B. "MIREX 2010 Submission: Tempo Detection With Bidirectional Long Short-Term Memory Neural Networks". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, August 2010.

[69] BÖCK, S., EYBEN, F., SCHULLER, B. "MIREX 2010 SUBMISSION: Beat Detection With Bidirectional Long Short-Term Memory Neural Networks". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, October 2010.

[70] BÖCK, S., SCHEDL, M. "Enhanced Beat Tracking With Context-Aware Neural Networks". In: *Proceedings of the 14th International Conference on Digital Audio Effects*, pp. 135–139, Paris, France, September 2011.

[71] AYLON, E., WACK, N. "Beat Detection Using PLP". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, October 2010.

[72] PEETERS, G. "Template-Based Estimation of Tempo: Using Unsupervised or Supervised Learning To Create Better Spectral Templates". In: *Proceedings of the 13th International Conference on Digital Audio Effects*, Graz, Austria, September 2010.

[73] MCKINNEY, M. F., MOELANTS, D. "Deviations from the resonance theory of tempo induction". In: *Proceedings of the Conference on Interdisciplinary Musicology*, Graz, Austria, April 2004.

[74] MOELANTS, D., MCKINNEY, M. F. "Tempo Perception and Musical Content: What Makes A Piece Fast, Slow Or Temporally Ambiguous?" In: *Proceedings of the 8th International Conference on Music Perception & Cognition*, Evanston, USA, August 2004.

[75] TRYFOU, G., HÄRMÄ, A., MOUCHTARIS, A. "Tempo Estimation Based On Linear Prediction And Perceptual Modelling". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference*, pp. 197–202, Miami, USA, October 2011.

[76] STARK, A. M., DAVIES, M. E. P., PLUMBLEY, M. D. "Real-Time Beat-Synchronous Analysis of Musical Audio". In: *Proceedings of the 12th International Conference on Digital Audio Effects*, Como, Italy, September 2009.

[77] GAINZA, M., COYLE, E. "Tempo Detection Using a Hybrid Multiband Approach", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 1, pp. 57–68, January 2011.

[78] HAINSWORTH, S. W. *Techniques for the Automated Analysis of Musical Audio.* Ph.d. Thesis, University of Cambridge, Cambridge, UK, September 2004.

[79] NUNES, L. O., BISCAINHO, L. W. P. "Tempo Estimation: Evaluation of Different Spectral Flux Computation Methods". In: *Anais do 10o Congresso / 16a Convenção Nacional da AES Brasil*, May 2012.

[80] GOUYON, F., KLAPURI, A., DIXON, S., et al. "An experimental comparison of audio tempo induction algorithms", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 14, n. 5, pp. 1832 –1844, September 2006.

[81] COHEN, L. *Time Frequency Analysis: Theory and Applications.* New York, USA, Prentice Hall, 1995.

[82] FITZGERALD, D. "Harmonic/Percussive Separation using Median Filtering and Amplitude Discrimination". In: *Proceedings of the 13th International Conferencen on Digital Audio Effects*, Graz, Austria, September 2010.

[83] LAURENTI, N., DE POLI, G. "A Nonlinear Method for Stochastic Spectrum Estimation in the Modeling of Musical Sounds", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 2, n. 15, pp. 531–541, February 2007.

[84] NUNES, L. O., ESQUEF, P. A. A., BISCAINHO, L. W. P. "Evaluation of Threshold-Based Algorithms for Detection of Spectral Peaks in Audio". In: *Anais do 5o Congresso de Engenharia de Áudio*, pp. 66–73, São Paulo, Brazil, May 2007.

[85] NUNES, L. O. *Modelagem Senoidal de Sinais Musicais: Técnicas de Análise e Avaliação*. M.Sc. Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, August 2009.

[86] GKIOKAS, A., KATSOUROS, V., CARAYANNIS, G., et al. "Music tempo estimation and beat tracking by applying source separation and metrical relations". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 421–424, Kyoto, Japan, March 2012.

[87] WU, F.-H. F., JANG, J.-S. R., LI LU, C. "A Dynamic Programming Approach With Positional Weighting Window To Beat Tracking For Audio Music". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Porto, Portugal, October 2012.

[88] MCFEE, B., ELLIS, D. P. W. "Better Beat Tracking through Robust Onset Aggregation". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2173–2177, Florence, Italy, May 2014.

[89] DIXON, S. "Evaluation of the Audio Beat Tracking System BeatRoot", *Journal of New Music Research*, v. 36, n. 1, pp. 39–50, March 2007.

[90] OLIVEIRA, J. L., DAVIES, M. E. P., GOUYON, F., et al. "MIREX 2012 Audio Beat Tracking Submission: IBT". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Porto, Portugal, October 2012.

[91] OLIVEIRA, J. L., DAVIES, M. E. P., GOUYON, F., et al. "Beat Tracking for Multiple Applications: A Multi-Agent System Architecture With State Recovery", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 20, n. 10, pp. 2696–2706, December 2012.

[92] SHIU, Y., KUO, C.-C. J. "A Hidden Markov Model Approach to Musical Beat Tracking", Available at `http://viola.usc.edu/Research/atoultaro_yu_icassp08_2_v2.pdf`, 2008.

[93] WHITELEY, N., CEMGIL, A., GODSILL, S. "Bayesian Modelling of Temporal Structure in Musical Audio". In: *Proceedings of the 7th International Conference on Music Information Retrieval*, pp. 29–34, Victoria, Canada, October 2006.

[94] KREBS, F., WIDMER, G. "Mirex 2012 Audio Beat Tracking Evaluation: Beat.E". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Porto, Portugal, October 2012.

[95] KREBS, F., BÖCK, S., WIDMER, G. "Rhythmic Pattern Modeling For Beat And Downbeat Tracking In Musical Audio". In: *Proceedings of the International Symposium on Music Information Retrieval*, Curitiba, Brazil, October 2013.

[96] PEETERS, G. "Beat-Tracking Using A Probabilistic Framework And Linear Discriminant Analysis". In: *Proceedings of the 12th International Conference on Digital Audio Effects*, Como, Italy, September 2009.

[97] PAPADOPOULOS, H., PEETERS, G. "Joint Estimation of Chords and Downbeats From an Audio Signal", *IEEE Transactions on Audio, Speech, and, Language Processing*, v. 19, n. 1, pp. 138–152, January 2011.

[98] PEETERS, G., PAPADOPOULOS, H. "Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation", *IEEE Transactions on Audio, Speech, and Language Processing*, v. 19, n. 6, pp. 1754–1769, August 2011.

[99] PEETERS, G. "MIREX-2012 "Audio Beat Tracking" Task: IRCAMBEAT Submission". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Porto, Portugal, October 2012.

[100] KHADKEVICH, M., FILLON, T., RICHARD, G., et al. "A Probabilistic Approach To Simultaneous Extraction Of Beats And Downbeats". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Porto, Portugal, October 2012.

[101] DURAND, S., DAVID, B., RICHARD, G. "Enhancing Downbeat Tracking Detection When Facing Different Music Styles". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3156–3160, Florence, Italy, May 2014.

[102] SRINIVASAMURTHY, A., SERRA, X. "A Supervised Approach to Hierarchical Metrical Cycle Tracking From Audio Music Recordings". In: *Pro-*

ceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 5254–5258, Florence, Italy, May 2014.

[103] BOUFOUNOS, P., EL-DIFRAWY, S., EHRLICH, D., et al. "Hidden Markov Models for DNA Sequencing". In: *Proceedings of Workshop on Genomic Signal Processing and Statistics*, Raleigh, October 2002.

[104] RABINER, L. R. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, v. 77, n. 2, pp. 257–286, February 1989.

[105] BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York, USA, Springer, 2007.

[106] MURPHY, K. P. *Dynamic Bayesian Networks: Representation, Inference and Learning*. Phd thesis, University of California, Berkeley, USA, Fall 2002.

[107] VITERBI, A. J. "A Personal History of the Viterbi Algorithm", *Signal Processing Magazine*, v. 23, n. 4, pp. 120–142, July 2006.

[108] FARISELLI, P., MARTELLI, P. L., CASADIO, R. "A new decoding algorithm for hidden Markov models improves the prediction of the topology of all-beta membrane proteins", *BMC Bioinformatics*, v. 6, n. 4, pp. 1–7, December 2005.

[109] OLIVEIRA, J. L., GOUYON, F., MARTINS, L. G., et al. "IBT: A Real-Time Tempo And Beat Tracking System". In: *Proceedings of the Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, Netherlands, October 2010.

[110] HINTON, G., DENG, L., YU, D., et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups", *IEEE Signal Processing Magazine*, v. 29, n. 6, pp. 82–97, November 2012.

[111] GUYON, I., ELISSEEFF, A. "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, v. 3, pp. 1157–1182, January 2003.

[112] SHLENS, J. *A Tutorial on Principal Component Analysis*. Technical report, New York University, March 2009.

[113] MEINSHAUSEN, N., BUHLMANN, P. "Stability selection", *Journal of the Royal Statistical Society: Series B*, v. 72, n. 4, pp. 417–473, September 2010.

[114] BURGES, C. J. C. "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, v. 2, n. 2, pp. 121–167, November 1998.

[115] BENNETT, K. P., CAMPBELL, C. "Support Vector Machines: Hype or Hallelujah?" *SIGKDD Explorations*, v. 2, n. 2, pp. 1–13, January 2000.

[116] DUAN, K.-B., KEERTHI, S. S. "Which Is the Best Multiclass SVM Method? An Empirical Study". In: *Proceedings of the Sixth International Workshop on Multiple Classifier Systems*, pp. 278–285, Seaside, USA, June 2005.

[117] NUNES, L. O., BISCAINHO, L. W. P., LEE, B., et al. "Degradation Type Classifier for Full Band Speech Contaminated With Echo, Broadband Noise, and Reverberation", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 19, n. 8, pp. 2516–2526, November 2011.

[118] BREIMAN, L. "Random Forests", *Machine Learning*, v. 45, n. 1, pp. 5–32, October 2001.

[119] DAVIES, M. E. P., DEGARA, N., PLUMBLEY, M. D. *Evaluation Methods for Musical Audio Beat Tracking Algorithms*. Technical report, Queen Mary University of London, London, UK, October 2009.

[120] FERREIRA, L. "An Afrocentric Approach to Musical Performance in the Black South Atlantic: The Candombe Drumming in Uruguay", *TRANS-Transcultural Music Review*, v. 11, pp. 12, January 2007.

[121] NUNES, L. O., BISCAINHO, L. W. P. "HMM-Based Beat-Tracking with Modeling of Tempo Variations", Submetido para o periódico IEEE Signal Processing Letters, 2014.

[122] DIELEMAN, S., SCHRAUWEN, B. "End-to-End Learning for Music Audio". In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 7014–7018, Florence, Italy, May 2014.