

Editado por

Célia A. Zorzo Barcelos

Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Eliana X.L. de Andrade

Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil

Maurílio Boaventura

Universidade Estadual Paulista - UNESP
São José do Rio Preto, SP, Brasil



A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.7)**, as figuras em **eps** e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
<http://www.sbmac.org.br/notas.php>

MÉTODOS MATEMÁTICOS E COMPUTACIONAIS EM MÚSICA

Paulo Cezar Carvalho
pcezar@impa.br

Luiz Velho
lvelho@impa.br

Sergio Krakowski
skrako@impa.br

Vision and Graphics Laboratory - VISGRAF
Instituto Nacional de Matemática Pura e Aplicada - IMPA
Rio de Janeiro, RJ, Brasil



Sociedade Brasileira de Matemática Aplicada e Computacional

São Carlos - SP, Brasil
2012

Coordenação Editorial: Sandra Mara Cardoso Malta

Coordenação Editorial da Série: Eliana Xavier Linhares de Andrade

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC

Copyright ©2012 by Paulo Cezar P. Carvalho, Luiz Velho, Marcelo Cicconet e Sergio Krakowski.

Direitos reservados, 2012 pela SBMAC. A publicação nesta série não impede o autor de publicar parte ou a totalidade da obra por outra editora, em qualquer meio, desde que faça citação à edição original.

Catálogo elaborado pela Biblioteca do IBILCE/UNESP
Bibliotecária: Maria Luiza Fernandes Jardim Froner

Carvalho, Paulo Cezar P.

Métodos Matemáticos e Computacionais em Música
Integrais - São Carlos, SP : SBMAC, 2012, 108 p.; 20,5cm
- (Notas em Matemática Aplicada; v. 38)

e-ISBN 978-85-86883-99-6

1. Computação Musical. 2. Extração de Informações Musicais.
3. Descritores de Áudio.

I. Velho, Luiz. II. Cicconet, Marcelo. III. Krakowski, Sergio.
IV. Título. V. Série.

CDD - 51

Esta é uma republicação em formato de e-book do livro original do mesmo título publicado em 2009 nesta mesma série pela SBMAC.

Conteúdo

1	Introdução	7
1.1	Matemática e Música	7
1.2	Um breve histórico	7
1.2.1	Da escala Pitagórica à escala bem-temperada	8
1.2.2	O advento da música digital	10
1.2.3	Inteligência Computacional Aplicada à Música	12
1.2.4	Sobre o restante do livro	13
2	Processamento de Sinais Sonoros	15
2.1	Paradigmas de Abstração	16
2.1.1	Níveis de Abstração	16
2.2	Modelos Matemáticos de Sinais	18
2.2.1	Aproximação de Sinais	18
2.2.2	Modelos Funcionais e Níveis de Abstração	20
2.2.3	Modelo Temporal de Sinais	22
2.2.4	Modelo Espectral de Sinais	24
2.3	Representação Linear de Sinais	29
2.3.1	Existência de Representação	32
2.4	Operações com Sinais	32
2.4.1	Filtragem	33
2.4.2	Transformadas	34
2.4.3	Filtragem e Freqüência	35
2.5	Teoria de Amostragem	37
2.5.1	Amostragem Pontual Uniforme	37
2.5.2	Amostragem Pontual e Transformada de Fourier	39
2.5.3	Teorema da Amostragem	40
2.5.4	Aliasing	44
2.6	Operações no Dominio Discreto	46
2.6.1	Convolução Discreta	47

2.6.2	Transformada de Fourier Discreta	48
2.7	Reconstrução de Sinais	51
2.7.1	Reconstrução e Aliasing	53
2.8	Filtros Polinomiais	54
3	Ritmo e Percussão	59
3.1	Abordagens	60
3.1.1	Abordagem Teórica	60
3.1.2	Abordagem Experimental	63
3.1.3	Abordagem Computacional	65
3.1.4	Abordagem Etnomusicológica	67
3.2	Análise do Sinal Percussivo	69
3.2.1	Frases Rítmicas	70
3.3	Resultados Musicais	72
4	Melodia e Harmonia	75
4.1	Análise	75
4.1.1	Descritores Harmônicos/Timbrais	76
4.1.2	Detecção de Frequência Fundamental	79
4.1.3	Detecção de Acordes	83
4.1.4	Análise Estrutural de Peças Musicais	83
4.2	Composição	85
4.3	Síntese	86
4.3.1	Síntese por Wavetable	86
4.3.2	Síntese Aditiva	88
4.3.3	Síntese Subtrativa	89
4.3.4	Síntese por Modulação	89
4.3.5	E o que mais?	91
5	Tendências e Aplicações	93
5.1	Localização e Identificação	93
5.2	Recomendação por Similaridade	95
5.3	Novas Interfaces	95
	Bibliografia	97

Capítulo 1

Introdução

1.1 Matemática e Música

O desenvolvimento da Música, uma das manifestações culturais e artísticas mais importantes do homem, esteve sempre atrelado à Matemática. Há, entretanto, alguns momentos, como o que vivemos agora, em que esta conexão fica especialmente importante. O desenvolvimento tecnológico dos últimos anos, que permite aos computadores produzir, adquirir, armazenar e processar sinais musicais, aliado ao desenvolvimento de técnicas de Inteligência Computacional, fortemente baseadas em áreas de Matemática Aplicada tais como Estatística e Otimização, está trazendo novas formas de produzir, classificar e mesmo de ouvir música. Estas notas procuram dar um panorama destas tendências, de modo a permitir ao leitor (que se supõe tem uma formação em Matemática) identificar possíveis oportunidades de utilização de conhecimento matemático em Música. Não partimos do princípio que o leitor tenha conhecimentos teóricos de Música, mas algum conhecimento musical certamente o ajudará a apreciar melhor o material.

1.2 Um breve histórico

É conveniente dividir a história da influência da Matemática na Música em três períodos marcantes:

- Da escala Pitagórica à escala bem-temperada
- O advento da música digital
- Inteligência computacional aplicada à Música

Descrevemos brevemente, a seguir, os principais eventos de cada um desses períodos.

1.2.1 Da escala Pitagórica à escala bem-temperada

Todos sabemos distinguir sons musicais de ruídos. O que caracteriza, fisicamente, sons musicais é a grande coerência de frequências harmônicas, ou seja, múltiplas umas das outras, presentes nestes sons. Isto ocorre, por exemplo, ao fazer vibrar uma corda esticada entre dois pontos, ou o ar contido em um tubo, o que levou à construção dos instrumentos musicais primitivos. O que entendemos por música é o sequenciamento de combinações de sons com essas características. Todas as culturas desenvolveram padrões para os sons a serem combinados, que constituem as diversas escalas musicais.

Na cultura ocidental, as escalas utilizadas são provenientes da Grécia Clássica, em particular da escola Pitagórica, que estabeleceu as primeiras conexões entre Música e Aritmética, ao propor que, para fornecerem combinações percebidas como agradáveis, os sons deveriam ser produzidos por instrumentos (cordas, tubos de ar, etc) cujas dimensões estivessem relacionadas por frações simples (ou seja, com denominadores pequenos). É claro que, hoje, sabemos que isto faz com que as frequências predominantes dos sons produzidos sejam relacionadas pelas mesmas frações.

O tipo mais simples de flauta grega, por exemplo, era formado por 4 tubos de mesmo diâmetro, com comprimentos proporcionais a 1, $3/4$, $2/3$ e $1/2$ (em conseqüência, as frequências das notas que produziam eram proporcionais a 1, $4/3$, $3/2$ e 2), conforme ilustrado na figura 1.1. Em relação à primeira, as três outras notas caracterizam intervalos de frequência chamados de quarta, quinta e oitava. Na escala musical ocidental moderna, as quatro notas produzidas pela flauta correspondem, por exemplo, às notas dó, fá, sol e, novamente, dó. A expressão “por exemplo” se refere ao fato de que qualquer nota pode ser tomada como base para a construção da escala, determinando sua tonalidade.

Observe que, na flauta da figura 1.1, a razão entre 2 e $4/3$ é novamente igual a $3/2$, mostrando que o intervalo entre a segunda e a quarta nota é novamente um intervalo de quinta. Este fato pode ser usado para criar novas notas na escala, sempre tomando-se uma nota e multiplicando-se ou dividindo sua frequência pelos fatores $3/2$ ou $4/3$, dividindo-se por 2, quando necessário, para retornar à mesma oitava. A figura 1.2 mostra a relação entre as frequências das notas – que vieram, muito mais tarde, a receber os nomes de dó(C), ré(D), mi(E), fá(F), sol(G), lá(A), si(B), dó(C) – da escala diatônica clássica.

A figura 1.3 ilustra o processo usado para construir a escala diatônica, e

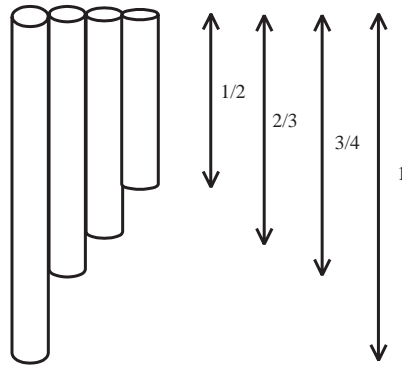


Figura 1.1: Uma flauta simples.

mostra que ele pode ser continuado infinitamente, sempre produzindo novas notas nas razões $2/3$ ou $3/4$ em relação às notas existentes (dividindo por 2, se necessário, para obter uma nota na mesma oitava). Deste modo, produzimos notas de frequências intermediárias às notas da escala diatônica, representadas pelos símbolos $\#$ (sustenido) ou b (bemol), conforme resultem da sequência de intervalos de quinta para cima ou para baixo. Note que as notas A_b e $G\#$ têm frequências ligeiramente diferentes (a razão entre suas frequências é igual a $\frac{531441}{524288} = 1,014$). O fato de que diferentes frequências são obtidas cada vez que se muda o ponto de partida da escala traz dificuldades para a produção de instrumentos musicais de uso universal e para a transposição tonal das obras musicais.

A solução encontrada para o problema, já no século 18, foi a criação da escala bem temperada, que instituiu intervalos idênticos entre as notas da escala, definidas por meio de uma progressão geométrica de frequências, na razão $2^{1/12}$, que interpola os números 1 e 2. Na escala bem temperada, há exatamente 12 notas em cada oitava: C, $C\#/D_b$, D, $D\#/E_b$, E, F, $F\#/G_b$, G, $G\#/A_b$, A, $A\#$, B_b , B (retornando a C a seguir), sempre com razão constante e igual a $2^{1/12}$ entre notas consecutivas. A figura 1.4 mostra as frequências relativas das notas da escala bem-temperada e da escala pitagórica.

É interessante observar que a tecnologia atual tem permitido a utilização, novamente, das escalas clássicas, já que instrumentos eletrônicos podem ser facilmente programados para trabalhar com qualquer escala. Para uma discussão destas tendências, sugerimos que o leitor consulte [6], onde tam-

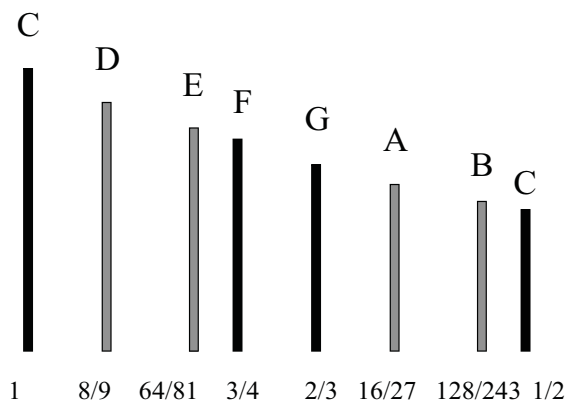


Figura 1.2: A escala diatônica.

bém vai encontrar uma discussão mais completa da evolução das escalas musicais ocidentais.

Também vale a pena notar que é muito mais recente o estabelecimento de um padrão absoluto de frequências para as notas musicais. Diferentes países e orquestras afinavam de modo diferente seus instrumentos (inclusive pela dificuldade, até o século 19, de se ter instrumentos precisos de aferição). Somente em 1939, foi estabelecido um padrão de 440 Hz para a nota lá situada acima do dó central de um piano (que também corresponde à nota produzida pela terceira corda do violão pressionada na segunda casa).

1.2.2 O advento da música digital

O advento da tecnologia digital de armazenamento de som pode ser considerado como um passo intermediário fundamental entre o desenvolvimento das escalas e o estado da arte atual na interação entre Matemática e Música. O primeiro ponto importante desta etapa foi a criação da representação digital utilizada nos CDs. Esta representação é uma transcrição direta do resultado da amostragem da intensidade de um sinal sonoro, ao longo do tempo. O padrão adotado foi o de coletar 44.100 amostras por segundo, quantizadas em 16 bits. A frequência de amostragem escolhida está relacionada com o fato de que os seres humanos são capazes de perceber sons cujas frequências estejam entre 20 Hz e 20000 Hz. Deste modo, é preciso uma taxa de amostragem em torno de 40000 Hz para que não haja perda na reconstrução destas frequências, segundo o Teorema de Nyquist (veja o capítulo 2). Os discos compactos foram inicialmente projetados para ar-

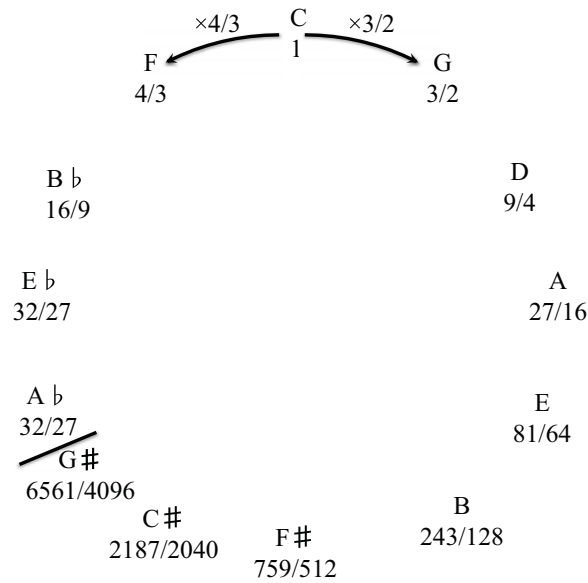


Figura 1.3: O círculo de quintas.

mazenar aproximadamente 1 hora de música, em 2 canais, o que corresponde a aproximadamente 600 MB de informação. A forma de representar som como intensidade ao longo do tempo é a mesma utilizada em vários formatos usuais de representação do som, como o .wav.

Embora métodos matemáticos não tenham desempenhado um papel muito importante na criação desta tecnologia, eles foram fundamentais para o próximo passo, que foi o de obter representações comprimidas de som, como o formato MP3. O estabelecimento deste padrão se baseou em experimentos psico-acústicos, que estabeleceram a ocorrência de um fenômeno conhecido como mascaramento de áudio, que revela, na verdade, uma limitação do sistema de percepção de áudio dos seres humanos: um som tem o poder de mascarar um outro, de frequência ligeiramente mais alta, que ocorra próximo a ele. Isto faz com que, no domínio da frequência (veja o capítulo 2), um som possa ser representado de modo muito mais compacto, sem que seja perceptualmente diferente do som original. Esta conversão para o domínio da frequência (isto é, em que a representação de intensidade ao longo do tempo é substituída pela descrição das frequências relevantes) é feita, no caso do MP3, pela chamada Transformada Discreta do Cosseno, que está relacionada à Transformada de Fourier (veja o capítulo 2). O de-

envolvimento do MP3 contribuiu para a popularização da música digital, facilitando sua distribuição, por proporcionar taxas de compressão da ordem de 1 para 10 (ou seja, uma música que, no formato de CD ocupa cerca de 40 MB pode ser codificada, sem perda perceptual, em um arquivo de cerca de 4 MB).

Um outro desenvolvimento importante deste momento intermediário ocorreu na produção de música eletrônica e, em particular, por meio do computador. Sintetizadores eletrônicos de som começaram a aparecer nos anos 20, mas a maturidade da música eletrônica só foi alcançada com a integração de sintetizadores eletrônicos e computadores. Nos anos 80, foi estabelecido o padrão MIDI (Musical Instrument Digital Interface), para a comunicação entre computadores digitais e sintetizadores de som. Uma sequência de comandos para esta interface resulta em uma representação musical no formato MIDI. Neste formato, são registrados os momentos de ocorrência de cada nota, sua altura, intensidade e o tipo de som a ser sintetizado. Deve-se observar que, enquanto os formatos .wav e .mp3 são formatos de representação de som em geral, o formato .mid fornece uma representação simbólica para uma peça musical.

1.2.3 Inteligência Computacional Aplicada à Música

A tecnologia desenvolvida na produção, representação e armazenamento de música, aliada ao aumento da velocidade dos computadores e suas interfaces, permitem tornar realidade o uso do computador como um participante no processo de composição, execução e escolha de peças musicais. Entre as muitas possibilidades já reais ou em desenvolvimento, podemos citar:

- composição automática de peças musicais
- acompanhamento automático de uma peça musical, acrescentando novos instrumentos a uma performance
- busca, por semelhança, de uma peça musical em bancos de dados
- classificação automática de peças musicais, detectando similaridades entre peças (por exemplo, para orientar um usuário na identificação de peças de que ele provavelmente vai gostar)

Nos capítulos seguintes, descreveremos com mais detalhe alguns destes problemas.

1.2.4 Sobre o restante do livro

No capítulo 2, apresentamos uma metodologia para o estudo de sinais sonoros em matemática aplicada computacional. Nesse contexto, o som é analisado a partir do universo físico, passando por seu modelo matemático até chegar na representação digital. O capítulo introduz também noções básicas de teoria de sinais, amostragem e reconstrução.

No capítulo 3, levando em conta a representação digital do som apresentada no capítulo anterior, discutimos aspectos específicos do sinal sonoro percussivo. Apresentamos as características básicas deste tipo de sinal e a forma de analisá-lo de modo rápido e com baixo custo computacional. A partir desta análise de baixo nível propomos uma representação que leve em conta o aspecto rítmico, essencial na compreensão do sinal percussivo. Apresentamos diversas aplicações que utilizam esta representação e exemplos reais de interação que permitiram a construção de peças musicais completas utilizando percussão e computador.

O capítulo 4 trata de métodos que lidam principalmente com as características tonais/timbrais de um som musical. Serão apresentados alguns *descritores* de áudio calculados a partir do sinal digital em estado bruto (ie, em formato PCM), bem como exemplos de problemas que podem ser resolvidos utilizando-os. Um dos problemas fundamentais a ser discutido é a obtenção da frequência fundamental de um segmento (ou seja, a nota musical predominante). Também serão abordadas algumas técnicas simples de síntese digital.

Finalmente, no capítulo 5, são apresentados os mecanismos que suportam alguns aplicativos musicais que têm se popularizado em anos recentes, devido principalmente à web interativa, como as rádios com programação sugerida pelo usuário e os buscadores de arquivos musicais.

Escala Pitagórica		Escala bem-temperada	
C	1,0000	1,0000	C
Db	1,0535	1,0595	Db/C#
C#	1,0679		
D	1,1250	1,1225	D
Eb	1,1852	1,1892	Eb/D#
D#	1,2014		
E	1,2656	1,2599	E
F	1,3333	1,3348	F
Gb	1,4047	1,4142	Gb/F#
F#	1,4238		
G	1,5000	1,4983	G
Ab	1,5802	1,5874	Ab/G#
G#	1,6018		
A	1,6875	1,6818	A
Bb	1,7778	1,7818	Bb/A#
A#	1,8020		
B	1,8984	1,8877	B
C	2,0000	2,0000	C

Figura 1.4: Escalas pitagórica e bem-temperada.

Capítulo 2

Processamento de Sinais Sonoros

Nossa interação diária com o ambiente se dá através de diversos tipos de sinais. Os sinais sonoros permitem a nossa percepção auditiva e nos propicia a apreciação da música e a utilização da comunicação falada; os sinais eletromagnéticos nos abrem as portas para o fantástico mundo das telecomunicações.

Com o advento do computador, e sua crescente presença no controle de diversos processos de nosso cotidiano, cresceu a importância da representação e do processamento de sinais em forma digital. Essa representação tem um papel relevante relacionado principalmente com o processo de síntese, processamento e análise de sons. O objetivo deste capítulo é fornecer ao leitor uma visão conceitual dos diversos problemas dessa área sob a ótica de sua utilização em música computacional.

A matemática envolvida no desenvolvimento da teoria de sinais pode se tornar bastante sofisticada. Nosso objetivo nesse capítulo é fazer com que o leitor entenda a conceituação matemática dessa teoria. Para tornar a leitura mais elementar fomos obrigados a abrir mão da formalização matemática rigorosa. Um bom exercício neste capítulo, para o leitor matematicamente sofisticado, consiste em descobrir os diversos pontos onde optamos por uma exposição mais simples porém pecando no rigor matemático.

O capítulo foi escrito com o objetivo de ser útil tanto para o leitor que já teve um primeiro contato com processamento digital de sinais, como para o leitor que está se iniciando no assunto. Procuramos atender a todo esse universo fazendo uma exposição coerente, simples e concisa.

2.1 Paradigmas de Abstração

Para estudar e analisar os diversos tipos de sinais no universo, devemos procurar descrições matemáticas desses sinais, determinar meios efetivos de construir representações discretas de sinais, e buscar algoritmos que permitam a implementação das diversas técnicas de síntese, análise e processamento de sinais no computador. Esse paradigma de abstração que é conhecido pelo nome de paradigma dos quatro universos, é ilustrado no diagrama da Figura 2.1.

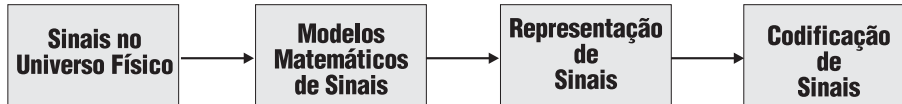


Figura 2.1: Paradigmas de abstração no estudo de sinais.

2.1.1 Níveis de Abstração

Baseado no paradigma descrito anteriormente, temos três níveis de abstração no processo de representação de um sinal: *sinais contínuos*, *sinais discretos* e *sinais codificados*. Cada um desses níveis corresponde a uma descrição do sinal que é conveniente para a formulação e solução dos diversos problemas da Matemática Aplicada Computacional. Associadas a esses níveis de abstração temos quatro operações que permitem a mudança entre eles: *discretização*, *codificação*, *decodificação* e *reconstrução*. Esse fato é ilustrado no diagrama da Figura 2.2.

Discretização e Reconstrução

Dá-se o nome de *discretização* ao processo de conversão de um sinal contínuo em uma representação discreta. O processo inverso, que consiste em obter um sinal contínuo a partir de uma representação discreta, é denominado de *reconstrução*.

Idealmente, a operação de reconstrução deveria ser uma operação inversa do processo de discretização, ou seja dado um determinado sinal s , e sua representação discreta s_d , deveria existir um processo de reconstrução que possibilitasse a obtenção do sinal original s , a partir de s_d :

$$s \rightarrow \text{Discretização} \rightarrow s_d \rightarrow \text{Reconstrução} \rightarrow s.$$

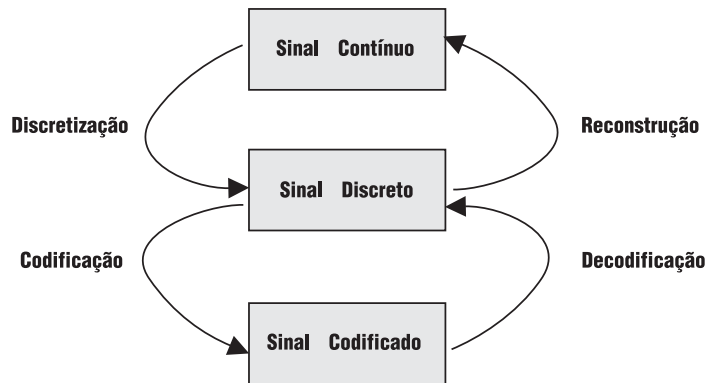


Figura 2.2: Níveis de abstração na representação de sinais.

Em geral, no entanto, o processo de reconstrução não recupera o sinal original, e temos na realidade três aspectos na representação de um sinal: o sinal contínuo s , o sinal discretizado s_d , e o sinal reconstruído s_r . Um dos problemas importantes em processamento de sinais consiste em se obter métodos de discretização e reconstrução de forma que o sinal reconstruído s_r seja uma “boa aproximação” do sinal original s . O significado do termo “boa aproximação” está diretamente relacionado com as diversas áreas de aplicação.

Codificação e Decodificação

O processo de *codificação* obtém, a partir da representação discreta do sinal, uma representação finita, que pode ser descrita por um conjunto finito de símbolos organizados de acordo com uma estrutura de dados. A *decodificação* possibilita o acesso às informações codificadas na forma de uma representação discreta. Dada uma representação discreta s_d de um determinado sinal s , temos

$$s_d \rightarrow \text{Codificação} \rightarrow s_c \rightarrow \text{Decodificação} \rightarrow \tilde{s}_d.$$

Se o sinal codificado for igual ao sinal discretizado, isto é $\tilde{s}_d = s_d$, temos um processo de *codificação sem perda*. Além desse problema de fidelidade no processo de codificação, vários outros fatores devem ser considerados na elaboração das estruturas que realizam um esquema de representação finito. Dentre eles, os mais importantes são o espaço ocupado pelo código, e o tempo de processamento para o processo de codificação e decodificação.

2.2 Modelos Matemáticos de Sinais

Dando início ao estudo do paradigma de abstração exposto na seção inicial, vamos determinar os modelos matemáticos utilizados para estudar os sinais do universo físico.

Um sinal se manifesta através da variação de uma determinada grandeza física. Essa variação pode ocorrer em relação ao espaço ou ao tempo. A propagação do sinal de áudio se dá através da variação da densidade de ar ao longo do tempo.

Baseados nos fatos explicados no parágrafo anterior podemos concluir que um sinal deve ser representado por um objeto matemático que estabeleça a variação de uma determinada grandeza. Se essa variação ocorre de forma determinística podemos utilizar uma função como modelo de um sinal, por outro lado se a variação da grandeza ocorre de forma não determinística o sinal pode ser descrito por um *processo estocástico*. No primeiro caso temos um *modelo funcional* de sinais, e no segundo caso temos um *modelo estocástico*.

Os modelos funcionais atendem bem o uso de sinais de audio e portanto será o modelo adotado. Nesse caso um sinal é representado por uma função $f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$. Ou seja, a grandeza física do sinal é representada por um vetor n -dimensional que varia em um espaço com m graus de liberdade. Chamamos de *espaço de sinais* a um subespaço de funções $\{f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n\}$ com U , m e n fixos. Um espaço de sinais é portanto um espaço de funções que é naturalmente um espaço vetorial em relação às operações usuais de soma de funções e de multiplicação de uma função por um número real:

$$(f + g)(t) = f(t) + g(t); \quad (\lambda f)(t) = \lambda \cdot f(t), \quad \lambda \in \mathbb{R}. \quad (2.2.1)$$

2.2.1 Aproximação de Sinais

Em diversos problemas envolvendo sinais é de extrema importância a existência de uma métrica de forma a permitir uma medida de aproximação entre dois sinais. Diversas métricas podem ser definidas em um espaço de sinais. A escolha da métrica adequada depende fundamentalmente dos problemas em cada área de aplicação. A nossa percepção de sinais se dá, de forma direta ou indireta, através dos sentidos (a audição percebe sinais de áudio, etc.). Uma métrica d no espaço de sinais é chamada de *métrica perceptual* se, dados dois sinais f e g , tem-se que $d(f, g) = 0$ se, e somente se, os sinais são indistinguíveis do ponto de vista perceptual. Por essa razão, e também por razões de eficiência computacional, é muito comum o uso de pseudo-métricas para medir a proximidade entre dois sinais. Nesse caso é

possível termos $d(f, g) = 0$ sem que se tenha, necessariamente, a igualdade funcional $f = g$.

Dois métricas bastante utilizadas no espaço de sinais são a *métrica uniforme* e a *métrica L^p* .

MÉTRICA UNIFORME.

Nessa métrica a distância $d(f, g)$ entre dois sinais f e g é definida por

$$d(f, g) = \sup\{|f(u) - g(u)| ; u \in U\}, \quad (2.2.2)$$

onde \sup indica o supremo de um conjunto de números reais. É claro que devemos supor que os sinais são limitados no conjunto U . Uma vizinhança de raio $\varepsilon > 0$ de um sinal f na métrica uniforme é mostrada na figura 2.3. A figura mostra também um sinal g nessa vizinhança, ou seja $d(f, g) < \varepsilon$.

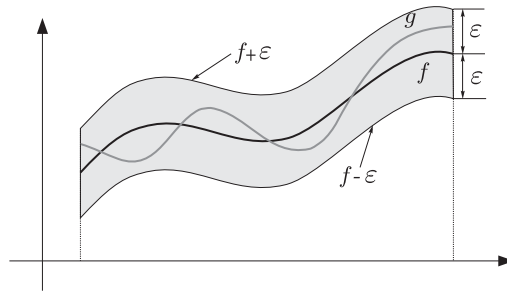


Figura 2.3: Vizinhança de um sinal na métrica uniforme.

MÉTRICA L^p .

Quando os sinais tomam valores em \mathbb{R} ou \mathbb{C} , podemos introduzir a métrica L^p , que define a distância d entre dois sinais por

$$d(f, g) = \left[\int_{-\infty}^{+\infty} |f(u) - g(u)|^p du \right]^{1/p}, \quad (2.2.3)$$

supondo que a integral existe. Um caso particular de grande importância na teoria de sinais ocorre para $p = 2$. Nesse caso a métrica é originada do produto interno L^2 no espaço de sinais definido por

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(u) \overline{g(u)} du. \quad (2.2.4)$$

A barra indica a operação de conjugação complexa. É claro que a métrica L^2 só está definida se

$$\int_{-\infty}^{+\infty} |f|^2 < \infty. \quad (2.2.5)$$

Fisicamente, essa condição significa que o sinal tem energia finita. A hipótese de ter energia finita é fisicamente adequada, e de grande importância do ponto de vista matemático. Neste capítulo, vamos supor que todos os sinais possuem energia finita, a menos que seja dito explicitamente ao contrário.

Deixamos ao leitor discutir a natureza perceptual da métrica uniforme e da métrica L^p .

2.2.2 Modelos Funcionais e Níveis de Abstração

Uma função $f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$, em um espaço de sinais, é chamada de *sinal contínuo*. Devemos ressaltar no entanto que o adjetivo contínuo é utilizado de modo a enfatizar que o domínio e o contra-domínio da função f é um “continuum” de números, ou seja, estamos trabalhando com o conjunto dos números reais. Isso não significa que a função f seja necessariamente uma função contínua no sentido da topologia. Em engenharia, os sinais contínuos são chamados de *sinais analógicos*. No computador o conjunto dos números reais é representado por um conjunto finito utilizando aritmética de ponto flutuante. Desse modo, para efeitos práticos considera-se que um sinal definido utilizando a representação de ponto flutuante dos números reais é um sinal contínuo.

Vamos agora revisar os três níveis de abstração na representação de sinais visto na seção 2.1 sob o ponto de vista do modelo funcional. Dado um sinal, definido por uma função f , o processo de representação consiste em discretizar o domínio da função f , enquanto que o processo de codificação requer a discretização tanto do domínio quanto do contra-domínio de f . O processo de discretização do domínio é conhecido classicamente pelo nome de *amostragem*, enquanto que o processo de discretização do contra-domínio é chamado de *quantização*. Existem portanto quatro diferentes aspectos relacionados com a representação funcional de um sinal no computador:

- Sinal Contínuo-Contínuo;
- Sinal Contínuo-Discreto;
- Sinal Discreto-Contínuo;
- Sinal Discreto-Discreto.

O sinal contínuo-contínuo corresponde ao que chamamos anteriormente simplesmente de sinal contínuo. O sinal contínuo-discreto tem o seu contra-domínio quantizado; o sinal discreto-contínuo tem o domínio discretizado. Finalmente, o sinal discreto-discreto é o sinal amostrado e quantizado. Esse sinal é chamado de *sinal digital*.

De um modo intuitivo, a operação de discretização de um sinal f consiste em calcular f em um número finito de pontos p_1, p_2, \dots, p_k , do conjunto U , que são chamados de amostras. A reconstrução do sinal corresponde a se utilizar um processo de interpolação de modo a obter o sinal original f , ou uma aproximação f_r do sinal f , a partir dos valores $f(p_1), f(p_2), \dots, f(p_k)$ de f nas amostras. A Figura 2.4 ilustra esse processo utilizando o método de interpolação linear na etapa de reconstrução.

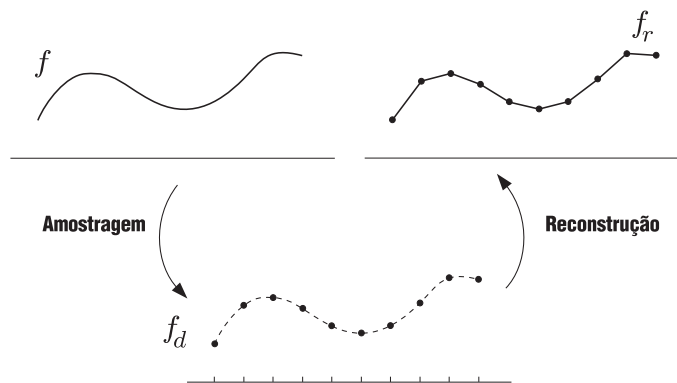


Figura 2.4: Amostragem e reconstrução de um sinal.

Se o método de interpolação recuperar o sinal originalmente amostrado, isto é $f_r = f$, temos um processo de *reconstrução ideal* ou *reconstrução exata*. No restante deste capítulo vamos estender e elaborar os conceitos de amostragem e reconstrução para modelos funcionais de sinais. O problema da codificação e decodificação de sinais será estudado nos próximos capítulos apenas para imagens.

A diferença entre os diversos modelos funcionais de sinais está na interpretação das grandezas no domínio e no contra-domínio da função que representa o sinal. Utilizaremos dois modelos funcionais de sinais: o *modelo temporal* e o *modelo espectral*. Cada um desses modelos será descrito em seguida.

2.2.3 Modelo Temporal de Sinais

No *modelo temporal*, dada uma função $f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$ o subconjunto U representa o intervalo de tempo no qual varia a grandeza física. Por essa razão U é chamado de *domínio do tempo*. A grandeza física é representada por um vetor do \mathbb{R}^n , e a lei de definição da função f nos fornece as regras de variação da grandeza. A dimensão m do domínio determina a dimensão do sinal. Se $m = 1$ dizemos que o sinal é unidimensional.

Voltando ao nosso exemplo anterior vemos que o sinal de áudio é um um sinal unidimensional; Nesse caso U é um subconjunto da reta \mathbb{R} , e a função f associa a cada ponto t o valor $f(t)$ que fornece a informação de som em p .

Exemplo 2.1. (*Sinal de pulso*). Dado $a > 0$, o sinal de pulso $p_a(t)$ unidimensional é um sinal de intensidade constante e duração finita. Esse sinal é representado pela função

$$p_a(t) = \begin{cases} 1 & \text{se } |t| \leq a; \\ 0 & \text{se } |t| > a, \end{cases} \quad (2.2.6)$$

cujos gráfico é mostrado na Figura 2.5(a).

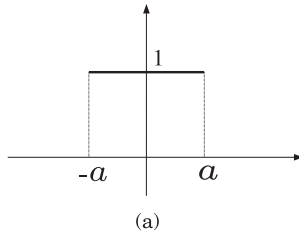


Figura 2.5: Sinal de pulso.

Infelizmente devemos observar que alguns sinais do universo físicos não podem ser representados no modelo funcional. Isso significa que a variação das grandezas definidas no sinal não podem ser descritas por uma função. Um exemplo de sinal dessa natureza é um *impulso físico*. Esse sinal consiste de uma *variação instantânea de uma determinada grandeza, com grande intensidade e energia finita*. Uma tentativa de representar esse sinal por uma função f poderia ser traduzida conforme indicamos abaixo:

- $f(t_0) \neq 0$, e $f(t) = 0$ se $t \neq t_0 \rightarrow$ Variação instantânea;

- $f(t_0) = +\infty \rightarrow$ Grande intensidade em t_0 ;
- $\int_{-\infty}^{+\infty} f(t)dt < \infty \rightarrow$ Energia finita.

É claro que não existe uma função satisfazendo as três propriedades acima. Poderíamos utilizar uma extensão adequada do conceito de função para o nosso modelo matemático de sinais. Uma dessas extensões, que atenderia às nossas necessidades, é o conceito de distribuição. Neste capítulo porém preferimos trocar o rigor matemático por uma exposição mais conceitual e, na medida do possível, elementar. Mostramos abaixo como podemos utilizar o modelo funcional de sinais para dar uma descrição aproximada do sinal impulso. O leitor deve ficar avisado que toda a manipulação matemática relacionada com o sinal impulso não está correta. No entanto ela pode ser formalizada corretamente no contexto do modelo de sinais usando distribuição.

O Sinal Impulso

O impulso $\delta(t)$, também chamado de *delta de Dirac*, é definido pelo limite

$$\delta(t) = \lim_{n \rightarrow +\infty} \frac{n}{2} p_{1/n}(t). \quad (2.2.7)$$

onde $p_{1/n}$ é a função pulso definida no exemplo 2.1. Note que quando $n \rightarrow +\infty$ o suporte dos sinais $\frac{n}{2} p_{1/n}$ converge para 0, e o seu valor converge para $+\infty$ (ver Figura 2.6). Portanto, no limite temos um sinal que atende às duas primeiras propriedades do sinal impulso vistas anteriormente.

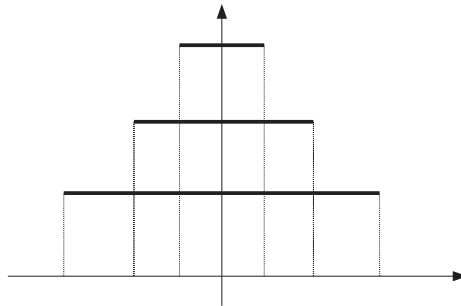


Figura 2.6: Sequência de pulsos convergindo para um impulso na origem.

Quanto à terceira propriedade, energia finita, observamos que

$$\int_{-\infty}^{+\infty} \delta(t) dt = \int_{-\infty}^{+\infty} \lim_{n \rightarrow +\infty} \frac{n}{2} p_{1/n}(t) dt = \lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} \frac{n}{2} p_{1/n}(t) dt = 1. \quad (2.2.8)$$

A função impulso é de grande importância na teoria dos sinais, tanto do ponto de vista teórico como do ponto de vista das aplicações. Uma de suas propriedades mais importantes é dada pela equação

$$f(x) = \int_{-\infty}^{+\infty} f(t) \delta(x - t) dt. \quad (2.2.9)$$

Essa equação mostra que todo sinal pode ser escrito como uma soma infinita de impulsos devidamente transladados e modulados pelo valor do sinal. Deixamos ao leitor a tarefa de se convencer da validade da equação (2.2.9) usando a definição de δ como limite de sinais do tipo pulso (equação (2.2.7)).

2.2.4 Modelo Espectral de Sinais

Vimos que no modelo temporal um sinal é determinado por uma função $f : U \subset \mathbb{R} \rightarrow \mathbb{R}$, que define a variação do sinal no domínio do tempo U . A interpretação física do significado das variáveis no domínio e no contradomínio de uma representação funcional de um sinal, apesar de ser irrelevante do ponto de vista matemático, é de fundamental importância nas aplicações onde diferentes modelos funcionais de sinais podem ser utilizados. Para melhor entender o modelo espectral de sinal, considere a sinal definido por uma senoide $f(t) = a \sin(2\pi\omega_0 t + \Phi)$. Esse sinal fica completamente caracterizado pela sua amplitude a , pela sua frequência ω_0 e pelo seu ângulo de fase Φ . A frequência nos dá uma medida da variação do sinal por unidade de comprimento (o sinal dá ω_0 ciclos completos por unidade de comprimento), conforme ilustramos na Figura 2.7 para frequências 2, 4 e 8.

Com base no exemplo acima podemos tentar caracterizar um sinal através de suas componentes de frequência. Ou seja, devemos obter um modelo funcional de um sinal que associa a cada frequência presente no sinal a amplitude e o ângulo de fase correspondentes. Considere, por exemplo, o sinal periódico $f(t) = a \cos(2\pi\omega_0 t)$ cujo gráfico é mostrado na Figura 2.8(a). Esse sinal é constituído por uma única componente de frequência ω_0 com amplitude a e ângulo de fase 0. Portanto a sua representação funcional usando frequências é dada pela função

$$\hat{f}(s) = \begin{cases} a & \text{se } s = \omega_0; \\ 0 & \text{se } s \neq \omega_0, \end{cases} \quad (2.2.10)$$

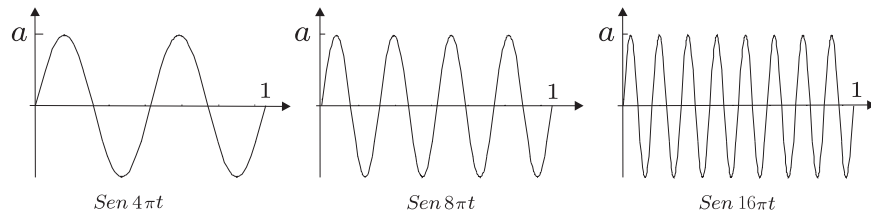


Figura 2.7: Sinais com frequências 2, 4 e 8.

cuja representação gráfica é feita conforme mostramos na Figura 2.8(b). Vamos agora fazer uma observação importante com relação à definição de $\hat{f}(s)$. Como

$$\cos(t) = \frac{1}{2}(e^{it} + e^{-it}), \tag{2.2.11}$$

podemos escrever o sinal $f(t)$ na forma

$$f(t) = \frac{a}{2}(e^{2\pi i\omega_0 t} + e^{2\pi i(-\omega_0)t}). \tag{2.2.12}$$

Vemos nessa representação que temos duas componentes de frequência simétricas, ω_0 e $-\omega_0$, ambas com amplitude $a/2$. É comum portanto utilizar a representação gráfica do modelo de frequências \hat{f} conforme mostramos na Figura 2.8(c).

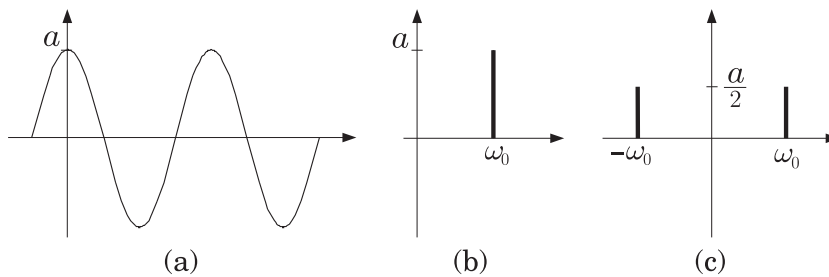


Figura 2.8: Sinal periódico e seus modelos de frequência.

A representação com o modelo funcional de frequências utilizada no sinal acima pode ser generalizada para obter uma representação no espaço de frequências para qualquer sinal periódico utilizando a série de Fourier. Com efeito, a teoria das séries de Fourier garante que um sinal periódico f

de período T_0 pode ser escrito como uma série

$$f(t) = \sum_{k=-\infty}^{+\infty} c_k e^{i2\pi k\omega_0 t}, \quad (2.2.13)$$

onde ω_0 é a *freqüência fundamental* do sinal, definida por $\omega_0 = 1/T_0$.

Esse desenvolvimento em série de Fourier mostra que um sinal periódico contém todas as freqüências múltiplas de sua freqüência fundamental ω_0 , e apenas essas freqüências. O modelo de freqüências do sinal pode então ser definido por uma função que associa a cada múltiplo $k\omega_0$, $k \in \mathbb{Z}$, da freqüência fundamental, a amplitude c_k . No caso de c_k ser um número complexo, podemos representar o modelo freqüências da parte real e imaginária em separado. Outra opção nesse caso é utilizar a forma polar do número complexo c_k representando, para cada múltiplo $k\omega_0$ da freqüência fundamental, o módulo de c_k e o seu ângulo de fase.

Example 2.2. (*Sinal dente de serra*). Considere o sinal periódico $f(t)$ cujo gráfico é mostrado na Figura 2.9(a) (*sinal dente de serra*). Tomando $\omega_0 = 1/T_0$, a série de Fourier de f é dada por

$$f(t) = \frac{1}{T_0} + \frac{8}{\pi^2 T_0} \left[\cos(2\pi\omega_0 t) + \frac{1}{3^2} \cos(6\pi\omega_0 t) + \frac{1}{5^2} \cos(10\pi\omega_0 t) + \dots \right]. \quad (2.2.14)$$

O gráfico da representação por freqüências do sinal f é mostrado na Figura 2.9(b)

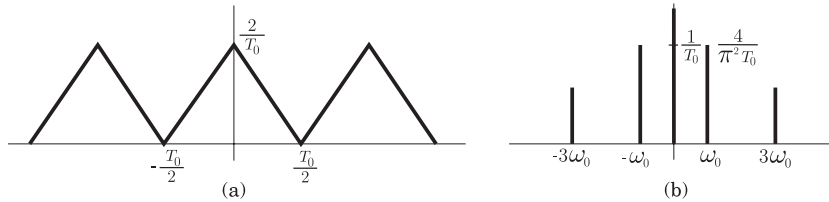


Figura 2.9: Modelo espacial (A), e modelo de freqüências (B), do sinal dente de serra.

A série de Fourier nos permitiu definir um modelo funcional de freqüências para sinais periódicos. Desejamos agora estender o modelo funcional de freqüências para sinais não-periódicos. Para isso, precisamos encontrar uma ferramenta que nos permita medir a ocorrência de uma determinada

freqüência em um sinal arbitrário. É claro que não podemos utilizar a série de Fourier para esse fim. No caso não-periódico o sinal pode conter freqüências arbitrárias, ou seja, podemos ter um “continuum” de freqüências. O método clássico para fazer essa medida é a *transformada de Fourier* conforme explicamos abaixo.

Transformada de Fourier

Dado um sinal $f : \mathbb{R} \rightarrow \mathbb{R}$, a sua *transformada de Fourier* é definida por

$$f \mapsto \hat{f}(s) = \int_{-\infty}^{+\infty} f(t)e^{-2\pi its} dt. \quad (2.2.15)$$

É imediato verificar que a transformada de Fourier define um operador linear $F : \mathcal{S} \rightarrow \mathcal{S}'$, $f \mapsto \hat{f} = F(f)$, entre dois espaços de sinais.

É fácil ver, de modo intuitivo, que a transformada de Fourier cumpre o papel de detectar freqüências no sinal f . Com efeito o núcleo $e^{-2\pi its}$ é um sinal periódico com freqüência s . Desse modo, para cada $s \in \mathbb{R}$ a modulação $f(t)e^{-2\pi its}$ ressalta as regiões da reta onde a freqüência do sinal f entra em ressonância com as freqüências do núcleo $e^{-2\pi its}$. A integral em (2.2.15) é uma medida da “densidade” da freqüência s no sinal f em todo o seu domínio. Portanto o valor $\hat{f}(s)$ indica a ocorrência da freqüência s no sinal f porém não nos dá informação sobre a localização dessa freqüência no domínio do sinal.

A transformada de Fourier F possui uma transformada inversa F^{-1} , definida por

$$f(t) = F^{-1}(\hat{f}(s)) = \int_{-\infty}^{+\infty} \hat{f}(s)e^{2\pi ist} ds. \quad (2.2.16)$$

Note que, intuitivamente, a equação (2.2.16) mostra que o sinal f é uma soma infinita de sinais com freqüência s , $s \in \mathbb{R}$, e amplitude $\hat{f}(s)$.

Portanto, um sinal pode ser caracterizado tanto pelo seu modelo espacial f , como pelo seu modelo de freqüências \hat{f} . A transformada de Fourier, e sua inversa, fazem a conversão entre os dois modelos. O modelo funcional, $\hat{f} = F(f)$, de freqüências de um sinal é chamado de *modelo espectral*. Desse modo, o modelo espacial nos dá informação sobre a variação do sinal no domínio do espaço, enquanto o modelo espectral nos dá informações sobre a variação do sinal no domínio da freqüência. De um modo geral o modelo espacial é utilizado para síntese enquanto que o modelo espectral é mais utilizado para a análise de sinais.

OUTROS MODELOS DE FREQUÊNCIA.

Existem outros modelos funcionais de sinais no espaço da frequência além do modelo espectral obtido através da transformada de Fourier. Para cada um desses modelos temos uma transformada associada que desempenha o mesmo papel da transformada de Fourier no modelo espectral. Dentre essas transformadas podemos citar a *transformada cosseno*, a *transformada de Fourier com janela*, e a *transformada de wavelets*. A existência de diversos modelos de sinais, e de transformadas, que permitem a conversão entre esses modelos e o modelo espacial, é de grande importância nas aplicações de síntese e análise de sinais, onde procuramos o modelo mais adequado para cada tipo de problema.

Função Sinc

Considere a função pulso

$$p_a(t) = \begin{cases} 1 & \text{se } |t| \leq a; \\ 0 & \text{se } |t| > a, \end{cases}$$

definida no exemplo 2.1. Um cálculo imediato mostra que sua transformada de Fourier é dada por

$$\hat{p}_a(s) = 2a \frac{\text{sen}(2\pi as)}{2\pi as}. \quad (2.2.17)$$

Usando a notação

$$\text{sinc}(t) = \begin{cases} \frac{\text{sen}(t)}{t} & \text{se } t \neq 0; \\ 1 & \text{se } t = 0, \end{cases}$$

podemos escrever

$$\hat{p}_a(s) = 2a \text{sinc}(2\pi as). \quad (2.2.18)$$

Na Figura 2.10 mostramos os gráficos da função pulso e de sua transformada de Fourier. A função sinc desempenha um papel importante no estudo de sinais.

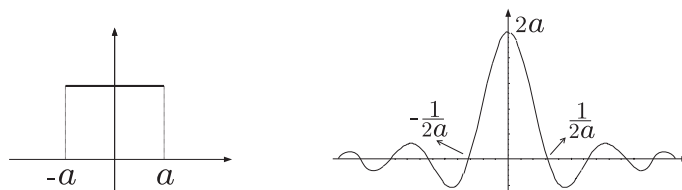


Figura 2.10: Função pulso e sua transformada de Fourier.

Observando a transformada de Fourier do sinal pulso vemos que se o suporte do pulso vai para 0, isto é $a \rightarrow 0$, a presença de altas frequências no sinal aumenta, uma vez que $1/(2a) \rightarrow \infty$. Reciprocamente, se $a \rightarrow \infty$, então a presença de altas frequências no sinal diminui. Esse é um caso particular de um resultado geral que estabelece uma relação entre o desvio padrão de um sinal no domínio do espaço e no domínio da frequência. Quanto menor for o desvio em um dos domínios, maior será o desvio da transformada de Fourier do sinal no outro domínio. Essa observação é a base do *princípio da incerteza* que é de grande importância no relacionamento dos modelos de um sinal no domínio do espaço da frequência. É possível dar uma explicação intuitiva para esse fato: quando uma determinada frequência de um sinal está distribuída em uma região grande do domínio do espaço, há um efeito de cancelamento de frequências no cálculo da densidade de frequências feita pela transformada de Fourier, resultando dessa forma um valor pequeno. O oposto ocorre quando temos um valor de frequência numa pequena região do espaço.

Função Impulso

A transformada de Fourier do sinal impulso pode ser obtida utilizando a nossa aproximação de um impulso por funções pulsos. O cálculo, incorreto é claro, é imediato:

$$\begin{aligned} \hat{\delta}(s) = F(\delta) &= F\left[\lim_{n \rightarrow \infty} \frac{n}{2} p_{1/n}(t)\right] = \lim_{n \rightarrow \infty} F\left[\frac{n}{2} p_{1/n}(t)\right] \\ &= \lim_{n \rightarrow \infty} \frac{\text{sen}\left(\frac{2\pi s}{n}\right)}{\frac{2\pi s}{n}} = 1. \end{aligned} \quad (2.2.19)$$

Vemos portanto que a transformada de Fourier do sinal impulso é a função constante $\hat{\delta}(s) = 1$.

2.3 Representação Linear de Sinais

Uma vez definidos os nossos modelos de sinais no universo matemático, voltamos agora ao segundo nível em nossa hierarquia de abstração: o problema de discretização. Conforme dissemos antes, classicamente a operação de discretizar um sinal é chamada de amostragem. Aqui vamos introduzir o conceito de *representação*, que estende, em um sentido a ser precisado, o conceito de amostragem.

Inicialmente definimos o espaço ℓ^2 das *seqüências de quadrado somável*. Esse espaço consiste de todas as seqüências

$$(\dots, c_{-2}, c_{-1}, c_0, c_1, c_2, \dots), \quad (2.3.20)$$

tais que

$$\sum_{i=-\infty}^{+\infty} c_i^2 < \infty. \quad (2.3.21)$$

A condição de convergência da série em (2.3.21) é imposta de modo a podermos definir um produto interno no espaço ℓ^2 . Isso é feito pondo

$$\langle (c_i), (b_j) \rangle = \sum_{i=-\infty}^{+\infty} c_i b_i. \quad (2.3.22)$$

Observamos que os espaços euclidianos \mathbb{R}^n são naturalmente mergulhados isometricamente no espaço ℓ^2 .

Uma *representação linear* de um espaço de sinais $\mathcal{S} = \{f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n\}$ é um operador linear contínuo $L : \mathcal{S} \rightarrow \ell^2$. A imagem $V = L(\mathcal{S})$ do espaço \mathcal{S} pela aplicação L é um subespaço de ℓ^2 , chamado de *espaço de representação* do espaço de sinais \mathcal{S} . Se V tem dimensão finita, temos uma *representação finita*. Quando é possível obter o sinal f a partir de sua representação $L(f)$, temos uma *representação exata* do sinal f . Se esse fato for verdadeiro para todo sinal $f \in \mathcal{S}$ dizemos que a representação é exata. O processo de representar um sinal tem por objetivo obter um sinal discreto a partir de seu modelo contínuo.

Example 2.3. (*Amostragem pontual finita*). Dado um espaço de sinais $\mathcal{S} = \{f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}^n\}$, fixemos k pontos u_1, u_2, \dots, u_k no domínio U . É imediato verificar que a aplicação $L : \mathcal{S} \rightarrow \mathbb{R}^k \subset \ell^2$ definida por

$$L(f) = (f(u_1), f(u_2), \dots, f(u_k)), \quad (2.3.23)$$

é uma *representação linear* de \mathcal{S} . Neste exemplo temos uma *representação linear finita*. Classicamente essa representação é conhecida pelo nome de *amostragem pontual finita*.

Example 2.4. (*Amostragem de Fourier*). A *série de Fourier* de um sinal periódico f , dada pela equação (2.2.13), define na realidade uma *representação linear*

$$f \mapsto (\dots, c_{-2}, c_{-1}, c_0, c_1, c_2, \dots), \quad (2.3.24)$$

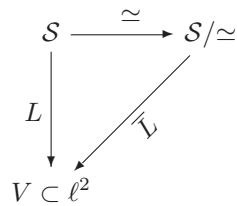
do sinal f . Essa *representação linear* obtém um modelo discreto do sinal no domínio da frequência, ou seja um modelo espectral discreto. Essa representação é chamada na literatura de *amostragem de Fourier*.

Uma *representação linear* tem por objetivo substituir um espaço de sinais, que em geral possui dimensão infinita, por um subespaço de ℓ^2 . Em

geral buscamos subespaços de dimensão finita de ℓ^2 de modo a obter uma posterior codificação do sinal representado. O método de representação linear permite uma identificação natural entre o espaço de sinais e o espaço de representação. Com efeito, dada uma representação $L : S \rightarrow \ell^2$, definimos uma relação de equivalência no espaço de sinais S pondo

$$f \simeq g \quad \text{se, e somente se,} \quad L(f) = L(g). \quad (2.3.25)$$

A aplicação \bar{L} definida de modo que o diagrama



seja comutativo, é um isomorfismo linear. Podemos então usar o espaço de representação V ao invés do espaço de sinais. Advertimos o leitor no entanto que se a representação L não for exata temos uma perda de informação no processo de representação. Esse fato é ilustrado na Figura 2.11 para a amostragem pontual finita introduzida no exemplo 2.3: dois sinais distintos com a mesma representação pontual, $f(u_i) = g(u_i)$, $i = 1, \dots, 6$.

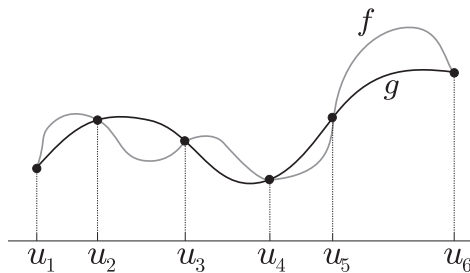


Figura 2.11: Sinais distintos com a mesma representação.

Em princípio, o isomorfismo \bar{L} do diagrama anterior estabelece apenas uma equivalência algébrica entre o espaço quociente e o espaço de representação. Para cada representação é importante que se faça uma análise detalhada da perda de informação do processo de passagem ao quociente.

2.3.1 Existência de Representação

No exemplo 2.3 introduzimos a representação pontual finita de um sinal. É claro que todo sinal admite essa representação, e isso mostra que sempre existe uma representação linear associada a um dado espaço de sinais. No entanto, conforme vimos acima, essa representação não é exata. O teorema abaixo garante a existência de representação exata de um espaço de sinais

Theorem 2.1. *(Teorema de representação.) Todo espaço de sinais \mathcal{S} , com energia finita, admite uma representação exata.*

ESBOÇO DA PROVA:

Considerando \mathcal{S} como um subespaço, não necessariamente fechado, de L^2 , existe um conjunto ortonormal $\{\dots, e_{-2}, e_{-1}, e_0, e_1, e_2, \dots\}$ de modo que para todo sinal f temos

$$f = \sum_{k=-\infty}^{\infty} c_k e_k, \quad (2.3.26)$$

onde $c_k = \langle f, e_k \rangle$. Além disso,

$$\sum_{k=-\infty}^{\infty} c_k^2 = \|f\|^2 < \infty, \quad (2.3.27)$$

e portanto $(c_k) \in \ell^2$. Definimos então uma representação pondo

$$f \mapsto L(f) = (\dots, c_{-2}, c_{-1}, c_0, c_1, c_2, \dots). \quad (2.3.28)$$

É imediato verificar que L é linear. \square

A amostragem de Fourier introduzida no exemplo 2.4 ilustra bem a construção da representação exata na demonstração acima. Nesse caso temos

$$e_k(t) = e^{2\pi i k \omega_0 t}. \quad (2.3.29)$$

2.4 Operações com Sinais

Uma *operação* em um espaço de sinais \mathcal{S} é uma aplicação

$$T : \mathbb{R}^m \times \mathcal{S}^n \rightarrow \mathcal{S}',$$

em um outro espaço de sinais \mathcal{S}' .

Conforme já ressaltamos anteriormente, sendo o espaço de sinais um espaço vetorial, duas operações comuns entre sinais são a *soma* de sinais,

$f + g : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{S}$, e o *produto* de um sinal por um escalar, $\lambda g : \mathbb{R} \times \mathcal{S} \rightarrow \mathcal{S}$, definidos por:

$$(f + g)(t) = f(t) + g(t); \quad (\lambda f)(t) = \lambda \cdot f(t), \quad \lambda \in \mathbb{R}. \quad (2.4.30)$$

Essas duas operações permitem, por exemplo, que se faça uma interpolação linear entre dois sinais f e g , obtendo um família h_u de sinais

$$h_u(t) = (1 - u)f(t) + ug(t), \quad u \in [0, 1]. \quad (2.4.31)$$

Essa família faz um efeito de transição contínua entre os sinais f e g , quando o parâmetro u varia de 0 até 1.

Quando os sinais tomam valores em \mathbb{R} , ou seja $f : U \subset \mathbb{R}^m \rightarrow \mathbb{R}$, ou em um outro espaço que admita uma estrutura de produto (como por exemplo o conjunto dos números complexos \mathbb{C}), podemos definir a operação de *produto* entre dois sinais, $fg : \mathcal{S} \times \mathcal{S} \rightarrow \mathcal{S}$, pondo

$$(fg)(x) = f(x) \cdot g(x), \quad (2.4.32)$$

onde \cdot é o produto em \mathbb{R} .

2.4.1 Filtragem

Uma operação unária $L : \mathcal{S} \rightarrow \mathcal{S}$ em um espaço de sinais é chamada de *filtragem*. A aplicação L é chamada de *filtro*. Se L for uma aplicação linear, ou seja,

$$L(f + g) = L(f) + L(g) \quad \text{e} \quad L(\lambda f) = \lambda L(f), \quad (2.4.33)$$

dizemos que L é um *filtro linear*. O filtro L é *espacialmente invariante* se

$$(Lf)(x - a) = L[f(x - a)], \quad (2.4.34)$$

ou seja, o comportamento do filtro L é o mesmo em qualquer ponto do domínio espacial do sinal.

Uma classe importante de filtros não-espacialmente invariantes, são os *filtros adaptativos*. Nesse caso, o valor do filtro em cada ponto depende do comportamento do sinal em uma vizinhança desse ponto.

Convolução e Resposta de Impulso

A *resposta de impulso* h do filtro L é a imagem $L(\delta)$ do sinal impulso δ pelo filtro. Da equação (2.2.9) temos que o sinal f é uma soma infinita de impulsos transladados e modulados convenientemente. Portanto se conhecermos

a resposta de impulso $h(t) = L(\delta)$ de um filtro linear e espacialmente invariante L , podemos afirmar que a resposta do filtro ao sinal f , $L(f)$, será uma soma infinita da função h transladada e modulada convenientemente, ou seja

$$Lf(x) = \int_{-\infty}^{+\infty} f(t)L[\delta(x-t)]dt = \int_{-\infty}^{+\infty} f(t)h(x-t)dt. \quad (2.4.35)$$

A segunda integral em (2.4.35) define uma operação chamada de *produto de convolução* de f e h . Mais precisamente, o produto de convolução $f * g$ entre dois sinais f e g é uma operação binária definida por

$$(f * g)(u) = \int_{-\infty}^{+\infty} f(t)g(u-t)dt.$$

O resultado acima pode ser resumido no teorema abaixo.

Theorem 2.2. *Um filtro linear e espacialmente invariante L fica determinado por sua função resposta de impulso h . Mais precisamente, para todo sinal f tem-se $L(f) = f * h$.*

A função resposta de impulso é também chamada de *núcleo do filtro*. Como em geral se utiliza filtros lineares e espacialmente invariantes, é muito comum uma identificação da operação de filtragem com a operação de convolução.

2.4.2 Transformadas

Uma transformação linear $T : \mathcal{S} \rightarrow \mathcal{S}'$ entre dois modelos funcionais distintos de um espaço de sinais é chamada de *transformada*. Em geral exigimos que essa transformação seja invertível, de modo a possibilitar uma conversão completa entre os dois modelos funcionais de sinais. Um exemplo importante de transformada é a *transformada de Fourier*, estudada anteriormente na seção 2.2.4. Essa transformada relaciona o modelo espacial com o modelo espectral de sinais, fornecendo informações sobre as frequências presentes em cada sinal.

O uso de transformadas permite a mudança entre diversos modelos de sinais de modo a se obter um modelo mais adequado à solução de determinado problema. Desse modo, é importante se obter as operações correspondentes em cada espaço de sinais através de uma transformada. Podemos ilustrar esse fato de modo mais preciso, usando uma operação binária como exemplo. Dada uma operação binária \oplus em um espaço de sinais \mathcal{S} , desejamos saber qual a operação correspondente \ominus no espaço \mathcal{S}' de forma que

o diagrama

$$\begin{array}{ccc}
 \mathcal{S} \times \mathcal{S} & \xrightarrow{(T, T)} & \mathcal{S}' \times \mathcal{S}' \\
 \oplus \downarrow & & \downarrow \ominus \\
 \mathcal{S} & \xrightarrow{T} & \mathcal{S}'
 \end{array}$$

seja comutativo, ou seja, dados $f, g \in \mathcal{S}$, $T(f \oplus g) = T(f) \ominus T(g)$.

Um caso particular, porém de grande importância, do problema acima, consiste em procurar a operação no modelo espectral de sinais que corresponde à operação de filtragem no modelo espacial. Se o filtro for linear e espacialmente invariante a resposta a essa pergunta é simples, e vem de uma relação clássica entre a transformada de Fourier e o produto de convolução. Mais precisamente, dados dois sinais f e g no domínio do espaço, tomando valores reais (ou complexos), tem-se

$$F(f * g) = F(f) \cdot F(g), \quad (2.4.36)$$

onde \cdot é o produto de sinais. Dessa forma, uma operação de filtragem com núcleo h no domínio do espaço corresponde no modelo espectral a uma operação de produto da transformada de Fourier do sinal pela transformada de Fourier $F(h) = \hat{h}$ do núcleo do filtro. A transformada \hat{h} do núcleo é chamada de *função de transferência* do filtro. Esse resultado é de grande importância tanto do ponto de vista de aplicações como do ponto de vista teórico. A função de transferência permite, por exemplo, uma análise do filtro com relação à sua atuação nos valores de frequência presentes no sinal. Estudaremos esse fato com maiores detalhes na seção seguinte.

2.4.3 Filtragem e Frequência

É comum dividir o domínio da frequência de um sinal em duas regiões: uma vizinhança da origem, que é chamada de *região de baixas frequências*, e o seu complemento, que é chamada de *região de altas frequências*. É claro que o conceito de baixas e altas frequências, e portanto o tamanho da vizinhança acima, depende diretamente da aplicação. Em algumas aplicações é útil se tomar uma partição do domínio de frequências, dividindo-o em diversas regiões disjuntas de forma a obter uma decomposição do sinal em diversas *bandas* do espectro de frequências.

O *suporte* $\text{supp}(f)$ de uma função f é o conjunto dos pontos do seu domínio onde f não se anula. Dizemos que um sinal f tem *banda limitada* quando o suporte de sua transformada de Fourier $\text{supp}\hat{f}$ está contido em um conjunto compacto. No caso unidimensional isso significa que

$\text{supp}(f) \subset [-\Omega, +\Omega]$, com $\Omega < \infty$. Quando o núcleo de um filtro possui suporte compacto, dizemos que o filtro tem *resposta de impulso finita*, ou é do tipo FIR (“finite impulse response”). Caso contrário, quando o núcleo de um filtro não tem suporte compacto, dizemos que o filtro tem *resposta de impulso infinita* ou é do tipo IIR (“infinite impulse response”).

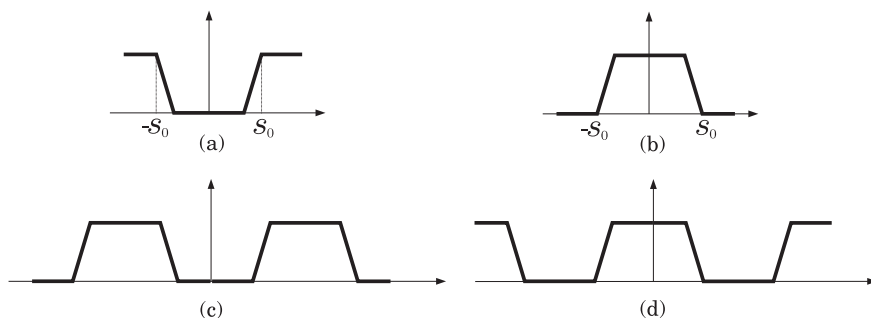


Figura 2.12: Funções de transferência.

Seja L um filtro e f um sinal. Dizemos que L é de *passa alta*, ou *filtro de agudização*, quando a operação de filtragem subtrai as baixas frequências do sinal. O filtro é de *passa baixa*, ou *filtro de suavização*, quando o sinal filtrado não possui altas frequências. O filtro L é de *para faixa*, ou *para banda*, quando ele elimina as frequências em uma determinada faixa do espectro. Finalmente, L é de *passa banda*, ou *passa faixa*, se $L(f)$ possui apenas frequências em uma determinada região do espectro. Os filtros de passa baixa também são conhecidos na literatura pelo nome de *filtros seletivos*. A função de transferência dos filtros de passa alta, passa baixa, passa banda e para banda possuem a forma mostrada na Figura 2.12 (a), (b), (c) e (d) respectivamente. É claro que um filtro de passa baixa é um filtro de passa banda para uma faixa de baixas frequências do sinal e, analogamente, um filtro de passa alta é um filtro de para banda para para uma faixa de baixas frequências do sinal.

O valor de frequência s_0 que delimita a região de baixas e altas frequências nos filtros de passa alta e passa baixa é chamado de *frequência de corte* (ver Figuras 2.12(a) e (b)).

2.5 Teoria de Amostragem

Nesta seção vamos revisar o conceito de amostragem com base na teoria de representação linear que estudamos anteriormente. Inicialmente vamos definir o conceito de *amostragem pontual*, que introduzimos brevemente no exemplo 2.3.

Seja $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ um sinal, e $\bar{U} = \{\dots, u_{-2}, u_{-1}, u_0, u_1, u_2, \dots\}$ um subconjunto enumerável de U . Uma amostragem pontual de f é a seqüência

$$(\dots, f(u_{-2}), f(u_{-1}), f(u_0), f(u_1), f(u_2), \dots). \quad (2.5.37)$$

Esse processo é mostrado na Figura 2.13. Cada valor $f(u_i)$ é chamado de *amostra* do sinal. A seqüência em (2.5.37) é chamada de *seqüência de amostras*.

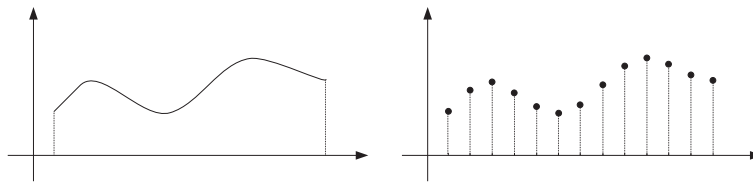


Figura 2.13: Amostragem pontual de um sinal.

2.5.1 Amostragem Pontual Uniforme

A amostragem pontual uniforme é bastante comum e importante na teoria de sinais. Além de permitir um tratamento matemático relativamente simples, ela é utilizada nos diversos dispositivos que manipulam sinais digitais.

Inicialmente precisamos definir o conceito de reticulado. Tomemos para cada eixo x_i de \mathbb{R}^n um segmento Δx_i não nulo. Um *reticulado* no espaço \mathbb{R}^n é o conjunto dos pontos de \mathbb{R}^n definido por

$$R_{\Delta} = R_{\Delta x_1, \dots, \Delta x_n} = \{(m_1 \Delta x_1, m_2 \Delta x_2, \dots, m_n \Delta x_n) ; m_i \in \mathbb{Z}\}. \quad (2.5.38)$$

Cada um dos pontos do reticulado, é chamado de *vértice do reticulado*. É comum nos referirmos ao reticulado através do vetor $\Delta = (\Delta x_1, \dots, \Delta x_n)$ que o define.

A amostragem pontual é dita *uniforme* quando o conjunto de pontos u_i em (2.5.37) são vértices de um reticulado do espaço. O número de amostras por unidade de espaço, é chamado de *taxa de amostragem*, ou *frequência de amostragem*.

Conforme vimos na seção 2.3 (ver Figura 2.11), em geral há uma perda no processo de amostragem pontual. Nesta seção estamos interessados em estudar dois problemas:

- Em que condições podemos evitar a perda no processo de amostragem pontual?
- No caso de amostragem pontual sem perda, que método pode ser utilizado para reconstruir o sinal original?

As duas perguntas acima podem ser reformuladas do seguinte modo:

- Quando é possível reconstruir o sinal f a partir de sua seqüência de amostras (2.5.37)?
- Que método de interpolação deve ser utilizado para recuperar o sinal a partir de suas amostras?

Ambas as perguntas podem ser reunidas em um único problema: *em que condições o método de amostragem pontual determina uma representação linear exata de um sinal?*. Vamos estudar essa pergunta para a amostragem uniforme.

Conforme podemos observar na Figura 2.13, o processo de amostragem pontual substitui o sinal por uma seqüência de “impulsos finitos”. Esse processo introduz altas freqüências no sinal amostrado. Essas altas freqüências se combinam com as freqüências do sinal original para formar o modelo espectral do sinal amostrado. Portanto, a solução de nosso problema pode ser obtida analisando o espectro do sinal amostrado. Com efeito, basta utilizarmos um filtro adequado para extrair o espectro do sinal original a partir do espectro do sinal amostrado. Obtemos assim o espectro do sinal original, e a partir desse modelo obtemos o modelo espacial do sinal original utilizando a transformada de Fourier inversa. Ou seja, se f é o sinal original e f_d o sinal amostrado, e F a transformada de Fourier, temos

$$F(f_d) \rightarrow \text{Filtragem} \rightarrow F(f) \rightarrow f = F^{-1}(F(f)).$$

Pelo visto acima devemos analisar as freqüências do sinal amostrado em busca da solução. Desse modo, devemos procurar uma representação desse sinal no domínio da freqüência. Com essa finalidade, vamos precisar de vários resultados sobre transformadas de Fourier. Vamos fazer uma exposição desses resultados sem demonstração.

2.5.2 Amostragem Pontual e Transformada de Fourier

A amostragem pontual do sinal impulso, δ de Dirac, é a função δ de Kronecker, definido por

$$\delta(p) = \begin{cases} 1 & \text{se } p = 0; \\ 0 & \text{se } p \neq 0. \end{cases} \quad (2.5.39)$$

Chamamos de *função pente* associada a um reticulado R_Δ , onde Δ é o vetor $(\Delta x_1, \dots, \Delta x_n)$, ao sinal definido por

$$\text{comb}_\Delta = \sum_{k, i_k \in \mathbb{Z}} \delta(x - (i_1 \Delta x_1, \dots, i_n \Delta x_n)), \quad x \in \mathbb{R}^n. \quad (2.5.40)$$

Na Figura 2.14 ilustramos a função pente, para os casos unidimensional e bidimensional.

Figura 2.14: Função pente.

A importância da função pente está no seguinte resultado: *fazer uma amostragem pontual uniforme de um sinal f no reticulado $R_{\Delta x_1, \dots, \Delta x_n}$ é equivalente a multiplicar f pela função pente associada ao reticulado.* No caso unidimensional, temos

$$f \cdot \text{comb}_{\Delta t} = (f(k\Delta t))_{k \in \mathbb{Z}}, \quad (2.5.41)$$

conforme ilustrado na Figura 2.15. Por essa razão a função pente é também chamada de *sinal de amostragem*.

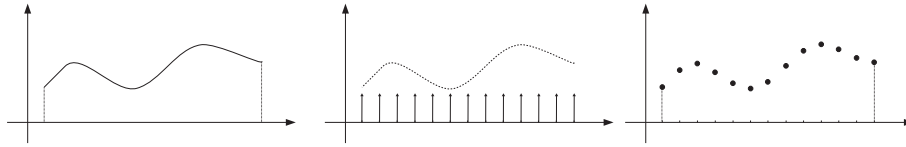


Figura 2.15: Amostragem pontual e função pente.

Um outro resultado importante para a análise de nosso problema consiste em se obter o modelo espectral da função pente. O resultado, não muito intuitivo, é uma outra função pente. Mais precisamente, temos

$$F(\text{comb}_{\Delta t}(t)) = \frac{1}{\Delta t} \text{comb}_{1/\Delta t}(s). \quad (2.5.42)$$

Finalmente, segue-se da equação (2.2.9) que $(f * \delta(t - t_0))(t) = f(t - t_0)$. Esse resultado pode ser generalizado para se obter o produto de convolução de um sinal f por uma função pente $\text{comb}_{\Delta t}(t)$. O resultado, no caso unidimensional, é o seguinte:

$$(f * \text{comb}_{\Delta t}) = \sum_{k=-\infty}^{+\infty} f(t - k\Delta t). \quad (2.5.43)$$

Ou seja, o sinal f é transladado para os vértices do reticulado e o resultado é obtido somando todos os transladados do sinal. Esse resultado é ilustrado na Figura 2.16. É claro que esse resultado é válido para sinais multidimensionais.

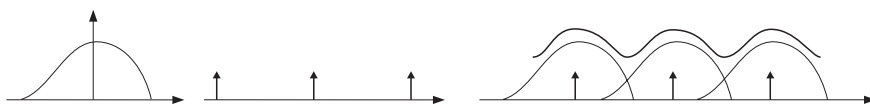


Figura 2.16: Convolução de um sinal F com uma função pente.

2.5.3 Teorema da Amostragem

Usando os resultados acima podemos finalmente obter informações sobre o modelo espectral de um sinal amostrado. Fazemos isso na Figura 2.17: no domínio do tempo a amostragem corresponde ao produto com a função pente; no domínio da frequência esse produto corresponde ao produto de convolução com uma transformada da função pente, obtida utilizando a equação (2.5.43). Na Figura 2.17(c) o leitor pode observar o sinal amostrado e a sua transformada de Fourier.

Concluimos que o modelo espectral de um sinal amostrado em um reticulado definido por Δt é formado a partir do modelo espectral do sinal original transladado por intervalos de comprimento $1/\Delta t$. A Figura 2.18 é uma versão ampliada do modelo espectral do sinal amostrado na Figura 2.17. Conforme o leitor pode observar, no processo de transladar e somar o espectro do sinal original f pode haver uma superposição de frequências. Mais precisamente, altas frequências do sinal aparecem como baixas frequências no espectro do sinal amostrado. No espectro final, segmento AB na figura, essas altas frequências se perdem tornando impossível a recuperação da informação correta das frequências presentes no sinal original.

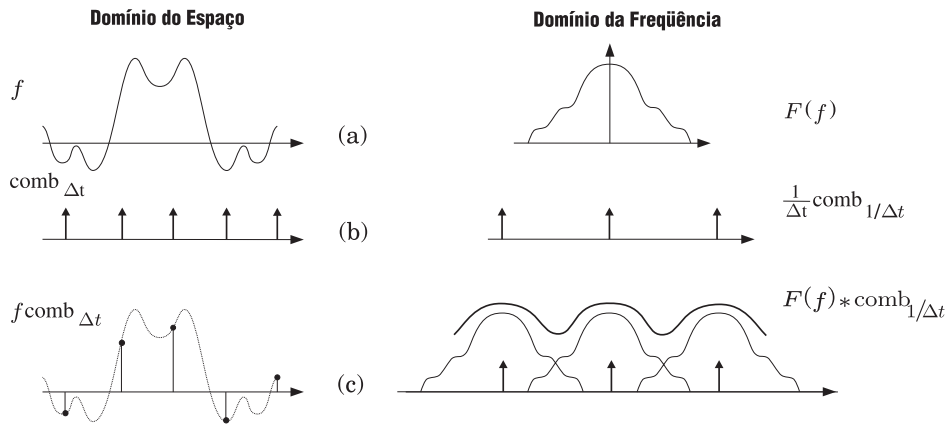


Figura 2.17: Amostragem do sinal no domínio do tempo e da frequência.

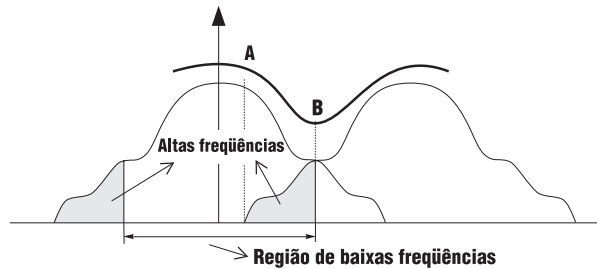


Figura 2.18: Superposição de frequências no sinal amostrado.

Considere agora um sinal f com banda limitada no domínio $[-\Omega, \Omega]$ de frequências. Conforme ilustramos na Figura 2.19, se

$$\frac{1}{\Delta t} > 2\Omega, \quad \text{ou seja,} \quad \Delta t < \frac{1}{2\Omega}, \quad (2.5.44)$$

no modelo espectral do sinal amostrado não haverá superposição de frequências, portanto o modelo espectral do sinal original pode ser obtido usando um filtro de passa baixa com ganho unitário, cuja função de transferência é a função pulso $p_\Omega(s)$, é mostrada na Figura 2.19. O sinal original pode então ser obtido usando a transformada de Fourier inversa após o processo de filtragem. Esse argumento demonstra o teorema da amostragem de Shannon-Whittaker conforme enunciamos abaixo.

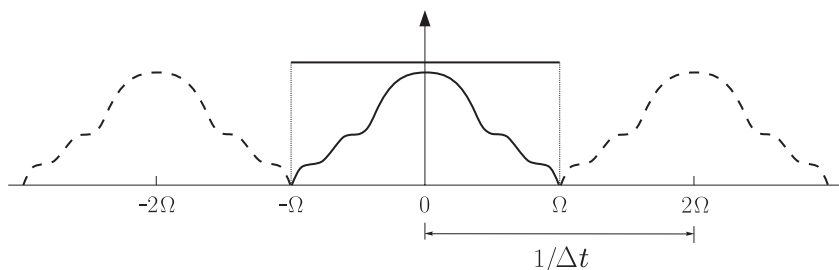


Figura 2.19: Modelo espectral de um sinal de banda limitada amostrado.

Theorem 2.3. (Teorema da amostragem de Shannon-Whittaker.) Sejam f um sinal de banda limitada, e Ω o menor valor de frequência tal que $\text{supp}(\hat{f}) \subset [-\Omega, \Omega]$. O sinal f pode ser reconstruído de forma exata a partir de um conjunto uniforme de amostras $\{m\Delta t; m \in \mathbb{Z}\}$ se $\Delta t < 1/(2\Omega)$.

Em outras palavras, o teorema afirma que se o sinal f é limitado com uma banda de 0 a Ω ciclos por unidade de tempo, então ele fica completamente determinado por uma amostragem pontual uniforme em um reticulado de frequência $\Delta t < 1/(2\Omega)$. Isso significa que devemos tomar pelo menos uma amostra a cada meio ciclo do sinal.

O limite da frequência de amostragem $1/(2\Omega)$ é conhecido como *limite de Nyquist*, em homenagem a H. Nyquist, que na década de 20 apontou para a importância do limite em telegrafia.

O teorema de Shannon-Whittaker faz uma relação entre a presença de altas frequências na imagem e a frequência de amostragem. De forma intuitiva, quanto maior for a presença de altas frequências na imagem, maior deve ser a frequência de amostragem de forma a garantir uma reconstrução sem problemas. Observamos que o teorema se estende para a reconstrução de sinais m -dimensionais.

Base de Reconstrução Ideal

O teorema de Shannon-Whittaker não fala sobre qual o método adequado de interpolação para se reconstruir o sinal de forma exata a partir de suas amostras feitas obedecendo o limite de Nyquist. Uma base de reconstrução ideal pode ser obtida explicitando os cálculos que nos levaram à demonstração do teorema de amostragem de Shannon-Whittaker, na seção anterior.

Com efeito, dado um sinal f , $\text{supp} f \subset [-\Omega, \Omega]$, e a sua seqüência de

amostragem $(f(k\Delta t))$, $k \in \mathbb{Z}$, tem-se que

$$(f(k\Delta t)) = f(t) \cdot \text{comb}_{\Delta t}(t) \quad (2.5.45)$$

Usando a transformada de Fourier, obtemos o modelo espectral do sinal amostrado:

$$F[f(t) \cdot \text{comb}_{\Delta t}(t)] = \hat{f}(s) * \hat{f}[\text{comb}_{\Delta t}(t)] = \frac{1}{\Delta t} \hat{f}(s) * \text{comb}_{1/\Delta t}(t). \quad (2.5.46)$$

As freqüências do sinal original podem ser obtida usando um filtro de passa baixa cuja função de transferência é o pulso $p_{\Omega}(s)$, com a finalidade de obter o modelo espectral $\hat{f}(s)$ do sinal f . Temos então

$$\hat{f}(s) = \left[\frac{1}{\Delta t} \hat{f}(s) * \text{comb}_{1/\Delta t}(t) \right] p_{\Omega}(s). \quad (2.5.47)$$

Para obter o modelo espacial do sinal, devemos usar a transformada de Fourier inversa na equação acima. Temos

$$f(t) = F^{-1} \left[\frac{1}{\Delta t} \hat{f}(s) * \text{comb}_{1/\Delta t}(t) \right] * F^{-1} [p_{\Omega}(s)]. \quad (2.5.48)$$

Usando a equação (2.4.36), juntamente com o fato de que

$$F^{-1}(p_{\Omega}(s) \frac{1}{2\Omega}) = 2\Omega \text{sinc}(2\pi\Omega t), \quad (2.5.49)$$

obtemos

$$f(t) = f(t) \cdot \text{comb}_{\Delta t}(t) * \text{sinc}(2\pi\Omega t). \quad (2.5.50)$$

Das equações (2.5.41) e (2.5.43), segue-se que

$$f(t) = \sum_{k=-\infty}^{+\infty} f(k\Delta t) 2\Omega \text{sinc}[2\pi\Omega(t - k\Delta t)]. \quad (2.5.51)$$

A equação (2.5.51) é exatamente a expressão de reconstrução exata do sinal f a partir das amostras $f(k\Delta t)$, $k \in \mathbb{Z}$. Pode-se mostrar que a série na equação (2.5.51) converge absolutamente, e a convergência é uniforme em partes compactas do domínio.

Na realidade pode-se mostrar que normalizando os elementos do conjunto

$$\{\text{sinc}[2\pi\Omega(t - k\Delta t)]\}, \quad k \in \mathbb{Z}, \quad (2.5.52)$$

obtemos uma base ortonormal completa em um espaço de sinais conveniente. Essa base é chamada de *base de Shannon*. Usando a notação da demonstração acima, o teorema de amostragem de Shannon-Whittaker pode então ser enunciado do seguinte modo:

Theorem 2.4. (*Teorema de reconstrução ideal.*) Se um sinal f tem banda limitada, $\text{supp} f \subset [-\Omega, \Omega]$, e f é amostrado pontualmente dentro do limite de Nyquist, então a seqüência de amostras ($f(k\Delta t)$) é uma representação exata do sinal. O sinal f pode ser reconstruído a partir de suas amostras usando a base de Shannon.

2.5.4 Aliasing

Consideremos um sinal f qualquer. Vimos que o problema de recuperar f a partir de suas amostras está diretamente relacionado com o problema de superposição de freqüências no processo de amostragem: Altas freqüências do sinal aparecem no sinal amostrado como componentes de baixas freqüências (ver Figura 2.18).

Essas componentes de baixas freqüências são chamadas de *alias* das componentes de altas freqüências, e o fenômeno é chamado de *aliasing*. Quando ocorre o fenômeno de aliasing, o processo de filtragem usando um filtro de passa baixa ideal, conforme feito na reconstrução ideal com a base de Shannon, não funciona, porque não conseguimos recuperar as altas freqüências que se manifestam como baixas freqüências. Isso significa que a série em (2.5.51) não converge para o sinal f , o que dá origem a um *erro de aliasing* calculado por

$$\varepsilon = \left| f(t) - \sum_{k=-\infty}^{+\infty} f(k\Delta t) \text{sinc}[2\pi\Omega(t - k\Delta t)] \right|. \quad (2.5.53)$$

Nesse caso o sinal construído pode não representar uma boa aproximação do sinal originalmente amostrado.

É bastante instrutivo olhar o fenômeno de aliasing no domínio do espaço. Para isso, vamos dar um exemplo usando um sinal periódico. Considere o sinal $f(t) = \text{sen}(2\pi\omega_0 t)$, que tem freqüência ω_0 . O gráfico do modelo espectral do sinal f é mostrado na Figura 2.8, página 25. É claro que $\text{supp} f \subset [-\omega_0, \omega_0]$. O limite de Nyquist é portanto dado por

$$\frac{1}{\Delta t} > 2\omega_0, \quad \text{ou seja } \Delta t < \frac{1}{2\omega_0}. \quad (2.5.54)$$

Isso significa que devemos tomar pelo menos uma amostra em cada meio ciclo do sinal, conforme ilustramos na Figura 2.20.

Na Figura 2.21(a) temos uma freqüência de amostragem quatro vezes superior ao limite de Nyquist, e vemos claramente que mesmo uma interpolação linear fornece uma boa aproximação do sinal. Na Figura 2.21(b) a freqüência de amostragem é exatamente igual ao limite de Nyquist (o dobro

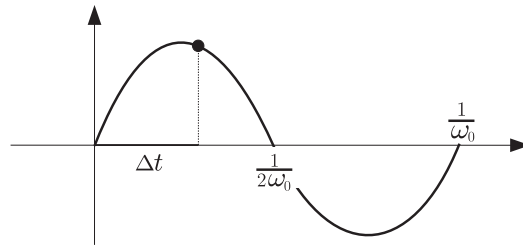


Figura 2.20: Taxa de amostragem e frequência.

da frequência do sinal). Nesse caso as amostras estão todas localizadas sobre o eixo- x , e, claramente, uma interpolação linear não é uma boa aproximação do sinal original. Na Figura 2.21(c) a frequência de amostragem é inferior ao limite de Nyquist. Observe que nessa figura fazemos uma interpolação das amostras obtidas e obtemos, na curva traçada com linha mais grossa, um sinal completamente distinto da senóide original.

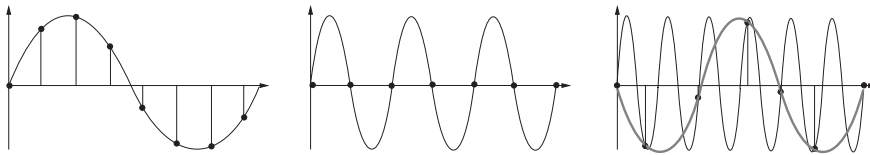


Figura 2.21: Fenômeno de aliasing no domínio do tempo.

O leitor deve observar que a reconstrução do sinal na Figura 2.21(c) obtém um sinal próximo de uma senóide, com uma frequência bem inferior à frequência do sinal original. Ou seja, as altas frequências do sinal original aparecem como baixas frequências no sinal reconstruído. Esse fato caracteriza o fenômeno de aliasing.

Temos dois casos onde a superposição de frequências, que dá origem ao fenômeno de aliasing, pode ocorrer:

- O sinal não possui banda limitada;
- O sinal possui banda limitada porém a amostragem é feita abaixo da frequência de Nyquist.

Para evitar, ou pelo menos minimizar, a ocorrência de aliasing na amostragem, devemos procurar satisfazer o limite de Nyquist

$$\frac{1}{\Delta t} > 2\Omega, \quad \text{ou seja,} \quad \Delta t < \frac{1}{2\Omega}. \quad (2.5.55)$$

Para satisfazer a desigualdade acima, devemos reduzir o valor de Δt , ou então diminuir o valor de Ω . Esse fato dá origem aos dois métodos possíveis de minimizar, ou evitar, o aliasing:

- A redução do valor de Δt , corresponde a aumentar a frequência de amostragem;
- A redução do valor de Ω , corresponde a reduzir a presença de altas frequências no sinal.

A frequência de amostragem está diretamente relacionada com a resolução do dispositivo de saída do sinal. Além disso, o aumento excessivo da frequência de amostragem traz problemas consideráveis de armazenamento e de processamento do sinal. Já a redução das altas frequências, pode ser convenientemente obtida através do uso de um filtro de passa baixa.

Finalizamos esta seção lembrando ao leitor que aliasing é um fenômeno estreitamente ligado à amostragem pontual de um sinal. Quando utilizamos uma representação linear associada a uma base ortonormal completa do espaço o problema de aliasing não se coloca. Temos, é claro, problemas de uma outra natureza conforme vamos estudar mais adiante na seção de reconstrução de sinais.

2.6 Operações no Dominio Discreto

Nesta seção vamos estender as operações com sinais estudadas na seção 2.4, para o universo discreto. Nesse caso, cada sinal f é dado por uma seqüência de representação

$$(f_n) = (\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots).$$

em ℓ^2 . A fim de facilitar a notação, é comum indicarmos f_i por $f(i)$. Chamamos a atenção do leitor que ao escrever (f_n) nos referimos à seqüência, portanto o índice n pode ser substituído por outra letra qualquer. Por outro lado, ao escrevermos f_n , sem o parênteses, nos referimos ao n -ésimo termo da seqüência.

As operações usuais de soma, produto, produto por um escalar, se estendem facilmente para a versão discreta de um sinal. Em geral dada uma

operação $L : \mathcal{S} \rightarrow \mathcal{S}'$ no domínio contínuo, devemos definir uma operação $L' : \mathcal{S}'_d \rightarrow \mathcal{S}'_d$ no domínio discreto, de modo que o diagrama

$$\begin{array}{ccc} \mathcal{S} & \xrightarrow{L} & \mathcal{S} \\ R \downarrow & & \downarrow R \\ \mathcal{S}'_d & \xrightarrow{L'} & \mathcal{S}'_d \end{array}$$

seja comutativo, ou seja $L'(R(f)) = R(L(f))$, onde R é a transformação de representação do sinal. Em outras palavras, a ordem em que fazemos a operação e a discretização do sinal é irrelevante. Vamos no que se segue estudar esse problema para a operação de filtragem e para a transformada de Fourier.

2.6.1 Convolução Discreta

A operação de filtragem com um filtro linear e espacialmente invariante, se reduz a um produto de convolução, conforme definimos na seção 2.4. Dados dois sinais f e g , e suas seqüências de discretização (f_m) , $m \in \mathbb{Z}$ e (g_n) , $n \in \mathbb{Z}$, o produto de convolução é o sinal discreto (h_k) tal que, para cada $k \in \mathbb{Z}$

$$h_k = h(k) = ((f_m) * (g_n))_k = \sum_{j=-\infty}^{+\infty} f_j g_{k-j}, \quad (2.6.56)$$

supondo que a série converge, o que sempre ocorre se f_k e g_k são seqüências em ℓ^2 .

Observamos que se os sinais (f_m) e (g_n) estão definidos em um conjunto finito de amostras, como por exemplo, $k = 0, 1, \dots, N-1$, o sinal (g_k) deve ser estendido além desse domínio uma vez que devemos calcular g_{k-j} para $k = 0, \dots, N-1$. Existem diversos modos de se estender um sinal (g_k) . Trataremos desse problema com mais detalhes no capítulo sobre operações com imagem digital.

Um caso importante de convolução, é quanto tomamos o produto de convolução da representação discreta $(f(t_k))$, $k \in \mathbb{Z}$, do sinal f , com um sinal contínuo h . O resultado é dado por

$$g(t) = (f(t_k) * h) = \sum_{j=-\infty}^{+\infty} f(t_j) h(t - t_j), \quad (2.6.57)$$

É interessante observar que essa equação nos fornece uma interpolação das amostras $f(t_j)$ do sinal f .

2.6.2 Transformada de Fourier Discreta

Vamos tratar da transformada de Fourier discreta apenas para o caso de seqüências finitas. Nesse caso poderíamos chama-la mais apropriadamente de *transformada de Fourier finita*. Dado uma seqüência $(f_0, f_1, \dots, f_{N-1})$, a transformada discreta, $(\hat{f}(j))$, dessa seqüência é a seqüência definida por

$$\hat{f}(j) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-\frac{2\pi i k j}{N}}, \quad j = 0, 1, \dots, N-1. \quad (2.6.58)$$

A transformada inversa $F_d^{-1}(F(j))$ é a seqüência (f_k) definida por

$$f(k) = \sum_{j=0}^{N-1} \hat{f}(j) e^{\frac{2\pi i k j}{N}}, \quad k = 0, 1, \dots, N-1. \quad (2.6.59)$$

Ao contrário do desenvolvimento da transformada de Fourier contínua, a matemática envolvida no desenvolvimento da teoria da transformada discreta se reduz à aritmética dos números reais e complexos. Sugerimos, a título de familiarização, que o leitor faça as contas para mostrar que, de fato, a transformada discreta inversa F_d^{-1} é realmente a inversa da transformada F_d .

Um fato importante relacionado com a transformada de Fourier discreta é que $F_{k+N} = F(k)$, ou seja a seqüência (F_j) é periódica de período N . Esse fato é bastante intuitivo para o leitor que entendeu bem a nossa discussão sobre a amostragem dada anteriormente: a transformada de Fourier de um sinal discreto, replica o espectro do sinal original. A demonstração pode ser feita diretamente a partir da equação (2.6.58).

Segue-se da periodicidade de F_k , e da existência da transformada inversa, que a seqüência (f_k) , $k = 0, \dots, N-1$, é uma subseqüência finita de uma seqüência infinita de período N , definida pela equação (2.6.59). Isso não significa que a seqüência (f_j) original deva ser periódica para calcular a transformada discreta. Na realidade essa seqüência é finita com N elementos. O que dissemos acima significa que ao usarmos a equação (2.6.59) para reobter o sinal discreto (f_k) obtemos na realidade uma extensão periódica desse sinal.

Aspectos Computacionais

Consideramos nesta seção o seguinte problema: *como calcular a transformada de Fourier $\hat{f} = F(f)$ do sinal f a partir de sua seqüência de amostras $(f(k))$?*

Suponhamos que o sinal f tem suporte contido no intervalo $[0, T]$, $T > 0$. Usando um reticulado uniforme Δt com frequência de amostragem $1/\Delta t$, de acordo com o limite de Nyquist, e tomando N amostras de modo que $N\Delta t \geq 2T$, pode-se mostrar que

$$F(j) = N\Delta t (\hat{f}(k)). \quad (2.6.60)$$

A equação acima relaciona a transformada de Fourier discreta (no segundo membro) das amostras do sinal, com a amostragem $F(j)$ da transformada de Fourier do sinal f . Esse fato nos fornece um método de calcular a transformada de Fourier (contínua) de um sinal, conforme ilustramos no diagrama da Figura 2.22:

- Discretizamos o sinal f ;
- Aplicamos a transformada de Fourier discreta;
- Obtemos a discretização da transformada de Fourier usando a equação (2.6.60);
- Reconstruímos o sinal discreto $F(j)$ para obter a transformada contínua $F(f)$.

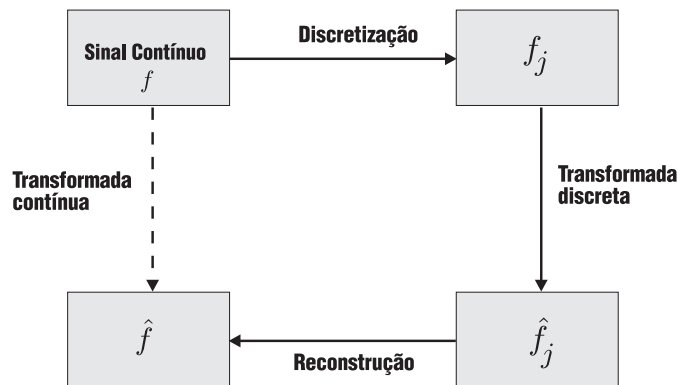


Figura 2.22: Cálculo da transformada de Fourier usando a transformada discreta.

Devemos chamar a atenção do leitor para três fatos importantes ao utilizar o método acima para calcular a transformada de Fourier contínua de um sinal:

MUDANÇA DE DOMÍNIO.

A equação 2.6.60 foi obtida supondo que o sinal f tem suporte compacto no intervalo $[0, N]$. Caso isso não seja verdadeiro, fazemos uma translação do intervalo de suporte do sinal por um fator constante A , e substituímos o sinal f pelo sinal transladado $g(t) = f(t - A)$. A transformada do sinal f é obtida utilizando a equação

$$\hat{f}(s) = \hat{g}e^{i2\pi sa}. \quad (2.6.61)$$

SIMETRIA.

A transformada de Fourier discreta de um sinal com N amostras é um outro sinal discreto no domínio da frequência, com N amostras nos pontos $0, 1, \dots, N - 1$. Usamos a periodicidade de \hat{f} , $\hat{f} = \hat{f}(-j)(N - j)$, de modo a obter a transformada no intervalo $[-N/2, N/2]$.

TRANSFORMADA FINITA.

Todos os cálculos anteriores foram feitos sob a suposição de que o suporte do sinal f amostrado é um intervalo compacto $[a, b]$. Se isso não ocorrer, devemos fazer um recorte de f , substituindo-o pelo sinal

$$h(x) = \begin{cases} f(x) & \text{se } x \in [a, b]; \\ 0 & \text{se } x \notin [a, b]. \end{cases}$$

Por essa razão alguns textos chamam a transformada de Fourier acima calculada de *transformada de Fourier finita*. É claro que, a menos que o sinal tenha suporte compacto, a transformada de Fourier calculada acima não coincide com a transformada de Fourier do sinal. O leitor deve sempre lembrar esse fato para evitar problemas.

A operação de recorte equivale a multiplicar o sinal f por uma função pulso. Esse recorte do sinal introduz altas frequências no modelo espectral do sinal. Esse fenômeno, chamado de *vazamento*, é o análogo no domínio da frequência ao conhecido fenômeno de Gibbs no domínio do espaço, que ocorre ao truncarmos a série no domínio da frequência.

O cálculo da transformada de Fourier discreta diretamente a partir da equação (2.6.58) é bastante caro computacionalmente. Esse cálculo pode ser feito de forma mais eficiente utilizando uma combinatória eficiente dos diversos índices que determinam o cálculo da transformada. Existem diversos algoritmos com essa finalidade. Esses algoritmos recebem a denominação genérica de *transformada de Fourier rápida* (“fast Fourier transform”).

2.7 Reconstrução de Sinais

Chamamos de *reconstrução de um sinal* ao problema de se obter um sinal contínuo a partir de uma representação discreta. Em geral temos três aspectos de um sinal em um sistema: o sinal contínuo f , o sinal discreto f_d , e o sinal reconstruído f_r . Idealmente, deveríamos ter $f = f_r$, porém nem sempre isso é possível conforme vimos na seção 2.5. As diversas técnicas de reconstrução têm por objetivo minimizar o erro $|f - f_r|$ entre o sinal contínuo f e o sinal reconstruído f_r .

As técnicas de reconstrução são muito importantes na manipulação de sinais no computador por diversas razões. Por um lado, na solução de alguns problemas necessitamos da representação contínua do sinal. Por outro lado, vale lembrar que os equipamentos de saída utilizados por um sistema para comunicar o sinal para o usuário fazem uma reconstrução do sinal digital. Um exemplo são os auto-falantes que reconstroem o sinal digital de voz que vem de um CD. O perfeito conhecimento das técnicas de reconstrução de cada dispositivo é importante na elaboração de algoritmos que processam sinais para serem enviados para um determinado dispositivo de saída.

Conforme estudamos na seção 2.3, conhecendo-se a base de representação linear $\{e_i\}$ o sinal pode ser reconstruído de forma exata a partir das amostras c_i , pondo

$$f(t) = \sum_{i=-\infty}^{+\infty} c_i e_i(t). \quad (2.7.62)$$

A equação (2.7.62) define uma interpolação entre os diversos valores c_i das amostras do sinal. Esse fato deve ser bem lembrado pelo leitor: *reconstrução* e *interpolação* são nomes distintos para o mesmo problema.

Na prática a reconstrução exata é, em geral, impossível de ser obtida devido aos seguintes problemas:

- Os elementos $e_i(t)$ da base podem não ter suporte compacto. Isso ocorre, por exemplo, quando as hipóteses do teorema de Shannon-Whittaker são satisfeitas e temos reconstrução exata no processo de amostragem pontual: os elementos da base de Shannon não possuem suporte compacto.
- A série em (2.7.62) pode possuir uma infinidade de termos não nulos. Nesse caso somos obrigados a truncar a série de reconstrução. Isso ocorre, por exemplo, na amostragem de Fourier (Exemplo 2.4), onde a soma de apenas um número finito de termos da série pode dar origem ao conhecido *fenômeno de Gibbs*.

- Pode não existir uma representação linear exata associada ao processo de amostragem utilizado. Isso ocorre, por exemplo, quando a amostragem pontual é feita fora do limite de Nyquist, conforme vimos anteriormente.

Nos dois primeiros casos, temos um *erro de truncamento*, que é dado por

$$\varepsilon = \left| f(t) - \sum_{k=-N}^{+N} c_k e_i(t) \right| = \left| \sum_{|k|>N} c_k e_i(t) \right|. \quad (2.7.63)$$

Um exemplo do terceiro caso ocorre quando fazemos amostragem pontual de um sinal sem obedecer ao limite de Nyquist. O resultado neste caso é a ocorrência de aliasing.

Quando temos a amostragem pontual de um sinal f , $(f_k) = (f(t_k))$, $k \in \mathbb{Z}$, a reconstrução é obtida mediante um processo de convolução com uma base de reconstrução h . Tem-se

$$f(t) = (f_k) * h = \sum_{k=-\infty}^{+\infty} f(t_k) h(t - t_k), \quad (2.7.64)$$

onde t_k , $k \in \mathbb{Z}$, são os pontos do domínio onde o sinal foi amostrado, e c_k são os valores das amostras nesses pontos, isto é $c_k = f(t_k)$. Geometricamente, essa equação indica que os elementos da base de reconstrução são obtidos por translação $h(t - t_k)$ da função h . Esse fato ocorre, por exemplo, no processo de reconstrução ideal com a base de Shannon definida pela equação (2.5.51).

O leitor deve lembrar da Seção 2.6.1, que a equação (2.7.64) é o produto de convolução do sinal discreto $(f_k) = (f(t_k))$ com a função h , chamada de *núcleo de interpolação* ou *núcleo de reconstrução* (ver equação (2.6.56)). Nesse caso temos uma relação imediata entre o processo de reconstrução e a operação de filtragem de sinais: *a reconstrução de um sinal f pode ser vista como um processo de filtragem do sinal discreto f_d utilizando um filtro linear e espacialmente invariante com núcleo h* . Nesse caso a comparação entre os diversos métodos de reconstrução se reflete na comparação dos núcleos de interpolação.

Inspirado na equação (2.7.64), podemos buscar definir núcleos de interpolação h para utilizar na reconstrução de sinais. Em geral no entanto o conjunto $\{h(t - t_k)\}$, $k \in \mathbb{Z}$, gera apenas um subespaço V do espaço de sinais \mathcal{S} . Quando $V = \mathcal{S}$ temos reconstrução exata, porém quando $V \neq \mathcal{S}$ temos reconstrução exata apenas para sinais $f \in V$. Na próxima seção analisaremos com detalhes um caso particular.

2.7.1 Reconstrução e Aliasing

Consideraremos nesta seção o processo de reconstrução com o filtro cujo núcleo é a função pulso $p_a(t)$ introduzida no exemplo 2.1. Nesse contexto a função pulso é chamada de *filtro retangular* ou *filtro "box"*. O sinal reconstruído é definido por

$$f_r(t) = \sum_k f(t_k) p_{\Delta t/2}(t - t_k). \quad (2.7.65)$$

Esse sinal é uma soma de funções pulso moduladas pelos valores das amostras, sendo portanto constante em cada intervalo da partição definida pelo reticulado. Dizemos que o sinal original é aproximado por um sinal constante por partes. É claro que se o sinal original não é constante por partes, a reconstrução não é exata. Esse método de reconstrução é ilustrado na Figura 2.23.

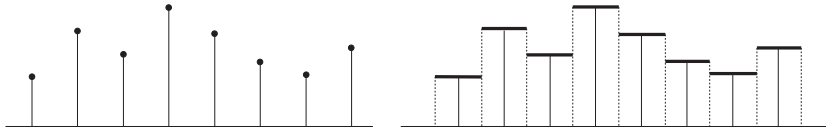


Figura 2.23: Reconstrução com o filtro retangular.

O sinal reconstruído possui descontinuidades que se traduzem na presença de altas freqüências. Podemos ver esse fato no domínio da freqüência. Com efeito, no domínio do espaço o sinal é reconstruído pelo produto de convolução na equação (2.7.65). No domínio da freqüência o modelo espectral do sinal reconstruído é dado então pelo produto do modelo espectral do sinal discreto (f_k) pela função de transferência $\hat{p}_{\Delta t/2}(t)$ do filtro retangular. Sabemos que

$$\hat{p}_{\Delta t/2}(t) = \Delta t \operatorname{sinc}(2\pi \Delta t s). \quad (2.7.66)$$

Na Figura 2.24 mostramos o modelo espectral de um sinal amostrado superposto ao gráfico da função de transferência $\hat{p}_{\Delta t}(s)$. Conforme mostra a figura, mesmo que o sinal tenha banda limitada e a amostragem seja feita de acordo com o limite de Nyquist, como o filtro não possui freqüência de corte, o modelo espectral do sinal reconstruído possui altas freqüências não existentes no sinal original. Essas altas freqüências se manifestam perceptualmente no sinal reconstruído de diferentes modos, dependendo, é claro, do tipo de sinal.

Como evitar, ou minimizar o aparecimento de altas freqüências no sinal reconstruído? Temos duas possíveis soluções para esse problema:

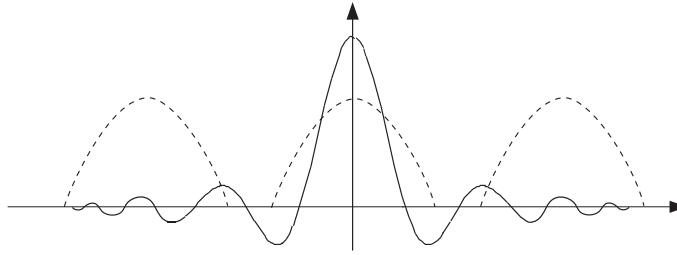


Figura 2.24: Altas frequências introduzidas na filtragem com o filtro retangular.

- Mudar o filtro de reconstrução;
- Aumentar a frequência de amostragem.

Com efeito, mudança do filtro de reconstrução consiste em se tomar um filtro cuja função de transferência tenha um maior decaimento no infinito de forma a minimizar a introdução de altas frequências. Por outro lado, o aumento da frequência de amostragem propicia uma maior separação entre os transladados do modelo espectral do sinal original, facilitando desse modo a tarefa do filtro em se livrar das altas frequências durante o processo de reconstrução.

Conforme vimos na seção 2.5.4, o aumento da frequência de amostragem é também uma das soluções recomendadas para minimizar o problema de aliasing.

Finalizando, lembramos ao leitor que pode haver problemas tanto no processo de amostragem, como na etapa de reconstrução. Desse modo, um sinal mesmo amostrado corretamente, pode sofrer um processo de reconstrução não adequado.

2.8 Filtros Polinomiais

O filtro retangular, ou *filtro box*, utilizado na seção anterior, é o mais simples de uma família de filtros chamada de *filtros polinomiais*, cujos núcleos são definidos por expressões polinomiais. Os outros filtros dessa família são obtidos a partir do filtro retangular através de um processo de convolução sucessiva

Retangular * Retangular = Linear;
 Retangular * Linear = Spline Quadrático;
 Retangular * (Spline Quadrático) = Spline Cúbico.
 ⋮

Segue do resultado acima que a função de transferência dos filtros polinomiais é obtida fazendo o produto sucessivo da função sinc, que é a função de transferência do filtro retangular, conforme indicamos na tabela abaixo:

Função de Transferência	Filtro
$\text{sinc}(s)$	Retangular
$\text{sinc}^2(s)$	Linear
$\text{sinc}^3(s)$	Quadrático
$\text{sinc}^4(s)$	Cúbico
⋮	⋮

Dessa tabela segue-se que os filtros polinomiais de grau mais alto possuem um decaimento mais rápido, de forma que as frequências mais altas são filtradas no processo de reconstrução. Desse modo, na seleção de um dos filtros de reconstrução acima devemos decidir entre um processo de interpolação mais suave, sem altas frequências, ou um filtro que propicie a ocorrência de altas frequências no sinal reconstruído.

O *filtro linear* faz uma interpolação linear entre as amostras e portanto o sinal reconstruído é contínuo. Mostramos esse método de reconstrução na Figura 2.25.

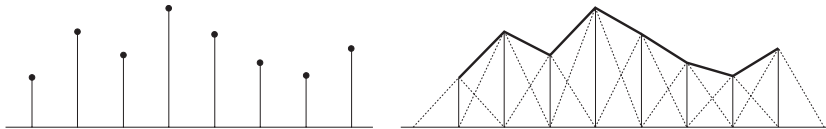


Figura 2.25: Reconstrução com filtro linear.

Os sinais reconstruídos com filtros polinomiais de ordem mais alta possuem um maior grau de diferenciabilidade, conforme ilustramos na Figura 2.26.

Observamos que os filtros na família polinomial buscam um processo de reconstrução que de alguma forma aproxima a reconstrução com o filtro



Figura 2.26: Reconstrução com filtro polinomial de ordem superior.

ideal dada pela função sinc da base de Shannon. Na realidade pelo teorema central do limite, o processo de convolução sucessiva acima converge para a função gaussiana, $g(t)$, definida por

$$g(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}},$$

cuja função de transferência é também dada por uma exponencial, ou seja, seja $F(g)(s) = e^{-2\sigma^2\pi^2s^2}$.

Na Figura 2.27 mostramos os filtros de reconstrução discutidos acima, juntamente com a função de transferência de cada filtro. Incluímos também o filtro de reconstrução ideal.

Conforme vimos anteriormente, mesmo quando temos a amostragem pontual de um sinal obedecendo ao limite de Nyquist, temos o erro de truncamento no processo de reconstrução ideal. Isso ocorre porque o filtro sinc não possui suporte compacto.

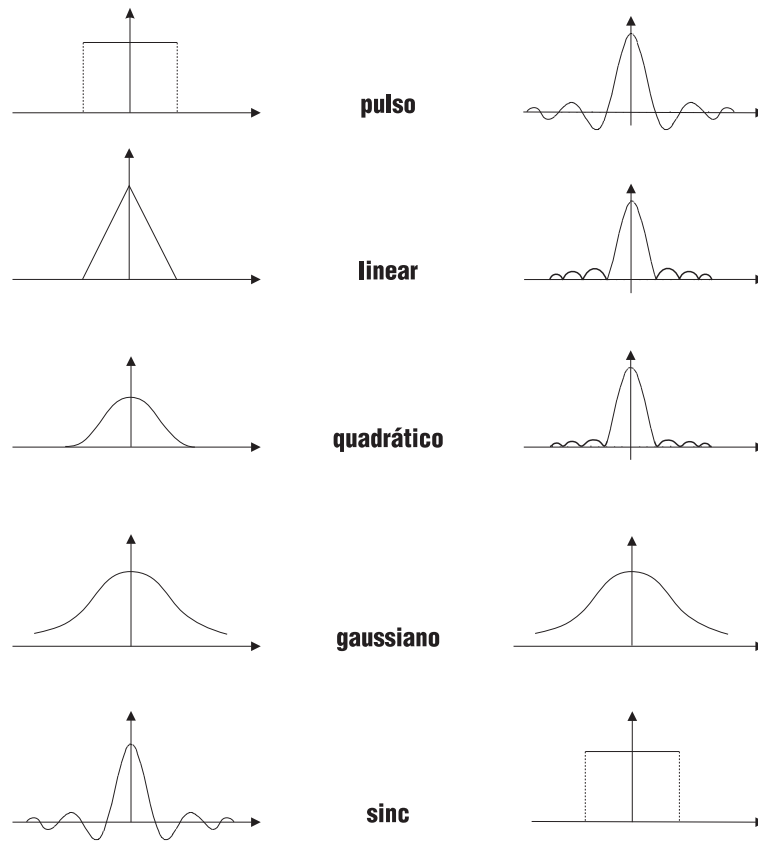


Figura 2.27: Núcleos de reconstrução e funções de transferência.

Capítulo 3

Ritmo e Percussão

Adotando uma visão ingênua, poderíamos dizer que fazer música significa distribuir sons no tempo. Os fenômenos que emergem desta distribuição temporal podem ser vistos como fenômenos rítmicos.

No entanto, já se sabe que não apenas os parâmetros temporais são fundamentais para o entendimento do conceito de ritmo. De fato, a experiência musical deve ser vista como um todo, e a tentativa de isolamento de alguma de suas partes pode em risco a sua compreensão completa, como atestam os princípios da *Gestalt*.

Paul Fraisse, um dos mais importantes estudiosos da psicologia da música, declara categoricamente que não existe uma definição precisa e universalmente aceita sobre o conceito ([38]). A origem do termo vem do grego *rhythmos* e significa medida de um fluxo ou movimento, ou ainda uma simetria [46].

Primeiramente vamos investigar diversas abordagens sobre o assunto no intuito de criar uma base conceitual que inclua o ritmo em seu contexto real, explicitando suas dependências culturais e sua interrelação com outros aspectos musicais e cognitivos. As referências [39, 48, 61] foram decisivas nesta investigação.

A partir desta análise, desenvolveremos o ferramental matemático e computacional para lidar com o sinal sonoro proveniente de um instrumento de percussão.

Fechando o capítulo, apresentaremos algumas questões a respeito de sistemas interativos musicais e uma solução para se fazer música com o computador controlando-o através da informação rítmica de instrumentos de percussão, utilizando-se o ferramental desenvolvido no segundo capítulo.

Antes de seguirmos no nosso discurso, porém, devemos ressaltar um

importante aspecto. Lidar com música significa lidar com cultura. Muitos textos científicos que desenvolvem ferramentas teóricas ou práticas para se lidar com música se pretendem universais e com isso não expõem seus pressupostos culturais, como se eles não existissem. Evidentemente que a teorização de características de mais baixo nível, como, e.g., a representação digital do sinal sonoro, são aplicáveis a quase todos os tipos de música, mas qualquer tipo de análise de mais alto nível, e.g. a extração de características harmônicas ou rítmicas, exige algum tipo de modelo que carrega implícita ou explicitamente algum pressuposto cultural.

Por outro lado, definir exatamente qual estilo musical será abordado pode ser uma tarefa difícil e limitante. No nosso caso, é suficiente dizer que estamos interessados em incluir determinados estilos de música popular que possuem o ostinato rítmico (repetição periódica de uma linha rítmica) como um de seus elementos. Podemos citar diversos exemplos tais como samba, choro, jazz, rumba, maracatu e funk, todos eles oriundos do encontro entre as culturas européia e africana que se deu no continente americano nos últimos cinco séculos.

3.1 Abordagens

Para começar a nossa investigação a respeito de ritmo dividiremos as abordagens em quatro grupos:

- Teórica: baseada em pressupostos da música de concerto da Europa ocidental do período do século XVII ao XIX (que chamaremos de *musica clássica* no presente trabalho), cria uma teorização útil para a análise de peças desse período.
- Experimental: utilizando-se de resultados de laboratório, pretende criar modelos de cognição do ritmo aplicáveis a situações musicais reais.
- Computacional: através de métodos de análise, representação e síntese do sinal sonoro, busca soluções implementáveis para atividades relacionadas a música e pode inspirar modelos cognitivos.
- Etnomusicológica: tenta criar modelos úteis para a compreensão de fenômenos musicais provenientes de diversas culturas.

3.1.1 Abordagem Teórica

Como colocado anteriormente, esta abordagem é fruto de alguns postulados desenvolvidos a partir da intuição dos seus autores, sem embasamento

experimental [39], visando modelar a música Européia do período já citado.

Apesar de pretenderem tratar de fenômenos intrínsecos a música na vida real, essas análises são todas baseadas na notação musical clássica (o pentagrama), pressupondo que esta representação dê conta de todos aqueles fenômenos, fato que não é consenso nem entre intérpretes daquele estilo musical.

Por outro lado, é também importante notar que muitas questões levantadas por essa abordagem de fato procedem e são aplicáveis, com algumas modificações, a diversos estilos musicais que não pertencem ao do enfoque inicial mas que possuem influências diretas, ou indiretas deste, e.g. os já citados samba, choro, jazz, rumba, e funk.

Portanto, nos é fundamental o entendimento de alguns pontos consensuais desta abordagem.

Iremos apresentar alguns conceitos provenientes de dois trabalhos que são considerados referências no assunto: “The rhythmic structure of music” de Grovesnor Cooper & Leonard Meyer ([43]) e “A Generative Theory of Tonal Music” de Fred Lerdahl & Ray Jackendoff ([42]).

Ambos aplicam conceitos de teorias linguísticas a análise musical. O primeiro utiliza-se de figuras da prosódia, estudo do ritmo, entonação e demais atributos da fala, para segmentar a informação sonora musical em grupos, que, quando organizados temporalmente, geram fenômenos rítmicos. Já o segundo trabalho utiliza a teoria de Gramática Gerativa que estuda ‘o que um ser humano sabe quando sabe falar uma língua’. Em outras palavras discute como uma pessoa pode gerar diversas sentenças, novas e por ela desconhecidas, em uma determinada língua. A idéia básica é de criar regras gramaticais as quais essas sentenças estão submetidas e com isso modelar a geração de novas expressões. O que Lerdahl e Jackendoff fizeram foi aplicar isto à música, gerando *regras de boa-formação* as quais as sentenças musicais estariam submetidas.

Desde já podemos perceber algumas discrepâncias nestas abordagens. A estrutura da linguagem falada certamente difere da estrutura da música, já que esta última pode gerar fenômenos tais como a dança, aspectos ritualísticos e emocionais, como e.g. o *groove*, fato que não possui paralelo nos modelos linguísticos ([49]).

Além disso, percebemos que a abordagem através da prosódia ainda guarda algum cuidado com relação as questões intrinsecamente sonoras da música, como a entonação, o que não ocorre na abordagem através de regras gramaticais, onde a informação musical é considerada puramente simbólica.

Como dissemos anteriormente essa abordagem é baseada na notação musical tradicional.

Definimos sucintamente alguns conceitos fundamentais:

- **Grupamento:** os princípios da Gestalt definem que a percepção humana possui mecanismos para segmentar a informação em partes. Um ouvinte que escuta um sinal musical, é capaz de segmentá-lo em notas, que se agrupam em motivos, temas, seções e, finalmente, peças musicais completas. Esta ação é chamada de Grupamento.
- **Métrica:** A música clássica tem como uma característica marcante o fato de que algumas notas de suas frases musicais são mais acentuadas que outras. Se observarmos bem, percebemos que essa acentuação é, muitas vezes, periódica e que essas notas são quase sempre a primeira de cada compasso. Além disso a duração de cada nota é normalmente uma subdivisão em partes iguais do período de acentuação. O fato subdividirmos os compassos em partes iguais, gerando duração de notas que por sua vez podem ser subdivididas em partes iguais, e assim sucessivamente cria uma estrutura hierárquica de níveis de subdivisão. É a chamada estrutura métrica.
- **Pulso ou tactus:** Muitas vezes esse conceito tem uma definição informal, tal como a frequência em que batemos o pé ao ouvirmos música, ou a frequência com que o maestro move a batuta. Como o próprio nome pulso indica, é uma série de estímulos isocrônicos (apesar de existirem abordagens que lidam com pulsos não isocrônicos) que preferencialmente tem um período próximo de 600 milissegundos, como apresentaremos adiante. Em relação a estrutura métrica, seria um nível especial de referencia, onde os eventos mais frequentes que o pulso seriam vistos como subdivisões deste e os menos frequentes, como grupamentos.

Como podemos perceber, ambas as estruturas de Grupamento e Métrica são hierárquicas e sugerem uma interpretação utilizando o conceito de multiresolução para ritmo. De fato, esta idéia foi explorada em [54] onde Smith desenvolve uma representação de ritmo utilizando wavelets de Morlet e Grossmann cujo núcleo da wavelet mãe é

$$g(t) = e^{-t^2/2} e^{i\omega_0 t}.$$

É interessante notar que ele consegue obter algumas interpretações através da congruência de fase que condizem com as expectativas das abordagens teórica e experimental.

Uma das características que fazem da notação musical algo distante da experiência musical real é o que podemos chamar de *timing* ou *expressividade*. Quando uma intérprete lê uma partitura, e por um processo de

cognição complexo, gera música através da interação de seus movimentos e de seu instrumento, esta música possui características que dependem das possibilidades motoras dessa intérprete, e das influências culturais que modelaram aquele processo cognitivo.

Tanto as imprecisões de movimento do braço de um percussionista, quanto o ritardando *ritardando* de um intérprete clássico (uma pequena desaceleração do fluxo musical que não é quantificada na notação), são casos de imposições motoras ou culturais que modificam a informação escrita na partitura e tornam a música mais humana. Nessa abordagem, portanto, a expressividade é considerada um desvio da notação, um erro, uma discrepância, apesar de muitos autores perceberem a importância central que ela tem, por exemplo, na criação de modelos de síntese mais realísticos [55].

Por final devemos notar que a abordagem teórica encara o fenômeno rítmico como apresentamos: Grupamento/Métrica como questões centrais e Expressividade aparecendo como um desvio da notação. Grande maioria dos trabalhos que lidam com este fenômeno, mesmo em outras áreas tais como a Psicologia da Música e a Computação Musical, usam esta estrutura como algo dado *a priori* sobre o qual se baseiam os seus discursos.

Como exemplo, podemos citar o trabalho teórico em psicologia da música [39] que utiliza essa estrutura como base narrativa, e o trabalho [56] que desenvolve um sistema interativo para improvisação, fundamentado nas idéias de Lerdahl e Jackendoff.

3.1.2 Abordagem Experimental

Diferentemente da abordagem teórica, a experimental pretende extrair de dados empíricos a afirmação de modelos cognitivos. Como a análise metódica de experiências com o som é uma tarefa difícil, os modelos a serem testados costumam focar em pontos específicos da cognição o que faz a literatura da área assumir um aspecto fragmentário.

Algumas sínteses foram bem sucedidas e permitiram um passo adiante na compreensão geral de ritmo. O trabalho que parece ser referência fundamental é o de Paul Fraise, que foi amplamente publicado [39].

Considerando a síntese feita em [39], podemos ressaltar alguns resultados de Fraise. Ele fez a distinção entre a *percepção do tempo*, que lida com fenômenos de duração de até 5 segundos aproximadamente; e a *estimação do tempo*, que lida com fenômenos de durações maiores e que são atribuídos a reconstrução de informações guardadas na memória. Além disso ele observou a importância do corpo na cognição do ritmo, a tendência a categorizar uma sequência de intervalos como tendo a relação de duração 2:1 e uma série de outros fenômenos importantes. Finalmente devemos ressaltar o es-

tudo sobre a importância do valor de 600 milissegundos, chamado de “valor de indiferença”, um limiar que separaria qualitativamente o conjunto de segmentos temporais e que serviria como valor preferencial para a marcação física do fluxo musical, como ressaltado na definição de pulso anteriormente.

Podemos citar, também, o trabalho de Richard Parncutt [50] que enfoca de maneira metódica a questão de encontrar o pulso em uma sequência rítmica periódica de estímulos. Mais uma vez o valor de 700 milissegundos foi ressaltado como preferencial (em acordo com os 600 milissegundos de Fraisse).

Em um de seus experimentos, no entanto, Parncutt se questiona a respeito de uma interpretação inesperada que muitos sujeitos deram de um determinado padrão rítmico, e [61] coloca que provavelmente essa interpretação foi influenciada pela difusão universal do rock'n'roll, fato que não é levado em conta na interpretação de Parncutt.

Neste sentido, o que Pfeleiderer está questionando é se seria possível extrairmos princípios universais da cognição sem levarmos em conta aspectos culturais. Ele cita, então, o trabalho pioneiro de Alf Gabrielsson que utiliza inclusive gravações de música popular em seus experimentos.

Em [41], o autor faz um apanhado de alguns resultados experimentais a respeito de ritmo. Sua motivação é de querer investigar a experiência musical para além do que ele chama de *aspectos estruturais*, aqueles relacionados com a notação. Ele chega a citar [43] como exemplo de trabalho que aborda apenas os aspectos estruturais.

Gabrielsson parte do pressuposto que a experiência musical é multidimensional. Tentando chegar a um acordo sobre quantas e quais dimensões seriam importantes para caracterizar essa experiência, ele utiliza técnicas de análise multivariada. Apesar de não explicitar os cálculos aplicados aos dados de suas experiências (ele o faz em outros trabalhos), o autor relata ter chegado a um espaço de 15 dimensões onde cada ciclo rítmico é visto como um ponto (esses ciclos são gerados não só com sons sintéticos, mas também tocados por um percussionista ou então extraídos de gravações de pedaços de diversos estilos de música popular).

Ele afirma que cada uma dessas dimensões podem ser vistas como adjetivos para caracterizar esses ciclos e que esses adjetivos podem ser facilmente agrupados em três tipos: estrutural, mocional (no original *motional*), e emocional. Os adjetivos estruturais dão conta de questões relacionadas a notação musical, como já foi dito; os mocionais estão relacionados a características de movimento associadas aos ciclos, e.g. caminhante, dançante, pulante, suingante etc; e os emocionais são, e.g. solene, calmo, rígido, vital, brincalhão etc.

Com esse trabalho, Gabrielsson prova ser possível lidar com modelos

de cognição do ritmo de maneira mais profunda ao incluir aspectos que normalmente são considerados supérfluos por serem subjetivos.

3.1.3 Abordagem Computacional

Muitos dos trabalhos que abordam a questão do ritmo computacionalmente tem por objetivo final resultados práticos tais como a catalogação automática de estilos musicais, a extração de informação de alto nível ou a criação de sistemas interativos em tempo real.

Alguns desses trabalhos se relacionam com as abordagens já apresentadas. Alguns servem para afirmar ou negar modelos experimentais e outros se baseiam ou questionam modelos teóricos.

Iremos elencar alguns exemplos de trabalhos computacionais voltados para o ritmo e que nos serão úteis ao apresentarmos o nosso ferramental para se lidar com o sinal percussivo monofônico.

Primeiramente iremos apresentar o importante trabalho [57] extensamente difundido na literatura sobre ritmo. Scheirer apresenta um método para a detecção do pulso em um sinal polifônico.

Neste trabalho, o autor demonstra interesse em abranger diversos estilos musicais, o que parece ser atualmente uma tônica na abordagem computacional, contrastando com a abordagem teórica e a maioria das abordagens empíricas.

Scheirer parte do pressuposto, que passou a ser largamente aceito, de que o conteúdo rítmico de um sinal polifônico pode ser integralmente representado pelos envelopes de amplitude de cada banda de frequência deste sinal. Ele propõe, portanto, que o sinal seja dividido em bandas (seis, no caso) cobrindo todo o espectro e que seja extraída a derivada de cada envelope de amplitude. Essa derivada alimenta um banco de filtros pente (*comb filter* em inglês) que servem como ressoadores. O período do filtro pente cujo sinal de saída tem mais energia passa a ser um bom candidato ao período do pulso deste sinal.

O autor pondera que os sinais analisados por este algoritmo são mais complexos do que os analisados em outros trabalhos da época, e portanto a comparação de eficiência quantitativa não cabe. Ao comentar sobre a ineficiência de seu algoritmo em detectar o pulso em alguns casos por ele denominados “up-tempo jazz”, Scheirer observa que os humanos têm a capacidade de “induzir” o pulso de sinais complexos mesmo que não haja acentuações periódicas explícitas.

Nesse mesmo sentido, Tristan Jehan, utilizando este mesmo algoritmo, encontrou dificuldades em determinar o pulso de um sinal polifônico produzido por um grupo de maracatú ([26]). Ao questionar seus colegas, perce-

beu que eles também tinham dificuldades em cumprir esta tarefa, já que eles não eram familiarizados com a complexa célula rítmica deste estilo.

Fica evidente que o fenômeno do pulso não é objetivo e que os aspectos subjetivos ligados a cultura devem ser incluídos na análise, como o faz o próprio [26].

Outro trabalho importante a respeito do ritmo é o de [62] onde Paulus & Klapuri apresentam uma medida de similaridade entre padrões rítmicos.

Primeiramente, eles desenvolvem uma maneira de segmentar um padrão rítmico calculando a periodicidade dos envelopes de amplitude de [57] sem utilizar o banco de filtros pente. A representação do sinal é feita utilizando-se três funções características, a energia, cujo correspondente perceptual seria o volume do sinal, o centróide espectral que corresponderia a percepção de claridade do som (*brightness* em inglês) e alguns coeficientes da MFCC.

Desta forma, dois padrões rítmicos quaisquer são segmentados, representados usando estas funções e comparados utilizando-se a técnica de warping temporal.

O método foi testado com sucesso em frases de bateria eletrônica tocadas por um músico amador e em pedaços de música real. Os autores chegam a conclusão que o centróide espectral é que fornece a melhor representação rítmica para esse cálculo.

Considerando que o conceito de recorrência é fundamental para a música, Foote & Cooper propuseram em [31] um importante método de visualização das auto-similaridades de um sinal de áudio polifônico.

Os autores constroem uma matriz cujas colunas e linhas representam os segmentos sequenciais do sinal musical. Cada segmento é comparado com todos os outros segmentos da peça utilizando-se uma representação espectral (eles citam coeficientes de Fourier ou da MFCC como sendo eficazes para a tarefa).

Desta forma eles obtêm uma visualização onde se pode perceber repetições rítmicas e de forma, explicitando a característica de multiresolução do sinal musical já citada anteriormente.

Além disso, os autores testam este método para peças de música clássica e jazz, obtendo resultados eficazes na detecção do tactus.

Como se percebe, a detecção do pulso se faz extensamente presente na literatura a respeito de ritmo. Outro campo extenso e importante é o da classificação de sons percussivos. Muitas abordagens podem ser encontradas, algumas visando problemas específicos como a transcrição de sons de tabla, [63], e outras apresentando métodos de representação e classificação tais como [64, 65, 58]. Chamamos a atenção para a técnica apresentada em [66] onde a programação genética aplicada a construção de funções características gera melhores resultados na classificação de sons de pandeiro do que

aqueles obtidos utilizando funções usuais.

3.1.4 Abordagem Etnomusicológica

Pode parecer estranho apresentarmos esta abordagem já que a etnomusicologia não está interessada em resultados computacionais práticos ou teorizações universalizantes como são os casos precedentes. No entanto, decidimos apresentar nesta seção algumas considerações sobre o ritmo que surge da análise de estilos musicais que tem o ostinato rítmico como característica importante. Estes estilos diferem da música clássica e são estudados, em sua maioria, por etnomusicólogos, mas acreditamos que os resultados por eles obtidos são de essencial importância já que o nosso interesse é em incluir estilos com essa característica.

Primeiramente, chamamos a atenção para a formalização feita por Kofi Agawu no livro “Representing African Music” ([40]). Um dos objetivos do autor, é de destruir o mito de que a música africana é interessante apenas pelas suas características rítmicas. Neste sentido, ele mostra que, até então, as análises musicais de gêneros africanos traz sempre algum tipo de exotismo que inviabiliza uma possível comparação com gêneros provenientes de outras etnias. Indo contra este fato, Agawu apresenta uma análise de estilos de dança/música da África do Oeste e Central utilizando-se da notação musical usual (ao invés de notações mirabolantes e incompreensíveis, como fizeram outros autores).

Uma característica central desta análise é que ela é focada no par dança-música, e as considerações tecidas sobre a última levam sempre em conta questões relacionadas ao corpo da dançarina ou dançarino como sendo fundamentais para compreensão completa do assunto. Para nós será de suma importância o conceito de *topos* utilizado por Agawu em seu trabalho.

Em suas palavras, *topos* é “uma figura rítmica distinta, de fácil memorização, e de duração modesta (por volta de um compasso ou ciclo), usualmente tocada no conjunto por um ‘cow-bell’ ou instrumento de som agudo e que serve de ponto de referência temporal”. Como ele coloca, e o próprio significado da palavra em latim induz, o *topos* fornece a referência para os músicos e dançarinos. Não é só nos estilos africanos que observamos a ocorrência desse elemento. A clave da rumba cubana e o surdo da escola de samba brasileira são exemplos de *topoi* em seus respectivos contextos.

Sincronização entre os músicos e entre os dançarinos é de fato um papel fundamental que o ostinato rítmico assume nessas culturas, mas podemos ir além. Podemos considerar que o *topos* não é explicitamente tocado por um instrumento. Por exemplo, a contagem ‘UM, dois, três, quatro, UM, dois, ...’ onde o um é acentuado para se marcar o início de cada compasso 4/4

pode ser vista como um *topos* implícito da música clássica. Uma flautista que lê sua partitura, sabe que deve acentuar a nota que estiver na posição do 'UM', o que cria um tipo de ostinato rítmico subjetivo (subjetivo pois não necessariamente existirá uma nota na posição 'UM' ou não necessariamente todas as notas acentuadas estarão nesta posição). Se entendermos o conceito de *topos* desta forma geral, obtemos que tanto os ostinatos rítmicos, quanto a métrica tradicional são casos particulares desta idéia de referência temporal. Esta generalização não é feita por Kofi Agawu e será adotada no presente trabalho.

Outra questão central que percebemos tanto em trabalhos como o de Agawu quanto o de Paul Fraisse ou até mesmo o de Lerdahl & Jackendoff (ver definição de pulso) é a importância do corpo na compreensão de ritmo. Neste sentido, chamamos a atenção para o trabalho [49] onde Vijay Iyer utiliza o ferramental da Cognição Encorporada (*embodied cognition* em inglês) para dar conta de fenômenos rítmicos. Ele argumenta que todas as experiências com o ritmo musical que um ser humano tenha em sua vida se relacionam de alguma forma com ritmos de funções corporais tais como a respiração, o pulso cardíaco, o andar, o falar, etc.

O interesse de Iyer é de compreender o que ele chama de Micro-timing, pequenos desvios temporais que surgem durante a geração de ciclos rítmicos por um músico profissional. A conclusão que ele chega é de que essas variações só podem ser concebidas através da Cognição Encorporada e são muito importantes como referência temporal, como é o caso do 'swing' no jazz.

Finalmente citamos um modelo de ritmo que se encaixa perfeitamente com a generalização de *topos* apresentada aqui. [51] propõe um modelo de ritmo musical baseado em quatro proposições:

- Ritmo concerne eventos musicais que ocorrem regularmente.
- Existe diferenciação entre eventos rítmicos.
- Eventos rítmicos são cíclicos.
- Existe um forte sentido de incorporação (*embodiment* no original) associado ao ritmo musical.

Neste trabalho o autor não desenvolve uma análise rigorosa de exemplos e argumenta que não é possível lidar com um conceito tão amplo e subjetivo de maneira puramente algorítmica.

Baseado em todas essas referências, definimos que um *topos* de um sinal musical é um *padrão recorrente* deste sinal. Tanto o termo 'padrão' quanto 'recorrente' são propositalmente subjetivos. Definimos que uma *instância do*

topos é um *padrão recorrente* de alguma característica computacionalmente calculável do sinal musical. Por exemplo, a acentuação periódica do 'um' ou a presença da clave são exemplos de instâncias dos respectivos *topoi*. Finalmente, ao ouvir uma *instância do topos* a pessoa familiarizada com o respectivo *topos*, experimenta algum tipo de incorporação. A possível resposta corporal normalmente possui também algum tipo de periodicidade, e esse período costuma ser o mesmo do *topos* em questão.

Utilizando estas definições, podemos perceber que, dado uma *instância de um topos* e o conhecimento cultural para se interpretar corporalmente o mesmo, é possível representar ostinatos rítmicos, como foi feito em [26]. Veremos a seguir como esses conceitos serão úteis para o desenvolvimento de sistemas interativos musicais,

Com relação a representações cíclicas, citamos ainda [67] como um trabalho que vai ao encontro das questões discutidas nesta seção.

3.2 Análise do Sinal Percussivo

Desenvolveremos nesta seção o ferramental para se lidar com o sinal percussivo no intuito de fornecer peças para a construção de sistemas interativos musicais. Neste sentido, o objetivo desta análise é ser simples, eficaz e computacionalmente leve.

Iremos nos basear nos trabalhos computacionais discutidos anteriormente mas, na prática, utilizaremos versões simplificadas daquelas abordagens, e que serão suficientes para as aplicações apresentadas na última seção deste capítulo.

Primeiramente consideramos que o sinal sonoro a ser analisado é monofônico (proveniente de apenas um instrumento), captado por um microfone e digitalizado a uma frequência de 44KHz, como foi discutido anteriormente. Além disso, consideramos que a informação digital é fornecida para a análise em grupos de 64 amostras que notaremos $b(i)$, $i \in \mathbb{N}$.

Como vimos anteriormente, [57] argumenta que o conteúdo rítmico de um sinal de áudio pode ser integralmente representado pelos envelopes de amplitude de cada banda de frequência deste sinal. No caso do sinal monofônico percussivo, o início de cada evento musical é quase sempre marcado por um pico súbito no envelope de amplitude do sinal como um todo, e não apenas em uma determinada banda. De fato, a separação em bandas se mostrou desnecessária e optamos por calcular a amplitude média sobre todo o espectro da onda sonora monofônica.

Chamamos esses picos de *ataques* e devemos agrupá-los em *frases rítmicas*.

3.2.1 Frases Rítmicas

Formalizamos a idéia de extrair informação de um sinal sonoro na seguinte definição:

O signo r é dito *extraído automaticamente de um sinal sonoro* se existe um algoritmo que calcula um valor característico f_r a partir do sinal digitalizado e fornece, como saída, aquele signo se esse valor estiver acima de um determinado limiar t_r . Chamamos de ε_r a probabilidade deste algoritmo extrair corretamente o signo r do sinal sonoro.

No caso de sistemas interativos em tempo real estamos interessados em signos que são extraídos periodicamente do sinal. Como dissemos anteriormente, trabalharemos com o sinal dividido em grupos de amostras e tipicamente, o valor característico f_r será calculado como uma média sobre cada um desses grupos.

Um exemplo de signo é a detecção de ataques no sinal percussivo. A cada grupo de 64 amostras, calcula-se a média do valor absoluto da amplitude da onda. Isto nos fornece uma curva $f_r(i)$ da energia sonora do sinal, indexada pelo mesmo índice i dos grupos $b(i)$. O algoritmo que extrai ataques procura por picos da derivada desta curva que sejam maiores que um determinado limiar t_r . Sempre que esses picos são encontrados, o algoritmo fornece o signo r . Este método é uma simplificação do utilizado em [26] onde é feita uma filtragem sobre a curva f_r e sua derivada, para garantir coerência com algumas características psicoacústicas da percepção do sinal sonoro.

Além disso, precisamos de ferramental para lidar com sequências de ataques detectados durante a experiência interativa. Como o tamanho desta sequência não pode ser definido a priori, precisamos de sequências de tamanho variável.

A *sequência de entrada* s é definida como a função

$$s : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}^+ \cup \{-1\}$$

tal que, dado o ataque atual $c \in \mathbb{N}$,

1. $s(i, c) < s(j, c)$, $\forall i < j \leq c \in \mathbb{N}$
2. if $i \leq c \implies s(i, c) \in \mathbb{R}^+$ e se $i > c \implies s(i, c) = -1$
3. $\forall k \leq c$, $s(i, c) = s(i, k)$, $\forall i \leq k$

Denotamos $s_c(i) = s(i, c)$.

Dizemos que a sequência s_c é a sequência de entrada definida até o ataque atual c e as posições $s_c(i) = -1, i > c$ são ditas indefinidas. Tipicamente, $s_c(i), i \leq c$ é dado em milissegundos.

Nosso interesse está na informação local contida em parte do vetor s_c .

A frase rítmica (ou apenas a frase) P_k de tamanho k começando na posição j da sequência de entrada s é definida como o vetor

$$P_k = (s_c(j), s_c(j+1), \dots, s_c(j+k-1)).$$

Poderíamos deixar claro na notação, a sequência de entrada s , a posição j onde começa a frase e o índice i desta forma $P_k(j, s)(i) = P_k(i)$ mas tentamos evitar isto para não sobrecarregar a notação. Quando possível, omitimos também o k e escrevemos apenas P .

É importante perceber que $P_k \in (\mathbb{R}^+)^k$. Algumas vezes trataremos de frases rítmicas que não foram encontradas na sequência de entrada e portanto serão escritas na forma $P_k = (0, a_1, a_2, \dots, a_{k-1})$ onde os a_i 's são os instantes onde ocorreram os ataques em milissegundos a partir do instante 0 (portanto $a_i < a_{i+1}$).

Como vimos no trabalho de [62], é possível comparar dois padrões rítmicos inclusive com tamanhos diferentes. No nosso caso já adotamos a simplificação de representarmos as frases rítmicas como uma sequência de ataques, e definiremos, então, a distância entre duas frases da seguinte forma:

A distância $d : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}^+$ é definida como

$$\left| 1 - \frac{l(P_k^1)}{l(P_k^2)} \right| + \sum_{i \in \{1, \dots, k\}} \left| \left| \frac{P_k^1(i) - P_k^1(i-1)}{l(P_k^1)} \right| - \left| \frac{P_k^2(i) - P_k^2(i-1)}{l(P_k^2)} \right| \right|$$

$\forall P_k^1, P_k^2 \in \wp_k$ onde $l : \mathbb{R}^k \rightarrow \mathbb{R}^+$ é o comprimento

$$l(P_k) = |P_k(k) - P_k(0)|$$

da frase P_k .

Apesar de nossa representação ser mais simplificada do que a de [62], se considerarmos apenas o segundo termo de d , estaremos calculando a distância normalizada entre as frases, o que pode ser visto como uma simplificação do warping temporal dinâmico apresentado naquele trabalho.

Usando esta distância, poderemos extrair dois signos importantes do sinal percussivo. O primeiro é a detecção de frases rítmicas pré-determinadas e o segundo é a detecção da repetição de frases.

Para detectarmos uma determinada frase rítmica $P_k^1 = (0, a_1, \dots, a_{k-1})$, devemos apenas calcular, a cada detecção de um novo ataque c , a distância

entre P_k^1 e a frase $P_k^2(c - k + 1, s) = (s_c(c - k + 1), s_c(c - k + 2), \dots, s_c(c))$, ou seja, a frase formada pelos últimos k ataques da sequência de entrada. Quando esta distância se torna suficientemente pequena, obtemos o signo de detecção da frase P_k^1 .

A detecção de repetição de frase pode ser vista como uma simplificação da auto-similaridade de [31] que vimos anteriormente. Neste caso, devemos calcular a distância entre a frase $P_k^1(c - k + 1, s) = (s_c(c - k + 1), s_c(c - k + 2), \dots, s_c(c))$ e a frase $P_k^2(c - 2k + 1, s) = (s_c(c - 2k + 1), s_c(c - k + 2), \dots, s_c(c - k))$. Ou seja, estamos calculando se a frase formada pelos ataques $c - 2k + 1$ ao $c - k$ está sendo repetida nos k últimos ataques da sequência de entrada (do $c - k + 1$ ao c). Podemos fazer este cálculo, a cada ataque, para alguns valores de k , obtendo assim um análogo da matriz de auto-similaridade de [31]. Quando a distância entre P_k^1 e P_k^2 , para algum desses ks é suficientemente pequena, detectamos a repetição desta frase.

Além da relação entre os instantes de cada ataque (conhecida na literatura por IOI, *inter onset intervals*), também poderíamos pesquisar quais as características timbrais dos sons que se seguem a eles, tanto de uma forma classificatória, e.g. [66], quanto utilizando funções características como em [62]. No entanto, mostraremos no próximo capítulo que este ferramental simples, já é suficiente para se obter importantes resultados de interação musical.

3.3 Resultados Musicais

Alguns dos primeiros sistemas musicais interativos em tempo real datam da década de oitenta, tais como o “Voyager” de George Lewis [52] e o “Cypher” de Robert Rowe [44]. Ambos utilizam o protocolo MIDI (ou um interpretador audio-MIDI) como entrada e saída. No entanto, com o desenvolvimento computacional já é possível construir sistemas que lidem diretamente com o sinal sonoro, o que permite a utilização de instrumentos de percussão como interface de entrada de informação para interação.

[44] define dois paradigmas a serem utilizados na construção de sistemas interativos. O primeiro é o paradigma sistema-músico, ou seja, sistemas que tentam simular o comportamento de um músico experiente. Neste sentido, em [44] e [45], este mesmo autor se baseia na abordagem teórica já discutida e considera que para interação rítmica acontecer, a máquina deve ser capaz de encontrar o pulso (beat-track) do sinal sonoro de entrada. Este trabalho parece ser o ponto de partida de muitos outros pesquisadores de sistemas interativos. Alguns como [56] e [59] usam uma representação transcritiva, seguindo ainda mais a risca a abordagem teórica. Outros tais como [60] e [26] desenvolvem sistemas que são baseados em características

da cognição da música e atacam a abordagem transcritiva, como propôs [57], mas ainda assumem que o primeiro passo é fazer a máquina encontrar o pulso automaticamente.

Por outro lado, outro paradigma possível é o de sistema-instrumento, ou seja, a idéia de que o sistema deve ser controlável como um instrumento musical e que desta interação o músico obtém resultados diferentes daqueles obtidos com instrumentos tradicionais. Podemos citar como referências dessa abordagem os trabalhos [68] e [69]. Devemos ainda considerar o “Continuator” de François Pachet ([53]), que aprende a partir da interação com o músico e o robô “Haile” de Gil Weinberg & Scott Driscoll ([70]), um dos poucos trabalhos voltados para percussão e que se utiliza do conceito de *modos de interação*.

Intuitivamente, um *modo de interação* pode ser visto como um jogo entre o músico e o computador, onde as regras são explícitas para ambos. Este conceito pode ser formalizado utilizando-se a Teoria de Autômatos. Por alto, definimos que um *modo de interação* é um autômato com um número definido de estados que por sua vez possui um número definido de ações. Quando um usuário aplica alguma dessas ações, o autômato muda de estado (podendo se manter no mesmo estado no caso de ações reflexivas) e gera uma resposta sonora.

No trabalho [71], Krakowski apresenta uma abordagem usando o paradigma sistema-instrumento e esta formalização dos *modos de interação*. O autor associa a cada ação de um determinado autômato, uma frase rítmica (definida na seção anterior). Um estudo de caso utilizando o pandeiro como instrumento-interface é apresentado. Como a detecção de frases é feita com suficiente precisão, o usuário pode ‘passear’ pelos estados de um *modo de interação*, e com isso criar um resultado musical formado pela resposta da máquina juntamente com as frases rítmicas do pandeiro. Outro resultado musical pode ser visto em [47].

O autor ainda deixa claro que a utilização de frases rítmicas como comandos, permite o usuário-músico a lidar com ciclos rítmicos de uma forma extremamente fácil e natural, fato que é de suma importância em uma performance voltada para música popular. Portanto, percebemos que esta abordagem está de acordo com o conceito de *topos* desenvolvido na primeira seção deste capítulo, e que este conceito gera soluções mais eficazes musicalmente e computacionalmente do que a abordagem métrico-tradicional que pressupõe o ‘beat-tracking’ como passo primordial.

Capítulo 4

Melodia e Harmonia

Dada a natureza artística da música, é impossível definir com precisão os conceitos de *melodia* e *harmonia* de um modo que não inclua outras características musicais, como *ritmo* e *percussão*. (Veja, por exemplo, a definição de *melodia* em [7].) É até mesmo difícil, dado um instrumento musical, classificá-lo como sendo harmônico ou percussivo¹.

De modo grosseiro, entretanto, pode-se afirmar que melodia e harmonia são características musicais atreladas à noção de *nota musical*². Assim, os métodos computacionais abordados neste capítulo tem forte relação o conceito físico correspondente: a frequência.

O capítulo é dividido em três partes: Análise, Composição e Síntese. Na primeira, discutimos algumas técnicas para extrair características tonais/timbrais de um arquivo de áudio. Na segunda seção tratamos da composição musical auxiliada por computador. Finalmente falamos um pouco sobre os algoritmos para síntese de som digital.

4.1 Análise

As técnicas de análise de áudio apresentadas nesta seção são ditas de *baixo nível*, no seguinte sentido: trabalham diretamente sobre a onda sonora em formato digital (não sobre representações simbólicas, do tipo MIDI por exemplo), e observam conceitos físicos do áudio, como *frequência* e *amplitude*. Uma análise de *alto nível* levaria mais em conta aspectos de Teoria Musical: escala, modo, acorde principal, etc.

¹Lenine, importante nome da Música Popular Brasileira, declarou, no Programa do Jô exibido em 17/10/2008, ser um “percussionista de violão”.

²Em inglês, *pitch*.

4.1.1 Descritores Harmônicos/Timbrais

Um *descriptor* é um vetor representando alguma propriedade (característica física) do áudio. Normalmente os descritores são obtidos de segmentos de tamanho constante (*janelas*) que se sobrepõem (Fig. 4.1). O tamanho da janela depende da característica em estudo, mas para sons amostrados a 44100 frames por segundo (fps) em geral vai de 512 a 4096 frames. A taxa de sobreposição³ também é variável. Ela determina a resolução temporal da característica extraída.

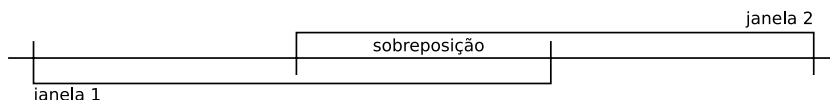


Figura 4.1: Duas janelas consecutivas com sobreposição.

Taxa de Eventos

É o mais simples dos descritores. Trata-se de olhar para o sinal de áudio e contar quantos eventos ocorreram numa certa unidade de tempo. Os principais eventos são a inversão de sinal em frames consecutivos⁴ e a anulação da derivada do sinal. Na seção 4.1.2 veremos uma aplicação.

Transformada de Fourier Discreta

Dado um sinal amostrado $x = (x_0, \dots, x_{N-1})$, sua transformada de Fourier Discreta (DFT), $\hat{x} = (\hat{x}_0, \dots, \hat{x}_{N-1})$, é dada pela fórmula $\hat{x}_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i kn/N}$. A DFT devolve a amplitude complexa com que as frequências entre zero e a metade da taxa de amostragem estão presentes no sinal. Por exemplo: se a taxa de amostragem é 44100 fps e o tamanho da janela é de 1024 frames, devido ao fato de a DFT de um sinal real ser simétrica em torno do ponto $(N - 1)/2$, o resultado serão 512 valores de frequência igualmente espaçados entre zero e 22050. Assim, um aumento no tamanho da janela não aumenta a gama de frequências avaliadas; aumenta a precisão da informação (resolução) sobre as frequências.

A maior parte dos descritores que veremos tem como base Transformada de Fourier (Discreta) com Janela, mas aqui *janela* significa mais do que o que foi dito anteriormente. Quando se aplica a DFT sobre um segmento de áudio, supõe-se que este contenha um período inteiro de um sinal periódico definido em toda a reta, o que não acontece na prática. Deste modo, se x_0 é muito diferente de x_{N-1} , a DFT vai encontrar amplitudes não nulas para altas frequências.

³Em inglês, *overlap*.

⁴O descriptor correspondente é chamado, em inglês, de *zero-crossing rate*.

Para contornar este problema multiplica-se o segmento x por uma função suave, com integral unitária, que se anula nos extremos x_0 e x_{N-1} (Fig. 4.2). Essa função também recebe o nome de *janela*.



Figura 4.2: Respectivamente: janela, sinal e produto entre ambos.

Existem vários tipos de janela. Uma das mais comuns é a de Hann (Fig. 4.2), dada por $h_n = \frac{1}{2} \left(1 - \cos \frac{2\pi n}{N-1}\right)$, para $n = 0, \dots, N-1$.

Pela fórmula da DFT, nota-se que seu custo computacional é quadrático no tamanho N do sinal x . Felizmente existem algoritmos mais eficientes, como o de *Cooley-Tukey*, que usa a estratégia de *divisão e conquista* para reduzir o custo computacional a $O(N \log N)$. Trata-se da Transformada Rápida de Fourier (FFT). Neste capítulo usaremos DFT e FFT indistintamente.

Espectro de Potências

A entrada \hat{x}_n da FFT carrega informação de amplitude e de fase da frequência correspondente ao índice n . Entretanto a informação de fase é quase sempre descartada, sendo usada somente a magnitude de \hat{x}_n , ou seu quadrado. O gráfico $\{(n, |\hat{x}_n|^2) : n = 0, \dots, N-1\}$ é conhecido como o *espectro de potências* (ou ainda, *espectrograma*, em analogia a histograma) do sinal x .

Amplitude⁵

Corresponde à média das entradas do espectro de potências, sendo uma medida da energia do sinal sonoro. Muitas vezes esse valor é apresentado em escala logarítmica, a fim de ser mais coerente com a intensidade sonora percebida pelo sistema auditivo humano.

Auto-Correlação

Dado um segmento de áudio $x = (x_0, \dots, x_{N-1})$, a auto-correlação c no ponto k é uma medida de quanto o sinal (x_0, \dots, x_{N-1-k}) é semelhante ao sinal (x_k, \dots, x_{N-1}) . Formalmente, $c_k = \sum_{n=0}^{N-1-k} x_n x_{n+k}$, para $k = 0, \dots, N-1$ (Fig. 4.3).

Note que o custo computacional da auto-correlação é quadrático em

⁵Loudness, em inglês.

N . Usando, porém, a FFT e teorema da *convolução circular*⁶ consegue-se calcular a auto-correlação em tempo $O(N \log N)$. O algoritmo é simples: defina $\bar{x} = (\bar{x}_0, \dots, \bar{x}_{2N-1})$ de modo que $\bar{x}_k = x_k$ para $k = 0, \dots, N-1$ e $\bar{x}_k = 0$ caso contrário. Aplique a FFT em \bar{x} , obtendo $\hat{\bar{x}}$. Aplique a FFT inversa⁷ sobre $(|\hat{\bar{x}}_0|^2, \dots, |\hat{\bar{x}}_{2N-1}|^2)$. A parte real da k -ésima entrada do resultado é justamente c_k .

Detalhes desse algoritmo podem ser encontrados em [8]. Na seção 4.1.2 apresentaremos uma aplicação deste conceito.

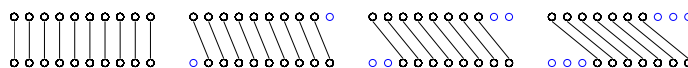


Figura 4.3: Ilustração da fórmula de auto-correlação.

Vetor de Croma

Este descritor, associado a métodos de Aprendizagem de Máquina, é bastante usado para detecção de acordes (Seção 4.1.3). Primeiro captura-se, no espectro de potências do sinal, as 84 amplitudes correspondentes às notas cujo código MIDI vai de 24 a 107 (o que cobre 7 oitavas de um piano, ou seja, quase todas as notas). Em seguida as amplitudes correspondentes às notas de mesmo nome (independentemente da oitava) são somadas, obtendo-se um vetor de 12 entradas⁸ (Fig. 4.4). Os elementos desse vetor são divididos pelo máximo, para que as diferenças de amplitude em diferentes janelas não sejam levadas em conta.

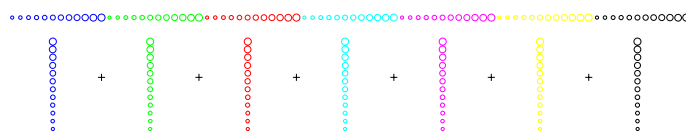


Figura 4.4: Ilustração de parte do cálculo do vetor de cromas.

Cepstrum

Alguns sons possuem espectro com energia concentrada em índices linearmente espaçados, múltiplos de um índice correspondente a uma certa frequência. Neste caso diz-se que o som tem muitos *harmônicos*. É o que

⁶A convolução entre dois sinais é igual à transformada de Fourier inversa dos produtos das transformadas de Fourier dos mesmos.

⁷A IDFT de um sinal (y_0, \dots, y_{N-1}) é dada por $\check{y}_n = \frac{1}{N} \sum_{k=0}^{N-1} y_k e^{2\pi i kn/N}$.

⁸Lembramos que em uma oitava há 12 notas.

ocorre, por exemplo, com a voz humana e com as cordas de um violão (Fig. 4.5).

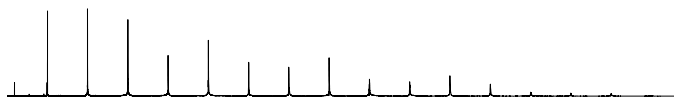


Figura 4.5: Magnitude do espectro do som da primeira corda do violão.

Assim sendo, os picos do espectro de potências ocorrem com uma certa periodicidade. A ideia do *cepstrum*⁹ é justamente capturar essa periodicidade. Formalmente, trata-se da magnitude da transformada de Fourier inversa do logaritmo da magnitude da transformada de Fourier do sinal.

4.1.2 Detecção de Frequência Fundamental

A noção de nota musical está intimamente ligada à de Frequência Fundamental (também conhecida por F_0). Quando um dispositivo digital renderiza uma onda senoidal com frequência de, digamos, 440 Hz, nosso ouvido percebe uma nota musical que convencionou-se chamar de A4 (código MIDI para a nota Lá central do piano). Entretanto, quando se percute, ao piano, a mesma nota, ou quando se toca a nota da quinta casa da primeira corda do violão, o espectro de potências do áudio revela não apenas um pico na frequência de 440 Hz, mas também picos em frequências correspondentes aos múltiplos de 440. Todas essas frequências (os chamados harmônicos) e suas respectivas amplitudes particularizam o som do instrumento musical. No caso recém descrito, 440 Hz é a *frequência fundamental* da nota, pois é a frequência tal que seus múltiplos inteiros explicam melhor o conteúdo do espectro do sinal [22].

Algoritmos para detecção de F_0 são muito importantes em Computação Musical. Eles são usados, por exemplo, nas seguintes aplicações: afinador eletrônico, transcrição automática de melodia, acompanhamento automático, reconhecimento de fala e transposição de notas musicais.

Existem muitos desses algoritmos, mas essencialmente duas categorias: uns atuam no domínio do tempo, outros no da frequência. Aqui apenas descreveremos brevemente alguns dos algoritmos encontrados na literatura, começando pelos da primeira categoria.

Taxa de Eventos

Se sabemos que um áudio é do tipo $x_n = a \sin bn$, para encontrar F_0 basta contar o número de vezes que o sinal cruzou a origem em um determinado

⁹O termo *cepstrum* vem do inglês *spectrum*, pela inversão das primeiras quatro letras.

intervalo de tempo. Por exemplo: se x_n cruza a origem 800 vezes em um segundo, a frequência fundamental é $400Hz$, pois a cada período de onda há dois cruzamentos. Poder-se-ia também usar outros eventos, como a anulação da derivada, os picos ou os vales. Obviamente esse método é falho para ondas não tão simples (Fig. 4.6).

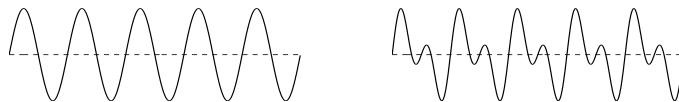


Figura 4.6: O método da contagem de eventos acertaria a frequência fundamental da onda da esquerda, mas falharia na da direita.

Método de McLeod

O método de McLeod [23] é baseado na função de auto-correlação, definida na seção 4.1.2, incluindo algumas heurísticas para contornar os problemas que aparecem quando a maior concentração de energia no espectro não está na frequência fundamental, mas num dos outros harmônicos¹⁰. Consideremos, por exemplo, o sinal $x_n = \sin(2\pi \cdot 2n) + \sin(2\pi \cdot 4n) + 3\sin(2\pi \cdot 6n)$ no intervalo $[0, 1]$ (Fig. 4.7, esquerda). A frequência fundamental é 2 (primeiro harmônico), mas a maior concentração está na frequência 6 (terceiro harmônico). Deste modo, se tomarmos o segundo maior pico (o máximo para índices posteriores ao do primeiro mínimo local) da função de auto-correlação, a frequência correspondente não será a fundamental, mas a do terceiro harmônico (Fig. 4.7, centro).

A normalização da função de auto-correlação proposta em [23] é

$$c_k = \frac{2 \sum_{n=0}^{N-1-k} x_n x_{n+k}}{\sum_{n=0}^{N-1-k} x_n^2 + x_{n+k}^2} \quad (4.1.1)$$

que no caso da figura 4.7 (direita) seria mais adequada ao procedimento de tomar a frequência correspondente ao segundo maior pico.

O algoritmo completo é como segue. Primeiro calcula-se a função de auto-correlação normalizada (Eq. 4.1.1). Em seguida encontra-se todos os *máximos chave*: máximos dos intervalos entre um cruzamento pela origem com inclinação positiva e um cruzamento pela origem com inclinação negativa. Toma-se então o primeiro dos tais máximos que está acima de um certo limiar (dado por uma fração do maior dos máximos) e calcula-se a frequência relacionada.

O Método YIN

¹⁰Este fenômeno não é raro.



Figura 4.7: Respectivamente: sinal, sua função de auto-correlação e a normalização desta no método de McLeod.

A ideia do método YIN [24] é semelhante à anterior, mas ao invés de se procurar por um máximo de uma função de auto-correlação, procura-se por um mínimo de uma *função diferença*.

Consideremos a seguinte função diferença:

$$d_k = \sum_{n=0}^{N-1} (x_n - x_{n+k})^2 \quad (4.1.2)$$

para $k = 0, \dots, N - 1$. Seus mínimos correspondem aos k 's tais que a janela com o respectivo translado está mais próxima à janela sem translado do que as janelas adjacentes.

O método descrito em [24] faz uso da seguinte normalização de d_k :

$$\bar{d}_k = 1_{[k=0]} + 1_{[k \neq 0]} \frac{d_k}{\frac{1}{k} \sum_{j=1}^k d_j} \quad (4.1.3)$$

onde $1_{[A]}$ vale 1 resp. 0 se a afirmação A for verdadeira resp. falsa. A figura 4.8 mostra um exemplo, para o mesmo sinal do método anterior.

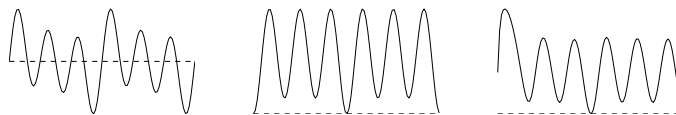


Figura 4.8: Respectivamente: sinal, d_k e \bar{d}_k .

No algoritmo YIN toma-se como frequência fundamental a que corresponde ao menor valor de k tal que \bar{d}_k é um mínimo local da função 4.1.3 abaixo de um certo limiar (maior que zero).

Descrevemos em seguida dois métodos para detecção de F_0 que atuam no domínio das frequências.

HPS

Digamos que a frequência fundamental de um sinal seja $100Hz$ e que o áudio seja rico em harmônicos, ie, tenha energia não nula nas frequências

de $200Hz$, $300Hz$, $400Hz$ e assim por diante. Num caso ideal a energia correspondente às outras frequências seria nula. Pode, contudo, haver outros parciais fortes, como, digamos, o de $90Hz$. Mas é pouco provável que haja muita energia nos múltiplos inteiros de $90Hz$: $180Hz$, $270Hz$, etc. Neste caso, sendo $E(f)$ a energia correspondente à frequência f , o produto $\prod_{j=1}^5 E(100j)$ deve ser maior que $\prod_{j=1}^5 E(90j)$.

O método HPS (Harmonic Product Spectrum), descrito em [25], faz exatamente isso: percorre um conjunto de frequências e verifica qual delas tem harmônicos cujo produto das energias é o maior. Mais precisamente, sendo R o número de fatores a considerar (normalmente $R = 5$) e \hat{x}_k a k -ésima entrada da FFT do sinal x , para k entre zero e o índice correspondente à frequência de $22050/R$ calcula-se $h(k) = \prod_{r=0}^{R-1} |\hat{x}_{(k+1)(r+1)-1}|$, e toma-se como frequência fundamental a que corresponde ao \bar{k} que realiza $\max_k h(k)$.

O principal problema deste método é a resolução da FFT. Se $F0$ vale $80Hz$ mas a resolução da FFT não permite avaliar com precisão frequências próximas desse valor e de seus múltiplos inteiros, o produto $\prod_{j=1}^5 E(80j)$ pode não ser maior que os outros avaliados pelo HPS. Isso pode ser contornado acrescentando zeros à janela avaliada¹¹, o que obviamente eleva o custo computacional.

Máxima Verossimilhança

Neste algoritmo, também descrito em [25], cria-se um banco de dados com os assim chamados “espectros ideais” e, dado um espectro de cuja onda se quer saber a frequência fundamental, procura-se na base de dados o espectro mais próximo¹², sendo devolvido o valor de $F0$ associado a este.

Para uma determinada $F0$, um espectro ideal (Fig. 4.9, direita) é construído a partir de uma “função pente” (Fig. 4.9, esquerda), com picos nos harmônicos correspondentes a $F0$, convolvida com um núcleo como a janela de Hann (Fig. 4.9, centro).

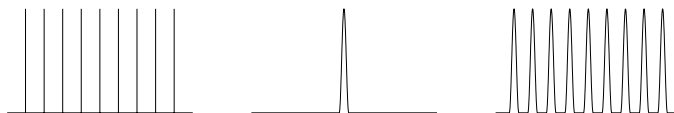


Figura 4.9: Respectivamente: Função pente, janela de Hann e convolução entre ambos.

Obviamente o banco de dados deve ser grande o suficiente para atender a todas as possíveis notas musicais a serem testadas. No caso de um piano, por exemplo, deve haver um “espectro ideal” para cada tecla. Este método

¹¹Em inglês, isso é chamado de *zero-padding*.

¹²De acordo com a norma Euclidiana.

funciona melhor para instrumentos que produzem uma gama discreta de notas musicais, como é o caso do piano e da flauta. Já para a guitarra o método teria problemas com *bends* e vibratos.

Os algoritmos que acabamos de descrever são mais adequados a sons monofônicos: uma única voz ou um único instrumento musical. Existem muitos métodos para sons polifônicos, sendo [1] (Cap. 7, 8 e 9) uma boa referência.

4.1.3 Detecção de Acordes

A técnica de detecção de acordes mais utilizada é muito semelhante ao método de máxima verossimilhança, visto anteriormente, para detecção de $F0$. Porém ao invés de um banco de dados com “acordes ideais”, cria-se uma coleção de vetores de croma de acordes reais, com várias amostras do mesmo acorde, as quais tendem a formar agrupamentos¹³. Dado um acorde a classificar, utiliza-se algum método de Aprendizagem de Máquina para associar o acorde a um dos agrupamentos.

No trabalho original [28] é utilizado o algoritmo k -NN¹⁴: entre os k vetores de croma mais próximos ao vetor de croma do acorde a classificar, faz-se a contagem do número de representantes de cada classe (acorde) e classifica-se o novo acorde como sendo da classe que tiver mais representantes.

4.1.4 Análise Estrutural de Peças Musicais

Digamos que não haja mudança de acorde entre os instantes t_1 e t_2 de uma peça musical. Sendo o vetor de croma um bom descritor de acordes, é razoável supor que, para duas janelas j_1 e j_2 distintas no intervalo $[t_1, t_2]$, os respectivos vetores de croma c_1 e c_2 estejam próximos. Por outro lado, se j_3 é uma janela de um intervalo $[t_3, t_4]$ cujo acorde é diferente daquele em $[t_1, t_2]$, espera-se que c_1 diste de c_3 mais do que de c_2 , ie, que c_1 seja menos similar a c_3 do que a c_2 .

Sendo s_1, \dots, s_K segmentos (janelas) consecutivos de um arquivo de áudio (possivelmente com sobreposição), a *Matriz de Auto-Similaridade* (SSM) deste áudio é $M_{K,K}$, onde $M(i, j)$ é a similaridade entre os segmentos s_i e s_j . Não há consenso na literatura sobre a medida de similaridade. Alguns exemplos são: (a) co-seno do ângulo entre os vetores que descrevem os segmentos s_i e s_j ; (b) $e^{-d(s_i, s_j)/\sigma^2}$, onde d é uma função distância e σ

¹³ *Clusters*.

¹⁴ *k*-Nearest Neighbor.

um parâmetro; (c) $1 - N(i, j)$, onde $N_{K,K}$ é uma matriz de distâncias, normalizada de modo a ter valores em $[0, 1]$.

Assim, dependendo do descritor no qual são baseadas, as SSM's apresentam blocos de elevada auto-similaridade ao longo da diagonal, um bloco para cada intervalo de áudio que apresente grande "coerência interna", ie, valores parecidos para o descritor. O número de blocos e a extensão de cada um também dependem do descritor, como se pode ver na Figura 4.10.



Figura 4.10: Matrizes de auto-similaridade correspondentes aos descritores de cromia e de amplitude, respectivamente, para o áudio da música Anything Goes, da banda AC/DC. A escala de cores vai do azul (baixa similaridade) para o vermelho.

A ideia de utilizar matrizes de auto-similaridade para observar estruturas em peças musicais apareceu em 2001 [31], e desde então tem sido o estado da arte. Uma de suas principais aplicações é o *resumo* de peças musicais¹⁵.

Em [29], por exemplo, propõe-se encontrar o refrão de uma música a partir das linhas diagonais (fora da diagonal principal) que ficam evidentes em matrizes de auto-similaridade. A ideia é a seguinte: nas músicas de estilos como *pop* e *rock*, o refrão é geralmente repetido várias vezes; assim, se a primeira vez que o refrão aparece é a partir do segmento s_i e a segunda é a partir do segmento s_j , então s_{i+k} deve ser muito similar a s_{j+k} , onde a cota superior de k depende da duração do refrão. Aliás, nesse trabalho o *thumbnail* é definido como sendo o próprio refrão, que por sua vez é o bloco com mais repetições.

Já em [30] a técnica de *thumbnailing* é mais sofisticada. Inicialmente calcula-se a SSM (onde o descritor é o espectro de potências) com uma janela estreita (50 mili-segundos). Em seguida a diagonal principal da SSM é convolvida com um núcleo que detecta as fronteiras entre consecutivos blocos com elevada auto-similaridade. O áudio é então re-segmentado com base nas fronteiras entre esses blocos, e uma matriz de auto-similaridade entre os mesmos é calculada. Essa matriz é decomposta em valores singulares, e estes são usados para agrupar os blocos. O *thumbnail* pode ser, por

¹⁵ *Thumbnailing*, em inglês.

exemplo, um dos blocos correspondentes ao *cluster* relacionado ao maior valor singular.

4.2 Composição

A música, em especial a ocidental, a despeito de ser uma arte, é consideravelmente estruturada. Dado um estilo musical qualquer, é possível encontrar um conjunto limitado de regras que o explicam, mesmo que este seja tão numeroso quanto o das canções que pertencentes ao tal estilo. E o compositor, conscientemente ou não, faz uso dessas regras para produzir novas peças musicais do gênero.

Uma vez que há regras, estas podem ser traduzidas em procedimentos, e estes, ser executados por máquinas, as quais são capazes, portanto, de fazer o trabalho do compositor. Isso é conhecido por *composição algorítmica*, ao termo sendo as vezes acrescentado *auxiliada por computador*.

Existem muitas técnicas, mas a maioria segue um mesmo protocolo, ilustrado na figura 4.11. A razão de serem muitas é simples: é grande a quantidade de processos determinísticos/estocásticos que se encontra na literatura matemática, e não há restrições para o mapeamento entre o resultado destes e o conjunto de notas musicais disponíveis.

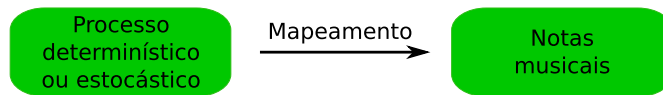


Figura 4.11: Essência da composição algorítmica.

Algumas das ferramentas classicamente usadas em composição automática são: cadeias de Markov, autômatos celulares, fractais e redes neurais.

Embora os trabalhos relacionados ao tema tenham o cuidado de justificar os algoritmos adotados, sendo a justificativa, muitas vezes, matematicamente baseada (como é o caso dos que usam Aprendizagem de Máquina), a avaliação dos resultados obtidos em geral é pautada por critérios estéticos. Por esta razão não daremos detalhes dos algoritmos. Ao leitor interessado recomendamos [2], capítulos 17, 18 e 19.

Dentre os métodos que fogem à regra da figura 4.11, destacam-se aqueles cuja saída não é uma nota musical apresentada em uma partitura ou enviada diretamente a um sintetizador digital, mas um segmento de áudio, extraído de um banco de dados. A composição consiste, neste caso, da concatenação desses segmentos, os quais são organizados com base em algum descritor como os apresentados na sub-seção 4.1.1. Por lidarem diretamente com o

áudio digital, esses algoritmos são computacionalmente mais caros que os anteriores, sendo talvez este o motivo de terem eles aparecido somente em anos recentes. Alguns exemplos podem ser encontrados em [32] e [26].

4.3 Síntese

Síntese de áudio é o processo de gerar som usando dispositivos eletrônicos, sejam eles especialmente desenhados para a tarefa (caso dos sintetizadores analógicos e digitais) ou não (caso dos computadores pessoais, onde a síntese é feita via software).

Nesta seção descreveremos algumas das principais técnicas de síntese, destacando o princípio básico de cada uma delas.

4.3.1 Síntese por Wavetable

Começamos pela mais simples das técnicas. Ela consiste basicamente na reprodução sucessiva da mesma forma de onda, armazenada numa “tabela”, em formato PCM¹⁶ (Fig. 4.12).

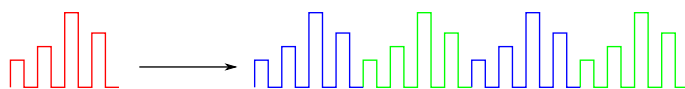


Figura 4.12: Formato básico da síntese por wavetable.

Esse modelo é particularmente eficiente quando se busca fidelidade com respeito ao som de um instrumento real. Exemplo clássico é o dos pianos digitais, cujos fabricantes gostam de dizer que os sons foram *amostrados* desse ou daquele piano de cordas. Por isso as vezes esse tipo de síntese é chamado *síntese por amostragem*.

Dentre as ondas *sintéticas*, ie, não amostradas de um instrumento real, destacam-se a senoidal, a quadrada, a triangular e a dente-de-serra (Fig. 4.13).

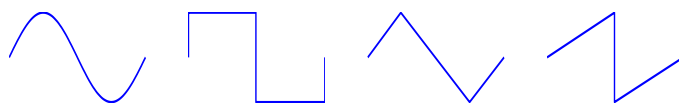


Figura 4.13: Ondas sintéticas mais comuns: senoidal, quadrada, triangular e dente-de-serra.

¹⁶Pulse Code Modulation.

As vezes é interessante pensar na síntese por Wavetable como um sistema constituído de três componentes: um *oscilador*, um controlador de amplitude e um controlador de frequência.

O oscilador é responsável por percorrer os índices da tabela contendo as amostras da onda sonora e devolver os respectivos valores de PCM. A velocidade com que o oscilador percorre a tabela e o conjunto de índices cujos valores PCM serão retornados são de responsabilidade do controlador de frequência. Por exemplo: se a tabela contém um segundo do áudio de uma nota de frequência $440Hz$ amostrada a $44100fps$, para se reproduzir uma nota de $880Hz$ com a mesma taxa de amostragem basta “pedir” ao oscilador para percorrer a tabela duas vezes devolvendo apenas as amostras cujos índices são números ímpares.

E se quiséssemos uma frequência de $450Hz$? O oscilador deveria passar por índices não inteiros, ie, inexistentes na tabela. Neste caso o que se faz é interpolar os valores PCM dos índices próximos, ou seja, reconstruir o sinal no índice inexistente. Este procedimento já foi discutido neste texto, no capítulo sobre processamento de sinais sonoros.

O controlador de amplitude é responsável pelo que se chama de *envelope* da onda, que é o gráfico da função de amplitude ao longo do tempo. O ciclo de vida de uma nota tocada por um instrumento musical pode ser dividido em quatro partes, correspondentes a quatro regiões distintas de seu envelope. A figura 4.14 nomeia essas regiões.



Figura 4.14: Partes de um envelope de onda: (1) Ataque, (2) decaimento inicial, (3) sustentação e (4) decaimento final.

O formato do envelope difere bastante entre os instrumentos musicais. Por exemplo: não há decaimento inicial e sustentação em uma nota tocada em uma corda, somente ataque e decaimento final; já no trompete há apenas ataque e sustentação. Na síntese por Wavetable, o ataque é uma parte delicada, pois sua forma de onda difere bastante das demais, principalmente pela existência de ruído. Logo a simples elevação abrupta da amplitude não vai soar realista. O que se faz normalmente é armazenar uma Wavetable contendo apenas o ataque.

4.3.2 Síntese Aditiva

Imaginemos agora um banco de N osciladores, cada qual com seus controladores (de frequência e de amplitude). Síntese aditiva é o processo de geração de áudio no qual se soma as formas de onda sintetizadas em cada um desses osciladores (Fig. 4.15).

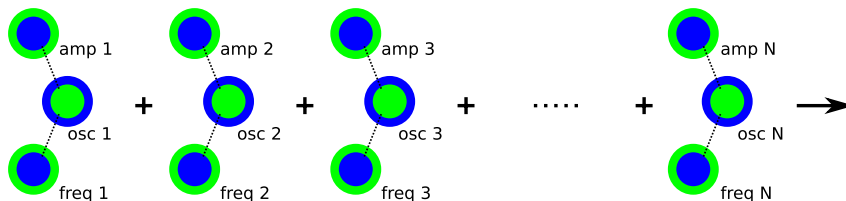


Figura 4.15: Processo de síntese aditiva.

Soa parecido com Transformada de Fourier, não? De fato: a teoria subjacente à transformada de Fourier diz que toda onda periódica pode ser sintetizada por um banco (em geral infinito) de osciladores senoidais. A figura 4.16 ilustra este fato para as ondas não suaves da figura 4.13.

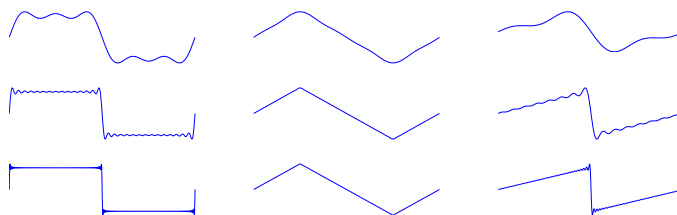


Figura 4.16: Reconstrução das ondas não suaves da figura 4.13 usando os 4, 16 e 64 primeiros osciladores cuja amplitude é não-nula.

Sons harmônicos, como o dos instrumentos tonais, podem ser modelados via síntese aditiva. Basta examinar o espectro do áudio e em seguida re-sintetizar o mesmo usando osciladores senoidais com as respectivas amplitudes sendo aquelas encontradas durante a análise espectral.

Um experimento interessante consiste em usar os valores das amplitudes e frequências de um determinado instrumento em osciladores que geram outra forma de onda, que pode ser, por exemplo, sintética, ou de um segundo instrumento musical. É o que se chama *síntese cruzada*¹⁷.

¹⁷*Cross synthesis*, em inglês.

4.3.3 Síntese Subtrativa

Originalmente a síntese subtrativa consiste em “esculpir” o som: dada uma fonte sonora de espectro rico (ie, com muitas componentes de frequência não nulas), aplica-se uma série de *filtros* de modo a reduzir o *ganho* correspondente à certas frequências do espectro.

Como o conceito de filtro evoluiu do simples processo de subtrair componentes indesejáveis de sinal para, literalmente, a execução de qualquer operação sobre o mesmo, atualmente talvez o termo mais adequado para descrever o que a síntese subtrativa faz com o áudio seja “moldar”. Em resumo, síntese subtrativa e filtragem se confundem.

A teoria dos filtros digitais¹⁸ é muito grande, existindo inúmeros algoritmos, técnicas e abordagens. Só para citar alguns exemplos, a figura 4.17 mostra as funções de transferência de alguns dos filtros disponíveis no software Garage Band. Atualmente a maioria dos editores de áudio implementa uma série de filtros deste tipo.



Figura 4.17: Curvas de resposta de frequência para os filtros (1) *Add Brightness*, (2) *Add Sharpness*, (3) *Bass Boost*, (4) *Stronger Bass*, (5) *Vocal Presence* e (6) *Telephone Voice*.

4.3.4 Síntese por Modulação

Recordemos o sistema de síntese por Wavetable descrito anteriormente, dado por um oscilador e dois controladores de parâmetros (figura 4.18).

Suponhamos que o próprio controlador de amplitude seja um oscilador, ie, que a amplitude mude ao longo do tempo de acordo com uma função, digamos, senoidal. Se a frequência desta senóide estiver dentro do campo audível (ie, acima de 20Hz) sua contribuição ao espectro do oscilador central será perceptível. Por exemplo: se a frequência do oscilador principal for 440Hz e a do controlador de amplitude for de 40Hz, o espectro do áudio

¹⁸O capítulo 2 deste texto apresenta uma introdução ao assunto.

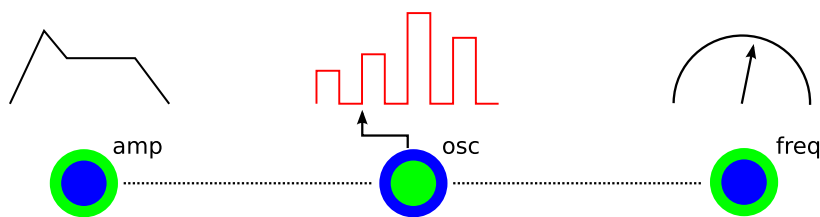


Figura 4.18: Simples unidade de síntese. Com velocidade dada pelo controlador de frequência, o oscilador percorre a Wavetable, cujo envelope é dado pelo controlador de amplitude.

resultante apresentará dois picos além do original, localizados nas frequências 400Hz e 480Hz. O leitor pode se convencer deste fato notando que o som resultante é simplesmente o produto da senóide do controlador de amplitude pela Wavetable original, e lembrando que o espectro do produto é igual à convolução dos espectros.

O que acabamos de descrever é o que se chama de *síntese por modulação de amplitude*, ou simplesmente *síntese AM*.

De modo análogo se define a *síntese FM* (*síntese por modulação de frequência*). Em sua forma mais simples, ela é dada pela seguinte equação:

$$y_n = A \sin \left(\frac{2\pi n f_c}{R} + I \sin \frac{2\pi n f_m}{R} \right) \quad (4.3.4)$$

onde A é a amplitude, f_c (em Hz) é a frequência base¹⁹, f_m (em Hz) é frequência moduladora e I é o índice de modulação, dado por $I = \Delta_f / f_m$, onde Δ_f é o desvio de frequência (em Hz) em relação a f_m .

Trata-se, portanto, de um sistema não-linear, e o entendimento do espectro do sinal modulado é um exercício não-trivial. O truque é escrever a expressão acima em termos das chamadas *Funções de Bessel do Primeiro Tipo*, o que pode ser visto em [5]. O que se encontra é o seguinte: além da frequência f_c , o espectro do sinal modulado (y) apresentará picos nas frequências $f_c \pm k f_m$, para k inteiro positivo. A quantidade de k 's para os quais as amplitude correspondentes às frequências $f_c \pm k f_m$ são significativas é aproximada por $I + 2$. A figura 4.19 mostra alguns exemplos.

A síntese FM foi descoberta por John Chowning, no final da década de 60. Por ser computacionalmente mais barata que as sínteses aditiva e subtrativa, esta modalidade alcançou grande popularidade, vindo a ser

¹⁹Carrier frequency, em inglês.

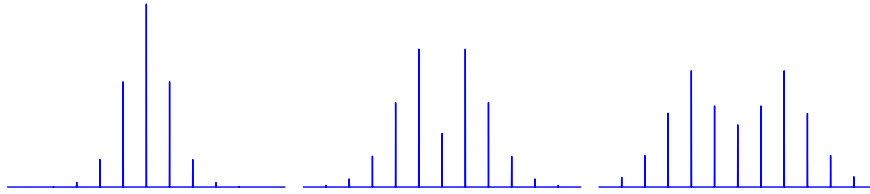


Figura 4.19: Espectros do áudio modulado de acordo com a equação 4.3.4, para $f_c = 600$, $f_m = 100$, $A = 1$ fixados e $I = 1, 2$ e 3 , respectivamente.

adotada pela indústria. Exemplo importante é o Yamaha DX-7 (fig. 4.20), lançado em 1983, o qual vendeu centenas de milhares de unidades ao redor do mundo [5].



Figura 4.20: Yamaha DX-7

4.3.5 E o que mais?

Usando um trocadilho, essa seção foi apenas uma *amostra* do que é feito em termos de síntese de som, a fim de apresentar alguns procedimentos clássicos de construção de áudio *from scratch*.

Faltou falar, por exemplo, de síntese granular e de síntese fisicamente baseada. A primeira consiste essencialmente na concatenação de pequenos fragmentos de áudio (na faixa de 5 a 100 mili-segundos). A segunda usa a caixa de ferramentas de Equações Diferenciais para modelar a onda sonora gerada por instrumentos musicais (ou outras fontes, como uma martelada em uma barra de ferro).

Uma excelente introdução ao assunto é [68]. Já a abordagem do capítulo 3 de [5] é mais detalhada. Outra boa referência é a parte II de [2], que, diga-se de passagem, é um clássico da literatura de Computação Musical.

Capítulo 5

Tendências e Aplicações

Neste capítulo trataremos de aplicações de métodos de computação musical que têm se tornado populares nos últimos anos, devido aos avanços na computação pessoal (de hardware, software e interface com o usuário) e à crescente velocidade da internet banda-larga.

5.1 Localização e Identificação

Não é novidade que a quantidade de informação produzida pelo ser humano sempre cresceu ano após ano. Mas só recentemente essa informação tem se tornado disponível em grande escala, tornando ainda mais importante o problema de *busca* neste enorme banco de dados.

Esse fenômeno atingiu, claro, o universo da música, onde uma pesquisa textual (que é algo bem resolvido) nem sempre é a estratégia mais adequada. A razão é que normalmente tudo que o usuário lembra de uma canção é um fragmento de sua melodia, que pode nem ter uma letra.

A solução é efetuar uma busca com argumentos cantados, cantarolados ou assobiados, o que é mais conhecido por *Query by Humming* (QbH).

Vistos *do alto*, os algoritmos de QbH têm duas etapas principais: primeiro o áudio é capturado e analisado para extração de informação melódica e rítmica; em seguida essa informação é comparada com a aquela correspondente a cada amostra do banco de dados. As melodias que apresentam maior similaridade com a entrada são apresentadas como saída.

Grande parte dos sistemas utilizam informação simbólica [27] (do tipo MIDI). Neste caso a primeira etapa consiste na detecção das notas musicais e no intervalo (temporal) entre as mesmas. Outras possibilidades são, ao invés das notas, representar o intervalo (tonal) entre notas consecutivas, ou

a direção (com respeito à frequência fundamental) para a qual uma nota se desviou da anterior (para cima ou para baixo), caso o tenha feito (Fig. 5.1). Quanto mais grosseira for a representação, mais robusto será o método: a representação por intervalos é imune a alterações no tom fundamental da entrada; a representação por direções é, além disso, imune a certos deslizes de afinação.

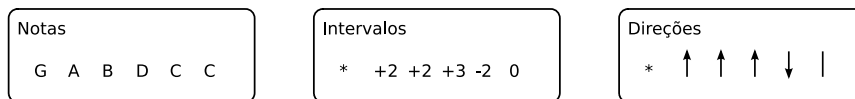


Figura 5.1: Representações melódicas em ordem crescente de robustez.

Quanto à busca de uma melodia similar, pode-se usar, por exemplo, um algoritmo de alinhamento de sequências, onde a função distância é o *custo de edição*; ou *n-gramas*, cujo princípio é buscar a sequência que compartilha com a da entrada a maior sub-sequência de notas consecutivas possível.

O leitor pode experimentar esses sistemas visitando os sites Midomi [9] e Musipedia [10].

Um caso particular do problema acima ocorre quando a entrada é a reprodução muito próxima (ie, a menos de algum ruído, por exemplo) de um segmento de áudio cuja versão inteira está no banco de dados.

Uma das técnicas adotadas neste caso está relacionada à “impressão digital” do áudio, que pode ser construída do seguinte modo: (1) o áudio é segmentado, e um descriptor é extraído de cada segmento; (2) um algoritmo de *clustering* é aplicado ao conjunto de descritores, sendo eleito um representante de cada grupo; (3) o conjunto de representantes define a “impressão digital”.

Em [33] a etapa (2) é realizada via *k-means clustering*, e em (1) os descritores testados são Amplitude¹, Medida de Achatamento Espectral² e Fator Crista Espectral³.

O áudio que se quer classificar passa pelo procedimento (1), e os descritores obtidos são comparados com as “impressões digitais” das amostras da base de dados, buscando-se aquela que permite maior aproximação.

Além da simples busca, uma aplicação interessante desta técnica é a monitoração de rádios. Ocorre que as rádios devem pagar pelos direitos autorais das músicas executadas. Assim, o sinal das mesmas deve ser monitorado, uma tarefa complexa considerando-se o número crescente de rádios.

¹Visto no capítulo 5.

²Razão entre a média geométrica e a média aritmética do espectro de potências.

³Razão entre o valor máximo e a média aritmética do espectro de potências.

Algoritmos de identificação de áudio tem sido aplicados com sucesso neste tipo de situação.

A título de curiosidade, o procedimento recém descrito foi desenvolvido em [11].

5.2 Recomendação por Similaridade

Silenciosa e sorrateiramente, muitos sites da Web 2.0 monitoram o comportamento dos usuários, coletando informações e produzindo estatísticas que visam não apenas tomar decisões que afetam os clientes como um todo, mas também (e principalmente) atingir cada indivíduo em particular.

No caso dos serviços que oferecem produtos, sejam eles bens físicos ou intelectuais (mídias em geral), é prática comum a oferta de produtos similares àqueles que interessam ao usuário. Estamos falando dos chamados *sistemas de recomendação*.

Em relação ao comércio eletrônico de conteúdo musical, talvez o exemplo mais importante seja o sistema Genius, do software iTunes [12], que recomenda canções da loja iTunes com base em uma música escolhida na biblioteca do usuário.

Os sistemas de recomendação também possibilitaram o aparecimento de uma nova classe de rádios on-line, nas quais a lista de execução é composta de músicas semelhantes à que o usuário sugere. As rádios LastFM [14] e Pandora [13] são exemplos.

Existem essencialmente duas abordagens para o problema de recomendação por similaridade [34]: a *colaborativa* e a *baseada em conteúdo*. Na primeira um produto é recomendado com base nas preferências de usuários com perfil semelhante ao do cliente. Na segunda a recomendação é feita simplesmente a partir das características do produto. Os métodos que obtêm melhores resultados são híbridos, embora priorizem a *filtragem colaborativa* [35].

Caso tenha interesse, o leitor encontrará em [36] uma boa lista de referências sobre o assunto.

5.3 Novas Interfaces

Não é surpreendente que as máquinas que chamamos computadores desde o princípio foram vistas por amantes da música como possíveis instrumentos musicais. Esse ponto de vista virou realidade na medida em que os processadores se tornaram capazes de lidar com sinais de áudio em tempo

real, algo que veio a ocorrer (com respeito à computação pessoal, não aos super-computadores dos centros de pesquisa) somente na virada do século.

Atualmente é possível encontrar orquestras constituídas inteiramente de músicos ao computador. Exemplo notório é a *PLOrk* (Princeton Laptop Orchestra) [15], fundada em 2005 pelos departamentos de Música e de Ciência da Computação da Universidade de Princeton.

A orquestra é composta por 15 *meta-instrumentos*, ie, 15 laptops, cada qual equipado com os ambientes Max/MSP [16], SuperCollider [17] e Chuck [18], uma interface *firewire*, um amplificador e um alto-falante hemisférico [37]. Na PLOrk os instrumentistas podem ser orientados por um maestro (como numa orquestra tradicional), dispensar o maestro (comunicando-se via *wireless*), ou uma combinação dos dois.

Revolução parecida com a que possibilitou o processamento de sinal em tempo real está ocorrendo nos dias atuais, desta vez relacionada à interação entre humano e máquina. Trata-se das interfaces multi-toque e tangíveis.

Elas constituem-se normalmente de mesas dotadas de monitores e sistemas de visão computacional que detectam vários toques ao mesmo tempo, além de identificar objetos (tangíveis) portando códigos especiais. No sistema *reactable* [19], por exemplo, essa interface, conectada ao software Pure Data [21], possibilita a performance em tempo real, executada por um ou mais músicos, em uma ou mais mesas. Em [20] o leitor encontrará uma extensa lista de projetos baseados neste tipo de tecnologia.

Bibliografia

- [1] A. Klapuri, M. Davy, “Signal Processing Methods for Music Transcription”. Springer, New York, 2006.
- [2] C. Roads, “The Computer Music Tutorial”. MIT Press, Cambridge, 1996.
- [3] M. Puckette, “The Theory and Technique of Electronic Music”. Versão Online: <http://crca.ucsd.edu/msp/techniques.htm>.
- [4] P. Cook, “Real Sound Synthesis for Interactive Applications”. A K Peters, Natick, 2002.
- [5] F. Moore, “Elements of Computer Music”. Prentice Hall, New Jersey, 1990.
- [6] D. Benson, “Music: A Mathematical Offering”. Versão Online: <http://www.maths.abdn.ac.uk/bensondj/html/maths-music.html>
- [7] “Virginia Tech Multimedia Music Dictionary”. URL: <http://www.music.vt.edu/musicdictionary/>. Consultado em 08/03/2009.
- [8] J. Romberg, “Circular Convolution and the DFT”. URL: <http://cnx.org/content/m10786/2.8/?format=pdf>. Consultado em 13/03/2009.
- [9] Midomi. URL: <http://www.midomi.com>. Consultado em 20/04/2009.
- [10] Musipedia. URL: <http://www.musipedia.org>. Consultado em 20/04/2009.
- [11] Fraunhofer Institute for Digital Media Technology. URL: <http://www.idmt.fraunhofer.de>. Consultado em 21/04/2009.

- [12] iTunes. URL: <http://www.apple.com/itunes>. Consultado em 23/04/2009.
- [13] Pandora Internet Radio. URL: <http://www.pandora.com>. Consultado em 23/04/2009.
- [14] LastFM. URL: <http://www.last.fm>. Consultado em 23/04/2009.
- [15] PLOrk: The Princeton Laptop Orchestra. URL: <http://plork.cs.princeton.edu>. Consultado em 24/04/2009.
- [16] Cycling '74 Max 5. URL: <http://www.cycling74.com/products/max5>. Consultado em 24/04/2009.
- [17] SuperCollider: A Real Time Audio Synthesis Programming Language. URL: <http://www.audiosynth.com>. Consultado em 24/04/2009.
- [18] ChuckK: Strongly-Timed, Concurrent, and On-The-Fly Audio Programming Language. URL: <http://chuck.cs.princeton.edu>. Consultado em 24/04/2009.
- [19] Reactable. URL: <http://www.reactable.com>. Consultado em 24/04/2009.
- [20] Tangible Music. URL: <http://modin.yuri.at/tangibles>. Consultado em 24/04/2009.
- [21] Pure Data. URL: <http://puredata.info>. Consultado em 24/04/2009.
- [22] G. Peeters, "A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project". IRCAM, Analysis/Synthesis Team, 2004.
- [23] P. McLeod, G. Wyvill, "A Smarter Way to Find Pitch". Department of Computer Science, University of Otago.
- [24] A. Cheveigné, H. Kawahara, "YIN, a Fundamental Frequency Estimator for Speech and Music". J. Acoust. Soc. Am. 111 (4), Abril de 2002.
- [25] P. de la Cuadra, A. Master, C. Sapp. "Efficient Pitch Detection Techniques for Interactive Music". Center For Computer Research in Music and Acoustics, Stanford University.
- [26] T. Jehan. "Creating Music by Listening". Tese de Doutorado, MIT, 2005.

- [27] G. Cabral. “Harmonization Automatique en Temps Reel”. Tese de Doutorado, Paris 6, 2008.
- [28] T. Fujishima. Real-Time Chord Recognition of Musical Sound. A System Using Common Lisp Music. Em “Proceedings of International Computer Music Conference”, 1999.
- [29] M. Bartsch, G. Wakefield. To Catch a Chorus: Using Chroma-Based Representations for Audio Thumbnailing. Em “Proceedings of IEEE Worksnop on Applications of Signal Processing to Audio and Acoustics”, 2001.
- [30] M. Cooper, J. Foote. Summarizing Popular Music via Structural Similarity Analysis. Em “Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics”.
- [31] J. Foote, M. Cooper. Visualizing musical structure and rhythm via self-similarity. Em “Proceedings of International Computer Music Conference”. La Habana, Cuba, 2001.
- [32] A. Zils, F. Pachet. Musical Mosaicing. Em “Proceedings of the Cost G-6 Conference on Digital Audio Effects”. Limerick, Ireland, 2001.
- [33] O. Hellmuth, E. Allamanche, J. Herre, T. Kastner, M. Cremer, W. Hirsch. Advanced Audio Identification Using MPEG-7 Content Description. “111th Convention of the Audio Engineering Society”. New York, USA, 2001.
- [34] A. Loizou, S. Dasmahapatra. Recommender Systems for the Semantic Web. Em “ECAI Recommender Systems Workshop”. Trento, Italy, 2006.
- [35] B. Sarwar, G. Karypis, J. Konstan, J. Riedl. Applications of Dimensionality Reduction in Recommender System - A Case Study. Em “ACM WebKDD 2000 Web Mining for E-Commerce Workshop”.
- [36] A. Schein, A. Popescul, L. Ungar, D. Pennock. Methods and Metrics for Cold-Start Recommendations. Em “Proceedings of the 25th Annual International ACM Conference on Research and Development in Information Retrieval”. Tampere, Finland, 2002.
- [37] D. Trueman, P. Cook, S. Smallwood, G. Wang. PLOrk: The Princeton Laptop Orchestra, Year 1. Proceedings of the International Computer Music Conference. New Orleans, USA, 2006.

- [38] P. Fraisse, Rhythm and Tempo. In D. Deutsch (Ed.) "The Psychology of Music" (pp.149-180). New York: Academic Press, 1982.
- [39] E. Clarke, Rhythm and Timing in Music. In D. Deutsch (Ed.) "The Psychology of Music" (pp.473-500). New York: Academic Press, 1999.
- [40] K. Agawu, "Representing African Music". Routledge, New York, 2003.
- [41] A. Gabrielsson, The Complexities of Rhythm. In T. Thige, W. Dowling (Ed.) "Psychology and Music: The Understanding of Melody and Rhythm" (pp. 93-120). Lawrence Erlbaum Associates, New Jersey, 1993.
- [42] F. Lerdahl, R. Jackendoff, "A Generative Theory of Tonal Music". The MIT Press, Cambridge, 1983.
- [43] G. Cooper, L. Meyer, "The Rhythmic Structure of Music". The University of Chicago Press, Chicago, 1960.
- [44] R. Rowe, "Interactive Music Systems". The MIT Press, Cambridge, 1993.
- [45] R. Rowe, "Machine Musicianship". The MIT Press, Cambridge, 2001.
- [46] Rhythm definition. URL: <http://en.wikipedia.org/wiki/Rhythm>. Consultado em 05/05/2009.
- [47] Oldham - Uk Performance in march 2007. URL: <http://www.youtube.com/watch?v=TqzvvcyAM0Q> Consultado em 05/05/2009.
- [48] C. Krumhansl, Rhythm and Pitch in Music Cognition, *Psychological Bulletin*, **126** (2000), 159-179.
- [49] V. Iyer, Embodied Mind, Situated Cognition, and Expressive Micro-timing in African-American Music, *Music Perception*, **19**, (2002), 387-414.
- [50] R. Parncutt, A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms, *Music Perception*, **11**, (1994), 409-464.
- [51] L. Zbikowski, Modelling the Groove: Conceptual Structure and Popular Music, *Journal of the Royal Musical Association*, **129**, (2004), 272-297.
- [52] G. Lewis, Too Many Notes: Computers, Complexity and Culture in Voyager, *Leonardo Music Journal*, **10**, (2000), 33-39.

- [53] F. Pachet, The Continuator: Musical Interaction With Style, *Journal of New Music Research*, **31**, (2002), **-**.
- [54] L. Smith. “A Multiresolution Time-Frequency Analysis and Interpretation of Musical Rhythm”. PhD Thesis, The University of Western Australia, 1999.
- [55] J. Bilmes. “Timing is of the Essence: Perceptual and Computational Techniques for Representing, Learning, and Reproducing Expressive Timing in Percussive Rhythm”. Masters Thesis, MIT, 1993.
- [56] B. Thom. “BoB: An Improvisational Music Companion”. PhD Thesis, Carnegie Mellon University, 2001.
- [57] E. Scheirer. “Music-Listening Systems”. PhD Thesis, MIT, 2000.
- [58] V. Sandvold. “Percussion Descriptors”. Master Thesis, University of Oslo, Norway, 2004.
- [59] D. Murray-Rust. “Musical Acts and Musical Agents: theory, implementation and practice”. PhD Thesis, University of Edinburgh, 2007.
- [60] N. Collins. “Towards Autonomous Agents for Live Computer Music: Realtime Machine Listening and Interactive Music Systems”. PhD Thesis, University of Cambridge, 2006.
- [61] M. Pfeleiderer, The Study of Rhythm in Popular Music: Approaches and Empirical Results. In “Proceedings of the 5th Triennial ESCOM Conference”. Hanover, Germany, 2003.
- [62] J. Paulus, A. Klapuri, Measuring the Similarity of Rhythmic Patterns. In “Proceedings of the 3rd ISMIR”. Paris, France, 2002.
- [63] O. Gillet, G. Richard, Automatic Labelling of Tabla Signals. In “Proc. of the 4th ISMIR Conf”. Maryland, USA, 2003.
- [64] P. Herrera, V. Sandvold, F. Gouyon, Percussion-Related Semantic Descriptors of Music Audio Files. In “Proceedings of 25th International AES Conference”. London, UK, 2004.
- [65] K. Yoshii, M. Goto, H. Okuno, Automatic Drum Sound Description for Real-World Music Using Template Adaptation and Matching Methods. In “Proc. of the 4th ISMIR Conf”. Maryland, USA, 2003.
- [66] P. Roy, F. Pachet, S. Krakowski, Improving the Classification of Percussive Sounds with Analytical Features: a Case Study. In “Proc. of the 8th ISMIR Conf”. Vienna, Austria, 2007.

- [67] V. Iyer, J. Bilmes, M. Wright, D. Wessel, A Novel Representation for Rhythmic Structure. In "Proceedings of the ICMC'97". Thessaloniki, Greece, 1997.
- [68] P. Cook, Principles for designing computer music controllers. In "Proc of NIME'01". Seattle, USA, 2001.
- [69] S. Jorda, Afasia: the Ultimate Homeric One-man-multimedia-band. In "Proceedings of the NIME'02 Conference". Dublin, Ireland, 2002.
m
- [70] G. Weinberg, S. Driscoll, Robot-Human Interaction with an Anthropomorphic Percussionist. In "Proceedings of the SIGCHI conference on Human Factors in computing systems". Montreal, Canada, 2006.
- [71] S. Krakowski, Pandeiro Funk: Experiments on Rhythm-Based Interaction. In "Proceedings of SIGGRAPH'09". New Orleans, USA, in press.

Índice

- alias, 44
- aliasing, 44
 - erro de, 44
- amostra, 37
- amostragem, 20
 - de Fourier, 30
 - pontual, 30, 37
 - uniforme, 37
- amplitude, 77
- auto-correlação, 77

- base
 - de Shannon, 43, 44, 51, 56

- cepstrum, 78
- codificação, 16, 17
 - sem perda, 17
- composição algorítmica, 85
- convolução, 34
 - no domínio discreto, 47
- convolução circular, 78

- decodificação, 16, 17
- delta
 - de Dirac, 23
 - de Kronecker, 39
- descritor, 76
- descritores, 13
- discretização, 16

- envelope, 87
- espectro de potências, 77
- espectrograma, 77

- fenômeno de Gibbs, 51
- filtragem
 - de sinais, 33
- filtragem colaborativa, 95
- filtro, 33
 - adaptativo, 33
 - box, 53, 54
 - especialmente invariante, 33
 - linear, 33, 55
 - para banda, 36
 - passa baixa, 36
 - passa banda, 36
 - passa faixa, 36
 - polinomial, 54
 - retangular, 53
 - seletivo, 36
- frequência
 - de corte, 36
 - vazamento de, 50
- frequência fundamental, 79
- função
 - de transferência, 35
 - pena, 39

- harmônicos, 78, 79
- harmonia, 75

- impulso
 - resposta de, 33
- interpolação, 51

- janela, 76, 77

- limite de Nyquist, 42
- métrica
 - L^p , 19
 - perceptual, 18
 - uniforme, 19
- matriz de auto-similaridade, 83
- melodia, 75
- núcleo de
 - interpolação, 52
 - reconstrução, 52
 - um filtro, 34
- nota musical, 75
- operação com sinais, 32
- passa alta, 36
- percussão, 75
- princípio da incerteza, 29
- produto de sinais, 33
- quantização, 20
- query by humming, 93
- reconstrução, 16, 51
 - erro de truncamento, 52
 - exata, 21
 - ideal, 21
- região de
 - altas frequências, 35
 - baixas frequências, 35
- representação
 - espaço de, 30
 - exata, 30
 - finita, 30
 - linear, 30
- reticulado, 37
 - vértice do, 37
- ritmo, 75
- síntese am, 90
- síntese cruzada, 88
- síntese fm, 90
- sinais
 - espaço de, 18
 - soma de, 32
- sinal
 - analógico, 20
 - codificado, 16
 - contínuo, 16, 20
 - de amostragem, 39
 - de banda limitada, 35
 - de pulso, 22
 - dente de serra, 26
 - digital, 21
 - discreto, 16
 - impulso, 22
 - modelo espectral, 27
 - modelo estocástico, 18
 - modelo funcional, 18
 - representação de um, 29
- sistemas de recomendação, 95
- soma
 - de sinais, 32
- suporte, 35
- taxa de eventos, 76
- teorema
 - de Shannon-Whittaker, 43, 51
- transformada, 34
 - cosseno, 28
 - de Fourier, 27, 34
 - de Fourier com janela, 28
 - de Fourier discreta, 48
 - de Fourier finita, 48, 50
 - de Fourier rápida, 50
 - de wavelets, 28
- transformada de Fourier discreta, 76
- vetor de croma, 78

NOTAS EM MATEMÁTICA APLICADA

Arquivos em pdf disponíveis em <http://www.sbmac.org.br/notas.php>

1. Restauração de Imagens com Aplicações em Biologia e Engenharia
Geraldo Cidade, Antônio Silva Neto e Nilson Costa Roberty
2. Fundamentos, Potencialidades e Aplicações de Algoritmos Evolutivos
Leandro dos Santos Coelho
3. Modelos Matemáticos e Métodos Numéricos em Águas Subterrâneas
Edson Wendlander
4. Métodos Numéricos para Equações Diferenciais Parciais
Maria Cristina de Castro Cunha e Maria Amélia Novais Schleicher
5. Modelagem em Biomatemática
Joyce da Silva Bevilacqua, Marat Rafikov e Cláudia de Lello
Courtouke Guedes
6. Métodos de Otimização Randômica: algoritmos genéticos e “simulated annealing”
Sezimária F. Pereira Saramago
7. “Matemática Aplicada à Fisiologia e Epidemiologia”
H.M. Yang, R. Sampaio e A. Sri Ranga
8. Uma Introdução à Computação Quântica
Renato Portugal, Carlile Campos Lavor, Luiz Mariano Carvalho
e Nelson Maculan
9. Aplicações de Análise Fatorial de Correspondências para Análise de Dados
Homero Chaib Filho

10. Modelos Matemáticos baseados em autômatos celulares para Geoprocessamento
Marilton Sanchotene de Aguiar, Fábria Amorim da Costa,
Graçaliz Pereira Dimuro e Antônio Carlos da Rocha Costa
11. Computabilidade: os limites da Computação
Regivan H. N. Santiago e Benjamín R. C. Bedregal
12. Modelagem Multiescala em Materiais e Estruturas
Fernando Rochinha e Alexandre Madureira
13. Modelagem em Biomatemática (Coraci Malta ed.)
 - 1 - “Modelagem matemática do comportamento elétrico de neurônios e algumas aplicações”
Reynaldo D. Pinto
 - 2 - “Redes complexas e aplicações nas Ciências”
José Carlos M. Mombach
 - 3 - “Possíveis níveis de complexidade na modelagem de sistemas biológicos”
Henrique L. Lenzi, Waldemiro de Souza Romanha e Marcelo Pelajo- Machado
14. A lógica na construção dos argumentos
Angela Cruz e José Eduardo de Almeida Moura
15. Modelagem Matemática e Simulação Numérica em Dinâmica dos Fluidos
Valdemir G. Ferreira, Hélio A. Navarro, Magda K. Kaibara
16. Introdução ao Tratamento da Informação nos Ensinos Fundamental e Médio
Marcilia Andrade Campos, Paulo Figueiredo Lima
17. Teoria dos Conjuntos Fuzzy com Aplicações
Rosana Sueli da Motta Jafelice, Laércio Carvalho de Barros,
Rodney Carlos Bassanezi
18. Introdução à Construção de Modelos de Otimização Linear e Inteira
Socorro Rangel

19. Observar e Pensar, antes de Modelar
Flavio Shigeo Yamamoto, Sérgio Alves, Edson P. Marques Filho,
Amauri P. de Oliveira
20. Frações Contínuas: Propriedades e Aplicações
Eliana Xavier Linhares de Andrade, Cleonice Fátima Bracciali
21. Uma Introdução à Teoria de Códigos
Carlile Campos Lavor, Marcelo Muniz Silva Alves, Rogério
Monteiro de Siqueira, Sueli Irene Rodrigues Costa
22. Análise e Processamento de Sinais
Rubens Sampaio, Edson Cataldo, Alexandre de Souza Brandão
23. Introdução aos Métodos Discretos de Análise Numérica de EDO e
EDP
David Soares Pinto Júnior
24. Representações Computacionais de Grafos
Lílian Markenzon, Oswaldo Vernet
25. Ondas Oceânicas de Superfície
Leandro Farina
26. Técnicas de Modelagem de Processos Epidêmicos e Evolucionários
Domingos Alves, Henrique Fabrício Gagliardi
27. Introdução à teoria espectral de grafos com aplicações
Nair Maria Maia de Abreu, Renata Raposo Del-Vecchio, Cybele
Tavares Maia Vinagre e Dragan Stevanović
28. Modelagem e convexidade
Eduardo Cursi e Rubens Sampaio
29. Modelagem matemática em finanças quantitativas em tempo discreto
Max Oliveira de Souza e Jorge Zubelli
30. Programação não linear em dois níveis: aplicação em Engenharia
Mecânica
Ana Friedlander e Eduardo Fancello

31. Funções simétricas e aplicações em Combinatória
José Plínio de Oliveira Santos e Robson da Silva
32. Semigrupos aplicados a sistemas dissipativos em EDP
Carlos Raposo da Cunha
33. Introdução à Simulação Estocástica para Atuária e Finanças Usando R
Hélio Côrtes Vieira, Alejandro C. Frery e Luciano Vereda
34. Modelos de Sustentabilidade nas Paisagens Amazônicas Alagáveis
Maurício Vieira Kritz, Jaqueline Maria da Silva e Cláudia Mazza
35. Uma Introdução à Dinâmica Estocástica de Populações
Leonardo Paulo Maia
36. Geometria de Algoritmos Numéricos
Gregorio Malajovich
37. Equações Diferenciais, Teorema do Resíduo e as Transformadas Integrais
Edmundo Capelas de Oliveira e Jayme Vaz Júnior