UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA POLITÉCNICA
DEPARTAMENTO DE ELETRÔNICA E DE COMPUTAÇÃO

Reconhecedor de Notas Musicais em Sons Polifônicos

Autor:
Angélica Soares Ogasawara

Orientador:
Sergio Lima Netto, Ph.D.

Examinador:
Luiz Wagner Pereira Biscainho, D.Sc.

Examinador:
Filipe Castello da Costa Beltrão Diniz, M.Sc.

DEL
Abril de 2008

Agradecimentos

Meus sinceros agradecimentos:

- aos meus pais, por tudo;
- ao meu irmão e à minha cunhada, pelas palavras de apoio nos momentos oportunos;
- ao meu namorado, Vitor, por estar sempre presente e ter me apoiado em todos os momentos, incondicionalmente;
- ao Professor Sergio Lima Netto, pela orientação dada no desenvolvimento deste projeto;
- ao Professor Carlos José Ribas D'Avila, por seu incansável trabalho na coordenação do curso;
- ao Professor Jomar Gozzi, por ser mais do que um mestre, mas amigo e inspiração, em todos os momentos durante o curso;
- ao Professor Mauros Campello Queiroz, que, com sua maneira simples e prática, indicou o caminho para diferenciar o que realmente importa do que "não serve para nada";
- aos amigos Rafael e Thaís, pela presença constante, paciência e apoio incondicional, pelas leituras das sucessivas versões, debates, sugestões, amostras de notas musicais e acordes de teclado;
- às amigas Juliana, Maraísi, Raquel e Débora, que contribuíram para a conclusão deste;
- à Turma November, a família de São José dos Campos que, de uma maneira ou outra, serviu de fonte de inspiração para o término deste. Em especial, agradeço aos amigos: Jonathas, por ter ensinado a utilizar o Finale e oferecido grande apoio em teoria musical; Paulo, por ter ouvido inúmeras vezes os rascunhos dos capítulos aqui presentes; e Bruna, pelas conversas, resenhas e palavras de conforto e de amizade; e
- a todos os funcionários e professores da Universidade Federal do Rio de Janeiro, por todos os serviços prestados.

Entendeu? 'Tendeu não, né?

Resumo

O objetivo deste projeto é desenvolver um sistema modularizado capaz de reconhecer as notas musicais presentes em um trecho de sinal contendo um som polifônico. A proposta é realizar o treinamento do sistema, que corresponde a gerar um banco de tons a partir de instrumentos sintetizados eletronicamente, e alimentá-lo com peças de música tocadas pelo mesmo instrumento para validação. A identificação das notas musicais componentes se dará pela segmentação em trechos com duração de uma nota, seguida pela identificação de suas freqüências fundamentais e parciais harmônicos. Testes foram realizados para validar o método e apontar os pontos críticos para abordagens futuras. O estudo realizado neste projeto é interessante para estudantes de música e para a preservação de peças musicais, através do registro das partituras.

Palavras-chave: Transcrição Musical, Modularização, Polifonia.

Abstract

The objective of this project is to develop a modularized system able to recognize musical note contained in a signal which represents a polyphonic sound. The proposal is to train the system, which corresponds to generate a tone model database for electronic synthesized musical instrumentals, and to feed it with musical pieces played by the same instrument in order to validate it. Musical-note identification is done by time segmentation, followed by search for their fundamental frequencies and harmonics partials. Tests were made to validate the method and to recognize critical points for further improvements. The proposed system is interesting for music students and for preservation of musical pieces, since it allows the storage of their musical scores.

Key-words: Musical transcription, Modularization, Polyphony.

Índice

Índice de Figuras

Índice de Tabelas

Capítulo 1 - Introdução

"Let's start at the very beginning
a very good place to start
When you read you begin with A-B-C
When you sing you begin with do-re-mi"
('Do Re Mi', música do filme "The sound of music")

A música é uma manifestação cultural de ampla difusão, capaz de levar a algum grau de compreensão qualquer pessoa, independente de século, etnia, formação cultural ou intelectual. Ela é uma das mais expressivas formas de se manifestar sentimentos, sejam agradáveis, dramáticos ou até mesmo de repulsa. Enquanto um quadro permanece estático, uma mesma peça musical é passível de constante renovação, uma vez que cada intérprete oferece sua própria versão, seja alterando a velocidade de sua execução, a intensidade com a qual toca os diferentes tons que a compõem ou mesmo o grupo de instrumentos utilizados. E, no fato de não

se precisar estudar música para ter sua própria percepção da mesma, reside a razão pela qual é aclamada como “linguagem universal”.

Há quem considere a música uma linguagem, com elementos como fonética, morfologia e sintaxe. Outros não a definem como tal, mas como um veículo ou metáfora. Não é de interesse discutir se a música é ou não uma linguagem, mas fazer uso de uma analogia para justificar a motivação deste trabalho.

Ao se pensar no tema do presente trabalho, depara-se com o que seria a “Fonologia da música”, ou seja, o estudo de seus sons básicos, aqueles que, agregados e trabalhados, geram as peças complexas.

Numa linguagem como o idioma Português, existem fonemas como /a/, /k/ ou /x/, que, comparados ao contexto musical, seriam equivalentes às notas musicais, os sons básicos. Mas como elas estão inseridas no contexto musical? Como este alfabeto ‘fonético’ é relacionado internamente?

As palavras poderiam ser seqüências de sons; os dígrafos, acordes; estilos musicais distintos, dialetos. Mas independente de classificações, ou destas comparações com uma linguagem, a música seria sempre formada por micropedaços chamados notas musicais, uma espécie de unidade básica, como um átomo, a partir da qual qualquer som pode ser reproduzido.

Foi devido à padronização de uma notação musical que hoje se desfruta do som secular de Bach, Tchaikovsky ou Beethoven. Através da leitura de partituras, músicos experientes podem reproduzir qualquer peça destes compositores. Embora cada intérprete seja capaz de tocar com intensidades e entonações diferentes, modificando a sensação que o som nos inflige, existe uma melodia padrão que é decifrada a partir de códigos.

Sendo, então, possível traduzir cada som musical em um conjunto de sinais grafados, é possível também preservá-lo e reproduzi-lo inúmeras vezes. Esta é a motivação de uma área chamada transcrição musical, definida como “o ato de ouvir uma peça musical e transcrever a notação musical para as notas que constituem a peça” [1].

A transcrição musical envolve procedimentos complexos. Mesmo os músicos mais experientes teriam dificuldade em transcrever peças com grande polifonia, ou seja, com mais de um som tocando por vez. Daí o interesse em desenvolver um método computacional (portanto, automático) visando à transcrição. No entanto, identificar os sons presentes em uma mistura complexa de notas, como é o caso de uma música, não é o suficiente. Seria fundamental a análise rítmica (para a compreensão da duração de cada nota, dos compassos empregados), além de sólidas noções de estilos musicais (o que, por si só, renderia inúmeras linhas de pesquisa). Esses assuntos contribuiriam bastante para resolver questões de ambigüidade e para detecção de erros.

Desta forma, o objetivo inicial seria buscar reconhecer as notas que fazem parte de um trecho polifônico de piano, através da análise das freqüências fundamentais e seus parciais harmônicos, comparando-os com um banco de notas previamente apresentadas ao sistema (na fase de treinamento), e buscando como resultado as notas e a intensidade das mesmas. Assim, o foco deste trabalho é mais um identificador de notas do que um transcritor propriamente dito.

Com o decorrer do projeto, à medida que a literatura apresentava novos métodos, chegou-se à conclusão de que, mais importante do que um *software* final ou algoritmos para sua execução, era a necessidade de modularizá-lo. É preciso ter em mente que a tecnologia se renova a uma velocidade impressionante, mas a maneira como se trata o sistema (seus blocos lógicos e como são interligados entre si) não se modifica de uma forma tão brusca. Desta forma, apesar de a motivação ser obter um código funcional, algo tangível, preocupou-se mais em organizar as idéias envolvidas neste projeto do que o programa em si. A idéia de modularização do reconhecedor, fragmentando-o em blocos com objetivos menores, entradas e saídas bem definidas e com função bem caracterizada, é essencial para que o sistema possa ser aprimorado.

1.1 Proposta de trabalho

Este Projeto Final propõe uma configuração de sistema capaz de identificar automaticamente notas musicais presentes em um trecho polifônico. A entrada do sistema é uma curta seqüência musical polifônica e a saída é composta de um vetor

com os índices de início de cada nota ou conjunto de notas simultâneas existentes no som, as notas musicais presentes em cada um destes segmentos, suas intensidades e o tipo de instrumento.

A notação musical expressa a freqüência fundamental de um som que deve ser tocado, fazendo uso de algum instrumento musical. O uso de tal representação como saída do sistema é pertinente e adequado, uma vez que é uma notação simbólica passível de ser novamente transformada em um sinal acústico.

A premissa básica do sistema é segmentar o processo em diversos módulos independentes de funções e parâmetros bem definidos. Para isso, é necessário analisar o fluxo de informações e estudar maneiras de como se obter os dados necessários para a modelagem de um tom e seu posterior reconhecimento, dadas tais características.

Identificam-se dois processos fundamentais: o de treinamento e o de reconhecimento. No treinamento, gravam-se todas as notas musicais possíveis de um piano, extraem-se as informações de freqüência fundamental e da intensidade de uma quantidade J de harmônicos, gerando então um banco de notas, contendo todos os tons possíveis a partir daquele instrumento. No processo de reconhecimento, uma peça musical é aplicada na entrada do sistema, que deverá identificar quais as notas (freqüências fundamentais) existentes naquele trecho. Três processos auxiliares são utilizados: um processo de segmentação temporal, que dividirá a seqüência de notas da peça musical em trechos com duração de uma única nota; uma transformada que levará o sinal do domínio do tempo para o domínio da freqüência, onde serão feitas as análises de tom; e um processo de criação de núcleo de filtro, que permitirá a implementação de um filtro capaz de extrair informações mais precisas a partir dos valores de freqüência e amplitude dos parciais harmônicos.

No sistema descrito, trabalham-se dois sinais de informação. A primeira análise é feita a partir da variação de energia (intensidade) no domínio do tempo, que permitirá a segmentação temporal em blocos menores. A segunda análise é feita a partir do espectro de freqüência do sinal, rastreando-se as possíveis freqüências

fundamentais existentes naquele som, em que a presença de cada parcial harmônico relativo a um candidato é considerada como evidência da existência do mesmo.

A validação do sistema é dada por três tipos fundamentais de teste. Primeiramente, efetua-se um teste monofônico, a fim de avaliar se o sistema é capaz de reconhecer uma nota sem a interferência de outros sons. A seguir, testam-se trechos polifônicos como acordes ou notas aleatórias tocadas ao mesmo tempo. O último teste consiste em analisar um trecho de uma peça musical, a fim de avaliar seu desempenho como um todo.

A partir do resultado, identificaram-se os principais pontos críticos do sistema e discutiram-se quais as melhorias a serem implementadas nele no futuro.

A principal ferramenta utilizada foi o Matlab©, versão 6.0 *release* 12, para execução de todo o processamento de áudio em busca do reconhecimento das notas. Todas as amostras musicais foram geradas pelo programa Finale©, versão 2005, sejam elas notas de treinamento do programa, acordes e músicas de teste, a partir da inserção das partituras e execução das mesmas, obtendo-se um arquivo *wave* amostrado a 44100 Hz. Desta forma, o material de treinamento e de teste deste projeto é formado por arquivos *wave* gerados a partir de arquivos MIDI.Também foi usado o programa Cool Edit Pro©, versão 2.1, para a edição dos sinais acústicos e também a título de comparação em alguns tópicos.

1.2 Base de dados

O método proposto neste projeto será construído realizando análises em sinais de áudio, gerados a partir do *software* Finale© 2005. Este programa permite a inserção direta das partituras, escolhendo-se os instrumentos, podendo-se salvar o resultado tanto na extensão **.mus*, própria, ou exportá-lo no formato *wave*, que foi o utilizado para a criação dos sons de treinamento, dos acordes de teste e das músicas.

A principal vantagem em fazer uso deste programa é o fato de se saber exatamente o que foi inserido no som, ou seja, a comparação dos resultados se dará diretamente entre o que há em uma partitura e o que resultou do sistema.

Para o treinamento, foi utilizado um piano sintetizado via *software*, sendo executado da faixa de 130 Hz (C3) até 1047,5 Hz (C6), totalizando 37 tons de alcance. A Tabela 1 resume os dados de treinamento.

Tabela 1 – Treinamento dos Sons.

Instrumento	Notas	Alcance Total
Piano	C3-C6	37

Para a primeira avaliação de desempenho, foi utilizada uma série de acordes e misturas polifônicas, que serão apresentados no Capítulo 9.

Para a segunda avaliação de desempenho, fazendo uso de trechos e pequenas peças musicais, foram utilizados sons monofônicos e polifônicos. A Tabela 2 apresenta os dados referentes a tais peças, de acordo com parâmetros como tipo de som (monofônico ou polifônico) e polifonia máxima, que corresponde ao máximo de notas tocadas simultaneamente.

Tabela 2 – Peças musicais utilizadas para avaliação de desempenho.

Música	**Tipo de Som**	**Polifonia Máxima**
Escala Maior C	Monofônico	1
Brothers	Polifônico	4
First Love	Polifônico	5

Trata-se de sinais digitais de áudio amostrado à taxa de 44.100 Hz com 16 bits por amostra em um único canal.

1.3 Organização do texto

O Capítulo 2 estuda as estreitas relações entre música e matemática, apresentando a análise do som no domínio do tempo e no domínio da freqüência, bem como quais características podem ser observadas em cada um deles.

O Capítulo 3 é dedicado a uma visão geral do projeto e de seus principais sistemas componentes. Neste, apresentam-se os módulos integrantes e o fluxo de dados, bem como a dedução lógica que levou a tal arquitetura. O objetivo é deixar clara a função de cada membro do conjunto e suas relações de entrada e saída.

O Capítulo 4 apresenta o módulo de segmentação temporal do som de entrada, que visa a dividir em trechos com a duração de cada uma das notas ou das misturas polifônicas existentes.

O Capítulo 5 trabalha a conversão dos dados do domínio do tempo para o domínio da freqüência. Para isso, será feita uma revisão inicial das relações matemáticas que definem as notas musicais e a natureza logarítmica da audição humana e, conseqüentemente, da musical ocidental. Com base nestas restrições, apresenta-se uma transformada que atende a uma resolução suficiente nas baixas freqüências e computacionalmente eficiente para o objetivo.

O Capítulo 6 debate sobre o grande problema de se lidar com sons polifônicos: a sobreposição de parciais harmônicos. Com base nisso, discute-se sobre a probabilidade de um sinal ser corrompido, ou seja, sobreposto. Em seguida, apresenta-se um algoritmo para a criação de um filtro capaz de extrair as informações a partir de todos os parciais harmônicos representativos do som e relacionam-se alguns momentos em que este filtro é utilizado.

No Capítulo 7, apresenta-se o processo de modelagem de tom. A partir de notas de treinamento, o sistema funciona como um identificador monofônico, extraindo os dados referentes a freqüência fundamental, parciais harmônicos e suas respectivas amplitudes, para serem armazenados e utilizados para consulta no módulo reconhecedor.

O Capítulo 8 descreve o módulo de reconhecimento polifônico. Algumas restrições lógicas são apresentadas e o processo é conduzido com base no rastreamento por candidatos à freqüência fundamental, em ordem ascendente de freqüência. Caso o som exista, calcula-se sua intensidade e ele é, então, subtraído, para que não interfira na identificação de sons em freqüências mais altas.

O Capítulo 9 apresenta os testes de validação do procedimento e a análise de desempenho do sistema. Busca-se explicar suas restrições e apontar suas principais falhas, a fim de que lhe possam ser feitas melhorias posteriores.

Finalmente, no Capítulo 10, as conclusões são conduzidas, apresentando as vantagens e desvantagens do método descrito, as principais falhas a serem exploradas e sugestões para trabalhos futuros.

Capítulo 2 - Elementos do Tempo e da Freqüência na Música

É de fundamental importância solidificar alguns conceitos relacionados à música e seus aspectos a serem observados. Este capítulo não tem a pretensão de suprir todos os temas envolvendo a análise matemática da música, mas de apresentar as características mais relevantes para a transcrição musical com as quais este projeto lida. Caso haja interesse em uma leitura mais abrangente, este assunto pode ser encontrado em [2], [3] e [4].

A Seção 2.1 discute a natureza física e perceptual de um som e como se pode descrevê-lo a fim de evidenciar alguns dos parâmetros extraídos, como a amplitude, a freqüência fundamental, a distribuição da energia pelos parciais harmônicos, caracterizando o timbre do instrumento, e a evolução da intensidade do som no

decorrer do tempo. Evidencia-se como a emissão de um instrumento pode estar associada à envoltória de amplitude de um sinal, colaborando com a detecção do início de um novo som, premissa a ser utilizada na segmentação temporal. Na Seção 2.2, revisam-se os conceitos e onde devem ser procuradas tais informações, dividindo o processamento em dois domínios: o do tempo, envolvendo todas as operações relacionadas a ritmo e segmentação; e o da freqüência, responsável pela identificação das notas musicais. Por fim, a Seção 2.3 concluirá o capítulo.

2.1 A física e a percepção do som

A onda sonora é um tipo de onda mecânica, consistindo numa perturbação que viaja através de um meio físico, transportando energia de um local para outro. A onda sonora produz áreas de altas e baixas pressões do ar, que movem o tímpano de modo que um ser humano é capaz de ouvi-la. A percepção desta é o som.

Os conceitos físicos de uma onda sonora estão intimamente ligados aos conceitos perceptuais de um som. Isso significa que, para cada um dos elementos capazes de descrever uma onda fisicamente, tem-se uma sensação de como esta característica se manifesta. A ciência que estuda a percepção de um som é a psicoacústica, área afim onde podem ser encontrados muitos trabalhos relacionados à música e à transcrição musical, como [5] e [6].

A forma mais simples de uma onda sonora é a senoidal, portanto, uma onda periódica, representada pela Figura 1, que apresenta os parâmetros com os quais pode ser descrita. Um exemplo é o tom de discagem de uma linha telefônica, que corresponde a uma senóide de aproximadamente 400Hz.

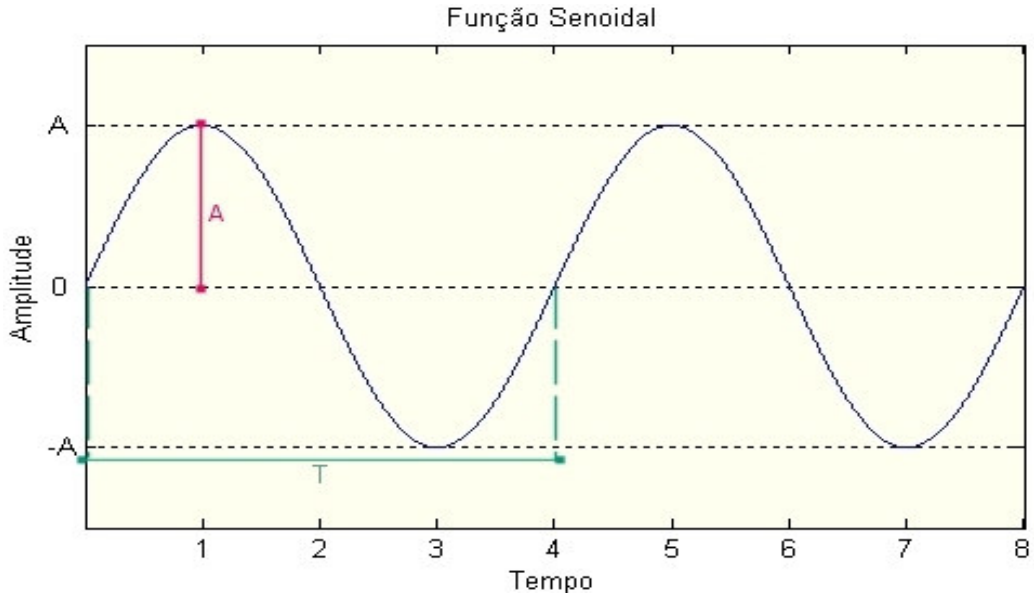


Figura 1 - A forma mais simples de onda sonora, a senóide, pode ser descrita por parâmetros como amplitude A e o período T

Uma senóide pode ser descrita como uma função do tipo:

$$x(t) = A.sen(2\pi f.t + \varphi) \tag{2.1}$$

Seus parâmetros:

- a amplitude máxima, ou simplesmente amplitude, *A*, que corresponde ao máximo valor de deslocamento positivo;
- uma freqüência *f*, que deriva do período T da onda, sendo o número de ciclos por unidade de tempo, normalmente medido em Hertz (Hz). No exemplo, temos uma onda completa a cada 4 segundos, o que resultaria em $f = 0{,}25$ Hz;
- uma fase φ, que determina a posição inicial de uma onda, medida em graus ou em radianos. No exemplo, a fase é nula.

Os parâmetros podem ser mapeados como qualidades sensoriais. A **amplitude** está relacionada com a maneira como percebemos a **intensidade** de um som. Quanto maior a amplitude, maior o deslocamento do ar e, portanto, maior a intensidade percebida. Já a **freqüência** é associada à **altura** de um som, se ele é grave (baixas freqüências) ou agudo (altas).

Na música ocidental, convenciona-se que alguns valores de freqüência são equivalentes às notas musicais. Esta consideração é a chave para a resolução da questão da transcrição musical.

Desta forma, a percepção do som inclui os aspectos de **intensidade** (amplitude) e de **tom** (freqüência). O terceiro aspecto é o **timbre**, que é o que difere o som produzido por um instrumento de outro, também chamado de *cor* ou *textura* de um tom [3]. Por exemplo, caso insiram um osciloscópio na saída de um teclado musical, as formas de onda de uma mesma nota terão grande complexidade, e cada instrumento sintetizado possuirá um desenho diferente, o seu timbre respectivo, como se pode observar na Figura 2, que apresenta as distintas formas de onda obtidas para a mesma nota executada por instrumentos diversos. No caso de existir apenas uma freqüência, o timbre é dito monótono.

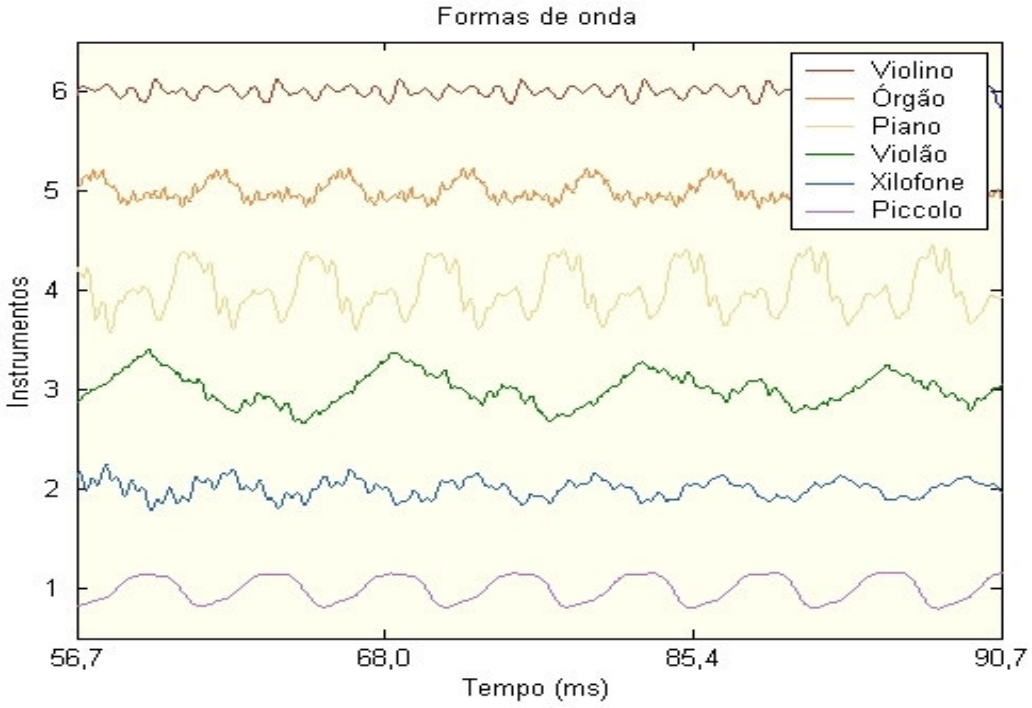


Figura 2 - Para uma mesma nota, cada instrumento possui formas de ondas distintas, o que caracteriza o timbre.

Qualquer forma de onda pode ser desmembrada como um somatório de inúmeras ondas mais simples – as senóides puras – de diferentes amplitudes e

freqüências. Assim, a forma de onda de um instrumento pode ser reescrita desta forma:

$$S(t) = A_1 sen(2\pi f_1 t + \varphi_1) + A_2 sen(2\pi f_2 t + \varphi_2) + A_3 sen(2\pi f_3 t + \varphi_3) + ... \quad (2.2)$$

Pode-se concluir que, para um som não monótono, existem dois tipos de intensidade. Uma delas é a **intensidade de cada uma de suas componentes** (ou seja, os valores de A_1, A_2, A_3, *etc*, da Equação (2.2), por exemplo). A outra é a **intensidade geral** do sinal, que, para o seu cálculo, deve-se considerar todas as amplitudes relativas às freqüências que compõem aquele som. Isso significa que existe uma operação matemática, envolvendo A_1, A_2, A_3, *etc*, que resultará neste valor único. Tal método será discutido no Capítulo 6. Por ora, é importante observar que para se identificar um som, busca-se a sua intensidade geral, mas antes será preciso identificar as intensidades individuais para efetuar o cálculo.

2.1.1 Parciais Harmônicos

Quando as freqüências das componentes de um som são relacionadas de forma simples, como múltiplos inteiros da freqüência fundamental, as componentes serão chamadas de **parciais harmônicos** ou somente **harmônicos**. Portanto, um som é harmônico quando existe uma relação múltipla e inteira entre as freqüências e uma fundamental chamada f_0.

Desta forma, ao se extrair o espectro de freqüência de um som harmônico, tem-se intervalos constantes de tamanho f_0, como visto na Figura 3, que ilustra os componentes espectrais de uma nota A (440 Hz) de um Oboé, em escala linear.

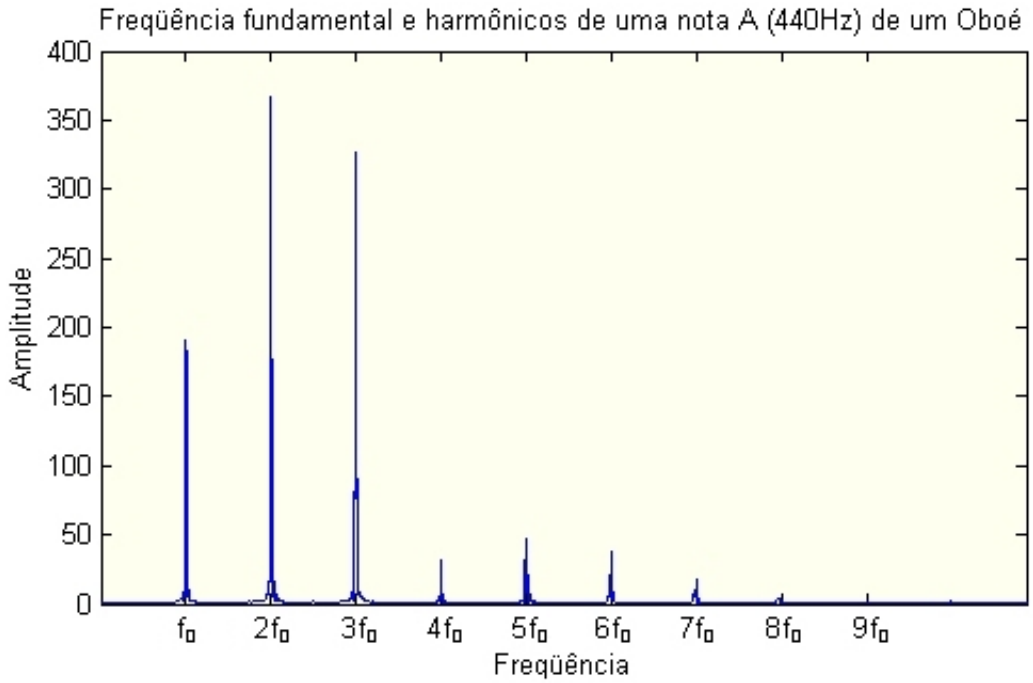


Figura 3 - O desmembramento de um sinal acústico em sua freqüência fundamental e harmônicos de uma nota musical A, 440 Hz, de um Oboe sintetizado por computador. Observe que o primeiro pico se dá na freqüência f_0, a fundamental, e as demais obedecem a uma relação inteira do valor de f_0.

Os harmônicos aparecem como picos no espectro de freqüência, sendo a menor parcial na freqüência f_0 (fundamental). Usa-se a palavra **tom** como sinônimo de freqüência fundamental, apesar de a primeira expressão estar normalmente associada à percepção da segunda.

Neste projeto, estudam-se apenas os sons harmônicos. Estritamente falando, para sons harmônicos, a freqüência de um harmônico h_j qualquer será $j.f_0$.

Os sons cujas parciais não obedecem a intervalos inteiros determinados são chamados usualmente de inarmônicos (do inglês, *inharmonic*). Normalmente estão associados a instrumentos de percussão, como pratos de bateria, gongos, xilofones e sinos [7].

Se uma mesma nota musical for tocada por dois instrumentos harmônicos distintos, a freqüência fundamental será praticamente a mesma, da mesma forma como os harmônicos. No entanto, à f_0 e a cada harmônico h_j estarão associados valores distintos de amplitude, de forma que o som será composto por intensidades

individuais diferentes, podendo até mesmo ser nulas em alguns harmônicos, como se pode observar na Figura 4, que apresenta o espectro de freqüência de uma mesma nota A (440 Hz) para um piano e um *piccolo*, (a) em escala linear, e (b) em escala logarítmica.

Não há notações para o timbre. Esse só pode ser especificado no sistema de notação musical tradicional a partir de instruções relacionadas a outros atributos (combinações de alturas; indicações de intensidade), ou por instruções verbais. A proposta inicial deste projeto limita-se a tentar descobrir o instrumento a partir de seu conteúdo espectral, não levando em consideração alterações devido ao modo de se tocar o instrumento, apesar de poderem ocorrer variações em função da interpretação. Mais informações sobre a relação entre variações de timbre e o conteúdo espectral podem ser encontradas em [8].

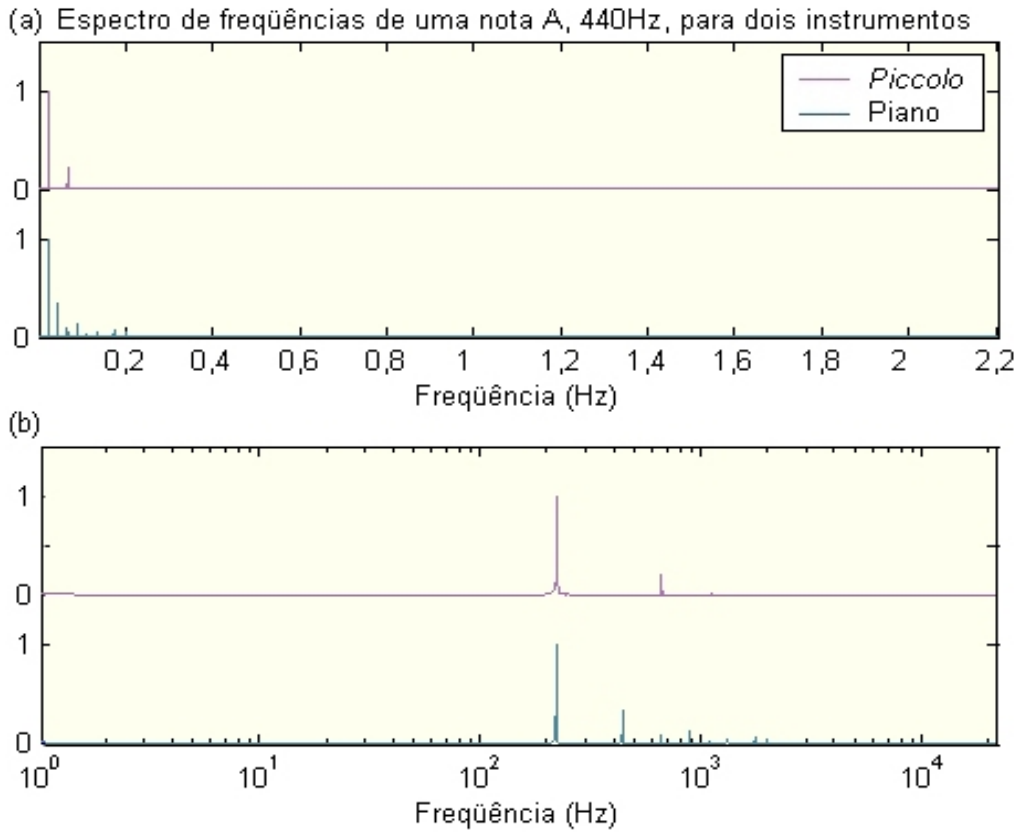


Figura 4 – Espectro de freqüência de dois instrumentos (a) em escala linear e (b) em escala logarítmica. Para uma mesma nota tocada por dois instrumentos, o espectro de freqüência tem distribuições diferentes de intensidade. Enquanto o espectro do piano apresenta mais de 10 harmônicos, o do *piccolo* possui apenas 3 picos expressivos, estando os demais ausentes ou com amplitudes quase desprezíveis.

2.1.2 Características no decorrer do tempo

As formas de onda apresentadas Figura 2 exibem eventos que ocorrem na ordem de 10^{-2} segundos, espaços de tempo muito curtos. Ao se observar um som, desde seu início até o seu fim, percebe-se que o timbre está associado também a como ele se manifesta no decorrer do tempo, com a maneira como sua intensidade está distribuída. A Figura 5 apresenta o comportamento no decorrer do tempo para uma mesma nota tocada por diferentes instrumentos.

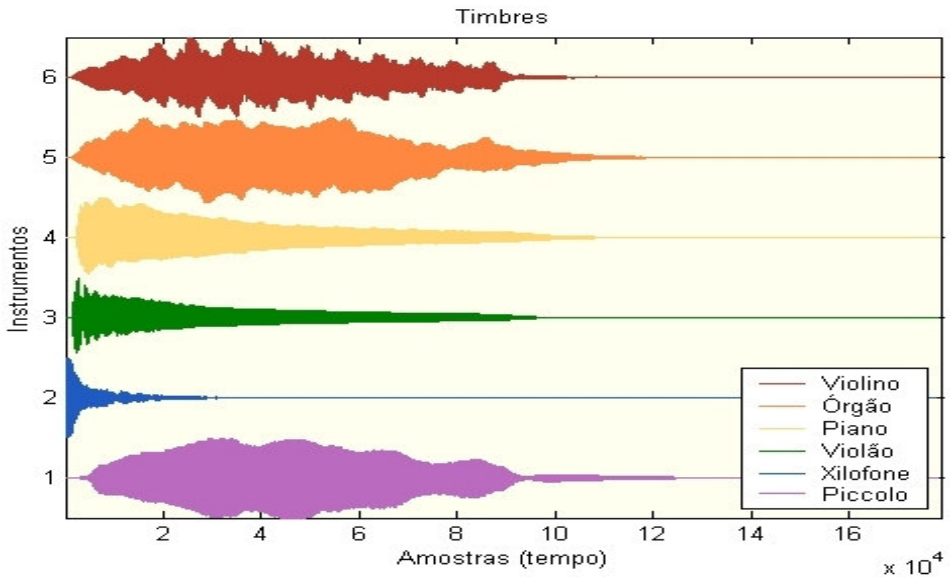


Figura 5 – Para uma mesma nota, com a mesma duração prevista, o som se comporta de maneiras distintas. Alguns sons concentram a intensidade em seus instantes iniciais e logo depois tendem a zero, como o xilofone, o violão e o piano. Outros apresentam intensidades mais distribuídas, como o *piccolo*, o violino e o órgão.

Há inúmeras formas de se observar esta variação. Uma delas é a utilização de **envoltórias**. Um dos tipos mais comuns de envoltórias é a de **amplitude**, que fornece o molde da variação intensidade de um sinal sonoro (volume ou energia) através do tempo. Para se obter a envoltória de amplitude de um sinal, pode-se aplicar um filtro passa-baixas, uma vez que são as baixas freqüências que geralmente dão o formato de um sinal.

A primeira parte de uma envoltória é chamada de ataque. Para uma envoltória de amplitude, isso significa quanto tempo demora para ir do estado de silêncio para o

volume máximo. Instrumentos de percussão possuem tempos de ataque curtos. A seção seguinte é denominada de decaimento, que é tempo em que um som demora para, a partir de sua intensidade máxima, atingir um certo grau de intensidade, que se manterá constante por alguns instantes, no que é chamado também de nível de sustentação, denominando a seção seguinte. A parte final de uma envoltória é o *release*, que é o tempo em que se gasta para um som sair do estado de sustentação para um estado de silêncio. A Figura 6 apresenta as partes de uma envoltória de uma nota tocada por um *piccolo*.

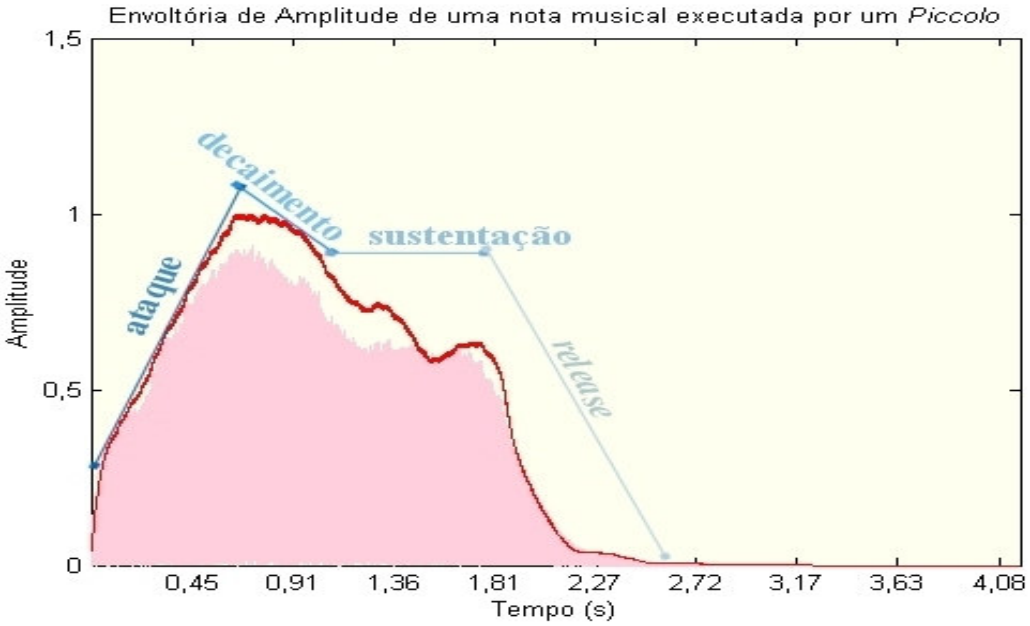


Figura 6 - Partes de uma envoltória de uma nota de piccolo. A primeira parte é o ataque, seguido pelo decaimento. Após, o som se mantém estável por um tempo de sustentação, até a fase de *release*, no qual se retorna ao silêncio.

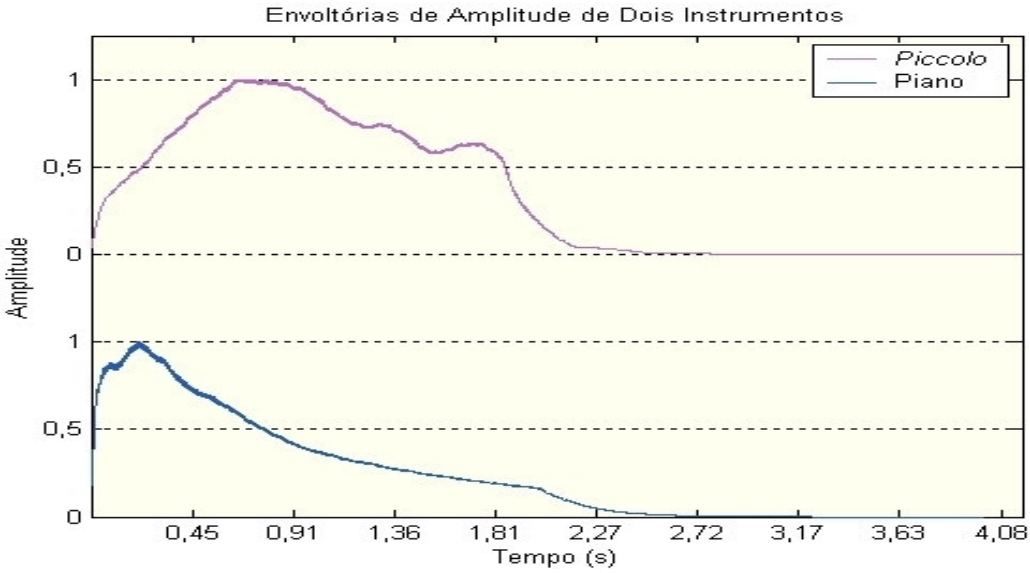


Figura 7 - Para uma mesma nota, as envoltórias de amplitude de dois instrumentos são diferentes. O piano possui ataque curto, período estável e queda longa, sua energia se concentra nos primeiros instantes do som, uma vez que resulta da batida do martelo na corda. Já o *piccolo*, por ser um instrumento controlado pelo sopro, possui ataque lento, quase linear, e mantém sua energia por um período mais longo (no caso, a amplitude é maior que 50% do seu valor total durante aproximadamente 1,5 segundo), decaindo rapidamente.

Diferentes sons possuem diferentes envoltórias de amplitude. Instrumentos como o *piccolo* e o piano possuem envoltórias bem características. Na Figura 7, observa-se que o piano tem ataque curto e queda longa, uma vez que resulta da batida do martelo na corda; já o *piccolo*, sendo um instrumento de sopro, tem ataque mais lento (o seu som não é percussivo, como o do piano), período estável de duração variável e uma queda curta. Na Figura 5, apresentada na página 16, nota-se que o xilofone, um instrumento percussivo, tem um ataque muito rápido, um decaimento igualmente rápido, um período mínimo de sustentação, até finalmente se silenciar.

2.1.3 Seqüência de sons

A música é formada por uma sucessão de sons, associados a notas musicais. Ao se juntar um ao outro, seqüencialmente, para se formar uma composição, determina-se que possuam duração e intensidade distintas.

Na notação musical, existe uma série de marcações diferentes para indicar a duração de cada nota, em relação a um referencial, chamado de **tempo**. No entanto, a duração real de uma nota, ou seja, aquela medida em segundos, depende do

compasso e do andamento (o grau de velocidade) utilizado. Também é possível que haja pausas dentro da música, intervalos em que nenhum som é produzido, e igualmente dependentes do compasso e do andamento.

O **compasso** é uma forma de ordenar e dividir em grupos os sons de uma composição musical, baseando-se em suas batidas e repousos. Alguns estilos musicais tradicionais já estão associados a um determinado compasso: a valsa, por exemplo, tem o compasso ¾, o que significa três semínimas por compasso. Os compassos facilitam a execução da música, uma vez que definem a unidade de tempo, o pulso e o ritmo.

O **ritmo** resulta de movimentos que se repetem a intervalos regulares. Na música, trata-se de um acontecimento sonoro que se repete com regularidade temporal, sendo uma forma de ordenação dos sons de acordo com padrões musicais estabelecidos. É resultado da variação de duração e de acentuação de uma série de sons ou eventos, estando, portanto, relacionado com a repetição de batidas (daí normalmente associados a instrumentos de percussão), ou com a variação da intensidade tocada. Por exemplo, um compasso ternário de uma valsa sempre terá o primeiro tempo mais forte que os demais. Essa diferença de intensidade fornece sensação de ritmo.

2.2 Domínio do tempo e domínio da freqüência

O seqüenciamento de notas musicais e o ritmo estão associados ao domínio do tempo. Nele, a representação do som se dá pela variação de sua amplitude instantânea ao longo do tempo.

Apresentaram-se algumas das características do som que permitem descrevê-lo e, a partir disso, modelá-lo e identificá-lo. Algumas das informações serão estudadas no domínio do tempo. Outras serão vistas no domínio da freqüência, no qual a representação se dá, em um determinado instante de tempo, pelos valores de amplitude relativos a cada uma de suas freqüências componentes.

A ordenação e o seqüenciamento de uma música devem ser analisados no domínio do tempo, uma vez que estão estritamente associados a elementos temporais, como o ritmo.

Da mesma forma, a freqüência fundamental do som, seus harmônicos, seu timbre e sua intensidade geral, serão analisados do domínio da freqüência. A intensidade geral depende da relação das intensidades de cada harmônico. O timbre informará o instrumento que foi utilizado. Necessita-se, assim, descobrir a freqüência fundamental daquele som, porque a ela estará associada a nota executada naquele trecho. Esta última relação na música ocidental será reapresentada no Capítulo 5.

2.3 Conclusões

O objetivo deste capítulo era agrupar os principais conceitos a respeito de música que serão aplicados neste projeto. Desta forma, visava-se a oferecer o conhecimento mínimo necessário para a compreensão do sistema. Portanto, foi apresentado o embasamento teórico necessário para entender amplitude, freqüência fundamental, harmônicos, definições relacionadas à transcrição musical, bem como se identificou onde serão buscadas tais informações.

O próximo capítulo discutirá a arquitetura do sistema, os requisitos básicos e como foi feita a modularização do mesmo.

Capítulo 3 - Visão Geral

É usual a idéia de se quebrar um problema muito grande em várias etapas, a fim de facilitar a sua resolução. Na área de programação, isso significa modularizar ao máximo um código, separando-o em rotinas, funções e objetos que desempenhem atividades específicas. Assim, a partir do momento em que se sabe exatamente o que aquele grupo de linhas executa, pode-se passar a chamar aquela função e trocá-la, eventualmente, para compará-la com novas versões. Essa idéia de modularização é muito aplicada em grandes sistemas que necessitam de constante atualização.

Desde o início do desenvolvimento deste projeto até a versão aqui apresentada, foram observadas várias mudanças e incrementos de novos conceitos na área de transcrição musical. Isso fez com que o objetivo do trabalho fosse sendo revisto aos poucos. O que era mais importante? Um sistema fechado e razoavelmente funcional não tolerante a mudanças e que seria, portanto, arquivado; ou um sistema

remodelável, que perderia em desempenho ao admitir mais falhas, mas que pudesse ser mais facilmente atualizável? Neste projeto, optou-se por configurar um sistema que fosse facilmente atualizável, desmembrado em funções mais simples e bem definidas, que oferecessem um bom suporte para quem vier a trabalhar futuramente nesta área.

Neste capítulo, apresenta-se a visão geral deste sistema reconhecedor de notas musicais em trechos polifônicos. Visando a uma abordagem lógica, na Seção 3.1 analisa-se um projeto de sistema monofônico, mais simples, de forma a se compreender os principais mecanismos envolvidos neste projeto. Na Seção 3.2, são acrescidas as funções necessárias para atender ao caso polifônico, o objetivo final. Por fim, a Seção 3.3 resumirá o capítulo e discutirá os efeitos da modularização.

3.1 Caso monofônico

Como som monofônico, entende-se um único instrumento tocando uma única nota. Não é necessário que seja um timbre monótono (ou seja, uma senóide pura). Pelo contrário, busca-se perceber como um som é composto por suas freqüências fundamental e seus parciais harmônicos, quais as informações mapeadas e analisadas, e como o fluxo de informações se dará.

3.1.1 Geração de um banco de notas

Por definição, para que se possa reconhecer algo, é necessário já possuir prévio conhecimento a seu respeito. Por exemplo, para se identificar um objeto, por exemplo, é necessário saber como ele é; e, para isso, precisam-se definir alguns parâmetros. Uma cadeira pode ser descrita pela quantidade de pés, se possui assento, encosto, o tipo do material, entre outros. Analogamente, é requisito conhecer as notas musicais, sabendo quais as características que definem uma nota e quais informações são necessárias.

Pode-se dizer que o mais importante a respeito de uma nota é:

- sua freqüência fundamental, f_0, que está associada ao um tom, representado pela nota em si; e

- seus parciais harmônicos e as relações de amplitude entre eles, ou seja, como a energia está distribuída em seus harmônicos, comparados à intensidade da fundamental. Isso, além de definir o perfil (timbre) do instrumento, oferece informações cruciais para a detecção em sons polifônicos.

A Figura 8 apresenta o bloco esquematizado do treinamento de um reconhecedor monofônico.

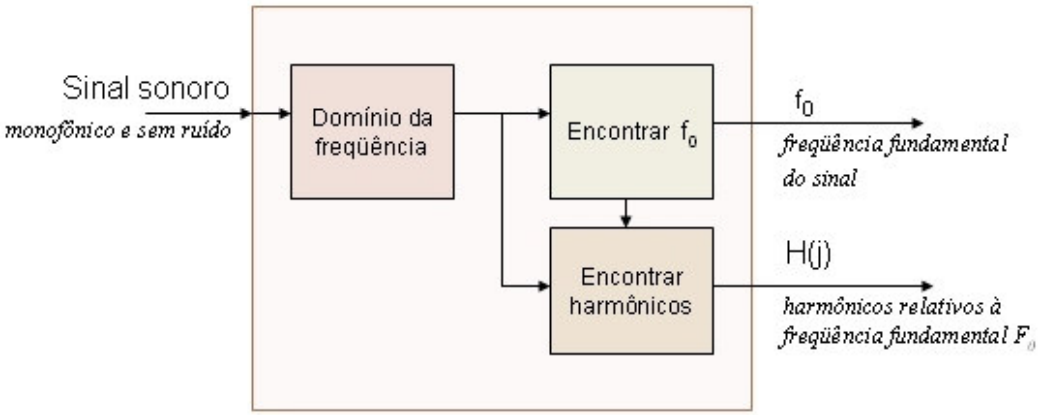


Figura 8 - Bloco esquematizado de treinamento de um reconhecedor monofônico.

Desta forma, tem-se um bloco denominado "treinamento" responsável por extrair de um sinal monofônico ideal (sem ruído) as informações de freqüência fundamental e as relações de amplitude entre seus harmônicos. Isso é feito normalizando a amplitude de f_0 em 1, mantendo a devida proporção às demais, de forma a se obter uma quantificação unitária das demais componentes do sinal. Para isso, faz-se necessário extrair o espectro do sinal e analisar seus picos na freqüência. Para um som monofônico, o comportamento será semelhante ao apresentado na Figura 3, no capítulo anterior. A maneira como esses dados são extraídos e tratados será explicada no Capítulo 7.

As informações que caracterizam cada nota serão armazenadas em um arquivo ou em um banco de dados, para serem consultadas na fase de reconhecimento.

3.1.2 Reconhecimento

Para o reconhecimento monofônico, é necessária a existência de um módulo de extração de dados, análogo ao utilizado para o treinamento, que registrará as

informações referentes à freqüência fundamental e aos harmônicos. Tal módulo apresenta-se esquematizado na Figura 9.

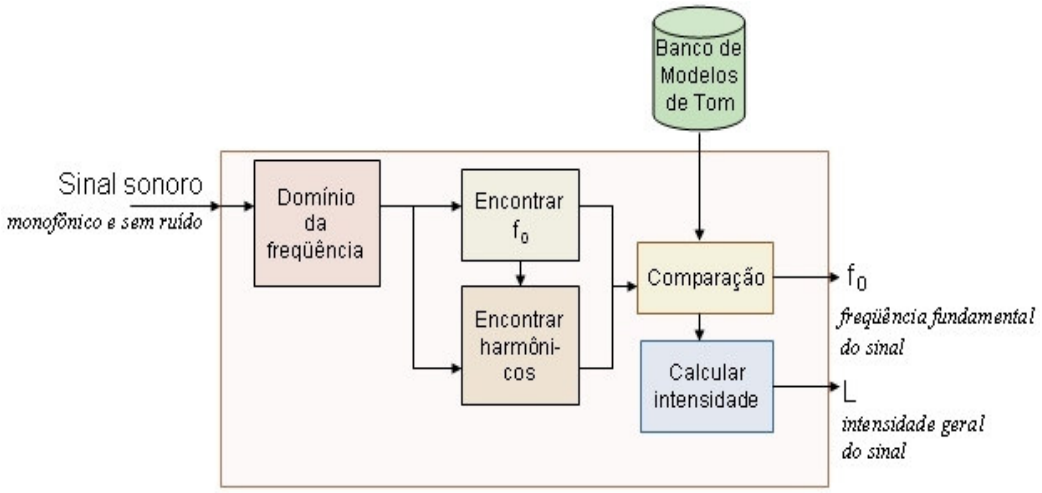


Figura 9 - Bloco esquematizado de um reconhecimento monofônico.

Tais informações serão, então, levadas a um módulo comparador responsável por identificar a nota mais próxima constante no banco, fornecendo, portanto, os dados de saída do sistema: freqüência fundamental, instrumento (de acordo com o timbre) e intensidade da nota.

Necessita-se passar o sinal sonoro a ser analisado para o domínio da freqüência, fazendo uso dos mesmos parâmetros que o caso de geração de modelos de tom. Pode-se, portanto, admitir que esta transformada seja um módulo em separado. Esse tipo de estudo deve ser minuciosamente feito para se entender a modularização do projeto. Um resumo com as principais funções será apresentado no final deste capítulo.

3.1.3 Divisão de um segmento musical em arquivos menores de uma nota apenas

A música é uma seqüência de notas. Portanto, a entrada do sistema será como tal. Desta forma, faz-se necessário um bloco que deverá dividir tal seqüência em arquivos menores, nos quais conste apenas uma nota, o bloco de segmentação.

Neste projeto, o princípio do funcionamento utilizado é a detecção de início de cada nota (*sound onset detection*), de acordo com a variação de energia do sinal. Isso será melhor explicado no Capítulo 4, referente ao assunto.

3.1.4 Módulo integrado

A Figura 10 apresenta o módulo integrado para o treinamento e a identificação de tons em se tratando de um sinal sonoro monofônico.

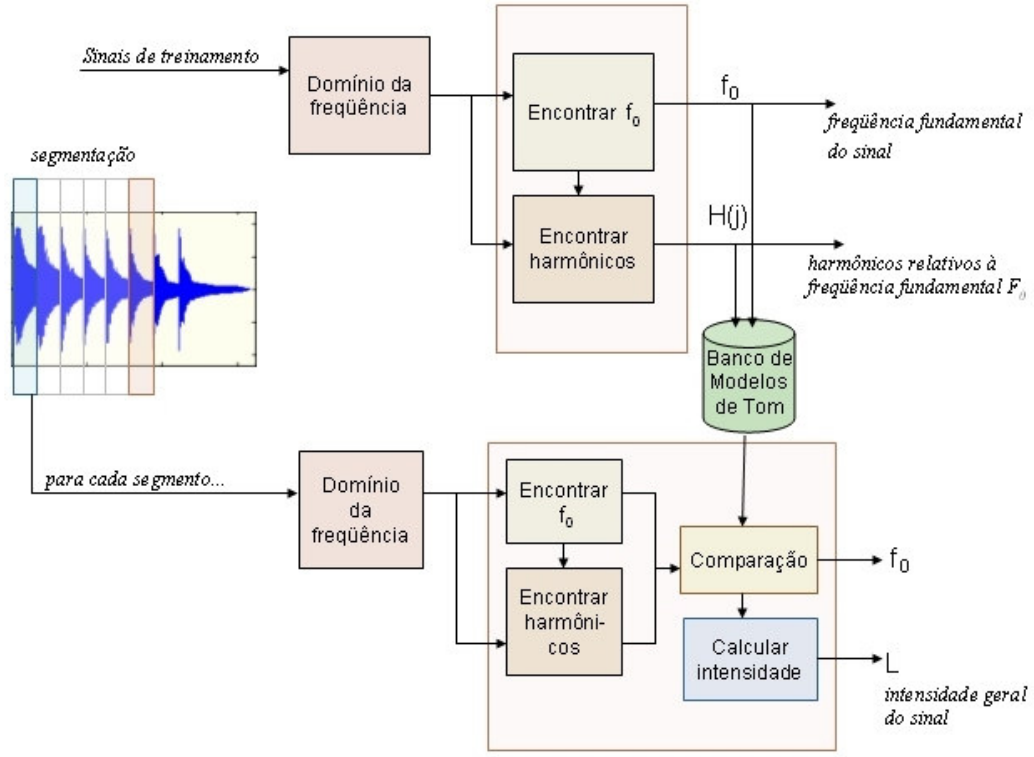


Figura 10 - Módulo integrado para a criação de modelos de tom (treinamento) e identificação (reconhecimento) de tons presentes em um sinal sonoro monofônico.

Basicamente, o princípio de funcionamento dos sistemas, seja monofônico, seja polifônico, é o mesmo. O que difere é o modo como serão tratadas as informações, sua extração e validação, uma vez que o sistema polifônico deve levar em consideração inúmeros fatores, como será visto na seção a seguir.

3.2 Caso polifônico

Existe uma série de bons métodos de medir freqüências, amplitudes e fases de parciais senoidais em um sinal. A grande questão para resolver uma mistura de sons

harmônicos é que estes harmônicos estendem-se por uma vasta faixa de freqüências, e faixas de sons diferentes costumam se sobrepor, fazendo com que seus diversos harmônicos também se sobreponham. Observa-se que, na música ocidental, as notas estão organizadas em escala logarítmica, em que a freqüência fundamental de uma nota *k* é definida por:

$$f_{0_k} = 440.2^{k/12} \text{ Hz} \tag{3.1}$$

Desta forma, identificam-se dois problemas sérios:

- Como as séries harmônicas de sons diferentes se estendem por bandas de freqüência comuns, é muito difícil atribuir os harmônicos às suas verdadeiras freqüências fundamentais;
- Se as parciais senoidais se sobrepõem, ou seja, possuem uma mesma freqüência, as envoltórias e as fases de duas senóides sobrepostas não podem mais ser deduzidas pelas suas somas.

Estas são as razões fundamentais por que conteúdos de sinais polifônicos não podem ser resolvidos pela aplicação direta dos algoritmos desenvolvidos para sinais monofônicos [9].

3.2.1 Processo de criação de modelo de tom

O processo de criação de modelo de tom tem o comportamento análogo ao apresentado no processo de geração de banco de notas, na Seção 3.1.1.

A entrada do sistema é um conjunto de sinais de treinamento, monofônicos, em que todas as notas do instrumento são tocadas uma única vez, em separado a fim de representar o alcance total de tons que podem ser produzidos por ele.

A seguir, é feita a extração de dados desta entrada. Primeiro o sistema procura pela freqüência fundamental f_0 existente naquele sinal. A seguir, ele vasculha os harmônicos, multiplicando o valor da freqüência por números inteiros de 1 até *J*, que seria a quantidade máxima de harmônicos a ser utilizada para descrever o som. Finalmente, a informação da freqüência e os respectivos valores de amplitude em relação aos seus harmônicos são armazenados no banco de modelos de tom. Uma vez

armazenados os dados, encerra-se este processo, que é executado apenas uma vez por instrumento.

3.2.2 Módulo de criação de um núcleo de filtro

As principais informações a serem obtidas no processo de caracterização de uma nota são as freqüências fundamentais e suas respectivas intensidades gerais. Para que se obtenha a intensidade geral de um sinal, precisa-se efetuar um certo cálculo a partir das intensidades de cada um dos harmônicos detectados.

O objetivo deste módulo é gerar um filtro capaz de extrair alguma característica de um sinal sonoro na observância de sons interferentes. O princípio básico do funcionamento deste filtro é identificar características singulares de cada uma das diferentes notas e, a partir deste perfil único, poder identificar uma nota mesmo quando houver a sobreposição de inúmeros harmônicos.

Este filtro será gerado apenas uma vez e utilizado na extração de dados de cada nota em cada segmento polifônico. Seus parâmetros serão explicados no Capítulo 6. Estão interligados com a quantidade J de harmônicos a serem contabilizados para a representação do som e dependem da tolerância que se espera que o sistema tenha.

3.2.3 Processo de reconhecimento

A entrada do processo de reconhecimento deve ser o espectro de freqüências de cada mistura polifônica a ser transcrita. Este bloco também deve ter acesso aos núcleos de filtro, obtidos no processo mencionado anteriormente, e aos modelos de tons, armazenados em um banco ou arquivo. A saída do processo consiste em freqüências fundamentais e intensidades respectivas de cada tom presente em cada segmento da peça musical analisada. A Figura 11 apresenta a organização esquematizada do módulo de reconhecimento.

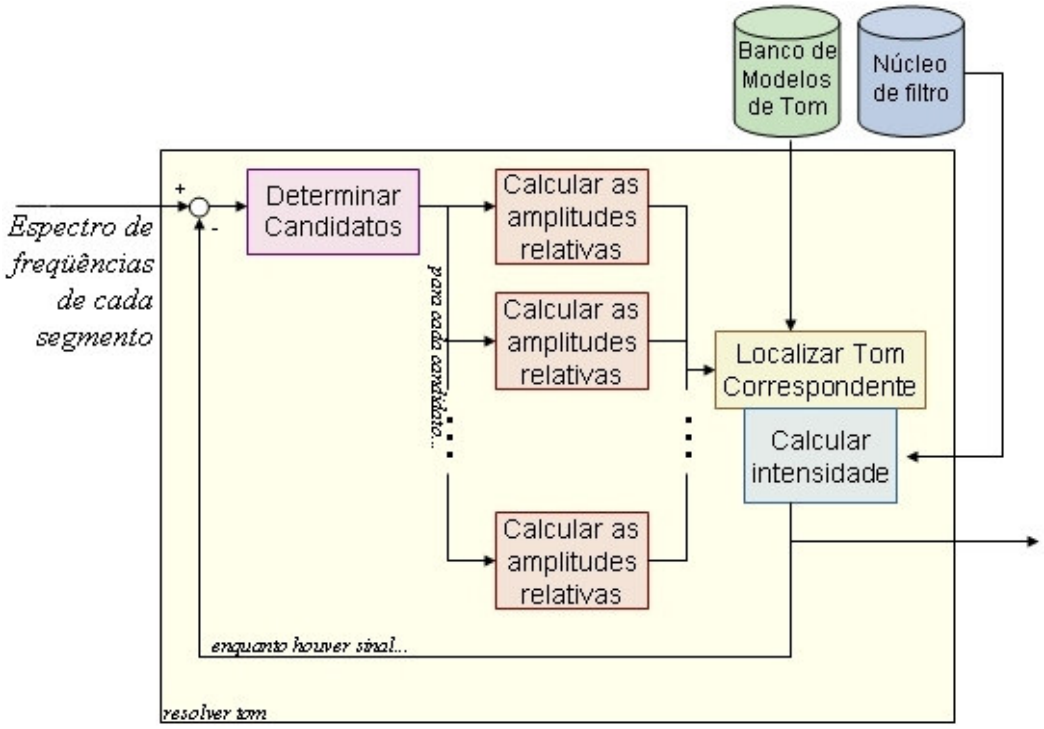


Figura 11 -Módulo de reconhecimento de notas musicais.

O coração do processo é uma sub-rotina denominada *resolver tons,* que decompõe o segmento de sinal em um conjunto de tons, extraindo suas freqüências fundamentais e intensidades gerais, da seguinte forma:

1) Detecção de todas as candidatas potenciais a serem freqüências fundamentais no espectro;
2) Todas as candidatas devem ser tratadas em ordem ascendente de freqüência e avaliadas; delas são obtidas as intensidades, sendo elas depois subtraídas do espectro, seguindo para a próxima candidata. Sons que realmente existirem terão intensidades significativas e claras, sendo exportados para a saída do sistema.

3.2.4 Módulo integrado

Os processos apresentados nos itens acima compõem os módulos básicos do reconhecedor de sons polifônicos. Com isso, tem-se a visão geral do projeto, com dois blocos atuantes, o de treinamento e o de reconhecimento.

Para o bloco de reconhecimento, a entrada é uma peça musical, a ser segmentada temporalmente em trechos polifônicos menores com a duração de uma única nota. A seguir, é feita a varredura destes sons e cada um destes segmentos é resolvido em suas freqüências fundamentais, instrumentos e intensidades gerais. Junto a isso, deve-se organizar a última saída do sistema, um segmento de tempos de início de nota, ou seja, a informação temporal de quando cada nota se inicia e, por conseguinte, suas durações. Na Figura 12, pode-se observar o desenho esquemático da integração.

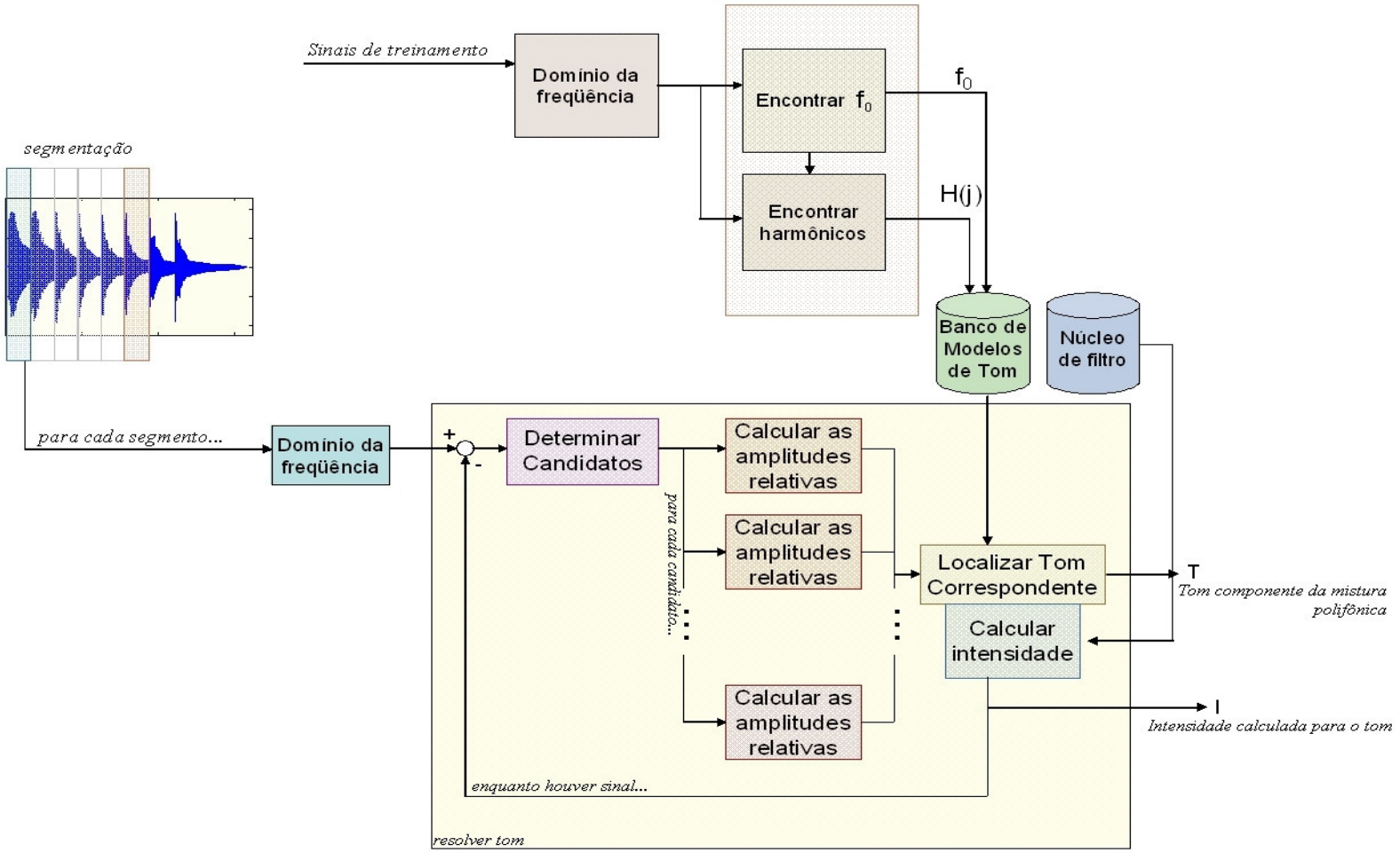


Figura 12 - Módulo integrado do sistema polifônico.

3.3 Conclusões

Neste capítulo foi vista a estrutura proposta do sistema, modelando-a passo a passo, identificando, assim, as necessidades em termos de funcionalidades e o fluxo de informações. A partir disso, é proposto o protótipo dos blocos integrantes e como eles estarão interligados.

Basicamente, o sistema proposto é dividido em dois blocos estruturais maiores, o de treinamento e o de reconhecimento. O treinamento se dá ao fornecer ao sistema o som de cada um das notas do instrumento, para caracterizá-lo. O som é submetido a uma função de transformada, obtendo-se o seu espectro de freqüências, a partir do qual se é capaz de calcular a freqüência fundamental e a distribuição de energia pelos seus harmônicos. Introduziu-se uma constante, *J*, que é a quantidade de harmônicos utilizados para representar o som. O valor associado a *J* será discutido no Capítulo 6. Após a caracterização dos tons possíveis deste instrumento, os valores obtidos são armazenados em um banco de dados ou arquivo, e ficarão disponíveis para a consulta pelo módulo reconhecedor.

O processo de reconhecimento se inicia pela segmentação temporal da peça musical introduzida no sistema. Cada um dos segmentos resultantes será submetido à transformada para o domínio da freqüência e se inicia a busca por uma nota. Rastreiam-se os picos e cada parcial harmônica relativa a um candidato é considerado como indicativo da existência daquela nota. Calcula-se a intensidade geral do candidato, e, se expressivo, é levado em consideração. Enquanto houver a possibilidade de um som existir, este procedimento é refeito em ordem ascendente do valor da freqüência fundamental.

O resumo dos métodos empregados em cada um dos módulos é descrito na Tabela 3, a seguir, onde foram usados os rótulos “T” para treinamento e “R” para reconhecimento.

A saída do sistema será, portanto, composta de: os índices que segmentam o som, temporalmente, em pedaços de duração de uma nota ou um conjunto de notas simultâneas, as notas musicais que compõem cada fragmento e suas intensidades, e o tipo de instrumento.

Nos próximos capítulos, seguirá uma descrição detalhada e o algoritmo proposto para a execução de cada um destes módulos.

Tabela 3 - Resumo das principais funções e parâmetros de entrada e saída do sistema.

Módulo	Domínio	Bloco	Função	Entrada	Saída
Segmentação Temporal	Tempo	-	Dividir uma seqüência de sons em um conjunto unitários de sons simultâneos	Seqüência de notas (a peça musical a ser transcrita)	Vetor contendo índices dos inícios de nota, e uma série de arquivos *waves* com os segmentos observados
Transformada	Tempo e Freqüência	-	Obter o espectro de freqüência de um sinal sonoro	Segmento de som	Um vetor contendo o espectro de freqüência do sinal
Encontrar f_0	Freqüência	T	Rastrear o espectro a procura do primeiro pico e analisá-lo	Espectro de freqüência	Posição no espectro e respectivo valor de freqüência do mais indicado candidato a f_0
Encontrar Harmônicos	Freqüência	T	A partir do valor de f_0, localizar as freqüências múltiplas (harmônicos) e seus respectivos valores de amplitude	Espectro de freqüência e posição/valor do mais indicado candidato a f_0	Vetores com os valores de amplitude, freqüência e posição dos harmônicos referentes a um som f_0
Determinar candidatos	Freqüência	R	Rastrear o espectro a procura dos primeiros picos e analisá-los.	Espectro de freqüência	Posição no espectro e respectivo valor de freqüência do mais indicado candidato a f_0
Calcular amplitudes relativas	Freqüência	R	A partir do valor da candidata a f_0, localizar as freqüências múltiplas (harmônicos) e seus respectivos valores de amplitude	Espectro de freqüência e posição/valor do mais indicado candidato a f_0	Vetores com os valores de amplitude, freqüência e posição dos harmônicos referentes a um candidato f_0
Localizar tom correspondente	Freqüência	R	Comparar os vetores de amplitude obtidos com os vetores de amplitude armazenados no banco de tons que tiverem freqüências fundamentais muito próximas às das candidatas	Vetores com os valores de amplitude dos harmônicos referentes a candidatos f_0 e Vetores com os valores de amplitude dos harmônicos do banco	f_0 mais provável
Calcular intensidade	Freqüência	R	Calcular a intensidade geral referente a f_0 mais provável	f_0 mais provável e vetor com os valores de amplitude dos seus harmônicos	Intensidade geral do som

Capítulo 4 - Segmentação Temporal: Detecção de Início de Nota

O princípio do funcionamento do bloco de segmentação utilizado neste projeto é a detecção de início de cada nota, e para isso é feita a análise da variação de energia (intensidade) do sinal. A idéia principal é tentar identificar qual o momento em que uma nota (ou um conjunto de notas simultâneas) se inicia, fazendo uso das informações contidas naquele sinal.

Na Seção 4.1, descreve-se o funcionamento geral do detector, apresentando o fluxo de informações, as operações realizadas e os parâmetros ajustados. A Seção 4.2 apresenta os testes de validação do procedimento, fazendo uso de sons monofônicos

e polifônicos. Finalmente, na Seção 4.3, há o resumo dos procedimentos, a avaliação final dos resultados e a conclusão.

4.1 Funcionamento geral do detector

A motivação deste bloco é identificar onde um som começa. Mesmo em uma única nota musical, os componentes de freqüência começam em instantes de tempo diferentes, alguns antes, alguns depois. Assim, separa-se o sinal em diferentes bandas de freqüência, que são analisadas separadamente.

Desta forma, um banco de filtros é projetado para dividir o sinal musical em sete bandas de freqüências distintas, sendo 127, 254, 508, 1016, 2032 e 4064 Hz os valores de corte.

4.1.1 Banco de filtros

Para muitos tipos de bancos de filtros, o sinal sonoro resultante possui percepção rítmica praticamente igual àquela do sinal de música original. Isso quer dizer que, em cada faixa, as formas serão semelhantes às do sinal composto. Para se ter uma idéia, com apenas quatro bandas de freqüência, o pulso e as características de métrica do som original são facilmente reconhecíveis [9]. No entanto, esta similaridade só é admissível para sons percussivos: tais simplificações não podem ser feitas, por exemplo, para considerável parte de músicas clássicas, muitas vezes interpretadas por instrumentos não percussivos, como violino, órgão e flauta.

Neste processo, foram usados filtros elípticos de sexta ordem, cujos coeficientes foram calculados pela função *ellip* do Matlab© na seguinte disposição:

- F_1: Passa-baixas com freqüência de corte 127 Hz;
- F_2: Passa-faixa com freqüências de corte 127 Hz (inferior) e 254 Hz (superior);
- F_3: Passa-faixa com freqüências de corte 254 Hz (inferior) e 508 Hz (superior);
- F_4: Passa-faixa com freqüências de corte 508 Hz (inferior) e 1016 Hz (superior);

- F_5: Passa-faixa com freqüências de corte 1016 Hz (inferior) e 2032 Hz (superior);
- F_6: Passa-faixa com freqüências de corte 2032 Hz (inferior) e 4064 Hz (superior);
- F_7: Passa-altas com freqüência de corte 4064 Hz;

Tais filtros estão apresentados na Figura 13. Os valores de freqüência de cortes são os mesmos citados em [9]. Não foi implementado nenhum tipo de tratamentos aos atrasos individuais gerados por cada um destes filtros elípticos.

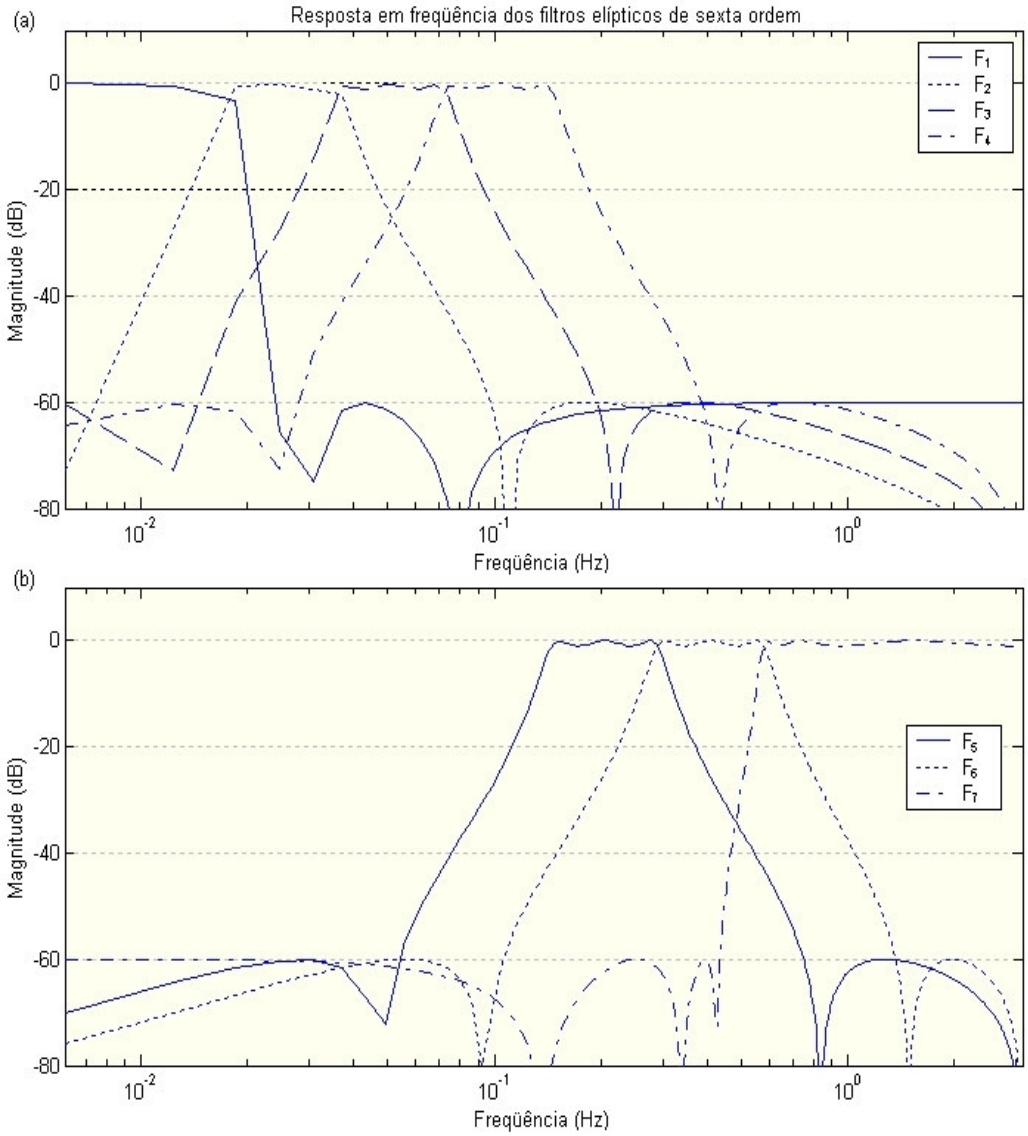


Figura 13 - (a) Filtros passa baixas F_1, passa faixa F_2, passa faixa F_3, e passa faixa F_4, (b) Filtros passa faixa F_5, passa faixa F_6, e passa altas F_7.

Na Figura 14, o sinal usado para exemplificar o método é um pequeno trecho de uma peça de piano, instrumento propício, uma vez que seu timbre apresenta uma grande quantidade quase pontual de energia, decaindo exponencialmente. Observa-se, então, que a quarta faixa de freqüência (de 508 a 1016 Hz), ainda é semelhante ao sinal original, podendo visualmente se identificar o início de cada uma das notas em ambos.

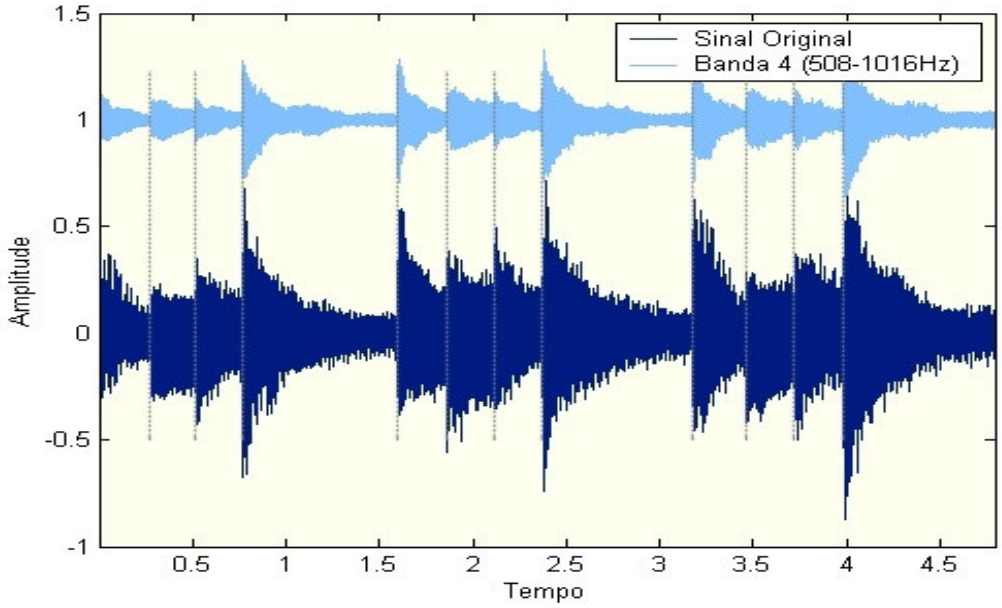


Figura 14 – Para um trecho polifônico tocado por um piano, o sinal sonoro resultante, com apenas os componentes de freqüência 508 a 1016Hz, possui percepção rítmica praticamente igual àquela do sinal de música original.

Na Figura 15, observa-se que, acima da freqüência de 4064 Hz (banda 7), quase não há componentes. Também há maior concentração de energia justamente nas bandas 2, 3, 4 e 5, que englobam as freqüências de 127 até 2032 Hz. As bandas 3 e 4 carregam a informação de início de notas mais clara. Isso significa que elas são boas candidatas a serem usadas como referência na detecção.

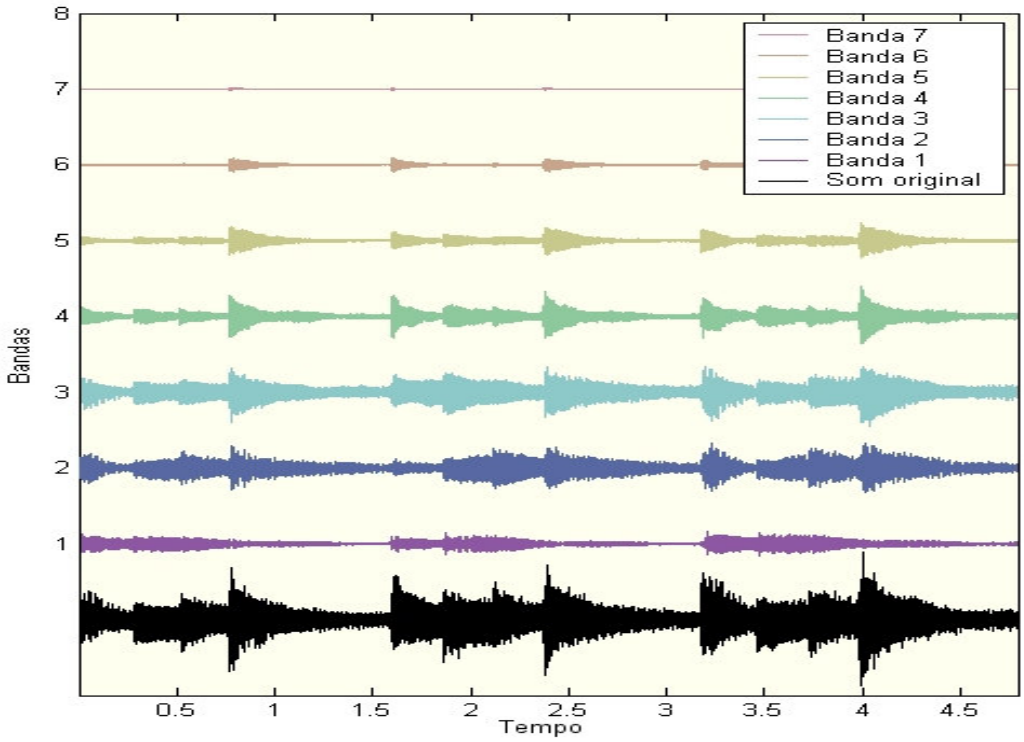


Figura 15 - O sinal original (em preto), após ser filtrado e separado nas sete faixas de freqüência trabalhadas. Observe que as bandas de freqüência 3 (254-508 Hz), 4 (508-1016 Hz) e 5 (1016-2032 Hz) são as que mais se assemelham ao sinal original, evidenciando os pontos de início de nota, ao terem mais energia concentrada naqueles instantes de tempo. Essa visualização só é possível por se tratar de um instrumento com alguma percussão, devido ao seu ataque curto.

4.1.2 Obtenção das envoltórias

Como visto no Capítulo 2, envoltórias são utilizadas para observar como um som muda no decorrer do tempo. E, para se obter a envoltória de amplitude de um sinal, pode-se pensar na aplicação de um filtro passa-baixas, uma vez que são as baixas freqüências que dão o formato ao sinal. Se as envoltórias de amplitude de cada uma das sete bandas forem obtidas, tem-se a variação da intensidade no decorrer do tempo e, sobre ela, efetuam-se cálculos diretos para a identificação do momento em que a nota se inicia.

Na aquisição das envoltórias, obteve-se o módulo das saídas de cada banda. Em seguida, calcularam-se as envoltórias de cada sinal, através da convolução com uma janela Half-Hanning de 50ms. A largura de 50ms foi proposta por [10], uma vez que executa quase a mesma integração de energia que o sistema auditivo humano, enfatizando as entradas mais recentes, mas mascarando modulação rápida.

4.1.3 Função diferencial de primeira ordem

A partir da envoltória, busca-se o ponto de início de um som. Para isso, é necessário analisar a variação da intensidade do som no decorrer do tempo. Portanto, é natural que se pense em calcular a função diferencial absoluta de primeira ordem, escolhendo-se o ponto em que a ascensão é máxima. No entanto, existe outro método mais eficiente para se identificar o início da nota.

Inicialmente, calcula-se a função diferencial de primeira ordem, anulando-se qualquer valor que se encontre abaixo de um determinado limite. Em seguida, divide-se este resultado pela função da envoltória, o que fornece uma análise quantitativa da variação de energia em relação ao nível do sinal absoluto. Isso é o mesmo que derivar o logaritmo da envoltória.

$$P = \frac{dy}{y}. \Rightarrow \int P = \int \frac{dy}{y} \Rightarrow \int P = \ln(y)$$

$$P = d(\ln(y)) \tag{4.1}$$

A função diferencial relativa de primeira ordem mostra-se mais eficiente para identificar o início da nota por duas razões [9]:

- Sons baixos poderiam demorar algum tempo para chegar ao ponto em que sua amplitude está em ascensão máxima, o que implicaria uma estimativa bastante atrasada em relação ao início da nota.
- A variação de energia que culmina no início de um som não é monótona; portanto há inúmeros máximos locais na função diferencial de primeira ordem próximo ao instante buscado.

A Figura 16 ilustra a comparação entre a envoltória da função derivada absoluta de primeira ordem (linha tracejada) e a envoltória da função derivada relativa de primeira ordem (linha sólida).

A função diferencial absoluta, apesar de não atender ao objetivo de se identificar o ponto inicial de uma nota, oferece uma boa medida para os valores de *proeminência*, que seria a força com a qual um evento de início de nota se destacaria

em relação ao sinal musical, e que dependeria de atributos como a freqüência do som iniciante, a sua mudança relativa de amplitude e a rapidez desta mudança.

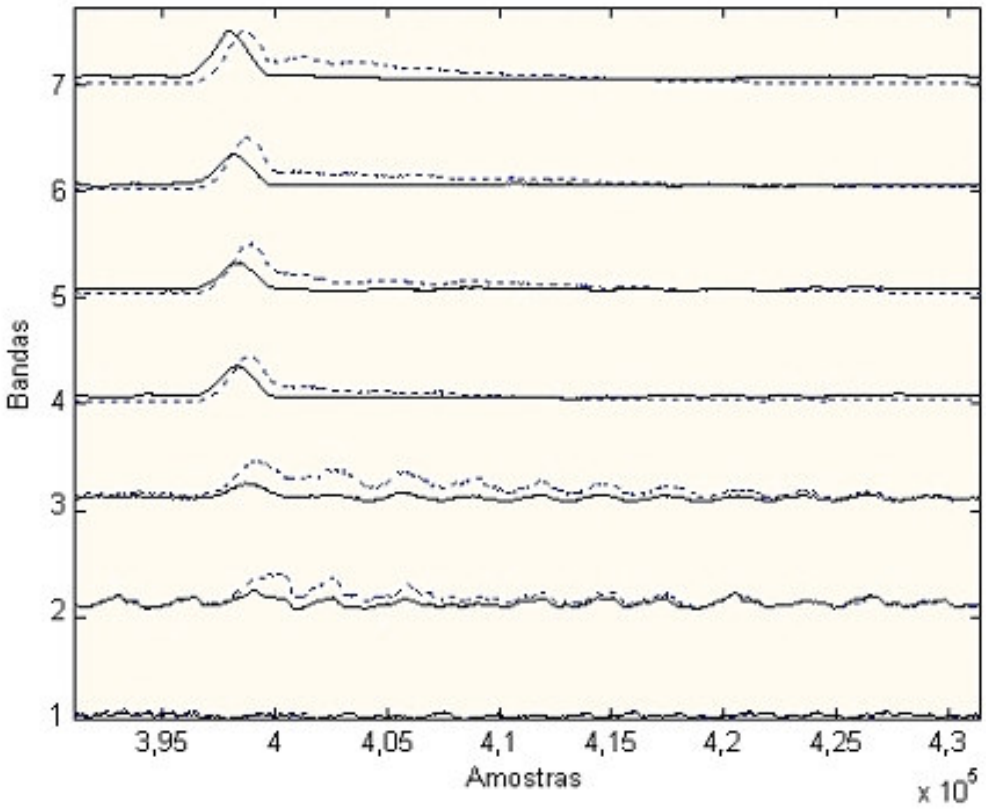


Figura 16 - Comparação entre a envoltória da função derivada absoluta de primeira ordem (função tracejada) e a envoltória da função derivada relativa de primeira ordem (linhas sólidas pretas) de uma nota musical de piano sintetizado por computador. Observa-se que o valor máximo da função relativa antecede o valor máximo da função absoluta, que estaria implicando uma estimativa atrasada do início do som. Fez-se uso da envoltória para que a diferença entre as duas funções ficasse visualmente clara para a comparação.

A Figura 17 ilustra derivada relativa de primeira ordem de um sinal de teste, dividido nas sete bandas propostas na Seção 4.1.1.

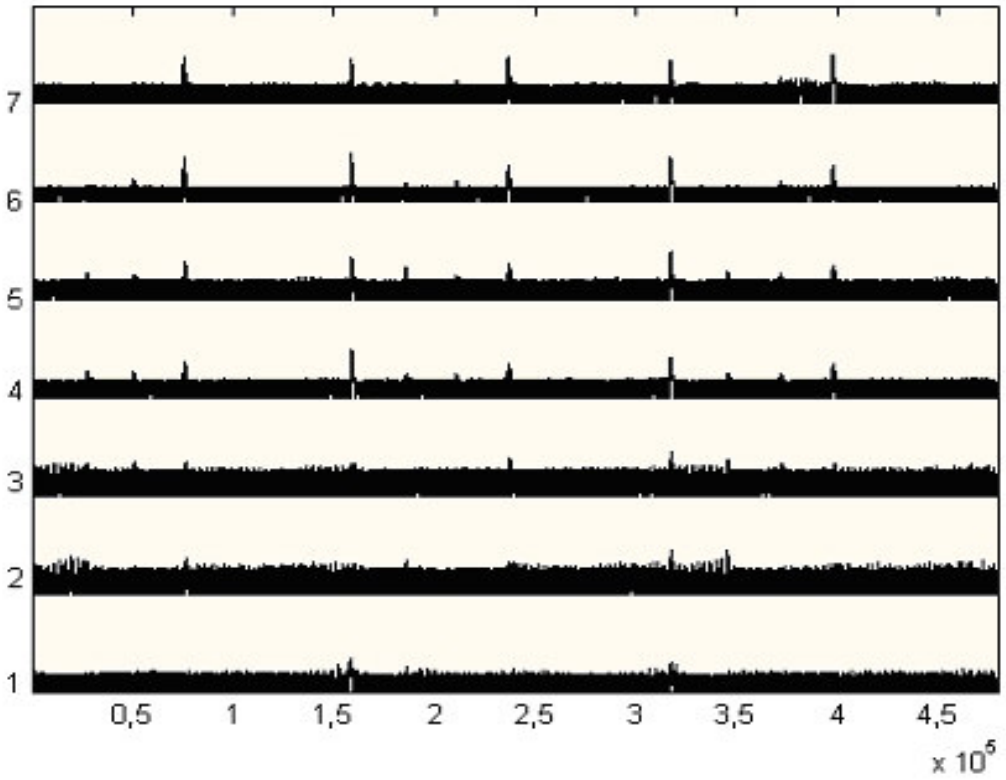


Figura 17 - Função derivada relativa de primeira ordem do sinal de teste (piano_01.wav), dividido nas sete bandas propostas. Se for feita a comparação com o sinal apresentado na figura 15, fica evidente que os máximos locais são os pontos em que o som se inicia em cada uma das bandas de freqüência. Observe também que são mais facilmente identificados nas bandas 4 e 5 (508-2032 Hz).

4.1.4 Descobrindo as posições de início em potencial

Procura-se, então, um conjunto de posições de início de nota em potencial, em cada uma das bandas, através de uma operação de escolha de pico, a partir da função diferencial relativa do sinal. Para isso, é preciso determinar um valor limite, para o qual se anulariam quaisquer pontos abaixo de seu valor.

Levantou-se a discussão do que seria um bom valor limite: se um valor estático ou se relativo ao sinal de entrada. Este último aparenta ser mais lógico, uma vez que os sinais sonoros podem ter diferentes intensidades e seria impossível encontrar um valor absoluto que satisfizesse a inúmeros sinais distintos.

O primeiro valor a ser cogitado é a média. No entanto, a média somente não se mostrou suficiente para distinguir alguns picos. Foram feitos extensos testes de ajuste fino para variados tipos de entrada a fim de se obter valores absolutos a serem

multiplicados à média, com o objetivo de atribuir pesos diferentes para distintas bandas de freqüência. As preocupações eram, basicamente, duas: permitir que os picos que indicassem o início de notas reais fossem exibidos e eliminar ao máximo picos que não fossem relacionados.

Conforme citado na Seção 4.1.1, as bandas de freqüência que oferecem informações rítmicas mais claras são a 3ª, a 4ª e a 5ª (ou seja, de 254 a 2032 Hz). As demais faixas oferecem poucos valores confiáveis, na maioria dos casos introduzem mais erros que notas verdadeiras. A escolha dos valores limite, para cada uma das bandas, foi realizada empiricamente. De um lado, busca-se identificar os picos verdadeiros. Por outro, também se deseja eliminar valores falsos. A Tabela 4 apresenta os valores-limite escolhidos.

Tabela 4 – Valores-limite escolhidos para a filtragem do sinal em cada uma das bandas. Valores abaixo destes limites serão zerados.

Banda	Valores-limite escolhidos
B1 (até 127 Hz)	3,8*média(B1)
B2 (127-254 Hz)	2,5*média(B2)
B3 (254-508 Hz)	1,8*média(B3)
B4 (508-1016 Hz)	1,6*média(B4)
B5 (1016-2032 Hz)	1,5*média(B5)
B6 (2032-4064 Hz)	1,8*média(B6)
B7 (acima de 4064 Hz)	2,0*média(B7)

A Figura 18 apresenta os picos escolhidos, ou seja, aqueles que estavam acima dos valores-limite apresentados na Tabela 4 para a filtragem do sinal nas diferentes bandas de freqüência, para o som de teste piano_01.wav.

Em geral, é preferível a detecção de picos inexistentes à negligência do início de sons presentes no sinal, uma vez que parciais de uma mesma nota apresentarão intensidades decrescentes no decorrer do tempo, o que poderia permitir o tratamento das mesmas no processo de reconhecimento. No entanto, com esta implementação, se forem admitidos valores mais tolerantes que os citados na Tabela 4, ou seja, se diminuirmos os pesos a serem multiplicados pelo valor da média, de forma a serem detectados todos os inícios de nota presentes, inclusive os de tons com intensidade

mais baixa, a quantidade de estimativas de início de nota obtida torna-se tão grande que não compensa as despesas computacionais envolvidas no pós-processamento.

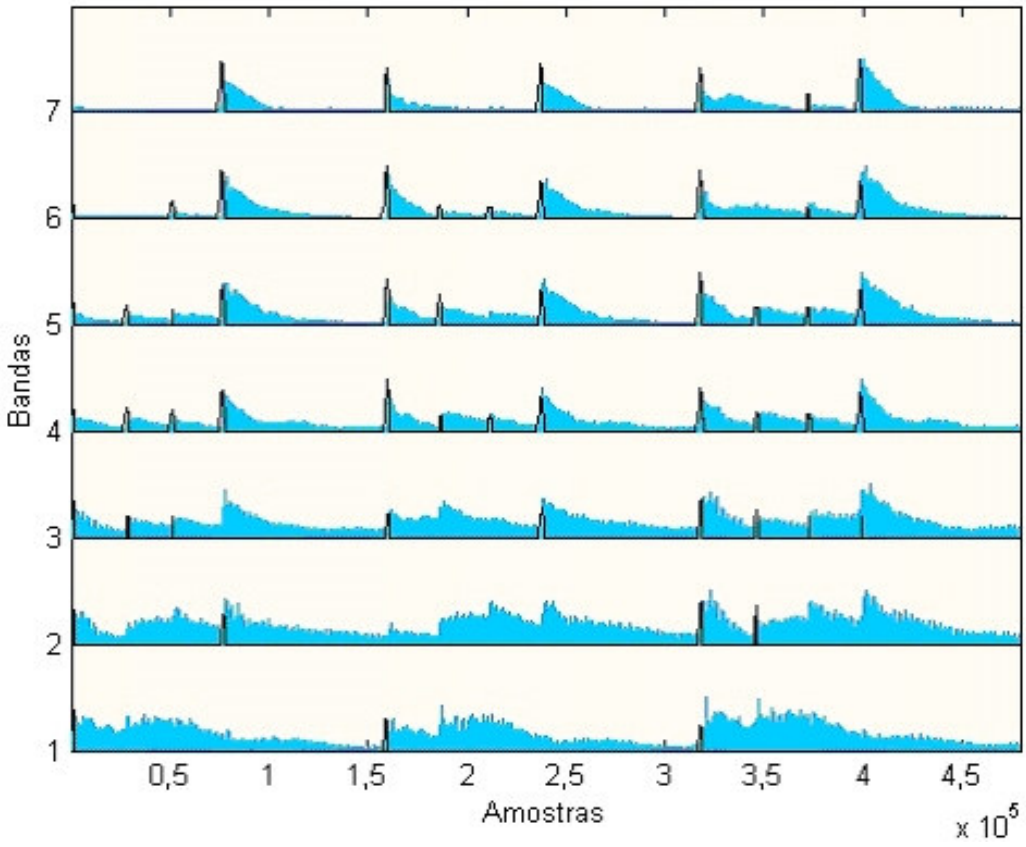


Figura 18 - Os picos selecionados das diferentes bandas de freqüência.

A partir de então, calcula-se a proeminência de cada candidato a início de nota, efetuando-se uma varredura até o próximo máximo relativo na função derivada **absoluta** e admitindo-se este valor como a proeminência daquele candidato.

4.1.5 Combinação dos resultados através das diferentes bandas

Na fase final, combinam-se os resultados das diferentes bandas de freqüência para revelar os tempos de início de nota e proeminências do sinal geral. Os valores de proeminência de cada banda são somados, gerando um único vetor, com diferentes candidatos de início de nota, uma vez que, em cada freqüência, a energia se inicia em tempos distintos, ainda assim próximos.

Em seguida, cada candidato é associado a um novo valor de proeminência, que corresponde à soma das proeminências dos candidatos dentro de uma janela de 50 ms ao redor deles. Esse valor é baseado nos estudos da percepção do ouvido humano, que seria capaz de distinguir sons que tivessem no mínimo 50 ms de intervalo de tempo entre eles [10].

Descartam-se as candidatas cujas proeminências estiverem abaixo de um certo limite. Neste desenvolvimento, a média deste novo vetor foi considerada como um bom valor de limite. A seguir, são descartadas as candidatas que estão perto demais (menos que 50 ms de distância temporal, ou seja, 2205 pontos se a freqüência de amostragem for 44,1 kHz) de um candidato mais apropriado. Dentre candidatas igualmente prováveis, porém perto demais umas das outras, a do meio é escolhida e as demais, abandonadas. As que sobrarem são aceitas como inícios de nota verdadeiros.

4.2 Testes de validação de procedimento

O procedimento apresentado foi validado através de testes de desempenho para localizar os inícios de nota em gravações de piano, monofônicas e polifônicas. Foi utilizado o *software* Finale©, cuja interface de edição de uma escala musical maior é apresentada na Figura 19. A vantagem em fazer uso de uma geração eletrônica é que a falta de prática do executor das músicas poderia comprometer o resultado final, bem como, na ausência de equipamentos de som de qualidade, haveria agravamento devido ao ruído.

Como exemplo monofônico, utilizou-se uma escala musical maior, tocando-se, pausadamente, as notas entre o C5 e o C6, totalizando oito notas. Este exemplo foi o primeiro utilizado para estimar os valores-limite apresentados na Tabela 4.

A seguir, partiu-se para o exemplo de uma música com leve polifonia, “Brothers”, da trilha sonora do anime “Fullmetal Alchemist”. Ela é interessante porque apresenta o compasso de três tempos bem marcado e notas de duração diferentes tocando simultaneamente.

O terceiro exemplo é o instrumental da música “First Love”, da cantora Utada Hikaru. A dificuldade desta peça são as notas de durações diferentes tocadas simultaneamente, alternando trechos rápidos, com muitas notas tocadas em pouco tempo, e lentos.

4.2.1 Exemplo monofônico - escala musical

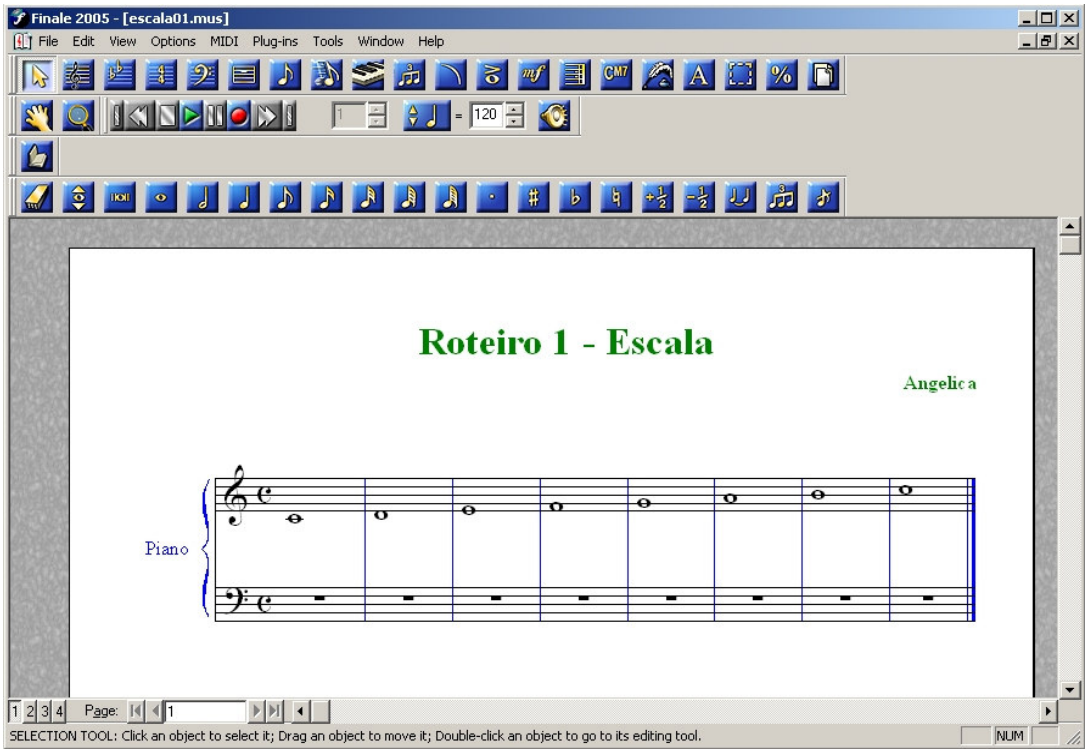


Figura 19 - Edição no *software* Finale© 2005 de uma escala musical maior.

O primeiro exemplo básico utilizado foi uma escala musical, do C central (C5, 523 Hz) até o C seguinte (C6, 1046 Hz). A seguir, o arquivo estéreo gerado com taxa de amostragem de 41,1 kHz foi convertido em mono, mantendo-se a precisão de 16 bits por amostra. Na execução desta rotina, também foram armazenados os sons obtidos na filtragem em sete bandas de freqüência.

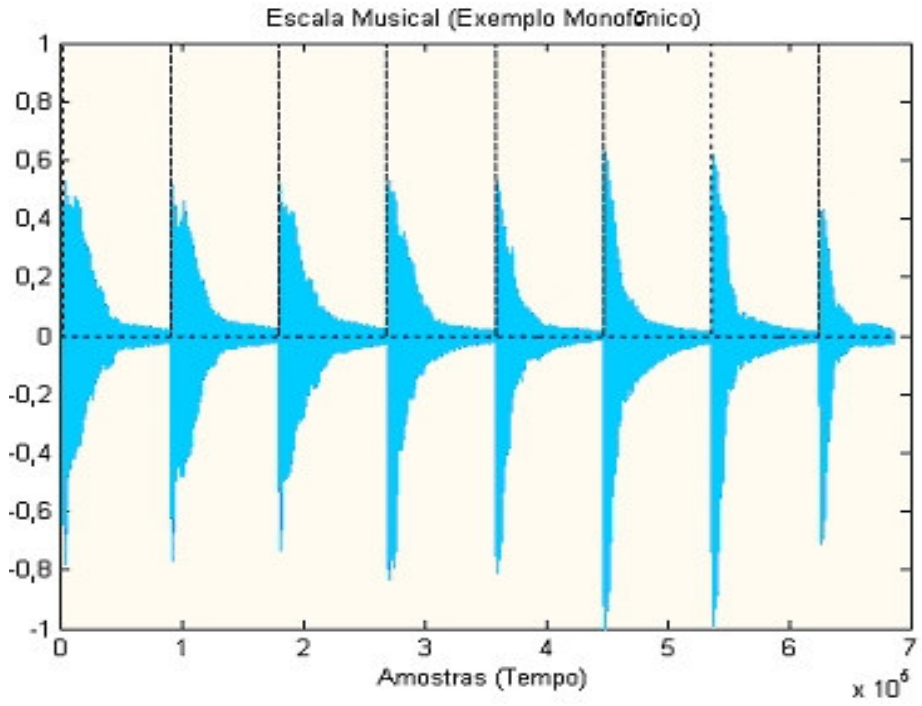


Figura 20 – Inícios de nota rastreados no sinal monofônico de uma escala musical, apresentando resultado satisfatório. Todas as notas foram detectadas e nenhum erro foi adicionado.

O resultado é visto na Figura 20. Por se tratar de um sinal monofônico e com bastante espaçamento entre as notas, o resultado foi facilmente obtido e considerado satisfatório, uma vez que todos os inícios foram detectados e não houve nenhum pico extra detectado.

4.2.2 Exemplo polifônico - fragmento da música "Brothers"

O segundo exemplo a ser apresentado aqui é um trecho de música simples, já com uma certa polifonia. Apresenta o compasso de três tempos bem marcado, notas de duração diferentes tocando simultaneamente. A partitura relativa a 12 compassos desta música é apresentada na Figura 21.

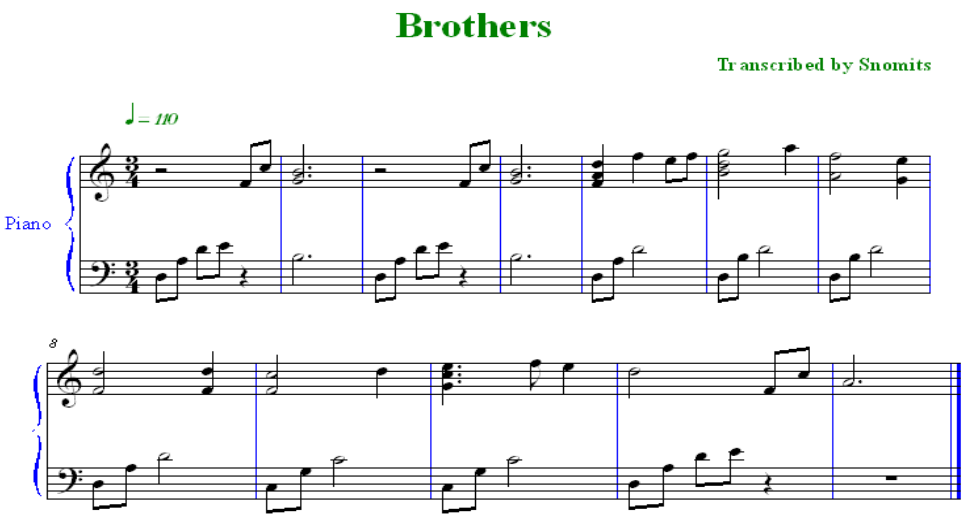


Figura 21 - Partitura do trecho inicial da música "Brothers".

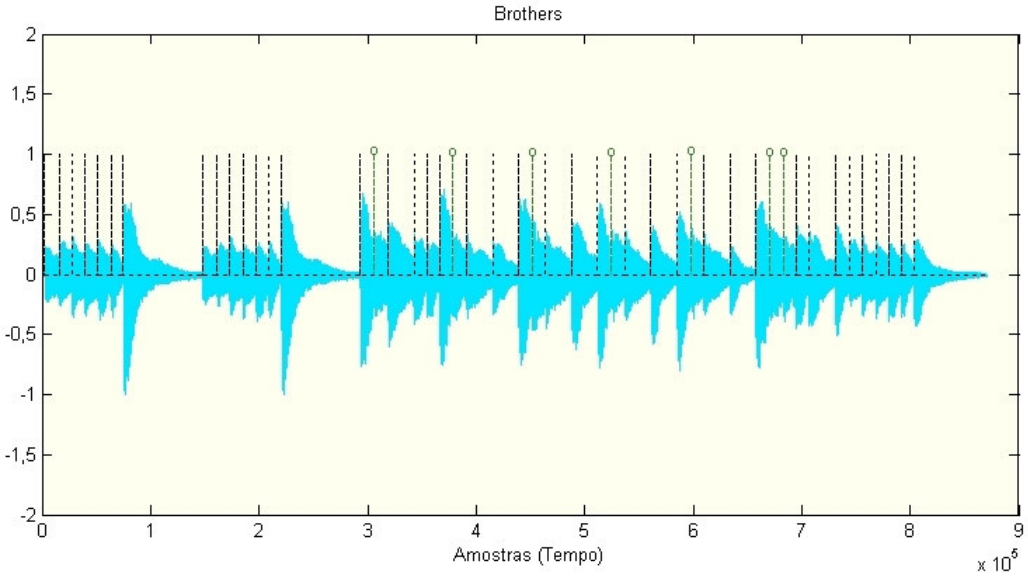


Figura 22 - Inícios de nota rastreados no sinal polifônico do trecho inicial da música Brothers. De um total de 47 notas, foram corretamente assinaladas 40 e 7 não percebidas. Nenhum erro foi adicionado.

O resultado é ilustrado na Figura 22. De um total de 47 notas, foram corretamente assinaladas 40, sendo que 7 não foram percebidas. O resultado não é exatamente satisfatório, indicando que, possivelmente, seja necessário rever os valores-limite na identificação de picos. Mas, ao se comparar ao resultado apresentado na Figura 23, que corresponde à detecção de *onset* obtida com um

software comercial existente (Cool Edit Pro©, versão 2.1), tem-se um bom resultado com o método proposto:

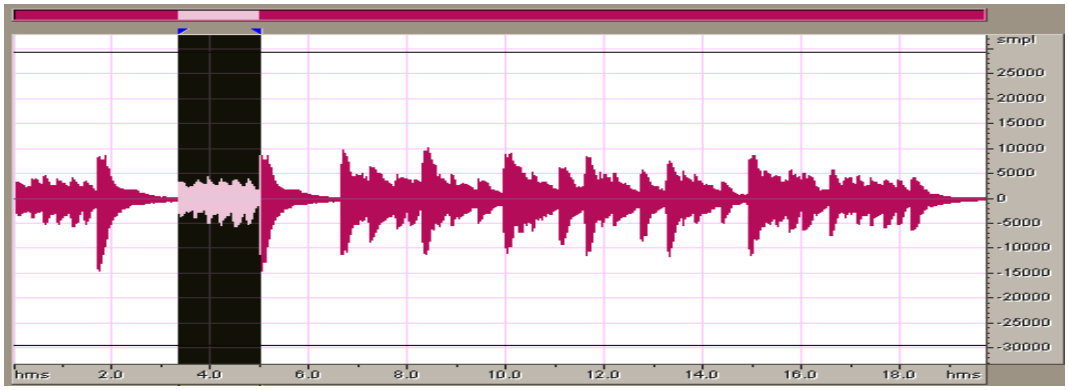


Figura 23 – Para título de comparação, um *software* comercial de edição de música considerou este bloco inteiro de 6 notas como apenas uma nota. O modelo comparado utiliza valores absolutos de diferença de decibéis como parâmetro.

4.2.3 Exemplo polifônico 2 – fragmento da música "First Love"

O terceiro exemplo a ser apresentado aqui é um trecho de uma música com uma certa polifonia de até 5 tons simultâneos. A dificuldade desta peça são as notas de durações diferentes tocadas simultaneamente, alternando trechos rápidos, com muitas notas tocadas em pouco tempo, e lentos. A partitura referente aos 8 primeiros compassos desta música se encontram na Figura 24.

Figura 24 - Partitura de um fragmento da música "First Love".

O resultado é apresentado na Figura 25. De um total de 33 notas, todos os inícios foram corretamente reconhecidos. No entanto, introduziram-se também cinco picos que não eram representativos de notas reais, assinalados com quadrados no gráfico.

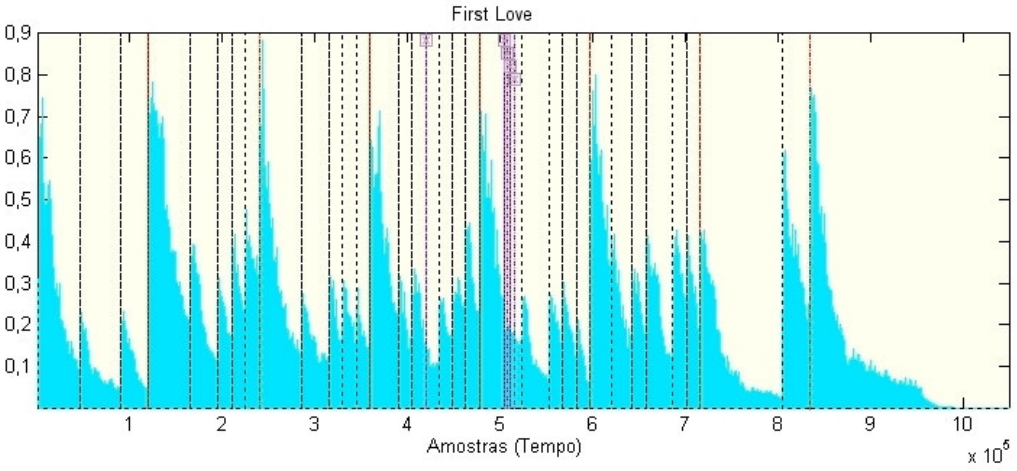


Figura 25 - Inícios de nota rastreados no sinal polifônico de um fragmento da música "First Love". Os 33 conjuntos de nota foram corretamente assinalados. No entanto, cinco notas inexistentes foram erroneamente assinaladas.

4.3 Conclusões

Este capítulo teve como objetivo descrever o funcionamento geral do módulo detector de início de nota, apresentando como seria tratado o fluxo de informações, as operações e os parâmetros ajustados.

Viu-se que, em uma única nota musical, os componentes de freqüência começam em instantes de tempo diferentes. Por isso, sugere-se que o sinal original seja dividido de acordo com bandas de freqüência. Utilizaram-se os valores de banda sugeridos em [9]. A seguir, calculou-se a envoltória de amplitude do sinal de cada uma das bandas, já retificado.

Discutiu-se sobre a utilização da função diferencial relativa de primeira ordem como sendo a principal dica para se identificar o início de uma nota, através da máxima variação relativa de energia. Assim sendo, de acordo com a Equação (4.1), deriva-se o logaritmo da envoltória, zerando os valores que estavam abaixo de determinados valores-limite. Estes foram calculados por meios empíricos, buscando-se eliminar candidatos falsos, mas sem mascarar possíveis inícios de sons existentes no sinal.

Finalmente, combinam-se os resultados de cada banda de freqüência, organizando-os temporalmente em um único vetor, eliminando aqueles cuja proeminência fosse menor que outra em uma janela de tempo de 50 ms, correspondente ao menor intervalo de tempo em que o ouvido humano é capaz de distinguir dois sons consecutivos.

Na Seção 4.2, apresentou-se o resultado obtido para três músicas de teste. A primeira, monofônica, era uma escala maior em C, que foi utilizada para ajuste fino dos valores-limite resumidos na Tabela 4. A segunda música, polifônica, apresentou falhas na detecção de 15% das notas. Em compensação, com este mesmo ajuste, a terceira música teste detectou 5 picos inexistentes, incluindo quase os mesmos 15% de erro.

Estes resultados sugerem que algum refinamento possa ser necessário para estes valores-limite. Por um lado, a detecção de notas inexistentes não é um

problema crítico, uma vez que podem ser futuramente tratadas como uma única, desenvolvendo-se um método que na fase posterior ao reconhecimento as reúna caso sejam de mesma freqüência, intensidades decrescentes e durações curtas. Por outro, ignorar sons presentes não pode ser tratado de forma análoga. No entanto, o excesso de notas fragmentadas pode comprometer o funcionamento dos algoritmos, uma vez que podem elevar consideravelmente o custo computacional. Desta forma, verificou-se que, com apenas estes parâmetros, torna-se difícil atender aos dois requisitos propostos: permitir que os picos que indicassem início de notas reais fossem exibidos e eliminar, ao máximo, picos que não fossem relacionados.

Este módulo, então, gera um conjunto de fragmentos de som que serão posteriormente analisados em separado. Desta forma, tem-se uma variável de saída, indicando o total de fragmentos localizados, e o seu valor correspondente em número de arquivos *wave* contendo tais sons.

No próximo capítulo, fala-se sobre a passagem de tais trechos musicais, ainda no domínio do tempo, para seus respectivos espectros de freqüência, e o que deve ser observado quando da sua implementação.

Capítulo 5 - Transformada de *Q* limitado

De uma maneira simplificada, esquematiza-se o reconhecimento de notas como sendo um processo fluindo hierarquicamente de uma representação de baixo nível (o sinal acústico, que nada mais é que a amostragem temporal de valores de pressão sonora) para alto nível (a representação simbólica de notas musicais), como a Figura 26 resume.

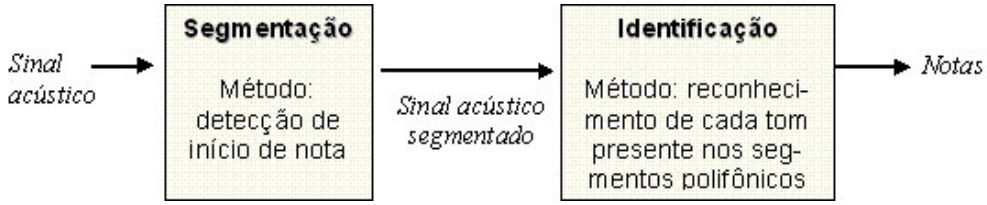


Figura 26 - Esquema simplificado do fluxo de dados do reconhecedor de notas musicais. Neste processo, tem-se inicialmente uma representação de baixo nível (sinal acústico) fluindo para uma representação de alto nível (notação musical).

Do sinal acústico, estruturado temporalmente, extraem-se informações como a estrutura rítmica e a divisão temporal de uma seqüência de notas e mistura de tons simultâneos, conforme visto no Capítulo 4. No entanto, não é possível extrair diretamente do domínio do tempo informações que irão distinguir as notas musicais entre si. Ou seja, entre essas duas representações, um nível de abstração intermediário se faz necessário, uma vez que a notação musical não pode ser diretamente deduzida a partir da informação presente no sinal acústico. A identificação da nota musical é feita a partir da obtenção de dados como a sua freqüência fundamental. Portanto, esta representação intermediária deve oferecer informações quanto às freqüências que compõem aquele sinal complexo, atentando-se para a necessidade de atender a requisitos como baixa complexidade e resolução satisfatória para os dados.

Usualmente, a análise espectral de sinais digitais é associada à aplicação da transformada discreta de Fourier (DFT) em sua implementação rápida (FFT). Mas, na análise de sinais musicais, a FFT apresenta pouca eficiência nos resultados, devido à distribuição linear de amostras no domínio da freqüência, enquanto a música sugere uma escala logarítmica, em função do espaçamento geométrico entre seus semitons. Assim, sugere-se uma transformada que ofereça resolução logarítmica.

Este capítulo versa sobre a transformação dos dados no domínio do tempo para uma representação no domínio da freqüência. Desta forma, a Seção 5.1 debate o aspecto logarítmico da audição humana e, portanto, da música ocidental. Esta discussão apresenta um dos requisitos fundamentais desta transformada, a questão da largura de banda. Na Seção 5.2, discute-se o uso da FFT e a falta de eficiência em sua resolução. A Seção 5.3 apresenta duas opções para que tal problema seja

superado: a transformada de Q constante (CQT) e a transformada de Q limitado (BQT). Apresenta-se como é efetuado o cálculo da BQT, que é computacionalmente menos complexo que a obtenção da CQT, justificando a escolha da BQT para a presente aplicação. Finalmente, na Seção 5.4, resumem-se os questionamentos.

5.1 O aspecto logarítmico da audição humana e da música ocidental

A audição humana é caracterizada por uma percepção logarítmica de freqüências, baseando-se em uma razão de valores. Para título de ilustração, ao se pensar em uma razão 1:2, isso significaria que a variação de uma freqüência f qualquer para $2f$ seria a mesma percebida de $2f$ para $4f$, ou de $4f$ para $8f$ [3]. Esta é a razão pela qual as freqüências escolhidas para compor as escalas, na música ocidental, estão espaçadas geometricamente.

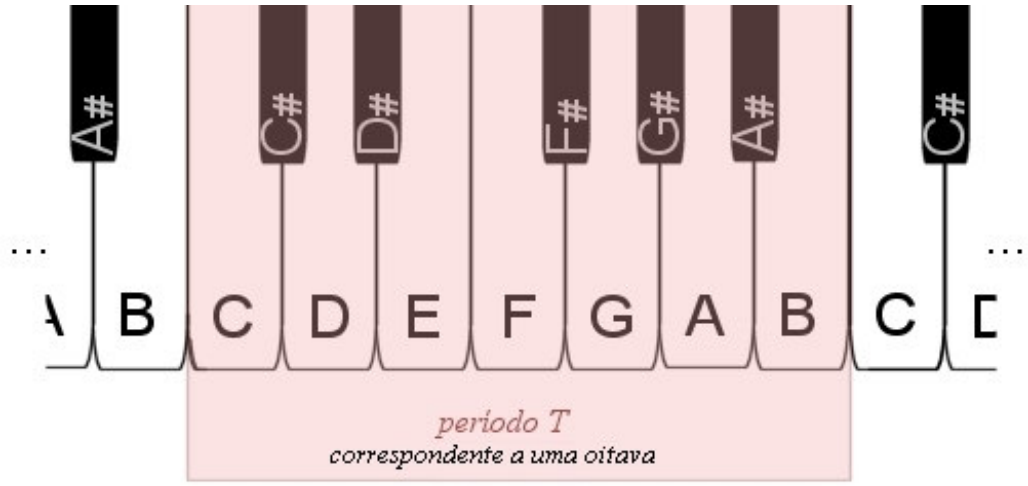


Figura 27 – Desenho esquematizado de algumas teclas de um piano. Ambas as teclas pretas e brancas representam notas. O teclado é periódico na direção horizontal, repetindo-se após uma seqüência de sete notas brancas e cinco pretas, o que corresponde a uma oitava. Este período representa o duplicamento da freqüência fundamental das notas em questão.

Desta forma, a relação entre duas oitavas equivale à razão 1:2. A nota C da quinta oitava tem como freqüência fundamental $f_{C5} = 2.f_{C4}$, sendo f_{C4} a freqüência fundamental de um C da quarta oitava.

Observa-se que a Figura 27 esquematiza as teclas de um piano. Pode-se dizer que o teclado é periódico na direção horizontal, repetindo-se após uma seqüência de

sete notas brancas (de C até B) e cinco pretas (de C# até A#), o que corresponde a uma oitava. Em um teclado moderno, cada um desses doze intervalos que compõem a oitava representa uma mesma razão de freqüência, o semitom.

A partir disso, pode-se arranjar as freqüências das 88 teclas de um piano, para k inteiro, $-48 \leq k \leq 39$, atribuindo $k = 0$ para o A central (A4 = 440 Hz), através da seguinte relação:

$$f_k = 440.2^{k/12} \tag{5.1}$$

5.2 O uso da FFT

A FFT divide o espectro em faixas da mesma largura, representando, portanto, os sinais em intervalos de freqüência constantes, seguindo uma progressão aritmética. A Figura 3, apresentada no Capítulo 2, ilustra os componentes de uma nota A, de freqüência fundamental 440 Hz, de um Oboé, na qual ficou claro o espaçamento constante entre os harmônicos. Destaca-se que aquela figura focava no intervalo até 4400 Hz, exibindo até o décimo harmônico. Observa-se que, a partir do décimo quinto harmônico (6600 Hz), as componentes seriam praticamente nulas, como a Figura 28 ilustra, apresentando o eixo das freqüências em resolução (a) linear e (b) logarítmica.

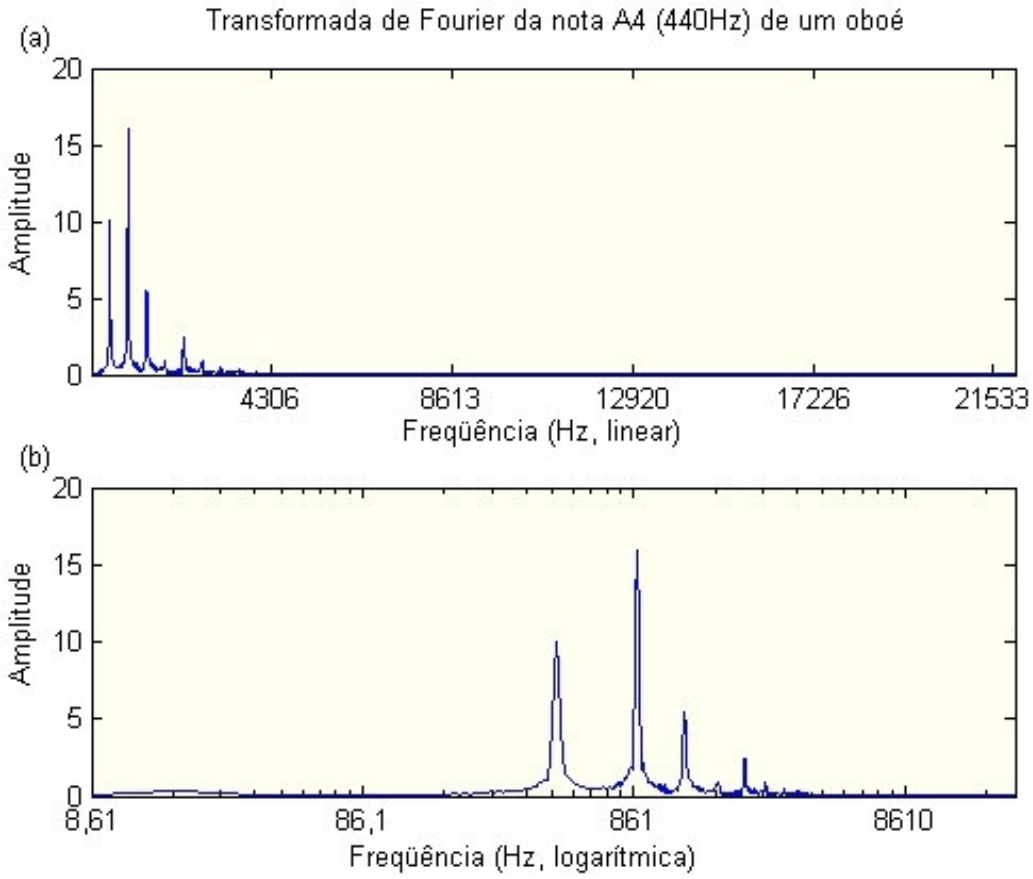


Figura 28 – Transformada de Fourier de uma nota A (440 Hz) de um oboé. Tem-se a informação do sinal concentrada até o 15º harmônico, ou seja, até 6600 Hz, o que equivale aproximadamente aos 3/20 iniciais do eixo das freqüências, para (a) resolução linear da transformada. Em (b), apresenta-se a Transformada de Fourier em escala logarítmica.

A informação do sinal acústico está concentrada até 6600 Hz, o que equivale a 0,15 da freqüência de amostragem de 44100 Hz. Isso significa que 17/20 do gráfico apresenta valores nulos ou muito próximos de zero, o que sugere que a distribuição da FFT não é a mais adequada para a análise de um sinal musical.

Observa-se a faixa de notas musicais geradas por um piano, conforme a Equação (5.1), compreendida entre 27,5 Hz (A0) até 4185 Hz (C8). Uma resolução suficiente para distinguir duas notas adjacentes quaisquer seria a metade do intervalo do seu menor semitom (27,50 Hz) até o segundo menor semitom (29,13 Hz) [15].

$$\frac{\Delta f_{\min}}{2} = \frac{f_{0_{A\#}} - f_{0_A}}{2} = \frac{29{,}13 - 27{,}50}{2} = 0{,}815 \text{ Hz} \qquad (5.2)$$

Para se obter tal resolução fazendo uso da FFT, são necessárias mais de 50000 amostras na taxa de amostragem, F_S, de 44100 Hz, uma vez que:

$$\frac{F_S}{resolução} = \frac{44100}{0{,}815} \approx 54110 \text{ amostras} \tag{5.3}$$

A escolha de uma resolução pior resultaria em insuficiência nas baixas freqüências. Em contrapartida, este valor oferece um zelo excessivo nas altas freqüências, uma vez que na sétima oitava, entre C7 (2093 Hz) e C8 (4186 Hz), a resolução de 62 Hz já é suficiente.

É interessante que a representação intermediária proposta atenda ao espaçamento geométrico da música. A *transformada rápida de Fourier* (FFT), apesar de ter baixa complexidade computacional, não é a melhor opção, uma vez que não mapeia as freqüências musicais eficientemente.

5.3 As transformadas de *Q* constante e *Q* limitado

5.3.1 O fator de qualidade *Q*

O fator de qualidade *Q* de cada faixa de freqüência está relacionado com a seletividade da transformada, de acordo com:

$$Q = \frac{f_k}{\Delta f_k} \tag{5.4}$$

onde f_k é a freqüência central e Δf_k é a largura da faixa de freqüências de cada canal.

Na DFT, esta largura Δf_k é constante e definida pela razão entre a freqüência de amostragem F_S e o número N de amostras do segmento do sinal sendo analisado. Desta forma, na DFT, o fator de qualidade *Q* varia com a freqüência em uma relação diretamente proporcional, o que provoca o comprometimento de qualidade nas baixas freqüências. Em contrapartida, nas freqüências mais altas, *Q* será desnecessariamente grande.

5.3.2 A transformada de *Q* constante

A C*Q*T, *transformada de* Q *constante*, é uma adaptação da DFT que oferece uma distribuição logarítmica das freqüências no domínio da transformada. Como o próprio nome diz, busca-se o mesmo valor de *Q* para todas as freqüências, o que faz com que a largura da faixa de freqüências Δf_k de cada canal seja variável de acordo com a f_k central analisada pela transformada.

Para a sua implementação, é necessário escolher as freqüências mínima e máxima e o fator *Q* desejado, que deve ser suficiente para a distinção de dois semitons adjacentes. Para tal, a razão entre as freqüências centrais de dois canais adjacentes deve ser $2^{1/24}$, que equivale à raiz quadrada da razão entre as freqüências de dois semitons adjacentes, uma vez que tais freqüências estão em progressão geométrica [15].

A sua forma de implementação eficiente [11] praticamente combina diversas amostras de FFTs para produzir a freqüência logarítmica. Já a sua realização direta da C*Q*T é trabalhosa demais para uso prático, o que pode comprometer a complexidade computacional.

5.3.3 A transformada de *Q* limitado

Uma técnica híbrida é a *transformada de* Q *limitado*, elaborada para adequar a distribuição de faixas de freqüência da FFT à música. A premissa é a divisão do espectro em oitavas, às quais são aplicadas FFTs com resoluções adequadas a cada uma delas, produzindo um número constante de amostras por oitava [12]. O nome "*Q* limitado" deriva do fato de a variação do fator de qualidade *Q* ser limitada dentro de uma oitava.

De acordo com a referência [9], para a obtenção da B*Q*T, a FFT é calculada e metade dos valores de freqüência é descartada, sendo armazenada apenas a oitava superior. A seguir, o sinal no domínio do tempo é decimado por um fator de dois. A FFT é novamente calculada com a mesma largura de janela, que agora fornecerá o dobro de resolução. Desta transformação, a metade superior de valores é armazenada, que agora corresponderá à segunda oitava mais alta e tendo número de

bins que a oitava superior. Este procedimento é repetido até que se alcance a menor oitava de interesse. Tem-se, portanto, uma espécie de FFT por partes, com o aumento da resolução crescendo das oitavas mais altas até as mais baixas, mas ainda constante dentro de cada oitava [15].

A operação de decimação por um fator de dois é caracterizada por duas operações em cascata. A primeira é a aplicação de um filtro passa-baixas. A segunda é a reamostragem do sinal. Tal processo dilata o espectro, como pode ser visto na Figura 29, que apresenta o efeito da decimação na resolução das faixas de freqüência.

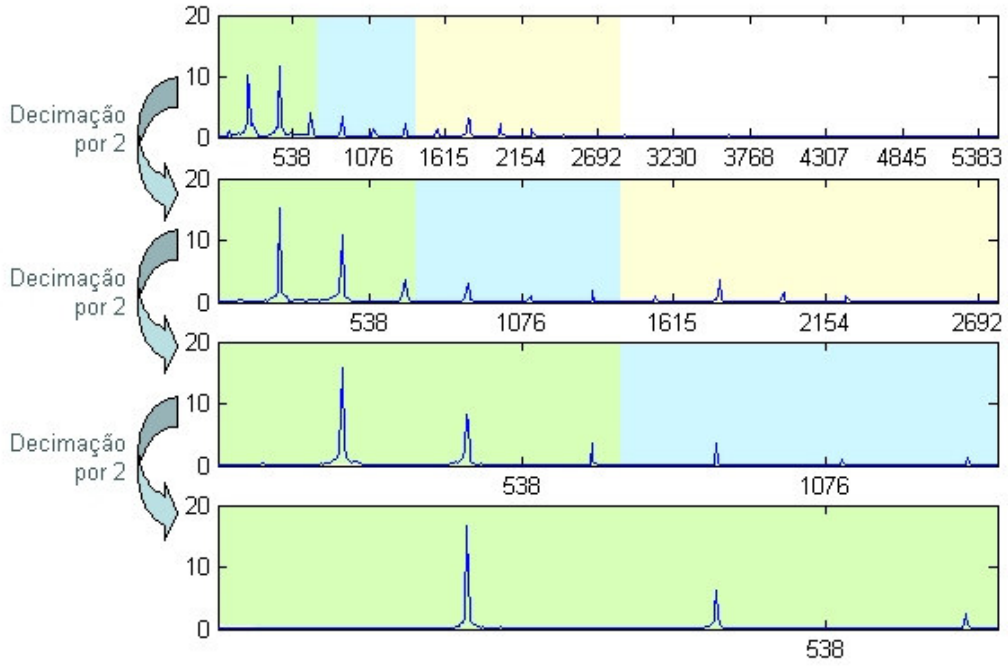


Figura 29 – Efeito da decimação por fator de 2. Tal processo dilata o espectro, fazendo com que a FFT correspondente possua o dobro de resolução.

Na implementação da B*Q*T, três parâmetros podem ser ajustados:

1. O tamanho da menor janela de tempo (para as freqüências mais altas);
2. O número de oitavas para iteração até alcançar as freqüências mais baixas de interesse, dobrando a precisão de freqüência e a janela de tempo para cada oitava;
3. O instante de tempo no qual a análise de freqüência e as janelas de tempo são centralizadas.

5.3.4 Comparação entre o uso da FFT e da B*Q*T

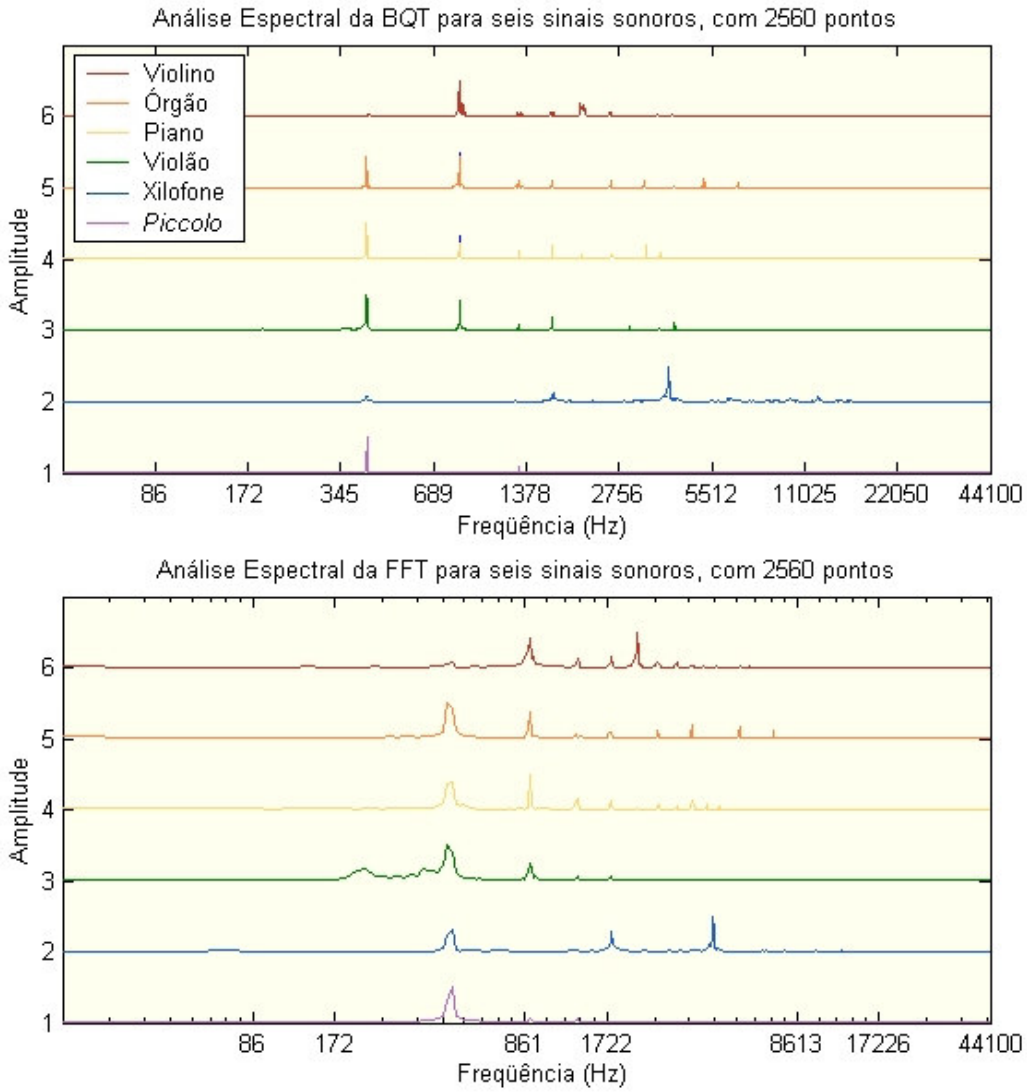


Figura 30 - Comparação da análise espectral para seis sinais sonoros, compostos por uma única nota lá (A4, 440Hz), de instrumentos diferentes, sintetizados via *software*.

Na Figura 30, uma nota A da quarta oitava (A4, 440Hz), tocada por diversos instrumentos, é submetida à FFT e à B*Q*T, sendo obtido o mesmo número de 2560 amostras para os dois casos, que é o valor utilizado neste projeto. A B*Q*T permite o eficiente rastreamento de senóides, oferecendo uma resolução maior em freqüências mais baixas e, ao mesmo tempo, diminuindo a precisão em altas freqüências. Observe atentamente os espectros referentes à nota A4 tocada no violão, a terceira linha do gráfico, em verde. Na FFT, só é possível observar quatro picos de

harmônicos até 4410 Hz. Já fazendo uso da B*Q*T, vêem-se pelo menos 7 harmônicos presentes até tal freqüência.

No exemplo citado na Figura 30, tem-se apenas uma nota sendo tocada. A resolução da cada faixa, na FFT, é de 17,22 Hz por amostra. Isso já seria o suficiente para sobrepôr as freqüências fundamentais de notas como A0 (27,5 Hz) e A0# (29,13 Hz), por exemplo. Na Figura 31, foi composto um sinal com a soma de três senóides puras: a primeira com freqüência $f_1 = 65{,}4$ Hz, que é a fundamental da nota C2; a segunda sendo o semitom adjacente $f_2 = 69{,}3$ Hz, correspondente a C2#; e a terceira sendo a maior freqüência fundamental observada em um piano, C8, $f_3 = 4186$ Hz., com taxa de amostragem, F_S, de 44100 Hz. A este sinal foram aplicadas a FFT e a B*Q*T, ambas com 2560 pontos. Para a FFT, isso significa uma resolução constante de 17,22 Hz/amostra, o que seria já menor que a diferença entre as notas C2 e C2#. Para a B*Q*T, tem-se a resolução nesta faixa de freqüência de 0,32 Hz/amostra. Já na faixa à qual $f_3 = 4186$ Hz pertence, a resolução é de 10,8 Hz/amostra. A Figura 31 ilustra portando que a baixa resolução da FFT em baixas freqüências acaba por considerar as duas senóides puras referentes a semitons adjacentes como apenas uma. Já a B*Q*T identifica os dois picos em separado. Na alta freqüência, a senóide de 4186 Hz é devidamente representada em ambos os casos.

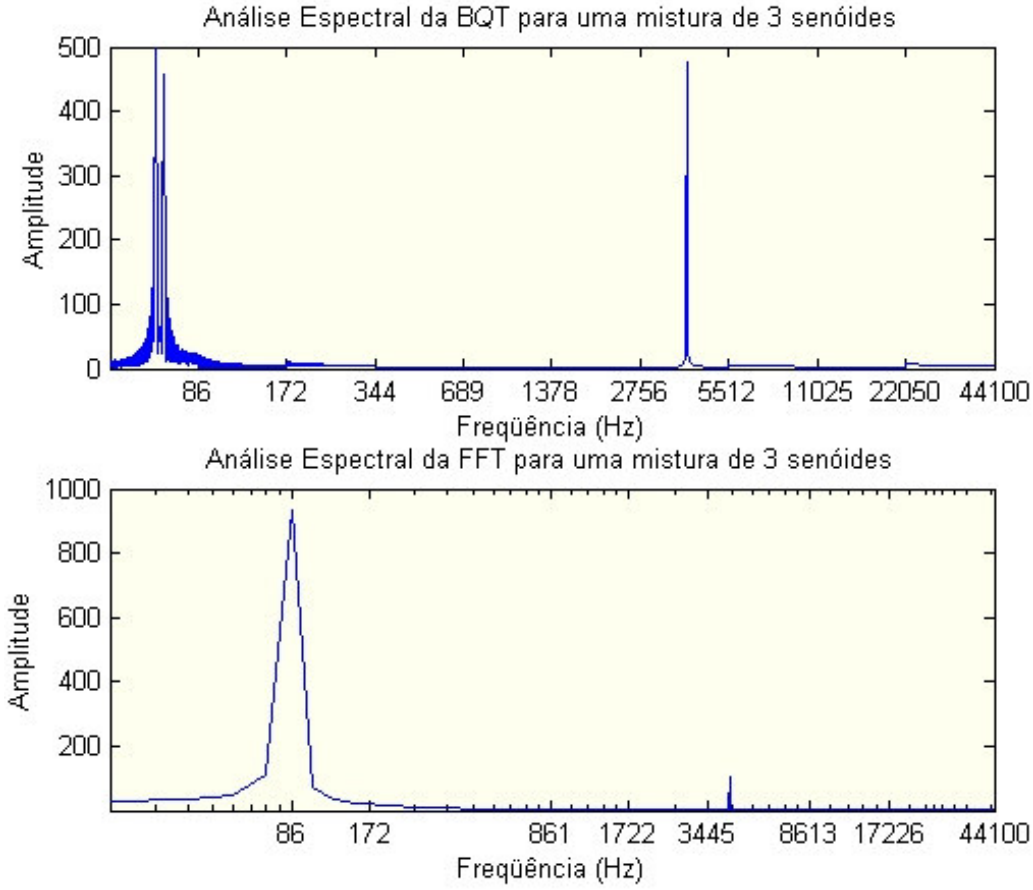


Figura 31 - Análise espectral da FFT e da B*Q*T para o sinal composto por três senóides puras com freqüências f_1 = 65,4 Hz, f_2 = 69,3 Hz e f_3 = 4187 Hz, fazendo uso de 2560 pontos. Observe que a baixa resolução da FFT nas baixas freqüências faz com que as duas senóides sejam consideradas como uma apenas, com freqüência f = 69 Hz. Já na B*Q*T, com a mesma quantidade de pontos, as senóides são vistas como picos separados. Nas altas freqüências, a senóide de 4186 Hz é identificada corretamente como um pico em ambos os casos.

Comparativamente, a B*Q*T apresentou-se o método mais eficiente, dentre os discutidos neste capítulo, para se representar um sinal musical no domínio da freqüência. Comparado com a FFT, seu custo computacional é próximo, uma vez que, apesar de serem necessárias 10 FFTs para se compor uma B*Q*T, estas são muito menores. Além disso, deve-se levar em consideração que a fraca resolução da FFT nas baixas freqüências exigiria uma quantidade muito maior de pontos para a representação do sinal, consumindo mais memória.

5.3.5 Parâmetros da implementação da B*Q*T

É importante ressaltar que os parâmetros de análise selecionados devem ser os mesmos tanto para o bloco de modelagem de tom quanto para o bloco de reconhecimento de notas.

Neste projeto, fixou-se a saída da transformada como um vetor de tamanho de 10 vezes uma constante arbitrária BIN, à qual foi atribuído o valor de 256, ou seja, 2^8, uma potência de 2, a fim de facilitar o cálculo da FFT. Com este valor, tem-se como resolução mínima uma janela de aproximadamente 86,12 Hz por amostra, para freqüências mais altas, e como resolução máxima uma janela de aproximadamente 0,32 Hz por amostra, para as freqüências mais baixas.

Desta forma, 10 é o número de oitavas em que foi dividido o sinal, de forma a se ter um alcance até a freqüência do C3, o que corresponde a 130 Hz, que é o limite inferior de freqüência obtido. Esta limitação será explicada no capítulo referente à modelagem de tom.

5.4 Conclusões

Sendo, na música, o espaçamento entre os semitons geométrico, na passagem para o domínio da freqüência de um sinal sonoro, é necessária uma escolha criteriosa de uma representação cujos principais requisitos são a distribuição eficiente de resolução de acordo com a freqüência e baixo custo computacional. Para uma transformada com intervalos fixos como a FFT, a fim de atender este requisito seria preciso trabalhar com vetores muito grandes, o que seria um desperdício nas altas freqüências, que necessitam de resoluções baixas.

Uma solução a ser buscada é uma transformada que ofereça uma escala logarítmica. A C*Q*T, *transformada de* Q *constante*, atende a este requisito, embora não seja indicada para a transcrição musical, uma vez que seu cálculo é altamente dispendioso, computacionalmente, uma vez que para se obter um fator de qualidade *Q* constante, seria necessária uma FFT para cada freqüência central f_k.

A B*Q*T, *transformada de* Q *limitado*, é uma transformada que atende razoavelmente aos dois requisitos. Ela é facilmente obtida, tem resolução crescente em relação às oitavas, mas distribuição linear dentro das mesmas, dobrando-se a resolução à medida que se deslocam para as freqüências mais baixas. Por esta razão, escolheu-se a B*Q*T como a forma a se representar o sinal do domínio da freqüência.

Este capítulo, portanto, tratou a questão da transformação dos dados, debatendo o aspecto logarítmico da percepção humana do som e, portanto, da música ocidental, e assim, propondo o uso da B*Q*T. O seu cálculo foi descrito e justificado por atender aos requisitos.

Capítulo 6 - Extração de dados a partir de uma mistura de sons harmônicos

Um som complexo, não importando se é ou não periódico, pode ser sempre decomposto em um número de componentes senoidais, cada um com freqüência, amplitude de pico e fase individuais. Conforme apresentado no Capítulo 2, um som é harmônico quando existe uma relação múltipla e inteira entre as freqüências e uma fundamental, chamada f_0. O grande problema na análise e obtenção de dados de um sinal polifônico é a sobreposição de seus harmônicos. Isso é agravado pelo fato de a música, muitas vezes, querer que o ouvinte sinta aquelas composições como um único som harmônico.

Neste capítulo, estuda-se um possível método a partir do qual se possam extrair dados os mais confiáveis possíveis. Para isso, deve-se descobrir quais os harmônicos menos suscetíveis a sobreposição. Assim, na Seção 6.1, é feita nova abordagem da questão da sobreposição dos harmônicos, apresentando exemplos de composição de acordes. A Seção 6.2 apresenta os harmônicos primos como possíveis fontes não deturpadas de informações do sinal, uma vez que eles estão menos sujeitos à sobreposição. No entanto, eles não podem ser utilizados separadamente para responder pelas características do sinal. A Seção 6.3 discute os requisitos do filtro de extração de dados, sendo complementada pela Seção 6.4, que apresenta o vetor de probabilidade de seleção de um harmônico como a saída do filtro de extração de dados. Tal vetor é utilizado como parâmetro de peso no filtro. Assim, justificam-se as escolhas a respeito da quantidade de harmônicos utilizados para descrever o sinal, bem como as limitações ao sistema em que isso irá resultar. A Seção 6.5 oferece a conclusão do capítulo.

6.1 A questão da sobreposição de harmônicos

Apesar de a escala musical ocidental ser logarítmica, são geralmente produzidos intervalos harmônicos que provocam a sobreposição de seus componentes[1]. Observe, por exemplo, o acorde C maior (ou simplesmente C), composto pelas notas C, E e G (dó-mi-sol). Para esta ilustração, foi analisada a quinta oitava:

$$F_{0_C} = 440.2^{(3/12)} \cong 523{,}25Hz \tag{6.1}$$

$$F_{0_E} = 440.2^{(7/12)} \cong 659{,}26Hz \tag{6.2}$$

$$F_{0_G} = 440.2^{(10/12)} \cong 784{,}00Hz \tag{6.3}$$

Na Tabela 5, apresentam-se a freqüência fundamental e os primeiros harmônicos referente às três notas que compõem o acorde C. Pode-se, então, observar que, ao serem somados os três sinais, haverá a sobreposição de harmônicos, destacados na tabela aos pares.

[1] Estritamente falando, em se tratando da Escala Temperada, definida matematicamente como esta progressão geométrica cujo primeiro termo é a freqüência da nota A4 e cuja razão é o valor numérico $2^{1/12}$, em decorrência da divisão de uma oitava em 12 intervalos, não há sobreposição. No entanto, no decorrer deste trabalho, admite-se esta aproximação.

Tabela 5 - Freqüências fundamentais e primeiros harmônicos das notas que compõem o acorde C

Nota	F_0	$2.F_0$	$3.F_0$	$4.F_0$	$5.F_0$	$6.F_0$
C	523,25	1046,50	1569,75	2093,00	2616,26	3139,51
E	659,26	1318.51	1977,77	2637,02	3296,28	3955,53
G	784,00	1567,98	2351,97	3135,96	3919,95	4703,95

Analisando com cuidado a Tabela 5, observa-se que todo o *3m*-ésimo harmônico da nota C será sobreposto pelo *2m*-ésimo harmônico da nota G. Da mesma forma, todo *5n*-ésimo harmônico de C será sobreposto pelo *4n*-ésimo harmônico de E; e todo *6i*-ésimo harmônico de E será sobreposto pelo *5i*-ésimo harmônico de G.

Dois tons *R* e *S* estão em relações harmônicas entre si se as suas freqüências fundamentais satisfazem a relação $F_{0_R} = \frac{m}{n} F_{0_S}$, para *m* e *n* naturais pequenos. Quanto menores os valores de *m* e de *n*, mais próxima é a relação harmônica e, portanto, melhores os dois tons soarão juntos. Uma aplicação direta desta regra está na formação de acordes, que são a combinação harmônica de três ou mais notas diferentes. Tais relações harmônicas tornam o sinal agradável de ser ouvido. Sons que não são regidos por relações harmônicas, em geral, causam estranheza e desconforto.

Os acordes menores são constituídos por freqüências fundamentais relacionadas entre si por $\frac{4}{6}f, \frac{4}{5}f$ e $\frac{4}{4}f$. Já os acordes maiores são constituídos por freqüências fundamentais relacionadas entre si por $\frac{4}{4}f, \frac{5}{4}f$ e $\frac{6}{4}f$. Como exemplo, novamente, o acorde maior C:

$$\left.\begin{array}{l} C = 440.2^{3/12} = 523{,}25Hz \\ E = 440.2^{7/12} = 659{,}25Hz \\ G = 440.2^{10/12} = 783{,}99Hz \end{array}\right\} \begin{array}{l} E = \dfrac{659{,}25}{523{,}25} \cong 1{,}25 = \dfrac{5}{4}C \\ G = \dfrac{783{,}99}{523{,}25} \cong 1{,}5 = \dfrac{6}{4} = \dfrac{3}{2}C \end{array} \qquad (6.4)$$

O conjunto de equações (6.4) conduz à mesma observação feita a partir dos dados da Tabela 5, sobre a relação da sobreposição dos harmônicos: cada harmônico múltiplo de 4 da freqüência fundamental da nota E se sobreporá a cada harmônico múltiplo de 5 da freqüência fundamental da nota C; da mesma forma, cada harmônico par da nota G se sobreporá a cada harmônico múltiplo de 3 da nota C.

A partir daí, pode-se dizer que 47% dos harmônicos da nota C sofrerão a interferência dos harmônicos das notas E ou G. Analogamente, tem-se que 33% dos harmônicos de E serão interferidos por C ou G; e 60% dos harmônicos de G serão alterados por C e E. Visualmente, a Figura 32 apresenta a *transformada de* Q *limitado* (B*Q*T) deste acorde C, em que as três notas foram tocadas com a mesma intensidade. A sobreposição dos harmônicos fica evidente.

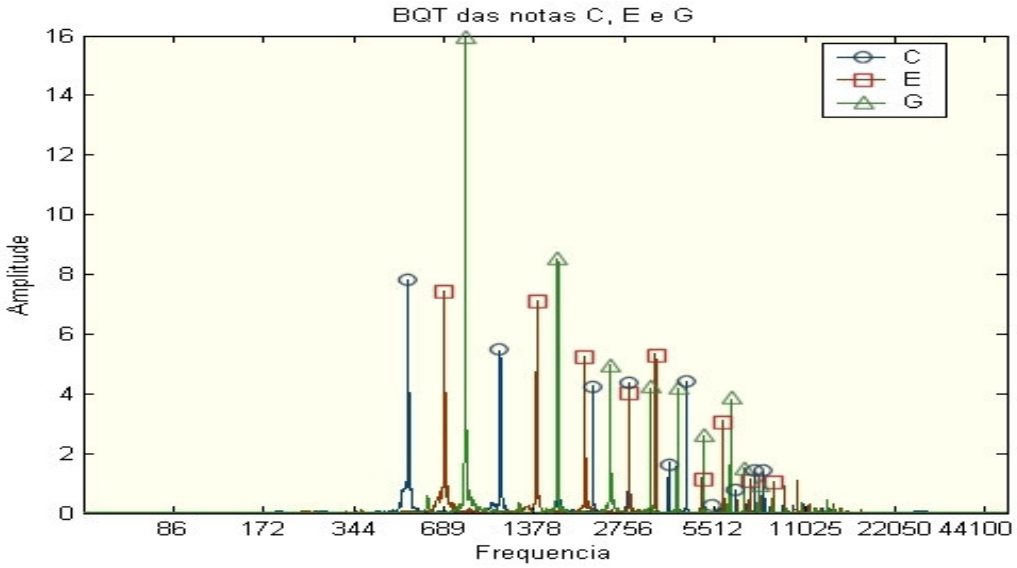


Figura 32 – A B*Q*T das três notas que compõem o acorde C em separado, plotadas no mesmo gráfico. Observe que a sobreposição dos harmônicos fica evidente.

6.2 O uso de números primos

Os parciais harmônicos mais suscetíveis à sobreposição são aqueles que possuem índice com a maior quantidade de divisores. Como exemplo, observe o 12º parcial harmônico de um tom *S*, $h_S(12)$, cuja freqüência é igual a $12.f_{0_S}$. O número 12 é divisível por {1,2,3,4,6,12}. Isso significa que este 12º harmônico de um tom *S* poderá ser sobreposto por qualquer n-ésimo parcial harmônico de um sinal interferente *R* cuja freqüência fundamental obedeça a quaisquer das seguintes relações:

$$f_{0_R} = \frac{1}{n} f_{0_S},\ f_{0_R} = \frac{2}{n} f_{0_S},\ f_{0_R} = \frac{3}{n} f_{0_S},\ f_{0_R} = \frac{4}{n} f_{0_S},\ f_{0_R} = \frac{6}{n} f_{0_S} \text{ ou } f_{0_R} = \frac{12}{n} f_{0_S} \quad (6.5)$$

Portanto, fica evidente que se basear em parciais harmônicos com índice divisível por muitos números para a extração de dados representativos de uma característica geral do tom não é aconselhável, uma vez que estes estão sujeitos a uma maior probabilidade de interferência.

Números primos são aqueles divisíveis somente por 1 e por eles mesmos. Desta forma, os harmônicos de índices primos são sujeitos a uma menor probabilidade de interferência. Isso porque um tom interferente *R* só pode sobrepor um único harmônico primo de um sinal *S*, a não ser que estejam em uma relação $f_{0_R} = \frac{1}{n} f_{0_S}$. Neste caso em especial, todos os harmônicos seriam sobrepostos.

Harmônicos primos de um tom *S* podem ser considerados evidências independentes da existência do tom *S* ou de suas características dedutíveis de seus harmônicos. No entanto, utilizar somente dados referentes aos harmônicos primos não é interessante, por se tratar de um número muito restrito de amostras, o que pode ocasionar desvios muito grandes, já que alguns harmônicos primos podem ter sido severamente perturbados por algum som interferente. Além disso, pode tornar o algoritmo muito sensível ao conteúdo tonal, uma vez que harmônicos primos podem não existir no tom *S*, dependendo do timbre do instrumento.

Esta é a principal motivação para o desenvolvimento de um filtro que extraia a informação desejada do conjunto formado por todos os harmônicos de um tom *S*, dando ênfase aos primos e desprezando os valores irrelevantes.

6.3 Filtro extrator de características

Quando se têm informações relativas ao conjunto dos harmônicos de um tom *S*, como, por exemplo, as amplitudes de cada um deles, levar em consideração a média das amostras não é satisfatório. Isso é devido ao fato de que um único valor errado dentre outros pode estar tão severamente comprometido a ponto de provocar um desvio inaceitável no valor médio. Em função disso, filtros de estatística de ordem e medianas são cogitados, pois são particularmente eficientes ao lidar com o tipo de dado caracterizado acima [9]. Tais filtros dependem da organização do conjunto de amostras. A Figura 33 ilustra um exemplo de como a ordenação e o uso da mediana podem prevenir que valores corrompidos influenciem na saída de um filtro.

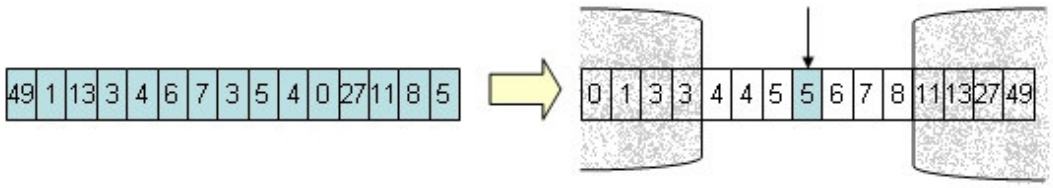


Figura 33 – Supondo-se um vetor de dados de amplitudes relativas a cada um dos harmônicos de um sinal sonoro, ordenar o conjunto de amostras é um bom modo de separar valores inválidos, selecionando-se a mediana. No exemplo, valores muito altos como 27 e 49 destoam do conjunto de amostras apresentado e podem ser resultado da sobreposição por outro tom interferente, *R*. Levá-los em consideração, como no cálculo da média das amostras, resultaria em 9,73, um valor quase o dobro do obtido pela ordenação e escolha da mediana.

Ao se ordenar os valores da amostra, dados errôneos, sejam subestimados ou superestimados, tendem a se concentrar nas extremidades. Se é possível admitir que a maioria dos valores de amostra é confiável, selecionar a mediana, ou seja, a amostra do meio do conjunto ordenado, é seguro.

Na busca por um filtro que se comporte conforme o desejado em uma aplicação em música, deve-se ter em mente que os harmônicos não são igualmente confiáveis em valor. Além da questão apresentada na Seção 6.2, que ilustrou que

parciais harmônicos primos são menos suscetíveis à sobreposição, é interessante que os parciais harmônicos de freqüências mais baixas sejam enfatizados, uma vez que a amplitude dos parciais harmônicos de freqüências mais altas tende a ser menor para a maioria dos instrumentos, o que os torna mais suscetíveis à interferência e ao ruído.

Variações como o tipo de instrumento e a velocidade com a qual as notas são tocadas acabam por alterar os valores das freqüências mais altas, ocasionando desvios que vêm a descaracterizar o modelo tonal utilizado. Isso foi observado na fase de criação de modelos de tom. De fato, uma nota com duração de quatro tempos apresentou a amplitude de freqüência fundamental mais intensa se comparada com os demais harmônicos. À medida que o som se tornava mais breve, como 1/16 de tempo, os harmônicos mais altos passaram a apresentar intensidades mais altas se comparado com as notas mais longas, o que faz com que o modelo não se ajuste perfeitamente ao som. No entanto, uma boa aproximação é obtida.

Com isso, caminha-se para um modelo de filtro que deve considerar a totalidade dos harmônicos, atribuindo-lhes pesos de acordo com o seu grau de confiabilidade. Filtros de estatística de ordem ponderada (WOS – *weighted order statistic filters*) possuem boas propriedades para este fim [14]. Neles, as amostras são organizadas em ordem crescente de valor. A seguir, cada amostra é repetida de acordo com os pesos fornecidos como parâmetros do filtro. Finalmente, o *T*-ésimo valor é escolhido dentro do conjunto ordenado obtido, sendo então *T* o limite e também um parâmetro do filtro. A Figura 34 ilustra o funcionamento do filtro WOS, para o qual $\nu\{49 \quad 1 \quad 13 \quad 3 \quad 4 \quad 6 \quad 7\} = 4$. Neste exemplo, o valor limite *T* e os pesos foram atribuídos arbitrariamente, em caráter apenas ilustrativo.

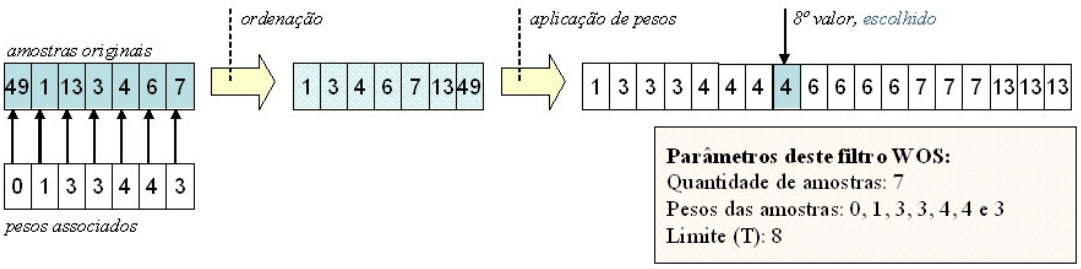


Figura 34 - Parâmetros e cálculos de exemplo de filtro WOS. A cada amostra é atribuído um peso, que será a quantidade de vezes que tal amostra será repetida. O vetor é ordenado. Em

seguida, os pesos são aplicados e escolhe-se o *T*-ésimo elemento como a saída do filtro. No exemplo, *T* = 8, sendo, portanto, o oitavo elemento selecionado.

Na composição do filtro, como parâmetro para a obtenção dos pesos será utilizado $P_s(j)$, que é o valor de probabilidade de um harmônico *j* ser escolhido como a saída do filtro, após a aplicação de uma função $e(j)$ de ênfase aos harmônicos mais baixos. Sua descrição e obtenção são detalhados na seção a seguir.

6.4 O vetor de probabilidades $P_s(j)$

Este vetor é definido tal que cada elemento $P_s(j)$ é a probabilidade de o *j*-ésimo harmônico de um conjunto de harmônicos $\{h_j\}$ ser escolhido como a saída do filtro $\nu\{h_j\}$. Para seu cálculo, deve-se estabelecer um limite λ de probabilidade para que um sinal interferente *R* gere, na saída, um parcial harmônico sobreposto (perturbado). Isso significa que, dado um número *N* de sons interferentes, existe um limite, λ, de probabilidade de se ter na saída do filtro ν um valor inválido procedente de um harmônico distorcido.

Dado um conjunto de harmônicos $\{h_j\}$, define-se E_m como o seu subconjunto que contém todos os harmônicos múltiplos de *m* a partir do *m*-ésimo harmônico:

$$E_m = \{h_{m.j}\}, \text{ para } j \text{ inteiro e positivo} \tag{6.6}$$

Uma vez que se percebe a presença de um tom interferente *R* se sobrepondo a harmônicos do tom observado, *S*, sabe-se que *R* alterará cada *m*-ésimo harmônico de *S*, ou seja, o subconjunto E_m deste som.

Para um número finito de harmônicos, *J*, pode-se dizer que os maiores subconjuntos E_m são aqueles formados pelos menores valores atribuídos a *m*, uma vez que o subconjunto E_1 engloba todos os harmônicos até *J*; E_2, os pares; E_3, os múltiplos de três; e assim por diante. A Figura 35 ilustra a relação entre o valor de *m* e a quantidade de parciais harmônicos presentes nos subconjuntos E_m, para *J* = 40.

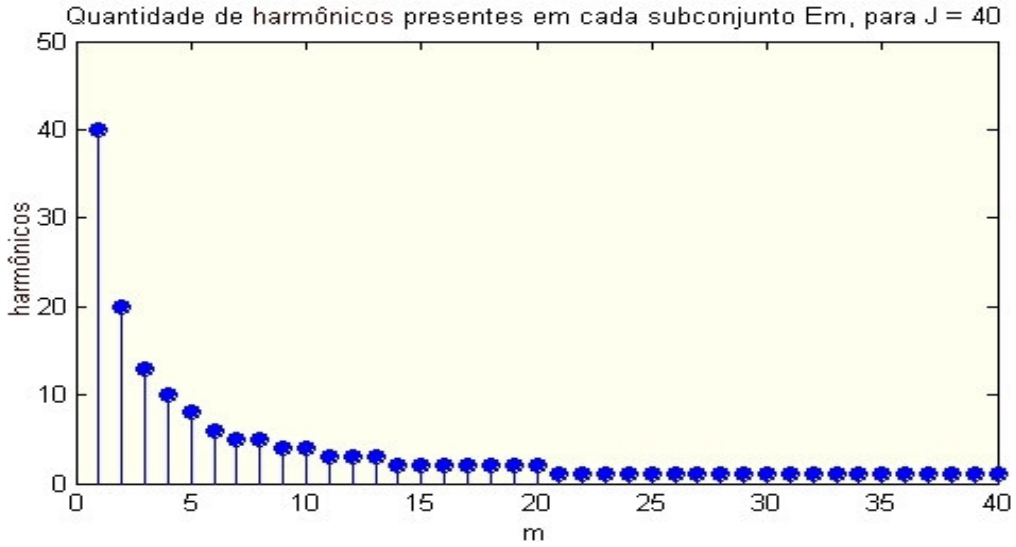


Figura 35 - Relação entre o valor de *m* e a quantidade de harmônicos presentes no subconjunto E_m para *J* = 40. Os maiores subconjuntos são aqueles para os quais *m* é menor, uma vez que haverá uma quantidade maior de múltiplos.

Desta forma, um dos requisitos do filtro ν é que a soma das probabilidades de seleção de amostra dos N maiores subconjuntos E_m atinja, no máximo, um certo limite de probabilidade, λ, ou seja:

$$\sum_{m=2}^{primos(N)} P_S^0(E_m) \leq \lambda, \tag{6.8}$$

em que *primos(N)* representa o N-ésimo número primo e $P_S^0(j)$ é o vetor inicial de probabilidade de seleção de um harmônico para a saída do filtro. O vetor $P_S(j)$ que será utilizado como peso do filtro será obtido a partir de $P_S^0(j)$ inicial, multiplicado por uma função de ênfase dos harmônicos mais baixos, a ser apresentada na Seção 6.4.2.

Nesta análise, consideram-se apenas os N maiores subconjuntos para valores de *m* primos, $\{E_m \mid m = 2,3,5,7...\}$, uma vez que qualquer outro valor de *m* seria múltiplo de um primo, fazendo com que seu respectivo subconjunto $\{E_m \mid m$ não primo $\}$ esteja contido em $\{E_m \mid m = 2,3,5,7...\}$.

A confiabilidade relativa não é a mesma para todos os harmônicos $\{h_j\}$. Desta forma, obtém-se $\tau^{D(j)}$, que representa a probabilidade de um harmônico $\{h_j\}$ ser confiável e, portanto, ser preferido na escolha como saída de filtro. O parâmetro τ representa a probabilidade geral de um som interferente se sobrepor a algum subconjunto E_m. $D(j)$ é o número de subconjuntos E_m ao qual um harmônico $\{h_j\}$ pertence, o que é o mesmo que o número de divisores de *j*. Assim, como a probabilidade de seleção de uma amostra deve estar de acordo com a probabilidade de cada harmônico ser confiável, pode-se dizer que:

$$P_S^0(j) = \tau^{D(j)}, \text{ para } j \geq 1 \tag{6.9}$$

Seja *J* o número total de harmônicos em um tom observado, a probabilidade geral de seleção dos harmônicos é dada por:

$$\sum_{j=1}^{J} \tau^{D(j)} = 1 \tag{6.10}$$

Seja *I* o conjunto formado pelos índices *j* dos harmônicos h_j que pertencem a alguns dos *N* maiores subconjuntos $\{E_m \mid m = 2,3,5,7...\}$. Desta forma, tem-se que, para $N = 1$, o conjunto *I* conterá os índices pares de 1 até *J*. Para $N = 2$, o conjunto *I* conterá os índices pares ou múltiplos de 3, dos harmônicos de 1 até *J*. E a soma de probabilidades de seleção dos harmônicos nestes *N* maiores subconjuntos é $\sum_{j \in I} \tau^{D(j)}$. Assim, o primeiro requisito do filtro ν a ser projetado pode ser reescrito como:

$$\sum_{j \in I} \tau^{D(j)} = \lambda \sum_{j=1}^{J} \tau^{D(j)} \Rightarrow \sum_{j \in I} P_S^0(j) = \lambda \tag{6.11}$$

Portanto, o vetor de probabilidade $P_S^0(j)$ inicial pode ser obtido a partir desse requisito.

6.4.1 Construção e parâmetros do vetor $P^0_S(j)$

O algoritmo que apresenta o cálculo de $P_S^0(j)$ é encontrado em [9]. Seus parâmetros são o limite λ, o número de harmônicos observados *J*, e a quantidade *N* de subconjuntos E_m a serem considerados. Satisfaz-se, então, o primeiro requisito do filtro ν (Equação (6.11)). Isso significa que *N* sons interferentes podem contribuir juntos com uma probabilidade máxima λ de que um harmônico sobreposto seja a saída do filtro. Um segundo requisito para o filtro é a utilização de todos os parciais harmônicos de forma a apresentarem amplitude tão próxima quanto possível para que o filtro seja robusto e aplicável a diversos tipos de sinal.

O algoritmo apresentado em [9] é bem flexível para atender a troca de prioridade entre os requisitos do filtro. Quanto menos robusto é o filtro na presença de sons interferentes (ou seja, quanto maior o limite λ), mais homogeneamente serão utilizados os harmônicos do som observado, dando ênfase ao segundo requisito. Da mesma forma, quanto menor λ, ou seja, quanto menos se tolerar a presença de interferência na saída, mais desigual será a distribuição de probabilidade de seleção, como podemos observar na Figura 36, que apresenta os vetores de probabilidade de seleção calculados para N = 1 e (a) λ = 0,5, (b) λ = 0,3 e (c) λ = 0,1.

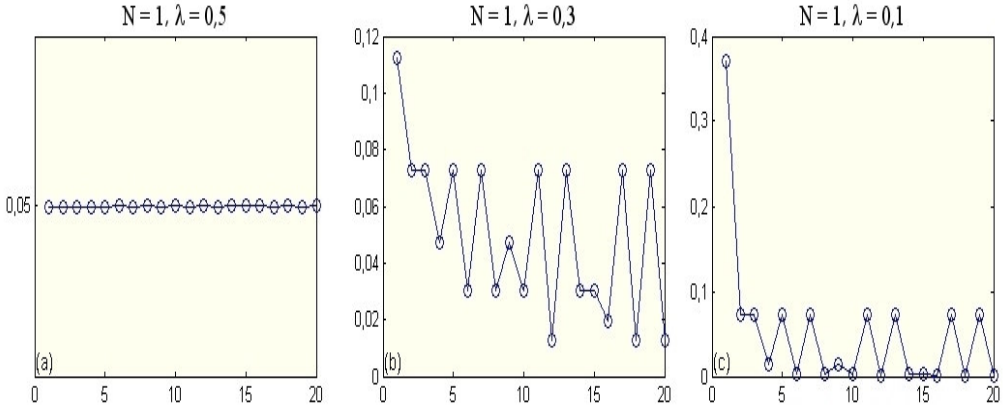


Figura 36 - Vetor $P^0_s(j)$ obtido para N = 1 e valores de λ variáveis. Quanto maior λ, mais desigual é a distribuição de $P^0_s(j)$.

Na Figura 36, observam-se três valores distintos para λ. No primeiro caso, $\lambda = 0{,}5$, a um som interferente é permitido perturbar metade da massa de $P_S^0(j)$,

resultando em uma distribuição igualitária, o que não o torna de forma alguma sensível, uma vez que até mesmo um único som é capaz de corromper a saída do filtro. No terceiro caso, $\lambda = 0{,}1$, a um som interferente é permitido perturbar, no máximo, 10% da massa de $P_S^0(j)$, o que resulta em praticamente só usar os harmônicos primos do som observado, fazendo com que o requisito de utilizar os harmônicos o mais igualmente possível seja ignorado.

A escolha de bons valores de parâmetro N e seu λ correspondente depende de quão rica for a polifonia interferente do outro som. Uma proposta mediana se torna mais interessante, uma vez que balanceia o atendimento de ambos os requisitos do filtro, como acontece no segundo caso, para $\lambda = 0{,}3$.

Um resultado exatamente igual, ilustrado na Figura 37, é obtido quando se alteram os parâmetros, permitindo que dois sons ($N = 2$) perturbem no máximo 45% ($\lambda = 0{,}45$) da massa de $P_S^0(j)$. Isso significa que dois sons não corromperiam a saída do filtro e, ainda assim, a distribuição $P_S^0(j)$ utilizaria razoavelmente bem os diferentes harmônicos do som observado.

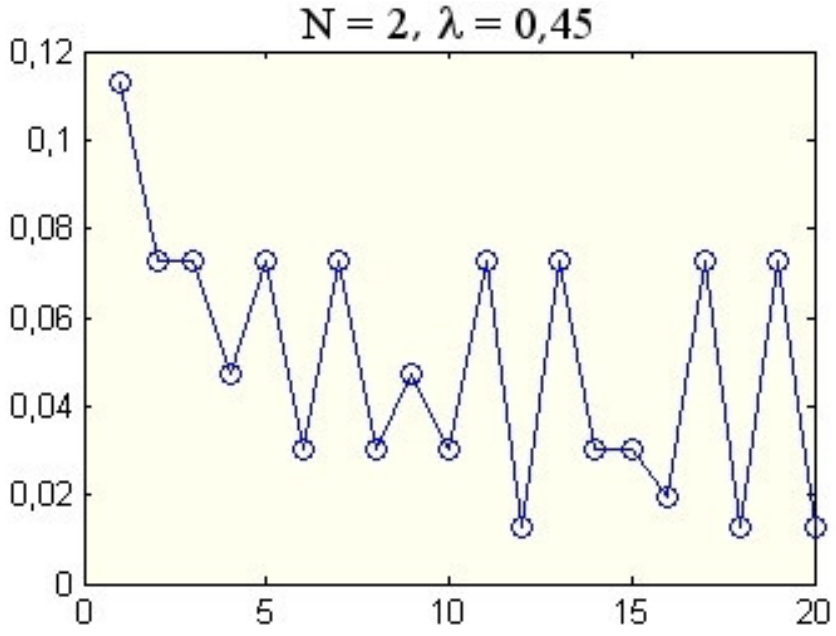


Figura 37 – P^0_S(j) para N = 2 e λ = 0,45

A referência [9] apresenta o resultado da comparação de porcentagem de massa perturbada para diversos parâmetros e diversos acordes, encontrando $N = 2$ e $\lambda = 0{,}40$ como bons valores para se obter um filtro robusto para dois sons interferentes. Para os valores citados, a saída do filtro praticamente não é alterada se a porcentagem for menor ou igual a 40%.

6.4.2 Ênfase aos harmônicos mais baixos

No início da Seção 6.3, foi dito que "os harmônicos não são igualmente confiáveis em valor" e que é necessário enfatizar os harmônicos mais baixos. Multiplicar os valores de $P_S^0(j)$ por uma função decrescente suave e monótona não altera substancialmente as somas de cada subconjunto E_m, uma vez que tais subconjuntos estão distribuídos por todo o comprimento de $P_S^0(j)$.

Esta função de ênfase, $e(j)$, deve ser selecionada de forma a seguir o nível geral dos harmônicos do som cujas características são observadas. Tal escolha não é crítica, e uma mesma ênfase pode ser dada a todos os sons de instrumentos.

A referência [9] apresentou uma função conveniente:

$$e(j) = \frac{1}{j+a} - \frac{1}{J+a+1}, \tag{6.12}$$

para a qual *J* é o número de amostras de harmônicos do sinal, ou seja, a quantidade de harmônicos que estão sendo considerados no sinal, *j* é o índice e está relacionado com *j*-ésimo harmônico, e *a* é o parâmetro de ajuste, que varia de acordo com o instrumento a ser trabalhado. Esta função se ajusta bem para a intensidade relativa dos harmônicos de sons diferentes. Para a observação de sons de piano, sugere-se definir $a = J$ [9], de forma a se obter:

$$e(j) = \frac{1}{j+J} - \frac{1}{2J+1}, \tag{6.13}$$

A Figura 38 ilustra como é o comportamento da função de ênfase $e(j)$, para $a=J$. A aplicação desta função resulta em enfatizar os primeiros harmônicos, mais baixos, em detrimento dos mais altos.

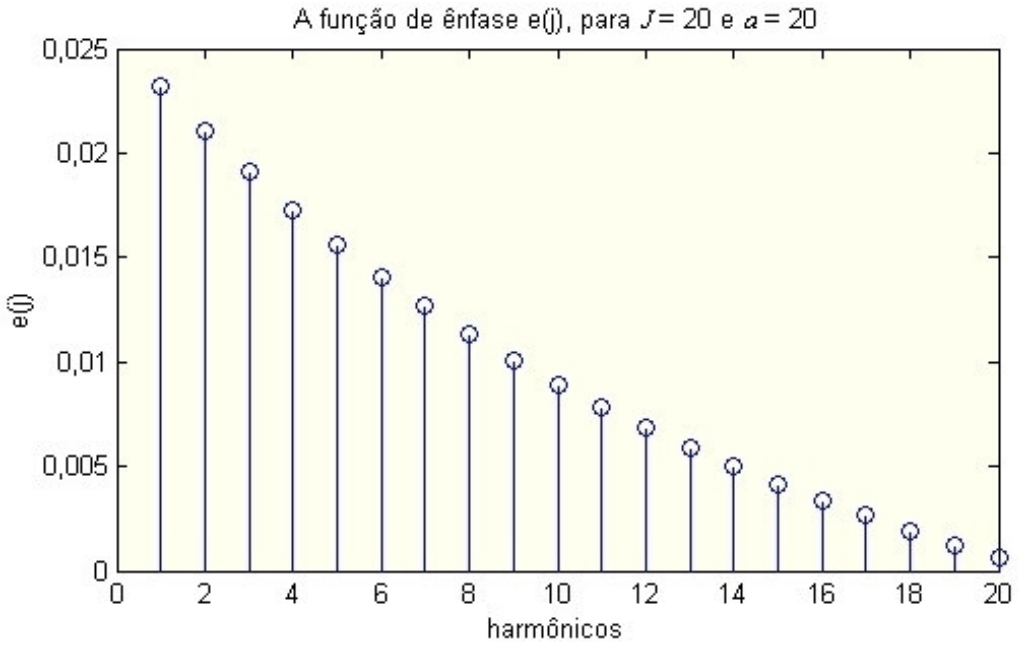


Figura 38 - A aplicação desta função resulta em enfatizar os primeiros parciais harmônicos, de freqüências mais baixas, em detrimento dos que possuem freqüências mais altas.

A partir da aplicação da função de ênfase, finalmente obtém-se um vetor final $P_S(j)$ de probabilidade de seleção de amostra para um filtro ν, que seleciona o valor estimado de uma característica de um som a partir do conjunto de características observadas em seus harmônicos, ignorando os valores irrelevantes:

$$P_S(j) = e(j).P_S^0(j), \tag{6.15}$$

para o qual $P_S^0(j)$ é o vetor normalizado obtido com o algoritmo 1 em [9], a partir dos parâmetros *J*, *N* e λ; e $e(j)$ é a função de ênfase. Após isso, a soma de $P_S(j)$ deve ser escalada novamente para a unidade, uma vez que o somatório da probabilidade deve ser 1. A Figura 39 ilustra o efeito da multiplicação por $e(j)$, comparando-se (a) $P_S^0(j)$ e (b) $P_S(j)$, para $J = 20$, $N = 2$, $\lambda = 0.4$ e $a = J$.

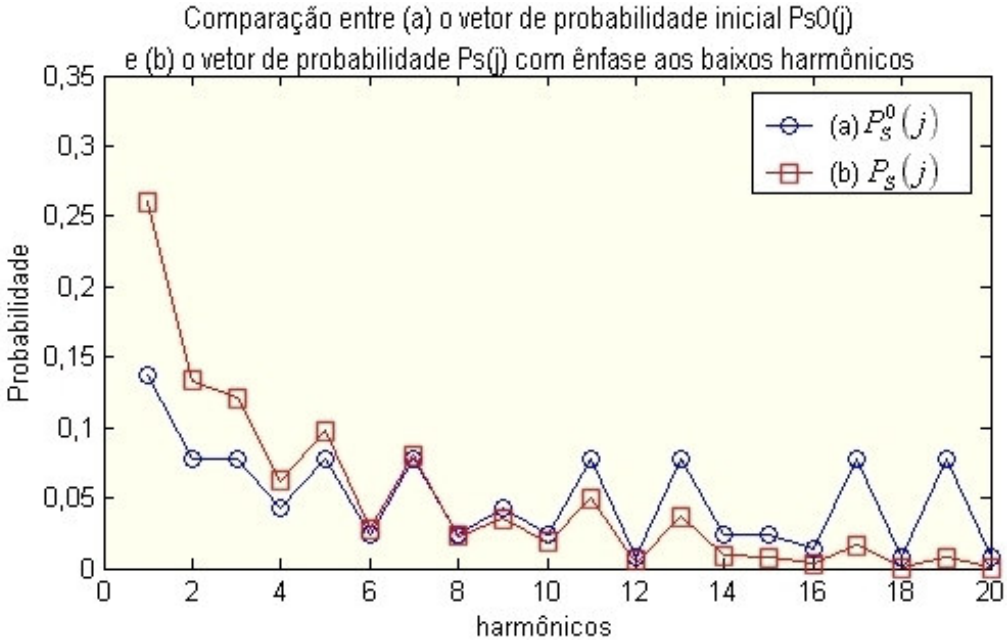


Figura 39 - Comparação entre o vetor de probabilidade inicial P^0_S(j) e o vetor de probabilidade de seleção P_S(j), resultante da aplicação da função de ênfase *e*(j) ao P^0_S(j). Observe que a probabilidade de seleção dos cinco primeiros harmônicos aumenta sensivelmente. Da mesma forma, a probabilidade de seleção dos harmônicos maiores que 10 diminui de forma considerável, principalmente nos harmônicos primos (11, 13, 17 e 19).

6.4.3 A implementação do filtro WOS a partir de P_S(j)

O próximo passo é projetar o filtro que implemente as propriedades estatísticas relativas aos vetor $P_S(j)$ quando aplicado a um conjunto de características dos harmônicos de um som. O princípio básico do filtro já foi apresentado na Figura 34 e a descrição a seguir se baseia no algoritmo 2, em [9].

A idéia básica é inicializar os pesos $w(j)$ do filtro da seguinte forma:

$$w(j) = (P_S(j))^{0,75} \tag{6.16}$$

Após, escala-se a soma geral de $w(j)$ para um valor *x*, que será o tamanho do conjunto, após a aplicação dos pesos, gerando $w(j)$ cópias de cada *j*-ésima amostra. A complexidade do filtro WOS independe de *x*, mas a complexidade e a precisão deste algoritmo, sim. Um valor razoável para *x* é 100 [9].

Se o número de amostras (ou seja, *J*, o número de harmônicos existentes em cada sinal) for maior ou igual a 15, a aproximação é precisa o suficiente. Caso

contrário, é necessário um procedimento cíclico de refinamento, encontrado em [9] e [13], até que o vetor de pesos $w(j)$ convirja, o que provoca um consumo de memória de complexidade $O(2^J)$, tornando o algoritmo pesado para grandes valores de *J*. Ao mesmo tempo, com o aumento de *J* a aproximação inicial, a equação (6.16), torna-se mais precisa, tornando desnecessários os passos seguintes para $J \geq 15$. Em função disso, neste projeto, assumimos o valor de *J* como sendo 20. Isso significa que, de cada nota, podem ser obtidos 20 harmônicos, o que reduz a complexidade computacional. Em compensação, o fato de utilizarmos 20 harmônicos limita o alcance de notas para, no máximo, a freqüência fundamental de 1046,5 Hz (considerando a referência em A = 440 Hz), uma vez que 20 vezes esta freqüência resulta em 20,930 kHz, quase no limite da freqüência de Nyquist de 22,050 kHz, associada à taxa de amostragem de 44,100 kHz, utilizada neste projeto. A partir desta freqüência, não é possível montar o vetor de amostras de tamanho 20.

A escolha do parâmetro de limite *T* do filtro WOS é baseada no conhecimento da distribuição estatística dentre dados errôneos muito pequenos ou muito grandes. Para fins de transcrição musical, o valor médio $T = \left|\frac{x}{2}\right|$ é comumente utilizado para a escolha [9], visto que o modelo impreciso de instrumentos tende a causar quantidades parecidas de erros de valores pequenos e grandes. O valor médio para o limite *T* foi considerado suficiente para a aplicação neste projeto. Como foi admitido $x = 100$, resulta $T = 50$.

6.4.4 Parâmetros de implementação

O filtro WOS foi construído utilizando os procedimentos descritos anteriormente, sendo usados como parâmetros:

- $N = 2$;
- $\lambda = 0{,}4$;
- $J = 20$, limitando a freqüência máxima a 1046,5 Hz, C6 do piano;
- $a = J$, para a função de ênfase $e(j)$; e
- $x = 100$, para o número de elementos na implementação do filtro WOS.

6.5 Conclusões

No início do projeto, foi mencionado que a maior dificuldade na resolução de um trecho polifônico é a sobreposição dos harmônicos. Como as bandas de freqüência são limitadas, freqüentemente as componentes de freqüências múltiplas da fundamental irão se sobrepor, o que acarreta em alterações na proporção dos valores esperados para determinadas freqüências, fazendo com que sejam gerados erros na saída do sistema, como, por exemplo, no cálculo da intensidade.

A partir de então, houve questionamentos sobre que tipos de informação seriam confiáveis e o que poderia ser usado para a obtenção de valores próximos aos reais. Inicialmente, discutiu-se a importância dos harmônicos primos, uma vez que, sendo um número primo divisível por ele mesmo ou por um, eles seriam menos sujeitos a sobreposição, apresentando amplitudes com menos probabilidade de interferência. Mas considerar apenas os harmônicos primos de um sinal também é perigoso, uma vez que, dependendo do timbre do instrumento, poderia ocorrer de a maioria deles serem inexistentes no espectro, além do fato de serem poucas amostras, o que não é aconselhável.

Apresentou-se então o conceito de um vetor de probabilidade de seleção, $P_S(j)$, representando a probabilidade de um determinado harmônico ser escolhido como a saída de um filtro seletor de característica. Este cálculo teria como parâmetros um valor máximo λ de limite de probabilidade e um valor N correspondente ao número de sons. Isso significa que N sons interferentes podem contribuir juntos com uma probabilidade máxima λ de que um harmônico sobreposto seja a saída do filtro. O parâmetro λ representa, portanto, o quanto se tolera a presença de interferência na saída. Valores muito altos de λ ($\geq 0{,}5$) fazem com que o sistema seja pouco eficiente, tornando-o pouco seletivo, uma vez que qualquer tipo de sinal seria capaz de corromper a saída. O oposto força o sistema a ser excessivamente seletivo, levando em consideração basicamente os harmônicos primos, o que não é aconselhável pelos fatores apresentados logo acima. A solução é procurar valores intermediários, sendo escolhidos $N = 2$, $\lambda = 0{,}4$ e $J = 20$.

Discutiu-se sobre a necessidade de enfatizar os harmônicos de freqüências mais baixas, uma vez que neles normalmente está concentrada a energia do som. As freqüências mais altas, por terem menor intensidade, estão mais sujeitas a ruídos e interferências. Utilizou-se uma função de ênfase $e(j)$, a ser multiplicada pelo vetor de probabilidade de seleção, apresentada na Equação (6.12). Assim, foi feito o cálculo do filtro WOS, que corresponde à aplicação dos pesos $w(j)$ na repetição dos valores de um determinado sinal, sua posterior ordenação e a seleção do valor mediano do vetor.

O vetor de probabilidade de seleção de amostra, $P_S(j)$, e o filtro WOS, $v\{.\}$, serão utilizados na determinação da freqüência fundamental monofônica (na escolha de um candidato dentre outros e na obtenção de um valor de freqüência fundamental f_0 mais preciso) e na determinação da freqüência fundamental polifônica (no cálculo do peso de um candidato, ou seja, o quanto um candidato é capaz de se encaixar em um modelo existente no banco de modelos de tom), sendo devidamente explicados em seus contextos no Capítulo 7 e no Capítulo 8, respectivamente.

Capítulo 7 - Criação dos modelos de tom

Neste capítulo, apresenta-se o módulo de criação de modelos de tom. Tal processo se define pelo registro de todas as notas, uma a uma, de um determinado instrumento, a fim de representar todas as possíveis *cores de tom* produzidas por tal instrumento. No caso do piano, por exemplo, isso significaria gravar cada uma de suas teclas, separadamente, correspondendo ao *material de treinamento*. Para a geração destes modelos, assume-se, então, que apenas um som é tocado por vez, determinando-se a freqüência fundamental e a cor de tom de cada som musical. Este bloco receberá o suporte do núcleo de filtro gerado no capítulo anterior.

Na Seção 7.1, apresenta-se a visão geral do processo de modelagem de tom, debatido superficialmente no Capítulo 3. Segue a Seção 7.2 ilustrando o algoritmo proposto e sua implementação. Na Seção 7.3, há um resumo e as conclusões do tema.

7.1 Visão geral do módulo de modelagem de tom

O módulo de modelagem de tom é o principal elemento no processo de treinamento, cujo esquema geral pode ser revisto na Figura 40, de forma a ilustrar o contexto no qual o bloco está inserido. A entrada do módulo será, portanto, a transformada B*Q*T de cada um dos sons de treinamento. Sua principal função é determinar a freqüência fundamental f_0 do sinal de entrada, obtendo-se também as amplitudes relativas a seus *J* harmônicos. Tais dados devem ser armazenados em um banco de dados ou em registros de arquivos, de forma a serem consultados quando no processo de reconhecimento. O número *J* de harmônicos será 20 e o processo de extração de dados aplicará o vetor de probabilidades $P_S(j)$, conforme estabelecido no capítulo anterior.

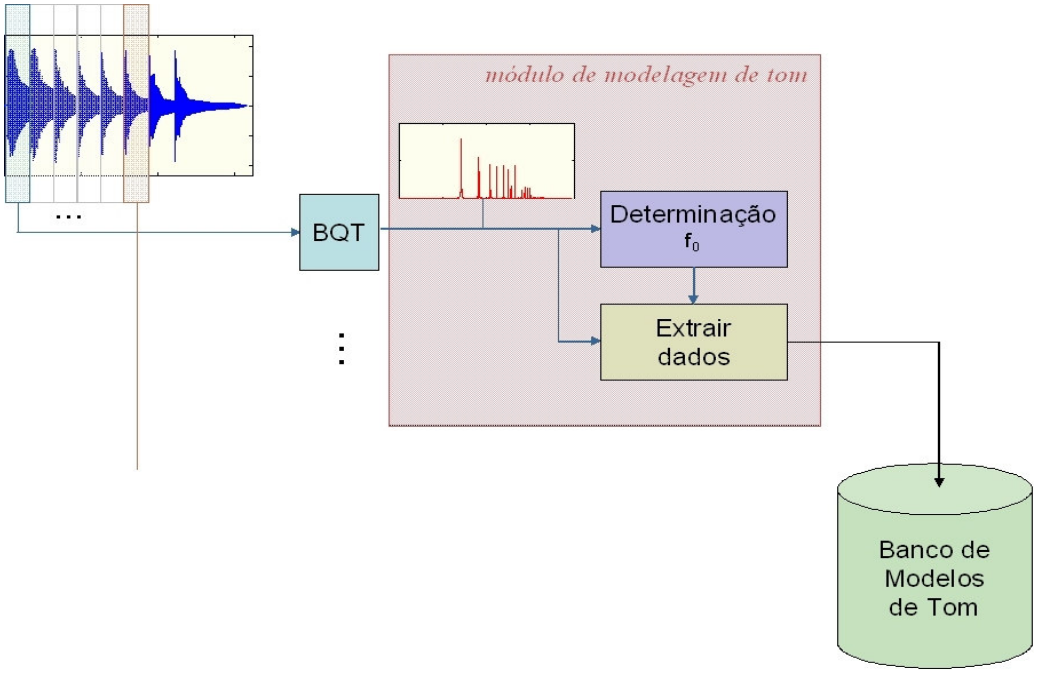


Figura 40 - Esquema geral do processo de treinamento de tom, visando mostrar o contexto no qual o módulo de modelagem de tom se encontra. Este será responsável pela determinação da freqüência fundamental f_0 e pela extração das amplitudes respectivas aos seus harmônicos, e seu posterior armazenamento no banco de modelos de tom, que pode ser um banco de dados ou registro em arquivos.

7.2 Implementação do Módulo de Modelagem Tonal

A determinação da freqüência fundamental monofônica não é trivial, mas mais fácil se comparado com a polifônica. Procuram-se picos no espectro, *P*, que sejam maiores que um determinado limite. A média da amplitude da B*Q*T do sinal foi utilizada como um bom parâmetro de limite para a procura de tais picos, sendo multiplicada por um fator de acordo com cada oitava musical, como ilustrado pela Tabela 6.

Tabela 6 – Valores-limite escolhidos para a filtragem do sinal em cada uma das oitavas musicais. Só são levados em consideração picos em potencial cuja amplitude esteja acima deste valor.

Oitava	Valores-limite escolhidos
De C2 até C3 (65,4-130,8 Hz)	4*média(B*Q*T)
De C3 até C4 (130,8-261,6 Hz)	9,5*média(B*Q*T)
De C4 até C5 (261,6-523,2 Hz)	9,5*média(B*Q*T)
De C5 até C6 (523,2-1046,5 Hz)	9,5*média(B*Q*T)

A partir disso, procura-se, em ordem crescente de freqüência, o primeiro pico que seja maior que os valores-limite definidos na Tabela 6. A partir deste pico inicial, admitem-se três candidatos a freqüência fundamental f_0: $C_1 = f_P$, $C_2 = f_P / 2$ e $C_3 = f_P / 3$, ou seja, a freqüência f_P do primeiro pico no espectro, a metade e a terça parte desta. O cálculo envolvendo tais freqüências se deve a possibilidade de a fundamental não possuir amplitude expressiva no tom analisado.

Para cada candidato, cria-se um vetor $H_A(j)$ com as amplitudes referentes aos seus harmônicos. Cada parcela harmônica pode ser considerada como um pedaço de evidência da existência deste candidato. Desta forma, enfatiza-se cada vetor $H_A(j)$ de harmônicos, aplicando-lhe diretamente como pesos os valores constantes no vetor de probabilidade $P_S(j)$, obtido conforme a descrição na Seção 6.4. Comparam-se os resultados obtidos, de forma que o candidato para o qual se encontrar o valor mais elevado é escolhido como a freqüência fundamental mais provável, f_C.

Por fim, para se encontrar um valor mais preciso para a freqüência f_0, deve-se procurar no espectro as posições múltiplas da freqüência fundamental f_C, ainda não precisa, formando, a partir disso, o vetor $H_F(j)$, com os valores das freqüências relativas a cada harmônico *j*, desde que tal harmônico seja um pico. Caso contrário, para tal harmônico será atribuído um valor nulo em $H_F(j)$. Calcula-se a média ponderada a partir deste vetor, fazendo uso de $P_S(j)$ como peso. Deve-se observar que o peso será zero para harmônicos que não representarem picos no espectro. A equação (7.1) resume este procedimento de procura de uma freqüência fundamental mais precisa:

$$f_0 = \frac{\sum_{j=1}^{J} \frac{H_F(j).P_S(j)}{j}}{\sum_{j=1}^{J} P_S(j)}, \text{ para valores de } j \text{ associados a picos no espectro.} \quad (7.1)$$

A freqüência f_0 mais precisa é o resultado da média ponderada, que é utilizada no lugar do filtro WOS por dois motivos. Primeiro porque dificilmente haverá algum valor corrompido no vetor de freqüências de harmônicos. Segundo porque nenhum valor relativo à freqüência a partir de um único harmônico é bom o suficiente para ser escolhido sozinho como um representante confiável da freqüência fundamental.

As amplitudes dos harmônicos de um tom são encontradas procurando-se as posições dos harmônicos e selecionando suas amplitudes respectivas, armazenando-as em um vetor $H_A(j)$.

7.2.1 Implementação

Algoritmo 1	createModel.m

- parâmetros de entrada: B*Q*T de um sinal monofônico, constituído por um único tom, para efeito de treinamento;

- saída:

freqüência fundamental f_0;

vetor contendo a amplitude dos harmônicos $H_A(j)$;

arquivo *snotes.dat*

- **Passo 1:** Calcular o primeiro pico em potencial para se obter o primeiro candidato a freqüência fundamental, f_P. Para evitar erros indesejados, a implementação desta função levou em consideração os limites do projeto (F_{min} = 97 Hz e F_{max} = 2093 Hz), e a procura se deu por picos a partir de F_{min}, em faixas divididas pelas oitavas.
- **Passo 2**: A partir do primeiro candidato, $C_1 = f_P$, obter os outros possíveis candidatos $C_2 = f_P / 2$ e $C_3 = f_P / 3$ e montar o vetor de harmônicos, com suas respectivas amplitudes;
- **Passo 3:** Calcular os valores do vetor de probabilidade de seleção $P_S(j)$, multiplicado pela função de ênfase $e(j)$. Em seguida aplicá-los nos vetores dos harmônicos dos possíveis candidatos para a escolha de f_0 ainda não precisa, a qual se chamou de f_C;
- **Passo 4:** Uma vez escolhido o melhor candidato, precisar o valor de f_C;
- **Passo 5:** Para este valor preciso de f_0, montar o novo vetor com as amplitudes dos harmônicos respectivos. Armazenar estes dados em um arquivo de notas.

  Numa melhoria futura, seria interessante fazer uso de um banco de dados para isso. No entanto, para uma aplicação pequena e com poucos modelos de tom, apenas um arquivo de gerenciamento se mostra suficiente.

Junto com o arquivo de gerenciamento *snotes.dat*, a B*Q*T de cada sinal de teste foi armazenada em um arquivo respectivo a cada nota.

A regra de formação do arquivo binário *snotes.dat* segue abaixo. Para cada nota, temos o seguinte conjunto:

- F_0 = um número com precisão *double*;
- *[Ha(J)]* = um vetor com vinte elementos com precisão *double*;

- *Nome da nota* = uma *string* com quatro caracteres, com o nome da nota, sendo a regra de formação:

1º caractere: definição de instrumento;

2º caractere: definição da oitava;

3º caractere: a letra representativa da nota;

4º caractere: _ para notas sem modificação;

para notas sustenidas;

Exemplo: P5C# significa: Dó sustenido da quinta oitava de piano.

- *Nome do arquivo com a BQT relativa* = uma *string* com oito caracteres, sendo os quatro primeiros o nome da nota e os quatro últimos ".dat"

A Figura 41 ilustra como ficaram os dados obtidos para notas de piano. Por se tratar de um arquivo binário, utilizou-se uma rotina criada especialmente para a leitura e exibição dos dados, como a seguir.

```
  F0   |  Nota  |                                HA
-------------------------------------------------------------------------------
130.2     P3C_     2.5e-003 - 1.8e-002 - 2.5e-003 - 1.0e-003 - 1.3e-003 - 2.0e-003 -
138.7     P3C#     2.0e-003 - 1.3e-002 - 4.8e-003 - 4.6e-003 - 6.4e-003 - 1.3e-002 -
 146      P3D_    2.1e-003 - 1.3e-002 - 1.9e-003 - 3.0e-004 - 7.2e-004 - 1.6e-003 -
155.7     P3D#     1.9e-003 - 1.1e-002 - 4.5e-003 - 4.0e-003 - 1.1e-002 - 1.2e-002 -
163.9     P3E_     2.3e-003 - 1.2e-002 - 2.5e-003 - 8.9e-005 - 6.8e-004 - 9.8e-004 -
 175      P3F_    6.6e-003 - 7.6e-003 - 8.5e-003 - 7.4e-003 - 2.4e-003 - 9.8e-003 -
185.1     P3F#     6.6e-003 - 6.8e-003 - 8.2e-003 - 7.3e-003 - 3.1e-003 - 3.9e-003 -
195.9     P3G_     6.5e-003 - 6.1e-003 - 9.4e-003 - 7.1e-003 - 2.2e-003 - 2.5e-003 -
207.9     P3G#     1.0e-002 - 1.7e-002 - 8.1e-003 - 1.0e-002 - 4.6e-003 - 1.5e-003 -
220.1     P3A_     1.4e-002 - 2.0e-002 - 8.8e-003 - 6.4e-003 - 8.3e-003 - 4.4e-004 -
232.2     P3A#     1.1e-002 - 7.9e-003 - 6.3e-003 - 1.2e-003 - 1.5e-003 - 1.6e-003 -
247.6     P3B_     1.1e-002 - 1.3e-002 - 1.4e-002 - 8.0e-003 - 7.6e-003 - 8.3e-003 -
261.8     P4C_     1.3e-002 - 1.8e-002 - 1.2e-002 - 5.0e-003 - 2.9e-003 - 2.3e-003 -
276.6     P4C#     1.0e-002 - 1.1e-002 - 2.5e-003 - 2.2e-003 - 6.1e-004 - 1.5e-003 -
294.1     P4D_     9.5e-003 - 1.9e-002 - 1.7e-002 - 6.7e-003 - 7.5e-003 - 2.2e-003 -
 311      P4D#    9.2e-003 - 1.7e-002 - 1.6e-002 - 2.4e-003 - 8.1e-004 - 1.4e-003 -
329.8     P4E_     9.5e-003 - 2.0e-002 - 1.9e-002 - 2.2e-003 - 6.2e-003 - 1.6e-003 -
 350      P4F_    2.9e-002 - 3.6e-002 - 8.1e-003 - 1.8e-003 - 6.3e-004 - 3.9e-004 -
371.5     P4F#     3.0e-002 - 3.1e-002 - 1.1e-002 - 7.0e-003 - 1.1e-002 - 2.6e-003 -
 393      P4G_    3.1e-002 - 3.4e-002 - 1.3e-002 - 6.6e-003 - 2.7e-003 - 1.3e-003 -
415.9     P4G#     2.9e-002 - 2.6e-002 - 1.2e-002 - 1.9e-003 - 1.5e-003 - 9.8e-004 -
441.5     P4A_     3.9e-002 - 3.8e-002 - 8.0e-003 - 8.0e-003 - 7.1e-003 - 3.3e-004 -
468.5     P4A#     3.0e-002 - 2.9e-002 - 2.0e-002 - 5.9e-003 - 9.3e-003 - 6.4e-003 -
493.9     P4B_     3.8e-002 - 3.0e-002 - 4.8e-003 - 2.8e-003 - 1.0e-003 - 4.1e-004 -
523.6     P5C_     3.8e-002 - 2.6e-002 - 1.8e-002 - 2.9e-003 - 1.5e-003 - 1.2e-004 -
554.5     P5C#     3.2e-002 - 3.0e-002 - 1.5e-002 - 3.0e-003 - 1.2e-003 - 3.8e-004 -
588.2     P5D_     3.8e-002 - 2.7e-002 - 2.8e-002 - 4.2e-003 - 2.0e-003 - 8.0e-005 -
621.7     P5D#     3.6e-002 - 2.0e-002 - 3.3e-003 - 5.2e-003 - 7.8e-004 - 4.8e-004 -
659.6     P5E_     3.8e-002 - 3.4e-002 - 2.3e-002 - 3.0e-003 - 2.0e-003 - 3.7e-004 -
699.9     P5F_     6.5e-002 - 3.4e-002 - 1.2e-002 - 6.4e-004 - 2.1e-004 - 3.5e-004 -
740.1     P5F#     6.8e-002 - 2.4e-002 - 4.7e-003 - 1.1e-003 - 2.4e-004 - 4.0e-004 -
783.3     P5G_     6.1e-002 - 1.9e-002 - 8.4e-003 - 1.3e-003 - 4.5e-004 - 2.7e-003 -
834.3     P5G#     5.3e-002 - 3.6e-002 - 1.8e-002 - 2.6e-003 - 6.6e-004 - 1.0e-003 -
882.9     P5A_     6.3e-002 - 4.6e-002 - 1.7e-002 - 1.5e-003 - 2.4e-004 - 7.9e-004 -
 934      P5A#    7.2e-002 - 3.0e-002 - 7.8e-003 - 1.0e-003 - 5.0e-004 - 4.9e-004 -
990.5     P5B_     6.8e-002 - 4.3e-002 - 2.0e-002 - 1.1e-003 - 5.6e-004 - 8.3e-004 -
1047      P6C_    5.6e-002 - 2.4e-002 - 7.9e-003 - 4.5e-004 - 2.1e-004 - 1.3e-004 -
>>
```

Figura 41 - f_0 e os valores de amplitude dos seus seis primeiros harmônicos para notas de piano.

7.3 Conclusões

O módulo de criação de modelos de tom se encontra inserido no processo de treinamento do sistema de reconhecimento de notas polifônicas. Seu principal objetivo é, a partir da transformada B*Q*T de um sinal monofônico, extrair suas informações relativas à freqüência fundamental, às freqüências dos harmônicos e à intensidade dos mesmos, para armazenamento, ordenação e posterior consulta. Desta forma, tem-se o registro de todas as notas, uma a uma, de um determinado instrumento, a fim de representar todas as possíveis *cores de tom* produzidas por ele. No caso do piano, por exemplo, isso significou gravar cada uma de suas teclas,

separadamente, dentro da faixa de operação deste sistema, que é de 97 Hz até 1046,5 Hz.

Apresenta-se a visão geral do funcionamento do módulo de criação de modelos de tom. Dentre os valores constantes na transformada BQT, procuram-se os seus picos, ou seja, os pontos nos quais o valor do meio é maior que o anterior e o posterior, simultaneamente. Em ordem crescente de freqüência, o primeiro pico com amplitude maior que determinados valores-limite, apresentados na Tabela 6, é então considerado um candidato a freqüência fundamental. A metade e a terça parte do valor desta freqüência são utilizados, também, para efeito de cálculo. Aquela para a qual se obter maior intensidade geral é admitida como a freqüência fundamental verdadeira.

A partir daí são armazenados os dados, constituídos basicamente em uma freqüência fundamental, as amplitudes de seus J-ésimos primeiros harmônicos (J=20), o nome da nota (fornecido como uma string na entrada do bloco) e o arquivo respectivo que contenha a transformada BQT do mesmo. Isto foi feito em um sistema de arquivos binários.

Capítulo 8 - Reconhecimento

Nos capítulos anteriores, montaram-se os blocos que suportam a principal parte do projeto, que é o módulo de reconhecimento de notas. Não obstante seja este o núcleo do sistema, é requisito fundamental que todos os seus módulos estejam devidamente integrados e associados, de forma que parâmetros comuns estejam igualmente definidos para todas as suas funções.

Neste capítulo, revisa-se o processo de reconhecimento, cuja entrada é o sinal musical a ser transcrito, junto com os modelos de tom e os núcleos de filtro dos dois processos de apoio citados anteriormente. Seu objetivo é conseguir resolver cada segmento de som, um a um, de forma a obter todas as notas integrantes de cada segmento, fazendo com que as saídas do processo, como um todo, sejam:

- Os tempos de cada segmento, o que implica conhecer o tempo de início de cada nota; e
- A freqüência fundamental, a amplitude e o tipo de instrumento em cada segmento de sinal.

O módulo de reconhecimento está inserido neste processo como a parte responsável por, a partir da transformada B*Q*T dos segmentos do sinal a ser analisado, identificar as freqüências fundamentais presentes, as amplitudes de seus respectivos harmônicos e a intensidade geral de cada um dos tons presentes naquele fragmento.

Na Seção 8.1 revisa-se a visão geral do módulo de reconhecimento. A partir disso, a Seção 8.2 apresenta o módulo de reconhecimento, determinações de seus parâmetros, implementação e algoritmo proposto, explicando as escolhas feitas e as limitações implicadas por cada uma delas. Finalmente, a Seção 8.3 apresenta as conclusões e o resumo deste capítulo.

As simulações do sistema serão apresentadas no Capítulo 9.

8.1 Visão geral do processo de reconhecimento

Os sinais de entrada são cada um dos segmentos de som, tratados em separado, já no domínio da freqüência, uma vez aplicada a *transformada de* Q *limitado* (B*Q*T). Desta forma, para cada segmento é realizada uma operação de resolução de tom, que é o coração deste processo, recebendo como entradas de apoio os modelos de tom e o núcleo de filtro.

É feita a varredura à procura de picos no espectro de freqüências, ou seja, pontos cujo valor é, simultaneamente, maior que o adjacente anterior e que o adjacente posterior. Além disso, para ser considerado um pico, é necessário que seu valor associado seja maior que um limite estabelecido e que não esteja muito próximo de outro com maior valor. Estas condições serão melhores descritas em momento oportuno.

Com isso, cria-se um vetor de candidatos a freqüências fundamentais existentes naquele segmento. Tais candidatos deverão ser tratados separadamente sempre da menor para a maior freqüência, para evitar que o harmônico de um sinal seja tratado como possível freqüência fundamental. Este erro é evitado quando se determina a intensidade do som da respectiva candidata f_0 analisada e a subtraímos do sinal analisado, retirando assim os seus harmônicos, antes de seguir adiante na procura.

A Figura 42 mostra o esquema geral do funcionamento interno do módulo de reconhecimento e o fluxo de dados.

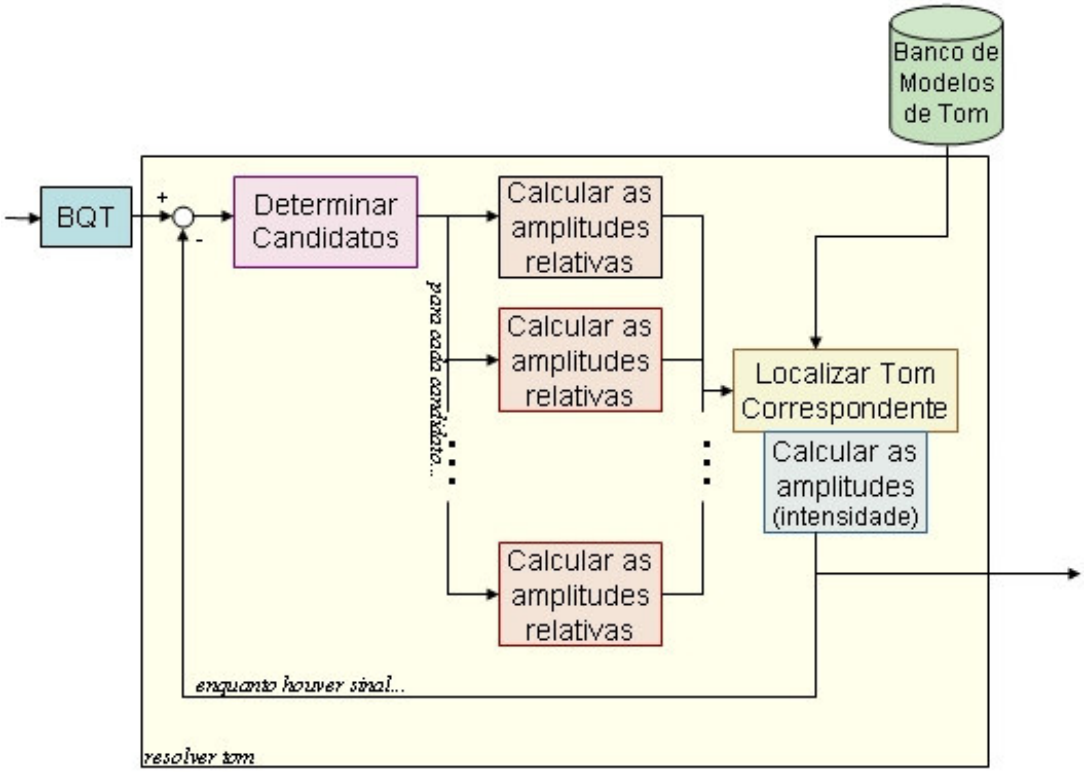


Figura 42 - Esquema geral do funcionamento do módulo de reconhecimento

A identificação do som é feita por uma comparação com o modelo existente no banco de tons, criado na ocasião de treinamento do sistema. Este processo se repete até que não haja mais nenhum valor expressivo indicando a existência de um tom ou até se atingir a máxima freqüência fundamental suportada pelo sistema (1046,5 Hz). Neste projeto, foi arbitrado que, quando a média do sinal resultante da subtração de tons identificados atingisse um valor inferior a 0,1 da média do sinal

analisado original, o restante seria admitido como resíduos dos tons identificados e que nenhum outro tom estaria presente.

8.2 Implementação e Algoritmos

8.2.1 A escolha dos candidatos

Para que se encontrem os candidatos a freqüências fundamentais, C_i, do espectro do sinal, é preciso coletar todos os picos do espectro, a fim de que não se ignorem os sons existentes. Os picos que estiverem abaixo de um valor mínimo de intensidade devem ser descartados, assim como aqueles que estiverem perto demais de um candidato mais forte. Para isso, é importante definir qual o intervalo mínimo entre duas freqüências fundamentais, $\Delta f_{0_{\min}}$.

Na música, a distância entre duas notas adjacentes é definida por:

$$\Delta f_{0_{\min}} = \frac{f_{0_{x+1}}}{f_{0_x}} = 2^{1/12} \text{ Hz} \tag{8.1}$$

Caso sons de intensidade muito baixa estejam sendo considerados, é necessário gerar os candidatos C_i com freqüência duas vezes menor que os menores candidatos, uma vez que o harmônico fundamental de sons baixos pode estar faltando ou possuir intensidade muito baixa se comparada com a dos harmônicos sobrepostos por outros sons.

8.2.2 Parâmetros para avaliação dos candidatos

A seguir, serão introduzidos dois parâmetros para que seja feita a identificação da presença de uma determinada freqüência fundamental em um som polifônico. Os dois conceitos e como são obtidos são explicados, e, então, é proposta a implementação do algoritmo apresentado em [9]. Os resultados são discutidos.

8.2.2.1 O peso de um determinado modelo tonal T_i no candidato C

Define-se um parâmetro $P(T_i, C)$ como o *peso*, o quanto um candidato C é capaz de se encaixar em um modelo de tom T_i no espectro e, conseqüentemente, nas

suas posições das séries harmônicas. Em termos gerais, a característica *peso* pode ser definida, individualmente, para cada harmônico pela seguinte relação:

$$P(h_j, T_i, C) = \frac{A(h_j, C)}{A(h_j, T_i)}, \tag{8.2}$$

em que $A(h_j, C)$ representa a amplitude do *j*-ésimo harmônico da nota candidata, escolhido na sua posição no espectro da mesma forma que no modelo de criação de tom, e $A(h_j, T_i)$ representa a amplitude do *j*-ésimo harmônico no modelo de tom.

Em um caso ideal, se o tom está sozinho no sinal, o ruído é ignorado e o modelo de tom perfeito $P(h_j, T_i, C)$ é o mesmo para todos os harmônicos, determinando o peso para esta instância particular do tom no sinal. Tais condições, no entanto, não são válidas para sons musicais verdadeiramente polifônicos, especialmente porque, na presença de um som interferente, tem-se a sobreposição dos *mj*-ésimos harmônicos do candidato, tornando-se necessário calcular o peso geral do modelo tonal T_i no candidato *C*:

$$P(T_i, C) = v\{P(h_j, T_i, C)\},$$

onde *v{.}* é o filtro WOS desenvolvido no Capítulo 6.

Como é previsto que a inspeção por candidatos seja em ordem crescente de valores de freqüência (evitando-se que um harmônico sobreposto seja confundido com uma outra freqüência fundamental já que os sons constatados como verdadeiros são subtraídos), pode-se confiar neste valor de *peso geral* obtido, uma vez que se pode admitir que não há nenhum som em uma relação harmônica *abaixo* daquele candidato.

8.2.2.2 Intensidade

Para os harmônicos h_j do tom na faixa de freqüência de 20 a 2000 Hz, a referência [9] propõe um valor de referência bruto para a intensidade de um modelo tonal calculado como:

$$I(T) = a.\sum F(h_j, T)A(h_j, T) \tag{8.3}$$

em que *a* é uma constante, *F(.)* representa a freqüência e *A(.)* a amplitude.

A intensidade dos modelos tonais precisa ser calculada apenas uma vez, já que a intensidade do som candidato C_i pode ser calculada encontrando-se o modelo tonal T_{best} em que o candidato melhor se enquadre, escalando a intensidade de T_{best} de acordo com seu encaixe $F(T_{best}, C_i)$ para C_i. Assim, a intensidade de um candidato C_i pode ser escrita como:

$$I(C_i) = P(T_{best}, C_i).I(T_{best}) \tag{8.4}$$

8.2.2.3 Algoritmo para o cálculo dos parâmetros *Peso* e *Intensidade*

Abaixo segue o algoritmo proposto para a obtenção destes valores de referência:

Algoritmo 2

- objetivo: descobrir o tipo do tom (instrumento) e intensidade de um candidato a freqüência fundamental C.

- parâmetros de entrada:

  transformada B*Q*T do sinal analisado;

  arquivo de registro de modelos de tom

- saída:

  tipo de instrumento;

  intensidade L;

- **Passo 1:** Para cada instrumento, achar o modelo de tom T_i, cuja freqüência fundamental esteja mais próxima à de *C*;
- **Passo 2:** Comparar os modelos a *C*, calculando o quanto *C* se encaixa ao modelo, pelo valor $P(T_i, C)$;

- **Passo 3:** Calcular a intensidade de *C* para diferentes T_i, substituindo seus valores de encaixe $P(T_i, C)$ e de intensidade única $I(T)$ na equação (8.4).

- **Passo 4**: Escolher o instrumento que oferece maior intensidade a *C*, caracterizando-o como o mais indicado e o que oferece melhor enquadramento. A intensidade *I* correspondente é a intensidade do candidato;

8.2.2.4 Discussão dos resultados

O algoritmo foi implementado conforme a descrição, com os valores calculados como proposto. No entanto, não se obteve êxito na identificação dos instrumentos musicais. A Figura 43 apresenta os timbres a serem discutidos. Tentou-se identificar o mesmo tom tocado por três instrumentos: piano (azul), que é o nosso principal objeto de estudo; violino (vermelho), por possuir um timbre mais diferenciado em relação ao piano; e violão (verde), cujo timbre se assemelha ao do piano, mas mais breve e percussivo, representados na Figura 43, no domínio do tempo, e na Figura 44, a B*Q*T relativa a cada um dos três instrumentos, para que se observe a diferente distribuição de energia entre os harmônicos.

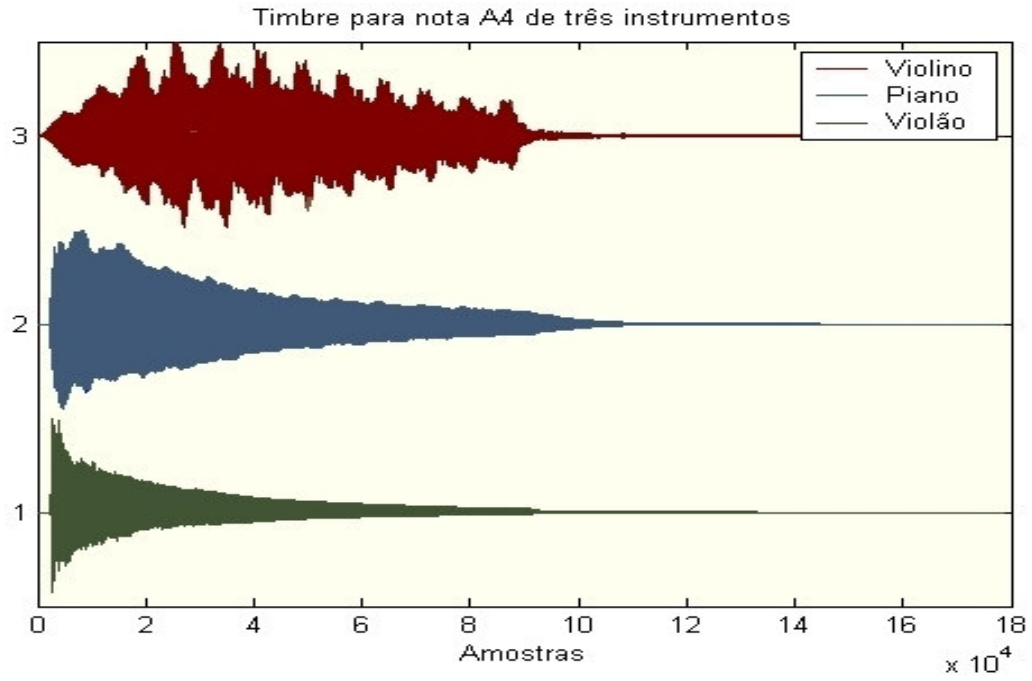


Figura 43 - Timbre para nota A4 para três instrumentos

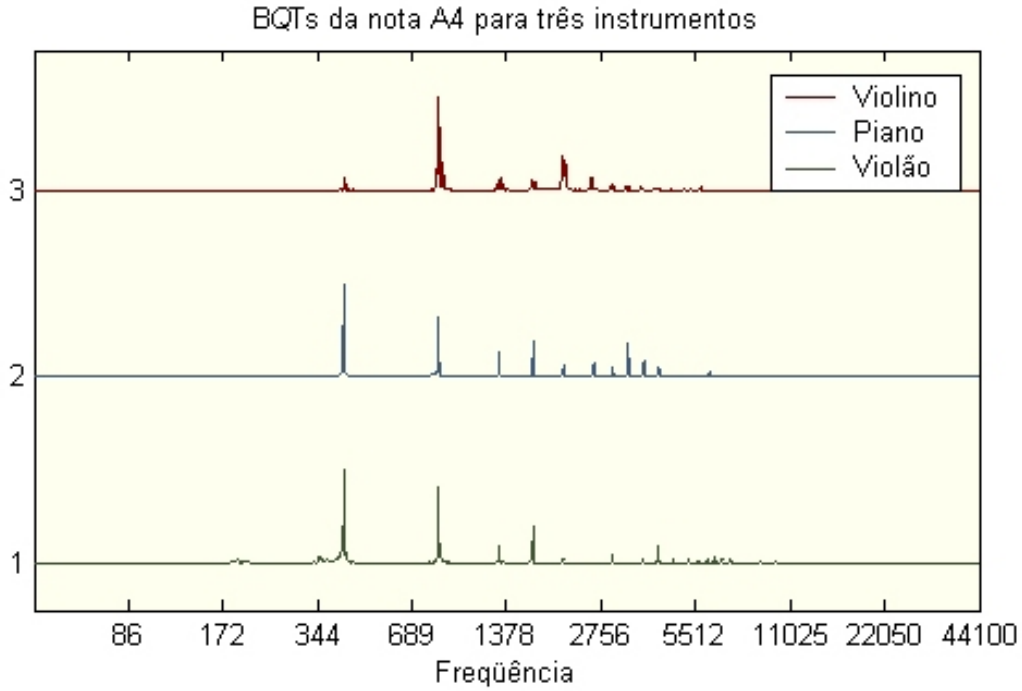


Figura 44 - B*Q*T relativa à nota A4 para três instrumentos. Observe que o violino é harmonicamente distorcido, com a freqüência fundamental com pouca energia, mas o segundo e o quinto harmônicos intensos; o piano concentra sua energia na freqüência fundamental, apresenta os harmônicos intermediários com intensidades moderadas; e o violão concentra sua energia na freqüência fundamental e no segundo harmônico, com vários harmônicos a seguir ausentes, e voltando a exibir energia em freqüências mais altas.

A razão de encaixe de um candidato em relação a um modelo de tom pré-existente deveria funcionar como um seletor em potencial. Assim, multiplicado pela intensidade do modelo, teria mais intensidade o modelo de tom que mais se adequasse ao candidato. Infelizmente, esta implementação não se mostrou eficiente. Sons, como o do violino, apresentavam no modelo de tom tamanha intensidade no quinto harmônico, que acarretava um valor obtido para intensidade independente do peso, com variação de até uma ordem de grandeza.

Sem reconhecer o instrumento correto, a intensidade não é um parâmetro bom o suficiente para se usar no algoritmo de subtração. Desta forma, a partir de então, limitou-se *a um instrumento por vez*, ou seja, só poderá haver no banco de modelos de tom um instrumento para efetuar a comparação. Isso foi necessário para que as demais particularidades deste projeto pudessem ser avaliadas.

8.2.3 Procedimento de subtração

Em música, é razoavelmente comum que duas notas estejam em uma razão de:

$$f_{0_{menor}} \propto \frac{m}{n} f_{0_{maior}} \text{, para } n \text{ inteiro.} \tag{8.5}$$

O método para a observação do som de freqüência menor já foi apresentado, garantido através da procura por picos em ordem crescente. Mas, para a detecção e a observação do tom de freqüência mais alta, é necessário um método que elimine a presença do tom menor.

As características utilizadas neste processo são *peso*, $P(T_i, C)$, e *intensidade*, $I(C_i)$, ambos baseados nas *amplitudes* dos harmônicos. Desta forma, a melhor maneira de compensar o efeito do som de freqüência mais baixa, é subtrair as amplitudes de seus harmônicos pelo espectro.

Inicialmente, a idéia de se subtrair todos os harmônicos h_j, enquanto estes forem menores que o maior harmônico audível de *C*, pode gerar problemas em sons de freqüência mais baixas. Isso reflete uma consideração feita na ocasião da modelagem de tom. Se o número de harmônicos *J* é a quantidade admitida para representar o sinal (e, em seu modelo gerado, nós tivermos registrado apenas *J* harmônicos), pode ocorrer que, para um valor de *j* maior que *J*, mas ainda menor que o maior som audível em *C*, não tenha $A(h_j, T_i)$, ou seja, não possua uma amplitude correspondente àquele harmônico no modelo.

Poder-se-ia, então, armazenar as B*Q*Ts relativas a cada uma das notas utilizadas no processo de modelagem de tom. Desta forma, foi armazenada a B*Q*Ts de cada um dos sinais de teste (os arquivos .dat com o nome da nota, citados na Seção 7.2). Calculou-se a intensidade relativa *I* do som candidato em relação ao modelo e ele assim o foi subtraído da transformada do sinal a ser analisado, zerando os valores que eventualmente se tornassem negativos nesta operação. Obviamente, esta solução não é a melhor, mas se mostrou razoavelmente eficiente ao se efetuarem

os testes de validação deste projeto. Seria interessante, portanto, que o número *J* de parciais harmônicos representativos de um tom não fosse fixo, mas variasse de acordo com a faixa de freqüências analisada, sendo os tons mais graves representados por mais parciais harmônicos e, analogamente, os tons mais agudos possuindo menos amostras.

8.3 Implementação

A implementação do processo de reconhecimento de notas musicais em sons polifônicos pode ser apresentada no Algoritmo 3, a seguir:

Algoritmo 3	Tom.m

- parâmetros de entrada:

  transformada B*Q*T do sinal analisado;

  arquivo de registro de modelos de tom

- saída:

  freqüências fundamentais F_0, presentes naquele sinal;

  intensidade associada a cada uma das freqüências fundamentais;

- **Passo 1:** Calcular os valores do vetor de probabilidade de seleção $P_S(j)$, A seguir, definir $w(j)$, como a função peso a ser utilizada pelo filtro WOS.
- **Passo 2:** Calcular *m1* como o valor médio da B*Q*T *S* do sinal a ser analisado. A motivação de armazenar o valor médio *m1* é usá-lo como referência para interromper um evento cíclico que se dará a partir do próximo passo. Enquanto a média do sinal *S* for maior que 10% do valor de *m1*, os passos a seguir serão realizados:
  - Procurar todos os picos no vetor *S*, armazenando-os no vetor *PEAK*. Outro vetor, *PPOS*, foi criado contendo os índices das posições indicadas como picos;

- A seguir, eliminar os picos que estejam a um intervalo de freqüência menor que $2^{1/12} Hz$ de picos mais altos. Também eliminar aqueles que estiverem abaixo de um determinado limite *TH*, em relação à média do sinal *S*.

- Tratar do primeiro pico a ser encontrado, desde que sua freqüência seja maior que 97 Hz, condição relacionada ao alcance mínimo de tons neste projeto. Também é necessário avaliar se este pico está a uma distância de, pelo menos, 99% do intervalo $2^{1/12}$ Hz da última nota reconhecida. E, finalmente, testa-se se o valor não é maior que 1055 Hz, visto que, o C6, aproximadamente 1046,5 Hz é o último valor em relação ao qual é possível satisfazer a condição de montar um vetor com 20 harmônicos. Enquanto não se achar um pico que satisfaça tais condições, procura-se no valor seguinte do vetor *PEAKS*. Caso, ao final da varredura de picos, estas condições não sejam atendidas, o programa é interrompido.

- Para um valor F0P qualquer que tenha satisfeito as condições expressas no sub-item anterior, monta-se o vetor de harmônicos, nos quais apenas são armazenados aqueles que representarem um pico no espectro. Multiplica-se pelo $P_S(j)$, de forma a encontrarmos outro parâmetro, *PROB*, de avaliação da presença ou não do sinal. Caso este valor obtido seja menor do que um valor limite, estabelecido por meio de tentativas, este candidato F0P é desconsiderado e busca-se o pico seguinte no vetor de picos.

- Uma vez definida a primeira freqüência candidata, F0P, procuram-se as freqüências próximas, comparando-as com os valores no banco de modelos de tom. Esta procura deverá retornar as freqüências que estejam entre $1{,}01.\frac{F_{0_p}}{2^{1/12}}$ e $0{,}99.F_{0_p}\,2^{1/12}$.

- Caso não seja encontrada nenhuma freqüência neste intervalo, deverá aparecer a mensagem de que não haveria mais nenhuma outra freqüência, interrompendo o ciclo do programa. Este seria o caso de

termos uma nota sem correspondente no banco ou quando o banco de modelos de tom estivesse vazio. Caso MF0 seja diferente de zero, o programa segue.

- Se for encontrada apenas uma freqüência fundamental MF0 próxima ao valor de F0, o programa avalia a intensidade do som, de acordo com o algoritmo proposto. De acordo com o valor encontrado, é feita a subtração do sinal e é exibida uma mensagem formada pela freqüência, pelo nome da nota (de acordo com o padrão apresentado no final do Capítulo 7) e pelo seu valor correspondente de intensidade.
- Se forem encontradas mais do que uma freqüência fundamental MF0 próxima ao valor de F0, o programa deve então comparar as intensidades obtidas para cada uma das candidatas. Aquela que apresentar a maior intensidade será considerada a verdadeira. De acordo com o valor encontrado, é feita a subtração do sinal é exibida uma mensagem formada pela freqüência, pelo nome da nota e pelo seu valor correspondente de intensidade.
- Atribuir a variável de controle, NANT, o valor da última freqüência analisada. Ela será utilizada para garantir que a próxima freqüência a ser considerada seja no mínimo $0{,}99.NANT.2^{1/12}$

- Este processo é repetido enquanto a média da BQT do sinal S, que é alterada cada vez que efetuamos a subtração do tom encontrado, foi maior que 0,1 do valor m_1, a média inicial da BQT do sinal S.

A Figura 45 apresenta o aspecto da saída deste processo, que ainda não é totalmente automatizado. O resultado apresenta, em ordem crescente de freqüência, conforme a varredura, todas as notas encontradas que satisfizeram as condições do algoritmo, ou seja, maior que 97 Hz e menor que 2099 Hz, cujos picos estejam acima de um limite *TH*. No Capítulo 9, referente às simulações do sistema como um todo, discutir-se-á sobre a necessidade de um pós-processamento a fim de validar as notas, ressaltando as verdadeiras e desprezando as falsas.

```
Digite o path e o nome do arquivo: ../casol/lcasoA.wav
Existem as possiveis frequencias:
 1.299988e+002
A nota P3C_ tem intensidade 1.229524e+001.
--------------------------------------------------
Existem as possiveis frequencias:
 1.553936e+002
A nota P3D# tem intensidade 5.902319e+000.
--------------------------------------------------
Existem as possiveis frequencias:
 1.845462e+002
A nota P3F# tem intensidade 1.336010e+001.
--------------------------------------------------
Existem as possiveis frequencias:
 2.194536e+002
A nota P3A_ tem intensidade 9.660241e+000.
--------------------------------------------------
```

Figura 45 - Resultado obtido na saída do processo. As notas são identificadas de acordo com o nome gravado no arquivo de registro de modelos de tom (apresentado na Seção 7.2). A procura é feita em ordem crescente de freqüência. O resultado do sistema não é, ainda, automatizado. São apresentadas todas as notas encontradas, sem efetuar um pós-processamento de validação.

8.4 Conclusões

O módulo de reconhecimento (ou de resolução de tom) se encontra inserido no processo de reconhecimento do sistema, sendo, basicamente o coração do mesmo. Seu principal objetivo é, a partir da transformada B*Q*T de um sinal polifônico, extrair suas informações relativas à freqüência fundamental, às freqüências dos harmônicos e à intensidade dos mesmos, de todas as notas que compuserem a mistura a ser analisada.

Apresentou-se a seguir a visão geral do funcionamento do módulo de reconhecimento. Dentre os valores constantes na transformada B*Q*T, procuram-se os seus picos, ou seja, os pontos nos quais o valor do meio é maior que o anterior e o posterior, simultaneamente. São analisados os picos em ordem crescente de freqüência.

O primeiro pico, a partir de um determinado valor-limite é então considerado um candidato à freqüência fundamental. Calcula-se sua intensidade geral, e o resultado para esta freqüência é exibido na tela. O sinal identificado é subtraído, de forma que ele não influencia no resultado quando as freqüências mais altas forem

analisadas. Feito isso, estuda-se o próximo pico, e assim sucessivamente até que se possa admitir que não há mais tons remanescentes a serem identificados no sinal.

Embora tenha sido dito que o Algoritmo 2, proposto na Seção 8.2.2.3, não fosse eficiente na identificação do instrumento musical, ainda são utilizados os valores de intensidade I e de peso P como parâmetros para a seleção da freqüência fundamental dos candidatos C_i.

Cada escolha feita durante o projeto trouxe limitações, críticas principalmente nesta fase, exigindo que alguns cuidados fossem tomados na implementação dos algoritmos propostos. O número de harmônicos J determinou a freqüência máxima suportada, e isso tem de ser refletido na execução. Houve limitações na freqüência mínima suportada, principalmente devido à imprecisão dos picos das fundamentais de tons mais graves.

No Algoritmo 3, foi necessário ter cautela com inúmeras situações que provocariam erros lógicos ou *loops* infinitos. Portanto, o que parece ser um exagero de precauções se mostrou necessário no decorrer dos testes. Não foi gerado nenhum algoritmo automático para validar os valores encontrados. A discussão a respeito de se atribuir um valor de referência que tornasse possível avaliar se um som é falso ou verdadeiro, tornando o algoritmo automático, é apresentada junto com as simulações e os testes de validação detalhados no Capítulo 9, uma vez que esse processo irá avaliar não apenas o módulo de reconhecimento, mas o sistema como um todo.

Capítulo 9 - Simulações

Todas as simulações foram feitas com apenas um instrumento por vez. A maioria dos testes apresentados aqui foi realizada com piano, por sua característica percussiva, tornando mais fácil o rastreamento de ritmo.

Utilizou-se o *software* Finale© para sintetizar os sons. Inicialmente, todas as notas foram gravadas uma a uma, para efeito de modelagem de tom, conforme descrito e apresentado no Capítulo 7. Para a segunda etapa, foram gravados acordes e trechos com rica polifonia, para revisão do desempenho do algoritmo em relação a casos particulares bem definidos. Na última etapa, trechos de algumas músicas foram analisados.

O reconhecimento de notas em todos os casos de simulação foi feito sem prévio conhecimento a respeito da polifonia envolvida. O alcance das notas foi

identificado para o sistema em seu código-fonte, que valida ou não a possibilidade de um tom ser um candidato em potencial. Na prática, tal alcance foi restrito a quatro oitavas, desde o terceiro C (130 Hz) até o sétimo C (1047 Hz), pelas razões apresentadas nos últimos três capítulos. Este intervalo compreende 37 teclas musicais, como ilustrado na Figura 46.

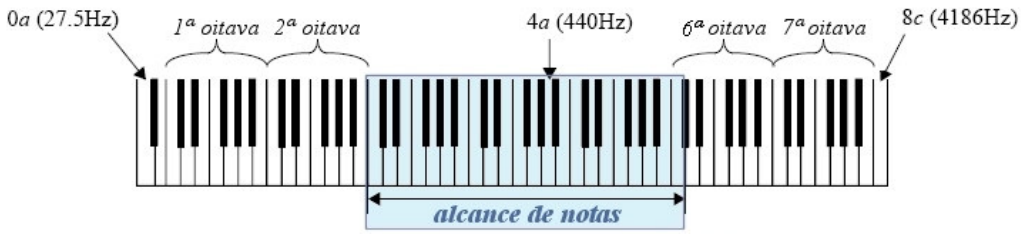


Figura 46 - Configuração das teclas de um piano, evidenciando o alcance de notas deste projeto, um total de 37 semitons, abrangendo as freqüências entre 130 Hz (C3) e 1046,5 Hz (C6).

Na Seção 9.1 revisa-se a questão do rastreamento de ritmo, explicando os momentos em que foram e que não foram implementados os algoritmos de segmentação temporal descritos no Capítulo 4. A partir disso, a Seção 9.2 apresenta as figuras de mérito inicialmente propostas para a avaliação deste processo, tentando se buscar um módulo decisório, a ser aplicado na saída do módulo de reconhecimento, a fim de validar o resultado, verificando se os tons apresentados na saída do sistema estavam presentes ou não no sinal original. A Seção 9.3 descreve brevemente a avaliação feita em cima de sons monofônicos. Na Seção 9.4, são analisados os resultados obtidos para um estudo específico de cinco tipos de casos diferentes envolvendo polifonias e acordes. Finalmente, implementam-se todos os blocos na Seção 9.5, com a avaliação em fragmentos de peças musicais. A Seção 9.6 apresenta as conclusões referentes às simulações do sistema.

9.1 Rastreamento de ritmo

Para a geração de modelo de tom e para a análise dos acordes individuais, a segmentação temporal foi feita manualmente a fim de evitar que erros gerados devido a este procedimento fossem potencializados quando da análise do sinal. Já para a análise dos fragmentos de música, na Seção 9.5, o módulo apresentado no Capítulo 4 foi utilizado, visto que é importante ressaltar a influência da segmentação

na identificação de notas, que apresentou notas falsas decorrentes da existência alguns agrupamentos.

É importante ressaltar que um dos motivos da escolha do piano como principal instrumento de simulação é devido a sua característica percussiva, o que permitiu a fácil identificação do início de cada nota. Não fosse um instrumento com estas configurações, o algoritmo de reconhecimento como um todo não teria funcionado como esperado.

9.2 Figuras de mérito

Certas misturas de notas foram tocadas separadamente e gravadas para análise da capacidade do algoritmo de resolver misturas polifônicas particularmente difíceis, em algum ponto de vista, e que, geralmente, representam falhas pontuais em sistemas de transcrição existentes.

Em todos os casos foram analisados os mesmos parâmetros, de forma que pudesse se obter uma forma fiel de avaliação da eficiência do reconhecedor de notas. Avaliou-se a saída do sistema, sem refinamento, de forma a ser possível a análise das relações entre tons identificados corretamente, tons identificados incorretamente e os tons não identificados no sinal, definidas a seguir:

- **T**ons identificados **c**orretamente (TC): aqueles que foram identificados e que estavam presentes no sinal analisado;
- **T**ons identificados **i**ncorretamente (TI): aqueles que foram identificados no sinal, mas que inexistem na gravação;
- **T**ons **n**ão identificados (TN): aqueles que existem na gravação, mas não foram identificados.

Um refinamento posterior deveria permitir que TIs sejam eliminados. A alternativa estudada baseou-se em seus valores de intensidade. Para TCs, ela deve ser mais alta que a obtida para TIs. Desta forma, apresenta-se a intensidade relativa, IR, de cada nota. Para isso, atribuiremos o valor de 100% para a nota de maior intensidade, e, assim, projetaremos o valor das demais intensidades para obtermos um valor relativo. É desejável que sons verdadeiros se manifestem com intensidades

relativas superiores às registradas por sons falsos, uma vez que, a partir disso, poderia ser obtido um parâmetro de intensidade relativa mínima, IR_{min}, para o qual sons abaixo deste percentual de intensidade seriam desconsiderados. No entanto, esta relação não foi satisfeita, conforme será visto adiante.

As figuras de mérito na avaliação do funcionamento do reconhecedor são o número absoluto de TIs, TCs e TNs e as suas intensidades relativas, IR, divididas em dois grupos: mínimo para TCs (Min TC) e máximo para TIs (Max TI). O objetivo destes valores é procurar um valor limite para ser utilizado em um algoritmo de pós-processamento, a fim de validar as notas obtidas, ressaltando as associadas a TCs e desprezando as associadas a TIs.

9.3 Avaliação dos sons monofônicos

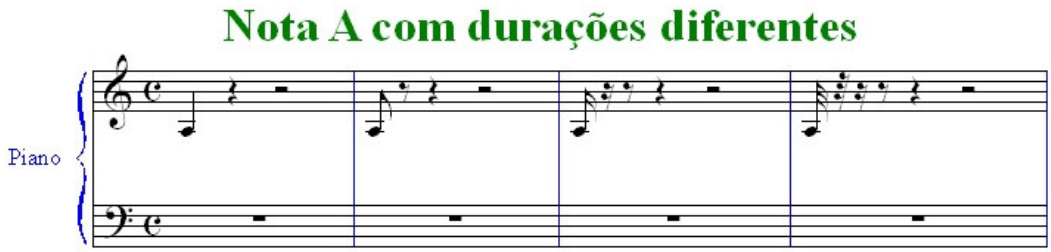
É interessante, antes de seguir para os exemplos polifônicos, analisar se o sistema é capaz de reconhecer suas próprias notas de treinamento, armazenadas no banco.

As simulações foram realizadas de duas formas. Na primeira, utilizaram-se as mesmas amostras que serviram de base para o banco de modelos de tom. Na segunda, foram geradas novas amostras, com durações diferentes.

```
Digite o path e o nome do arquivo: ../p/P4CB.wav
Existem as possiveis frequencias:
 2.609974e+002
A nota P4C_ tem intensidade 4.019092e+000.
--------------------------------------------------
```

Figura 47 - Interface do resultado apresentado pela rotina de reconhecimento de som. Neste caso, submeteu-se o arquivo "P4CB.wav" que corresponde ao som musical da nota C4 de piano, utilizado para a geração do modelo de tom. O programa efetua a procura por tom em ordem crescente de freqüência fundamental, comparando a freqüência do candidato com as freqüências de notas conhecidas armazenadas no arquivo de registro. O programa então exibe a intensidade encontrada para a nota mais próxima.

A primeira análise, como esperado, apresentou total compatibilidade, sendo encontrados os mesmos valores de intensidade geral. A Figura 47 ilustra o resultado do reconhecimento da nota C4 do piano, sendo o mesmo arquivo utilizado para o treinamento do sistema, na criação de modelo de tom.




```
>> tom
Digite o path e o nome do arquivo: ../pianolB.wav
Existem as possiveis frequencias:
 4.399189e+002
A nota P4A_ tem intensidade 1.102024e+001.
--------------------------------------------------
>> tom
Digite o path e o nome do arquivo: ../pianol_2B.wav
Existem as possiveis frequencias:
 4.399189e+002
A nota P4A_ tem intensidade 8.493731e+000.
--------------------------------------------------
>> tom
Digite o path e o nome do arquivo: ../pianol_4B.wav
Existem as possiveis frequencias:
 4.399189e+002
A nota P4A_ tem intensidade 6.569001e+000.
--------------------------------------------------
>> tom
Digite o path e o nome do arquivo: ../pianol_8B.wav
Existem as possiveis frequencias:
 4.399189e+002
A nota P4A_ tem intensidade 5.280085e+000.
--------------------------------------------------
```

Figura 48 - Nota A4 tocada com durações diferentes, conforme a representação em partitura, diminuindo seu tempo de execução da esquerda para a direita. A seguir, apresenta-se o resultado do programa para estas quatro execuções da nota. Observe que, quando menor o tempo, menor é a intensidade calculada.

Na segunda análise, ilustrada pela Figura 48, todos os tons foram corretamente identificados, sendo que os executados mais brevemente apresentaram pior encaixe e valores menores de intensidade geral, o que ocorre devido à duração de referência utilizada para a geração dos modelos de tom foi de uma semínima, e ao fato de quanto mais breve a duração, maior a intensidade relativa dos seus parciais harmônicos de freqüências mais altas, como pode ser visto na Figura 49. Isso é importante porque ilustra a questão apresentada no Capítulo 6, que menciona o fato de que a velocidade e o modo de execução das notas pode interferir nos resultados finais.

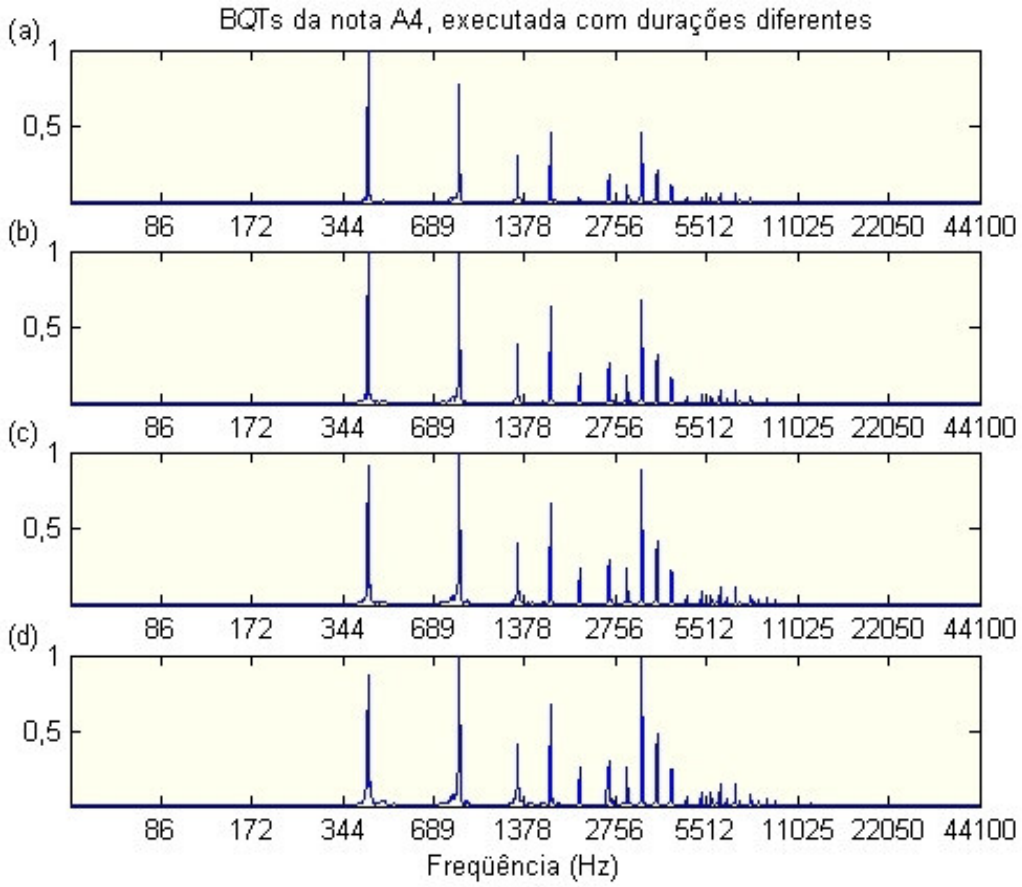


Figura 49 – B*Q*T da nota A4 executada com durações diferentes, conforme a representação em partitura (Figura 48), diminuindo seu tempo de execução de (a) semínima, (b) colcheia, (c) semicolcheia e (d) fusa. Observe que, quando menor o tempo, maior é influência dos parciais harmônicos de freqüências mais altas.

9.4 Avaliação de misturas polifônicas

A seguir, são apresentados cinco casos especiais que foram considerados como bons parâmetros de teste na avaliação do sistema proposto. Cada caso de mistura polifônica oferece um nível de dificuldade diferente. Para ilustrar como são tratados os resultados a partir da saída do sistema e como eles serão representados nas próximas subseções, três casos serão detalhados na sub-seção a seguir.

9.4.1 Exemplos de tratamento de dados

O primeiro exemplo é o caso 1.1, da Sub-seção 9.4.2, em que o acorde $C^0$3 é executado na terceira oitava (130-261 Hz). Ele é composto pelas notas: C3, D3#, F3# e A3. A Figura 50 ilustra o resultado apresentado pelo sistema.

```
>> tom
Digite o path e o nome do arquivo: ../casol/lcasoA.wav
Existem as possiveis frequencias:
 1.302003e+002
A nota P3C_ tem intensidade 3.330147e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 1.557228e+002
A nota P3D# tem intensidade 1.794266e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 1.851187e+002
A nota P3F# tem intensidade 1.946895e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 2.200640e+002
A nota P3A_ tem intensidade 3.899225e+002.
-------------------------------------------------
```

Figura 50 - Resultado apresentado na saída do sistema para a resolução do caso 1.1, em que o acorde C^0 é executado na terceira oitava.

Monta-se a primeira tabela de dados ilustrada pela Tabela 7. A primeira coluna é preenchida com as notas correspondentes aos tons presentes no sinal analisado, ou seja, aqueles de que se tem conhecimento. A segunda coluna é preenchida com os tons encontrados, sendo colocados na mesma linha caso existam no modelo. Na terceira coluna, preenche-se I, a intensidade absoluta obtida na saída do sistema, para cada um dos tons encontrados. Finalmente a intensidade relativa, IR, é calculada a partir da atribuição de 100% ao tom cuja intensidade registrada for a mais elevada, sendo normalizadas as demais intensidades a partir deste valor. No caso 1.1, todos os resultados obtidos são TCs.

Tabela 7 – Primeira análise de dados a partir da saída do sistema, para o caso 1.1. Neste caso, não há notas falsas ou faltantes, consistindo o caso ideal. A coluna "T. Componentes" apresenta os tons realmente existentes no sinal analisado. "T. Encontrados" e suas respectivas intensidades "I" são resultantes da saída do sistema. A intensidade relativa "IR" é calculada atribuindo-se 100% para a nota de maior intensidade e normalizando o resultado das intensidades das demais a partir disso.

T. Componentes	T. Encontrados	I	IR
C3	C3	333,0	85,4%
D3#	D3#	179,4	46,0%
F3#	F3#	194,7	50,0%
A3	A3	389,9	100,0%

O segundo exemplo é o caso 3.3, da Sub-seção 9.4.4, em que sete notas são tocadas em posições diversas em uma oitava, sendo C5, D5, E5, G5, A5, A5# e C6 as notas utilizadas para o teste. O resultado obtido pelo sistema é exibido na Figura 51. Monta-se a tabela de dados ilustrada pela Tabela 8, de acordo com o padrão estabelecido. No caso 3.3, o tom correspondente à nota A5 não foi detectado, não havendo correspondente a ele na coluna 2.

```
>> tom
Digite o path e o nome do arquivo: ../caso3/3casoC.wav
Existem as possiveis frequencias:
 5.236218e+002
A nota P5C_ tem intensidade 5.701079e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 5.882363e+002
A nota P5D_ tem intensidade 1.793149e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 6.596448e+002
A nota P5E_ tem intensidade 1.619183e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 7.832956e+002
A nota P5G_ tem intensidade 1.503883e+001.
-------------------------------------------------
Existem as possiveis frequencias:
 9.339616e+002
A nota P5A# tem intensidade 2.491380e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 1.047101e+003
A nota P6C_ tem intensidade 7.358603e+001.
-------------------------------------------------
```

Figura 51 - Resultado apresentado pelo sistema à resolução do caso 3.3, constituído por sete notas tocadas em posições diferentes em uma mesma oitava. O som analisado é constituído pelas notas C5, D5, E5, G5, A5, A5# e C6. Como se pode observar, as notas C5, D5, E5, G5, A5# e C6 são identificadas corretamente, mas a nota A5 está faltando.

Tabela 8 – Primeira análise de dados a partir da saída do sistema, para o caso 3.3. Neste caso, há um tom não identificado, que é visível por não existir correspondência na linha referente à nota A5 na coluna 2.

T. Componentes	T. Encontrados	I	IR
C5	C5	570,1	100,0%
D5	D5	179,3	31,4%
E5	E5	161,9	28,4%
G5	G5	15,0	2,6%
A5	-	-	-
A5#	A5#	249,1	43,7%
C6	C6	73,6	12,9%

O terceiro exemplo é o caso 5.1, da Sub-seção 9.4.6, em que as notas F3, F3#, G3, G3#, A3 e A3# são utilizadas para o teste. O resultado obtido pelo sistema é exibido na Figura 52.

```
>> tom
Digite o path e o nome do arquivo: ../caso5/5casoA.wav
Existem as possiveis frequencias:
 1.851187e+002
A nota P3F# tem intensidade 3.180114e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 2.200640e+002
A nota P3A_ tem intensidade 4.488086e+002.
-------------------------------------------------
Existem as possiveis frequencias:
 3.499645e+002
A nota P4F_ tem intensidade 7.201855e+001.
-------------------------------------------------
Existem as possiveis frequencias:
 4.684593e+002
A nota P4A# tem intensidade 5.030880e+001.
-------------------------------------------------
Existem as possiveis frequencias:
 6.217095e+002
A nota P5D# tem intensidade 3.532922e+001.
-------------------------------------------------
Existem as possiveis frequencias:
 6.999083e+002
A nota P5F_ tem intensidade 2.087782e+001.
-------------------------------------------------
```

Figura 52 - Resultado apresentado pelo sistema à resolução do caso 5.1, constituído pelas notas F3, F3#, G3, G3#, A3 e A3#. Observa-se que os tons correspondentes às notas F3# e A3 foram devidamente identificados. Os tons F4, A4#, D5# e F5 foram incorretamente identificados. E os tons correspondentes às notas F3, G3, G3# e A3# não foram identificados.

Monta-se a sua respectiva tabela de dados ilustrada pela Tabela 9 No caso 5.1, as notas F3, G3, G3# e A3# não foram detectadas, não havendo correspondente a elas na coluna 2, enquanto as notas F4, A4#, D5# e F5 não estão presentes no sinal original, sendo portanto TIs.

Tabela 9 – Primeira análise de dados a partir da saída do sistema, para o caso 5.1. Neste caso, há quatro tons não identificados, que é visível por não existir correspondência nas linhas referente às notas F3, G3, G3# e A3# na coluna 2, e quatro tons incorretamente identificados, F4, A4#, D5# e F5, o que é observável pela ausência de correspondentes na coluna 1.

T. Componentes	T. Encontrados	I	IR
F3	-	-	-
F3#	F3#	318,0	70,8%
G3	-	-	-
G3#	-	-	-
A3	A3	448,8	100%
A3#	-	-	-
-	F4	72,0	16,0%
-	A4#	50,3	11,2%
-	D5#	35,3	7,9%
-	F5	20,9	4,7%

Observe que a intensidade relativa mínima detectada para um TC (Min TC) é de 70,8%, maior que a máxima registrada para um TI (Max TI), que é de 16,0%. Neste caso, poder-se-ia ajustar um parâmetro de validação de resultado com qualquer valor limite entre 16,0% e 70,8%, uma vez que abaixo destes valores só existem sons falsos e, acima, somente verdadeiros. Esta situação, no entanto, não é observada na maioria dos casos exibidos a seguir, não sendo possível ajustar um único parâmetro de validação sem haver perdas em TCs ou admissão de TIs.

A seguir, resumem-se, na Tabela 10, os dados obtidos para estes três exemplos no formato a ser exibido na discussão dos casos apresentados abaixo. De cada caso, apresentam-se os valores mínimos de intensidade relativa para notas verdadeiras, os valores máximos de intensidade relativa para as falsas, de modo a se buscar um valor intermediário para definir um parâmetro de validação do resultado apresentado na saída do sistema. A seguir, apenas a título de informação, exibe-se a quantidade de notas falsas com IR acima de 20% e de 10%. Finalmente, um resumo quantitativo é oferecido pela divulgação do número de tons identificados

corretamente (TCs), de tons identificados incorretamente (TIs) e tons não identificados (TNs).

Tabela 10 - Tabela simplificada com os valores de mérito a serem observados.

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
1 (1.1)	46,0%	-	-	-	4	-	-
2 (3.3)	2,6%	-	-	-	6	-	1
3 (5.1)	70,8%	16,0%	-	2	2	4	4

Com base nestes três casos, não é possível escolher um valor de validação sem que haja perdas de TCs ou admissão de TIs. Escolher um valor menor que 2,6%, IR mínima para um TC no caso 3.3, faria com que fossem admitidos os 4 TIs do caso 5.1. Desta forma, tem-se em mente que a abordagem de validação do resultado baseado diretamente na intensidade não é um método eficaz, uma vez que haverá perdas e erros. Neste projeto, admitiu-se que qualquer valor diferente de zero era o suficiente como resultado. Não foi dada ênfase na elaboração de uma rotina de validação de resultados.

9.4.2 Caso 1 – Acordes tocados em diferentes oitavas

Acordes maiores e acordes diminutos foram tocados em diferentes posições do teclado e submetidas à análise. Tais combinações são relativamente comuns, mas não são triviais para transcrição. A dificuldade envolvida se deve ao fato de as freqüências fundamentais se encontram em relações numéricas racionais.

Foram utilizados os seguintes acordes:

1. $C^0$3, na terceira oitava: C3, D3#, F3# e A3;
2. E, na terceira oitava: E3, G3# e B3;
3. $C^0$4, na quarta oitava: C4, D4#, F4# e A4;
4. F, na quarta oitava: C4, F4 e A4;
5. $C^0$5, na quinta oitava: C5, D5#, F5# e A5;
6. G, na quinta oitava: D5, G5 e B5;

Na Tabela 11, apresenta-se o resumo dos principais dados observados, organizados conforme explicado na Seção 9.4.1.

Tabela 11 – Síntese dos dados observados para o caso de notas musicais em relações numéricas racionais.

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
1.1	46,0%	-	-	-	4	-	-
1.2	12,5%	-	-	-	3	-	-
1.3	51,0%	-	-	-	4	-	-
1.4	70,8%	-	-	-	3	-	-
1.5	47,8%	-	-	-	4	-	-
1.6	50,2%	-	-	-	3	-	-

O resultado para o caso de notas musicais em relações numéricas racionais é bastante otimista, uma vez que, na música, a maioria das notas que são tocadas em conjunto correspondem a acordes observando tais tipos de relação. Como neste exemplo não foi detectado nenhum TI, não haveria necessidade de se definir um parâmetro de validação de resultado.

9.4.3 Caso 2 – Duas notas tocadas juntas com freqüências fundamentais múltiplas entre si

Neste caso, as freqüências fundamentais estão em uma relação de $f_{0_2} = m.f_{0_1}$, o que faz com que os harmônicos da nota mais baixa estejam sobrepostos a todos os harmônicos da freqüência mais alta. A nota 'raiz' é o Ré da terceira oitava (D3), e a segunda nota é tocada de acordo com um valor numérico para *m*. Foram utilizadas as seguintes duplas:

1. D3 e D4: $m = 2$;
2. D3 e A4: $m = 3$;
3. D3 e D5: $m = 4$;
4. D3 e F5#: $m = 5$;
5. D3 e A5: $m = 6$;
6. D3 e C6: $m = 7$;

Dificuldade: A nota de freqüência mais baixa sobrepõe todos os parciais harmônicos da nota de freqüência mais alta.

Tabela 12 - Resultados obtidos para notas em que suas freqüências fundamentais sejam múltiplas entre si.

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
2.1*	100,0%	16,8%	-	1	1	1	1
2.2	45,6%	-	-	-	2	-	-
2.3	0,7%	-	-	-	2	-	-
2.4	41,1%	-	-	-	2	-	-
2.5	27,2%	-	-	-	2	-	-
2.6	36,7%	-	-	-	2	-	-

A Tabela 12 oferece o resumo dos resultados obtidos para o caso em que as freqüências fundamentais de suas notas sejam múltiplas entre si. O exemplo 2.1 foi assinalado com um asterisco devido a características interessantes que sucederam. A nota raiz, D3, foi devidamente identificada, com um valor de intensidade considerável se comparado com os demais exemplos, sendo portanto subtraído do espectro. No entanto, como ilustrado pela Figura 53, que apresenta a B*Q*T do tom correspondente à nota D3, a freqüência fundamental possui baixa amplitude se comparada com os seus parciais harmônicos, gerando um desvio no cálculo da intensidade do sinal. Tal desvio faz com que seja subtraída a energia correspondente à fundamental do tom correspondente à nota D4, que passa a não ser identificada. Com isso, seus harmônicos não foram subtraídos e continuaram a se fazer presentes no sinal, fazendo com que TIs sejam encontrados. Também, devemos ressaltar uma particularidade observada nos sinais de baixa freqüência:

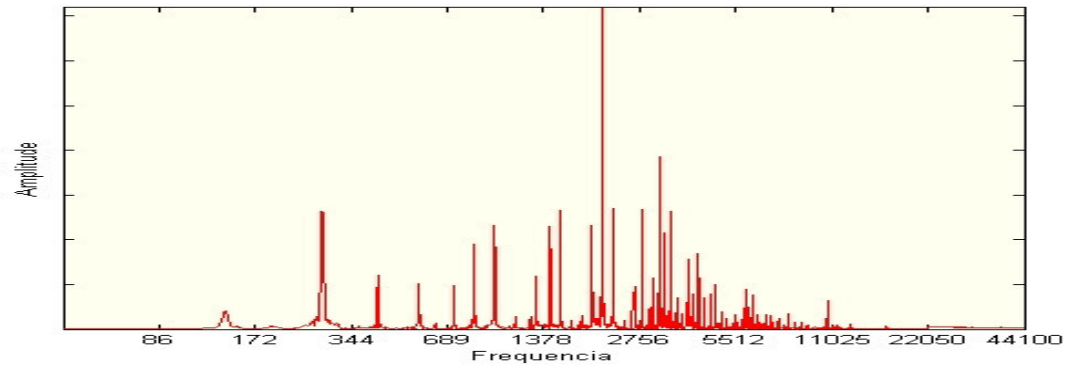


Figura 53 - B*Q*T da nota D3, note que o primeiro pico, correspondente à freqüência fundamental, encontra-se muito espalhado. As maiores intensidades são encontradas em freqüências harmônicas mais altas (na faixa de 1378 a 2756 Hz). Isso gera um desvio no cálculo da intensidade do sinal, fazendo com que o som resultante da subtração acabe eliminando o som presente na freqüência múltipla da nota D4.

A transformada B*Q*T de uma nota de piano na oitava mais baixa de alcance neste projeto, ou seja, entre 130 Hz e 261 Hz, possui um comportamento atípico, conforme pode ser observado na Figura 53. Enquanto as demais apresentam uma relação de amplitude x freqüência caindo quase exponencialmente, as amplitudes obtidas para harmônicos muito altos tiveram valores elevados, alguns inclusive muito maiores que o da freqüência fundamental. Tais valores modificam a intensidade média do sinal, fazendo com que, durante a subtração da nota identificada, a energia das freqüências referentes aos primeiros parciais harmônicos seja totalmente subtraída, fazendo com que os tons em relação 2:1 sejam apagados.

9.4.4 Caso 3 – Sete notas tocadas em diversas posições

As seguintes notas foram tocadas em cada uma das oitavas de alcance do projeto: C, D, E, G, A, A# e o C da oitava seguinte. Temos, desta forma:

1. C3, D3, E3, G3, A3, A3# e C4;
2. C4, D4, E4, G4, A4, A4# e C5;
3. C5, D5, E5, G5, A5, A5# e C6;

Dificuldade: Polifonia rica. Muitas freqüências fundamentais estão em relações numéricas racionais.

Tabela 13 - Resultado obtido para sete notas tocadas em diversas posições

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
3.1	14,3%	-	-	-	6	-	1
3.2	20,7%	-	-	-	6	-	1
3.3	10,6%	-	-	-	6	-	1

A Tabela 13 apresenta o resumo dos resultados obtidos para as sete notas tocadas em diversas posições dentro de uma mesma oitava. Na terceira oitava (3.1), não foi possível identificar o C da oitava seguinte no som polifônico, uma vez que seus harmônicos podem ser facilmente derivados das outras notas, especialmente o C mais baixo. Na quarta (3.2) e na quinta (3.3) oitavas, não se obtiveram TIs, mas as notas A4 e A5, respectivamente, não foram identificadas.

9.4.5 Caso 4 – Notas raízes mais acordes

As freqüências fundamentais das notas que constituem um acorde maior correspondem ao quarto, ao quinto e ao sexto harmônicos de uma nota raiz do acorde, localizada duas oitavas abaixo do acorde. Devido às suas propriedades harmônicas, a nota raiz é freqüentemente tocada junto com o acorde. Neste caso, os parciais harmônicos da nota raiz sobrepõem todas as notas mais altas. Na Tabela 14, apresentaremos o resultado obtido para os seguintes casos:

1. Nota raiz 3D e acorde D: D3, D5, F5# e A5;
2. Nota raiz 3D e acorde Dm: D3, D5, F5 e A5;
3. Nota raiz 3E e acorde C: E3, C5, E5 e G5;

Dificuldade: Notas dos acordes estão totalmente sobrepostas pela nota raiz

Tabela 14 - Resultado obtido para sons formados por uma nota raiz e seu acorde maior ou menor

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
4.1	15,1%	-	-	-	4	-	-
4.2	20,9%	-	-	-	4	-	-
4.3	5,5%	-	-	-	4	-	-

Não existe nenhuma nota real na posição do primeiro e do segundo harmônicos da nota raiz. No entanto, se, na subtração da nota raiz, sobrarem resquícios que possam induzir o sistema a identificar, erroneamente, um candidato fundamental nestas posições, tal TI será subtraído, e, conseqüentemente, as amplitudes dos parciais harmônicos deste TI que coincidirem com os de um tom que realmente esteja presente no som original também serão diminuídas, alterando o valor de intensidade dos sons que realmente existem.

9.4.6 Caso 5 – Diversas notas adjacentes

Este exemplo consiste em considerar diversas notas adjacentes, de forma que estejam próximas umas das outras no tom, mas não há relação harmônica entre eles, o que deveria torná-los facilmente detectados. Os conjuntos de nota foram:

1. 3ª oitava: F3, F3#, G3, G3#, A3 e A3#;
2. 3ª oitava: E3, F3, G3 e G3#;

3. 4ª oitava: G4, G4#, A4#, B4;
4. 5ª oitava: F5, F5#, G5 e G5#

Dificuldade: Pode haver mascaramento em freqüência, uma vez que se tratam de dois tons de freqüências fundamentais próximas.

Tabela 15 - Resultado obtido para sons formados por uma nota raiz e seu acorde maior ou menor

Exemplo	Intensidade Relativa		TIs acima de		Tons		
	Min TC	Max TI	20%	10%	TC	TI	TN
5.1	70,8%	16,0%	-	2	2	4	4
5.2	66,4%	-	1	-	2	-	1
5.3	81,2%	17,9%	-	1	2	2	2
5.4	42,3%	-	-	-	3	-	1

A Tabela 15 apresenta o resumo dos resultados obtidos para a análise de sons de diversas notas adjacentes, o que caracteriza o fenômeno de mascaramento em freqüência ou mascaramento simultâneo, no qual um som cuja freqüência fundamental esteja muito próxima a de outro tende a não ser percebido. Observa-se que aí reside a maior dificuldade encontrada pelo sistema.

9.5 Avaliação de trechos musicais

Dois fatores adicionais devem ser considerados nestas avaliações. Primeiro, fez-se uso do algoritmo de segmentação temporal, apresentado no Capítulo 4, o que ocasionalmente fez com que notas com intervalos de tempo muito curtos entre si fossem agrupadas em segmentos maiores. Segundo, porque algumas notas podem continuar a tocar durante diversos segmentos de sinal e seus tons alterarem levemente no decorrer do tempo.

Para tais simulações, as tabelas foram configuradas de modo a facilitar a visualização da evolução temporal das notas musicais obtidas. Nelas, as colunas representam os segmentos de notas, sendo que os índices superiores correspondem à segmentação esperada (segmentos originais), enquanto os índices inferiores denotam os segmentos obtidos. As linhas representam a totalidade do alcance de tons nestes trechos, de modo que um bloco em azul representa um TC; em vermelho, um TI; e, em amarelo, um TN, conforme a Tabela 16, que exemplifica um caso, que seriam as notas iniciais da música “First Love”, tratada na Sub-seção 9.5.2. As últimas linhas

correspondem à síntese dos dados de acordo com o segmento obtido, totalizando TCs, TIs e TNs. No final, calculam-se as intensidades relativas mínimas para TCs, e máximas, para TIs.

Ressalta-se que a faixa de operação deste projeto está restrita aos tons localizados entre C3 e C6. Nas peças musicais apresentadas a seguir, há tons mais agudos e, desta forma, não serão identificados.

Tabela 16 – Exemplo de apresentação dos resultados do sistema para os trechos de peças musicais analisadas. As colunas representam a evolução temporal da música, exibindo os segmentos nos quais elas foram divididas. O índice superior representa a numeração esperada (original). O inferior, a numeração do segmento obtido. TCs são assinaladas em azul; TIs, em vermelho; e TNs, em amarelo. No final da tabela, há um resumo dos dados, apresentando a quantidade de TCs, TIs e TNs, bem como as intensidades relativas mínimas para TCs, e máximas para TIs.

Segmentos originais	1	2	3	4
Segmentos obtidos	1	2	3	4
D6				
G5				
F5#				
D5				
B4				
A4				
G4				
F4#				
D4#				
TCs	4	2	2	4
TIs				1
TNs		2	2	
Min TC (%)	59,2	36,7	20,3	12,3
Max TI (%)				21,5

9.5.1 Trecho 1 – Brothers

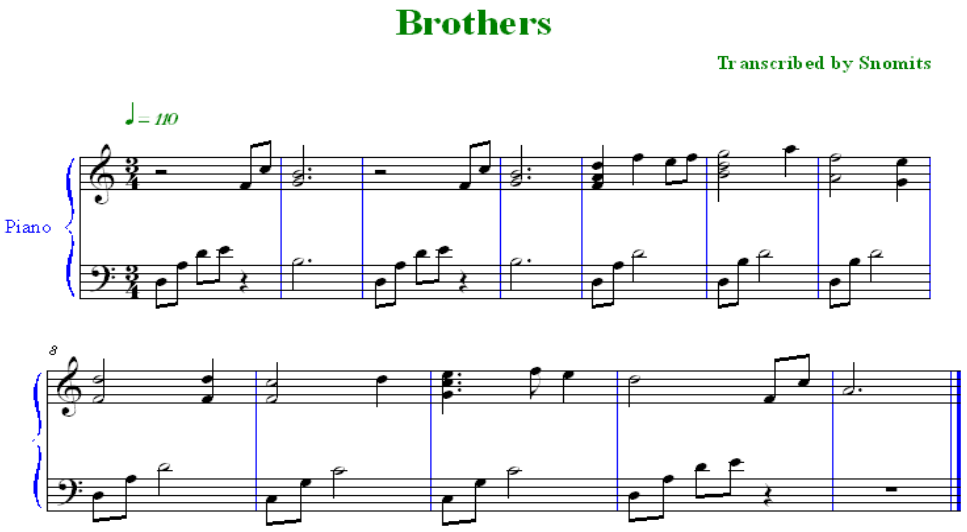


Figura 54 – Os primeiros doze compassos da composição "Brothers". A segmentação temporal das mesmas foi discutida na Seção 4.2.2.

O primeiro exemplo a ser apresentado aqui é um trecho de música simples, polifonia de até quatro tons simultâneos. Apresenta o compasso de três tempos bem marcado, notas de duração diferentes tocando simultaneamente. A partitura referente a 12 compassos desta música é apresentada na Figura 54.

Tabela 17 – Resultado obtido para o trecho inicial da música "Brothers", até a nota 27. Observe os agrupamentos nos segmentos 15 e 16, 20 e 21, 24 e 25, o que justifica a existência de alguns dos TIs na saída do sistema. Ressalta-se que a grande quantidade de TNs na parte superior da tabela resulta de tais tons estarem fora da faixa de operação do reconhecedor proposto neste projeto.

Segmentos originais	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Segmentos obtidos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		16	17	18	19		20	21	22		23	24
A6																											
G6																											
F6																											
E6																											
D6																											
C6																											
B5																											
A5																											
G5																											
F5																											
E5																											
D5																											
B4																											
A4																											
D4																											
TCs	1	1	1	1	1	1	3	1	1	1	1	1	1	3	3	3	1			2	2	1	1	2	2	2	1
TIs			1			1				1			1		1	1	2			1	1	1		1	1		
TNs															1	1	1	2	2	2	2	3	1	1	1	2	2
Min TC (%)							46,9							46,9	13,8	47,3				26,2	18,1			11,0	12,3	23,5	
Max TI (%)			49,2			55,3				47,2			51,2		51,5	13,8	17,5			27,7	26,2	27,5		13,1	11,0		

Tabela 18 – Complemento da Tabela 17. Resultado obtido para "Brothers", entre as notas 28 e 47. Observe novamente a existência de agrupamentos nos segmentos 28 e 29, 32 e 33, e entre os segmentos 36,37 e 38, o que justifica a existência de alguns dos TIs na saída do sistema.

Segmentos originais	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
Segmentos obtidos	25		26	27	28		29	30	31			32	33	34	35	36	37	38	39	40
F6																				
E6																				
D6																				
C6																				
A5																				
G5																				
F5																				
E5																				
D5																				
C5																				
A4																				
G4																				
D4																				
C4																				
TCs	2	2	2	2	3	3	2		3	3	2	1		1	1	1	1	1	1	1
TIs	1	1	1		1	1			1	1	2					1			1	1
TNs	1	1	1	1			1	2	1	1	2	1	2	1	1	1	1			
Min TC (%)	22,6	18,1	34,7	53,3	19,8	23,1	35,8		3,4	3,4	3,4	8,2		98,6						
Max TI (%)	18,8	22,6	38,7		23,1	19,8			21,3	25,8	25,8		41,9			62,6			84,2	14,7

As Tabelas 17 e 18 apresentam o resultado conforme estabelecido no início desta seção. É interessante ressaltar algumas características encontradas nesta música.

A segmentação temporal, conforme discutido na Seção 4.2.2, apresentou algumas falhas ao agrupar alguns conjuntos de notas. O segmento obtido 15, por exemplo, engloba os segmentos esperados 15 e 16. Este tipo de agrupamento acarreta na identificação de notas falsas dentro dos subconjuntos. Nesta situação em questão, a nota A4 não existe no primeiro segmento original 15, tornando-se um TI para esta parcela, mas existe no segmento original 16, tornando-se um TC. Além disso, pode-se observar os segmentos 3 e 6, por exemplo, que apresentam TIs como continuações de um tom existente anteriormente. Isso pode significar que a segmentação também falhou naquele instante, uma vez que houve resquícios suficientes no segmento seguinte para ser possível a identificação deste som.

Excetuando os casos em que os tons correspondem a notas cujas fora da faixa de operação deste projeto, os TNs, em sua maioria, correspondem a notas de duração de 2 a 3 tempos, dentro de um compasso, o que significa que haveria parciais de intensidades decrescentes no decorrer do tempo, nos segmentos seguintes, uma vez que são tocadas em conjunto com colcheias, de duração ½ semínima. O uso do piano facilita a segmentação temporal devido à sua característica percussiva, com a energia do tom concentrando-se em um ataque rápido e decaimento exponencial. Esta mesma característica faz com que uma nota longa tenha queda de intensidade rápida, de forma que haja pouca representatividade da nota nos segmentos a seguir. Isso justifica em parte os TNs, uma vez que muitas delas decorrem desta característica. Este problema fica mais claro na música a seguir, “First Love”, que apresenta o toque de acordes de piano de quatro tempos, enquanto se tocam outras notas em tempos curtos, como ilustrado pela Figura 55.

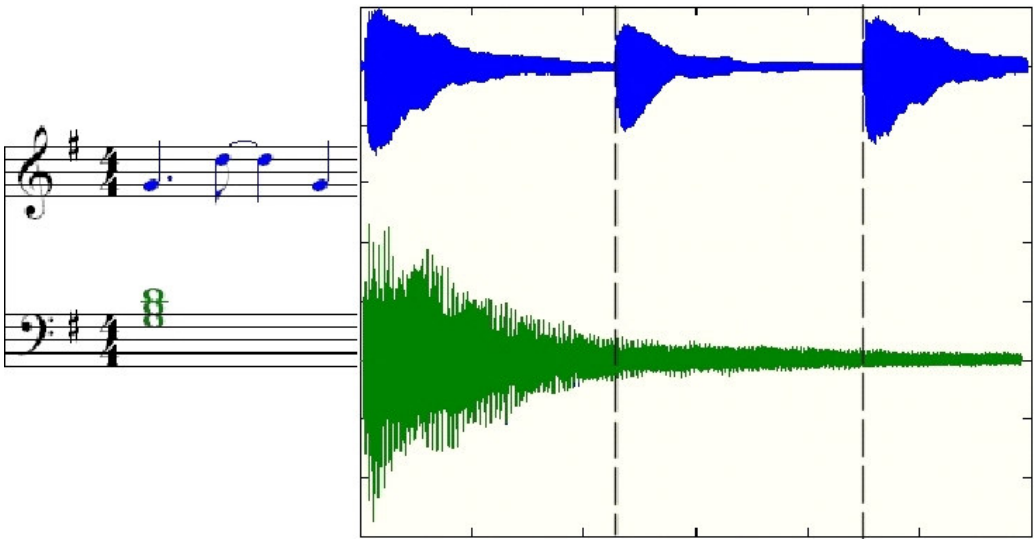


Figura 55 - O uso do piano facilita a segmentação temporal devido à sua característica percussiva. Esta mesma característica faz com que uma nota longa tenha queda de intensidade rápida, de forma que haja pouca representatividade da nota nos segmentos a seguir. Isso justifica em parte os TNs, uma vez que muitas delas decorrem desta característica.

9.5.2 Trecho 2 – "First Love"

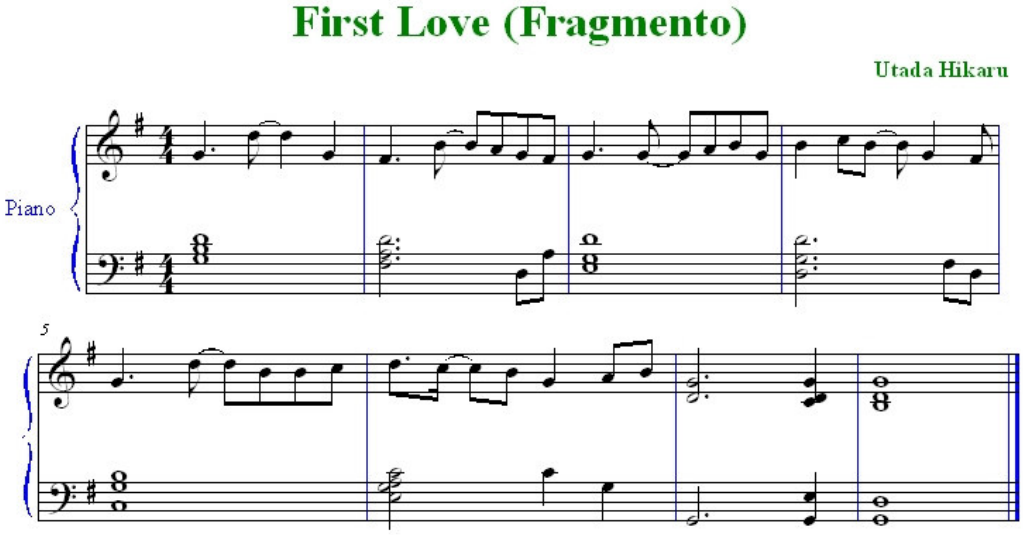


Figura 56 – Oito compassos da composição "First Love". A segmentação temporal das mesmas foi discutida na Seção 4.2.3.

O segundo exemplo a ser apresentado aqui é um trecho da música "First Love". A dificuldade desta peça são as notas de durações diferentes tocadas simultaneamente, alternando trechos rápidos, com muitas notas tocadas em pouco tempo, e lentos. Sua polifonia máxima é de cinco tons simultâneos. A partitura referente aos 8 primeiros compassos desta música se encontram na Figura 56.

Tabela 19 – Resultado obtido para "First Love", até o segmento original 19. Conforme dito na Seção 4.2.3, em vez de agrupamento de notas, houve o caso em que um mesmo segmento foi dividido em dois (segmento original 16 gerou os segmentos 16 e 17 obtidos).

Segmentos originais	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		17	18	19
Segmentos obtidos	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D6																				
C6																				
B5																				
A5																				
G5																				
F5#																				
E5																				
D5																				
B4																				
A4																				
G4																				
F4#																				
E4																				
D4#																				
D4																				
TCs	4	2	2	4	3	2	2	2	4	2	1	1	2	4	2	2	1	2	1	2
TIs				1					1	1		1		1						1
TNs		2	2		1	2				2	3	3	2		2	2	3	2	1	
Min TC (%)	59,0	42,7	20,9	9,6	21,1	65,9	37,9	48,6	34,9	60,5			10,6	28,6	30,3	20,3		28,6		44,8
Max TI (%)				24,5					12,9	61,3		21,5		2,3						100

Tabela 20 - Complementação da Tabela 19, com o resultado obtido para "First Love", do segmento original 20 até o 33

Segmentos originais	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Segmentos obtidos	21	26	27	28	29	30	31	32	33	34	35	36	37	38
D6														
C6														
B5														
A5														
G5														
E5														
D5														
C5														
B4														
A4														
G4														
E4														
D4														
C4														
G3														
TCs	4	2	1	1	1	4	4	1	2	2	2	2	4	5
TIs					1		1	1					1	
TNs		2	3	3	3	1		3				1	1	
Min TC (%)	34,8	22,2				78,3	63,8		71,3	51,1	38,8	32,8	17,0	20,8
Max TI (%)					55,6		31.8	19,0					100	

As Tabelas 19 e 20 apresentam o resultado obtido para a música “First Love”. Conforme dito na Seção 4.2.3, a segmentação temporal desta música apresentou cinco segmentos extras, dos quais 4 foram percebidos como silêncio e portanto não foram incluídos na tabela (entre os segmentos esperados 20 e 21). Um quinto segmento extra surgiu a partir da divisão do segmento original 16 em dois, nos quais a principal diferença e a não observância da nota D5. Esta nota era integrante de um acorde de quatro tempos, cuja intensidade decaiu com o decorrer dos segmentos, até se tornar tão baixa a ponto de não ser reconhecida no sinal. Isso corrobora a discussão apresentada em função da música “Brothers”, e ressaltado pela Figura 55, de que o timbre do piano é determinante para que tais sons não sejam observados, não se caracterizando uma falha de severa gravidade.

9.6 Conclusões

O sistema se mostrou razoavelmente robusto para casos monofônicos e no gerenciamento de acordes. Ficaram visíveis as principais falhas do mesmo, como a dificuldade em lidar com sons em baixa freqüência, revelando a necessidade de algum tratamento especial para enfatizar suas freqüências fundamentais, amenizando o efeito de serem desprezados pelos seus parciais harmônicos, que apresentam intensidades mais altas.

Também é necessário prestar atenção no caso de sons formados por diversas notas adjacentes, ou seja, cujas freqüências fundamentais estão próximas entre si. O fato de não estarem em relação harmônica deveria torná-los facilmente detectados, mas uma nota intensa facilmente mascara uma nota próxima mais fraca. Além disso, acredita-se que parte das falhas apresentadas tenha sido ocasionada pela lógica utilizada no controle de freqüências de candidatos, dentro do algoritmo do módulo de reconhecimento.

Destaca-se sobre a necessidade de desenvolvimento de um pós-processamento para validar o resultado obtido na saída do sistema. Durante a discussão dos resultados, buscou-se um valor limite de percentagem de intensidade relativa, a partir do qual valores acima seriam considerados verdadeiros e abaixo, falsos. No entanto, esta escolha deve ser criteriosa, uma vez que alguns sons falsos

apresentam intensidades muito elevadas se comparadas com sons verdadeiros. É preciso ter em mente que um valor elevado para o corte faria com que sons verdadeiros fossem ignorados, ao passo que um valor muito baixo faria com que sons falsos fossem levados em consideração. É preciso pesar o que é mais crítico antes de se implementar tal validação. Por ora, sugere-se considerar todas as notas de intensidade obtida não-nulas.

A avaliação do sistema como um todo, possível a partir da apresentação dos resultados da Seção 9.5, é incentivadora. Percebe-se que é necessário implementar melhorias na segmentação temporal de notas, uma vez que alguns conjuntos aglutinados refletiram no tratamento de tais misturas. Felizmente, também se observou que, apesar disso, as notas reconhecidas se mantiveram dentro do padrão esperado, com poucos desvios.

Também é importante ressaltar que não é possível gerar um valor padrão para o descarte dos sons falsos baseados em suas intensidades relativas, uma vez que muitas vezes a intensidade de um TI era maior do que a de um TC. No entanto, se admitirmos que o resultado do sistema corresponde a todos os valores de intensidade diferentes de zero, o resultado ainda assim é otimista.

Capítulo 10 - Conclusões

Neste projeto, desenvolveu-se um sistema capaz de identificar as notas musicais componentes de um som polifônico. Neste capítulo, evidencia-se o que foi feito, avaliam-se as contribuições do projeto, e, por fim, são indicadas as sugestões para novos caminhos visando a aprimorar o que é proposto aqui.

10.1 Contribuições

A idéia de transcrição automática aqui defendida é deveras complexa. Não é à toa que existem muitos métodos e a literatura tem se renovado constantemente nestes últimos anos.

O sistema aqui apresentado foi desenvolvido visando, principalmente, a segmentar um problema complexo em blocos individuais de fácil integração e construir tais módulos, mesmo que de forma rudimentar, para que se pudesse ter uma idéia da totalidade do processo. Desde o início, não se tinha como interesse um processo completo e fechado, mas um esqueleto funcional sobre o qual poderiam ser realizados outros projetos de aprimoramento.

Desta forma, mais do que descrever procedimentos internos, a maior contribuição deste projeto é organizar os requisitos, apresentando os conceitos relacionados, os objetivos e definir os seus componentes e como estarão interligados, construindo um sistema geral com funções bem definidas.

A partir das propostas apresentadas e de um sistema modulado funcional, melhoras podem ser implementadas e o resultado de determinadas alterações poderá ser analisado no contexto geral, efetuando-se comparações mais visíveis, uma vez que o impacto final de cada modificação poderá ser obtido.

Além disso, outro objetivo deste era oferecer uma documentação acessível. Assim, mapearam-se todos os pontos de melhorias imediatas a serem implementadas, e buscou-se identificar as principais fontes de erro.

10.2 Retrospectiva

O Capítulo 1 apresentou o tema, justificando a motivação para o desenvolvimento do projeto. Relatou a proposta do trabalho e a base de dados utilizada para testes de execução deste projeto. Detalhou também o formato de apresentação deste documento.

O Capítulo 2 pincelou levemente as estreitas relações entre música e matemática, apresentando a análise do som no domínio do tempo e no domínio da freqüência, bem como quais características podem ser observadas em cada um deles.

O Capítulo 3 apresentou uma visão geral do projeto e seus principais sistemas componentes. Assim, apresentaram-se os dois processos, de treinamento e de reconhecimento, e segmentou-os em módulos reaproveitáveis com finalidades

específicas. Apresentou-se o fluxo de dados, bem como a dedução lógica que levou a tal arquitetura, além de suas relações de entrada e saída.

O Capítulo 4 tratou o módulo de segmentação temporal do som de entrada, baseando-se no método de identificação de início de nota. Tal método, no entanto, restringe o funcionamento do sistema a sons de instrumentos percussivos, já que a identificação do ponto inicial de uma nota leva em consideração as variações relativas significativas da energia do sinal no decorrer do tempo.

O Capítulo 5 trabalhou a conversão dos dados do domínio do tempo para o domínio da freqüência. Assim, procurou-se uma transformada que apresentasse característica logarítmica, oferecendo diferentes níveis de resolução e, ao mesmo tempo, baixa complexidade computacional. A solução encontrada para tal questão foi a implementação da *transformada de* Q *limitado*.

O Capítulo 6 debateu sobre o grande problema de se lidar com sons polifônicos: a sobreposição de harmônicos. Discutiu-se sobre a probabilidade de um sinal ser corrompido, ou seja, sobreposto, propondo o cálculo de um vetor de probabilidade de seleção de um determinado harmônico para representar alguma característica do som completo. Em seguida, mostraram-se os requisitos para a criação de um filtro capaz de extrair as informações a partir de todos os harmônicos representativos do som, a partir da aplicação de tal vetor como peso. Relacionaram-se, então, alguns momentos em que este filtro seria utilizado.

O Capítulo 7 apresentou o módulo de modelagem de tom, integrante exclusivo do processo de treinamento, comportando-se como um identificador monofônico, uma vez que extrai os dados referentes à freqüência fundamental, a seus parciais harmônicos e a suas respectivas amplitudes, para serem armazenados e utilizados para consulta no módulo reconhecedor. Foi proposto um algoritmo e discutido como tais dados seriam armazenados.

O Capítulo 8 descreveu o módulo de reconhecimento polifônico. O processo é conduzido com base no rastreamento por candidatos a freqüência fundamental, em ordem ascendente de freqüência. Caso o som exista, calcula-se sua intensidade e ele é, então, subtraído, para que não interfira na identificação de sons em freqüências

mais altas. Algumas restrições lógicas para o funcionamento do algoritmo foram apresentadas, como a limitação da banda de operação do sistema, ou seja, a limitação das freqüências nas quais se procuram sons, para evitar que o tratamento dos mesmos gere muitos erros na saída.

O Capítulo 9 apresentou os testes de validação do procedimento e a análise de desempenho do sistema. Buscou-se explicar as restrições e apontar as principais falhas, a fim de que possam ser feitas melhorias posteriores.

10.3 Propostas para Trabalhos Futuros

Este Projeto Final apresenta o esqueleto estrutural de um Transcritor Musical, no qual apenas foi implementada a operação de Reconhecimento de Sinais Polifônicos. Desse modo, certamente novos estudos podem ser feitos, novas abordagens podem ser agregadas e melhorias efetuadas. A seguir, uma listagem com algumas propostas:

- Implementar uma interface gráfica para o sistema;
- Otimizar os códigos utilizados para a implementação das funções descritas para reduzir o custo computacional. O desenvolvimento deste foi executado no Matlab©, o que, por si só, já é um fator crítico no quesito desempenho. Uma implementação em uma linguagem compilável reduziria, em muito, o custo computacional. No entanto, deve-se observar que tal implementação estaria condicionada a manter a idéia de modularização aqui proposta;
- Implementar melhorias na segmentação temporal. Isso ficou evidente ao se exibir os resultados da Seção 9.5, em que pequenos desvios na detecção de início de nota foram o suficiente para a identificação de resíduos de notas que executadas em momento anterior e que já deveriam ter sido dadas como encerradas;
- Enfatizar e aperfeiçoar as técnicas de tratamento de sinais de baixa freqüência;
- Implementar melhorias na identificação de notas muito próximas sendo tocadas simultaneamente;

- Aperfeiçoar o algoritmo de decisão de nota, mediante dois possíveis candidatos, fazendo uso de algum parâmetro que possibilite diferenciar o timbre dos instrumentos;
- Utilizar outras abordagens do estudo de processamento digital de áudio e áreas afins, a fim de obter mais um meio de determinar as notas musicais;
- Aumentar o alcance da faixa de alcance de tons nas freqüências mais altas, através de um processamento de anti-*aliasing*, de forma a se aproveitar a faixa entre as notas C6 e C7.

Referências bibliográficas

[1]. Martin, K., "Automatic Transcription of Simple Polyphonic Music: Robust Front End Processing", MIT Media Laboratory, Perceptual Computing Section, Technical Report No. 385, Third Joint Meeting of the Acoustical Societies of America and Japan, Dezembro 1996.

[2]. Just Intonation Network [http://www.justintonation.net/]. Acesso em 10 de fevereiro de 2008.

[3]. Lazzarini, V. E. P., "Elementos de Acústica", apostila do Departamento de Artes da UEL, Londrina, 1998.

[4]. Benson, D., "Music: A Mathematical Offering", Cambridge University Press, 2006.

[5]. Bregman, A., "Auditory Scene Analysis", MIT Press, 1990.

[6]. Horward, D. M., Angus, J., "Acoustics & Psychoacoustics", Focal Press, Oxford, 1995.

[7]. Caderno de Leitura e Escrita Musical I, Escola de Música Villa-Lobos, Rio de Janeiro, Fevereiro, 2007.

[8]. Loureiro, M. A., Paula, H. B., "Timbre de um instrumento musical", Per Musi, Belo Horizonte, n.14, pp. 57-81, 2006.

[9]. Klapuri, A., "Automatic Transcription of Music", MSc. thesis, Department of Information Technology, Tampere University of Technology, 1998.

[10]. Scheirer, E. D., "Tempo and Beat Analysis of Acoustic Musical Signals", Journal of Acoustical Society of Americam, Volume 103 (1), pp. 588-601, 1998.

[11]. Brown, J. C., "An efficient algorithm for the calculation of a constant Q transform", The Journal of the Acoustical Society of America, Volume 92 (5), pp. 2698-2701, 1992.

[12]. Kashino, K., Nakadai K., Kinoshita T., Tanaka, H., "Application of Bayesian Probability Network to Music Scene Analysis", Computational Auditory Scene Analysis, Lawrence Erlbaum Associates, pp. 21-26, 1998.

[13]. Kuosmanen. P., “Statistical analysis and optimization of stack filters”, Ph.D. dissertation, Acta Polytechnica Scandinavia, Helsinki, Finland, 1994.

[14]. Astola, J., Kuosmanen, P., “Fundamentals of Nonlinear Digital Filtering”, CRC Press LLC, 1997.

[15]. Santos, C. N., “Representação espectral de sinais para transcrição musical automática”, tese de Mestrado, Programa de Engenharia Elétrica, COPPE, Universidade Federal do Rio de Janeiro, 2004.