UNIVERSIDADE FEDERAL DA BAHIA

ESCOLA DE POLITÉCNICA

PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA

GABRIEL SIMÕES GONÇALVES DA SILVA

UM MODELO COMPUTACIONAL PARA A TRANSCRIÇÃO AUTOMÁTICA DE MELODIAS PARA PARTITURA

SALVADOR
2008

UM MODELO COMPUTACIONAL PARA A TRANSCRIÇÃO AUTOMÁTICA DE MELODIAS PARA PARTITURA

GABRIEL SIMÕES GONÇALVES DA SILVA

Dissertação submetida à coordenação do curso de pós-graduação em Engenharia Elétrica da Universidade Federal da Bahia como parte dos requisitos para obtenção do grau de mestre em Engenharia Elétrica.

Área de concentração: Processamento de Sinais

Orientador: Antonio Cezar de Castro Lima

Banca examinadora:

Prof. Dr. Antonio Cezar de Castro Lima - UFBA
Prof. Dr. Márcio Fontana – UFBA
Prof. Dr. Luiz Wagner Pereira Biscainho - UFRJ

Salvador – Julho de 2008

AGRADECIMENTOS

Ao amigo e prof. Antonio Cezar de Castro Lima pela orientação deste trabalho e também por sua paciência, dedicação, comprometimento, motivação e ensinamentos a mim oferecidos não somente durante o desenvolvimento deste projeto, mas desde o meu primeiro contato com o programa, ainda como ouvinte em uma de suas disciplinas. Agradeço a ele, em especial, pela confiança depositada em mim e neste projeto, quando muitos se mostraram descrentes em relação ao seu potencial como pesquisa acadêmica multidisciplinar.

Aos meus pais e meu irmão por tudo o que pudemos compartilhar ao longo das nossas vidas, influência que sem dúvida foi decisiva para me moldar como o homem que sou hoje. Pelo amor, carinho, paciência, dedicação e compreensão a mim dispensados desde o dia 29 de outubro de 1983, e por tudo mais que eu nunca conseguiria demonstrar aqui apenas através de palavras; muito obrigado.

Ao meu sobrinho, Rafael Pacheco Gonçalves, pelo sopro alegre de vida trazido num momento tão difícil, me fazendo compreender que é necessário se manter forte frente aos obstáculos da vida, pois, ao final, não faltarão razões para reerguer a cabeça e seguir em frente.

Aos professores e colegas do mestrado em Engenharia Elétrica da Universidade Federal da Bahia pelos ensinamentos e experiências, técnicas e profissionais, compartilhadas durante os últimos dois anos. Aos professores e colegas da Faculdade Ruy Barbosa por todo o incentivo e acompanhamento mesmo após o término do meu curso de graduação.

A minha família e aos meus amigos, por todo o apoio e compreensão durante essa jornada, em especial nos momentos em que precisei abdicar da companhia de vocês para poder realizar este projeto. Obrigado por existirem na minha vida!

HOMENAGEM

Este projeto é dedicado a memória de Arthur de Assis Gonçalves da Silva; pai, amigo, companheiro, "professor", escritor, meu exemplo para toda a vida. A participação dele, seja através da sua simples presença, do seu humor característico, do seu incentivo e dos seus conselhos e conversas, foi de fundamental importância para a minha formação como pessoa e como profissional, para fomentar a minha crença em mim, em minhas capacidades e nos meus sonhos, para definir as minhas trilhas pessoas e profissionais (entre elas a acadêmica) e, conseqüentemente, também para o desenvolvimento deste trabalho.

Espero, pai, que onde quer que você esteja, esteja contente com o resultado de todo este esforço; e que ao final desta jornada eu tenha conseguido atingir o meu grande e eterno objetivo: o seu orgulho.

Te amo, "meu velho", fica com Deus.

"... A nau frágil enfrenta a tormenta, se rompe ao meio
A naufrágio alguém se entrega, se atira de corpo, esquece que é vida e que vida é pra ser vivida
Longe de estar eu perdido nesse mar, perto de estar eu com quem me queira amar ..."

Arthur de Assis Gonçalves da Silva

RESUMO

Mesmo já sendo um tema de estudo da computação musical há mais de 35 anos, a transcrição automática de sinais de áudio continua a ser um assunto em aberto, fomentando o desenvolvimento de pesquisas multidisciplinares por todo o mundo. O presente trabalho propõe um novo modelo computacional capaz de transcrever melodias, não obrigatoriamente monofônicas no sentido estrito, utilizando a partitura como notação de saída. Para atingir este objetivo, o modelo foi desenvolvido com base em uma arquitetura modular, capaz de extrair informações sobre as notas e pausas que compõem a melodia do sinal em análise, transcrevendo-as para estruturas musicais de mais alto nível. Os ótimos índices de acerto obtidos por meio dos testes efetuados indicam que é possível atingir desempenho satisfatório através da transcrição assistida por computador, porém esse é ainda um tema longe de poder ser rotulado como um desafio superado.

ABSTRACT

More than 35 years after the publication of its first dated research, automatic transcription of musical audio signals still figures as one of the most important topics in the computer music field, being the absense of a robust, unique and complete solution the strongest motivator for the development of new studies around the world. This work proposes a new computational model for transcribing melodies, which may or may not be strictly monophonic, using the score as the main output musical notation. For archiving this goal, the model must be able to extract raw information related to the notes and rests which compose the melody under analysis, transcribing them to higher level musical structures. The good results obtained by the application of test scenarios to a prototyped software based on the proposed model reveal that computer assisted transcription of melodies can archive acceptable performance, mostly comparable to human made transcriptions; however this topic is still far from reaching a unique and definitive solution.

SUMÁRIO

1 INTRODUÇÃO

Seguindo a tendência enunciada através da lei de Moore (MOORE, 1965), as capacidades de processamento numérico e de armazenamento de dados dos computadores digitais evoluíram rapidamente nas últimas décadas, ao ponto de permitir que processos antes ditos improváveis de serem executados sem a interferência da inteligência humana, fossem, nos dias de hoje, completamente automatizados. Porém, mesmo com todo o desenvolvimento tecnológico, estes mesmos computadores ainda apresentam desempenho insatisfatório quando utilizados como substituto do homem no processamento de dados sensoriais, como a visão ou a audição (SMITH, 1997).

O desenvolvimento da percepção computacional da música tem como base conceitos e ferramentas matemáticas e lógicas aplicadas à análise de sinais de áudio digitalizados. Através de comparações entre os resultados providos por estes e características específicas dos sinais provenientes da execução de instrumentos musicais buscam-se associações que levem a conclusões embasadas em conhecimentos de teoria musical. Dentro desta área, a transcrição automática de sinais de áudio tem atraído o interesse de músicos e cientistas da computação por mais de 35 anos (MARTIN, 1996).

De acordo com (NAGARAJ, 2003), transcrever um trecho de música pode ser definido como o ato de escutar e escrever o conteúdo correspondente ao que se ouviu utilizando uma notação adequada e requer que se extraia deste as notas tocadas, as suas respectivas alturas e durações e a classificação dos instrumentos utilizados. Assim, a transcrição automática pode ser definida como o processo computacional de análise de sinais de áudio digitalizados e conseqüente extração de informações simbólicas que possam ser relacionadas com estruturas musicais em representações de mais alto nível (SCHEIRER, 1995). Os principais problemas correspondentes a esse processo podem ser resumidos à detecção do ritmo ou andamento da música, da altura e duração de cada uma das notas e pausas e a identificação e separação dos instrumentos musicais utilizados.

No processo de transcrição e escrita musical, a notação utilizada figura como um item crítico, já que esta representa a linguagem através da qual o músico poderá ter acesso às informações referentes à execução da peça. Por isso, a sua escolha deve sempre levar em consideração o músico e o instrumento a ser utilizado, além da precisão necessária entre a

representação e o resultado sonoro decorrente da sua execução. Notações resumidas e direcionadas a conjuntos de instrumentos específicos, como a tablatura, tendem a ser mais simples e de entendimento mais fácil, porém não são exatas na sua forma de representar, além de, muitas vezes, não serem adequadas para uso por diferentes instrumentistas (WEST; HOWEL; CROSS, 1991).

Do início das primeiras pesquisas na área até os dias de hoje, o campo de aplicação de modelos de transcrição automática de melodias se expandiu, transcendendo o cunho puramente acadêmico. Exemplo disso é o surgimento freqüente de novas ferramentas comerciais baseadas em transcrição por computador como sistemas de acompanhamento musical, auto-afinação, performance interativa e *e-learning*, além dos sistemas de auxílio a transcrição e escrita musical (SIMÕES; LIMA, 2008). É importante destacar, porém, a dificuldade de utilização de boa parte destas ferramentas por muitos músicos devido à utilização de notações musicais não padronizadas, restritivas e incompletas como linguagem de saída de informações.

Este trabalho tem por objetivo propor um modelo computacional capaz de efetuar a transcrição automática de melodias, não obrigatoriamente monofônicas[1] no sentido estrito, representadas através de sinais de áudio digitalizados. A notação musical escolhida para a representação das informações resultantes do processo de análise do modelo é a partitura, devido ao potencial de precisão da sua representação e, principalmente, a sua condição de notação musical universal, podendo ser aplicada à execução de qualquer instrumento (acompanhada ou não por uma bula).

A motivação para a realização deste projeto de pesquisa advém da popularização das plataformas de análise e processamento digital de áudio e conseqüente aumento do uso de sistemas de computação musical baseados em algoritmos de transcrição automática, aliados à inexistência, até a presente data, de uma solução única capaz de transcrever com perfeição as informações referentes à execução de uma melodia, independente do instrumento ou técnica de execução utilizada.

Algumas restrições de escopo foram definidas para o desenvolvimento deste trabalho, devido ao grande número de aspectos técnicos e informações existentes no processo de transcrição musical e também às dificuldades decorrentes do uso da partitura como notação para

[1] Monofônica: constituída por uma única voz melódica, sem qualquer acompanhamento. Assim, em uma melodia estritamente monofônica duas ou mais notas nunca soam simultaneamente.

representação das transcrições a serem efetuadas. Assim, o objetivo do modelo proposto estará limitado à segmentação temporal de melodias de acordo com a posição, duração e altura das notas e pausas presentes nesta, e associação de tais informações com estruturas musicais que possam ser representadas em uma partitura. Informações referentes ao instrumento utilizado, dinâmica e técnicas de execução serão desconsideradas. Por fim, ficará a cargo do usuário a definição da clave e das unidades de tempo e compasso a serem utilizadas pelo modelo, além do andamento a ser tomado como referência para efetuar as associações entre "duração real" e "duração musical".

O processo de desenvolvimento desta pesquisa está subdividido em três etapas. A primeira é elaborar a definição do referencial teórico através do levantamento de fontes de estudo sobre transcrição musical, modelos de transcrição automática e notações musicais que permitam compreender os seus conceitos. A segunda etapa é pesquisar sobre diferentes heurísticas de análise e de transcrição de sinais de áudio, e assimilar a teoria musical necessária a escrita e leitura de partituras. Por fim, compreender como integrar todos esses conceitos para segmentar as melodias em função de notas e pausas e transcrever os resultados de acordo com a notação escolhida.

A elaboração do modelo computacional requer um estudo sobre diferentes métodos de análise e identificação de freqüências fundamentais, *onsets*[2] e *offsets*[3], e heurísticas que possibilitem a transcrição efetiva dos dados levantados por estes para a partitura. Com as necessidades devidamente atendidas, um sistema será implementado tomando como base o modelo definido. Ao final serão realizados testes de validação através da análise de um conjunto de melodias e comparação dos resultados com transcrições-gabarito, criadas por músicos devidamente treinados.

Esta dissertação está organizada da seguinte forma: o Capítulo 2 detalha os conceitos sobre modelos para transcrição automática, detecção de *onsets*, *offsets*, extração de contornos de f_0[4] e notações musicais, sendo apresentado como referencial teórico para esta dissertação. O Capítulo 3 descreve o modelo proposto, a sua arquitetura, o sistema resultante, os testes efetuados e a análise dos seus resultados. Por fim, o Capítulo 4 apresenta uma conclusão sobre as questões abordadas e trabalhos futuros relativos a esta pesquisa.

[2] *Onset*: instante que marca o **início** da execução de uma nota.
[3] *Offset*: instante que marca o **fim** da execução de uma nota.

2 TRANSCRIÇÃO AUTOMÁTICA DE MELODIAS POR MEIO DE COMPUTADORES

A computação musical, com destaque para o processo de transcrição automática, é um tema por natureza multidisciplinar, reunindo conhecimentos de diferentes áreas como a computação, a música, a matemática, a física, a acústica e a psico-acústica, entre outras. Este capítulo tem como objetivo apresentar conceitos básicos sobre sinais de áudio, transcrição musical e sobre o processo de transcrição automática de melodias, suas dificuldades e propostas encontradas na literatura para a resolução das mesmas.

Devido à complexidade do problema abordado e, conseqüentemente, do seu perfeito entendimento, faz-se necessária uma introdução com foco interdisciplinar nas teorias relacionadas ao processamento digital de sinais de áudio e à segmentação automática de melodias, e na ligação destes processos com a teoria musical, suas notações e representações gráficas. Fazem parte do conteúdo abordado nas seções a seguir características dos sinais de áudio provenientes da execução de instrumentos musicais, modelos de segmentação automática de sinais melódicos, heurísticas de análise e detecção de freqüências fundamentais, *onsets* e *offset* e notações musicais, com destaque para a partitura.

2.1 Sinais

Segundo Smith (1997), um sinal é uma descrição de como dois diferentes parâmetros estão relacionados um ao outro. Assim, para a audição humana o som pode ser analisado como a representação do comportamento da pressão sonora exercida no sistema auditivo no tempo, enquanto em sistemas elétricos como microfones o mesmo sinal é descrito pela variação de tensão elétrica em um determinado ponto do circuito no tempo.

Existem na literatura diversas classificações para sinais; porém, no contexto desta pesquisa será necessário apenas analisar a sua classificação como periódicos ou não periódicos, de acordo com a sua definição no domínio da freqüência, e como contínuo ou discreto, com base na sua representação no tempo.

[4] f_0: sinônimo de freqüência fundamental.

Através dos estudos apresentados por Jean Baptiste Fourier, pode-se afirmar que um sinal é composto pela adição de diferentes senóides puras, cada qual com a sua própria freqüência, amplitude e fase (STEIGLITZ, 1996) (figura 1). Partindo desse princípio, um sinal é classificado como periódico quando este é composto por um número finito de harmônicos, enquanto um sinal dito não periódico é constituído pela soma de infinitas senóides.

Figura 1. Ilustração do processo de composição de um sinal a partir de diferentes senóides puras

Segundo (HAYKIN; VEEN, 1999), um sinal é dito contínuo quando a sua amplitude (ou outro parâmetro qualquer) varia continuamente em relação ao tempo, enquanto num sinal discreto os valores para esse mesmo parâmetro são definidos apenas em instantes discretos do tempo (figura 2). Mesmo os computadores modernos não são capazes de processar ou armazenar sinais contínuos de qualquer natureza, sendo então necessário que estes sofram um processo de discretização antes que possam ser manipulados digitalmente, denominado digitalização.

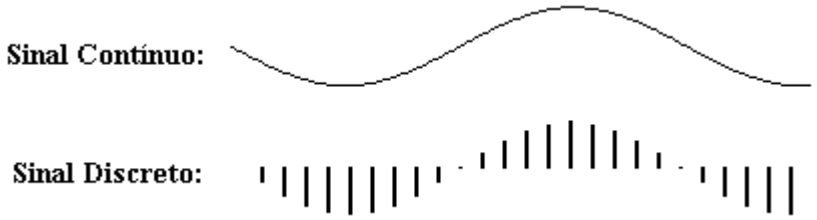

Figura 2 - Representação visual de um mesmo sinal nas formas contínua e discreta no tempo
Fonte: BORES, 2004

O processo de digitalização é realizado por meio de um circuito elétrico chamado conversor analógico-digital (A/D). Através deste, a tensão elétrica do sinal a ser digitalizado é mensurada e armazenada em intervalos regulares de tempo, sendo o valor da sua amplitude quantizado de acordo com a resolução numérica determinada, gerando uma amostra. Assim, fica claro que a digitalização ocorre no domínio do tempo em função da amplitude (WATKINSON, 1994), transformando um sinal analógico e contínuo em um sinal digital e discreto.

O período entre a captura de amostras utilizado no processo de digitalização define a taxa de amostragem do sinal (*Frequency Sampling* ou FS), ou seja, o número de amostras digitalizadas em 1 segundo. Com base no teorema de amostragem de Nyquist (também conhecido como teorema de Shannon), a menor taxa de amostragem que possibilite a digitalização da maior freqüência encontrada em um sinal sem criar distorções por superposição espectral, chamadas de *aliasing*, é maior que o dobro desta mesma freqüência (IFEACHOR; JERVIS, 2002). Já a resolução numérica, definida como o número de *bits* utilizados para representar e armazenar a quantização de cada amostra, influencia diretamente na precisão e, conseqüentemente, na fidelidade do sinal digitalizado em relação ao original.

2.1.1 Sinais de áudio provenientes da execução de instrumentos musicais

A análise de sinais de áudio provenientes da execução de instrumento musicais possibilita a identificação de características próprias destes que permitem uma associação direta com informações relativas à teoria musical como, por exemplo, a altura das notas.

Segundo (SMITH, 1997), a percepção do som pode ser dividida em três componentes básicas: volume, altura e timbre. O volume é definido como a característica relacionada à potência do sinal, sendo medido em decibéis. A altura está diretamente ligada a características de freqüência, sendo facilmente compreendida quando relacionada com notas graves ou agudas, sendo quantificada em *hertz* (Hz). Por fim, entende-se como timbre a característica sonora peculiar de cada instrumento musical que permite a sua diferenciação auditiva e identificação, mesmo quando em meio a outras fontes sonoras.

Através do estudo de sinais referentes à execução de uma única nota musical é possível identificar um padrão de comportamento aproximado, composto por quatro diferentes estágios chamados ataque, amortecimento, sustentação e dissipação. Este modelo de evolução de notas musicais, conhecido como envelope ADSR (*attack*, *decay*, *sustain* e *release*), leva em consideração não apenas informações relacionadas ao comportamento da energia do sinal no domínio do tempo, mas também características relativas ao caráter do seu espectro de freqüência, apresentadas na tabela 1.

Tabela 1 – Modelo ADSR para a evolução de notas musicais no tempo

Estágio	Energia	Caráter
Ataque	Cresce rapidamente	Transitório
Amortecimento	Decresce	Em estabilização
Sustentação	Permanece constante ou decresce lentamente	Periódico (*quasi*-harmônico)
Dissipação	Decresce rapidamente	Periódico (*quasi*-harmônico)

O espectro de freqüência do sinal de áudio de uma nota musical nos estágios de sustentação e dissipação (já estabilizado) pode ser decomposto, basicamente, em uma componente localizada na freqüência fundamental (f_0), também chamada de primeiro harmônico, e numa série de harmônicos mais altos, múltiplos inteiros da freqüência fundamental (STEIGLITZ, 1996). Utilizando como exemplo a escala dodecafônica[5], considere a nota A3 (nota lá, oitava três) cuja f_0 é 220 Hz e as harmônicas são 440 Hz, 660 Hz, 880 Hz, etc., conforme apresentado na figura 3.

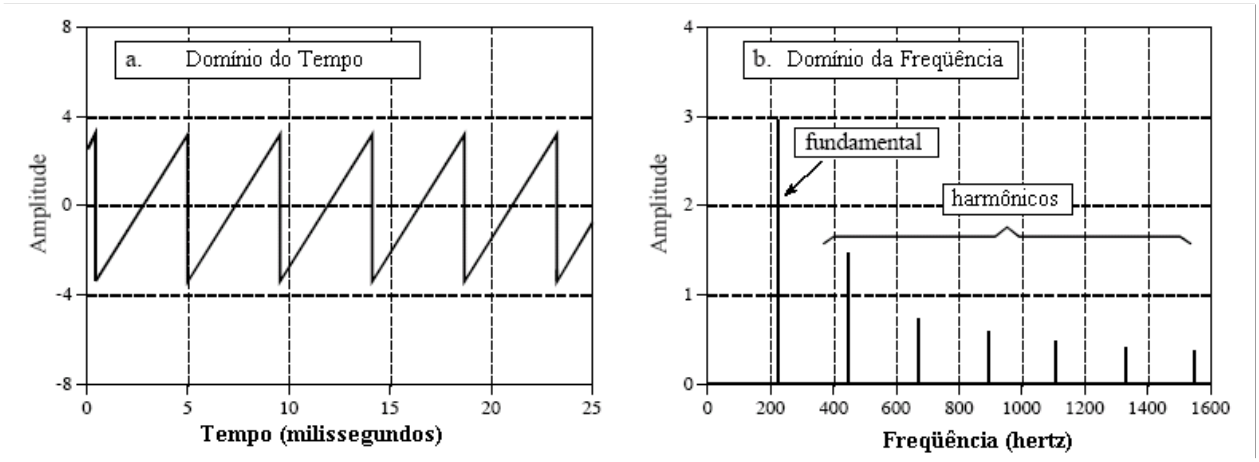

Figura 3 – (a) trecho do sinal de uma nota no domínio do tempo; (b) espectro de freqüência com destaque para a componente fundamental e às freqüências harmônicas
Fonte: SMITH, 1997

Segundo (SMITH, 1997), o termo oitava significa fator de dois em relação à freqüência fundamental e, no piano, poderia ser representado como uma distância de oito teclas brancas, contadas a partir da inicial. De forma mais clara, na distribuição das notas segundo a escala dodecafônica a f_0 referente a uma nota qualquer é exatamente o dobro da f_0 da nota distanciada uma oitava justa abaixo dessa. Considere como exemplo a nota A4 (nota lá, oitava

[5] Escala dodecafônica: escala musical composta por 12 notas, padrão do sistema musical ocidental.

quatro): a sua freqüência fundamental é 440 Hz, o dobro em relação à da nota A3 (220 Hz) e metade em relação à da nota A5 (880 Hz) (ver figura 4).

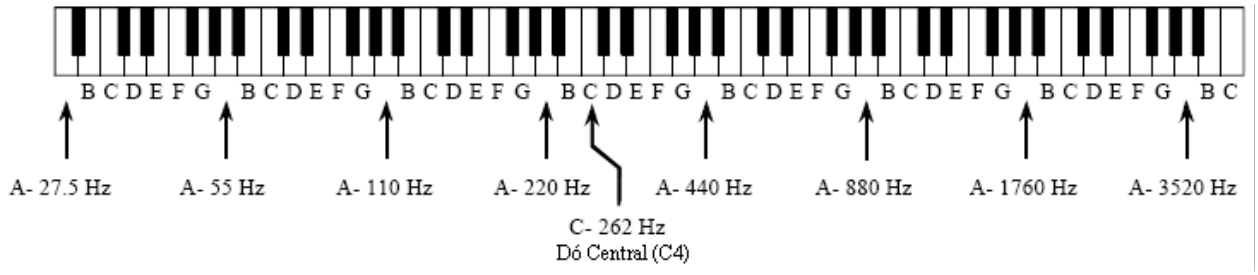

Figura 4 - Representação das teclas de um piano, com destaque para as notas lá em diferentes oitavas
Fonte: SMITH, 1997

2.2 Modelos de transcrição automática de melodias

Durante as mais de 3 décadas que sucederam os primeiros estudos da área, pesquisadores propuseram diferentes abordagens para solucionar, mesmo que em parte, os problemas intrínsecos da transcrição automática de sinais melódicos. O estudo destes modelos, suas características e funcionamento básico, permite evidenciar os pontos fortes e fracos de cada um, servindo assim como fonte de conhecimento para novos projetos.

Todo modelo de transcrição automática deve ser capaz de, primeiramente, segmentar o sinal no tempo em função dos seus componentes (pausas e notas, de acordo com as durações e alturas) e, em seguida, traduzir os dados levantados para a notação musical à qual se propôs. Tomando como base as diferentes abordagens para a segmentação temporal de sinais, os modelos encontrados na literatura podem ser classificados em 3 diferentes grupos:

- Grupo 1: segmentação através da identificação de *onsets* e *offsets*;
- Grupo 2: segmentação através de contornos de freqüência fundamental;
- Grupo 3: segmentação baseada em *onsets*, *offsets* e contornos de freqüência fundamental.

Na análise sobre abordagens anteriores a sua proposta de *framework* para sistemas da transcrição automática de melodias, (MITRE; QUEIROZ, 2007) apresenta uma breve descrição sobre os dois primeiros grupos acima citados, denominados abordagens clássicas. Segundo esta, no primeiro grupo, os algoritmos de detecção de *onsets* determinam o início das notas ou fim das

pausas, enquanto os detectores de *offsets* apontam o início das pausas ou fim das notas; a identificação das alturas fica a cargo da aplicação de um estimador de freqüências fundamentais nos trechos do sinal referentes às notas já segmentadas. No segundo grupo, o processo de segmentação é realizado com base única e exclusivamente no contorno resultante da detecção de f_0, onde os intervalos de tempo em que a melodia mantém a mesma freqüência fundamental são considerados notas e os intervalos onde nenhuma f_0 é detectada são considerados pausas. A tabela 2 lista os pontos fortes e pontos fracos de cada uma das abordagens.

Tabela 2 – Lista de pontos fortes e fracos dos modelos de segmentação dos grupos 1 e 2

	Pontos Fortes	Pontos Fracos
Grupo 1	• Possibilidade de uso de janelas de tamanho ótimo para a detecção de f_0; • Aumento da resolução espectral no processo de detecção de freqüências fundamentais.	• Erros de segmentação podem implicar diretamente erros de detecção de f_0; • Extrema dependência de detectores de *onsets*, módulo que, na média, apresentar a pior taxa de acerto entre os módulos de análise.
Grupo 2	• Dependência de uma heurística que, na média, apresenta boa taxa de acerto; • Uso de um único algoritmo de análise para a detecção de notas e pausas.	• Imprecisão temporal devido ao tamanho da janela e tamanho do salto (*hop size*); • Tendência de incorrer em atrasos em relação à posição real dos *onsets* e *offsets*.

O processo de segmentação temporal dos modelos pertencentes ao terceiro grupo pode ser visto, de forma geral, como uma fusão das duas primeiras abordagens acima apresentadas. Neste, o início das notas pode ser definido tanto por *onsets* como por variações no contorno de f_0, enquanto as pausas são segmentadas com base nos *offsets* e trechos onde nenhuma freqüência fundamental tenha sido detectada. Como vantagem, esse método permite a segmentação correta de eventos mesmo na presença de uma falha de análise, já que a redundância de informação permite que inconsistências sejam detectadas e corrigidas durante o processo. Em (SIMÕES; FREITAS; SOUZA, 2006), a utilização dessa redundância é controlada através da priorização das informações de maior precisão em cada uma das análises (ex. precisão temporal das detecções de *onsets* e *offsets*), criando assim uma hierarquia de escolha de quais informações

serão utilizadas para cada necessidade. Como desvantagens, esta abordagem apresenta maior complexidade computacional e maior dependência em relação à taxa de acerto dos módulos de maior hierarquia.

Independentemente da classificação dos modelos estudados, os processos de segmentação temporal de melodias em função de notas e pausas tem como base informações providas por 3 diferentes famílias de heurísticas de análise de sinais:

- Análise e identificação de freqüências fundamentais: utilizada para extrair do sinal o contorno de f_0 relativo às alturas das notas da melodia a ser transcrita;
- Análise e identificação de *onsets*: capaz de identificar o momento da execução de cada nota da melodia;
- Análise e identificação de *offsets* ou silêncios: utilizada para identificar trechos de silêncio ou o momento relativo ao fim da execução de notas que precedem pausas.

As subseções a seguir apresentam cada uma destas com um maior nível de detalhamento, em conjunto com alguns dos exemplos encontrados na literatura que serviram como base para o desenvolvimento desta pesquisa.

Considerando os conceitos da engenharia de *software*, os modelos de transcrição automática podem ser também classificados segundo a sua arquitetura como acoplados ou modulares. De modo geral, os sistemas de arquitetura modular apresentam como benefício a facilidade na implementação e substituição de heurísticas relativas ao processo de transcrição, enquanto o uso de arquiteturas acopladas possibilita uma melhor integração e otimização dos algoritmos implementados, acarretando ganho em desempenho computacional em detrimento da facilidade de manutenção, atualização e evolução.

2.2.1 Análise e identificação de freqüências fundamentais

A identificação de freqüências fundamentais, também chamada de detecção de altura (*pitch detection*), tem como objetivo extrair o contorno de f_0 de sinais de áudio (figura 5) e, a partir deste, determinar a altura das notas presentes na melodia analisada.

A altura de notas em um sinal de áudio é uma característica puramente perceptual, relevante apenas para o contexto de quem escuta este sinal (GERHARD, 2003). Porém, mesmo

sabendo que altura e f_0 são conceitos diferentes, em algumas situações eles são tão correlatos que é possível utilizá-los sem efetuar distinção. Segundo (MITRE; QUEIROZ, 2007), devido às características físicas da vibração de cordas e colunas de ar, a voz humana e a maioria dos instrumentos musicais produzem sons *quasi*-harmônicos dos quais a altura é determinada essencialmente pela freqüência fundamental. Assim, como na maioria dos trabalhos encontrados na literatura, nesta pesquisa os dois conceitos serão utilizados de forma indistinta.

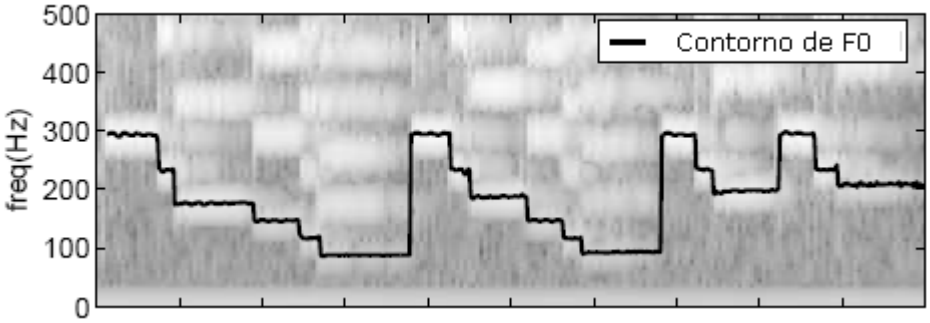

Figura 5 – Ilustração de um contorno de freqüências fundamentais sobre um espectrograma.
Fonte: RÖBEL; YEH; RODET, 2006

Do início das pesquisas no tema até os dias de hoje, diversos estimadores de f_0 para melodias já foram propostos, porém apenas alguns foram considerados robustos (RÖBEL; YEH; RODET, 2006). Uma das razões para isto é que a maioria das soluções propostas não apresenta taxas de acerto consistentes quando aplicadas a sinais de áudio de domínios diferentes daqueles nos quais foram projetadas para trabalhar. São exemplos de domínio de estimadores de f_0: música (monofônica ou polifônica, cantada ou proveniente da execução de instrumentos musicais), fala, etc.

As diferentes heurísticas pesquisadas se baseiam em inúmeros princípios matemáticos aplicados às informações do sinal, estejam elas no domínio do tempo ou da freqüência. No seu estudo sobre estimadores de f_0, (GERHARD, 2003) faz uma análise sobre diferentes abordagens efetuadas em cada um dos dois domínios. De forma resumida, e já ilustrando as classificações utilizadas para agrupar as heurísticas de acordo com o seu funcionamento, as conclusões apresentadas neste trabalho seguem descritas nas subseções a seguir.

2.2.1.1 Métodos de análise no domínio do tempo

- **Detecção da taxa de repetição de eventos no tempo:**

Família de métodos para a estimação de freqüências fundamentais através da detecção do período no qual um trecho da forma de onda do sinal se repete completamente. Os processos se baseiam na idéia de que se um sinal é periódico, então existem eventos que se repetem no tempo e podem ser contados. O número de vezes que estes acontecem em um segundo é inversamente proporcional a sua freqüência.

A maior dificuldade desta abordagem é que sinais com espectros de freqüência mais complexos raramente possuem apenas um evento por ciclo. Já como vantagens, (GERHARD, 2003) destaca sua extrema simplicidade de entendimento e implementação, e também o seu baixo custo computacional.

- **Autocorrelação:**

Com base na idéia de que sinais periódicos apresentam funções de autocorrelação também periódicas, estes algoritmos procuram identificar a freqüência de repetição de eventos relativos à forma da onda em sinais de áudio.

O maior problema desta abordagem é identificar o valor da freqüência fundamental em sinais harmonicamente complexos. Nestes casos, a função autocorrelação apresenta como resultado um conjunto de picos relativos a cada um dos harmônicos que o compõem o espectro do sinal analisado, necessitando de alguma inteligência para definir um valor de f_0 com base nas informações levantadas.

Leitores com interesse especial na estimação de freqüências fundamentais por autocorrelação devem se sentir encorajados a ler (CHEVEIGNÉ, KAWAHARA, 2002).

2.2.1.2 Métodos de análise no domínio da freqüência

- **Relação entre componentes de freqüência:**

Métodos deste conjunto efetuam uma análise de todos os picos de freqüência presentes no espectro de cada janela do sinal e, com base nestes, identificam os possíveis candidatos a f_0, determinando o vencedor a partir de algum tipo de inteligência, como a definição de pesos de acordo com a magnitude dos parciais que compõem a série de cada candidato.

Entre as vantagens desta abordagem destaca-se a sua capacidade de identificar freqüências fundamentais mesmo quando ausentes no espectro analisado, quando presentes com baixa magnitude em relação às demais componentes, ou cuja série harmônica esteja incompleta (MITRE; QUEIROZ; FARIA, 2006). Como desvantagem pode-se indicar a necessidade e complexidade dos algoritmos de inteligência utilizados para determinar os possíveis candidatos e também selecionar a f_0 entre todos eles.

- **Métodos baseados em filtros combinados:**

O funcionamento deste método se baseia no fato de que ao se aplicar um filtro passa-faixa a um sinal, a amplitude da sua saída apresentará valores mais altos quando a sua freqüência central estiver alinhada com a freqüência de uma das componentes do espectro em análise. Dessa forma, é possível identificar as freqüências dos componentes espectrais de maior magnitude através de uma comparação entre as saídas de um conjunto de filtros passa-faixa igualmente espaçados ou de apenas um único filtro variando a sua freqüência central. Caso os componentes encontrados estejam regularmente espaçados, a distância entre eles definirá o valor da freqüência fundamental a ser determinada (LANE, 1990).

- ***Cepstrum*:**

Este processo é uma forma de análise espectral cujo resultado é a transformada de Fourier do logaritmo da magnitude do espectro de freqüência de janelas do sinal. O nome

cepstrum foi originado a partir da inversão das quatro primeiras letras da palavra *spectrum*, indicando um modelo de espectro modificado (BOGERT; *et al*, 1963).

A idéia por trás desse método parte do fato de que a transformada de Fourier de sinais de áudio *quasi*-harmônicos com altura definida apresenta um conjunto de picos igualmente espaçados. O cálculo do logaritmo da magnitude do espectro reduz os picos a uma escala utilizável, gerando como resultado uma forma de onda periódica no domínio da freqüência, com período equivalente a freqüência fundamental do sinal original. Assim, a transformada de Fourier desta produz um espectro com um único pico na f_0 a ser estimada.

Este método assume que o sinal apresenta parciais regularmente espaçados. Assim, quando aplicados a sinais com grande número de componentes inarmônicos ou compostos por apenas um parcial, este tende a apresentar resultados incorretos.

2.2.2 Análise e identificação de *onsets*

Dentre os diferentes processos de recuperação de informações através da análise de sinais de áudio, a identificação de *onsets* é o responsável por detectar os instantes equivalentes ao início da execução de cada uma das notas presentes na melodia. Para o melhor entendimento deste problema, (BELLO; *et al*, 2005) apresenta como referencial teórico inicial a distinção entre os diferentes conceitos correlatos:

- Ataque: intervalo de tempo em que a energia do envelope do sinal apresenta crescimento;
- Transitório: curto intervalo de tempo no qual o sinal evolui de forma não trivial ou imprevisível;
- *Onset*: instante que marca o início do transitório, equivalente ao início da execução de uma nota.

O correto entendimento destes e das suas similaridades permite abordar o problema da detecção de *onsets* sob diferentes pontos de vista, o que amplia o seu escopo de possíveis soluções.

Segundo (COLLINS, 2005), o processo de detecção de *onsets* é freqüentemente dividido em dois componentes, chamados função de detecção (em inglês *detection function* ou *novelty function*) e heurísticas de escolha de picos (*peak picking)* (figura 6). O primeiro tem como objetivo construir um sinal que represente as mudanças de estado (presença de transitório) de um

sinal musical, tipicamente em uma taxa de amostragem bem inferior à deste. Já o segundo pode ser descrito como o algoritmo aplicado à função de detecção com o intuito de identificar nesta os picos de variação equivalentes aos *onsets* do sinal. Também, de forma opcional, pré-processamentos podem ser utilizados no sinal para acentuar ou atenuar algumas de suas características, de acordo com a relevância destas em relação ao funcionamento dos componentes supracitados.

A função de detecção é o componente mais importante do processo de identificação de *onsets*, sendo a metodologia empregada na sua construção a maior diferença entre as diversas heurísticas já propostas para a resolução do problema. Segundo (BELLO; *et al*, 2005), este processo, chamado de método de redução, pode ser efetuado com base em 2 diferentes modelos de análise:

- Redução com base em características do sinal: estes métodos buscam ressaltar as variações de estado do sinal a partir da oscilação de características próprias deste, estudadas através de análises em diferentes domínios, como o tempo (variações de energia) e a freqüência (variações de magnitude e/ou fase);
- Métodos baseados em modelos probabilísticos de sinal: métodos estatísticos para a detecção da ocorrência de transitório, baseados no pressuposto que sinais musicais podem ser descritos por modelos probabilísticos.

Um estudo detalhado sobre o funcionamento de diferentes métodos pertencentes a cada um dos grupos acima listados foge ao escopo desta seção. Em caso de interesse, (COLLINS, 2005) e (BELLO; *et al*, 2005) apresentam um grande conjunto de análises descritivas e testes comparativos sobre algumas das heurísticas consagradas na literatura, tendo sido este conteúdo utilizado como referência para o desenvolvimento desta pesquisa.

Caso a Função de Detecção (FD) tenha sido bem construída, os momentos relativos ao início dos transientes no sinal serão ressaltados de forma a possibilitar a detecção de variações de comportamento bem localizadas. Assim, o objetivo dos algoritmos de *peak picking* é conseguir identificar na FD os picos de tais variações referentes aos *onsets* do sinal analisado. O processo de escolha parte da definição de um limiar inferior (*threshold*), acima do qual os momentos referentes a qualquer pico identificado são considerados como o início de uma nova nota. Outros dados como distância entre os picos podem ser levados em consideração como

forma de diminuir a detecção de falsos *onsets* e, conseqüentemente, aumentar a taxa de acerto do processo de identificação.

Segundo (BELLO; et al, 2005), existem duas abordagens principais para a cálculo da função limiar: estática e adaptativa ou dinâmica. Métodos da abordagem estática definem como *onsets* picos onde a função de detecção excede um valor fixo único, previamente definido. Já os métodos de abordagem adaptativa buscam ajustar o valor do limiar às variações de potência existentes ao longo do sinal em analise.

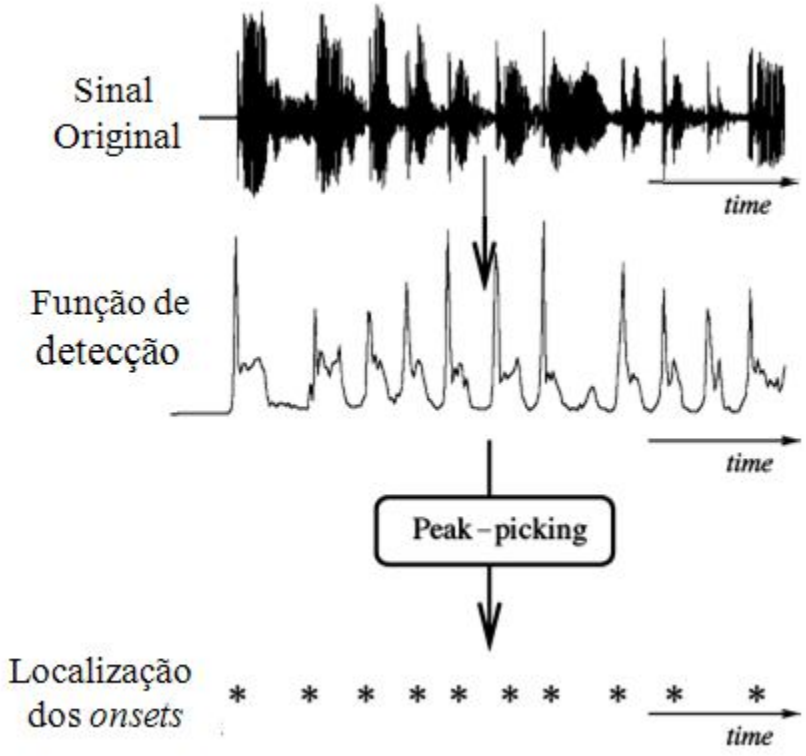

Figura 6 – Componentes do processo de análise e identificação de *onsets*.
Fonte: BELLO; et al, 2005

2.2.3 Análise e identificação de *offsets*

Considerando o ponto de vista da música, as pausas (silêncios) são tidas como componentes tão relevantes quanto as notas (sons). Assim, pode-se concluir que em uma melodia qualquer a correta transcrição e execução das suas pausas é tão importante quanto a correta transcrição e execução das suas notas.

No processo de transcrição automática, as diferentes heurísticas de análise e identificação de *offsets* são aplicadas a um sinal com o objetivo de detectar neste os momentos que marcam o fim da execução de cada uma das suas notas. No escopo da transcrição melódica, a identificação de *offsets* pode ser dividida, basicamente, em duas situações distintas (MONTI; SANDLER, 2000):

- *Offset* de notas sucedidas por notas: neste caso o *offset* da nota que se encerra coincide no tempo com o *onset* da nova nota executada, o que torna a sua identificação redundante;
- *Offsets* de notas sucedidas por pausas: este momento marca não só o fim da execução da nota em questão, mas também o início da execução da pausa que a sucede.

Para o bom entendimento desta pesquisa faz-se necessário esclarecer a distinção entre a identificação de *offsets* e a detecção de trechos de silêncio. Enquanto, na análise de sinais melódicos, a primeira busca identificar os momentos que marcam o fim da execução de notas que precedem pausas, a segunda tem como objetivo encontrar trechos em que a energia do sinal cai abaixo de um nível considerado audível. A diferença entre as duas análises pode ser compreendida como o tempo que o estágio de dissipação leva para que o sinal alcance um nível de energia considerado inaudível, sendo o tamanho deste intervalo diretamente influenciado por fatores como a sustentação natural do instrumento utilizado e a reverberação do ambiente no qual a execução da melodia foi captada.

Desde o início das pesquisas no tema “transcrição musical”, já foram propostas diferentes heurísticas com o objetivo de permitir a correta transcrição de pausas. Em (BROSSIER; BELLO; PLUMBEY, 2004) um *silence gate* foi utilizado como ferramenta capaz de identificar trechos de silêncio em sinais melódicos. O funcionamento desta consiste, inicialmente, no cálculo da envoltória de energia do sinal (através da sua retificação em onda completa seguida pela aplicação de um filtro passa-baixas) e acompanhamento da sua evolução no tempo. Para todos os trechos em que o nível de energia da envoltória cair abaixo de um limiar previamente definido, qualquer nota detectada deverá ser desconsiderada, tornando pausas as transcrições dos segmentos referentes a estes intervalos. Em (BROSSIER; *et al*, 2004) a mesma abordagem é utilizada, porém a envoltória do sinal é gerada através do cálculo da energia média

de janelas do sinal. O uso do *silence gate* em (MONTI; SANDLER, 2000) tem como intuito identificar, de forma aproximada, *offsets* de notas que precedem pausas.

Uma segunda abordagem para a identificação de silêncios ou *offsets* é o estudo de contornos de freqüências fundamentais. Em um sinal melódico, a detecção de f_0 em trechos de silêncio apresenta como resultado, longos "vales" cujo valor estimado é zero. Assim, ao estudar contornos de freqüência pode-se deduzir que os seus trechos sem freqüência fundamental definida e com duração considerada grande o suficiente caracterizam as pausas da melodia a ser transcrita.

Do ponto de vista técnico, as duas heurísticas supracitadas possibilitam a transcrição de pausas presentes em sinais melódicos; porém, a utilização destas não é adequada a modelos de transcrição associados a notações como a partitura devido a problemas de imprecisão temporal no processo de segmentação (SIMÕES; FREITAS; SOUZA, 2006). Em suma, além das imprecisões inerentes às heurísticas de cada processo, ambas as abordagens buscam identificar os momentos em que o sinal cai abaixo do limite audível (silêncio) e não os reais *offsets* que marcam o início da execução das pausas, sendo a diferença entre estes a principal causa do problema assinalado.

Visando melhorar a precisão temporal do processo de segmentação e assim permitir a correta transcrição das pausas e das notas que as antecedem, (SIMÕES; FREITAS; SOUZA, 2006) propôs uma heurística para a identificação de *offsets* com base nas características do envelope ADSR. A análise parte do pressuposto que o *offset* de uma nota é equivalente ao momento em que esta entra no estágio de dissipação. Dessa forma, o algoritmo procura identificar na envoltória do sinal o início de estágios de queda de abrupta energia com comportamento *quasi*-exponencial, sendo estes marcados como *offsets* caso o valor da envoltória caia abaixo de um limiar definido como limite audível. Segundo os autores, a solução proposta se mostrou adequada a um escopo restrito, apresentando bons resultados apenas quando aplicada a sinais sintéticos.

2.3 Notações Musicais

Notação musical é o nome genérico dado a sistemas de escrita utilizados para representar graficamente informações referentes a peças musicais, de maneira que o intérprete possa executá-la da forma mais próxima à planejada pelo compositor. Em linhas gerais, as

notações musicais buscam encontrar o melhor compromisso entre a riqueza de informações e a sua legibilidade. Quanto mais rica a notação, mais precisas podem ser as representações escritas com base nesta; porém, como conseqüência, estas também tendem a ser menos legíveis e mais complexas (WEST; HOWEL; CROSS, 1991).

Seguindo uma tendência também apontada por West, Howel e Cross (1991), uma parte crescente dos músicos tem buscado utilizar representações mais simples, o que levou à separação dos diversos elementos musicais em diferentes notações de cunho mais específico. Dentro desta tendência, destacam-se algumas notações:

- Cifra: concentra-se em informar sobre os componentes harmônicos da música, supondo o conhecimento da melodia, do ritmo e da dinâmica por parte do músico.
- Tablatura: contextualizada apenas para instrumentos de cordas, tem como maiores adeptos os guitarristas e baixistas. Dá ênfase às notas na ordem a serem tocadas e leva em consideração que o músico já conhece o ritmo e a dinâmica referentes à sua execução.
- *MIDI*: padrão desenvolvido para controlar instrumentos eletrônicos como teclados e sintetizadores. É escrita em linguagem de máquina, dificultando o seu entendimento e uso como notação musical para seres humanos. É a notação musical mais utilizada por sistemas de transcrição automática.

Caminhando no sentido contrário ao desta tendência, a partitura figura como uma notação musical completa, capaz de representar informações referentes à harmonia, melodia, ritmo, andamento e dinâmica, entre tantas outras. O seu domínio de aplicação engloba os mais diversos instrumentos, mesmo que, em alguns casos, seja necessário utilizar um guia de referência de execução, conhecido como bula.

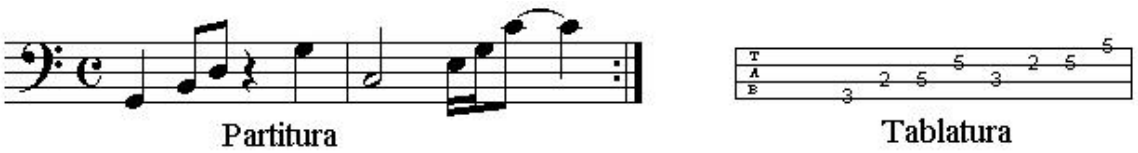

Figura 7 – a esquerda um exemplo de partitura; a direta a sua tablatura equivalente

Tomando como base a metodologia tradicional de escrita, a partitura é composta por um conjunto finito de símbolos básicos, apresentados de forma simples e ilustrativa a seguir (ver figura 8).

- Pentagrama: conjunto de 5 linhas e quatro espaços entre elas, utilizados como referência para indicar a altura das notas.
- Linhas e espaços suplementares: linhas e espaços posicionados acima ou abaixo do pentagrama, desenhados apenas quando necessário, com a mesma função e significado das linhas e espaços do pentagrama.
- Clave: símbolo que indica a extensão representada pelo pentagrama. Através desta os valores de altura referentes às linhas e pausas da partitura são definidos.
- Armadura de clave: conjunto opcional de sinais utilizado para simbolizar variações de altura que não podem ser representadas apenas pelo uso do pentagrama associado a uma clave.
- Unidade de compasso: unidade que define a figura relativa à pulsação do andamento de referência para execução da peça.
- Unidade de tempo: define o tamanho do compasso como um determinado múltiplo da unidade de compasso.
- Notas e pausas: figuras posicionadas diretamente no pentagrama. Seu desenho define a sua duração e, no caso das notas apenas, sua posição vertical define a sua altura.
- Barras de compasso: barras simples denotam o fim de um compasso e início do seguinte; barras duplas marcam o final do último compasso da partitura.
- Alterações: símbolos posicionados junto às notas para indicar alguma alteração relativa à sua altura ou duração.
- Sinais diversos: responsáveis por representar informações relativas ao andamento, dinâmica, efeitos e técnicas de execução, entre outros.

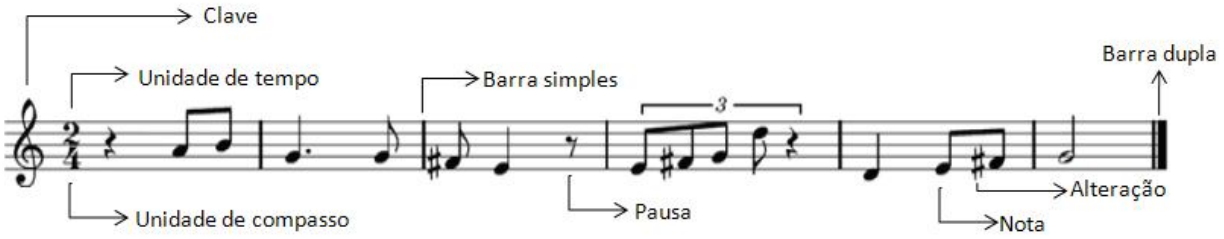

Figura 8 – Exemplo de partitura com destaque para alguns dos seus símbolos básicos

De modo geral, as regras básicas da partitura são genéricas o suficiente para serem aplicadas a todo e qualquer instrumento, respeitando as suas características peculiares. Porém, em casos específicos, os significados de alguns dos símbolos básicos da partitura podem ser alterados de forma a facilitar a escrita e a leitura das informações e/ou também a mecânica de execução de certos grupos de instrumentos. Uma destas exceções é o caso dos instrumentos transpostos, em que a altura das notas representadas na partitura é propositalmente diferente (por um intervalo fixo) da emitida pelo instrumento durante a execução da música.

2.3.1 Partitura

A música é criada através da união de sons e silêncios. Enquanto ambos apresentam a duração como uma de suas propriedades, apenas os primeiros possuem altura, que precisa ser definida durante a escrita em qualquer notação musical.

A altura das notas musicais em uma partitura é representada através de uma relação entre o pentagrama, a clave e os símbolos de alteração.

O pentagrama serve como base para a escrita das notas, de acordo com as suas respectivas alturas. Cada linha ou espaço irá representar uma das sete notas musicais (dó, ré, mi, fá, sol, lá, si) em uma determinada oitava. Como linhas e espaços suplementares, superiores ou inferiores, podem ser adicionados de acordo com a necessidade, a extensão da nota mais grave até a mais aguda que poderá ser transcrita é, em teoria, infinita. Porém, é importante ressaltar que a adição de linhas e espaços suplementares dificulta a leitura por parte do músico.

A clave tem como função, na partitura, atrelar o significado de uma linha específica do pentagrama a uma determinada altura. Por exemplo, a clave sol é utilizada para indicar que a segunda linha inferior do pentagrama representa a altura sol 4 ou G4. Através dessa informação, é

possível calcular a relação entre a linha já identificada e as demais linhas e espaços, visando determinar as suas respectivas alturas relativas. Dando prosseguimento ao exemplo anterior, o espaço acima da segunda linha inferior será um lá 4, a linha central um si 4, o espaço abaixo da segunda linha inferior um fá 4, a primeira linha inferior um mi 4 e assim por diante.

Os símbolos de alteração são utilizados para possibilitar a escrita de todas as notas da escala dodecafônica na partitura, modificando a altura de uma ou mais notas em uma determinada linha ou espaço. Podem aparecer em duas posições: em uma armadura de clave, indicando que a alteração é válida para toda a música até que outro sinal venha a anulá-la; e também antes de uma nota, assinalando que todas as repetições da mesma serão alteradas até o fim desse compasso.

De forma geral, as músicas têm as suas execuções atreladas a uma determinada contagem de tempo, também chamada de andamento ou pulsação. Os casos em que esse conceito não se enquadra são conhecidos como execuções de tempo livre e não trazem, na sua representação musical, figuras que determinem a duração de cada nota, mas apenas algo que marque a sua altura. Nos demais casos, o valor de duração de cada uma das figuras será determinado pela relação entre elas em conjunto com os valores de andamento e unidade de compasso.

Existem duas maneiras de grafar o andamento de uma música na sua notação: através de palavras que dêem idéia de velocidade para a execução, ou utilizando uma medida de pulsação fixa. No primeiro caso, não existe uma relação pré-definida entre a rapidez da execução da música e o tempo real. O andamento é definido de forma subjetiva, através do entendimento de uma palavra e posterior exteriorização de tal velocidade. Normalmente, cabe ao músico fazer tal análise e definir o andamento final para que a música seja executada. Alguns desses termos, em ordem crescente de velocidade, são: *largo*, *larghetto*, *adagio*, *andante*, *moderato*, *allegro*, *presto*, etc.

No segundo caso, o andamento está diretamente atrelado ao tempo real, já que a unidade utilizada como medida (BPM – batidas por minuto) se relaciona diretamente com a grandeza de tempo “minuto”. A representação, nesses casos, é feita através de um dos símbolos de duração de uma nota, um sinal de igualdade e um valor em BPM, indicando que essa figura terá a duração de um minuto dividido pelo valor do andamento. A exatidão de tempo dessa forma de representar permite converter computacionalmente o tempo real em musical, podendo ser utilizada, de maneira associativa, com outros algoritmos para a transcrição musical automática.

Uma das grandes vantagens da partitura, quando comparada a outras notações musicais, como a tablatura e a cifra, é a capacidade de representar não apenas a altura das notas, mas também a duração da execução de cada uma delas, além da duração das pausas.

A representação dessas durações é feita através de três grupos simbólicos: grupo das notas, grupo das pausas e figuras de alteração. É necessário destacar que essas figuras definem apenas uma razão de entre as suas durações, enquanto a relação entre o tempo real e o tempo musical será obtida apenas através da definição do andamento.

O grupo das notas e o grupo das pausas estão diretamente relacionados entre si. Cada um possui sete figuras, sendo que, para cada nota, existe uma pausa com duração correspondente. A tabela 3 ilustra as notas, pausas, suas figuras e a razão entre as suas durações.

Diversos símbolos podem ser utilizados para alterar as durações relativas às figuras de notas ou pausas. Os pontos de aumento simples ou duplos e a tercina podem ser empregados para ambos, enquanto a barra de ligadura é utilizada apenas para notas.

Tabela 3 - Notas e pausas, figuras e durações relativas

Nota	Figura	Pausa Relativa	Figura	Duração em relação a semibreve
Semibreve		Pausa de semibreve		----------
Mínima		Pausa de mínima		$\frac{1}{2}$ de semibreve
Semínima		Pausa de semínima		$\frac{1}{4}$ de semibreve
Colcheia		Pausa de colcheia		$\frac{1}{8}$ de semibreve
Semicolcheia		Pausa de semicolcheia		$\frac{1}{16}$ de semibreve
Fusa		Pausa de fusa		$\frac{1}{32}$ de semibreve
Semifusa		Pausa de semifusa		$\frac{1}{64}$ de semibreve

3 UM MODELO COMPUTACIONAL PARA A TRANSCRIÇÃO AUTOMÁTICA DE MELODIAS PARA PARTITURA

O modelo computacional proposto nesta pesquisa tem como objetivo transcrever melodias, não obrigatoriamente provenientes da execução de instrumentos monódicos ou gravadas em ambientes anecóicos, representadas por meio de sinais de áudio digitalizados e não comprimidos.

Através da utilização de heurísticas de análise e processamento das informações presentes nas fontes a serem transcritas, o modelo buscará segmentar o sinal no tempo em função das notas e pausas que compõem a melodia e, em seguida, transcrever os dados levantados para estruturas musicais de mais alto nível. Ao final do processo, a visualização da partitura será obtida através da formatação das informações já transcritas segundo o padrão definido pelo *Lilypond* (LILYPOND, 2008), um *software* livre sob licença *GNU* capaz de interpretar uma linguagem própria de descrição de dados para desenhar partituras.

O processo de transcrição musical envolve mais do que a segmentação do que se ouve em sons e silêncios. Existe também um conjunto de outras informações referentes a detalhes de execução como a dinâmica, técnicas e efeitos que precisam ser levados em consideração para que a reprodução do material transcrito seja a mais próxima possível da música original em termos de fidelidade sonora. Devido à subjetividade e também às dificuldades inerentes ao processo para a estimação precisa de tais informações, o escopo do modelo proposto foi restrito à transcrição da altura, posição e duração das notas presentes nas melodias dos sinais analisados, ignorando quaisquer outras informações, sejam estas relativas à dinâmica ou a sinais de execução.

Devido à escolha da partitura como notação musical padrão para a saída de dados, o modelo proposto precisa ser capaz de converter os valores de tempo referentes à posição e à duração das notas e pausas, de real para musical. Para isso, é necessário que este identifique e tome como base o andamento relativo à execução da melodia a ser transcrita, o que também foge ao escopo desta pesquisa. Assim, fica a cargo do usuário definir o andamento a ser utilizado como referência para a transcrição, além de outras informações ligadas à notação musical, como a clave, as unidades de tempo e compasso e uma duração mínima referente ao menor evento a ser levado em consideração pelo processo de transcrição.

3.1 Arquitetura

O modelo proposto se baseia em uma arquitetura modular e desacoplada, composta por quatro diferentes módulos com funções bem definidas e distintas (ver figura 9). Como indicado em (MITRE; QUEIROZ, 2007), esta arquitetura apresenta como vantagens a facilidade de avaliação independente de cada um dos módulos que a compõem e também a escalabilidade do modelo como um todo, permitindo a inserção, substituição e também remoção de módulos segundo as necessidades identificadas durante o processo de desenvolvimento.

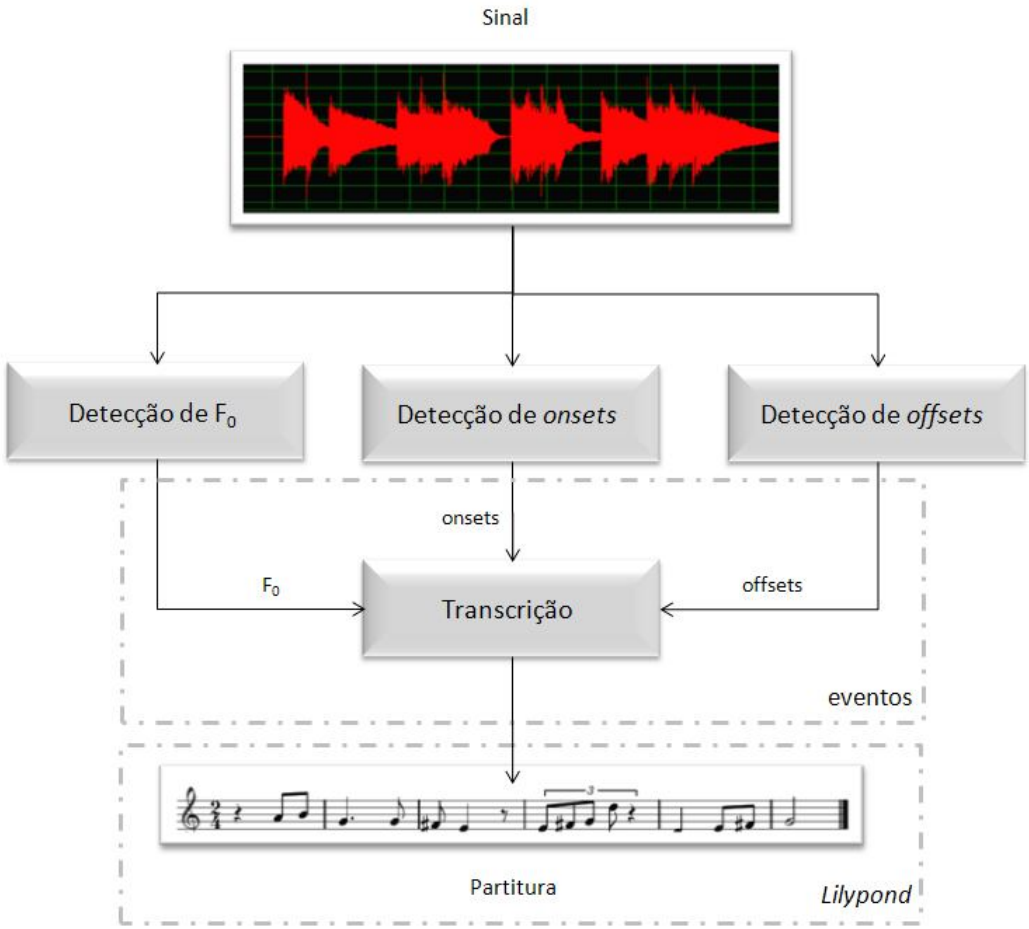

Figura 9 – Arquitetura do modelo proposto

O módulo detecção de f_0 tem como objetivo extrair do sinal o contorno de freqüências fundamentais referentes à altura das notas presentes na melodia a ser transcrita, aproximá-lo segundo o padrão ISO (A4 = 440 Hz) e, por fim, agrupar trechos com alturas

similares em eventos. O processo de análise se baseia na magnitude e na relação entre componentes de freqüência, a partir do espectro do sinal calculado através da aplicação da STFT (*Short Time Fourier Transform*) a blocos seqüenciais do mesmo, selecionados por meio de uma janela deslizante.

O segundo módulo de análise, chamado detecção de *onsets*, busca identificar o momento referente ao início da execução de cada nota presente no sinal. Já o módulo detecção de *offsets* é utilizado com o intuito de identificar apenas o momento em que notas que precedem silêncios deixam de ser executadas, sendo este instante em sinais melódicos equivalente ao início da execução das pausas.

É importante ressaltar que os três módulos de análise e extração de informações são executados de modo independente, não exercendo quaisquer influências uns sobre os resultados dos outros. Esta decisão visa evitar que erros de análise de um dos módulos se propaguem para os demais, deteriorando os resultados obtidos no final do processo.

Por fim, com base nas informações fornecidas pelos três módulos já citados, o módulo de transcrição efetua a segmentação do sinal e, em seguida, transcreve as notas e pausas para estruturas musicais de mais alto nível diretamente relacionadas à partitura, de acordo com os dados disponibilizados pelo usuário (andamento e unidades de tempo e compasso). O resultado produzido por este pode ser então processado pelo *Lilypond*, que ao final da transcrição irá desenhar a partitura, atingindo o objetivo traçado no início desta pesquisa.

As subdivisões desta seção apresentam com maior nível de detalhe as diferentes heurísticas e soluções aplicadas em cada um dos módulos acima apresentados.

3.1.1 Detecção de f_0

O módulo detecção de f_0 é o encarregado de extrair informações do sinal que permitam determinar a altura das notas que compõem a melodia em análise. Através da sua heurística de estimação, o espectro de freqüência de janelas do sinal é processado e analisado em busca de relações harmônicas que permitam identificar a freqüência fundamental da série de parciais que o compõem.

O processamento efetuado pelo módulo pode ser dividido basicamente em quatro etapas seqüenciais: escolha de parciais, estimação precisa de parciais, estimação de f_0 e análise da

evolução de parciais. A ferramenta matemática utilizada para possibilitar a análise do sinal no domínio da freqüência foi a STFT, aplicada a janelas deslizantes de tamanho e salto definidos pelo usuário de acordo com as características do sinal a ser transcrito.

O número de picos de freqüência que compõem o espectro de uma janela de um sinal de áudio complexo pode chegar a centenas ou até milhares. A funcionalidade básica da etapa de escolha de parciais é selecionar, dentre tantos picos, apenas as freqüências que sejam de interesse do modelo, eliminando a influência das demais. Assim, o módulo busca diminuir a interferência de ruídos e picos espúrios no processo de identificação da altura das notas, melhorando a taxa final de acerto.

Para selecionar os picos em meio a tantos candidatos, o modelo utiliza duas regras distintas. A primeira delas se baseia na extensão de altura de cada instrumento musical e visa delimitar o espectro a ser utilizado pelo modelo. Dessa forma são definidos os limites de freqüência ϕ_{min} e ϕ_{max}, dentro dos quais todos os candidatos, o que inclui a f_0 a ser estimada e toda a sua série harmônica, deverão estar localizados.

A segunda regra leva em consideração a magnitude de cada um dos picos de freqüência localizados dentro dos limites acima apresentados. Devido à variação da escala de magnitudes de acordo com o nível de entrada do sinal no momento da gravação, o módulo compara o valor absoluto da magnitude de cada parcial com o da maior magnitude encontrada no espectro de cada janela e seleciona apenas aquelas que estiverem entre 0 dB e λ dB abaixo desta. Os valores de ϕ_{min}, ϕ_{max} e λ são arbitrários, devendo ser escolhidos de forma a maximizar a taxa de acerto do módulo.

Sendo **V** o vetor resultante do processamento de uma janela deslizante do sinal através da STFT, **k** o seu índice de acesso, **FS** a taxa de amostragem do sinal e **N** o número de amostras da janela utilizada no cálculo da STFT, segue abaixo a formalização das regras apresentadas.

$$\phi_{\min} < k * \left(\frac{FS}{N}\right) < \phi_{\max} \tag{1}$$

$$10 * \log_{10}\left(\frac{V_{magnitude}[k]}{Max(V_{magnitude})}\right) + \lambda > 0 \tag{2}$$

$$\forall\, k \mid \begin{cases} k \in \left[0, \left(\frac{N}{2} + 1\right)\right] \\ V_{magnitude}[k] > V_{magnitude}[k-1] \\ V_{magnitude}[k] > V_{magnitude}[k+1] \end{cases}$$

O processo da estimação precisa de parciais está diretamente ligado aos problemas de resolução espectral decorrentes do uso da STFT como ferramenta para levar as informações do sinal do domínio do tempo para o da freqüência. Sinais analisados com baixa resolução espectral dificultam a diferenciação entre baixas freqüências, enquanto o aumento do número de amostras para incremento da resolução espectral leva ao crescimento do custo computacional e também à possibilidade de análise de duas ou mais notas numa mesma janela, somando assim as energias dos seus espectros.

Como forma de minimizar os efeitos decorrentes deste problema, foi implementado um algoritmo de interpolação quadrática proposto em Grandke (1983), capaz de precisar o valor da freqüência dos parciais com base na relação entre a sua magnitude e a maior magnitude entre as freqüências adjacentes. A interpolação é feita através da forma da janela de Hann (equação 3), sendo então obrigatório o uso desta como função de suavização das janelas do sinal antes do cálculo da STFT. Através desta técnica é possível identificar, com uma margem de erro de aproximadamente 3%, o valor real de uma freqüência **F** no espectro cuja duração seja maior ou igual ao dobro do período de **F**.

$$\mathrm{w}(n) = 0.5 * \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right), n\, \epsilon [0, N-1] \tag{3}$$

Informações comparativas sobre diversos algoritmos de interpolação para a estimação precisa de parciais podem ser encontradas em (JACOBSEN, 1994), (KEILER; MARCHAND, 2002) e (HAINSWORTH; MACLEOD, 2003).

Para ilustrar o efeito do processo de interpolação na resolução da STFT, imagine o seguinte cenário: a necessidade de extração de freqüências fundamentais de um sinal proveniente da execução de um contrabaixo-elétrico de 4 cordas, onde a nota mais grave segundo a afinação padrão é um Mi 1 (E1 – 41,2 Hz). Considerando a taxa de amostragem utilizada em gravações em

CD (44100 Hz), a não utilização de algoritmos de interpolação implicaria na necessidade de análise de janelas do sinal com duração mínima de 435 ms aproximadamente, enquanto através do uso da heurística de interpolação de (GRANDKE, 1983) esse período poderia ser reduzido para pouco menos de 50 ms.

Após selecionar e estimar a freqüência dos parciais é necessário identificar uma relação harmônica entre estes que caracterize a f_0 relativa à altura da nota em análise. A heurística de inteligência utilizada no modelo parte do pressuposto que a freqüência de maior amplitude presente no espectro faz parte da série harmônica da nota em análise (BROSSIER; BELLO; PLUMBEY, 2004). Assim, sendo $\mathbf{F_{max}}$ a freqüência do componente de maior amplitude, a lista de f_0 candidatas é definida de acordo com a equação 4.

$$F_0 \in \left\{ \frac{F_{max}}{j} \middle| 1 \leq j \leq \frac{F_{max}}{\phi_{min}} \right\} \tag{4}$$

Dentre todos os candidatos resultantes da regra supracitada, apenas um poderá ser escolhido. Para efetuar essa tarefa, a heurística do modelo partiu inicialmente do algoritmo proposto em (MITRE; QUEIROZ; FARIA, 2006), cuja idéia é, basicamente, definir notas para cada candidato a partir da atribuição de pesos às magnitudes dos parciais que compõem a sua série harmônica. Considerando apenas o conjunto de parciais anteriormente selecionados e sendo **Φ** a função proeminência harmônica do candidato a freqüência fundamental **Fc**, a sua nota será definida de acordo com a equação 5.

$$\Phi(n) = \sum_{i=1}^{\mathrm{I(n)}} \mathrm{Fc}[i]_{magnitude} * \psi(i) \tag{5}$$

$$\mathrm{I}(n) = max\{j : \mathrm{Fc}[j]_{magnitude} > 0\}$$

Onde: $\mathbf{Fc[i]_{magnitude}}$ indica a magnitude do *i*-ésimo parcial que compõe a serie harmônica do candidato **Fc** e **ψ[i]** denota a função fração de banda crítica, ou seja, peso com base psico-

acústica, fundamentada em Klapuri (2004), aplicado a magnitude do *i*-ésimo parcial da série harmônica em estudo, formalizada através da equação 6.

$$\psi(\mathrm{i}) = \begin{cases} 1, \mathrm{i} \leq 4 \\ \Gamma[\mathrm{i}] - \Gamma[\mathrm{i}-1], \mathrm{i} > 4 \end{cases} \tag{6}$$

$$\Gamma(n) = \log_{\frac{1}{2^{\frac{1}{3}}}}\left(n * \sqrt{\frac{n+1}{n}}\right)$$

Segundo o algoritmo como proposto originalmente, seria necessário ainda efetuar uma segunda rodada de estimação utilizando apenas os candidatos com um valor de proeminência harmônica de $\beta \in [0,1]$ em relação à maior proeminência encontrada. Essa necessidade se deve, em muito, a igualdade dos pesos aplicados aos quatro primeiros parciais das séries harmônicas, o que pode resultar em mais de um candidato com a mesma e maior proeminência harmônica no estudo de séries compostas por 3 ou menos parciais.

Assim, para evitar a necessidade de uma segunda rodada de estimação, diminuindo a complexidade do algoritmo, seu custo computacional e evitando a relatividade da escolha de β, foi efetuada uma alteração no cálculo da função fração de banda crítica através da diferenciação numérica dos seus quatro primeiros valores, como apresentado na equação 7.

$$\psi(i) = \begin{cases} 1.000, & i = 1 \\ 0.999, & i = 2 \\ 0.998, & i = 3 \\ 0.997, & i = 4 \\ \Gamma\,[\mathrm{i}] - \Gamma\,[\mathrm{i}-1], & i > 4 \end{cases} \tag{7}$$

$$\Gamma(n) = \log_{\frac{1}{2^{\frac{1}{3}}}}\left(n * \sqrt{\frac{n+1}{n}}\right)$$

Ao final do processo, a freqüência do candidato com maior proeminência harmônica será definida como f_0 da janela analisada. Caso exista um empate, vence o candidato com a menor freqüência.

Devido à natureza da solução proposta em (MITRE; QUEIROZ; FARIA, 2006), o valor da freqüência determinada como f_0 do espectro da janela do sinal em análise é estimado com base em apenas um parcial: o de maior magnitude no espectro. Dessa forma, qualquer erro na estimação da freqüência deste parcial é automaticamente propagado para o valor da f_0 identificada. Visando minimizar este efeito e tornar a solução ainda mais robusta, (MITRE; QUEIROZ; FARIA, 2006) propuseram uma etapa de re-estimação da freqüência fundamental com base em todos os componentes da sua série harmônica. Sendo **Fe** o candidato escolhido como freqüência fundamental com base nos valores de proeminência harmônica, a equação 8 descreve o processo de re-estimação para identificação da freqüência fundamental final.

$$F_0 = \frac{\sum_{i=1}^{I(n)} \frac{Fe[i]_{frequencia}}{i} * Fe[i]_{magnitude} * \psi(i)}{\sum_{i=1}^{I(n)} Fe[i]_{magnitude} * \psi(i)} \quad (8)$$

O resultado encontrado é, então, aproximado de acordo com os valores padrão definidos pela ISO, apresentados na tabela 4.

Tabela 4 – Valores da primeira oitava segundo a definição da ISO (A4 = 440,00 *Hz*).

Nota	Freqüência (Hz)	Nota	Freqüência (*Hz*)
A	27.50	D#	38.89
A#	29.14	E	41.20
B	30.87	F	43.65
C¹	32.70	F#	46.25
C#	34.65	G	49.00
D	36.71	G#	51.91

Fonte: INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 1975

Em situações perfeitas, onde todas as melodias seriam estritamente monofônicas, a heurística de estimação de freqüências fundamentais, acima apresentada, seria suficiente para extrair do sinal o seu contorno de f_0, atingindo ótimas taxas de acerto. Porém, como apresentado em (RÖBEL; YEH; RODET, 2006), devido a fatores não incomuns como o uso de instrumentos não monódicos ou a reverberação, o espectro da melodia analisada pode apresentar trechos de

polifonia em momentos relativos à execução de novas notas. Nestes casos, algoritmos de estimação de f_0 que esperam sinais estritamente monofônicos tendem a identificar como resultado um valor de freqüência fundamental que tenha na sua série os harmônicos pertencentes tanto ao espectro da nota atual quanto ao da nota anterior.

Para ilustrar o problema, imagine a execução de duas notas em seqüência: um Sol 4 (G4 – 392,00 Hz) e um Ré 5 (D5 – 587,36 Hz). Na existência de um curto trecho de polifonia no início da execução da segunda nota, o conjunto de parciais formado pela sobreposição dos espectros poderia enganar o estimador, favorecendo a identificação de um Sol 3 (G3 – 196,00 Hz) como freqüência fundamental já que todos os componentes das séries harmônicas das notas sobrepostas também fazem parte da série desta.

Visando aumentar a robustez da análise e ampliar o escopo de atuação do modelo proposto, um sistema de acompanhamento de evolução de parciais no tempo (SAEPT) foi adaptado à heurística de estimação de freqüências fundamentais apresentada em (MITRE; QUEIROZ; FARIA, 2006). Através deste, o módulo de detecção de f_0 buscará identificar a sobreposição de notas executadas em seqüência e bloquear o efeito dos componentes espectrais remanescentes, permitindo a correta detecção da freqüência fundamental referente à nova nota.

O funcionamento do SAEPT parte de dois diferentes pressupostos. O primeiro aponta que, caso existam trechos de polifonia que favoreçam a identificação de um sub-harmônico das duas notas sobrepostas, o valor da f_0 estimada será menor do que o da freqüência fundamental referente à nota anterior. O segundo diz respeito à evolução de energia dos parciais que compõem a série harmônica da nota em análise: em ambientes sem reverberação, a partir do estágio de estabilização, as magnitudes de todos os parciais da série apresentarão comportamento decrescente, apenas. Sob o efeito de fraca reverberação a energia dos parciais poderá apresentar oscilações positivas de baixa magnitude, devendo essas também ser levadas em consideração.

Inicialmente o módulo executa o algoritmo de estimação anteriormente apresentado, comparando o resultado da janela atual com a f_0 identificada na janela anterior: freqüências fundamentais maiores ou iguais às anteriores são automaticamente consideradas corretas, enquanto valores menores indicam a possibilidade de estimação de um sub-harmônico em conseqüência do prolongamento de componentes da nota anterior. Assim, para distinguir entre uma nova nota com f_0 mais baixa e uma detecção incorreta, o SAEPT precisa ser capaz de

identificar a presença de componentes da série harmônica da janela anterior em meio ao espectro da janela atual, e analisar a sua influência.

Para diferenciar entre componentes que realmente compõem a série harmônica da nota atual e parciais da série harmônica da janela anterior prolongados no tempo, foi definido um parâmetro chamado fator de oscilação **γ**. Sendo **k** o índice da janela do sinal, **i** o *i*-ésimo parcial do espectro da mesma, **SH** a série harmônica relativa à f_0 de cada janela e **j** o *j*-ésimo parcial da série, o método de seleção de parciais do SAEPT foi definido de acordo com a equação 9.

$$Janela[k][i]_{magnitude} = 0 \;\forall\; i, j \;|$$
$$\begin{cases} \left(SH[k-1][j]_{frequencia} * 0{,}97\right) < Janela[k][i]_{frequencia} < \left(SH[k-1][j]_{frequencia} * 1{,}03\right) \\ \dfrac{Janela[k][i]_{magnitude}}{SH[k-1][j]_{magnitude}} < \gamma \end{cases} \quad (9)$$

A definição do valor de **γ** deve ser de tal forma que os efeitos decorrentes da sustentação natural do instrumento ou reverberação sejam removidos, porém todos os parciais que compõem a série harmônica da freqüência fundamental da nova nota, coincidentes com os da nota anterior, sejam mantidos.

Ao final do processo de seleção, o algoritmo de estimação de f_0 é novamente executado, levando em consideração apenas o novo conjunto de parciais. Caso o resultado estimado seja menor ou igual ao da primeira detecção, o sistema considerará que não houve sobreposição espectral ou identificação de um sub-harmônico da freqüência fundamental, sendo a primeira f_0 identificada mantida como correta. Caso contrário, estará comprovada a influência da série harmônica da nota anterior sobre a janela atual, o que resultará na utilização da segunda f_0 estimada e no bloqueio de todos os parciais eliminados pelo SAEPT.

O conceito de parciais bloqueados tem como objetivo anular os efeitos da sobreposição espectral de notas na capacidade de identificação dos estimadores de freqüência fundamental não apenas na janela em que esta foi detectada, mas até que a influência da nota anterior se dissipe. Quando um parcial é marcado como bloqueado, este fica impedido de compor o espectro de qualquer janela subseqüente até que uma nova nota, de cuja série harmônica este faça parte, seja executada. Ao final de cada etapa de escolha e estimação precisa, cada parcial é confrontado com a lista de parciais bloqueados. Para todo caso de identificação positiva, o

SAEPT analisa a evolução da magnitude do parcial ainda de acordo com a equação 9. Caso a relação entre as magnitudes das janelas atual e anterior seja menor do que γ, o parcial será mantido como bloqueado. Caso contrário, o sistema entenderá que uma nova nota foi executada e que este parcial faz parte da sua série harmônica, devendo então ser removido da lista de parciais bloqueados.

Ao finalizar a análise de todo o sinal e já de posse do seu contorno de freqüências fundamentais aproximado segundo a definição da ISO, o módulo agrupa os trechos com mesmo resultado e distanciados por um período menor do que o relativo à menor nota/pausa a ser transcrita (definido pelo usuário) em eventos com início e duração definidos, e nunca sobrepostos. A lista de eventos resultante é, então, disponibilizada para o módulo de transcrição.

3.1.2 Detecção de *onsets*

A detecção de *onsets* é um tema de estudo de grande relevância na computação musical, tendo sido o foco de diversas pesquisas da área durante as últimas décadas. A partir destas já foram propostas muitas diferentes heurísticas para a solução do problema de identificação de *onsets*, cada qual aplicada a domínios distintos como sinais PNP (*Pitched Non Percussive* – não percussivos com altura definida) ou NPP (*Non Pitched Percussive* – percussivos sem altura definida).

Infelizmente, devido a fatores como a falta de uma metodologia de avaliação padrão para o problema e a inexistência de uma biblioteca comum de áudio e gabaritos de identificação que permitam a execução de testes em ambientes equivalentes, não é incomum encontrar inconsistências ao comparar resultados apresentados em diferentes pesquisas da área. Um exemplo disto pode ser identificado ao confrontar (BELLO; *et al*, 2005) e (COLLINS, 2005), onde os autores desenvolvem um estudo semelhante sobre heurísticas para o cálculo de FD´s, porém alcançam resultados divergentes em algumas das avaliações apresentadas.

Devido à existência de soluções robustas para a resolução do problema de detecção de *onsets* em sinais PNP e buscando aumentar a confiabilidade e o escopo de aplicação da solução, foram implementados neste modelo 4 diferentes abordagens para o cálculo da função de detecção, a serem escolhidas livremente pelo usuário no momento da transcrição:

- FD´s baseadas na variação de energia no espectro:

    - *Equal Loudness Contour* (COLLINS, 2005);
    - *Log Spectral Power* (COLLINS, 2005);
    - *Hi Frequency Content* (BELLO; *et al*, 2005);
- FD´s baseadas na variação de fase dos componentes do espectro:
    - *Phase deviation* (BELLO; *et al*, 2005).

A decisão de não escolher, inicialmente, uma única abordagem visa aumentar a robustez da solução proposta e também diminuir os riscos da escolha de uma heurística não tão apropriada para uma utilização generalista com base nos resultados apresentados em apenas um único artigo.

Depois de calculada, a função de detecção passa por um *Envelope Follower* com um filtro IIR passa-baixas, ativado sempre que a diferença de primeira ordem do sinal apresenta valores negativos. O objetivo de utilização do *Envelope Follower* na FD é suavizar os trechos de decaimento pós-*onset*, reduzindo o número de máximos locais de pequena expressão com o intuito de aumentar a taxa de acerto do módulo através da redução do número de falsos *onsets* identificados. A escolha do uso do *Envelope Follower* em detrimento de um filtro passa-baixas convencional se deve a sua característica de manter a precisão dos valores de energia e posição no tempo dos grandes picos da função, não distorcendo as informações referentes à posição dos *onsets*.

Por fim, a seleção dos picos e identificação dos *onsets* é realizada por meio de um algoritmo de *peak-picking* com base em uma função de limiar estática (valor fixo) ou adaptativa (BELLO; *et al*, 2005), ficando a escolha entre estas também a cargo do usuário.

3.1.3 Detecção de *offsets*

Os modelos de transcrição automática devem ser capazes de segmentar o sinal não apenas em função das notas que compõem a melodia em análise, mas também considerando as suas pausas. Dessa forma, a identificação precisa de *offsets* tem papel fundamental no processo, pois esta permite determinar os momentos em que notas que precedem silêncios deixam de ser executadas pelo músico, sendo estes momentos equivalentes ao início da execução das pausas.

Como apontado anteriormente na Seção 2.2.3, nenhuma das heurísticas de identificação de *offsets* ou silêncios encontradas na literatura durante o desenvolvimento desta pesquisa se mostrou adequada para o propósito deste projeto. Considerando a partitura como

notação de saída, a conseqüência direta da segmentação do sinal com base em informações imprecisas seria não apenas a transcrição incorreta da duração das pausas como também o encurtamento ou alongamento da duração das notas que a precedem.

Com o objetivo de suprir as necessidades do modelo, foi proposta uma nova heurística de detecção capaz de identificar com precisão os momentos relativos aos *offsets* de notas que antecedem silêncios. O funcionamento do módulo se baseia na análise do comportamento da envoltória do sinal a ser transcrito, comparando-a com os estágios do envelope *ADSR* e suas características. Dessa forma, diferentemente da solução proposta em (SIMÕES; FREITAS; SOUZA, 2006), o processo de detecção se torna genérico o suficiente para analisar tanto sinais sintéticos quanto reais, contanto que o comportamento do envelope das notas que os compõem siga o modelo *ADSR*.

Partindo do pressuposto que um *offset* marca também o início do estágio de dissipação de notas que precedem pausas (SIMÕES; FREITAS; SOUZA, 2006), o módulo busca inicialmente identificar os trechos de silêncio presentes na melodia. Para esta tarefa escolheu-se aplicar um algoritmo de *Silence Gate* à envoltória do sinal, calculado através da sua retificação de onda completa com posterior processamento por um filtro IIR (*Infinite Impulse Response* – Resposta de Impulso Infinita) passa-baixas. O modelo de filtro escolhido foi o Bessel de ordem 2, por sua característica de apresentar *overshooting* mínimo na sua resposta no domínio do tempo, atendendo as necessidades de atenuação. O valor da freqüência de corte foi definido empiricamente como 12,5 Hz. Em seguida, o módulo calcula a diferença de primeira ordem da envoltória filtrada com o intuito de ressaltar suas variações negativas de energia. O sinal resultante é também filtrado por um passa-baixas do tipo Bessel de ordem 2, com uma freqüência de corte um pouco mais baixa (10,00 Hz), suavizando-o de forma a minimizar possíveis erros de detecção decorrentes da presença de oscilações no mesmo.

Com base na determinação dos períodos de silêncio que compõem o sinal por parte do *Silence Gate*, o módulo busca identificar o *offset* associado a cada um destes através de um processo de análise do comportamento de energia do sinal diferença de primeira ordem. Partindo do momento equivalente ao final de cada silêncio (ponto de aumento de energia em que a envoltória do sinal e o limiar do *Silence Gate* se cruzam) e retrocedendo no tempo, a heurística proposta busca identificar o momento mais próximo que marque o início do crescimento (em módulo) da aceleração de decréscimo do envelope. Relacionando com o modelo *ADSR*, este pode

ser visto como o ponto que marca o início da queda *quasi*-exponencial da energia do sinal a níveis abaixo do limiar que delimita o nível do silêncio, comportamento que define o estágio de dissipação.

O processo de identificação do *offset* a partir desta etapa pode ser dividido na identificação seqüencial de três estágios de energia característicos. O primeiro deles é o mínimo local que antecede o ponto de início da busca, equivalente à maior variação negativa de energia do sinal durante o estágio de dissipação, também interpretado como fim da aceleração ou início da desaceleração do comportamento decrescente de energia do sinal.

Tomando o mínimo local acima descrito como um novo ponto de partida, o módulo inicia o processo de identificação dos dois estágios restantes, referentes ao comportamento da aceleração de decaimento da energia do sinal, através do cálculo ponto a ponto da diferença de segunda ordem do envelope. Com base nesta, o módulo procura os instantes referentes ao início da diminuição da aceleração (mínimo local que antecede o novo ponto de início) e, em seguida, o início do crescimento da aceleração (máximo local que antecede o ponto referente ao estágio anterior).

A partir das informações referentes ao último estágio identificado, o *offset* será assinalado de acordo com conjunto de regras abaixo:

- Caso no instante que marca o início do crescimento da aceleração a diferença de primeira ordem apresente valor positivo, o *offset* deverá ser assinalado no momento mais próximo a seguir em que esta assuma valor menor ou igual a zero;
- Caso no instante do último estágio a envoltória do sinal apresente valor abaixo do limiar do *Silence Gate,* o *offset* será assinalado, retrocedendo no tempo, no instante mais próximo em que o valor deste e da envoltória se equiparem.
- Caso as regras acima não se apliquem, o *offset* será assinalado no mesmo instante identificado como o início do crescimento da aceleração.

Considerando estas regras, é possível afirmar que apenas no pior dos cenários a heurística proposta apresentará um resultado tão impreciso quanto o obtido através do uso exclusivo de um *Silence Gate*.

3.1.4 Transcrição

Após extrair o contorno de freqüências fundamentais e identificar os instantes relativos aos *onsets* e *offsets* do sinal, o modelo proposto precisa ser capaz de segmentar este em função das suas pausas e notas, a partir das suas durações e alturas, e também transcrever os resultados obtidos de acordo com a notação musical a ser utilizada.

O processo de segmentação de um sinal pode ser resumido como a divisão deste de acordo com as informações fornecidas pelos módulos de análise. Como ilustrado na figura 10, o módulo busca alinhar os eventos de forma a possibilitar a identificação do início e também da duração de cada nota (segmento iniciado por um *onset* ou mudança de freqüência fundamental) ou pausa (segmento iniciado por um *offset*).

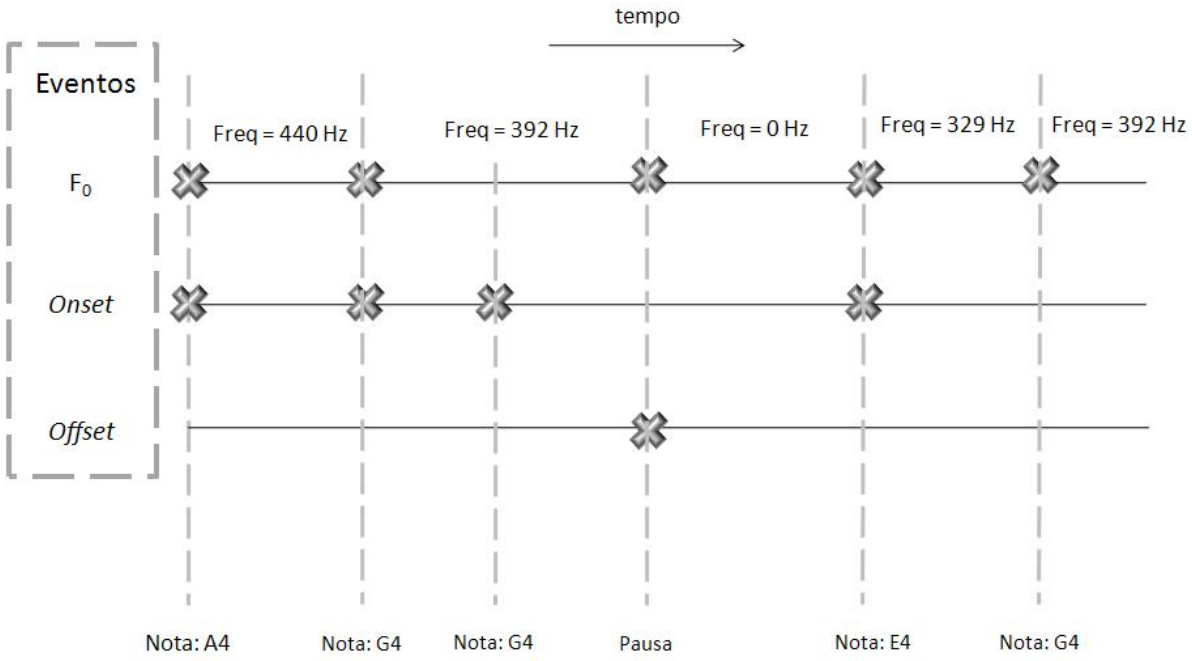

Figura 10 – Segmentação do sinal baseada nos eventos fornecidos pelos módulos de análise

A mecânica do processo de segmentação funciona como uma máquina de estados finitos (também chamada autômato finito), tendo como entrada o próximo evento entre os resultados dos módulos de análise de acordo com a sua posição no tempo. Caso a distância entre o evento anterior e o atual seja maior que o período mínimo definido pelo usuário, um novo

segmento deverá ser criado tendo como posição no tempo a mesma do evento anterior e como duração a distância entre este e o evento atual. Caso contrário, o módulo deverá aplicar um modelo de prioridades de acordo com a informação conflitante para decidir como proceder:

- Devido à maior precisão temporal do módulo de detecção de *onsets* em comparação com o módulo detecção de f_0, caso dois eventos, um associado a cada um destes, estejam distanciados de um período menor do que o mínimo definido pelo usuário, o evento de nova freqüência fundamental deverá ser posicionado no mesmo instante do evento de *onset*.
- Caso os eventos analisados sejam conflitantes, como, por exemplo, um *offset* e um *onset* ou uma nova f_0, a máquina de estados finitos deverá priorizar a utilização do evento *offset*.

Após segmentar o sinal e assim gerar a lista de notas e pausas que compõem a melodia, o módulo precisa ser capaz de transcrever as informações obtidas para a notação musical a ser utilizada, neste caso a partitura. O processo de transcrição da altura das notas é realizado através da comparação direta com os valores definidos no padrão ISO, apresentados na tabela 4, e das relações de altura apresentadas na Seção 2.1.1.

Para a transcrição da duração dos segmentos relativos às notas e pausas foi utilizada inicialmente uma abordagem de comparação direta com as figuras da partitura, de acordo com o andamento definido pelo usuário. Para todas as figuras possíveis, o módulo escolheria como resultado aquela com duração mais próxima ao valor real de duração do segmento. Porém, essa heurística apresentou uma taxa de acerto apenas razoável devido às imprecisões temporais dos módulos de análise e também da execução dos próprios músicos, gerando erros no resultado transcrito como compassos com duração incompleta ou que extrapolam a duração definida pelas unidades de tempo e compasso.

Visando aumentar a taxa final de acerto do modelo, foi proposta uma nova heurística para a transcrição da duração das notas e pausas com base em um processo iterativo de cálculo e análise de combinações. Assim, para cada período equivalente à duração de um compasso, o módulo busca encontrar de forma recursiva a melhor combinação entre as figuras e a duração dos segmentos transcritos.

Sendo esta uma abordagem executada compasso a compasso, para cada segmento da melodia pertencente ao compasso em análise o modelo identifica as duas melhores aproximações

entre a sua duração real e as durações fixas das possíveis figuras, guardando o valor referente ao módulo da diferença entre estas (módulo da aproximação) para uso futuro. Na seqüência, para cada uma das duas figuras escolhidas o algoritmo calcula todas as possibilidades de combinação com as aproximações das demais notas que a sucedem no mesmo compasso (figura 11). Ao final, será escolhida como transcrição correta a combinação cuja soma das durações das figuras seja a mais próxima da duração total do compasso, sem ultrapassá-la. Caso existam duas ou mais combinações com esta mesma duração, o critério de desempate será a menor soma dos módulos das aproximações.

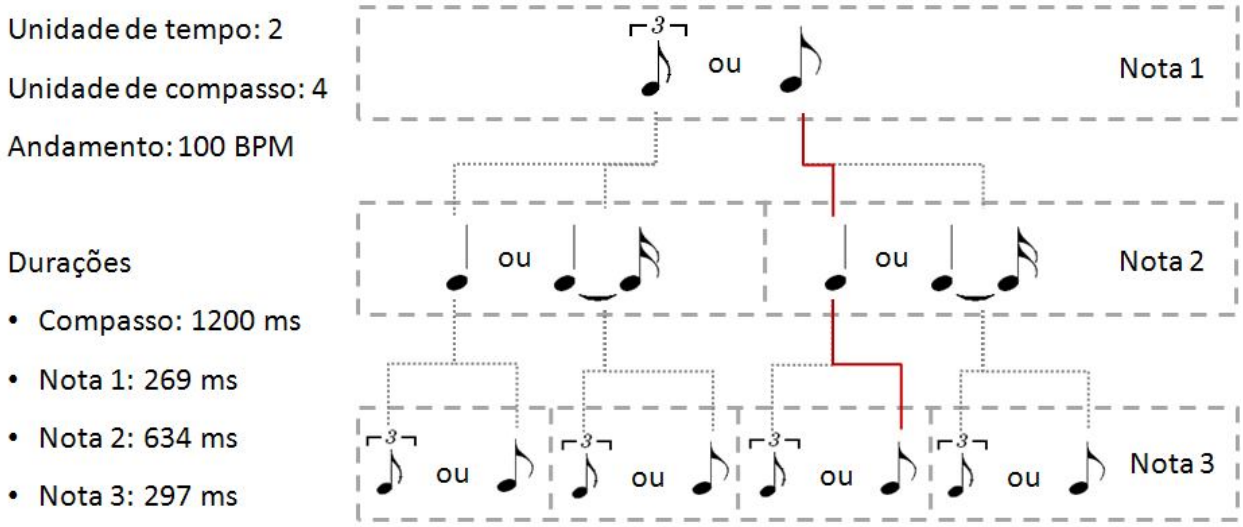

Figura 11 – Ilustração do processo de transcrição da duração de segmentos de um compasso. As linhas vermelhas indicam a combinação escolhida pelo módulo

3.2 Testes

Com o intuito de submeter o modelo a um cenário de testes e quantificar os resultados obtidos, foi implementado um protótipo de um sistema de transcrição a partir das heurísticas descritas na Seção 3.1. A linguagem escolhida para o desenvolvimento foi a C++, devido às suas características que facilitam o trabalho num nível de abstração tão baixo como o necessário para o processamento digital de sinais. Além disso, a C++ possibilita a programação orientada a objetos, sendo este paradigma adequado a implementação do modelo proposto. É importante destacar também que a biblioteca FFTW (FFTW, 2008), utilizada para efetuar o

cálculo da STFT de forma otimizada, é o único trecho de código não desenvolvido especificamente para esta etapa da pesquisa.

Os resultados obtidos através da aplicação do modelo proposto são o reflexo do resultado conjunto das análises de cada um dos módulos que compõem a sua arquitetura. Dessa forma, para validar o funcionamento de cada um destes o sistema foi inicialmente submetido a testes preliminares com o intuito de garantir que o funcionamento das heurísticas implementadas estivesse de acordo com o comportamento esperado. Devido à natureza qualitativa de tais testes e também ao objetivo final desta pesquisa, é suficiente afirmar que os resultados obtidos apresentaram consistência quando confrontados com gabaritos de transcrição criados por músicos profissionais sem o auxílio de qualquer sistema computacional.

Após confirmar o funcionamento adequado de cada um dos módulos anteriormente apresentados, o modelo de transcrição como um todo foi avaliado através da análise das transcrições de diferentes melodias. Devido à inexistência de um conjunto de critérios padrão para a avaliação de sistemas de transcrição automática, em especial para aqueles que utilizam a partitura como notação musical de saída, foi definido um conjunto de 10 critérios com o intuito de ilustrar não somente a precisão do resultado final das transcrições, mas também de refletir o desempenho de cada um dos módulos de análise:

- Notas corretas: número de notas transcritas com altura e figura de duração corretas;
- Notas com erro de altura: número de notas transcritas com altura incorreta;
- Notas com erro de duração: número de notas transcritas com altura correta, porém com figura de duração incorreta;
- Notas inexistentes: número de notas transcritas, porém inexistentes na melodia original;
- Notas não transcritas: número de notas existentes na melodia original, porém não transcritas;
- Notas com duração unida: número de notas cuja duração foi somada à da nota anterior;
- Pausas corretas: número de pausas transcritas com figura de duração correta;
- Pausas com erro de duração: número de pausas transcritas com figura de duração incorreta;

- Pausas inexistentes: número de pausas transcritas, porém inexistentes na melodia original;
- Pausas não transcritas: número de pausas existentes na melodia original, porém não transcritas.

Em decorrência da não existência de uma base de sinais melódicos padrão para a análise e validação de sistemas de transcrição monofônica (MITRE; QUEIROZ; FARIA, 2006), em especial composta apenas por melodias gravadas utilizando um andamento fixo como referência, foi criado um conjunto de sinais proprietário com o intuito de permitir a formalização e quantificação dos testes realizados. Este é composto por 308 notas e 54 pausas subdivididas em 22 melodias, gravadas com uma taxa de amostragem de 44100 Hz, agrupadas da seguinte maneira:

- 12 melodias sintetizadas a partir de amostras de sinais reais de diferentes instrumentos de diferentes famílias (baixo e guitarra elétricos, violino e violoncelo executados por meio do uso de arco, flauta, oboé, piano, piano rhodes, órgão, saxofone, trombone e trompete);
- 10 melodias gravadas a partir da execução de instrumentos reais (baixo e guitarra elétricos).

Para a realização dos testes, os diferentes parâmetros de configuração dos módulos que compõem o modelo foram ajustados da seguinte forma:

- Tamanho da janela do módulo de detecção de f_0: 2048 amostras;
- Tamanho do salto da janela do módulo de detecção de f_0: 1024 amostras;
- Fator de oscilação (γ): 2,5;
- Tamanho da janela do módulo de detecção de *onsets*: 2048 amostras;
- Tamanho do salto da janela do módulo de detecção de *onsets*: 256 amostras;
- Limiar do módulo de detecção de *onsets*: dinâmico, com valor de piso fixo igual a 19% do valor máximo da função de detecção;
- Limiar do módulo de detecção de *offsets* (*Silence Gate*): estático em 8% do valor máximo da envoltória;
- Tamanho do menor evento a ser transcrito: 115 milissegundos.

Para a detecção de *onsets*, definiu-se a utilização da heurística de *Equal Loudness Contours* como função de detecção seguida pela aplicação de um limiar dinâmico devido à taxa de acertos apresentada nos testes iniciais, superior a das demais heurísticas implementadas (listadas na Seção 3.1.2).

Com o objetivo de permitir uma melhor aferição dos resultados obtidos através de comparações diretas, o mesmo cenário de testes (base de melodias e critérios de avaliação) foi aplicado ao sistema *AudioScore Professional 3* (NEURATRON, 2008), *software* comercial especializado na transcrição de melodias para partitura. Para igualar as condições de execução dos testes e evitar distorções no processo de comparação dos resultados, a função de autodetecção de andamento do *AudioScore* foi desabilitada, sendo este valor definido manualmente de forma similar ao modelo proposto.

Seguem abaixo as informações obtidas através da análise e comparação entre as transcrições geradas pelos sistemas e transcrições gabarito, criadas por um músico profissional devidamente treinado. A tabela 5 apresenta os resultados obtidos durante a transcrição do conjunto de melodias sintetizadas, enquanto a tabela 6 ilustra os resultados das transcrições do conjunto de melodias gravadas através da execução de instrumentos reais.

Além dos testes descritos acima, um terceiro estudo foi efetuado com o intuito de analisar a capacidade do sistema de transcrever melodias não estritamente monofônicas. Para isto, foram escolhidas dentre o conjunto de melodias sintetizadas duas das quais o modelo proposto foi capaz de transcrever com perfeição (piano e flauta). Para montar o novo cenário de testes, estas foram processadas por um algoritmo de reverberação “leve” com o intuito de criar curtos trechos de polifonia nos momentos referentes ao início da execução de cada nova nota. É importante destacar que a intenção deste teste não é provar que através do SAEPT o modelo seria capaz de transcrever sinais polifônicos ou gravados em ambiente com forte reverberação, mas sim analisar a robustez do mesmo ao transcrever melodias que apresentem curtos trechos de polifonia em decorrência da execução de instrumentos não monódicos ou de suave e curta reflexão sonora do ambiente.

Tabela 5 – Consolidação dos resultados da base de dados de melodias sintetizadas de acordo com os critérios de avaliação previamente definidos

Melodias Sintetizadas		
Critérios de Avaliação	Modelo Proposto	*AudioScore*
Número de melodias	12	
Número total de notas	168	
Número total de pausas	24	
Total de notas transcritas corretamente	162	102
Total de notas transcritas com erro de f_0	0	0
Total de notas transcritas com erro de duração	4	57
Total de notas inexistentes transcritas	0	8
Total de notas não transcritas	1	3
Total de notas com duração unida	1	6
Total de pausas transcritas corretamente	24	12
Total de pausas transcritas com erro de duração	0	0
Total de pausas inexistentes transcritas	0	1
Total de pausas não transcritas	0	12

Tabela 6 – Consolidação dos resultados da base de dados de melodias reais de acordo com os critérios de avaliação previamente definidos

Melodias provenientes da execução de instrumentos reais		
Critérios de avaliação	Modelo Proposto	*AudioScore*
Número de melodias	10	
Número total de notas	140	
Número total de pausas	30	
Total de notas transcritas corretamente	133	85
Total de notas transcritas com erro de f_0	0	1
Total de notas transcritas com erro de duração	7	53
Total de notas inexistentes transcritas	0	6
Total de notas não transcritas	0	1
Total de notas com duração unida	0	0
Total de pausas transcritas corretamente	29	10
Total de pausas transcritas com erro de duração	1	10
Total de pausas inexistentes transcritas	0	0
Total de pausas não transcritas	0	10

As figuras 12 e 13 ilustram os resultados obtidos pela aplicação do módulo de detecção de f_0 ao analisar primeiramente os sinais sem a adição de reverberação e, em seguida, com a adição de reverberação utilizando ou não o sistema de acompanhamento de evolução de parciais no tempo. Já a tabela 6 apresenta um comparativo entre os resultados das transcrições criadas pelo modelo nos dois últimos cenários.

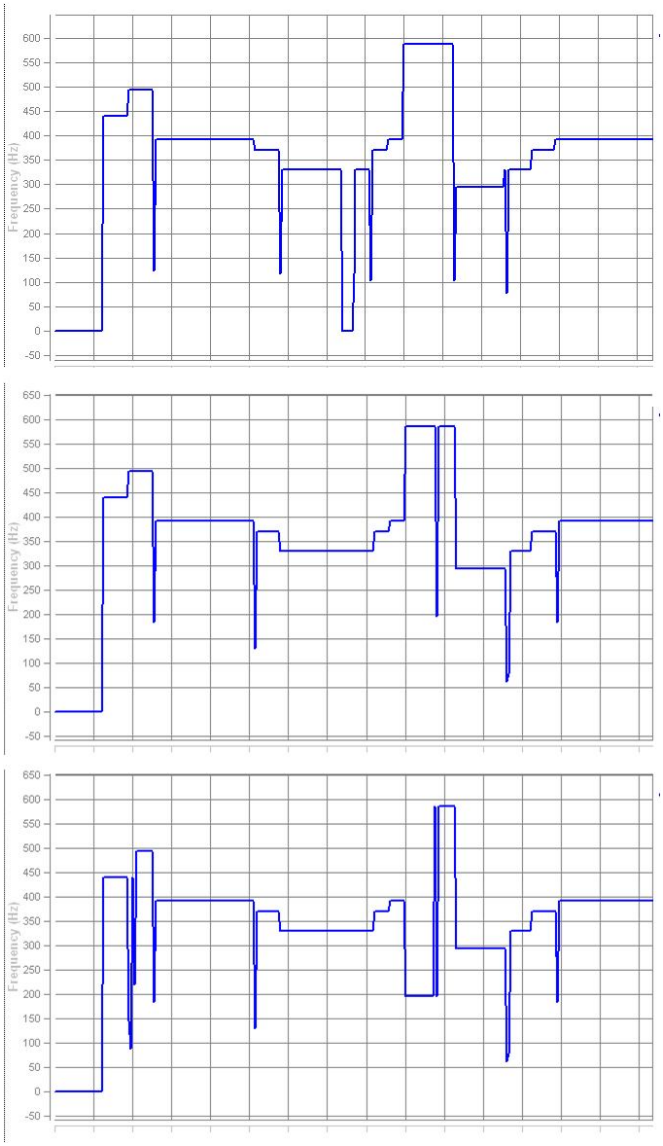

Figura 12 – Resultado da extração de contorno de f_0 da melodia de piano pelo módulo de detecção de f_0: Em cima a análise da melodia sem reverberação e utilizando o SAEPT; no centro a análise da melodia com reverberação e utilizando o SAEPT; em baixo a análise da melodia com reverberação e não utilizando o SAEPT

Figura 13 – Resultado da extração de contorno de f_0 da melodia de flauta pelo módulo de detecção de f_0: à esquerda a análise da melodia sem reverberação e utilizando o SAEPT; no centro a análise da melodia com reverberação e utilizando o SAEPT; à direita a análise da melodia com reverberação e não utilizando o SAEPT

Tabela 7 – Consolidação dos resultados da base de dados de melodias sintéticas com reverberação de acordo com os critérios de avaliação previamente definidos

Melodias sintéticas com adição de reverberação		
Critérios de avaliação	com SAEPT	sem SAEPT
Número de melodias	2	
Número total de notas	28	
Número total de pausas	4	
Total de notas transcritas corretamente	28	25
Total de notas transcritas com erro de f_0	0	0
Total de notas transcritas com erro de duração	0	3
Total de notas inexistentes transcritas	0	2
Total de notas não transcritas	0	0
Total de notas com duração unida	0	0
Total de pausas transcritas corretamente	4	4
Total de pausas transcritas com erro de duração	0	0
Total de pausas inexistentes transcritas	0	0
Total de pausas não transcritas	0	0

3.2.1 Análise dos resultados

Os resultados produzidos pelo modelo a partir dos testes relatados na Seção 3.2 superaram em muito as expectativas iniciais do projeto. O protótipo desenvolvido se mostrou estável durante todo o processo, apresentando qualidade de transcrição satisfatória para quase todos os sinais analisados, o que indica a adequação do modelo para o processamento de sinais provenientes de instrumentos de diferentes famílias.

Com base nos valores ilustrados nas tabelas 4 e 5, é possível afirmar que o modelo proposto se mostrou mais robusto e eficiente que o sistema *AudioScore 3* em todas as etapas do cenário de testes aos quais ambos foram submetidos. Tanto na transcrição de sinais sintetizados quanto na transcrição de sinais provenientes da execução de instrumentos reais, o modelo

apresentou índices iguais ou superiores em todos os critérios de avaliação considerados, o que reforça a idéia de adequação da solução proposta ao objetivo final do projeto.

O método de extração de contornos de freqüência fundamental implementado no módulo de detecção de f_0 apresentou altas taxas de acerto, identificando corretamente freqüências entre 82,4 Hz até 1174 Hz (freqüências fundamentais das notas extremas que compõem a base de dados analisada). Como conseqüência do bom funcionamento deste e também do algoritmo de agrupamento de janelas de f_0 em eventos, não foram encontradas nos resultados do modelo proposto notas transcritas com erro de altura, o mesmo não acontecendo com o *AudioScore 3*. A partir das análises efetuadas não foi possível identificar qualquer tipo de deturpação dos resultados obtidos pelo módulo em decorrência das alterações propostas à heurística apresentada em (MITRE; QUEIROZ; FARIA, 2006), fazendo valer a sua utilização em função dos seus benefícios de redução de custo computacional e também de diminuição da complexidade algorítmica.

Em relação aos testes ilustrados nas figuras 12 e 13, é possível perceber que a adição de reverberação ao sinal foi capaz de criar vales (erros de detecção) em alguns trechos referentes à execução de novas notas devido à polifonia criada pela sobreposição temporal destas em relação a suas antecessoras. Considerando tais trechos, a aplicação do SAEPT não se mostrou suficiente para eliminar por completo os efeitos provocados na capacidade de análise do módulo de detecção de f_0, porém conseguiu atenuá-los o suficiente para possibilitar a transcrição correta dos sinais analisados (vide tabela 6), aumentando a robustez da solução proposta. Os autores desta pesquisa compreendem a necessidade de se efetuar testes mais conclusivos em relação aos efeitos da utilização do SAEPT na detecção de freqüências fundamentais em janelas seqüenciais, porém acreditam, com base na sua fundamentação e nos resultados das análises preliminares, que em aspectos gerais o seu uso no pior caso leva a resultados semelhantes aos da sua não aplicação.

A definição do algoritmo de *Equal Loudness Contour* como solução para o módulo detecção de *onsets* permitiu a correta identificação do momento de execução da maioria das notas que compõem os sinais da base de dados analisada. As maiores distorções da função de detecção resultaram da análise dos sinais sintetizados a partir de amostras de sinais de violino, violoncelo e flauta, tendo sido interpretadas da seguinte maneira:

- O ruído branco gerado pelo arco ao friccionar as cordas do violino e do violoncelo cria uma variação freqüente de energia em diferentes faixas de

freqüência do espectro do sinal, fator este que distorce a função de detecção, dificultando a identificação dos máximos locais referentes aos *onsets*;

- No caso da flauta, a execução seqüencial de notas com uma mesma freqüência fundamental não resulta em variações relevantes de energia dos componentes do espectro (*soft onsets*), o que dificulta a diferenciação de tais momentos e, conseqüentemente, a identificação de tais *onsets*.

A heurística proposta para a solução do problema de detecção de *offsets* de notas que precedem pausas se mostrou não só adequada como extremamente eficiente, apresentando a maior taxa de acerto entre os três módulos de análise que compõem o modelo proposto. Em comparação direta com a solução descrita em (SIMÕES; FREITAS; SOUZA, 2006), sua predecessora na identificação de *offsets*, a heurística proposta apresenta como melhoria a sua adequação e bom funcionamento tanto para a análise de sinais reais quanto para sinais sintetizados. Já em relação às demais abordagens descritas na Seção 2.2.3, a solução proposta apresenta como vantagem a precisão superior dos seus resultados. É importante destacar também que mesmo durante a análise de sinais melódicos com adição de reverberação, onde os trechos de silêncio foram sobrepostos pela extensão do som das notas que os precediam, a heurística proposta para a identificação de *offsets* se mostrou capaz de identificar o momento do início das pausas com precisão suficiente para permitir sua correta transcrição.

O módulo de transcrição se mostrou como componente fundamental do modelo para alavancar o seu índice final de acertos. A inteligência aplicada à transcrição das durações dos segmentos do sinal, leia-se notas e pausas da melodia, funcionou como um sistema de recuperação de inconsistências no preenchimento dos compassos, não somente identificando erros, mas também propondo soluções de acordo com as diferentes possibilidades encontradas. O maior ganho identificado foi no processo de transcrição de sinais reais, onde as durações das notas e pausas executadas não são perfeitas, o que muitas vezes leva à associação a figuras de duração incorreta quando utilizados apenas sistemas de aproximação direta.

Mesmo não tendo sido levados em consideração processos referentes à otimização computacional durante o seu desenvolvimento, o protótipo implementado foi capaz de analisar e transcrever todos os sinais da base de dados em um período de tempo inferior à duração original das melodias. Baseado neste fato, é possível afirmar que, mesmo apresentando variações de custo computacional de acordo com a complexidade do espectro do sinal a ser analisado, o modelo

proposto neste projeto apresenta potencial para ser utilizado também em soluções que necessitem de transcrições *online*. Para estes casos é necessário destacar que, devido à heurística de transcrição das durações das notas e pausas, o período mínimo de atraso da saída do modelo é equivalente à duração de um compasso.

4 CONCLUSÕES

O presente trabalho apresentou um modelo computacional para a transcrição automática de melodias, não obrigatoriamente estritamente monofônicas, para partitura. Estudos na área vêm sendo realizados há mais de 35 anos, objetivando criar soluções robustas através da união de diferentes soluções pontuais para o conjunto de tarefas que constituem o processo de transcrição musical sem restrições.

Em decorrência de diferentes fatores como o surgimento de novas tecnologias e demandas, e também do avanço das pesquisas nas diferentes áreas da computação musical, o leque de aplicações amparadas em modelos de transcrição automática tem crescido ano após ano, sendo este fato comprovado pelo surgimento freqüente de novas ferramentas baseadas em transcrição com o auxílio do computador. Infelizmente, a maioria destas ferramentas utiliza notações musicais incompletas, imprecisas ou simplesmente não adequadas ao uso por qualquer instrumentista.

A escolha da partitura como notação de saída do modelo proposto agregou complexidade ao projeto, devido à necessidade de transcrever as informações decorrentes da segmentação do sinal para estruturas musicais de mais alto nível, em especial às dificuldades inerentes à transcrição das durações das notas e pausas. Por outro lado, o uso da partitura apresenta diversas vantagens como a exatidão das informações representadas através dela, a sua aplicabilidade para um grande número de instrumentos (necessitem estes ou não do acompanhamento de uma bula) e também por esta ser considerada a notação musical universal.

Para possibilitar a realização deste projeto, foi necessário unir conhecimentos de diferentes áreas como processamento digital de sinais, computação, física, música, matemática, acústica e psicoacústica. Como resultado, foi proposto um novo modelo de transcrição automática de melodias baseado em uma arquitetura modular, capaz de extrair de sinais de áudio provenientes da execução de instrumentos musicais informações como o seu contorno de f_0, *onsets* e *offsets* e, em seguida, segmentar o sinal, transcrevendo os dados resultantes de modo que o *Lilypond* possa desenhar a partitura referente à melodia analisada.

Através da realização deste projeto de pesquisa, foi possível concluir que a transcrição automática de sinais de áudio, monofônicos ou não, ainda não pode ser considerada um desafio superado. Isso porque, para atingir boas taxas de acerto, os escopos das soluções

propostas acabam sendo fortemente restringidos, ignorando informações importantes para a execução, como sinais de dinâmica, expressão e andamento, entre outros. Um segundo argumento é que ainda não foram propostas soluções para o problema capazes de garantir transcrições sem erros para sinais provenientes de todo e qualquer instrumento, em todas as suas possibilidades de execução.

A dependência de informações definidas pelo usuário, sejam estas ajustes manuais de algum parâmetro de um dos módulos ou dados necessários aos processos de transcrição (valor de andamento e unidades de tempo e compasso), pode ser considerada como o ponto fraco do modelo proposto. Dessa forma, trabalhos futuros devem focar na automatização da captura dessas informações com o intuito de simplificar o uso do modelo e torná-lo mais confiável, sendo menos suscetível a erros humanos.

Em relação ao desempenho das heurísticas propostas nesta pesquisa para o módulo de detecção de f_0, estudos mais aprofundados sobre os efeitos da utilização SAEPT poderão ser conduzidos com o intuito de identificar oportunidades de melhoria e de aumento de desempenho e robustez do modelo na análise de sinais não estritamente monofônicos. Também deverão ser avaliadas soluções com maior embasamento teórico para o processo de agrupamento de janelas em eventos, sendo a estratégia baseada em predição linear apresentada em (RÖBEL; YEH; RODET, 2006) um bom ponto de partida para este estudo.

O desempenho de cada um dos módulos também deve ser constantemente reavaliado através de comparações com as soluções tidas como *benchmark* para cada um dos problemas abordados. Em virtude da arquitetura utilizada, o processo de substituição, adição ou remoção de um dos módulos não implica diretamente na necessidade de alteração dos demais, o que facilita a evolução do modelo em acompanhamento ao desenvolvimento das pesquisas relacionadas às tarefas que constituem o processo de transcrição musical automática.

Por fim, seguindo a evolução lógica das pesquisas na área, novos estudos podem ter como foco a transcrição de sinais polifônicos, tema tido como o maior desafio das pesquisas na área de transcrição musical assistida por computador, ainda sem solução.

REFERENCIAS BIBLIOGRÁFICAS

BELLO, Juan; *et al*. A tutorial on onset detection in music signals. *IEEE Transaction on speech and audio processing*. [S.l], v.13, nº 5, p. 1025-1047, 2005.

BOGERT, B; *et al*. *The frequency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and shape cracking*. In: Proceedings of the Symposium of Time Series Analysis. New York:Wiley, c. 15, p. 209-243, 1963.

BORES SIGNAL PROCESSING. *Training in DSP and media processing*. Disponível em: <http://www.bores.com>. Acesso em: 26 abr.2008.

BROSSIER, Paul; BELLO, Juan; PLUMBEY, Mark. *Fast labelling of notes in music signals*. In: INTERNATIONAL CONFERENCE ON MUSIC INFORMATION RETRIVIAL, 05. 2004, Catalunha. Disponível em: <http://aubio.piem.org/articles/brossier04fastnotes.pdf>. Acesso em: 07 jun.2005.

BROSSIER, Paul; *et al*. *Real-time temporal segmentation of note objects in music signals*. In Proceedings of the International Computer Music Conference (ICMC 2004), 11. 2004, Florida. Disponível em <http://aubio.org/articles/brossier04realtimesegmentation.pdf>. Acesso em: 11 mai.2008.

CHEVEIGNÉ, Alain; KAWAHARA, Hideki. Yin, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, p. 111-115, 2002.

COLLINS, Nick. *A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions*. In: AES CONVENTION, 118. 2005, Barcelona. Disponível em: <http://www.aes.org/e-lib/browse.cfm?elib=13079>. Acesso em: 15 mar.2006.

FFTW; *Fastest Fourrier Transform of the West*. Disponível em: <www.fftw.org>. Acesso em: 02 jun.2008.

GERHARD, David. Pitch Extraction and Fundamental Frequency: History and Current Techniques. Technical Report TR-CS 2003-06. Department of Computer Science, University of Regina. 2003, Regina. Disponível em: <http://www2.cs.uregina.ca/~gerhard/publications/TR dbg-Pitch.pdf>. Acesso em: 19 mai.2008.

GRANDKE, Thomas. Interpolation algorithms for discrete Fourier transforms of weighted signals. *IEEE Transaction on Instruments and Measurements*. [S.l], v.32, nº 2, p. 350-353, 1983.

HAINSWORTH, Stephen; MACLEOD, Malcolm. On sinusoidal parameter estimation. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 06. 2003, Londres. Disponível em: <http://www.elec.qmul.ac.uk/dafx03/proceedings/pdfs/dafx04.pdf>. Acesso em: 13 nov.2005

HAYKIN, Simon; VEEN, Barry Van. *Signals and systems*. [S.l]: John Wiley & Sons, 1999.

IFEACHOR, Emmanuel; JERVIS, Barrie. *Digital Signal Processing, a practical approach*. [S.l]: Prentice Hall, 2002.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Acoustics -- Standard tuning frequency (Standard musical pitch)*. Disponível em <http://www.iso.org>. Acesso em: 20 set.2005.

JACOBSEN, Eric. On Local Interpolation of DFT Outputs. EF Data Corp, 1994. Disponível em < http://www.ericjacobsen.org/FTinterp.pdf>. Acesso em: 23 mai.2008.

KEILER, Florian; MARCHAND, Sylvain. Survey on extraction of sinusoids in stationary sounds. In: INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 05. 2002, Hamburgo. Disponível em: <http://www2.hsu-hh.de/ant/dafx2002/papers/DAFX02_Keiler_Marchand_sine_extract_compare.pdf >. Acesso em: 13 nov.2005.

KLAPURI, Anssi. *Signal processing methods for the automatic transcription of music*. 2004. Tese de PhD – Tampere University of Technology, Tampere, 2004.

LANE, John. Pitch detection using tunable IIR filter. *Computer Music Journal*. [S.l], v. 14, nº 3, p. 46-57, 1990.

LILYPOND. *Music notation for everyone*. Disponível em: <http://www.lilypond.org>. Acesso em: 2 mai.2008.

MARTIN, Keith. Automatic transcription of simple polyphonic music: robust front end processing. *M.I.T. Media Laboratory Perceptual Computing Section Technical Report*. Disponível em: <http://xenia.media.mit.edu/~kdm/research/papers/kdm-TR399.pdf>. Acesso em: 21 abr.2005.

MITRE, Adriano; QUEIROZ, Marcelo. A framework for low-latency music transcription. Departamento de Ciências da Computação, Universidade de São Paulo, 2007.

MITRE, Adriano; QUEIROZ, Marcelo; FARIA, Régis. Accurate and Efficient Fundamental Frequency Determination from Precise Partial Estimates. In: AES BRASIL CONFERENCE, 04. 2006, São Paulo. *Anais* - [S.l.]:[S.n], 2006, 113-118.

MONTI, Giuliano; SANDLER, Mark. *Monophonic transcription with autocorrelation*. In: THE COST G-6 CONFERENCE ON DIGITAL AUDIO EFFECTS, 01. 2000, Verona. Disponível em: <http://profs.sci.univr.it/~dafx/FinalPapers/pdf/Monti_DAFX00poster.pdf>. Acesso em: 07 jun.2005.

MOORE, Gordon. Cramming more components onto integrated circuits. *Proceedings of the IEEE*. [S.l]:[S.n], v. 38, nº 8, p. 114-117, 1965.

NAGARAJ, Keerthi. Toward automatic transcription - pitch tracking in polyphonic. *EE381K - Multidimensional digital signal processing*. Disponível em: <http://www.ece.utexas.edu/~bevans/courses/ee381k/projects/spring03/nagaraj/LitSurveyReport.pdf>. Acesso em: 21 abr.2005.

NEURATRON. *AudioScore Professional 3*. Disponível em <www.neuratron.com/audioscore.htm>. Acesso em: 17 abr.2008.

RÖBEL, Axel; YEH, Chunghsin; RODET, Xavier. Multiple f_0 Tracking of monodic instrument solo recordings. In: Audio Engineering Society Convention, 120. 2006, Paris. Disponível em: <http://mediatheque.ircam.fr/articles/textes/Yeh06b/>. Acesso em: 19 mai.2008.

SCHEINER, Eric. *Extracting Expressive Musical Performance Information from Recorded Music*. Tese de Mestrado – MIT, Cambridge, 1995.

SIMÕES, Gabriel; FREITAS, Allan; SOUZA, Hercules. Desenvolvimento de um sistema computacional de transcrição de melodias monofônicas para partitura. In: I Workshop de Computação e Aplicações – Congresso da SBC, 26. 2006, Campo Grande. *Anais* - [S.l]:[S.n], 2006.

SIMÕES, Gabriel; LIMA, Antonio. Um modelo para a transcrição automática de melodias para partitura. In: CONGRESSO DA AES BRASIL, 06; CONVENÇÃO NACIONAL DA AES BRASIL, 12. 2008, São Paulo. *Anais* - [S.l.]:[S.n], 2006.

SMITH, Steven. *The scientist and engineer's guide to digital signal processing*. [S.l.]: Califórnia Technical Publishing, 1997.

STEIGLITZ, Ken. *A digital signal processing primer, with applications to digital audio and computer music*. Menio Park: Addison-Wesley Publishing Company, 1996.

WEST, Robert; HOWELL, Peter; CROSS, Ian. *Musical Structure and Knowledge Representation*. In: *Representing Musical Structure*. Londres: Academic Press, 1991.