

Universidade Federal de Pernambuco

Centro de Informática

Graduação em Engenharia da Computação

*Benchmark* para Avaliação de Algoritmos de Reconhecimento de Acordes

Pedro Augusto Lopes Barbosa

Trabalho de Graduação

Recife

13 de dezembro de 2011

Universidade Federal de Pernambuco

Centro de Informática

Pedro Augusto Lopes Barbosa

*Benchmark* para Avaliação de Algoritmos de Reconhecimento de Acordes

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: Geber Lisboa Ramalho

Recife

13 de dezembro de 2011

Agradecimentos

Agradeço a Ilson Barbosa e Eunice Barbosa, meus queridos pais e aos meus irmãos Ilson Barbosa Jr. e Marcelle Barbosa, pelo apoio incondicional e pela fé que depositaram no decorrer da minha graduação e sempre. Agradeço à minha querida amiga e companheira Adryanna Andrade por me apoiar, incentivar e ajudar a estruturar e revisar este trabalho; ao meu amigo Arnóbio Ferreira Jr. por me ajudar na escolha do material para experimentação sem o qual este trabalho não existiria; aos pesquisadores que me ajudaram enviando materiais para estudo e pesquisa, bem como tirando dúvidas: Christopher Harte, Laurent Oudre, Maksim Khadkevich, John Ashley Burgoyne, Mert Bay, Cédric Févotte, Ricardo Vasconselos e Giordano Cabral. Agradeço ao meu orientador Geber Ramalho por me guiar na execução deste trabalho e me indicar os caminhos corretos. Agradeço aos meus amigos da graduação e da universidade pelo apoio e pela partilha de conhecimentos e experiências. E aos amigos do peito, que estão sempre por perto para apoiar e render boas gargalhadas e alegrias.

Resumo

O reconhecimento de acordes é um processo complexo que toma como entrada uma informação de áudio e extrai informações simbólicas que representam os acordes. As aplicações possíveis para o reconhecimento de acordes vão desde o estudo da Harmonia até aplicações como busca de músicas, similaridade musical. Vários algoritmos para essa finalidade têm sido propostos ao longo dos anos e comparar a precisão destes algoritmos tornou-se um problema. Este trabalho propõe uma solução para este problema com a implementação de um sistema de avaliação - benchmark - de algoritmos para reconhecimento de acordes utilizando uma base de dados com variações de timbre baseadas no estilo e formação.

Abstract

The chord recognition task is a complex process that receives as input audio information and extracts from it a symbolic information representing chords. There are several applications for chord recognition like harmony study and musical queries. Several algorithms have been proposed over the years and their evaluation has become a problem. This work propose an implementation of an evaluation system – benchmark – for chord recognition algorithms using a database with timbre variations based on style and formation.

Sumário

[Introdução 1](#_Toc311586906)

[1.1 Objetivos 3](#_Toc311586907)

[1.2 Estrutura 3](#_Toc311586908)

[Acústica e Teoria Musical Básica 4](#_Toc311586909)

[2.1 Som e Timbre 4](#_Toc311586910)

[2.1.1 O Som 4](#_Toc311586911)

[2.1.2 A Onda Senoidal 5](#_Toc311586912)

[2.1.3 A Onda Complexa 6](#_Toc311586913)

[2.1.4 Análise Espectral 7](#_Toc311586914)

[2.1.5 Percepção Sonora 8](#_Toc311586915)

[2.1.6 O Conceito de Timbre 9](#_Toc311586916)

[2.2 Teoria Musical Básica 10](#_Toc311586917)

[2.2.1 Altura 10](#_Toc311586918)

[2.2.2 Escala e Intervalos 11](#_Toc311586919)

[2.2.3 Acordes 12](#_Toc311586920)

[Estado da Arte 15](#_Toc311586921)

[3.1 MIREX 15](#_Toc311586922)

[3.1.1 Padronização da Representação 16](#_Toc311586923)

[3.1.2 Corpus 20](#_Toc311586924)

[3.1.3 Métricas de Avaliação 21](#_Toc311586925)

[Experimentos 23](#_Toc311586926)

[4.1 Diversidade da Base de Dados 23](#_Toc311586927)

[4.2 Proposta 23](#_Toc311586928)

[4.2.1 Formação e o Timbre 24](#_Toc311586929)

[4.2.2 Estilo e o Timbre 24](#_Toc311586930)

[4.3 Procedimento para anotação da base 25](#_Toc311586931)

[Conclusão 27](#_Toc311586932)

[Referências 28](#_Toc311586933)

Lista de Figuras

[Fig. 1 – Representação simplificada das ondas sonoras geradas por diferentes instrumentos. (Representação no domínio do tempo) 15](#_Toc311587460)

[Fig. 2 – Na parte superior da figura podemos ver as contrações e rarefações do ar. Na parte de baixo vemos a representação gráfica dessa onda senoidal. 16](#_Toc311587461)

[Fig. 3 – Ilustração dos parâmetros do som: amplitude e comprimento de onda. 16](#_Toc311587462)

[Fig. 4 – Espectro do sinal sonoro captado de um Cello. O eixo da frente expressa as frequências e o eixo lateral o tempo. 19](#_Toc311587463)

[Fig. 5 – As alturas e as frequências associadas mostradas no teclado de um piano convencional. A parte em cinza não faz parte de um piano comum. 21](#_Toc311587464)

[Fig. 6 – Todos os intervalos da escala de Dó-Maior. 23](#_Toc311587465)

[Fig. 7 – Notações para representação da harmonia: a) Partitura, b) Baixo cifrado barroco, c) Notação clássica por números romanos, d) notação clássica por letras, e) Notação da música popular, f) Notação do *jazz*. 24](#_Toc311587466)

[Fig. 8 – Representação do acorde de Dó menor com sétima na sua primeira inversão. 27](#_Toc311587467)

[Fig. 9 – a) Dois exemplos de representação de acordes utilizando notação abreviada e de mapeamento em lista de intervalos. b) Exemplo de utilização da lista de intervalos extra, e da omissão de intervalo através do símbolo \*. 28](#_Toc311587468)

[Fig. 10 – Lista de abreviações mais comuns separadas por tipos de acordes. Na última coluna encontra-se o mapeamento em forma de lista de intervalos. 29](#_Toc311587469)

[Fig. 11 – Gramática na notação BNF. 29](#_Toc311587470)

[Fig. 12 – Arquivo de anotação de um trecho da música Lucy In The Sky With Diamonds da banda The Beatles, mais precisamente dos primeiros 16 segundos de música . Podemos ver que o padrão tempo\_inicial tempo\_final rótulo é seguido. 30](#_Toc311587471)

[Fig. 13 – A contagem de frames pode ser observada na parte inferior da figura. Cada *frame* acertado pelo algoritmo está marcado com o valor 1; os erros com o valor 0. 32](#_Toc311587472)

Capítulo 1

Introdução

**Nas últimas décadas temos visto um contínuo aumento da quantidade de música digital disponível, principalmente pelo aumento da demanda que está diretamente relacionado com o desenvolvimento de tecnologias ubíquas que permitem aos usuários estarem consumindo e ouvindo música em quaisquer lugares onde estiverem. Outros fatores também estão envolvidos como toda uma indústria e um negócio voltado para a música investindo na produção de novo trabalhos, principalmente nos gêneros pop e rock.** Torna-se, então, necessário uma maneira eficiente de manipular e recuperar essas informações (CHENG, YANG, *et al.*, 2008).

**Para lidar com essa grande massa de informação surgiu, recentemente, uma área de pesquisa conhecida como *Music Information Retrieval* (MIR) cujo principal objetivo é melhorar ainda mais a experiência do usuário na manipulação de conteúdo musical digital. A maioria das abordagens utilizadas em MIR é do tipo *content-based* cuja ideia principal é que um documento possui características que podem ser computadas diretamente do próprio documento** (NICOLA, 2006)**.**

**Uma peça musical pode ser caracterizada segundo uma série de dimensões segundo as quais o usuário pode classifica-la como relevante para si. Para a MIR o problema da multidimensionalidade musical foi chamado de desafio multifacetado (*multifaceted challenge*):** faceta das alturas, faceta temporal, harmônica, *timbral*, editorial e bibliográfica (DOWNIE, 2003). Dentre essas facetas, a dimensão harmônica de uma peça musical pode ser extremamente útil para comparar ou para classificar e *clusterizar* documentos musicais.

**Em música, a sobreposição de duas ou mais alturas no tempo constitui um evento harmônico caracterizado como acorde. As sucessões desses eventos no tempo formam a estrutura harmônica de uma obra musical. Durante séculos, os teóricos da música têm usado várias formas de representação baseadas nas alturas constituintes do evento, das relações intervalares entre as alturas ou da função desses intervalos e alturas dentro da estrutura formal da obra musical** (DOWNIE, 2003)**.**

Para resolver os problemas concernentes à questão harmônica, surgiram as técnicas de reconhecimento ou detecção de acorde e transcrição musical automática. **O reconhecimento de acordes é um processo complexo que toma como entrada uma informação de áudio e extrai informações simbólicas que representam os acordes. As aplicações possíveis para o reconhecimento de acordes vão desde o estudo da Harmonia até aplicações como busca de músicas, e similaridade musical. Vários algoritmos para essa finalidade têm sido propostos ao longo dos anos e comparar a precisão destes algoritmos tornou-se um problema** (LEE e SLANEY, 2006)**.**

Não havia uma forma padronizada de avaliação de resultados, ficando a cargo de cada pesquisador utilizar suas próprias métricas de avaliação. A partir do ano de 2001, quando ocorreu o segundo ISMIR (*International Conference on Music Information Retrieval*), os pesquisadores começaram a concentrar esforços para padronizar as métricas de avaliação das atividades relacionadas a MIR. O ponto de partida foi o paradigma de avaliação e teste desenvolvido pelo *National Institute of Standards and Technology* para a área de Recuperação de Informação Textual chamado TREC (*Textual REtrieval Conference*). Sobre este paradigma, cada equipe de recuperação de informação textual teria acesso a uma grande quantidade de dados para testes, de consultas padronizadas, e avaliações padronizadas dos resultados de cada time. Olhando para este modelo, foi construído um paradigma para recuperação de informações musicais chamado de TREC-like (DOWNIE, 2003).

Nesse contexto surge o MIREX (*Music Information Retrieval Evaluation eXchange*) em 2005, mantido pelo ISMIRSEL (*International Music Information Retrieval Systems Evaluation Laboratory*). É um *framework* baseado na cooperação da comunidade científica para avaliação formal de algoritmos e técnicas relacionadas às atividades relacionadas à MIR (DOWNIE, 2008). Apenas em 2008 o MIREX introduziu a atividade de *Audio Chord Detection* para realizar a avaliação de algoritmos de reconhecimento de acordes utilizando uma base de dados contendo toda a discografia da banda *The Beatles*. Os resultados dos algoritmos são colhidos e comparados segundo métricas de avaliação já bem estabelecidas na área de *Information Retrieval* como *recall* e *precision* .

O problema do MIREX é que ele não utiliza uma base dados variada para aferir o desempenho dos algoritmos submetidos, o inviabiliza ao pesquisador perceber como a sua técnica se comporta sob variados padrões de dados.

* 1. Objetivos

O principal objetivo deste trabalho é construir um *benchmark* para avaliação de algoritmos de reconhecimento de acorde baseado em um banco de dados com uma representativa variação de estilos musicais, bem como variadas formações musicais.

O sistema fará a avaliação de acordo com uma base de dados montada a partir de categorias de timbre separadas por dois parâmetros: formação e estilo. Além disso, o *frame-based recall­* será utilizado como métrica de avalição.

* 1. Estrutura

Este trabalho está dividido da seguinte forma:

* Capítulo 2: este capítulo tem o objetivo de sumarizar os conceitos básicos necessário para um completo entendimento do trabalho e está dividido em dois tópicos principais: o som e a teoria musical.
* Capítulo 3: descreve como a avaliação é realizada atualmente pelo principal meio de avaliação, o MIREX.
* Capítulo 4: possui a proposta deste trabalho, bem como a avaliação dos resultados.
* Capítulo 5: este capítulo encerra o trabalho descrevendo as conclusões proporcionadas pelo mesmo.

Capítulo 2

Acústica e Teoria Musical Básica

* 1. Som e Timbre

Primeiramente precisamos entender alguns princípios e técnicas da acústica para que possamos ter um pleno entendimento dos fatores que podem influênciar nos resultados dos algoritmos.

De uma maneira mais ampla, como definido em (KISNLER e FREY, 1962), a acústica compreende a geração, transmissão e recepção de energia na forma de ondas de vibração. Estas ondas são produzidas quando átomos de fluidos ou sólidos são tiradas da sua posição natural gerando uma força de restauração, que por sua vez implicará em um movimento oscilatório.

Desta área de estudo complexa e ampla, desejamos extrair apenas alguns conceitos e representações úteis para entendermos o timbre.

* + 1. O Som

Seguindo a mesma linha da definição de acústica dada acima, segundo Beranek (BERANEK, 1996) podemos dizer que existe som no momento em que uma perturbação é propagada através de um material elástico ocasionando variações de pressão ou deslocamentos das partículas do material capazes de serem percebidos pelo homem, ou através de algum instrumento de medição ou captação, tal qual um microfone.

Em suma, podemos dizer que através da definição dada acima que o som, na música, é uma perturbação propagada através do ar ocasionada por um instrumento musical, pela voz, ou por síntese sonora computacional, passível de ser captada pelo ouvido humano.

Podemos visualizar melhor na Fig. 1 uma representação gráfica simplificada de diferentes padrões de ondas sonoras gerados a partir de diferentes instrumentos musicais, incluindo a voz.

* + 1. A Onda Senoidal

A mais simples onda sonora é representada através de uma curva senoidal e representa o som mais puro. Na realidade, isto é apenas uma idealização sendo impossível encontrar na natureza uma onda sonora com tamanha pureza. Nos últimos anos com o aperfeiçoamento dos sistemas computacionais digitais, através da síntese sonora, foi possível chegar mais próximo da onda senoidal pura e mesmo esta, sofre de ruídos e limitações físicas dos componentes envolvidos no processo de síntese.

Na figura Fig. 2, vemos, na parte de cima, uma representação esquemática da compressão e rarefação do ar deslocado por um alto-falante, e na parte de baixo vemos a representação gráfica deste acontecimento, onde o eixo horizontal representa uma posição no espaço e o eixo vertical representa a pressão associada àquela posição no espaço.



Fig. – Representação simplificada das ondas sonoras geradas por diferentes instrumentos. (Representação no domínio do tempo)



Fig. – Na parte superior da figura podemos ver as contrações e rarefações do ar. Na parte de baixo vemos a representação gráfica dessa onda senoidal.

O som possui alguns parâmetros que podem ser observados na Fig. 3. A amplitude diz respeito à distância máxima da curva até o nível zero e a frequência está relacionada a quantas vezes o sinal se repete dentro de uma unidade de tempo.



Fig. 3 – Ilustração dos parâmetros do som: amplitude e comprimento de onda.

* + 1. A Onda Complexa

Como foi mencionado na subseção anterior, as ondas sonoras senoidais não existem na natureza, porém possuem crucial importância no entendimento dos sons que percebemos e na área em que este trabalho se insere: na música.

Na natureza as perturbações e vibrações possuem uma infinidade de padrões e formas muito mais complexas do que a forma senoidal. Mas uma forte característica das ondas sonoras musicais é a conservação de certa periodicidade, ou seja, repetição de um mesmo padrão ao longo do tempo.

Ao longo dos séculos XVIII e XIX, desenvolveu-se o que ficou conhecida como a Série de Fourier, através de D’Alambert ao tentar resolver o problema da oscilação das cordas de um violino, e através J. Fourier ao estudar o problema dos fluxos de calor. A ideia por traz da série é que qualquer função periódica complexa pode ser representada como uma soma de funções periódicas mais simples (GENARO, 1997), no caso mais comum e mais interessante para a compreensão dos sons, essa função mais simples é a senoidal, daí porque a importância de estudá-la.

Podemos, então, afirmar que um som complexo periódico é composto por vários sons mais simples, mais precisamente, as senoidais. Cada onda senoidal componente possui determinada frequência e é chamada de harmônico superior sendo a onda de mais baixa frequência chamada de fundamental. Quando as frequências das senóides não são múltiplas da frequência fundamental, chamamos de parciais em vez de harmônicos.

* + 1. Análise Espectral

Como uma extensão da Série de Fourier, que descreve uma função periódica como uma soma de outras funções periódicas mais simples, surge a Transformada de Fourier que leva funções, não apenas as periódicas, para o chamado domínio da frequência.

Quando a Transformada de Fourier é aplicada à determinada função, obtemos uma nova representação onde o domínio da função não é mais o tempo, e sim a frequência (GOMES e VELHO, 1999). Então, é um artifício extremamente importante na analise do sinal sonoro, revelando as frequências que o compõem.

* + 1. Percepção Sonora

É importante estudar como o ouvido humano percebe um som e o diferencia, pois na área de reconhecimento de acordes o que se pretende fazer é uma interpretação do sinal sonoro a fim de identificar padrões correspondentes aos acordes conhecidos na literatura musical que por sua vez é baseada na percepção do som.

O processo se inicia quando a onda sonora, propagada através do ar, atinge o ouvido (RABINER e JUANG, 1993):

1. A onda vibra a membrana basilar no ouvido interno a qual realza uma analise espectral do sinal de entrada.
2. Na saída da membrana basilar acontece uma transdução do do sinal espectral em sinais de atividade no nervo auditivo. Esse ponto se assemelha há extração de características.
3. Por último os sinais neurais são interpretados pelo cérebro.

O que de fato nos importa é que o ouvido humano é sensível ao conteúdo espectral do sinal sonoro. E de uma maneira geral os algoritmos seguem o caminho da análise espectral e extração de características a fim de classificar determinado padrão em uma classe de acordes.

Para finalizar esta seção segue abaixo alguns parâmetros do som do ponto de vista musical relacionados a percepção sonora:

Intensidade: trata-se da característica do som associada a volume e está relacionada com a amplitude.

Altura: trata-se da diferenciação de sons graves e agudos. Está associada à frequência (e.g. Convencionou-se que a nota Lá corresponde a uma frequência de 440Hz).

Timbre: “atributo do sentido auditivo em termos do qual o ouvinte pode julgar que dois sons similarmente apresentados com a mesma intensidade e altura, são dissimilares.” (RISSET e WESSEL, 1999).

* + 1. O Conceito de Timbre

Vimos ao longo deste capítulo, de uma maneira sucinta, alguns conceitos que nos dão condições de definirmos com maior detalhe o timbre.

Como descrito na subseção anterior, o timbre está relacionado com as características sonoras de determinada fonte e é o que possibilita ao ouvido diferenciar, por exemplo, o som de um piano do som de uma flauta.

Podemos dizer que o timbre tem forte correlação com o conteúdo espectral do sinal sonoro, ou seja, pela contribuição das parciais na composição de um complexo sonoro.

Para exemplificar, abaixo na Fig. 4 podemos observar que, na realidade, o sinal sonoro, geralmente, é ainda mais complexo do que o exposto na seção 2.1.3 possuindo um conteúdo harmônico variável no tempo. O eixo mais à frente indica as parciais do som do violoncelo, e o eixo lateral indica como elas variam ao longo do tempo.



Fig. 4 – Espectro do sinal sonoro captado de um Cello. O eixo da frente expressa as frequências e o eixo lateral o tempo.

Além do conteúdo espectral, o timbre é um conceito que agrupa outras dimensões tais quais: o volume do som, o envelope da amplitude – sendo o ataque de especial importância na variação do timbre -, flutuações da altura ocasionadas por vibratos e trêmulos, etc. (LOUREIRO e DE PAULO, 2006).

Podemos concluir desta discussão que o timbre é uma dimensão do universo sonoro que apresenta alto grau de complexidade, principalmente quando diversas fontes – instrumentos, sintetizadores, vozes -, com as mais variadas características harmônicas são agrupadas a fim de produzir música.

Fica claro que as técnicas de reconhecimento de acordes, a partir de um sinal de entrada – uma música -, deve realizar a extração de características que permitam a definição de estruturas conhecidas como acordes, sendo o timbre o fator primordial na taxa de acerto ou de erro de determinada técnica.

* 1. Teoria Musical Básica

Já que estamos no domínio musical quando tratamos de técnicas de reconhecimento de acordes, então devemos ter uma noção básica dos conceitos musicais que estão por trás da transcrição.

* + 1. Altura

Como vimos na seção 2.1.5, a grosso modo, a altura de um som é percebida pelo ouvido como sendo uma variação do grave para o agudo. De uma maneira mais precisa, “Altura é um atributo perceptual que permite a ordenação de sons de acordo com uma escala – relacionada com a frequência – que se estende do baixo ao alto.” (KLAPURI, 2006). A representação de uma altura na notação padrão é feita através de um nome seguido de zero ou mais indicações de alteração – ♭ ou ♯ -, seguido de um número representando a oitava. Podemos entender melhor observando a Fig. 5, que cada altura tem uma frequência relacionada e a relação das alturas pertencentes a uma mesma classe – e.g. A0, A1, A2, etc. – é uma relação de oitava, ou seja, A1 possui o dobro da frequência de A0; A2 possui o dobro da frequência de A3.

O que, de fato, nos interessa é que o ouvido é capaz de identificar que duas alturas diferentes estão em relação de oitavas (SHEPARD, 1964). Portanto, dois acordes que possuem a mesma combinação de classes de altura são considerados equivalentes – exceto em alguns casos no baixo, onde as inversões devem ser consideradas. Esta última observação ficará mais clara na 2.2.3.

O uso do termo nota, altura e classe de altura serão usados a seguir de maneira intercambiável.

* + 1. Escala e Intervalos

Uma escala musical é composta por uma sucessão de alturas. A relação entre essas alturas determina os chamados intervalos. A escala cromática, por exemplo, é formada por 12 classes de altura sendo elas Dó, Ré, Mi, Fá, Sol, La, Si e mais os acidentes Dó♯ / Ré♭, Ré♯ / Mi♭, Fá♯/ Sol♭, Sol♯/ La♭, La ♯/ Si♭. Na música ocidental, essas alturas formam todas as possibilidades de sons possíveis dentro de uma oitava e a menor distância entre duas dessas alturas é chamada de semitom; dois semitons formam um tom.

Em uma determinada escala a ordem em que determinada altura se encontra na progressão indica o seu grau. Por exemplo, na escala de Dó-Maior – Dó, Ré, Mi, Fá, Sol, Lá e Si – a nota Dó é o primeiro grau da escala denotado pelo algarismo romano I, Fá é o IV grau da escala e assim por diante.

Um intervalo é a distância em graus entre duas alturas – dentro de determinada escala - tomando como referência uma delas. Para exemplificar, tomando Ré como referência na escala de Dó-Maior, Ré e Fá constituem um intervalo de segunda; Ré e Si constituem um intervalo de sexta; e assim por diante. Na Fig. 6 podemos ver todos os intervalos de Dó-Maior de uma maneira mais detalhada.



Fig. – As alturas e as frequências associadas mostradas no teclado de um piano convencional. A parte em cinza não faz parte de um piano comum.

* + 1. Acordes

Segundo o *Oxford Dictionary of Music* (KENNEDY e KENNEDY, 1994), um acorde pode ser caracterizado pela combinação simultânea de quaisquer notas, normalmente não menos do que três. A partir da sucessão de acordes é que se constrói a estrutura harmônica de uma peça musical.

Podemos combinar alturas de diversas formas produzindo assim diversas famílias de acordes. Dois tipos comuns de acordes são os acordes maiores e os acordes menores. O primeiro tipo é formado pela nota fundamental mais o quinto e terceiro grau da escala maior enquanto o acorde menor é formado pela fundamental mais o quinto e terceiro grau da escala menor. Outros dois tipos comuns de acordes são os aumentados e os diminutos, sendo o primeiro formado por duas terças maiores e o segundo por duas terças menores.

Em suma, na música ocidental, um acorde é composto pela nota fundamental sobre o qual o acorde é composto; pelos graus, relativos à fundamental, que compõem o acorde; e por uma possível inversão (HARTE, SANDLER, *et al.*, 2005). A inversão é definida pelo grau que está sobre o baixo – a nota mais grave. Por exemplo, se temos a tríade composta por Dó, Mi, Sol, sua primeira inversão é a tríade Mi, Sol, Dó, onde o dó foi elevado uma oitava acima.

Um dos componentes estruturais principais de uma peça musical é a harmonia que, por sua vez é composta por sucessões de acordes. Nas música tonal ocidental, existe uma série de regras e procedimentos para organizar os acordes a fim de conceber a estrutura formal harmônica de uma pela musical. A transcrição musical, comumente se refere ao ato de representar por escrito o som que se houve. No contexto em que este trabalho se insere, a transcrição é feita de maneira automática através de um algoritmo que processa o som de entrada e gera uma transcrição.

Por fim, devemos fazer uma análise das formas de representação do acordes. Ao longo dos séculos, músicos e teóricos têm empregado formas distintas de representar os acordes dentre as quais as mais relevantes são: baixo cifrado (Fig. 7(b)), a clássica através de números romanos (Fig. 7(c)), a clássica através de letras (Fig. 7(d)), a forma popular e a notação do *jazz* (Fig. 7(d)) (HARTE, SANDLER, *et al.*, 2005). A representação de um acorde depende do tipo dele, ou da família a qual ele pertence, sendo a tríade do tipo maior representada por *M* na representação popular e a tríade menor representada por *m*, também na notação popular.



Fig. – Todos os intervalos da escala de Dó-Maior.



Fig. – Notações para representação da harmonia: a) Partitura, b) Baixo cifrado barroco, c) Notação clássica por números romanos, d) notação clássica por letras, e) Notação da música popular, f) Notação do *jazz*.

Capítulo 3

Estado da Arte

* 1. MIREX

Até o ano de 2004, todo processo de avaliação dos algoritmos ficava a cargo do pesquisador, com cada grupo de pesquisa possuindo seu próprio conjunto de dados e *queries*. As métricas de avaliações usadas por cada um também divergiam entre si.

Os primeiros esforços para consolidar um *framework* com o intuito de unificar as avaliações foram desempenhados, ainda em 2004, pelo *Music Technology Group of the Audiovisual Institute* da Universidade Pompeu Fabra, em Barcelona. Era uma competição chamada *Audio Descriptor Contest*, separada em seis categorias diferentes: classificação de gênero, identificação de artista, similaridade entre artistas, classificação do ritmo, indução do *tempo* e extração de melodia. O procedimento adotado, que ainda hoje é utilizado na área de avaliação em MIR, foi a submissão dos algoritmos em vez da submissão dos resultados, ficando a cargo dos organizadores todo o processo de compilar, rodar, avaliar resultados..

Mas foi através das iniciativas do Dr. J. Stephen Downie, que fundou o projeto ISMIRSEL, que pode ser criado um *framework* para avaliação de larga escala. A ideia era criar um ambiente seguro e acessível para que os grupos de pesquisa pudessem ter acesso às coleções de músicas e poderem avaliar os resultados de suas pesquisas. A escolha e composição da coleção e as métricas de avaliação foram definidos de uma maneira democrática e todos os possíveis participantes opinaram sobre todos os aspectos.

A primeira campanha surgida do IRMIRSEL foi realizada em 2005 e apresentada no mesmo ano no ISMIR sob o nome de MIREX, e foi dividida em nove tarefas distintas: classificação de gênero, identificação de artista, detecção de bateria, detecção de *onsets*, extração de *tempo*, extração de melodia e três outras tarefas associadas a processamento de informação simbólica em vez de processamento de áudio (NICOLA, 2006).

Em 2008, surge como uma nova tarefa no MIREX o reconhecimento de acorde sob o nome de *Audio Chord Detection* e a partir de 2010, passou a se chamar de *Audio Chord Estimation*. No ano em que esta tarefa surgiu, houve uma intensa discussão sobre todos os tópicos concernentes à tarefa que podem ser consultadas na wiki do MIREX (2008).

Três importantes tópicos merecem atenção: padronização da representação, o *corpus*, e as métricas de avaliação.

* + 1. Padronização da Representação

Para poder compatibilizar a avaliação de diversos algoritmos é necessário que tanto a entrada quanto a saída sejam padronizadas. No caso de algoritmos de reconhecimento de acordes, as entradas são arquivos de áudio os quais serão processados pelo algoritmo que por sua vez gerará uma saída contendo a estrutura harmônica da música. Na seção 3.1.2 será explicado como é composta a entrada, ou seja, a base de dados. Nesta seção trataremos da representação padrão, desenvolvida por Christopher Harte (HARTE, 2010), dos acordes que os algoritmos submetidos ao MIREX devem seguir ao fazerem a transcrição de um arquivo de áudio.

Vimos na seção 2.2.3 que existem diversas notações que foram desenvolvidas ao longo dos séculos para representar a estrutura harmônica de uma peça musical. Precisamos, então, definir qual representação é mais adequada para o contexto do MIREX dentre as possibilidades existentes listadas naquela seção. Na realidade, as formas tradicionais não podem ser utilizadas sem algumas modificações necessárias, pois apresentam alguns problemas que podem atrapalhar a exatidão de todo o *framework*.

O primeiro problema encontrado está relacionado com a ambiguidade da sintaxe: a representação A♭7 quer dizer A♭~ 7 (lá bemol com sétima) ou A ~ ♭7 (tríade maior de Lá mais a sétima “bemolizada”). Outro cuidado que devemos ter é em relação à utilização de letras minúsculas e maiúsculas: algumas representações utilizam letras minúsculas para representar acordes do tipo menor e letras maiúsculas para representar acordes do tipo maior. Por exemplo, uma tríade maior de Sí seria representada pela letra B, e a tríade menor de Sí seria representada pela letra b. Porém, se nesta mesma notação a letra b é utilizada para representar a alteração bemol (♭) poderá haver confusão. No caso da harmonia clássica, temos o problema do contexto: o símbolo C7, dependendo da tonalidade, pode representar acordes diferentes. Na tonalidade de Fá-Maior pode ser composto pelas notas C, E, G e B♭, enquanto que na tonalidade de Dó-Maior, pode ser composto pelas notas C, E, G e B.

Considerando todo esse contexto, Christopher Harte (HARTE, 2010), desenvolveu e formalizou uma notação livre de ambiguidade, própria para facilitar o processamento do texto e também fácil e intuitivo no contexto musical. Além disso, é uma notação flexível – possibilitando a representação de toda sorte de acordes - e livre de contexto, ou seja, cada símbolo de acorde representa um acorde de maneira única.



Fig. – Representação do acorde de Dó menor com sétima na sua primeira inversão.

Podemos ver na Fig. 8 o modelo lógico dessa representação. Um acorde é formado por uma fundamental de A-G, por uma lista de intervalos e uma possível inversão caracterizada pelo *baixo*. Lembrando que a fundamental, a lista de intervalos e o baixo podem sofrer alterações – bemol ou sustenido. Podemos derivar dessa lógica uma notação em texto simples para representar qualquer acorde:

fundamental : (intervalo1, intervalo2,...) / baixo

As alterações são representadas pelos símbolos b e # e quando são utilizadas do lado direito quando se trata de um natural – um símbolo dentre A-G – e do lado esquerdo quando se trata de um intervalo. Por exemplo: Cb : (1, b3, #5). Esta notação já é capaz de descrever qualquer tipo de acorde, mas para que ela seja facilmente lida por humanos é necessário introduzir a notação abreviada (*shorthand notation*) que nada mais é do que um conjunto de mnemônicos que podem ser mapeados em uma lista de intervalos. Então podemos escrever um acorde de outra forma:

fundamental : abreviação (intervalos extra) / baixo

A parte (intervalos extra) permite que seja adicionada uma lista de intervalos que se deseje acrescentar ao acorde representado pelo mnemônico. Mais ainda, pode especificar uma lista de intervalos os quais desejamos remover do acorde através do símbolo \*. Na Fig. 9 abaixo podemos ver alguns exemplos de como um acorde pode ser mapeado em uma lista de intervalos e alguns exemplos da utilização da lista de intervalos extra.



Fig. 9 – a) Dois exemplos de representação de acordes utilizando notação abreviada e de mapeamento em lista de intervalos. b) Exemplo de utilização da lista de intervalos extra, e da omissão de intervalo através do símbolo \*.

Para finalizar, abaixo, na fig. 9 e fig. 10, encontram-se, respectivamente, um quadro contento as abreviações mais comuns e um quadro com a gramática completa, na forma BNF (*Backus-Naus Form*), que representa a notação desenvolvida (HARTE, 2010).



Fig. – Lista de abreviações mais comuns separadas por tipos de acordes. Na última coluna encontra-se o mapeamento em forma de lista de intervalos.



Fig. – Gramática na notação BNF.

* + 1. Corpus

Desde a edição de 2008 do MIREX, a coleção de músicas utilizadas para treino e avaliação dos algoritmos é a discografia completa da banda The Beatles composta por 13 álbuns contendo, ao todo, 180 faixas gravadas com qualidade de CD – 16 bits a 44.1KHz – totalizando pouco mais de 8 horas de música, organizada por Christopher Harte (2008). Ao longo das edições que se seguiram outras coleções foram agregadas ao *corpus* do MIREX como a coleção da banda Queen e Zweiech de Matthia Mauch (MAUCH, 2010).

Cada música desta base de dados possui uma anotação chamada *ground truth* que indica com a maior fidelidade possível quais são os acordes corretos para toda a extensão da música. É com estes arquivos – que possuem a extensão .lab - de anotação que as respostas dos algoritmos podem ser comparadas e avaliadas. Os arquivos .lab possuem o seguinte formato:

tempo\_inicial tempo\_final rótulo

O tempo\_inicial indica o tempo em que o acorde de nome rótulo começou a soar e o tempo\_final indica o tempo em que o acorde cessou. Abaixo na fig. 12, temos o exemplo de um trecho anotado de uma das músicas dos Beatles presente na base.



Fig. – Arquivo de anotação de um trecho da música Lucy In The Sky With Diamonds da banda The Beatles, mais precisamente dos primeiros 16 segundos de música . Podemos ver que o padrão tempo\_inicial tempo\_final rótulo é seguido.

* + 1. Métricas de Avaliação

Por último, e igualmente essencial, é necessário estabelecer métricas para comparar as anotações com os resultados obtidos dos algoritmos avaliados.

Existem duas medidas bastante comuns na área da IR (*Information Retrieval)* chamadas de *precision* e *recall*. Foram definidas no contexto tradicional da IR que é extração de informações em documentos textuais e são definidas como se segue (MANNING, RAGHAVAN e SCHÜTZE, 2008):

$$precision=\frac{Número de itens relevantes recuperados}{Número de itens recuperados}$$

$$recall=\frac{Número de itens relevantes recuperados}{Número de itens relevantes}$$

No contexto do reconhecimento de acordes, o *chord symbol recall* é referido muitas vezes como *chord average overlap* ou *relative correct overlap (RCO)* (MAUCH, 2010) e pode ser definido de uma maneira similar. O MIREX, para calcular o RCO, convencionou utilizar o chamado *frame-based recall* (2011) que consiste em, baseado nos arquivos .lab, fragmentar toda a música em *frames* de igual duração e contar quantos *frames* foram acertados. Para exemplificar a fragmentação, se temos uma arquivos .lab com a seguinte configuração

0.000000 4.000000 A

4.000000 6.000000 C

e desejamos fragmentá-lo em *frames* com 1 segundo de duração, teremos 6 *frames*: A, A, A, A, C, C. A partir daí calculamos o *recall* comparando os frames estimados pelo algoritmo com os frames gerados a partir do *ground truth* [20]:

$$recall=\frac{frames acertados}{total de frames} $$

Abaixo, na fig. 13 ,encontra-se uma ilustração da contagem dos frames. No caso do MIREX utiliza-se um *frame* de 10ms.



Fig. – A contagem de frames pode ser observada na parte inferior da figura. Cada *frame* acertado pelo algoritmo está marcado com o valor 1; os erros com o valor 0.

No MIREX, a avaliação é feita levando em conta um dicionário de acordes separados por dídades – apenas duas notas juntas -, tríades e tétrades – quatro notas juntas. O dicionário é guardado em um arquivo de texto a partir do qual outro arquivo .lab é criado substituindo-se os acordes que não encontram-se no dicionário pelo valor “X”.

Caso esteja-se usando um dicionário de tríades, apenas os três primeiros intervalos serão considerados e, por exemplo, major e major7 serão considerados iguais, pois major possui os intervalos (1, 3, 5) e major7 possui os intervalos (1, 3, 5, 7). Já numa análise a partir de um dicionário de tétrades estes acordes serão considerados diferentes. Apenas após essa filtragem dos acordes, é que o *frame-based recall­* é utilizado (2011).

Capítulo 4

Experimentos

* 1. Diversidade da Base de Dados

Como vimos, o MIREX é uma avaliação anual de algoritmos da área de *Musical Information Retrieval*, incluindo a avaliação de algoritmos de reconhecimento de acordes sob o título de *Chord Detection*.

O MIREX utiliza, no processo de avaliação dos algoritmos de reconhecimento de acordes, uma base de dados composta pela discografia completa da banda The Beatles composta por Christopher Harte (Verificar Histórico.) e mais recentemente, a partir de 2010, um conjunto de músicas da banda Queen e Zweieck montado por Matthias Mauch.

O problema que reside neste procedimento é que se avaliarmos o desempenho dos algoritmos sob um conjunto restrito de variações de timbre fica difícil apontar as falhas e possíveis caminhos para melhora dos algoritmos avaliados. Como vimos, o timbre é uma característica importantíssima e determinante no desempenho na tarefa de identificar acordes dentro de uma estrutura sonora complexa.

* 1. Proposta

Uma maneira de tornar a avaliação mais significativa é testar os algoritmos em cima de uma base de dados o mais representativa possível sendo o que define a representatividade é a diversidade de padrões de dados.

No caso particular dos algoritmos de reconhecimento de acordes a variação é conseguida quanto maior for a diversidade de conteúdo musical. E como vimos, o sinal sonoro musical possui uma alta complexidade e quantidade de informação que atribuímos à qualidade chamada timbre. Podemos, então, propor que o nosso benchmark possua uma base de dados contendo músicas com os mais variados timbres para poder realizar uma avaliação representativa e precisa. Porém, agrupar instâncias de uma base de dados através do timbre é uma tarefa complicada, como visto seção 2.1.6, timbre é um conceito abstrato e multidimensional do ponto de vista físico e é um dos desafios da psicoacústica (LOUREIRO e DE PAULO, 2006).

Por esse motivo, para realizar o agrupamento das instâncias fragmentamos o conceito de timbre em duas outras qualidades: a formação e o estilo. Nas seções que se seguem, entenderemos melhor como essas medidas são úteis para medir de uma maneira qualitativa a variação de timbre.

* + 1. Formação e o Timbre

A formação diz respeito à quais elementos estão presentes no som, ou seja, quais elementos foram utilizados para produzir determinada música. E temos uma infinidade de possíveis elementos que podem ser utilizados para os fins musicais como instrumentos tradicionais - cello, piano, violão, dentre outros -, voz, sintetizadores, e quaisquer elementos que se deseje utilizar para produzir som.

O fato de o timbre estar fortemente relacionado ao conteúdo espectral e o este último estar fortemente relacionado à fonte sonora, faz da formação um meio significativo do ponto de vista qualitativo para a classificação de diferentes instâncias. Então podemos supor que uma banda de Jazz cuja formação seja piano, baixo e bateria tem o timbre mais próximo de outra banda que possui a mesma formação do que de uma banda de Jazz que é formada por uma guitarra, trompete e baixo.

* + 1. Estilo e o Timbre

Outro aspecto qualitativo escolhido para a seleção e agrupamento de instâncias foi o estilo musical, pois além da relação com o conteúdo espectral que o timbre apresenta, ele possui outras características tal qual o envelope de amplitude, por exemplo. A forma com a qual um instrumento é tocado também influencia no timbre do sinal observado e podemos verificar isso comparando as formas de tocar de João Gilberto e Bob Dylan. Ambos possuem gravações com a mesma formação – voz e violão – porém a forma de ataque e padrões rítmicos são diferentes, além das vozes que também diferem entre si, sendo uma mais limpa e a outra mais distorcida. Essas características se refletem no estilo musical com Gilberto representando a Bossa Nova e Dylan representando a *Folk Music*.

* 1. Procedimento para anotação da base

A base de dados foi montada levando em conta as considerações acima e 15 músicas foram escolhidas baseadas na variação de estilo e de formação. Estas músicas forma convertidas do formato MP3 para Wave em 44.1KHz.

Para realizar a transcrição das músicas dos Beatles, Harte (HARTE, 2010) criou um processo que consiste em quatro etapas:

1. Familiarização;
2. *Aural transcription*;
3. Definição dos limites dos acordes;
4. Rotulação do segmento.

Na primeira fase, o responsável pela transcrição deve ouvir a música até se acostumar com a estrutura harmônica. Na segunda fase, cada acorde da música é identificado e escrito. Na terceira fase, os limites de cada acorde são definidos, ou seja, nesta etapa devemos prestar atenção apenas à mudança de acorde, registrando os momentos em que isto ocorre e criando os segmentos que serão rotulados na fase 4. Nesta última fase, os acordes que foram identificados na fase dois, são atribuídos aos segmentos criados na fase 3.

Seguindo um procedimento parecido a escrita do arquivo de anotação – *ground truth* - .lab de cada uma das músicas utilizas neste trabalho, foram utilizadas partituras através do *software* Guitar Pro 6 que possui um dicionário de acordes. Ou seja, para as finalidades deste projeto, a etapa 1 não foi necessária e a etapa 2 foi realizada através das partituras e não “de ouvido”, restando apenas as etapas 3 e 4 para serem feitas. Para realizar a etapa 3, foi criado um programa em Matlab que, ao apertar a tecla “enter”, marca o tempo já utilizando o formato de um arquivo .lab escrevendo “labeln” para a n-ésima marcação. Por exemplo:

0.000000 0.232059 label1

0.232059 1.529933 label2

1.529933 4.000000 label3

...

Ao final desta etapa, temos um arquivo .lab com todo preenchido especificando os limites de cada acorde. Então, na etapa 4, varremos o arquivo substituindo todas as n *labels* pelos acordes listados na etapa 2 concluindo a transcrição.

Capítulo 5

Conclusão

Para o a realização da transcrição musical vários algoritmos com técnicas diferentes tem sido empregados. Para acompanhar esta evolução é necessário construir uma ambiente para manter e avaliar estes algoritmos.

Para isto surgiram vários ambientes e *frameworks* para esta finalidade culminando no mais famoso *framework*: o MIREX. Porém este *benchmark* utiliza uma base de dados com apenas 3 artistas diferentes limitando, de certa forma, a variação de timbres.

Fazendo uma análise utilizando uma base de dados baseada na variação de timbres foi possível perceber que este é um fator importante e que deve ser considerado na montagem da base de dados de qualquer *benchmark*. Esta variação permite ao pesquisador visualizar em quais casos o seu algoritmo é mais efetivo provendo informações importantes para a evolução de suas técnicas.

Referências

**MIREX**, 2008. Disponivel em: <http://www.music-ir.org/mirex/wiki/2008:Audio\_Chord\_Detection>. Acesso em: 10 dez. 2011.

**MIREX**, 2011. Disponivel em: <http://www.music-ir.org/mirex/wiki/2011:Audio\_Chord\_Estimation >. Acesso em: 11 dez. 2011.

BERANEK, L. L. **Acoustics**. Woodburry: American Instute of Physics, 1996.

CHENG, H.-T. et al. **Automatic chord recognition for music classification and retrieval**. Multimedia and Expo, 2008 IEEE International Conference on. Hannover: [s.n.]. 2008. p. 1505 - 1508.

DOWNIE, J. S. Music Information Retrieval. **Annual Review of Information Science and Technology**, v. 37, n. 1, p. 295-340, 2003.

DOWNIE, J. S. **The TREC-like evaluation of music IR systems**. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. New York: ACM. 2003.

DOWNIE, J. S. The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. **Acoustical Science and Technology**, 29, 2008. 247-255.

GENARO, G. Series de Fourier, Transformadas de Fourier y Aplicaciones. **Divulgaciones Matemáticas**, Maracaibo, v. 5, p. 43-60, 1997. ISSN 1315-2068.

GOMES, J.; VELHO, L. **From Fourier Analysis to Wavelets**. Rio de Janeiro: [s.n.], 1999. Notas de Curso do Instituto de Matemática Pura e Aplicada (IMPA).

HARTE, C. **Towards automatic extraction of harmony information from music signals**. [S.l.]: [s.n.], 2010. Disponivel em: <https://qmro.qmul.ac.uk/jspui/handle/123456789/534>. Tese de PhD.

HARTE, et al. **Symbolic representation of musical chords:** a proposed syntax for text annotations. Proceedings of 6th International Conference on Music Information Retrieval. Londres: [s.n.]. 2005.

KENNEDY, M.; KENNEDY, J. B. **The Oxford dictionary of music**. New York: Oxford University Press, 1994.

KISNLER, L. E.; FREY, A. R. **Fundamentals of Acoustics**. 2ª. ed. Monterey: John Wiley & Sons, 1962.

KLAPURI, A. Introduction to music transcription. In: KLAPURI, A.; DAVY, M. **Signal Processing Methods for Music Transcription**. New York: Springer, 2006. Cap. 1.

LEE, K.; SLANEY, M. **Automatic chord recognition from audio using an HMM with supervised learning**. Proceedings of the 1st ACM workshop on Audio and music computing multimedia. [S.l.]: citeseer. 2006. p. 2-6.

LOUREIRO, M. A.; DE PAULO, H. B. Timbre de um instrumento musical: caracterização e representação. **Per Musi**, Belo Horizonte, n. 14, p. 57-81, 2006. ISSN 1517-7599.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. 1ª. ed. Cambridge: Cambridge University Press, 2008.

MAUCH, M. **Automatic chord transcription from audio using computational models of musical context**. [S.l.]: [s.n.], 2010. Disponivel em: <http://qmro.qmul.ac.uk/jspui/handle/123456789/451>. Tese de PhD.

NICOLA, O. Music Retrieval: A Tutorial and Review. **Foundations and Trends® in Information Retrieval**, v. 1, n. 1, p. 1-90, 2006.

RABINER, L.; JUANG, B.-H. **Fundamentals of Speech Recognition**. Englewood Cliffs: PTR Prentice-Hall, 1993.

RISSET, J.-C.; WESSEL, D. Exploration of Timbre by Analysis and Synthesis. In: DEUTSCH, D. **Psychology of Music**. 2ª. ed. San Diego: Academic Press, 1999. p. 25-28.

SHEPARD, R. N. Circularity in Judgments of Relative Pitch. **Journal of the Acoustical Society of America**, 36, n. 12, 23 Julho 1964. 2346-2353.