Pedro Rodrigues Nacione Pedruzzi
Ricardo A. Redder Jr.

Reconhecimento e Busca Adaptativos de Padrões Musicais

São Paulo – SP

Dezembro / 2008

Pedro Rodrigues Nacione Pedruzzi
Ricardo A. Redder Jr.

Reconhecimento e Busca Adaptativos de Padrões Musicais

Dissertação apresentada à Comissão de Graduação em Engenharia da Computação da Escola Politécnica da Universidade de São Paulo para a obtenção da graduação no curso de Engenharia da Computação.

Orientador:
Prof. Doutor João José Neto

Graduação em Engenharia da Computação
Departamento de Sistemas e Sinais
Escola Politécnica da Universidade de São Paulo

São Paulo – SP

Dezembro / 2008

"A música escondida não tem valor."

Aulo Gélio

Resumo

Abstract

Sumário

Lista de Figuras

Lista de Tabelas

Introdução

1 Motivação

Os sistemas de busca atuais evoluíram rapidamente desde suas primeiras versões, e hoje se tornaram parte fundamental do dia-a-dia de grande parte da população. Em alguns casos chega a ser difícil se imaginar navegando na internet sem utilizar algum sistema de busca. Pode-se citar como exemplo destes sistemas: Google.com ™, Yahoo.com ™ ou Live Search ™.

Porém tais sistemas de busca em geral baseiam-se sobre os mesmos princípios e métodos de busca, em sua maioria se aplicando a documentos de texto. Mesmo alguns sistemas que, por exemplo, efetuam buscas por imagens continuam em sua essência baseando-se nos mesmos princípios, já que se apóiam na categorização textual destas imagens.

A popularização destes sistemas de busca mostrou as possibilidades de expansão e a importância que tais sistemas podem adquirir no cotidiano das pessoas. Além disso, suas limitações de contexto (restrição a textos apenas) imediatamente levantam a necessidade de novas técnicas com o fim de ampliar os domínios de aplicação destes sistemas.

Um dos domínios de extrema importância com relação a tal tema é o domínio de áudio, devido a fatores como sua popularidade no meio virtual, quantidade de conteúdo disponível, facilidade de produção de conteúdo, etc. Músicas em formatos digitais, tais como MP3, MIDI, WAV, etc, podem ser facilmente encontradas na internet e já se tornaram parte do cotidiano de grande parte da população. Além disso, a capacidade de reproduzir tais formatos torna-se cada vez mais um padrão nos aparelhos eletrônicos de áudio. Paralelamente aos conteúdos de áudio digitais, podem-se citar ainda conteúdos de áudio-visual, que recentemente adquiriram grande popularidade e, analogamente aos conteúdos de áudio, estão cada vez mais se tornando parte importante do ambiente virtual.

Apesar desta popularidade e importância de conteúdos multimídia, há uma deficiência em métodos de busca para lidar com tais conteúdos. Os métodos de busca tradicionais seriam capazes apenas de encontrar trechos idênticos em arquivos multimídia. Fato que

inviabiliza a criação de sistemas capazes de lidar com incertezas ou variações nos dados de entrada das buscas, analogamente ao que sistemas de busca textuais realizam atualmente.

Ambos os domínios, áudio e áudio-visual, poderiam ser beneficiados por um sistema de buscas sobre áudio capaz de lidar com incertezas e variações nos dados de entrada. Um caso típico, e extremamente frequente, é o caso onde um indivíduo é capaz de reproduzir apenas um trecho de uma música e deseja encontrar a mesma. A reprodução deste trecho pelo indivíduo está inerentemente sujeita a variações de diversas ordens com relação à música original. O indivíduo pode reproduzir a música com divergências na frequência, ou no jargão musical, fora do tom, pode divergir com relação ao tempo, ou pode ainda cometer erros com relação às próprias notas musicas, por exemplo, esquecendo uma nota ou inserindo uma nota inexistente. Este caso pode manifestar-se a partir de diversas situações do cotidiano, como ao tentar encontrar uma música ouvida em um filme, um programa de TV, uma rádio, etc.

Este tipo de problema é extremamente difícil para um indivíduo comum devido à falta de ferramentas que possam auxiliar tal busca, já que a grande quantidade de músicas existentes torna inviável uma varredura completa dos repositórios existentes, ou seja, ouvir todas as músicas, uma a uma. Além disso, as formas mais comuns de organização de repositórios musicais recaem sobre estilos musicais e nomes, o que não é o suficiente para endereçar, ou ajudar no problema apresentado, já que, em geral, tais informações não são suficientes para encontrar a música buscada.

Atualmente, a forma mais comum de tentar encontrar uma solução para tal problema é utilizando-se a ajuda de um especialista, que em alguns casos, é capaz de identificar o trecho reproduzido. O problema imediato com tal abordagem é o fato da mesma não ser escalável e de difícil acesso. Além de enfrentar limitações impostas pela própria natureza do especialista, como dificuldade de lidar com grandes quantidades de músicas, diversidades de estilo, etc.

Estes fatos demonstram a demanda por uma ferramenta que possa auxiliar tal tipo de busca, ou seja, uma ferramenta que seja capaz de identificar uma música, ou conteúdo de áudio, que o usuário esteja buscando, utilizando para isso apenas um trecho reproduzido pelo próprio usuário. Tal ferramenta seria análoga aos sistemas de busca correntes, porém aplicado ao domínio de áudio.

2 Objetivo

Através deste projeto deseja-se abordar uma questão extremamente ampla, que é a busca sobre conteúdos de áudio, porém, por ser um tema vasto, uma redução de escopo para os fins deste projeto se faz necessária. Assim decidiu-se por focar-se nos pontos de maior relevância para o problema.

Um dos pontos que estabelece a maior barreira para a construção de um sistema como o vislumbrado é a dificuldade de comparação entre dois conteúdos de áudio. Esta comparação não é bem definida, assim não há uma forma consistente de se estabelecer o conceito de distância entre dois conteúdos de áudio, conceito que por sua vez é fundamental para a utilização dos modelos de busca e indexação existentes.

Assim, pretende-se com este trabalho promover um avanço sobre tal questão de comparação de conteúdos de áudio, e para tanto, serão adotadas técnicas pouco exploradas em tal domínio. O projeto se apoiará sobre dois pilares importantes: a utilização de conceitos do domínio musical e o uso de técnicas adaptativas, ambos aplicados à comparação de conteúdos de áudio.

Diversas técnicas têm sido utilizadas para análise e comparação de sinais de áudio, porém, em sua maioria, tais métodos recaem sobre princípios de processamento de sinais, por serem genéricos e possuírem grande aplicabilidade, tendendo a ignorar conceitos específicos do domínio musical, ou seja, conceitos de notas, tempos, etc. Entretanto, tais princípios podem ser de grande relevância quando se deseja efetuar uma busca por músicas. Assim, com este trabalho, aspira-se a utilização de tais conceitos dentro do contexto de comparação de conteúdos de áudio, com o fim de obter melhores resultados.

Paralelamente a isto, deseja-se utilizar técnicas adaptativas no esforço de se obter um método de comparação entre trechos de áudio. A idéia é criar um autômato adaptativo baseado em um trecho de áudio, e assim, um segundo trecho de áudio que submetido a este autômato geraria uma saída que conteria uma indicação da distância entre os dois trechos. Esta distância, reconhecida pelo autômato, entre dois trechos de áudio poderia

servir como base de entrada para outros algoritmos tradicionais de ranking. Da junção destes conceitos e ferramentas espera-se criar um sistema capaz de receber um trecho de áudio de um usuário e identificar dentro de um repositório limitado qual música se corresponde ao trecho reproduzido pelo usuário.

3 Histórico

Nesta seção serão apresentados os históricos dos principais temas relacionados ao projeto desenvolvido. Mostrando sucintamente a história que se desenrolou paralelamente dos sistemas de busca e dos avanços dos métodos de manipulação de conteúdos de áudio através dos computadores.

3.1 Histórico dos sistemas de busca

Métodos de busca baseados em texto já são antigos e utilizados há um longo tempo, entre os métodos mais simples pode-se citar o uso índices remissivos. Apesar de simples, este método é extremamente útil e eficiente quando se deseja buscar por uma palavra dentro de um conjunto de documentos. Além deste método simples, existem ainda outras formas de se indexar um documento e efetuar uma busca sobre o mesmo. Porém a aplicação manual destes métodos sempre apresentou dificuldades, pelas dificuldades de indexação de palavras, lentidão de busca, etc.

Como advento dos computadores, tais método passaram a ser implantados por computadores, o que obviamente aumentou sua capacidade, e facilidade de uso. Assim nasceram os primeiros sistemas de busca, juntamente com os computadores. Porém estes métodos eram em sua essência muito simplistas, considerando em geral apenas buscas por trechos exatos.

A idéia dos sistemas de busca da forma como conhecemos hoje surgiu algum tempo depois, já por volta da década de 60, e foi se aprimorando ao longo dos anos. Conceitos como modelo de espaço vetorial, frequência inversa no documento (IDF), frequência do termo (TF), discriminação de termos, relevância e feedback começaram a ser galgados nesta época. Tais técnicas evoluíram muito ao longo dos anos, provendo ferramentas extremamente importantes para efetuar indexação e buscas sobre conjuntos extensos de documentos.

Conforme o tamanho do espaço de busca cresce, maior importância tais técnicas assumem, assim, com o advento da internet, estes métodos adquiriram um papel especial no mundo da computação. Isso porque a internet abriu a possibilidade de se criar espaços de buscas muito maiores do que até então construídos, já que a superfície de busca poderia ser virtualmente todo documento disponível na rede. Por volta da década de 90 começam então a surgir os sistemas web de indexação e busca, o primeiro sistema deste tipo foi o Archie, porem efetuava buscas apenas sobre nomes de arquivos, e não sobre seus conteúdos. Pouco tempo depois surgiram os primeiros crawlers, componente dos sistemas de busca que se tornou indispensável aos sistemas atuais.

Desde seu início até o presente a ciência de Recuperação de Informação (Information Retrieval) evolui muito, e diversos métodos de busca, além de variações, foram criados ao longo destes anos, e hoje se podem citar dois modelos que assumiram importância fundamental nesta ciência o modelo espaço vetorial, e o modelo probabilístico.

Tais modelos e técnicas dão o tom do estudo desta ciência no mundo acadêmico, porém apenas tais conceitos não são o suficiente para se construir um sistema de busca similar aos que encontramos atualmente. Além destes princípios, considerações diversas relacionadas a desempenho, propriedade intelectual, conteúdo impróprio, conteúdo falso, tentativas de manipulação de resultados, etc. devem ser levadas em conta. Hoje, uma das tendências mais fortes de desenvolvimento desta área é a especialização dos sistemas de buscas, levando em conta, por exemplo, aspectos semânticos do tema que constitui o espaço de busca. Relacionado a isso, há também um grande interesse em expandir as fronteiras da ciência de recuperação de informação para outros tipos de conteúdo, como conteúdos de áudio e vídeo. Recentemente, a TREC - Text REtrieval conference, uma das maiores conferências sobre recuperação de texto, incorporou o tema de busca sobre áudio como uma sub-tarefa.

3.2 Histórico de técnicas de estimação de frequência

O problema de estimar a frequência de trecho de áudio é um problema estudado há um longo tempo, diversas técnicas e métodos já foram desenvolvidos sobre o tema, porém até o presente momento estas técnicas ainda apresentam fortes deficiências e não são capazes de atingir o nível desejado de qualidade. Frente a um sinal único claro, diversas técnicas apresentam um bom desempenho, porém quando testadas com sinais ruidosos, ou contendo mais de uma linha melódica estas técnicas tendem a falhar. Diferentes conceitos

podem ser aplicados na tentativa de estimar a frequência de um trecho de áudio, entre as principais técnicas tem-se: métodos que se baseiam no domínio do tempo, métodos que utilizam o domínio da frequência e métodos estatísticos.

Métodos baseados na análise do domínio do tempo se valem do fato que os sinais são periódicos, o que faz alguns eventos também serem periódicos, e, portanto, podem ser contados.

Taxa de cruzamento pelo zero (ZCR - Zero-crossing rate). A idéia deste método consiste em contar o número de vezes que o sinal de áudio cruza o eixo dos tempos, imaginando-se que a principal componente de frequência responsável por este cruzamento será a frequência fundamental. A Figura 1 exemplifica o fato, onde uma componente de frequência mais alta não exerce grande influência sobre o número de cruzamentos do sinal com o eixo dos tempos.

Figura 1: Influência da frequência fundamental na taxa de cruzamento pelo zero

Taxa de picos. Este método consiste em contar o número de picos por segundo em um sinal, sabendo que através do número de picos é possível inferir a frequência do sinal, tem-se então a estimativa do frequência. Analogamente ao ZCR, a frequência fundamental será a componente de frequência que mais contribuirá para a ocorrência de picos no sinal, assim é possível dizer que a estimativa obtida corresponde à estimativa da frequência fundamental.

Taxa de eventos de inclinação. Devido ao fato do sinal ser periódico a inclinação do sinal também irá variar periodicamente, assim observar picos e zeros da inclinação do sinal pode ser mais informativo do que observar picos e zeros do sinal original.

Correlação. Existem ainda métodos que se baseiam na correlação entre duas amostras de áudio, definindo assim a similaridade entre os dois sinais. Formas de onde similares apresentariam uma correlação alta, enquanto formas de onda muito diferentes teriam uma baixa correlação.

Além de métodos baseados no domínio do tempo, existem também diversas técnicas baseadas no domínio da frequência. Estas, por sua vez, recaem sobre o fato de que o sinal pode ser modelado como uma soma de séries harmônicas, guardando um alto grau de informação sobre a frequência fundamental.

Proporção de componentes de frequência Em 1979, Martin Piszczalski (PISZCZALSKI, 1986) (Piszczalski; Galler, 1979) trabalhava em um sistema capaz de transcrever músicas automaticamente, assim, necessariamente um dos componentes deste sistema era o componente de extração de notas. O procedimento adotado se valia do cálculo do espectro do sinal, da detecção de picos deste espectro, e de uma análise probabilística destes picos.

Métodos baseados em filtros Estes métodos utilizam a idéia de aplicar diferentes filtros ao sinal, e analisar sua saída. Por exemplo, caso um sinal possua uma saída alta após a aplicação de um filtro passa-faixa, pode-se afirmar que este sinal possui entre suas componentes a frequência do filtro. Em 1977 James A. Moorer (MOORER, 1977), propos um algoritmo denominado Filtro Comb Ótimo, baseado nestes conceitos. Uma tentativa mais recente foi proposta por John E. Lane (LANE, 1990), denominado Filtro IIR Ajustável. Existem ainda diversas técnicas que se apóiam sobre a análise cepstrum, que corresponde ao resultado da tranformada de Fourier do log do espectro de magnitude so sinal de entrada.

Diversos métodos estatísticos sobre o domínio da frequência também foram desenvolvidos, dentre estes se deve destacar duas abordagens importantes: redes neurais e estimadores de Máxima Verossimilhança.

3.3 Histórico de técnicas de reconhecimento de sinais de áudio

Há uma diversidade de métodos que podem ser úteis na tentativa de se extrair algum tipo de informação de um conteúdo de áudio, simplificadamente, podem ser divididos entre

os que assume algum tipo de conteúdo relacionado à fala, e os que não assumem. Uma diversidade de métodos é direcionado a reconhecer automaticamente a fala, traduzindo um discurso para texto, por exemplo. Este trabalho porém está mais próximo de métodos que abordam o áudio de uma forma mais genérica, sendo mais adequado para conteúdos relacionados a melodias, por exemplo.

Conteúdos de áudio podem armazenar diversas classes de áudio, como melodias, efeito sonoros, sons de animais, e etc., isso deixa claro que métodos baseados na fala são suficientes para uma tarefa geral de busca de áudio. Além desta variedade de sons, um grande complicador é que estes diversos tipos de sons muitas vezes estão misturados, e até simultâneos em um conteúdo de áudio, por exemplo, em uma música, diversos instrumentos contribuem para gerar uma melodia, enquanto pode ainda haver a parte cantada da música.

Um dos problemas mais básicos da análise de áudio é diferenciar um conteúdo constituído por fala, de um conteúdo não vocal. Esta tarefa é importante, pelo fato de existirem diferentes técnicas adequadas para cada um dos tipos, assim, aplicar uma técnica baseada na fala sobre um conteúdo constituído por uma melodia não produzirá resultados úteis. John Saunders (SAUNDERS, 1996) apresentou uma técnica baseada em estatísticas do contorno de energia e da taxa de cruzamentos pelo zero, e reportou uma um acerto de 98% na classificação de comerciais de rádio. Eric Scheirer e Malcolm Slaney utilizaram uma técnica baseada na combinação de diversas características, energia de modulação, "centróide espectral", e taxa de cruzamento pelo zero, além de se valer de diversos classificadores, e reportam uma taxa de erro de 1.4% sobre uma grande coleção de radiodifusão FM. Michelle Spina and Victor Zue (SPINA; ZUE, 1996) utilizaram noticiários de rádio, e foram capazes de atingir 80,9% de acerto na classifcação dos conteúdos em sete categorias: limpo, telefone, fala ruidosa, silêncio, música e fala mais música.

O próximo passo após a distinção de um conteúdo de áudio seria permitir a busca sobre conteúdos, isto requer alguma medida de similaridade entre conteúdos de áudio, o que é um assunto extremamente complicado. Pode-se usar conceitos simples para similaridades de texto, como o número de palavras em comum, a ordem que as palavras aparecem, etc., porém, no domínio do áudio, estas medidas não estão tão claras. Para tentar evitar definir tal conceito, técnicas de inteligência artificial como redes neurais e mapas auto-organizáveis foram utilizadas, pelo fato de serem capazes de lidar com conceitos que não estão formalmente definidos.

Em 1996, um grupo em Muscle Fish LLC (WOLD et al., 1996) produzio um interes-

sante trabalho, que se valia de características psico-acústicas para caracterizar arquivos de áudio. Um classificador Gaussiano foi utilizado para analisar os arquivos, e uma distância de Mahalanobis foi utilizada para estabelecer a similaridade entre os conteúdos.

Simplificações do problema foram feitas para se permitir um entendimento melhor da área, assim alguns trabalhos utilizaram arquivos MIDI (Musical Instrument Digital Interface), que já possuem uma representação das notas da música. Pesquisadores em Cornell conseguiram bons resultados ao estabelecer três níveis de quantização, dependendo se uma nota seguinte fosse mais alta, mais baixa, ou similar à nota anterior (GHIAS et al., 1995). Algoritmos de busca baseados em cadeias também foram utilizados em tal tarefa, como nos trabalhos produzidos pela Universidade de Waikato na Nova Zelândia (MCNAB et al., 1996).

Alguns sistemas que se propõem a realizar esta tarefa de busca de áudio podem ser encontrados, diversas formas de entrada para a busca são utilizadas, além é claro da diversidade de algoritmos que são empregados. Entre os principais podem ser citados: midomi e musipedia.

4 Resenha Bibliográfica

Na presente seção serão apresentados as diversas técnicas nas quais este trabalho se fundamenta. Além de uma visão geral dos aspectos fundamentais de cada assunto, são dadas referências para material detalhado para o completo entendimento destes.

4.1 Processamento de sinais

4.1.1 Transformada rápida de Fourier

4.2 Método dos mínimos quadrados

O método dos mínimos quadrados surgiu no início do século XIX e é uma técnica matemática de otimização amplamente utilizada até os dias atuais. Este método permitem encontrar os parâmetros de uma função f modelo que melhor representam uma relação entre grandezas. Esta relação é usualmente definida a partir de pares ordenados (x_i, y_i) e, neste caso, deseja-se ajustar os parâmetros de f de modo a minimizar a soma dos quadrados dos resíduos:

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

A teoria mostra que para o caso em que a função modelo f é linear com relação aos seus parâmetros, a solução para a aproximação é única e ocorre quando as derivadas parciais do erro quadrático com relação a cada parâmetro é zero. A solução é obtida resolvendo o sistema linear resultante.

Referência: X

4.3 Adaptatividade

5 Conceitos

5.1 Processamento de sinais

5.2 Adaptatividade

6 Técnicas e procedimentos usados

7 Resultados

8 Análise

9 Crítica

10 Melhorias e Trabalhos futuros

11 Contribuições

Referências

BRøNDSTED, T. et al. A system for recognition of hummed tunes. In: *COST G-6 Conference on Digital Audio Effects*. [S.l.: s.n.], 2001.

FOOTE, J. An overview of audio information retrieval. *ACM Multimedia Systems*, v. 7, p. 2–10, 1999.

GERHARD, C. D.; GERHARD, D.; GERHARD, D. *Pitch Extraction and Fundamental Frequency: History and Current Techniques*. [S.l.], 2003.

GHIAS, A. et al. Query by humming: Musical information retrieval in an audio database. In: *ACM Multimedia*. [S.l.: s.n.], 1995. p. 231–236.

LANE, J. E. Pitch detection using a tunable iir filter. In: *Computer Music Journal*. [S.l.: s.n.], 1990. v. 14, p. 46–57.

MCNAB, R. J. et al. Towards the digital music library: tune retrieval from acoustic input. In: *DL '96: Proceedings of the first ACM international conference on Digital libraries*. New York, NY, USA: ACM, 1996. p. 11–18. ISBN 0-89791-830-4.

MIDOMI. `http://www.midomi.com/`.

MOORER, J. A. On the transcription of musical sound by computer. In: *Computer Music Journal*. [S.l.: s.n.], 1977. v. 3, p. 32–38.

PARSONS, D. Book. *The directory of tunes and musical themes*. [S.l.]: S. Brown, Cambridge, Eng. :, 1975. 288 p. : p. ISBN 090474700.

PISZCZALSKI, M. *A computational model of music transcription*. Tese (Doutorado) — University of Stanford, Ann Arbor, MI, USA, 1986.

Piszczalski, M.; Galler, B. A. Predicting musical pitch from component frequency ratios. *Acoustical Society of America Journal*, v. 66, p. 710–720, set. 1979.

SAUNDERS, J. Real-time discrimination of broadcast speech/music. In: *ICASSP '96: Proceedings of the Acoustics, Speech, and Signal Processing, 1996. on Conference Proceedings., 1996 IEEE International Conference*. Washington, DC, USA: IEEE Computer Society, 1996. p. 993–996. ISBN 0-7803-3192-3.

SPINA, M. S.; ZUE, V. Automatic transcription of general audio data: Preliminary analyses. In: *Proc. ICSLP '96*. Philadelphia, PA: [s.n.], 1996. v. 2, p. 594–597.

THE Open Music Encyclopedia. `http://www.musipedia.org/`.

WALL, A. *History of Search Engines: From 1945 to Google 2007*. `http://www.searchenginehistory.com/`.

WIERING, F. Can humans benefit from music information retrieval? In: *Adaptive Multimedia Retrieval*. [S.l.: s.n.], 2006. p. 82–94.

WOLD, E. et al. Content-based classification, search, and retrieval of audio. *IEEE MultiMedia*, IEEE Computer Society Press, Los Alamitos, CA, USA, v. 3, n. 3, p. 27–36, 1996. ISSN 1070-986X.