Transcrição Automática de Sinais de Áudio Monofônicos

Narciso Trevilatto Junior[1], Jayme Garcia Arnal Barbedo[1], Amauri Lopes[1]

[1]Departamento de Comunicações - FEEC - UNICAMP
Caixa Postal 6101, CEP: 13.083-970, Campinas - SP - Brasil

{narciso,jgab,amauri}@decom.fee.unicamp.br

Abstract. *This paper describes a method to capture pitches and durations of musical notes. It is applied on monophonic digital signals generated by only one sound source. The method is based on the estimation of the fundamental frequency over time by calculating the signal autocorrelation. Processing techniques are applied to adjust the method to different temporal and spectral features of the signals. The validation of the method is made by using a data bank specially developed to include the widest range of real possibilities the method can face. Finally, advantages and weaknesses are pointed out and suggestions are presented to further improvement of this work.*

Resumo. *Este artigo propõe um método para extração das alturas e durações de notas musicais. A aplicação do método é voltada a sinais gerados por uma única fonte sonora. O método baseia-se na estimação da freqüência fundamental ao longo do tempo através do cálculo da autocorrelação do sinal. Técnicas de processamento são aplicadas a fim de sintonizar o método às diferentes características temporais e espectrais apresentadas pelos sinais. A validação do método é realizada usando um banco de dados especialmente desenvolvido para incluir a mais ampla gama de condições reais com as quais o método poderá se deparar. Por fim, pontos fortes e fracos são destacados e sugestões para continuidade do trabalho são apresentadas.*

1. Introdução

Transcrição automática de música é o ato de se extrair uma linguagem simbólica de um sinal de áudio, contendo todas as notas com suas alturas, durações, intensidades, timbres, etc. Apesar desse processo poder ser realizado sem ajuda computacional, uma ferramenta eficaz para esse propósito seria de grande utilidade no meio musical.

Estudos em transcrição automática começaram na década de 70 [Moorer 1975], mas resultados satisfatórios só têm sido alcançados por métodos atuando em sinais de áudio sujeitos a restrições. A proposta deste trabalho é a transcrição automática de sinais monofônicos (uma única fonte sonora). Apesar de relativamente mais simples, uma vez que só há necessidade de se encontrar uma nota por vez, a solução deste problema é essencial para o desenvolvimento de métodos mais sofisticados.

2. Estimativa da Freqüência Fundamental

Na transcrição de sinais de áudio monofônicos, é imprescindível a obtenção da altura de cada nota. A essas alturas estão associadas freqüências, que normalmente coincidem com a freqüência fundamental (*f0*) do sinal. O método proposto se baseia no cálculo de *f0* para estimar as alturas das notas.

Existem muitas possibilidades para se detectar a freqüência fundamental. Um estudo mais específico é apresentado em [Gerhard 2003], [Gerhard 2002] e [Middleton 2003]. O método utilizado neste trabalho é o da autocorrelação, pois foi considerado a melhor opção para maximizar a relação entre eficiência e complexidade. Esse método é estudado em [Bello, Monti e Sandler 2000] e [Hamidi-Toosi e Laska 2004].

2.1. Método da Autocorrelação

O método da autocorrelação (AC) é uma medida do grau de semelhança entre o sinal e ele mesmo deslocado no tempo. Se o sinal tem uma periodicidade, essa semelhança será maior quando o deslocamento equivaler ao período da onda. A freqüência fundamental do sinal é obtida através de

$$f_0 = F_s / n_d \ , \tag{1}$$

onde F_s é a freqüência de amostragem e n_d é o deslocamento do primeiro pico da AC. Para a obtenção das notas musicais, é necessário rastrear *f0* ao longo do tempo. Adota-se o seguinte procedimento: uma janela de 50 ms é deslocada com passos de 25 ms por toda a duração do sinal. Para cada janela aplica-se a AC e calcula-se a *f0* correspondente. Esse tamanho de janela permite a detecção de ondas com período máximo de 25 ms (40 Hz). Existe então um compromisso entre o limite inferior da freqüência e a menor duração de uma nota que pode ser detectada.

2.2. Extração do Pico Correspondente à Freqüência Fundamental

O comportamento da AC pode apresentar grande variação, dependendo da fonte sonora considerada. A análise da AC pode apresentar problemas devidos, por exemplo, à presença de harmônicas cuja energia relativa é grande se comparada a *f0*. Nesse caso podem surgir outros picos na autocorrelação antes daquele que corresponde a *f0*.

Para detecção do pico correto, a função autocorrelação é expandida no tempo através de sobre-amostragem e subtraída da própria função de autocorrelação original. O efeito deste procedimento, tomando-se como referência um pico com índice de tempo *n*, será o de eliminar ou atenuar o possível pico com índice de tempo *2n*. O procedimento elimina também o pico localizado no índice de tempo zero.

O máximo dessa função de autocorrelação modificada geralmente corresponde ao primeiro pico (deslocamento de um período). Porém, podem existir picos maiores que o primeiro. O procedimento então é procurar por um novo pico entre o máximo e a origem. Se a amplitude desse novo pico estiver acima de um certo limiar relativo ao máximo, este é assumido como o atual candidato. O procedimento é realizado recursivamente até que não se satisfaça a condição do novo pico ultrapassar o limiar.

3. Estimativa das Alturas (Freqüências) e Durações das Notas

Nesta seção são descritos os processamentos necessários para que o programa forneça em sua saída estimativas confiáveis para a freqüência e duração de cada nota. Devem ser definidas todas as notas e pausas presentes no sinal, assim como suas alturas e instantes inicial e final.

O resultado do método da autocorrelação é um vetor contendo os valores de *f0* encontrados para todas as janelas. São necessárias algumas manipulações para se extrair desse vetor os parâmetros desejados. Essas manipulações são detalhadas a seguir.

Normalmente, os instrumentos musicais são afinados usando alguma nota como referência. Podem ocorrer situações em que determinado instrumento é afinado usando uma freqüência ligeiramente diferente da esperada. Quando isso ocorre, todas as notas do instrumento são deslocadas em freqüência e um ajuste deve ser feito para a obtenção das notas corretas. Primeiramente, o vetor das freqüências é recalculado de acordo com:

$$m = 12\log_2\left(f_0/440\right) + 69, \quad (2)$$

onde k_n é a nota MIDI correspondente a cada freqüência. A seguir, o algoritmo busca por segmentos de 125 ms (quatro janelas) em que as quatro notas k_n associadas têm uma variância menor que um certo limiar. Satisfeita essa condição, a média das quatro notas é calculada. Essas médias são subtraídas dos valores inteiros mais próximos resultando um conjunto de valores entre -0,5 e 0,5. Por fim o vetor de notas é subtraído da média desses valores. Com isso, as notas devem assumir posições mais próximas a valores inteiros, reduzindo a chance de se arredondar valores para notas erradas.

Uma última manipulação é realizada sobre o vetor com o propósito de explicitar os instantes de início e término de cada nota. Isso é alcançado definindo-se uma duração mínima de 125 ms (quatro janelas) para as notas e pausas. A seqüência final será então composta por subseqüências com quatro ou mais valores iguais. Segmentos com valor "zero" indicam ausência de periodicidade (nenhuma nota executada).

O método para detecção do início e final de cada nota segue os seguintes passos: 1) procura-se uma seqüência de quatro pontos consecutivos com o mesmo valor, o primeiro dos quais definindo o início de uma nota; 2) o final da nota é dado pelo último ponto após o qual surgem quatro pontos consecutivos com valores diferentes do seu; 3) se o algoritmo determinar que a próxima nota se inicia até quatro pontos após o final da anterior, os pontos de transição são considerados pertencentes a essa nova nota, caso contrário o valor nulo é associado a esses pontos, caracterizando a ausência de nota.

4. Resultados

O banco de dados usado é composto por 100 arquivos de 15 segundos, amostrados a 44,1 kHz e quantizados com 16 bits. Foram avaliados sinais provenientes de cordas, sopros, piano e voz masculina e feminina. Os resultados obtidos são apresentados na Tabela 1.

Tabela 1. Resultado das simulações.

Fonte Sonora	Total de Notas	Notas Detectadas com Sucesso	Notas Falsas Detectadas	Índice *I*
Cordas	507	484	45	0,87
Sopros	1805	1712	93	0,90
Voz	492	463	69	0,80
Total	2804	2659	224	0,87

Obs.: $I = (NotasCorretas - NotasFalsas) / TotalNotas$

Na tabela, quanto mais próximo de 1 o valor de *I*, mais preciso o método. Como se pode observar, a menor eficiência ocorreu na análise de sinais de voz. Dentre os motivos encontrados, pode-se citar: a) em sinais de voz a transição entre notas é freqüentemente realizada de um modo suave (*legato*), tornando mais freqüente a ocorrência de erros nessas regiões; b) este tipo de sinal possui sons fricativos mais longos do que nos instrumentos, tornando possível a detecção de uma nota falsa nesses

intervalos de tempo; c) estes sinais possuem uma maior quantidade de desafinações, associadas à liberdade de interpretação do cantor.

A saída do programa apresenta notas discretizadas. Por esse motivo, efeitos relativos a variações contínuas de freqüência (*glissando, vibrato*) não são apontados. A detecção do pico correspondente a *f0* é outro ponto que deverá ser aperfeiçoado, especialmente na análise de sinais com harmônicas de grande intensidade. Apesar disso, o método apresentou um desempenho próximo daqueles encontrados na literatura. Além disso, é importante frisar que as técnicas aqui apresentadas são apenas a primeira etapa na implementação de um método de transcrição de sinais de música confiável e robusto. Novas técnicas de rejeição de harmônicas vêm sendo testadas com sucesso. Estas técnicas, em conjunto com a incorporação de lógicas baseadas em teoria musical, deverão melhorar significativamente os resultados alcançados.

5. Conclusão

Neste artigo foi proposto um método para transcrição automática de sinais de áudio monofônicos. O algoritmo faz uso do método da autocorrelação para detecção de *f0*. Uma quantização da *f0* de cada nota é realizada através de sua transformação para a escala MIDI. Este procedimento aumenta a precisão das estimativas, embora faça com que efeitos como *vibrato* ou *glissando* sejam ignorados. As durações das notas foram estimadas tendo como base os instantes em que o sinal se torna ou deixa de ser periódico.

O método obteve o desempenho esperado para esta primeira etapa de desenvolvimento do transcritor automático de música. Os resultados alcançados mostram boa confiabilidade do sistema, principalmente nos casos em que os instrumentos geram notas com valores de freqüência discretos. Porém, existem pontos que precisam ser mais bem explorados. Dentre as medidas que vêm sendo adotadas para melhorar a precisão do método, pode-se citar o aperfeiçoamento do algoritmo de detecção do pico da autocorrelação de modo a evitar a detecção de picos referentes a harmônicas da freqüência fundamental. Adicionalmente, versões futuras do algoritmo deverão incorporar a visualização de vibratos e glissandos, fazendo com que o conjunto de informações fornecidas se torne mais completo. Por fim, as notas deverão ter suas probabilidades de ocorrência alteradas através da implementação de uma lógica baseada em teoria musical.

6. Referências Bibliográficas

Bello, J. P., Monti, G. and Sandler, M. (2000) “Techniques for Automatic Music Transcription”, Int. Symp. on Music Inf. Retrieval (ISMIR), Plymouth, Mass., USA

Gerhard, D. (2003) “Pitch Extraction and Fundamental Frequency: History and Current Techniques”, University of Regina Technical Report TR-CS 2003-06

Gerhard, D. (2002) “Computer Music Analysis”, Simon Fraser University, School of Computing Science, Technical Report CMPT TR 97-13, Burnaby, BC, July.

Hamidi-Toosi, F. e Laska, J. (2004) “Pitch Detection for the Next Millenium and Beyond” www.ews.uiuc.edu/~laska/ece320/files/Report.pdf

Middleton, G. (2003) “Pitch Detection Algorithms”, http://cnx.rice.edu/content/m11714

Moorer, J. (1975) “On the Segmentation and Analysis of Continuous Musical Sound by Digital Computer”. Ph.D. thesis. Department of Music, Stanford University.