



Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 5º Congresso de Engenharia de Áudio
11ª Convenção Nacional da AES Brasil
21 a 23 de Maio de 2007, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Avaliação do Bounded- Q Fast Filter Bank na Transcrição Automática de Música

Filipe C. da C. B. Diniz,¹ Luiz W. P. Biscainho¹ e Sergio L. Netto¹

¹ Universidade Federal do Rio de Janeiro, COPPE, Laboratório de Processamento de Sinais
Caixa Postal 68504, Rio de Janeiro, RJ, CEP 21941-972, Brasil

filiped@lps.ufrj.br, wagner@lps.ufrj.br, sergioln@lps.ufrj.br

RESUMO

Este artigo apresenta um estudo do desempenho da ferramenta de análise “*Bounded- Q Fast Filter Bank*” como base para um sistema de transcrição automática de música (TAM). Esse banco de filtros possui alta seletividade, espaçamento não-uniforme dos filtros e complexidade computacional reduzida, o que o torna atrativo para uso em TAM. São utilizados sinais musicais com características escolhidas para avaliar a resposta do sistema completo a problemas específicos relacionados com essa aplicação.

0 INTRODUÇÃO

A transcrição de música consiste em escrevê-la em uma partitura ou similar a partir de sua execução. Tradicionalmente, é realizada por uma pessoa com elevado nível de treinamento musical, apenas com o uso de sua audição.

Por questão de comodidade e praticidade, hoje a transcrição musical tende a ser realizada de forma automatizada, possivelmente por meio de algum *software*, a partir de uma gravação. A este processo, dá-se o nome de transcrição automática de música (TAM) [8].

Em termos gerais, a TAM pode ser dividida em 4 etapas:

1. Segmentação do sinal musical no domínio do tempo.

2. Identificação das notas presentes nos segmentos.

3. Reconhecimento do timbre das notas musicais identificadas.

4. Avaliação da coerência dos resultados obtidos.

A base para esse sistema é a análise espectral dos sinais de áudio. Uma das formas de a realizar é utilizar um banco de filtros [12] que separe o sinal em faixas estreitas na frequência. As saídas dos filtros descrevem ao longo do tempo a distribuição espectral do sinal sob análise.

O presente trabalho afere a capacidade de análise do *bounded- Q fast filter bank* (BQFFB), recentemente proposto [3, 4] num contexto de TAM. O artigo está assim organizado: na Seção 1, revisam-se soluções para

análise espectral por bancos de filtros; o BQFFB é rapidamente descrito na Seção 2; na Seção 3 são descritos, realizados e avaliados alguns experimentos com o BQFFB aplicado a sinais musicais de características específicas; por fim, as conclusões são resumidas na Seção 4.

1 ANÁLISE ESPECTRAL POR BANCOS DE FILTROS

Um tipo de banco de filtros muito difundido é a *sliding Fast Fourier Transform* (sFFT), cuja implementação requer apenas uma multiplicação complexa por amostra de saída [5]. A sFFT corresponde a uma FFT [2] aplicada sobre janelas deslocadas de amostra em amostra sobre o sinal de entrada. Vista sob outra perspectiva, ela equivale a um banco de filtros cujas saídas, a cada nova amostra de entrada, são as raiais da FFT calculada sobre o bloco corrente. Esse esquema resulta numa descrição dinâmica do espectro do sinal. Sua desvantagem está na baixa seletividade dos filtros associados (cerca de 13 dB em relação aos canais adjacentes), obtidos a partir de filtros-base de ordem muito reduzida [10].

Para atacar esse problema, uma nova ferramenta de análise foi proposta em [10]. Nesta nova técnica, denominada de *fast filter bank* (FFB), combina-se a estrutura modular da FFT com filtros-base de ordem maior, gerando filtros longos, porém com muitos coeficientes nulos (interpolados), como na técnica de *frequency-response masking* (FRM) [9] para projeto de filtros digitais não-recursivos. Com isso, obtém-se um banco de filtros de complexidade reduzida, mas alta seletividade (cerca de 56 dB entre canais adjacentes), que permite maior distinção entre as várias componentes do sinal musical, e conseqüente melhor identificação das notas. O espaçamento freqüencial dos filtros do FFB é constante, o que o torna ineficiente em TAM: na escala de temperamento igual, predominante na música ocidental [13], o espaçamento freqüencial entre as fundamentais das notas musicais aumenta geometricamente com a altura. Assim, se o espaçamento entre canais do FFB for dimensionado para baixas freqüências, será superdimensionado para altas freqüências, gerando uma quantidade desnecessária de dados para análise. Em contrapartida, se o espaçamento entre canais do FFB for escolhido de acordo com os requisitos de altas freqüências, acarretará resolução insuficiente em baixas freqüências.

No domínio da transformada, foi proposto em Brown [1] um sucedâneo da FFT com espaçamento geométrico entre as componentes espectrais de análise, a chamada *constant-Q transform* (CQT). Contudo, esta compartilha a baixa seletividade da FFT. Tal dificuldade foi superada por Graziosi et al. [6] [11] no chamado CQFFB, que combina o espaçamento geométrico da CQT com a alta seletividade do FFB. Embora implementando um banco de filtros altamente seletivos espaçados geometricamente, a CQFFB perde a característica de baixa complexidade computacional do FFB.

Uma outra transformada propõe um espaçamento quase-geométrico entre as componentes espectrais: analisam-se as oitavas (distribuídas de forma geométrica) subdivididas por filtros com espaçamento uniforme. Esse esquema, chamado *bounded-Q transform* (BQT) [7], também herda a seletividade reduzida da FFT, na qual se baseia. Associada ao FFB, a BQT foi a base do *bounded-Q fast filter bank* (BQFFB), que é objeto deste trabalho.

2 BOUNDED-Q FAST FILTER BANK

A versão de alta seletividade da BQT é denominada de *bounded-Q fast filter bank* (BQFFB), e foi apresentada por Diniz et al. [3] [4].

A implementação do BQFFB consiste em submeter o sinal a ser analisado, inicialmente, a uma CQFFB de 10 canais com espaçamento geométrico de 2, correspondendo às 10 oitavas percebidas pelo sistema auditivo humano, entre 20 Hz e 20 kHz. Essa separação inicial possui alta seletividade e, por apresentar poucos canais, moderado custo computacional. Cada oitava é decimada para que seu espectro ocupe toda a faixa de freqüências; nesta nova escala, é submetida a um FFB de 32 canais, resultando numa separação linear de no mínimo um quarto de tom entre os sub-canais. Tal resolução é suficiente para distinguir freqüências afastadas de meio tom, que é, convencionalmente, o menor afastamento entre notas da escala musical de igual temperamento. Esse procedimento pode ser visualizado na Figura 1.

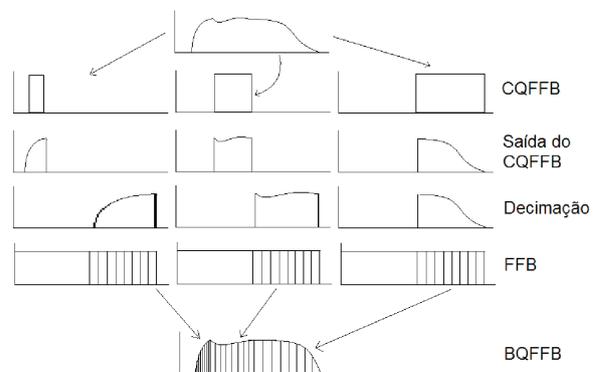


Figura 1: Procedimento para implementação de uma BQFFB a partir de um CQFFB seguido de um FFB.

Dessa maneira, tem-se uma ferramenta que une a propriedade de alta seletividade do FFB com a de espaçamento quase-geométrico da BQT. Deve-se notar, contudo, que para sua implementação, o BQFFB associa um CQFFB (que realiza a separação inter-oitavas) com um FFB (que realiza a separação intra-oitavas). Isso é análogo a implementar a BQT associando uma CQT seguido de uma FFT.

O resultado final é uma ferramenta de análise espectral que consegue conjugar espaçamento geométrico, alta seletividade e complexidade moderada. Uma tabela

comparativa com as propriedades de cada ferramenta de análise espectral e uma análise de complexidade podem ser encontradas em [3].

3 EXPERIMENTOS

Nesta seção, são realizados testes para investigar a capacidade de análise do BQFFB para sinais musicais reais de características variadas. Foram utilizados os seguintes sinais de teste:

- Exemplo 1 - Sinal com frequências baixas: trecho de Os Elefantes, de O Carnaval dos Animais, de Saint-Saëns.
- Exemplo 2 - Sinal com frequências altas: trecho do 3º Movimento do Concerto para Flautim RV 443, de Vivaldi.
- Exemplo 3 - Sinal com passagem rápida: trecho de O Vôo do Besouro, de Rimski-Korsakoff, em arranjo para piano de Rachmaninoff.
- Exemplo 4 - Sinal com passagem lenta: trecho do 3º Movimento da Marcha Fúnebre, 3º Movimento da Sonata nº 2 para piano de Chopin.
- Exemplo 5 - Sinal contendo glissando: trecho da Rhapsody in Blue, de Gershwin.

3.1 Visualização dos resultados

Os resultados dos experimentos descritos nesta seção precisam ser adequadamente apresentados, para permitirem avaliar o desempenho do sistema. Uma métrica largamente usada na literatura associada é a contagem das notas presentes na partitura do sinal original que foram corretamente detectadas no sinal analisado. Esse será o critério aplicado na Seção 3.7 a todos os experimentos, de forma comparativa.

Para facilitar a identificação das notas, foi feita uma normalização no sinal de saída dos filtros dentro de cada oitava do BQFFB, dividindo-se as amplitudes das saídas destes filtros pelo valor máximo de amplitude na oitava correspondente. Isso faz com que se perca a proporcionalidade entre as oitavas, o que, contudo, não altera a capacidade de detecção de notas do sistema como um todo.

Em seguida, para cada oitava, recorre-se à representação gráfica de uma matriz em que cada linha representa a saída de um filtro e cada coluna representa um instante de tempo. As amplitudes são codificadas em tons de cinza: o branco corresponde a amplitude zero e o preto, a amplitude máxima. Essa matriz, conhecida como sonograma, fornece uma representação tempo-frequência de fácil visualização.

Contudo, esse tipo de figura ainda é deficiente para representar os transitórios dos filtros. Para contornar esse problema, lança-se mão de um gráfico tridimensional conhecido como *waterfall*. Ele exibe uma

representação tempo-frequência menos compacta, favorecendo a análise do regime transitório.

Para facilitar a visualização dos resultados obtidos com o BQFFB, foi feita uma análise preliminar para determinar em que oitavas o espectro do sinal musical estava concentrado. Tal análise levou em conta que instrumentos que estavam presentes na execução de cada trecho. E a seguir, também por economia de espaço, são representadas apenas as oitavas com informação significativa. Na ordenação das 10 oitavas analisadas, a décima contém as frequências mais altas.

3.2 Exemplo 1 - Frequências baixas

Este experimento tem como objetivo estudar o desempenho do BQFFB em baixas frequências. O sinal usado neste Exemplo 1 (ver lista acima) é uma peça onde se pode ouvir uma linha melódica executada por um contrabaixo, instrumento capaz de gerar notas que vão, tipicamente, do E1 (43,65 Hz) até o B3 (246,94 Hz). A partitura do trecho da música aqui utilizado é visualizada na Figura 2(a).

As oitavas 2, 3 e 4 do sonograma deste sinal são mostradas na Figura 2(b), enquanto que o comportamento transitório da oitava 3 é representado na Figura 2(c).

Pela Figura 2(b), pode-se verificar que a melodia executada pelo contrabaixo se concentra nas oitavas 2 e 3. Na oitava 4 há apenas parciais superiores das notas emitidas pelo instrumento. A análise do gráfico de *waterfall* da Figura 2(c) mostra que a detecção precisa do início e do fim das notas fica efetivamente comprometida: o diagrama apresenta um aspecto suavizado e pouco contraste temporal entre notas consecutivas.

Isso se dá porque os filtros das oitavas inferiores do BQFFB, cujas frequências centrais são mais baixas, apresentam banda muito estreita e, por isso, ordem mais elevada. Como consequência, seus transitórios de saída se tornam muito espalhados no tempo, dificultando com isso a detecção de início (*onset*) ou de fim (*offset*) de cada nota a ser identificada. Visto sob outro prisma, o transitório, sendo concentrado no tempo, envolve informação de alta frequência, que os filtros das oitavas baixas não conseguem detectar.

3.3 Exemplo 2 - Frequências altas

Este exemplo verifica o desempenho do BQFFB na identificação de notas de altas frequências. O sinal deste Exemplo 2 é uma peça onde se pode ouvir uma linha melódica executada por um flautim (também chamado de *piccolo*). Esse instrumento é capaz de gerar notas que vão, tipicamente, do C5 (523,25 Hz) até o D8 (4698,60 Hz). A partitura do trecho da música analisado nesse experimento é mostrada na Figura 3(a).

Para a peça deste Exemplo 2, montaram-se o sonograma das oitavas de 5 a 10, visto na Figura 3(b), e o gráfico *waterfall* da oitava 7, dado na Figura 3(c).

Obsevando a Figura 3(b), pode-se verificar que a melodia executada pelo flautim se concentra nas oitavas

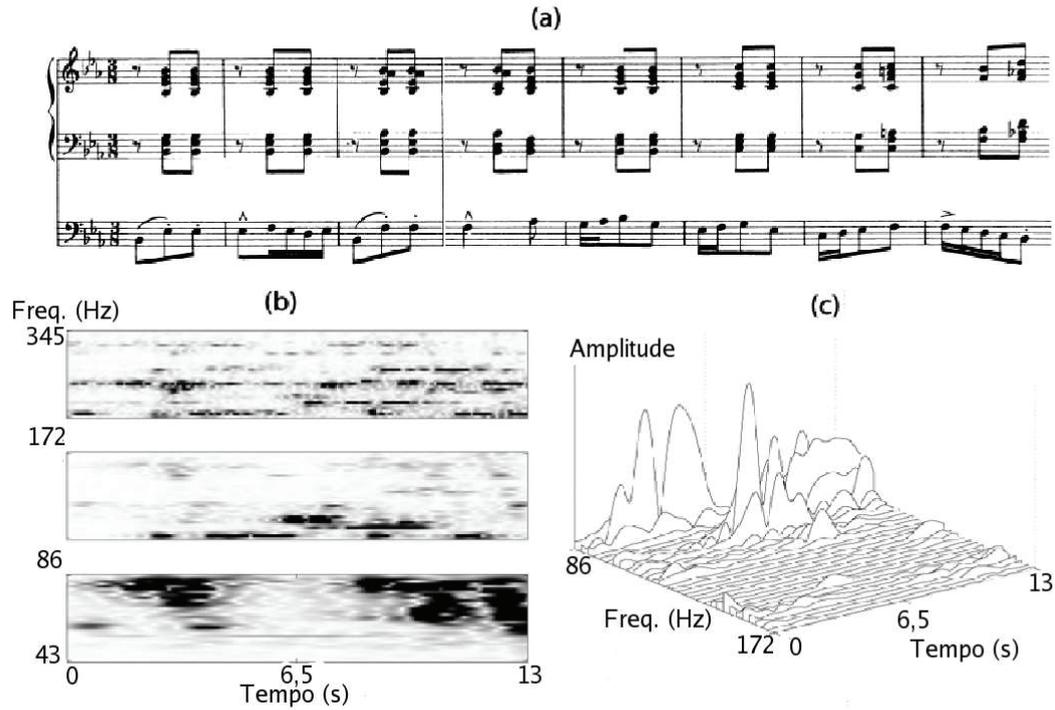


Figura 2: (a) Partitura do trecho da peça O Carnaval dos Animais usado no Exemplo 1. (b) Sonograma das oitavas 2, 3 e 4 no Exemplo 1. (c) Dinâmica temporal da oitava 3 no Exemplo 1.

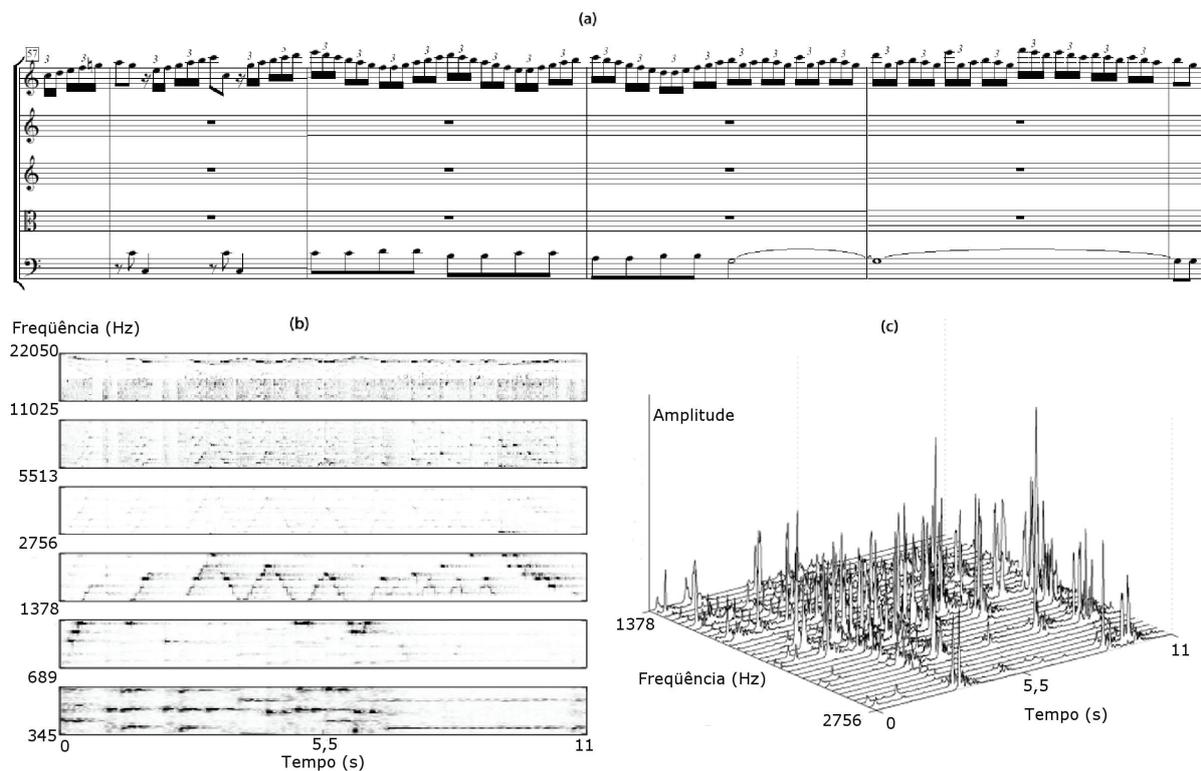


Figura 3: (a) Partitura do trecho do Concerto para Flautim usado no Exemplo 2. (b) Sonograma das oitavas de 5 a 10 no Exemplo 2. (c) Dinâmica temporal da oitava 7 no Exemplo 2.

6, 7 e 8, sendo fácil perceber-se o subir e descer da melodia. Nesse caso é possível, ainda, observar boa separação temporal das notas, se comparadas com os resultados do Exemplo 1, sendo possível precisar seus instantes de *onset* e de *offset*.

3.4 Exemplo 3 - Passagens rápidas

Este experimento analisa o desempenho do BQFFB sobre um sinal com passagens rápidas, ou seja, com rápida mudança entre as notas, que soam durante um curto intervalo de tempo.

O sinal usado neste Exemplo 3 é uma peça que apresenta notas tocadas ao piano em grande velocidade pela mão direita (notas de frequências mais altas), enquanto a mão esquerda executa o acompanhamento (em frequências mais baixas). A partitura do trecho analisado nesse exemplo é exibida na Figura 4(a).

O sonograma das oitavas de 4 a 8 deste sinal é representado na Figura 4(b), enquanto que a representação dinâmica da oitava 7 é vista na Figura 4(c).

Por essas figuras, pode-se verificar que a melodia se concentra nas oitavas 6, 7 e 8, enquanto que o acompanhamento se concentra na oitava 4. Com base no que foi explicado nos dois exemplos anteriores, tais notas rápidas recaem no caso de sinais transitórios. Estando associadas, nesse caso, a frequências relativamente altas, não apresentam problemas de detecção. De fato, o acompanhamento na faixa média ainda consegue ser bem descrito, pois ocorre em notas de maior valor, isto é, mais lentas.

Isso permite traçar limites para o bom funcionamento do sistema.

3.5 Exemplo 4 - Passagem lenta

Neste exemplo analisamos o desempenho do BQFFB para um sinal com passagens lentas, ou seja, com mudança lenta entre notas, que soam durante períodos mais prolongados.

O sinal usado neste exemplo, cuja partitura é vista na Figura 5(a), é uma peça que apresenta notas na região grave do piano, executadas em andamento muito lento.

A Figura 5(b) representa o sonograma das oitavas de 2 a 7 do sinal usado neste exemplo, enquanto que a Figura 5(c) representa o desenvolvimento dinâmico da oitava 3 deste sinal. Por essas figuras, verifica-se que a melodia e seus harmônicos se espriam pelas oitavas de 3 a 7, e que é possível fazer a separação temporal de todas as notas. Deve-se ressaltar que só foi possível distinguir notas de tão baixas frequências porque elas foram tocadas de forma lenta.

3.6 Exemplo 5 - Trinado e *Glissando*

Este experimento tem como objetivo estudar o desempenho do BQFFB sobre um sinal que apresente efeitos de trinado (repetição alternada e rápida de duas

notas vizinhas) ou *glissando* (seqüência de notas executadas em rápida sucessão, de modo a não serem individualmente articuladas, mas sim tocadas em um movimento contínuo). As notas envolvidas nestes dois efeitos tendem a cair em filtros adjacentes e, por serem tocadas rapidamente, devem apresentar os mesmos problemas associados a sinais de passagem rápida.

O sinal usado neste exemplo é uma peça que apresenta ambos os recursos, como indicado pela partitura dada na Figura 6(a).

O sonograma das oitavas de 3 a 9 é visto na Figura 6(b), enquanto que as Figuras 6(c) e 6(d) representam as dinâmicas das oitavas 4 (com o trinado) e 7 (com o *glissando*), respectivamente.

Pela Figura 6(b), pode-se verificar que o *trinado* é executado na primeira metade do trecho na oitava 4, e o *glissando* na segunda metade do trecho, envolvendo as oitavas 4, 5 e 6. Ambos os recursos foram devidamente detectados. No gráfico *waterfall* da Figura 6(c), pode-se ver as notas bem definidas alternando-se no trinado. Já o *glissando* fica muito bem caracterizado na Figura 6(d), onde se pode acompanhar as notas deslizando ao longo dos diferentes canais.

3.7 Contagem de notas

Para avaliar a capacidade de análise do sistema, realizou-se a contagem das notas coincidentes entre as prescritas na partitura e as reconhecidas no sinal como executadas pelo instrumentista (obviamente, escolheram-se interpretações suficientemente acuradas). Nesse processo, são identificados algumas questões:

1. Quando considerar que uma nota foi efetivamente tocada? Isso pode ser feito aplicando-se um limiar à saída dos filtros, acima do qual decide-se pela presença da nota.
2. Como decidir se, em um instante que a saída do filtro ultrapassa o limiar estabelecido, trata-se de uma nota tocada ou de um parcial de uma outra nota? Analisando-se as outras parciais associadas.
3. Como mapear os 32 canais de cada oitava nos 12 semitons da escala temperada? Quantizando-os para as fundamentais mais próximas.

Como o presente trabalho é uma etapa inicial do estudo da transcrição automática de música, o processo de contagem de notas foi feito de forma manual, com base na leitura direta dos sonogramas e dos gráficos tipo *waterfall*. Dessa maneira, foi possível montar a Tabela 1, onde pode-se ver o percentual de acertos entre as notas reconhecidas como executadas e as descritas na partitura. Deve-se lembrar que as notas nesta etapa foram detectadas de forma subjetiva, levando-se em consideração os parciais e buscando apenas verificar a presença das notas contidas na partitura. O sistema

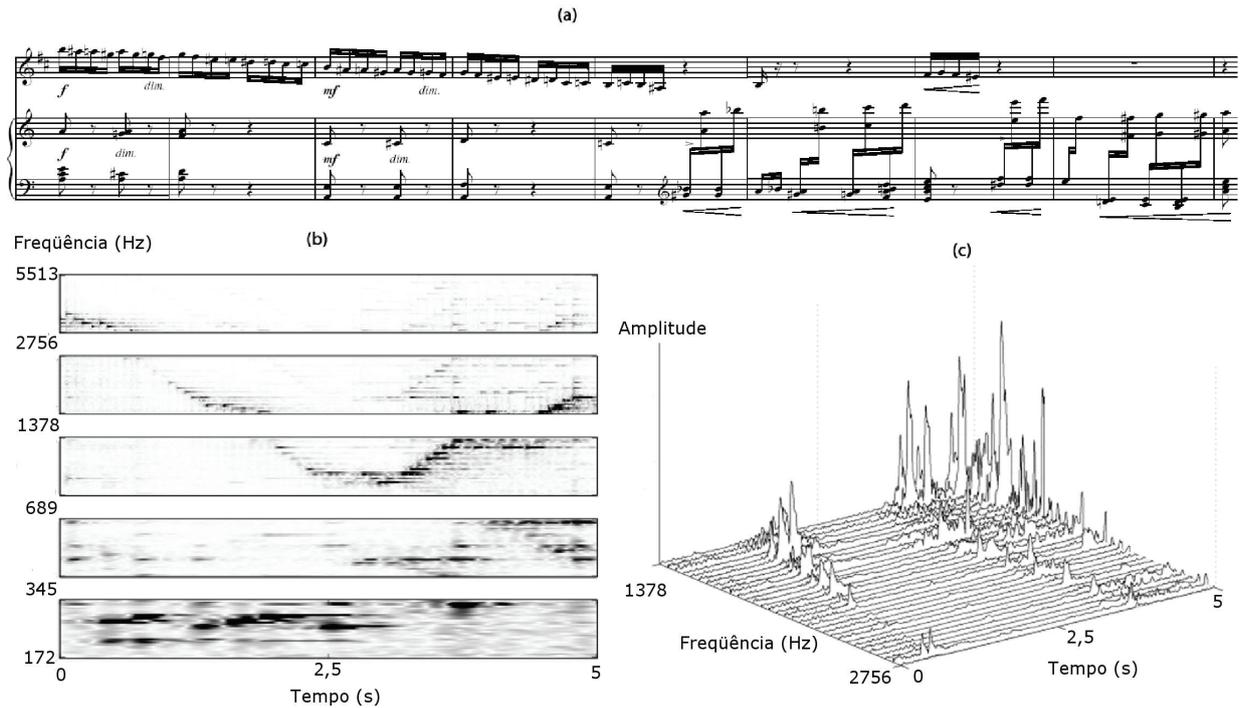


Figura 4: (a) Partitura do trecho de O Vôo do Besouro usado no Exemplo 3. (b) Sonograma das oitavas de 4 a 8 no Exemplo 3. (c) Dinâmica temporal da oitava 7 no Exemplo 3.

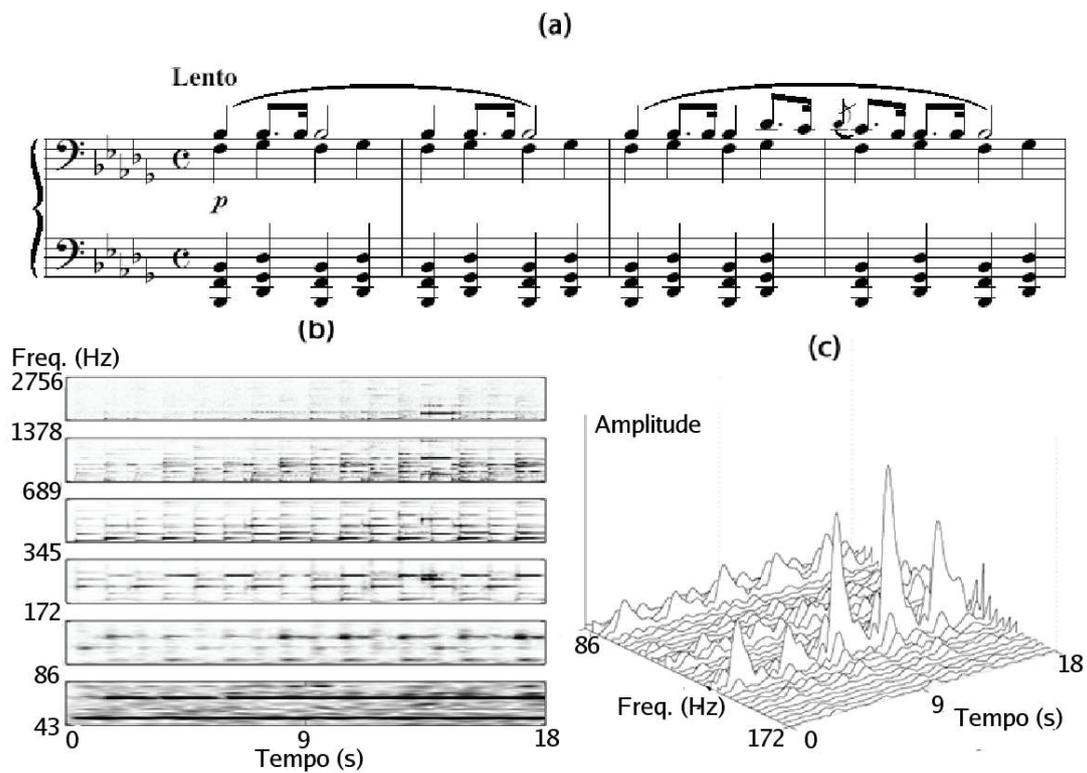


Figura 5: (a) Partitura de trecho da Marcha Fúnebre usado no Exemplo 4. (b) Sonograma das oitavas de 2 a 7 no Exemplo 4. (c) Dinâmica temporal da oitava 3 no Exemplo 4.

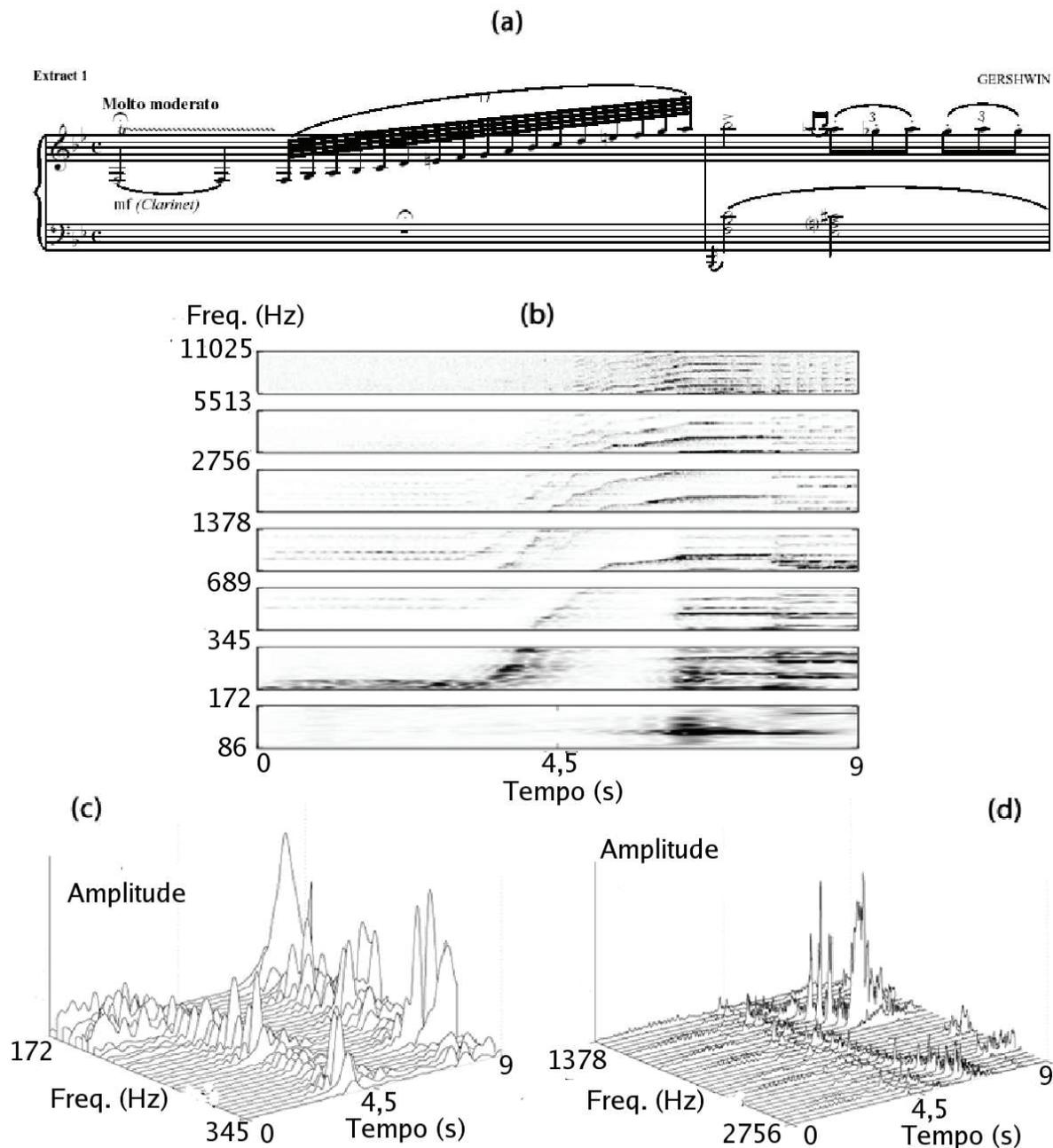


Figura 6: (a) Partitura de trecho de “Rhapsody in Blue” usado no Exemplo 5. (b) Sonograma das oitavas de 5 a 10 no Exemplo 5. (c) Dinâmica temporal da oitava 4 no trinado do Exemplo 5. (d) Dinâmica temporal da oitava 7 no *glissando* do Exemplo 5.

completo de TAM incluirá uma heurística sofisticada para esse fim.

Pode-se notar que o pior desempenho do sistema se deu nas frequências mais baixas, onde ainda não há um bom compromisso entre as resoluções temporais e frequenciais. Também como já observado qualitativamente na experiência de altas frequências, tal região do espectro apresentou o melhor desempenho. O desempenho médio na experiência de passagens lentas se deveu à presença de acordes ricos e respectivas parciais, en-

quanto que as passagens rápidas tiveram sua detecção prejudicada devido ao fato de a resolução temporal não ter sido suficiente (a acentuada reverberação presente na gravação certamente contribuiu para problema).

Já que o mapeamento entre os canais do banco de filtros e as frequências das notas ainda está em um estágio inicial, foi admitida uma tolerância de 1 semitom e efetuada nova contagem. Percebe-se que, considerando tal tolerância, os indicadores de desempenho sobem, o que sugere que o mapeamento pode ser

Experiência	Percentual de acertos exatos	Percentual de acertos +/- 1 semitom
Frequências baixas	36,67 %	73,34 %
Frequências altas	62,90 %	80,64 %
Passagens rápidas	41,93 %	51,61 %
Passagens lentas	60 %	60 %
Trinado	100 %	100 %
Glissando	100%	100 %

Tabela 1: Percentuais de acerto de notas para as experiências realizadas.

melhorado reduzindo-se o número de canais por oitava. É importante lembrar que, como existem 32 canais mas apenas 12 semitons por oitava, há em média quase 3 canais por semitom. Formas mais elaboradas de mapeamento serão objeto das próximas etapas da pesquisa. Esse mapeamento, aliado a uma forma robusta de detecção de *onset* e *offset* das notas e a um sistema que permita discriminar automaticamente notas verdadeiras de parciais elevadas será o fundamento de um sistema completo para transcrição automática de música.

4 CONCLUSÃO

Neste artigo foi feita uma análise do desempenho do método de análise BQFFB como base de um sistema de transcrição musical automática (TMA). Ao sistema foram submetidos sinais musicais com características escolhidas para avaliar sua resposta a problemas específicos envolvidos em TMA. Foi possível verificar que o BQFFB apresenta um bom desempenho nas altas frequências, permitindo detectar com precisão o início e o fim das notas (a detecção automática de *onset* e *offset* será implementada no sistema completo de TAM). Contudo, também foi visto que o desempenho em baixas frequências fica prejudicado, em alguns casos específicos, dado que as estreitas bandas de passagem dos filtros com frequências centrais baixas não são capazes de comportar as largas bandas dos transitórios. Efeitos como o trinado e o *glissando* foram bem analisados, com resultados bastante positivos. Este estudo ilustra o funcionamento de um sistema TMA como um todo e serve como base para a aperfeiçoamento do sistema de TMA usando a BQFFB.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] J. C. Brown, "Calculation of a constant Q spectral transform," *Journal Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, 1991.
- [2] J. W. Cooley and J. W. Tukey, "An algorithm for the machine computation of complex Fourier series," *Mathematics of Computation*, vol. 19.
- [3] F. C. C. B. Diniz, I. Kothe, S. L. Netto and L. W. P. Biscainho, "High-Selectivity Filter Banks

for Spectral Analysis of Music Signals," *EURASIP Journal Advances on Signal Processing*, vol. 2007, pp. 1–12, paper ID: 94704, 2007.

- [4] F. C. C. B. Diniz, I. Kothe, L. W. P. Biscainho, and S. L. Netto, "A bounded- Q fast filter bank for audio signal analysis," in *Proc. IEEE International Telecommunications Symposium*, pp. 971–975, Fortaleza, Brazil, Sept. 2006.
- [5] B. Farhang-Boroujeny and Y. C. Lim, "A comment on the computational complexity of sliding FFT," *IEEE Trans. Circuits and Systems - II: Analog and Digital Signal Processing*, vol. 39, no. 12, pp. 875–876, 1992.
- [6] D. B. Graziosi, C. N. dos Santos, S. L. Netto, and L. W. P. Biscainho, "A constant- Q spectral transformation with improved frequency response," in *Proc. IEEE International Symposium on Circuits and Systems*, vol. V, pp. 544–547, Vancouver, Canada, May 2004.
- [7] K. L. Kashima and B. Mont-Reynaud, "The bounded- Q approach to time-varying spectral analysis," Stanford University, Dep. of Music, Tech. Rep. STAN-M-28, 1985.
- [8] A. Klapuri and M. Davy (Editores), *Signal Processing Methods for Music Transcription*, Nova Iorque, EUA: Springer-Verlag, 2006.
- [9] Y. C. Lim, "Frequency-response masking approach for the synthesis of sharp linear phase digital filters," *IEEE Trans. Circuits and Systems*, vol. 33, no. 4, pp. 357–364, 1986.
- [10] Y. C. Lim and B. Farhang-Boroujeny, "Fast filter bank (FFB)," *IEEE Trans. Circuits and Systems-II: Analog and Digital Signal Processing*, vol. 39, no. 5, pp. 316–318, 1992.
- [11] C. N. dos Santos, S. L. Netto, L. W. P. Biscainho, and D. B. Graziosi, "A modified constant- Q transform for audio signals," in *Proc. IEEE International Conference on Audio, Speech, and Signal Processing*, vol. 2, pp. 469–472, Montreal, Canada, May 2004.
- [12] P. P. Vaidyanathan, *Multirate Systems and Filter Banks* Upper Saddle River, USA: Prentice-Hall, 1992.
- [13] D. William and E. Brown, *Theoretical Foundations of Music*, Belmont, USA: Wadsworth, 1978.