

Sampler de áudio em tempo real baseado no reconhecimento de sons musicais¹

*Thomas Koch²
Daniel Faistauer³*

Resumo

Atualmente, os semicondutores são elementos imprescindíveis com aplicações que vão desde uma simples calculadora a coisas mais sofisticadas como um MP3 player. O som tem características que só podem ser estudadas com o advento dos semicondutores e mais especificamente dos processadores de sinais digitais denominados DSPs. A criação destes processadores de sinais desencadeou a era da comunicação digital, sendo que suas contribuições para a música têm proporcionado a criação de novos equipamentos como os samplers digitais, os filtros digitais e os equalizadores digitais. O estudo dos samplers de música tem sido uma constante, na busca de torná-los uma ferramenta adequada para substituir um montante de instrumentos. Este trabalho aborda a elaboração de um sampler em tempo real, baseado na amostragem de sinais audíveis e na sua conversão em áudio de um instrumento qualquer em tempo real.

Palavras-chave: Processamento de áudio; Tempo real; Reconhecimento.

Abstract

Semiconductors are essential elements whose applications can be from a simple calculator until sophisticated things like a MP3 player. The sound has characteristics that only can be studied with the advent of the semiconductors and more specifically the signal processors called DSPs. The creation of these signal processors started the age of the digital communication and their contributions have provided to the development of new equipment such as music samplers, digital filters, and equalizers. The studies about music samplers have been a constant in the pursuit of making them an appropriated tool to substitute many different instruments. This paper aims the development of a real-time sampler, based on the audible signals and its conversion in sound of any other instrument.

Keywords: Signal processing; Real time; Recognition.

1 Introdução

Com o advento dos semicondutores e, mais especificamente, dos processadores de sinais digitais, tornou-se possível criar equipamentos sofisticados e precisos na reprodução

de sons. Os instrumentos musicais eletrônicos da atualidade são tão precisos que quase não existe diferença entre um sinal proveniente de um instrumento musical real e sua réplica vinda de um instrumento eletrônico. Contudo esses instrumentos eletrônicos não têm a

¹ Artigo elaborado a partir da monografia apresentada como requisito parcial para a obtenção do título de Engenheiro Eletricista na Universidade do Vale do Rio dos Sinos – UNISINOS.

² Engenheiro Eletricista – UNISINOS <superleiter@yahoo.com.br>.

³ Professor Mestre da Universidade do Vale do Rio dos Sinos – UNISINOS. Av. Unisinos, 950, Fiação, São Leopoldo, RS, Brasil, 93022-000. <danielf@unisinos.br>

Recebido em 01/2009 e aceito em 03/2009.

flexibilidade de sintetizar sons com base em outros, pois foram construídos com o propósito de reconhecerem códigos específicos para sua reprodução.

Atualmente, o reconhecimento de sons musicais por processos computacionais pode tornar-se uma realidade, devido, principalmente, à criação dos equipamentos específicos para processamentos de sinais digitais. Contudo, os processos envolvidos neste tipo de reconhecimento têm grande complexidade, já que fazem uso das mais diversas ferramentas matemáticas desenvolvidas.

Além disso, as aplicações em tempo real também se tornaram possíveis com o desenvolvimento dos processadores de sinais digitais. Para exemplificar a importância das atividades em tempo real, pode ser citado o caso dos veículos controlados por computação embarcada que devem ser monitorados e processados a cada instante para que decisões importantes sejam tomadas.

Na música, aplicações em tempo real também podem ser de grande utilidade para profissionais do ramo ou até mesmo apreciadores e *hobbistas*. Portanto, a partir dos processadores de sinais digitais e do estudo dos sinais sonoros musicais, podem ser desenvolvidos equipamentos e sistemas para aplicações musicais em tempo real.

Alguns equipamentos eletrônicos dedicados à música trabalham com códigos específicos para a identificação das notas a serem tocadas como, por exemplo, o formato *MIDI*. Um exemplo deste tipo de equipamento são os teclados eletrônicos e os *samplers*. Todavia, equipamentos eletrônicos capazes de reconhecer notas musicais ainda não existem no mercado devido à sua complexidade e custos altos de desenvolvimento. Outra lacuna no mercado é a falta de dispositivos eletrônicos capazes de reconhecer tais sons musicais e realizar a transposição deste instrumento para outro qualquer. Isto possibilitaria o aproveitamento do talento de um músico num determinado instrumento para gerar sons dos mais variados timbres.

Este trabalho tem por objetivo criar uma ferramenta que seja capaz de trabalhar com

o reconhecimento dos sinais de entrada provenientes de um instrumento musical (original) para ser posteriormente processado e reproduzido. Em linhas gerais, busca-se implementar um *sampler* em tempo real que será capaz de reconhecer notas aplicadas à sua entrada e à posterior reprodução de som de característica sonora diferente.

Para o desenvolvimento será utilizado um DSP (Processador de Sinais Digitais) da família C6000 da *Texas Instruments*. Além deste dispositivo, o sistema será composto de algoritmos que serão responsáveis pelo reconhecimento dos sinais musicais de entrada. Posteriormente, será sintetizado o som de um instrumento de outra natureza escolhido pelo usuário a partir das informações extraídas do sinal de entrada. Por fim, o som sintetizado será aplicado à saída do dispositivo para a apreciação do usuário, sendo que todo o processo visa ser executado em tempo real.

2 Desenvolvimento da aplicação

Sabe-se que um instrumento não mantém suas características sonoras constantes ao longo de todo seu espectro de frequências devido a aspectos intrínsecos como o material em que foi construído, a forma como é tocado e a sua geometria (HENRIQUE, 2002). Portanto seria impossível a replicação fiel de um instrumento musical a partir de sua digitalização a começar pela limitação da amostragem e principalmente pelo fato de que o comportamento dos instrumentos é não-linear e variante. Tendo isso em vista, os fabricantes de instrumentos musicais eletrônicos como *samplers* e teclados digitais procuraram reproduzir os sons com o máximo de fidelidade, inserindo também as grandezas que têm natureza não-linear de maneira intercalada, tornando-os mais interessantes. Estes equipamentos utilizam um banco de dados com informações de instrumentos os quais foram destinados a reproduzir, sendo que estas informações nada mais são do que grandes conjuntos de timbres pré-armazenados para posterior processamento.

Para a criação de novos sons, faz-se necessária a transposição dos timbres armazenados para as frequências desejadas. Isto pode ser realizado aplicando a propriedade de *frequency-shifting* de uma *DFT* (Transformada de Fourier Discreta) no espectro sonoro relativo ao timbre do som previamente armazenado (CARTINHO, 2000). A transposição é conhecida no ramo da música como sendo o efeito de *Pitch-Shift* que, na verdade, é apenas um deslocamento de todo espectro para outra frequência maior ou menor. De maneira geral, para reconstruir um som em frequência diferente da original, é tomado um período amostrado qualquer, no tempo e, logo após, é aplicado à transposição sobre seu espectro de frequências. Depois de realizada essa transposição, o novo sinal no domínio do tempo é obtido e posto em *loop* para se tornar um sinal periódico e audível.

Além da transposição do som em frequências, os samplers inserem alguns efeitos como *reverb*, *vibrato* e *sustain* que transmitem uma sensação de realidade para o ouvinte. Lembrando que, dessa maneira, a integridade do instrumento a ser reproduzido vai ficando comprometida, pois, conforme vai se aumentando a frequência de reprodução do instrumento, vão sendo anexadas ou retiradas informações não exatas ao sinal sintetizado.

Em linhas gerais, o desenvolvimento para sintetização dos sons seguirá o mesmo roteiro de *samplers* tradicionais, acrescido à função de reconhecimento da frequência do som tocado e o processamento em tempo real.

No diagrama geral (figura 1), podem ser verificadas as etapas principais do processo de funcionamento do *sampler* em tempo real proposto.

Ao iniciar o sistema, é necessário informar os seguintes parâmetros de utilização:

- seleção do número de amostragens para cada *DFT*;
- seleção da frequência de amostragem;
- seleção do instrumento a ser sintetizado.

O primeiro parâmetro define a quantidade de amostras que se serão utilizadas para calcular a *DFT*. Em tese, para o bom funcionamento e sem atrasos de propagação significativos, esse valor deve estar entre 32 e 2048 pontos sendo sempre potências de dois. O segundo parâmetro define a frequência de amostragem para a operação em tempo real, entre as taxas de 44,1 kHz e 96 kHz. Essas frequências altas foram estipuladas tendo em vista o teorema da amostragem, considerando que o ser humano só pode ouvir até a frequência de 20 kHz. Por fim, o terceiro parâmetro de configuração seleciona o instrumento a ser reproduzido entre os que compõem a biblioteca. Na biblioteca devem estar representados os dados de amplitude, frequência e timbre obtidos previamente da amostra de cada instrumento. Estas informações se encontram na forma de pares complexos obtidos de uma *DFT* realizada em outro dispositivo.

No momento em que o DSP é inicializado, via comunicação USB, é solicitada pelo mesmo uma cópia da biblioteca que irá compor a estrutura de reconstrução dos sinais resultantes. As bibliotecas para teste foram criadas a partir de uma rotina *DFT* desenvolvida em Turbo C++. O programa elaborado se destina a montar uma estrutura de pares complexos resultantes de uma *DFT* para que seja possível a sua execução no núcleo do DSP. A esses pares de números complexos será aplicada posteriormente uma *IDFT*

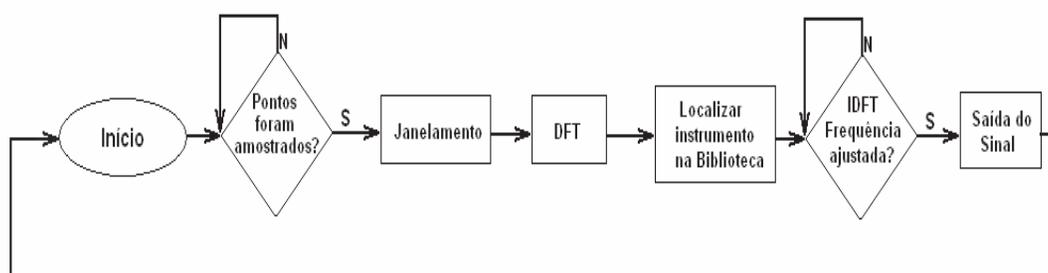


Figura 1 – Diagrama geral de funcionamento do equipamento

(Transformada Inversa de Fourier Discreta), para se retornar ao domínio do tempo com a frequência devidamente alterada.

O programa desenvolvido para a elaboração da biblioteca de instrumentos processa uma DFT radix-2 do som de determinado instrumento, com a frequência padrão definida de 130,81 Hz. Esta frequência equivale a uma nota Dó da segunda oitava dos sons audíveis (aqui chamada de Dó₂). Com esta nota é possível realizar a sintetização satisfatória até duas oitavas acima ou abaixo da frequência original até mesmo para instrumentos como o violino que tem um espectro de difícil reprodução.

2.1 Método de janelamento

Para o processamento dos sinais na entrada, faz-se necessário o condicionamento do sinal através do método de janelamento. O janelamento é uma técnica utilizada para minimizar os efeitos de perda espectral do sinal amostrado (OPPENHEIM, 1989). Ao se trabalhar com amostragem de sinais, ou seja, sinais discretos, estes têm duração finita, o que ocasiona uma perda espectral. Isso ocorre porque o sinal amostrado não possui toda informação que deveria representar. Neste momento, surge o método de janelamento que

proporciona um ganho espectral nos sinais amostrados como é o caso da janela *Hanning*, da janela *Hamming* e da triangular.

A janela *Hanning* tem uma característica que a coloca a frente de todas as demais pela sua facilidade de execução e, principalmente, pelos poucos cálculos envolvidos, o que é de grande relevância para o sistema (ANDRADE; SOARES, 2006).

Sinais sem e com janelamento podem ser observados através das diferenças na figura 2 e na figura 3, tanto no domínio do tempo quanto no domínio da frequência. Em se tratando de uma fonte de sinal composta por poucos sinais puramente senoidais, esperava-se que o espectro de frequência apontasse apenas as frequências de interesse, como definido no gerador de sinais, porém os valores de frequência, na figura 2, não são condizentes com os valores esperados visto que aparece maior número de componentes de frequência além das fundamentais. Este efeito trata-se de um equívoco conhecido como efeito *leakage* (invasão), pois, como citado anteriormente, o sinal foi truncado o que provoca uma perda de informação ou até a incorporação de frequências inválidas, contribuindo consideravelmente para uma resposta equivocada.

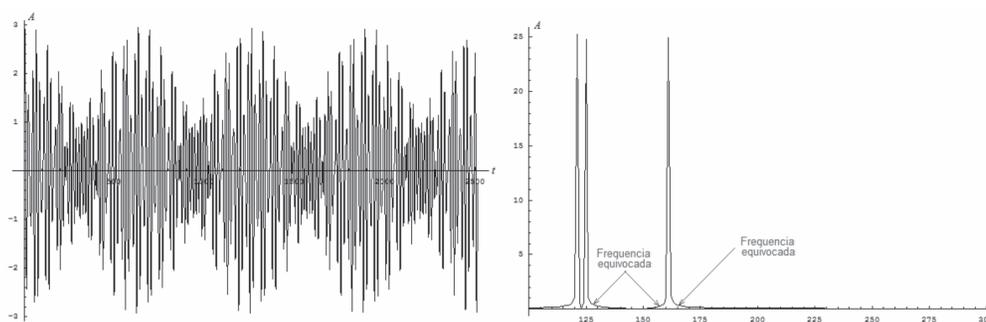


Figura 2 – Sinal amostrado sem aplicação do janelamento e o seu respectivo espectro

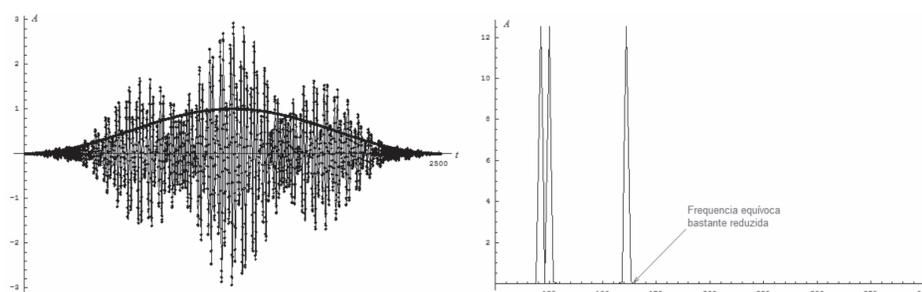


Figura 3 – Aplicação do janelamento Hanning no domínio do tempo e o espectro resultante

Por outro lado, a figura 3 mostra que através de janelamento *Hanning* é possível diminuir esse fenômeno de distorção de informação, fazendo com que essas informações sejam até mesmo eliminadas quase que por completo na resposta final.

2.2 Algoritmo de detecção

Além do janelamento, outro método interessante é a utilização da média ponderada para determinação exata das frequências, tornando o efeito de *leakage* insignificante. A média ponderada entre as frequências resultantes da DFT mostra excelentes resultados neste tipo de aplicação, pois ela é capaz de mesclar as condições de quantidade de energia contida no espectro juntamente com a frequência contida neste espectro. Na figura 4, há um exemplo no qual foi amostrado um sinal de 1720 Hz, com 512 pontos. A interpretação para o processador (DSP) será de que existem mais componentes dentro do sinal, além deste de 1720 Hz e que estão muito próximas, portanto não está definida a frequência a ser considerada. Neste caso, seria necessário uma DFT com, no mínimo, 2048 pontos para que o processador seja capaz de distingui-los. Portanto ao se aplicar apenas um algoritmo de detecção de magnitude, percebe-se que será realizada uma medição equivocada, uma vez que não se trata, necessariamente, da frequência contida na maior magnitude que caracteriza 1808 Hz. Por outro lado, ao aplicar o algoritmo de detecção em combinação com a média ponderada, onde são utilizadas as duas medições mais relevantes junto à maior amplitude encontrada, o resultado torna-se bem mais preciso. Com a média ponderada aplicada, o resultado obtido, 1722 Hz, representa um erro de 0,116 % na frequência identificada.

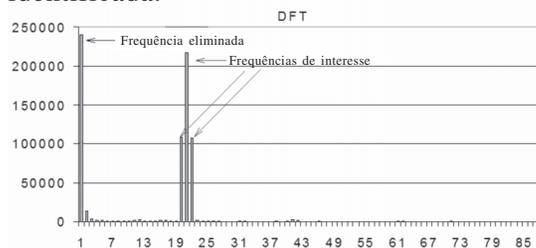


Figura 4 – Pontos assinalados indicando as máximas amplitudes

O valor indicado na figura 4 como sendo “frequência eliminada” é uma componente que aparece em determinadas transformações que envolvem sinais com níveis DC, ou seja, valores que estão com um deslocamento para cima ou para baixo em relação ao eixo do tempo (t). Como neste trabalho estamos apenas tratando de sinais puramente AC, então esta componente não deverá aparecer ou será desconsiderada.

2.3 Biblioteca de instrumentos

Ao ser tomado um sinal, deve-se observar o tamanho da janela a ser utilizada para que as frequências de interesse não sejam prejudicadas ou, até mesmo, não sejam identificadas pelo processador de sinais. Com base nisso, imaginemos um sinal proveniente de um piano, que tem complexidade tão alta que nem mesmo amostragens da ordem de 2048 pontos a uma taxa de 44,1 kHz consigam amostrar o sinal com fidelidade. Um som vindo de um piano tem tamanha complexidade que ao ser reproduzida uma nota e se elevar meio-tom o som tem espectro completamente diferente. Isso se deve a muitos fatores que tornam este instrumento único quanto a sua sonoridade.

As principais características que favorecem os diferentes tipos de som que o piano reproduz são necessariamente suas cordas, o martelo e principalmente a caixa acústica, sendo que é notada, também, diferença na reprodução de um som com a caixa aberta e fechada. A figura 5 exibe a complexidade presente numa nota executada em Mi_0 , ($f=41$ Hz) que apresenta um rico espectro de frequência.

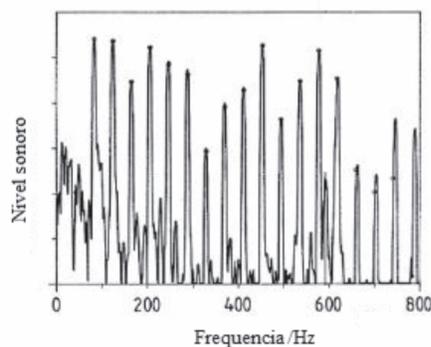


Fig. 05 – Sinal proveniente de um piano em Mi_0 (CONKLIN, 1990)

Essas características espectrais fazem com que o piano seja descartado como instrumento gerador de sinal para o processador de sinais, uma vez que, com o algoritmo de detecção adotado, é quase impossível aplicar tal sinal nessa situação, fazendo-se necessária a aplicação de outros meios computacionais para tal. Portanto, os instrumentos utilizados como fontes de sinal sonoro trabalham com frequências mais altas e bem definidas onde o efeito de ressonância é bastante minimizado como, por exemplo, uma flauta, um órgão e uma guitarra, todos apresentam espectros de sinais comportados e as amplitudes da componente fundamental e suas vizinhanças bastante evidenciadas.

Para o estudo dos instrumentos mencionados, levou-se em conta a relação de nível de distorção ao longo do tempo, conforme fossem aumentadas as frequências dos instrumentos que compõem a biblioteca. A escolha desta frequência foi estabelecida pelo fato de que em duas oitavas acima ou duas oitavas abaixo de 130,81 Hz, as notas apresentam maior fidelidade de reprodução, e o timbre característico de cada instrumento é mais notável.

A frequência de operação ideal, ou simplesmente razoável para a aplicação de 130,81 Hz, foi obtida com base na amostragem sucessiva de várias frequências do som o qual se queria tomar. Este som era, então, gravado na biblioteca na forma de 1024 pontos de DFT de pares complexos. Para efetuar a busca e classificação das frequências que seriam guardadas na memória, constituindo a biblioteca de instrumentos, foi utilizado um sistema simples que media, além das regiões de maior amplitude de frequência, os sinais de amplitude próximos a este, fazendo com que o algoritmo detecte se o sinal em questão contém realmente a frequência de interesse (som tocado) ou não (figura 4).

2.4 Sintetização do sinal sonoro

O método de expansão ou compressão do sinal é uma maneira simples e eficaz de obter as demais frequências do som sem muita

complexidade e gastos computacionais em relação à ocupação de memória extra. Existem vários métodos de expansão e compressão, destacando-se o de deslocamento em frequência, com base na Transformada de Fourier, e o método de eliminação de pontos ou acréscimo de pontos.

A diferença entre os dois métodos está justamente na qualidade do som resultante. Ou seja, se ao partir de um som, com frequência de 130,81 Hz, for obtida uma frequência de 200 Hz, praticamente teria que se reduzir à metade a quantidade de pontos amostrados, aplicando o método de corte de pontos, o que diminuiria drasticamente a qualidade do sinal, e por questão de fidelidade, não é adequado para a aplicação deste trabalho. Portanto, para a compressão do sinal, que implica diretamente o aumento da frequência, é necessário fazer o *frequency-shift* da DFT para a frequência desejada. Em linhas gerais, o espectro é totalmente deslocado, tendo o cuidado de localizar a frequência fundamental, para que esta assuma o valor desejado, como apresentado nas figuras 6 e 7.

Porém, este método tem problemas quanto ao deslocamento do espectro, pois se imaginarmos, por exemplo, um som de violão, que depende muito da caixa acústica, apareceriam problemas quanto à ressonância conforme fossem sendo aumentadas as frequências. Em suma, a reprodução de uma nota diferente, com base numa nota qualquer, não consiste simplesmente em deslocar o espectro, mas deve-se ter o cuidado para não serem levadas frequências provenientes de ressonâncias em outras frequências.

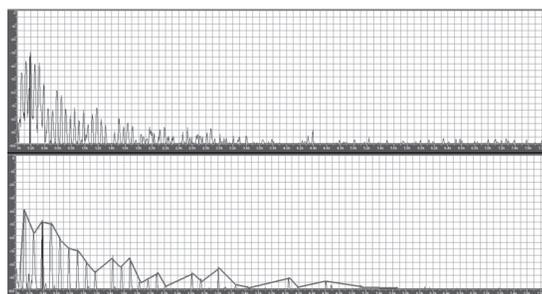


Figura 6 – Espectro de um violão em D_2 e D_3 , respectivamente de um instrumento real indicando suas respectivas fundamentais

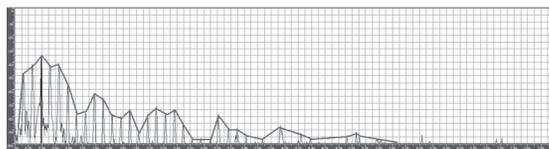


Figura 7 – Dó₃ obtido a partir de um Dó₂ pelo método do *Frequency-Shift*

Ao observar a figura 6, nota-se que quase não existe relação entre o espectro das duas notas, o que prova de maneira simples que o resultado obtido na figura 7 é um equívoco, se for feita uma análise criteriosa, pois nela, foram simplesmente excursionadas as componentes referentes à nota Dó₂ até a frequência correspondente à frequência de um Dó₃. A presença de elementos de ressonância marcantes, como o caso das caixas acústicas do violão e piano, não é de fácil compreensão nem de fácil modelagem matemática, portanto, devem ser feitos testes mais criteriosos em relação a seu comportamento ao longo do espectro de frequência antes de utilizá-los para a biblioteca de instrumentos desta aplicação.

Portanto, os instrumentos que compõem a biblioteca devem permitir que o deslocamento em frequência seja aplicado sem grandes distorções no espectro, devido ao fato de as componentes referentes à ressonância serem minimizadas e até imperceptíveis, como pode ser observado nos espectros de uma flauta bixel mostrados na figura 8 e na figura 9.

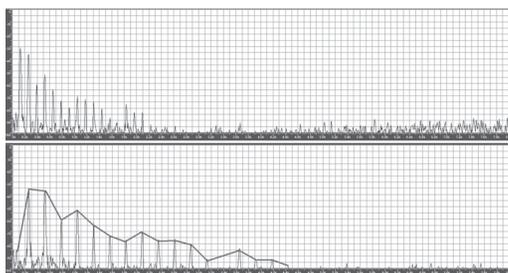


Figura 8 - Espectro de uma flauta bixel em Dó₂ e Dó₃ respectivamente de um instrumento real

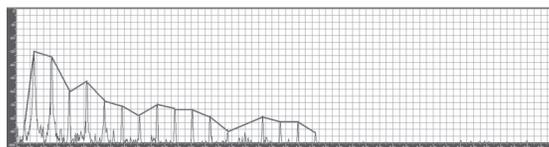


Figura 9 – Dó₃ obtido a partir de um Dó₂ pelo método do *Frequency-Shift*

O caso da flauta bixel é um bom exemplo para mostrar que não há alterações significativas ao efetuar-se o deslocamento no espectro. Observando as figuras 8 e 9, verifica-se que a diferença marcante é apenas na amplitude. Quanto ao som reproduzido, pode-se afirmar que um ser humano comum não seria capaz de perceber diferença, pois as frequências que realmente provocam alterações no timbre são quase imperceptíveis ao ouvido humano.

Entretanto a rotina normalmente utilizada para a transformada inversa não oferece a possibilidade de mudança de frequência. Para isso ser possível, o conjunto das equações tradicionais é levemente modificado para permitir o deslocamento em frequência.

Sabe-se que:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] \cdot W_N^{-kn} \quad (1)$$

(OPPENHEIM; SCHAFER, 1989)

então, supondo que se tenha uma DFT de 8 pontos armazenada, para se obter a posição x[3] seria necessário realizar o seguinte procedimento :

$$x[3] = \frac{1}{8} \sum_{k=0}^7 X[k] \cdot W_8^{-kn}$$

$$x[3] = \frac{1}{8} \left(X[0] \cdot e^0 + X[1] \cdot e^{\frac{\pi}{2}i} + \dots + X[7] \cdot e^{\frac{7\pi}{2}i} \right)$$

(OPPENHEIM; SCHAFER, 1989).

Por fim, para se modificar a frequência, é necessário apenas utilizar a variável com índice fr de acordo com a relação:

$$x[3] = \frac{1}{8} \sum_{k=0}^7 X[k] \cdot W_8^{-kn(fr)}$$

$$x[3] = \frac{1}{8} \left(X[0] \cdot e^{0(fr)} + X[1] \cdot e^{\frac{\pi}{2}i(fr)} + \dots + X[7] \cdot e^{\frac{7\pi}{2}i(fr)} \right)$$

(OPPENHEIM; SCHAFER, 1989).

Sendo que fr é o fator multiplicador para obter a frequência desejada a partir da nota padrão Dó₂.

3 Análise dos resultados

Com o objetivo de identificar o número de amostras adequado para a DFT e a frequência de amostragem ideal, foram realizados testes de identificação de frequência com a variação dos parâmetros investigados. Na figura 10, estão indicados os resultados obtidos para o aumento sucessivo do número de amostragens nas frequências de 44,1 kHz e de 96 kHz. Observa-se claramente que, ao serem tomadas 512 amostras, o sinal apresenta a tendência de se igualar à frequência real. Observa-se, também, que para frequências baixas uma taxa de amostragem muito alta não é viável porque se utilizam menos pontos dentro de um período, provocando distorções na identificação.

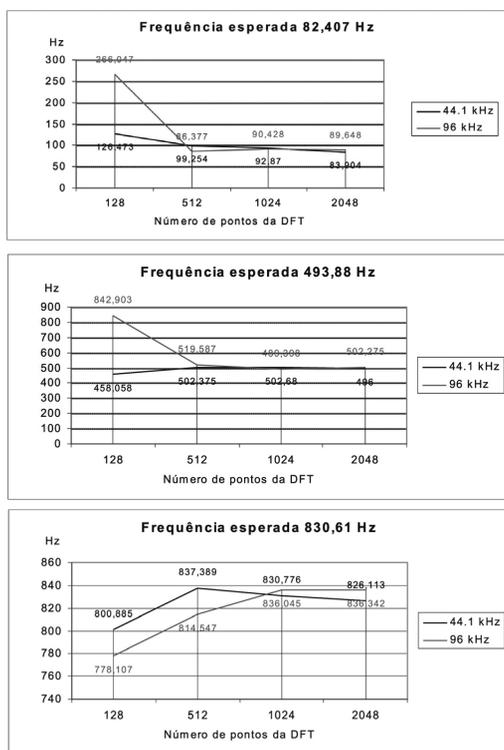


Figura 10 – Frequência identificada em função do número de amostras usadas para a DFT (em 82,407 Hz, 493,88 Hz e 830,61 Hz)

Portanto, a utilização de 512 pontos para a realização da DFT, com frequência de amostragem de 44,1 kHz mostra-se suficiente para identificar boa parte dos sinais sonoros audíveis. Lembrando que a identificação de sinais sonoros de baixa frequência é mais compli-

cada devido aos longos comprimentos dos períodos relacionados, a utilização de DFT de 512 pontos torna-se suficiente minimizando o tempo para aquisição dos dados e, conseqüentemente, melhorando a performance do equipamento.

A figura 11 representa o período de som de violino obtido tomando como entrada uma nota executada em Ré₂ pela guitarra. Após este processamento no DSP, foram implementadas repetições deste período de amortecimento nas amplitudes do mesmo.

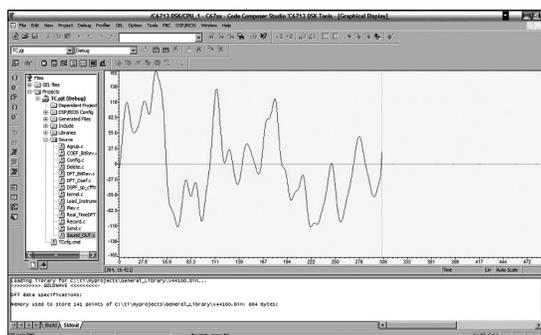


Figura 11 – Sinal de violino em Ré₂ obtido a partir de um Ré₂ fornecido por uma guitarra

3.1 Aprimoramentos do sinal sonoro sintetizado

Na figura 12, é observado um sinal de flauta originado a partir de uma guitarra executando Ré₄, neste sinal foi inserida uma modulação chamada de *vibrato*, já mencionada anteriormente. O *vibrato* é um efeito intencional conhecido pelo seu traço característico de modulação em frequência e amplitude, introduzido pelo músico em momentos desejados.

Por exemplo, o som da flauta tem características muito particulares que estão relacionadas ao músico que irá tocá-lo, como é o caso da maioria dos instrumentos. Mas este merece atenção especial, pois o envelope tem certas dificuldades de ser implementado pelo simples fato de o transitório de ataque depender, exclusivamente, do sopro do músico e, portanto, não é uma atividade trivial programá-lo digitalmente. Para manter uma fidelidade com este instrumento, fez-se o que se chama de *fade-in* e *fade-out* que é justamente uma entrada e uma saída suave através de exponenciais.

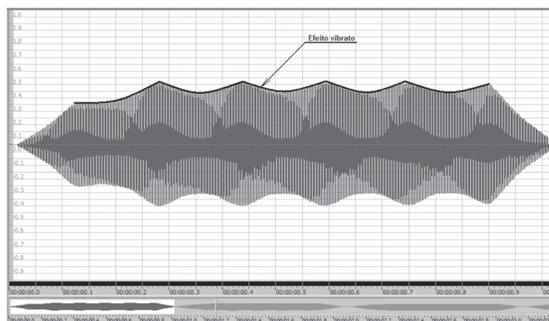


Figura 12 – Som de flauta bisel em C_4 , submetido ao envelope evidenciando o efeito *vibrato*

Os sons sintetizados de órgão também apresentaram algumas dificuldades no transitório de ataque por este ser uma rampa de entrada muito íngreme, além do efeito de entrada ser semelhante ao de um martelo de piano. Contudo, o som obtido apresentava boa qualidade e clareza na definição e identificação do instrumento tocado. Na figura 13, pode ser observado o resultado do som sintetizado de órgão com seu respectivo envelope.

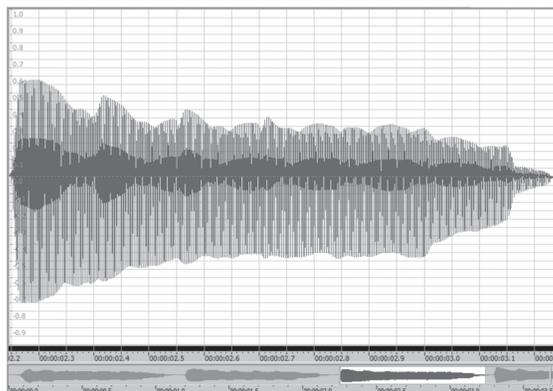


Figura 13 – Som de órgão em Mi_4 , submetido ao envelope

Outra aplicação testada com o algoritmo de detecção, com o fim de experimentação, foi a identificação da frequência fundamental contida na voz de uma pessoa emitindo algum sinal sonoro, sendo o sinal da saída um instrumento da biblioteca com a mesma frequência da voz.

Em suma, uma pessoa com uma voz afinada seria capaz de cantarolar ao microfone e ouvir sua voz modificando as notas da saída do equipamento. Para validação deste teste de reconhecimento das notas cantadas,

o algoritmo foi submetido a um sinal (voz) de aproximadamente 600 Hz e convertido num violino na saída como mostrado nas figuras 14 e 15.

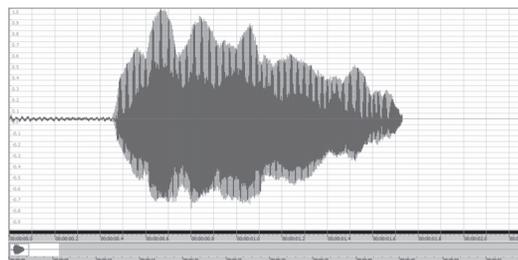


Figura 14 – Som de “A” sendo cantarolado

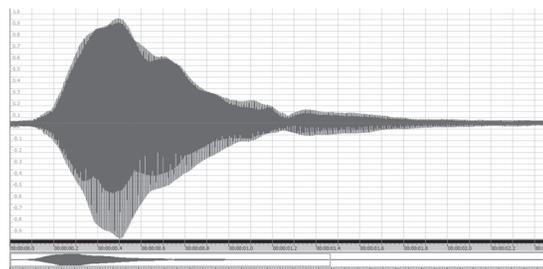


Figura 15 – Som de violino gerado a partir da voz de 600 Hz

Nota-se que o som reproduzido segue o transitório de ataque e o transitório de decaimento, conforme a voz vai se atenuando. Este experimento foi realizado num ambiente desprovido de ruído para que o algoritmo não cometesse o equívoco de apresentar uma nota na saída de um som que não foi emitido.

3.2 Discussões

Na figura 16, é observado um fenômeno de parada para atendimento de interrupção que ocorre ao longo da reprodução do sinal processado e que é indesejável na execução. Esse problema pode ser facilmente solucionado através de um condicionamento do sinal através de outro processador e outro conjunto de memória (*buffer*) os quais irão sustentar o tempo que o DSP realiza outras atividades como o cálculo da DFT.

Este efeito decorre do processamento das atividades executadas pelo processador como, por exemplo, o tempo que se leva para acessar, processar, gravar, etc. O caso mais crítico ocorre no acesso do DSP ao CODEC

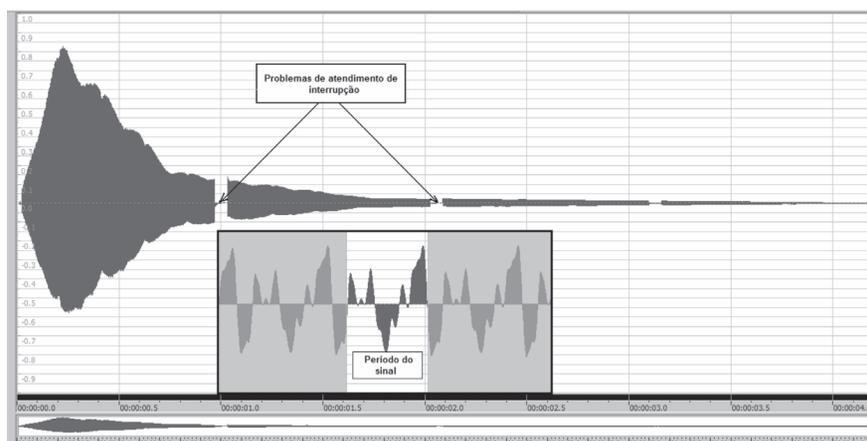


Figura 16 – Sinal de violino sendo submetido ao envelope e o problema do atendimento a interrupções.

para aquisição, processamento e reprodução no qual este periférico fica sem um *buffer* de dados, provocando uma descontinuidade no som que dura cerca de 62 ms. Este é exatamente o tempo que o DSP leva para a aquisição, processamento e devolução de 2048 valores ao CODEC.

4. Considerações finais

A busca de automação, no que se diz respeito à música, é uma atividade difícil e exige conhecimentos que, muitas vezes, estão longe do nosso cotidiano como, por exemplo, análises espectrais e de timbre.

A tecnologia disponível para tratamento de sinais evolui em passos largos, pois, como dito anteriormente, há a necessidade constante de aperfeiçoamentos das máquinas e equipamentos em geral. O DSP é um exemplo destas novas tecnologias que, muitas vezes, passam despercebidas, uma vez que já está presente em muitos itens do cotidiano como celulares, câmeras digitais e tecnologias embarcadas em veículos, por exemplo.

Os resultados obtidos com o desenvolvimento deste trabalho foram bastante satisfatórios no que se refere ao funcionamento do algoritmo de detecção de notas, devido ao fato de que se obteve fidelidade e estabilidade na identificação das notas musicais tocadas através de uma guitarra elétrica. Além dos testes realizados com os sons de guitarra, também foram testados sinais provenientes

de um ser humano como, por exemplo, o solfejo de notas musicais, apresentando os mesmos resultados.

No processo de sintetização dos sons a partir da identificação da nota tocada, pode ser constatado que os sons gerados apresentavam razoável semelhança com os instrumentos reais. Isto foi possível, uma vez que foram reproduzidos os períodos dos sinais comprimidos e expandidos a partir do deslocamento espectral realizado no sinal original armazenado na biblioteca de instrumentos. Além disso, o sinal foi submetido a um processo de modulação a partir da informação do envelope característico do som do instrumento a ser reproduzido.

Considerando as constatações apresentadas, pode-se afirmar que estão distantes os limites das aplicações de processamento de sinais em tempo real. Neste contexto, pode ser citado o auxílio a deficientes visuais, a segurança em geral ou até mesmo uma próxima geração de equipamentos musicais eletrônicos como os *samplers* em tempo real baseados em reconhecimento de sinais musicais diversos.

Por fim, muitos pontos podem ser melhorados para que o sistema proposto possa ser utilizado em aplicações profissionais. A seguir, apresentam-se sugestões de trabalhos futuros e aperfeiçoamentos para o sistema desenvolvido:

- Aprimoramento das informações armazenadas no que diz respeito ao instrumento

musical a ser sintetizado e, conseqüentemente, maior fidelidade no som sintetizado: Os sons sintetizados até então apresentam o mínimo de informação necessária para a reprodução do instrumento desejado. Como sugestão para o melhoramento deste trabalho, deve ser feito um estudo mais apurado através da aquisição de informações espectrais e das características do envelope de diversas notas musicais (razoavelmente distanciadas em frequência) do instrumento desejado. Os resultados deste estudo servirão para que o som sintetizado se aproxime mais do instrumento real.

- Eliminação da interrupção da reprodução do som durante a aquisição de dados para reconhecimento da nota tocada: as interrupções de reprodução decorrem do fato de que o componente conversor A/D é o mesmo que realiza a conversão D/A. Por consequência, no instante em que se está digitalizando dados, é interrompida a reprodução do som sintetizado. Para resolver este problema, pode-se utilizar um kit de desenvolvimento com conversores A/D e D/A distintos ou lançar mão de outro dispositivo que irá somente reproduzir os sinais sintetizados através do DSP.

- Diminuição do tempo entre aquisição de dados para reconhecimento de nota: devido às interrupções de reprodução de sons, o tempo entre as aquisições de dados é relativamente longo. Ao realizar a melhoria da segunda sugestão, também seria possível fazer a digitalização constante dos sinais de entrada em bloco de dados, conforme a configuração

do *software* desenvolvido, no caso otimizado, 2048 dados. Assim, no *sample rate* de 44,1 kHz, obteremos a informação da nota tocada em 46 ms mais o tempo necessário para o processamento dos dados que é de aproximadamente 62 ms. Através deste processo, o tempo de atraso máximo entre identificação de notas tocadas se reduziria de 1,06 s para aproximadamente 0,1 s.

5 Referências

ANDRADE, A.O; SOARES, A.B. Técnicas de janelamento de sinais. *In: SEMINÁRIO DOS ESTUDANTES DE ENGENHARIA ELÉTRICA DA UFU*, 2000, Uberlândia. **Anais...** Uberlândia: UFU, 2000.

CARTINHO, Jack. **Digital signal processing**. New Jersey: Prentice Hall, 2000. 398p.

H. A. CONKLIN, Jr. **Piano design factors - their influence on tone and acoustical performance**. Stockholm: Royal Swedish Acad. of Music, 1990.

HENRIQUE, Luís L. **Acústica Musical**. Lisboa: Fundação Calouste Gulbenkian, 2002. 1130p.

OPPENHEIM, Alan V.; SCHAFER, Ronald W. **Discrete-time signal processing**. New Jersey: Prentice Hall, 1989. 879p.

PAPOULIS, Athanasios. **The Fourier Integral and its Applications**. New York: McGraw-Hill, 1962. 318p.

