

IDENTIFICAÇÃO EXPERIMENTAL DE NOTAS MUSICAIS

Sanderson Manoel da Conceição¹

smconceicao@aluno.feis.unesp.br

Douglas Domingues Bueno²

ddbueno@dem.feis.unesp.br

Vicente Lopes Junior²

vicente@dem.feis.unesp.br

GMSINT – Grupo de Materiais e Sistemas Inteligentes, Departamento de Matemática¹, Departamento de Engenharia Mecânica², Faculdade de Engenharia de Ilha Solteira, UNESP, Av. Brasil n.º 56, CEP 15385000, Ilha Solteira, SP, <http://www.dem.feis.unesp.br/gmsint>.

Resumo: A capacidade de acumular conhecimento é uma das mais importantes habilidades do homem. Esta capacidade permite o desenvolvimento tecnológico, avanços na área médica, entre outros. Embora existam fontes ilimitadas de conhecimento, alguns assuntos são de grande complexidade de aprendizado, pelo menos para parte das pessoas. O reconhecimento do som, através do ouvido humano, é fundamental no processo de aprendizagem de um instrumento musical. Assim, simplificar este aprendizado é o interesse de muitos, incluindo das indústrias deste setor. Em geral, para um estudante em fase inicial, reconhecer as notas musicais emitidas por um instrumento é tarefa bastante complexa. Assim, o desenvolvimento de um sistema de aquisição, compacto e de baixo custo, que avalie um som e indique a nota emitida é algo consideravelmente interessante. Neste contexto, neste trabalho é apresentado o uso da densidade espectral de potência para classificar algumas notas musicais de um violino. A aquisição dos sinais foi realizada utilizando o *software* Matlab[®] e a placa de som de um micro computador com processador AMD Sempron 2800.

Palavras-chave: *identificação experimental, notas musicais, densidade espectral de potência.*

1. INTRODUÇÃO

A música constitui-se de uma sucessão de sons e silêncio organizada ao longo do tempo. É considerada por diversos autores como uma prática cultural e humana. Atualmente não se conhece nenhuma civilização ou agrupamento que não possua manifestações musicais próprias. Embora nem sempre seja feita com esse objetivo, a música pode ser considerada como uma forma de arte, considerada por muitos como sua principal função. Também pode ter diversas outras utilidades, tais como a militar ou educacional.

Há evidências de que a música é conhecida e praticada desde a pré-história. Provavelmente a observação dos sons da natureza tenha despertado no homem, através do sentido auditivo, a necessidade ou vontade de uma atividade que se baseasse na organização de sons. Embora nenhum critério científico permita estabelecer seu desenvolvimento de forma precisa, a história da música confunde-se com a própria história do desenvolvimento da inteligência e da cultura humanas.

O ouvido humano é o principal responsável pelo desenvolvimento da música e da prática instrumental, pois as oscilações produzidas pela vibração de um corpo, como a corda de um violão, por exemplo, propagam-se pelo ar sob a forma de ondas e atingem o ouvido. O ouvido pode perceber como “sons” as ondas que tenham de 20 até 20.000 oscilações por segundo. As oscilações abaixo dessa faixa são chamadas de subsônicas, enquanto que as acima da faixa são chamadas de ultra-sônicas. Uma corda ao vibrar oscila n ciclos por segundo em seu modo fundamental, também oscila $2n$ ciclos por segundo no modo de segundo harmônico, $3n$ ciclos por segundo no modo de terceiro harmônico, e assim por diante, como mostra a Figura 1 (Matras, 1991).

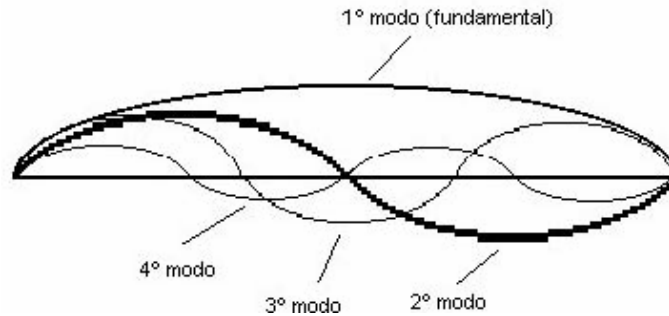


Figura 1: Ilustração dos quatro primeiros modos de vibrar de uma corda.

Entre os problemas existentes na fase inicial da aprendizagem da música, em particular, na prática de um instrumento musical, há dificuldade de se reconhecer os sons emitidos por violinos, trompetes, saxofones, trombones, entre outros vários. Ainda, para boa parte dos músicos, mesmo com certa experiência, ouvir uma melodia e representá-la pelas notas musicais em uma pauta não é tarefa trivial. Notas musicais são sinais utilizados para a representação musical, ou seja, uma nota musical é uma discretização de um fenômeno físico buscando facilitar a análise (Nattiez, 1990). Neste contexto, desenvolver um sistema capaz de automatizar o reconhecimento das notas musicais é consideravelmente interessante.

Contudo, o atual trabalho descreve um problema de identificação sonora. O desenvolvimento dos computadores e dos equipamentos de aquisição e de processamento de dados; além da aplicação da Transformada Rápida de Fourier (FFT) estimularam estes estudos (Maia et al., 1997). Diversos trabalhos apresentam estudos similares, no entanto, em sua maioria, são direcionados ao reconhecimento da fala humana (Tse et al., 2006; Raj e Whittaker, 2003; Coelho, 1990). O uso de redes neurais artificiais acelerou o desenvolvimento de trabalhos nesta linha, em particular, após o trabalho intitulado *The “Neural” Phonetic Typewriter* do pesquisador finlandês Teuvo Kohonen (Kohonen, 1992). Este trabalho, publicado originalmente em 1988 na *IEEE computer magazine*, demonstra uma pesquisa realizada inteiramente por Kohonen sobre o reconhecimento de fala. Para este trabalho foi utilizada a densidade espectral de potência (PSD do inglês *Power Spectrum Density*) do sinal para se obter as frequências de alguns sons emitidos por um violino. A PSD foi calculada utilizando o método de Welch e como placa de aquisição dos sinais foi utilizada a placa de som de um micro computador com processador AMD Sempron 2800.

2. METODOLOGIA

O universo é um sistema dinâmico definido a cada momento pelas diversas formas que a energia apresenta. O som é uma dessas formas de energia, e como tal obedece às leis fundamentais da física. O som pode ser descrito como aumentos e reduções periódicas da densidade do ar, ou seja, compressão e rarefação. Recorrendo à definição dada pela ABNT – Associação Brasileira de Normas Técnicas, em 1959, o som é “toda e qualquer vibração ou onda mecânica em um meio elástico dentro da faixa de áudio-frequência”. Ainda, é possível acrescentar que o som se propaga através de um meio elástico por uma onda longitudinal. Estas ondas transportam uma perturbação

vibratória ao longo do espaço em todas as direções (em material elástico) e são formadas por regiões alternantes de compressão e de rarefação das moléculas do ar.

Na prática, a geração do som se dá, geralmente, pela variação da pressão ou da velocidade das partículas do meio. Imaginando que uma perturbação tenha sido criada na pressão estática de um ambiente, este novo componente de pressão poderá receber a denominação de pressão sonora desde que seja captado pelo ouvido humano. A Figura 2 ilustra este processo, que pode ser especificado pela Equação 1, mostrada a seguir:

$$P(t) = P_0 + p(t) \quad (1)$$

onde $P(t)$ é a pressão total atmosférica, P_0 é a pressão atmosférica local, $p(t)$ é a pressão sonora instantânea e t é o tempo.

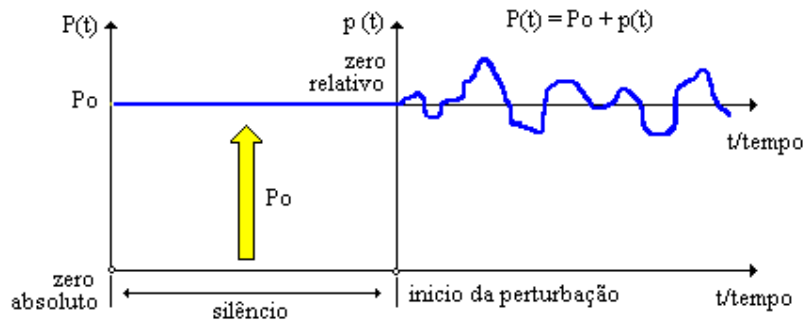


Figura 2: Som como uma perturbação da pressão do meio ambiente.

O uso de um micro computador para a execução de processamento do sinal de som é quase que fundamental. O computador não apenas processa informações matemáticas relativas ao sinal, mas, também, pode capturar o sinal sonoro transformando-o em informação digital através de um conversor analógico/digital, ou simplesmente, conversor A/D (Malvino, 1985). Isto é feito através do uso de interfaces especiais, normalmente chamadas de *placas de som*. Acopladas ao computador, essas placas convertem o sinal analógico em sinal digital (sinal reconhecido pelo computador), tornando-o, assim, um sinal discreto, e, portanto, passível de ser manuseado dentro de um sistema computadorizado. A conversão A/D procura transformar uma tensão elétrica em uma informação correspondente digital, como mostra a Figura 3 (Newland, 1986).



Figura 3: Conversão de um sinal analógico para um sinal digital.

Para converter o som em um sinal digital é necessário antes que o som seja convertido em um sinal elétrico. Para esta primeira conversão existem os transdutores acústico-elétricos, sendo o microfone o mais conhecido (Matras, 1991). Este transdutor, excitado por ondas sonoras, produz ondas elétricas substancialmente equivalentes.

2.1 Densidade Espectral de Potência – Método do Periodograma

Um procedimento possível para obter a densidade espectral de potência (PSD do inglês *Power Spectrum Density*) de um sinal determinístico é através do cálculo do módulo ao quadrado da sua transformada de Fourier (TF). Quando o sinal é aleatório, do ponto de vista teórico, a sua TF não existe visto que o integral da TF não converge com probabilidade 1. No entanto, num intervalo temporal de observação limitado, a probabilidade de falta de convergência é praticamente nula e assim $X(f)$, portanto a PSD é dada por

$$P_{xx}(f) = |X(f)|^2 \quad (2)$$

Este método de cálculo direto recebe o nome de *sample mean*, e foi utilizado durante alguns anos, apesar de algumas vezes apresentar resultados inconsistentes. Na realidade pode-se provar que a relação que permite obter um resultado equivalente ao do correlograma* é o chamado periodograma de Schuster (*Schuster's periodogram*), dado por

*o Correlograma também é um estimador não paramétrico clássico da densidade espectral de potência, e pode ser definido a partir do Estimador Periodograma (R. do Val e Lopes, 2006).

$$P_{xx}(f) = \lim_{T \rightarrow \infty} E \left[\left| \frac{1}{T} \int_0^T x(t) \exp(-j2\pi ft) dt \right|^2 \right] \quad (3)$$

ou seja, a PSD de um sinal $x(t)$ é igual à esperança matemática do módulo ao quadrado da TF de $x(t)$ calculada num intervalo suficientemente longo; sendo $\exp(\)$ é a função exponencial. Nota-se que, à parte a duração do intervalo, a diferença essencial em relação à Equação 2 é a introdução da esperança matemática. Pode-se efetivamente observar que a relação mostrada na Equação 3 se deduz do correlograma da seguinte forma: desdobrando o módulo ao quadrado a partir da Equação 3, tem-se

$$\begin{aligned} P_{xx}(f) &= \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \int_0^T \int_0^T x(t_1) x^*(t_2) \exp[-j2\pi f(t_1 - t_2)] dt_1 dt_2 \right] \\ P_{xx}(f) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \int_0^T E[x(t_1) x^*(t_2)] \exp[-j2\pi f(t_1 - t_2)] dt_1 dt_2 \\ P_{xx}(f) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \int_0^T \tau_{xx}(t_1 - t_2) \exp[-j2\pi f(t_1 - t_2)] dt_1 dt_2 \end{aligned} \quad (4)$$

onde; utilizando integral dupla

$$\int_a^b \int_a^b g(t_1 - t_2) dt_1 dt_2 = (b-a) \int_{a-b}^{b-a} \left(1 - \frac{|t|}{b-a} \right) g(t) dt \quad (5)$$

É possível escrever a PSD da seguinte forma (Urkowitz, 1983)

$$P_{xx}(f) = \lim_{T \rightarrow \infty} \int_{-T}^T \left(1 - \frac{|t|}{T} \right) \tau_{xx}(\tau) \exp(-j2\pi ft) d\tau \quad (6)$$

Quando $T \rightarrow \infty$ a função triangular aproxima-se cada vez mais de uma constante, assim:

$$P_{xx}(f) = \lim_{T \rightarrow \infty} \int_{-\infty}^{\infty} \tau_{xx}(\tau) \exp(-j2\pi f\tau) d\tau \quad (7)$$

ou seja, a densidade espectral de potência pode ser obtida através do correlograma (Teorema de Wiener-Khintchine). O Teorema de Wiener-Khintchine mostra explicitamente que a informação veiculada pela função de correlação cruzada no domínio do tempo é a mesma informação veiculada pela DEPX (Densidade Espectral de Potência Cruzada) no domínio de frequência (Lin, 1967; Davenport e Root, 1958).

2.2 Caso discreto

A implementação do periodograma no caso discreto faz-se substituindo a TF limitada no tempo pela TFD também limitada no tempo ou simplesmente pela TFD (Transformada de Fourier Discreta). Assim, a Equação 2 para o caso discreto é dada por

$$P_{xx}(f) = \left| \sum_{n=0}^{N-1} x[n] \exp(-j2\pi fn) \right|^2 \quad (8)$$

Claro que esta relação sofre do mesmo problema de instabilidade do que no caso contínuo. Utilizando a mesma demonstração do que no caso contínuo, prova-se que a PSD é dada por:

$$P_{xx}(f) = \lim_{N \rightarrow \infty} E \left[\frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j2\pi ft) \right|^2 \right] \quad (9)$$

Então, é possível se obter efetivamente uma relação idêntica àquela obtida pela TFD da autocorrelação. Para verificar da inconsistência do *sample-mean* como estimador da PSD, torna-se necessário mencionar alguns resultados respeitantes à sua variância e covariância. Os resultados são mostrados a seguir sem demonstração, pois pode ser encontrada em textos da especialidade (Therrien, 1992).

Para um processo $X(z) = TZ[x[n]] = \sum_{n=-\infty}^{+\infty} x[n]z^{-n}$ de PSD

$\text{COV}[\hat{P}_{xx}^p(\omega_1)\hat{P}_{xx}^p(\omega_2)] = P_{xx}(\omega_1)P_{xx}(\omega_2) \frac{\sin(N(\omega_1 - \omega_2)/2)}{N \sin(\omega_1 - \omega_2)/2}$, temos que a covariância do *sample mean* se escreve:

$$\omega_1 = \omega_2 \quad (10)$$

Portanto, para $V[\hat{P}_{xx}^p(\omega)] \approx P_{xx}^2(\omega)$, tem-se a variância dada por:

$$\frac{\omega_1 - \omega_2}{2} = \frac{k\pi}{N} \quad (11)$$

de onde se pode concluir que:

(a) A covariância passa por valores nulos sempre que:

$$P_{xx}(\omega) = \lim_{N \rightarrow \infty} E \left[\frac{1}{2N+1} \left| \sum_{n=-N}^N x[n] \exp(-j\omega n) \right|^2 \right] \quad (12)$$

ou seja, que as frequências para as quais o estimador é não correlacionado ficam mais próximas quanto mais elevado for N;

- (b) A variância é não enviesada, mas assintoticamente inconsistente, pois não depende de N. Por outras palavras, um aumento de N não produz uma diminuição da variância do periodograma. Pode-se notar que no que diz respeito ao *sample mean*, à parte a questão do limite $V\{\hat{t}_{xx}[m]\} = q_N[m]V\{\hat{t}_{xx}[m]\}$, também foi ignorado a $E[s(t)] = s(t)$. O que significa dizer que na prática o *sample mean* vai sofrer de uma grande inconsistência caso não seja realizada qualquer tipo de média. A média terá forçosamente de ser temporal admitindo a hipótese de ergodicidade e por falta de realizações do processo $X(z) = TZ[x[n]] = \sum_{n=-\infty}^{+\infty} x[n]z^{-n}$.

Devido a estes problemas do *sample-mean* e a necessidade de efetuar algum “alisamento” (*smoothing*) do espectro, surgiram várias sugestões de formas de calcular a PSD, baseadas no periodograma e que constituem outros tantos estimadores, tais como o periodograma de Daniell, o periodograma de Bartlett e o próprio periodograma de Welch utilizado neste trabalho.

Periodograma de Welch

Welch propôs uma versão semelhante ao periodograma de Bartlett no qual utilizou janelas temporais e a possibilidade de sobreposição dos intervalos de estimação dos espectros de ordem k. Se a série temporal $S \leq L$ for dividida em L segmentos com $N = KL$ amostras cada, cada segmento está atrasado em relação ao anterior de $[(N - L)/S + 1]$ amostras e, então, o número de segmentos L é igual à parte inteira de $x^{(k)}[n] = \omega[n]x[n + kS]$. No periodograma de Welch o segmento de ordem k é

$$\omega[n] \quad (13)$$

onde $\hat{P}_w(\omega)$ é a função janela. Em seguida o estimador escreve-se exatamente como o de Bartlett utilizando as L estimativas. As expressões definitivas e mais detalhes podem ser encontrados em Marple (1987).

Pode-se notar que o procedimento de Welch permite, para um mesmo valor de N, uma maior estabilidade porque L é mais elevado mantendo uma resolução constante, valor de $N = KL$. Prova-se, além disto, que uma sobreposição de 50% oferece o melhor compromisso estabilidade resolução para valores de N e $N = KL$ fixos (Nuthall e Carter, 1980).

3. APLICAÇÃO EXPERIMENTAL

A metodologia proposta foi aplicada para a identificação das notas musicais de um violino do tipo Anton Breton modelo 4/4. A Figura 4 apresenta o instrumento para clareza de entendimento. A aquisição do som de cada corda do violino, vibrando apenas em função do efeito do arco, foi realizada através da placa de som (*on board*) de um micro computador com processador AMD Sempron 2800, utilizando um algoritmo implementado em ambiente Matlab®. A aquisição foi realizada por três segundos, com frequência máxima de aquisição igual a 8 kHz. A Figura 5 mostra os sinais no tempo para cada uma das quatro cordas do instrumento.



O violino tem uma extensão de quatro oitavas e suas cordas são afinadas em intervalos de quinta, sendo a mais grave afinada pelo sol abaixo do dó central (isto é, sol-ré-lá-mi). Os primeiros violinos foram feitos na Itália em meados do século XVI (1700), evoluindo de antecessores como arebec, a vielle e a lira da braccio. A arte de fabricar violinos de primeira classe foi, por 200 anos, apanágio de três famílias de Cremona - os Amati, Guarneri e Stradivari (de que a latinização deu Stradivarius). Embora o violino propriamente dito tenha-se mantido inalterado por 400 anos (excetuando-se a utilização de cordas mais finas e de um cavalete mais alto no século XIX), a forma atual do arco só se consolidou no século XIX (1800). Originalmente convexo em relação ao violino, o arco é agora côncavo.

Figura 4: Detalhes de um violino para clareza de compreensão do trabalho.

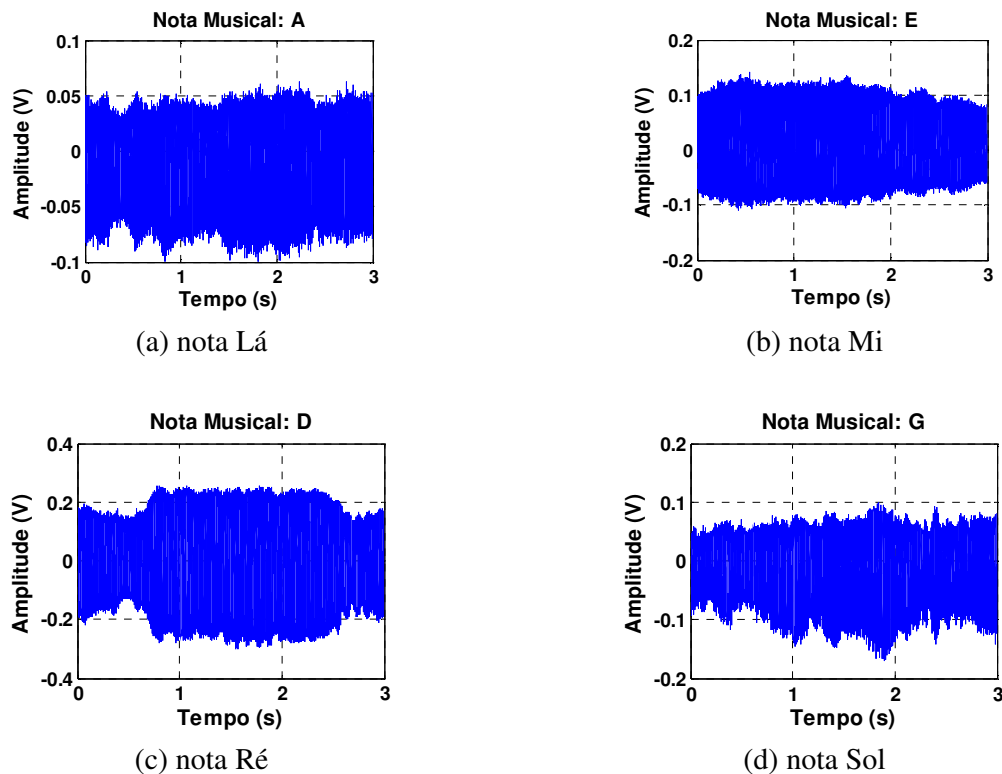


Figura 5: Sinais no domínio do tempo - notas musicais de um violino.

Para cada sinal mostrado na Figura 5 foi calculada a densidade espectral de potência utilizando o método de Welch, como mostram as Figuras 6, 7, 8 e 9. Antes da aquisição dos sinais o violino foi afinado com relação a um diapasão, e então, foi possível comparar com os valores teóricos disponíveis na literatura. Para clareza de compreensão, a Tabela 1 mostra as frequências identificadas e os valores teóricos esperados.

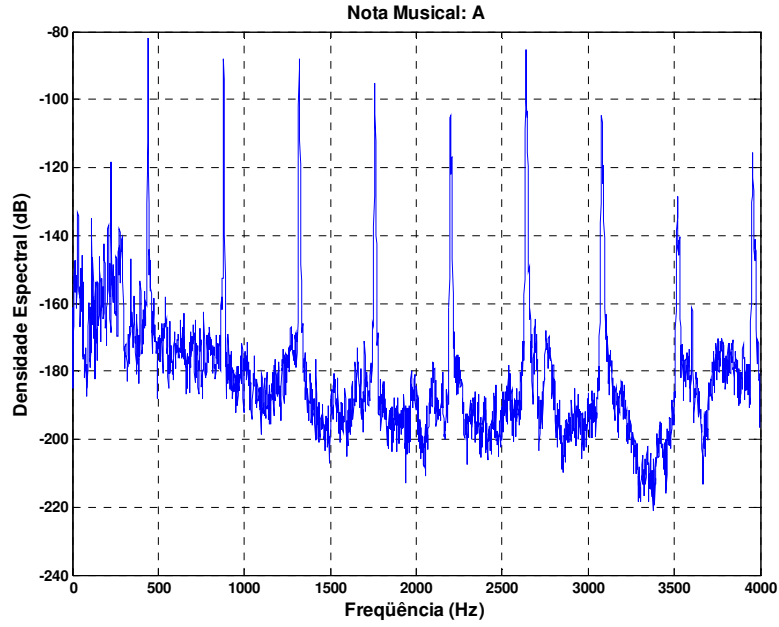


Figura 6: Densidade espectral da nota lá (A).*

*Note que as figuras 6, 7, 8 e 9 mostram a densidade espectral de potência apenas até 4 kHz, no entanto, aquisição foi feita até 8 kHz.

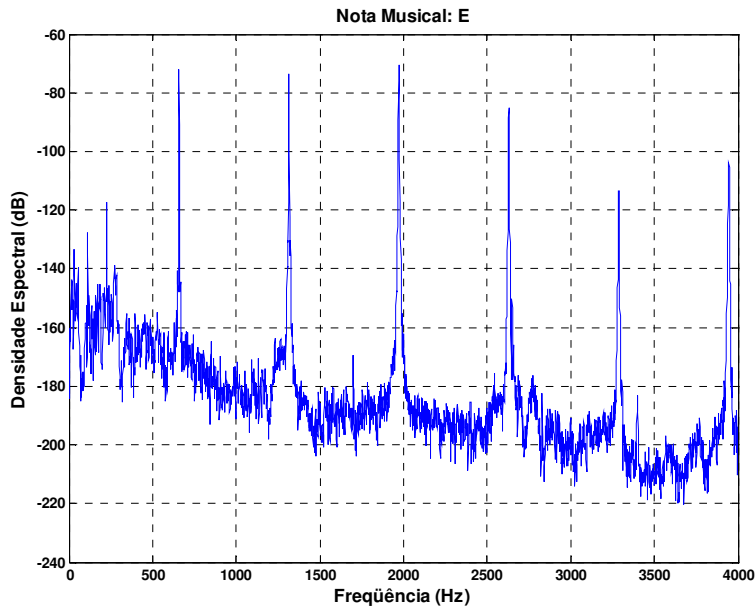


Figura 7: Densidade espectral da nota Mi (E).

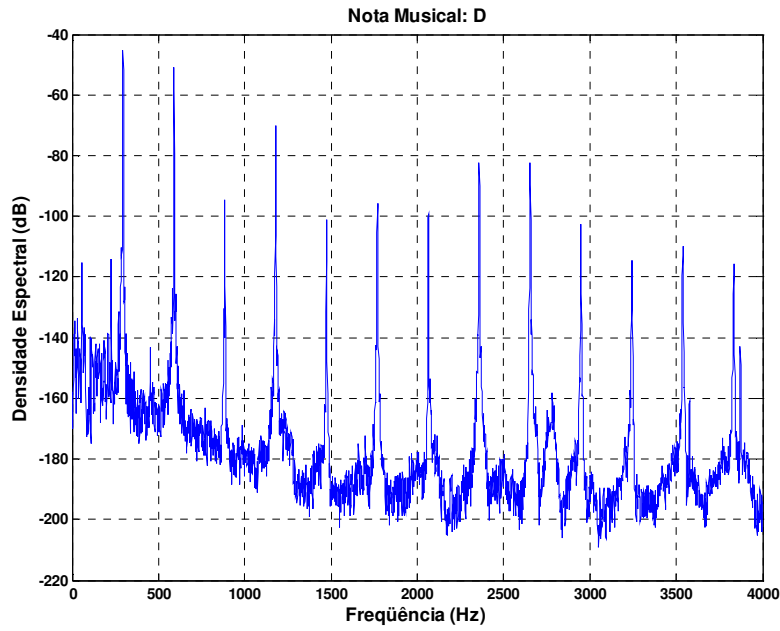


Figura 8: Densidade espectral da nota Ré (D).

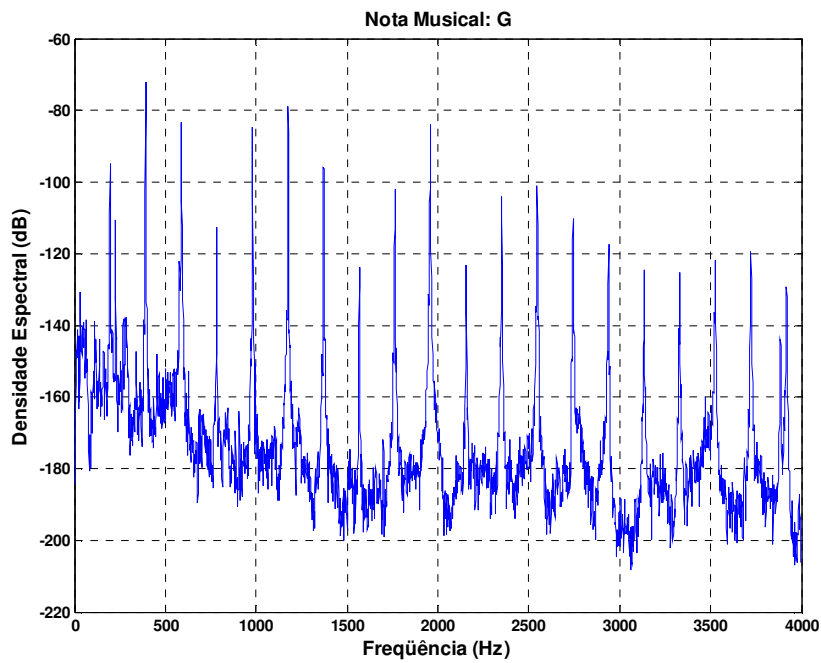


Figura 9: Densidade espectral da nota Sol (G).

Tabela 1: Valores das frequências identificadas; usando PSD via método de Welch.

Nota Musical	Valor Experimental [Hz]	Valor Teórico [Hz]*	Erro Relativo (%)
G	196	196	0,00
D	295	294	0,34
A	440	440	0,00
E	657	660	0,45

*Valores de referência obtidos em: <http://en.wikipedia.org/wiki/Note>, acesso em 05/04/2007.

4. CONSIDERAÇÕES FINAIS

O desenvolvimento de um sistema compacto que reconheça uma nota musical certamente atrai o interesse de grande parte dos estudantes de um instrumento musical em fase inicial. Ainda, atrai também o interesse de todos que eventualmente precisem representar uma melodia através das notas musicais em uma pauta, como por exemplo, os regentes de corporações como bandas e fanfarras; bastante comuns no Brasil. Neste contexto, o atual trabalho apresentou a utilização da densidade espectral de potência de um sinal, obtida através do método de Welch, para a identificação das notas musicais de um violino. A partir deste trabalho, a idéia é desenvolver um sistema compacto e de baixo custo, que não utilize um micro computador, mas que faça aquisição de um som e reconheça as notas musicais. Ainda, para se obter robustez em relação a variação do timbre dos instrumentos será incorporado ao trabalho alguma técnica de inteligência artificial, como por exemplo as redes neurais artificiais já amplamente utilizadas em trabalhos de identificação da voz humana.

5. REFERÊNCIAS

- Coelho, H., 1990, *Inteligência artificial*. Portugal, Lisboa: Caminho.
- Davenport, W. B. Jr. e Root, W. L., 1958, *An Introduction to the Theory of Random Signals and Noise*. New York, McGraw-Hill.
- Kohonen, T., 1992, *The "neural" phonetic typewriter*. In: Sánchez-Sinencio, Edgar e Lau, Clifford. Artificial neural networks. USA, New York: IEEE Press, p. 369-387.
- Lin, Y. K., 1967, *Probabilistic Theory of Structural Dynamics*. New York, McGraw-Hill.
- Malvino, A. P., 1985, *Microcomputadores e microprocessadores*. São Paulo: McGraw-Hill.
- Marple, L., 1987, *Digital Spectral Analysis with Applications*, Prentice-Hall, Signal Processing Series, USA 1987.
- Matras, Jean-Jacques, 1991, *O som*. São Paulo: Martin Fontes.
- Nattiez, Jean-Jacques, 1990, "*Music and Discourse: Toward a Semiology of Music (Musicologie générale et sémiologie*". Translated by Carolyn Abbate (1990). ISBN 0-691-02714-5.
- Newland, D. E., 1986, *Random vibrations and spectral analysis*. 2 ed. New York, Longman Inc.
- Nuthall, A. H. e Carter, G. C., 1980, *A generalized framework for power spectral estimation*, IEEE Trans. Acoust. Speech and Signal Processing, vol. ASSP 28, pp 334-335, June 1980.
- R. do Val, J. B. e Lopes, A., 2006, *Métodos Não-Paramétricos*, Cap. 2 da Enciclopédia de Controle, Sociedade Brasileira de Automática – SBA, Texto Preliminar.
- Raj, B. e Whittaker, E. W. D., 2003, "*Lossless Compression of Language Model Structure and Word Identifiers*", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ISSN: 1520-6149, Vol. 1, pp. 388-391, April.
- Therrien, C. W., 1992, *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall.
- Tse, E., Greenberg, S. e Shen, C., 2006, "*GSI Demo: Multiuser Gesture/Speech Interaction over Digital Tables by Wrapping Single User Applications*", International Conference on Multimodal Interfaces (ICMI), pp 76-83, November.
- Urkowitz, H., 1983, *Signal Theory and Random Processes*, Artech House Inc.
- Welch, P.D., 1967, *The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms*, IEEE Trans. Audio Electroacoust, vol. AU-15, pp. 70-73.

6. AGRADECIMENTOS

Os autores agradecem à Unesp; à FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo); e a Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro concedido às pesquisas.

7. DIREITOS AUTORAIS

Os autores são os únicos responsáveis pelo conteúdo deste trabalho.

EXPERIMENTAL MUSICAL NOTES IDENTIFICATION

Sanderson Manoel da Conceição¹

smconceicao@aluno.feis.unesp.br

Douglas Domingues Bueno²

ddbueno@dem.feis.unesp.br

Vicente Lopes Junior²

vicente@dem.feis.unesp.br

GMSINT – Grupo de Materiais e Sistemas Inteligentes, Mathematical Department¹, Mechanical Engineering Department², Faculdade de Engenharia de Ilha Solteira, UNESP, Av. Brasil n.º 56, ZIP CODE 15385000, Ilha Solteira, SP, <http://www.dem.feis.unesp.br/gmsint>.

Abstract: Constantly, the man is trying to improve some processes intending the minimization of costs, to get high speed and precision of results and, still, to help some learning process. To simplify the musical instrument learning is very interesting for all the people. In general, for a student in an initial step, to recognize the musical notes emitted by a musical instrument is a sufficiently complex task. Thus, the development of an acquisition system, compact and with low cost, which it can evaluate a sound and indicates the corresponding musical note is very interesting. In this context, in this work is used the power spectral density to classify some musical notes of a violin. The acquisition of the signals was realized through an algorithm implemented using the Matlab® software and the computer sound board (AMD Sempron 2800).

Key-words: *experimental identification, musical notes, power spectrum density.*