



# 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 42<sup>nd</sup> Street, New York, New York 10165-2520, USA, [www.aes.org](http://www.aes.org). Informações sobre a seção Brasileira podem ser obtidas em [www.aesbrasil.org](http://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 de uma Representação Compacta por “Matching Pursuit” em um Esquema de Análise/Síntese de Áudio

Michel P. Tcheou,<sup>1</sup> Lisandro Lovisolo,<sup>2</sup> Luiz W. P. Biscainho<sup>3</sup> e Eduardo A. B. da Silva<sup>3</sup>

<sup>1</sup> UFRJ, LPS - PEE/COPPE, CP 68504, Rio de Janeiro, RJ, 21941-972, Brasil

<sup>2</sup> UERJ, DETEL/FEN, R. São Francisco Xavier 524, sala 5036A, Rio de Janeiro, RJ, 20559-900, Brasil

<sup>3</sup> UFRJ, LPS - DEL/Poli e PEE/COPPE, CP 68504, Rio de Janeiro, RJ, 21941-972, Brasil

[pompeu@lps.ufrj.br](mailto:pompeu@lps.ufrj.br), [lisandro@uerj.br](mailto:lisandro@uerj.br), [wagner@lps.ufrj.br](mailto:wagner@lps.ufrj.br), [eduardo@lps.ufrj.br](mailto:eduardo@lps.ufrj.br)

### RESUMO

A decomposição de sinais por “Matching Pursuit” (MP) vem sendo aplicada em áudio. Na literatura, mostrou-se que um dicionário de janelas moduladas nas frequências da escala musical cromática possibilita boa resolução para análise tempo-frequência. Este trabalho implementa um esquema de análise/síntese de áudio utilizando MP cromático. Seu desempenho para diversas famílias de sinais é avaliado por uma medida objetiva descrita pelo ITU e por testes subjetivos informais, em função do número de coeficientes da representação. É sugerido, ainda, um critério para parada da decomposição.

### 0 INTRODUÇÃO

Recentemente, o estudo de técnicas capazes de obter representações compactas de sinais tem-se intensificado [1]. Essas técnicas têm obtido sucesso no processamento de sinais de áudio para fins de transcrição musical [2, 3] e codificação [4, 5]. A codificação de áudio, que é o enfoque deste trabalho, envolve diversas etapas tais como análise, mascaramento no tempo e na frequência, quantização, codificação por entropia,

quantização inversa e síntese. Para que o codificador apresente bom desempenho é fundamental utilizar um método de análise capaz de aproximar o sinal com o menor número possível de coeficientes.

Normalmente, essa aproximação é realizada através de decomposições que aproximam os sinais que compõem um dado espaço, usando uma combinação linear de formas de onda pré-definidas (ou átomos) provenientes de um dicionário dito redundante por conter mais elementos que os necessários para gerar o espaço. As-

sim, pode-se escolher um subconjunto dos átomos de um dicionário redundante e associar-lhes diferentes pesos a fim de representar um dado sinal. Esse paradigma de representação de sinais é também conhecido por decomposição atômica. Existem diversos métodos que realizam decomposições atômicas; dentre eles, há o popular *Matching Pursuit* [6].

Uma representação compacta pressupõe alto grau de similaridade entre os elementos do dicionário e os fenômenos presentes no sinal. O uso de dicionários redundantes permite alcançar maior nível de compacidade na representação que o uso de bases ortonormais reais, tais como as *wavelets* discretas, visto que um sinal pode apresentar estruturas complexas que não são bem representadas utilizando-se somente uma classe de formas de onda. Por sua vez, o dicionário redundante pode ser formado por uma união de bases ortogonais, por transformadas redundantes (*wavelet packet* e *frames de Gabor*) ou por formas de ondas parametrizadas [4, 7, 8, 9].

A Figura 1 ilustra o esquema de codificação de sinais utilizando decomposições atômicas. O *codificador* analisa o sinal de forma a encontrar uma boa representação para este, com base em um dicionário  $D$ . Nesse dicionário são modelados os fenômenos existentes nos sinais que se deseja aproximar. Em seguida, os coeficientes e os índices dos átomos da representação do sinal são transmitidos/armazenados. O *decodificador*, com base no mesmo dicionário, sintetiza/reconstrói o sinal a partir dos átomos e pesos provindos do codificador.

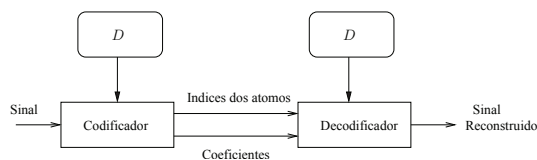


Figura 1: Codificação de sinais usando dicionários redundantes / representações atômicas

Se soubermos de que forma os sinais de uma classe são produzidos, podemos utilizar representações atômicas para tentar modelar o sistema físico que gera o sinal. Se conseguirmos um bom modelo e o codificarmos eficientemente, poderemos obter altas taxas de compressão. Esse é o contexto deste trabalho. Nele discutimos a utilização de representações atômicas obtidas usando o algoritmo conhecido por *Matching Pursuit* [6, 10] para sinais de áudio. Abordagens similares a esta já foram propostas na literatura [2, 3, 4, 5]. No presente trabalho, avaliam-se potencialidades e deficiências de tais métodos, e implementa-se um esquema simples de análise/síntese de áudio utilizando o MP cromático. Seu desempenho é avaliado para diversas famílias de sinais por uma medida objetiva descrita pelo ITU, em função do número de coeficientes da representação. É proposto, ainda, um critério para parada da decomposição.

A próxima seção apresenta aspectos relevantes de representações atômicas e como elas podem fornecer representações compactas e coerentes – relacionadas ao processo físico gerador dos sinais observados. A seção seguinte explica o dicionário aqui empregado, composto por senóides geradas a partir das frequências da escala cromática. A seguir, detalha-se o algoritmo de decomposição empregado, que é uma variante do *Matching Pursuit*. Após isso, são apresentados resultados referentes ao desempenho do método para diversas famílias de sinais, que são também avaliadas usando o PEAQ [11]. Por fim, as conclusões são apresentadas.

## 1 REPRESENTAÇÕES ATÔMICAS

Numa representação atômica, um sinal  $\mathbf{x}$  é descrito como uma combinação linear de poucas formas de onda selecionadas a partir de um conjunto pré-definido, o dicionário  $\mathcal{D}$ . Os sinais  $\mathbf{g}_\gamma$  que compõem  $\mathcal{D}$  são as formas de onda que podem ser utilizadas na decomposição, chamadas de átomos ou elementos. Algoritmos que obtêm representações atômicas de sinais escolhem um subconjunto de  $M$  elementos  $\mathbf{g}_{\gamma(i)}$  do dicionário  $\mathcal{D}$  que aproximam um sinal  $\mathbf{x}$  na forma

$$\mathbf{x} \approx \hat{\mathbf{x}} = \sum_{i=1}^M \alpha_i \mathbf{g}_{\gamma(i)}, \quad \mathbf{g}_{\gamma(i)} \in \mathcal{D}. \quad (1)$$

Cada átomo  $\mathbf{g}_{\gamma(i)}$  (indexado por  $\gamma(i)$ ) utilizado na aproximação de  $\mathbf{x}$  pode ser interpretado como uma característica intrínseca do sinal  $\mathbf{x}$  de “peso”  $\alpha_i$ . A representação mais compacta de  $\mathbf{x}$  é aquela que utiliza o menor número de átomos [4] com distorção nula. Na prática, não se procura, necessariamente, o menor número de átomos, pois um pequeno conjunto de átomos que aproxime  $\mathbf{x}$  com uma distorção aceitável pode ser suficiente [7, 8]. Assim, a compacidade de uma representação está relacionada ao número de termos  $M$ , e à medida que  $M$  cresce menos compacta é a representação.

Para que combinações lineares de  $M$  átomos de um dicionário  $\mathcal{D}$  sejam capazes de representar qualquer sinal  $\mathbf{x} \in \mathbb{X}$  com distorção  $d(\mathbf{x}, M, \mathcal{D})$  arbitrariamente pequena,  $\mathcal{D}$  tem que ser completo em  $\mathbb{X}$  [10]. Assim, deverá existir pelo menos uma combinação linear de elementos de  $\mathcal{D}$  que produza  $\hat{\mathbf{x}} = \mathbf{x}$ ,  $\forall \mathbf{x} \in \mathbb{X}$ , isto é,  $\mathcal{D}$  deve gerar o espaço  $\mathbb{X}$ . Dicionários são ditos sobre-completos ou redundantes quando possuem mais elementos que os necessários para tal fim. A redundância permite que diferentes combinações lineares de elementos de  $\mathcal{D}$  expressem um mesmo  $\mathbf{x}$ .

Quando lidamos com processos físicos, o sinal observado é uma mistura de componentes  $\mathbf{p}_i$ , que representam fenômenos físicos, dada por

$$\mathbf{x} = \sum_i \beta_i \mathbf{p}_i + \mathbf{n}, \quad (2)$$

onde  $\mathbf{n}$  é o ruído inerente à observação. Quanto mais parecidos forem os átomos  $\mathbf{g}_{\gamma(i)}$  e seus pesos  $\alpha_i$  utilizados na representação de um sinal  $\mathbf{x}$  com as componentes  $\mathbf{p}_i$  e seus pesos  $\beta_i$  presentes em  $\mathbf{x}$ , melhor será a

representação obtida para modelagem do sinal e reconhecimento de padrões, e a dizemos coerente.

Assim, vemos que para se poder obter uma representação compacta ou coerente, um dos aspectos-chave reside no dicionário, isto é, na definição de seus átomos.

## 2 DICIONÁRIO DE ESCALA MUSICAL CROMÁTICA

Os elementos de um dicionário redundante podem ser gerados escalonando-se, transladando-se e modulando-se uma função janela real  $h(t)$ . No caso de um dicionário de átomos complexos, cada átomo  $\mathbf{g}_\gamma$  é definido como [6]

$$g_\gamma(n) = K_\gamma h\left(\frac{n-u}{s}\right) e^{i\xi n}, \quad (3)$$

onde  $K_\gamma$  é tal que  $\mathbf{g}_\gamma$  possui norma unitária. O índice  $\gamma$  é composto pelo escalonamento  $s$ , pelo deslocamento no tempo  $u$  e pela frequência de modulação  $\xi$ , ou seja,  $\gamma = (s, u, \xi)$ , onde  $\gamma \in \Gamma = \mathbb{R}^+ \times \mathbb{R}^2$  e  $\Gamma$  é conjunto de todos os parâmetros do dicionário.

No trabalho original de Mallat e Zhang [6], propõe-se um dicionário formado por átomos gerados a partir de uma janela gaussiana, que possuem a melhor localização tempo-frequência. Entretanto, esses átomos possuem baixo grau de similaridade com relação aos fenômenos presentes em sinais de áudio, comprometendo, assim, a compacidade na representação de sinais de áudio, fator fundamental para o bom desempenho de um codificador. Em [2], verifica-se que o uso de um dicionário composto por janelas de Hanning moduladas nas frequências da escala musical cromática possibilita boa resolução para análise tempo-frequência. Além disso, seus elementos são mais parecidos com os padrões de sinais de áudio. Dessa maneira, opta-se por utilizá-lo com o propósito de obter representações mais compactas.

O dicionário de escala musical cromática é obtido, de acordo com a Eq. (3), a partir de uma janela de Hanning  $h(t)$  definida por

$$h(t) = 1 + \cos(2\pi t). \quad (4)$$

Os parâmetros de  $\gamma$ ,  $s$  (escala) e  $u$  (de deslocamento no tempo), são discretizados da seguinte maneira (ver Eq. (3))

$$s = 2^j, \quad j \in \{j_0, \dots, \log_2(N)\}, \quad (5)$$

$$u = p\left(\frac{s}{2}\right), \quad p \in \left\{0, \dots, \frac{2N}{s} - 1\right\}, \quad (6)$$

onde  $N$  é comprimento do sinal. É importante destacar que evita-se que parte de um átomo fique "fora" do suporte do sinal (no caso, para  $u = 0$ ) realizando-se a extensão periódica do sinal a ser decomposto. A frequência, por sua vez, é discretizada da seguinte forma

$$\xi = \xi_0 2^{\frac{k}{12}}, \quad k \in \{0, \dots, k_{\max}\}, \quad (7)$$

onde  $k$  é o índice da frequência correspondente a um semitom da escala musical cromática entre a frequência mínima  $\xi_0$  e a frequência máxima associada a  $k_{\max}$ . Com o intuito de representar sinais provenientes de instrumentos de afinação imperfeita, adicionam-se frequências laterais em torno de cada semitom, resultando em uma escala mais fina

$$\xi \in \left\{ \xi_0 2^{\frac{k-\delta}{12}}, \xi_0 2^{\frac{k}{12}}, \xi_0 2^{\frac{k+\delta}{12}} \right\}, \quad k \in \{0, \dots, k_{\max}\}, \quad (8)$$

O parâmetro  $\delta \in ]0, 1/3]$  ajusta a razão custo-benefício entre sensibilidade e seletividade. Utilizou-se  $\delta = 1/3$  para manter constante a razão geométrica entre as frequências.

Observe que não é possível analisar de maneira eficaz componentes de frequências que possuem menos de 1 ciclo dentro de uma escala  $s$ . Portanto para impedir a inserção de um número maior que o necessário de elementos no dicionário, seleciona-se, para uma dada escala,  $\xi_0$  de modo que as frequências apresentem mais que 1 ciclo.

Uma vez definidos os átomos que iremos empregar, passemos à descrição do algoritmo empregado para decompor os sinais.

## 3 ALGORITMO DE DECOMPOSIÇÃO

O método de decomposição proposto, inspirado em [12] e [13], utiliza o algoritmo de "Matching Pursuit" (MP) com base no dicionário de escala musical cromática. O MP realiza aproximações sucessivas de um sinal  $\mathbf{x}$  buscando em um dicionário redundante a melhor aproximação, a cada iteração.

O sinal de áudio é dividido em quadros  $\mathbf{x}_k$  ( $k$  é o índice do quadro) de tamanho  $N$ , ponderados por uma janela suavizadora, em saltos de  $H$  amostras. Aqui, usou-se a janela de Hanning com  $H = N/2$ , garantindo a reconstrução exata do sinal com a sobreposição dos quadros. Antes de se janelar o sinal, é necessário inserir  $H/2$  amostras antes de seu início e após o seu fim. A decomposição é realizada em cada um dos quadros separadamente.

Dado que o sinal é real e passa-banda, os elementos do dicionário são sinais analíticos (ver Eq. (3)), e como supomos que a janela compreenderá pelo menos 1 ciclo (como determinado na seção anterior) então o resíduo inicial é ajustado como

$$\mathbf{r}_{\mathbf{x}_k}^0(n) = h_N(n) x_k(n) + j h_N(n) \mathcal{H}\{x_k(n)\}, \quad (9)$$

onde  $h_N(n)$  corresponde à janela de Hanning de comprimento  $N$  e  $\mathcal{H}\{\cdot\}$  representa a transformada de Hilbert. É importante ressaltar que a convergência na aproximação de um sinal analítico garante a convergência da aproximação do sinal real correspondente [13]. A cada iteração da decomposição, o MP busca o átomo  $\mathbf{g}_{\gamma(i)}$  de maior correlação com o resíduo da iteração anterior,  $\mathbf{r}_{\mathbf{x}_k}^{i-1}$ . O átomo escolhido é subtraído do resíduo, gerando um novo resíduo

$$\mathbf{r}_{\mathbf{x}_k}^i = \mathbf{r}_{\mathbf{x}_k}^{i-1} - \alpha_i \mathbf{g}_{\gamma(i)}, \quad \text{onde } \alpha_i = \left\langle \mathbf{r}_{\mathbf{x}_k}^{i-1}, \mathbf{g}_{\gamma(i)} \right\rangle. \quad (10)$$

Este processo se repete até que o energia do resíduo seja suficientemente pequena ou que algum outro critério de parada seja atendido. Ao final, obtém-se a seguinte aproximação linear de  $\mathbf{x}_k$ :

$$\hat{\mathbf{x}}_k = \Re \left\{ \sum_{i=1}^M \alpha_i \mathbf{g}_{\gamma(i)} \right\}, \quad (11)$$

onde  $\mathbf{g}_{\gamma(i)}$  é o átomo complexo do dicionário associado ao índice de parâmetros  $\gamma_i$ ,  $\alpha_i$  corresponde ao coeficiente complexo e  $M$  é o número de termos na representação. A seqüência de pares  $\{\alpha_i, \gamma(i)\}$ ,  $i = 1, \dots, M$ , forma o que denomina-se de Livro de Estruturas [6].

### 3.1 CRITÉRIO DE PARADA

Deseja-se que a decomposição seja interrompida quando a aproximação do sinal for suficientemente boa, significando que os átomos escolhidos para a aproximação são coerentes, ou seja, altamente correlacionados com as características intrínsecas do sinal. Normalmente, utiliza-se a norma do resíduo ou um número máximo de iterações como critério de parada [6, 4, 2]. A norma do resíduo é uma medida do erro de aproximação, contudo com base apenas nela não é possível averiguar se o resíduo ainda é correlacionado significativamente com algum átomo do dicionário [9]. Para isso, propõe-se o uso da taxa de aproximação [6, 14], definida como

$$\lambda(i) = \frac{\left| \left\langle \mathbf{r}_x^{i-1}, \mathbf{g}_{\gamma(i)} \right\rangle \right|}{\|\mathbf{r}_x^{i-1}\|}, \quad (12)$$

que mede o quanto do resíduo  $\mathbf{r}_x^{i-1}$  é aproximado na iteração  $i$ . A taxa de aproximação possui um comportamento decrescente exponencial, como é demonstrado em [6]. A derivada da envoltória dessa exponencial diminui à medida que o número de iterações aumenta até que  $\lambda(i)$  convirja para um valor médio estacionário. Quando a taxa de aproximação atinge esse valor, poder-se dizer que não há mais elementos do dicionário que sejam coerentes com o resíduo, portanto a decomposição pode ser interrompida. Esse critério apresenta um bom desempenho quando aplicado a sinais elétricos [7, 8, 9].

## 4 RESULTADOS

Nos experimentos, realizam-se a análise/síntese, utilizando o MP cromático, de notas de diferentes famílias de instrumentos: clarinete (A3), piano (A3), tímpano (E3) e violão (A3). Os sinais possuem 1 segundo de duração e são amostrados a uma taxa de 44,1 kHz, logo têm 44100 amostras cada. O sinal é dividido em quadros de  $N = 1024$  amostras, o que corresponde a 0,023 segundo, com sobreposição de  $H = 512$  amostras. Dessa forma, têm-se um total de 88 quadros. O dicionário foi gerado para valores do parâmetro de escala  $s \in \{4, 8, \dots, 1024\}$ , resultando em 95420 elementos. Definiu-se a escala de frequências cromáticas variando

de 27,5 Hz (A0) a 15,6 kHz (B9). Esse limite superior corresponde à frequência da escala cromática mais próxima da média geométrica dos extremos da última oitava representável com a taxa de amostragem de 44,1 kHz.

Primeiramente, avaliou-se o esquema de análise/síntese proposto através do PEAQ (*Perceptual Evaluation of Audio Quality*) [11]. O PEAQ é um método de medição objetiva da qualidade de sinais de áudio descrito pelo ITU (*International Telecommunications Union*), originalmente destinado à avaliação de codecs com perdas. Ele utiliza diversas medidas de base psicoacústica que, combinadas, resultam em uma medida de qualidade relativa entre duas instâncias de um sinal – um sinal de referência (original) e um sinal de teste (reconstruído) [11]. Essa medida de qualidade é classificada nas seguintes faixas:

- **-4 a -3**: muito perturbador;
- **-3 a -2**: perturbador;
- **-2 a -1**: pouco perturbador;
- **-1 a 0**: não perturbador.

As Figuras de 2 a 5 apresentam resultados da avaliação do desempenho do MP cromático através do PEAQ. Todas apresentam o mesmo padrão, segundo o qual o avaliador, naturalmente, premia as representações mais completas. O comportamento anômalo para poucos coeficientes se explica pelo fato de o PEAQ se destinar à avaliação de áudio de alta qualidade: sinais muito prejudicados geram notas pouco confiáveis.

Observa-se para o clarinete (Fig. 2), o número mínimo de coeficientes para que o sinal reconstruído fosse avaliado como não perturbador (PEAQ entre -1 e 0) foi 14178; no caso do piano (Fig. 3), foram necessários 9776 coeficientes; no do tímpano (Fig. 4), 6903; e no do violão (Fig. 5), 6818. Em ordem decrescente de compactação, teríamos violão, tímpano, piano e clarinete.

Em seguida, fizeram-se testes informais de audição a fim de verificar o número mínimo de coeficientes considerados necessários para representar fielmente as notas de cada instrumento. Usou-se como referência o percentual da norma inicial de cada quadro que determinava o resíduo de parada. Os resultados são apresentados na Tabela 1, juntamente com o percentual de energia representado. O número máximo de iterações do algoritmo de decomposição foi fixado em 400 por quadro.

Os testes subjetivos atribuíram o maior nível de compactação ao violão, seguido do piano, do tímpano e do clarinete. Observa-se que, em todos os casos, as quantidades de coeficientes encontradas foram menores que as sugeridas pelo PEAQ, o que numa primeira análise poderia indicar pequeno grau de exigência dos ouvintes. Mas além disso, somente o

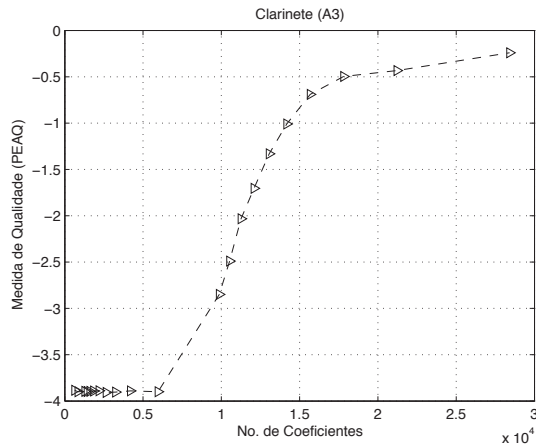


Figura 2: Avaliação de Qualidade através do PEAQ para diferentes números de coeficientes na representação do sinal de clarinete (A3).

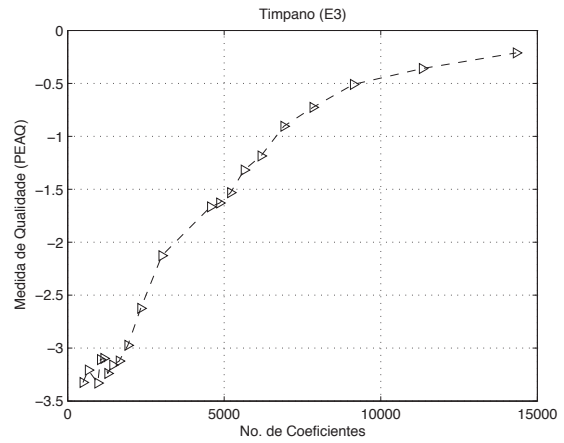


Figura 4: Avaliação de Qualidade através do PEAQ para diferentes números de coeficientes na representação do sinal de timpano (E3).

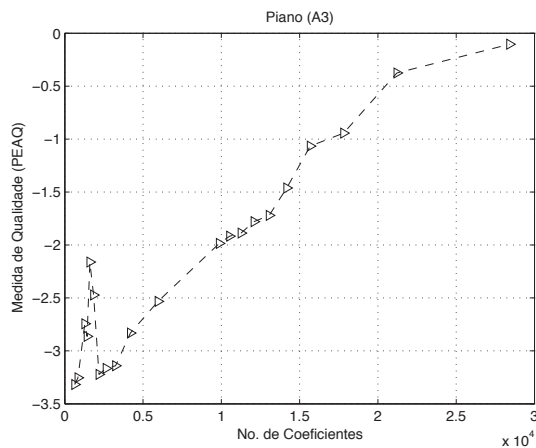


Figura 3: Avaliação de Qualidade através do PEAQ para diferentes números de coeficientes na representação do sinal de piano (A3).

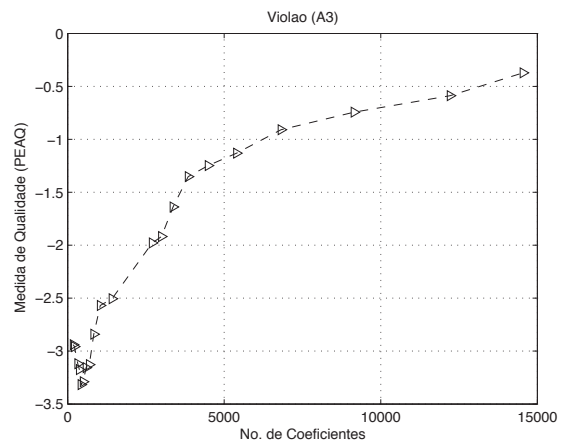


Figura 5: Avaliação de Qualidade através do PEAQ para diferentes números de coeficientes na representação do sinal de violão (A3).

Tabela 1: Resultados de testes informais de audição.

	Clarinete (A3)	Piano (A3)	Tímpano (E3)	Violão (A3)
Percentual da norma inicial como critério de parada (%)	1	2	1	4
Número mínimo de coeficientes	≈ 10000	≈ 3000	≈ 4500	≈ 800
Percentual de energia representado (%)	99,90	99,83	99,89	99,84

clarinete e o violão foram classificados na mesma ordem, quanto à compactação. A discrepância entre os resultados dos testes subjetivos e objetivos pode ser causada em parte pelo fato de o PEAQ ter sido treinado com base em codificadores de fundamentação psicoacústica. A adaptação dessa medida objetiva de qualidade para aplicação a sinais que sofreram outras for-

mas de modificação é um tópico a ser investigado.

As Figuras de 6 a 9 mostram as formas de onda originais das notas dos 4 instrumentos e o número de coeficientes usados, quadro a quadro, para reconstruir cada um deles no contexto da Tabela 1.

Nota-se que os trechos de baixa energia no início da Figura 6 (clarinete) e no final da Figura 9 (violão) precisaram de uma grande quantidade de coeficientes para serem representados. Isso sugere que nessas porções de sinal já resta pouca correlação entre o resíduo e os átomos do dicionário.

A Figura 10 ilustra o comportamento da norma do resíduo em função das iterações para dois quadros do sinal de clarinete: 4 e 50. Verifica-se o mesmo comportamento exponencial. Contudo, uma derivada reduzida da envoltória exponencial da norma indica que os átomos escolhidos do dicionário aproximam cada vez menos o resíduo. Isso explica que sejam utilizados mais átomos para o quadro 4, de baixa energia, que para o quadro

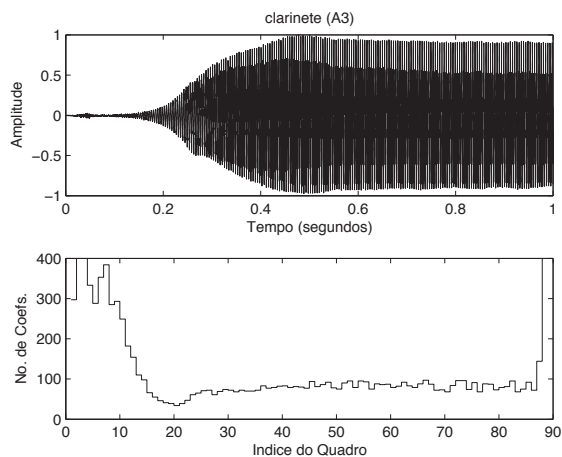


Figura 6: Sinal do clarinete (A3) e número de coeficientes na representação dos quadros.

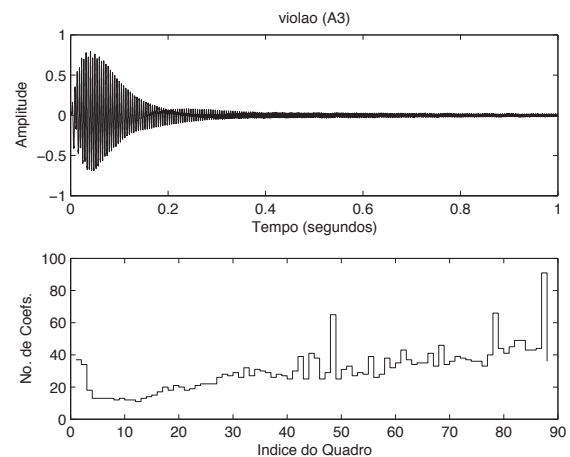


Figura 9: Sinal do violão (A3) e número de coeficientes na representação dos quadros.

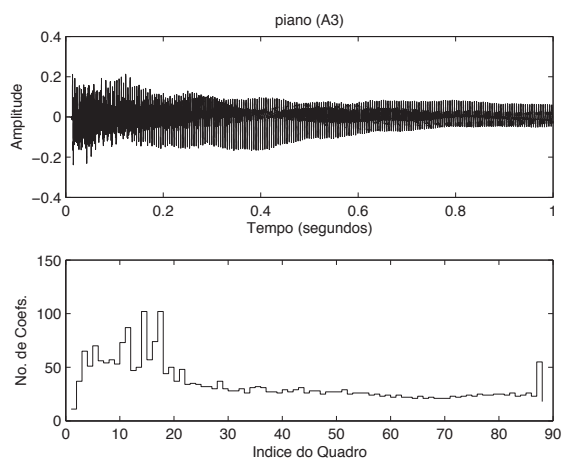


Figura 7: Sinal do piano (A3) e número de coeficientes na representação dos quadros.

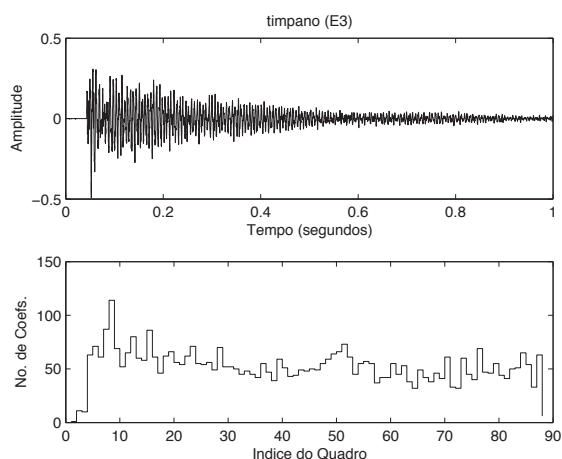


Figura 8: Sinal do tímpano (E3) e número de coeficientes na representação dos quadros.

50, de alta energia. Portanto, adotar um percentual da energia do resíduo inicial como energia do resíduo que decide a parada da decomposição não parece ser adequado.

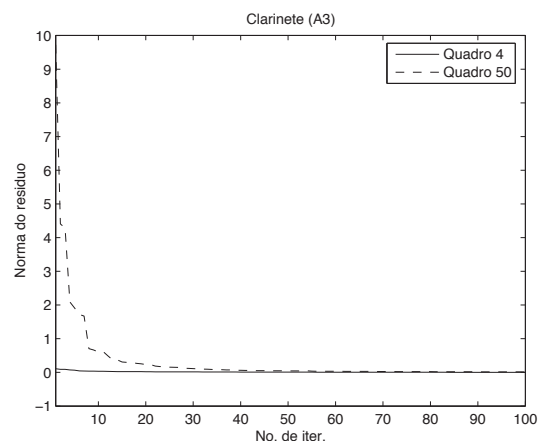


Figura 10: Comportamento da norma do resíduo nos quadros 4 e 50 do sinal de clarinete (A3)

A Figura 11 ilustra a taxa de aproximação (Eq. (12)) em função do número de iterações do MP. Observa-se que esta figura de mérito consegue traduzir a capacidade do dicionário de aproximar o resíduo a cada iteração. Essa capacidade diminui exponencialmente da mesma forma que a norma do resíduo. Quando a sua taxa de variação média se torna pequena é porque os elementos do dicionário passaram a ter baixa correlação com o resíduo, e, portanto, não é mais preciso prosseguir com a decomposição. Isso indica que pode ser interessante utilizar a taxa de aproximação como critério de parada para o MP cromático. Esse critério apresentou um bom desempenho quando aplicado a sinais elétricos [7, 8, 9], e é um tópico correntemente sob investigação.

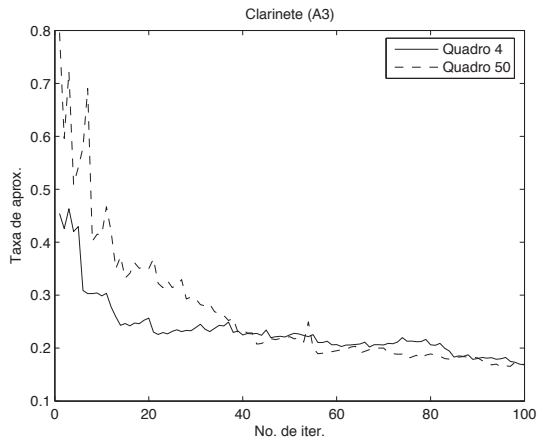


Figura 11: Taxa de aproximação dos quadros 4 e 50 do sinal de clarinete (A3)

## 5 CONCLUSÃO

Neste artigo, avaliaram-se potencialidades e deficiências do método de *Matching Pursuit* quando aplicado a áudio. Implementou-se um esquema de análise/síntese de sinais de áudio utilizando MP cromático. Seu desempenho para diversas famílias de sinais foi avaliado pelo *Perceptual Evaluation of Audio Quality* (PEAQ) em função do número de coeficientes empregados na representação e por testes informais de audição, com alguma discrepância, sugerindo a necessidade de se modificar o PEAQ para ampliar seu alcance. Por fim, verificou-se que a taxa de aproximação possui um bom potencial como critério de parada da decomposição.

Na continuação imediata deste trabalho, deverão ser realizados testes sistemáticos sobre um conjunto mais extenso de sinais, contemplando outros naipes e instrumentos, e com inclusão controlada de ruído. À medida que outros desenvolvimentos forem incorporados ao método apresentado, seu desempenho deve ser comparado ao de outros da literatura.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. on Information Theory*, vol. 50, no. 10, pp. 2231–2242, Out. 2004.
- [2] O. Derrien, "Multi-scale frame-based analysis of audio signals for musical transcription using a dictionary of chromatic waveforms," in *IEEE ICASSP'06 - Int. Conf. on Acoustics, Speech, and Signal Processing*, Toulouse, França, Maio 2006, vol. 5, pp. 57–60.
- [3] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. on Signal Processing*, vol. 51, no. 1, pp. 101–111, Jan. 2003.
- [4] M.M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal model based on recursive filter banks," *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 1890–1902, Jul. 1999.
- [5] L. Daudet, "Sparse and structured decompositions of signals with molecular matching pursuit," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1890–1902, Set. 2006.
- [6] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dez. 1993.
- [7] L. Lovisolo, E.A.B. da Silva, M.A.M. Rodrigues, and P.S.R. Diniz, "Efficient coherent representations of power systems signals using damped sinusoids," *IEEE Trans. on Signal Processing*, vol. 53, no. 10, pp. 3831–3846, Out. 2005.
- [8] M.P. Tcheou, L. Lovisolo, E.A.B. da Silva, M.A.M. Rodrigues, and P.S.R. Diniz, "Optimum rate-distortion dictionary selection for compression of atomic decompositions of electric disturbance signals," *IEEE Signal Processing Letters*, vol. 14, no. 2, pp. 81–84, Fev. 2007.
- [9] L. Lovisolo, M.P. Tcheou, E.A.B. da Silva, M.A.M. Rodrigues, and P.S.R. Diniz, "Modeling of electric disturbance signals using damped sinusoids via atomic decompositions and its applications," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. Article ID 29507, 15 pp., 2007, doi:10.1155/2007/29507.
- [10] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.
- [11] P. Kabal, "An examination and interpretation of ITU-R BS.1387: perceptual evaluation of audio quality," Tech. Rep., TSP Lab Tech. Report, Dept. Elec. and Comp. Eng., McGill Univ., vers. 2, Montreal, Canadá, Dez. 2003.
- [12] T.V. Verma and T.H. Meng, "Sinusoidal modeling using frame-based perceptually weighted matching pursuits," in *IEEE ICASSP'99 - Int. Conf. on Acoustics, Speech, and Signal Processing*, Phoenix, EUA, Mar. 1999, vol. 2, pp. 981–984.
- [13] M.G. Christensen and S. van de Par, "Efficient parametric coding of transients," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1340–1351, Jul. 2006.
- [14] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency approximations with matching pursuits," Tech. Rep., TR1994-657 Tech. Report, Computer Science Dept., New York Univ., Nova Iorque, EUA, Mar. 1994.