Codificação perceptiva de áudio por meio de decomposições atômicas em exponenciais complexas

Valmir dos Santos Nogueira Junior [1], Michel Pompeu Tcheou [2], Flávio Rainho Ávila[3]

[1] vsnjunior@gmail.com. [2] mtcheou@uerj.br. [3] flavio.avila@uerj.br. Universidade do Estado do Rio de Janeiro (UERJ) / Laboratório de Processamento de Sinais, Aplicações Inteligentes e Comunicações (PROSAICO).

RESUMO

A decomposição atômica de sinais por algoritmo da classe *Matching Pursuit* (MP) vem sendo aplicada à compressão de áudio. De acordo com a literatura, a utilização de critérios psicoacústicos possibilita uma representação mais compacta do sinal, sem perda de qualidade percebida. Neste trabalho é apresentada a implementação de um sistema de análise por síntese de sinais de áudio utilizando MP associado ao uso de limiar de mascaramento global psicoacústico, inspirado na camada I do MPEG, além de Dicionários de Exponenciais Complexas (DEC). Para a compressão do sinal, utiliza-se a otimização taxa-distorção por curvas operacionais, ajustando-se o multiplicador de Lagrange. O desempenho do método de compressão para diferentes tipos de sinais é avaliado por uma medida objetiva padronizada pela *International Telecommunications Union* (ITU), o *Perceptual Evaluation of Audio Quality* (PEAQ) em função da taxa de bits por amostra, obtendo-se resultados satisfatórios.

Palavras-chave: Matching Pursuit. Decomposição atômica de sinais. Psicoacústica.

ABSTRACT

The atomic decomposition of signals by algorithm of the class "Matching Pursuit" (MP) has been applied in audio compression. Literature review suggests that, the use of psychoacoustic criteria allows a more compact representation of the signal, without loss of perceived quality. This work presents the implementation of an analysis system by synthesis of audio signals using MP associated with the use of psychoacoustic global masking threshold, inspired by MPEG layer I, as well as Complex Exponential Dictionaries (DEC). For the compression of the signal, we used the optimization of rate-distortion by operational curves, adjusting the Lagrange multiplier. The performance of the compression method for different types of signals is evaluated by an objective measurement standardized by the International Telecommunications Union (ITU), the PEAQ (Perceptual Evaluation of Audio Quality) based on the bit rate per sample, obtaining satisfactory results.

Keywords: *Matching Pursuit. Atomic decomposition of signals. Psychoacoustic.*

SUBMETIDO 02/01/2019 > APROVADO 12/04/2019

1 Introdução

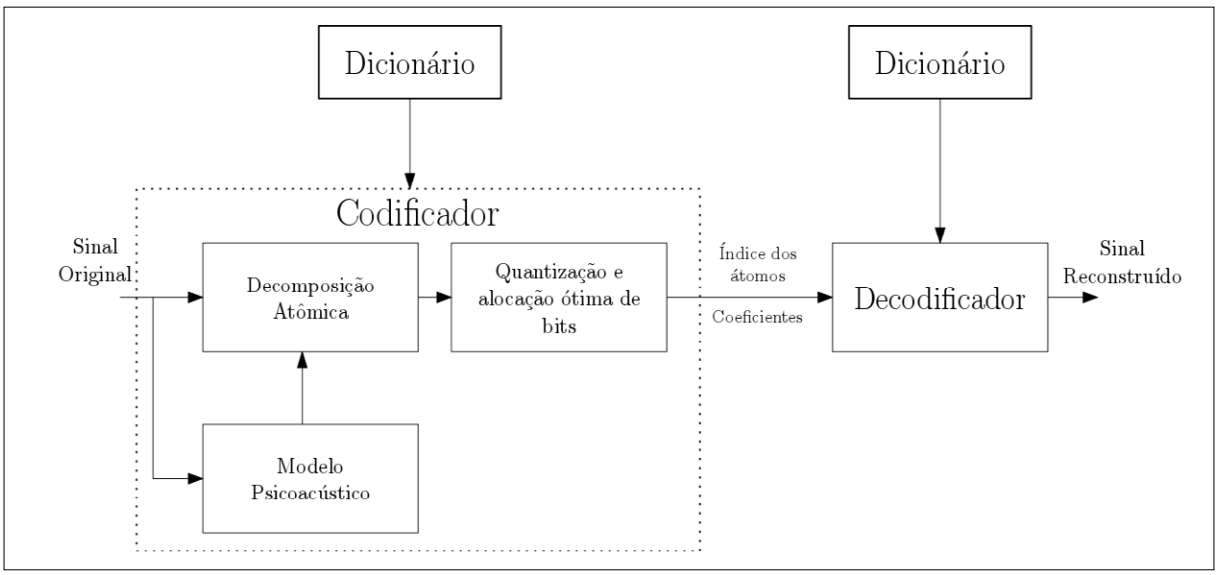
No processo de obtenção de representações compactas de sinais, deve-se buscar expandir esses sinais em funções-base que apresentem uma grande similaridade com suas estruturas complexas (MALLAT; ZHANG, 1993). A modelagem de sinais torna possível descrever matematicamente, e de forma suficientemente precisa, seus fenômenos intrínsecos por meio de ferramentas que propiciam a análise e a síntese desses sinais. Uma poderosa técnica de decomposição de sinais, introduzida por Mallat e Zhang (1993), é o *Matching Pursuit* (MP). O algoritmo MP calcula iterativamente a decomposição do sinal em funções (ou átomos), selecionando, a cada iteração, dentre um conjunto de funções que formam um dicionário, aquelas que melhor se correlacionam com o sinal em análise. Na aplicação apresentada neste artigo, relacionada à representação compacta de sinais de áudio, o sinal é decomposto em formas de ondas selecionadas desse dicionário de átomos de características tempo-frequenciais, que, por sua vez, é gerado, normalmente, a partir de dilatações, translações e modulações de uma função janela simples (PETROVSKY; HERASIMOVICH; PETROVSKY, 2015).

Para a obtenção de uma representação maximamente compacta do sinal de áudio original, é fundamental explorar aspectos perceptivos da audição, os quais informam que componentes tempo-frequenciais serão realmente audíveis (MALLAT; ZHANG, 1993). Para tanto, é preciso entender conceitos de psicoacústica, que é a ciência da percepção sonora humana, em especial os conceitos de audibilidade e de mascaramento (TOUMI; DERRIEN, 2015).

Ainda com o objetivo de tornar a representação compacta, é preciso garantir um número baixo de bits e alto desempenho, assegurando a máxima fidelidade do sinal. Nesse sentido, a quantização é responsável pela compressão dos coeficientes das representações atômicas dos sinais de áudio. Para se obter tal objetivo, é utilizada a otimização taxa-distorção por curvas operacionais, ajustando-se o multiplicador de Lagrange.

A Figura 1 ilustra um esquema de codificação e decodificação de sinais utilizando decomposições atômicas com base em um dicionário de funções elementares, incluindo aspectos psicoacústicos para a seleção dos átomos.

Figura 1 – Esquema de codificação e decodificação de sinais com base em decomposição atômica psicoacústica



Fonte: Elaboração própria

O codificador analisa o sinal de forma a obter uma boa representação com base em um dicionário e no modelo psicoacústico, encontrando os coeficientes e os índices da representação atômica da decomposição de cada quadro do sinal. O modelo psicoacústico é obtido a partir do sinal original e serve como referência para a seleção de átomos audíveis por meio da decomposição

atômica. Portanto, ele é calculado preliminarmente ao processo de decomposição. Os coeficientes e os índices da representação do sinal são quantizados e transmitidos ao decodificador, onde é efetuada a reconstrução do sinal com base no mesmo dicionário usado no codificador.

O presente trabalho tem como objetivo apresentar o desenvolvimento de um sistema para decomposição de sinais de áudio, com o auxílio do algoritmo do *Matching Pursuit*, utilizando o princípio de relevância psicoacústica dos componentes do sinal. A elaboração do trabalho foi baseada no artigo de Verma e Meng (1999), no qual os autores utilizam o MP com dicionário de exponenciais complexas e ponderação psicoacústica. No lugar da função de ponderação, a curva psicoacústica obtida pelo modelo MPEG-1 (camada I) (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 1993) foi utilizada neste artigo. À medida que a energia do resíduo obtida no processo iterativo do MP passa a estar abaixo da máscara psicoacústica em determinadas faixas espectrais, as exponenciais complexas de frequências referentes a essas faixas são descartadas nas iterações posteriores. Neste trabalho, a aferição dos resultados obtidos foi realizada com uma ferramenta de avaliação perceptiva de qualidade de áudio, o *Perceptual Evaluation of Audio Quality* –PEAQ (THIEDE *et al.*, 2000).

Em trabalho recente (PETROVSKY; HERASIMOVICH; PETROVSKY, 2016), é proposto um codificador de sinais de áudio e de voz, em que são empregados o algoritmo de *Matching Pursuit* e uma estratégia de seleção de átomos com base em psicoacústica. Nesse caso, o sinal é dividido em blocos, e a cada bloco processado, um dicionário de *Wavelet-Packet* é perceptivamente otimizado. Dessa maneira, o dicionário se altera a cada bloco a partir de uma seleção preliminar dos átomos possivelmente audíveis antes da decomposição. Além disso, reporta-se o uso de um codificador de entropia agregado, alcançando-se uma taxa mínima de codificação em 62 kbps sem que a distorção inserida seja perceptível ao ouvido humano.

Em contraste a Petrovsky, Herasimovich e Petrovsky (2016), no codificador proposto no presente trabalho o dicionário não se altera a cada quadro processado, e o discernimento entre os átomos audíveis ocorre durante a decomposição do sinal. Ademais, realiza-se uma alocação ótima por meio da otimização da taxa-distorção, porém sem agregar um codificador de entropia.

Na seção 2, o referencial teórico é apresentado, incluindo noções de psicoacústica, o detalhamento do modelo psicoacústico, bem como a descrição do método de decomposição atômica baseado no algoritmo de *Matching Pursuit* e na seleção de átomos de relevância psicoacústica a partir de um dicionário fixo de exponenciais complexas. Além disso, apresenta-se a estratégia de alocação ótima de bits por meio de quantização uniforme e otimização taxa-distorção. Na seção 3, a configuração e o procedimento experimentais são descritos. Na seção 4, os resultados e avaliação de desempenho do codificador de áudio proposto são apresentados. Por fim, na seção 5, as conclusões do trabalho são apresentadas.

2 Referencial teórico

Nesta seção, o referencial teórico necessário é apresentado para o desenvolvimento do codificador de sinais de áudio proposto neste artigo. Além disso, o próprio codificador é descrito em detalhes, abrangendo o método de decomposição atômica, com base em psicoacústica, e a estratégica de alocação ótima de bits.

2.1 Noções de psicoacústica

A psicoacústica é a ciência da percepção sonora que possui como objetivo principal o estudo das relações entre as magnitudes dos estímulos físicos e as magnitudes das sensações por eles produzidos (FASTL; ZWICKER, 2007). Ela apresenta características fundamentais na concepção de um codificador perceptivo de áudio, entre as quais se podem destacar as unidades de medida de níveis de pressão sonora (*Sound Pressure Level* – SPL), os limiares da audição humana, fenômenos de mascaramento e a escala Bark (LIN; ABDULLA, 2015).

A SPL é a grandeza que representa a intensidade de um determinado som, expressa em dB, como (BOSI; GOLDBERG, 2002):

$$SPL = 10log_{10}\left(\frac{P}{P_0}\right)^2 \quad (1)$$

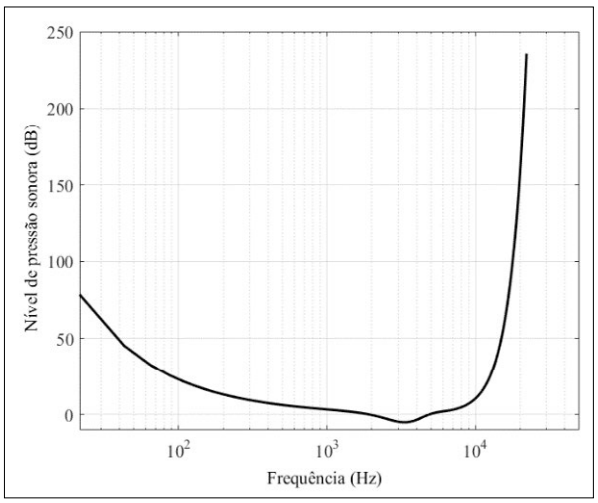
em que P é a pressão sonora no ponto em questão, e $P_0 = 20$ μPa é aproximadamente igual à pressão sonora no limiar de audição na frequência por volta de 2 kHz.

A sensação de audição que se relaciona com SPL é a sonoridade (*loudness*), expressa em *phon*

(LIN; ABDULLA, 2015). O nível de sonoridade é definido como o nível de um tom sonoro de 1 kHz, que é percebido tão alto quanto o som em análise para campos planos frontais incidentes (BOSI; GOLDBERG, 2002). A audição humana é capaz de responder a valores SPL em frequências que vão de 20 Hz até 20 kHz.

O limiar absoluto da audição (*Absolute Threshold of Hearing* – ATH) humana, representado na Figura 2, caracteriza a intensidade sonora necessária em um tom puro que pode ser detectado por um ouvinte em um ambiente silencioso (FASTL; ZWICKER, 2007), isto é, representa o menor nível de pressão sonora em decibéis que se pode ouvir em uma dada frequência. Os componentes de frequência do sinal que estejam abaixo desse nível serão irrelevantes para a percepção de sons e, portanto, não precisam ser armazenados ou transmitidos.

Figura 2 – Curva de limiar de audição humana



Fonte: Elaboração própria

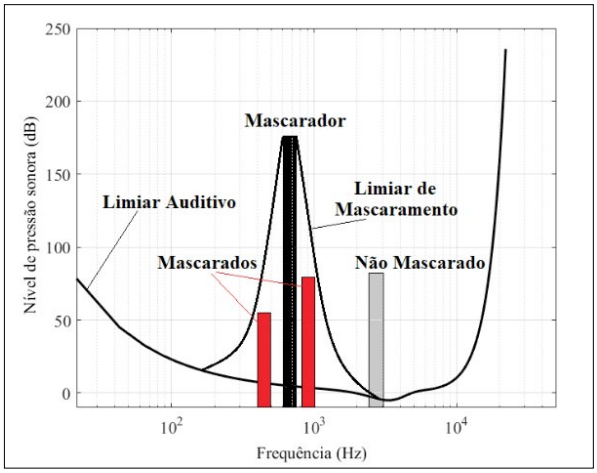
O limiar do silêncio, *A(f)*, dado em dB e dependente da frequência, pode ser representado como (BOSI; GOLDBERG, 2002):

$$A(f) = 3{,}64f^{(-0{,}8)} - 6{,}5e^{-0{,}6(f-3{,}3)^2} + 10^{-3}f^4 \quad (2)$$

Em que *f* é a frequência em kHz.

O fenômeno de mascaramento é de extrema importância para a codificação de sinais de áudio. Devido a esse fenômeno, a percepção de um som está relacionada não apenas com a sua própria frequência e intensidade, mas também com as de seus componentes vizinhos (LIN; ABDULLA, 2015). A Figura 3 apresenta um componente de sinal mascarador, alterando o limiar absoluto da audição e impedindo que componentes de frequências vizinhas com menor SPL sejam percebidos.

Figura 3 – Ilustração do fenômeno de mascaramento



Fonte: Elaboração própria

Existe uma faixa de frequências em torno da frequência do sinal mascarador na qual o limiar de mascaramento é plano. Essa faixa de mascaramento plana é conhecida como a banda crítica e está intimamente relacionada com a escala Bark. A equação que relaciona a escala Bark à frequência em Hertz é (BOSI; GOLDBERG, 2002):

$$z = 13arctan\left(\frac{0{,}76f}{1000}\right) + 3{,}5arctan\left(\frac{f}{7500}\right) \quad (3)$$

Em que *f* é a frequência em Hz.

O modelo psicoacústico adotado neste trabalho foi inspirado na camada I do padrão MPEG-1 (*Moving Pictures Experts Group*), que representa o primeiro padrão internacional que especifica um formato digital para áudio de alta qualidade (SPANIAS; PAINTER; ATTI, 2006).

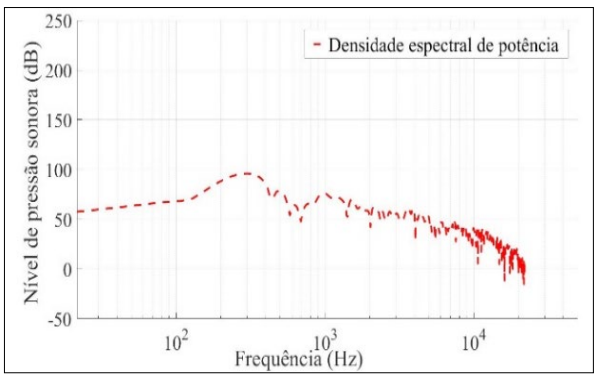
O padrão MPEG-1, conhecido como ISO/IEC 11172-3, descreve um algoritmo perceptivo de codificação de áudio projetado para sinais genéricos.

O MPEG-1 Áudio especifica três camadas que oferecem diversos níveis de qualidade de áudio com complexidade variada (BOSI; GOLDBERG, 2002).

O cálculo realizado pelo algoritmo do limiar de mascaramento psicoacústico global utilizado neste trabalho é consolidado em nove etapas:

1) Divisão do sinal temporal de entrada *x* do algoritmo em *Q* quadros com *N* pontos cada;
2) Multiplicação de quadro de sinal no domínio do tempo por uma janela de Hanning *w[n]* para atenuar os efeitos espectrais causados por transições abruptas;
3) Conversão do sinal do domínio do tempo para o domínio da frequência, por meio de uma transformada rápida de Fourier (*Fast Fourier Transform* – FFT) de *M* pontos, em que *M* é o tamanho máximo do dicionário a ser definido;
4) Obtenção das bandas críticas e da curva de limiar de silêncio;
5) Cálculo da densidade espectral de potência $X_i[k] = SPL_{PCM16} + 20\log_{10}|S_i[k]|$, em que *Si[k]* é a transformada de discreta de Fourier do *i*-ésimo quadro, e SPL_{PCM16} corresponde ao nível de pressão sonora ao se reproduzir um sinal senoidal de amplitude unitária codificado por um PCM (*Pulse Code Modulation*) de 16 bits com fundo de escala de -1 a 1, sendo igual a 96 dB (BOSI; GOLDBERG, 2002). Essa premissa é adotada, posto que não se tem conhecimento a priori do volume de reprodução do áudio codificado, sendo que quanto maior o volume, maior é o nível de pressão sonora. Nesse caso, a densidade espectral de potência, definida dessa forma, corresponde ao nível de pressão sonora (BOSI; GOLDBERG, 2002). A Figura 4 ilustra o nível de pressão sonora (SPL) obtido a partir da densidade espectral de potência de um quadro do sinal obtido pelo algoritmo.

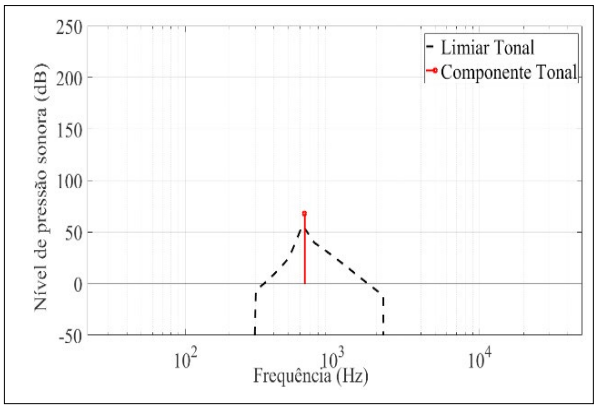
Figura 4 – Representação da densidade espectral de potência do sinal



Fonte: Elaboração própria

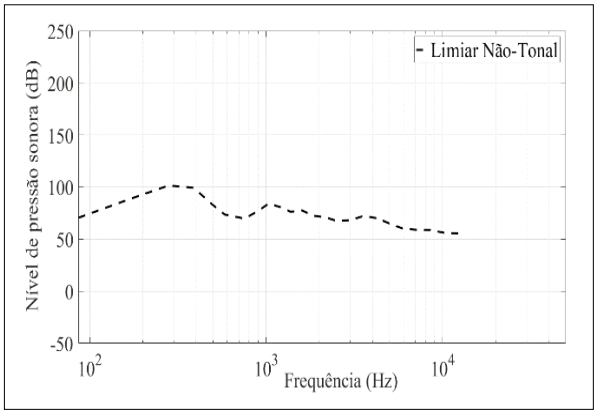
6) Determinação dos componentes tonais $XTM_i[k]$ e não tonais $XNT_i[k]$ do i-ésimo quadro em análise;
7) Determinação dos componentes mascaradores principais;
8) Cálculo do limiar de mascaramento para cada componente mascarador.
   A Figura 5 ilustra o limiar de mascaramento de um mascarador tonal, LT_{TM}, devido a um componente com frequência em 646 Hz e SPL de 68,11 dB.
   Na Figura 6, está ilustrado o limiar de mascaramento não tonal LT_{NT} consequente de diferentes mascaradores não tonais encontrados pelo algoritmo para o bloco em análise;
9) Determinação do limiar de mascaramento global.

Figura 5 – Representação do limiar de mascaramento tonal



Fonte: Elaboração própria

Figura 6 – Representação do limiar de mascaramento não tonal.



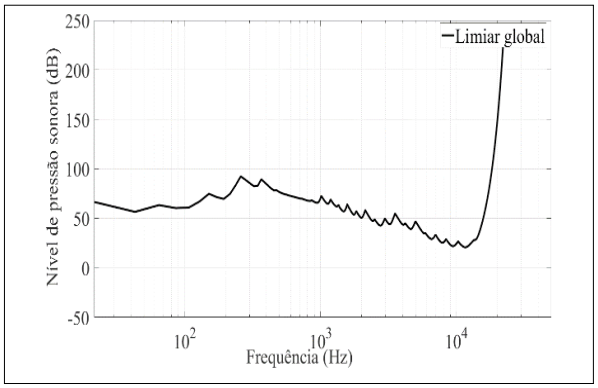
Fonte: Elaboração própria

O limiar global de mascaramento , representado na Figura 7, é obtido por uma combinação dos limiares individuais e do limiar de silêncio, dada por:

$$L_{TG}[k] = 10log_{10}(10^{\frac{ATH[k]}{10}} + \sum_{j}^{N_{TM}} 10^{\frac{LT_{TM}[Z(j),Z(k)]}{10}} + \sum_{j}^{N_{NT}} 10^{\frac{LT_{NT}[Z(j),Z(k)]}{10}}) \quad (4)$$

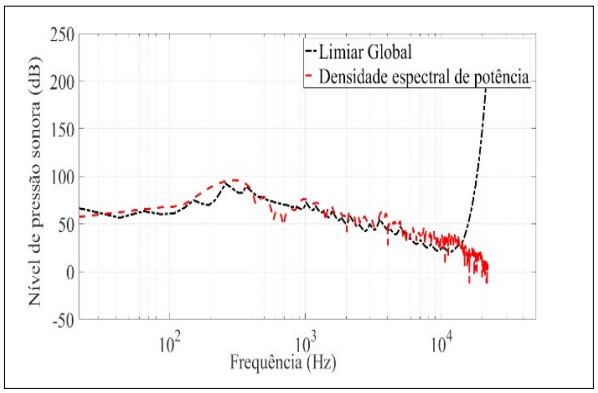
Em que *ATH[k]* é o SPL do limiar de silêncio na linha espectral, *k*, N_{TM} e N_{NT} são os números de mascaradores tonais e não tonais listados, e LT_{TM} e LT_{NT} são seus limiares de mascaramento individuais correspondentes. É possível observar, ao analisar a Figura 7, a influência de cada curva componente na sua estrutura. Nota-se, também, a existência de um número maior de componentes nas frequências mais altas do espectro em análise.

Figura 7 – Representação do limiar de mascaramento global



Fonte: Elaboração própria

Figura 8 – Resultado final do cálculo do limiar de mascaramento global



Fonte: Elaboração própria

A Figura 8 retrata o resultado final do cálculo do limiar de mascaramento global realizado pelo algoritmo para um determinado trecho de sinal de áudio, com a densidade espectral de potência de um sinal arbitrário.

2.2 Decomposição atômica psicoacústica

Os algoritmos de compressão de sinais são utilizados para produzir representações compactas de sinais de alta qualidade, quando se quer expandir esses sinais por meio de funções que apresentem alto nível de similaridade com as suas estruturas complexas (MALLAT, 2008).

As decomposições atômicas têm como objetivo selecionar um subconjunto de elementos, denominados átomos ou estruturas, a partir de um dicionário de formas de onda predefinidas, a fim de aproximar um sinal como uma combinação linear desses elementos (DAVIS; MALLAT; ZHANG, 1994).

Considerando que o sinal pode ser aproximado por átomos que compõem uma família de vetores de um dicionário , pertencentes ao espaço de Hilbert , tem-se que (FERRANDO; KOLASA; KOVACEVIC, 2002):

$$\boldsymbol{x} \approx \sum_{k=0}^{K-1} \alpha_{m_k} \boldsymbol{g}_{m_k} \quad (5)$$

Os átomos possuem e são indexados por , que é definido como ; é o número de elementos do dicionário , portanto . O parâmetro é o coeficiente que pondera e corresponde ao número de átomos selecionados para representar .

A utilização de dicionários altamente redundantes possibilita a extração direta de uma variedade maior de padrões e fenômenos presentes em sinais, resultando em representações mais compactas e eficientes (MALLAT; ZHANG, 1993).

2.2.1 *Matching Pursuit*

O *Matching Pursuit* (MP) é um algoritmo que decompõe um sinal e o representa como uma expansão linear de formas de onda ou funções (MALLAT; ZHANG, 1993). A cada etapa, o algoritmo procura em seu dicionário uma função que combina melhor com o sinal atual e extrai deste uma versão escalada daquela do sinal corrente, produzindo o resíduo. O *Matching Pursuit* continua a ser aplicado nesse sinal residual até que seu critério de parada seja atendido.

Desejamos representar um sinal de dimensão *N* em um conjunto de *M* coeficientes, em que *M* < *N* (MALLAT; ZHANG, 1993). Um dicionário redundante apresenta uma cardinalidade maior que a dimensão *N* do sinal, propiciando alto grau de liberdade na construção da expansão de funções.

Em nossa aplicação, cada átomo g_m é caracterizado por parâmetros de amplitude *A*, frequência f e fase θ, por meio de aproximações sucessivas de x, a partir de projeções ortogonais envolvendo os elementos do dicionário.

Seja o sinal inicial x, com o resíduo inicial sendo definido como $r_x^0 = x$; supondo que o resíduo da *k*-ésima ordem r_x^k já está calculado para $k \geq 0$; a próxima escolha de m_k é tal que:

$$\alpha_k = argmax_{k \epsilon M} |\langle r_x^k, g_{m_k} \rangle| \quad (6)$$

e projetar r_x^k em g_{m_k} e subtrair de r_x^k:

$$r_x^k = \langle r_x^k, g_{m_k} \rangle g_{m_k} + r_x^{k+1} \quad (7)$$

em que r_x^{k+1} é o resíduo da *"k + 1"*-ésima iteração, r_x^k é o resíduo do sinal sendo na *k*-ésima iteração decomposto e o operador $< y, z >$ representa o produto interno entre os vetores y e z.

Assim, o sinal original x é decomposto em uma soma ponderada de elementos de dicionário escolhidos que melhor correspondem aos resíduos obtidos de forma iterativa.

2.2.2 Dicionário de exponenciais complexas

Um dicionário com alta correlação com as formas dos sinais analisados permite baixo erro de aproximação. Neste trabalho, o dicionário de exponenciais complexas (DEC) foi utilizado, cujos elementos são definidos da seguinte maneira (VERMA; MENG, 1999):

$$g_m = \{g_m[n] = \frac{1}{N} e^{j2\pi \frac{m}{M} n}\} \quad (8)$$

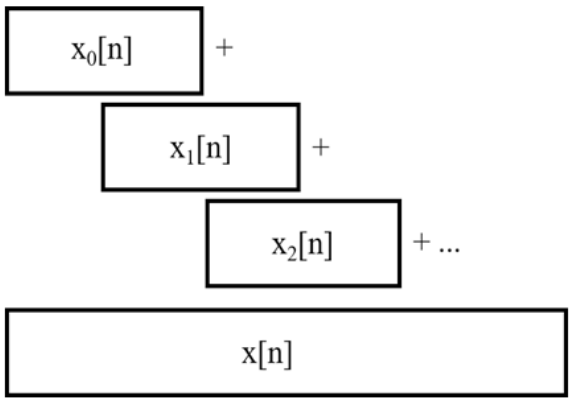
em que $n = 0, 1, \ldots, N-1$ e $m = 0, 1, \ldots, M-1$. Em caso de sinais reais, os coeficientes de correlação aparecem em pares conjugados; assim, somente metade dos coeficientes de correlação é pesquisada (VERMA; MENG, 1999).

A escolha desse dicionário se deve ao fato de que seus elementos apresentam alto grau de similaridade com relação aos padrões encontrados em sinais de áudio.

2.2.3 Algoritmo de decomposição

O algoritmo desenvolvido neste trabalho é baseado em Verma e Meng (1999), que propõem o algoritmo de *Matching Pursuit*, incorporando características perceptivas do ouvido humano. O processamento do sinal é realizado bloco a bloco, sendo que cada bloco corresponde a um trecho janelado do sinal, podendo haver sobreposição entre blocos adjacentes, conforme ilustrado na Figura 9. Nesse caso, o sinal é reconstruído por meio de um procedimento de sobreposição e adição entre os blocos.

Figura 9 – Esquema da representação da divisão do sinal em blocos



Fonte: Elaboração própria

Seja o *i*-ésimo bloco do sinal $x[n]$ em que $i = 0,1,\ldots,Q$, Q é o número de blocos, l é o comprimento da largura da janela e $w[n]$ possui comprimento *N*. É importante que $\sum_{l=0}^{Q-1} w[n - il] = 1$, de modo que a reconstrução seja perfeita.

$$x_l[n] = w[n]x[n - lp] \quad (9)$$

Em Verma e Meng (1999), propõe-se uma implementação baseada em Transformada Discreta de Fourier (DFT), que permite utilizar algoritmos rápidos como a *Fast Fourier Transform* (FFT). Para aproveitar ao máximo a FFT, a cardinalidade *M* do dicionário deve ser potência de 2.

O processo de inserir zeros no final do sinal é usado para que se possa utilizar a FFT, sendo denominado de *zero-padding*. Inicialmente, introduz-

se uma generalização do produto interno para um produto interno ponderado, , em que é uma matriz positiva definida simétrica, de formato diagonal, na qual a diagonal é composta pelas amostras da janela . Na *k*-ésima iteração do MP, calcula-se, para (VERMA; MENG, 1999):

$$\frac{|\langle \boldsymbol{g}_m, \boldsymbol{r}_k \rangle_{\boldsymbol{w}}|}{\langle \boldsymbol{g}_m, \boldsymbol{g}_m \rangle_{\boldsymbol{w}}} = \frac{R_k^w[\frac{m}{M}]}{W[\frac{0}{M}]} \quad (10)$$

em que

$$R_k^w\left[\frac{m}{M}\right] = \sum_{n=0}^{M-1} w[n] r_k[n] e^{-j2\pi\frac{m}{M}n} \quad (11)$$

e

$$W\left[\frac{0}{M}\right] = \sum_{n=0}^{M-1} w[n] \quad (12)$$

Em seguida, busca-se o máximo produto interno normalizado, dado por:

$$m_k = argmax_m \frac{|\langle \boldsymbol{g}_m, \boldsymbol{r}_k \rangle|_{\boldsymbol{w}}}{\langle \boldsymbol{g}_m, \boldsymbol{g}_m \rangle_{\boldsymbol{w}}} \quad (13)$$

Portanto, o *k*-ésimo coeficiente é dado por:

$$\alpha_k = \frac{R_k^w\left[\frac{m_k}{M}\right]}{W\left[\frac{0}{M}\right]} = \frac{A_k e^{j\theta_k}}{W\left[\frac{0}{M}\right]} \quad (14)$$

em que $\boldsymbol{A}_k$ e $\boldsymbol{\theta}_k$ são o módulo e a fase de $R_k^w[m_k/M]$, respectivamente.

Dado que sinais de áudio são reais, $\boldsymbol{x}_l[\boldsymbol{n}]$ também será real. Nesse caso, $R_k^w[m_k/M]$ possui simetria Hermitiana, sendo necessário calcular somente metade das correlações a cada iteração, ou seja, variando $m = 0, \ldots, M/2$.

No algoritmo de decomposição atômica psicoacústica implementado neste trabalho, inicialmente a cada bloco, o limiar de mascaramento global é calculado, a FFT (de comprimento *M*) da janela é armazenada, $W\left[\frac{m}{M}\right]$; $r_0[n] = x_l[n]$ é definido e a FFT (de comprimento *M*) de $r_0[n]w[n]$ é calculada, ou seja, $R_0^w[m/M]$. No processo do MP, o seguinte procedimento é realizado a cada iteração:

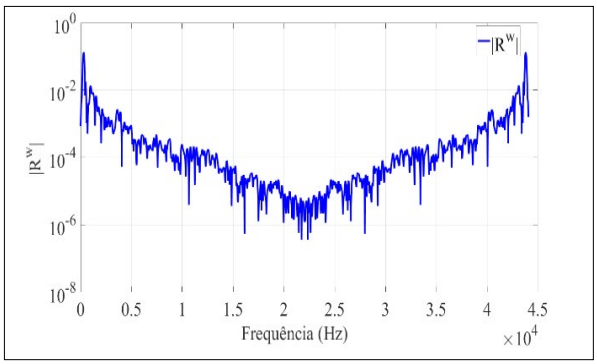
1) Calcula-se a transformada do resíduo corrente janelado, $R_k^w[m/M]$, (ver Equação 11);
2) Encontra-se o valor máximo absoluto do conjunto resultante, obtendo-se os parâmetros senoidais de amplitude $\boldsymbol{A}_k$, frequência $\boldsymbol{f}_k$ e fase $\boldsymbol{\theta}_k$. As Figuras 10 e 11 ilustram as respostas em frequência do resíduo e do átomo selecionado, respectivamente;
3) Calcula-se a transformada do resíduo janelado da próxima iteração:

$$R_{k+1}^w\left[\frac{m}{M}\right] = R_k^w\left[\frac{m}{M}\right] - \frac{A_k(e^{i\theta_k}W[\frac{m-m_k}{M}])}{2} - \frac{A_k(e^{-i\theta_k}W[\frac{m+m_k}{M}])}{2} \quad (15)$$

A Figura 12 na página seguinte ilustra a remoção átomo com máxima correlação encontrado no sinal;

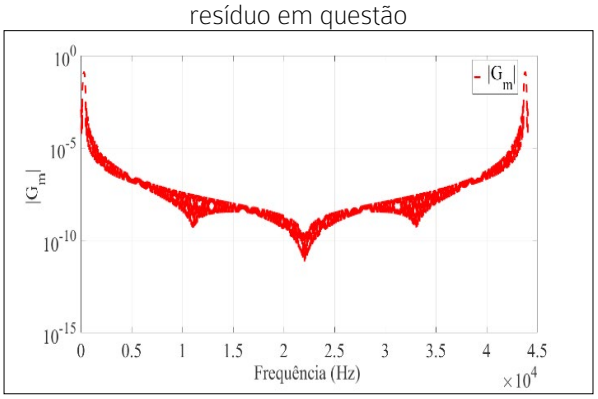
4) Se $R_{k+1}^w[m/M]$ em dB_{SPL} estiver abaixo do limiar global de mascaramento para todas as frequências, interrompe-se o processo iterativo;
5) Os elementos correspondentes às frequências com níveis de pressão (SPL) abaixo do limiar global de mascaramento são removidas do dicionário.

Figura 10 – Resposta em frequência do resíduo correspondente a um quadro de um sinal de áudio



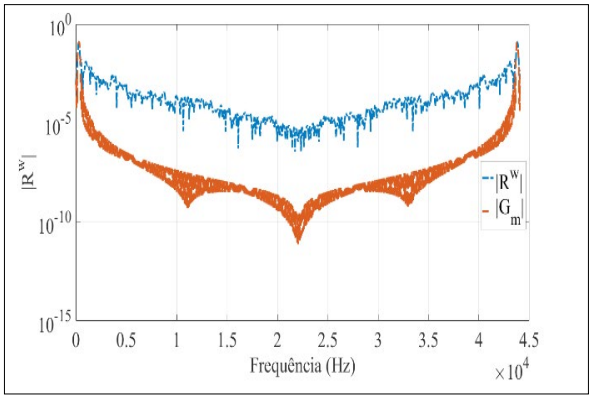
Fonte: Elaboração própria

Figura 11 – Resposta em frequência do átomo do dicionário de maior correlação tendo em vista o resíduo em questão



Fonte: Elaboração própria

Figura 12 – Remoção dos máximos encontrados no sinal de entrada em escala logarítmica



Fonte: Elaboração própria

Ao final do processo iterativo, obtém-se uma representação do bloco

$$x_l[n] = \sum_{k}^{N_{iter}-1} 2a_k \cos[2\pi f_k n + \theta_k] \qquad (16)$$

em que N_{iter} é o número de iterações, $a_k = A_k/M[0/M]$ e $f_k = m_k/M$. Observe que, a cada bloco, a complexidade de inicialização é $0(2M log 2M)$ em função das duas FFTs e, a cada iteração, $0(2M log 2M + 2M)$ em função das subtrações da Equação 15. A função O(·) denota uma medida de execução de algoritmo, em geral relacionada a tempo e memória, dado um problema de tamanho M (CORMEN *et al.*, 2009). Nesse caso, mais especificamente, corresponde ao número de multiplicações.

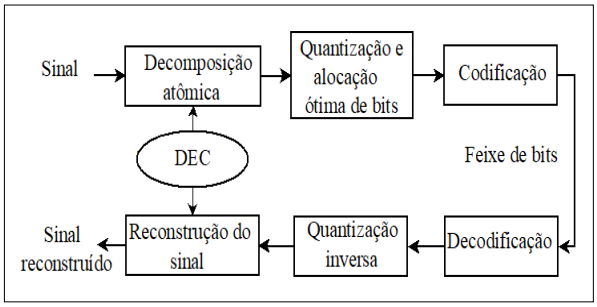
2.3 Alocação ótima de bits

O objetivo principal da compressão ou codificação é representar o sinal com o número mínimo de bits. O processo de representar um número infinito de valores com um conjunto finito de símbolos é denominado de quantização (SAYOOD, 2017).

Altos níveis de compressão são possíveis de serem obtidos ao custo da representação imperfeita da fonte. A troca entre a fidelidade da fonte e a taxa de codificação é exatamente o compromisso taxa-distorção (ORTEGA; RAMCHANDRAN, 1998).

O sistema de compressão de áudio proposto neste trabalho é apresentado na Figura 13. Inicialmente, o método descrito de decomposição atômica psicoacústica é utilizado, no qual o algoritmo seleciona um dicionário de exponenciais complexas (DEC) parametrizadas para um subconjunto de átomos, que são os mais correlacionados com os padrões existentes no sinal. Definindo-se $\boldsymbol{x}$ como o sinal original, obtém-se o sinal aproximado $\hat{\boldsymbol{x}}$ com K termos, conforme representado na Equação 5. Ao final da decomposição, tem-se a sequência dos pares (a_{m_k}, m_k) com $k = 1, 2, \ldots, K$, que formam o livro de estruturas, cujos parâmetros são dados por $m_k = (a_k; f_k; \phi_k)$.

Figura 13 – Compressão de sinais de áudio via decomposição atômica do sinal utilizando o DEC e o binômio taxa-distorção por meio de curvas operacionais



Fonte: Elaboração própria

Após a etapa de quantização, a otimização da taxa-distorção é realizada por meio de curvas operacionais, que permitem a alocação ótima de bits. Definida a alocação ótima de bits entre os coeficientes e parâmetros dos átomos, o livro de estruturas é quantizado, produzindo os símbolos que são codificados e transmitidos ao decodificador. No decodificador, o feixe de bits é decodificado, gerando os símbolos, que, por sua vez, sofrem o processo de

quantização inversa, produzindo o livro de estruturas quantizado. Por fim, com base neste livro de estruturas, o sinal é reconstruído.

2.3.1 Quantização

O tipo de quantizador mais simples é o escalar uniforme, em que todos os intervalos possuem tamanho único. Os quantizadores podem ser do tipo *midrise*, que não possuem zero no seu nível de saída, ou do tipo *midtread*, que possuem um nível de saída zero. Se R for o número de bits, o quantizador *midtread* permite utilizar 2^R - 1 códigos diferentes em relação aos 2^R códigos permitidos pelo quantizador *midrise*. Os quantizadores do *midtread* produzem melhores resultados (BOSI; GOLDBERG, 2002), pois é possível representar períodos de silêncio.

O coeficiente α e cada parâmetro de m_k são quantizados utilizando-se um quantizador escalar uniforme definido como (SAYOOD, 2017):

$$x_q = I_x \Delta_{q(x)}, onde\ I_x = \left\lfloor \frac{x + \Delta_{q(x)}/2}{\Delta_{q(x)}} \right\rfloor \quad (17)$$

em que x é qualquer parâmetro, x_q representa a sua versão quantizada, $\Delta_{q(x)}$ é o passo de quantização, e l_x corresponde ao símbolo associado a x. Os parâmetros são quantizados de acordo com um intervalo dinâmico definido por seus respectivos valores máximo e mínimo (SAYOOD, 2017):

$$\Delta_{q(x)} = \begin{cases} \frac{x_{max} - x_{min}}{2^{b_x} - 1}; \text{se for } midrise \\ \frac{x_{max} - x_{min}}{2^{b_x} - 2}; \text{se for } midtread \end{cases} \quad (18)$$

Em que b_x é o número de bits alocados a x.

Os parâmetros de amplitude (a_k) são quantizados de acordo com seus respectivos valores de amplitude máxima (a_{max}) e mínima (a_{min}). A fase Φ_k é uniformemente quantizada fazendo-se o seu valor máximo igual a $\Phi_{max} = 2\pi$ e o seu valor mínimo igual a $\Phi_{min} = 0$. A frequência f_x é quantizada de acordo com a discretização da frequência do dicionário de exponenciais complexas. O número de bits alocados para a frequência é $r_f = log2(M/2)$ *bits*, em que $k = 0, 1, ..., M$ e M é a cardinalidade do dicionário.

Assim, pode-se definir o número de bits associados à representação (ou codificação) de um átomo como:

$$r = r_a + r_f + r_\Phi \quad (19)$$

em que r_a é a quantidade de bits alocados à amplitude, r_f é a quantidade de bits alocados à fase, e r_Φ é a quantidade de bits alocados para a frequência. Dessa forma, o número total de bits gastos é rN_{iter}, em que N_{iter} é o número de iterações necessário para a decomposição do sinal do *i*-ésimo bloco.

2.3.2 Otimização taxa-distorção

O objetivo da otimização taxa-distorção é obter a melhor reprodução do sinal para uma dada taxa de compressão alvo (ORTEGA; RAMCHANDRAN, 1998). O critério de medida de distorção do trabalho é a diferença quadrática média entre a entrada e a saída do quantizador.

A distorção total pode ser expressa como uma função de:

$$d_s = f(r_a, r_f, r_\Phi) \quad (20)$$

Considere o quantizador uniforme definido pela Equação 17 e os comprimentos de bits de cada parâmetro na tripla $b_k = (r_a, r_f, r_\Phi) \epsilon \mathcal{B}$, em que $\mathcal{B}$ representa o conjunto de todas as possíveis combinações permitidas de taxas de bits dentro do intervalo definido por cada elemento $b_{k'}$ com $k = [1, 2, ..., K_\mathcal{B}]$ e $K_\mathcal{B}$ o número de elementos em $\mathcal{B}$. Afim de se obter o melhor compromisso taxa-distorção, deve-se buscar o que minimiza a distorção total inserida no processo de codificação, dada uma quantidade de bits disponíveis r_{alvo}. A solução é obtida por meio da resolução do seguinte problema de otimização (ORTEGA; RAMCHANDRAN, 1998):

$$\min_{b_k \in \mathcal{B}} d_S \quad \text{sujeito a } Mr \geqslant r_{alvo} \quad (21)$$

A solução clássica para esse problema é baseada na introdução de um número real e não negativo denominado de multiplicador de Lagrange $\lambda \geq 0$, que auxilia a minimização da função-custo Lagrangeana (ORTEGA; RAMCHANDRAN, 1998):

$$J = d_S + \lambda(Mr - r_{alvo}) \quad (22)$$

Em que M é o número de elementos do livro de estruturas, com r definido na Equação 19.

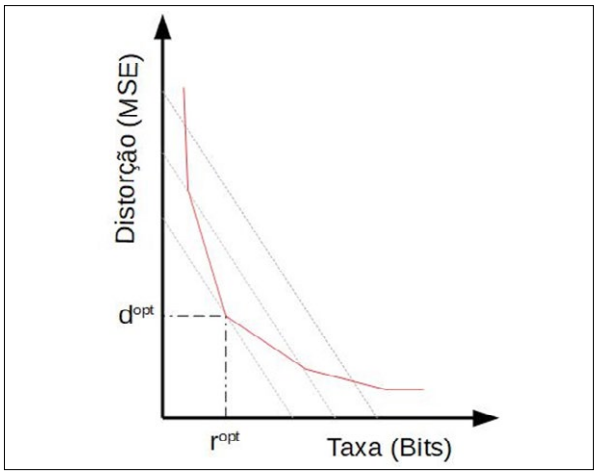
Para um dado , é possível encontrar o par $(d_S^{opt}; r_{alvo}^{opt})$ para o qual J é mínimo. O problema se

soluciona por meio da resolução do seguinte sistema de equações:

$$\frac{\partial J}{\partial r_a} = 0; \frac{\partial J}{\partial r_f} = 0; \frac{\partial J}{\partial r_\phi} = 0 \quad (23)$$

A fronteira ótima é então definida pelo fecho convexo do conjunto de pontos operacionais representados na Figura 14. Quando não existir forma fechada para d_s em função das taxas (r_a, r_f, r_ϕ), é possível adotar uma abordagem empírica para se obter as curvas operacionais. Para cada $b_k = (r_a, r_f, r_\phi)$, e para um dado sinal, o par taxa-distorção (r_k, d_k) é calculado, resultando no gráfico taxa-distorção (T-D).

Figura 14 – Interpretação gráfica da otimização da função Lagrangeana



Fonte: Elaboração própria

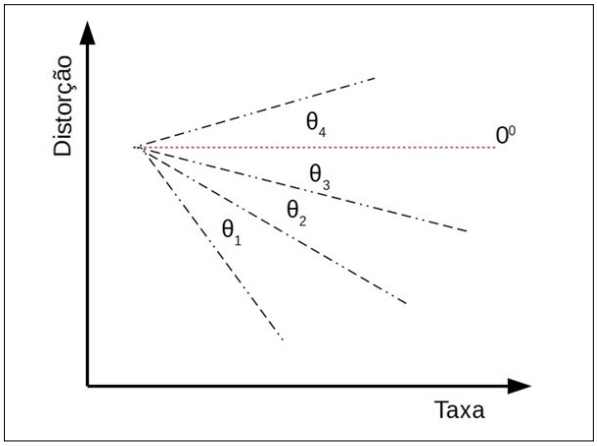
O procedimento para se obter o fecho convexo é definido da seguinte forma:

1) Busca-se; $b_{k0} = argmin_{b_k \in \mathcal{B}} r_k$

   $[r_{katual}; d_{katual}] = [r_{k0}; d_{k0}]$

2) Atribui-se $b_{katual} = b_{k0}$, portanto ;
3) Traça-se uma reta do ponto $[r_{katual}; d_{katua}$ a todos os outros pontos $[r_k; d_k]$, em que $r_k > r_{katual}$, como ilustrado na Figura 15. Cada reta possui um ângulo θ_k com a horizontal calculada por $\theta_k = arctan[(d_k - d_{katual})/(r_k - r_{katual})]$;
4) Obtêm-se $b_{kproximo}$, cujo par correspondente $[r_{kproximo}; d_{kproximo}]$ possui o menor ângulo $\theta_{k'}$ ou seja, $\theta_{min} = min_k \theta_k$;
5) Se $\theta_{min} \leq 0$, inclui-se b_{katual} na curva operacional e atualiza-se $b_{katual} = b_{kproximo}$;
6) Se $\theta_{min} > 0$, interrompe-se o procedimento;
7) Repetem-se os procedimentos de 3 a 6 até alcançar-se o par de r_k máximo;
8) Ao fim, os pontos pertencentes à curva operacional correspondem aos b_k ótimos do sinal.

Figura 15 – Traçando o fecho convexo



Fonte: Elaboração própria

A otimização taxa-distorção é realizada quadro a quadro, de forma independente, e considerando a taxa-alvo. Desse modo, a alocação de bits é localmente ótima a cada quadro, não sendo globalmente ótima em relação a todo o sinal. Para que a alocação seja globalmente ótima, o quadro de quantização deve possuir o mesmo comprimento do sinal. No entanto, como os sinais de áudio normalmente possuem longa duração, portanto inúmeras amostras, o processo de otimização taxa-distorção nessa situação torna-se impraticável em termos computacionais.

O que se faz é obter as curvas operacionais de taxa-distorção dos blocos e encontrar um multiplicador de Lagrange $\hat{\lambda}$, associado a um ângulo $\hat{\theta}$ que resulte em uma taxa $\hat{r}$ próxima da taxa desejada.

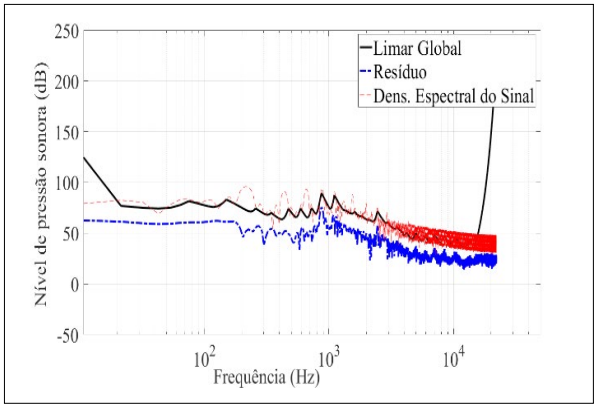
3 Configuração experimental

O sinal de áudio foi dividido em Q quadros de tamanho N, ponderados por uma função janela com saltos de l amostras. Neste trabalho, duas funções janela foram testadas: a retangular, com saltos de $l = N = 512$ amostras, e de Hanning, com saltos de $l = N/2 = 256$ amostras. A escolha da janela de Hanning se deve ao fato dela oferecer boa resolução em frequência e dispersão espectral reduzida.

É interessante destacar que a janela de Hanning é capaz de fornecer uma estimação da densidade espectral de potência do sinal mais precisa, permitindo assim melhor identificação dos componentes tonais e não tonais no processo de obtenção do limiar global e na decomposição por blocos do sinal. Esse fato pode ser verificado nas Figuras 16 e 17, nas quais estão representadas as densidades espectrais de potência do sinal e o seu limiar global para as janelas retangular e de Hanning, respectivamente.

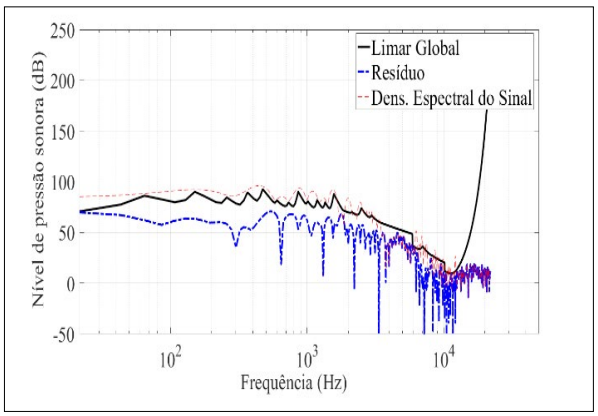
O critério de parada da decomposição utilizado se baseia no limiar global de mascaramento psicoacústico. Muitas vezes é necessário fazer uso de uma margem, a ser subtraída do limiar global, de modo a garantir que o resíduo da decomposição seja inaudível. Dessa forma, mais átomos são extraídos para compor a aproximação do sinal. A Figura 18 ilustra algumas margens que serão subtraídas no limiar global psicoacústico.

Figura 16 – Representação de um bloco de um sinal de áudio utilizando a janela retangular



Fonte: Elaboração própria

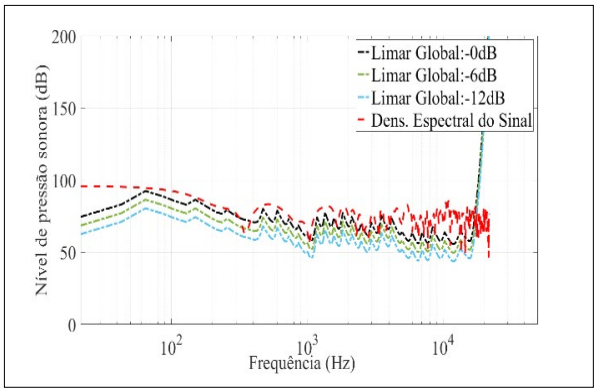
Figura 17 – Representação de um bloco de um sinal de áudio utilizando a janela de Hanning



Fonte: Elaboração própria

A quantização do livro de estruturas é realizada pelos quantizadores escalares uniformes *midrise* e *midtread* (SAYOOD, 2017). A alocação ótima de bits é realizada por meio da otimização taxa-distorção, ajustando-se o multiplicador de Lagrange.

Figura 18 – Exemplos de diferentes limiares psicoacústicos em dB



Fonte: Elaboração própria

A avaliação dos sinais reconstruídos é realizada por meio do algoritmo *Perceptual Evaluation of Audio Quality* (PEAQ), que mede objetivamente a qualidade de sinais de áudio, padronizado pelo *International Telecommunications Union* (ITU), na recomendação ITU-R BS.1387 (INTERNATIONAL TELECOMMUNICATION UNION, 2001). 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.

Os sinais de áudio utilizados nos experimentos se referem a notas de instrumentos musicais: nota A3 de um *piano*, nota A4 de *flauta*, nota A4 de um *violoncelo*, nota A4 de um *fagote* e dois trechos de bateria denominados *Bateria* A e *Bateria* B. Dicionários de exponenciais complexas foram utilizados com redundância de quatro e oito vezes o tamanho do bloco. Os sinais possuem 1 segundo de duração e taxa de amostragem de 44,1 kHz, com um total de 44.100 amostras. Para a decomposição, os sinais são divididos em blocos de 512 amostras com sobreposição de 256 amostras.

Os códigos e arquivos de áudio utilizados estão disponíveis em https://github.com/NogueiraJunior/Decomposicoes-Atomicas-em-Exponenciais-Complexas.

4 Resultados

Os resultados experimentais são apresentados em duas partes: na primeira, estão os resultados referentes à decomposição atômica e, na segunda, estão os referentes à alocação ótima de bits.

4.1 Decomposição atômica

Os resultados da avaliação PEAQ para decomposições realizadas nos sinais de áudio, em que o dicionário possui redundância de quatro e oito vezes o tamanho dos quadros (N) e as margens a serem subtraídas no limiar global psicoacústico variam de 0 a 10 dB, estão apresentados na Tabela 1 (NOGUEIRA JUNIOR; TCHEOU; ÁVILA, 2017).

Existe uma tendência geral de melhora utilizando um dicionário com maior redundância, pois o aumento da redundância do dicionário possibilita a representação do sinal de maneiras diferentes. Assim, é possível caracterizar melhor as várias formas e padrões presentes no sinal. Outro ponto observado é que, quanto maior for a margem subtraída do limiar global psicoacústico, melhor será o resultado perceptivo da decomposição. Esse fato ocorre porque o aumento da margem acarreta o aumento do número de iterações necessárias para se alcançar o critério de parada, aumentando também o número de elementos que descrevem o sinal.

Com o auxílio da Tabela 1 é possível observar que os sinais decompostos pelo algoritmo atingem a nota de PEAQ superior a 1 com a margem de 6 dB.

Na Tabela 2 (página seguinte) está ilustrado o número de iterações médio utilizado para cada uma das margens subtraídas do limiar psicoacústico de 0 dB até 10 dB para diferentes instrumentos usados no trabalho, com dicionário de redundâncias de quatro e oito vezes o tamanho do quadro, .

Os diferentes tipos de instrumentos apresentam comportamentos sonoros variados. Essa distinção altera o número de iterações necessárias para se obter uma boa avaliação perceptiva da decomposição atômica, conforme o limiar psicoacústico utilizado. Esse comportamento sugere que instrumentos cujos sons têm variações mais acentuadas na amplitude e na fase da densidade espectral de frequência – como os sinais de bateria A e bateria B durante o quadro em análise – contribuem para o surgimento de mais fenômenos de mascaramento, tornando a decomposição psicoacústica mais eficaz. O número de fenômenos de mascaramento afeta a quantidade de iterações necessárias para que o Dicionário de Exponenciais Complexas possa representar adequadamente as estruturas que compõem cada quadro do sinal em questão.

4.2 Alocação ótima de bits

Tabela 1 – Valores do PEAQ para diferentes sinais decompostos com dicionários que possuem redundâncias de quatro e oito vezes o número de amostras por bloco

Redundância / Margem(dB)	Piano A3		Violoncelo A4		Fagote A4		Flauta A4		Bateria A		Bateria B	
	4	8	4	8	4	8	4	8	4	8	4	8
0	-0,55	-0,61	-1,58	0,91	-1,53	-0,74	-2,74	-1,33	-0,39	-0,36	-0,27	0,28
1	0,01	-0,58	-1,38	-1,2	-1,43	-0,76	-2,62	-1,17	-0,31	0,27	-0,22	-0,23
2	-0,41	-0,55	-1,25	-1,03	-0,99	-0,66	-2,29	-0,88	-0,26	-0,23	-0,16	-0,15
3	-0,33	-0,4	-1,07	-0,84	-1,14	-0,58	-1,93	-0,71	-0,17	-0,16	-0,13	-0,13
4	-0,2	-0,28	-0,85	-0,68	-0,74	-0,47	-1,52	-0,52	-0,15	-0,14	-0,09	-0,1
5	-0,11	-0,19	-0,68	-0,51	-0,66	-0,55	-1,29	-0,4	-0,12	-0,09	-0,09	-0,06
6	-0,09	-0,09	-0,59	-0,39	-0,39	-0,37	-0,96	-0,25	-0,08	-0,061	-0,06	-0,06
7	-0,03	-0,05	-0,47	-0,27	-0,48	-0,24	-0,68	-0,16	-0,03	-0,037	-0,04	-0,04
8	0,01	-0,01	-0,34	-0,17	-0,28	-0,1	-0,56	-0,08	-0,03	-0,021	-0,01	-0,01
9	0,03	0,04	-0,19	-0,05	-0,21	-0,04	-0,42	-0,01	-0,02	-0,019	0,01	0,01
10	0,06	0,08	-0,06	0,02	-0,11	-0,04	-0,22	0,05	-0,01	0,011	0,01	0,01

Fonte: Elaboração própria

Tabela 2 – Valores do número médio de iterações para os diferentes sinais decompostos com dicionários que possuem redundâncias de quatro e oito vezes o número de amostras por bloco

Redundância / Margem (dB)	Piano A3		Violoncelo A4		Fagote A4		Flauta A4		Bateria A		Bateria B	
	4	8	4	8	4	8	4	8	4	8	4	8
0	52,08	51,95	51,98	52,23	34,37	32,29	66,99	61,1	105,51	104,1	144,42	141,13
1	56,03	55,67	55,7	55,71	39,1	34,38	72,78	66,07	113,48	112,11	155,23	152,03
2	59,81	59,36	59,59	59,35	44,78	36,78	80,5	71,25	122,02	119,85	165,87	162,27
3	64,05	63,39	63,73	63,07	50,63	39,97	90,89	77,02	131,23	128,74	176,54	173,05
4	68,53	67,69	68,16	67,26	55,95	43,96	101,94	82,98	140,88	138,55	188,81	184,65
5	73,15	72,07	73,22	71,83	61,54	48,06	117,15	89,82	151,69	149,22	202,69	197,39
6	77,95	76,53	78,53	76,61	69,3	52,19	120,97	96,83	164,21	161,2	217,27	211,38
7	83,04	81,14	84,29	81,58	78,99	56,7	126,19	104,26	177,93	174,5	233,86	227,63
8	87,95	86,03	90,86	87,25	68,74	62,76	138,87	112,8	192,59	189,03	251,38	245,44
9	93,78	91,39	97,91	93,4	73,91	68,7	151,54	121,69	211,53	206,17	271,43	264,58
10	100,93	97,66	105,71	100,39	79,63	75,6	167,67	132,02	232,09	225,45	292,86	286,09

Fonte: Elaboração própria

Nesta subseção, o desempenho da codificação dos sinais realizada com quantizadores escalares uniformes e alocação ótima de bits por meio de taxa-distorção (*T - D*) via multiplicador de Lagrange foi avaliado. O objetivo é conseguir a menor distorção dada uma da taxa de bits desejada.

O número de bits utilizados para obter a curva de otimização da taxa-distorção para cada áudio é definido da seguinte forma:

- A taxa de amplitude r_a varia de 2 bits até 16 bits;
- A taxa de fase r_Φ varia de 2 bits até 32 bits;
- Na frequência, o número de bits r_f é dado por $log_2(M/2)$. Assim, o número de bits utilizados pelos componentes de frequência é fixo e dependente da cardinalidade do dicionário M, com M = 2048, tem-se $r_f = 10$ bits.

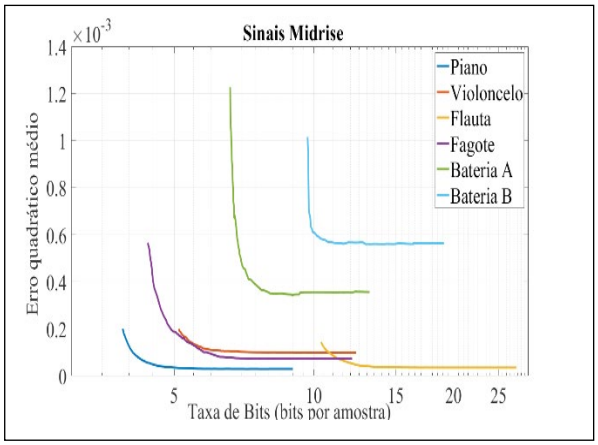
O conjunto $\mathcal{B}$ de todas as possíveis combinações permitidas para a alocação ótima de bits por meio de taxa-distorção (*T - D*) via multiplicador de Lagrange é igual a: $\mathcal{B}$ = 15 x 31 x 10 = 4.650 bits por quadro.

Os sinais submetidos ao processo de quantização foram decompostos com o DEC, utilizando redundância de $M = 4N$ (para $N = 512$, $M = 2048$) e diferentes margens psicoacústicas. Os valores das margens psicoacústicas escolhidas foram: *piano* A3 com margem de 3 dB, *violoncelo* A4 com margem de 7 dB, *fagote* A4 com margem de 9 dB, *flauta* A4 com margem de 11 dB, *bateria* A com margem de 0 dB e *bateria* B com margem de 0 dB.

A otimização de taxa-distorção por meio de curvas operacionais foi realizada para os quantizadores *midrise* e *midtread*. Para a avaliação da qualidade psicoacústica, uma curva de qualidade taxa-PEAQ por instrumento foi elaborada, isto é, para cada taxa de bits obtida pela curva de otimização da taxa-distorção, é gerada uma curva de avaliação de desempenho taxa-PEAQ.

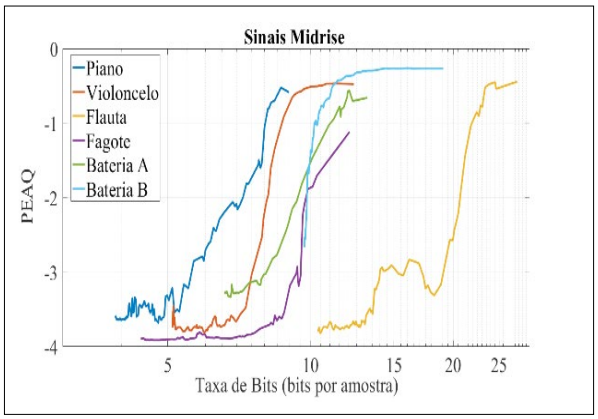
As Figuras 19 a 22 (página seguinte) são referentes as curvas de taxa-distorção e taxa-PEAQ de todos os sinais utilizados no trabalho. Por elas é possível ver que as características dos sons de cada instrumento influenciam no desempenho da alocação ótima de bits. O sinal *piano* A3 apresentou a menor distorção, com a melhor avaliação por PEAQ, utilizando, para isso, a menor taxa de bits para ambos os quantizadores. Os sinais *bateria* A, *bateria* B e *flauta* A4 apresentaram bons desempenhos em suas codificações. No sinal *flauta* A4, os quantizadores conseguem alcançar boas avaliações psicoacústicas, mas a custo de grandes taxas de bits. Com o sinal *fagote* A4, os quantizadores não conseguem obter bons resultados na avaliação por PEAQ.

Figura 19 – Curvas de otimização da taxa-distorção do quantizador *midrise*



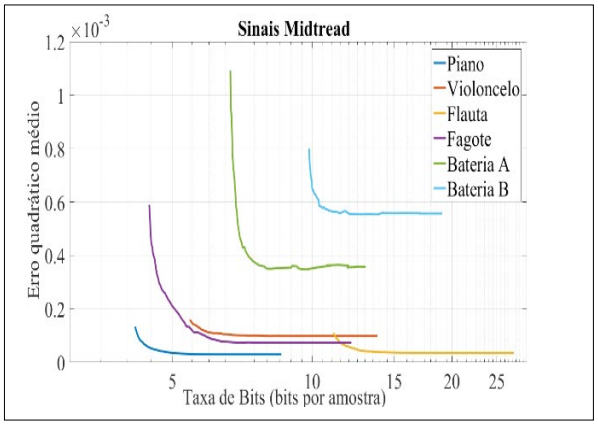
Fonte: Elaboração própria

Figura 20 – Curvas de otimização da taxa-PEAQ do quantizador midrise



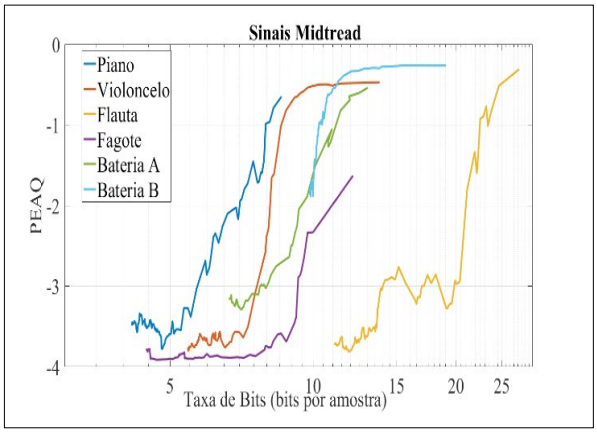
Fonte: Elaboração própria

Figura 21 – Curvas de otimização da taxa-distorção do quantizador *midtread*



Fonte: Elaboração própria

Figura 22 – Curvas de otimização da taxa-PEAQ do quantizador *midtread*



Fonte: Elaboração própria

A Tabela 3 apresenta os valores de taxas de bits por segundo (kbps) com suas faixas de qualidade superior para cada sinal.

Tabela 3 – Valores de taxa de bits por segundo com boa qualidade

Sinal	Midrise		Midtread	
	Taxa (kbps)	PEAQ	Taxa (kbps)	PEAQ
Piano A3	352,8	-1	352,8	-1
Violoncelo A4	374,85	-1	374,85	-1
Fagote A4	432,18	-2	485,1	-2
Flauta A4	926,1	-1	970,2	-1
Bateria A	485,10	-1	493,92	-1
Bateria B	454,23	-1	454,23	-1

Fonte: Elaboração própria

O padrão MPEG-1 Layer 3 (MP3) alcança qualidade de CD com taxas de bits de 192 kbps. Observa-se que o codificador proposto apresenta taxas maiores que 192 kbps para que o sinal seja avaliado pelo PEAQ como não perturbador. O MP3 é um padrão desenvolvido e aperfeiçoado por vários pesquisadores no decorrer do tempo. Uma das diferenças entre o padrão proposto e o MP3 é que o último utiliza o código de Huffman e codificação por entropia. Esse tipo de codificação permitiria melhorar o desempenho do codificador proposto.

Outra maneira de aperfeiçoar o método proposto é utilizar alguma forma de alocação ótima de bits que leve em consideração aspectos psicoacústicos do sinal em análise.

5 Conclusão/Considerações

Através dos resultados obtidos neste trabalho, nota-se que existe uma tendência geral de melhora dos resultados perceptivos quando se usa um dicionário com maior redundância, resultado do maior número de possibilidades de representação do sinal. Outro ponto observado é que, quanto maior a margem subtraída do limiar global psicoacústico, melhor é o resultado perceptivo da decomposição. O aumento da margem acarreta no acréscimo do número de iterações necessário para se alcançar o critério de parada, incrementando assim o número de elementos que descrevem o sinal.

Não foi observada uma margem psicoacústica global que possa responder adequadamente a todos os sinais analisados com o dicionário usado (DEC). Cada sinal necessita de uma margem correspondente às suas características.

Com o critério de parada psicoacústico, é possível reduzir o número de iterações necessárias para a decomposição de cada sinal, tornando o algoritmo mais rápido e aumentando o grau de compressão obtido.

A quantização escalar uniforme alcançou uma boa qualidade psicoacústica, mas o número de bits necessários para tal fim ainda é bastante elevado. Melhores desempenhos podem ser atingidos com o desenvolvimento de codificadores que levem em consideração as informações psicoacústicas de cada sinal na etapa de alocação de bits. Neste trabalho, as informações psicoacústicas foram consideradas na etapa de decomposição, mas não na alocação de bits.

Propõem-se, para trabalhos futuros, a implementação do princípio de relevância psicoacústica para múltiplos dicionários, com o uso de um critério de escolha do melhor dicionário, a aplicação da codificação por entropia, o que permitirá melhorar o desempenho do codificador proposto, e a utilização de alocação ótima de bits que leve em consideração aspectos psicoacústicos do sinal em análise.

REFERÊNCIAS

BOSI, M.; GOLDBERG, R. E. **Introduction to digital audio coding and standards**. New York: Springer, 2002.

CORMEN, T. H. *et al.* **Introduction to algorithms**. Cambridge, MA: MIT press, 2009.

DAVIS, G.; MALLAT, S.; ZHANG, Z. Adaptive time-frequency approximations with matching pursuits. **Wavelet Analysis and Its Applications,** v. 5, p. 271-293, 1994. DOI: 10.1016/B978-0-08-052084-1.50018-1. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780080520841500181. Acesso em: 3 dez. 2018.

FASTL, H.; ZWICKER, E. **Psychoacoustics**: facts and models. Berlin: Springer, 2007.

FERRANDO, S. E.; KOLASA, L. A.; KOVACEVIC, N. Algorithm 820: a flexible implementation of matching pursuit for Gabor functions on the interval. **ACM Transactions on Mathematical Software (TOMS),** v. 28, n. 3, p. 337-353, 2002. DOI: 10.1145/569147.569151. Disponível em: https://dl.acm.org/citation.cfm?id=569151. Acesso em: 3 dez. 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO/IEC 11172-3:1993:** Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s. Part1: Systems, Part2: Video, Part3: Audio. Geneva, Switzerland: [*s. n.*], 1993.

LIN, Y.; ABDULLA, W. H. **Audio watermark**: a comprehensive foundation using MATLAB. Cham: Springer, 2015.

MALLAT, S. **A wavelet tour of signal processing**: the sparse way. 3. ed. Burlington: Academic Press, 2008.

MALLAT, S. G.; ZHANG, Z. Matching pursuits with time-frequency dictionaries. **IEEE Transactions on Signal Processing**, v. 41, n. 12, p. 3397-3415, 1993. Disponível em: https://pdfs.semanticscholar.org/0b6e/98a6a8cf8283fd76fe1100b23f11f4cfa711.pdf. Acesso em: 3 dez. 2018.

NOGUEIRA JUNIOR, V. S.; TCHEOU, M. P.; ÁVILA, F. R. Decomposição psicoacústica de sinais de áudio com base em dicionários redundantes e exponenciais complexas. *In:* SIMPÓSIO DE PROCESSAMENTO DE SINAIS, 7., 2017, São Bernardo do Campo. **Anais** [...]**.** São Bernardo do Campo: UFABC, 2017. Disponível em: http://eventos.ufabc.edu.br/siimsps/files/id151.pdf. Acesso em: 3 dez. 2018.

ORTEGA, A.; RAMCHANDRAN, K. Rate-distortion methods for image and video compression. **IEEE Signal Processing Magazine**, v. 15, n. 6, p. 23-50, 1998. DOI: 10.1109/79.733495. Disponível em: https://ieeexplore.ieee.org/abstract/document/733495. Acesso em: 3 dez. 2018.

PETROVSKY, A.; HERASIMOVICH, V.; PETROVSKY, A. Scalable parametric audio coder using sparse approximation with frame-to-frame

perceptually optimized wavelet packet based dictionary. *In*: AES CONVENTION, 138., 2015, Varsóvia. **Proceedings** [...]. Varsóvia: AES, 2015. Disponível em: http://www.aes.org/e-lib/online/browse.cfm?elib=17688. Acesso em: 3 dez. 2018.

PETROVSKY, A.; HERASIMOVICH, V.; PETROVSKY, A. Audio/speech coding using the matching pursuit with frame-based psychoacoustic optimized time-frequency dictionaries and its performance evaluation. *In:* IEEE INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING: ALGORITHMS, ARCHITECTURES, ARRANGEMENTS, AND APPLICATIONS, 20., 2016, Poznan, Poland. **Proceedings** [...]. Poznan: IEEE Xplore, 2016. DOI: 10.1109/SPA.2016.7763617. Disponível em: https://ieeexplore.ieee.org/abstract/document/7763617. Acesso em: 3 dez. 2018.

INTERNATIONAL TELECOMMUNICATION UNION. **BS. 1387**: method for objective measurements of perceived audio quality. Geneva, Switzerland: International Telecommunication Union, 2001.

SAYOOD, K. **Introduction to data compression**. 5. ed. Cambridge, MA: Morgan Kaufmann, 2017.

SPANIAS, A.; PAINTER, T.; ATTI, V. **Audio signal processing and coding**. New Jersey: Wiley, 2006.

THIEDE, T. *et al*. PEAQ-The ITU standard for objective measurement of perceived audio quality. **Journal of the Audio Engineering Society**, v. 48, n. 1/2, p. 3-29, 2000. Disponível em: http://www.aes.org/e-lib/browse.cfm?elib=12078. Acesso em: 3 dez. 2018.

TOUMI, I.; DERRIEN, O. Sparse decomposition of audio signals using a perceptual measure of distortion. Application to lossy audio coding. *In:* INTERNATIONAL CONFERENCE ON DIGITAL AUDIO EFFECTS, 18., 2015, Trondheim, Norway. **Proceedings** [...]. Trondheim, Norway: Norwegian University of Science and Technology, 2015. Disponível em: https://hal.archives-ouvertes.fr/hal-01240863/. Acesso em: 3 dez. 2018.

VERMA, T. S.; MENG, T. H. Y. Sinusoidal modeling using frame-based perceptually weighted matching pursuits. *In:* IEEE International Conference on Acoustics, Speech, and Signal Processing, 24., Phoenix, USA, 1999. **Proceedings** [...]. Phoenix, USA: IEEE, 1999. p. 981-984. DOI: 10.1109/ICASSP.1999.759861. Disponível em: https://ieeexplore.ieee.org/abstract/document/759861. Acesso em: 3 dez. 2018.