Uma Visão Geral do Sistema de Compressão de Áudio Para HDTV (MPEG) – Sistema Europeu

Edgard Luciano Oliveira da Silva, Yuzo Iano, Guilhermo Leopoldo Kemper Vásques, José Geraldo Chiquito

DECOM – FEEC - UNICAMP

***Resumo*: Este artigo cobre a teoria utilizada na compressão MPEG/Áudio. MPEG (Moving Picture Experts Groups) é um grupo técnico que produz normas para compressão de vídeo e áudio digitais, reunidos na forma de sub-comite ISO (International Organization for Standardization). A norma de compressão MPEG/áudio é uma parte da norma que inclui ainda a norma para compressão de vídeo e a norma para sincronização de áudio e vídeo e cadeias de dados relacionados. A norma MPEG/Áudio também pode ser usada em aplicações somente de áudio para comprimir áudio de alta-fidelidade a baixa taxa de bits.**

No processo de compressão MPEG/áudio ocorre perda de informação, embora essa perda normalmente seja transparente, perceptivamente sem perdas, mesmo com fatores de compressão de 6/1 ou maiores. O algoritmo trabalha explorando as propriedades perceptivas do sistema auditivo humano, eliminando as ondas sonoras que não sensibilizem o ouvido humano. Este artigo irá cobrir as bases do modelo psico-acústico e os métodos utilizados pelo algoritmo MPEG/Áudio para comprimir sinais de áudio com a menor degradação perceptiva.

I. INTRODUÇÃO

MPEG é o acrônimo para *Moving Picture Experts Group* (Grupo de Especialistas de Imagens em Movimento). Trata-se de um grupo técnico de trabalho, que produz normas para compressão de vídeo e áudio digitais, reunidos na forma de sub-comite ISO/IEC (*International Standards Organization / International Electrotechnical Commission*). MPEG é o primeiro padrão internacional no domínio de compressão de áudio de alta-fidelidade. Em particular, MPEG define a sintaxe da seqüência de bits (*bit-stream*) de vídeo e áudio codificados a baixas taxas de bits. O algoritmo de codificação não é definido pela norma MPEG. Isso permite um aprimoramento contínuo dos codificadores e sua adaptação a aplicações específicas, seguindo a definição da sintaxe do *bit-stream*. Além da codificação de áudio e vídeo, MPEG também define os meios para multiplexar seqüências de vídeo e áudio sincronamente em uma única seqüência de bits, descrevendo também métodos para testes de conformidade. Assim, a norma MPEG divide-se em três partes (áudio, vídeo e sistemas). Embora a compressão de áudio MPEG seja uma parte das três partes que compõem a norma, MPEG/áudio é perfeitamente compatível com aplicações somente de áudio. A norma MPEG/áudio é o resultado de mais de 3 anos de trabalho conjunto de um comitê internacional de especialistas em compressão de áudio de alta-fidelidade. O comitê MPEG trabalha em fases distintas, normalmente denotadas por números seqüênciais (MPEG-1, MPEG-2, MPEG-4). As atividades da primeira fase (MPEG-1) resultaram na norma ISO/IEC 11172-3 de 1993. Essa norma foi desenvolvida para codificar sinais de áudio em formatos mono (1ª fase) e estéreo (2ª fase).

A segunda fase (MPEG-2) resultou na norma internacional ISO/IEC 13818-3, estabelecendo padrões apropriados para HDTV. As diferenças com relação a MPEG-1/Áudio são 5 canais de áudio (*full bandwidth*) + 1 canal de baixa freqüência (canal LFE - *Low Frequency Enhancement*), amostrado a 1/96 da freqüência de amostragem adotada e com faixa: 15 Hz a 120 Hz (por esta razão é conhecido como "MC 5+1") contra 2 canais do MPEG-1, novas freqüências de amostragem, novas taxas de bits, até 24 bit/amostra/canal, novas tabelas de quantização, melhoria em codificação de fatores de escalonamento, canais *surround* e suporte a múltiplos idiomas.

A utilização de 5 canais de áudio permite uma representação estereofônica mais realista. Os primeiros experimentos com som estéreo datam da década de 30 nos laboratórios Bell e utilizavam três canais. O público pôde ouvir som estéreo pela primeira vez no início da década de 50 nos cinemas que utilizavam nada menos que quatro canais e algumas vezes até sete canais Quando finalmente o som estéreo pôde entrar nos lares poucos anos depois, utilizava apenas dois canais, pois era o máximo de canais que um gravador de discos de vinil podia registrar. Essa limitação tecnológica, e não a preferência do ouvinte, fez com que dois canais estéreo se tornassem o padrão para a reprodução do som doméstico.

Produtores de cinema, entretanto, continuaram utilizando quatro canais (esquerdo, direito, central e *surround*) como o mínimo necessário para criar uma reprodução convincente. ITU-R e outros grupos internacionais têm recomendado uma configuração de cinco alto-falantes, pois tal configuração oferece um campo de sonorização *surround* com uma imagem de som e um aumento da área de audição estáveis. Estes canais são denominados *L*(*left*), *R*(*right*), *C*(*center*), e dois canais laterais/*surround Ls*(*left surround*) e *Rs*(*right surround*). Essa configuração também

denominada "3/2 stereo", já que faz uso de três alto-falantes na frente do ouvinte e dois atrás.

De modo a compatibilizar os padrões MPEG-1 e MPEG-2, os sinais desses cinco canais são combinados formando dois canais denominados *L*0 e *R*0, numa operação denominada *matrixing*. Dessa forma, um decodificador MPEG-1 interpretará os sinais *L*0 e *R*0 como os sinais dos canais esquerdo e direito e os decodificará corretamente (sinal *stereo*); já o MPEG-2 operará de modo inverso, procedendo o *dematrixing*, recuperando os cinco canais originais. A Figura mostra este processo. As equações de transformação para os canais *L*0 e *R*0 em função dos canais *R*, *L*, *C*, *Ls* e *Rs* são assim representadas pelas Equações (1), (2), (3) e (4).

$$\alpha = \frac{1}{1+\sqrt{2}} \quad (1)$$

$$\beta = \delta = \sqrt{2} \quad (2)$$

$$L0 = \alpha(L + \beta C + \delta Ls) \quad (3)$$

$$R0 = \alpha(R + \beta C + \delta Rs) \quad (4)$$

Outras escolhas são possíveis , incluindo *L*0 = *L* e *R*0 = *R*. Os fatores α, β e δ atenuam o sinal para evitar *overload*, quando se está calculando o sinal estéreo compatível (*L*0 e *R*0). *L*0 e *R*0 são transmitidos no formato MPEG-1 nos canais *T*1 e *T*2. Os canais *T*3, *T*4 e *T*5 formam juntos o sinal da extensão multi-canal (Figura 1). Eles devem ser escolhidos de modo que um decodificador possa recalcular o sinal multi-canal estéreo 3/2 completo. Redundâncias intercanal e efeitos de mascaramento são considerados para a melhor escolha. Um exemplo simples é *T*3=*C*, *T*4=*Rs* e *T*5=*Ls*. A operação de *matrixing* no MPEG-2 pode ser feita de uma maneira bem flexível.

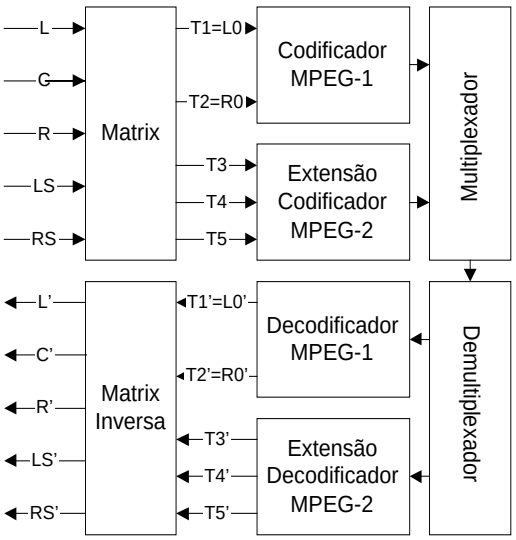

Figura 1. *Compatibilidade do bit-stream de áudio multicanal MPEG-2*

Essas facilidades podem ser utilizadas na geração de *bit-stream* a serem decodificados por decoders MPEG-1. Assim, no codificador MPEG-2/Áudio, o par estéreo (Esquerdo e Direito) é transmitido como no MPEG-1 e os canais adicionais são enviados em campos de dados auxiliares da sintaxe MPEG-1. Um decodificador MPEG-1 só poderá decodificar uma parte do sinal codificado por um codificador MPEG-2 (a parte do sinal compatível). Essa compatibilidade recebe o nome de MPEG-2 BC (*Backward Compatible*). Com algumas informações adicionais é possível a reprodução de até 7+1 canais (MC 7).

		Decodificador		
		Estéreo	5+1	7+1
Fonte	Estéreo	Estéreo	Estéreo	Estéreo
	5+1	Estéreo	5+1	5+1
	7+1	Estéreo	5+1	7+1

Tabela 1. *Compatibilidade multi-canais (MC).*

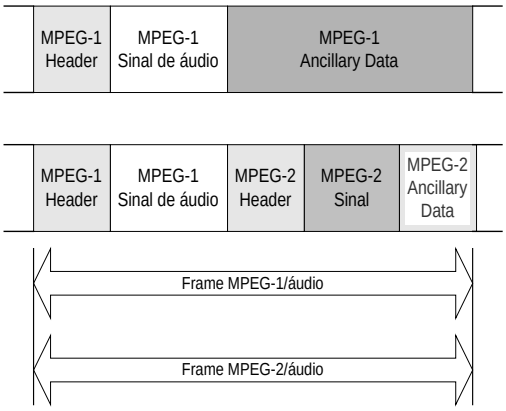

Figura 2. *Formato dos dados dos bit-streams do MPEG/áudio, frame MPEG-1 e frame MPEG-2 compatível com MPEG-1*

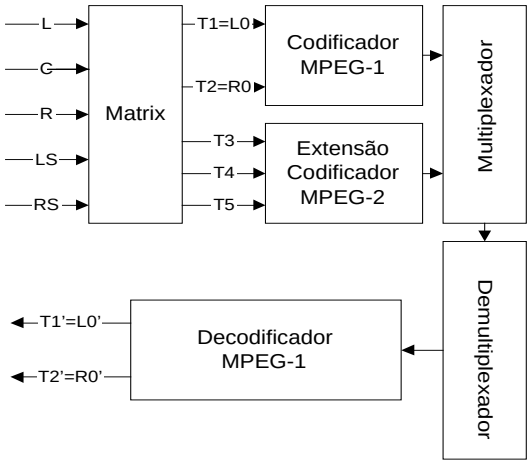

Figura 3. *Decodificador estéreo MPEG-1 de um bit-stream multi-canal MPEG-2*

Uma evolução da norma ISO/IEC 13818-3 foi o padrão MPEG-2 AAC (*Advanced Audio Coding*) (ISO 13818-7), também conhecido como codificação MPEG-2 NBC (*Non – Backward - Compatible*). Trata-se de um sistema sem compatibilidade regressiva com MPEG-1/áudio, ou seja, nem todo *bit-stream* MPEG-2 pode ser decodificado por decoder MPEG-1, mas apresenta melhor desempenho do que MPEG-2 BC.

II. MPEG/ÁUDIO: CARACTERÍSTICAS E APLICAÇÕES

MPEG/áudio foi criado para compressão de áudio genérico, isto é, todos os tipos de sinais da fala, sinais de música e ruídos sonoros. O codificador MPEG/áudio realiza a compressão sem conhecer a natureza do sinal de áudio. Ele, o codificador, explora a limitação perceptiva do sistema auditivo humano. A maior parte da compressão resulta da remoção da parte perceptivamente irrelevante do sinal de áudio. Nesse tipo de codificação, o sinal decodificado (saída) não é exatamente igual ao sinal codificado. O objetivo, no entanto, é assegurar que o sinal decodificado tenha o mesmo som do sinal codificado. A remoção dessas partes do sinal resultam em distorções inaudíveis, assim MPEG/áudio pode comprimir um sinal de áudio genérico.

A primeira fase ou MPEG-1/áudio, mantendo sua natureza genérica, oferece diversos modos de compressão com diferentes freqüências de amostragem que são: 32 kHz, 44,1 kHz e 48 kHz.

A seqüência de bits comprimida pode ter uma das várias taxas de bits, medida em *kilobitspersecond* (kbps), de 32 a 224 kbps/canal. Para se ter uma idéia, o Compact Disk (CD) é hoje um padrão para a representação de áudio digital. Com uma taxa de amostragem de 44,1 kHz, a taxa resultante para 16 bit/amostra/canal x 44.100 amostra/s = 705,6 kbps/canal. Como temos 2 canais, a taxa resultante será 1,41 Mbps, sem *overhead*. Dependendo da taxa de amostragem do sinal de áudio, a faixa do fator de compressão pode variar de 2,7 a 24. Os testes mostraram que mesmo com uma taxa de compressão de 6:1 (2 canais, 16 bits por amostra, freqüência de amostragem de 48 kHz comprimidos para 256 kbps) e sob ótimas condições auditivas, ouvintes experientes não foram capazes de distinguir o sinal de áudio original do sinal codificado com significância estatística [5]. Além disso, os sinais de áudio foram escolhidos de modo a serem difíceis de serem comprimidos. As principais motivações para uma codificação em baixa taxa de bits são a necessidade de minimizar a custo das transmissões e/ou para proporcionar o custo eficiente de armazenamento.

Ambas as fases (MPEG-1 e MPEG-2)/áudio possibilitam a escolha de três *layers* (camadas) independentes de compressão. Para cada *layer* a norma especifica o formato do *bit-stream* (seqüência de bits). Diferentemente das camadas do modelo OSI, os *layers* MPEG não são encadeados e sim autônomos e principalmente compatíveis. Os *layers* são compatíveis hierarquicamente, fazendo com que o decodificador do *layer N* possa decodificar seqüências codificadas no *layer N* ou inferiores. Assim , o decodificador do *layer* III é capaz de decodificar seqüências codificadas nos *layers* III, II e I, enquanto o decodificador do *layer* II é capaz de decodificar seqüências codificadas no *layer* II e I.

Esses diferentes *layers* foram definidos pois cada um possui suas vantagens. Basicamente, a complexidade do codificador e decodificador, o atraso do codificador/decodificador e a eficiência de codificação aumenta quando vamos do *layer* I, via *layer* II para o *layer* III.

A necessidade de se ter mais de um *layer* advém da possibilidade dada aos fabricantes de equipamentos de escolherem parâmetros como: a qualidade do áudio transmitido, o seu tempo de processamento, a taxa de transmissão, etc, que se adequem ao custo da implementação.

O *layer* I apresenta menor complexidade e apresenta qualidade de CD em 384 kbps. Como exemplo o *Phillips Digital Compact Cassette* (DCC) utiliza a compressão MPEG-1 *layer* I e uma taxa de bits de192 kbps/canal;

O *layer* II apresenta uma complexidade intermediária e qualidade CD em 256 kbps (128 kbps/canal). Uma possível aplicação para essa camada inclui *Digital Audio Broadcasting* (DAB) ou o armazenamento de seqüências sincronizadas de vídeo e áudio em CD-ROM. Comparado com o *layer* I, o *layer* II é capaz de remover mais sinal redundante e capaz de aplicar os modelos psico-acústico mais eficientemente;

O *layer* III apresenta uma complexidade elevada e qualidade CD em 128 kbps (64 kbps/canal). Este *layer* pode ser utilizado para transmissão sobre ISDN. O MPEG-1 *layer* III ficou popularmente conhecido como MP3.

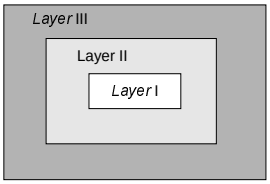

Figura 4. *Hierarquia entre os layers I, II e III do padrão MPEG*

A norma MPEG é rígida somente onde for necessário para garantir uma interoperabilidade. Ela determina a sintaxe do *bit-stream* codificado, define o processo de codificação e proporciona testes de distorção para assegurar a precisão do decodificador. Isso garante que, dependendo da origem, um decodificador MPEG/áudio estará apto a decodificar um *bit-stream* MPEG/áudio com um resultado previsível. Uma grande aceitação da norma irá permitira fabricantes produzirem e venderem, a um custo razoável, um grande número de codificadores MPEG.

Onde é possível, a norma está aberta para aperfeiçoamentos. Projetistas estão livres para experimentarem novas e diferentes implementações de codificadores e decodificadores nos limites da norma. Há, assim, um bom potencial para diversificar os codificadores.

III. VISÃO GERAL

A chave para a compressão MPEG/áudio é a quantização. Embora seja uma quantização feita em partes, como veremos, o algoritmo pode dar uma compressão transparente, ou seja, aparentemente sem perdas. O comitê MPEG/áudio realizou extensivos

testes subjetivos auditivos durante o desenvolvimento da norma.

A Figura 5 mostra o diagrama de blocos do codificador e a Figura 6 o decodificador MPEG/áudio. A seqüência de áudio de entrada passa através de um banco de filtros que divide o sinal de entrada em sub-bandas de freqüência. Simultaneamente a seqüência de áudio de entrada passa através de um modelo psico-acústico que determina a razão da energia do sinal com relação ao limiar de enmascaramento de cada sub-banda. O bloco de alocação de bit ou ruído usa a relação sinal-enmascaramento (SMR) para decidir como dividir o número total de bits disponível na sub-banda de sinal para minimizar a audição do ruído de quantização. Finalmente o último bloco pega a representação das amostras quantizadas em sub-banda e outras informações e forma o *bit-stream.* Dados auxiliares não necessariamente relacionados com a seqüência de áudio podem ser inseridos no *bit-stream* codificado. O decodificador decifra esse *bit-stream*, restaura os valores quantizados em sub-banda e reconstrói o sinal de áudio a partir dos valores da sub-banda.

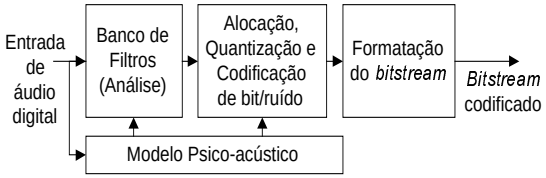

Figura 5. *Codificador MPEG/Áudio* (*Layers* I *e* II)

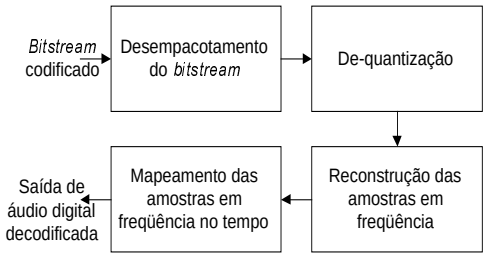

Figura 6. *Decodificador MPEG/áudio* (*Layers* I *e* II)

Neste trabalho vamos apresentar uma visão geral do formato de codificação MPEG/Áudio: *layers* I e II (utilizados em HDTV), a fim de esclarecer os procedimentos de codificação especificados nas respectivas normas ISO/IEC para uma futura implementação desses sistemas de compressão.

IV. BANCO DE FILTROS

Esta seção irá dar algumas idéias do comportamento do banco de filtros polifásicos apresentando um exame detalhado do banco de filtros de análise no codificador. Uma análise similar aplica-se ao banco de filtros de síntese no decodificador.

O filtro de análise polifásico é o componente chave comum em todos os *layers* da compressão MPEG/áudio. Esse filtro divide o sinal de áudio em 32 sub-bandas com larguras de banda iguais através de um banco de filtros passa-faixa, sendo que a saída de cada filtro é decimada por um fator de 32. Os filtros são relativamente simples e proporcionam boa resolução temporal com uma resolução razoável em freqüência. Entretanto as larguras de banda iguais para as sub-bandas não representam precisamente o sistema auditivo humano.

Se são assumidos filtros ideais com uma resposta em freqüência retangular, podemos garantir através do teorema de *Nyquist* que o sinal original pode ser reconstruído exatamente interpolando-se os sinais correspondentes a cada sub-banda segundo suas freqüências de amostragem originais e somando-se os resultados.

No entanto, desde que não é possível se construir filtros com resposta em freqüência perfeitamente plana na banda de passagem e zero na banda de rejeição, o efeito de *aliasing* pode ser introduzido durante o processo de decimação, o qual resultaria em perda de informação.

O projeto é um compromisso com três concessões. As 32 larguras de banda não representam com precisão as bandas críticas do ouvido humano. As larguras de banda do banco de filtros são muito largas para as baixas freqüências como pode ser visto na Tabela 2. Segundo, o banco de filtro de análise no codificador não é perfeitamente o inverso do banco de filtros de síntese no decodificador. Mesmo sem quantização a transformação inversa pode não recuperar perfeitamente o sinal de entrada original. Felizmente, o erro introduzido pelo banco de filtros é pequeno e inaudível. Finalmente, há uma sobreposição de freqüências entre bandas adjacentes. Um sinal em uma determinada freqüência pode afetar a saída de duas bandas adjacentes.

Existem duas teorias que têm sido estudadas e desenvolvidas para reduzir estes efeitos: a ***Frequency Domain Alias Cancellation (FDAC)*** e a ***Time Domain Alias Cancellation (TDAC)*** utilizada pelo sistema MPEG através de uma variante da mesma.

O codificador MPEG utiliza filtros FIR de 512 *taps* tanto para a parte de análise como para a parte de síntese. Os filtros foram projetados a fim de se obter cancelamento de *aliasing* e uma considerável atenuação fora da sua faixa.

A fim de se obter uma eficiente implementação do processo de filtragem, o MPEG emprega a estrutura polifase, no entanto, foi recentemente mostrado que uma implementação mais eficiente é conseguida através da ***Fast Discrete Cosine Transform (FDCT)***.

Nas Figuras 7 e 8 são ilustradas, respectivamente, a implementação polifase do banco de filtros de análise e síntese. Durante a fase de análise, a seqüência de amostras do sinal de áudio é deslocada em 32 amostras e armazenada em um *buffer* de 512 amostras. Assim, em cada processo de análise, as 512 amostras de processo anterior armazenadas no *buffer* são deslocadas em 32 amostras para permitir a entrada de 32 novas amostras.

Logo, o conteúdo do *buffer* é multiplicado por uma janela **C** (*analysis window*) de 512 amostras que se encontra tabulada na norma respectiva ISO/IEC. Os

resultados da multiplicação são armazenados num buffer **Z**.

Banda Número	Banda (Hz)		Banda Número	Banda (Hz)	
	Crítica	MPEG		Crítica	MPEG
0	50	750	14	1970	11250
1	95	1500	15	2340	12000
2	140	2250	16	2720	12750
3	235	3000	17	3280	13500
4	330	3750	18	3840	14250
5	420	4500	19	4690	15000
6	560	5250	20	5440	15750
7	660	6000	21	6375	16500
8	800	6750	22	7690	17250
9	940	7500	23	9375	18000
10	1125	8250	24	11625	18750
11	1265	9000	25	15375	19500
12	1500	9750	26	20250	20250
13	1735	10500			

Tabela 2. *Comparação entre os limites superiores de freqüência para um codificador MPEG (48kHz) e os limites aproximados para as bandas críticas do ouvido humano.*

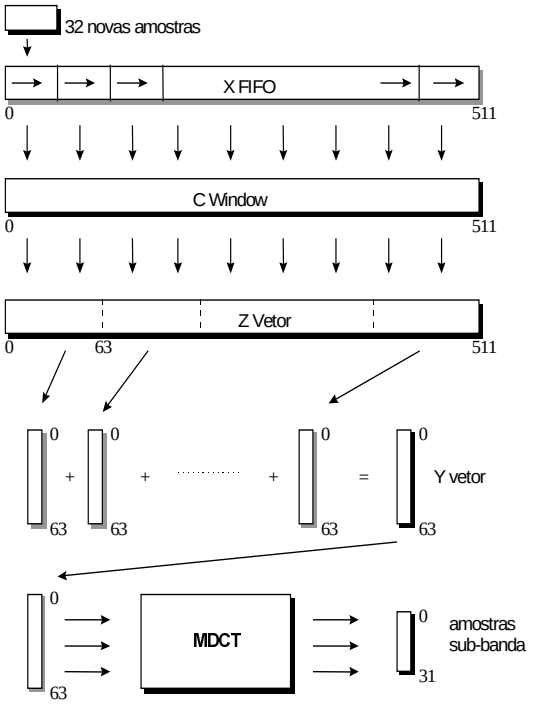

Figura 7.- *Implementação MPEG do Banco de Filtro de Análise Polifase.*

O conteúdo do buffer **Z** é dividido em 8 vetores de 64 elementos cada um, os quais são somados para formar o vetor **Y**. O vetor **Y** é transformado usando-se uma variante da ***Modified Discrete Cosine Transform*** (**MDCT**) para gerar finalmente as amostras correspondentes às 32 sub-bandas.

De outra maneira,

$$S_t[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] * (C[k+64j] * x[k+64j]) \quad (5)$$

Onde:

i é o índice da sub-banda e sua faixa é de 0 a 31

$S_t[i]$ é a amostra de saída para a sub-banda *i* em um tempo *t,* onde *t* é um múltiplo inteiro de um intervalo de 32 amostras.

$C[n]$ é um dos 512 coeficientes da janela de análise definida na norma, $x[n]$ é uma amostra de entrada de áudio lida do buffer de 512 amostras, e

$$M[i][k] = \cos\left[\frac{(2*i+1)*(k-16)*\pi}{64}\right] \quad (6)$$

são os coeficientes da matriz de análise.

As equações são parcialmente otimizadas para reduzir o número de cálculos computacionais. Devido à função entre parênteses ser independente do valor de *i*, e $M[i][k]$ ser independente de *j*, a saída dos 32 filtros necessitam apenas de 512+32*64 = 2560 multiplicações e 64*7+32*63=2464 adições, ou grosseiramente pouco mais de 80 multiplicações por saída.

Note que nessa implementação do banco de filtros o sinal de entrada é perfeitamente amostrado. Para cada 32 amostras de entrada, o banco de filtros produz 32 amostras de saída. De fato, cada sub-amostra das 32 sub-bandas do filtro produz 32 saídas que por sua vez produzirão 32 novas amostras de áudio.

Na fase de síntese, as amostras das 32 sub-bandas são transformadas no seu vetor original de 64 amostras (denominado na fase de síntese como vetor **V**) usando uma variante da ***Inverse Modified Discrete Cosine Transform*** (**IMDCT**). O vetor **V** é colocado dentro de um buffer FIFO, o qual armazena os últimos 16 vetores V. Logo, um vetor U é formado extraindo-se de forma alternada blocos de 32 amostras do *buffer* FIFO. Em seguida, o vetor U é multiplicado por uma janela **D** (*window synthesis* , tabulado na norma ISO/IEC) para gerar o vetor **W**. As amostras reconstruídas são obtidas de **W** decompondo o mesmo em 16 vetores de 32 elementos cada um. Finalmente os 16 vetores são somados para gerar as amostras de áudio reconstruídas.

As respostas impulsivas dos filtros não apresentam fase linear, e elas aparentemente introduziriam distorção no sinal reconstruído. No entanto, a implementação em cascata dos filtros análise-síntese gera um filtro de fase linear de 1023 amostras permitindo recuperar o sinal livre de distorção.

O codificador MPEG transmite as amostras de áudio codificadas através de *frames* de sincronização. Cada *frame* leva um número fixo de amostras (384 ou 1152) de áudio dependendo do *layer* de compressão. Dentro de cada *frame,* as amostras correspondentes a cada sub-banda são escalonadas e quantizadas de acordo com um modelo psico-acústico. Assim mesmo, desde que o fator de escala e o número de níveis de quantização variem com o tempo e com o número de sub-bandas, ambos devem ser transmitidos junto com as amostras quantizadas de cada uma das bandas transmitidas em cada *frame*.

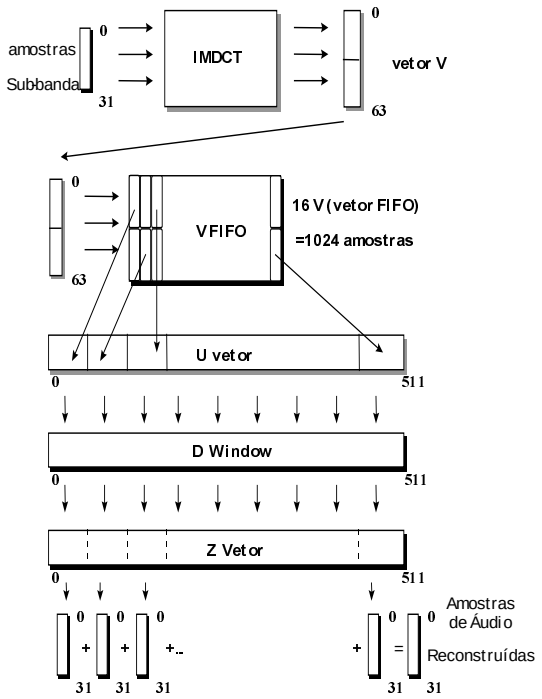

Figura 8. *Implementação Polifase da Síntese Filter Bank MPEG*

V. FORMATO DO FRAME MPEG/ÁUDIO

O codificador MPEG utiliza um algoritmo de alocação de bits adaptativo onde os fatores de escala e o número de bits alocados para cada amostra variam de *frame* a *frame* , de acordo com o modelo psico-acústico utilizado. No *layer* I, o quadro é codificado com 384 amostras de sub-banda ou 12 amostras de sinal de áudio por sub-banda. Nos *layers* II e III, o quadro contém 1152 amostras ou 36 amostras por sub-banda.

Cada quadro começa com um bloco de *header* (cabeçalho) de 32 bytes, o qual é formado por campos que levam informação necessária para estabelecer a sincronização com o receptor, assim como para informar ao mesmo os parâmetros básicos de codificação, em qualquer instante de tempo durante a transmissão.

Alguns desses parâmetros são a freqüência de amostragem, a taxa de bits, o número do *layer*, o formato dos canais de áudio (mono, estéreo, multicanal, etc.), e a existência de códigos corretores de erros no *frame* (CRC - *Cyclic Redundancy Check*) para proteção da informação que está sendo transmitida dentro da mesma.

A freqüência de amostragem é de 32 kHz, 44.1 kHz, e 48 kHz para MPEG-1 e de 16 kHz, 22.05 kHz e 24 kHz para MPEG-2. A taxa de bits é também restrita a certos números. Cada *layer* e freqüência de amostragem têm disponíveis várias taxas de bits. A escolha da taxa de bits, medida em kbps (*kilobits per second*) depende da qualidade de áudio e do modo (mono ou éstereo) desejados.

Se é especificada a presença de códigos de correção de erros no cabeçalho do frame, então após o mesmo é codificado um campo de 16 bits levando um código CRC. Em seguida são transmitidos os blocos de áudio levando a informação de som codificada dentro do frame.

Freqüência de Amostragem	LAYER	
	I	II
32 kHz e 44.1 kHz e 48 kHz	32, 64, 96 , 128, 160, 192, 224, 256, 288, 320, 352, 384, 416 e 448	32, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320, e 384
16kHz e 22.05kHz e 24 kHz	32, 48, 56, 64, 80, 96, 112, 128, 144, 160, 176, 192, 224, 256	8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 11, 128, 144 e 160

Tabela 3. *Taxa de bits (kbps) disponível para taxas de amostragem padrão e baixas taxas de amostragem*

Após, encerrando cada frame, são codificados campos que levam dados auxiliares definidos pelo usuário, os quais podem conter informação relativa a algum tipo de aplicação.

O campo de sincronização localizado no início do cabeçalho do frame é formado por um código de 12 bits onde todos são iguais a um. O codificador MPEG evita transmissão de qualquer outro código dentro do frame similar ao código de sincronização.

A informação a ser codificada pode ser mono ou estéreo. O modo mono é a escolha óbvia para se trabalhar com apenas um canal.

A informação de áudio em formato estéreo pode ser codificada em três modos diferentes : stereo, dual e joint stereo.

Nos modos stereo e dual, os dois canais são transmitidos independentemente no mesmo frame sem remoção de qualquer tipo de redundância. O modo stereo é utilizado para transmitir os canais esquerdo (*left*) e direito (*right*) nas aplicações broadcasting, enquanto que o modo dual é utilizado para transmitir diferentes tipos de informação nos dois canais de áudio, como por exemplo a radiodifusão bilingüe. O modo joint stereo apresenta dois métodos para retirar a redundância contida no formato stereo broadcast com o objetivo de otimizar a codificação. O primeiro método chamado de intensity stereo é a única opção para os *layers* I e II. Nesses *layers*, os fatores de escala são transmitidos em forma normal para ambos os canais, enquanto que para o *layer* III, só os fatores de escala do canal direito são transmitidos; sendo que as amostras em sub-banda quantizadas são também tratadas de forma diferente. A qualidade subjetiva desse modo varia com a imagem estéreo do sinal codificado. Contudo, esse método é particularmente indicado para baixas taxas de bits, apresentando uma qualidade melhor do que os outros modos. Assim, deve ser reservado para aplicações, como transmissão, onde as baixas taxas de bits têm prioridade.

Para todas as sub-bandas acima de certo limiar, e dependendo do modo de extensão, os canais esquerdo e direito são transmitidos separadamente, e abaixo desse limiar é transmitida a soma de ambos os canais. Por outro lado, no modo MS stéreo disponível só no *layer* III, a redundância entre ambos os canais é explorada transformando os canais esquerdo e direito em somas e diferenças dependendo de qual operação resulta em maior quantidade de energia.

VI. LAYER I

Cada *frame* contém as últimas 12 amostras decimadas de cada uma das 32 sub-bandas resultantes da análise *filter bank*. Para cada uma das sub-bandas codificadas dentro do frame, as 12 amostras são escalonadas de forma que o máximo valor das mesmas não exceda de um. O modelo psico-acústico e a taxa de bits são utilizados para computar o número de bits alocados para cada sub-banda. Logo, as amostras escalonadas são quantizadas de acordo com o número de níveis de quantização determinado pelo algoritmo de alocação de bits.

Os bits alocados, os fatores de escala e as amostras quantizadas são codificadas e colocadas em três áreas designadas dentro do *frame*. Na Figura 9 apresentamos o diagrama de fluxo de codificação de um *frame* correspondente ao *layer* I.

O algoritmo de alocação de bits indica o número de bits alocados para cada amostra contida em uma determinada sub-banda, e assim determina também o número de níveis de quantização do quantizador linear a ser utilizado. Incrementando o número de níveis, reduz-se o ruído de quantização, mas aumenta-se o número de bits requerido para codificar as amostras, assim como a taxa de bits. A seção de alocação de bits dentro do *frame* contém 32 valores (para 32 sub-bandas) de 4 bits cada um que permitem escolher entre 15 diferentes quantizadores para cada sub-banda (para evitar conflito com o código de sincronização o valor '1111' é definido como sendo ilegal).

A seção dos fatores de escala contém 32 valores de 6 bits cada um, os quais indexam um dos 63 valores tabulados na tabela especificada no padrão MPEG (o código '111111' é ilegal). O fator de escala é utilizado para multiplicar a amostra requantizada de uma sub-banda. Esse valor é dependente da amplitude do sinal filtrado nessa sub-banda. Os fatores de escala apresentados na tabela se incrementam em fator de $\sqrt[3]{2}$ ou aproximadamente 2dB.

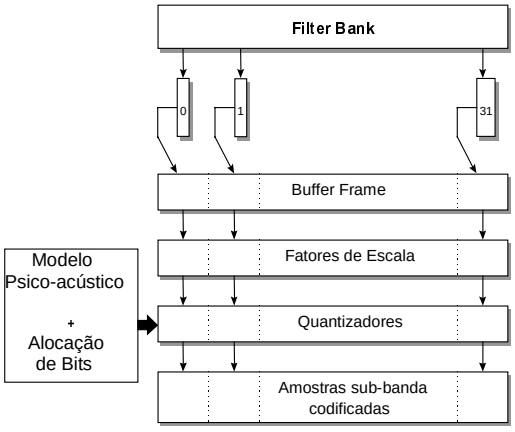

Figura9. *Diagrama de fluxo para a codificação de um frame MPEG*

Cada uma das 12 amostras correspondentes às 32 sub-bandas é codificada e alocada na seção de amostras dentro do *frame*. O padrão MPEG utiliza um quantizador linear *mid-tread* de *n* níveis, onde *n* é igual a um menos o antilogaritmo na base 2 do número de bits designados para a sub-banda. O método que é empregado no padrão para se realizar a quantização e a de-quantização das amostras evita que se transmita uma longa seqüência de 1's, evitando assim confusão com o código de sincronização do *frame*.

Na Figura 10 apresentamos o formato do *frame* MPEG para o *layer* I.

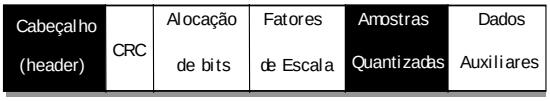

Figura 10. *Formato de um frame MPEG layer* I

VII. LAYER II

Nesse formato de compressão, o *frame* consiste de 36 amostras por sub-banda e é dividido em três partes: parte 0, parte 1 e parte 2. Cada parte contém 12 amostras por sub-banda da mesma forma que o *frame* do *layer* I. A área de alocação de bits é aplicável às três partes; o fator de escala pode ser transmitido separadamente para cada uma das três partes ou um único fator de escala pode ser aplicado a duas ou mais partes de uma sub-banda.

A seção de fatores de escala apresenta um novo campo de 2 bits denominado *scale fator selection information* (SFSI), o qual indica se um, dois ou três fatores de escala são transmitidos para uma determinada sub-banda e como eles são aplicados. Assim, mediante a utilização de SFSI, a *layer* II pode detectar possíveis transitórios dentro do sinal de áudio que está sendo codificado. A seção de alocação de bits também tem sido reduzida limitando a utilização do número de quantizadores para as sub-bandas das altas freqüências e para baixas taxas de bits. Assim, em lugar de se transmitir 4 bits por sub-banda para especificar o número de bits alocados à mesma, o número de bits varia de 0 a 4 como uma função do número da sub-banda tal como é especificado nas tabelas apresentadas no padrão MPEG para uma dada freqüência de amostragem e taxa de bits.

As amostras correspondentes a cada sub-banda são em seguida quantizadas da mesma forma que no *layer* I. No entanto, existe a possibilidade de quantizar três amostras consecutivas em um único código para certos quantizadores. Isso reduz o número de bits perdidos quando o quantizador não é exatamente uma potência de dois.

Na Figura 11, apresenta-se o formato do *frame* de áudio MPEG correspondente à *layer* II.

Figura 11. *Formato de uma frame MPEG-layer* II.

Na Figura 12 mostra-se o formato do frame de sincronização correspondente ao *layer* II e sua extensão ao formato multicanal do sistema MPEG-2

especificado na norma ISO/IEC 13818-3 MPEG-2 - *Layer* II. Pode-se observar na Figura 12 e na Figura 2 que a seqüência de codificação do *frame* é feita de forma que um decodificador estéreo MPEG-1 possa interpretar a informação contida nela caso seja MPEG-2 BC. Dessa forma, na área de extensão são codificados os canais restantes do formato multicanal do sistema MPEG-2.

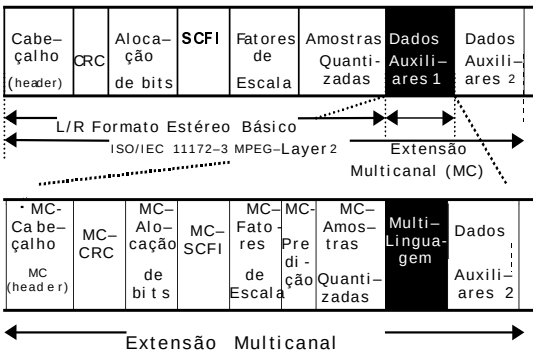

Figura 12. *Extensão de um frame de sincronização do sistema MPEG-1 layer* II *estéreo para o sistema multicanal MPEG-2 layer II (ISO/IEC 13818-3)*

VIII. CODIFICAÇÃO MPEG: LAYERS I E II

O sistema de compressão MPEG é capaz de manter a qualidade do sinal de áudio reproduzido no receptor (CD-*Compact Disk*) com uma taxa de compressão de aproximadamente 5 para 1 ou mais. Essa taxa de bits é equivalente a se codificar cada amostra de áudio com aproximadamente 3 bits por amostra.

Para um codificador PCM (*Pulse Code Modulation*), isso corresponde a uma relação nominal sinal/ruído (SNR) de 18 dB.

O sistema MPEG alcança esses níveis de compressão alocando ruído de quantização nas sub-bandas onde o ouvido humano apresenta menor sensibilidade. Dessa forma, o modelo psico-acústico determina, a partir do sinal de entrada, o nível de ruído perceptível (nível de enmascaramento) para cada uma das sub-bandas resultantes da análise *filter bank*.

Devido ao fato de que a quantidade de ruído de quantização é proporcional ao número de bits usados pelo quantizador, o algoritmo de alocação de bits aloca os bits disponíveis de uma forma que minimize a distorção audível.

IX. FATOR DE ESCALA

Como é descrito no padrão MPEG, o valor máximo absoluto das 12 amostras de cada sub-banda é definido como o fator de escala. Logo, esse valor é mapeado em uma tabela de especificação MPEG a fim de se codificar o mesmo através de um número binário de bits. Dessa forma, existem 64 possíveis combinações diferentes que indicam ao decodificador o valor aproximado do fator de escala calculado para uma determinada sub-banda. Esse valor dentro da tabela é mapeado como sendo o maior número inteiro mais próximo do valor calculado pelo codificador.

No *layer* II, três fatores de escala são calculados por cada sub-banda; um para cada parte de 12 amostras. Em seguida são determinadas as diferenças entre o primeiro fator de escala com o segundo, e o segundo com o terceiro. Logo, dependendo do valor dessas diferenças é decidido se é transmitido um, dois ou três fatores de escala para a atual sub-banda.

X. ALOCAÇÃO DE BITS

Os dois modelos psico-acústicos descritos no padrão MPEG retornam como resultado a relação sinal/enmascaramento (SMR) para cada sub-banda. O algoritmo de alocação de bits calcula a relação enmascaramento/ruído (MNR) a partir da SMR e da relação sinal/ruído (SNR) usando as seguintes expressões (A SNR é dada como uma função do número de bits alocados nas tabelas do padrão MPEG):

$$MNR = SMR - SNR \tag{7}$$

O número de bits disponíveis para a codificação de um *frame* é determinado a partir da taxa de bits desejada e da freqüência de amostragem a ser utilizada. O número de bits disponíveis para se codificar as amostras de áudio sub-banda é obtido subtraindo do número de bits por *frame* os 32 bits do cabeçalho, os 16 bits do CRC caso seja usado, o número de bits usado para indicar a alocação de bits para cada sub-banda e finalmente o número de bits utilizado para se codificar os dados auxiliares. O algoritmo de alocação de bits pretende maximizar logo o mínimo MNR de todas as sub-bandas alocando o número restante de bits aos fatores de escalas e às amostras sub-banda. Se são alocados zero bits a uma determinada sub-banda, então o fator de escala dessa sub-banda não é codificado; de outra forma dependendo do *layer*, é transmitido um, dois ou três fatores de escala para uma sub-banda. Inicialmente o algoritmo assume que zero bits são alocados para cada sub-banda. Em seguida, é computada a *SNR* e a *MNR* de cada sub-banda e é determinada a sub-banda com a mínima *MNR* cujo número de bits alocados não tenha ultrapassado o limiar máximo.

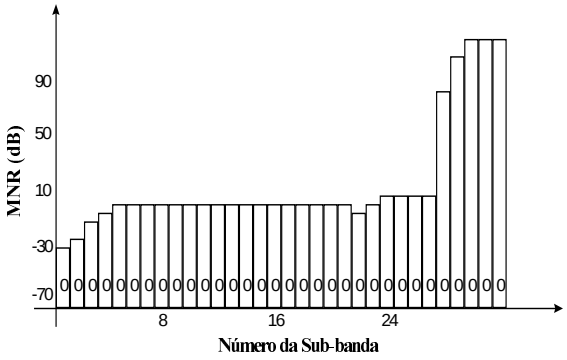

Figura 13. *MNR e Alocação de Bits antes da execução do Algoritmo*

A alocação de bits é incrementada em um nível e o número necessário de bits adicionais é subtraído do número de bits ainda disponíveis para se codificar as

amostras de áudio na sub-banda. O processo é repetido até que o número de bits utilizado para todas as sub-bandas tenha ultrapassado seu limiar máximo. Na Figura 13 e Figura 14 mostram-se respectivamente, a função *MNR* antes e depois da execução do algoritmo de alocação de bits.

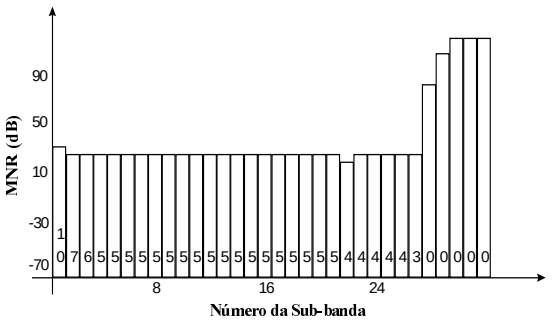

Figura 14. *MNR e Alocação de Bits após a execução do Algoritmo*

XI. MODELO PSICO-ACÚSTICO

O algoritmo MPEG/áudio comprime os dados de áudio em grande parte removendo as partes acústicas irrelevantes do sinal de áudio, isto é, ele tira proveito da incapacidade do sistema auditivo humano em escutar o ruído de quantização sobre condições de enmascaramento auditivo. Esse enmascaramento é uma propriedade perceptiva do sistema auditivo humano que ocorre na presença de um forte sinal de áudio que possui uma vizinhança espectral ou temporal comum a um sinal fraco de áudio.

A tarefa do modelo psico-acústico é analisar o sinal de áudio de entrada e determinar onde, no espectro, o sinal irá ser mascarado e a extensão do efeito.

Um modelo psico-acústico descreve principalmente as características da resposta em freqüência e da resposta no tempo de nosso sistema de audição.

Dessa forma, através de um modelo matemático, é possível descrever a maneira como são percebidas as distintas componentes de freqüência que conformam um sinal de áudio num determinado instante de tempo.

Uma das características mais importantes de nosso sistema de audição consiste do fenômeno de enmascaramento de componentes de baixa potência pela ocorrência simultânea de componentes de maior potência que ficam muito próximos dentro do espectro do sinal. Assim os níveis de ruído permitidos para a componente de baixa freqüência podem ser consideráveis, uma vez que a mesma é percebida em menor intensidade pelo nosso sistema de audição.

O enmascaramento de uma componente de baixa potência pode acontecer de duas formas: devido à ocorrência simultânea de componentes tonais isoladas de alta potência (enmascaramento tonal), ou devido à ocorrência simultânea de um grupo de componentes tonais de alta potência muito próximas entre si (enmascaramento não tonal).

À medida que uma determinada componente de freqüência fica a uma maior distância de componentes tonais ou não tonais de alta potência, o efeito de enmascaramento é menor.

O padrão MPEG descreve a implementação de dois modelos psico-acústicos. O primeiro modelo foi projetado a fim de que seja computacionalmente simples e para prover adequada precisão em altas taxas de bits. O segundo modelo é mais complexo e é recomendado para baixas taxas de bits. Ambos os modelos requerem o cálculo do espectro de potência através da *FFT (Fast Fourier Transform)*, mapeando o mesmo no domínio das bandas críticas do ouvido humano, distinguindo componentes tonais e não tonais, aplicando a *spreading function* (função que determina o nível de enmascaramento de componentes de freqüência devido à presença de componentes tonais e não tonais) nessas componentes, computando a ***função de enmascaramento (masking function)***, mapeando isso no domínio do espectro da transformada de Fourier e no domínio das 32 sub-bandas de codificação.

As tabelas das funções de mapeamento do limiar de sensibilidade do ouvido humano em função da freqüência, das representações paramétricas para a *spreading function* e para a função de enmascaramento são fornecidas por cada modelo psico-acústico para todas as freqüências de amostragem.

XII. CONCLUSÕES

O formato de compressão MPEG/áudio constitui atualmente um dos sistemas de melhor desempenho dentre os diversos sistemas que estão sendo considerados para transmissões de Televisão Digital, *FM Broadcasting*, HDTV, etc.

Os níveis de compressão e de qualidade alcançado por esse sistema devem-se principalmente à utilização de um processo de codificação sub-banda, baseado num modelo psico-acústico do ouvido humano que permite enmascarar a percepção de qualquer ruído introduzido no sinal de áudio após o processo de compressão. Assim mesmo, os *layers* I e II estudados neste trabalho apresentam algoritmos de complexidade intermediária que podem facilmente ser implementados em plataformas DSP's (Digital Signal Processors) e ser ainda mais adequados para sua utilização em aplicações *broadcasting*.

Como mencionamos anteriormente, a evolução do sistema MPEG-1 para o sistema MPEG-2 foi a utilização de um formato multicanal adequado para aplicações em HDTV. No entanto, os formatos de compressão utilizados por esses sistemas são praticamente os mesmos.

A codificação do *frame* de sincronização utilizada no sistema MPEG-2 BC constitui uma extensão do *frame* utilizado no sistema MPEG-1 devido à utilização de um maior número de canais de áudio. No entanto, a formatação do *frame* MPEG-2 BC é feita de tal forma que a mesma possa ser interpretada por qualquer decodificador MPEG-1. O sistema MPEG-2 AAC é uma evolução do sistema MPEG-2 BC, mas não garante uma compatibilidade com o sistema MPEG-1.

O *layer* III que não foi abordado neste trabalho se apresenta atualmente como o sistema de compressão

de áudio mais avançado da linha MPEG. O formato de compressão utilizado nesse sistema é diferente e muito mais complexo do que aqueles utilizados nos *layers* I e II. No entanto, os níveis de compressão alcançados são maiores e adequados para transmissões sobre sistemas **ISDN**.

XIII. AGRADECIMENTOS

Este trabalho contou com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (**FAPESP**), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (**CAPES**), da Fundação Centro de Pesquisa e Desenvolvimento em Telecomuicações (**CPqD**), do Fundo de Apoio ao Ensino e Pesquisa da UNICAMP (**FAEP/UNICAMP**), da Universidade Estadual de Campinas (**UNICAMP**) e da Universidade Particular Antenor Orrego de Trujillo-Peru (**UPAO**).

REFERÊNCIAS BIBLIOGRÁFICAS

[1] Seymour Shlien, "Guide to MPEG-1 Audio Standard", IEEE Trans. *Broadcasting*, vol 40, no. 4, 1998.

[2] Charles D. Murphy and K. Anandakumar, "Real-Time MPEG-1 Audio Coding and Decoding on a DSP Chip", *IEEE Trans. Consumer Electronics*, vol. 43. No. 1, 1997.

[3] ISO/IEC JTCI/SC29, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1,5 Mbits/s – IS 11172 (Part 3, Audio)", 1993

[4] ISO/IEC JTCI/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information – IS 13818 (Part 3, Audio)", 1995

[5] Pan, Davis Yen, "Digital Audio Compression", Digital Technical Journal, vol 5, nº 2, Spring 1993

[6] Pan, Davis Yen, "A Tutorial on MPEG/Audio Compression," IEEE Trans.on Multimidia, Vol 2, nº 2, 1995, pp 60-74

[7] Valle, André, "MP3. A revolução do som via Internet", Reichmann & Affonso Editores, 1999

[8] Brandenburg, K.;Stoll,G,"The ISO-MPEG Audio Codec: A Generic Standard for Coding of High Quality Digital Audio" . JAES, Vol.42,NG 10,1994 October, pp.780-792

[9] Noll, Peter, "MPEG Digital Audio Coding", IEEE Signal Processing Magazine, pp 59-81, september 1997

Edgard Luciano Oliveira da Silva nasceu em Poços de Caldas, MG, em 22 de dezembro de 1967. Formou-se em Engenharia Elétrica na Escola Federal de Engenharia de Itajubá (EFEI) em 1990. Obteve o grau de Mestre em Engenharia Elétrica pela Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, SP (FEEC - UNICAMP) em 1996. Atualmente está no programa de Doutorado da FEEC-UNICAMP, onde continua suas pesquisas na área de processamento digital de sinais, codificação e sinais da fala.

Yuzo Iano nasceu em 29 de abril 1950. Recebeu os títulos de Engenheiro Eletricista, Mestre e Doutor pela UNICAMP (1972,1974 e 1986). Atualmente é Professor MS-5 do DECOM/FEEC/UNICAMP (Adjunto). É também o Coordenador do curso de graduação de Engenharia Elétrica da FEEC/UNICAMP. Trabalha com transmissão digital de sinais desde 1973, inicialmente com telefonia e posteriormente com televisão e em especial com HDTV desde 1986. Os interesses atuais abrangem processamento e tratamento digital de áudio e vídeo (som e imagem).

Guillermo Kemper Vásques nasceu em Trujillo-Perú em 1971. Recebeu o grau de Engenheiro Eletrônico na Universidade Particular Antenor Orrego de Trujillo em 1994, e o de Mestre em Eletrônica e Comunicações na Universidade Estadual de Campinas (UNICAMP) em 1996. Atualmente faz doutorado na UNICAMP desenvolvendo trabalhos na área de compressão de sinais áudio orientado a sistemas de televisão de alta definição HDTV.

José Geraldo Chiquito recebeu o título de Engenheiro Eletrônico pelo ITA (Instituto Tecnológico da Aeronáutica) em 1974. Recebeu o título de Doutor em Engenharia Elétrica através da UNICAMP (Universidade Estadual de Campinas) em 1983. Atualmente é o responsável pelo Laboratório de Processamento de Sinais do DECOM/FEEC/UNICAMP. Seus interesses estão voltados para Transmissão e Modulação Digitais, Processamento de Sinais, Instrumentação Eletrônica e televisão de alta definição.

e-mail: edgard@decom.fee.unicamp.br
yuzo@decom.fee.unicamp.br;
gkemper@decom.fee.unicamp.br