

Compressão de Sinais de Áudio para HDTV: Sistema DOLBY AC-3 (U.S.A)

Guilherme Leopoldo Kemper Vásques, Yuzo Iano, Edgard Luciano Oliveira da Silva

Faculdade de Engenharia Elétrica e de Computação – Universidade Estadual de Campinas
Departamento de Comunicações

Resumo: Os sistemas de televisão HDTV que se encontram em processo de padronização no mundo inteiro consideram em todos os casos a transmissão de um sinal digital de áudio submetido a processos de compressão com grande eficiência. Esse critério é levado em conta com o objetivo de se obter sinais de áudio com som de alto nível de fidelidade e naturalidade associados a taxas de bits adequadas para sistemas de televisão de alta definição. Esse é o caso do sistema de televisão HDTV proposto pela Grande Aliança dos Estados Unidos da América do Norte, onde foi considerado o sistema de compressão de áudio DOLBY SURROUND AC-3 para a codificação dos sinais de som associados às imagens de alta definição. O codificador AC-3 suporta a transmissão de 5.1 canais de áudio que permitem obter um som de alta qualidade para taxas de transmissão na faixa de 384 Kbit/s. O sistema AC-3 é conformado por vários blocos funcionais de codificação que serão objeto de estudo neste artigo.

I. INTRODUÇÃO

Os sistemas de televisão HDTV que se encontram em processo de padronização no mundo inteiro consideram em todos os casos a transmissão de um sinal digital de áudio submetido a processos de compressão de grande desempenho. Esse critério é levado em conta com o objetivo de se obter sinais de som com alto nível de fidelidade e naturalidade associados a taxas de bits adequadas para sistemas de televisão de alta definição. Esse é caso do sistema de televisão HDTV proposto pela Grande Aliança dos Estados Unidos da América do Norte (aliança política e técnica que envolve a participação de 3 grupos de pesquisa selecionados para se gerar um sistema de televisão avançada HDTV para os Estados Unidos), onde foi considerado o sistema de compressão de áudio DOLBY SURROUND AC-3 para a codificação dos sinais de som associados às imagens de alta definição.

O sistema AC-3 suporta a transmissão do sinal através de 5 canais de áudio mais um canal de baixas frequências (considerado formato de 5.1 canais de áudio). Isso resulta em um melhor desempenho dos sons fortes como explosões, motores e efeitos especiais.

A frequência de amostragem considerada pelo sistema AC-3 é de 48 KHz por canal, o que demanda uma taxa de aproximadamente 5 Mbit/s para a transmissão dos 5.1 canais de áudio. Isso constitui um

problema devido às limitações do canal de transmissão designado para sistemas HDTV (6 MHz, U.S.A; 8MHz Europa), pois, de fato, no sistema completo, deve-se levar junto com o sinal de áudio a informação do sinal de vídeo cuja taxa é de 18.2 Mbit/s (formato MPEG-2, americano).

Para solucionar esse problema é necessário realizar a compressão do sinal a fim de se reduzir a taxa de bits a transmitir conservando os padrões de qualidade do sinal original. O sistema Dolby AC-3 oferece um potente sistema de compressão baseado na sensibilidade do ouvido humano que é diferente para as diversas componentes de frequência que formam o espectro da audição. Isso permitirá designar os níveis de quantização adequados, de acordo com o nível de percepção de uma componente de frequência em um determinado instante de tempo. Assim, as degradações que porventura o sinal vier a sofrer não serão percebidas pelo sistema de audição. Com a aplicação desses critérios, o codificador AC-3 é capaz de transmitir os 5.1 canais de áudio através de uma taxa de 384 Kbit/s, conservando os níveis de qualidade de áudio exigidos para os sistemas HDTV.

Os diagramas de blocos do codificador e do decodificador AC-3 são mostrados nas Figuras 1 e 2, respectivamente. Note que os mesmos são formados por vários blocos funcionais de codificação que serão objeto de estudo neste artigo.

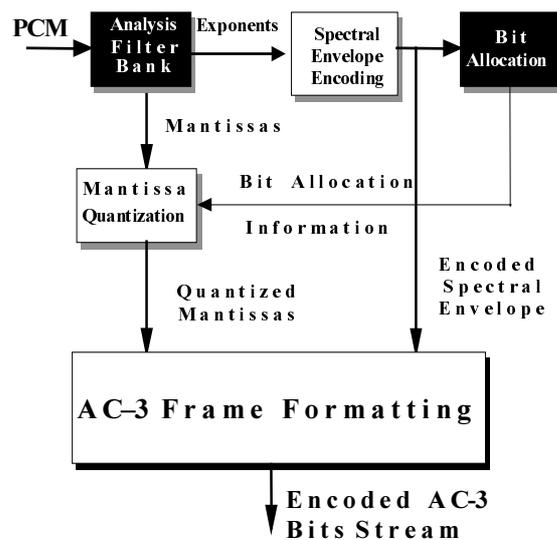


Figura 1. Codificador DOLBY AC-3

II. CANAIS DE ÁUDIO

O sistema de compressão Dolby AC-3 trabalha com até seis canais de áudio de entrada no sistema de codificação (geralmente os seis canais são considerados como 5.1 canais de áudio, pois o canal de baixas frequências é considerado como 0.1). O sistema de canalização é feito de acordo com a recomendação ITU-R BS-775, denominado “sistema de som estéreo multicanal com e sem acompanhamento de vídeo”. A largura de faixa designada para cada canal é mostrada na Tabela 1.

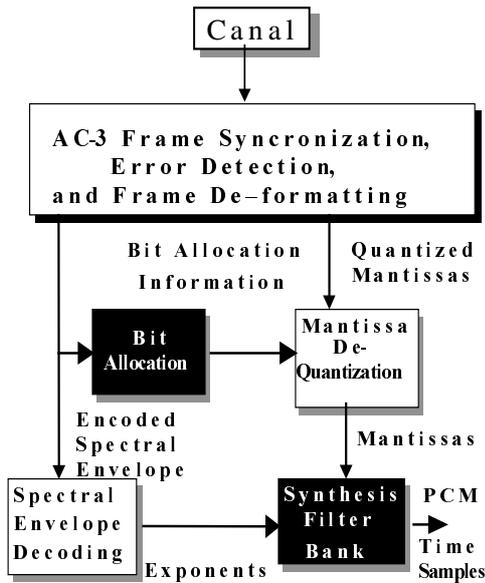


Figura 2. Decodificador DOLBY AC-3

| Canal de Áudio | Largura de Faixa |
|------------------------------------|------------------|
| Esquerdo | 20 KHz |
| Direito | 20 KHz |
| Central | 20 KHz |
| Esquerdo Adjacente (left surround) | 20 KHz |
| Direito Adjacente (right surround) | 20 KHz |
| LFE (Canal de Baixas Frequências) | 120 Hz |

Tabela 1. Canais de áudio suportados pelo codificador AC-3

A disposição de cada canal dentro do ambiente de reprodução é atribuída pelo sistema AC-3 tal como mostra a Figura 3. Cada sinal de entrada no sistema pode estar na forma analógica ou na forma digital. No caso de estar em formato analógico, deve-se fazer a conversão analógico-digital a fim de que se possa processar o sinal usando-se o algoritmo de compressão e codificação.

III. CONVERSÃO ANALÓGICO-DIGITAL

Cada canal de entrada no codificador AC-3 pode ser amostrado a uma frequência de 32 KHz (FM Broadcasting), 44.1 KHz ou 48 KHz (HDTV). As frequências de 44.1 KHz e 48 KHz são utilizadas para áudio com acompanhamento de vídeo, e portanto constituem um compromisso entre as frequências de

varredura de quadro e de linha, a fim de se assegurar um número inteiro de amostras de áudio por quadro ou campo. Para o caso de HDTV, é utilizada uma frequência de amostragem de 48 KHz, que é gerada a partir de um oscilador padrão de 27 MHz, a fim de se sincronizar o codificador de áudio, o codificador de vídeo e o sistema de transporte.

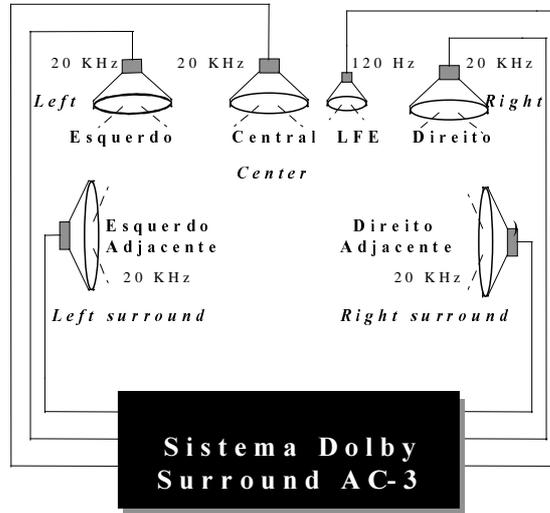


Figura 3. Disposição Espacial dos Canais de Áudio do Sistema AC-3

Se o sinal de entrada no codificador estiver em formato digital e a uma frequência de amostragem diferente, o codificador faz a conversão a 32 KHz, 44.1 KHz ou 48 KHz (dependendo do caso), a fim de que todo o sinal seja codificado a partir do mesmo formato de amostragem.

Cada amostra de áudio pode ser quantizada e codificada com uma precisão de 16 bits a 24 bits. Dessa forma, na saída do conversor analógico digital se tem disponíveis amostras PCM (Pulse Code Modulation) quantizadas nessa faixa de precisão.

A filtragem dos canais de áudio e o processo de conversão A/D são mostrados na Figura. 4. Note que no caso do canal LFE, a filtragem passa-baixas é feita por um filtro com frequência de corte em 120 Hz. O nível DC contínuo é eliminado mediante um filtro passa-altas com uma frequência de corte de 3 Hz (-3dB).

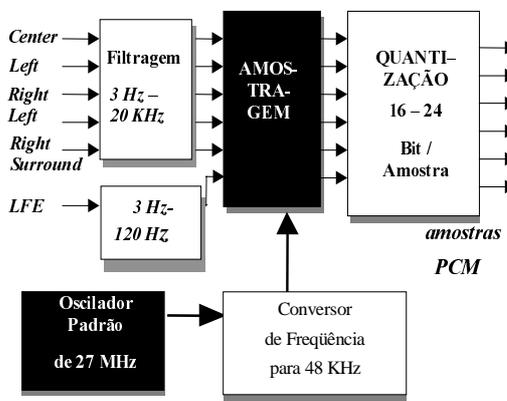


Figura. 4. Conversão Analógico-Digital dos Sinais de Áudio.

IV. ANÁLISE/SÍNTESE FILTER BANK

O sistema AC-3 explora as características do sinal dentro do domínio da frequência (Análise Espectral *Filterbank*), a fim de se fazer o tratamento digital com base na sensibilidade que cada componente de frequência tem dentro do sistema de audição. A vantagem disso é que, de acordo com a importância da componente, o valor do coeficiente correspondente pode ser quantizado com maior ou menor número de níveis de quantização, de forma que o número de bits designado para a representação de cada amostra seja variável e muito menor do que aquele correspondente ao do sinal original.

No codificador AC-3, a transformada de um sinal de áudio para o domínio da frequência é feita a partir da formação de blocos de áudio de 512 amostras com uma superposição de 256 amostras com blocos adjacentes (50%). Assim, um bloco de áudio é formado (cada 5.33 ms, para 48 KHz) a partir de 256 amostras prévias e de 256 amostras novas.

Essa forma de segmentação dobra a quantidade de informação a ser processada, mas sua vantagem está no fato de que, quando o sinal é decodificado, o som reproduzido fica livre do efeito de blocos (ruído “click”) que geralmente aparece quando um sinal de infinitas amostras é processado através de blocos com um número finito de amostras.

Cada bloco de áudio é passado por um processo de janelamento que envolve a multiplicação das 512 amostras de um bloco por uma janela de 512 elementos cuja forma é apresentada na Figura 5. A forma da janela determina a seletividade da transformada utilizada na análise *filter bank* e permite que as amostras resultantes do processo de superposição/soma apresentado na decodificação de blocos de áudio adjacentes se encontrem livres do efeito de bloco.

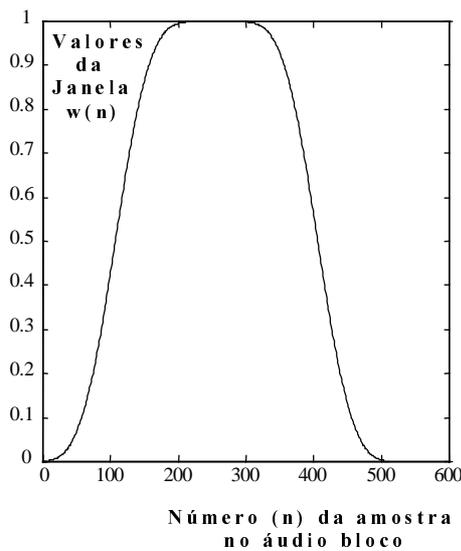


Figura 5. Janela aplicada ao sinal de áudio pelo sistema AC-3.

A análise *filter bank* AC-3 é implementada através da MDCT (*Modified Discrete Cosine Transform*) e é

baseada na teoria de Cancelamento de Aliasing no Domínio do Tempo (TDAC :*Time Domain Aliasing Cancellation*).

Isso implica em que a transformada seria feita intercalando-se a DCT (*Discrete Cosine Transform*) e a DST (*Discrete Sine Transform*) em blocos adjacentes de 512 amostras, sendo que para cada bloco de áudio são transmitidos somente 256 coeficientes. Esse procedimento logicamente vai gerar *aliasing* e distorção nas amostras de áudio decodificadas no receptor, mas por outro lado, vai eliminar a redundância introduzida no sinal devido à formação dos blocos de áudio. No entanto, a multiplexação DCT/DST permite que a superposição de espectros introduzida nas amostras de áudio decodificadas no receptor possa ser eliminada. Isso é conseguido aproveitando-se a característica de que a DCT e a DST apresentam o mesmo tipo de *aliasing* em blocos adjacentes, mas de sinal contrário.

A superposição de espectros é apresentada como um *overlap/add*, tal como mostram as seguintes expressões :

$$X(k) = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi k}{N}(n+n_0)\right) \quad \text{(DCT) (1)}$$

$$k = 0 \dots N/2 - 1$$

$$y(n) = \sum_{k=0}^{N/2-1} X(k) \cos\left(\frac{2\pi k}{N}(n+n_0)\right) \quad \text{(IDCT) (2)}$$

$$n = 0 \dots N - 1$$

$$= \frac{x(n)}{2} + \frac{x(N-n-2n_0)}{2} \quad \text{(aliasing-DCT) (3)}$$

onde $x(n)$ é o sinal original, $y(n)$ é o sinal resultante, N é o número de amostras que conformam o bloco do sinal original, e n_0 é um fator de fase opcional.

Fazendo-se os mesmos cálculos mediante a DST, o sinal resultante é dado pela seguinte equação:

$$y(n) = \frac{x(n)}{2} - \frac{x(N-n-2n_0)}{2} \quad \text{(aliasing-DST) (4)}$$

Um ajuste adequado do fator de fase n_0 permitirá a obtenção do mesmo tipo de *aliasing* entre o segundo grupo de 256 amostras de um determinado bloco de áudio e o primeiro grupo de 256 amostras do bloco de áudio consecutivo. A eliminação ou cancelamento do *aliasing* poderá ser feita fazendo-se a soma entre esses grupos de amostras. Dessa forma, será eliminada a distorção devido ao *aliasing* originado pela transformada, e o sinal original é finalmente recuperado (sem considerar o processo de quantização das componentes de frequência).

O sistema AC-3 aplica esses mesmos critérios utilizando a DCT IV e fazendo um adequado ajuste do fator de fase n_0 . É construída assim uma MDCT que permite a obtenção do mesmo tipo de *aliasing* entre o segundo grupo de 256 amostras de um bloco de áudio, e o primeiro grupo de 256 amostras do bloco consecutivo. Novamente, a eliminação ou cancelamento da superposição poderá ser feita fazendo-se a soma entre esses grupos de amostras. Esse processo implicará no armazenamento das

amostras do bloco de áudio atual até que o próximo bloco consecutivo seja decodificado.

A MDCT utilizada no sistema AC-3 é dada pela seguinte expressão:

$$X(k) = \frac{-2}{N} \sum_{n=0}^{N-1} x_j(n) \cos\left(\frac{2\pi}{4N} (2n+1 + 2\underbrace{[(N/4)(\alpha+1)]}_{n_0}) (2k+1)\right) \tag{5}$$

$$k = 0 \dots N/2$$

onde $x_j(n)$ representa as amostras de áudio passadas através da janela da Figura 5, $n_0 = (N/4)(1+\alpha)$ e α pode assumir os valores de -1, 0 e +1.

A MDCT expressa pela Equação (5) permite obter *aliasing* negativo entre as amostras da primeira metade do bloco de entrada, e *aliasing* positivo entre as amostras da segunda metade. Se $\alpha = 0$, isso acontece porque não tem sido detectado um transitório dentro do bloco de áudio que está sendo transformado. Para esse caso, $n_0 = N/4$ e $N = 512$, pois as 512 amostras do bloco de áudio vão ser transformadas num único conjunto de 512 amostras no domínio da frequência.

Quando um transitório (exemplo: um impulso ou um som de castanhetas) é detectado no bloco atual de áudio (a detecção é aplicada ao segundo grupo de 256 amostras de áudio), então existe a possibilidade de que o erro de quantização dos coeficientes MDCT seja audível e ocasione uma distorção no som reproduzido no receptor (ruído de pré-eco). Para corrigir esse problema, a transformação do bloco de amostras onde foi detectado o transitório é feita em dois grupos. No primeiro grupo são transformadas as primeiras 256 amostras desse bloco de áudio para o qual $\alpha = -1$, $N = 256$ e $n_0 = 0$. Para o segundo grupo o valor de $\alpha = +1$ e $n_0 = N/2$.

Para esse caso, são calculados 128 coeficientes no primeiro grupo e em seguida eles são intercalados com os 128 coeficientes obtidos do segundo grupo para se formar um bloco de 256 coeficientes que é codificado da mesma forma que os blocos onde não aconteceram transitórios. Os coeficientes gerados em cada grupo representam frequências de até 24 KHz. Esse tipo de processamento é feito, a fim de se evitar o pré-eco ocasionado pelo transitório detectado num determinado bloco de áudio.

Os coeficientes MDCT gerados a partir da Equação (5) são calculados a partir da FFT (*Fast Fourier Transform*), a fim de se aumentar a eficiência computacional do codificador AC-3. Os resultados de pesquisas realizadas por **Zelinski** e **Noll** sobre o tipo de transformada aplicado a sinais de voz e áudio indicam que existe uma preferência pela DCT em relação à DFT. Isso é devido ao fato de que a primeira apresenta um desempenho maior de cerca de 4 a 5 dB de SNR (relação sinal/ruído) sobre a segunda, para blocos de 128 amostras, sendo que essa diferença diminui à medida que o tamanho do bloco a ser transformado aumenta.

V. FRAME DE SINCRONIZAÇÃO AC-3

Um *frame* de sincronização é formado a partir do agrupamento de seis blocos codificados de áudio, com campos adicionais de sincronização, sinalização e de detecção de erros. Cada *frame* representa 1536 amostras de áudio codificadas e constitui a unidade de codificação e de acesso aleatório do sistema.

O formato de um *frame* de sincronização AC-3 é mostrado na Figura 6. Note que a informação de áudio, formada pelos blocos de áudio AB0..AB5, está acompanhada de informação adicional constituída pelos campos SI, (sincronização), BSI (informação) e CRC (detecção de erro).

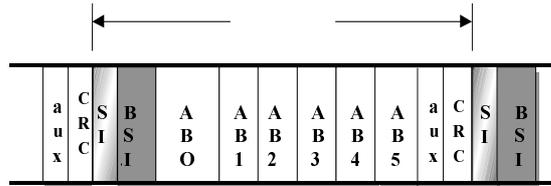


Figura 6. Formato de um Frame de Sincronização AC-3.

VI. CODIFICAÇÃO DOS COEFICIENTES MDCT

Após a obtenção do sinal de áudio no domínio da frequência, cada coeficiente MDCT é representado por números reais cujos valores são escalonados na faixa de -1 a 1. O formato empregado no sistema AC-3 para a representação desse tipo de números é a notação em ponto flutuante. Assim, um coeficiente com valor 0.000101011000100_2 expresso em sistema binário e com precisão de 16 bits pode ser representado em formato ponto flutuante como sendo $2^{-3} \times 0.1101011000100_2$; onde o número de zeros depois do ponto decimal representa o valor do expoente que neste caso é -3. Os 13 bits consecutivos (101011000100) depois do ponto decimal representam o valor da mantissa normalizada que será posteriormente quantizada e codificada junto com o expoente. Além disso, acrescenta-se também um bit de sinal para identificar valores positivos e negativos.

VI.1. CODIFICAÇÃO DOS EXPOENTES

Cada bloco de áudio contido num *frame* é formado por um conjunto de expoentes e seu correspondente conjunto de mantissas. Dessa forma, um *frame* de sincronização é formado por seis conjuntos de expoentes e seis conjuntos de mantissas. Para a codificação dos conjuntos de expoentes, o sistema explora o grau de diferença que existe entre eles, a fim de se aplicar os seguintes critérios:

- Se a diferença entre os conjuntos é pequena então apenas um conjunto é transmitido no primeiro bloco de áudio (campo AB0), e reutilizado pelos outros blocos de áudio que pertencem a um determinado canal.
- Se a diferença é grande, então apenas um conjunto de expoentes é transmitido para blocos onde a diferença é pequena, de tal forma que grupos de

blocos compartilhem o mesmo conjunto de expoentes.

- Cada conjunto de expoentes é codificado explorando-se a vantagem de que a diferença entre expoentes de frequências adjacentes é de no máximo ± 2 (± 12 dB). Dessa forma, o sistema transmite completamente o primeiro valor de cada conjunto e os seguintes valores consecutivos em forma diferencial. Cada expoente é codificado com valores de 5 bits, a fim de permitir que todos os possíveis valores (0...24) possam ser codificados. O primeiro expoente de cada conjunto é codificado com 4 bits. Para o caso de canais independentes, o valor limite para a codificação do primeiro expoente é de 15 (4 bits) considerando o bit mais significativo (mas não transmitido) igual a zero. No caso de canais acoplados (canais formados pelo acoplamento de outros canais em certas faixas de frequências) o bit menos significativo é considerado sempre zero e portanto podem ser transmitidos os valores pares do intervalo 0,2,4,...,24. Para a codificação de cada conjunto de expoentes, o sistema AC-3 utiliza três estratégias de codificação : D15, D25 e D45.

A estratégia D15 codifica a diferença entre os expoentes individuais de frequências adjacentes. A estratégia D25 codifica a diferença entre cada par de expoentes consecutivos, sendo que o codificador considera o segundo expoente de cada par igual ao primeiro. Finalmente, a estratégia D45 codifica a diferença entre grupos consecutivos de 4 coeficientes, sendo que o codificador considera os 3 últimos expoentes de cada grupo iguais aos do primeiro.

VI.2. ALOCAÇÃO DE BITS

O processo de alocação de bits consiste de um algoritmo que tem como finalidade a determinação do número de níveis de quantização necessário para a codificação de cada mantissa. A execução do algoritmo é feita a partir de um modelo psico-acústico do ouvido humano, a fim de evitar que a distorção introduzida no sinal seja percebida pelo nosso sistema de audição. O sistema AC-3 especifica o seguinte algoritmo de alocação de bits:

VI.3. CÁLCULO DA DENSIDADE ESPECTRAL DE POTÊNCIA (PSD)

Neste procedimento, o codificador calcula a densidade espectral de potência **PSD** com base nos valores dos expoentes de cada coeficiente. A função PSD é expressa através de valores de 13 bits com sinal e é calculada através da seguinte equação:

$$psd[k] = 3072 - (\exp[k] \ll 7) \quad (6)$$

onde o índice k indica o coeficiente MDCT correspondente, enquanto que $\exp(k)$ é o expoente do coeficiente k . A função $(v \ll n)$ indica um deslocamento de n bits para a esquerda do valor v .

Desde que $\exp(k)$ assume valores inteiros no intervalo de 0 a 24, então os valores da densidade espectral de potência calculados logaritmicamente através da Equação (6) estarão na faixa dinâmica de 0 (para níveis baixos do sinal) a 3072 (para níveis altos do sinal de áudio). Portanto, a **PSD(k)** é uma referência do nível de intensidade que as componentes do espectro de um determinado bloco de áudio apresentam.

VI.4. INTEGRAÇÃO DA PSD(K) EM BANDAS

O codificador AC-3 faz a divisão do espectro MDCT em 50 bandas de largura variável denominadas bandas críticas ou *critical bands*, sendo que as bandas correspondentes às componentes de baixas frequências são mais estreitas do que aquelas que agrupam componentes de alta frequência (Figura 7). Esse formato de divisão do espectro auditivo aproxima-se da forma como o sinal de áudio é percebido pelo nosso sistema de audição. Assim, as componentes agrupadas em uma mesma banda são percebidas aproximadamente com a mesma intensidade, e portanto codificadas com o mesmo número de níveis de quantização.

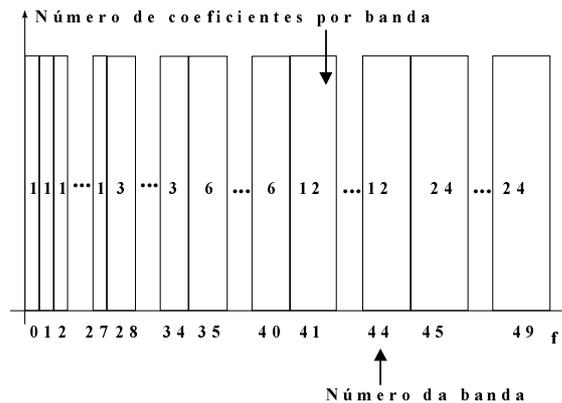


Figura 7. Divisão do Espectro Auditivo em Bandas Críticas

A integração do mapeamento da $PSD(k)$ nas bandas apresentadas na Figura 7 é feita logaritmicamente mediante a utilização de uma operação chamada de *log-add*, a qual é dada pela seguinte equação:

$$\log(a + b) = \max[\log(a), \log(b)] + \log(1 + e^d) \quad (7)$$

onde :

$$d = |\log(a) - \log(b)| \quad (8)$$

O segundo termo do lado direito da equação (8) é implementado como um subtração $\log(a) - \log(b)$, seguida de uma operação de valor absoluto e uma *lookup table* (tabela de valores armazenados na memória). O conteúdo da endereço de memória “d” é: $\log(1 + e^d)$. Assim, uma *completa log-addition* é calculada com uma instrução de adição, comparação e *lookup table*.

Após a realização da integração de $PSD(k)$ em cada banda através da Equação (7), são obtidos os valores $\text{banpsd}(k)$, onde desta vez, k varia de 0 à 49,

especificando as 50 bandas em que foi dividido o espectro auditivo.

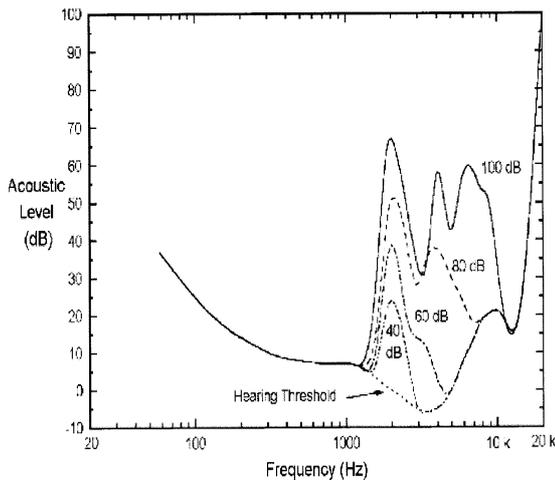


Figura 8. Spreading Function para um tom de 2 KHz com intensidades de 40, 60, 80 e 100 dB (Fonte [2]).

VI.5. CÁLCULO DA FUNÇÃO DE ESPALHAMENTO

A função de espalhamento ou *spreading function* estima os limiares de ruído permitidos para cada componente de frequência, segundo um modelo paramétrico que é ajustado de acordo com a forma do espectro MDCT expresso por $banpsd(k)$.

Os parâmetros que definem a *spreading function* são ajustados continuamente até se alcançar uma estimação ótima da função. Os valores finais dos parâmetros são transmitidos para o decodificador, a fim de que a rotina seja executada apenas uma vez no receptor. A *spreading function* é gerada a partir da exploração do fenômeno de enmascaramento de componentes de baixa potência, que se tornam inaudíveis devido à ocorrência simultânea de componentes de alta potência. Essa característica de nosso sistema de audição é apresentada na Figura 8. Note que para esse caso são calculadas as *spreading functions* para uma componente enmascaradora de 2 KHz com vários níveis de intensidade (40, 60, 80 e 100dB). As curvas indicam o nível de enmascaramento das componentes localizadas na vizinhança da componente enmascaradora. Se normalizamos as *spreading functions* da Figura 8 relativo ao nível da componente enmascaradora, obtemos as funções apresentadas na Figura 9.

A partir desse conjunto de curvas uma *spreading function* é construída. Ela especifica o nível de enmascaramento das componentes espectrais para o pior caso (menor nível de enmascaramento). A curva resultante (desenhada em negrito) é a *spreading function* utilizada no algoritmo de alocação de bits.

O sistema AC-3 faz a aproximação dessa função mediante dois segmentos lineares especificados pela sua pendente negativa e pelo seu *offset*. Assim mesmo, tal como mostra a Figura 10, a fim de se simplificar e generalizar o modelo, a componente enmascaradora é sempre localizada na posição zero. O cálculo da *spreading function* é feito através da convolução do

modelo especificado pelos dois segmentos lineares com a função $banpsd(k)$ calculada no passo anterior. A aproximação é feita ponto a ponto sobre os dois segmentos, escolhendo-se para cada *critical band* o máximo valor em cada ponto.

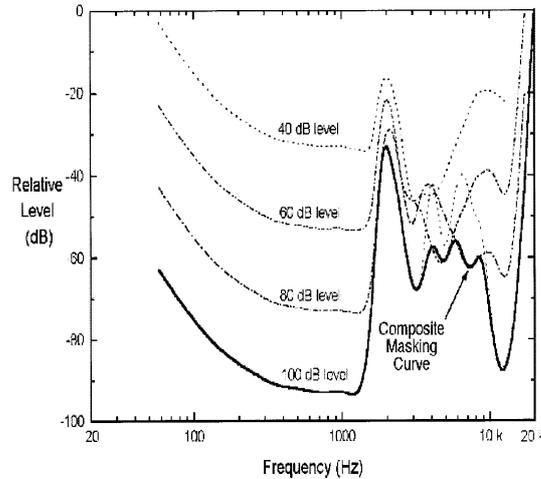


Figura 9. Spreading Functions Relativas à intensidade da Componente Enmascaradora (Fonte:[2])

O codificador executa a rotina várias vezes, a fim de calcular a inclinação e o *offset* ótimo (menor alocação de bits) de cada segmento. Logo esses parâmetros são transmitidos ao decodificador onde o algoritmo será executado uma única vez.

A convolução é feita através das seguintes expressões:

$$x_0 = (x_0 - d_0) \oplus (banpsd(k) - g_0) \tag{9}$$

$$x_1 = (x_1 - d_1) \oplus (banpsd(k) - g_1) \tag{10}$$

$$E(k) = \max(x_0, x_1) \tag{11}$$

onde d_0 e d_1 são as inclinações do primeiro e segundo segmento expressas em dB; g_0 e g_1 são os *offsets* expressos em dB do primeiro e segundo segmento, respectivamente. O símbolo \oplus indica a operação *log-addition* expressa pela Equação (7). Para cada uma das 50 bandas, os valores dos dois acumuladores x_0 e x_1 são calculados computando-se a *log-addition* dos acumuladores prévios diminuídos das inclinações d_0 e d_1 , e a densidade espectral de potência integrada $banpsd(k)$ escalonada pelos *offsets* g_0 e g_1 . A função resultante da aproximação $E(k)$ é chamada de função de excitação.

VI.6. CÁLCULO DA CURVA DE ENMASCARAMENTO

Depois que os valores da função de excitação são calculados, faz-se uma comparação entre os níveis de ruído especificados pela mesma com os limiares de ruído permitidos pelo ouvido humano para cada componente de frequência. Em seguida é escolhido o valor máximo, gerando finalmente a curva de enmascaramento **mask(k)** (*masking curve*). Essa operação permite assegurar que nunca se demandará

uma precessão mais alta do que o nível máximo de sensibilidade do nosso sistema de audição.

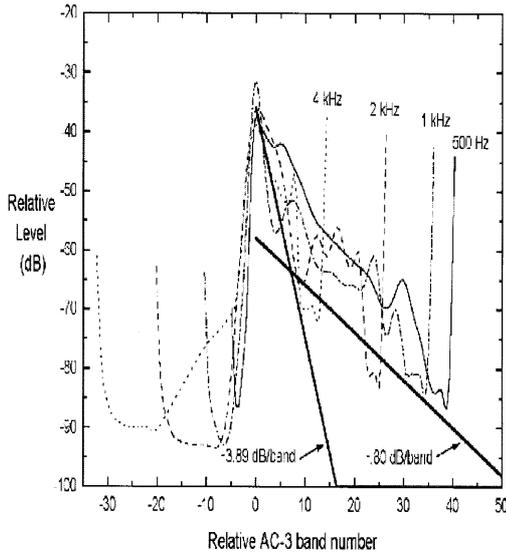


Figura 10. Aproximação das Spreading Functions Relativas mediante dois segmentos lineares para tons entre 500Hz e 4KHz (Fonte[2])

VI.7. CÁLCULO DOS PONTEIROS DE ALOCAÇÃO DE BITS (BAP)

Nessa parte final da rotina, o codificador calcula o tipo de quantizador designado para cada coeficiente. O cálculo é feito subtraindo-se o valor correspondente da curva de enmascaramento $mask(k)$ dos valores $PSD(k)$. Logo, o resultado obtido dessa diferença é deslocado para a direita em 5 bits e o valor resultante é tomado como o endereço do ponteiro bap . Note que esse procedimento equivale ao cálculo da $SNR(k)$ em dB. Dessa forma, quanto maior for a SNR de um determinado coeficiente maior também será o número de níveis de quantização designado para a mantissa correspondente.

A soma do número de bits alocados para cada mantissa deve ser menor ou igual ao total de bits disponíveis para a codificação de um determinado *frame*. Dessa forma, o sistema pode fazer o ajuste da curva de enmascaramento através dos valores $csnroffst$ (6 bits) e $fsnroffst[ch]$ (4 bits) (disponíveis no algoritmo) que são transmitidos ao decodificador. O valor $csnroffst$ indica ganhos múltiplos de ± 6 dB no nível de ruído para todas as bandas da curva de referência. O valor $fsnroffst$ indica ganhos de ruído múltiplos de $\pm 3/8$ dB para todas as bandas da curva de enmascaramento de um determinado canal (ch). Quando o $csnroffst$ indica um ganho de 6 dB, isso corresponde a um incremento de um bit na resolução dos valores de cada mantissa, pois isso está indicando um menor nível de ruído para todos os coeficientes do espectro.

Os resultados finais da rotina são portanto os valores $csnroffst$, $fsnroffst$ e os ponteiros bap . Finalmente, depois que esses resultados são obtidos, o sistema tem a possibilidade de aplicar um segundo modelo psico-acústico de maior desempenho que pode

melhorar a qualidade do som no receptor mediante a inserção de bits no *frame*. Esses bits adicionais indicam um ajuste na curva de enmascaramento de referência quando o processo de decodificação está sendo realizado. Esse processo denomina-se alocação *delta* de bits e indica ajustes de ± 6 dB na curva de enmascaramento a fim de que o som reproduzido no receptor seja de melhor qualidade. Na Figura 11 apresentamos finalmente o diagrama de blocos do algoritmo de alocação de bits AC-3.

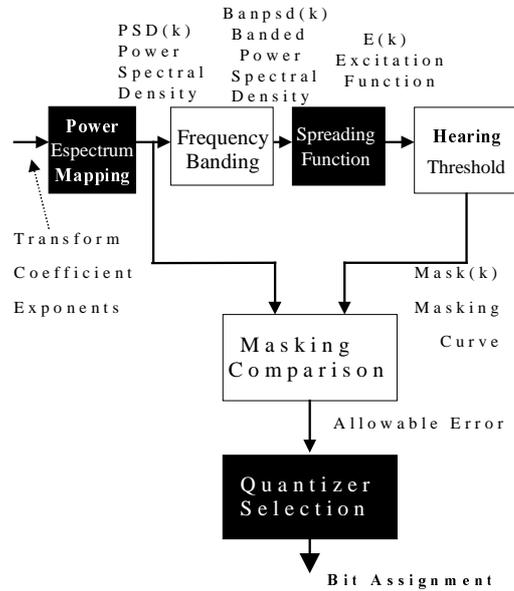


Figura 11. Diagrama de Blocos do Algoritmo de Alocação de Bits AC-3.

VI.8. QUANTIZAÇÃO DAS MANTISSAS

Após serem calculados os ponteiros *bap* para cada componente de frequência, é feita a quantização do valor da mantissa correspondente.

Assim, o número de níveis de quantização pode variar de 0 até 65536, o que corresponde a uma precisão de 0 a 16 bits na representação do valor de cada mantissa. A forma de quantização pode ser feita de maneira simétrica ou assimétrica de acordo com o número de níveis de quantização especificado para cada mantissa.

VI.9. QUANTIZAÇÃO ASSIMÉTRICA

A quantização assimétrica é considerada quando os valores do ponteiro *bap* estão no intervalo $6 \leq bap \leq 15$, e denomina-se assimétrica devido ao fato de que o número de níveis de quantização para valores positivos e negativos de cada mantissa não é o mesmo. De acordo com isso, os valores quantizados de cada mantissa vão a variar de: $(1.0 - 2^{qntzizb[bap]})$ a -1.0 , sendo os valores negativos quantizados com um nível a mais do que os valores positivos.

VI.10. QUANTIZAÇÃO SIMÉTRICA

A quantização simétrica é considerada quando os valores do ponteiro *bap* estão no intervalo $1 \leq bap \leq 5$.

Esse tipo de quantização é caracterizado por considerar o mesmo número de níveis de quantização para os valores positivos e negativos das mantissas.

Finalmente, a fim de salvar alguns bits, o sistema faz a codificação dos valores designados para cada mantissa nos casos em que o *bap* é igual a 1, 2 ou 4. No caso em que o *bap* é igual a um (três níveis de quantização), três mantissas consecutivas são agrupadas em um código de 5 bits de comprimento. Para o caso onde *bap* é igual a dois, três mantissas consecutivas são agrupadas em códigos com 7 bits de comprimento. Finalmente, no caso em que o *bap* = 4, duas mantissas consecutivas são agrupadas num código com 7 bits de comprimento.

VII. NORMALIZAÇÃO DO SISTEMA DE VOLUME

Continuamente nos sistemas de televisão convencional enfrentamos o problema da variação do nível (volume) de áudio quando mudamos de canal ou quando assistimos diferentes programas de um mesmo canal. Isso é devido ao fato de que alguns canais transmitem o áudio em níveis mais baixos ou mais altos do que outros, de tal forma que o usuário deve variar repetidamente o volume do seu receptor a fim de escutar os canais em um nível uniforme de acordo com a sua escolha. Geralmente o ajuste do nível de volume do receptor é feito tomando-se como referência o nível de diálogo do programa que se está assistindo.

O sistema AC-3 oferece uma solução para esse problema introduzindo o sub-campo *dialnorm* de 5 bits dentro do campo BSI o qual indica o número de dB's com que foi codificado o nível de diálogo em relação ao nível máximo de codificação que é 0 dB (100%). Por exemplo, se num determinado momento o *dialnorm* apresenta um valor igual a 19, isso indica que o nível de diálogo foi codificado com -19 dB (19 dB abaixo do valor máximo que é 0 dB). Os valores *dialnorm* se encontram na faixa de 1 a 31, sendo o valor zero reservado pelo sistema AC-3. Com essa informação o decodificador pode dar um nível uniforme de áudio para todos os canais tomando como referência o nível de diálogo com que foi transmitido.

O valor *dialnorm* é utilizado pelo sistema de controle de volume do receptor, a fim de determinar o nível de reprodução do áudio de acordo com o nível de volume ajustado pelo ouvinte. Em seguida mostramos um exemplo para esclarecer mais o procedimento de normalização: suponhamos que um determinado ouvinte ajusta o volume do seu receptor no valor de 61 dB. Nesse instante ele começa a assistir a um programa cujo áudio está sendo transmitido com um nível de diálogo de -25 dB que corresponde a um valor *dialnorm* igual a "11001". Em seguida, o sistema ajusta o nível de reprodução do áudio a 86 dB (61 dB + 25 dB), de forma que se possa compensar os -25 dB com que foi codificado o sinal original.

No caso de se ter um serviço principal de áudio com serviços associados, o receptor modifica o *dialnorm* de cada um deles antes de fazer uma mistura dos mesmos.

VIII. SERVIÇOS DE ÁUDIO OFERECIDOS PELO SISTEMA AC-3

O sistema AC-3 classifica a reprodução de formas distintas de sons na forma de serviços. Isso tem por objetivo oferecer ao usuário maiores facilidades e vantagens na reprodução do áudio que está sendo decodificado. Os serviços oferecidos pelo sistema são classificados em dois serviços principais e seis serviços associados. Cada serviço associado é sempre reproduzido junto com algum serviço principal.

Os serviços principais se classificam em dois grupos: o serviço normal (CM) que contém todos os tipos de sons que estão dentro de um programa de TV (transmitido pelos sistemas de TV convencionais) e o serviço de música e efeitos especiais que produz todos os tipos de som do programa, com exceção do som de diálogo. Ambos os serviços podem utilizar os 6 canais de áudio e serem transmitidos a uma taxa de 320 até 384 Kbit/s.

Os serviços associados, por outro lado, são serviços adicionais que oferecem algumas facilidades para pessoas com problemas na vista e com problemas na audição. Além disso, oferecem alguns serviços de informação importantes para o ouvinte. Existem seis tipos de serviços associados: serviço para pessoas com visão diminuída (VI) que utiliza um único canal de áudio, o serviço para pessoas com audição deficiente (HI) que também utiliza um único canal de áudio; o serviço de diálogo (D) o qual provê o diálogo que vai ser misturado com o serviço principal CM para a formação de um serviço completo de áudio e para facilitar a transmissão multi-linguagem; o serviço de comentário (C) que é similar ao serviço de diálogo e oferece comentários adicionais sobre o programa atual que está sendo decodificado; o serviço de emergência (E) que é um tipo de serviço orientado à transmissão de avisos de emergência, pois tanto o sistema de transporte quanto o receptor dão prioridade à transmissão e recepção desse tipo de serviço; o serviço *voice-over* (VO) é um tipo de serviço similar ao serviço de emergência. A diferença está no fato de que os serviços atuais não são apagados quando a reprodução é feita no receptor.

Cada serviço de áudio é codificado através de uma seqüência de *frames* que constitui uma *elementary bit stream*, os quais posteriormente são multiplexados pelo sistema de transporte junto com o sinal de vídeo para serem transmitidos dentro de um determinado programa de TV.

IX. CONCLUSÕES

Através do uso da codificação perceptual adaptativa, chegou-se finalmente a resultados ótimos em relação à taxa de transmissão versus a qualidade do sinal resultante. A taxa de bits alcançada na transmissão dos seis canais num serviço normal de áudio é de 384 Kbit/s, onde o nível de compressão alcançado encontra-se na relação de 13:1.

Durante todo o processo de codificação, tem-se levado em conta a diminuição da taxa de bits

conservando-se a qualidade do sinal resultante. Dessa forma, desde a codificação dos expoentes até a quantização das mantissas, foram utilizadas técnicas de codificação definidas com base nas características de resposta em frequência do sistema de audição humano.

Por outro lado, além de se obter um sinal de alta qualidade no receptor, a utilização de um sistema de normalização e compressão da faixa do nível de volume permite solucionar um problema apresentado na televisão convencional, onde diversos programas de um mesmo canal ou de canais diferentes são reproduzidos com diferentes níveis de volume. A normalização permite a reprodução de qualquer programa de televisão codificado através do sistema AC-3 mediante um som de nível constante, inclusive quando é feita uma mudança de canal.

A compatibilidade do sistema Dolby AC-3 com sistemas de áudio estéreo convencionais é assegurada mediante a detecção de dois canais de áudio L, R, codificados em cada *frame* de sincronização.

O sistema Dolby AC-3 vem sendo utilizado para o processamento de áudio de alguns filmes de cinema desde o ano de 1992. A alta qualidade de som obtido para esses casos, bem como a taxa de compressão alcançada, permitiram a consideração do sistema AC-3 dentro do sistema HDTV proposto pela Grande Aliança nos Estados Unidos.

X. 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 Telecomunicações (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] United States Advanced Television Systems Committee, “Digital Audio Compression (AC-3) Standard”, Doc A/52, 10 Nov. 94 -12 Apr 95.
- [2] Grant A. Davidson, Lois D. Fielder, and Brian D. Link, “Parametric Bit Allocation in a Perceptual

Audio Coder”, AES 97th Convention, November, 1994.

- [3] Craig C. Todd, Grant A. Davidson, and Mark F. Davis, “AC-3: Flexible Perceptual Coding for Audio Transmission and Storage”, AES 96th Convention, February-March, 1994.
- [4] Guillermo Leopoldo Kemper Vásquez, “Aspectos Relevantes para a Definição de um Sistema de Televisão HDTV”, Campinas: FEEC, UNICAMP, Outubro 1996, Mestrado (Comunicações) Faculdade de Engenharia Elétrica, Universidade Estadual de Campinas.
- [5] United States Advanced Television Systems Committee, “Guide to the Use of the Digital Television Standard For HDTV Transmission”, Doc. A/54 12 April 1995.

Guillermo Leopoldo Kemper Vásques nasceu em Trujillo-Peru em 1971. Recebeu o grau de Engenheiro Eletrônico na Universidade Particular Antenor Orrego de Trujillo (UPAO) 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 de áudio orientado a sistemas de HDTV.

Yuzo Iano recebeu os títulos de Engenheiro Eletrônico, Mestre e Doutor em Engenharia Elétrica através da UNICAMP. Atualmente é Professor Adjunto do Departamento de Comunicações da Faculdade de Engenharia Elétrica e de Computação da UNICAMP. É o responsável pelo Laboratório de Comunicações Visuais do DECOM. Tem interesses em processamento digital de sinais de áudio e vídeo. Trabalha com pesquisas em televisão e em especial com a HDTV.

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 (FEEC/UNICAMP) em 1996. Atualmente está no programa de Doutorado da FEEC-UNICAMP, onde continua suas pesquisas na área de processamento digital de sinais, sinais da fala e sinais

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