



República Federativa do Brasil  
Ministério da Economia  
Instituto Nacional da Propriedade Industrial

**(11) BR 112013010754-5 B1**



**(22) Data do Depósito:** 26/10/2011

**(45) Data de Concessão:** 15/06/2021

---

**(54) Título:** ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDENS ELEVADAS, MÉTODO PARA CODIFICAR E DISPOR DADOS PARA UMA ESTRUTURA DE DADOS, MÉTODO PARA APRESENTAÇÃO DE ÁUDIO E APARELHO PARA APRESENTAÇÃO DE ÁUDIO

**(51) Int.Cl.:** G10L 19/008; H04S 3/00.

**(30) Prioridade Unionista:** 05/11/2010 EP 10306211.3.

**(73) Titular(es):** DOLBY INTERNATIONAL AB.

**(72) Inventor(es):** FLORIAN KEILER; SVEN KORDON; JOHANNES BOEHM; HOLGER KROPP; JOHANN-MARKUS BATKE.

**(86) Pedido PCT:** PCT EP2011068782 de 26/10/2011

**(87) Publicação PCT:** WO 2012/059385 de 10/05/2012

**(85) Data do Início da Fase Nacional:** 30/04/2013

**(57) Resumo:** ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDENS ELEVADAS A invenção está relacionada com uma estrutura de dados para dados de áudio HOA Ambisonics de Ordens Elevadas, estrutura de dados esta que inclui dados de conteúdo de áudio espacial 2D ou 3D para uma ou mais diferentes descrições de fluxo de dados de áudio HOA. Os dados de áudio HOA podem possuir uma ordem maior do que "3", e a estrutura de dados em adição pode incluir dados de fonte única de sinal de áudio e/ou dados de áudio de conjunto de microfones a partir de posições espaciais fixas ou variando com o tempo.

Relatório Descritivo da Patente de Invenção para  
**"ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS  
DE ORDENS ELEVADAS, MÉTODO PARA CODIFICAR E DISPOR  
DADOS PARA UMA ESTRUTURA DE DADOS, MÉTODO PARA  
APRESENTAÇÃO DE ÁUDIO E APARELHO PARA  
APRESENTAÇÃO DE ÁUDIO".**

[0001] A invenção se relaciona com uma estrutura de dados para dados de áudio Ambisonics de Ordens Elevadas, a qual inclui dados de conteúdo de áudio espacial 2D e/ou 3D e que também é adequada para dados de áudio HOA possuindo uma ordem de mais do que "3".

Antecedentes

[0002] O Áudio 3D pode ser realizado utilizando uma descrição de campo sonoro por uma técnica chamada Ambisonics de Ordens Elevadas (HOA) como descrita abaixo. Armazenar dados HOA requer algumas convenções e condições de como estes dados devem ser utilizados por um decodificador especial para estarem aptos a criar sinais de alto-falante para reprodução em uma dada configuração do alto-falante de reprodução. O formato de armazenamento não existente define todas estas condições para o HOA. O formato B (baseado na estrutura "Riff/wav" extensível) com sua realização de formato de arquivo \*.amb como descrito, por exemplo, em "File Format for B-Format" de 30 de março de 2009, por Martin Leese, <http://www.ambisonia.com/Members/etienne/Members/mleese/file-format-for-b-format>, é o formato mais sofisticado disponível atualmente.

[0003] De 16 de julho de 2010, uma vista geral dos formatos de arquivo existentes é revelada no Ambisonics Xchange Site: "Existing formats", <http://ambisonics.iem.at/xchange/format/existing-formats>, e uma proposta para um formato de troca Ambisonics também é revelada neste site: "A first proposal to specify, define and determine

the parameters for an Ambisonics exchange format", <http://ambisonics.iem.at/xchange/format/a-first-proposal-for-the-format>.

### Invenção

[0004] Com respeito aos sinais HOA, para 3D, uma coleção de  $M = (N + 1)^2 (2N + 1)$  para 2D diferente de objetos de Áudio a partir de diferentes fontes de som, todas na mesma frequência, podem ser gravados (codificados) e reproduzidos como diferentes objetos de som posto que eles são espacialmente igualmente distribuídos. Isto significa que um sinal Ambisonics de primeira ordem pode transportar quatro objetos de áudio 3D ou três 2D e estes objetos precisam ser separados de forma uniforme ao redor de uma esfera para 3D ou ao redor de um círculo em 2D. A sobreposição espacial e mais do que  $M$  sinais na gravação irá resultar em indistinção – somente os sinais mais altos podem ser reproduzidos como objetos coerentes, os outros sinais difusos irão degenerar um pouco os sinais coerentes dependendo da sobreposição no espaço, da frequência e da similaridade de sonoridade.

[0005] Com respeito à situação acústica em um cinema, a precisão de localização do som espacial alto é requerida para a área frontal da tela de modo a combinar com a cena visual. A percepção dos objetos de som circundante é menos crítica (reverberação, objetos de som sem conexão com a cena visual). Aqui, a densidade dos alto-falantes pode ser menor comparada com a área frontal.

[0006] A ordem HOA dos dados HOA, relevante para a área frontal, precisa ser grande para permitir a reprodução holofônica à vontade. Uma típica ordem é  $N = 10$ . Isto requer  $(N + 1)^2 = 121$  coeficientes HOA. Na teoria, também poderíamos codificar  $M = 121$  objetos de áudio, se estes objetos de áudio fossem igualmente espacialmente distribuídos. Porém, no nosso cenário, eles estão constrictos à área frontal (devido a aqui somente precisarmos de altas

ordens). Na verdade, somente podemos codificar ao redor de  $M = 60$  objetos de Áudio sem som confuso (a área frontal é no máximo metade de uma esfera de direções, assim,  $M/2$ ). Com respeito ao formato B mencionado acima, ele somente permite uma descrição até uma ordem de Ambisonics de 3, e o tamanho do arquivo é restrito a 4 GB. Outros itens de informação especial estão ausentes, como o tipo de onda ou o raio de referência de decodificação que são vitais para os decodificadores modernos. Não é possível utilizar formatos de amostra diferentes (larguras de palavra) e larguras de bandas para diferentes componentes (canais) Ambisonics. Também não existe padronização para armazenar informação adicional e metadados para Ambisonics.

[0007] Na técnica conhecida, gravar sinais Ambisonics utilizando um conjunto de microfones é restrito às ordens de um. Isto pode alterar no futuro se protótipos experimentais de microfones HOA forem desenvolvidos. Para a criação de conteúdo 3D, uma descrição do campo sonoro ambiente poderia ser gravada utilizando um conjunto de microfones em Ambisonics de primeira ordem, por meio do que as fontes direcionais são capturadas utilizando microfones modo de aproximação ou microfones altamente direcionais juntos com a informação direcional (isto é, a posição da fonte). Os sinais direcionais podem então ser codificados em uma descrição HOA, ou isto pode ser executado por um decodificador sofisticado. De qualquer forma, um novo formato de arquivo Ambisonics precisa estar apto a armazenar mais do que uma descrição de campo de som ao mesmo tempo, mas parece que nenhum formato existente pode encapsular mais do que uma descrição Ambisonics.

[0008] Um problema a ser resolvido pela invenção é proporcionar um formato de arquivo Ambisonics que seja capaz de armazenar duas ou mais descrições de campo de som ao mesmo tempo, onde a ordem

de Ambisonics pode ser maior do que 3. Este problema é solucionado pela estrutura de dados revelada na reivindicação 1 e pelo método revelado na reivindicação 12.

[0009] Para recriar Áudio 3D realístico, decodificadores Ambisonics da próxima geração irão exigir várias convenções e condições juntas com os dados armazenados a serem processados, ou um único formato de arquivo onde todos os parâmetros e elementos de dados relacionados podem ser coerentemente armazenados.

[00010] O formato de arquivo da invenção para conteúdo de som espacial pode armazenar um ou mais sinais HOA e/ou sinais modo direcionais juntos com a informação direcional, onde ordens Ambisonics maiores do que 3 e arquivos > 4 GB são viáveis. Adicionalmente, o formato de arquivo da invenção proporciona elementos adicionais que os formatos existentes não oferecem:

1) Informação vital requerida para decodificados HOA da próxima geração é armazenada dentro do formato de arquivo:

- Informação de onda Ambisonics (plano, esférica, tipos de mistura), região de interesse (fontes fora ou dentro da área de escuta), e raio de referência (para decodificação de ondas esféricas)

- Sinais mono direcionais relacionados podem ser armazenados. Informação de posição destes sinais direcionais pode ser descrita utilizando a informação de ângulo ou de distância ou um vetor de codificação de coeficientes Ambisonics.

2) Todos os parâmetros definindo os dados Ambisonics estão contidos na informação adicional, para garantir clareza sobre a gravação:

- Dimensionamento e normalização de Ambisonics (SN3D, N3D, Furse Malham, Formato B, ..., definido pelo usuário), informação de ordem misturada.

3) O formato de armazenamento de dados Ambisonics é estendido para permitir um armazenamento flexível e econômico de dados:

- O formato da invenção permite armazenar dados relacionados com a ordem Ambisonics (canais Ambisonics) com resolução de tamanho de PCM – palavra bem como utilizando largura de banda restrita.

4) Meta campos permitem armazenar informação acompanhante sobre o arquivo tal como informação de gravação para sinais do microfone:

- Sistema de coordenadas de referência de gravação, microfone, posições do microfone, da fonte e de ouvintes virtuais, características direcionais do microfone, informação sobre ambiente e fonte.

[00011] Este formato de arquivo para conteúdo de áudio 2D e 3D cobre o armazenamento tanto das descrições de Ambisonics de Ordens Elevadas (HOA) bem como fontes únicas com posições fixas ou variando com o tempo, e contém todas as informações permitindo aos decodificadores de áudio de próxima geração proporcionar Áudio 3D realístico.

[00012] Utilizando parâmetros apropriados, o formato de arquivo da invenção também é adequado para reprodução contínua de conteúdo de áudio. Assim, informação adicional dependente de conteúdo (dados de cabeçalho) pode ser enviada em instâncias de tempo como selecionadas pelo criador do arquivo. O formato de arquivo da invenção também serve como descrição de cena onde trilhas de uma cena de áudio podem iniciar e terminar a qualquer momento.

[00013] Em princípio, a estrutura de dados da invenção é adequada para dados de áudio Ambisonics de Ordens Elevadas HOA, estrutura de dados esta que inclui dados de conteúdo espacial 2D e/ou 3D para

uma ou mais diferentes descrições de fluxo de dados de áudio HOA, e estrutura de dados esta que também é adequada para dados de áudio HOA que possuem uma ordem de mais do que "3", e estrutura de dados esta que em adição pode incluir dados de fonte única de sinal de áudio e/ou dados de áudio de conjunto de microfones a partir de posições espaciais fixas ou variando com o tempo.

[00014] Em princípio, o método da invenção é adequado para apresentação de áudio, onde um fluxo de dados de áudio HOA contendo pelo menos dois sinais de dados de áudio HOA diferentes é recebido e pelo um primeiro dos mesmos é utilizado para apresentação com uma disposição densa de alto-falantes localizada em uma área distinta de um local de apresentação, e pelo menos um segundo e diferente dos mesmos é utilizado para apresentação com uma disposição de alto-falantes menos densa circundando o dito local de apresentação.

[00015] Concretizações vantajosas adicionais são reveladas nas respectivas reivindicações dependentes.

#### Desenhos

[00016] Concretizações ilustrativas da invenção são descritas com referência aos desenhos acompanhantes, nos quais:

[00017] A Fig. 1 é uma reprodução holofônica no cinema com disposições densas de alto-falantes na região frontal e densidade esparsa de alto-falantes circundando a área de escuta;

[00018] A Fig. 2 é um sistema de decodificação sofisticado;

[00019] A Fig. 3 é a criação de conteúdo HOA a partir da gravação do conjunto de microfones, da gravação de fonte única e da geração de campo de som simples e complexo;

[00020] A Fig. 4 é a criação de conteúdo imersível de próxima geração;

[00021] A Fig. 5 é a decodificação 2D de sinais HOA para a

configuração simples de alto-falantes de efeito surround, e a decodificação de sinais HOA para uma configuração de alto-falantes holofônicos para o estágio frontal e uma configuração mais esparsa de alto-falantes de efeito surround 3D;

[00022] A Fig. 6 é um problema de domínio interior, onde as fontes estão fora da região de interesse / validade;

[00023] A Fig. 7 é a definição de coordenadas esféricas;

[00024] A Fig. 8 é o problema de domínio exterior, onde as fontes estão dentro da região de interesse / validade;

[00025] A Fig. 9 é um exemplo simples do formato de arquivo HOA;

[00026] A Fig. 10 é um exemplo para um arquivo HOA contendo múltiplos quadros com múltiplas trilhas;

[00027] A Fig. 11 é um arquivo HOA com múltiplos MetaDataChunks;

[00028] A Fig. 12 é o processamento de codificação de TrackRegion;

[00029] A Fig. 13 é o processamento de decodificação de TrackRegion;

[00030] A Fig. 14 é a implementação da redução de largura de banda utilizando o processamento MDCT;

[00031] A Fig. 15 é a implementação da Reconstrução de Largura de Banda utilizando o processamento MDCT.

#### Concretizações Ilustrativas

[00032] Com a crescente expansão de Vídeo 3D, tecnologias de áudio imersível estão se tornando um aspecto interessante para diferenciação. Ambisonics de Ordens Elevadas (HOA) é uma destas tecnologias que pode proporcionar um meio para introduzir Áudio 3D de um modo progressivo em cinemas. Utilizando trilhas de som HOA e decodificadores HOA, um cinema pode começar com as configurações de áudio de alto-falantes de efeito surround existentes e investir em



mais alto-falantes gradualmente, melhorando a experiência imersível com cada etapa.

[00033] A Fig. 1a apresenta a reprodução holofônica em cinema com a disposição densa de alto-falantes 11 na região frontal e a densidade esparsa de alto-falantes 12 em volta da área de escuta ou de assentos 10, proporcionando um meio para reprodução precisa de sons relacionados com a ação visual e de precisão suficiente dos sons ambientes reproduzidos. A Fig. 1b apresenta a direção de chegada percebida das ondas sonoras frontais reproduzidas, onde a direção de chegada de ondas planas combina com diferentes posições da tela, isto é, ondas planas são adequadas para reproduzir profundidade. A Fig. 1c apresenta a direção de chegada percebida de ondas esféricas reproduzidas, o que leva a melhor consistência de direção percebida do som e da ação visual 3D ao redor da tela.

[00034] A necessidade de dois fluxos HOA diferentes é causada pelo fato de que a ação visual principal em um cinema acontece na região frontal dos ouvintes. Além disso, a precisão de percepção de detecção da direção de um som é maior para as fontes de som frontais do que para as fontes ao redor. Portanto, a precisão da reprodução de som espacial frontal precisa ser maior do que a precisão espacial para sons ambientes reproduzidos. Dispositivos holofônicos para reprodução de som, um alto número de alto-falantes, um decodificador dedicado e acionadores de alto-falante relacionados são requeridos para a região frontal da tela, enquanto tecnologia menos onerosa é necessária para a reprodução de som ambiente (menor densidade de alto-falantes rodeando a área de audição e tecnologia de decodificação menos perfeita).

[00035] Devido às tecnologias de criação de conteúdo e de reprodução de som, é vantajoso fornecer uma representação HOA para os sons ambientes e uma representação HOA para os sons de

ação de primeiro plano, conforme a Fig. 4. Um cinema utilizando uma configuração simples com um equipamento de som simples de reprodução esparsa pode misturar ambos os fluxos antes da decodificação (conforme Fig. 5, parte superior). Um cinema mais sofisticado equipado com dispositivos de reprodução totalmente imersível pode utilizar dois decodificadores – um para decodificar os sons ambientes e um decodificador especializado para posicionamento de alta precisão de fontes virtuais de som para ação principal em primeiro plano, como apresentado no sistema sofisticado de decodificação na Fig. 2 e na parte de baixo da Fig. 5. Um arquivo HOA especial contém pelo menos duas trilhas que representam campos de som HOA para sons ambientes  $A_n^m(t)$  e para sons frontais relacionados com a ação visual principal  $C_n^m(t)$ . Fluxos opcionais para efeitos direcionais podem ser proporcionados. Dois sistemas decodificadores correspondentes juntos com um posicionador proporcionam sinais para um sistema denso de alto-falantes holofônicos 3D frontal 21 e para o sistema de efeito surround 3D menos denso (isto é, esparsa) 22. O sinal de dados HOA do fluxo da Trilha 1 representa os sons ambiente e é convertido em um conversor HOA 231 para entrada em um Decodificador 1 232 especializado para reprodução de ambiente. Para o fluxo de dados da Trilha 2, os dados de sinal HOA (sons frontais relacionados com a cena visual) é convertido em um conversor HOA 241 para entrada para um filtro de distância corrigida (Eq. (26)) 242 para melhor colocação de fontes de som esféricas ao redor da área da tela com um Decodificador 2 dedicado 243. Os fluxos de dados direcionais são diretamente posicionados para os L alto-falantes. Os três sinais de alto-falante são misturados PCM para reprodução conjunta com o sistema de alto-falantes 3D.

[00036] Parece que não existe formato de arquivo dedicado para tal

cenário. Gravações de campo sonoro 3D utilizam descrições completas de cena com trilhas de som relacionadas, ou uma única descrição de campo sonoro quando armazenando para posterior reprodução. Exemplos para o primeiro tipo são os formatos WFS (Síntese de Campo de Ondas) e vários formatos container. Os exemplos para o segundo tipo são os formatos Ambisonics como o formato B ou formatos AMB, de acordo com o artigo mencionados acima "File Format for B-Format". O último restringe as ordens Ambisonics a três, um formato de transmissão fixo, um modelo de decodificador fixo e a campos de som único.

#### Criação e Reprodução de Conteúdo HOA

[00037] O processamento para gerar descrições de campo de som HOA é representado na Fig. 3. Na Fig. 3a, gravações naturais de campos de som são criadas por utilizar conjuntos de microfones. Os sinais de cápsula são colocados em matriz e equalizados de modo a formar sinais HOA. Sinais de Ordens Elevadas (Ambisonics de ordem  $> 1$ ) normalmente são filtrados com filtro de passa baixa para reduzir artefatos devido aos efeitos de distância de cápsula: filtrados com filtro de passa baixa para reduzir interferência espacial em altas frequências, e filtrados em filtro de passa alta para reduzir níveis excessivos de baixa frequência com ordem Ambisonics crescente  $n$  ( $h_n(kr_{d_{mic}})$ ), veja Eq. (34). Opcionalmente, a filtragem de codificação de distância pode ser aplicada, veja as Eqs. (25) e (27). Antes do armazenamento, a informação de formato HOA é adicionada para o cabeçalho da trilha.

[00038] Representações de campo de som artístico normalmente são criadas utilizando múltiplos fluxos direcionais de fonte única. Como apresentado na Fig. 3b, um sinal de fonte única pode ser capturado como uma gravação PCM. Isto pode ser feito por microfones de aproximação ou por utilizar microfones com alta direcionalidade. Em

adição, os parâmetros direcionais ( $r_s, \theta_s, \phi_s$ ) da fonte de som em relação a uma posição virtual de melhor escuta são gravados (sistema de coordenadas HOA, ou qualquer ponto de referência para mapeamento posterior). A informação de distância também pode ser criada por artisticamente colocar os sons quando sintetizando cenas para filmes. Como apresentado na Fig. 3c, a informação direcional ( $\theta_s, \phi_s$ ) é então utilizada para criar o vetor de codificação  $\Psi$ , e o sinal de fonte direcional é codificado em um sinal Ambisonics, veja a Eq. (18). Isto é equivalente a uma representação de onda plana. Um processo de filtragem final pode utilizar a informação de distância  $r_s$  para imprimir uma característica de fonte esférica no sinal Ambisonics (Eq. (19)), ou para aplicar filtragem de codificação de distância, Eqs. (25), (27). Antes do armazenamento, a informação de formato HOA é adicionada para o cabeçalho da trilha.

[00039] Descrições mais complexas de campo de onda são geradas por sinais Ambisonics de mistura HOA como representados na Fig. 3d. Antes do armazenamento, a informação de formato HOA é adicionada para o cabeçalho da trilha.

[00040] O processo de geração de conteúdo para cinema 3D é representado na Fig. 4. Sons frontais relacionados com a ação visual são codificados com alta precisão espacial e misturados com um sinal HOA (campo de onda)  $C_n^m(t)$  e armazenados como a Trilha 2. Os codificadores envolvidos codificam com uma alta precisão espacial e tipos de onda especiais necessárias para melhor associar com a cena visual. A Trilha 1 contém o campo sonoro  $A_n^m(t)$  que está relacionado com os sons ambientes codificados sem restrição de direção de fonte. Normalmente, a precisão espacial dos sons ambiente não precisa ser tão alta quanto para os sons frontais (por consequência, a ordem Ambisonics pode ser menor) e a modelagem do tipo de onda é menos crítico. O campo de som ambiente também pode incluir partes

reverberantes dos sinais de som frontais. Ambas as trilhas são multiplexadas para armazenamento e/ou troca. Opcionalmente, os sons direcionais (por exemplo, Trilha 3) podem ser multiplexados para o arquivo. Estes sons podem sons de efeitos especiais, diálogos ou informação esportiva como uma fala narrativa para deficientes visuais.

[00041] A FIG. 5 apresenta os princípios de decodificação. Como representado na parte superior, um cinema com configuração esparsa de alto-falantes pode misturar ambos os sinais HOA a partir da Trilha 1 e da Trilha 2 antes da decodificação HOA simplificada, e pode truncar a ordem da Trilha 2 e reduzir a dimensão de ambas as trilhas para 2D. NO caso de um fluxo direcional estar presente, ele é codificado para HOA 2D. Então, todos os três fluxos são misturados para formar uma única representação HOA que é então decodificada e reproduzida. A parte de baixo corresponde à Fig. 2. Um cinema equipado com um sistema holofônico para o estágio frontal e um sistema de efeito surround 3D esparsa irá utilizar decodificadores sofisticados dedicados e misturar as alimentações para o alto-falante. Para o fluxo de dados da Trilha 1, os dados HOA representando os sons ambiente são convertidos para o Decodificador 1 especializado para reprodução do ambiente. Para o fluxo de dados da Trilha 2, HOA (sons frontais relacionados com a cena visual) é convertido e a distância corrigida (Eq. (26)) para melhor colocação de fontes de som esféricas ao redor da área da tela com um Decodificador 2 dedicado. Os fluxos de dados direcionais são diretamente posicionados para os alto-falantes L. Os três sinais de alto-falante são misturados PCM para reprodução conjunta com o sistema de alto-falantes 3D.

Descrições de campo de som utilizando Ambisonics de Ordens Elevadas

Descrição de campo de som utilizando Harmônicas Esféricas (SH)

[00042] Quando utilizando descrições de Harmônica / Bessel

esférica, a solução da equação de onda acústica é proporcionada na Eq. (1), de acordo com M.A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics", Journal of Audio Engineering Society, 53(11), pp.1004–1025, Novembro, 2005, e Earl G. Williams, "Fourier Acoustics", Academic Press, 1999. A pressão do som é uma função das coordenadas esféricas  $r, \Theta, \Phi$  (veja a Fig. 7 para

sua definição) e da frequência espacial  $k = \frac{\omega}{c} = \frac{2\pi f}{c}$ .

[00043] A descrição é válida para fontes de som de áudio fora da região de interesse ou de validade (problema de domínio interior, como apresentado na Fig. 6) e assume as Harmônicas Esféricas Ortogonais – Normalizadas:

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n A_n^m(k) j_n(kr) Y_n^m(\theta, \phi) \quad (1)$$

[00044] Os  $A_n^m(k)$  são chamados Coeficientes Ambisonics,  $j_n(kr)$  é a função esférica de Bessel do primeiro tipo,  $Y_n^m(\theta, \phi)$  são chamados de Harmônicos Esféricos (SH),  $n$  é o índice da ordem Ambisonics, e  $m$  indica o grau.

[00045] Devido à natureza da função de Bessel que possui valores significativos somente para valores  $Kr$  pequenos (distâncias pequenas a partir da origem ou baixas frequências), a série pode ser Prada na mesma ordem  $n$  e restrita a um valor  $N$  com precisão suficiente. Quando armazenando dados HOA, normalmente, os coeficientes Ambisonics  $A_n^m, B_n^m$  ou alguns derivados (detalhes são descritos abaixo) são armazenados até esta ordem  $N$ .  $N$  é chamada de ordem Ambisonics.  $N$  é chamado ordem Ambisonics, e o termo "ordem" também normalmente é utilizado em combinação com  $n$  nas funções de Bessel  $j_n(kr)$  e de Hankel  $h_n(kr)$ .

[00046] A solução das equações de onda para o caso exterior, onde as fontes se situam dentro de uma região de interesse ou de validade como representado na Fig. 8, é expressa para  $r > r_{Source}$  na Eq. (2):

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n B_n^m(k) h_n^{(1)}(kr) Y_n^m(\theta, \phi) \quad (2)$$

[00047] Os  $B_n^m(k)$  novamente são chamados de coeficientes Ambisonics e  $h_n^{(1)}(kr)$  denota a função esférica de Henkel do primeiro tipo e da n-ésima ordem. A fórmula assume SH ortogonal-normalizada. Observação: Geralmente, a função esférica de Hankel de primeiro tipo  $h_n^{(1)}$  é utilizada para descrever ondas de saída (relacionadas com  $e^{ikr}$ ) para frequências positivas e a função esférica de Hankel do segundo tipo  $h_n^{(2)}$  é utilizada para ondas entrantes (relacionadas com  $e^{-ikr}$ ), de acordo com o livro "Fourier Acoustics" mencionado acima.

### Harmônicas Esféricas

[00048] As harmônicas esféricas  $Y_n^m$  podem ser com valores complexos ou reais. O caso geral para HOA utiliza harmônicas esféricas com valor real. Uma descrição unificada de Ambisonics utilizando harmônicas esféricas com valores reais e complexos pode ser vista em Mark Poletti, "Unified description of Ambisonics using real and complex spherical harmonics", Proceedings of the Ambisonics Symposium 2009, Gras, Austria, Junho, 2009.

[00049] Existem diferentes modos de normalizar as harmônicas esféricas (as quais são independentes das harmônicas esféricas sendo reais ou complexas), de acordo com as páginas da Rede seguintes com respeito às harmônicas esféricas (reais), e aos esquemas de normalização: <http://www.ipgp.fr/~wiecsor/SHTOOLS/www/conventions.html>, [http://en.citiseum.org/wiki/Spherical\\_harmonics](http://en.citiseum.org/wiki/Spherical_harmonics). A normalização corresponde à relação de ortogonalidade entre  $Y_n^m$  e  $Y_n^{m'}$ .

[00050] Observação:

$$\int_{S_2} Y_n^m(\Omega) Y_n^{m'}(\Omega)^* d\Omega = \frac{N_{n,m}}{\sqrt{\frac{(2n+1)(n-|m|)!}{4\pi(n+|m|)!}}} \frac{N_{n',m'}}{\sqrt{\frac{(2n'+1)(n'-|m'|)!}{4\pi(n'+|m'|)!}}} \delta_{nn'} \delta_{mm'}$$

[00051] onde  $s^2$  é a esfera unitária e o delta de Kroneker  $\delta_{aa'}$  é igual a 1 para  $a = a'$ , senão 0.

[00052] Harmônicas esféricas complexas são descritas por:

$$Y_n^m(\theta, \phi) = s_m \Theta_n^m(\theta) e^{im\phi} = s_m N_{n,m} P_{n,|m|}(\cos(\theta)) e^{im\phi} \quad (3)$$

[00053] onde  $i = \sqrt{-1}$  e  $s_m = \begin{cases} (-1)^m & m > 0 \\ 1 & \text{else} \end{cases}$  para um sinal alternativo para  $m$  positivo como no livro "Fourier Acoustic" mencionado acima. (Observação:  $s_m$  é um termo de convenção e pode ser omitido para SH somente positivo).  $N_{n,m}$  é um termo de normalização que assume a forma para uma representação ortogonal – normalizada (! denota fatorial):

$$N_{n,m} = \sqrt{\frac{(2n+1)(n-|m|)!}{4\pi (n+|m|)!}} \quad (4)$$

[00054] A Tabela 1 abaixo apresenta alguns esquemas de normalização normalmente utilizados para as harmônicas esféricas com valores complexos.  $P_{n,|m|}(x)$  são as funções de Legendre associadas, onde é seguida a notação com  $|m|$  a partir do artigo acima "Unified description of Ambisonics using real and complex spherical harmonics" que evita o termo de fase  $(-1)^m$  chamado de fase de Condon-Shortley, e que algumas vezes é incluído dentro da representação de  $P_n^m$  dentro de outras notações. As funções de Legendre associadas  $P_{n,|m|}: [-1,1] \rightarrow \mathbb{R}, n \geq |m| \geq 0$  podem ser expressas utilizando a fórmula de Rodriguez como:

$$P_{n,|m|}(x) = \frac{1}{2^n n!} (1-x^2)^{\frac{|m|}{2}} \frac{d^{n+|m|}}{dx^{n+|m|}} (x^2-1)^n \quad (5)$$

$N_{n,m}$ , Esquemas comuns de normalização para SH complexa			
Não normalizada	Semi-Normalizada de Schmidt, SN3D	$4\pi$ normalizada, N3D, geodésia $4\pi$	Orto-Normalizada
1	$\sqrt{\frac{(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2n+1)(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2n+1)(n- m )!}{4\pi (n+ m )!}}$



**Tabela 1 – Fatores de Normalização para harmônicas esféricas com valores complexos**

[00055] Numericamente é vantajoso derivar  $P_{n,|m|}(x)$  de uma maneira progressiva a partir de uma relação de recorrência, veja William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, "Numerical Recipes in C", Cambridge University Press, 1992. Aas funções de Legendre associadas até  $n = 4$  são dadas na Tabela 2:

nm	0	1	2	3	4
0	$P_0^0(\cos\theta) = 1$	$P_1^0(\cos\theta) = \cos\theta$	$P_2^0(\cos\theta) = \frac{1}{2}(3\cos^2\theta - 1)$	$P_3^0(\cos\theta) = \frac{1}{2}(5\cos^3\theta - 3\cos\theta)$	$P_4^0(\cos\theta) = \frac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3)$
1		$P_1^1(\cos\theta) = \sin\theta$	$P_2^1(\cos\theta) = 3\cos\theta\sin\theta$	$P_3^1(\cos\theta) = \frac{3}{2}(5\cos^2\theta - 1)\sin\theta$	$P_4^1(\cos\theta) = \frac{5}{2}(7\cos^3\theta - 3\cos\theta)\sin\theta$
2			$P_2^2(\cos\theta) = 3\sin^2\theta$	$P_3^2(\cos\theta) = 15\cos\theta\sin^2\theta$	$P_4^2(\cos\theta) = \frac{15}{2}(7\cos^2\theta - 1)\sin^2\theta$
3				$P_3^3(\cos\theta) = 15\sin^3\theta$	$P_4^3(\cos\theta) = 105\cos\theta\sin^3\theta$
4					$P_4^4(\cos\theta) = 105\sin^4\theta$

**Tabela 2 – Alguns Primeiros Polinomiais de Legendre**

$$P_{n,|m|}(\cos\theta), n=0\dots 4$$

[00056] SH com valor real são derivadas por combinar o conjugado complexo  $Y_n^m$  correspondendo aos valores opostos de  $m$  (o termo  $(-1)^m$  na definição (6) é introduzido para obter expressões não sinalizadas para a SH real, o que é o caso normal na Ambisonics):

$$S_{n,m}(\theta, \phi) = \begin{cases} \sqrt{2} Y_n^m(\theta, \phi) & m > 0 \\ Y_n^0(\theta, \phi) & m = 0 \\ \sqrt{2} Y_n^{-m}(\theta, \phi) & m < 0 \end{cases} \quad (6)$$

[00057] que pode ser reescrito como Eq. (7) para realçar a conexão com harmônicas circulares com  $\Phi_m(\phi) = \Phi_{n=|m|}^m(\phi)$  apenas retendo o termo azimuthal:

$$S_n^m(\theta, \phi) = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \Phi_m(\phi) \quad (7)$$

$$\Phi_{n=|m|}^m(\phi) = \begin{cases} \cos(m\phi), & m > 0 \\ 1, & m = 0 \\ \sin(|m|\phi), & m < 0 \end{cases} \quad (8)$$

[00058] O numero total de componentes esféricas  $S_n^m$  para uma dada ordem Ambisonics  $N$  é igual  $(N+1)^2$ . Esquemas comuns de normalização das harmônicas esféricas com valores reais são dados na Tabela 3.

$\tilde{N}_{n,m}$ , Esquemas Comuns de Normalização para SH real		
Semi-Normalizada de Schmidt, SN3D	$4\pi$ normalizada, N3D, geodésia $4\pi$	Orto-Normalizada
$\sqrt{\frac{(2 - \delta_{0,m})(n -  m )!}{(n +  m )!}}$	$\sqrt{\frac{(2 - \delta_{0,m})(2n + 1)(n -  m )!}{(n +  m )!}}$	$\sqrt{\frac{(2 - \delta_{0,m})(2n + 1)(n -  m )!}{4\pi (n +  m )!}}$

**Tabela 3 – Esquemas de normalização de SH real 3D,  $\delta_{0,m}$  possui um valor de 1 para  $m = 0$  e senão 0**

### Harmônicas Circulares

[00059] Para representações bidimensionais somente um subconjunto de harmônicas é necessário. O grau SH pode somente pegar os valores  $m \in \{-n, n\}$ . O número total de componentes para um dado N se reduz para  $2N+1$  devido aos componentes representando a inclinação  $\theta$  se tornarem obsoletos e as harmônicas esféricas poderem ser substituídas pelas harmônicas circulares dadas na Eq. (8).

[00060] Existem diferentes esquemas de normalização  $N_m$  para harmônicas circulares, os quais precisam ser considerados quando convertendo coeficientes Ambisonics 3D para coeficientes 2D. A fórmula mais geral para harmônicas circulares se torna:

$$\check{\Phi}_{n=|m|}^m(\phi) = N_m \Phi_m(\phi) = \begin{cases} N_m \cos(m\phi), & m > 0 \\ N_m & m = 0 \\ N_m \sin(|m|\phi) & m < 0 \end{cases} \quad (9)$$

[00061] Alguns fatores de normalização comuns para harmônicas circulares são proporcionados na Tabela 4, onde o termo de normalização é introduzido pelo fator antes do termo horizontal  $\phi_m(\phi)$ :

$N_m$ , Esquemas Comuns de Normalização para Harmônicas Circulares			
Não normalizadas	SN2D	Normalizadas 2D, N2D	Orto-Normalizadas
$\sqrt{\frac{2 - \delta_{0,m}}{2}}$	1	$\sqrt{(2 - \delta_{0,m})}$	$\sqrt{\frac{(2 - \delta_{0,m})1}{2\pi}}$

**Tabela 4 – Esquemas de normalização CH 2D,  $\delta_{0,m}$  possui um valor de 1 para  $m = 0$  e senão 0**

[00062] A conversão entre diferentes normalizações é direta. Em geral, a normalização tem um efeito sobre a notação descrevendo a

pressão (de acordo com as Eqs. (1), (2)) e sobre todas as considerações derivadas. O tipo de normalização também influencia os coeficientes Ambisonics. Também existem pesos que podem ser aplicados para dimensionar estes coeficientes, por exemplo, pesos de Furse-Malhan (FuMa) aplicados para coeficientes Ambisonics quando armazenado um arquivo utilizando o formato AMB.

[00063] Com respeito à conversão 2D para 3D, a conversão CH para SH e vice-versa também pode ser aplicada para coeficientes Ambisonics, por exemplo, quando decodificando uma representação Ambisonics 3D (gravação) com um decodificador 2D para uma configuração de alto-falante 2D. A relação entre  $S_n^m$  e  $\check{\Phi}_{n=|m|}^m$  para a conversão 3D para 2D é representada no esquema seguinte até uma ordem de Ambisonics de 4:

				$S_0^0$				
			$S_1^{-1}$	$S_1^0$	$S_1^1$			
		$S_2^{-2}$	$S_2^{-1}$	$S_2^0$	$S_2^1$	$S_2^2$		
	$S_3^{-3}$	$S_3^{-2}$	$S_3^{-1}$	$S_3^0$	$S_3^1$	$S_3^2$	$S_3^3$	
$S_4^{-4}$	$S_4^{-3}$	$S_4^{-2}$	$S_4^{-1}$	$S_4^0$	$S_4^1$	$S_4^2$	$S_4^3$	$S_4^4$

[00064] O fato de conversão 2D para 3D pode ser derivado para o posicionamento horizontal em  $\theta = \frac{\pi}{2}$  como a seguir:

$$\alpha_{\frac{2D}{3D}} = \frac{S_{n=m}^m\left(\theta = \frac{\pi}{2}, \Phi\right)}{\check{\Phi}_{n=|m|}^m(\phi)} = \frac{\tilde{N}_{m,m} (2m)!}{N_m m! 2^m} \quad (10)$$

[00065] A conversão de 3D para 2D utiliza  $\frac{1}{\alpha_{\frac{2D}{3D}}}$ . Detalhes são apresentados em conexão com as Eqs. (28), (29), (30) abaixo. Uma conversão 2D normalizada para normalizado ortogonal se torna:

$$\alpha_{\frac{N2D}{ortho3D}} = \sqrt{\frac{(2m+1)!}{4 \pi m!^2 2^{2m}}} \quad (11)$$

### Coeficientes Ambisonics

[00066] Os coeficientes Ambisonics possuem a escala unitária da

pressão do som:  $1Pa = 1 \frac{N}{m^2} = 1 \frac{kg}{s^2} \frac{m}{m^2}$ . Os coeficientes Ambisonics formam o sinal Ambisonics e em geral são em função de tempo separado. A tabela 5 apresenta a relação entre a representação dimensional, Ambisonics de ordem N e o número de coeficientes (canais) Ambisonics.

Dimensão	Numero de coeficientes Ambisonics			
	Número de Canais Ambisonics			
	N=1	N=2	N=3	N
2D	3	5	7	2 N+1
3D	4	9	16	(N+1) <sup>2</sup>

**Tabela 5 – Número de coeficientes Ambisonics**

[00067] Quando lidando com representações de tempo separado normalmente os coeficientes Ambisonics são armazenados de uma maneira intercalada como representações de canal PCM para gravações com múltiplos canais (canal 1 = coeficiente Ambisonics  $A_n^m$  da amostra  $v$ ). A sequência de coeficientes sendo uma questão de convenção. Um exemplo para 3D, N = 2 é:

$$A_0^0(v) \ A_1^{-1}(v) \ A_1^0(v) \ A_1^1(v) \ A_2^{-2}(v) \ A_2^{-1}(v) \ A_2^0(v) \ A_2^1(v) \ A_2^2(v) \ A_0^0(v+1) \ ... \quad (12)$$

[00068] e para 2D, N = 2:

$$A_0^0(v) \ A_1^{-1}(v) \ A_1^1(v) \ A_2^{-2}(v) \ A_2^2(v) \ A_0^0(v+1) \ A_1^{-1}(v+1) \ ... \quad (13)$$

[00069] O sinal  $A_0^0(n)$  pode ser considerado como uma representação mono da gravação Ambisonics, não possuindo informação direcional mas sendo representativo para a impressão de timbre geral da gravação. A normalização dos coeficientes Ambisonics é geralmente executada de acordo com a normalização da SH (como será aparente abaixo, veja a Eq. (15)), que deve ser considerada quando decodificando uma gravação externa ( $A_n^m$  são baseados na SH com fator de normalização  $N_{n,m}$ ,  $\tilde{A}_n^m$  são baseados na SH com fator de normalização  $\tilde{N}_{n,m}$ ):

$$A_n^m = \frac{N_{n,m}}{\tilde{N}_{n,m}} \tilde{A}_n^m, \quad (14)$$

[00070] que se torna  $A_{N3D}^m = \sqrt{(2n+1)} \tilde{A}_{SN3D}^m$  para o caso de SN3D para N3D.

[00071] O formato B e o formato AMB utilizam pesos adicionais (Pesos de Gerson, de Furse-Malhan (FuMa), MaxN) que são aplicados para os coeficientes. A normalização de referência então normalmente é SN3D, de acordo com Jerome Daniel, "Représentation de champs acoustiques, application à La transmission et à La reproduction de scènes sonores complexes dans um contexte multimédia", PhD thesis, Université Paris, 6, 2001, e Dave Malham, "3-D acoustic space and its simulation using ambisonics", [http://www.dxarts.washington.edu/courses/567/current/malham\\_3d.pdf](http://www.dxarts.washington.edu/courses/567/current/malham_3d.pdf).

[00072] As duas realizações específicas seguintes das equações de onda para ondas planas ou ondas esféricas ideais apresentam mais detalhes sobre os coeficientes Ambisonics:

#### Ondas Planas

[00073] Solucionando a equação de onda para ondas planas  $A_n^m$  se torna independente de  $k$  e  $r_s$ ;  $\theta_s, \phi_s$  descrevem o ângulos fonte, '\*' denota o complexo conjugado:

$$A_{n\ plane}^m(\theta_s, \phi_s) = 4\pi i^n P_{s_0} Y_n^m(\theta_s, \phi_s)^* = 4\pi i^n d_n^m(\theta_s, \phi_s) \quad (15)$$

[00074] Aqui,  $P_{s_0}$  é utilizado para descrever a pressão do sinal de dimensionamento da fonte medida na origem do sistema de coordenadas de descrição que pode ser uma função de tempo e se

torna  $\frac{A_{0\ plane}^0}{\sqrt{4\pi}}$  para harmônicas esféricas normalizada ortogonais. Geralmente, Ambisonics assume ondas planas e coeficientes Ambisonics  $d_n^m(\theta_s, \phi_s) = \frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n} = P_{s_0} Y_n^m(\theta_s, \phi_s)^*$ , são transmitidos ou armazenados. Esta suposição oferece a possibilidade de sobreposição de diferentes sinais direcionais bem como um projeto de decodificador simples. Isto também é verdadeiro para sinais de um microfone

Soundfield™ gravado no formato B de primeira ordem (N = 1), que se torna óbvio quando comparando a progressão de fase dos filtros de equalização (para progressão teórica, veja o artigo mencionado acima "Unified description of Ambisonics using real and complex spherical harmonics", capítulo 2.1, e para uma progressão protegida por patente veja a US 4042779. A Eq. (1) se torna:

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) Y_n^m(\theta, \phi) 4\pi i^n P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (17)$$

[00075] Os coeficientes  $d_n^m$  podem ser derivados por sinais de conjunto de microfones processados posteriormente ou podem ser criados sinteticamente utilizando um sinal mono  $P_{S_0}(t)$ , caso em que as harmônicas esféricas direcionais  $Y_n^m(\theta_s, \phi_s, t)^*$  podem ser igualmente dependentes do tempo (fonte móvel). A Eq. (17) é válida para cada instância de amostragem temporal  $v$ . O processo de codificação sintética pode ser reescrito (para cada instância de amostra  $v$ ) em uma forma de vetor / matriz para Ambisonics de ordem N selecionada:

$$d = \Psi P_{S_0} \quad (18)$$

[00076] onde d é um sinal Ambisonics, mantendo  $d_n^m(\theta_s, \phi_s)$ , (exemplo para N = 2:  $d(t) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]^T$ ), tamanho(d) = (N+1)<sup>2</sup> × 1 = 0 × 1,  $P_{S_0}$  é a pressão do sinal fonte na origem de referência, e  $\Psi$  é o vetor de codificação, mantendo  $Y_n^m(\theta_s, \phi_s)^*$ , tamanho( $\Psi$ ) = 0 × 1. O vetor de codificação pode ser derivado a partir das harmonias esféricas para a direção específica da fonte  $\theta_s, \phi_s$  (igual à direção da onda plana).

### Ondas Esféricas

[00077] Coeficientes Ambisonics descrevendo ondas periféricas de entrada geradas por fontes de ponto (fontes de campo próximo) para  $r < r_s$  são :

$$A_{n\text{ spherical}}^m(k, \theta_s, \phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (19)$$

[00078] Esta equação é derivada em conexão com as Eqs. (31) até (36) abaixo.  $P_{s_0} = p(\mathbf{0}|r_s)$  descreve a pressão do som na origem e

novamente se torna idêntico a  $\frac{A_0^0}{\sqrt{4\pi}}$ ,  $h_n^{(2)}$  é a função esférica de Hankel de zero-ésima ordem de segundo tipo. A eq. (19) é similar à instrução em Jerome Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format", AES 23rd International Conference, Dinamarca, Maio de 2003. Aqui,

$$\frac{h_n(kr_s)}{h_0(kr_s)} = i^n \sum_{a=0}^n \frac{(n+a)!}{(n-a)! a!} \left( -\frac{ic}{2r_s \omega} \right)^a, \quad \text{a} \quad \text{propósito}$$

$(h_1(kr_s))/(h_0(kr_s)) = i(1 - ic/(r_s \omega))$  que, tendo a Eq. (11) em mente, pode ser encontrado em M.A. Gerson, "General metatheory of auditory localisation", 92th AES Convention, 1992, Preprint 2206, onde Gerson descreve o efeito de proximidade para sinais de primeiro grau. A criação sintética de sinais Ambisonics esféricos é menos comum para Ambisonics de ordens mais elevadas N devido às respostas de

frequência de  $\frac{h_n(kr_s)}{h_0(kr_s)}$  serem difíceis de manusear numericamente para baixas frequências. Estes problemas numéricos podem ser superados por considerar um modelo esférico para decodificação / reprodução como descrito abaixo.

#### Reprodução de campo de som

#### Decodificação de Onda Plana

[00079] Em geral, Ambisonics assume uma reprodução do campo sonoro por L alto-falantes que são uniformemente distribuídos em um círculo ou em uma esfera. Quando assumindo que os alto-falantes são colocados distantes o suficiente a partir da posição do ouvinte, um modelo de decodificação de onda plana é válido no centro ( $r_s > \lambda$ ). A pressão do som gerada por L alto-falantes é descrita por:

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) Y_n^m(\theta, \phi) 4\pi i^n \sum_{l=1}^L w_l Y_n^m(\theta_l, \phi_l)^* \quad (20)$$

[00080] com  $w_l$  sendo o sinal para o alto-falante  $l$  e possuindo a escala unitária de uma pressão de som, 1Pa.  $w_l$  é frequentemente chamada de função de acionamento do alto-falante  $l$ . É desejável que esta pressão do som da Eq. (20) seja idêntica à pressão descrita pela Eq. (17). Isto leva a:

$$\sum_{l=1}^L w_l Y_n^m(\theta_l, \phi_l)^* = d_n^m(\theta_s, \phi_s) = \frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n} \quad (21)$$

[00081] Isto pode ser reescrito na forma de matriz, conhecida como "fórmula de recodificação" (compare com a Eq. (18)):

$$d = \Psi y \quad (22)$$

[00082] Onde  $d$  é um sinal Ambisonics, mantendo  $d_n^m(\theta_s, \phi_s)$  ou  $\frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n}$ , (exemplo para  $N = 2$ :  $d(n) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]$ ), tamanho( $d$ ) =  $(N+1)^2 \times 1 = O \times 1$ ,  $\Psi$  é a matriz (recodificação), mantendo  $Y_n^m(\theta_l, \phi_l)^*$ , tamanho ( $\Psi$ ) -  $O \times L$ , e  $y$  são os sinais do alto-falante  $w_l$ , tamanho ( $y(n)$ , 1) =  $L$ .  $y$  pode então ser derivado utilizando alguns métodos conhecidos, por exemplo, correspondência de modo, ou por métodos que otimizam função especiais de posicionamento de alto-falante.

#### Decodificação para o modelo de onda esférica

[00083] Um modelo de decodificação mais geral novamente assume alto-falantes igualmente distribuídos ao redor da origem com um ponto de irradiação de distância  $r_l$  como ondas esféricas. Os coeficientes Ambisonics  $A_n^m$  são fornecidos pela descrição geral a partir da Eq. (1) e a pressão do som gerada pelos  $L$  alto-falantes é dada de acordo com a Eq. (19):

$$A_n^m = \sum_{l=1}^L 4\pi \frac{h_n(kr_l)}{h_0(kr_l)} w_l Y_n^m(\theta_l, \phi_l)^* \quad (23)$$

[00084] Um decodificador mais sofisticado pode filtrar os coeficientes Ambisonics  $A_n^m$  de modo a recuperar  $C_n^m = A_n^m \frac{h_0(kr_l)}{4\pi h_n(kr_l)}$  e



depois disso aplicar a Eq. (17) com  $d=[C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots]$ , para derivar os pesos dos alto-falantes. Com este modelo, os sinais de alto-falante  $w_l$  são determinados pela pressão na origem. Existe uma abordagem alternativa que utiliza a abordagem de fonte simples primeiro descrita no artigo mencionado acima "Three-dimensional surround sound systems based on spherical harmonics". É assumido que os alto-falantes são igualmente distribuídos na esfera e possuem características sonoras secundárias. A solução é derivada em Jens Ahrens, Sascha Spors, "Analytical driving functions for higher order ambisonics", Proceedings of the ICASSP, páginas 373 até 376, 2008, Eq. (13), a qual pode ser reescrita para truncamento na ordem Ambisonics  $N$  e para um ganho de alto-falante  $g_l$  como uma generalização:

$$w_l = \sum_{n=0}^N \sum_{m=-n}^n g_l \frac{A_n^m}{kr_l h_n^{(2)}(kr_l)} Y_n^m(\theta_l, \phi_l) \quad (24)$$

#### Sinais Ambisonics Codificados à Distância

[00085] Criar  $C_n^m$  no codificador Ambisonics utilizando uma distância de referência do alto-falante  $r_{l_{ref}}$  pode resolver os problemas numéricos de  $A_n^m$  quando modelando ou gravando ondas esféricas (utilizando a Eq. (18)):

$$C_n^m = A_n^m \frac{h_0(kr_{l_{ref}})}{4\pi h_n(kr_{l_{ref}})} = \frac{h_0(kr_{l_{ref}})}{h_n(kr_{l_{ref}})} \frac{h_n(kr_s)}{h_0(kr_s)} P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (25)$$

[00086] São transmitidos ou armazenados  $C_n^m$ , a distância de referência  $r_{l_{ref}}$  e um indicador de que os coeficientes codificados de distância esférica são utilizados. No lado do decodificador, um processamento de decodificação simples como dado na Eq. (22) é viável contanto que a distância real do alto-falante  $r_l \approx r_{l_{ref}}$ . Se esta

diferença for muito grande, uma correção  $D_n^m = C_n^m \frac{h_n(kr_{l_{ref}})}{h_n(kr_l)}$  (26) por

filtragem antes da decodificação Ambisonics é requerida.

[00087] Outros modelos de decodificação tal como a Eq. (24) resultam em fórmulas diferentes para Ambisonics codificados por distância:

$$\tilde{C}_n^m = \frac{A_n^m}{kr_{l_{ref}} h_n(kr_{l_{ref}})} = \frac{1}{kr_{l_{ref}} h_n(kr_{l_{ref}})} \frac{h_n(kr_s)}{h_0(kr_s)} P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (27)$$

[00088] Além disso, a normalização das Harmônicas Esféricas pode ter a influência da fórmula de Ambisonics codificado por distância, isto é, os coeficientes de Ambisonics codificado por distância precisam de um contexto definido.

[00089] Os detalhes para a conversão 2D para 3D mencionada acima são como a seguir: O fator de conversão  $\frac{\alpha_{2D}}{3D}$  para converter um componente circular 2D para um componente esférico 3D por multiplicação pode ser derivado como a seguir:

$$\alpha_{\frac{2D}{3D}} = \frac{S_{n=m}^m\left(\theta = \frac{\pi}{2}, \Phi\right)}{\tilde{\Phi}_{n=|m|}^m(\phi)} = \frac{\tilde{N}_{m,m} P_{|m|,|m|}\left(\cos\left(\theta = \frac{\pi}{2}\right)\right) \Phi_m(\phi)}{N_m \Phi_m(\phi)} \quad (28)$$

[00090] Utilizando a identidade comum (de acordo com Wikipédia de 12 de outubro de 2010, "Associated Legendre polynomials", <http://en.wikipedia>

[.org/w/index.php?title=Associated\\_Legendre\\_polynomials&oldid=3630](http://en.wikipedia.org/w/index.php?title=Associated_Legendre_polynomials&oldid=3630)

01511),  $P_{l,l}(x) = (2l-1)!! (1-x^2)^{\frac{l}{2}}$ , onde  $(2l-1)!! = \prod_{i=1}^l (2i-1)$  é fatorial duplo e  $P_{|m|,|m|}$  pode ser expresso como:

$$P_{|m|,|m|}\left(\cos\left(\theta = \frac{\pi}{2}\right)\right) = (2m-1)!! = \frac{(2m)!}{m! 2^m} \quad (29)$$

[00091] A Eq. (29) inserida na Eq. (28) leva à Eq. (10). A conversão de 2D para orto-3D é derivada por

$$\alpha_{\frac{N2D}{ortho3D}} = \sqrt{\frac{(2m+1)}{4\pi}} \frac{(2m)!}{(2m)!} \frac{1}{m! 2^m} = \sqrt{\frac{(2m+1)(2m)!}{4\pi m!^2 2^{2m}}} = \sqrt{\frac{(2m+1)!}{4\pi m!^2 2^{2m}}} \quad (30)$$

[00092] utilizando a relação  $l! = \frac{(l+1)!}{l+1}$  e substituindo  $l = 2m$ .

[00093] Os detalhes para a expansão de Onda Esférica mencionada acima são como a seguir:

[00094] Solucionar a Eq. (1) para ondas esféricas, as quais são geradas por fontes pontuais para  $r < r_s$  e onda de entrada é mais complicado devido às fontes pontuais com tamanho infinitesimal de desaparecimento precisarem ser descritas utilizando um fluxo de volume  $Q_s$ , onde a pressão irradiada para um ponto do campo em  $r$  e fonte posicionada em  $r_s$  ser dado por (de acordo com o livro "Fourier Acoustics") mencionado acima:

$$p(r|r_s) = -i \rho_0 c k Q_s G(r|r_s) \quad (31)$$

[00095] com  $\rho_0$  sendo a densidade específica e  $G(r|r_s)$  sendo a

função de Green  $G(r|r_s) = \frac{e^{-ik|r-r_s|}}{4\pi |r-r_s|} \quad (32)$ ,  $G(r|r_s)$  também pode ser expressa nas harmônicas esféricas para  $r < r_s$  por

$$G(r|r_s) = i k \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) h_n^{(2)}(kr_s) Y_n^m(\theta, \phi) Y_n^m(\theta_s, \phi_s)^* \quad (33), \text{ onde } h_n^{(2)} \text{ é a}$$

função de Hankel de segundo tipo. Observe que a função de Green

possui uma escala de  $\text{metro}^{-1}$  unitária ( $\frac{1}{m}$  devido a  $k$ ). As Eqs. (31), (33) podem ser comparadas com a Eq. (1) para derivar os coeficientes Ambisonics de ondas esféricas:

$$A_{n\text{ spherical}}^m(k, \theta_s, \phi_s, r_s) = \rho_0 c k^2 Q_s h_n^{(2)}(kr_s) Y_n^m(\theta_s, \phi_s)^* \quad (34) \text{ onde}$$

$Q_s$  é o fluxo de volume em  $\text{m}^3\text{s}^{-1}$  unidades, e  $\rho_0$  é a densidade específica em  $\text{kg m}^{-3}$ . Para estar apto a sinteticamente criar sinais Ambisonics e a relacionar as considerações de onda plana acima, é sensível expressar a Eq. (34) utilizando a pressão sonora gerada na origem do sistema de coordenadas:

$$P_{S_0} = p(0|r_s) = \frac{-i \rho_0 c k Q_s}{4\pi} \frac{e^{-ikr_s}}{r_s} = \frac{\rho_0 c k^2 Q_s}{4\pi} h_0^{(2)}(kr_s) \quad (35)$$

[00097] o que leva a

$$A_{n_{spherical}}^m(k, \Theta_s, \phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{S_0} Y_n^m(\Theta_s, \phi_s)^* \quad (36)$$

### Formato de Armazenamento de Troca

[00098] O formato de armazenamento de acordo com a invenção permite armazenar mais do que uma representação HOA e fluxos direcionais adicionais juntos em um container de dados. Isto permite formatos diferentes de descrições HOA que permitem que decodificadores otimizem a reprodução, e oferece um armazenamento de dados eficiente para tamanhos > 4 GB. Vantagens adicionais são:

A) Pelo armazenamento de várias descrições HOA utilizando diferentes formatos juntos com a informação relacionada de formato de armazenamento, um decodificador Ambisonics está apto a misturar e decodificar ambas as representações.

B) Itens de informação requeridos para decodificadores HOA de próxima geração são armazenados como informação de formato:

- Dimensionalidade, região de interesse) (fontes fora ou dentro da área de escuta), normalização de funções com base esférica;

- Empacotamento de coeficiente Ambisonics e informação de dimensionamento;

Tipo de onda Ambisonics (plana, esférica), raio de referência (para decodificação de ondas esféricas);

- Sinais mono direcionais relacionados podem ser armazenados. Informação de posição destes sinais direcionais pode ser descrita utilizando informação de ângulo ou de distância ou um veto de codificação de coeficientes Ambisonics.

C) O formato de armazenamento de dados Ambisonics é estendido para permitir o armazenamento flexível e econômico de dados:

- Armazenar dados Ambisonics relacionados com

componentes Ambisonics (canais Ambisonics) com diferente resolução de tamanho de palavra PCM;

- Armazenar dados Ambisonics com largura de banda reduzida utilizando reamostragem ou processamento MDCT.

D) Campos de metadados são disponíveis para associar trilhas para decodificação especial (frontal, ambiente) e para permitir o armazenamento de informação acompanhante sobre o arquivo, tal como informação de gravação para sinais de microfone:

- Gravar sistema de coordenadas de referência, microfone, posições da fonte e do ouvinte virtual, características direcionais do microfone, informação de ambiente e de fonte.

E) O formato é adequado para armazenamento de múltiplos quadros contendo diferentes trilhas, permitindo alterações de cena de áudio sem uma descrição da cena. (Observação: uma trilha contém uma descrição de campo de som HOA ou uma única fonte com informação de posição. Um quadro é a combinação de uma ou mais trilhas paralelas). As trilhas podem iniciar no começo de um quadro ou terminar no fim de um quadro, portanto, nenhum código de tempo é requerido.

F) O formato facilita o acesso rápido aos dados de trilha de áudio (rápido para frente ou salto para pontos de deixa) e determinar um código de tempo relativo ao tempo do começo dos dados do arquivo.

#### Parâmetros HOA para troca de dados HOA

[00099] A Tabela 6 resume os parâmetros requeridos a serem definidos para uma troca não ambígua de dados de sinal HOA. A definição das harmônicas esféricas é fixa para os casos com valores complexo e para os casos com valores reais, de acordo com as Eqs. (3) e (6).

Contexto	Dimensionalidade	2D/3D, influencia também empacotamento de coeficientes Ambisonics (AC)
	Região de Interesse	Fig.6, Fig.8, Eqs.(1)(2)

Coeficientes Ambisonics	Tipo SH	Complexo, com valor real, circular para 2D
	Normalização SH	SN3D, N3D, orto-normalizado
	Ponderação AC	Formato B, FuMa, maxN, nenhuma ponderação, definido pelo usuário
	Sequência e resolução de amostra AC	Exemplos nas Eqs.(12)(13), resolução 16/24 bit ou tipos flutuantes.
	Tipo AC	$A_n^m$ não especificado, onda plana Tipo $d_n^m$ , Eq.(16), tipos codificados por distância $D_n^m$ ou $\tilde{C}_n^m$ , Eqs.(26)(27)

**Tabela 6 – Parâmetros para troca não ambígua de gravações HOA**  
**Detalhes do Formato de Arquivo**

[000100] No dito a seguir, o formato do arquivo para armazenar cenas de áudio compostas de Ambisonics de Ordens Elevadas (HOA) ou de fontes únicas com informação de posição é descrito em detalhes. A cena de áudio pode conter múltiplas sequência HOA que podem utilizar diferentes esquemas de normalização. Assim, um decodificador pode calcular os sinais de alto-falante correspondentes para a configuração de alto-falante desejada como uma sobreposição de todas as trilhas de áudio a partir de um arquivo corrente. O arquivo contém todos os dados requeridos para decodificar o conteúdo de áudio. O formato do arquivo de acordo com a invenção oferece o aspecto de armazenar mais do que um HOA ou sinal de fonte única no arquivo único. O formato do arquivo utiliza uma composição de quadros, cada um podendo conter várias trilhas, onde os dados de uma trilha são armazenados em um ou mais pacotes chamados TrackPackets.

[000101] Todos os tipos de número inteiro são armazenados na ordem byte de extremidade menor primeiro de modo que o byte menos significativo vem primeiro. A ordem dos bits é sempre os bits mais significativos primeiro. A notação para tipos de dados inteiros é "int". Um "u" na frente indica inteiro não sinalizado. A resolução no bit é

escrita no fim da definição. Por exemplo, um campo de número inteiro com 16 bits não sinalizado é definido como "uint16". Amostras PCM e coeficientes HOA no formato de número inteiro são representados como número de ponto fixo com o ponto decimal no bit mais significativo. Todos os tipos de dados de ponto flutuante estão de acordo com a Especificação IEEE IEEE-754, "Standard for binary floating-point arithmetic", <http://grouper.ieee.org/groups/754/>. A notação para o tipo de dado de ponto flutuante é "float". A resolução no bit é escrita no fim da definição. Por exemplo, um campo de ponto flutuante com 32 bits é definido como "float32". Identificadores de constante ID, os quais identificam o começo de um quadro, trilha ou bloco, e de cadeias são byte de tipo de dado. A ordem dos bytes de arranjos de bytes é byte e bit mais significativo primeiro. Portanto, o ID "TRACK" é definido em um campo de byte com 32 bits onde os bytes são gravados na ordem física "T", "R", "C" e "K" (<0x54; 0x52; 0x4b>). Os valores hexadecimais iniciam com "0x" (por exemplo, 0xAB64C5). Bits únicos são colocados entre aspas (por exemplo "1") e múltiplos valores binários começam com "0b" (por exemplo, 0b0011 = 0x3). Os nomes de campo de cabeçalho sempre começam com o nome do cabeçalho seguido pelo nome do campo, onde a primeira letra de cada palavra é colocada em maiúscula (por exemplo, TrackHeaderSize). Abreviações de nomes de campo ou de cabeçalho são criadas pela utilização de letras somente em maiúscula (por exemplo, TrackHeaderSize = THS). O formato de arquivo HOA pode incluir mais do que um Quadro, Pacote ou Trilha. Para a discriminação de múltiplos campos de cabeçalho, um número pode seguir o nome do campo ou do cabeçalho. Por exemplo, o segundo TrackPacket da terceira Trilha é denominado "Track3Packet2". O formato do arquivo HOA pode incluir campos com valores complexos. Estes valores complexos são armazenados como a parte real e imaginária onde a

parte real é gravada primeira. O número complexo  $1+i2$  no formato "int8" seria armazenado como "0x01" seguido por "0x02". Por consequência, campos ou coeficientes em um tipo de formato com valores complexos requer duas vezes o tamanho de armazenamento se comparado com o tipo de formato com valores reais correspondente.

### Estrutura do Formato de Arquivo de Ambisonics de Ordens Elevadas

#### Formato de Trilha Única

[000102] O formato de arquivo de Ambisonics de Ordens Elevadas inclui pelo menos um FileHeader, um FrameHeader, um TrackHeader e um TrackPacket como representado na Fig. 9, a qual apresenta um arquivo de formato de arquivo HOA simples ilustrativo que contém uma Track em um ou mais Packets. Portanto, a estrutura básica de um arquivo HOA é um FileHeader seguido por um Frame que inclui pelo menos uma Track. Uma Track consiste sempre de um TrackHeader e de um ou mais TrackPacket.

#### Formato de Múltiplas Trilhas e Quadros

[000103] Em contraste com o FileHeader, o Arquivo HOA pode conter mais do que um Frame, onde um Frame pode conter mais do que uma Track. Um novo FrameHeader é utilizado se o tamanho máximo de um Frame for excedido ou se Tracks forem adicionadas, ou removidas a partir de um Frame para outro. A estrutura de um Arquivo HOA com Múltiplas Trilhas e Quadros é apresentada na Fig. 10. A estrutura de um Múltiplas Trilhas e Quadros começa com FrameHeader seguido por todos os TrackHeaders do Frame. Por consequência, os TrackPackets de cada Track são enviados sucessivamente para os FrameHeaders, onde os TrackPackets são intercalados na mesma ordem como TrackHeaders. Em um arquivo com múltiplas trilhas e quadros, o comprimento de um Pacote em amostras é definido no FrameHeader e é constante para todas as



Trilhas. Adicionalmente, as amostras de cada Trilha são sincronizadas, por exemplo, as amostras de Track1Packet1 são síncronas com as amostras de Track2Packet1. TrackCodingTypes específicos podem causar um atraso no lado do decodificador, e tal atraso específico precisa ser conhecido no lado do decodificador, ou é para estar incluso na parte dependente TrackCodingType do TrackHeader, devido ao decodificador sincronizar todos os TrackPackets com o atraso máximo de todas as Trilhas de um Quadro.

#### Metadados Dependentes do Arquivo

[000104] Metadados que se referem ao arquivo HOA completo podem opcionalmente ser adicionados após o FileHeader nos MetaDataChunks. Um MetaDataChunk começa com um ID Geral do Usuário (GUID) seguido pelo MetaDataChunkSize. A essência do MetaDataChunk, por exemplo, a informação de Metadados, é empacotada em um formato XML ou em qualquer formato definido pelo usuário. A Fig. 11 apresenta a estrutura de um formato de arquivo HOA utilizando vários MetaDataChunks.

#### Tipos de Trilha

[000105] Uma Trilha do Formato HOA se diferencia entre uma HOATrack e uma SingleSourceTrack. A HOATrack inclui o campo de som completo codificado como HOACoefficients. Portanto, uma descrição de cena, por exemplo, as posições das fontes codificadas, não é requerida para decodificar os coeficientes no lado do decodificador. Em outras palavras, uma cena de áudio é armazenada dentro do HOACoefficients. Contrário à HOATrack, a SingleSourceTrack inclui somente uma fonte codificada como amostras PCM junto com a posição da fonte dentro de uma cena de áudio. Ao longo do tempo, a posição da SingleSourceTrack pode ser fixa ou variável. A posição da fonte é enviada como TrackHOAEncodingVector ou como TrackPositionVector. O

TrackHOAEncodingVector contém os valores de codificação HOA para obter os HOACoefficients para cada amostra. O TrackPositionVector contém a posição da fonte como ângulo e distância com respeito à posição central de escuta.

#### Cabeçalho do Arquivo (FileHeader)

Nome do Campo	Tamanho/ bits	Tipo de Dado	Descrição
FileID	32	byte	Identificador constante do arquivo para o Formato de Arquivo HOA
FileVersionNumber	8	uint8	Número da versão do formato HOA 0-255
FileSampleRate	32	uint32	Taxa de Amostragem em Hz constante para todos os Quadros e Trilhas
FileNumberOfFrames	32	uint32	Número Total de Quadros, pelo menos '1' é requerido
Reserved	8	byte	
Número Total de Bits	112		

[000106] O FileHeader inclui toda as informações constantes para o Arquivo HOA completo. O FileID é utilizado para identificar o Formato de Arquivo HOA. A taxa de amostragem é constante para todas as Trilhas, mesmo se ela for enviada no FrameHeader. Arquivos HOA que alteram sua taxa de amostragem a partir de um quadro para outro são inválidos. O número de Quadros é indicado no FileHeader para indicar a estrutura do quadro para o decodificador.

#### Blocos de Metadados

Nome do campo	Tamanho / Bits	Tipo de Dado	Descrição
ChunkID	32	byte	ID Geral do Usuário (ainda não definido)
ChunkSize	32	uint32	Tamanho do bloco em bytes excluindo os campos ChunkID e o ChunkSize

Nome do campo	Tamanho / Bits	Tipo de Dado	Descrição
ChunkData	8 * ChunkSize	byte	Campos definidos pelo usuário ou estrutura XML dependendo de ChunkID
Número Total de Bits	64 + 8 * ChunkSize		

*Cabeçalho do Quadro (FrameHeader)*

Nome do Campo	Tamanho / Bits
FrameID	32
FrameSize	32
FrameNumber	32
FrameNumberOfSamples	32
FrameNumberOfTracks	8
FramePacketSize	32
FrameSampleRate	32
Número Total de Bits	200

[000107] O FrameHeader mantém a informação constante de todas as Trilhas de um Quadro e indica alterações dentro do Arquivo HOA. O FrameID e o FrameSize indicam o começo de um Quadro e o comprimento do Quadro. Estes dois campos permitem um fácil acesso a cada quadro e uma verificação cruzada da estrutura do Quadro. Se o comprimento do Quadro exibir mais do que 32 bits, um Quadro pode ser separado em vários Quadros. Cada Quadro possui um FrameNumber único. O FrameNumber deve iniciar com 0 e deve ser incrementado por um para cada novo Quadro. O número de amostras do Quadro é constante para todas as Trilhas de um Quadro. O número de Trilhas dentro do Quadro é constante para o Quadro. Um novo

cabeçalho de Quadro é enviado para terminar ou iniciar Trilhas em uma posição de amostragem desejada. As amostras de cada Trilha são armazenadas em Pacotes. O tamanho destes TrackPackets é indicado nas amostras e é constante para todas as Trilhas. O número de pacotes é igual ao número inteiro que é requerido para armazenar o número de amostras do Quadro. Portanto, o último pacote de uma Trilha pode conter menos amostras do que o tamanho de pacote indicado. A taxa de amostragem de um quadro é igual à FileSampleRate e é indicada no FrameHeader para permitir decodificação de um Quadro sem conhecimento do FileHeader. Isto pode ser utilizado quando decodificando a partir do meio de um arquivo com múltiplos quadros sem conhecimento do FileHeader, por exemplo, para aplicações de transmissão contínua.

**Cabeçalho da Trilha (TrackHeader)**

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackID	32	byte	O Identificador constante para todo o TrackHeader: <"T"; "R"; "A"; "C"> ou <0x54; 0x52; 0x41; 0x43>
TrackNumber	16	uint16	Um TrackNumber único para a identificação de Trilhas coerentes nos vários quadros
TrackHeaderSize	32	uint32	Tamanho do TrackHeader excluindo os campos TrackID e TrackNumber (Deslocamento até o começo do próximo TrackHeader ou primeiro TrackPacket)
TrackMetaDataOffset	32	uint32	Deslocamento a partir do fim deste campo até o começo do campo TrackMetaData. Zeros é igual a nenhum TrackMetaData incluído.
TrackSourceType	1	binário	'0' = HOATrack e '1' = SingleSourceTrack

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
reserved	7	binário	0b0000000
Condição: TrackSourceType == '0'			TrackHeader para Trilhas HOA
<HOATrackHeader>	din	byte	Veja seção HOA TrackHeader
Condição: TrackSourceType == '1'			TrackHeader para SingleSourceTracks
<SingleSourceTrackHeader>	din	byte	veja seções de Cabeçalho de Trilha de Posição fixa de Fonte Única e de Cabeçalho de Trilha de Posição Móvel de Fonte Única
Condição: TrackMetaDataOffset > 0			
TrackMetadata	din	byte	Campo XML para Metadados dependentes. Veja tabela de TrackMetaData
Número Total de Bits	120 + din		

[000108] O termo "din" se refere a um tamanho de campo dinâmico devido aos campos condicionais. O TrackHeader mantém a informação constante para os Pacotes da Trilha específica. O TrackHeader é separado em uma parte constante e em uma parte variável para dois TrackSourceTypes. O TrackHeader começa com um TrackID constante para verificação e identificação do começo do TrackHeader. Um TrackNumber único é designado para cada Trilha para indicar Trilhas coerentes através das bordas dos Quadros. Assim, uma trilha com o mesmo TrackNumber pode ocorrer no quadro seguinte. O TrackHeaderSize é proporcionado para saltar para o próximo TrackHeader e ele é indicador como um deslocamento a partir do fim do campo TrackHeaderSize. O TrackMetaDataOffset

proporciona o número de amostras para saltar diretamente para o começo do campo TrackMetaData, o qual pode ser utilizado para saltar a parte de comprimento variável do TrackHeader. Um TrackMetaDataOffset de zero indica que o campo TrackMetaData não existe. Dependendo do TrackSourceType, o HOATrackHeader ou o SingleSourceTrackHeader é proporcionado. O HOATrackHeader proporciona a informação adicional para coeficientes HOA padrão que descreve o campo de som completo. O SingleSourceTrackHeader mantém informação para as amostras de uma trilha PCM mono e a posição da fonte. Para SingleSourceTracks, o decodificador tem que incluir as Trilhas dentro da cena. NO fim do TrackHeader, um campo TrackMetaData opcional é definido, o qual utiliza o formato XML para proporcionar Metadados dependentes da trilha, por exemplo, informação adicional para a transmissão no formato A (sinais de conjunto de microfones).

**Cabeçalho da Trilha HOA (HOATrackHeader)**

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackComplex ValueFlag	2	binário	0b00: somente parte real 0b01: parte real e imaginária 0b10: somente parte imaginária 0b11 reservado
TrackSampleFormat	4	binário	0b0000Inteiro de 8 bits não sinalizado 0b0001Inteiro de 8 bits sinalizado 0b0010Inteiro de 16 bits sinalizado 0b0011Inteiro de 24 bits sinalizado 0b0100Inteiro de 32 bits sinalizado 0b0101Inteiro de 64 bits sinalizado 0b0110Flutuante de 32 bits (prec. unico binário) 0b0111Flutuante de 64 bits (prec. duplo binário)

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
			0b1000 Flutuante de 128 bits (prec. quad. binário) 0b1001-0b1111 reservado
reserved	2	binário	Preenche bits
TrackHOAParams	din	Bytes	Veja TrackHOAParams
TrackCodingType	8	uint8	,0'Os coeficientes HOA são codificados como amostras PCM com resolução de bit constante e resolução de frequência constante. ,1'Os coeficientes HOA são codificados com uma resolução de bit e resolução de frequência dependente da ordem senão reservado para tipos adicionais de codificação

Condição: TrackCodingType == '1'			Informação adicional para tipo de codificação 1
TrackBandwidthReductionType	8	uint8	0Largura de banda total para todas as ordens 1Redução de largura de banda via MDCT 2Redução de largura de banda via filtro no domínio de tempo
TrackNumberOfOrderRegions	8	uint8	A largura de banda e a resolução de bit podem ser adaptadas para uma série de regiões onde cada número possui uma ordem final e inicial. As regiões TrackNumberOfOrderRegions indicam o número de regiões definidas.

Grava os seguintes campos para cada região			
TrackRegionFirstOrder	8	uint8	Primeira ordem da região
TrackRegionLastOrder	8	uint8	Última ordem desta região
TrackRegionSampleFormat	4	binário	0b0000 Número inteiro de 8 bits não sinalizado 0b0001 Número inteiro de 8 bits sinalizado 0b0010 Número inteiro de 16 bits sinalizado 0b0011 Número inteiro de 24 bits sinalizado 0b0100 Número inteiro de 32 bits sinalizado 0b0101 Número inteiro de 64 bits sinalizado 0b0110 Flutuante de 32 bits (prec. único binário) 0b0111 Flutuante de 64 bits (prec. duplo binário) 0b1000 Flutuante de 128 bits (prec, quad. binário) 0b1001-0b1111 reservado
TrackRegionUseBandwidthReduction	1	binário	'0' Largura de banda total para esta região '1' reduz largura de banda para esta região com TrackBandwidthReductionType
reserved	3	binário	Preenche bits

Condição: TrackRegionUseBandwidthReduction == '1'			Largura de banda é reduzida nesta região
Condição: TrackBandwidthReductionType == 1			Redução de largura de banda via MDCT informação adicional
TrackRegionWindowType	8	uint8	0: Janela seno: $W(t) = \sin\left(\frac{\pi(t + 0.5)}{N}\right)$ senão: reservado



TrackRegionFirstBin	16	uint16	Depósito MDCT primeiro codificado (frequência de corte inferior)
TrackRegionLastBin	16	uint16	Depósito MDCT codificado por último (frequência de corte superior)
Condição: TrackBandwidthReductionType == 2			Redução de largura de banda via informação adicional do filtro no domínio de tempo
TrackRegionFilterLength	16	uint16	Número de coeficientes do filtro de passa-baixa
<TrackRegionFilterCoefficients>	din	float32	Coeficientes do filtro de passa-baixa do TrackRegionFilterLength
TrackRegionModulationFreq	32	float32	Frequência de modulação normalizada $\frac{\Omega_{mod}}{\pi}$ requerida para deslocar os espectros de sinal
TrackRegionDownsampleFactor	16	uint16	Fator de amostragem para frequência inferior M, deve ser um divisor de FramePacketSize
TrackRegionUpsampleFactor	16	uint16	Fator de amostragem para frequência superior K < M
TrackRegionFilterDelay	16	uint16	Atraso nas amostras (de acordo com FileSampleRate) do processamento de redução de largura de banda de codificação / decodificação

[000109] O HOATrackHeader é uma parte do TrackHeader que mantém informação para decodificação de uma HOATrack. Os TrackPackets de uma HOATrack transferem coeficientes HOA que codificam todo o campo de som de uma Trilha. Basicamente, o HOATrackHeader mantém todos os parâmetros HOA que são requeridos no lado do decodificador para decodificar coeficientes HOA para dada configuração de alto-falantes. O TrackComplexValueFlag e o

TrackSampleFormat definem o tipo do formato dos coeficientes HOA de cada TrackPacket. Para coeficientes codificados ou compactados, o TrackSampleFormat define o formato dos coeficientes decodificados ou descompactados. Todos os tipos de formato podem ser números reais ou complexos. Mais informação sobre números complexos é proporcionada na seção Detalhes do Formato do Arquivo acima. Toda a informação dependente do HOA é definida nos TrackHOAParams. Os TrackHOAParams são reutilizados em outros TrackSourceTypes. Portanto, os campos dos TrackHOAParams são definidos e descritos na seção de Parâmetros HOA da Trilha (TrackHOAParams). O campo TrackCodingType indica o formato de codificação (compactação dos coeficientes HOA. A versão básica do formato de arquivo HOA inclui, por exemplo, dois CodingType. Um CodingType é o tipo de codificação PCM (TrackCodingType == "0"), onde os coeficientes reais ou complexos não compactados são gravados em pacotes no TrackSampleFormat selecionado. A ordem e a normalização dos coeficientes HOA são definidas nos campos TrackHOAParams. Um segundo CodingType permite uma alteração do formato da amostra e limitar a largura de banda dos coeficientes de cada ordem HOA. Uma descrição detalhada deste CodingType é proporcionada na seção Codificação da TrackRegion, e a seguir segue uma curta explicação: O TrackBandwidthReductionType determina o tipo de processamento que foi utilizado para limitar a largura de banda de cada ordem HOA. Se a largura de banda de todos os coeficientes ficar inalterada, a redução da largura de banda pode ser desativada por estabelecer o campo TrackBandwidthReductionType para zero. Dois outros tipos de processamento de redução de largura de banda são definidos. O formato inclui um processamento MDCT no domínio de frequência e opcionalmente um processamento de filtro no domínio de tempo. Para mais informações sobre o processamento MDCT veja a seção Redução

de Largura de Banda via MDCT. As ordens HOA podem ser combinadas em regiões do mesmo formato de amostra e largura de banda. O número de regiões é indicado pelo campo `TrackNumberOfOrderRegions`. Para cada região, o primeiro e último índice de ordem, o formato da amostra e a informação opcional de redução de largura de banda têm que ser definidos. Uma região irá obter pelo menos uma ordem. Ordens que não são cobertas por qualquer região são codificadas com largura de banda total utilizando o formato padrão indicado no campo `TrackSampleFormat`. Um caso especial é o uso de nenhuma região (`TrackNumberOfOrderRegions == 0`). Este caso pode ser utilizado para retirar a intercalação de coeficientes HOA no formato PCM, onde os componentes HOA não são intercalados por amostra. Os coeficientes HOA das ordens de uma região são codificados no `TrackRegionSampleFormat`. O `TrackRegionUseBandwidthReduction` indica a utilização de processamento de redução de largura de banda para os coeficientes das ordens da região. Se o indicador `TrackRegionUseBandwidthReduction` for estabelecido, a informação adicional de redução de largura de banda irá seguir. Para o processamento MDCT, o tipo de janela e o primeiro e o último depósito MDCT codificado são definidos. Por meio disto, o primeiro depósito é equivalente à frequência de corte inferior e o último depósito define a frequência de corte superior. Os depósitos MDCT também são codificados no `TrackRegionSampleFormat`, de acordo com a seção Redução de Largura de Banda via MDCT.

### Tipo de Fonte Única

[000110] Fontes únicas são subdivididas em fontes de posição fixa e de posição móvel. O tipo de fonte é indicado no `TrackMovingSourceFlag`. A diferença entre o tipo de fonte de posição móvel e fixa é que a posição da fonte fixa é indicada somente no `TrackHeader` e em cada `TrackPacket` para fontes móveis. A posição de uma fonte pode ser indicada

explicitamente com o vetor de posição nas coordenadas esféricas ou implicitamente como vetor de codificação HOA. A própria fonte é uma trilha mono PCM que tem que ser codificada para coeficientes HOA no lado do decodificador no caso de utilizar um decodificador Ambisonics para reprodução.

**Cabeçalho de Trilha de Posição Fixa de Fonte Única**

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackMovingSourceFlag	1	binário	Constante '0' para fontes fixas
TrackPositionType	1	binário	'0' Posição é enviada como Posição de ângulo TrackPositionVector [R, theta, phi] '1' Posição é enviada como vetor de comprimento de codificação HOA TrackHOAParamNumberOfCoeffs
TrackSampleFormat	4	binário	0b0000 Inteiro de 8 bits não sinalizado 0b0001 Inteiro de 8 bits sinalizado 0b0010 Inteiro de 16 bits sinalizado 0b0011 Inteiro de 24 bits sinalizado 0b0100 Inteiro de 32 bits sinalizado 0b0101 Inteiro de 64 bits sinalizado 0b0110 Flutuante de 32 bits (prec. único binário) 0b0111 Flutuante de 64 bits (prec. duplo binário) 0b1000 Flutuante de 128 bits (prec. quad. binário) 0b1001-0b1111 reservado
reserved	2	binário	Preenche bits

Nome do Campo	Tamanho / bits	Tipo de dado	Descrição
TrackMovingSourceFlag	1	binário	Constante '0' para fontes fixas
TrackPositionType	1	binário	'0' Posição é enviada como posição de ângulo TrackPositionVector [R, theta, phi] '1' Posição é enviada como vetor de comprimento de codificação HOA
TrackSampleFormat	4	binário	0b0000 Inteiro de 8 bits não sinalizado 0b0001 Inteiro de 8 bits sinalizado 0b0010 Inteiro de 16 bits sinalizado 0b0011 Inteiro de 24 bits sinalizado 0b0100 Inteiro de 24 bits sinalizado 0b0101 Inteiro de 64 bits sinalizado 0b0110 Flutuante de 32 bits (prec. único binário) 0b0111 Flutuante de 64 bits (prec. duplo binário) 0b1000 Flutuante de 128 bits prec. quad. Binário) 0b1001-0b1111 reservado
reserved	2	binário	Preenche bits

Condição: TrackPositionType == '0'			Posição como ângulo TrackPositionVector a seguir
TrackPositionTheta	32	float32	Inclinação em rad [0..pi]
TrackPositionPhi	32	float32	azimute (sentido anti-horário) in rad [0..2pi]
TrackPositionRadius	32	float32	Distancia a partir do ponto de referência em metros
Condição: TrackPositionType == '1'			Posição como vetor de codificação HOA

TrackHOAParams	din	bytes	veja TrackHOAParams
TrackEncodeVectorComplexFlag	2	binário	0b00: somente parte real 0b01: parte real e imaginária 0b10: somente parte imaginária 0b11: reservado Tipo de Número para vetor de codificação
TrackEncodeVectorFormat	1	binário	'0'float32 '1'float64
reserved	5	binário	Preenche bits
Condição: TrackEncodeVectorFormat == '0'			Vetor de codificação como float32
<TrackHOAEncodingVector>	din	float32	Entradas TrackHOAParamNumberOfCoeffs do vetor de codificação HOA na ordem TrackHOAParamCoeffSequence
Condição: TrackEncodeVectorFormat == '1'			Vetor de codificação como float64
<TrackHOAEncodingVector>	din	float64	Entradas TrackHOAParamNumberOfCoeffs do vetor de codificação HOA na ordem TrackHOAParamCoeffSequence

[000111] O tipo de fonte de posição fixa é definido por um TrackMovingSourceFlag de zero. O Segundo campo indica o TrackPositionType que fornece a codificação da posição da fonte como vetor nas coordenadas esféricas ou como vetor de codificação HOA. O formato de codificação das amostras PCM mono é indicado pelo campo TrackSampleFormat. De a posição da fonte for enviada como TrackPositionVector, as coordenadas esféricas da posição da fonte são definidas nos campos TrackPositionTheta (inclinação a partir do eixo x até o plano x, plano y), TrackPositionPhi (partida azimuth no sentido anti-horário) e TrackPositionRadius. Se a posição da fonte for definida como um vetor de codificação HOA, os TrackHOAParams são definidos primeiro. Estes parâmetros são definidos na seção TrackHOAParams e

indicam as normalizações e definições utilizadas do vetor de codificação HOA. Os campos `TrackEncodeVectorComplexFlag` e `TrackEncodeVectorFormat` definem o tipo de formato do vetor `TrackHOAEncoding` seguinte. O `TrackHOAEncodingVector` consiste de valores `TrackHOAParamNumberOfCoeffs` que são codificados no formato "float32" ou "float64".

**Cabeçalho de Trilha de Posição Móvel de Fonte Única**

Nome do Campo	Tamanho / bits	Tipo de Dado	Descrição
<code>TrackMovingSourceFlag</code>	1	binário	Constante '1' para fontes móveis
<code>TrackPositionType</code>	1	binário	'0' Posição é enviada como ângulo <code>TrackPositionVector</code> [R, theta, phi] '1' Posição é enviada como vetor de comprimento de codificação HOA <code>TrackHOAParamNumberOfCoeffs</code>
<code>TrackSampleFormat</code>	4	binário	0b0000 Inteiro de 8 bits não sinalizado 0b0001 Inteiro de 8 bits sinalizado 0b0010 Inteiro de 16 bits sinalizado 0b0011 Inteiro de 24 bits sinalizado 0b0100 Inteiro de 32 bits sinalizado 0b0101 Inteiro de 64 bits sinalizado 0b0110 Flutuante de 32 bits (prec. único binário) 0b0111 Flutuante de 64 bits (prec. duplo binário) 0b1000 Flutuante de 128 bits (prec. quad. binário) 0b1001-0b1111 reservado
reserved	2	binário	Preenche bits
Condição: <code>TrackPositionType == '1'</code>			Posição como vetor de codificação HOA
<code>TrackHOAParams</code>	din	bytes	veja <code>TrackHOAParams</code>

Nome do Campo	Tamanho / bits	Tipo de Dado	Descrição
TrackEncodeVectorComplexFlag	2	binário	0b00: somente parte real 0b01: parte real e imaginária 0b10: somente parte imaginária 0b11: reservado Tipo de Número para vetor de codificação
TrackEncodeVectorFormat	1	binário	'0'float32 '1'float64
reserved	5	binário	Preenche bits

[000112] O tipo de fonte de posição móvel é definido por um TrackMovingSourceFlag de "1". O cabeçalho é idêntico ao cabeçalho de fonte fixa, exceto que os campos de dados de posição da fonte TrackPositionTheta, TrackPositionPhi, TrackPositionRadius e TrackHOAEncodingVector estão ausentes. Para fontes móveis, estes estão localizados no TrackPackets para indicar nova posição (móvel) de fonte em cada Pacote.

#### Tabelas de Trilha Especiais

##### TrackHOAParams

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackHOAParamDimension	1	binário	'0' = 2D and '1' = 3D



Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackHOAParamRegionOfInterest	1	binário	'0' coeficientes HOA foram calculados para fontes fora da região de interesse (interior) '1' coeficientes HOA foram calculados para fontes dentro da região de interesse (exterior) (A região de interesse não contém quaisquer fontes).
TrackHOAParamSphericalHarmonicType	1	binário	'0' real '1' complexo
TrackHOAParamSphericalHarmonicNorm	3	binário	0b000 não normalizado 0b001 semi-normalizado Schmidt 0b010 normalizado 4 $\pi$ ou normalizado 2D 0b011 Orto- normalizado 0b100 Dimensionamento Dedicado other Reservado
TrackHOAParamFurseMalhamFlag	1	binário	Indica que coeficiente HOA estão normalizados pelo dimensionamento de Furse-Malham

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackHOAParamDecoderType	2	binário	0b00 dimensionamento do decodificador de ondas planas: $1/(4\pi i^n)$ 0b01 Dimensionamento do decodificador de ondas esféricas (codificação de distância): $1/(ik h_n(kr_{ls}))$ 0b10 Dimensionamento do decodificador de ondas esféricas (codificação de distância para pressão de som medida): $h_0(kr_{ls})/(ik h_n(kr_{ls}))$ 0b11 coeficientes HOA simples
TrackHOAParamCoefficientSequence	2	binário	0b00 ordem do formato B 0b01 ascendente numérico 0b10 descendente numérico 0b11 Reservado
reserved	5	binário	Preenche bits
TrackHOAParamNumberOfCoeffs	16	uint16	Número de Coeficientes HOA por amostra menos 1
TrackHOAParamHorizontalOrder	8	uint8	Ordem Ambisonics no plano X/Y
TrackHOAParamVerticalOrder	8	uint8	Ordem Ambisonics para a dimensão 3D ('0' para coeficientes HOA 2D)
Condição: TrackHOAParamSphericalHarmonicNorm == "dedicated" <0b101>			Campo para Valores de Dimensionamento Dedicado para cada Coeficiente HOA

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackComplexValueScalingFlag	2	binário	0b00: somente parte real 0b01: parte real e imaginária 0b10: somente parte imaginária 0b11: reservado Tipo de Número para TrackScalingValues dedicados
TrackScalingFormat	1	binário	'0': float32 '1': float64
reserved	5	binário	Preenche bits
Condição: TrackScalingFormat = '0'	TrackScalingFactors como float32		
<TrackScalingFactors>	din	float32	Fatores de Dimensionamento TrackHOAParamNumberOfCoeffs se TrackComplexValueScalingFlag == 0b01 a ordem das partes de número complexo é < [real1, imaginario1], [real2, imaginario2], ..., [realN, imaginario] >
Condição: TrackScalingFormat = '1'	TrackScalingFactors como float64		
<TrackScalingFactors>	din	float64	Fatores de Dimensionamento TrackHOAParamNumberOfCoeffs se TrackComplexValueScalingFlag == 0b01 a ordem das partes de número complexo é < [real1, imaginario1], [real2, imaginario2], ..., [realN, imaginario] >
Condição: TrackHOAParamDecoderType == 0b01    TrackHOAParamDecoderType == 0b10		O raio de referência do alto-falante para a codificação de distância é definido	

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
TrackHOAParamReferenceRadius	16	uint16	Este é o raio de referência do alto-falante $r_{ls}$ em mm que foi aplicado para coeficientes HOA para um decodificador de onda esférica de acordo com Poletti ou Daniel.

[000113] Várias abordagens para a codificação e decodificação HOA foram discutidas no passado. Entretanto, sem qualquer conclusão ou acordo para codificar coeficientes HOA. De forma vantajosa, o formato de acordo com a invenção permite o armazenamento da maioria das representações HOA conhecidas. Os TrackHOAParams são definidos para esclarecer que tipo de normalização e sequência de ordem de coeficientes foi utilizada no lado do codificador. Estas definições têm que ser consideradas no lado do decodificador para a mistura de trilhas HOA e para aplicação da matriz do decodificador. Os coeficientes HOA podem ser aplicados para o campo de som tridimensional completo ou somente para o plano z/y bidimensional. A dimensão da HOATrack é definida pelo campo TrackHOAParamDimension. O TrackHOAParamRegionOfInterest reflete duas expansões de pressão do som em série, por meio do que as fontes residem dentro ou fora da região de interesse, e a região de interesse não contém quaisquer fontes. O cálculo da pressão do som para os casos interior e exterior é definido nas equações (1) e (2) acima, respectivamente, por meio do que a informação de direção do sinal HOA  $A_n^m(k)$  é determinada pela função harmônica esférica complexa conjugada  $Y_n^m(\theta, \phi)^*$ . Esta função é definida em uma versão de número complexo e real. O codificador e o decodificador têm que aplicar a função harmônica esférica do tipo de número equivalente. Portanto, o TrackHOAParamSphericalHarmonicType indica que tipo e função harmônica esférica foi aplicada no lado do decodificador. Como

mentionado acima, basicamente a função harmônica esférica é definida pelas funções de Legendre associadas e por uma função trigonométrica complexa ou real. As funções de Legendre associadas são definidas pela Eq. (5). A representação harmônica esférica com valor complexo é

$$Y_n^m(\theta, \phi) = N_{n,m} P_{n,|m|}(\cos(\theta)) e^{im\phi} \begin{cases} (-1)^m & ; m \geq 0 \\ 1 & ; m < 0 \end{cases}$$

[000114] onde  $N_{n,m}$  é o fator de dimensionamento (de acordo com a Eq. (3)). Esta representação com valor complexo pode ser transformada em uma representação com valor real utilizando a seguinte equação:

$$S_{n,m}(\theta, \phi) = \begin{cases} (-1)^m / \sqrt{2} (Y_n^m + Y_n^{-m}) & ; m > 0 \\ Y_n^0 & ; m = 0 \end{cases} = N_{n,m} P_{n,|m|}(\cos(\theta)) \cos(m\phi)$$

[000115] onde o fator de dimensionamento modificado para as harmônicas esféricas com valor real é

$$\tilde{N}_{n,m} = \sqrt{2 - \delta_{0,m}} N_{n,m}, \quad \delta_{0,m} = \begin{cases} 1 & ; m = 0 \\ 0 & ; m \neq 0 \end{cases}$$

[000116] Para representações 2D, a função Harmônica circular tem que ser utilizada para codificação e decodificação dos coeficientes HOA. A representação com valor complexo da harmônica circular é definida por  $\tilde{Y}_m(\phi) = \tilde{N}_m e^{im\phi}$ . A representação com valor real da harmônica circular é definida por

$$\tilde{S}_m(\phi) = \tilde{N}_m \begin{cases} \cos(m\phi) & ; m \geq 0 \\ \sin(|m|\phi) & ; m < 0 \end{cases}$$

[000117] Vários fatores de normalização  $N_{n,m}$ ,  $\tilde{N}_{n,m}$ ,  $\tilde{N}_m$  e  $\tilde{\tilde{N}}_m$  são utilizados para as funções harmônicas esféricas ou circulares para as aplicações ou requerimentos específicos. Para garantir a decodificação correta dos coeficientes HOA, a normalização da função harmônica esférica utilizada no lado do codificador tem que ser conhecida no lado do decodificador. A Tabela 7 seguinte define as normalizações que podem ser selecionados com o campo TrackHOAParamSphericalHarmonicNorm.

Normalizações de harmônica esférica com valor complexo 3D $\tilde{N}_{n,m}$			
Não normalizada 0b000	Semi-normalizada Schmidt, SN3D 0b001	Normalizada $4\pi$ , N3D, Geodésia $4\pi$ 0b010	Orto-Normalizada 0b011
1	$\sqrt{\frac{(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2n+1)(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2n+1)(n- m )!}{4\pi (n+ m )!}}$
Normalizações de harmônica esférica com valor real 3D $\tilde{N}_{n,m}$			
Não normalizada 0b000	Semi-normalizada Schmidt, SN3D 0b001	Normalizada $4\pi$ , N3D, Geodésia $4\pi$ 0b010	Orto-Normalizada 0b011
$\sqrt{2 - \delta_{0,r}}$	$\sqrt{\frac{(2 - \delta_{0,m})(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2 - \delta_{0,m})(2n+1)(n- m )!}{(n+ m )!}}$	$\sqrt{\frac{(2 - \delta_{0,m})(2n+1)(n- m )!}{4\pi (n+ m )!}}$
Normalizações de harmônica circular com valor complexo 2D $\tilde{N}_m$			
Não normalizada 0b000	Semi-normalizada Schmidt, SN2D 0b001	Normalizada 2D, N2D, 0b010	Orto-Normalizada 0b011
$\sqrt{\frac{1}{2}}$	$\sqrt{\frac{1 + \delta_{0,m}}{2}}$	1	$\sqrt{\frac{1}{2\pi}}$
Normalizações de harmônica circular com valor real 2D $\tilde{N}_m$			
Não normalizada 0b000	Semi-normalizada Schmidt, SN2D 0b001	Normalizada 2D, N2D, 0b010	Orto-Normalizada 0b011
$\sqrt{\frac{2 - \delta_{0,r}}{2}}$	1	$\sqrt{(2 - \delta_{0,m})}$	$\sqrt{(2 - \delta_{0,m}) \frac{1}{2\pi}}$

**Tabela 7 – Normalização de funções harmônicas esféricas e circulares**

[000118] Para normalizações futuras, o valor dedicado do campo TrackHOAParamSphericalHarmonicNorm está disponível. Para uma

normalização dedicada, o fato de dimensionamento para cada coeficiente HOA é definido no fim do TrackHOAParams. Os fatores de dimensionamento dedicados TrackScalingFactors podem ser transmitidos como valores "float32" ou "float64" reais ou complexos. O formato do fator de dimensionamento é definido nos campos TrackCompleteValueScalingFlag e TrackScalingFormat no caso de dimensionamento dedicado.

[000119] A normalização de Furse-Malham pode ser aplicada adicionalmente para os coeficientes HOA codificados para equalização das amplitudes dos coeficientes de diferentes ordens HOA para valores absolutos de menos do que "um" para uma transmissão em nos tipos de formato de número inteiro. A normalização Furse-Malham foi projetada para a função harmônica esférica com valor real SN3D até coeficientes de ordem três. Portanto, é recomendado utilizar a normalização Furse-Malham somente em combinação com a função harmônica esférica com valor real SN3D. Além disso, o TrackHOAParamFurseMalhamFlag é ignorado para Trilhas com uma ordem HOA maior do que três. A normalização Furse-Malham tem que ser invertida no lado do decodificador para decodificar os coeficientes HOA. A Tabela 8 define os coeficientes Furse-Malham.

n	m	Pesos Furse-Malham
0	0	$1/\sqrt{2}$
1	-1	1
1	0	1
1	1	1
2	-2	$2/\sqrt{3}$
2	-1	$2/\sqrt{3}$
2	0	1
2	1	$2/\sqrt{3}$
2	2	$2/\sqrt{3}$

n	m	Pesos Furse-Malham
3	-3	$\sqrt[3]{8/5}$
3	-2	$3/\sqrt{5}$
3	-1	$\sqrt{45/32}$
3	0	1
3	1	$\sqrt{45/32}$
3	2	$3/\sqrt{5}$
3	3	$\sqrt[3]{8/5}$

**Tabela 9 -= Fatores de normalização Furse-Malham a serem aplicados no lado do codificador**

[000120] O TrackHOAParamDecoderType define que tipo de decodificador está no lado do codificador assumido como estando presente no lado do decodificador. O tipo do decodificador determina o modelo de alto-falante (onda esférica ou plana) que é para ser utilizado no lado do decodificador para produzir o campo sonoro. Desse modo, a complexidade computacional do decodificador pode ser reduzida por deslocar partes da equação do decodificador para a equação do codificador. Adicionalmente, questões numéricas no lado do codificador podem ser reduzidas. Adicionalmente, o decodificador pode ser reduzido para um processamento idêntico para todos os coeficientes HOA devido a todas as inconsistências no lado do decodificador poderem ser movidas para o codificador. Entretanto, para ondas esféricas, uma distância constante dos alto-falantes a partir da posição de escuta tem que ser assumida. Portanto, o tipo de decodificador assumido é indicado no TrackHeader, e o raio dos alto-falantes  $r_{ls}$  para os tipos de decodificador de onda esférica é transmitido no campo opcional TrackHOAParamReferenceRadius em milímetros. Um campo adicional no lado do decodificador pode



equalizar as diferenças entre o raio dos alto-falantes assumido e real. A normalização TrackHOAParamDecoderType dos coeficientes HOA  $C_n^m$  depende da utilização da expansão de campo interior ou exterior nas séries selecionadas na TrackHOAParamRegionOfInterest. Observação: coeficientes  $d_n^m$  na Eq. (18) e nas equações seguintes correspondem aos coeficientes  $C_n^m$  no dito a seguir. No lado do codificador, os coeficientes  $C_n^m$  são determinados a partir dos coeficientes  $A_n^m$  ou  $B_n^m$  como definido na Tabela 9, e são armazenados. A normalização utilizada é indicada no campo TrackHOAParamDecoderType do cabeçalho TrackHOAParams:

TrackHOAParamDecoderType	Coeficientes HOA Interior	Coeficientes HOA Exterior
0b00: onda plana	$C_n^m = A_n^m / (4\pi i^n)$	---
0b01: onda esférica	$C_n^m = A_n^m / (i k h_n(kr_{ls}))$	$C_n^m = A_n^m / (i k j_n(kr_{ls}))$
0b10: pressão do som medida da onda esférica	$C_n^m = A_n^m h_0(kr_{ls}) / (h_n(kr_{ls}))$	$C_n^m = A_n^m h_0(kr_{ls}) / (j_n(kr_{ls}))$
0b11: não normalizada	$C_n^m = A_n^m$	$C_n^m = B_n^m$

**Tabela 9 – Coeficientes HOA transmitidos para várias normalizações de tipo de decodificador**

[000121] Os coeficientes HOA para uma amostra de tempo compreendem o número de coeficientes TrackParamNumberOfCoeffs(0)  $C_n^m$ .  $N$  depende da dimensão dos coeficientes HOA. Para campos de som 2D  $O'$  é igual a  $2N + 1$  onde  $N$  é igual ao campo TrackHOAParamHorizontalOrder a partir do cabeçalho TrackHOAParams. Os coeficientes HOA 2D são definidos como  $C_{|m|}^m = C_m$  com  $-N \leq m \leq N$  e podem ser representados como um subconjunto dos coeficientes 3D como apresentado na Tabela 10. Para campos de som 3D  $O$  é igual a  $(N + 1)^2$  onde  $N$  é igual ao campo TrackHOAParamVerticalOrder a partir do cabeçalho TrackHOAParam. Os coeficientes HOA 3D  $C_n^m$  são definidos para  $0 \leq n \leq N$  e  $-n \leq m \leq n$ . Uma representação comum dos coeficientes HOA é dada na Tabela

10:

				$C_0^0$				
			$C_1^{-1}$	$C_1^0$	$C_1^1$			
		$C_2^{-2}$	$C_2^{-1}$	$C_2^0$	$C_2^1$	$C_2^2$		
	$C_3^{-3}$	$C_3^{-2}$	$C_3^{-1}$	$C_3^0$	$C_3^1$	$C_3^2$	$C_3^3$	
$C_4^{-4}$	$C_4^{-3}$	$C_4^{-2}$	$C_4^{-1}$	$C_4^0$	$C_4^1$	$C_4^2$	$C_4^3$	$C_4^4$

**Tabela 10 – Representação de coeficientes HOA até a quarta ordem apresentando os coeficientes 2D em negrito como um subconjunto dos coeficientes 3D**

[000122] No caso de campos de som 3D e `TrackHOAParamHorizontalOrder` maior do que `TrackHOAParamVerticalOrder`, a decodificação de ordem misturada será executada. Nos sinais de ordem misturada, alguns coeficientes de ordem elevada são transmitidos somente em 2D. O campo `TrackHOAParamVerticalOrder` determinar a ordem vertical onde todos os coeficientes são transmitidos. A partir da ordem vertical para a `TrackHOAParamHorizontalOrder`, somente os coeficientes 2D são utilizados. Assim, a `TrackHOAParamHorizontalOrder` é igual ou maior do que a `TrackHOAParamVerticalOrder`. Um exemplo para uma representação de ordem misturada de uma ordem horizontal de quatro e uma ordem vertical de dois é representado na Tabela 11:

				$C_0^0$				
			$C_1^{-1}$	$C_1^0$	$C_1^1$			
		$C_2^{-2}$	$C_2^{-1}$	$C_2^0$	$C_2^1$	$C_2^2$		
	$C_3^{-3}$						$C_3^3$	
$C_4^{-4}$								$C_4^4$

**Tabela 11 – Representação de coeficientes HOA para uma representação de ordem misturada de ordem vertical dois e ordem horizontal quatro.**

[000123] Os coeficientes HOA  $C_n^m$  são armazenados nos Pacotes de uma Trilha. A sequência dos coeficientes, por exemplo, qual

coeficiente vem primeiro e qual vem a seguir, foi definida diferentemente no passado. Portanto, o campo TrackHOAParamCoeffSequence indica três tipos de sequências de coeficientes. As três sequências são derivadas a partir da disposição de coeficientes HOA da Tabela 10.

[000124] A sequência do formato B utiliza um fraseado especial para os coeficientes HOA até a ordem de três como apresentado na Tabela 12:

			W			
		Y	S	X		
	V	T	R	S	U	
Q	O	M	K	L	N	P

**Tabela 12. Convenção de denominação de coeficientes HOA do Formato B**

[000125] Para o Formato B, os coeficientes HOA são transmitidos a partir da menor para a maior ordem, onde os coeficientes HOA de cada ordem são transmitidos em ordem alfabética. Por exemplo, os coeficientes de uma configuração 3D da ordem HOA três são armazenados na sequência W, X, Y, S, R, S, T, U, V, K, L, M, N, O, P e Q. O formato B é definido somente até a terceira ordem HOA. Para a transmissão dos coeficientes horizontais (2D), os coeficientes 3D suplementares são ignorados, por exemplo, W, X, Y, U, V, P, Q.

[000126] Os coeficientes  $C_n^m$  para HOA 3D são transmitidos na TrackHOAParamCoeffSequence de uma maneira numericamente crescente ou decrescente a partir do menor para a menor ordem HOA ( $n = 0 \dots N$ ). A sequência numérica crescente começa com  $n = 0 \dots N$  e aumenta até

[000127]  $m = n$  (

$C_0(0)^0, C_1(1)^{-1}, [C_1(1)^0, C_1(1)^1, C_2(2)^{-2}, C_2(2)^{-1}, C_2(2)^0, C_2(2)^1, [C_2(2)^2, \dots]$ , que

é a sequência "CG" definida em Chris Travis, "Four candidate sequences",

<http://ambisonics.googlegroups.com/web/Four+candidate+component+>

sequences+V09.pdf, 2008. A sequência numérica decrescente  $m$  corre ao contrário de  $m = n$  até

$m = 1$  n (

$C_1(0)^*0, C_1(1)^*1, C_1(1)^*0, C_1(1)^*(-1), C_1(2)^*2, C_1(2)^*1, C_1(2)^*0, C_1(2)^*(-1), C_1(2)^*(-2), \dots$ ) que é a sequência "QM" definida nesta publicação.

[000128] Para coeficiente HOA 2D, as sequências numéricas crescente e decrescente TrackHOAParamCoeffSequence são iguais ao caso 3D, mas onde os coeficientes não utilizados  $|m| \neq n$  (Isto é, somente os coeficientes setoriais  $C_{|m|}^m = C_m$  da Tabela 10) são omitidos.

Assim, a sequência numérica crescente leva a  $(C_1(0)^*0, C_1(1)^*(-1), C_1(1)^*1, C_1(2)^*(-2), C_1(2)^*2, \dots)$  e a sequência numérica decrescente a  $(C_1(0)^*0, C_1(1)^*1, C_1(1)^*(-1), C_1(2)^*2, C_1(2)^*(-2), \dots)$ .

### Pacotes da Trilha

### Pacotes da Trilha HOA

### Pacote do Tipo de Codificação PCM

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
<PacketHOACoeffs>	din	din	Coeficientes HOA de Canal Intercalados armazenados em TrackSampleFormat e TrackHOAParamCoeffSequence, por exemplo, <[W(0), X(0), Y(0), S(0)], [W(1), X(1), Y(1), S(1)], ..., S(FrameNumberOfSamples - 1)] >

[000129] Este Pacote contém os coeficientes HOA  $C_n^m$  na ordem definida na TrackHOAParamCoeffSequence, onde todos os coeficientes de uma amostra de tempo são transmitidos sucessivamente. Este pacote é utilizado para Trilhas HOA padrão com uma TrackSourceType de zero e um TrackCodingType de zero.

### Pacote do Tipo Codificação com Resolução Dinâmica

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
<PacketHOACoeffsCoded>	din	din	Coeficientes HOA de Canal com Intercalação Retirada armazenados de acordo com o TrackCodingType, por exemplo, < [W(0), W(1), W(2), ...], [X(0), X(1), X(2), ...], [Y(0), Y(1), Y(2), ...], [S(0), S(1), S(2), ...] >

[000130] O pacote de resolução dinâmica é utilizado para um TrackSourceType de "zero" e um TrackCodingType de "um". As diferentes resoluções das TrackOrderRegions leva a diferentes tamanhos de armazenamento para cada TrackOrderRegion. Portanto, os coeficientes HOA são armazenados de uma maneira com intercalação retirada, por exemplo, todos os coeficientes de uma ordem HOA são armazenados sucessivamente.

#### Pacotes de Trilha de Fonte Única

##### Pacotes de Posição Fixa de Fonte Única

Nome do Campo	Tamanho / Bits	Tipo de Dados	Descrição
<PacketMonoPCMTrack>	din	din	Amostras PCM de fonte única de áudio armazenadas em TrackSampleFormat

[000131] O Pacote de Posição Fixa de Fonte Única é utilizado para um TrackSourceType de "um" e um TrackMovingSourceFlag de "zero". O Pacote mantém as amostras PCM de uma fonte mono.

##### Pacote de Posição Móvel de Fonte Única

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
PacketDirectionFlag	1	binário	Estabelecido para '1' se a direção tiver alterado. '1' é obrigatório para o primeiro

Nome do Campo	Tamanho / Bits	Tipo de Dado	Descrição
			Pacote de um Quadro.
reserved	7	binário	Bits de preenchimento

Condição: PacketDirectionFlag == '1'			Seguem novos dados de posição
Condição: TrackPositionType == '0'			Posição TrackPositionVector como ângulo TrackPositionVector
theta	32	float32	Inclinação em rad [0..pi]
phi	32	float32	Azimute (anti-horário) em rad [0..2pi]
radius	32	float32	Distancia a partir do ponto de referência em metros

Condição: TrackPositionType == '1'			Posição como vetor de codificação HOA
Condição: TrackEncodeVectorFormat == '0'			Vetor de codificação como float32
<TrackHOAEncoding Vector>	din	float32	Entradas TrackHOAParamNumberOfCoeffs do vetor de codificação HOA na ordem TrackHOAParamCoeffSequence
Condição: TrackEncodeVectorFormat == '1'			Vetor de codificação como float64
<TrackHOAEncoding Vector>	din	float64	Entradas TrackHOAParamNumberOfCoeffs do vetor de codificação HOA na ordem TrackHOAParamCoeffSequence

<PacketMonoPCMTra ck>	din	din	Amostras PCM da fonte única de áudio armazenadas em TrackSampleFormat
--------------------------	-----	-----	---

[000132] O Pacote de Posição Móvel de Fonte Única é utilizado para um TrackSourceType de "um" e um TrackMovingSourceFlag de "um". Ele mantém as amostras PCM mono e a informação de posição para a amostra do TrackPacket. O TrackDirectionFlag indica se a direção do Pacote foi alterada ou a direção do Pacote anterior deve ser utilizada. Para garantir decodificação a partir do começo de cada Quadro, o PacketDirectionFlag é igual a "um" para o primeiro TrackPacket de fonte móvel de um Quadro. Para um TrackDirectionFlag de "um", a informação de direção da fonte de amostra PCM seguinte é transmitida. Dependendo do TrackPositionType, a informação de direção é enviada como TrackPositionVector nas coordenadas esféricas ou como TrackHOAEncodingVector com o TrackEncodingVectorFormat definido. O TrackEncodingVector gera coeficientes HOA que estão de acordo com as definições do campo HOAParamHeader. Sucessivamente à informação de direção, as Amostras PCM mono do TrackPacket são transmitidas.

#### Processamento de Codificação

##### Codificação de TrackRegion

[000133] Sinais HOA podem ser derivados a partir dos registros de campo de som com conjuntos de microfones. Por exemplo, o Eigenmike revelado na WO 03/061336 A1 pode ser utilizado para obter gravações HOA de ordem três. Entretanto, o tamanho finito do conjunto de microfones leva às restrições para os coeficientes HOA gravados. Na WO 03/061336 A1 e no artigo mencionado acima "Three-dimensional surround sound systems based on spherical harmonics", questões causadas pelos conjuntos finitos de microfones são discutidas. A distância das cápsulas de microfone resulta em um limite superior de frequência dado pelo teorema de amostragem

espacial. Acima desta frequência superior, o conjunto de microfones pode não produzir coeficientes HOA corretos. Adicionalmente, a distância finita do microfone a partir da posição de escuta HOA requer um filtro de equalização. Estes filtros obtêm altos ganhos para baixas frequências que igualmente aumentam com cada ordem HOA. Na WO 03/061336 A1, uma frequência de corte inferior para coeficientes de ordem mais elevada é introduzida de modo a manipular a faixa dinâmica do filtro de equalização. Isto apresenta que a largura de banda de coeficientes HOA de diferentes ordens HOA pode ser diferente. Portanto, o formato de arquivo HOA oferece a TrackRegionBandwidthReduction que permite a transmissão somente da largura de banda requerida da frequência para cada ordem HOA. Devido à faixa altamente dinâmica do filtro de equalização e devido ao fato de que o coeficiente de ordem zero basicamente é a soma de todos os sinais de microfone, os coeficientes de diferentes ordens HOA podem possuir diferentes faixas dinâmicas. Portanto, o formato de arquivo HOA também oferece o aspecto de adaptar o tipo de formato para a faixa dinâmica de cada ordem HOA.

#### Processamento de Codificação da TrackRegion

[000134] Como apresentado na Fig. 12, os coeficientes HOA intercalados são alimentados para a primeira etapa ou estágio de retirada de intercalação 1211, a qual é designada para a primeira TrackRegion e separa todos os coeficientes HOA da TrackRegion em memórias temporárias com intercalação retirada em amostras FramePacketSize. Os coeficientes da TrackRegion são derivados a partir do campo TrackRegionLastOrder e do campo TrackRegionFirstOrder. Retirar a intercalação significa que os coeficientes  $C_n^m$  para uma combinação de  $n$  e  $m$  são agrupados em uma memória temporária. A partir da etapa ou estágio de retirada de intercalação 1211, os coeficientes HOA com intercalação retirada são



passados para a seção de codificação de TrackRegion. Os coeficientes HOA intercalados restantes são passados para a etapa ou estágio seguinte de retirada de intercalação de TrackRegion, e assim por diante até a etapa ou estágio de retirada de intercalação 121N. O número N de etapas ou estágios de retirada de intercalação é igual à TrackNumberOfOrderRegions mais "um". A etapa ou estágio de retirada de intercalação adicional 125 retira a intercalação dos coeficientes restantes que não são parte da TrackRegion em um caminho de processamento padrão incluindo uma etapa ou estágio de conversão de formato 126. O caminho de codificação da TrackRegion inclui uma etapa ou estágio opcional de redução de largura de banda 1221 e uma etapa ou estágio de conversão de formato 1231 e executa um processamento paralelo para cada memória temporária de coeficiente HOA. A redução de largura de banda é executada se o campo TrackRegionUseBandwidthReduction for estabelecido para "um". Dependendo do TrackBandwidthReductionType selecionado, um processamento é selecionado para limitar a faixa de frequências dos coeficientes HOA e para criticamente reduzir os mesmos. Isto é executado de modo a reduzir o número de coeficientes HOA para o número mínimo requerido de amostras. A conversão de formato converte o formato do coeficiente HOA corrente para o TrackRegionSampleFormat definido no cabeçalho HOATrack. Isto é a única etapa / estágio no caminho do processamento padrão que converte os coeficientes HOA para o TrackSampleFormat do Cabeçalho de Trilha HOA. A etapa ou estágio multiplexador de TrackPacket 124 multiplexa as memórias temporárias de coeficiente HOA em fluxos de arquivos de dados TrackPacket como definido no campo TrackHOAParamCoeffSequence selecionado, onde os coeficientes  $C_n^m$  para uma combinação de  $n$  e  $m$  índices permanece com a intercalação retirada (dentro de uma memória temporária).

Processamento de Decodificação de TrackRegion

[000135] Como apresentado na Fig. 13, o processamento de decodificação é inverso ao processamento de codificação. A etapa ou estágio demultiplexador 134 demultiplexa o arquivo ou fluxo de dados TrackPacket a partir da TrackHOAParamCoeffSequence em memórias temporárias (não representadas) de coeficiente HOA com intercalação retirada. Cada memória temporária contém coeficientes FramePacketLength  $C_n^m$  para uma combinação de  $n$  e  $m$ .

[000136] A etapa / estágio 134 inicializa caminhos de processamento TrackNumberOfOrderRegion mais "um" e passa o conteúdo das memórias temporárias de coeficiente HOA com intercalação retirada para o caminho de processamento apropriado. Os coeficientes de cada TrackRegion são definidos pelos campos TrackRegionLastOrder e TrackRegionFirstOrder do cabeçalho da Trilha HOA. As ordens HOA que não são cobertas pelas TrackRegions selecionadas são processadas no caminho de processamento padrão incluindo uma etapa ou estágio de conversão de formato 136 e uma etapa ou estágio de intercalação de coeficiente restantes 135. O caminho de processamento padrão corresponde a um caminho TrackProcessing sem uma etapa ou estágio de redução de largura de banda.

[000137] Nos caminhos TrackProcessing, uma etapa / estágio de conversão de formato 1331 até 133N converte os coeficientes HOA que são codificados no TrackRegionSampleFormat no formato de dados que é utilizado para o processamento do decodificador. Dependendo do campo de dado TrackRegionUseBandwidthReduction, uma etapa ou estágio opcional de reconstrução de largura de banda 1321 até 132N vem a seguir, no qual os coeficientes HOA com banda limitada e criticamente amostrados são reconstruídos para a largura de banda plena da Trilha. O tipo de processamento de reconstrução é definido no campo TrackBandwidthReductionType do cabeçalho da

Trilha HOA. Na etapa ou estágio seguinte de intercalação 131 até 131N, os conteúdos das memórias temporárias de coeficientes HOA com intercalação retirada são intercalados por agrupar os coeficientes HOA em uma amostra de tempo, e os coeficientes HOA da *TrackRegion* corrente são combinados com os coeficientes HOA das *TrackRegions* anteriores. A sequência resultante dos coeficientes HOA pode ser adaptada para o processamento da Trilha. Adicionalmente, as etapas / estágios de intercalação lidam com os atrasos entre as *TrackRegions* utilizando a redução de largura de banda e as *TrackRegions* não utilizando redução de largura de banda, atraso este que depende do processamento *TrackBandwidthReductionType* selecionado. Por exemplo, o processamento MDCT adiciona um atraso de amostras *FramePacketSize* e, portanto, as etapas / estágios de intercalação de caminhos de processamento sem redução de largura de banda irão atrasar sua saída em um pacote.

#### Redução da Largura de Banda via MDCT

##### Codificação

[000138] A Fig. 14 apresenta a redução de largura de banda utilizando o processamento MDCT (transformada de co-seno discreta modificada). Cada coeficiente HOA da *TrackRegion* de amostras *FramePacketSize* passa via uma memória temporária 141 até 141M por uma etapa ou estágio de adição de janela MDCT correspondente 1421 até 142M. Cada memória temporária de entrada contém os coeficientes HOA sucessivos temporais  $C_n^m$  de uma combinação de  $n$  e  $m$ , isto é, uma memória temporária é definida como

$$^{(buffer)}C_n^m = [C_n^m(0), C_n^m(1), \dots, C_n^m(FramePacketSize - 1)].$$

[000139] O número M de memórias temporárias é o mesmo que o número de componentes Ambisonics  $(N + 1)^2$  para um campo de som pleno 3D de ordem N. A manipulação da memória temporária executa uma sobreposição de 50% para o processamento MDCT seguinte por

combinar o conteúdo da memória temporária anterior com o conteúdo corrente da memória temporária em um novo conteúdo para o processamento MDCT nas etapas ou estágios correspondentes 1431 até 143M, e armazena o conteúdo corrente da memória temporária para o processamento do conteúdo seguinte da memória temporária. O processamento MDCT reinicia no começo de cada Quadro, o que significa que todos os coeficientes de uma Trilha do Quadro corrente podem ser decodificados sem conhecimento do Quadro anterior, e seguindo o último conteúdo da memória temporária do Quadro corrente, um conteúdo adicional de memória temporária de zeros é processado. Portanto, as *TrackRegions* processadas com MDCT produzem um *TrackPacket* extra. Nas etapas / estágios de adição de janela, o conteúdo da memória temporária correspondente é multiplicado pela função de janela selecionada  $w(t)$ , que é definida no *TrackRegionWindowType* do campo de cabeçalho *HOATrack* para cada *TrackRegion*. A transformada de Co-Seno Discreta Modificada é primeiro mencionada em J.P. Princen, A.B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Transactions on Acoustics, Speech and Signal Processing, vol.ASSP-34, no.5, paginas 1153–1161, Outubro, 1986. A MDCT pode ser considerada como representando um banco de filtros criticamente amostrado de subbandas *FramePacketSize*, e ela requer uma sobreposição de 50% da memória temporária de entrada. A memória temporária de entrada possui um comprimento de duas vezes o tamanho da subbanda. A MDCT é definida pela equação seguinte com  $T$  igual a *FramePacketSize*:

$$C'_n{}^m(k) = \sum_{t=0}^{2T-1} w(t)C_n^m(t) \cos \left[ \frac{\pi}{T} \left( t + \frac{T+1}{2} \right) \left( k + \frac{1}{2} \right) \right] \text{ for } 0 \leq k < T$$

[000140] Os coeficientes  $C'_n{}^m(k)$  são chamados de depósitos MDCT. O cálculo da MDCT pode ser implementado utilizando a Transformada

Rápida de Fourier. Nas etapas ou estágios seguintes de corte de região de frequência 1441 até 1444, a redução da largura de banda é executada por remover todos os depósitos MDCT  $c_n^m(k)$  com  $k < TrackRegionFirstBin$  and  $k > TrackRegionLastBin$ , para a redução do comprimento da memória temporária para  $TrackRegionLastBin - TrackRegionFirstBin + 1$ , onde,  $TrackRegionFirstBin$  é a frequência de corte inferior para a  $TrackRegion$  e  $TrackRegionLastBin$  é a frequência de corte superior. A omissão dos depósitos MDCT pode ser considerada como representando um filtro de passa-faixa com frequências de corte correspondendo às frequências  $TrackRegionLastBin$  e  $TrackRegionFirstBin$ . Portanto, somente os depósitos MDCT requeridos são transmitidos.

#### Decodificação

[000141] A Fig. 15 apresenta a decodificação ou reconstrução de largura de banda utilizando o processamento MDCT, no qual coeficientes HOA de  $TrackRegions$  com largura de banda limitada são reconstruídos para larguras de banda plenas da Trilha. Esta reconstrução de largura de banda processa o conteúdo da memória temporária de coeficientes HOA temporariamente com intercalação retirada em paralelo, onde cada memória temporária contém  $TrackRegionlastBin - TrackRegionFirstBin + 1$  depósitos MDCT de coeficientes  $c_n^m(k)$ . As etapas ou estágios de adição de regiões de frequência ausentes 1541 até 154M reconstroem o conteúdo completo da memória temporária MDCT de tamanho  $FramePacketLength$  por complementar os depósitos MDCT recebidos com os depósitos MDCT ausentes  $k < TrackRegionFirstBin$  e  $k > TrackRegionLastBin$  utilizando zeros. Depois disso, a MDCT inversa é executada nas etapas ou estágios de MDCT inversa correspondentes 1531 até 153M de modo a reconstruir os coeficientes HOA no domínio do tempo  $c_n^m(t)$ . A MDCT inversa pode ser interpretada como um banco de filtros de síntese

onde os depósitos MDCT com *FramePacketLength* são convertidos para duas vezes os coeficientes no domínio de tempo com *FramePacketLength*. Entretanto, a reconstrução completa das amostras no domínio de tempo requer uma multiplicação com a função de janela  $w(t)$  utilizada no codificador e uma sobreposição – adição da primeira metade do conteúdo corrente da memória temporária com a segunda metade do conteúdo anterior da memória temporária. A MDCT inversa é definida pela seguinte equação:

$$c_n^m(t) = \frac{w(t)}{2T} \sum_{k=0}^{T-1} c_n^m(k) \cos \left[ \frac{\pi}{T} \left( t + \frac{T+1}{2} \right) \left( k + \frac{1}{2} \right) \right] \text{ for } 0 \leq t < T$$

[000142] Da mesma forma que para a MDCT, a MDCT inversa pode ser implementada utilizando a Transformada Rápida de Fourier.

[000143] As etapas ou estágios de adição de janela MDCT 1521 até 152M multiplicam os coeficientes no domínio do tempos reconstruídos pela função de janela definida pelo *TrackRegionWindowType*. As memórias temporárias seguintes 151 até 151M adicionam a primeira metade do conteúdo corrente da memória temporária do *TrackPacket* com a segunda metade do último conteúdo da memória temporária do *TrackPacket* de modo a reconstruir os coeficientes no domínio do tempo com *FramePacketSize*. A segunda metade do conteúdo corrente da memória do *TrackPacket* é armazenada para o processamento do *TrackPacket* seguinte, processamento de sobreposição / adição este que remove os componentes de distorção contrários de ambos os conteúdos de memória temporária.

[000144] Para arquivos HOA com múltiplos Quadros, o codificador é proibido de utilizar o último conteúdo da memória temporária do quadro anterior para o procedimento de sobreposição / adição no começo de um novo Quadro. Portanto, nas bordas do Quadro ou no começo de um novo Quadro, o conteúdo de sobreposição / adição da memória temporária está ausente, e a reconstrução do primeiro *TrackPacket* de um Quadro pode ser executada no segundo

*TrackPacket*, por meio do que um atraso de um *FramePacket* e a decodificação de um *TrackPacket* extra é introduzida como comparada com os caminhos de processamento sem redução de largura de banda. Este atraso é manipulado pelas etapas / estágios de intercalação descritos em conexão com a Fig. 13.

## REIVINDICAÇÕES

1. Estrutura de dados para dados de áudio Ambisonics de Ordens Elevadas (HOA), incluindo coeficientes Ambisonics, cuja estrutura de dados inclui dados de conteúdo de áudio espacial 2D ou 3D ou ambos 2D e 3D para uma ou mais diferentes descrições de fluxo de dados de áudio HOA, e cuja estrutura de dados também é adequada para dados de áudio HOA que possuem uma ordem maior do que "3", e estrutura de dados esta que, em adição, pode incluir dados de fonte única de sinal de áudio, ou dados de áudio de conjunto de microfones, ou ambos os dados de fonte única de sinal de áudio e dados de áudio de conjunto de microfones a partir de posições espaciais fixas ou variando com o tempo,

**caracterizada pelo fato de que** as diferentes descrições de fluxo de dados de áudio HOA estão relacionadas com pelo menos dois dentre diferentes densidades de posição de alto-falante, tipos de onda HOA codificada, ordens HOA e dimensionalidade HOA, e

em que uma descrição de fluxo de dados de áudio HOA contém dados de áudio para uma apresentação com uma dada disposição de alto-falantes (11, 21) localizada em uma área distinta de um local de apresentação (10), e uma outra descrição de fluxo de dados de áudio HOA contém dados de áudio para uma apresentação com uma disposição de alto-falantes (12, 22) diferente rodeando o local da apresentação (10), em que a disposição de alto-falantes (12, 22) diferente possui uma densidade de posição de alto-falantes que é menor do que aquela da dada disposição de alto-falantes (11, 21).

2. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** os dados de áudio para a dada disposição de alto-falantes (11, 21) representa ondas esféricas e uma primeira ordem Ambisonics, e os dados de áudio para a disposição de alto-falantes (12, 22) diferente representa ondas planas, ou uma



segunda ordem Ambisonics, ou ambas as ondas planas e a segunda ordem Ambisonics, sendo que a segunda ordem Ambisonics é menor do que a primeira ordem Ambisonics.

3. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** a estrutura de dados serve como descrição de cena, onde trilhas de uma cena de áudio podem iniciar e terminar a qualquer momento.

4. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** a estrutura de dados inclui itens de dado com respeito a um ou mais dentre:

- região de interesse relacionada a fontes de áudio fora ou dentro de uma área de escuta;
- normalização de funções com base esférica;
- direcionalidade de propagação;
- informação de dimensionamento de coeficiente Ambisonics;
- tipo de onda Ambisonics, por exemplo, plana ou esférica;
- no caso de ondas esféricas, raio de referência para decodificação.

5. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** os coeficientes Ambisonics são coeficientes complexos.

6. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** a estrutura de dados inclui pelo menos um dentre metadados com respeito às direções e às características para um ou mais microfones, e um vetor de codificação para sinais de entrada de fonte única.

7. Estrutura de dados, de acordo com uma das reivindicações 1 até 6, **caracterizada pelo fato de que** pelo menos parte dos coeficientes Ambisonics tem a largura de banda reduzida, de

modo que, para diferentes ordens HOA, a largura de banda dos coeficientes Ambisonics relacionados é diferente (1221-122N).

8. Estrutura de dados, de acordo com a reivindicação 7, **caracterizada pelo fato de que** a redução de largura de banda é baseada em processamento de transformada discreta de cosseno modificada (MDCT) (1431-143M).

9. Estrutura de dados, de acordo com a reivindicação 1, **caracterizada pelo fato de que** o local de apresentação é uma área de escuta ou de sentar em um cinema.

10. Método para codificar e dispor dados para uma estrutura de dados **caracterizada pelo fato de que** estrutura de dados é uma estrutura de dados conforme definida na reivindicação 1.

11. Método para apresentação de áudio, **caracterizado pelo fato de que** compreende a etapa de:

receber um fluxo de dados de áudio Ambisonics de Ordens Elevadas (HOA) contendo pelo menos dois sinais de dados de áudio HOA diferentes,

em que pelo menos um primeiro dentre os sinais é utilizado (231, 232) para apresentação com uma dada disposição de alto-falantes (11, 21) localizada em uma área distinta de um local de apresentação (10), e

em que pelo menos um segundo e diferente sinal dentre os sinais é utilizado (241, 242, 243) para apresentação com uma disposição de alto-falantes (12, 22) diferente rodeando o local de apresentação (10), em que a disposição de alto-falantes (12, 22) diferente possui uma densidade de posição de alto-falantes que é menor do que aquela da dada disposição de alto-falantes (11, 21).

12. Método, de acordo com a reivindicação 11, **caracterizado pelo fato de que** os dados de áudio para a dada disposição de alto-falantes (11, 21) representam ondas esféricas e

uma primeira ordem Ambisonics, e os dados de áudio para a disposição de alto-falantes (12, 22) diferente representam ondas planas, ou uma segunda ordem Ambisonics, ou ambas as ondas planas e a segunda ordem Ambisonics, sendo que a segunda ordem Ambisonics é menor do que a primeira ordem Ambisonics.

13. Método, de acordo com a reivindicação 11, **caracterizado pelo fato de que** o local de apresentação é uma área de escuta ou de sentar em um cinema.

14. Aparelho para apresentação de áudio, **caracterizado pelo fato de que** compreende:

meios para receber um fluxo de dados de áudio Ambisonics de Ordens Elevadas (HOA) contendo pelo menos dois sinais de dados de áudio HOA diferentes;

meios para processar pelo menos um primeiro dentre os sinais (231, 232) para apresentação com uma dada disposição de alto-falantes (11, 21) localizada em uma área distinta de um local de apresentação (10); e

meios para processar pelo menos um segundo e diferente sinal dentre os sinais (241, 242, 243) para apresentação com uma disposição de alto-falantes (12, 22) diferente rodeando o local de apresentação (10), em que a disposição de alto-falantes (12, 22) diferente possui uma densidade de posição de alto-falantes que é menor do que aquela da dada disposição de alto-falantes (11, 21).

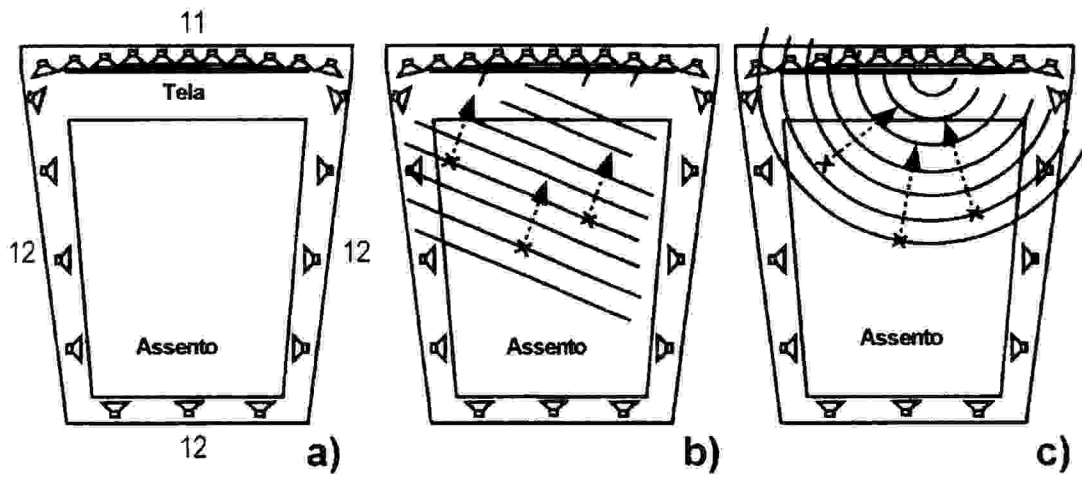


Fig. 1

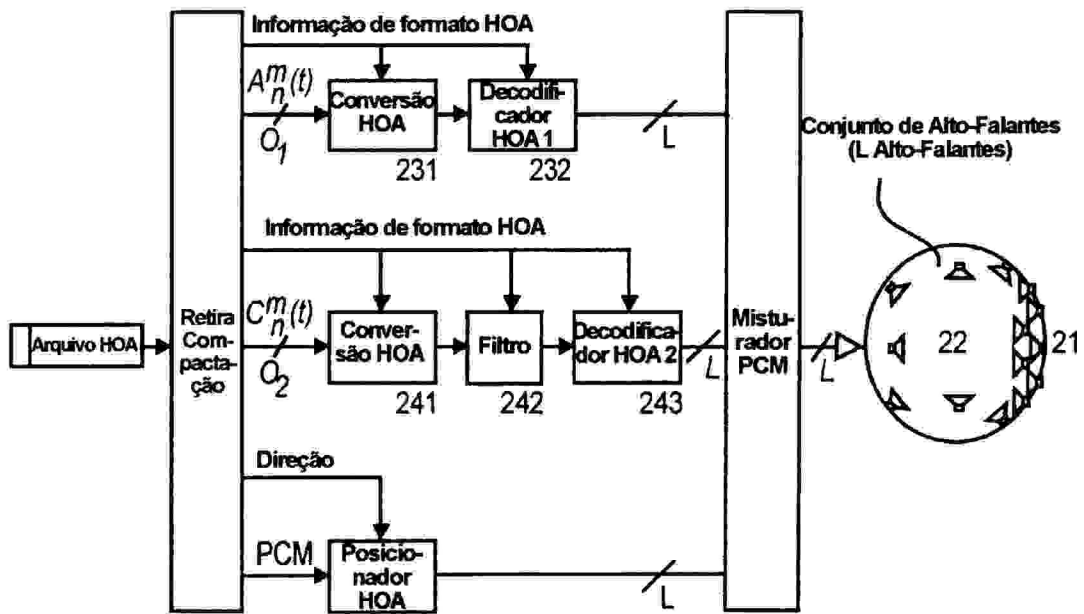


Fig. 2

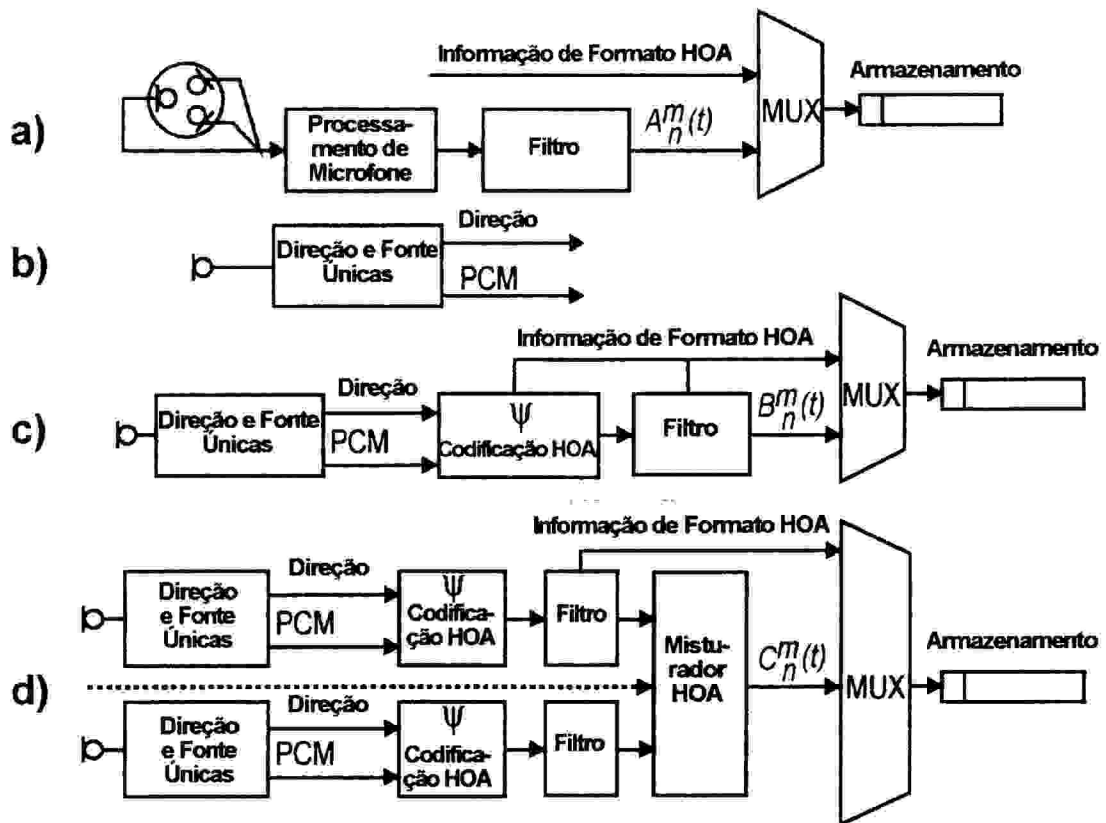


Fig. 3

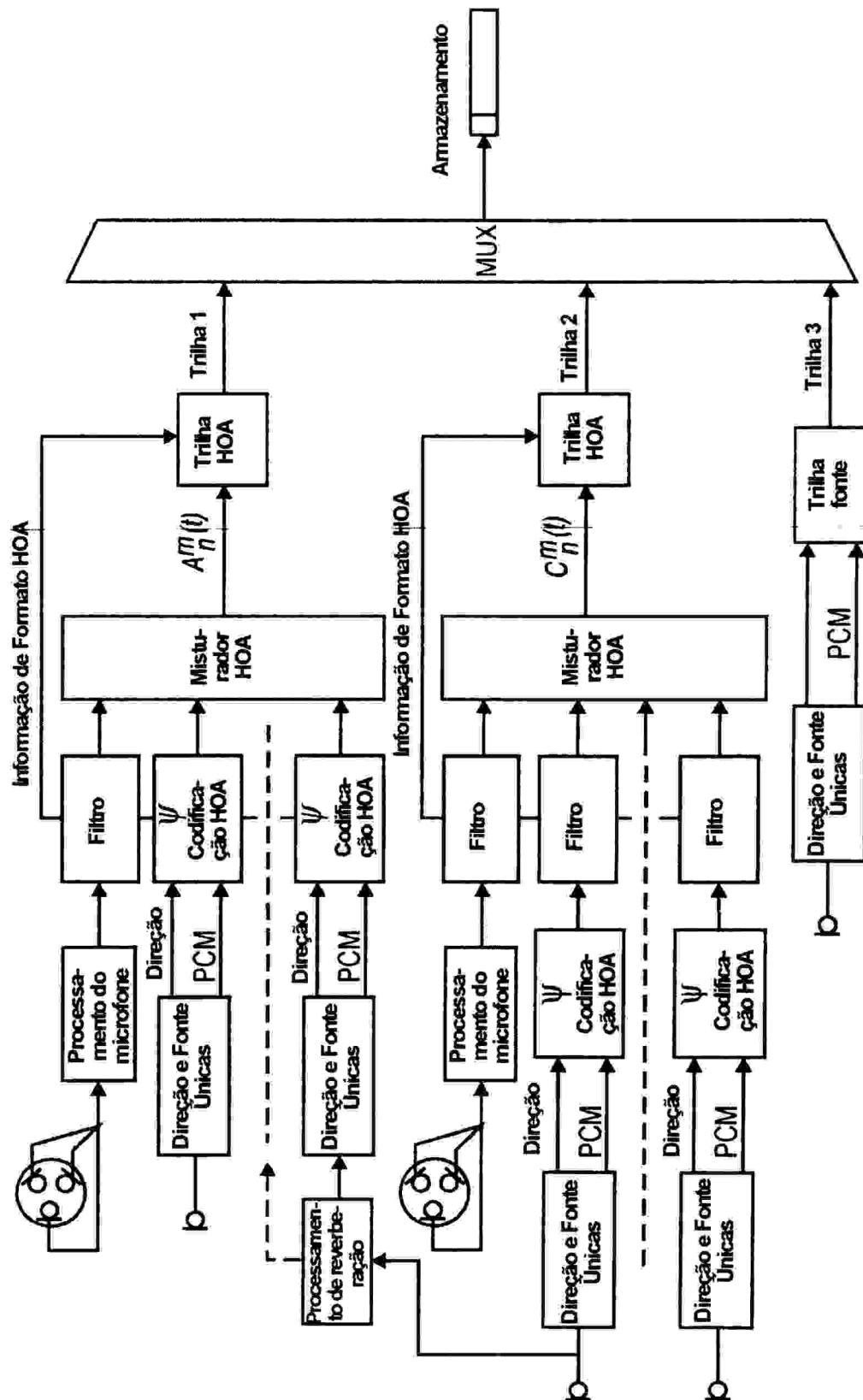


Fig. 4

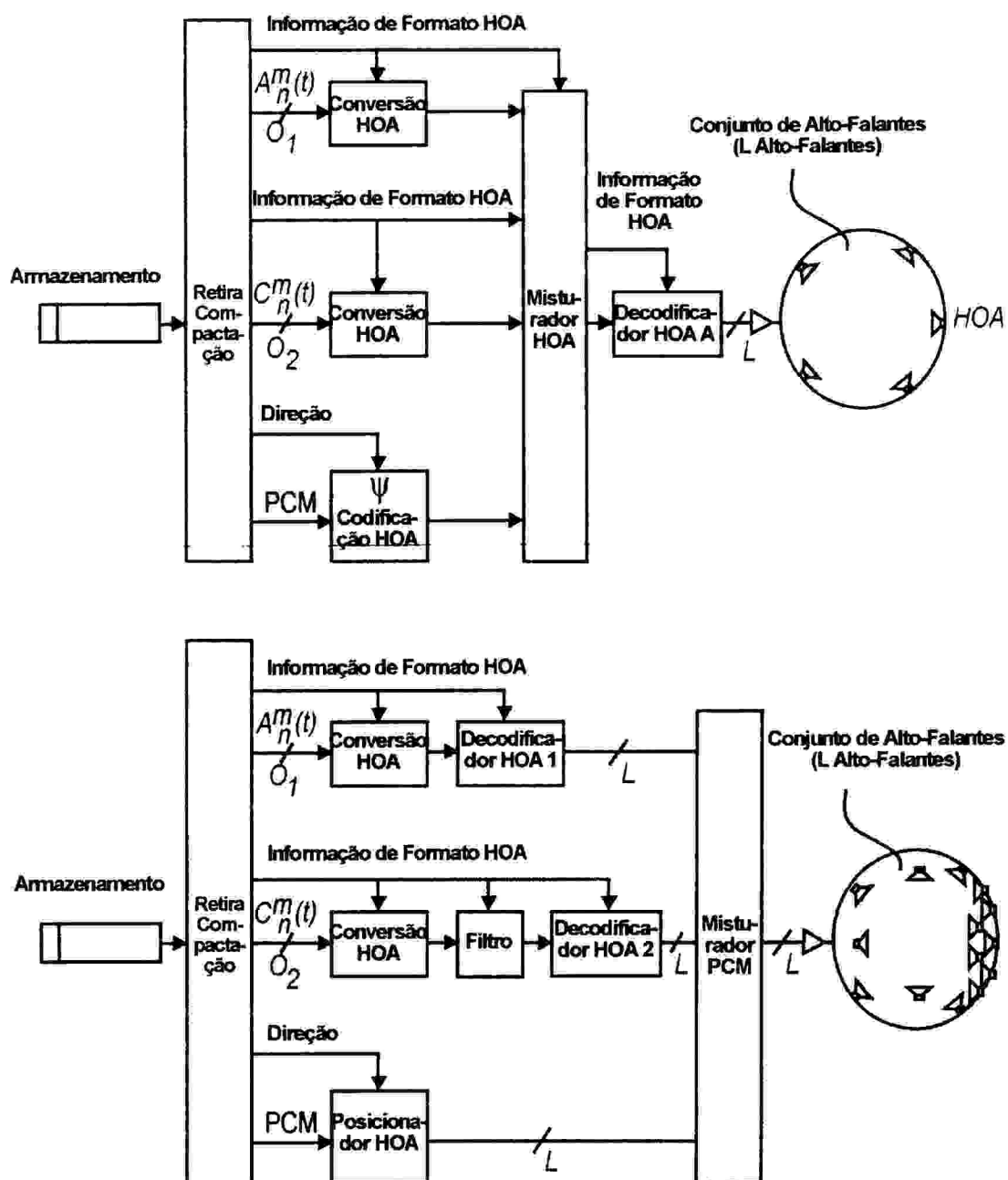


Fig. 5

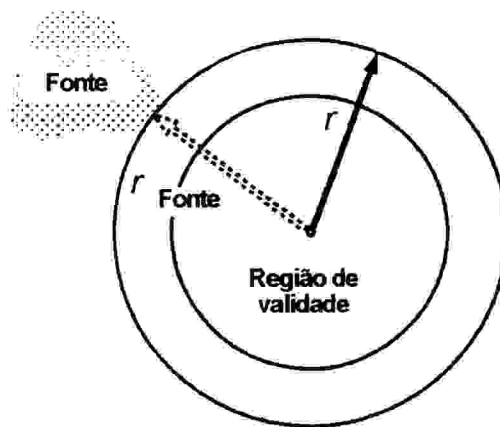


Fig. 6

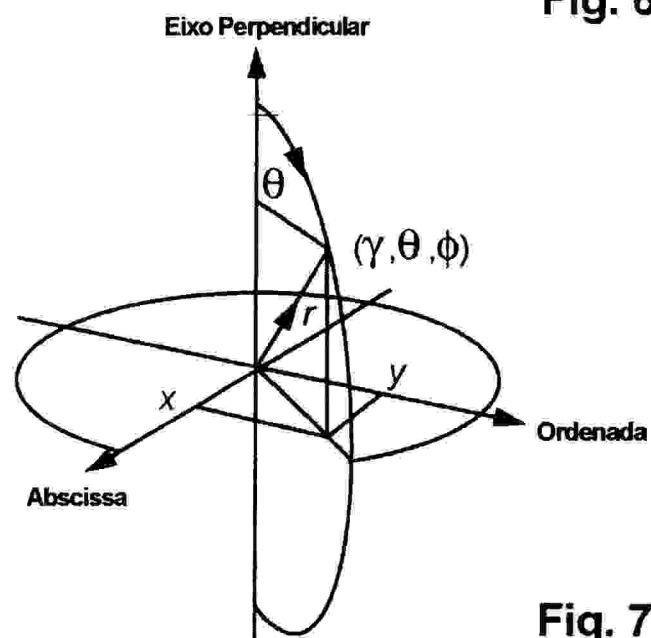


Fig. 7

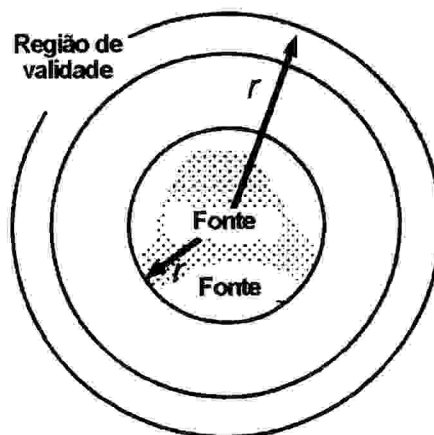


Fig. 8



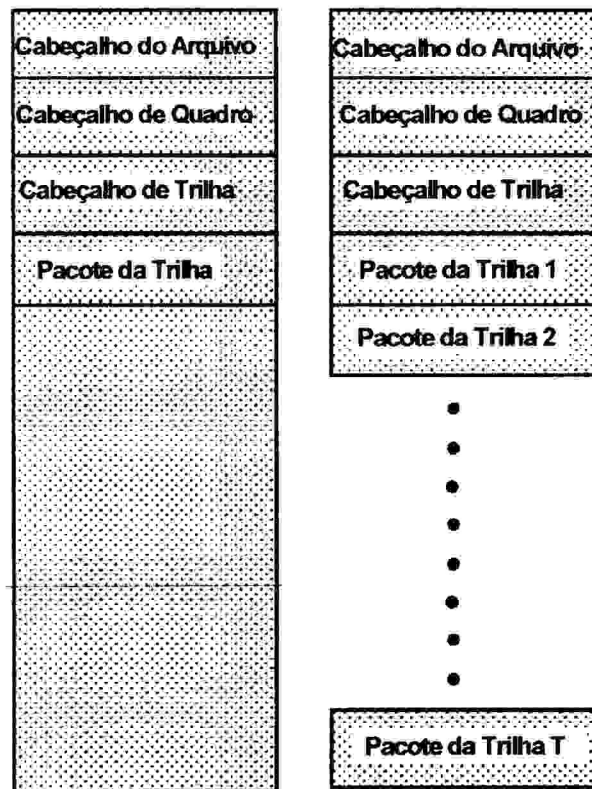


Fig. 9

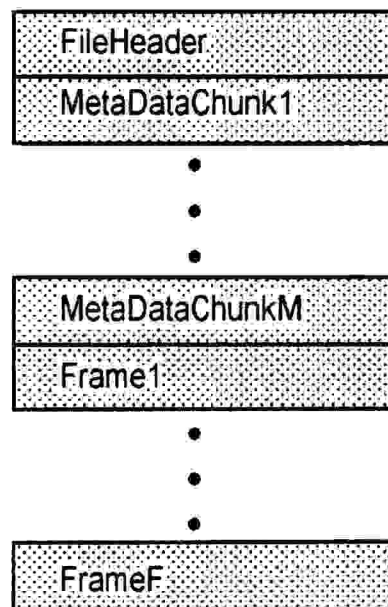
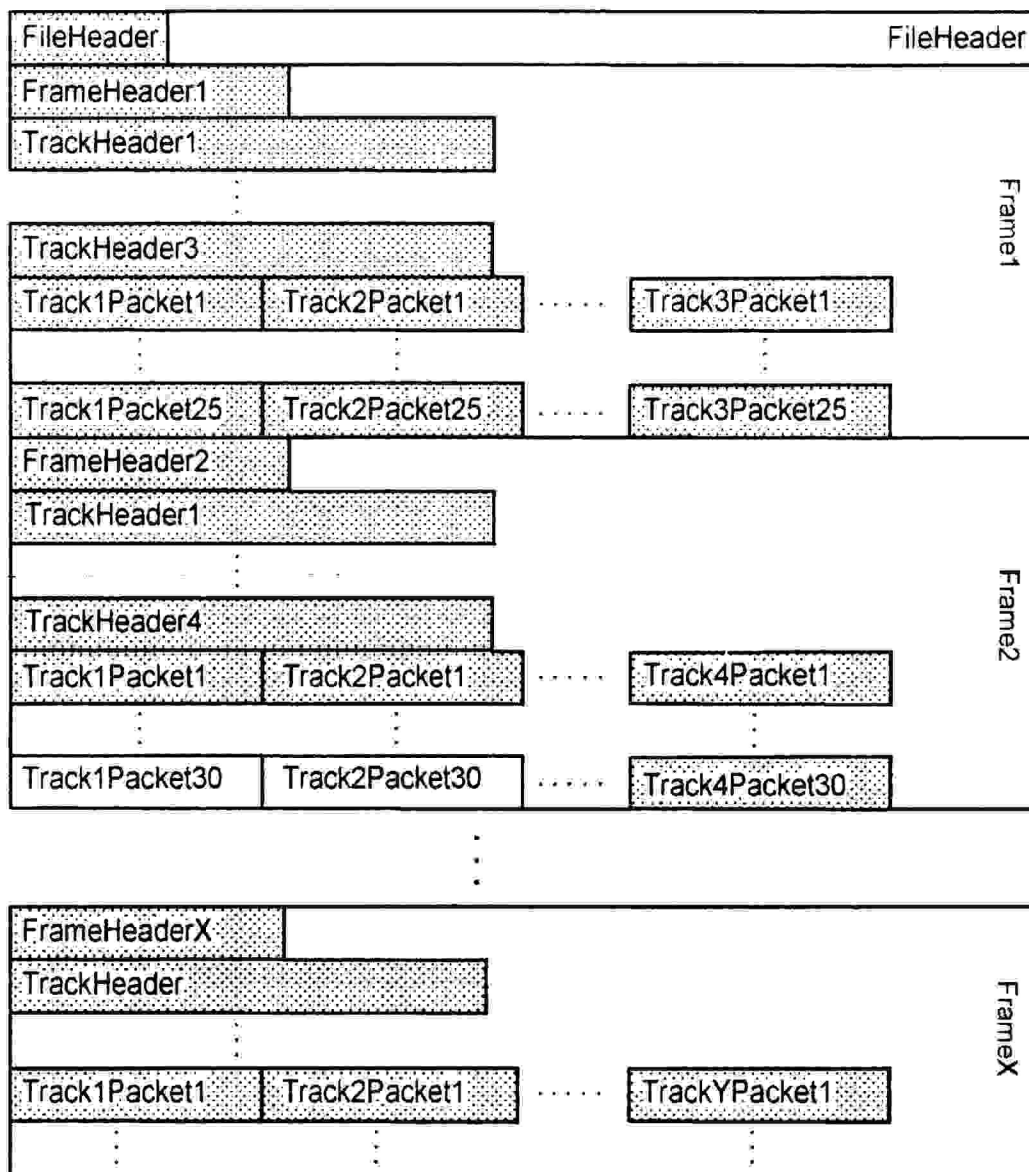


Fig. 11

**Fig. 10**

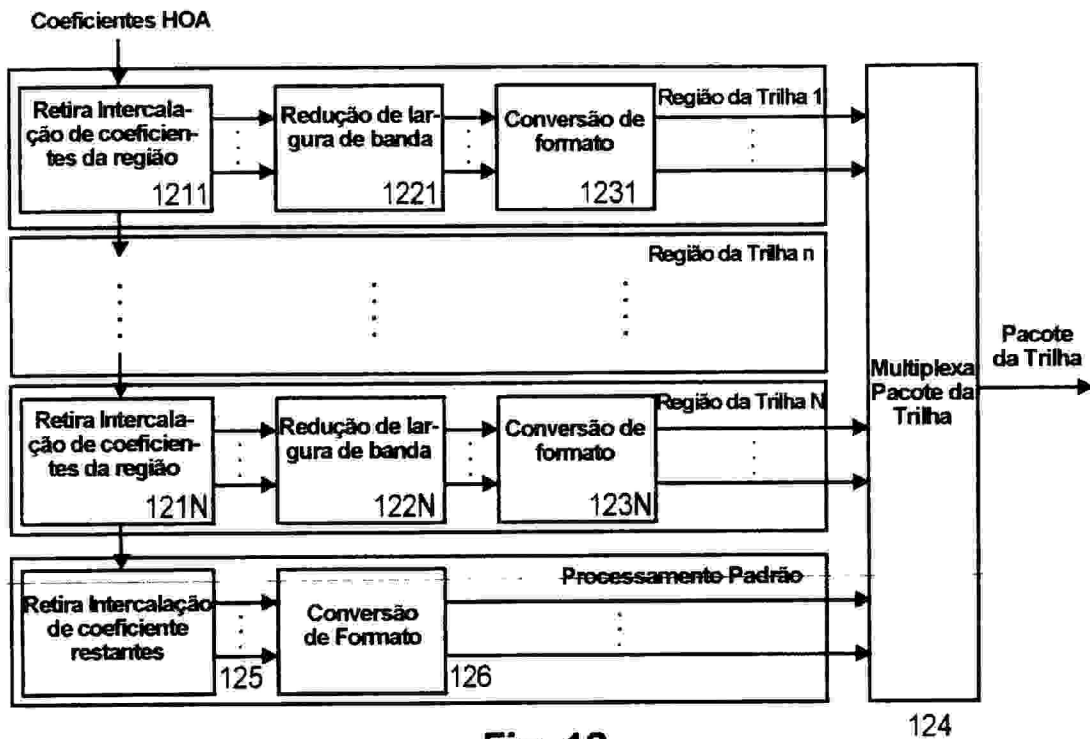


Fig. 12

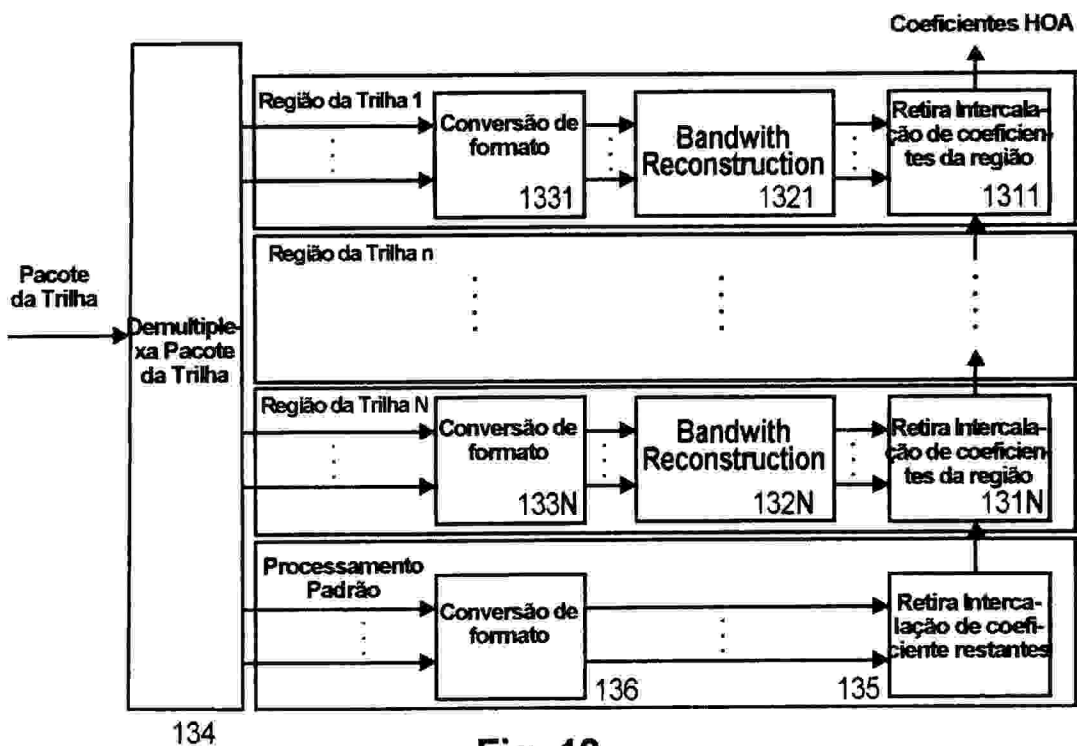


Fig. 13

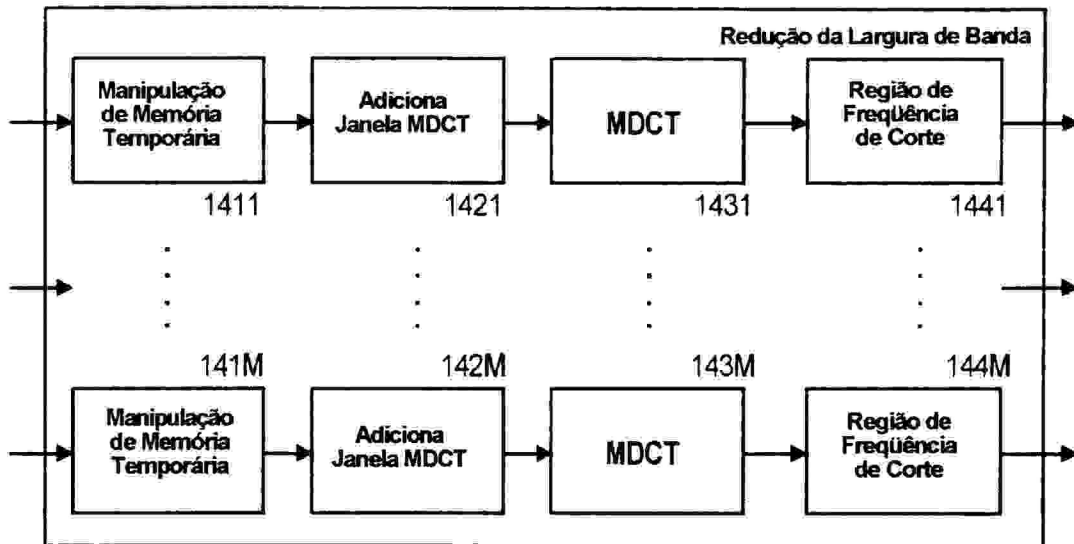


Fig. 14

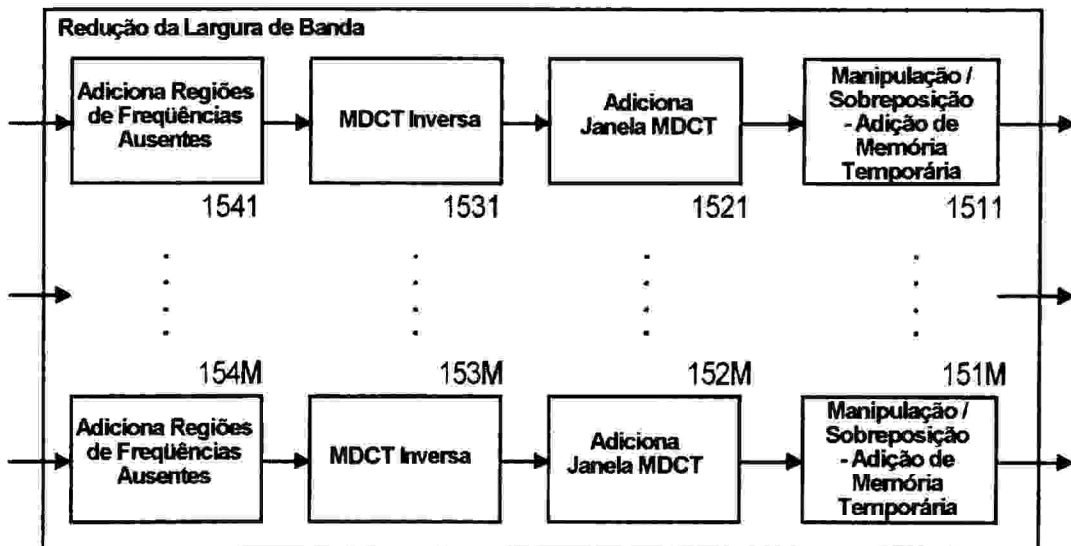


Fig. 15