

**(12) FASCÍCULO DE PATENTE DE INVENÇÃO**

(22) Data de pedido: <b>2011.10.26</b>	(73) Titular(es): <b>THOMSON LICENSING</b> <b>1-5 RUE JEANNE D`ARC 92130 ISSY-LES-</b> <b>MOULINEAUX</b>	<b>FR</b>
(30) Prioridade(s): <b>2010.11.05 EP 10306211</b>		
(43) Data de publicação do pedido: <b>2013.09.11</b>	(72) Inventor(es):	
(45) Data e BPI da concessão: <b>2014.08.27</b> <b>196/2014</b>	<b>FLORIAN KEILER</b> <b>JOHANN-MARKUS BATKE</b> <b>JOHANNES BOEHM</b> <b>SVEN KORDON</b> <b>HOLGER KROPP</b>	<b>DE</b> <b>DE</b> <b>DE</b> <b>DE</b> <b>DE</b>
	(74) Mandatário: <b>LUÍS MANUEL DE ALMADA DA SILVA CARVALHO</b> <b>RUA VÍCTOR CORDON, 14 1249-103 LISBOA</b>	<b>PT</b>

(54) Epígrafe: **ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDEM SUPERIOR**

(57) Resumo:

A INVENÇÃO ESTÁ RELACIONADA COM UMA ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDEM SUPERIOR HOA, ESTRUTURA DE DADOS QUE INCLUI DADOS DE CONTEÚDOS DE ÁUDIO ESPACIAIS EM 2D E 3D PARA UMA OU MAIS DESCRIÇÕES DE CORRENTES DE DADOS DE ÁUDIO DE HOA DIFERENTES. OS DADOS DE ÁUDIO DE HOA PODEM TER UMA ORDEM MAIOR DO QUE  $n$ , E A ESTRUTURA DE DADOS PODE INCLUIR ADICIONALMENTE DADOS DE UMA ÚNICA FONTE DE SINAL DE ÁUDIO E/OU DADOS DE ÁUDIO DE UM SISTEMA DE MICROFONES A PARTIR DE POSIÇÕES ESPACIAIS FIXAS OU VARIANTES NO TEMPO.

**RESUMO****"ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDEM SUPERIOR"**

A invenção está relacionada com uma estrutura de dados para dados de áudio Ambisonics de Ordem Superior HOA, estrutura de dados que inclui dados de conteúdos de áudio espaciais em 2D e 3D para uma ou mais descrições de correntes de dados de áudio de HOA diferentes. Os dados de áudio de HOA podem ter uma ordem maior do que "3", e a estrutura de dados pode incluir adicionalmente dados de uma única fonte de sinal de áudio e/ou dados de áudio de um sistema de microfones a partir de posições espaciais fixas ou variantes no tempo.

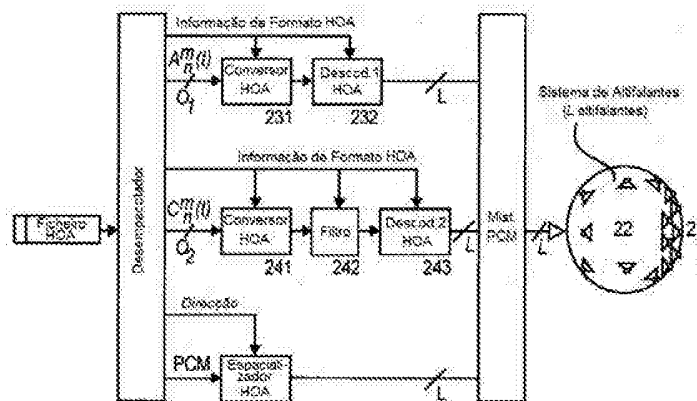


Fig. 2

## DESCRIÇÃO

### "ESTRUTURA DE DADOS PARA DADOS DE ÁUDIO AMBISONICS DE ORDEM SUPERIOR"

A invenção relaciona-se com uma estrutura de dados para dados de áudio de Ambisonics de Ordem Mais Elevadas, que incluem dados de conteúdos de áudio espacial de 2D e/ou 3D e que está também adequada para dados de áudio tendo uma ordem maior do que "3".

#### Antecedentes

O áudio em 3D pode ser realizado utilizando uma descrição de campo sonoro por uma técnica chamada Ambisonics de Ordem Superior (HOA) como descrito abaixo. O armazenamento de dados HOA requer algumas convenções e estipulações sobre como estes dados deverão ser utilizados por um decodificador especial para que este seja capaz de criar sinais para altifalante para a reprodução por uma dada configuração de altifalantes para reprodução. Nenhum formato de armazenamento existente define todas estas estipulações para HOA. O Formato-B (baseado na estrutura extensível "Riff/wav") com a sua realização de formato de ficheiro \*.amb como descrito a 30 de Março de 2009 por exemplo em Martin Leese, "File Format for B-Format", <http://www.ambisonia.com/Members/etienne/Members/mleese/fi>

le-format-for-b-format, é o formato mais sofisticado disponível actualmente. O formato de ficheiro .amb foi apresentado em 2000 por R. W. Dobson, "Developments in Audio File Formats", no ICMC Berlim 2000.

Em 16 de Julho de 2010, estava divulgada uma visão geral dos formatos de ficheiro existentes na página da Ambisonics Xchange: "Existing formats", <http://ambisonics.iem.at/xchange/format/existing-formats>, e está também divulgada nessa página uma proposta para um formato de permuta Ambisonics: "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>.

A Patente EP 2 205 007 A1 divulga codificação de material de áudio numa estrutura de dados atribuindo o material de áudio a dois grupos: Um primeiro grupo contém o áudio que necessita de uma localização altamente direccionada, um segundo grupo contém áudio para o qual é suficiente o Ambisonics de baixa ordem. Podem ser utilizados mais do que um conjunto de pistas de áudio no segundo grupo, por exemplo, para futuro re-versionamento (veja-se o parágrafo [0016], [0017] e [0051] da Patente EP 2 205 007 A1.

### Invenção

Relativamente aos sinais HOA, para 3D uma colecção de  $M = (N+1)^2$  ((2N+1) para 2D) objectos Áudio diferentes provenientes de diferentes fontes sonoras, todos com a mesma frequência, podem ser gravados (codificados) e reproduzidos como diferentes objectos sonoros desde que eles estejam uniformemente distribuídos espacialmente. Isto significa que um sinal Ambisonics de 1ª ordem pode conter quatro 3D ou três 2D objectos Áudio e estes objectos necessitam de estar separados uniformemente em torno de uma esfera para 3D ou em torno de um círculo em 2D. A sobreposição espacial e mais do que M sinais na gravação resultarão em indistinção - apenas os sinais mais fortes podem ser reproduzidos como objectos coerentes, os outros sinais difusos irão de alguma forma degenerar os sinais coerentes dependendo da sobreposição no espaço, frequência e similaridade na intensidade sonora.

Relativamente à situação acústica num cinema, é requerida uma elevada precisão na localização espacial do som para a área frontal do ecrã de forma a corresponder à cena visual. A percepção dos objectos sonoros envolventes é menos crítica (reverberação, objectos sonoros sem ligação à cena visual). Aqui, a densidade dos altifalantes pode ser menor em comparação com a área frontal.

A ordem de HOA dos dados de HOA, relevante para a área frontal, necessita de ser grande para permitir a

reprodução holofónica à escolha. Uma ordem típica é  $N = 10$ . Isto requer  $(N+1)^2 = 121$  coeficientes de HOA. Em teoria, podemos codificar também,  $M = 121$  objectos áudio, se estes objectos áudio estivessem uniformemente distribuídos espacialmente. Mas no nosso cenário, eles estão limitados à área frontal (porque apenas aqui necessitamos de tais ordens elevadas). De facto, podemos apenas codificar cerca de  $M = 60$  objectos Áudio sem indistinção (a área frontal é no máximo meia esfera de direcções, portanto  $M/2$ ).

Relativamente ao Formato-B mencionado acima, ela permite uma descrição apenas até uma ordem Ambisonics de 3, e a dimensão do ficheiro está restringido a 4 GB. Faltam outros itens de informação especiais, tais como o tipo de onda ou o raio de descodificação de referência que são vitais para descodificadores modernos. Não é possível utilizar diferentes formatos de amostragem (larguras de palavras) e larguras de banda para os diferentes componentes Ambisonics (canais). Também não há padronização para o armazenamento de informação suplementar e metadados para o Ambisonics.

Na arte conhecida, a gravação de sinais Ambisonics utilizando um sistema de microfones está restringido à ordem um. Isto poderá mudar no futuro se forem desenvolvidos protótipos experimentais de microfones HOA. Para a criação de conteúdo 3D uma descrição poderá ser gravada uma descrição do campo sonoro de ambiência utilizando um sistema de microfones em Ambisonics de

primeira ordem, pelo que as fontes direccionais são captadas utilizando microfones mono de proximidade ou microfones altamente direccionais conjuntamente com informação direccional (isto é, a posição da fonte). Os sinais direccionais podem então ser codificados numa descrição HOA, ou isto poderá ser realizado por um descodificador sofisticado. De qualquer forma, um novo formato de ficheiro Ambisonics necessita de ser capaz de armazenar mais do que uma descrição de campo sonoro de cada vez, mas parece que nenhum formato existente pode encapsular mais do que uma descrição Ambisonics.

Um problema a ser resolvido pela invenção é o de proporcionar um formato de ficheiro Ambisonics que é capaz de armazenar duas ou mais descrições de campo sonoro ao mesmo tempo, em que a ordem de Ambisonics pode ser maior do que 3. Este problema é solucionado pela estrutura de dados divulgada na Reivindicação 1, pelos métodos divulgados nas Reivindicações 9 e 10, e pelos aparelhos postos em evidência na Reivindicação 13.

Para recriar Áudio 3D realista, os descodificadores Ambisonics da próxima geração requererão tanto um conjunto grande de convenções e estipulações conjuntamente com dados armazenados a serem processados, como um único formato de ficheiro onde todos os parâmetros relacionados e elementos de dados possam ser armazenados coerentemente.

O formato de ficheiro da invenção para conteúdo sonoro espacial pode armazenar um ou mais sinais HOA e/ou sinais mono direccionais conjuntamente com informação direccional, em que ordens d Ambisonics maiores do que 3 e ficheiros >4 GB são exequíveis. Além disso, o formato de ficheiro da invenção proporciona elementos adicionais que os formatos existentes não oferecem:

1) Informação vital requerida para os descodificadores HOA da próxima geração é armazenada no formato de ficheiro:

Informação de onda de Ambisonics (plana, esférica, tipos mistos), região de interesse (fontes fora ou dentro da área de escuta), e raio de referência (para a descodificação de ondas esféricas).

Podem ser armazenados sinais mono direccionais relacionados. A informação de posição destes sinais direccionais pode ser descrita quer utilizando informação de ângulo e distância quer utilizando um vector de codificação de coeficientes Ambisonics.

2) Todos os parâmetros definindo os dados Ambisonics estão contidos na informação complementar, para assegurar clareza acerca da gravação:

Escala e normalização Ambisonics (SN3D, N3D, Furse Malham, Formato-B, ..., definido pelo utilizador), informação de ordem mista.



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

O formato da invenção permite armazenar dados relacionados com a ordem de Ambisonics (canais de Ambisonics) com diferentes resoluções de dimensão das palavras PCM bem como utilizando larguras de banda restringidas.

4) Metacampos que permitem armazenar informação acompanhante acerca do ficheiro tais como informação de gravação para sinais de microfone:

Sistema de coordenadas de referência de gravação, microfone, posições da fonte e virtual do ouvinte, características direccionais do microfone, informação sobre a sala e a fonte.

Este formato de ficheiro para conteúdo de áudio 2D e 3D cobre o armazenamento tanto de descrições Ambisonics de Ordem Superior (HOA) como fontes simples com posições fixas ou variantes no tempo, e contém toda a informação que permite aos descodificadores de áudio de próxima geração proporcionarem Áudio 3D realista.

Utilizando definições apropriadas, o formato de ficheiro da invenção é também adequado para a produção de correntes de bits com conteúdo áudio. Assim, pode ser

enviada informação complementar dependente do conteúdo (dados de cabeçalho) em posições no tempo conforme o seleccionado pelo criador do ficheiro. O formato de ficheiro da invenção serve também como uma descrição de cena onde pistas de uma cena de áudio podem começar e terminar em qualquer momento.

Em princípio, a estrutura de dados da invenção está adequada para dados de áudio Ambisonics de Ordem Superior HOA, cuja estrutura de dados inclua dados de conteúdo de áudio espacial 2D e/ou 3D para uma ou mais diferentes descrições para correntes de bits de dados de áudio de HOA, e cuja estrutura de dados é também adequada para dados de áudio de HOA que tenham uma ordem maior do que "3", e cuja estrutura de dados possa adicionalmente incluir dados de uma única fonte de sinal de áudio e/ou dados de áudio de um sistema de microfones provenientes de posições fixas ou variantes no tempo.

Em princípio, o método da invenção é adequado para apresentações de áudio, em que uma corrente de bits de dados de áudio de HOA contendo pelo menos dois diferentes sinais de dados de áudio de HOA é recebida e pelo menos um primeiro deles é utilizado para a apresentação com uma configuração de altifalantes densa localizada numa área distinta de um local da apresentação, e pelo menos um segundo e diferente deles é utilizado para a apresentação com um sistema de altifalantes menos denso envolvendo o referido local da apresentação.

Estão divulgadas formas de realização adicionais vantajosas da invenção nas respectivas Reivindicações dependentes.

### **Desenhos**

Estão descritas formas de realização exemplificadoras da invenção com referência aos desenhos acompanhantes, que apresentam na:

Figura 1 - representação holofónica em cinema com sistema denso de altifalantes na região frontal e densidade inferior de altifalantes em torno da área de audição;

Figura 2 - sistema de descodificação sofisticado;

Figura 3 - criação de conteúdo HOA a partir de gravação com sistema de microfones, gravação com fonte única, geração de campo sonoro simples e complexo;

Figura 4 - criação de conteúdo imersivo de próxima geração;

Figura 5 - descodificação 2D de sinais HOA para arranjo simples de altifalantes envolventes, e descodificação 3D de sinais HOA para um arranjo de altifalantes holofónico para o palco frontal e um menos denso arranjo 3D de altifalantes de envolvimento;

Figura 6 - problema do domínio interior, em que as fontes estão fora da região de interesse/validade;

Figura 7 - definição de coordenadas esféricas;

Figura 8 - problema do domínio exterior, em que as fontes estão dentro da região de interesse/validade;

Figura 9 - exemplo simples de formato de ficheiro HOA;

Figura 10 - exemplo para um ficheiro HOA contendo múltiplas tramas com múltiplas pistas;

Figura 11 - ficheiro HOA com múltiplos Grupos de Meta Dados;

Figura 12 - processamento da codificação da Região da Pista;

Figura 13 - processamento da descodificação da Região da Pista;

Figura 14 - Implementação da Redução de Largura de Banda utilizando o processamento MDCT;

Figura 15 - Implementação da Reconstrução da Largura de Banda utilizando o processamento MDCT.

### Formas de Realização Exemplificadoras

Com a expansão crescente do vídeo 3D, as técnicas de áudio imersivo estão a tornar-se uma característica interessante para a diferenciação. O Ambisonics de Ordem Superior (HOA) é uma destas tecnologias que podem proporcionar uma forma de introduzir o Áudio 3D de uma forma incremental nos cinemas. Utilizando pistas sonoras HOA e descodificadores HOA, um cinema pode começar com sistemas de altifalantes de áudio de envolvimento existentes e investir em mais altifalantes passo-a-passo, melhorando a experiência imersiva com cada passo.

A Figura 1a apresenta a reprodução holofónica no cinema com sistemas densos de altifalantes 11 na região frontal e menor densidade de altifalantes 12 em torno da área de audição ou dos lugares sentados 10, proporcionando uma forma de reprodução precisa de sons relacionados com a acção visual e com suficiente precisão dos sons ambientes reproduzidos.

A Figura 1b apresenta a direcção percebida da chegada de ondas sonoras frontais reproduzidas, em que a direcção de chegada de ondas planas corresponde a diferentes posições no ecrã, isto é, as ondas planas são adequadas para reproduzirem a profundidade.

A Figura 1c apresenta a direcção percebida da chegada de ondas esféricas reproduzidas, o que conduz a

melhor consistência na direcção percebida do som e da acção visual a 3D em torno do ecrã.

A necessidade de duas diferentes correntes de bits de HOA é causada pelo facto de que a acç visual principal num cinema tem lugar na região frontal aos ouvintes. Também, a precisão perceptiva na detecção da direcção de um som é mais elevada para fontes sonoras frontais do que para fontes envolventes. Portanto, a precisão da reprodução do som espacial frontal necessita de ser maior do que a precisão espacial para os sons ambientes reproduzidos. Holofónico significa que, para a reprodução sonora, são necessários um elevado número de altifalantes, um descodificador dedicado e alimentadores de altifalantes correspondentes, para a região frontal do ecrã, enquanto é necessária tecnologia menos dispendiosa para a reprodução do som ambiente (menor densidade de altifalantes envolvendo a área de audição e tecnologia de descodificação menos perfeita).

Devido às tecnologias de criação de conteúdo e de reprodução do som, é vantajoso disponibilizar uma representação de HOA para os sons de ambiente e uma representação de HOA para os sons de acção do primeiro plano, cf. a Figura 4. Um cinema utilizando um arranjo simples com um equipamento simples de reprodução sonora inferior pode misturar ambas as correntes de bits antes da descodificação (cf. Figura 5, parte superior). Um cinema mais sofisticado equipado com meios de reprodução

completamente imersivos pode utilizar dois descodificadores - um para descodificar os sons ambientes e um descodificador especializado para o posicionamento de elevada precisão de fontes sonoras virtuais para a acção principal no plano principal, como se apresenta no sistema de descodificação sofisticado na Figura 2 e na parte inferior da Figura 5. Um ficheiro HOA especial contém pelo menos duas pistas que representam campos sonoros HOA para sons ambientes  $A_n^m(t)$  e para sons frontais relacionados com a acção principal visual  $C_n^m(t)$ . Podem ser disponibilizadas correntes de bits opcionais para efeitos direccionais. Dois sistemas de descodificadores correspondentes conjuntamente com um espacializador proporcionam sinais para um sistema frontal denso holofónico 3D de altifalantes 21 e para um sistema de envolvimento 3D menos denso (isto é, inferior) 22. O sinal de dados HOA da corrente de bits da Pista 1 representa os sons de ambiente e é convertido num conversor de HOA 231 para entrada para um Descodificador1 232 especializado para a reprodução da ambiência. Para a corrente de bits da Pista 2, os dados de sinal HOA (sons frontais relacionados com a cena visual) são convertidos num conversor de HOA 241 para entrada num filtro 242 corrigido para a distância (Eq.(26)) para a melhor localização de fontes sonoras esféricas em torno da área do ecrã com um Descodificador2 dedicado 243. As correntes de bits de dados direccionais são espacializadas directamente para L altifalantes. Os três sinais de altifalante são misturados por PCM para a reprodução conjunta com o sistema de altifalantes 3D.

Parece que não existe um formato de ficheiro conhecido dedicado a tal cenário. As gravações de campos sonoros 3D conhecidas utilizam quer descrições de cenas completas com pistas sonoras relacionadas, quer uma única descrição de campo sonoro quando armazenada para posterior reprodução. Os exemplos do primeiro tipo são os formatos WFS (Síntese de Campo de Ondas) e numerosos formatos de contentores. Os exemplos para o segundo tipo são os formatos Ambisonics tais como os formatos B ou AMB, veja-se o artigo mencionado acima "File Format for B-Format". O último restringe a ordens de Ambisonics de três, um formato de transmissão fixo, um modelo de descodificador fixo e campos sonoros únicos.

### ***Criação e Reprodução de Conteúdos HOA***

O processamento para gerar descrições de campos sonoros HOA está descrito na Figura 3. Na Figura 3a, são criadas gravações naturais de campos sonoros utilizando sistemas de microfones. Os sinais das cápsulas são sujeitos à aplicação de matrizes e igualizados de maneira a formar sinais HOA. Os sinais de ordem superior (Ambisonics de ordem  $>1$ ) são usualmente sujeitos a um filtro passa-banda para reduzir artefactos devidos a efeitos de distância da cápsula: sujeitos a um filtro passa-baixas para reduzir o serrilhamento espacial a altas frequências, e sujeito a um filtro passa-altos para reduzir níveis excessivos de baixa frequência com a ordem  $n$  de Ambisonics crescente ( $h_n(kr_{d\_mic})$ , veja-se a Eq.(34). Opcionalmente pode ser



aplicada a filtragem de codificação de distância, vejam-se as Eqs.(25) e (27). Antes do armazenamento, é adicionada informação de formato HOA ao cabeçalho da pista.

As representações de campos sonoros artísticos são geralmente criadas utilizando múltiplas correntes de bits de fontes únicas direccionais. Como se apresenta na Figura 3b, um sinal de fonte única pode ser capturado como uma gravação PCM. Isto pode ser efectuado por microfones de proximidade ou utilizando microfones com alta direccionalidade. Adicionalmente, são gravados os parâmetros direccionais ( $r_s$ ,  $\theta_s$ ,  $\phi_s$ ) da fonte sonora em relação a uma melhor posição de audição virtual (sistema de coordenadas HOA, ou qualquer ponto de referência para posterior mapeamento). A informação de distância pode também ser criada colocando artisticamente sons quando se geram cenas a partir de filmes. Como se apresenta na Figura 3c, a informação direccional ( $\theta_s$ ,  $\phi_s$ ) é então utilizada para criar o vector de codificação  $\Psi$ , e o sinal da fonte direccional é codificado num sinal Ambisonics, veja-se a Eq.(18). Isto é equivalente a uma representação de uma onda plana. Um processo de filtragem de terminação pode utilizar a informação de distância  $r_s$  para inculcar 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 ao cabeçalho da pista.

São geradas descrições de campos de ondas mais complexas pela mistura pelo HOA de sinais Ambisonics como representado na Figura 3d. Antes do armazenamento, a informação do formato HOA é adicionada ao cabeçalho da pista.

O processo de geração de conteúdo para cinema 3D está representado na Figura 4. Os sons frontais relacionados com a acção visual são codificados com precisão espacial e misturados a um sinal HOA (campo de onda)  $C_n^m(t)$  e são armazenados como Pista 2. Os codificadores envolvidos codificam com uma elevada precisão espacial e tipos de onda especiais necessários para a melhor correspondência com a cena visual. A Pista 1 contém o campo sonoro  $A_n^m(t)$  que está relacionado com sons ambientes codificados sem restrição da direcção da fonte. Usualmente, a precisão espacial dos sons ambientes não necessita de ser elevada como para os sons frontais (consequentemente, a ordem de Ambisonics pode ser mais pequena) e a modelação do tipo de onda é menos crítica. O campo sonoro de ambiente pode também incluir partes reverberantes dos sinais do som frontal. Ambas as pistas são multiplexadas para armazenamento e/ou permuta. Opcionalmente, podem ser multiplexados sons direccionais (por exemplo Pista 3) no ficheiro. Estes sons podem ser sons de efeitos especiais, diálogos ou informação desportiva tal como um discurso narrativo para deficientes visuais.

A Figura 5 apresenta os princípios da descodificação. Como representado na parte superior, um cinema com um arranjo de altifalantes pouco denso pode misturar ambos os sinais HOA da Pista 1 e Pista 2 antes da descodificação HOA simplificada, e pode truncar a ordem da Pista 2 e reduzir a dimensão de ambas as pistas para 2D. No caso de estar presente uma corrente de bits direccional, é codificada para HOA 2D. Então, todas as três correntes de bits são misturadas para formarem uma única representação HOA que é então descodificada e reproduzida. A parte inferior corresponde à Figura 2. Um cinema equipado com um sistema holofónico para o palco frontal e um sistema de envolvimento 3D menos denso utilizará descodificadores sofisticados dedicados e misturará as alimentações dos altifalantes. Para a corrente de bits de dados da Pista 1, os dados HOA representando os sons de ambiente são convertidos para o Descodificador1 especializado para a reprodução de ambiência. Para a corrente de bits de dados da Pista 2, o HOA (sons frontais relacionados com a cena visual) é convertido e corrigido para a distância (Eq.(26)) para a melhor localização de fontes sonoras esféricas em torno da área do ecrã com um Descodificador2 dedicado. As correntes de bits de dados direccionais são directamente especializadas para L altifalantes. Os três sinais de altifalante são misturados por PCM para a reprodução conjunta com o sistema de altifalantes 3D.

***Descrições de campo sonoro utilizando Ambisonics de Ordem Superior***

**Descrição de campo sonoro utilizando Harmônicas Esféricas (SH)**

Quando se utilizam descrições Harmônicas esféricas/Bessel, a solução da equação de onda acústica é proporcionada na Eq.(1), veja-se M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics", Journal of Audio Engineering Society, 53(11), pág. 1004-1025, Novembro de 2005, e Earl G. Williams, "Fourier Acoustics", Academic Press, 1999. A pressão sonora é uma função das coordenadas esféricas  $r$ ,  $\Theta$ ,  $\Phi$  (veja-se a Figura 7 para a sua definição) e da frequência espacial  $k = \frac{w}{c} = \frac{2\pi f}{c}$ . A descrição é válida para fontes sonoras de áudio fora da região de interesse ou de validade (problema do domínio interior, como se apresenta na Figura 6) e assume Harmônicas Esféricas ortogonais normalizadas:

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

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 chamadas Harmônicas Esféricas (SH),  $n$  é o índice de ordem de Ambisonics, e  $m$  indica o grau.

Devido à natureza da função de Bessel que tem valores significativos apenas para valores de  $kr$  pequenos (pequena distância da origem ou frequências baixas), a

série pode ser parada em alguma ordem  $n$  e restringida a um valor  $N$  com precisão suficiente. Quando se armazenam dados HOA, usualmente os coeficientes de Ambisonics  $A_n^m$ ,  $B_n^m$  ou alguns derivados (estão descritos detalhes abaixo) são armazenados até essa ordem  $N$ .  $N$  é chamado a ordem Ambisonics.  $N$  é chamado a ordem Ambisonics, e o termo "ordem" é usualmente também utilizado em combinação com o  $n$  nas funções de Bessel  $j_n(kr)$  e de Hankel  $h_n(kr)$ .

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 Figura 8, é expressa para  $r > r_{\text{Fonte}}$  na Eq. (2):

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

Os  $B_n^m(k)$  são novamente chamados coeficientes de Ambisonics e  $h_n^{(1)}(kr)$  denomina a função esférica de Hankel do primeiro tipo e de ordem  $n^{\text{a}}$ . A fórmula assume SH ortogonais normalizadas. Nota: Geralmente a função esférica de Hankel do 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 de entrada (relacionadas com  $e^{-ikr}$ ), veja-se o livro mencionado acima "Fourier Acoustics".

### **Harmónicas Esféricas**

As harmónicas esféricas  $Y_n^m$  podem ter valores tanto complexos como reais. O caso geral para HOA utiliza harmónicas esféricas de valores reais. Uma descrição unificada de Ambisonics utilizando harmónicas esféricas reais e complexas pode ser revista em Mark Poletti, "Unified description of Ambisonics using real and complex spherical harmonics", Proceedings of the Ambisonics Symposium 2009, Gras, Áustria, Junho de 2009.

Existem diferentes maneiras de normalizar as harmónicas esféricas (o que é independente das harmónicas esféricas serem reais ou complexas), vejam-se as seguintes páginas na internet a respeito de harmónicas esféricas (reais), e esquemas de normalização:

<http://www.ipgp.fr/~wiecsor/SHTTOOLS/www/conventions.html>,

[http://en.citiseidum.org/wiki/Spherical\\_harmonics](http://en.citiseidum.org/wiki/Spherical_harmonics).

A normalização corresponde à relação de ortogonalidade entre  $Y_n^m$  e  $Y_{n'}^{m'}$ .

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

Nota:

em que  $S^2$  é a esfera unitária e delta de Kroneker  $\delta_{aa'}$  é igual a 1 para  $a=a'$ , igual a 0 noutro caso. As harmónicas esféricas complexas são descritas por:

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

em que  $i = \sqrt{-1}$  e  $s_m = \begin{cases} (-1)^m & m > 0 \\ 1 & \text{caso contrário} \end{cases}$  para um sinal alternante para m positivo tal como no livro "Fourier Acoustics" mencionado acima. (Nota: o  $s_m$  é um termo convencional e pode ser omitido para SH apenas positivas).  $N_{n,m}$  é um termo de normalização que toma forma para uma representação ortogonal normalizada (! designa factorial):

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

Abaixa, o Quadro 1 apresenta alguns esquemas de normalização comumente utilizados para as harmónicas esféricas de valores complexos.  $P_{n,|m|}(x)$  são as funções de Legendre associadas, em que 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 por vezes está incluído na representação de  $P_n^m$  noutras 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 Rodrigues como:

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

<b><math>N_{n,m}</math></b> Esquemas comuns de normalização para SH complexas			
Não	Schmidt	$4\pi$ normalizado,	Orto-normalizado

normalizado	semi-normalizado, SN3D	N3D, geodesia 4π	
1	$\sqrt{\frac{(n -  m )!}{(n +  m )!}}$	$\sqrt{\frac{(2n + 1)(n -  m )!}{(n +  m )!}}$	$\sqrt{\frac{(2n + 1)(n -  m )!}{4\pi (n +  m )!}}$

Quadro 1 - Factores de normalização para Harmónicas esféricas de valores complexos

Numericamente, é vantajoso derivar  $P_{n, |m|}(x)$  de forma progressiva a partir de uma relação de recorrência, veja-se William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery, "Numerical Recipes in C", Cambridge University Press, 1992. As funções de Legendre associadas até n=4 estão dadas no Quadro 2:

n \ m	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		$P_1^1(\cos\theta) = \sin\theta$	$P_2^1(\cos\theta) = 3\cos\theta\sin\theta$	$P_3^1(\cos\theta) = \frac{3}{2}(5\cos^2\theta - 1)\sin\theta$	$P_4^1(\cos\theta) = \frac{3}{2}(7\cos^3\theta - 3\cos\theta)\sin\theta$
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					$P_4^4(\cos\theta) = 105\sin^4\theta$

Quadro 2 - Alguns dos primeiros Polinómios de Legendre

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

As SH de valores reais são derivadas combinando o complexo conjugado  $Y_n^m$  correspondente a valores opostos de m (o termo  $(-1)^m$  na definição (6) é introduzido para se obterem expressões sem sinal para a SH real, o que é o caso



usual em Ambisonics):

$$S_n^m(\theta, \phi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}} (Y_n^m + Y_n^{m*}) & = \Theta_n^m(\theta) \sqrt{2} \cos(m\phi), & m > 0 \\ Y_n^0 & = \Theta_n^0(\theta), & m = 0 \\ \frac{(-1)^m}{i\sqrt{2}} (Y_n^{|m|} - Y_n^{|m|*}) & = \Theta_n^{|m|}(\theta) \sqrt{2} \sin(|m|\phi), & m < 0 \end{cases} \quad (6)$$

a qual pode ser reescrita como a Eq.(7) para salientar a ligação a harmónicas circulares com  $\Phi_m(\phi) = \Phi_{n=|m|}^m(\phi)$  mantendo apenas o termo de azimute:

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

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

O número total de componentes esféricas  $S_n^m$  para uma dada ordem N de Ambisonics é igual a  $(N+1)^2$ . Estão dados no Quadro 3 esquemas de normalização comuns das harmónicas esféricas de valores reais.

$\tilde{N}_{n,m}$ , Esquemas de normalização comuns para SH reais			
Não normalizado	Schmidt semi-normalizado, SN3D	$4\pi$ normalizado, N3D, geodesia $4\pi$	Orto-normalizado

$\sqrt{2 - \delta_{0,m}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(n -  m )!}{(n +  m )!}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(2n + 1)(n -  m )!}{(n +  m )!}}$	$\sqrt{(2 - \delta_{0,m}) \frac{(2n + 1)(n -  m )!}{4\pi (n +  m )!}}$
---------------------------	---	---	--

Quadro 3 - Esquemas de normalização de SH reais 3D,  $\delta_{0,m}$  tem um valor de 1 para  $m=0$  e 0 caso contrário

### *Harmónicas Circulares*

Para representações bidimensionais, apenas é necessário um subconjunto de harmónicas. O grau das SH pode apenas tomar valores de  $m \in \{-n, n\}$ . O número total de componentes para um dado  $N$  reduz-se para  $2N+1$  porque os componentes representando a inclinação  $\theta$  tornam-se obsoletos e as harmónicas esféricas podem ser substituídas pelas harmónicas circulares dadas na Eq.(8). Existem vários esquemas de normalização diferentes  $N_m$  para harmónicas circulares, que precisam de ser considerados quando se convertem coeficientes Ambisonics 3D para coeficientes 2D. A fórmula mais geral para harmónicas circulares torna-se:

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

São disponibilizados no Quadro 4 alguns factores de normalização comuns para as harmónicas circulares, em que o termo de normalização é introduzido pelo factor antes do termo horizontal  $\phi_m(\phi)$ :

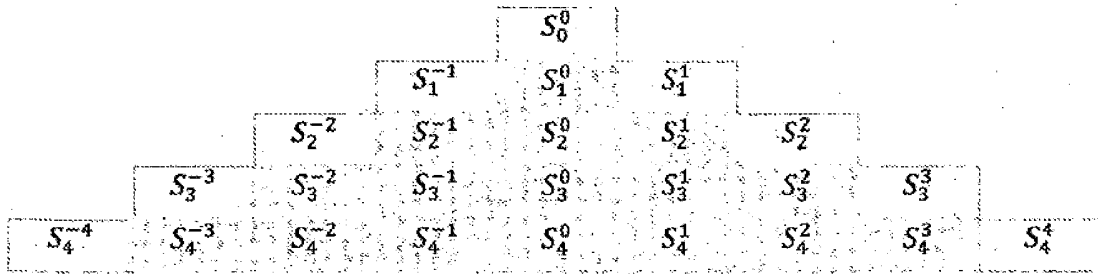
$N_m$ , Esquemas de normalização comuns para Harmónicas
---

Circulares			
Não normalizado	SN2D	2D Normalizado, N2D	Orto-normalizado
$\sqrt{\frac{2 - \delta_{0,m}}{2}}$	1	$\sqrt{2 - \delta_{0,m}}$	$\sqrt{(2 - \delta_{0,m}) \frac{1}{2\pi}}$

Quadro 4 - Esquemas de normalização de CH 2D,  $\delta_{0,m}$  tem um valor de 1 para  $m=0$  e 0 caso contrário

A conversão entre diferentes normalizações é fácil. Em geral, a normalização tem um efeito sobre a notação descrevendo a pressão (vejam-se as Eqs.(1), (2)) e todas as considerações derivadas. O tipo de normalização também influencia os coeficientes Ambisonics. Há também pesos que podem ser aplicados para redimensionar estes coeficientes, por exemplo pesos de Furse-Malham (FuMa) aplicados a coeficientes Ambisonics quando se armazena um ficheiro utilizando o formato AMB.

Em relação à conversão 2D-3D, a conversão de CH para SH e vice-versa pode também ser aplicada a coeficientes Ambisonics, por exemplo quando se descodifica uma representação Ambisonics 3D (gravação) com um descodificador 2D para um sistema de altifalantes 2D. A relação entre  $S_n^m$  e  $\check{\Phi}_{n=|m|}^m$  para a conversão 3D-2D está representada no esquema seguinte para uma ordem de Ambisonics de até 4:



O factor de conversão de 2D para 3D pode ser derivado para o plano horizontal em  $\theta = \frac{\pi}{2}$  como se segue:

$$\frac{\alpha_{2D}}{3D} = \frac{S_{n=m}^m(\theta=\pi/2, \Phi)}{\Phi_{n=|m|}^m(\phi)} = \frac{N_{m,m} (2m)!}{N_m m! 2^m} \tag{10}$$

A conversão de 3D para 2D utiliza  $\frac{1}{\alpha_{2D}^{3D}}$ . Os detalhes estão apresentados em ligação com as Eqs. (28) (29) (30) abaixo. Uma conversão 2D normalizada para ortogonal normalizada torna-se:

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

**Coefficientes Ambisonics**

Os coeficientes Ambisonics têm a escala de unidades da pressão sonora:  $1Pa = 1 \frac{N}{m^2} = 1 \frac{kg m}{s^2 m^2}$ . Os coeficientes Ambisonics formam o sinal Ambisonics e, em geral, são uma função de tempo discreto. O Quadro 5 apresenta a relação entre representação dimensional, ordem N de Ambisonics e número de coeficientes Ambisonics (canais):

Dimensão	Número 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>

Quadro 5 - Número de Coeficientes Ambisonics

Quando se opera com representações em tempo discreto, os coeficientes Ambisonics são, usualmente, armazenados de uma forma interlaçada tal como as representações de canal em PCM para gravações multicanal (canal=coeficiente Ambisonics  $A_n^m$  da amostra  $v$ ), sendo a sequência de coeficientes uma questão de convenção. Um exemplo para 3D, N=2 é:

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

e para 2D, N=2:

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

O sinal  $A_0^0(n)$  pode ser visto como uma representação mono da gravação Ambisonics, não tendo informação direccional mas sendo uma representante da impressão de timbre geral da gravação.

A normalização dos coeficientes Ambisonics é geralmente realizada de acordo com a normalização da SH (como se tornará evidente abaixo, veja-se a Eq.(15)), que

deverá ser tida em conta quando se descodifica uma gravação externa ( $A_n^m$  estão baseados na SH com o factor de normalização  $N_{n,m}$ ,  $\check{A}_n^m$  estão baseados na SH com o factor de normalização  $\check{N}_{n,m}$ ):

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

que se torna  $A_{N3D_n}^m = \sqrt{(2n+1)} \check{A}_{SN3D_n}^m$  para o caso SN3D para N3D.

O Formato B e o formato AMB utilizam pesos adicionais (pesos Gerson, Furse-Malham (FuMa), MaxN) que são aplicados aos coeficientes. A normalização de referência é então usualmente SN3D, veja-se Jérôme Daniel, "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia", Tese de PhD, Universidade de Paris 6, 2001, e Dave Malham, "3-D acoustics space and its simulation using ambisonics", <http://www.dxarts.washington.edu/courses/567/current/malham3d.pdf>.

As duas seguintes realizações específicas das equações de onda para ondas planas ou ondas esféricas ideais apresentam mais detalhes acerca dos coeficientes Ambisonics:

### ***Ondas Planas***

Resolvendo a equação de onda para ondas planas,

$A_n^m$  torna-se independente de  $k$  e  $r_s$ ;  $\theta_s, \phi_s$  descrevem os ângulos das fontes, '\*' designa o complexo conjugado:

$$A_n^m \text{ plana}(\theta_s, \phi_s) = 4\pi i^n P_{S_0} Y_n^m(\theta_s, \phi_s)^* = 4\pi i^n d_n^m(\theta_s, \phi_s) \quad (15)$$

Aqui  $P_{S_0}$  é utilizado para descrever a pressão do sinal de redimensionamento da fonte medida na origem do sistema de coordenadas descritor que pode ser uma função do tempo e se torna  $A_{\text{plana}}^0 / \sqrt{4\pi}$  para harmónicas esféricas ortogonais normalizadas. Geralmente, o Ambisonics assume ondas planas e são transmitidos ou armazenados os coeficientes

$$d_n^m(\theta_s, \phi_s) = \frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n} = P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (16)$$

Esta assunção oferece a possibilidade de sobreposição de sinais direccionais diferentes bem como um único desenho de decodificador. Isto é também verdade para sinais de um microfone Soundfield<sup>TM</sup> gravados no Formato B de primeira ordem (N=1), o que se torna óbvio quando se compara a progressão de fase dos filtros de igualização (para progressão teórica, veja-se o artigo mencionado acima "Unified description of Ambisonics using real and complex spherical harmonics", capítulo 2.1, e para um progressão protegida por patente veja-se a Patente dos Estados Unidos US 4042779. A Eq.(1) torna-se:

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

Os coeficientes  $d_n^m$  podem tanto ser derivados por sinais de sistemas de microfones pós-processados como podem ser criados sinteticamente utilizando um sinal mono  $P_{s_0}(t)$  caso em que as harmónicas esféricas direccionais  $Y_n^m(\theta_s, \phi_s, t)^*$  podem também ser dependentes do tempo (fonte móvel). A Eq.(17) é válida para cada instância de amostragem temporal  $v$ . O processo da codificação sintética pode ser reescrito (para cada instância de amostragem  $v$ ) na forma de vector/matriz para uma ordem  $N$  de Ambisonics seleccionada:

$$\mathbf{d} = \Psi P_{s_0} \quad (18)$$

em que  $d$  é um sinal Ambisonics, mantendo  $d_n^m(\theta_s, \phi_s)$ , (exemplo para  $N=2$ :  $\mathbf{d}(t) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]^T$ ), dimensão( $d$ ) =  $(N+1)^2 \times 1 = 0 \times 1$ ,  $P_{s_0}$  é a pressão do sinal fonte na origem de referência, e  $\Psi$  é o vector de codificação, mantendo  $Y_n^m(\theta_s, \phi_s)^*$ , dimensão( $\Psi$ ) =  $0 \times 1$ . O vector de codificação pode ser derivado a partir das harmónicas esféricas para a direcção da fonte específica  $\theta_s, \phi_s$  (igual à direcção da onda plana).

### **Ondas Esféricas**

Os coeficientes Ambisonics descrevendo ondas esféricas de chegada geradas por fontes pontuais (fontes de campo próximo) para  $r < r_s$  são:



$$A_n^m(k, \theta_s, \phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{s_0} Y_n^m(\theta_s, \phi_s)^* \quad (19)$$

Esta equação é derivada em ligação com as Eqs.(31) a (36) abaixo.  $P_{s_0} = p(0|r_s)$  descreve a pressão sonora na origem e, novamente, torna-se idêntica a  $A_0/\sqrt{4\pi}$ ,  $h_n^{(2)}$  é a função esférica de Hankel do segundo tipo e ordem  $n$ , e  $h_0^{(2)}$  é a função esférica de Hankel de ordem zero do segundo tipo. A Eq.(19) é similar aos ensinamentos em Jérôme Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonics format", AES 23<sup>a</sup> Conferência Internacional, Dinamarca, Maio de 2003. Aqui  $\frac{h_n(kr_s)}{h_0(kr_s)} = i^n \sum_{a=0}^n \frac{(n+a)!}{(n-a)!a!} \left(-\frac{ic}{2r_s\omega}\right)^a$ , btw  $\frac{h_1(kr_s)}{h_0(kr_s)} = i\left(1 - \frac{ic}{r_s\omega}\right)$  que, tendo a Eq.(11) em mente, pode ser encontrado em M. A. Gerson, "General metatheory of auditory localisation", 92<sup>a</sup> Convenção AES, 1992, Pré-impressão 3306, onde Gerson descreve o efeito de proximidade para sinais do primeiro grau.

A criação sintética de sinais Ambisonics esféricos é menos comum para ordens mais elevadas de Ambisonics N porque as respostas em frequência de  $\frac{h_n(kr_s)}{h_0(kr_s)}$  são difíceis de manusear numericamente para frequências baixas. Estes problemas numéricos podem ser ultrapassados considerando um modelo esférico para a descodificação/reprodução como descrito abaixo.

### Reprodução do campo sonoro

### Descodificação de Onda Plana

Em geral, o Ambisonics assume uma reprodução do campo sonoro por L altifalantes que estão uniformemente distribuídos num círculo ou numa esfera. Quando se assume que os altifalantes estão posicionados suficientemente afastados da posição de audição, é válido um modelo de descodificação de onda plana no centro ( $r_s > \lambda$ ). A pressão sonora gerada por L altifalantes é descrita por:

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

com  $w_l$  sendo o sinal para o altifalante  $l$  e tendo a unidade de escala de uma pressão sonora, 1 Pa.  $w_l$  é frequentemente chamado de função de alimentação do altifalante  $l$ .

É desejável que esta Eq.(20) de pressão sonora seja idêntica à pressão descrita pela Eq.(17). Isto conduz a:

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

Isto pode ser reescrito na forma de matriz, conhecida como "fórmula de re-codificação" (compare-se com a Eq.(18)):  $\mathbf{d} = \Psi_y$  (22)

em que  $\mathbf{d}$  é um sinal Ambisonics, mantendo  $\mathbf{d}_n^m(\theta_s, \phi_s)$  ou  $\frac{A_n^m(\theta_s, \phi_s)}{4\pi i^n}$ , (exemplo para N=2:  $\mathbf{d}(n) = [d_0^0, d_1^{-1}, d_1^0, d_1^1, d_2^{-2}, d_2^{-1}, d_2^0, d_2^1, d_2^2]^T$ ),

dimensão ( $\mathbf{d}$ ) =  $(N+1)^2 \times 1 = 0 \times 1$ ,  $\Psi$  é a matriz (de re-codificação), mantendo  $Y_n^m(\theta_l, \phi_l)^*$ , dimensão( $\Psi$ ) =  $0 \times L$ , e  $\mathbf{y}$  são os sinais dos altifalantes  $w_l$ , dimensão( $\mathbf{y}(n)$ , 1) = L.

Pode então ser derivado  $\mathbf{y}$  utilizando um par de métodos conhecidos, por exemplo correspondência de modo, ou por métodos que otimizam em relação a funções de espacialização de altifalantes especiais.

### **Descodificação para o modelo de onda esférica**

Um modelo de descodificação mais geral assume novamente altifalantes igualmente distribuídos em torno da origem com uma distância  $r_l$  do ponto radiante tal como as ondas esféricas. Os coeficientes Ambisonics  $A_n^m$  são dados pela descrição geral a partir da Eq.(1) e a pressão sonora gerada por L altifalantes é dada de acordo com a Eq.(19):

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

Um descodificador mais sofisticado pode filtrar os coeficientes Ambisonics  $A_n^m$  de forma a recuperar  $C_n^m = A_n^m \frac{h_0(kr_l)}{4\pi h_n(kr_l)}$  e consequentemente aplicar a Eq.(17) com  $\mathbf{c} = [C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots]'$  para derivar os pesos dos altifalantes. Com este modelo, os sinais dos altifalantes  $w_l$  são determinados pela pressão na origem. Existe uma abordagem alternativa que utiliza a abordagem de fonte única descrita a primeira vez no artigo mencionado acima "Three-dimensional surround sound systems based on

spherical harmonics". É assumido que os altifalantes estejam igualmente distribuídos sobre a esfera e que tenham características de fonte secundária. A solução é derivada em Jens Ahrens, Sascha Sports, "Analytical driving functions for higher order ambisonics", Proceedings of the ICASSP, páginas 373-376, 2008, Eq.(13), que pode ser reescrita para truncatura em Ambisonics de ordem N e um ganho dos altifalantes de  $g_l$  como uma generalização:

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

### **Sinais Ambisonics Codificados com Distância**

Criar  $C_n^m$  no codificador Ambisonics utilizando uma distância ao altifalante de referência  $r_{l,ref}$  pode resolver problemas numéricos de  $A_n^m$  quando se modelam ou gravam ondas esféricas (utilizando a Eq.(18)):

$$C_n^m = A_n^m \frac{h_0(k r_{l,ref})}{4\pi h_n(k r_{l,ref})} = \frac{h_0(k r_{l,ref})}{h_n(k r_{l,ref})} \frac{h_n(k r_s)}{h_0(k r_s)} P_{S_0} Y_n^m(\theta_s, \phi_s)^* \quad (25)$$

São transmitidos ou gravados  $C_n^m$ , a distância de referência  $r_{l,ref}$  e um indicador de que são utilizados coeficientes codificados com distância esférica. No lado do decodificador, é realizável um processamento de descodificação simples como o dado na Eq.(22) desde que a distância do altifalante real  $r_l \approx r_{l,ref}$ . Se esta diferença for muito grande, é requerida uma correcção

$$D_n^m = C_n^m \frac{h_n(k r_{l,ref})}{h_n(k r_l)} \quad (26)$$

por filtragem antes da descodificação Ambisonics.

Outros modelos de descodificação tais como a Eq.(24) resultam em diferentes formulações para Ambisonics codificado com distância:

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

Também a normalização das Harmónicas Esféricas pode ter uma influência sobre a formulação de Ambisonics codificado com distância, isto é, os coeficientes de Ambisonics Codificado com Distância necessitam de um contexto definido.

Os detalhes para a conversão 2D-3D mencionada acima são como se segue:

O factor de conversão  $\frac{\alpha_{2D}}{3D}$  para converter uma componente circular de 2D numa componente esférica de 3D por multiplicação, pode ser derivado como se segue:

$$\frac{\alpha_{2D}}{3D} = \frac{S_{n=m}^m(\theta=\pi/2, \Phi)}{\Phi_{n=\{m\}}^m(\Phi)} = \frac{N_{m,m} P_{\{m\},\{m\}}(\cos(\theta=\pi/2)) \Phi_m(\Phi)}{N_m \Phi_m(\Phi)} \quad (28)$$

Utilizando a identidade comum (veja-se Wikipedia de 12 de Outubro de 2010, "Associated Legendre polynomials",

[http://en.wikipedia.org/w/index.php?title=Associated Legendre polynomials&oldid=363001511](http://en.wikipedia.org/w/index.php?title=Associated_Legendre_polynomials&oldid=363001511)),

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

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

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

$$\alpha_{\frac{N2D}{orto3D}} = \sqrt{\frac{(2m+1)}{4\pi (2m)!}} \frac{(2m)!}{m! 2^m} = \sqrt{\frac{(2m+1)(2m)!}{4\pi m!^2 2^{2m}}} = \sqrt{\frac{(2m+1)!}{4\pi m!^2 2^{2m}}}, \quad (30)$$

utilizando a relação  $l! = \frac{(l+1)!}{l+1}$  e fazendo a substituição  $l = 2m$ .

Os detalhes para a expansão de Onda Esférica acima mencionada são como se segue:

Resolver a Eq.(1) para ondas esféricas, que são geradas por fontes pontuais para  $r < r_s$  e ondas de chegada, é mais complicado porque as fontes pontuais com dimensões infinitesimais que tendem para zero necessitam de ser descritas utilizando um fluxo de volume  $Q_s$ , em que a pressão radiada para um ponto do campo em  $r$  e a fonte posicionada em  $r_s$  é dada por (veja-se o livro mencionado acima "Fourier Acoustics"):

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

com  $\rho_0$  sendo a densidade específica e  $G(r|r_s)$  sendo a função de Green

$$G(r|r_s) = \frac{e^{-ik|r-r_s|}}{4\pi |r-r_s|} \quad (32)$$

$G(r/r_s)$  pode também ser expressa em harmônicas esféricas para  $r < r_s$  por

$$G(r/r_s) = i k \sum_{n=0}^{\infty} \sum_{m=-n}^n j_n(kr) h_n^{(2)}(kr_s) Y_n^m(\theta, \phi) Y_n^m(\Theta_s, \Phi_s)^* \quad (33)$$

em que  $h_n^{(2)}$  é a função de Hankel do segundo tipo. Note-se que a função de Green tem uma unidade de escala de metro<sup>-1</sup> (1/m devido a  $k$ ). As Eqs.(31), (33) podem ser comparadas com a Eq.(1) para derivar os coeficientes Ambisonics de ondas esféricas:

$$A_n^m \text{ esférica}(k, \Theta_s, \Phi_s, r_s) = \rho_0 c k^2 Q_s h_n^{(2)}(kr_s) Y_n^m(\Theta_s, \Phi_s)^* \quad (34)$$

onde  $Q_s$  é o fluxo de volume em unidades de m<sup>3</sup>s<sup>-1</sup>, e  $\rho_0$  é a densidade específica em kg m<sup>-3</sup>.

Para se ser capaz de criar sinteticamente sinais Ambisonics e para relacionar com as considerações de onda plana acima, é razoável exprimir a Eq.(34) utilizando a pressão sonora gerada na origem do sistema de coordenadas:

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

que conduz a

$$A_n^m \text{ esférica}(k, \Theta_s, \Phi_s, r_s) = 4\pi \frac{h_n^{(2)}(kr_s)}{h_0^{(2)}(kr_s)} P_{S_0} Y_n^m(\Theta_s, \Phi_s)^* \quad (36)$$

O formato de armazenamento de acordo com a invenção permite armazenar mais do que uma representação HOA e correntes de bits direccionais adicionais conjuntamente num contentor de dados. Ele permite formatos diferentes de descrições HOA o que permite aos decodificadores optimizarem a reprodução, e ele oferece um eficiente armazenamento de dados para dimensões >4 GB. Vantagens adicionais, são:

A) Pelo armazenamento de várias descrições HOA utilizando diferentes formatos conjuntamente com informação do formato de armazenamento relacionada um decodificador Ambisonics é capaz de misturar e decodificar ambas as representações.

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

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

- Informação de empacotamento e de dimensionamento de coeficiente de Ambisonics;

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



- Podem ser armazenados sinais mono direccionais relacionados. A informação de posição destes sinais direccionais pode ser descrita utilizando quer informação de ângulo e distância quer um vector de codificação de coeficientes Ambisonics.

C) O formato de armazenamento de dados Ambisonics é estendido para permitir um armazenamento de dados flexível e económico:

- Armazenamento de dados Ambisonics relacionados com as componentes do Ambisonics (canais Ambisonics) com diferentes resoluções de tamanho das palavras de PCM;

- Armazenar dados de Ambisonics com largura de banda reduzida utilizando quer o re-amostragem quer um processamento MDCT.

D) Estão disponíveis campos de metadados para associar pistas para descodificação especial (frontal, de ambiente) e para permitir o armazenamento de informação acompanhante acerca do ficheiro, tal como informação da gravação para sinais de microfone:

- Sistema de coordenadas de referência de gravação, microfone, posições da fonte e virtual do ouvinte, características direccionais do microfone, informação sobre a sala e a fonte.

E) O formato é adequado para armazenamento de múltiplas tramas contendo diferentes pistas, permitindo alterações na cena de áudio sem uma descrição de cena. (Nota: uma pista contém uma descrição de campo sonoro HOA ou uma única fonte com informação de posição. Uma trama é a combinação de uma ou mais pistas paralelas.) As pistas podem começar no início de uma trama ou terminar no final da trama, portanto não é requerido código de tempo.

F) O formato facilita o acesso rápido a dados de pista de áudio (avanço rápido ou salto até posições assinaladas) e a determinação de um código de tempo relativo ao tempo do início dos dados do ficheiro.

#### ***Parâmetro de HOA para permuta de dados de HOA***

O Quadro 6 resume os parâmetros requeridos para serem definidos para uma permuta não ambígua de dados de sinal HOA. A definição das harmónicas esféricas é fixada para os casos de valores complexos ou de valores reais, vejam-se as Eqs.(3) (6).

Cont	Dimensionalidade	2D/3D, influencia também o empacotamento de coeficientes
------	------------------	--

		Ambisonics (AC)
	Região de Interesse	Fig.6, Fig.8, Eqs.(1) (2)
	Tipo de SH	Valores complexos, reais, circular para 2D
	Normalização da SH	SN3D, N3D, orto-normalizada
Coeficiente Ambisonics	Ponderação de AC	Formato-B, FuMa, maxN, sem ponderação, definida pelo utilizador
	Sequência e resolução de amostragem de AC	Exemplos nas Eqs.(12) (13), resolução de 16/24 bits ou tipos flutuantes.
	Tipo de AC	$A_n^m$ não especificado, $d_n^m$ tipo onda plana, Eq.(16), tipos de codificação de distância $D_n^m$ ou $\tilde{C}_n^m$ , Eqs.(26) (27)

Quadro 6 - Parâmetros para a permuta não ambígua de gravações HOA

### **Detalhes do Formato de Ficheiro**

No que se segue, é descrito em detalhe o formato do ficheiro para armazenar cenas de áudio compostas por Ambisonics de Ordem Superior (HOA) ou por fontes únicas com informação de posição. A cena de áudio pode conter múltiplas sequências de HOA que podem utilizar diferentes esquemas de normalização. Assim, um descodificador pode calcular os correspondentes sinais de altifalante para o desejado arranjo de altifalantes como uma sobreposição de

todas as pistas de áudio a partir de um ficheiro actual. O ficheiro contém todos os dados requeridos para descodificar o conteúdo de áudio. O formato de ficheiro de acordo com a invenção oferece a característica de armazenar mais do que um sinal HOA ou de fonte única num único ficheiro. O formato de ficheiro utiliza uma composição de tramas, cada uma das quais pode conter várias pistas, em que os dados de uma pista estão armazenados num ou mais pacotes chamados *TrackPackets*.

Todos os tipos de inteiros estão armazenados pela ordem "little-endian" dos bytes de forma que o byte menos significativo vem primeiro. A ordem dos bits é sempre a do bit mais significativo primeiro. A notação para tipos de dados inteiros é 'int'. Um 'u' de início indica um inteiro sem sinal. A resolução em bit está escrita no final da definição. Por exemplo, um campo de inteiro de 16 bit sem sinal é definido como 'uint16'. As amostras PCM e os coeficientes HOA no formato inteiro são representados como números em vírgula fixa com a vírgula decimal no bit mais significativo.

Todos os tipos de dados em vírgula flutuante estão conformes com a especificação IEEE-754, "Standard for binary floating-point arithmetic", <http://grouper.ieee.org/groups/754/>. A notação para o tipo de dados de vírgula flutuante é 'float'. A resolução em bit está escrita no final da definição. Por exemplo, um campo de vírgula flutuante de 32 bit é definido como 'float32'.

Os identificadores de constantes ID, que identificam o início de uma trama, pista ou grupo, e as correntes de bits são definidas como tipo de dados de byte. A ordem dos bytes de conjuntos de bytes é a do byte e bit mais significativos primeiro. Portanto, o ID 'TRCK' é definido num campo de 32 bits em que os bytes estão escritos pela ordem física 'T', 'R', 'C' e 'K' (<0x54; 0x52; 0x42; 0x4b>).

Os valores em hexadecimal começam com '0x' (por exemplo, 0xAB64C5). Bits simples são colocados entre plicas (por exemplo, '1'), e múltiplos valores binários começam com '0b' (por exemplo, 0b0011 = 0x3).

Os nomes de campos de cabeçalho começam sempre com o nome do cabeçalho seguido do nome do campo, em que a primeira letra de cada palavra é maiúscula (por exemplo, *TrackHeaderSize*). As abreviaturas dos nomes de campos ou de cabeçalhos são criados utilizando apenas letras maiúsculas (por exemplo, *TrackHeaderSize* = THS).

O Formato de Ficheiro HOA pode incluir mais do que uma Trama, Pacote ou Pista. Para a discriminação de múltiplos campos de cabeçalho um número pode seguir o nome do campo ou cabeçalho. Por exemplo, o segundo *TrackPacket* da terceira Pista é nomeado '*Track3Packet2*'.

O formato de ficheiro HOA pode incluir campos de valores complexos. Estes valores complexos são armazenados

como a parte real e a imaginária, em que a parte real é escrita primeiro. O número complexo  $1+i2$  em formato 'int8' poderá ser armazenado como '0x01' seguido de '0x02'. Por isso, os campos ou coeficientes num tipo de formato de valor complexo requer o dobro do espaço de armazenamento em comparação com o tipo de formato de valor real correspondente.

### ***Estrutura do Formato de Ficheiro de Ambisonics de Ordem Superior***

#### ***Formato de Pista Única***

O formato de ficheiro de Ambisonics de Ordem Superior inclui pelo menos um *FileHeader*, um *FrameHeader*, um *TrackHeader* e um *TrackPacket* como representado na Figura 9, que apresenta um exemplo simples de formato de ficheiro de ficheiro HOA que transporta uma *Track* em um ou mais *Packets*.

Portanto, a estrutura básica de um ficheiro HOA é um *FileHeader* seguido de uma *Frame* que inclui pelo menos uma *Track*. Uma *Track* consiste sempre de um *TrackHeader* e de um ou mais *TrackPackets*.

#### ***Trama Múltipla e Formato de Pista***

Em contraste com o *FileHeader*, o Ficheiro HOA pode conter mais do que uma Trama, em que uma Trama pode

conter mais do que uma Pista. É utilizado um novo *FrameHeader* se a dimensão máxima de uma Trama é excedida ou são adicionadas Pistas, ou removidas de uma Trama para outra. A estrutura de uma Pista múltipla e de um Ficheiro de Trama HOA está apresentada na Figura 10.

A estrutura de uma Trama de Pista múltipla começa com o *FrameHeader* seguido por todos os *TrackHeaders* da Trama. Consequentemente, os *TrackPackets* de cada Pista são enviados sucessivamente para os *FrameHeaders*, em que os *TrackPackets* são interlaçados pela mesma ordem como os *TrackHeaders*.

Numa Trama de Pista múltipla, o comprimento de um Pacote em amostras é definido no *FrameHeader* e é constante para todas as Pistas. Além disso, as amostras de cada Pista são sincronizadas, por exemplo, as amostras do *Track1Packet1* são síncronas com as amostras do *Track2Packet1*. *TrackCodingTypes* específicos podem causar um atraso no lado do descodificador, e tais atrasos específicos necessitam de ser conhecidos no lado do descodificador, ou são para serem incluídos na parte dependente do *TrackCodingType* do *TrackHeader*, porque o descodificador sincroniza todos os *TrackPackets* para o atraso máximo de todas as *Pistas* da *Trama*.

### ***Meta Dados dependentes do Ficheiro***

Os Meta Dados que se referem ao Ficheiro HOA

completo podem opcionalmente ser adicionados após o *FileHeader* em *MetaDataChunks*. Um *MetaDataChunks* começa com um Id de Utilizador Geral específico (*GUID*) seguido do *MetaDataChunkSize*. A essência do *MetaDataChunck*, por exemplo, a informação Meta Dados, é empacotada num formato XML ou num qualquer formato definido pelo utilizador. A Figura 11 apresenta a estrutura de um formato de ficheiro HOA utilizando vários *MetaDataChuncks*.

### ***Tipos de Pistas***

Uma *Pista* do Formato HOA diferencia-se de entre uma *HOATrack* geral e uma *SingleSourceTrack*. A *HOATrack* inclui o campo sonoro completo codificado como *HOACoefficients*. Portanto, uma descrição de cena, por exemplo as posições das fontes codificadas, não é requerida para descodificar os coeficientes no lado do descodificador. Por outras palavras, uma cena de áudio é armazenada com os *HOACoefficients*.

Contrariamente ao *HOATrack*, o *SingleSourceTrack* inclui apenas uma fonte codificada como amostras PCM conjuntamente com as posições 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 *TrackPositionVectro*. O *TrackHOAEncodingVector* contém os valores de codificação de HOA para obter o *HOACoefficient* para cada amostra. O *TrackPositionVector* contém as posições da fonte como ângulo



e distância em relação à posição central de audição.

### ***Cabeçalho de Ficheiro***

<b>Nome do Campo</b>	<b>Tamanho /Bit</b>	<b>Tipo de Dados</b>	<b>Descrição</b>
FileID	3	byte	O identificador de ficheiro constante para o Formato de Ficheiro HOA:<"H"; "O"; "A"; "F"> ou <0x48; 0x4F; 0x41; 0x46>
FileVersionNumber	8	uint8	Número da versão do Formato HOA 0-255
FileSampleRate	32	uint32	Taxa de Amostragem em Hs para todas as Tramas e Pistas
FileNumberOfFrames	32	uint32	É requerido um Número Total de Tramas de pelo menos '1'
reservado	8	byte	
Número Total de Bits	112		

O *FileHeader* inclui toda a informação constante para o Ficheiro HOA completo. O *FileID* é utilizado para identificar o Formato de Ficheiro HOA. A taxa de amostragem é constante para todas as *Pistas* mesmo se ele é enviado no *FrameHeader*. Os Ficheiros HOA que alterem a sua taxa de amostragem de uma trama para outra são inválidos. O número de Tramas está indicado no *FileHeader* para indicar a estrutura da Trama ao descodificador.

*Grupos de Meta Dados*

<b>Nome do Campo</b>	<b>Tamanho /Bit</b>	<b>Tipo de Dados</b>	<b>Descrição</b>
ChunkID	32	byte	ID de Utilizador Geral (não definido ainda)
ChunkSize	32	uint32	Dimensão do grupo em bytes excluindo o ChunkID e o campo ChunkSize
ChunkData	8*ChunkSize	byte	Campos definidos pelo utilizador ou estrutura XML dependendo do ChunkID
Número Total de Bits	64+ 8*ChunkSize		

**Cabeçalho de Trama**

Nome do Campo	Tamanho /Bit	Tipo de Dados	Descrição
FrameID	32	byte	O identificador constante para todos os FrameHeader: <"F"; "R"; "A"; "M"> ou <0x46; 0x52; 0x41; 0x4D>
FrameSize	32	uint32	Tamanho da Trama em Byte excluindo o FrameID e o campo FrameSize
FrameNumber	32	uint32	Um único FrameNumber que começa com 0 para a primeira Trama e aumenta para Tramas seguintes. A última Trama tem o FrameNumber FileNumberOfFrame-1.
FrameNumberOfSamples	32	uint32	Número de amostras armazenadas em cada Pista da Trama
FrameNumberOfTracks	8	uint8	Número de Pistas armazenadas na Trama
FramePacketSize	32	uint32	Tamanho de um Pacote em amostras. O tamanho do pacote é constante para todas as Pistas.
FrameSampleRate	32	uint32	A Taxa de Amostragem em Hz constante para todas as Tramas e Pistas tem que ser idêntico ao FileSampleRate (Redefinição para aplicações de correntes de bits onde o FileHeader poderá ser desconhecido)
Número Total de Bits	200		

O *FrameHeader* contém a informação constante de todas as Pistas de uma Trama e indica alterações no Ficheiro HOA. O *FrameID* e o *FrameSize* indicam o início de

uma Trama e o comprimento da Trama. Estes dois campos permitem um fácil acesso de cada trama e uma verificação cruzada da estrutura da Trama. Se o comprimento da Trama requer mais do que 32 bits, uma Trama pode ser separada em várias Tramas. Cada Trama tem um único *FrameNumber*. O *FrameNumber* deverá começar com 0 e deverá ser incrementado de um para cada nova Trama.

O número de amostras da Trama é constante para todas as Pistas de uma Trama. O número de Pistas numa Trama é constante para a Trama. É enviado um novo Cabeçalho de Trama para terminar ou começar Pistas numa desejada posição da amostra.

As amostras de cada Pista são armazenadas em Pacotes. A dimensão destes *TrackPackets* está indicada em amostras e é constante para todas as Pistas. O número de Pacotes é igual ao número inteiro que é requerido para armazenar o número de amostras da Trama. Portanto, o último Pacote de uma Pista pode conter menos amostras do que a dimensão do Pacote indicada. A taxa de amostragem de uma trama é igual ao *FileSampleRate* e está indicada no *FrameHeader* para permitir a descodificação de uma Trama sem o conhecimento do *FileHeader*. Isto pode ser utilizado quando se faz a descodificação a partir do meio de um ficheiro com múltiplas tramas sem o conhecimento do *FileHeader*, por exemplo para aplicações de correntes de bits.

**Cabeçalho de Pista**

Nome do Campo	Tamanho /Bit	Tipo de Dados	Descrição
TrackID	32	byte	0 identificador constante para todos os TrackHeader: <"T"; "R"; "A"; "C"> ou <0x54; 0x52; 0x41; 0x43>
TrackNumber	16	uint16	Um único TrackNumber para a identificação de Pistas coerentes em várias Tramas
TrackHeaderSize	32	uint32	Tamanho do TrackHeader excluindo o TrackID e o campo TrackNumber (Desfasamento para o início do próximo TrackHeader ou primeiro TrackPacket)
TrackMetaDataOffset	32	uint32	Desfasamento a partir do fim deste campo para o início do campo TrackMetaData. Zeros é igual a TrackMetaData não incluído.
TrackSource Type	1	binário	'0'=PistaHOA e '1'=Pista de SingleSource
reservado	7	binário	0b0000000
Condição: TrackSource Type == '0'			TrackHeader para Pistas HOA
<HOATrackHeader>	dyn	byte	veja-se a secção HOA TrackHeader
Condição: TrackSourceType == '1'			TrackHeader para Pistas SingleSource
<Pista SingleSource-Header>	dyn	byte	veja-se a secção Cabeçalho de Pista de Fonte Única de Posição fixa e Cabeçalho de Pista de Fonte Única de Posição móvel
Condição: TrackMetaDataOffset>0			
TrackMetaData	dyn	byte	campo XML para Meta Dados dependente da Pista veja-se o Quadro TrackMetaData
Número Total de Bits	120+ dyn		

O termo 'dyn' refere-se a um tamanho de campo dinâmico devido a campos condicionais. O *TrackHeader* contém a informação constante para os Pacotes da Pista especificada. O *TrackHeader* é separado numa parte constante e numa parte variável para dois *TrackSourceTypes*. O *TrackHeader* começa com um *TrackID* constante para verificação e identificação do início do *TrackHeader*. É atribuído um único *TrackNumber* a cada Pista para indicar Pistas coerentes através de limites de Trama. Assim, uma pista com o mesmo *TrackNumber* pode ocorrer na trama seguinte. O *TrackHeaderSize* é proporcionado para saltar para o *TrackHeader* seguinte e é indicado como um desfasamento a partir do fim do campo *TrackHeaderSize*. O *TrackMetaDataOffset* proporciona o número de amostras a saltar directamente para o início do campo *TrackMetaData*, que pode ser utilizado para saltar a parte de comprimento variável do *TrackHeader*. Um *TrackMetaDataOffset* de zero indica que o campo *TrackMetaData* não existe. Confiando no *TrackSourceType*, são proporcionados os *HOATrackHeader* ou o *SingleSourceTrackHeader*. O *HOATrackHeader* proporciona a informação de tamanho para coeficientes HOA padronizados que descrevem o campo sonoro completo. O *SingleSourceTrackHeader* contém informação para as amostras de uma pista PCM mono e a posição da fonte. Para *SingleSourceTracks* o descodificador tem que incluir a Pista na cena.

No fim do *TrackHeader* é definido um campo opcional *TrackMetaData* que utiliza o formato XML para

proporcionar Meta Dados dependentes da pista, por exemplo informação adicional para a transmissão no formato A (sinais de sistema de microfones).

### *Cabeçalho de Pista HOA*

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackComplexValueFlag	2	binário	0b00: apenas parte real 0b01: parte real e imaginária 0b10: apenas parte imaginária 0b11: reservado
TrackSampleFormat	4	binário	0b0000 Inteiro sem sinal 8 bit 0b0001 Inteiro com sinal 8 bit 0b0010 Inteiro com sinal 16 bit 0b0011 Inteiro com sinal 24 bit 0b0100 Inteiro sem sinal 32 bit 0b0101 Inteiro com sinal 64 bit 0b0110 Flutuante 32 bit (binário precisão simples) 0b0111 Flutuante 64 bit (binário precisão dupla) 0b1000 Flutuante 128 bit (binário precisão quádrupla) 0b1001-0b1111 reservado
reservado	2	binário	bits de enchimento
TrackHOAParams	dyn	bytes	veja-se TrackHOAParams
TrackCodingType	8	uint8	,0' Os coeficientes HOA são codificados como amostras PCM com resolução de bits constante e constante resolução em frequência. ,1' Os coeficientes HOA são codificados com uma resolução de bits e resolução em frequência dependentes da ordem outro caso reservado para tipos de codificação adicionais
Condição: TrackCodingType == '1'			Informação adicional para codificação do tipo 1
TrackBandwidthReductionType	8	uint8	0 toda a largura de banda para todas as ordens 1 Redução de largura de banda por via de MDCT 2 Redução de largura de banda por via de filtro de domínio do tempo

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackNumberOfOrderRegions	8	uint8	A largura de banda e resolução de bit podem ser adaptadas para um número de regiões em que cada número tem uma ordem de início e de fim. TrackNumberOfOrderRegions indica o número de regiões definidas.
<b>Escrever os seguintes campos para cada região</b>			
TrackRegionFirstOrder	8	uint8	Primeira ordem da região
TrackRegionLastOrder	8	uint8	Última ordem da região
TrackRegionSampleFormat	4	binário	0b0000 Inteiro sem sinal 8 bit 0b0001 Inteiro com sinal 8 bit 0b0010 Inteiro com sinal 16 bit 0b0011 Inteiro com sinal 24 bit 0b0100 Inteiro com sinal 32 bit 0b0101 Inteiro com sinal 64 bit 0b0110 Flutuante 32 bit (binário precisão simples)  0b0111 Flutuante 64 bit binário precisão dupla)  0b1000 Flutuante 128 bit (binário precisão quádrupla)  0b1001-0b1111 reservado
TrackRegionUseBandwidthReduction	1	binário	'0' Totalidade da largura de banda para esta região '1' largura de banda reduzida para esta região com TrackBandwidthReductionType
reservado	3	binário	bits de enchimento
Condição: TrackRegionUseBandwidthReduction == '1'			A largura de banda está reduzida nesta região
Condição: TrackBandwidthReductionType == 1			Redução de largura de banda por via de informação adicional MDCT
TrackRegionWindowType	8	uint8	0: Janela seno: $W(t) = \sin\left(\frac{\pi(t+0.5)}{N}\right)$ noutro caso: reservado
TrackRegionFirstBin	16	uint16	primeiro bin MDCT codificado (frequência de corte inferior)
TrackRegionLastBin	16	uint16	último bin MDCT codificado (frequência de corte superior)
Condição: TrackBandwidthReductionType == 2			Redução de largura de banda por via de informação adicional de filtro no domínio do tempo
TrackRegionFilterLength	16	uint16	Número de coeficientes do filtro passa baixos
<TrackRegionFilterCoefficients>	dyn	float32	Coefficientes do filtro passa baixos do <i>TrackRegionFilterLength</i>
TrackRegionModulationFreq	32	float32	Modulação de frequência normalizada $\Omega_{mod}/\pi$ requerida para deslocar o espectro do sinal
TrackRegionDownsampleFactor	16	uint16	Factor de subamostragem <i>M</i> , deverá ser um divisor de <i>FramePacketSize</i>
TrackRegionUpsampleFactor	16	uint16	Factor de sobre amostragem <i>K</i> < <i>M</i>



Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackRegionFilterDelay	16	uint16	Atraso nas amostras (de acordo com <i>FileSampleRate</i> ) do processamento da redução de largura de banda de codificação/descodificação

O *HOATrackHeader* é uma parte do *TrackHeader* que contém informação para descodificar uma *HOATrack*. O *TrackPackets* de uma *HOATrack* transfere coeficientes de HOA que codificam o campo sonoro inteiro de uma Pista. Basicamente, o *HOATrackHeader* contém todos os parâmetros de HOA que são requeridos no lado do descodificador para descodificar os parâmetros de HOA para o arranjo de altifalantes dado. O *TrackComplexValueFlag* e o *TrackSampleFormat* definem o tipo de formato dos coeficientes de HOA de cada *TrackPacket*. Para coeficientes codificados ou comprimidos, o *TrackSampleFormat* define o formato dos coeficientes descodificados ou descomprimidos. Todos os tipos de formato podem ser números reais ou complexos. É disponibilizada mais informação sobre números complexos na secção acima *Detalhes de Formato de Ficheiro*.

Toda a informação dependente do HOA está definida no *TrackHOAParams*. Os *TrackHOAParams* são reutilizados noutros *TrackSourceTypes*. Portanto, os campos dos *TrackHOAParams* estão definidos e descritos na secção *TrackHOAParams*.

O campo *TrackCodingType* indica o formato de codificação (compressão) dos coeficientes de HOA. A versão básica do formato de ficheiro HOA inclui, por exemplo, dois

*CodingTypes*. Um *CodingType* é o tipo de codificação PCM (*TrackCodingType* == '0'), em que os coeficientes reais ou complexos descomprimidos são escritos nos pacotes no *TrackSampleFormat* seleccionado. A ordem e a normalização dos coeficientes de HOA estão definidos nos campos *TrackHOAParams*.

Um segundo *CodingType* permite uma alteração do formato de amostragem e limitar a largura de banda dos coeficientes de cada ordem de HOA. Está disponibilizada uma descrição detalhada deste *CodingType* na secção *Codificação de TrackRegion*, segue-se uma explicação mais curta: O *TrackBandwidthReductionType* determina o tipo de processamento que tem sido utilizado para limitar a largura de banda de cada ordem de HOA. Se a largura de banda de todos os coeficiente está inalterada, a redução de largura de banda pode ser desligada atribuindo o valor zero ao campo *TrackBandwidthReductionType*. Estão definidos os outros tipos de processamento de redução de largura de banda. O formato inclui um processamento MDCT no domínio da frequência e opcionalmente um processamento de filtro no domínio do tempo. Para mais informação sobre o processamento MDCT, veja-se a secção *Redução da largura de banda por via de MDCT*. As ordens de HOA podem ser combinadas em regiões do mesmo formatos de amostragem e de largura de banda. O número de regiões está indicado pelo campo *TrackNumberOfOrderRegions*. Para cada região tem que ser definida a informação do primeiro e do último índice, do formato de amostragem e da redução de largura de banda

opcional. Uma região obterá pelo menos uma ordem. As ordens que não sejam cobertas por qualquer região são codificadas com a totalidade da largura de banda utilizando o formato padronizado indicado no campo *TrackSampleFormat*. Um caso especial é a não utilização de região (*TrackNumber0-OrderRegions* == 0). Este caso pode ser utilizado para coeficientes de HOA desinterlaçados no formato PCM, em que os componentes de HOA não estão interlaçados por amostra. Os coeficientes de HOA das ordens de uma região são codificados no *TrackRegionSampleFormat*. O *TrackRegionUseBandwidthReduction* indica a utilização do processamento de redução de largura de banda para os coeficientes das ordens da região. Se o sinalizador do *TrackRegionUseBandwidthReduction* está activado, segue-se a informação adicional de redução de largura de banda. Para o processamento MDCT são definidos o tipo de janela e o primeiro e último bin MDCT codificados. Desta forma, o primeiro bin é equivalente à frequência de corte inferior e o último bin define a frequência de corte superior. Os bins MDCT são também codificados no *TrackRegionSampleFormat*, veja-se a secção *Redução de largura de banda por via de MDCT*.

### ***Tipo Fonte Única***

As fontes únicas estão subdivididas em fontes de posição fixa e de posição móvel. O tipo de fonte está indicado no *TrackMovingSourceFlag*. A diferença entre os tipos de fonte de posição móvel e fixa é a de que a posição

da fonte fixa está indicada apenas uma vez no *TrackHeader* e em cada *TrackPackage* para fontes móveis. A posição de uma fonte pode ser indicada explicitamente com o vector de posição em coordenadas esféricas ou implicitamente como vector de codificação de HOA. A fonte propriamente dita é uma pista mono PCM que tem que ser codificada para coeficientes de HOA no lado do decodificador no caso da utilização de um decodificador Ambisonics para a reprodução.

#### ***Cabeçalho de Pista de Fonte Única de Posição Fixa***

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackComplexValueFlag	2	binário	0b00: apenas parte real 0b01: parte real e imaginária 0b10: apenas parte imaginária 0b11: reservado
TrackSampleFormat	4	binário	0b0000 Inteiro sem sinal 8 bit 0b0001 Inteiro com sinal 8 bit 0b0010 Inteiro com sinal 16 bit 0b0011 Inteiro com sinal 24 bit 0b0100 Inteiro sem sinal 32 bit 0b0101 Inteiro com sinal 64 bit 0b0110 Flutuante 32 bit (binário precisão simples) 0b0111 Flutuante 64 bit (binário precisão dupla) 0b1000 Flutuante 128 bit (binário precisão quádrupla) 0b1001-0b1111 reservado
reservado	2	binário	bits de enchimento
TrackHOAParams	dyn	bytes	veja-se TrackHOAParams
TrackCodingType	8	uint8	,0' Os coeficientes HOA são codificados como amostras PCM com resolução de bits constante e constante resolução em frequência. ,1' Os coeficientes HOA são codificados com uma resolução de bits e resolução em frequência dependentes da ordem outro caso reservado para tipos de codificação adicionais
Condição: TrackCodingType == '1'			Informação adicional para codificação do tipo 1

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackMovingSourceFlag	1	binário	constante '0' para fontes fixas
TrackPositionType	1	binário	'0' É enviada a posição como ângulo TrackPositionVector [R, Theta, Phi] '1' É enviada a posição como vector de codificação de HOA do comprimento TrackHOAParamNumberOfCoeffs
TrackSampleFormat	4	binário	0b0000 Inteiro sem sinal 8 bit 0b0001 Inteiro com sinal 8 bit 0b0010 Inteiro com sinal 16 bit 0b0011 Inteiro com sinal 24 bit 0b0100 Inteiro com sinal 32 bit 0b0101 Inteiro com sinal 64 bit 0b0110 Flutuante 32 bit (binário precisão simples) 0b0111 Flutuante 64 bit (binário precisão dupla) 0b1000 Flutuante 128 bit (binário precisão quádrupla) 0b1001-0b1111 reservado
reservado	2	binário	bits de enchimento
Condição: TrackPositionType == '0'			Posição como ângulo TrackPositionVector a seguir
TrackPositionTheta	32	float32	Inclinação em rad [0, pi]
TrackPositionPhi	32	float32	azimute (sentido contrário ao dos relógios) em rad [0, 2pi]
TrackPositionRadius	32	float32	Distância ao ponto de referência em metros
Condição: TrackPositionType == '1'			Posição como vector de codificação HOA
TrackHOAParams	dyn	bytes	veja-se TrackHOAParams
TrackEncodeVectorComplexFlag	2	binário	0b00: apenas parte real 0b01: parte real e imaginária 0b10: apenas parte imaginária 0b11: reservado Tipo de número para Vector de codificação
TrackEncodeVectorFormat	1	binário	'0' float32 '1' float64
reservado	5	binário	bits de enchimento
Condição: TrackEncodeVectorFormat == '0'			vector de codificação como float32
<TrackHOAEncodingVector>	dyn	float32	Entradas de TrackHOAParamNumberOfCoeffs do vector de codificação HOA pela ordem TrackHOAParamCoeffSequence
Condição: TrackEncodeVectorFormat == '1'			vector de codificação como float64
<TrackHOAEncodingVector>	dyn	float64	Entradas de TrackHOAParamNumberOfCoeffs do vector de codificação HOA pela ordem TrackHOAParamCoeffSequence

O tipo de fonte de posição fixa é definido por uma *TrackMovingSourceFlag* de zero. O segundo campo indica o

*TrackPositionType* que dá a codificação da posição da fonte como um vector em coordenadas esféricas ou como um vector de codificação de HOA. O formato de codificação das amostras PCM mono está indicado pelo campo *TrackSampleFormat*. Se a posição da fonte é enviada como *TrackPositionVector*, as coordenadas esféricas da posição da fonte são definidas nos campos *TrackPositionTheta* (inclinação a partir do eixo *s* para o plano *x,y*), *TrackPositionPhi* (azimute no sentido contrário ao dos relógios começando no eixo *x*) e *TrackPositionRadius*.

Se a posição da fonte é definida como um vector de codificação de HOA, os *TrackHOAParams* são definidos primeiro. Estes parâmetros estão definidos na secção *TrackHOAParams* e indicam as normalizações e definições utilizadas do vector de codificação de HOA. A *TrackEncodeVectorComplexFlag* e o campo *TrackEncodeVectorFormat* definem o tipo de formato do vector *TrackHOAEncoding* seguinte. O *TrackHOAEncodingVector* consiste de valores *TrackHOAParamNumberOfCoeffs* que estão codificados no formato ou 'float32' ou 'float64'.

**Cabeçalho de Pista de Fonte Única de Posição****Móvel**

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackMovingSourceFlag	1	binário	constante '1' para fontes móveis
TrackPositionType	1	binário	0' É enviada a posição como ângulo TrackPositionVector [R, Theta, Phi] 1' É enviada a posição como vector de codificação de HOA do comprimento TrackHOAParamNumberOfCoeffs
TrackSampleFormat	4	binário	0b0000 Inteiro sem sinal 8 bit 0b0001 Inteiro com sinal 8 bit 0b0010 Inteiro com sinal 16 bit 0b0011 Inteiro com sinal 24 bit 0b0100 Inteiro com sinal 32 bit 0b0101 Inteiro com sinal 64 bit 0b0110 Flutuante 32 bit (binário precisão simples) 0b0111 Flutuante 64 bit (binário precisão dupla) 0b1000 Flutuante 128 bit (binário precisão quádrupla) 0b1001-0b1111 reservado
reservado	2	binário	bits de enchimento
Condição: TrackPositionType == '1'			Posição como vector de codificação de HOA
TrackHOAParams	dyn	bytes	veja-se TrackHOAParams
TrackEncodeVectorComplexFlag	2	binário	0b00: apenas parte real 0b01: parte real e imaginária 0b10: apenas parte imaginária 0b11: reservado Tipo de número para Vector de codificação
TrackEncodeVectorFormat	1	binário	'0' float32 '1' float64
reservado	5	binário	bits de enchimento

O tipo de fonte de posição móvel é definido por uma *TrackMovingSourceFlag* de '1'. O cabeçalho é idêntico ao cabeçalho das fontes fixas excepto que os campos de dados de posição da fonte *TrackPoaitionTheta*, *TrackPositionPhi*, *TrackPositionRadius* e *TrackHOAEncodingVector* estão ausentes. Para fontes móveis estes estão localizados nos *TrackPackets* para indicarem a nova (móvel) posição da fonte

em cada *Pacote*.

### Quadros de Pistas Especiais

#### TrackHOAParams

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackHOAParamDimension	1	binário	'0' = 2D e '1' = 3D
TrackHOAParamRegionOfInterest	1	binário	'0' são calculados os coeficientes HOA para fontes fora da região de interesse (interior)  '1' são calculados os coeficientes HOA para fontes dentro da região de interesse (exterior)  (A região de interesse não contém nenhuma fontes)
TrackHOAParamSphericalHarmonicType	1	binário	'0' real '1' complexa
TrackHOAParamSphericalHarmonicNorm	3	binário	0b000 não normalizado  0b001 Schmidt semi-normalizado  0b010 normalizado a $4\pi$ ou normalizado a 2D  0b011 orto-normalizado  0b100 Redimensionamento Dedicado  outros Reservado
TrackHOAParamFurseMalhamFlag	1	binário	Indica que os coeficientes de HOA estão normalizados pelo redimensionamento de Furse-Malham
TrackHOAParamDecoderType	2	binário	0b00 Redimensionamento do decodificador de ondas planas $1/(4\pi i^n)$  0b01 Redimensionamento de decodificador de ondas esféricas (codificação de distância): $1/(ikh_n(kr_{is}))$  0b10 Redimensionamento de decodificador de ondas esféricas (codificação de distância para pressão sonora medida): $h_0(kr_{is})/(ikh_n(kr_{is}))$  0b11 coeficientes de HOA planos



Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
TrackHOAParamCoeffSequence	2	binário	0b00 ordem de Formato-B 0b01 numérico ascendente 0b10 numérico descendente 0b11 Reservado
Reservado	5	binário	bits de enchimento
TrackHOAParamNumberOfCoeffs	16	uint16	Número de Coeficientes de HOA por amostra menos 1
TrackHOAParamHorizontalOrder	8	uint8	Ordem de Ambisonics no plano X/V
TrackHOAParamVerticalOrder	8	uint8	Ordem de Ambisonics para a dimensão 3D ('0' para coeficientes de HOA de 2D)
Condição: TrackHOAParamSphericalHarmonicNorm == "dedicado" <0b101>			Campo para Valores de Redimensionamento dedicado para cada Coeficiente de HOA
TrackComplexValueScalingFlag	2	binário	0b00: apenas parte real 0b01: parte real e imaginária 0b10: apenas parte imaginária 0b11: reservado Tipo de número para TrackScalingValues dedicados
TrackScalingFormat	1	binário	'0' float32 '1' float64
Reservado	5	binário	bits de enchimento
Condição: TrackScalingFormat = '0'			TrackScalingFactors como float32
<TrackScalingFactors>	dyn	float32	Factores de Redimensionamento TrackHOAParamNumberOfCoeffs se TrackComplexValueScalingFlag == 0b01 a ordem das partes complexas dos números é <[real1, imaginário1], [real2, imaginário2], ..., [realN, imaginário]>
Condição: TrackScalingFormat = '1'			TrackScalingFactor como float64
<TrackScalingFactors>	dyn	float64	Factores de Redimensionamento TrackHOAParamNumberOfCoeffs se TrackComplexValueScalingFlag == 0b01 a ordem das partes complexas dos números é <[real1, imaginário1], [real2, imaginário2], ..., [realN, imaginário]>
Condição: TrackHOAParamDecoderType == 0b01    TrackHOAParamDecoderType == 0b10			É definido o raio dos altifalantes de referência para a codificação de distância
TrackHOAParamReferenceRadius	16	uint16	Este é o raio dos altifalantes de referência $r_i$ , em mm que tem sido aplicado para os coeficientes HOA para um decodificador de ondas esféricas de acordo com Poletti ou Daniel.

Têm sido, no passado, discutidas várias abordagens para a codificação e descodificação de HOA. No entanto, sem qualquer conclusão ou acordo sobre os coeficientes de HOA. Vantajosamente, o formato de acordo com a invenção permite o armazenamento da maioria das representações de HOA conhecidas. Os *TrackHOAParams* são definidos para clarificar que tipo de normalização e sequência de ordem dos coeficientes tem sido utilizada no lado do codificador. Estas definições têm que ser tidas em conta no lado do descodificador para a mistura de pistas de HOA e para se aplicar a matriz de descodificação.

Os coeficientes de HOA podem ser aplicados para o campo sonoro tridimensional completo ou apenas para o plano x/y bidimensional. A dimensão da *HOATrack* é definida pelo campo *TrackHOAParamDimension*.

O *TrackHOAParamRegionOfInterest* reflecte duas expansões de pressão sonora em séries pelo que as fontes residem dentro ou fora da região de interesse, e a região de interesse não contém quaisquer fontes. O cálculo da pressão sonora para os casos interior e exterior está definida nas equações (1) e (2) acima, respectivamente, pelo que a informação direccional do sinal de HOA  $\mathbf{A}_n^m(\mathbf{k})$  é determinada pela harmónica esférica complexa conjugada  $\mathbf{Y}_n^m(\theta, \phi)$ . Esta função é definida numa versão de número complexo e de número real. Os codificadores e descodificadores têm que aplicar a função de harmónica esférica do tipo de número equivalente. Portanto, o

*TrackHOAParamSphericalHarmonicType* indica que tipo de função de harmónica esférica tem sido aplicado no lado do codificador.

Como mencionado acima, basicamente, a função de harmónica 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 das harmónicas esféricas de valor complexo é:

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

onde  $N_{n,m}$  é um factor de redimensionamento (veja-se a Eq.(3)). Esta representação de valor complexo pode ser transformada numa representação de valor real utilizando a equação seguinte:

$$S_n^m(\theta, \phi) = \begin{cases} \frac{(-1)^m}{\sqrt{2}} (Y_n^m + Y_n^{m*}) & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \cos(m\phi), & m > 0 \\ Y_n^0 & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) & m = 0 \\ \frac{-1}{i\sqrt{2}} (Y_n^m - Y_n^{m*}) & = \tilde{N}_{n,m} P_{n,|m|}(\cos(\theta)) \sin(|m|\phi), & m < 0 \end{cases}$$

onde o factor de redimensionamento modificado para as harmónicas esféricas de valor real é

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

Para representações 2D, tem que ser utilizada a função Harmónica circular para a codificação e

descodificação dos coeficientes de HOA.

A representação de valor complexo da harmónica circular é definida por  $\tilde{Y}_m(\phi) = \tilde{N}_m e^{im\phi}$ .

A representação de valor real da harmónica circular é definida por  $\tilde{S}_m(\phi) = \tilde{N}_m \begin{cases} \cos(m\phi) & ; m \geq 0 \\ \sin(|m|\phi) & ; m < 0 \end{cases}$ .

São utilizados vários factores de normalização  $N_{n,m}$ ,  $\tilde{N}_{n,m}$ ,  $\tilde{N}_m$  e  $\tilde{\tilde{N}}_m$  para adaptar as funções harmónicas esféricas ou circulares às aplicações ou requisitos específicos. Para assegurar a descodificação correcta dos coeficientes de HOA a normalização da função harmónica esférica utilizada no lado do codificador tem que ser conhecida no lado do descodificador. O Quadro 7 seguinte define as normalizações que podem ser seleccionadas com o campo *TRackHOAParamSphericalHarmonicNorm*.

Normalizações de harmônicas esféricas de valores complexos 3D $N_{n,m}$			
Não normalizada 0b000	Schmidt semi-normalizada, SN3D 0b001	Normalizada a $4\pi$ , N3D, Geodesia $4\pi$ 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 harmônicas esféricas de valores reais 3D $\tilde{N}_{n,m}$			
Não normalizada 0b000	Schmidt semi-normalizada, SN3D 0b001	Normalizada a $4\pi$ , N3D, Geodesia $4\pi$ 0b010	Orto-normalizada 0b011
$\sqrt{2-\delta_{0,m}}$	$\sqrt{(2-\delta_{0,m})\frac{(n- m )!}{(n+ m )!}}$	$\sqrt{(2-\delta_{0,m})\frac{(2n+1)(n- m )!}{(n+ m )!}}$	$\sqrt{(2-\delta_{0,m})\frac{(2n+1)(n- m )!}{4\pi(n+ m )!}}$
Normalizações de harmônicas circulares de valores complexos 2D $\tilde{N}_m$			
Não normalizada 0b000	Schmidt semi-normalizada, SN2D 0b001	Normalizada a 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 harmônicas circulares de valores reais 2D $\tilde{N}_m$			
Não normalizada 0b000	Schmidt semi-normalizada, SN2D 0b001	Normalizada a 2D, N2D 0b010	Orto-normalizada 0b011
$\sqrt{\frac{2-\delta_{0,m}}{2}}$	1	$\sqrt{2-\delta_{0,m}}$	$\sqrt{(2-\delta_{0,m})\frac{1}{2\pi}}$

Quadro 7 - Normalizações de funções harmônicas esféricas e circulares

Para futuras normalizações, está disponível o valor dedicado do campo `TrackHOAParamSphericalHarmonicNorm`.

Para uma normalização dedicada o factor de redimensionamento para cada coeficiente de HOA está definido no fim do *TrackHOAParams*. Os factores de redimensionamento dedicados *TrackScalingFactors* podem ser transmitidos como valores reais ou complexos `'float32'` ou `'float64'`. O formato do factor de redimensionamento está definido nos campos *TrackComplexValueScalingFlag* e *TrackScalingFormat* no caso de redimensionamento dedicado.

A normalização de Furse-Malham pode ser aplicada adicionalmente aos coeficientes de HOA codificados para igualizar as amplitudes dos coeficientes de diferentes ordens de HOA para valores absolutos de menos do que `'um'` para uma transmissão em tipos de formato inteiro. A normalização de Furse-Malham foi desenhada para a função harmónica esférica de valores reais SN3D até coeficientes de ordem três. Portanto, é recomendada a utilização da normalização de Furse-Malham apenas em combinação com a função harmónica esférica de valores reais SN3D. Além disso, a *TrackHOAParamFurseMalhamFlag* é ignorada para Pistas com uma ordem de HOA maior do que três. A normalização de Furse-Malham tem que ser invertida no lado do descodificador para descodificar os coeficientes de HOA. O Quadro 8 define os coeficientes de Furse-Malham.

n	m	Pesos de 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	$2/\sqrt{3}$
3	-3	$\sqrt{8/5}$
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	$\sqrt{8/5}$

Quadro 8 - Factores de normalização de Furse-Malham a serem aplicados no lado do codificador

O *TrackHOAParamDecoderType* define que tipo de descodificador está no lado do codificador assumindo estar presente no lado do descodificador. O tipo de descodificador **determina** o modelo de altifalante (onda esférica ou plana) que é **para ser** utilizada no lado do descodificador para gerar o campo sonoro. Desse modo a complexidade computacional do descodificador pode ser

reduzida deslocando partes da equação do descodificador para a equação do codificador. Adicionalmente, podem ser reduzidas questões numéricas no lado do codificador. Adicionalmente, o descodificador pode ser reduzido para **um processamento** igual para todos os coeficientes de HOA porque todas as inconsistências no lado do descodificador podem ser movidas para o codificador. No entanto, para ondas esféricas, tem que ser assumida **uma** distância constante dos **altifalantes** à posição de audição. Portanto, o tipo de descodificador assumido é indicado no *TrackHeader*, e o raio dos altifalantes  $r_{ts}$  para os tipos de descodificador de onda esférica é transmitido no campo opcional *TrackHOAParamReferenceRadius* em milímetros. Um filtro opcional no lado do descodificador pode igualizar as diferenças **entre** o raio dos altifalantes assumido e o real.

A normalização *TrackHOAParamDecoderType* dos coeficientes de HOA  $C_n^m$  depende da utilização da expansão em séries do campo sonoro interior ou exterior seleccionada em *TrackHOAParamRegionOfInterest*. Nota: Os coeficientes  $d_n^m$  na Eq.(18) e nas equações seguintes correspondem aos coeficientes  $C_n^m$  no que se segue. No lado do codificador, os coeficientes  $C_n^m$  são determinados a partir dos coeficientes  $A_n^m$  ou  $B_n^m$  como definido no Quadro 9, e são armazenados. A normalização utilizada está indicada no campo *TrackHOAParamDecoderType* do cabeçalho *TrackHOAParam*:



<i>TrackHOAParamDecoderType</i>	Coeficientes de HOA	
	Interior	Exterior
0b00: onda plana	$C_n^m = A_n^m / (4\pi i^n)$	--
0b01: onda esférica	$C_n^m = A_n^m / (ikh_n(kr_{ls}))$	$C_n^m = A_n^m / (ikj_n(kr_{ls}))$
0b10: pressão sonora 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	$C_n^m = A_n^m$	$C_n^m = B_n^m$

Quadro 9 - Coeficientes de HOA transmitidos para várias normalizações de tipos de decodificador

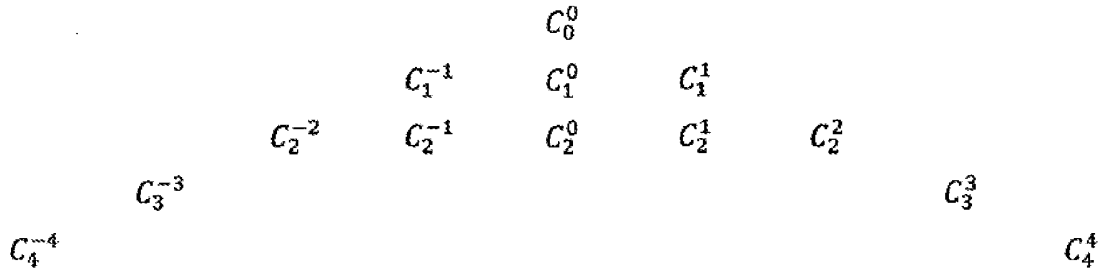
Os coeficientes de HOA para uma amostra no tempo incluem o *TrackHOAParamNumberOfCoeffs(0)* número de coeficientes  $C_n^m$ . O  $N$  depende da dimensão dos coeficientes de HOA. Para campos sonoros 2D 'O' é igual a  $2N + 1$  onde  $N$  é igual ao campo *TrackHOAParamHorizontalOrder* do cabeçalho *TrackHOAParam*. Os coeficientes de 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 se apresenta no Quadro 10.

Para campos sonoros 3D,  $O$  é igual a  $(N + 1)^2$  onde  $N$  é igual ao campo *TrackHOAParamVerticalOrder* do cabeçalho *TrackHOAParam*. Os coeficientes de HOA 3D  $C_n^m$  são definidos para  $0 \leq n \leq N$  e  $-n \leq m \leq n$ . É dada no Quadro 10 uma representação comum dos coeficientes de HOA:

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

Quadro 10 - Representação de coeficientes de HOA até à quarta ordem apresentando os coeficientes 2D em negrito como um subconjunto dos coeficientes 3D

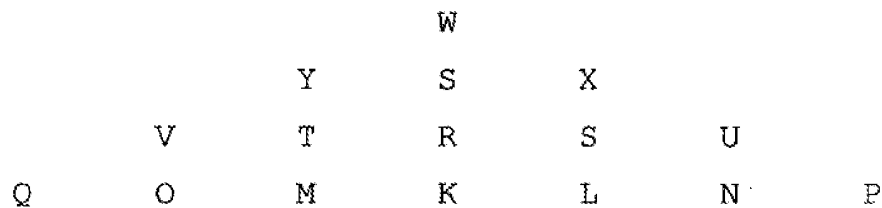
No caso de campos sonoros 3D e com *TrackHOAParamHorizontalOrder* maior do que *TrackHOAParamVerticalOrder*, será realizada a descodificação de ordem mista. Em sinais de ordem mista, alguns coeficientes de ordem superior são transmitidos apenas em 2D. O campo *TrackHOAParamVerticalOrder* determina a ordem vertical onde todos os coeficientes são transmitidos. A partir da ordem vertical para a *TrackHOAParamHorizontalOrder* apenas são utilizados os coeficientes 2D. Assim, a *TrackHOAParamHorizontalOrder* é igual ou maior do que a *TrackHOAParamVerticalOrder*. Está representado no Quadro 11 um exemplo de uma representação de ordem mista de uma ordem horizontal de quatro e uma ordem vertical de dois:



Quadro 11 - Representação de coeficientes de HOA para uma representação de ordem mista de ordem vertical dois e ordem horizontal quatro.

Os coeficientes de HOA  $C_n^m$  são armazenados nos Pacotes de uma Pista. A sequência dos coeficientes, por exemplo qual coeficiente vem primeiro e qual se segue, tem sido definida diferentemente no passado. Portanto, o campo *TrackHOAParamCoeffSequence* indica três tipos de sequências de coeficientes. As três sequências são derivadas a partir do arranjo de coeficientes de HOA do Quadro 10.

A sequência no Formato B utiliza uma expressão especial para os coeficientes HOA até à ordem de três como se apresenta no Quadro 12:



Quadro 12 - Convenções de designação dos coeficientes de HOA no Formato B.

Para o Formato B, os coeficientes de HOA são transmitidos a partir da ordem mais baixa para a mais elevada, em que os coeficientes de HOA de cada ordem são transmitidos por ordem alfabética. Por exemplo, os coeficientes de um arranjo 3D da ordem de HOA três são armazenados pela sequência W, X, Y, S, R, S, T, U, V, K, L, M, N, O, P e Q. O formato B é definido apenas até à terceira ordem de HOA. Para a transmissão dos coeficientes horizontais (2D) os coeficientes suplementares de 3D são ignorados, por exemplo W, X, Y, U, V, P, Q.

Os coeficientes  $C_n^m$  para HOA 3D são transmitidos em *TrackHOAParamCoeffSequence* de uma maneira numérica ascendente ou descendente a partir do ordem de HOA mais baixa para a mais elevada ( $n = 0 \dots N$ ). A sequência numérica ascendente começa com  $m = -n$  e cresce até  $m = n$  ( $C_0^0, C_1^{-1}, C_1^0, C_1^1, C_2^{-2}, C_2^{-1}, C_2^0, C_2^1, C_2^2, \dots$ ), que é a sequência 'CG' definida em Chris Travis, "Four candidate component sequences",

<http://ambisonics.googlegroups.com/web/Four+candidate+component+sequences+V09.pdf>, 2008. A sequência numérica descendente  $m$  desenvolve-se pela ordem inversa de  $m = n$  até  $m = -n$  ( $C_0^0, C_1^1, C_1^0, C_1^{-1}, C_2^2, C_2^1, C_2^0, C_2^{-1}, C_2^{-2}, \dots$ ), que é a sequência 'QM' definida naquela publicação.

Para coeficientes de HOA 2D as *TrackHOAParamCoeffSequence* sequências numéricas ascendente e descendente são como no caso 3D, mas em que os coeficientes não utilizados com  $|m| \neq n$  (isto é, apenas os

coeficientes de HOA sectoriais  $C_{|m|}^m \equiv C_m$  do Quadro 10) são omitidos. Assim, a sequência numérica ascendente conduz a  $(C_0^0, C_1^{-1}, C_1^1, C_2^{-2}, C_2^2, \dots)$  e a sequência numérica descendente a  $(C_0^0, C_1^1, C_1^{-1}, C_2^2, C_2^{-2}, \dots)$ .

### ***Pacotes de Pista***

### ***Pacotes de Pista HOA***

### ***Pacote de Tipo de Codificação PCM***

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
<PacketHOACoeffs>	dyn	dyn	Coeficientes de HOA interlaçados de canal 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)]>

Este Pacote contém os coeficientes de HOA  $C_n^m$  pela ordem definida no *TrackHOAParamCoeffSequence*, em que todos os coeficientes de uma amostra no tempo são transmitidos sucessivamente. Este Pacote é utilizado para Pistas HOA padronizadas com um *TrackSourceType* de zero e um *TrackCodingType* de zero.

### ***Pacote de Tipo de Codificação de Resolução***

**Dinâmica**

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
<PacketHOACoeffsCoded>	dyn	dyn	Coefficientes de HOA desinterlaçados de canal armazenados de acordo com o <i>TrackCodingType</i> , 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), ...]>

O empacotamento de resolução dinâmica é utilizado para um *TrackSourceType* de 'zero' e um *TrackCodingType* de 'um'. As diferentes resoluções das *TrackOrderRegions* conduz a diferentes dimensões de armazenamento para cada *TrackOrderRegion*. Portanto, os coeficientes de HOA são armazenados de uma maneira desentrelaçada, por exemplo todos os coeficientes de uma ordem de HOA são armazenados sucessivamente.

**Pacotes de Pista de Fonte Única****Pacote de Fonte Única de Posição Fixa**

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
<PacketMonoPCMTrack>	dyn	dyn	Amostras PCM da fonte de áudio única armazenadas em <i>TrackSampleFormat</i>

O Pacote de Fonte Única de Posição Fixa é utilizado para um *TrackSourceType* de 'um' e um *TrackMovingSourceFlag* de 'zero'. O Pacote contém as amostras PCM de uma fonte mono.

**Pacote de Fonte Única de Posição Móvel**

Nome do Campo	Tam. /Bit	Tipo de Dados	Descrição
PacketDirectionFlag	1	binário	Definido como '1' se a direcção foi alterada. '1' é mandatório para o primeiro Pacote de uma Trama.
Reservado	7	binário	Bits de enchimento
Condição: PacketDirctionFlag == '1' segue-se novos dados de posição			
Condição: TrackPositionType == '0' Posição TrackPositionVector como ângulo TrackPositionVector			
Theta	32	float32	inclinação em rad [0, pi]
Phi	32	float32	azimute (sentido contrário ao do relógio) em rad [0, 2pi]
Raio	32	float32	Distância ao ponto de referência em metros
Condição: TrackPositionType == '1' Posição como vector de codificação de HOA			
Condição: TrackEncodeVectorFormat == '0' Vector de codificação como float32			
<TrackHOAEncodingVector>	dyn	float32	Entradas de TrackHOAParamNumberOfCoeffs do vector de codificação de HOA pela ordem TrackHOAParamCoeffSequence
Condição: TrackEncodeVectorFormat == '1' Vector de codificação como float64			
<TracHOAEncodingVector>	dyn	float64	Entradas de TrackHOAParamNumberOfCoeffs do vector de codificação de HOA pela ordem TrackHOAParamCoeffSequence
<PacketMonoPCMTrack>	dyn	dyn	Amostras PCM da fonte de áudio única armazenadas em TrackSampleFormat

O Pacote de Fonte Única de Posição Móvel é utilizado para um *TrackSourceType* de 'um' e uma *TrackMovingSourceFlag* de 'um'. Ele contém as amostras PCM mono e a informação de posição para a amostra do *TrackPacket*.

A *PacketDirectionFlag* indica se a direcção do

Pacote mudou ou se deverá ser utilizada a direcção do Pacote anterior. Para assegurar a descodificação a partir do início de cada Trama, a *PacketDirectionFlag* é igual a 'um' para o primeiro *TrackPacket* de fonte móvel de uma Trama.

Para uma *PacketDirectionFlag* de 'um', é transmitida a informação de direcção da seguinte amostra PCM da fonte. Dependendo do *TrackPositionType*, é enviada a informação de direcção como *TrackPositionVector* em coordenadas esféricas ou como *TrackHOAEncodingVector* com o *TrackEncodingVectorFormat* definido. O *TrackEncodingVector* gera Coeficientes HOA que estão conformes com as definições do campo *HOAParamHeader*. Consecutivamente à informação direccional são transmitidas as Amostras PCM mono do *TrackPacket*.

### ***Processamento da Codificação***

#### ***Codificação do TrackRegion***

Os sinais HOA podem ser derivados a partir de gravações Sounfield com sistemas de microfones. Por exemplo, o Eigenmike divulgado na Patente WO 03/061336 A1 pode ser utilizado para obter gravações de HOA de ordem três. No entanto, a dimensão finita do sistema de microfones conduz a restrições para os coeficientes de HOA gravados. Na Patente WO 03/061336 A1 e no artigo mencionado acima "Three-dimensional surround sound systems based on



spherical harmonics” são discutidas questões causadas por sistemas de microfones finitos.

A distância das cápsulas dos microfones resultam num limite de frequência superior dado pelo teorema de amostragem espacial. Acima desta frequência superior o sistema de microfones não pode produzir coeficientes de HOA correctos. Além disso, a distância finita dos microfones à posição de audição HOA requer um filtro de equalização. Estes filtros obtêm altos ganhos para frequências baixas que ainda aumentam com cada ordem de HOA. Na Patente WO 03/061336 A1, é introduzida uma frequência de corte baixa para os coeficientes de ordem superior de forma a dominar a gama dinâmica do filtro de equalização. Isto mostra que a largura de banda de coeficientes de HOA de diferentes ordens de HOA podem diferir. Portanto, o formato de ficheiro HOA oferece a *TrackRegionBandwidthReduction* que permite a transmissão de apenas a largura de banda requerida para cada ordem de HOA. Devido à elevada gama dinâmica do filtro de equalização e devido ao facto de que o coeficiente de ordem zero é basicamente a soma de todos os sinais de microfone, os coeficientes de diferentes ordens de HOA podem ter diferentes gamas dinâmicas. Portanto, o formato de ficheiro HOA oferece também a característica de adaptar o tipo de formato à gama dinâmica de cada ordem de HOA.

Como se apresenta na Figura 12, os coeficientes de HOA entrelaçados alimentam o primeiro passo de desentrelaçamento ou andar 1211, que está atribuído ao primeiro *TrackRegion* e separa todos os coeficientes de HOA da *TrackRegion* em memórias tampão desentrelaçadas para amostras *FramePacketSize*. Os coeficientes da *TrackRegion* são derivados a partir do *TrackRegionLastOrder* e do campo *TrackRegionFirstOrder* do Cabeçalho de Pista HOA. Desentrelaçar significa que coeficientes  $C_n^m$  para uma combinação de  $n$  e  $m$  são agrupados numa memória tampão. A partir do passo de desentrelaçamento ou andar 1211 os coeficientes de HOA desentrelaçados são passados para a secção de codificação *TrackRegion*. Os restantes coeficientes de HOA entrelaçados são passados para o seguinte passo ou andar de desentrelaçamento *TrackRegion*, e por aí fora até ao passo ou andar de desentrelaçamento 121N. O número N de passos ou andares de desentrelaçamento é igual ao *TrackNumberOfOrderRegions* mais 'um'. O passo ou andar adicional de desentrelaçamento 125 desentrelaça os coeficientes restantes que não fazem parte do *TrackRegion* num percurso padronizado de processamento incluindo um passo ou andar 126 de conversão de formato.

O percurso de codificação de *TrackRegion* inclui um passo ou andar 1221 opcional de redução de largura de banda e um passo ou andar 1231 de conversão de formato e realiza um processamento paralelo para cada memória tampão de coeficientes de HOA. A redução da largura de banda é realizada se o campo *TrackRegionUseBandwidthReduction* está

definido como 'um'. Dependendo do *TrackBandwidthReductionType* seleccionado, é seleccionado um processamento para limitar a gama de frequência dos coeficientes de HOA e para os sub-amostrar criticamente. Isto é realizado de forma a reduzir o número de coeficientes de HOA ao número mínimo requerido de amostras. A conversão de formato converte o corrente formato dos coeficientes de HOA para o *TrackRegionSampleFormat* definido no cabeçalho do *HOATrack*. Este é o único passo/andar no percurso de processamento padronizado que converte os coeficientes de HOA para o *TrackSampleFormat* indicado do Cabeçalho de Pista HOA.

O passo ou andar 124 de multiplexação de *TrackPacket* multiplexa as memórias tampão dos coeficientes de HOA para a corrente de bits do ficheiro de dados *TrackPacket* como definido no campo *TrackHOAParamCoeffSequence* seleccionado, em que os coeficientes  $C_n^m$  para uma combinação de  $n$  e  $m$  índices permanecem desentrelaçados (dentro de uma memória tampão).

#### ***Processamento da Descodificação de TrackRegion***

Como se apresenta na Figura 13, o processamento de descodificação é inverso do processamento de codificação. O passo ou andar 134 de desmultiplexação desmultiplexa a corrente de bits ou ficheiro de dados *TrackPacket* a partir do *TrackHOAParamCoeffSequence* indicado para memórias tampão de coeficientes de HOA desentrelaçados

(não representado). Cada memória tampão contém *FramePacketLenght* coeficientes  $C_n^m$  para uma combinação de  $n$  e  $m$ .

O passo/andar 134 inicializa *TRackNumberOfOrderRegion* mais 'um' percursos de processamento e passa o conteúdo das memórias tampão dos coeficientes de HOA desentrelaçados para o percurso de processamento apropriado. Os coeficientes de cada *TrackRegion* são definidos pelos campos *TrackRegionLastOrder* e *TRackRegionFirstOrder* do Cabeçalho de Pista HOA. As ordens de HOA que não sejam cobertas pelas *TrackRegions* seleccionadas são processadas no percurso de processamento padrão incluindo um passo ou andar 136 de conversão de formato e um restante passo ou andar 135 de entrelaçamento de coeficientes. O percurso padrão de processamento corresponde a um percurso *TrackProcessing* sem um passo ou andar de redução de largura de banda.

Nos percursos de *TrackProcessing*, um passo/andar de conversão de formato de 1331 a 133N converte os coeficientes de HOA que estão codificados no *TrackRegionSampleFormat* no formato de dados que é utilizado para o processamento do descodificador. Dependendo do campo de dados *TrackRegionUseBandwidthReduction*, segue-se um passo ou andar 1321 a 1132N opcional de reconstrução da largura de banda no qual os coeficientes de HOA limitados em banda e amostrados criticamente são reconstruídos para a total largura de banda da Pista. O tipo de processamento de

reconstrução é definido no campo *TrackBandwidthReductionType* do Cabeçalho de Pista HOA. No passo ou andar 1311 a 131N seguinte de desentrelaçamento, o conteúdo das memórias tampão desentrelaçadas dos coeficientes de HOA são entrelaçadas agrupando coeficientes de HOA de uma amostra no tempo, e os coeficientes de HOA da corrente *TrackRegion* são combinados com os coeficientes de HOA das anteriores *TrackRegions*. A sequência resultante dos coeficientes de HOA pode ser adaptada ao processamento da Pista. Além disso, os passos/andares de entrelaçamento operam com os atrasos entre as *TrackRegions* utilizando a redução de largura de banda e com as *TRackRegions* que não utilizam a redução de largura de banda, atrasos que dependem do processamento *TrackBandwidthReductionType* seleccionado. Por exemplo, o processamento MDCT adiciona um atraso de *FramePacketSize* amostras e, portanto, os passos/andares de entrelaçamento dos percursos de processamento sem redução da largura de banda atrasarão as suas saídas em um pacote.

### ***Redução da largura de banda por via de MDCT***

#### ***Codificação***

A Figura 14 apresenta a redução da largura de banda utilizando o processamento MDCT (transformada de cosseno discreta modificada). Cada coeficiente de HOA da *TrackRegion* das amostras *FramePacketSize* passa por via de uma memória tampão 1411 a 141M um passo ou andar 1421 a 142M de adição de uma janela MDCT correspondente. Cada

memória tampão de entrada contém os coeficientes de HOA sucessivos temporais  $C_n^m$  de uma combinação de  $n$  e  $m$ , isto é, uma memória tampão é definida como

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

O número  $M$  de memórias tampão é o mesmo do número de componentes Ambisonics  $((N + 1)^2$  para um campo sonoro 3D completo de ordem  $N$ ). O tratamento pela memória tampão realiza uma sobreposição em 50% para o seguinte processamento de MDCT combinando o conteúdo da memória tampão anterior com o conteúdo da memória tampão actual para um novo conteúdo para o processamento de MDCT em passos ou andares 1431 a 143M correspondentes, e ele armazena o conteúdo da memória tampão actual para o processamento do conteúdo da memória tampão seguinte. O processamento de MDCT re-inicia no início de cada Trama, o que significa que todos os coeficientes de uma Pista da Trama corrente podem ser descodificados sem o conhecimento da Trama anterior, e a seguir ao conteúdo da última memória tampão da Trama corrente é processado um conteúdo adicional de uma memória tampão de zeros. Portanto, as *TrackRegions* processadas por MDCT produzem um *TrackPacket* extra.

Nos passos/andares de adição de janela o correspondente conteúdo de memória tampão é multiplicado com a função de janela  $w(t)$  seleccionada, que é definida no campo de cabeçalho *HOATrack TrackRegionWindowType* para cada *TrackRegion*.

A Transformada de Cosseno Discreta Modificada é mencionada a primeira vez 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, páginas 1153-1161, Outubro de 1986. A MDCT pode ser considerada como representando um banco de filtros amostrado criticamente das sub-bandas *FramePacketSize*, e ela requer uma sobreposição em 50% da memória tampão de entrada. A memória tampão de entrada tem um comprimento de duas vezes a dimensão da sub-banda. A MDCT é definida pela seguinte equação com  $T$  igual a *FramePacketSize*:

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

Os coeficientes  $C'_n{}^m(k)$  são chamados bins de MDCT. O cálculo da MDCT pode ser implementado utilizando a Transformada Rápida de Fourier. Nos passos ou andares 1441 a 144M seguintes de corte de região de frequência é realizada a redução de largura de banda removendo todos os bins de MDCT  $C'_n{}^m(k)$  com  $k < TrackRegionFirstBin$  e  $k > TrackRegionLastBin$ , para a redução do comprimento da memória tampão para  $TrackRegionLastBin - TrackRegionFirstBin + 1$ , em que  $TrackRegionFirstBin$  é a frequência de corte inferior para a *TrackRegion* e  $TrackRegionLastBin$  é a frequência de corte superior. A falta de bins de MDCT pode ser vista como representando um filtro passa-banda com frequências de corte correspondentes

às frequências *TrackRegionLastBin* e *TrackRegionFirstBin*. Portanto, apenas são transmitidos os bins de MDCT requeridos.

### **Descodificação**

A Figura 15 apresenta a descodificação ou reconstrução da largura de banda utilizando o processamento de MDCT, no qual os coeficientes de HOA de largura de banda limitada *TrackRegions* são reconstruídos para as completas larguras de banda da Pista. Esta reconstrução da largura de banda processa o conteúdo da memória tampão de coeficientes de HOA desentrelaçados temporalmente em paralelo, em que cada memória tampão contém *TrackRegionLastBin* - *TrackRegionFirstBin* + 1 bins de MDCT de coeficientes  $\mathbf{C}'_n{}^m(\mathbf{k})$ . Os passos ou andares 1541 a 154M de adição das regiões de frequência em falta reconstroem o conteúdo completo da memória tampão de MDCT de dimensão *FramePacketLenght* complementando os bins de MDCT recebidos com os bins de MDCT em falta  $\mathbf{k} < \textit{TrackRegionFirstBin}$  e  $\mathbf{k} > \textit{TrackRegionLastBin}$  utilizando zeros. Depois disso, é realizada a transformada inversa de MDCT em passos ou andares 1531 a 153M correspondentes de inversa de MDCT de forma a reconstruir os coeficientes de HOA no domínio do tempo  $\mathbf{C}_n{}^m(\mathbf{t})$ . A MDCT inversa pode ser interpretada como um banco de filtro de síntese em que *FramePacketLenght* bins de MDCT são convertidos para duas vezes *FramePacketLenght* coeficientes no domínio do tempo. No entanto, a reconstrução completa das amostras no domínio do tempo



requer uma multiplicação com a função de janela  $w(t)$  utilizada no codificador e uma adição de sobreposição da primeira metade do conteúdo da memória tampão corrente com a segunda metade do conteúdo da anterior memória tampão. A MDCT inversa é definida pela seguinte equação:

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

Tal como a MDCT, a MDCT inversa pode ser implementada utilizando a Transformada Rápida de Fourier inversa.

Os passos ou andares 1521 a 152M de adição de janela de MDCT multiplicam os coeficientes no domínio do tempo reconstruídos com a função janela definida por *TrackRegionWindowType*. As memórias tampão 1511 a 151M seguintes adicionam a primeira metade do conteúdo da memória tampão *TrackPacket* corrente com a segunda metade do conteúdo da última memória tampão *TrackPacket* de forma a reconstruir os *FramePacketSize* coeficientes no domínio do tempo. A segunda metade do conteúdo da memória tampão *TrackPacket* corrente é armazenada para o processamento do *TrackPacket* seguinte, cujo processamento de adição de sobreposição remove as componentes de serrilhamento contrárias de ambos os conteúdos das memórias tampão.

Para ficheiros HOA com multi-Tramas, o codificador está proibido de utilizar o conteúdo da última

memória tampão da trama anterior para o procedimento de adição de sobreposição no início de uma nova Trama. Portanto, nos limites das Tramas ou no início de uma nova Trama falta o conteúdo da memória tampão de adição de sobreposição, e a reconstrução do primeiro *TrackPacket* de uma Trama pode ser realizada no segundo *TrackPacket*, pelo que é introduzido um atraso de um *FramePacket* e de descodificação de um *TrackPacket* extra em comparação com os percursos de processamento sem redução de largura de banda. Este atraso é tratado pelos passos/andares de entrelaçamento descritos em relação à Figura 13.

Lisboa, 3 de Outubro de 2014

**REIVINDICAÇÕES**

1. Estrutura de dados para dados de áudio HOA Ambisonics de Ordem Superior incluindo coeficientes de Ambisonics, estrutura de dados que inclui dados com conteúdo de áudio espacial 2D e/ou 3D para diferentes descrições em correntes de bits de dados de áudio HOA, e estrutura de dados que é também adequada para dados de áudio de HOA que tenham uma ordem maior do que '3', e estrutura de dados que pode adicionalmente incluir dados de fonte de sinal áudio única e/ou dados de áudio de sistema de microfones provenientes de posições espaciais fixas ou variantes no tempo,

em que as referidas diferentes descrições de correntes de bits de dados de áudio de HOA estão relacionadas com diferentes densidades de posicionamento de altifalantes e pelo menos um de tipos de onda HOA codificada, ordens de HOA e dimensionalidade de HOA,

e em que uma descrição em corrente de bits de dados de áudio de HOA contém dados de áudio para uma representação com um arranjo denso de altifalantes (11, 21) localizado numa área distinta de um local de apresentação (10), e uma outra descrição em corrente de bits de dados de áudio de HOA contém dados de áudio para uma apresentação com um arranjo de altifalantes menos denso (12, 22) envolvendo o referido local de apresentação (10).

2. Estrutura de dados de acordo com a Reivindicação 2, em que os referidos dados de áudio para o referido arranjo de altifalantes denso (11, 21) representam ondas esféricas e uma primeira ordem de Ambisonics, e os referidos dados de áudio para o referido arranjo de altifalantes menos denso (12, 22) representam ondas planas e/ou uma segunda ordem de Ambisonics mais pequena do que a referida primeira ordem de Ambisonics.

3. Estrutura de dados de acordo com a Reivindicação 1 ou 2, em que a referida estrutura de dados serve como descrição de cena onde pistas de uma cena de áudio podem começar e terminar em qualquer altura.

4. Estrutura de dados de acordo com uma das Reivindicações de 1 a 3, em que a referida estrutura de dados inclui itens de dados relacionados com:

- região de interesse relacionada com fontes de áudio fora ou dentro de uma área de audição;
- normalização de funções de base esférica;
- directividade de propagação;
- informação de redimensionamento de coeficientes de Ambisonics;
- tipo de onda de Ambisonics, por exemplo plana ou

esférica;

- no caso de ondas esféricas, raio de referência para a descodificação.

5. Estrutura de dados de acordo com uma das Reivindicações de 1 a 4, em que os referidos coeficientes de Ambisonics são coeficientes complexos.

6. Estrutura de dados de acordo com uma das Reivindicações de 1 a 5, incluindo a referida estrutura de dados metadados relacionados com as direcções e características de um ou mais microfones, e/ou incluindo pelo menos um vector de codificação para sinais de entrada de fonte única.

7. Estrutura de dados de acordo com uma das Reivindicações de 1 a 6, em que pelo menos parte dos referidos coeficientes de Ambisonics estão com a largura de banda reduzida, de forma que para diferentes ordens de HOA a largura de banda dos coeficientes de Ambisonics relacionados é diferente (1221-122N).

8. Estrutura de dados de acordo com a Reivindicação 7, em que a referida redução de largura de banda está baseada no processamento de MDCT (1431-143M).

9. Método para codificar e organizar dados numa

estrutura de dados de acordo com uma das Reivindicações de 1 a 8.

10. Método para apresentação de áudio, em que uma corrente de bits de dados de áudio de HOA Ambisonics de Ordem Superior contendo pelo menos dois diferentes sinais de dados de áudio de HOA é recebida e é utilizada pelo menos uma primeira delas (231, 232) para apresentação com um arranjo denso de altifalantes (11, 21) localizado numa área distinta de um local de apresentação (10), e é utilizada pelo menos uma segunda e diferente delas (241, 242, 243) para apresentação com um arranjo menos denso de altifalantes (12, 22) envolvendo o referido local de apresentação (10).

11. Método de acordo com a Reivindicação 10, em que os referidos dados de áudio para o referido arranjo denso de altifalantes (11, 21) representam ondas esféricas e uma primeira ordem de Ambisonics, e os referidos dados de áudio para o referido arranjo de altifalantes menos denso (12, 22) representam ondas planas e/ou uma segunda ordem de Ambisonics mais pequena do que a referida primeira ordem de Ambisonics.

12. Estrutura de dados de acordo com a Reivindicação 1 ou 2, ou método de acordo com a Reivindicação 10 ou 11, em que o referido local de apresentação é uma área de audição ou de lugares sentados num cinema.

13. Dispositivo estando adaptado para levar a cabo o método da Reivindicação 10 ou 11.

Lisboa, 3 de Outubro de 2014

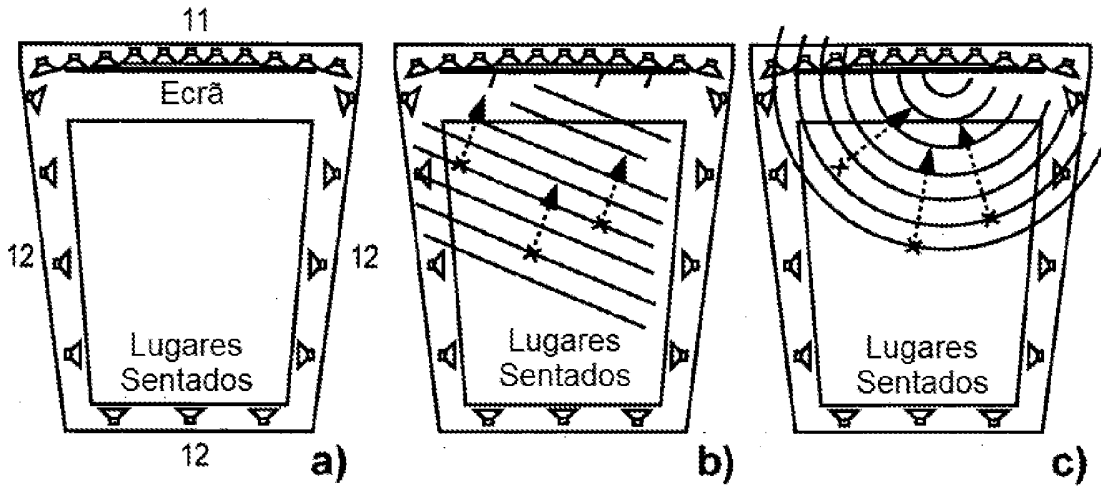


Fig. 1

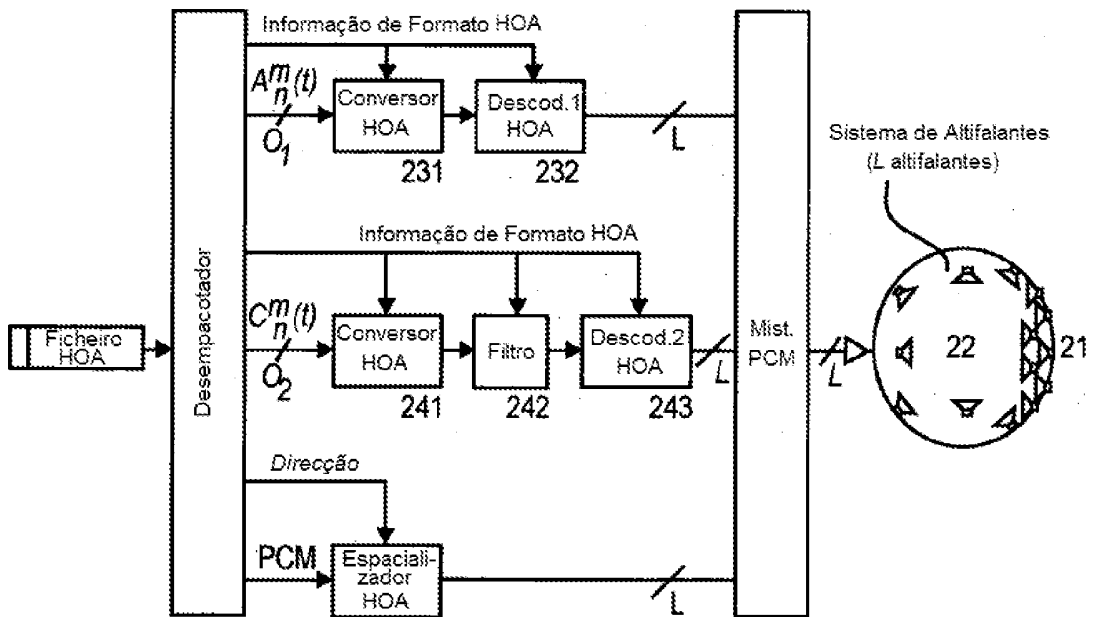


Fig. 2



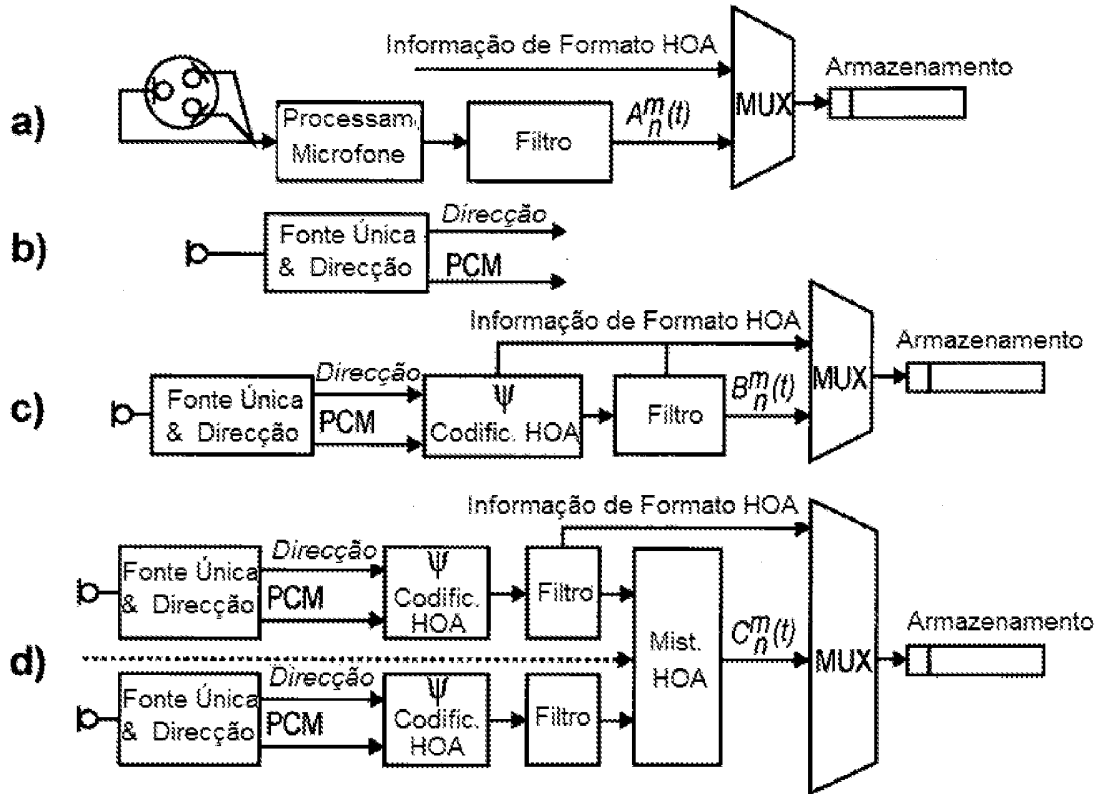


Fig. 3

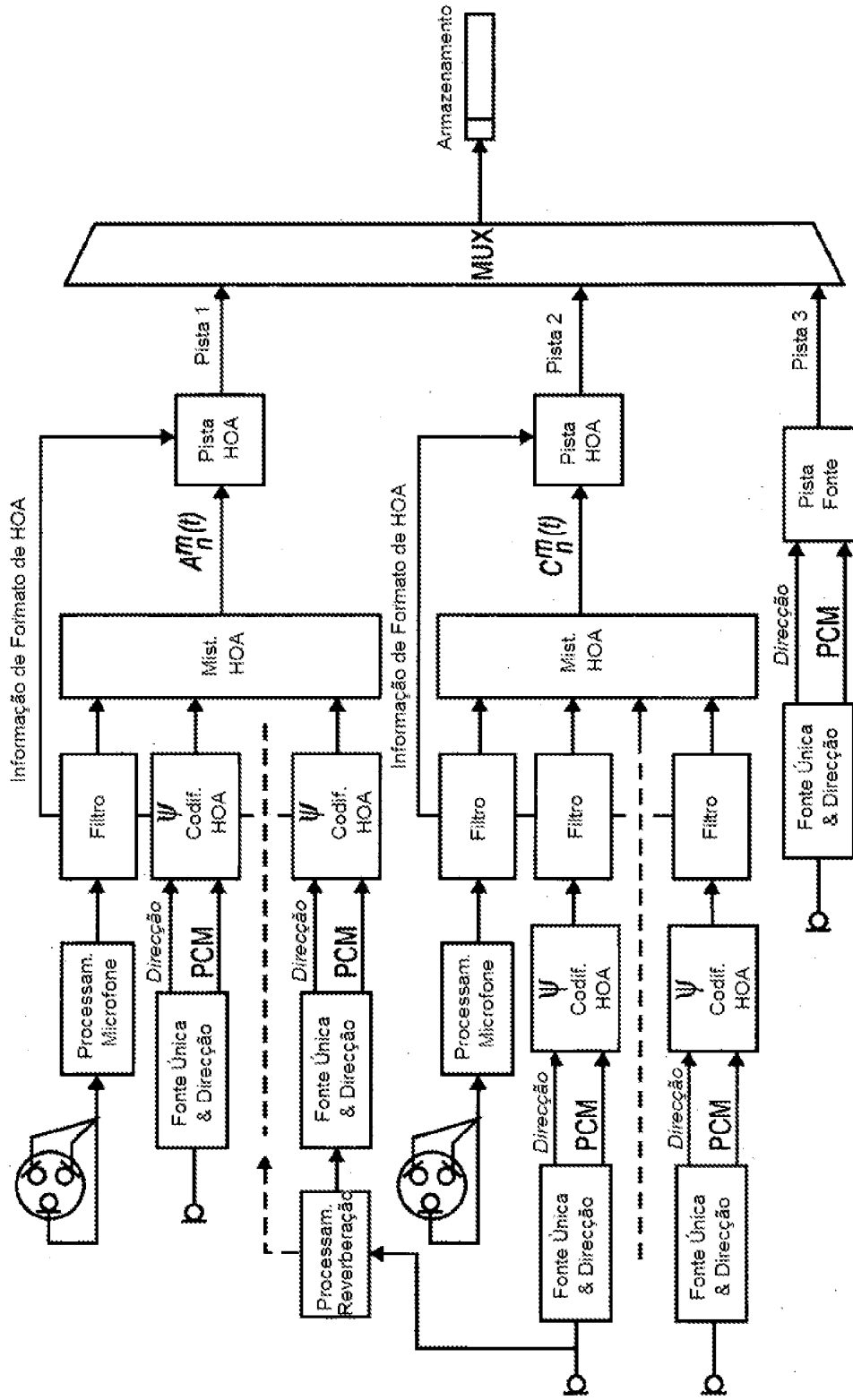


Fig. 4

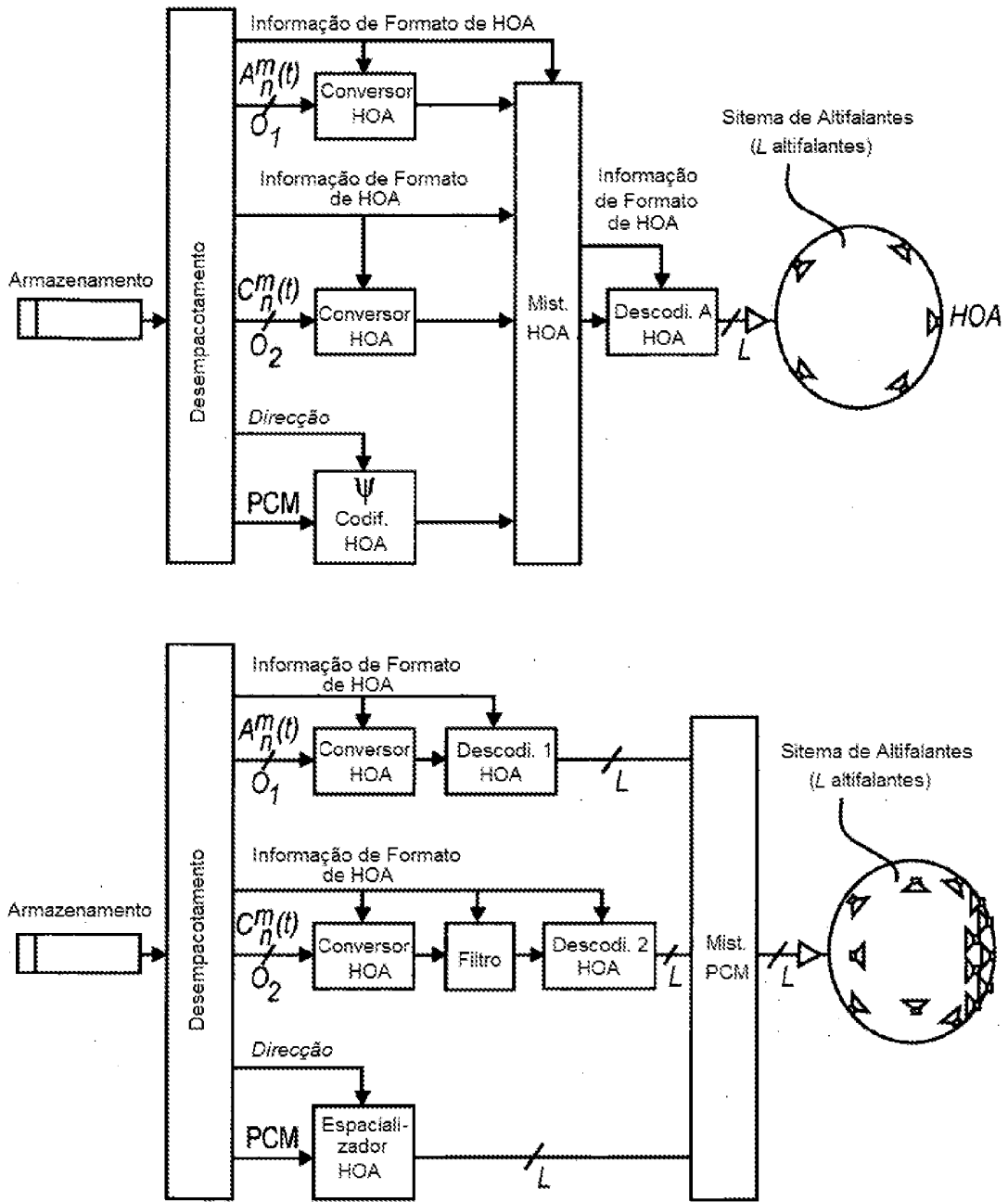


Fig. 5

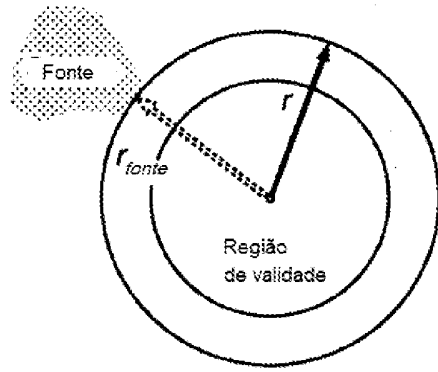


Fig. 6

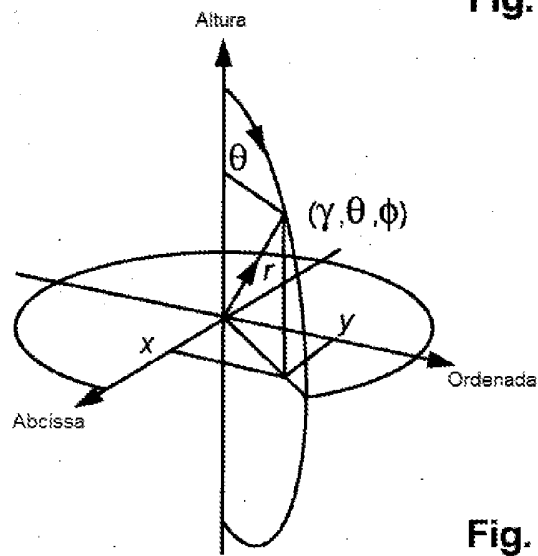


Fig. 7

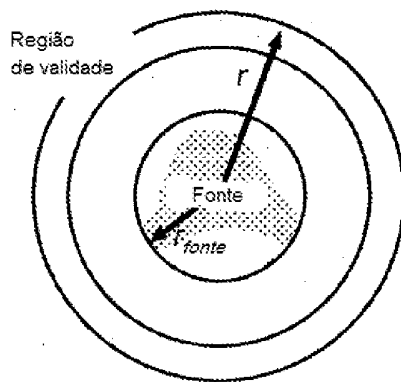


Fig. 8

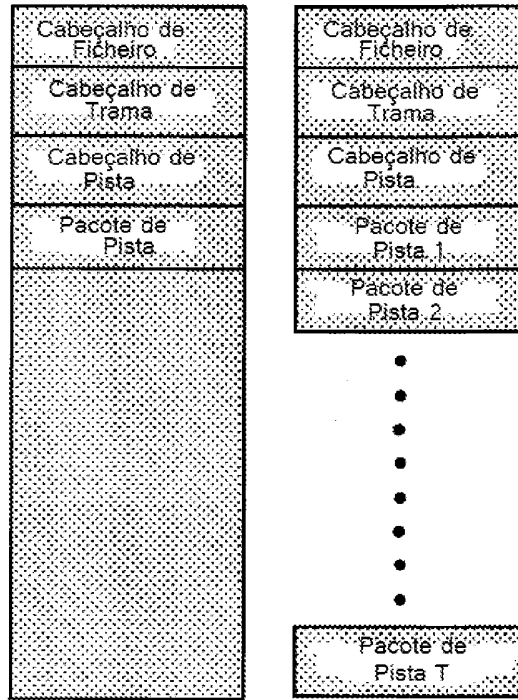


Fig. 9

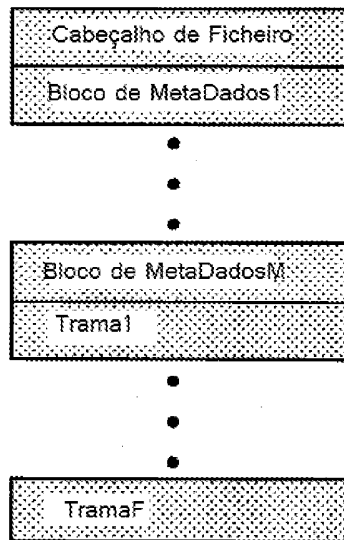


Fig. 11

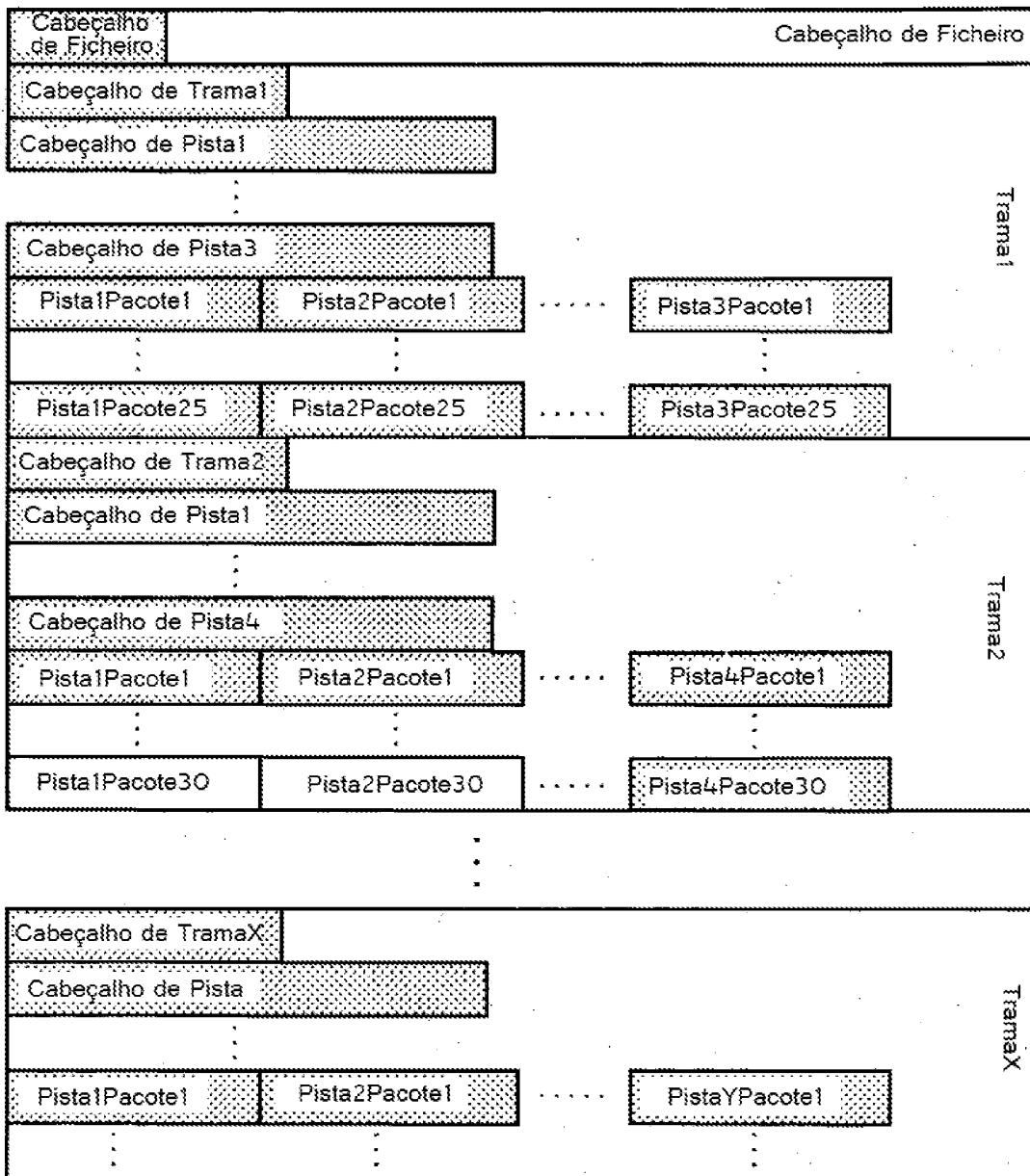


Fig. 10

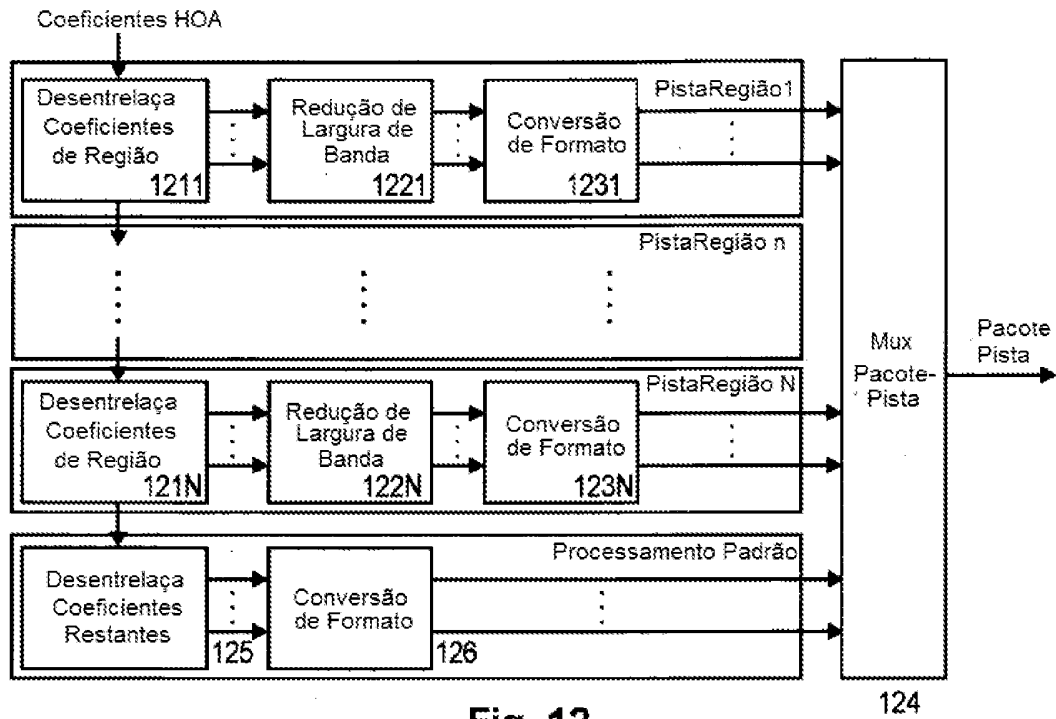


Fig. 12

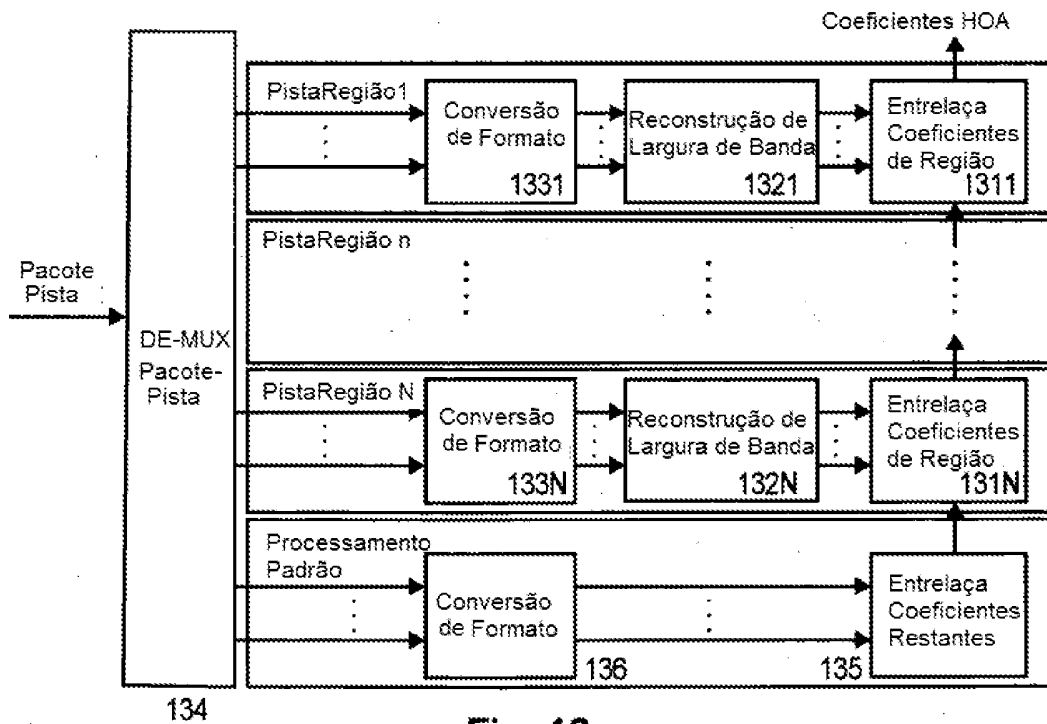


Fig. 13

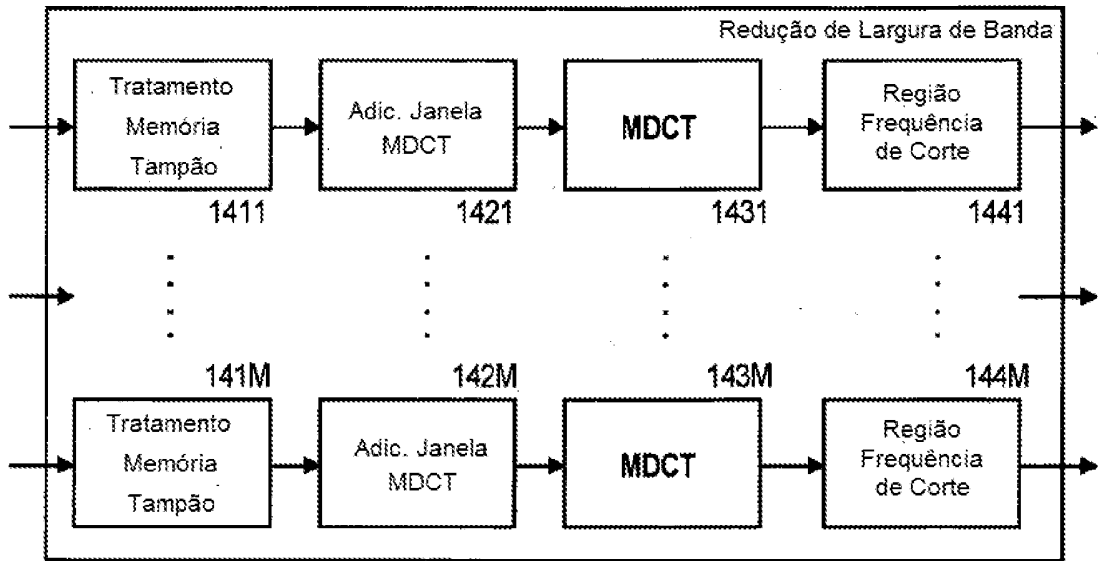


Fig. 14

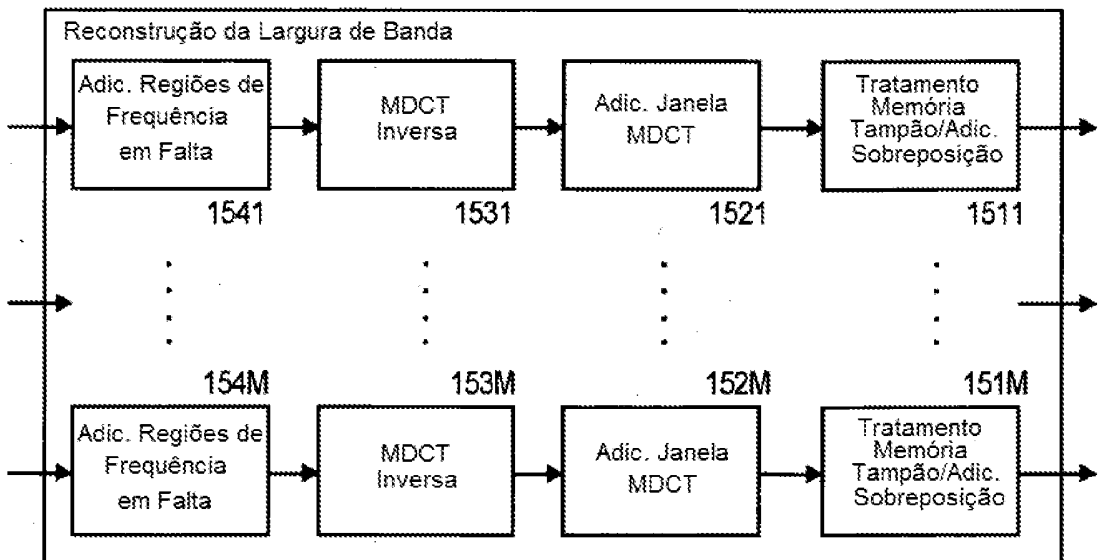


Fig. 15



## REFERÊNCIAS CITADAS NA DESCRIÇÃO

*Esta lista de referências citadas pelo requerente é apenas para conveniência do leitor. A mesma não faz parte do documento da patente Europeia. Ainda que tenha sido tomado o devido cuidado ao compilar as referências, podem não estar excluídos erros ou omissões e o IEP declina quaisquer responsabilidades a esse respeito.*

### Documentos de patentes citadas na Descrição

- EP 2205007 A1
- US 4042779 A
- WO 03061336 A1

### Literatura que não é de patentes citada na Descrição

- **MARTIN LEESE.** *File Format for B-Format*, 30 March 2009, <http://www.ambisonics.com/Members/eliemne/Members/mleese/file-format-for-b-format>
- **R.W. DOBSON.** *Developments in Audio File Formats. ICMC Berlin, 2000*
- *Existing formats. Ambisonics Xchange Site, <http://ambisonics.iem.at/xchange/format/existing-formats>*
- **M.A. POLETTI.** *Three-dimensional surround sound systems based on spherical harmonics. Journal of Audio Engineering Society, November 2005, vol. 53 (11), 1004-1026*
- **EARL G. WILLIAMS.** *Fourier Acoustics.* Academic Press, 1999
- **MARK POLETTI.** *Unified description of Ambisonics using real and complex spherical harmonics. Proceedings of the Ambisonics Symposium 2009, June 2009*
- **WILLIAM H. PRESS ; SAULA TEUKOLSKY ; WILLIAM T. VETTERLING ; BRIAN P. FLANNERY.** *Numerical Recipes in C.* Cambridge University Press, 1992
- **JÉRÔME DANIEL.** *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia. PhD thesis, Université Paris 6, 2001*
- **DAVE MALHAM.** *3-D acoustic space and its simulation using ambisonics, [http://www.dxarts.washington.edu/courses/587/current/malham\\_3d.pdf](http://www.dxarts.washington.edu/courses/587/current/malham_3d.pdf)*
- **JÉRÔME DANIEL.** *Spatial sound encoding including near field effect: introducing distance coding filters and a viable, new ambisonic format. AES 23rd International Conference, Denmark, 2003*
- **M.A. GERSON.** *General meta-theory of auditory localisation. 92th AES Convention, 1992*
- **JENS AHRENS ; SASCHA SPORS.** *Analytical driving functions for higher order ambisonics. Proceedings of the ICASSP, 2008, 373-376*
- *Associated Legendre polynomials, 12 October 2010, [http://en.wikipedia.org/w/index.php?title=Associated Legendre polynomials&oldid=363001511](http://en.wikipedia.org/w/index.php?title=Associated_Legendre_polynomials&oldid=363001511)*
- **CHRIS TRAVIS.** *Four candidate component sequences, 2008, [http://ambisonics.google-groups.com/web/Four\\_candidate\\_component\\_sequences+V09.pdf](http://ambisonics.google-groups.com/web/Four_candidate_component_sequences+V09.pdf)*
- **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, October 1986, vol. ASSP-34 (5), 1153-1161*