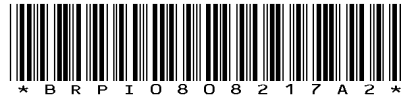


República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e da Comércio Exterior
Instituto Nacional de Propriedade Industrial

(21) PI 0808217-0 A2



* B R P I 0 8 0 8 2 1 7 A 2 *

(22) Data de Depósito: 01/02/2008
(43) Data da Publicação: 01/07/2014
(RPI 2269)

(51) Int.Cl.:
G10L 19/00
H04S 3/02

(54) Título: "MÉTODO E EQUIPAMENTO PARA CONVERSÃO ENTRE FORMATOS DE ÁUDIO MULTICANAL"

(57) Resumo:

(30) Prioridade Unionista: 30/04/2007 US 11/742,502, 21/03/2007 US 60/896,184, 21/03/2007 US 60/896,184

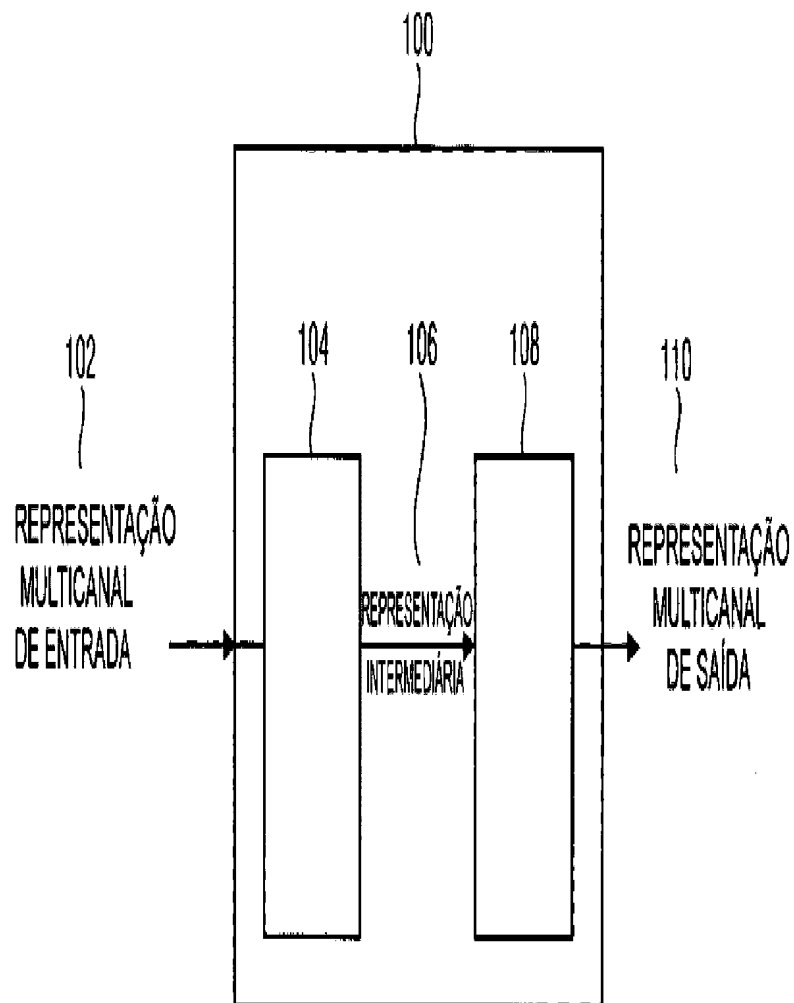
(73) Titular(es): FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V.

(72) Inventor(es): HERRE, JURGEN, VILLE PULKKI

(74) Procurador(es): Cruzeiro Newmarc Patentes e Marcas Ltda.

(86) Pedido Internacional: PCT EP2008000830 de 01/02/2008

(87) Publicação Internacional: WO 2008/113428de 25/09/2008



**"MÉTODO E EQUIPAMENTO PARA CONVERSÃO ENTRE
FORMATOS DE ÁUDIO MULTICANAL"**

Campo da Invenção

A presente invenção refere-se a uma técnica de
5 como converter entre diferentes formatos de áudio multicanal com a
qualidade mais alta possível sem limitar-se a representações
multicanal específicas, isto é, a presente invenção refere-se a
uma técnica que permite a conversão entre formatos multicanal
arbitrários.

10 Histórico da Invenção e método anterior

Em geral, em reprodução e escuta em multicanal, o
ouvinte é circundado por múltiplos alto-falantes. Existem vários
métodos para captar sinais de áudio para configurações
específicas. Um objetivo geral na reprodução é reproduzir a
15 composição espacial do evento de som registrado originalmente,
isto é, as origens das fontes de áudio individuais, como por
exemplo, a localização de um trompete dentro de uma orquestra.
Várias configurações de alto-falantes são bastante comuns, e podem
criar diferentes impressões espaciais. Sem usar técnicas de pós-
20 produção especiais, as configurações em estéreo de dois canais
comumente conhecidas podem somente recriar eventos auditivos em
uma linha entre os dois alto-falantes. Isto é conseguido
principalmente pela assim chamada "panorâmica de amplitude", onde
a amplitude do sinal associado a uma fonte de áudio é distribuída
25 entre os dois alto-falantes, dependendo da posição da fonte de
áudio em relação aos alto-falantes. Isto normalmente é feito
durante a gravação ou mixagem subsequente. Isto é, uma fonte de
áudio que vem da extremidade esquerda em relação à posição de

escuta, será reproduzida principalmente pelo alto-falante esquerdo, enquanto uma fonte de áudio em frente à posição de escuta será reproduzida com amplitude (nível) idêntica por ambos os alto-falantes. No entanto, o som que emanar de outras direções
5 não pode ser reproduzido.

Consequentemente, quando se usa mais alto-falantes que são distribuídos ao redor do ouvinte, mais direções podem ser abrangidas, e uma impressão espacial mais natural pode ser criada. O layout de alto-falante multicanal provavelmente
10 melhor conhecido é o padrão 5.1 (ITU-R775-1), que é composto de 5 alto-falantes cujos ângulos azimutais em relação à posição de escuta são pré-determinados em 0° , $\pm 30^\circ$ e $\pm 110^\circ$. Isso significa que durante a gravação ou mixagem, o sinal é customizado para aquela configuração específica de alto-falante, e desvios do
15 padrão de uma configuração de reprodução resultarão em uma redução na qualidade de reprodução.

Foram propostos também vários outros sistemas, com números variáveis de alto-falantes localizados em diferentes direções. Sistemas profissionais e especiais, especialmente em
20 teatros e instalações de som, também incluem alto-falantes em diferentes alturas.

Um sistema de reprodução de áudio universal chamado DirAC foi recentemente proposto, o qual é capaz de gravar e reproduzir som para configurações arbitrárias de alto-falante. A
25 finalidade do DirAC é reproduzir a impressão espacial de um ambiente acústico existente o mais precisamente possível, usando um sistema de alto-falante multicanal com configuração geométrica arbitrária. Dentro do ambiente de gravação, as respostas do

ambiente (que podem ser som gravado contínuo ou respostas de impulso) são medidas com um microfone onidirecional (W), e com um conjunto de microfones que permitem medir a direção de chegada do som e a difusibilidade do som. Nos parágrafos a seguir e dentro da aplicação, o termo "difusibilidade" deve ser compreendido como uma medida para a não-diretividade do som, isto é, o som que chega à posição de escuta ou gravação com potência igual de todas as direções, é maximamente difuso. Uma maneira comum de quantificar a difusão é usar valores de difusibilidade do intervalo $[0, \dots, 1]$, onde um valor de 1 descreve som maximamente difuso e um valor de 0 descreve som perfeitamente direcional, isto é, som que emana de somente uma direção claramente distinguível. Um método comumente conhecido de medir a direção de chegada do som é aplicar 3 microfones figura de oito (XYZ) alinhados com eixos de coordenada cartesiana. Microfones especiais, os chamados "microfones SoundField", foram projetados, os quais produzem diretamente todas as respostas desejadas. No entanto, conforme foi mencionado acima, os sinais W, X, Y e Z podem também ser computados de um conjunto de microfones onidirecionais discretos.

Outro método para armazenar formatos de áudio para número arbitrário de canais em um ou dois canais de áudio de "downmix" com dados direcionais acompanhados foi recentemente proposto por Goodwin e Jot. Este formato pode ser aplicado a sistemas de reprodução arbitrários. Os dados direcionais, isto é, os dados que contêm informações sobre a direção de fontes de áudio são computados usando-se "vetores Gerzon", que são compostos de um vetor de velocidade e um vetor de energia. O vetor de velocidade é uma soma ponderada de vetores voltados para alto-falantes da

posição de escuta, onde cada peso é a magnitude de um espectro de frequência em um determinado momento / "tile" de frequência de um alto-falante. O vetor de energia é uma soma de vetor ponderada de maneira semelhante. No entanto, os pesos são estimativas de energia de curto prazo dos sinais de alto-falante, isto é, eles descrevem um sinal de alguma maneira suavizado ou a íntegra da energia do sinal contida no sinal dentro de intervalos de tempo de extensão finita. Estes vetores compartilham a desvantagem de não estarem relacionados a uma quantidade física ou perceptual de maneira bem embasada. Por exemplo, a fase relativa dos alto-falantes um relação ao outro não é devidamente levada em conta. Isso significa, por exemplo, que se um sinal de banda larga for fornecido aos alto-falantes de um conjunto estereofônico à frente de uma posição de escuta com fase oposta, um ouvinte perceberia o som pela direção ambiente, e o campo de som na posição de escuta teria oscilações de energia sonora de um lado para o outro (por exemplo, do lado esquerdo para o lado direito). Nessas condições, os vetores Gerzon estariam apontando para a direção frontal, que obviamente não está representando a situação física ou perceptual.

Naturalmente, com múltiplos formatos multicanal ou representações no mercado, existe uma exigência da capacidade de converter entre as diferentes representações, de maneira que as representações individuais possam ser reproduzidas com conjuntos originalmente desenvolvidos para a reconstrução de uma representação multicanal alternativa. Isto é, por exemplo, pode ser necessária uma transformação entre os canais 5.1 e canais 7.1 ou 7.2 para usar-se uma configuração de reprodução de canal 7.1 ou 7.2 existente para reproduzir a representação multicanal 5.1

comumente usada em DVD. A grande variedade de formatos de áudio torna a produção de conteúdo de áudio difícil, pois todos os formatos requerem mixes e formatos de armazenamento / transmissão específicos. Assim, é necessária a conversão entre diferentes formatos de gravação para reprodução em diferentes configurações de reprodução.

Existem alguns métodos propostos para converter áudio de um formato de áudio específico em outro formato de áudio. No entanto, estes métodos são sempre customizados para formatos multicanal ou representações específicos. Isto é, são somente aplicáveis à conversão de uma representação multicanal pré-determinada específica em outra representação multicanal específica.

Em geral, uma redução no número de canais de reprodução (chamada "downmix") é mais simples de implementar que um aumento no número de canais de reprodução ("upmix"). Para algumas configurações padrão de reprodução de alto-falante, são feitas exigências, por exemplo, a ITU de como fazer "downmix" em configurações de reprodução com um número menor de canais de reprodução. Nestas assim chamadas equações de "downmix" "ITU", os sinais de saída são derivados como simples combinações lineares estáticas de sinais de entrada. Normalmente, uma redução do número de canais de reprodução leva a uma degradação da imagem espacial percebida, isto é, uma qualidade de reprodução degradada de um sinal de áudio espacial.

Para um possível benefício de um alto número de canais de reprodução ou alto-falantes de reprodução, foram desenvolvidas técnicas de "upmixing" para tipos específicos de

conversões. Um problema frequentemente investigado é como converter áudio estereofônico de 2 canais para reprodução com sistemas de alto-falante surround de 5 canais. Uma abordagem ou implementação para esse tipo de "upmix" 2-a-5 é usar um assim
5 chamado decodificador de "matriz". Esses decodificadores tornaram-se comuns para prover ou fazer "upmix" de som multicanal 5.1 em infraestruturas de transmissão em estéreo, especialmente no início do som surround para cinema e home theaters. A idéia básica é reproduzir componentes de som que estejam em fase no sinal estéreo
10 na frente da imagem do som, e pôr os componentes fora de fase nos alto-falantes traseiros. Um método de "upmixing" 2-a-5 alternativo propõe extrair os componentes ambientes do sinal estéreo e reproduzir esses componentes pelos alto-falantes traseiros da configuração 5.1. Uma abordagem que segue as mesmas idéias básicas
15 de maneira perceptivamente mais justificada e usando uma implementação matematicamente mais elegante foi recentemente proposta por C. Faller em "Parametric Multi-channel Audio Coding: Synthesis of Coherence Cues", IEEE Trans. On Speech and Audio Proc., vol. 14, no. 1, January 2006.

20 O padrão recentemente publicado de surround MPEG executa um "upmix" a partir de um ou dois canais com "downmix" e transmitidos, para os canais finais usados em reprodução ou playback, que normalmente é 5.1. Isto é implementado usando-se informações de lado espacial (informação de lado semelhante à
25 técnica BBC) ou sem informações de lado, usando as relações de fase entre os dois canais de um "downmix" estéreo ("modo não guiado" ou "modo de matriz ampliada").

Todos os métodos de conversão de formato

descritos nos parágrafos anteriores são especializados para serem aplicados a configurações específicas do formato de reprodução tanto de origem como de destino, não sendo, portanto, universais. Isto é, uma conversão entre representações multicanal de entrada 5 arbitrárias em representações multicanal de saída arbitrárias não pode ser executada. Isto significa que as técnicas de transformação do método anterior são especificamente projetadas para o número de alto-falantes e sua posição exata para a representação de áudio multicanal de entrada, bem como para a 10 representação multicanal de saída.

O pedido de patente internacional 2004/077884 propõe utilizar a codificação DirAC para gravar respostas de impulso de sinais de áudio dentro de ambientes de escuta. Usando essas respostas de impulso gravadas, sinais de áudio podem ser 15 reproduzidos com a impressão espacial do ambiente de escuta.

O trabalho da convenção AES 6658 é direcionado à codificação de áudio DirAC e propõe um método para criar uma representação codificada eficiente de sinais gravados por microfones de formato b.

O pedido de patente internacional 01/82651 refere-se a técnicas de masterização de surround e reprodução multicanal. Uma técnica de codificação espacial particular é proposta, a fim de possibilitar a transmissão de uma representação codificada compacta. A representação codificada pode então ser 20 decodificada por um decodificador especialmente projetado na extremidade de recepção.

Naturalmente é desejável ter um conceito para transformação multicanal que seja aplicável a combinações

arbitrárias de representações multicanal de entrada e de saída.

Sumário da Invenção

De acordo com uma configuração da presente invenção, equipamento para conversão de uma representação
5 multicanal de entrada em uma representação multicanal de saída diferente de um sinal de áudio espacial, composto de: analisador para derivar uma representação intermediária do sinal de áudio espacial, sendo que a representação intermediária contém parâmetros de direção que indicam uma direção de origem de uma
10 porção do sinal de áudio espacial; e um compositor de sinal para gerar a representação multicanal de saída do sinal de áudio espacial usando a representação intermediária do sinal de áudio espacial.

Como é usada uma representação intermediária, a
15 qual possui parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial, a conversão pode ser obtida entre representações multicanal arbitrárias, contanto que a configuração de alto-falante da representação multicanal de saída seja conhecida. É importante observar que a configuração de
20 alto-falante da representação multicanal de saída não precisa ser conhecida com antecedência, isto é, durante o projeto do equipamento de conversão. Como o equipamento de conversão e o método são universais, uma representação multicanal provida como
25 representação multicanal de entrada e projetada para uma configuração de alto-falante específica pode ser alterada no lado de recepção, para adequar-se à configuração de reprodução disponível, de maneira que a qualidade de uma reprodução de um sinal de áudio espacial seja melhorada.

De acordo com uma outra configuração da presente invenção, a direção de origem de uma porção do sinal de áudio espacial é analisada dentro de diferentes bandas de frequência. Assim, diferentes parâmetros de direção são derivados para finito
5 com porções de frequência do sinal de áudio espacial. Para derivar as porções de frequência de largura finita, pode ser usado, por exemplo, um banco de filtros ou um transformador de Fourier. De acordo com outra configuração, as porções de frequência ou bandas de frequência, para as quais a análise é realizada
10 individualmente, são escolhidas de maneira a corresponderem à resolução de frequência do processo auditivo humano. Estas configurações podem ter a vantagem de que a direção de origem das porções do sinal de áudio espacial é executada tão bem quanto o próprio sistema auditivo humano é capaz de determinar a direção de
15 origem dos sinais de áudio. Portanto, a análise é realizada sem uma perda potencial de precisão na determinação da origem de um objeto de áudio ou uma porção de sinal, quando esse sinal analisado é reconstruído e reproduzido através de uma configuração de alto-falante arbitrária.

20 De acordo com uma outra configuração da presente invenção, um ou mais canais de "downmix" são também derivados, pertencendo à representação intermediária. Isto é, os canais com "downmix" são derivados de canais de áudio correspondentes aos alto-falantes associados à representação multicanal de entrada,
25 que pode então ser usada para gerar a representação multicanal de saída, ou para gerar canais de áudio correspondentes aos alto-falantes associados à representação multicanal de saída.

Por exemplo, um "downmix" monofônico de um canal

pode ser gerado pelos canais de entrada 5.1 de um sinal de áudio de canal 5.1 comum. Isso poderia, por exemplo, ser realizado computando-se a soma de todos os canais de áudio individuais. Com base nesse canal de "downmix" monofônico derivado, um compositor de sinal pode distribuir essas porções do canal de "downmix" monofônico correspondentes às porções analisadas da representação multicanal de entrada nos canais da representação multicanal de saída, conforme indicado pelos parâmetros de direção. Isto é, uma frequência / tempo ou porção de sinal analisada como proveniente da extremidade esquerda de um sinal de áudio espacial será redistribuída para os alto-falantes da representação multicanal de saída, que estão localizados no lado esquerdo em relação à posição de escuta.

De maneira geral, algumas configurações da presente invenção permitem distribuir porções do sinal de áudio espacial com maior intensidade em um canal correspondente a um alto-falante mais próximo da direção indicada pelos parâmetros de direção que a um canal mais longe dessa direção. Isto é, independentemente de como a localização dos alto-falantes usados para reprodução estiver definida na representação multicanal de saída, será obtida uma redistribuição espacial adequando-se a configuração de reprodução disponível da melhor maneira possível.

De acordo com algumas configurações da presente invenção, uma resolução espacial, com a qual pode ser determinada uma direção de origem de uma porção do sinal de áudio espacial, é muito mais alta que o ângulo do espaço tridimensional associado a um único alto-falante da representação multicanal de entrada. Isto é, a direção de origem de uma porção do sinal de áudio espacial

pode ser derivada com uma melhor precisão que uma resolução espacial que pode ser obtida simplesmente redistribuindo-se os canais de áudio de uma configuração distinta para outra configuração específica, como por exemplo, redistribuindo-se os
5 canais de uma configuração 5.1 em uma configuração 7.1 ou 7.2.

Em resumo, algumas configurações da invenção permitem a aplicação de um método melhorado para conversão de formato, o qual é universalmente aplicável e não depende de um layout / configuração-alvo de alto-falante particular desejado(a).
10 Algumas configurações convertem um formato (representação) de áudio multicanal de entrada com canais N1 em um formato (representação) multicanal de saída com canais N2 extraíndo parâmetros de direção (semelhantes ao DirAC), os quais são então utilizados para sintetizar o sinal de saída com canais N2. Além
15 disso, de acordo com algumas configurações, alguns canais de "downmix" N0 são computados a partir dos sinais de entrada N1 (canais de áudio correspondentes a alto-falantes de acordo com a representação multicanal de entrada), os quais são então utilizados como base para um processo de decodificação usando-se
20 os parâmetros de direção extraídos.

Breve descrição dos desenhos

Várias configurações da presente invenção serão descritas a seguir, com referência aos desenhos anexos.

A Fig. 1 mostra uma ilustração da derivação de
25 parâmetros de direção que indicam uma direção de origem de uma porção de um sinal de áudio; e

A Fig. 2 mostra uma outra configuração de derivação de parâmetros de direção com base em uma representação

de canal 5.1;

A Fig. 3 mostra um exemplo de geração de uma representação multicanal de saída;

5 A Fig. 4 mostra um exemplo de conversão de áudio a partir de uma configuração de canal 5.1 em uma configuração de canal 8.1; e

A Fig. 5 mostra um exemplo de um equipamento inventivo para conversão entre formatos de áudio multicanal.

10 Algumas configurações da presente invenção derivam uma representação intermediária de um sinal de áudio espacial com parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial. Uma possibilidade é derivar um vetor de velocidade que indique a direção de origem de uma porção de um sinal de áudio espacial. Um exemplo para fazer
15 isso será descrito nos parágrafos a seguir, com referência à Fig. 1.

Antes de detalhar o conceito, pode-se observar que a análise a seguir pode ser aplicada a múltiplas porções de frequência ou tempo individuais do sinal de áudio espacial
20 subjacente simultaneamente. Para manter a simplicidade, no entanto, a análise será descrita para somente uma frequência ou tempo específicos ou porção de tempo / frequência. A análise baseia-se em uma análise energética do campo de som gravado em uma posição de gravação 2, localizada no centro de um sistema de
25 coordenadas, como indica a Fig. 1.

O sistema de coordenadas é um Sistema de Coordenadas Cartesianas, com um eixo x 4 e um eixo y 6, perpendiculares um ao outro. Usando-se um sistema de mão direita,

o eixo z, não mostrado na Fig. 1, aponta para a direção fora do plano do desenho.

Para a análise de direção, presume-se que os sinais 4 (conhecidos como sinais de formato B) são gravados. Um sinal onidirecional w é gravado, isto é, um sinal que recebe sinais de todas as direções com (idealmente) igual sensibilidade. Além disso, três sinais direcionais X, Y e Z são gravados, com uma distribuição de sensibilidade apontando na direção dos eixos do Sistema de Coordenadas Cartesianas. Exemplos de possíveis padrões de sensibilidade dos microfones usados são dados na Fig. 1, mostrando dois padrões "figura de oito" 8a e 8b, apontando nas direções dos eixos. Duas possíveis fontes de áudio 10 e 12 são ainda ilustradas na projeção bidimensional do sistema de coordenadas mostrado na Fig. 1.

Para a análise de direção, um vetor de velocidade instantâneo (no índice de tempo n) é composto para diferentes porções de frequência (descritas pelo índice i) por

$$v(n,i) = X(n,i) e_x + Y(n,i) e_y + Z(n,i) e_z. \quad (1)$$

Isto é, um vetor é criado com os sinais de microfone gravados individualmente dos microfones associados ao eixo do sistema de coordenadas como componentes. Na equação anterior e nas próximas, as Quantidades são indexadas em Tempo (n) e também em frequência (i) por dois índices (n,i). Isto é,

e_x , e_y e e_z representam vetores de unidade cartesiana.

Usando-se o sinal onidirecional gravado simultaneamente W, uma intensidade I instantânea é computada como

$$I(n,i) = w(n,i) v(n,i), \quad (2)$$

a energia instantânea é derivada de acordo com a seguinte fórmula:

$$E(n, i) = w^2(n, i) + \|\mathbf{v}\|^2(n, i), \quad (3)$$

onde $\|\cdot\|$ denota norma de vetor.

5 Isto é, uma quantidade de intensidade é derivada, permitindo uma possível interferência entre dois sinais (pois podem ocorrer amplitudes positivas e negativas). Além disso, é derivada uma quantidade de energia, a qual naturalmente não permite interferência entre dois sinais, pois a quantidade de
10 energia não contém valores negativos que permitam um cancelamento do sinal.

Estas propriedades da intensidade e os sinais de energia podem ser usados vantajosamente para derivar uma direção de origem de porções de sinal com alta precisão, preservando uma
15 correlação virtual de canais de áudio (uma fase relativa entre os canais), como será detalhado abaixo.

Por outro lado, o vetor de intensidade instantânea pode ser usado como vetor que indica a direção de origem de uma porção do sinal de áudio espacial. No entanto, este
20 vetor pode passar por alterações rápidas, causando, assim, artefatos dentro da reprodução do sinal. Portanto, alternativamente, pode ser computada uma direção instantânea usando-se média de curto prazo, utilizando-se uma janela de Hanning W_2 de acordo com a seguinte fórmula:

$$25 \quad D(n, i) = - \sum_{m=-M/2}^{M/2} I(n+m, i) W_2(m), \quad (4)$$

onde W_2 é a janela de Hanning para fazer a média de curto prazo D .

Isto é, opcionalmente, pode ser derivado um vetor de direção com média de curto prazo com parâmetros que indicam uma direção de origem do sinal de áudio espacial.

Opcionalmente, uma medida de difusividade ψ pode ser computada como segue:

$$\psi(n,i) = 1 - \frac{\sqrt{\sum_{m=-M/2}^{M/2} \|I(n+m,i)\|^2 W_1(m)}}{\sum_{m=-M/2}^{M/2} E(n+m,i) W_1(m)} \quad (5)$$

onde $W_1(m)$ é uma função de janela definida entre $-M/2$ e $M/2$ para média de curto prazo.

Deve-se novamente observar que a derivação é realizada de maneira a preservar a correlação virtual dos canais de áudio. Isto é, as informações de fase são devidamente consideradas, o que não é o caso para estimativas de direção baseadas somente em estimativas de energia (como por exemplo, vetores de Gerzon).

O exemplo simples a seguir deve servir para explicar isto em mais detalhes. Considere um sinal perfeitamente difuso que seja reproduzido por dois alto-falantes de um sistema estéreo. Como o sinal é difuso (origina-se de todas as direções), ele deve ser reproduzido por ambos os alto-falantes com igual intensidade. No entanto, como a percepção será difusa, é necessário um deslocamento de fase de 180 graus. Nesse panorama, uma estimativa de direção baseada puramente em energia produziria um vetor de direção que apontaria exatamente para o meio, entre os dois alto-falantes, o que certamente é um resultado indesejável que não reflete a realidade.

De acordo com o conceito inventivo detalhado acima, a correlação virtual dos canais de áudio é preservada, ao

mesmo tempo em que estima os parâmetros de direção (vetores de direção). Neste exemplo particular, o vetor de direção seria zero, indicando que o som não se origina de uma direção distinta, o que claramente não é o caso na realidade. De maneira correspondente, o parâmetro de difusividade da equação (5) é 1, correspondendo
5 perfeitamente à situação real.

As janelas de Hanning nas equações acima podem ainda ter diferentes extensões para diferentes bandas de frequência.

10 Como resultado desta análise, para cada fatia de tempo de uma porção de frequência, um vetor de direção ou parâmetros de direção são derivados, indicando uma direção de origem da porção do sinal de áudio espacial, para o qual a análise foi realizada. Opcionalmente, um parâmetro de difusibilidade pode
15 ser derivado, indicando a difusibilidade da direção de uma porção do sinal de áudio espacial. Como foi descrito anteriormente, um valor de difusão de um derivado de acordo com a equação (4) descreve um sinal de difusibilidade máxima, isto é, originário de todas as direções com igual intensidade.

20 Contrariamente, valores de difusibilidade pequenos são atribuídos a porções de sinal originárias predominantemente de uma direção.

A Fig. 2 mostra um exemplo para a derivação de parâmetros de direção de uma representação multicanal de entrada
25 com cinco canais, de acordo com ITU-775-1. O sinal de áudio de entrada multicanal, isto é, a representação multicanal de entrada, é primeiramente transformado em formato B, simulando-se uma gravação anecóica da configuração de áudio multicanal

correspondente. Em relação a um centro 20 do Sistema de Coordenadas Cartesianas com um eixo x 22 e y 24, um alto-falante traseiro direito 26 está localizado em um ângulo de 110°. Um alto-falante frontal direito 28 está localizado a +30°, um alto-falante central a 0°, um alto-falante frontal esquerdo 32 a 31°, e um alto-falante traseiro esquerdo 34 a -110°. Na prática, uma gravação anecóica pode ser simulada aplicando-se operações simples de matriz, a configuração geométrica da representação multicanal de entrada é conhecida.

Um sinal onidirecional w pode ser obtido fazendo-se uma soma direta de todos os sinais de alto-falante, isto é, de todos os canais de áudio correspondentes aos alto-falantes associados à representação multicanal de entrada. O dipolo ou sinais de "figura de oito" X, Y e Z podem ser formados adicionando-se os sinais de alto-falante ponderados pelo co-seno do ângulo entre o alto-falante e os eixos cartesianos correspondentes, isto é, a direção de máxima sensibilidade do microfone dipolo a ser simulada. Suponhamos que L_n seja o vetor cartesiano 2-D ou 3-D que aponta na direção do n ésimo alto-falante e V seja o vetor de unidade que aponta para a direção do eixo cartesiano correspondente ao microfone do dipolo. Assim, o fator ponderante é $\cos(\text{ângulo}(L_n, V))$. O sinal direcional X seria, por exemplo, escrito como

$$X = \sum_{n=1}^N C_n \cdot \cos(\text{ângulo}(L_n, V)),$$

quando C_n denota o sinal do alto-falante do n ésimo canal e N é o número de canais. O termo ângulo deve ser interpretado como um operador, computando o ângulo espacial entre

os dois vetores dados. Isto é, por exemplo, o ângulo 40° (Θ) entre o eixo Y 24 e o alto-falante frontal esquerdo 32 no caso bidimensional ilustrado na Fig. 2.

A derivação adicional de parâmetros de direção 5 poderia, por exemplo, ser feita de acordo com a ilustração da Fig. 1, e detalhada na descrição correspondente, isto é, os sinais de áudio X, Y e Z podem ser divididos em bandas de frequência de acordo com a resolução de frequência do sistema auditivo humano. A direção do som, isto é, a direção de origem das porções do sinal 10 de áudio espacial e, opcionalmente, a difusibilidade, são analisadas, dependendo do tempo em cada canal de frequência. Opcionalmente, uma substituição para difusibilidade de som usando-se outra medida de dessemelhança de sinal que não seja a difusibilidade também pode ser utilizada, como por exemplo a 15 coerência entre canais (estéreo) associados ao sinal de áudio espacial.

Se, em um exemplo simplificado, uma fonte de áudio 44 estiver presente, como indica a Fig. 2, onde essa fonte somente contribua para o sinal dentro de uma banda de frequência 20 específica, um vetor de direção 46 que aponta para a fonte de áudio 44 seria derivado. O vetor de direção é representado por parâmetros de direção (componentes de vetor) que indicam a direção da porção do sinal de áudio espacial originária da fonte de áudio 44. Na configuração de reprodução da Fig. 2, esse sinal seria 25 reproduzido principalmente pelo alto-falante frontal esquerdo 32, conforme ilustrado pela onda simbólica associada a este alto-falante. No entanto, pequenas porções de sinal também serão reproduzidas pelo alto-falante traseiro esquerdo 32. Assim, o

sinal direcional do microfone associado à coordenada X 22 receberia os componentes de sinal do canal frontal esquerdo 32 (o canal de áudio associado ao alto-falante frontal esquerdo 32) e do canal traseiro esquerdo 34.

5 Como, de acordo com a implementação acima, o sinal direcional Y associado ao eixo y receberá também porções de sinal reproduzidas pelo alto-falante frontal esquerdo 32, uma análise direcional baseada em sinais direcionais X e Y poderá reconstruir o som que vem do vetor de direção 46 com alta
10 precisão.

 Para a conversão final na representação multicanal desejada (formato multicanal), os parâmetros de direção que indicam a direção de origem de porções dos sinais de áudio são usados. Opcionalmente, um ou mais (N0) canais de "downmix" de
15 áudio adicionais podem ser usados. Esse canal de "downmix" pode, por exemplo, ser o canal onidirecional W ou qualquer outro canal monofônico. No entanto, para a distribuição espacial, o uso de somente um canal único associado à representação intermediária é de impacto negativo pequeno. Isto é, vários canais de "downmix",
20 como por exemplo, um mix estéreo, os canais W, X e Y ou todos os canais de um formato B podem ser usados, contanto que os parâmetros de direção ou os dados direcionais tenham sido derivados e possam ser usados para a reconstrução ou geração da representação multicanal de saída. É também alternativamente
25 possível usar os 5 canais da Fig. 2 diretamente, ou qualquer combinação de canais associada à representação multicanal de entrada como uma substituição de possíveis canais de "downmix". Quando somente um canal é armazenado, pode haver uma degradação da

qualidade da reprodução do som difuso.

A Fig. 3 mostra um exemplo da reprodução do sinal da fonte de áudio 44 com uma configuração de alto-falante que difere significativamente da configuração de alto-falante da Fig. 2, a qual era a representação multicanal de entrada da qual os parâmetros tinham sido derivados. A Fig. 3 mostra, como exemplo, seis alto-falantes 50a a 50f, igualmente distribuídos ao longo de uma linha na frente de uma posição de escuta 60, definindo o centro de um sistema de coordenadas com um eixo x 22 e um eixo y 24, conforme introduzido na Fig. 2. Como uma análise anterior proveu parâmetros de direção que descrevem a direção do vetor de direção 46 que aponta para a fonte do sinal de áudio 44, uma representação multicanal de saída adaptada à configuração de alto-falante da Fig. 3 pode ser facilmente derivada, redistribuindo-se a porção do sinal de áudio espacial a ser reproduzido para os alto-falantes próximos à direção da fonte de áudio 44, isto é, pelos alto-falantes próximos à direção indicada pelos parâmetros de direção. Isto é, os canais de áudio correspondentes aos alto-falantes na direção indicada pelos parâmetros de direção são enfatizados em relação aos canais de áudio correspondentes aos alto-falantes que estão distantes desta direção. Isto é, os alto-falantes 50a e 50b podem ser orientados (por exemplo, usando-se panorâmica de amplitude) para reproduzir a porção do sinal, enquanto os alto-falantes 50c e 50f não reproduzem essa porção específica do sinal, mas podem ser usados para reprodução de som difuso ou outras porções de sinal de diferentes bandas de frequência.

O uso de um compositor de sinal para gerar a

representação multicanal de saída do sinal de áudio espacial usando-se os parâmetros de direção pode também ser interpretado como sendo uma decodificação do sinal intermediário no formato de saída de multicanal desejado, com canais de saída N2. Os canais de "downmix" de áudio ou sinais gerados são tipicamente processados na mesma banda de frequência na qual eles foram analisados. A decodificação pode ser realizada de maneira semelhante ao DirAC. Na reprodução opcional de som difuso, o uso de áudio para representar uma corrente não difusa é tipicamente um dos dois sinais de canal de "downmix" N0 opcionais ou combinações lineares deles.

Para a criação opcional de uma corrente difusa, existem várias opções de síntese para criar a parte difusa dos sinais de saída ou dos canais de saída correspondentes aos alto-falantes de acordo com a representação multicanal de saída. Se houver somente um canal de "downmix" transmitido, esse canal tem que ser usado para criar sinais não-difusos para cada alto-falante. Se houver mais canais transmitidos, existem mais opções para a maneira pela qual o som difuso pode ser criado. Se, por exemplo, um "downmix" estéreo for usado no processo de conversão, um método obviamente adequado é aplicar o canal de "downmix" esquerdo aos alto-falantes à esquerda, e o canal de "downmix" direito aos alto-falantes do lado direito. Se vários canais de "downmix" forem usados para a conversão (isto é, $N0 > 1$), a corrente difusa de cada alto-falante pode ser computada como uma soma diferentemente ponderada destes canais de "downmix". Uma possibilidade seria, por exemplo, transmitir um sinal de formato B (canais X, Y, Z e W conforme descrito anteriormente) e computar o

sinal de um microfone cardióide virtual para cada alto-falante.

O texto a seguir descreve um possível procedimento para a conversão de uma representação multicanal de entrada em uma representação multicanal de saída como uma lista.

5 Neste exemplo, o som é gravado com um microfone de formato B simulado e depois continua a ser processado por um compositor de sinal para escuta ou reprodução com uma configuração de alto-falante multicanal ou monofônica. As etapas únicas são explicadas com referência à Fig. 4, mostrando a conversão de uma

10 representação multicanal de entrada de canal 5.1 em uma representação multicanal de saída de canal 8. A base é um formato de áudio de canal N1 (N1 sendo 5 no exemplo específico). Para converter a representação multicanal de entrada em uma representação multicanal de saída diferente, as etapas a seguir

15 devem ser executadas.

1. Simule uma gravação anecóica de uma representação de áudio multicanal arbitrária com canais de áudio N1 (5 canais), conforme ilustrado na seção de gravação 70 (com um microfone de formato B simulado em um centro 72 do layout).

20 2. Em uma etapa de análise 74, os sinais de microfone simulados são divididos em bandas de frequência, e em uma etapa 76 de análise direcional, é derivada a direção de origem de porções dos sinais de microfone simulados. Além disso, opcionalmente, a difusibilidade (ou coerência) pode ser

25 determinada em uma etapa de término de difusibilidade 78.

Como foi anteriormente mencionado, uma análise de direção pode ser realizada sem a utilização de uma etapa intermediária de formato B. Isto é, em geral, uma representação

intermediária do sinal de áudio espacial tem que ser derivada com base em uma representação multicanal de entrada, onde a representação intermediária possui parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial.

3. Em uma etapa de "downmix" 80, são derivados sinais de áudio de "downmix" N0, para serem usados como base para a conversão / criação da representação multicanal de saída. Em uma etapa de composição 82, os sinais de áudio de "downmix" são decodificados ou é feito "upmix" para uma configuração de alto-falante arbitrária que requer canais de áudio N2 por um método de síntese apropriado (por exemplo, usando panorâmica de amplitude ou técnicas igualmente adequadas).

O resultado pode ser reproduzido por um sistema de alto-falante multicanal, tendo por exemplo 8 alto-falantes, conforme indicado no exemplo de reprodução 84 da Fig. 4. No entanto, graças à universalidade do conceito, uma conversão pode também ser feita para uma configuração de alto-falante monofônica, provendo um efeito como se o sinal de áudio espacial tivesse sido gravado com um único microfone direcional.

A Fig. 5 mostra um esquema de princípio de um exemplo de um equipamento para conversão entre formatos de áudio multicanal 100.

O Equipamento 100 recebe uma representação multicanal de entrada 102.

O Equipamento 100 é composto de um analisador 104 para derivar uma representação intermediária 106 do sinal de áudio espacial, sendo que a representação intermediária 106 possui

parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial.

O Equipamento 100 é composto ainda de um compositor de sinal 108 para gerar uma representação multicanal de saída 110 do sinal de áudio espacial usando-se a representação intermediária (106) do sinal de áudio espacial.

Em resumo, as configurações dos equipamentos de conversão e métodos de conversão descritas anteriormente provêm algumas grandes vantagens. Primeiramente, virtualmente qualquer formato de áudio de entrada pode ser processado desta maneira. Além disso, o processo de conversão pode gerar saída para qualquer layout de alto-falante, inclusive layout / configurações de alto-falante não-padrão, sem a necessidade de customizar especificamente novas relações para novas combinações de layout / configurações de alto-falante de entrada e layout / configurações de alto-falante de saída. E ainda, a resolução espacial de reprodução de áudio aumenta quando o número de alto-falantes é aumentado, ao contrário das implementações do método anterior.

Dependendo de determinados requisitos de implementação dos métodos inventivos, os métodos inventivos podem ser implementados em hardware ou em software. A implementação pode ser feita utilizando-se um meio de armazenamento digital, em particular um disco, DVD ou CD com sinais de controle legíveis eletronicamente armazenados neles, os quais funcionam em conjunto com um sistema de computador programável para que os métodos inventivos sejam executados. Em geral, a presente invenção é, portanto, um produto de programa de computador com um código de programa armazenado em um portador legível em máquina, sendo que o

código de programa funciona para executar os métodos inventivos quando o programa de computador roda em um computador. Em outras palavras, os métodos inventivos são, portanto, um programa de computador com um código de programa para executar pelo menos um dos métodos inventivos quando o programa de computador roda em um computador.

Apesar da revelação acima ter sido particularmente demonstrada e descrita com referência a configurações particulares, será compreendido pelos técnicos no assunto que várias outras alterações na forma e detalhes podem ser feitas sem sair do espírito e escopo da invenção. Deve-se entender que várias alterações podem ser feitas na adaptação a diferentes configurações sem sair dos conceitos mais amplos revelados neste documento e abrangidos pelas reivindicações a seguir.

REIVINDICAÇÕES

1. Equipamento para conversão de uma representação multicanal de entrada em uma representação multicanal de saída diferente de um sinal de áudio espacial, 5 caracterizado pelo fato de que compreende: decodificador de representação de entrada para derivar alguns canais de áudio correspondentes aos alto-falantes associados à representação multicanal de entrada; analisador para derivação, usando o número de canais de áudio correspondentes aos alto-falantes associados à 10 representação multicanal de entrada, uma representação intermediária do sinal de áudio espacial, sendo que a representação intermediária possui parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial; e compositor de sinal para gerar a representação 15 multicanal de saída do sinal de áudio espacial usando a representação intermediária do sinal de áudio espacial.

2. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona derivando parâmetros de direção, dependendo de uma correlação virtual dos 20 canais de áudio associados à representação multicanal de entrada.

3. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona derivando parâmetros de direção que preservam as informações de fase relativa dos canais de áudio associados à representação multicanal 25 de entrada

4. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona derivando diferentes parâmetros de direção para porções de frequência de

largura finita do sinal de áudio espacial.

5. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona derivando diferentes parâmetros de direção para porções de tempo de extensão
5 finita do sinal de áudio espacial.

6. Equipamento, de acordo com a reivindicação 4, caracterizado pelo fato de que o analisador funciona derivando os diferentes parâmetros de direção para porções de tempo de extensão finita do sinal de áudio espacial associado às porções de
10 frequência, onde a extensão de uma primeira porção de tempo associada a uma primeira porção de frequência difere da extensão de uma associação de uma segunda porção de tempo a uma segunda porção de frequência diferente do sinal de áudio espacial.

7. Equipamento, de acordo com a reivindicação 1,
15 caracterizado pelo fato de que o analisador funciona derivando parâmetros de direção que descrevem um vetor que aponta para a direção de origem da porção do sinal de áudio espacial.

8. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona também
20 derivando um ou mais canais de áudio associados à representação intermediária.

9. Equipamento, de acordo com a reivindicação 8, caracterizado pelo fato de que o analisador funciona derivando canais de áudio correspondentes aos alto-falantes associados à
25 representação multicanal de entrada.

10. Equipamento, de acordo com a reivindicação 8, caracterizado pelo fato de que o analisador funciona derivando um canal de "downmix" como a soma dos canais de áudio correspondentes

aos alto-falantes associados à representação multicanal de entrada.

11. Equipamento, de acordo com a reivindicação 8, caracterizado pelo fato de que o analisador funciona derivando pelo menos um canal de áudio associado à direção de um eixo de um Sistema de Coordenadas Cartesianas.

12. Equipamento, de acordo com a reivindicação 11, caracterizado pelo fato de que o analisador funciona derivando pelo menos um canal de áudio que perfaz a soma ponderada dos canais de áudio correspondentes aos alto-falantes associados à representação multicanal de entrada.

13. Equipamento, de acordo com a reivindicação 11, caracterizado pelo fato de que o analisador funciona de maneira que a derivação de pelo menos um canal de áudio x associado à direção V de um eixo do Sistema de Coordenadas Cartesianas possa ser descrita por uma combinação de canais de áudio n C_n correspondente aos alto-falantes n associados à representação multicanal de entrada e direcionado a L_n , de acordo com a seguinte fórmula:

$$X = \sum_{n=1}^N C_n \cdot \cos(\text{ângulo}(L_n, V)).$$

14. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o analisador funciona ainda derivando um parâmetro de difusibilidade que indica uma difusibilidade da direção de origem da porção do sinal de áudio espacial.

15. Equipamento, de acordo com a reivindicação 1, caracterizado pelo fato de que o compositor de sinal funciona

distribuindo a porção do sinal de áudio espacial por um número de canais correspondente a um número de alto-falantes associado à representação multicanal de saída.

16. Equipamento, de acordo com a reivindicação 5 15, caracterizado pelo fato de que o compositor de sinal funciona de maneira que a porção do sinal de áudio espacial é distribuída com maior intensidade para um canal correspondente a um alto-falante mais próximo da direção indicada pelos parâmetros de direção que para um canal correspondente a um alto-falante mais 10 longe dessa direção.

17. Equipamento, de acordo com a reivindicação 14, caracterizado pelo fato de que o compositor de sinal funciona de maneira que a porção do sinal de áudio espacial é distribuída com mais intensidade uniforme para canais correspondentes a alto-falantes associados à representação multicanal de saída quando o 15 parâmetro de difusibilidade indica difusibilidade mais alta que quando o parâmetro de difusibilidade indica difusibilidade mais baixa.

18. Equipamento, de acordo com a reivindicação 1 20 caracterizado pelo fato de que compreende ainda: uma interface de entrada para receber a representação multicanal de entrada.

19. Equipamento, de acordo com a reivindicação 15, caracterizado pelo fato de que o compositor de sinal compreende ainda um codificador de canal de saída para derivar a 25 representação multicanal de saída com base nos canais de áudio correspondentes aos alto-falantes associados à representação de canal de saída.

20. Equipamento, de acordo com a reivindicação 1

caracterizado pelo fato de que compreende ainda uma interface de saída para prover a representação multicanal de saída.

21. Método para conversão de uma representação multicanal de entrada em uma representação multicanal de saída diferente de um sinal de áudio espacial, sendo que o método caracterizado pelo fato de que compreende: derivar um número de canais de áudio correspondente aos alto-falantes associados à representação multicanal de entrada; derivar, usando o número de canais de áudio correspondente aos alto-falantes associados à representação multicanal de entrada, uma representação intermediária do sinal de áudio espacial, sendo que a representação intermediária possui parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial; e gerar a representação multicanal de saída do sinal de áudio espacial usando a representação intermediária do sinal de áudio espacial.

22. Programa de computador para, quando for rodado em um computador, implementar o método para conversão de uma representação multicanal em uma representação multicanal de saída diferente de um sinal de áudio espacial, caracterizado pelo fato de que o método compreende: derivar um número de canais de áudio correspondente aos alto-falantes associados à representação multicanal de entrada; derivar, usando o número de canais de áudio correspondente aos alto-falantes associados à representação multicanal de entrada, uma representação intermediária do sinal de áudio espacial, sendo que a representação intermediária possui parâmetros de direção que indicam uma direção de origem de uma porção do sinal de áudio espacial; e gerar a representação

multicanal de saída do sinal de áudio espacial usando a representação intermediária do sinal de áudio espacial.

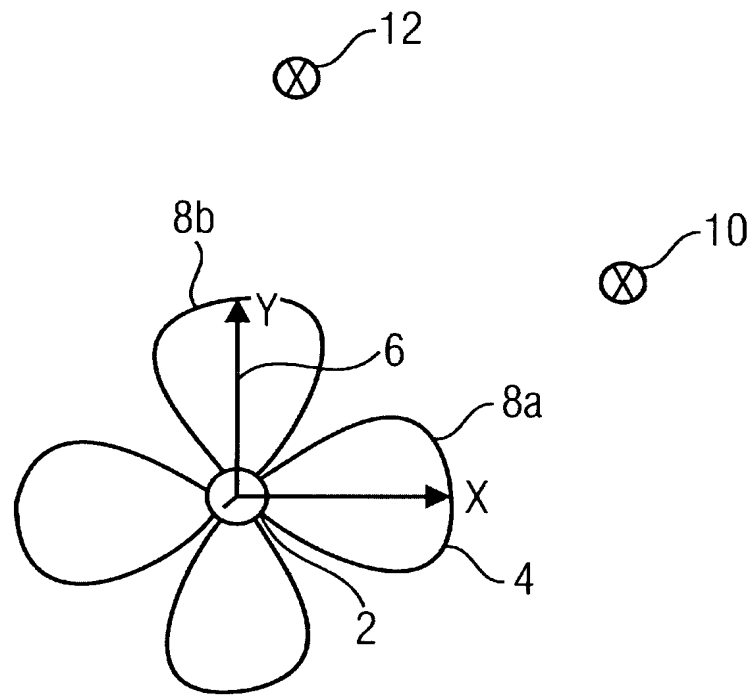


FIG 1

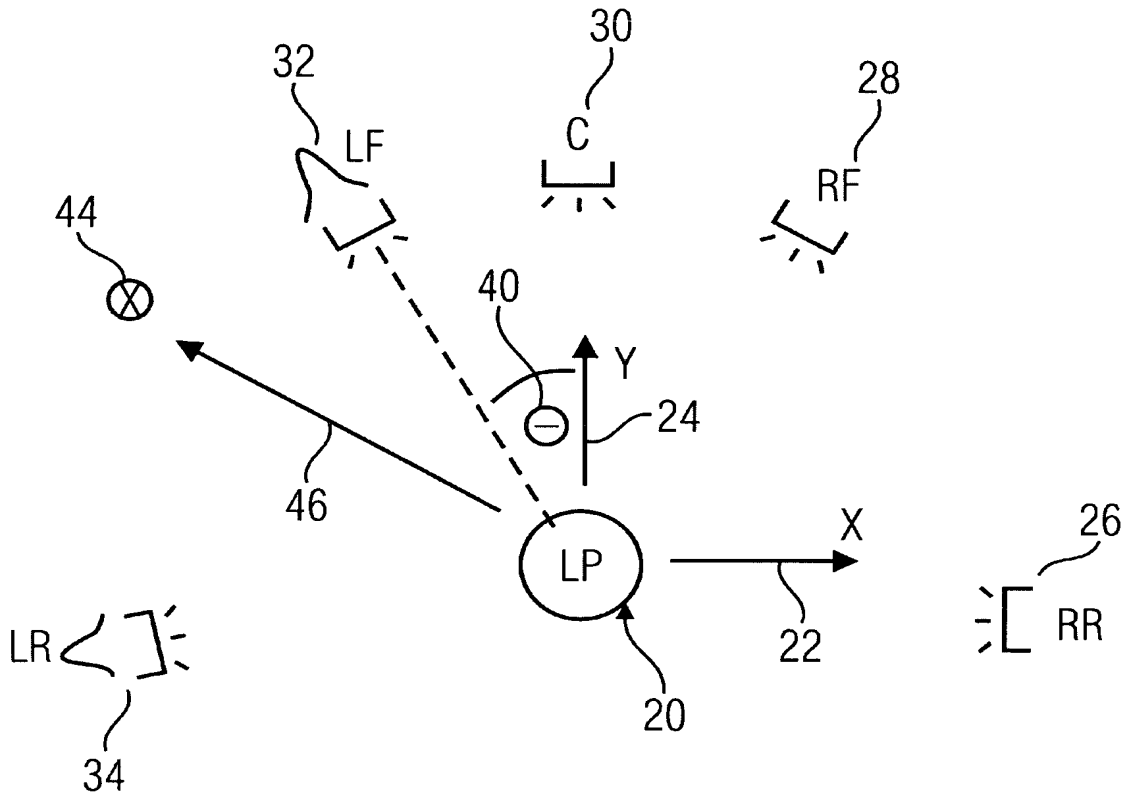


FIG 2

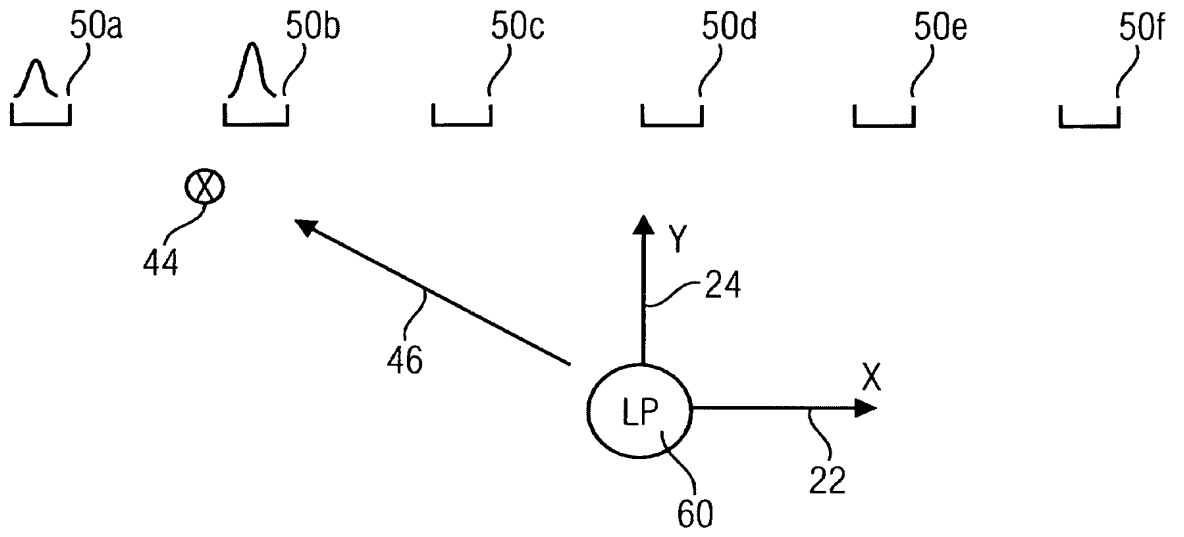


FIG 3

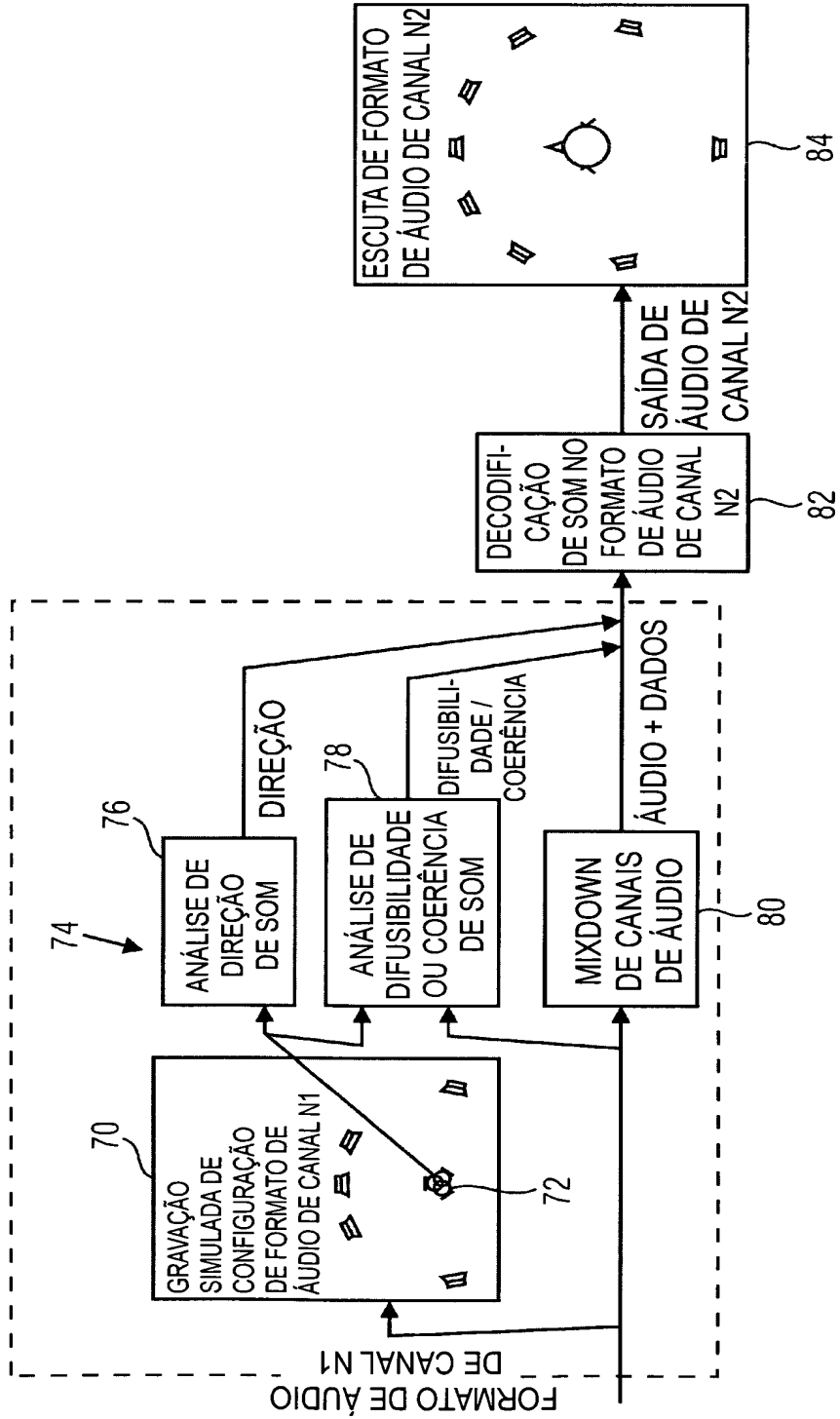


FIG 4

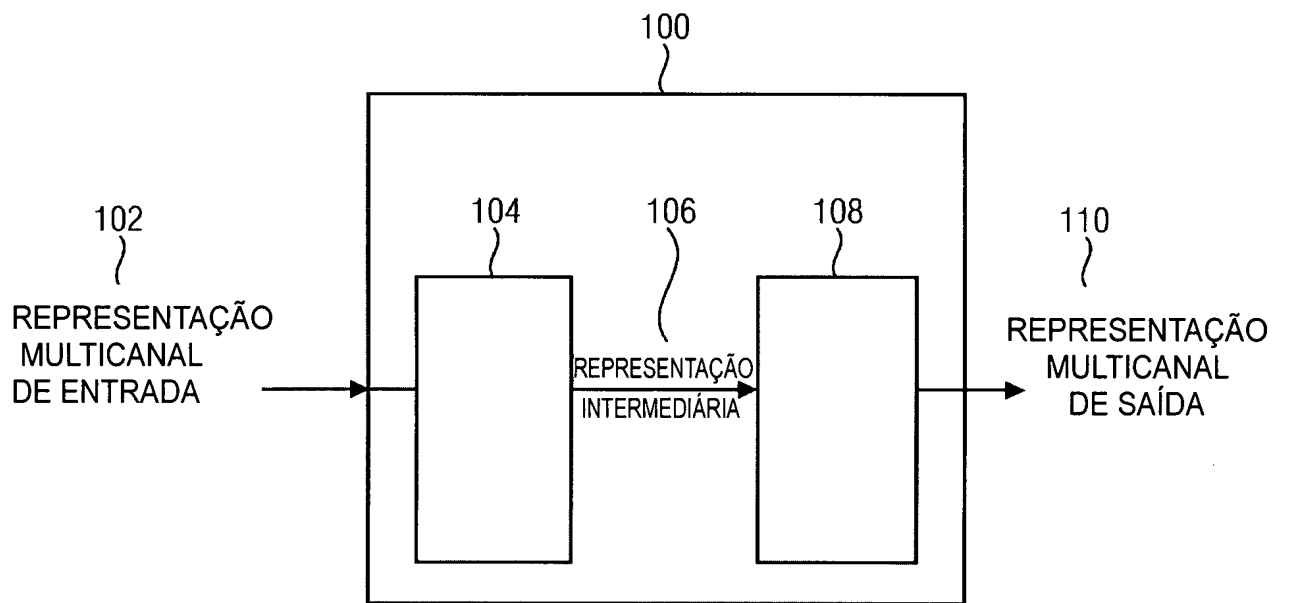


FIG 5

RESUMO

"MÉTODO E EQUIPAMENTO PARA CONVERSÃO ENTRE FORMATOS DE ÁUDIO MULTICANAL"

Uma representação multicanal de entrada é
5 convertida em uma representação multicanal de saída diferente de
um sinal de áudio espacial, na qual uma representação
intermediária do sinal de áudio espacial é derivada, sendo que a
representação intermediária possui parâmetros de direção que
indicam uma direção de origem de uma porção do sinal de áudio
10 espacial; e na qual a representação multicanal de saída do sinal
de áudio espacial é gerada usando-se a representação intermediária
do sinal de áudio espacial.