



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

(11) PI 0816618-8 B1

(22) Data do Depósito: 30/09/2008

(45) Data de Concessão: 10/11/2020



(54) Título: MÉTODO E APARELHO PARA GERAR SINAL DE ÁUDIO BINAURAL

(51) Int.Cl.: G10L 19/008; H04S 3/02.

(52) CPC: G10L 19/008; H04S 3/02; H04S 2400/01; H04S 2420/01; H04S 2420/03.

(30) Prioridade Unionista: 09/10/2007 EP 07118107.7.

(73) Titular(es): KONINKLIJKE PHILIPS ELECTRONICS N.V.; DOLBY INTERNATIONAL AB..

(72) Inventor(es): VILLEMOES, LARS FALCK; BREEBAART, DIRK JEROEN.

(86) Pedido PCT: PCT EP2008008300 de 30/09/2008

(87) Publicação PCT: WO 2009/046909 de 16/04/2009

(85) Data do Início da Fase Nacional: 07/04/2010

(57) Resumo: MÉTODO E APARELHO PARA GERAR SINAL DE BINAURAL. Aparelho para geração de um sinal de áudio binaural compreende um desmultiplexador (402) e decodificador (403) que recebe dados de áudio compreendendo um sinal de áudio do canal M de áudio que é um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N. Um processador de conversão (411) converte parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural. Um processador de matriz (409) converte o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais. Um filtro estéreo (415, 417) gera o sinal de áudio binaural pela filtragem do primeiro sinal estéreo. Os coeficientes de filtro para o filtro estéreo são determinados em resposta a pelo menos uma função de transferência perceptual binaural por um processador de coeficiente (419). A combinação de conversão/processamento de parâmetro e filtragem permite que um sinal binaural de alta qualidade seja gerado com baixa complexidade.

"MÉTODO E APARELHO PARA GERAR SINAL DE ÁUDIO BINAURAL"

CAMPO DA INVENÇÃO

A invenção se refere a um método e aparelho para
5 gerar um sinal de áudio binaural e, especificamente, mas não exclusivamente, para geração de um sinal de áudio binaural de um sinal de downmix mono.

HISTÓRICO DA INVENÇÃO

Na última década houve uma tendência para
10 multicanais de áudio e, especificamente, na direção da extensão de áudio espacial além de sinais estéreo convencionais. Por exemplo, gravações estéreas tradicionais apenas compreendem dois canais enquanto sistemas de áudio modernos, tipicamente, usam cinco ou seis canais, como nos sistemas de som "surround" 5.1 populares.
15 Isto provê uma experiência de audição mais envolvente onde o usuário pode ser envolvido pelas fontes de som.

Várias técnicas e padrões foram desenvolvidos para comunicação destes sinais de multicanais. Por exemplo, seis canais discretos representando um sistema surround 5.1 podem ser
20 transmitidos de acordo com padrões tais como padrões de Codificação de Áudio Avançada (AAC) ou Dolby Digital.

Entretanto, de modo a prover compatibilidade retroativa, é conhecido como fazer o downmix do número mais elevado de canais para um número menor, e especificamente, ele é
25 frequentemente usado para fazer o downmix de um sinal de som surround 5.1 para um sinal estéreo que permita a um sinal estéreo ser reproduzido por decodificadores "legacy" (estéreo) e um sinal 5.1 por decodificadores de som surround.

Um exemplo é o método de codificação compatível retroativamente MPEG2. É feito o downmix de um sinal de multicanais para um sinal estéreo. Sinais adicionais são codificados na porção de dados auxiliares permitindo que um

5 decodificador de multicanais MPEG2 gere uma representação do sinal de multicanais. Um decodificador MPEG1 irá desconsiderar os dados auxiliares e, dessa maneira, decodificar apenas o downmix estéreo.

Existem vários parâmetros que podem ser usados para descrever as propriedades espaciais de sinais de áudio. Um

10 destes parâmetros é a correlação cruzada entre canais, tal como a correlação cruzada entre o canal esquerdo e o canal direito para sinais estéreos. Um outro parâmetro é a proporção de energia dos canais. Nos codificadores de áudio espacial assim denominados (paramétricos), estes e outros parâmetros são extraídos do sinal

15 de áudio original de modo a produzir um sinal de áudio tendo um número reduzido de canais, por exemplo, apenas um canal único, mais um conjunto de parâmetros descrevendo as propriedades espaciais do sinal de áudio original. Nos decodificadores de áudio espaciais (paramétricos), as propriedades espaciais conforme

20 descrito pelos parâmetros espaciais transmitidos são reintegradas.

Posicionamento de fonte de som 3D está atualmente em evidência, especialmente no domínio móvel. Efeitos de som e reprodução de música em jogos móveis podem adicionar valor significativo à experiência do consumidor quando posicionado em

25 3D, efetivamente criando um efeito 3D "externo à cabeça". Especificamente, é sabido como gravar e reproduzir sinais de áudio binaurais que contêm informação direcional específica à qual o ouvido humano é sensível. Gravações binaurais são tipicamente

feitas usando dois microfones montados em uma cabeça humana de um manequim, de modo que o som gravado corresponda ao som capturado pelo ouvido humano e inclua quaisquer influências devido ao formato da cabeça e dos ouvidos. Gravações binaurais diferem de gravações estéreas (isto é, estereofônicas) pelo fato da reprodução de uma gravação binaural ser geralmente destinada para um conjunto de fones de ouvido, enquanto uma gravação estérea é geralmente feita para reprodução por alto-falantes. Embora uma gravação binaural permita a reprodução de uma informação espacial usando apenas dois canais, uma gravação estérea não proveria a mesma percepção espacial.

Gravações de canal duplo regular (estereofônico) ou canais múltiplos (por exemplo, 5.1) podem ser transformadas em gravações binaurais pela convolução de cada sinal regular com um conjunto de funções de transferência perceptual. Estas funções de transferência perceptual modelam a influência da cabeça humana, e possivelmente outros objetos no sinal. Um tipo bem conhecido de função de transferência perceptual espacial é a Função de Transferência Relacionada à Cabeça (HRTF). Um tipo alternativo de função de transferência perceptual espacial, que também considera reflexões causadas pelas paredes, teto e piso de uma sala, é a Resposta de Impulso de Sala Binaural (BRIR).

Tipicamente, algoritmos de posicionamento 3D empregam HRTFs (ou BRIRs), que descrevem a transferência de certa posição da fonte de som para os tímpanos por meio de uma resposta de impulso. Posicionamento de fonte de som 3D pode ser aplicado a sinais de multicanais por meio de HRTFs, portanto permitindo que um sinal binaural proveja informação de som espacial para um

usuário, por exemplo, usando um par de fones de ouvido.

Um algoritmo de síntese binaural convencional é descrito na Figura 1. Um conjunto de canais de entrada é filtrado por um conjunto de HRTFs. Cada sinal de entrada é dividido em dois
5 sinais (um componente esquerdo "L", e um direito "R"); cada um destes sinais é subsequentemente filtrado por uma HRTF correspondendo à posição de fonte de som desejada. Todos os sinais do ouvido esquerdo são subsequentemente somados para gerar o sinal de saída binaural esquerdo, e os sinais do ouvido direito são
10 somados para gerar o sinal de saída binaural direito.

Sistemas de decodificador são conhecidos por poderem receber um sinal codificado de som surround e gerar uma experiência de som surround de um sinal binaural. Por exemplo, sistemas de fones de ouvido são conhecidos por permitirem que um
15 sinal de som surround seja convertido em um sinal binaural de som surround para prover uma experiência de som surround ao usuário dos fones de ouvido.

A Figura 2 ilustra um sistema onde um decodificador surround MPEG recebe um sinal estéreo com dados
20 paramétricos espaciais. A corrente de bits de entrada é desmultiplexada por um desmultiplexador (201) resultando em parâmetros espaciais e uma corrente de bits de downmix. A última corrente de bits é decodificada usando um decodificador mono ou estéreo convencional (203). O downmix decodificado é decodificado
25 por um decodificador especial (205), que gera uma saída de multicanal baseada nos parâmetros espaciais transmitidos. Finalmente, a saída de multicanais é, então, processada por um estágio de síntese binaural (207) (similar àquela da Figura 1)

resultando em um sinal de saída binaural provendo uma experiência de som surround ao usuário.

Entretanto, esta abordagem é complexa e requer recursos computacionais substanciais e pode reduzir ainda mais a
5 qualidade de áudio e introduzir artefatos audíveis.

De modo a superar algumas destas desvantagens foi proposto que um decodificador de áudio de multicanais paramétricos pode ser combinado com um algoritmo de síntese binaural, de modo que um sinal de multicanais pode ser transmitido em fones de
10 ouvido sem requerer que o sinal de multicanais seja primeiramente gerado do sinal de downmix transmitido seguido por um downmix do sinal de multicanais usando filtros HRTF.

Nestes decodificadores, os parâmetros espaciais de upmix para recriar o sinal de multicanais são combinados com os
15 filtros HRTF de modo a gerar parâmetros combinados que podem, diretamente, serem aplicados ao sinal de downmix para gerar o sinal binaural. De modo a fazer isto, os filtros HRTF são parametrizados.

Um exemplo deste decodificador é ilustrado na
20 Figura 3 e ainda descrito por Breebaart, J. "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround", Proc. ICME, Beijing, China (2007) e Breebaart, J., Faller, C. "Spatial audio processing: MPEG Surround and other applications", Wiley & Sons, Nova York (2007).

25 Uma corrente de bits de entrada contendo parâmetros espaciais e um sinal de downmix são recebidos por um desmultiplexador 301. O sinal de downmix é decodificado por um decodificador convencional 303 resultando em um downmix mono ou

estéreo.

Além disso, dados HRTF são convertidos no domínio de parâmetro por meio de uma unidade de extração de parâmetro HRTF 305. Os parâmetros HRTF resultantes são combinados em uma unidade de conversão 307 para gerar parâmetros combinados referidos como parâmetros binaurais. Estes parâmetros descrevem o efeito combinado dos parâmetros espaciais e do processamento HRTF.

O decodificador espacial sintetiza o sinal de saída binaural pela modificação do sinal de downmix decodificado dependente dos parâmetros binaurais. Especificamente, o sinal de downmix é transferido para um domínio de transformada ou banco de filtro por uma unidade de transformada 309 (ou o decodificador convencional 303 pode prover diretamente o sinal de downmix decodificado como um sinal de transformada). A unidade de transformada 309 pode especificamente compreender um banco de filtro QMF para gerar sub-bandas QMF. O sinal de downmix de sub-banda é alimentado a uma unidade de matriz 311 que executa uma operação de matriz 2x2 em cada sub-banda.

Se o downmix transmitido é um sinal estéreo, os dois sinais de entrada para a unidade de matriz 311 são os dois sinais estéreos. Se o downmix transmitido é um sinal mono, um dos sinais de entrada para a unidade de matriz 311 é o sinal mono e o outro sinal é um sinal descorrelacionado (similar ao upmix convencional de um sinal mono para um sinal estéreo).

Para ambos os downmixes, mono e estéreo, a unidade de matriz 311 executa a operação:

$$\begin{bmatrix} y_{L_B}^{n,k} \\ y_{R_B}^{n,k} \end{bmatrix} = \begin{bmatrix} h_{11}^{n,k} & h_{12}^{n,k} \\ h_{21}^{n,k} & h_{22}^{n,k} \end{bmatrix} \begin{bmatrix} y_{L_0}^{n,k} \\ y_{R_0}^{n,k} \end{bmatrix},$$

onde k é o número de índice da sub-banda, n é o número de índice de fenda (intervalo de transformada), $h_{ij}^{n,k}$ são os elementos de matriz para a sub-banda k , $y_{L_0}^{n,k}, y_{R_0}^{n,k}$, os dois sinais de entrada para a sub-banda k , e $y_{L_B}^{n,k}, y_{R_B}^{n,k}$ as amostras de sinal de saída binaural.

A unidade de matriz 311 alimenta as amostras de sinal de saída binaural para uma unidade de transformada inversa 313 que transforma o sinal de volta para o domínio de tempo. O sinal binaural de domínio de tempo resultante pode ser, então, alimentado para os fones de ouvido para prover uma experiência de som surround.

A abordagem descrita tem uma série de vantagens:

O processamento HRTF pode ser executado no domínio de transformada, o que, em muitos casos, pode reduzir o número de transformadas que é requerido visto que o mesmo domínio de transformada pode ser usado para decodificação do sinal de downmix.

A complexidade do processamento é muito baixa (ela usa apenas multiplicação por matrizes 2×2) e é virtualmente independente do número de canais de áudio simultâneos. Isto pode ser aplicado a ambos os downmixes, mono e estéreo;

HRTFs são representadas de uma maneira bastante compacta e, assim, podem ser transmitidas e armazenadas muito eficientemente.

Entretanto, a abordagem também tem algumas desvantagens. Especificamente, a abordagem é apenas adequada para HRTFs tendo respostas de impulso relativamente muito curtas

(geralmente menos que o intervalo de transformada) visto que respostas de impulso mais longas não podem ser representadas pelos valores HRTF de sub-banda parametrizadas. Dessa maneira, a abordagem não é utilizável para ambientes de áudio tendo ecos ou reverberações longos. Especificamente, a abordagem tipicamente não funciona com HRTFs ou Respostas de Impulso de Sala Binaural (BRIRs) ecóicas que podem ser longas e, dessa maneira, muito difíceis de modelar corretamente com a abordagem paramétrica.

Dessa maneira, um sistema melhorado para gerar sinal de áudio binaural seria vantajoso e, especificamente, um sistema que permita flexibilidade aumentada, desempenho melhorado, implementação facilitada, uso reduzido de recursos e/ou aplicabilidade melhorada a diferentes ambientes de áudio seria vantajoso.

SUMÁRIO DA INVENÇÃO

Consequentemente, a invenção busca preferivelmente mitigar, aliviar ou eliminar uma ou mais das desvantagens mencionadas acima sozinhas ou em qualquer combinação.

De acordo com um primeiro aspecto da invenção é provido um aparelho para gerar um sinal de áudio binaural, o aparelho compreendendo: meios para receber dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M para o sinal de áudio do canal N; meios de dados de parâmetro para converter parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural; meios de conversão para converter o sinal de

áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais; um filtro estéreo para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo; e meios de coeficiente para determinar coeficientes de filtros para o filtro estéreo em resposta à função de transferência perceptual binaural.

A invenção pode permitir que um sinal de áudio binaural melhorado seja gerado. Em especial, configurações da invenção podem usar uma combinação de processamento de frequência e tempo para gerar sinais binaurais refletindo ambientes de áudio ecóicos e/ou HRTF ou BRIRs com respostas de impulso longo. Uma implementação de baixa complexidade pode ser atingida. O processamento pode ser implementado com baixa demanda de recursos computacionais e/ou de memória.

O sinal de downmix de áudio do canal M pode, especificamente, ser um sinal mono ou estéreo compreendendo um downmix de um número maior de canais espaciais, tal como um downmix de um sinal surround 5.1 ou 7.1. Os dados de parâmetros espaciais podem especificamente compreender diferenças entre canais e/ou diferenças de correlação cruzada para o sinal de áudio do canal N. A(s) função(ões) de transferência perceptual binaural pode(m) ser função(ões) de transferência HRTF ou BRIR.

Consequentemente a uma característica opcional da invenção, o aparelho compreende ainda meios de transformada para transformar o sinal de áudio do canal M de um domínio de tempo para um domínio de sub-banda e onde os meios de conversão e o filtro estéreo são arrançados para processar individualmente cada sub-banda do domínio de sub-banda.

A característica pode prover implementação facilitada, demandas de recurso reduzidas e/ou compatibilidade com muitas aplicações de processamento de áudio, tais como algoritmos de decodificação convencional.

5 De acordo com uma característica opcional da invenção, uma duração de uma resposta de impulso da função de transferência perceptual binaural excede um intervalo de atualização de transformada.

10 A invenção pode permitir um binaural melhorado para o sinal a ser gerado e/ou pode reduzir a complexidade. Em especial, a invenção pode gerar sinais binaurais correspondendo a ambientes de áudio com características de eco ou reverberação longas.

15 De acordo com uma característica opcional da invenção, os meios de conversão são arranjos para gerar, para cada sub-banda, amostras de saída estérea substancialmente como:

$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_I \\ R_I \end{bmatrix},$$

20 onde pelo menos um dentre L_I e R_I é uma amostra de um canal de áudio do sinal de áudio do canal M na sub-banda e os meios de conversão são arranjos para determinar coeficientes de matriz h_{xy} em resposta a ambos, os dados de parâmetros espaciais e a pelo menos uma função de transferência perceptual binaural.

A característica pode permitir um binaural melhorado ao sinal a ser gerado e/ou pode reduzir a complexidade.

25 De acordo com uma característica opcional da invenção, os meios de coeficiente compreendem: meios para prover uma representação de sub-banda de respostas de impulso de uma

pluralidade de funções de transferência perceptual binaural correspondendo a diferentes fontes de som no sinal do canal N; meios para determinar os coeficientes de filtro por uma combinação pesada de coeficientes correspondentes das representações de sub-
5 banda; e meios para determinar pesos para as representações de sub-banda para a combinação pesada em resposta aos dados de parâmetro espaciais.

A invenção pode permitir um sinal binaural melhorado a ser gerado e/ou pode reduzir a complexidade. Em
10 especial, coeficientes de filtro de baixa complexidade, mas com alta qualidade podem ser determinados.

De acordo com uma característica opcional da invenção, os primeiros parâmetros binaurais compreendem parâmetros de coerência indicativos de uma correlação entre canais do sinal
15 de áudio binaural.

A característica pode permitir um sinal binaural melhorado a ser gerado e/ou pode reduzir complexidade. Em especial, a correlação desejada pode ser eficientemente provida por uma operação de baixa complexidade antes da filtragem.
20 Especificamente, uma multiplicação de matriz de sub-banda de baixa complexidade pode ser executada para introduzir a correlação desejada ou propriedades de coerência ao sinal binaural. Estas propriedades podem ser introduzidas antes da filtragem e sem requerer que os filtros sejam modificados. Assim, a característica
25 pode permitir características de correlação ou de coerência sejam controladas eficientemente e com baixa complexidade.

De acordo com uma característica opcional da invenção, os primeiros parâmetros binaurais não compreendem pelo

menos um dos parâmetros de localização indicativos de uma localização de qualquer fonte de som do sinal de áudio binaural e parâmetros de reverberação indicativos de uma reverberação de qualquer componente de som do sinal de áudio binaural.

5 A característica pode permitir um binaural melhorado para o sinal a ser gerado e/ou pode reduzir a complexidade. Em especial, a característica pode permitir que a informação de localização e/ou parâmetros de reverberação sejam controlados exclusivamente pelos filtros, portanto facilitando a
10 operação e/ou provendo qualidade melhorada. A coerência ou correlação dos canais estéreos binaurais pode ser controlada pelos meios de conversão, portanto permitindo que a correlação/coerência e localização e/ou reverberação sejam controladas independentemente e onde for mais prático ou eficiente.

15 De acordo com uma característica opcional da invenção, os meios de coeficiente são arranjos para determinar os coeficientes de filtro para refletir pelo menos um dos sinais de som de localização e sinais de som de reverberação para o sinal de áudio binaural.

20 A característica pode permitir que um sinal binaural melhorado seja gerado e/ou pode reduzir a complexidade. Em especial, a localização desejada ou propriedades de reverberação podem ser eficientemente providas por filtragem de sub-banda, portanto provendo qualidade melhorada e, em especial,
25 permitindo, por exemplo, que ambientes de áudio ecóticos sejam eficientemente simulados.

De acordo com uma característica opcional da invenção, o sinal de áudio do canal M é um sinal de áudio mono e

os meios de conversão são arranjos para gerar um sinal descorrelacionado do sinal de áudio mono e para gerar o primeiro sinal estéreo por uma multiplicação de matriz aplicada a amostras de um sinal estéreo compreendendo o sinal descorrelacionado e o
5 sinal de áudio mono.

A característica pode permitir um binaural melhorado ao sinal a ser gerado de um sinal mono e/ou pode reduzir a complexidade. Em especial, a invenção pode permitir que todos os parâmetros requeridos para gerar um sinal de áudio binaural de
10 alta qualidade sejam gerados de parâmetros espaciais tipicamente disponíveis.

De acordo com um outro aspecto da invenção, é provido um método de geração de um sinal de áudio binaural, o método compreendendo: recepção de dados de áudio compreendendo um
15 sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N; convertendo parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função
20 de transferência perceptual binaural; convertendo o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais; gerando o sinal de áudio binaural por filtragem dos primeiros sinais estéreos; e determinando coeficientes de filtro para o filtro estéreo em resposta a pelo
25 menos uma função de transferência perceptual binaural.

De acordo com um outro aspecto da invenção, é provido um transmissor para transmitir sinal de áudio binaural; o transmissor compreendendo: meios para recepção de dados de áudio

compreendendo um sinal de áudio do canal M sendo um downmix de um
sinal de áudio do canal N e dados de parâmetros espaciais para
upmix do sinal de áudio do canal M para o sinal de áudio do canal
N; meios de dados de parâmetro para converter parâmetros espaciais
5 dos dados de parâmetros espaciais em primeiros parâmetros
binaurais em resposta a pelo menos uma função de transferência
perceptual binaural; meios de conversão para converter o sinal de
áudio do canal M em um primeiro sinal estéreo em resposta aos
primeiros parâmetros binaurais; um filtro estéreo para gerar o
10 sinal de áudio binaural pela filtragem do primeiro sinal estéreo;
meios de coeficiente para determinar coeficientes de filtro para o
filtro estéreo em resposta à função de transferência perceptual
binaural; e meios para transmitir o sinal de áudio binaural.

De acordo com um outro aspecto da invenção, é
15 provido um sistema de transmissão para transmitir um sinal de
áudio, o sistema de transmissão incluindo um transmissor
compreendendo: meios para receber dados de áudio compreendendo um
sinal de áudio do canal M sendo um downmix de um sinal de áudio do
canal N e dados de parâmetros espaciais para upmix do sinal de
20 áudio do canal M para o sinal de áudio do canal N, meios de dados
de parâmetros para converter parâmetros espaciais dos dados de
parâmetros espaciais nos primeiros parâmetros binaurais em
resposta a pelo menos uma função de transferência perceptual
binaural, meios de conversão para converter o sinal de áudio do
25 canal M em um primeiro sinal estéreo em resposta aos primeiros
parâmetros binaurais, um filtro estéreo para gerar o sinal de
áudio binaural pela filtragem do primeiro sinal estéreo, meios de
coeficiente para determinar coeficientes de filtro para o filtro

estéreo em resposta à função de transferência perceptual binaural, e meios para transmitir o sinal de áudio binaural; e um receptor para receber o sinal de áudio binaural.

De acordo com um outro aspecto da invenção, é
5 provido um dispositivo de gravação para gravar um sinal de áudio binaural, o dispositivo de gravação de áudio compreendendo meios para recepção de dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M
10 para o sinal de áudio do canal N; meios de dados de parâmetro para converter parâmetros espaciais dos dados de parâmetro espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural; meios de conversão para converter o sinal de áudio do canal M em um primeiro sinal
15 estéreo em resposta aos primeiros parâmetros binaurais; um filtro estéreo para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo; meios de coeficiente (419) para determinar coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural; e meios para gravação do
20 sinal de áudio binaural.

De acordo com um outro aspecto da invenção, é provido um método de transmissão de um sinal de áudio binaural, o método compreendendo: recepção dos dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do
25 canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M para o sinal de áudio do canal N; conversão dos parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função

de transferência perceptual binaural; convertendo o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais; geração do sinal de áudio binaural pela filtragem do primeiro sinal estéreo em um filtro estéreo; 5 determinação dos coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural; e transmissão do sinal de áudio binaural.

De acordo com um outro aspecto da invenção, é provido um método para transmitir e receber um sinal de áudio 10 binaural, o método compreendendo: um transmissor executando as etapas de: recepção de dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N, convertendo parâmetros 15 espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural, convertendo o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais, gerando o sinal de áudio binaural pela 20 filtragem do primeiro sinal estéreo em um filtro estéreo, determinando os coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural, e transmitindo o sinal de áudio binaural; e um receptor executando a etapa de recepção do sinal de áudio binaural.

25 De acordo com um outro aspecto da invenção, é provido um produto de programa de computador para executar o método de qualquer um dos métodos descritos acima.

Estes e outros aspectos, características e

vantagens da invenção ficarão aparentes e elucidadas com referência às configurações descritas abaixo.

BREVE DESCRIÇÃO DOS DESENHOS

Configurações da invenção serão descritas, apenas
5 como exemplo, com referência aos desenhos, nos quais:

A Figura 1 é uma ilustração de uma abordagem para geração de um sinal binaural de acordo com o estado da técnica;

A Figura 2 é uma ilustração de uma abordagem para geração de um sinal binaural de acordo com o estado da técnica;

10 A Figura 3 é uma ilustração de uma abordagem para geração de um sinal binaural de acordo com o estado da técnica;

A Figura 4 ilustra um dispositivo para gerar um sinal de áudio binaural de acordo com algumas configurações da invenção;

15 A Figura 5 ilustra um fluxograma de um exemplo de um método para gerar um sinal de áudio binaural de acordo com algumas configurações da invenção; e

A Figura 6 ilustra um exemplo de um sistema de transmissão para comunicação de um sinal de áudio de acordo com
20 algumas configurações da invenção.

DESCRIÇÃO DETALHADA DAS CONFIGURAÇÕES

A descrição a seguir está focada nas configurações da invenção aplicáveis à síntese de um sinal estéreo binaural de um downmix mono de uma pluralidade de canais
25 espaciais. Em especial, a descrição será apropriada para geração de um sinal binaural para reprodução de fone de ouvido de uma corrente de bits de som surround MPEG codificada usando uma configuração denominada "5151" que tem 5 canais como entrada

(indicada pelo primeiro '5'), um downmix mono (o primeiro 'um'), uma reconstrução de 5 canais (o segundo '5') e parametrização espacial de acordo com a estrutura de árvore '1'. Informações detalhadas sobre diferentes estruturas de árvore podem ser encontradas em Herre, J., Kjörling, K., Breebaart, J., Faller, C., Disch, S., Purnhagen, H., Koppens, J., Hilpert, J., Róden, J., Oomen, W., Linzmeier, K., Chong, K. S. "MPEG Surround - The ISO/MPEG standard for efficient and compatible multi-channel audio coding", convenção Proc. 122 AES, Viena, Áustria (2007) e Breebaart, J., Hotho, G., Koppens, J., Schuijers, E., Oomen, W., van de Par, S. "Background, concept, and architecture of the recent MPEG Surround standard on multi-channel audio compression" J. Audio Engineering Society, 55, pág. 331-351 (2007). Entretanto, será observado que a invenção não está limitada a esta aplicação, mas pode, por exemplo, ser aplicada a muitos outros sinais de áudio incluindo, por exemplo, sinais de som surround nos quais foi feito downmix para um sinal estéreo.

Nos dispositivos do estado da técnica como aqueles da Figura 3, HRTFs ou BRIRs longas não podem ser eficientemente representadas pelos dados parametrizados e operação de matriz executada pela unidade de matriz 311. Na verdade, as multiplicações de matriz de sub-banda são limitadas para representar respostas de impulso de domínio de tempo tendo uma duração que corresponde ao intervalo de tempo de transformada usado para a transformação no domínio de tempo de sub-banda. Por exemplo, se a transformada é uma Transformada de Fourier Rápida (FFT), cada intervalo FFT de N amostras é transferido em N amostras de sub-bandas que são alimentadas à unidade de matriz.

Entretanto, respostas de impulso mais longas que N amostras não serão adequadamente representadas.

Uma solução para este problema é usar uma abordagem de filtragem de domínio de sub-banda onde a operação de matriz é substituída por uma abordagem de filtragem de matriz onde as sub-bandas individuais são filtradas. Dessa maneira, nestas configurações, o processamento de sub-bandas pode, ao invés de uma multiplicação de matriz simples, ser provida como:

$$\begin{bmatrix} y_{L_B}^{n,k} \\ y_{R_B}^{n,k} \end{bmatrix} = \sum_{i=0}^{N_q-1} \begin{bmatrix} h_{11}^{n-i,k} & h_{12}^{n-i,k} \\ h_{21}^{n-i,k} & h_{22}^{n-i,k} \end{bmatrix} \begin{bmatrix} y_{L_0}^{n-i,k} \\ y_{R_0}^{n-i,k} \end{bmatrix},$$

onde N_q é o número de torneiras usado para o filtro para representar a(s) função(ões) HRTF/BRIR.

Esta abordagem efetivamente corresponde à aplicação de quatro filtros para cada sub-banda (um para cada permuta de canal de entrada e canal de saída da unidade de matriz 311).

Embora esta abordagem possa ser vantajosa em algumas configurações, ela possui também algumas desvantagens associadas. Por exemplo, o sistema requer quatro filtros para cada sub-banda, o que significativamente aumenta a complexidade e os requisitos de recursos para o processamento. Além do mais, em muitos casos pode ser complicado, difícil ou mesmo impossível gerar os parâmetros que correspondem precisamente com as respostas de impulso HRTF/BRIR desejadas.

Especificamente, para a multiplicação de matriz simples da Figura 3, a coerência do sinal binaural pode ser estimada com a ajuda de parâmetros HRTF e parâmetros espaciais transmitidos devido a ambos os tipos de parâmetro existirem no

mesmo domínio (parâmetro). A coerência do sinal binaural depende da coerência entre sinais de fonte de som individuais (conforme descrito pelos parâmetros espaciais), e o caminho acústico das posições individuais para os tímpanos (descrito pelas HRTFs). Se os níveis de sinal relativos, valores de coerência aos pares, e funções de transferência HRTF são todos descritos de uma maneira estatística (paramétrica), a coerência líquida resultante do efeito combinado de transmissão espacial e processamento HRTF pode ser estimada diretamente no domínio de parâmetro. Este processo é descrito por Breebaart, J. "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround", Proc. ICME, Beijing, China (2007) e Breebaart, J., Faller, C. "Spatial audio processing: MPEG Surround and other applications", Wiley & Sons, Nova York (2007). Se a coerência desejada é conhecida, um sinal de saída com uma coerência de acordo com o valor especificado pode ser obtido por uma combinação de um sinal de descorrelacionador e o sinal mono por meio de uma operação de matriz. Este processo é descrito por Breebaart, J., van de Par, S., Kohlrausch, A., Schuijers, E. "Parametric coding of stereo audio", EURASIP J. Applied Signal Proc. 9, pág. 1305-1322 (2005) e Engdegård, J., Purnhagen, H., Rödén, J., Liljeryd, L. "Synthetic ambience in parametric stereo coding", Proc. da 116^a AES convenção, Berlin, Alemanha (2004).

Como um resultado, as entradas de matriz do sinal do descorrelacionador (h_{12} e h_{22}) seguem de relações relativamente simples entre parâmetros HRTF e espaciais. Entretanto, para respostas de filtro tais como aquelas descritas acima, é significativamente mais difícil calcular a coerência líquida

resultante da decodificação espacial e síntese binaural devido ao valor da coerência desejado ser diferente para a primeira parte (o som direto) da BRIR em relação à parte remanescente (a última reverberação).

5 Especificamente, para BRIRs, as propriedades requeridas podem mudar consideravelmente com o tempo. Por exemplo, a primeira parte de uma BRIR pode descrever o som direto (sem efeitos da sala). Esta parte é, portanto, altamente direcional (com propriedades de localização distintas refletidas, por exemplo, pelas diferenças de nível e diferenças de tempo de chegada, e uma coerência elevada). As reflexões iniciais e a última reverberação, por outro lado, são freqüente e relativamente menos direcionais. Dessa maneira, as diferenças de nível entre os ouvidos são menos pronunciadas, as diferenças de tempo de chegada são difíceis de determinar precisamente devido à natureza estocástica destes, e a coerência é, em muitos casos, muito baixa. Esta mudança de propriedades de localização é bastante importante para capturar precisamente, mas isto pode ser difícil porque isto iria requerer que a coerência das respostas de filtro fosse alterada dependendo da posição dentro da resposta de filtro real, enquanto, simultaneamente, a resposta de filtro completa deve depender dos parâmetros espaciais e dos coeficientes HRTF. Esta combinação de requisitos é muito difícil de preencher com um número limitado de etapas de processamento.

25 Em resumo, determinar a coerência correta entre os sinais de saída binaurais e garantir seu comportamento temporal correto é muito difícil para um downmix mono e é tipicamente impossível usando as abordagens conhecidas para a abordagem de

multiplicação de matriz do estado da técnica.

A Figura 4 ilustra um dispositivo para gerar um sinal de áudio binaural de acordo com algumas configurações da invenção. Na abordagem descrita, multiplicação de matriz paramétrica é combinada com filtragem de baixa complexidade para permitir ambientes de áudio com eco ou reverberação longo a ser emulado. Em especial, o sistema permite que HRTFs/BRIRs longas sejam usadas enquanto uma baixa complexidade e uma implementação prática são mantidas.

O dispositivo compreende um desmultiplexador 401 que recebe uma corrente de dados de áudio que compreende um sinal de áudio do canal M de áudio que é um downmix de um sinal de áudio do canal N. Além disso, os dados compreendem dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N. No exemplo específico, o sinal de downmix é um sinal mono, isto é, $M=1$ e o sinal de áudio do canal N é um sinal surround 5.1, isto é, $N=6$. Os dados de áudio são especificamente uma codificação MPEG Surround de um sinal surround e os dados espaciais compreendem parâmetros de Diferenças entre Níveis (ILDS) e Correlação Cruzada Entre Canais (ICC).

Os dados de áudio do sinal mono são alimentados a um decodificador 403 acoplado ao desmultiplexador 401. O decodificador 403 decodifica o sinal mono usando um algoritmo de decodificação convencional adequado como é bem conhecido por uma pessoa especializada na técnica. Dessa maneira, no exemplo, a saída do decodificador 403 é um sinal de áudio mono decodificado.

O decodificador 403 é acoplado a um processador de transformada 405 que é operável para converter o sinal mono

decodificado do domínio de tempo a um domínio de sub-banda de frequência. Em algumas configurações, o processador de transformada 405 pode ser arranjado para dividir o sinal em intervalos de transformada (correspondendo a blocos de amostra compreendendo um número adequado de amostras) e executar uma Transformada de Fourier Rápida (FFTO) em cada intervalo de tempo de transformada. Por exemplo, a FFT pode ser uma FFT de 64 pontos com as amostras de áudio mono sendo divididas em 64 blocos de amostra aos quais a FFT é aplicada para gerar 64 amostras de sub-banda complexas.

No exemplo específico, o processador de transformada 405 compreende um banco de filtro QMF operando com um intervalo de transformada de 64 amostras. Dessa maneira, para cada bloco de 64 amostras de domínio de tempo, 64 amostras de sub-banda são geradas no domínio de frequência.

No exemplo, o sinal recebido é um sinal mono que deve passar por upmix para um sinal estéreo binaural. Consequentemente, o sinal mono de sub-banda de frequência é alimentado a um descorrelacionador 407 que gera uma versão descorrelacionada do sinal mono. Será observado que qualquer método adequado de gerar um sinal descorrelacionado pode ser usado sem sair da invenção.

O processador de transformada 405 e descorrelacionador 407 são alimentados a um processador de matriz 409. Dessa maneira, o processador de matriz 409 é alimentado à representação de sub-banda do sinal mono, assim como a representação de sub-banda do sinal descorrelacionado gerado. O processador de matriz 409 prossegue para converter o sinal mono em

um primeiro sinal estéreo. Especificamente, o processador de matriz 409 executa uma multiplicação de matriz em cada sub-banda provida por:

$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_I \\ R_I \end{bmatrix},$$

5 onde L_I e R_I são a amostra dos sinais de entrada no processador de matriz 409, isto é, no exemplo específico L_I e R_I são as amostras de sub-banda do sinal mono e do sinal descorrelacionado.

 A conversão executada pelo processador de matriz
10 409 depende dos parâmetros binaurais gerados em resposta às HRTFs/BRIRs. No exemplo, a conversão também depende dos parâmetros espaciais que se relacionarem com o sinal mono recebido e os canais espaciais (adicionais).

 Especificamente, o processador de matriz 409 é
15 acoplado a um processador de conversão 411 que é ainda acoplado ao desmultiplexador 401 e uma armazenagem HRTF 413 compreendendo os dados representando a HRTF(s) desejada(s) (ou equivalentemente a BRIR(s) desejada(s). A seguir, apenas por uma questão de brevidade, será feita referência apenas a HRTF(s), mas será
20 observado que BRIR(s) podem ser usadas ao invés (ou também como) HRTFs). O processador de conversão 411 recebe os dados espaciais do desmultiplexador e os dados representando a HRTF da armazenagem HRTF 413. O processador de conversão 411, então, prossegue para gerar os parâmetros binaurais usados pelo processador de matriz
25 409 pela conversão dos parâmetros espaciais nos primeiros parâmetros binaurais em resposta aos dados HRTF.

Entretanto, no exemplo, a parametrização completa

da HRTF e parâmetros espaciais necessários para gerar um sinal binaural de saída não são calculados. Ao invés disso, os parâmetros binaurais usados na multiplicação de matriz apenas refletem parte da resposta HRTF desejada. Em especial, os parâmetros binaurais são estimados para a parte direta (excluindo reflexões iniciais e última reverberação) da HRTF/BRIR apenas. Isto é atingido usando o processo de estimativa de parâmetro convencional, usando o primeiro pico da resposta de impulso de tempo-domínio HRTF apenas durante o processo de parametrização HRTF. Apenas a coerência resultante para a parte direta (excluindo sinais de som de localização tais como diferenças de nível e/ou tempo) é subsequentemente usada na matriz 2×2 . Na verdade, no exemplo específico, os coeficientes de matriz são gerados apenas para refletir a coerência ou correlação desejada do sinal binaural e não incluir consideração das características de localização ou reverberação.

Dessa maneira, a multiplicação de matriz apenas executa parte do processamento desejado e a saída do processador de matriz 409 não é o sinal binaural final, mas é, ao invés disso, um sinal intermediário (binaural) que reflete a coerência desejada do som direto entre os canais.

Os parâmetros binaurais na forma dos coeficientes de matriz h_{xy} estão no exemplo gerado pelo primeiro cálculo das energias de sinal relativas nos diferentes canais de áudio do sinal do canal N baseado nos dados espaciais e especificamente baseado nos parâmetros de diferença de nível contidos no mesmo. As energias relativas em cada um dos canais binaurais são então calculadas com base nestes valores e nas HRTFs associadas com cada

um dos canais N. Também, um valor esperado para a correlação cruzada entre os sinais binaurais é calculado com base nas energias de sinal em cada um dos canais N e nas HRTFs. Com base na correlação cruzada e na energia combinada do sinal binaural, uma
 5 medição de coerência para o canal é subsequentemente calculada e os parâmetros de matriz são determinados para prover esta correlação. Detalhes específicos de como os parâmetros binaurais podem ser gerados serão descritos posteriormente.

O processador de matriz 409 é acoplado a dois
 10 filtros 415, 417 que são operáveis para gerar o sinal de áudio binaural de saída pela filtragem do sinal estéreo gerado pelo processador de matriz 409. Especificamente, cada um dos dois sinais é filtrado individualmente como um sinal mono e nenhum acoplamento cruzado de qualquer sinal de um canal para o outro é
 15 introduzido. Consequentemente, apenas dois filtros monos são empregados, portanto reduzindo a complexidade se comparado, por exemplo, com abordagens requerendo quatro filtros.

Os filtros 415, 417 são filtros de sub-bandas onde cada sub-banda é individualmente filtrada. Especificamente,
 20 cada um dos filtros podem ser filtros de Resposta de Impulso Finito (FIR), em cada sub-banda executando uma filtragem provida substancialmente por:

$$z^{n,k} = \sum_{i=0}^{N_g-1} c_i^k \cdot y_0^{n-i,k}$$

onde y representa as amostras de sub-banda
 25 recebidas do processador de matriz 409, c são os coeficientes de filtro, n é o número de amostra (correspondendo ao número de intervalo de transformada), k é a sub-banda e N é o comprimento da

resposta de impulso do filtro. Dessa maneira, em cada sub-banda individual, uma filtragem de "domínio de tempo" é executada, portanto, estendendo o processamento de um intervalo de transformada único para considerar as amostras de sub-banda de uma pluralidade de intervalos de transformada.

As modificações de sinal de surround MPEG são executadas no domínio de um banco de filtro modulado complexo, o QMF, que não é amostrado criticamente. Seu projeto específico permite que um filtro de domínio de tempo provido seja implementado com alta precisão pela filtragem de cada sinal de sub-banda na direção de tempo com um filtro separado. A SNR geral resultante para implementação de filtro está na faixa de 50 dB com a parte de introdução de erro significativamente menor. Além do mais, estes filtros de domínio de sub-banda podem ser derivados diretamente do filtro de domínio de tempo provido. Um método especificamente atraente para computar o filtro de domínio de sub-banda correspondendo a um filtro de domínio $h(\nu)$ é usar um segundo banco de filtro de análise modulado complexo com um filtro protótipo FIR $q(\nu)$ derivado do filtro protótipo do banco de filtro QMF. Especificamente,

$$c_i^k = \sum_{\nu} h(\nu + iL) q(\nu) \exp\left(-j \frac{\pi}{L} \left(k + \frac{1}{2}\right) \nu\right),$$

onde $L=64$. Para o banco MPEG Surround QMF, o filtro protótipo do conversor de filtro $q(\nu)$ tem 192 torneiras. Como um exemplo, o filtro do domínio de tempo com 1024 torneiras será convertido em um conjunto de 64 filtros de sub-bandas todos tendo 18 torneiras na direção de tempo.

As características do filtro estão no exemplo

gerado para refletir ambos os aspectos dos parâmetros espaciais, assim como aspectos das HRTFs desejadas. Especificamente, os coeficientes de filtro são determinados em resposta às respostas de impulso HRTF e aos sinais de som de local espacial, de modo que a reverberação e características de localização do sinal binaural desejada sejam introduzidas e controladas pelos filtros. A correlação ou coerência da parte direta dos sinais binaurais não são afetadas pela filtragem, assumindo que a parte direta dos filtros é (quase) coerente e, dessa maneira, a coerência do som direto da saída binaural é totalmente definida pela operação de matriz precedente. A parte da última-reverberação dos filtros, por outro lado, é pressuposta estar não correlacionada entre os filtros do ouvido esquerdo e direito e, dessa maneira, a saída da parte específica estará sempre não correlacionada, independente da coerência do sinal alimentado nestes filtros. Assim, nenhuma modificação é requerida para os filtros em resposta à coerência desejada. Dessa maneira, a operação de matriz processando os filtros determina a coerência desejada da parte direta, enquanto a parte de reverberação remanescente irá, automaticamente, ter a correlação (baixa) correta, independente dos valores de matriz reais. Dessa maneira, a filtragem mantém a coerência desejada introduzida pelo processador de matriz 409.

Dessa maneira, no dispositivo da Figura 4, os parâmetros binaurais (na forma dos coeficientes de matriz) usados pelo processador de matriz 409 são parâmetros de coerência indicativos de uma correlação entre canais do sinal de áudio binaural. Entretanto, estes parâmetros não compreendem parâmetros de localização indicativos de um local de qualquer fonte de som do

5 sinal de áudio binaural ou parâmetros de reverberação indicativos de uma reverberação de qualquer componente de som do sinal de áudio binaural. Ao invés disto, estes parâmetros/características são introduzidos pela filtragem de sub-banda subsequente pela determinação dos coeficientes de filtro, de modo que eles refletem os sinais de som de localização e sinais de som de reverberação para sinal de áudio binaural.

Especificamente, os filtros são acoplados a um processador de coeficiente 419 que é ainda acoplado ao desmultiplexador 401 e à armazenagem HRTF 413. O processador de coeficiente 419 determina os coeficientes de filtro para o filtro estéreo 415, 417 em resposta à(s) função(ões) de transferência perceptual binaural. Além disso, o processador do coeficiente 419 recebe os dados espaciais do desmultiplexador 401 e usa isto para determinar os coeficientes de filtro.

Especificamente, as respostas de impulso HRTF são convertidas no domínio de sub-banda e como a resposta de impulso excede um intervalo de transformada único, isto resulta em uma resposta de impulso para cada canal em cada sub-banda ao invés de em um coeficiente de sub-banda único. As respostas de impulso para cada filtro HRTF correspondendo a cada um dos canais N são então somadas em uma soma pesada. Os pesos que são aplicados a cada uma das respostas de impulso de filtro HRTF N são determinados em resposta aos dados espaciais e são especificamente determinados para resultar na distribuição de energia apropriada entre os diferentes canais. Detalhes específicos de como os coeficientes de filtro podem ser gerados serão descritos posteriormente.

A saída dos filtros 415, 417 é, dessa maneira,

uma representação de sub-banda estéreo de um sinal de áudio binaural que efetivamente emula um sinal surround completo quando apresentado em fones de ouvido. Os filtros 415, 417 são acoplados a um processador de transformada inverso 421 que executa uma transformada inversa para converter o sinal de sub-banda no domínio de tempo. Especificamente, o processador de transformada inversa 421 pode executar uma transformada QMF inversa.

Dessa maneira, a saída do processador de transformada inversa 421 é um sinal binaural que pode prover uma experiência de som surround de um conjunto de fones de ouvido. O sinal pode, por exemplo, ser codificado usando um codificador estéreo convencional e/ou pode ser convertido ao domínio analógico em um conversor analógico para digital para prover um sinal que pode ser alimentado diretamente a fones de ouvido.

Dessa maneira, o dispositivo da Figura 4 combina processamento de matriz HRTF paramétrico e filtragem de sub-banda para prover um sinal binaural. A separação de uma multiplicação de matriz de correlação/coerência e uma localização baseada em filtro e filtragem de reverberação provê um sistema onde os parâmetros requeridos podem ser prontamente computados, por exemplo, para um sinal mono. Especificamente, em contraste a uma abordagem de filtragem pura onde o parâmetro de coerência é difícil ou impossível determinar e implementar, a combinação de diferentes tipos de processamento permite que a coerência seja eficientemente controlada, mesmo para aplicações baseadas em um sinal de downmix mono.

Dessa maneira, a abordagem descrita tem a vantagem de que a síntese da coerência correta (por meio da

multiplicação de matriz) é a geração de sinais de som de localização e reverberação (por meio dos filtros) são completamente separadas e controladas independentemente. Além do mais, o número de filtros é limitado a dois visto que nenhuma
 5 filtragem de canal é requerida. Como os filtros são tipicamente mais complexos que a multiplicação de matriz simples, a complexidade é reduzida.

A seguir, um exemplo específico de como os parâmetros binaurais de matriz requeridos e coeficientes de filtro
 10 podem ser calculados, será descrito. No exemplo, o sinal recebido é uma corrente de bits surround MPEG codificada usando uma estrutura de árvore '5151'.

Na descrição, os acrônimos a seguir serão usados:

15 l ou L: Canal esquerdo
 r ou R: Canal direito
 f: canal(is) frontal(is)
 s: Canal(is) surround
 c: Canal central
 ls: Surround esquerdo
 20 rs: Surround direito
 lf: Frontal esquerdo
 lr: Direito esquerdo

Os dados espaciais compreendidos na corrente de dados MPEG incluem os parâmetros a seguir:

Parâmetro	Descrição
CLD_{fs}	Diferença de nível frontal vs surround
CLD_{fc}	Diferença de nível frontal vs central

CLD_f	Diferença de nível frontal esquerdo vs frontal direito
CLD_s	Diferença de nível surround esquerdo vs surround direito
ICC_{fs}	Correlação frontal vs surround
ICC_{fc}	Correlação frontal vs central
ICC_f	Correlação frontal esquerda vs frontal direita
ICC_s	Correlação surround esquerda vs surround direita
CLD_{lfe}	Diferença de nível central vs LFE

Primeiramente, a geração dos parâmetros binaurais usados para a multiplicação de matriz pelo processador de matriz 409 será descrita.

O processador de conversão 411 primeiro calcula
 5 uma estimativa da coerência binaural que é um parâmetro refletindo a coerência desejada entre os canais dos sinais de saída binaural. A estimativa usa os parâmetros espaciais assim como parâmetros HRTF determinados para as funções HRTF.

Especificamente, aos parâmetros HRTF a seguir são
 10 usados:

P_l que é a energia rms dentro de certa banda de frequência da HRTF correspondendo ao ouvido esquerdo;

P_r que é a energia rms dentro de certa banda de frequência de uma HRTF correspondendo ao ouvido direito;

15 ρ que é a coerência dentro de certa banda de frequência entre a HRTF do ouvido esquerdo e direito para certa posição de fonte de som virtual.

ϕ que é a diferença de fase média dentro de certa banda de frequência entre HRTF do ouvido esquerdo e direito para certa posição de fonte de som virtual

Pressupondo representação HRTF de domínio de frequência $H_l(f)$, $H_r(f)$, para os ouvidos esquerdo e direito, respectivamente, e f o índice de frequência, estes parâmetros podem ser calculados de acordo com:

$$P_l = \sqrt{\sum_{f=f(b)}^{f=f(b+1)-1} H_l(f) H_l^*(f)}$$

$$P_r = \sqrt{\sum_{f=f(b)}^{f=f(b+1)-1} H_r(f) H_r^*(f)}$$

$$\phi = \arg \left(\sum_{f=f(b)}^{f=f(b+1)-1} H_l(f) H_r^*(f) \right)$$

$$\rho = \frac{\left| \sum_{f=f(b)}^{f=f(b+1)-1} H_l(f) H_r^*(f) \right|}{P_l P_r}$$

Onde soma através de f é executada para cada banda de parâmetro para resultar em um conjunto de parâmetros para cada banda de parâmetro b . Mais informação sobre este processo de parametrização HRTF pode ser obtida de Breebaart, J. "Analysis and synthesis of binaural parameters for efficient 3D audio rendering in MPEG Surround", Proc. ICME, Beijing, China (2007) e Breebaart, J., Faller, C. "Spatial audio processing: MPEG Surround and other applications", Wiley & Sons, Nova York (2007).

O processo de parametrização acima é executado independentemente para cada banda de parâmetro e cada posição de alto-falante virtual. A seguir, a posição do alto-falante é denotada por $P_l(X)$, com X o identificador do alto-falante (lf, rf, c, ls ou ls).

Como uma primeira etapa, as energias relativas (com relação à energia do sinal de entrada mono) do sinal de canal 5.1 são computadas usando os parâmetros CLD transmitidos. A energia relativa do canal esquerdo-frontal é provida por:

$$\sigma_{lf}^2 = r_1(\text{CLD}_{fs})r_1(\text{CLD}_{fc})r_1(\text{CLD}_f),$$

com

$$r_1(\text{CLD}) = \frac{10^{\text{CLD}/10}}{1 + 10^{\text{CLD}/10}},$$

e

$$r_2(\text{CLD}) = \frac{1}{1 + 10^{\text{CLD}/10}}.$$

10 Similarmente, as energias relativas dos outros canais são providas por:

$$\sigma_{rf}^2 = r_1(\text{CLD}_{fs})r_1(\text{CLD}_{fc})r_2(\text{CLD}_f)$$

$$\sigma_c^2 = r_1(\text{CLD}_{fs})r_2(\text{CLD}_{fc})$$

$$\sigma_{ls}^2 = r_2(\text{CLD}_{fs})r_1(\text{CLD}_s)$$

$$15 \quad \sigma_{rs}^2 = r_2(\text{CLD}_{fs})r_2(\text{CLD}_s)$$

Providas as energias \square de cada falante virtual, os parâmetros ICC que representam valores de coerência entre certos pares de falantes, e os parâmetros HRTF P_l , P_r , \square , e ϕ para cada alto-falante virtual, os atributos estatísticos do sinal 20 binaural resultante podem ser estimados. Isto é atingido pela adição da contribuição em termos de energia \square para cada alto-falante virtual, multiplicado pela energia da HRTF P_l , P_r para cada ouvido individualmente para refletir a alteração na energia introduzida pelo HRTF. Termos adicionais são requeridos para 25 incorporar o efeito de correlações mútuas entre sinais de alto-

falante virtual (ICC) e as diferenças de comprimento de caminho da HRTF (representadas pelo parâmetro ϕ) (vide, por exemplo, Breebaart, J., Fallor, C. "Spatial audio processing: MPEG Surround and other applications", Wiley & Sons, Nova York (2007)).

- 5 O valor esperado da energia relativa do canal de saída binaural esquerdo σ_L^2 (com relação ao canal de entrada mono) é provido por:

$$\begin{aligned}\sigma_L^2 = & P_l^2(C)\sigma_c^2 + P_l^2(Lf)\sigma_{lf}^2 + P_l^2(Ls)\sigma_{ls}^2 + P_l^2(Rf)\sigma_{rf}^2 + P_l^2(Rs)\sigma_{rs}^2 + \dots \\ & 2P_l(Lf)P_l(Rf)\rho(Rf)\sigma_{lf}\sigma_{rf}ICC_f \cos(\phi(Rf)) + \dots \\ & 2P_l(Ls)P_l(Rs)\rho(Rs)\sigma_{ls}\sigma_{rs}ICC_s \cos(\phi(Rs))\end{aligned}$$

- 10 Similarmente, a energia (relativa) para o canal direito é provida por:

$$\begin{aligned}\sigma_R^2 = & P_r^2(C)\sigma_c^2 + P_r^2(Lf)\sigma_{lf}^2 + P_r^2(Ls)\sigma_{ls}^2 + P_r^2(Rf)\sigma_{rf}^2 + P_r^2(Rs)\sigma_{rs}^2 + \dots \\ & 2P_r(Lf)P_r(Rf)\rho(Lf)\sigma_{lf}\sigma_{rf}ICC_f \cos(\phi(Lf)) + \dots \\ & 2P_r(Ls)P_r(Rs)\rho(Ls)\sigma_{ls}\sigma_{rs}ICC_s \cos(\phi(Ls))\end{aligned}$$

Baseado em pressupostos similares e usando técnicas similares, o valor esperado para o produto cruzado $L_B R_B^*$ do par de sinal binaural pode ser calculado de

$$\begin{aligned}\langle L_B R_B^* \rangle = & \sigma_c^2 P_l(C)P_r(C)\rho(C)\exp(j\phi(C)) + \dots \\ & \sigma_{lf}^2 P_l(Lf)P_r(Lf)\rho(Lf)\exp(j\phi(Lf)) + \dots \\ & \sigma_{rf}^2 P_l(Rf)P_r(Rf)\rho(Rf)\exp(j\phi(Rf)) + \dots \\ & \sigma_{ls}^2 P_l(Ls)P_r(Ls)\rho(Ls)\exp(j\phi(Ls)) + \dots \\ 15 & \sigma_{rs}^2 P_l(Rs)P_r(Rs)\rho(Rs)\exp(j\phi(Rs)) + \dots \\ & P_l(Lf)P_r(Rf)\sigma_{lf}\sigma_{rf}ICC_f + \dots \\ & P_l(Ls)P_r(Rs)\sigma_{ls}\sigma_{rs}ICC_s + \dots \\ & P_l(Rs)P_r(Ls)\sigma_{ls}\sigma_{rs}ICC_s \rho(Ls)\rho(Rs)\exp(j(\phi(Rs) + \phi(Ls))) + \dots \\ & P_l(Rf)P_r(Lf)\sigma_{lf}\sigma_{rf}ICC_f \rho(Lf)\rho(Rf)\exp(j(\phi(Rf) + \phi(Lf)))\end{aligned}$$

A coerência da saída binaural (ICC_B) é, então, provida por:

$$ICC_B = \frac{\left| \langle L_B R_B^* \rangle \right|}{\sigma_L \sigma_R},$$

Baseado na coerência determinada do sinal de saída binaural ICC_B (e ignorando os sinais de som de localização e características de reverberação) os coeficientes de matriz
 5 requeridos para reintegrar os parâmetros ICC_B podem, então, serem calculados usando métodos convencionais conforme especificado em Breebaart, J., van de Par, S., Kohlrausch, A., Schuijers, E. "Parametric coding of stereo audio", EURASIP J. Applied Signal Proc. 9, pág. 1305-1322 (2005):

$$10 \quad h_{11} = \cos(\alpha + \beta)$$

$$h_{12} = \sin(\alpha + \beta)$$

$$h_{21} = \cos(-\alpha + \beta)$$

$$h_{22} = \sin(-\alpha + \beta)$$

com

$$15 \quad \alpha = 0.5 \arccos(ICC_B)$$

$$\beta = \arctan\left(\frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L} \tan(\alpha)\right)$$

A seguir a geração dos coeficientes de filtro pelo processador de coeficiente 419 será descrita.

Primeiramente, representações de sub-banda de
 20 respostas de impulso da função de transferência perceptual binaural correspondendo a fontes de som diferentes no sinal de áudio binaural são geradas.

Especificamente, as HRTFs (ou BRIRs) são convertidas no domínio QMF resultando em representações de domínio
 25 QMF $H_{L,X}^{n,k}, H_{R,X}^{n,k}$ para as respostas de impulso do ouvido esquerdo e

ouvido direito, respectivamente, usando o método do conversor de filtro descrito acima na descrição da Figura 4. Na representação, X denota o canal fonte (X=Lf, Rf, C, Ls, Rs), R e L denotam o canal binaural esquerdo e direito, respectivamente, n é o número do bloco de transformada e k denota a sub-banda.

O processador do coeficiente 419, então, prossegue para determinar os coeficientes de filtro como uma combinação pesada de coeficientes correspondentes das representações de sub-banda $H_{L,X}^{n,k}, H_{R,X}^{n,k}$. Especificamente, os coeficientes de filtro para os filtros FIR 415, 417 $H_{L,M}^{n,k}, H_{R,M}^{n,k}$ são providos por:

$$H_{L,M}^{n,k} = g_L^k \cdot (t_{Lf}^k H_{L,Lf}^{n,k} + t_{Ls}^k H_{L,Ls}^{n,k} + t_{Rf}^k H_{L,Rf}^{n,k} + t_{Rs}^k H_{L,Rs}^{n,k} + t_C^k H_{L,C}^{n,k}),$$

$$H_{R,M}^{n,k} = g_R^k \cdot (s_{Lf}^k H_{R,Lf}^{n,k} + s_{Ls}^k H_{R,Ls}^{n,k} + s_{Rf}^k H_{R,Rf}^{n,k} + s_{Rs}^k H_{R,Rs}^{n,k} + s_C^k H_{R,C}^{n,k}).$$

O processador do coeficiente 419 calcula os pesos t^k e s^k conforme descrito a seguir.

Primeiramente, o módulo' dos pesos de combinação linear é escolhido de modo que:

$$|t_X^k| = \sigma_X^k, \quad |s_X^k| = \sigma_X^k$$

Dessa maneira, o peso para uma HRTF provida correspondendo a um canal espacial dado é selecionado para corresponder ao nível de energia daquele canal.

Em segundo lugar, os ganhos de escalonamento g_Y^k são computados como segue.

Deixe a energia de saída binaural de alvo normalizado para a banda híbrida k ser denotada por $(\sigma_Y^k)^2$ para o canal de saída $Y=L,R$, e deixar o ganho de energia $H_{Y,M}^{n,k}$ ser

denotado por $(\sigma_{Y,M}^k)^2$, então os ganhos de escalonamento g_Y^k são ajustados de modo a atingir

$$\sigma_{Y,M}^k = \sigma_Y^k .$$

Observe aqui que se isto pode ser atingido
5 aproximadamente com ganhos de escalonamento que são constantes em cada banda de parâmetro, então o escalonamento pode ser omitido a partir da mudança do filtro e executado pela modificação dos elementos de matriz da seção anterior a

$$h_{11} = g_L \cos(\alpha + \beta)$$

$$10 \quad h_{12} = g_L \sin(\alpha + \beta)$$

$$~~h_{21} = g_R \cos(-\alpha + \beta)~~$$

$$h_{22} = g_R \sin(-\alpha + \beta) .$$

Para isto ser verdadeiro, é um requisito que a combinação pesada não escalonada

$$15 \quad t_{Lf}^k H_{L,Lf}^{n,k} + t_{Ls}^k H_{L,Ls}^{n,k} + t_{Rf}^k H_{L,Rf}^{n,k} + t_{Rs}^k H_{L,Rs}^{n,k} + t_C^k H_{L,C}^{n,k}$$

$$s_{Lf}^k H_{R,Lf}^{n,k} + s_{Ls}^k H_{R,Ls}^{n,k} + s_{Rf}^k H_{R,Rf}^{n,k} + s_{Rs}^k H_{R,Rs}^{n,k} + s_C^k H_{R,C}^{n,k}$$

tenha ganhos de energia que não variem muito dentro das bandas de parâmetro. Tipicamente, uma contribuição principal para estas variações surge de diferenças de atraso
20 importantes entre as respostas HRTF. Em algumas configurações da presente invenção, um pré-alinhamento no domínio de tempo é executado para os filtros HRTF de domínio e os pesos de combinação de valor real simples podem ser aplicados:

$$t_X^k = s_X^k = \sigma_X^k .$$

25 Em outras configurações da presente invenção, aquelas diferenças de atraso são contra atacadas adaptativamente

nos pares HRTF dominantes, por meio da introdução de pesos de valor complexos. No caso de pares frontal/traseiro isto resulta no uso dos pesos a seguir.

$$t_{Lf}^k = \sigma_{Lf}^k \exp \left[-j\phi_{Lf,Ls}^{L,k} \frac{(\sigma_{Ls}^k)^2}{(\sigma_{Lf}^k)^2 + (\sigma_{Ls}^k)^2} \right],$$

5

$$t_{Ls}^k = \sigma_{Ls}^k \exp \left[j\phi_{Lf,Ls}^{L,k} \frac{(\sigma_{Lf}^k)^2}{(\sigma_{Lf}^k)^2 + (\sigma_{Ls}^k)^2} \right],$$

e $t_X^k = \sigma_X^k$ para $X=C, Rf, Rs$.

$$s_{Rf}^k = \sigma_{Rf}^k \exp \left[-j\phi_{Rf,Rs}^{R,k} \frac{(\sigma_{Rs}^k)^2}{(\sigma_{Rf}^k)^2 + (\sigma_{Rs}^k)^2} \right],$$

$$s_{Rs}^k = \sigma_{Rs}^k \exp \left[j\phi_{Rf,Rs}^{R,k} \frac{(\sigma_{Rf}^k)^2}{(\sigma_{Rf}^k)^2 + (\sigma_{Rs}^k)^2} \right],$$

e $s_X^k = \sigma_X^k$ para $X=C, Lf, Ls$.

10

Aqui $\phi_{Xf,Xs}^{X,k}$ é o ângulo de fase não envolvida da correlação cruzada complexa entre os filtros de sub-banda $H_{X,Xf}^{n,k}$ e $H_{X,Xs}^{n,k}$. Esta correlação cruzada é definida por

$$(CIC)_k = \frac{\sum_n (H_{X,Xf}^{n,k})(H_{X,Xs}^{n,k})^*}{\left(\sum_n |H_{X,Xf}^{n,k}|^2 \right)^{1/2} \left(\sum_n |H_{X,Xs}^{n,k}|^2 \right)^{1/2}},$$

onde a estrela denota conjugação complexa.

15

O objetivo do não envolvimento de fase é usar a liberdade de escolha de um ângulo de fase até múltiplos de 2π de modo a obter uma curva de fase que está variando tão lentamente quanto possível como uma função do índice de sub-banda k .

O papel dos parâmetros de ângulo de fase nas fórmulas de combinação acima é duplo. Primeiramente, ele realiza

20

uma compensação de atraso dos filtros frontal/posterior antes da sobreposição que conduz a uma resposta combinada que modela um tempo de atraso principal correspondendo a uma posição de fonte entre os falantes frontais e traseiros. Em segundo lugar, ele
 5 reduz a variabilidade dos ganhos de energia dos filtros não escalonados.

Se o ICC_M de coerência dos filtros combinados $H_{L,M}, H_{R,M}$ em uma banda de parâmetro ou uma banda híbrida é menor que um, a saída binaural pode se tornar menos coerente do que o
 10 pretendido, como ocorre da relação

$$ICC_{B,Out} = ICC_M \cdot ICC_B .$$

A solução para este problema de acordo com algumas configurações da presente invenção é usar um valor ICC_B modificado para a definição do elemento de matriz, definido por

15
$$ICC'_B = \min \left\{ 1, \frac{ICC_B}{ICC_M} \right\} .$$

A Figura 5 ilustra um fluxograma de um exemplo de um método para gerar um sinal de áudio binaural de acordo com algumas configurações da invenção.

O método inicia na etapa 501 onde os dados de
 20 áudio são recebidos compreendendo um sinal de áudio do canal M de áudio sendo um downmix de um sinal de áudio do canal N e dados de parâmetro espaciais para upmix do sinal de áudio do canal M para o sinal de áudio do canal N.

A etapa 501 é seguida pela etapa 503, onde os
 25 parâmetros espaciais dos dados de parâmetro espaciais são convertidos nos primeiros parâmetros binaurais em resposta a uma

função de transferência perceptual binaural.

A etapa 503 é seguida pela etapa 505, onde o sinal de áudio do canal M é convertido em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais.

5 A etapa 505 é seguida pela etapa 507, onde os coeficientes de filtro são determinados para um filtro estéreo em resposta à função de transferência perceptual binaural.

10 A etapa 507 é seguida pela etapa 509, onde o sinal de áudio binaural é gerado por filtragem do primeiro sinal estéreo no filtro estéreo.

O aparelho da Figura 4 pode, por exemplo, ser usado em um sistema de transmissão. ~~A Figura 6 ilustra um exemplo~~ de um sistema de transmissão para comunicação de um sinal de áudio de acordo com algumas configurações da invenção. O sistema de
15 transmissão compreende um transmissor 601 que é acoplado a um receptor 603 através de uma rede 605 que especificamente pode ser a Internet.

No exemplo específico, o transmissor 601 é um dispositivo de gravação de sinal e o receptor 603 é um dispositivo
20 de reproduzidor de sinal, mas será observado que em outras configurações, um transmissor e receptor pode ser usado em outras aplicações e para outros objetivos. Por exemplo, o transmissor 601 e/ou o receptor 603 pode ser parte de uma funcionalidade de transcodificação e pode, por exemplo, prover interface para outras
25 fontes ou destinos de sinal. Especificamente, o receptor 603 pode receber um sinal de som surround codificado e gerar um sinal binaural codificado emulando o sinal de som surround. O sinal binaural codificado pode, então, ser distribuído para outras

fontes.

No exemplo específico onde uma função de gravação de sinal é suportada, o transmissor 601 compreende um digitalizador 607 que recebe um sinal de multicanal analógico (surround) que é convertido para um sinal PCM digital (Código de Pulso Modulado) por amostragem e conversão de analógico-para-digital.

O digitalizador 607 é acoplado ao codificador 609 da Figura 1, que codifica o sinal de multicanal PCM de acordo com um algoritmo de codificação. No exemplo específico, o codificador 609 codifica o sinal como um sinal de som surround codificado MPEG. O codificador 609 é acoplado a um transmissor de rede 611 que recebe o sinal codificado e faz a interface com a Internet 605. O transmissor de rede pode transmitir o sinal codificado para o receptor 603 através da Internet 605.

O receptor 603 compreende um receptor de rede 613 que faz a interface com a Internet 605 e que é arranjado para receber o sinal do transmissor 601.

O receptor de rede 613 é acoplado a um decodificador binaural 615 que, no exemplo, é o dispositivo da Figura 4.

No exemplo específico onde uma função de reprodução de sinal é suportada, o receptor 603 compreende ainda um reproduutor de sinal 1617 que recebe o sinal de áudio binaural do decodificador binaural 615 e apresenta este para o usuário. Especificamente, o reproduutor de sinal 117 pode compreender um conversor de digital-para-analógico, amplificadores e falantes conforme requerido para envio do sinal de áudio binaural para um

conjunto de fones de ouvido.

Será apreciado que a descrição acima para clareza descreveu configurações da invenção com referência a unidades funcionais e processadores diferentes. Entretanto, ficará aparente
5 que qualquer distribuição adequada de funcionalidade entre unidades funcionais ou processadores diferentes pode ser usada sem sair da invenção. Por exemplo, a funcionalidade ilustrada a ser executada por processadores ou controladores separados pode ser executada pelos mesmos processadores ou controladores. Dessa
10 maneira, referências a unidades funcionais específicas são apenas para serem vistas como referências a meios adequados para prover a funcionalidade descrita ao invés de indicativas de uma lógica estrita ou estrutura física ou organização.

A invenção pode ser implementada em qualquer
15 forma adequada incluindo hardware, software, firmware ou qualquer combinação destes. A invenção pode opcionalmente ser implementada pelo menos parcialmente como software de computador operando em um ou mais processadores de dados e/ou processadores de sinal digital. Os elementos e componentes de uma configuração da
20 invenção podem ser fisicamente, funcionalmente e logicamente implementados de qualquer maneira adequada. Na verdade, a funcionalidade pode ser implementada em uma unidade única, em uma pluralidade de unidades ou como parte de outras unidades funcionais. Como tal, a invenção pode ser implementada em uma
25 unidade única ou pode ser fisicamente e funcionalmente distribuída entre unidades e processadores diferentes.

Embora a presente invenção tenha sido descrita em conexão com algumas configurações, não é objetivo estar limitado à

forma específica definida aqui. Ao invés disso, o escopo da presente invenção está limitado apenas pelas reivindicações em anexo. Além disso, embora uma característica possa parecer estar descrita em conexão com configurações específicas, alguém com
5 especialização na técnica reconheceria que várias características das configurações descritas podem ser combinadas de acordo com a invenção. Nas reivindicações, o termo "compreendendo" não exclui a presença de outros elementos ou etapas.

Além disso, embora listados individualmente, uma
10 pluralidade de meios, elementos ou etapas de métodos pode ser implementada, por exemplo, por uma unidade ou processador único. Além disso, embora ~~características individuais~~ possam estar incluídas em reivindicações diferentes, estas podem ser possivelmente vantajosamente combinadas, e a inclusão em
15 diferentes reivindicações não implica que uma combinação de características não seja possível e/ou vantajosa. Também, a inclusão de uma característica em uma categoria de reivindicações não implica em uma limitação a esta categoria, mas ao invés disso, indica que a característica é igualmente aplicável a outras
20 categorias de reivindicação, conforme apropriado. Além disso, a ordem de características nas reivindicações não implica em qualquer ordem específica na qual as características devem ser operadas e, em especial, a ordem de etapas individuais em uma reivindicação de método não implica que as etapas devem ser
25 executadas nesta ordem. Ao invés disso, as etapas podem ser executadas em qualquer ordem adequada. Além disso, referências únicas não excluem uma pluralidade. Dessa maneira, referências a "um", "uma", "primeira", "segunda" etc. não impossibilita uma

pluralidade. Sinais de referência nas reivindicações são providos meramente como um exemplo explicativo e não devem ser considerados como limitativos do escopo das reivindicações de qualquer maneira.

REIVINDICAÇÕES

1. Aparelho para gerar um sinal de áudio binaural, o aparelho compreendendo:

- meios (401, 403) para receber dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M para o sinal de áudio do canal N;

- meios de dados de parâmetro (411) para converter parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural;

- meios de conversão (409) para converter o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais;

Caracterizado por um filtro estéreo (415, 417) para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo; e

- meios de coeficiente (419) para determinar coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural.

2. Aparelho, de acordo com a reivindicação 1, caracterizado pelo fato de que compreende ainda:

- meios de transformada (405) para transformar o sinal de áudio do canal M de um domínio de tempo para um domínio de sub-banda e onde os meios de

conversão e o filtro estéreo são arranjos para processar individualmente cada sub-banda do domínio de sub-banda.

3. Aparelho, de acordo com a reivindicação 2, caracterizado pelo fato de que uma duração de uma resposta de impulso da função de transferência perceptual binaural excede um intervalo de atualização de transformada.

4. Aparelho, de acordo com a reivindicação 2, caracterizado pelo fato de que os meios de conversão (409) são arranjos para gerar, para cada sub-banda, amostras de saída estéreo substancialmente como:

$$\begin{bmatrix} L_o \\ R_o \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} L_I \\ R_I \end{bmatrix},$$

onde pelo menos um dentre L_I e R_I é uma amostra de um canal de áudio do sinal de áudio do canal M na sub-banda e os meios de conversão são arranjos para determinar coeficientes de matriz h_{xy} em resposta a ambos os dados de parâmetro espacial e a pelo menos uma função de transferência perceptual binaural.

5. Aparelho, de acordo com a reivindicação 2, caracterizado pelo fato de que os meios de coeficiente (419) compreendem:

- meios para prover representações de sub-banda de respostas de impulso de uma pluralidade de funções de transferência perceptual binaural correspondendo a diferentes fontes de som no sinal do canal N;
- meios para determinar os coeficientes de

filtro por uma combinação pesada de coeficientes correspondentes das representações de sub-banda; e

- meios para determinar pesos para as representações de sub-banda para as combinações pesadas em resposta aos dados de parâmetros espaciais.

6. Aparelho, de acordo com a reivindicação 1, caracterizado pelo fato de que os primeiros parâmetros binaurais compreendem parâmetros de coerência indicativos de uma correlação entre canais do sinal de áudio binaural.

7. Aparelho, de acordo com a reivindicação 1, caracterizado pelo fato de que os primeiros parâmetros binaurais não compreendem pelo menos um dos parâmetros de localização indicativos de um local de qualquer fonte de som do sinal do canal N e parâmetros de reverberação indicativos de uma reverberação de qualquer componente de som do sinal de áudio binaural.

8. Aparelho, de acordo com a reivindicação 1, caracterizado pelo fato de que os meios de coeficiente (419) são arranjos para determinar os coeficientes de filtro para refletir pelo menos um dos sinais de som de localização e sinais de som de reverberação para o sinal de áudio binaural.

9. Aparelho, de acordo com a reivindicação 1, caracterizado pelo fato de que o sinal de áudio do canal M de áudio é um sinal de áudio mono e os meios de conversão (407, 409) são arranjos para gerar um sinal descorrelacionado do sinal de áudio mono e para gerar o primeiro sinal estéreo por

uma multiplicação de matriz aplicada a amostras de um sinal estéreo compreendendo o sinal descorrelacionado e o sinal de áudio mono.

10. Método de geração de um sinal de áudio binaural, o método compreendendo:

- recepção (501) de dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M para o sinal de áudio do canal N;

- conversão (503) de parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural;

- conversão (505) do sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais;

caracterizado por compreender ainda a geração (509) do sinal de áudio binaural pela filtragem do primeiro sinal estéreo; e

- determinação (507) dos coeficientes de filtro para o filtro estéreo em resposta a pelo menos uma função de transferência perceptual binaural.

11. Transmissor para transmitir um sinal de áudio binaural, o transmissor compreendendo:

- meios (401, 403) para receber dados de

áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N;

- meios de dados de parâmetro (411) para converter parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural;

- meios de conversão (409) para converter o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais;

caracterizado por compreender ainda um filtro estéreo (415, 417) para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo;

- meios de coeficiente (419) para determinar coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural; e

- meios para transmitir o sinal de áudio binaural.

12. Sistema de transmissão para transmitir um sinal de áudio, o sistema de transmissão que inclui um transmissor compreendendo:

- meios (401, 403) para receber dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de

áudio do canal N,

- meios de dados de parâmetro (411) para converter parâmetros espaciais dos dados de parâmetro espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural,

- meios de conversão (409) para converter o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais,

- caracterizado por compreender ainda um filtro estéreo (415, 417) para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo,

- meios de coeficientes (419) para determinar coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural, e

- meios para transmitir o sinal de áudio binaural; e

- um receptor para receber o sinal de áudio binaural.

13. Dispositivo de gravação de áudio para gravar um sinal de áudio binaural, o dispositivo de gravação de áudio que compreende:

- meios (401, 403) para receber dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N;

- meios de dados de parâmetro (411) para converter parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural;

- meios de conversão (409) para converter o sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais;

caracterizado por compreender um filtro estéreo (415, 417) para gerar o sinal de áudio binaural pela filtragem do primeiro sinal estéreo;

- meios de coeficiente (419) para determinar coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural; e

- meios para gravação do sinal de áudio binaural.

14. Método de transmissão de um sinal de áudio binaural, o método que compreende:

- a recepção dos dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N;

- conversão dos parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural;

- conversão do sinal de áudio do canal M no

primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais;

- geração do sinal de áudio binaural pela filtragem do primeiro sinal estéreo em um filtro estéreo;

- caracterizado pela determinação dos coeficientes de filtro para o filtro estéreo em resposta à função de transferência perceptual binaural; e

- transmissão do sinal de áudio binaural.

15. Método de transmissão e recepção de um sinal de áudio binaural, o método que compreende:

- um transmissor executando as etapas de:

- recepção de dados de áudio compreendendo um sinal de áudio do canal M sendo um downmix de um sinal de áudio do canal N e dados de parâmetros espaciais para upmix do sinal de áudio do canal M no sinal de áudio do canal N,

- conversão de parâmetros espaciais dos dados de parâmetros espaciais nos primeiros parâmetros binaurais em resposta a pelo menos uma função de transferência perceptual binaural,

- conversão do sinal de áudio do canal M em um primeiro sinal estéreo em resposta aos primeiros parâmetros binaurais,

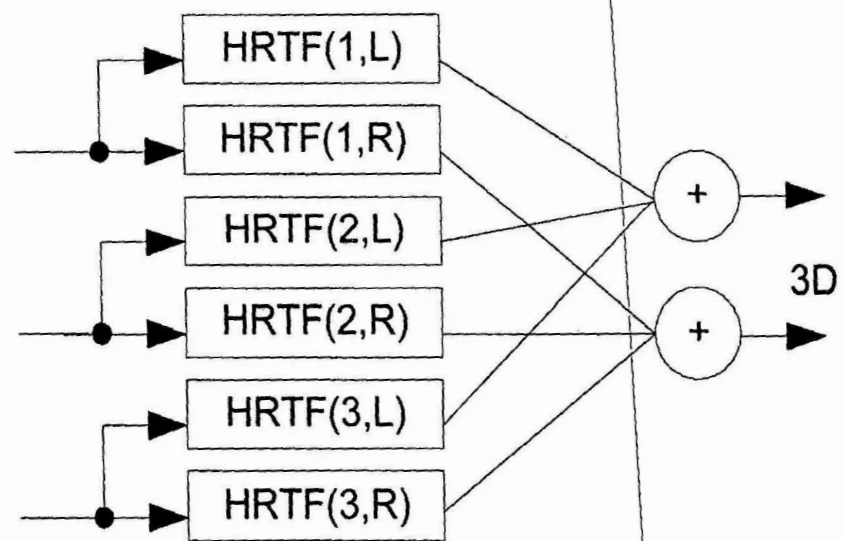
- caracterizado pela geração do sinal de áudio binaural pela filtragem do primeiro sinal estéreo em um filtro estéreo,

- determinação dos coeficientes de filtro

para o filtro estéreo em resposta à função de transferência perceptual binaural, e

- transmissão do sinal de áudio binaural; e

um receptor executando a etapa de recepção do sinal de áudio binaural.



Estado da Técnica

FIG. 1

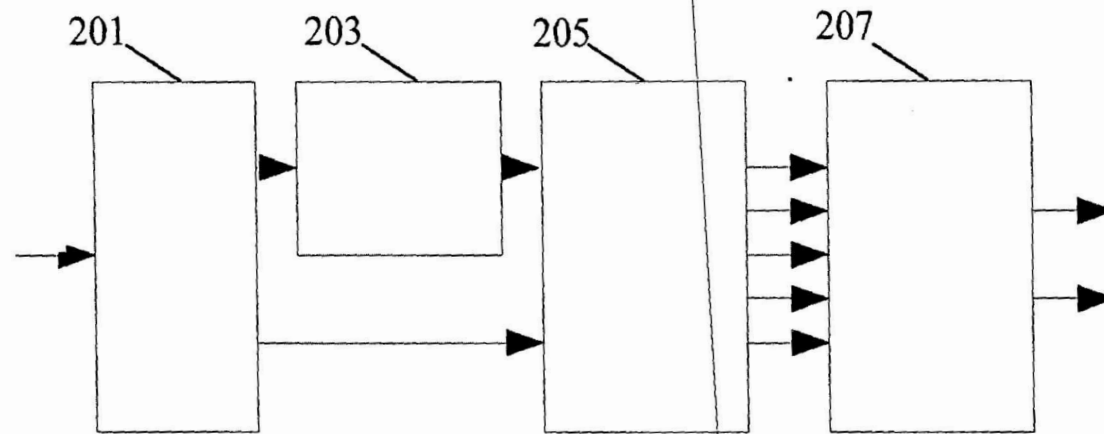


FIG. 2

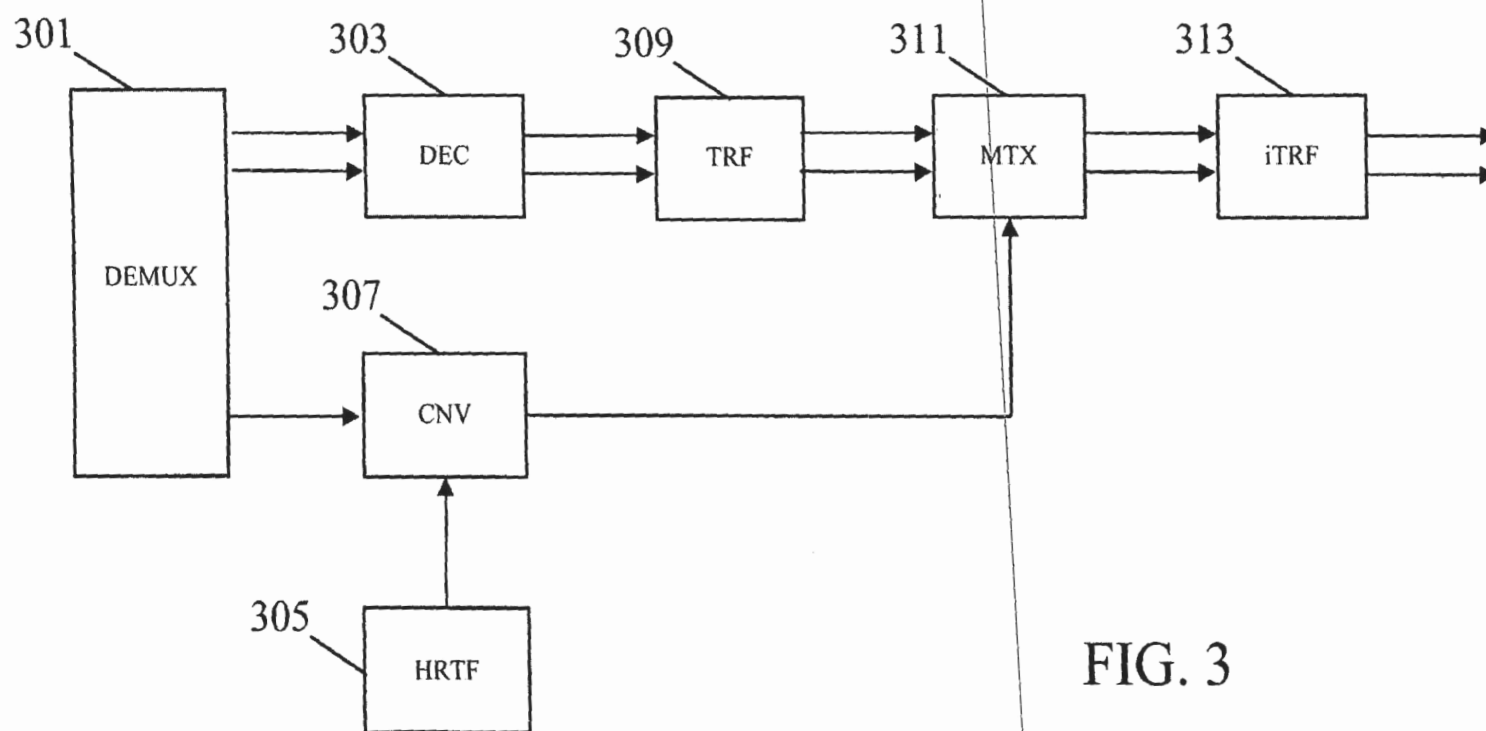


FIG. 3

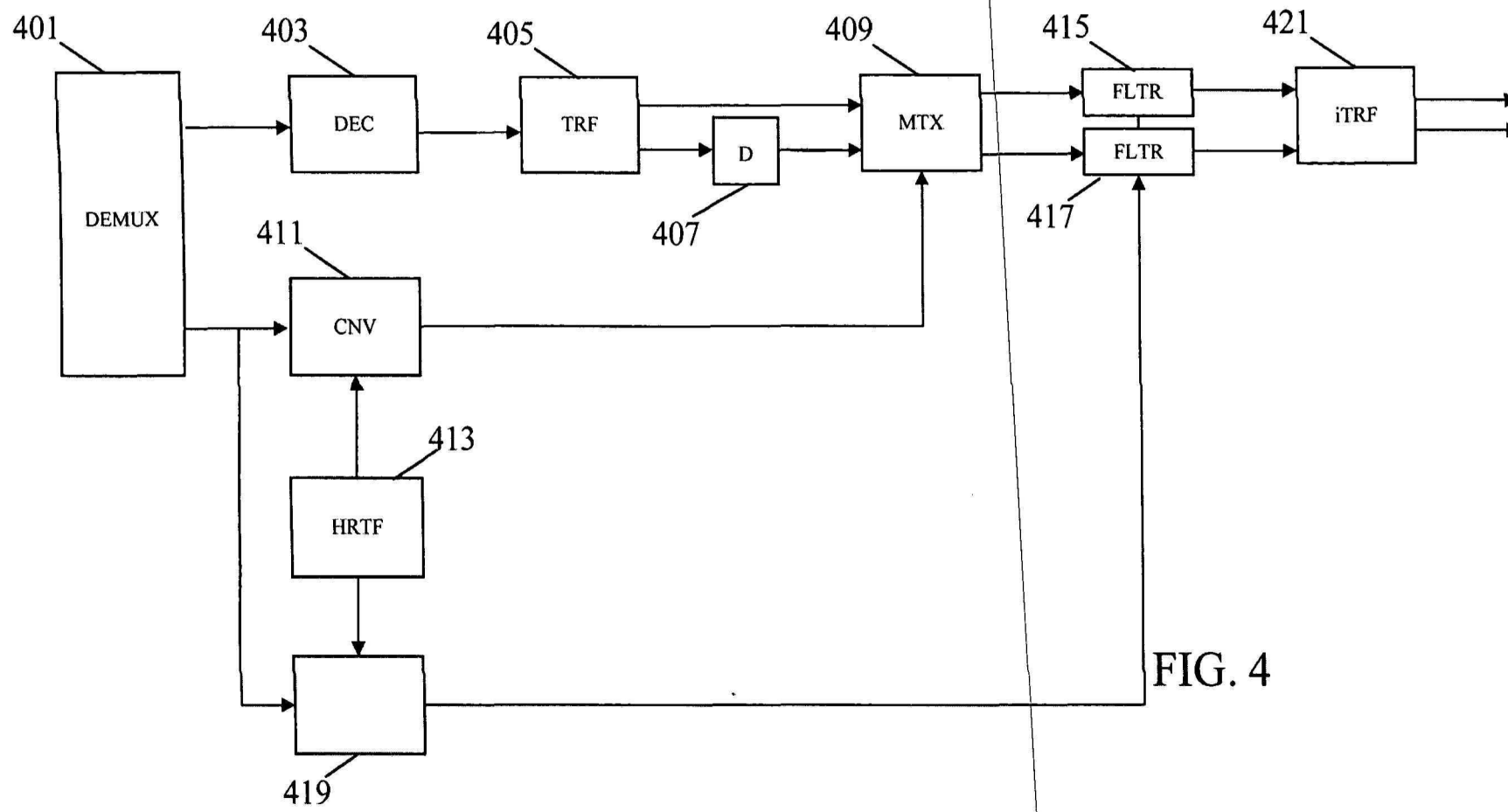
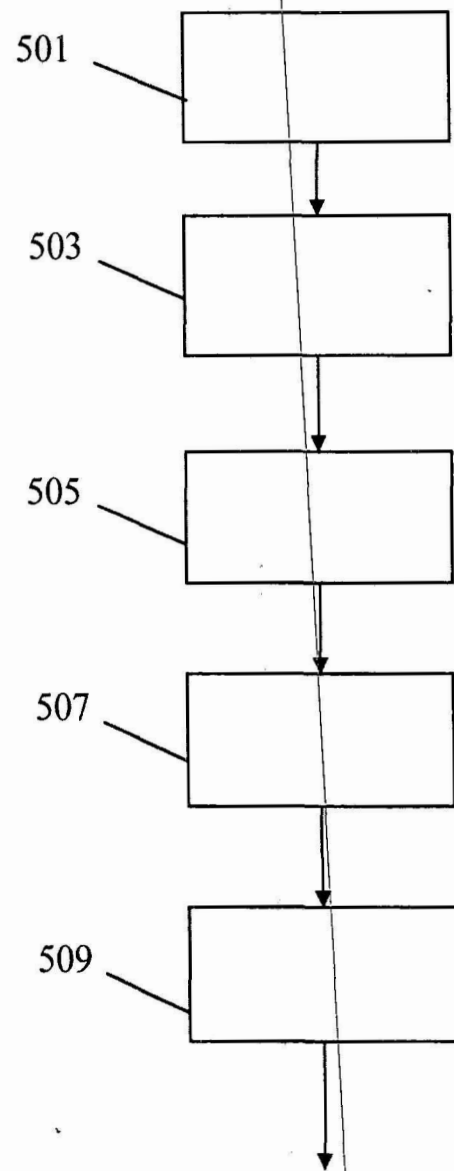


FIG. 4

FIG. 5



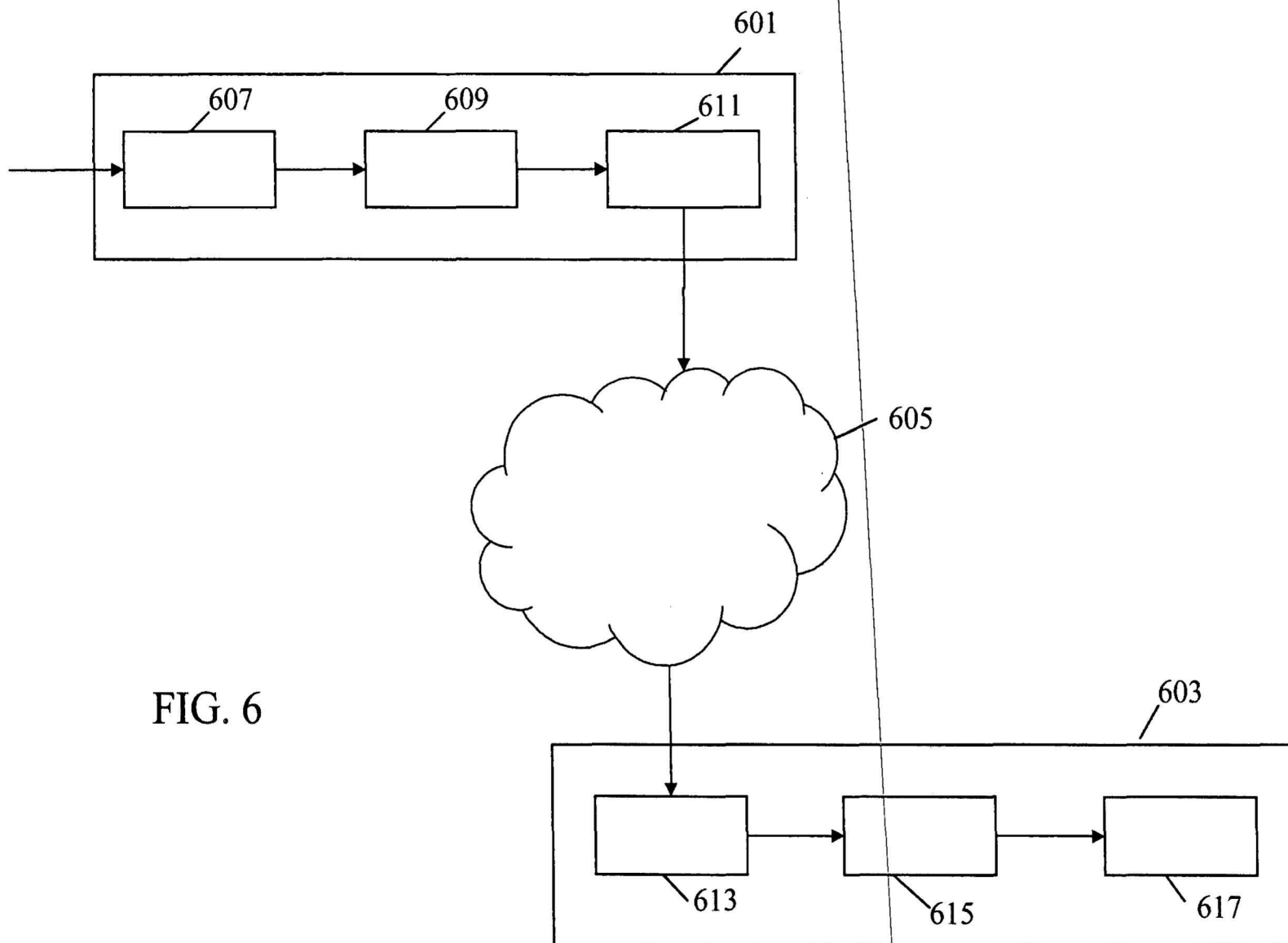


FIG. 6