



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

(11) BR 112015019958-5 B1



(22) Data do Depósito: 27/11/2014

(45) Data de Concessão: 14/12/2021

(54) Título: DESCODIFICADOR, CODIFICADOR E MÉTODO DE ESTIMATIVA INFORMADA DA PERCEPÇÃO SONORA EMPREGANDO SINAIS DE OBJETOS DE ÁUDIO DE BYPASS EM SISTEMAS DE CODIFICAÇÃO DE ÁUDIO BASEADA EM OBJETOS

(51) Int.Cl.: G10L 19/008.

(30) Prioridade Unionista: 27/11/2013 EP 13194664.2.

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

(72) Inventor(es): JOUNI PAULUS; SASCHA DISCH; HARALD FUCHS; BERNHARD GRILL; OLIVER HELLMUTH; ADRIAN MURTAZA; FALKO RIDDERBUSCH; LEON TERENTIV.

(86) Pedido PCT: PCT EP2014075801 de 27/11/2014

(87) Publicação PCT: WO 2015/078964 de 04/06/2015

(85) Data do Início da Fase Nacional: 19/08/2015

(57) Resumo: DESCODIFICADOR, CODIFICADOR E MÉTODO DE ESTIMATIVA INFORMADA DA PERCEPÇÃO SONORA EMPREGANDO SINAIS DE OBJETOS DE ÁUDIO DE BYPASS EM SISTEMAS DE CODIFICAÇÃO DE ÁUDIO BASEADA EM OBJETOS. É apresentado um decodificador para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio. O decodificador compreende um interface receptor (110) para receber um sinal de entrada de áudio, que compreende uma pluralidade de sinais de objetos de áudio, para receber informação de percepção sonora nos sinais de objetos de áudio e para receber informação de representação que indica se um ou mais sinais de objetos de áudio deverão ser amplificados ou atenuados. Além disso, o decodificador compreende um processador de sinal (120) para gerar um ou mais canais de saída de áudio do sinal de saída de áudio. O processador de sinal (120) está configurado para determinar um valor de compensação de percepção sonora dependendo da informação de percepção sonora e dependendo da informação de representação. Além disso, o processador de sinal (120) está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio, dependendo da informação de representação e dependendo do valor (...).

Relatório Descritivo da Patente de Invenção para: "**DESCODIFICADOR, CODIFICADOR E MÉTODO DE ESTIMATIVA INFORMADA DA PERCEPÇÃO SONORA EMPREGANDO SINAIS DE OBJETOS DE ÁUDIO DE BYPASS EM SISTEMAS DE CODIFICAÇÃO DE ÁUDIO BASEADA EM OBJETOS**".

[001] A presente invenção refere-se à codificação, processamento e decodificação de sinais de áudio e particularmente a um decodificador e codificador e método de estimativa informada da percepção sonora em sistemas de codificação de áudio baseada em objetos.

[002] Recentemente têm sido propostas técnicas paramétricas de transmissão de taxa de bit eficiente/armazenamento de cenas áudio compreendendo múltiplos sinais de objetos de áudio no domínio da codificação áudio [BCC, JSC, SAOC, SAOC1, SAOC2] e separação informada de fontes [ISS1, ISS2, ISS3, ISS4, ISS5, ISS6]. Estas técnicas procuram reconstruir uma cena áudio ou objeto fonte áudio com base em informação lateral adicional que descreve a cena áudio transmitida/armazenada e/ou os objetos fonte na cena áudio. Esta reconstrução decorre no decodificador, utilizando um esquema de separação informada da fonte. Os objetos reconstruídos podem ser combinados para produzir a cena áudio de saída. Dependendo da forma como os objetos são combinados, a sonoridade perceptual da cena de saída pode variar.

[003] Nas emissões de TV e rádio, os níveis de volume das faixas áudio dos vários programas podem ser normalizados com base em vários aspectos, tais como o nível de sinal de pico ou o nível de sonoridade. Dependendo das propriedades dinâmicas dos sinais, dois sinais com o mesmo nível de pico podem ter um nível completamente diferente de sonoridade perceptível. Nas trocas entre programas ou canais, as diferenças entre a percepção sonora do sinal são muito incômodas e têm vindo a ser uma fonte substancial de queixas da parte do utilizador final no âmbito das emissões.

[004] Na técnica anterior, foi proposto normalizar todos os programas em todos os canais de forma similar a um nível de referência comum, utilizando

uma escala baseada na sonoridade perceptual do sinal. Uma recomendação neste sentido, na Europa, é a recomendação R128 [EBU] (doravante designada R128).

[005] A recomendação diz que a "percepção sonora do programa", por ex. a percepção sonora média durante um programa (ou publicidade ou outra entidade significativa) deve corresponder a um nível especificado (com pequenos desvios admissíveis). Com cada vez mais emissoras a respeitar esta recomendação e a normalização necessária, as diferenças de percepção sonora média entre programas e canais deve ser minimizada.

[006] A estimativa da percepção sonora pode ser executada de várias formas. Existem vários modelos matemáticos para estimar a percepção sonora de um sinal áudio. A recomendação EBU R128 baseia-se no modelo apresentado em ITU-R BS.1770 (posteriormente designado BS.1770) (*vide* [ITU]) para a estimativa da sonoridade.

[007] Como indicado anteriormente, por ex. de acordo com a recomendação EBU R128, a percepção sonora de um programa, por ex. a percepção sonora média ao longo de um programa deve corresponde a um nível especificado com pequenos desvios admissíveis. No entanto, este aspecto levanta problemas significativos quando se executa a representação na técnica anterior e que até hoje ainda não tiveram solução. A execução de uma representação áudio do lado do decodificador tem um efeito significativo na percepção sonora global/total do sinal de entrada de áudio recebido. No entanto, apesar de ser executada a representação da cena, a percepção sonora total do sinal de áudio recebido permanece igual.

[008] Atualmente não existe uma solução específica do lado do decodificador para este problema.

[009] A EP 2 146 522 A1 ([EP]) refere-se a conceitos para a geração de sinais de saída de áudio utilizando meta-dados baseados em objetos. É gerado pelo menos um sinal de saída de áudio representando uma sobreposição de pelo menos dois sinais de objetos de áudio diferentes, mas não proporciona

uma solução para este problema.

[010] A WO 2008/035275 A2 ([BRE]) descreve um sistema áudio compreendendo um codificador que codifica objetos de áudio em uma unidade codificadora que gera um *downmix* do sinal de áudio e dados paramétricos que representa os vários objetos de áudio. O *downmix* de sinal de áudio e dados paramétricos são transmitidos para um decodificador que compreende uma unidade decodificadora, a qual gera réplicas aproximadas dos objetos de áudio e uma unidade de representação que gera um sinal de saída a partir dos objetos de áudio. O decodificador contém ainda um processador para a geração de dados de modificação codificadores, que é enviado para o codificador. O codificador depois modifica a codificação dos objetos de áudio e, em particular, modifica os dados paramétricos em resposta aos dados de modificação da codificação. A abordagem permite a manipulação dos objetos de áudio controlada pelo decodificador, mas executada na totalidade ou em parte pelo codificador. Assim, a manipulação pode ser executada nos objetos de áudio independentes reais em vez de em réplicas aproximadas, proporcionando assim um melhor desempenho.

[011] A EP 2 146 522 A1 ([SCH]) apresenta um aparelho para a geração de pelo menos um sinal de saída de áudio, representando uma sobreposição de pelo menos dois diferentes objetos de áudio, compreende um processador para o processamento de um sinal de entrada de áudio para proporcionar uma representação do objeto do sinal de entrada de áudio, podendo esta representação do objeto ser gerada por uma aproximação com orientação paramétrica dos objetos originais, utilizando um sinal *downmix* do objeto. Um manipulador do objeto manipula individualmente objetos, utilizando meta-dados baseados em objetos de áudio, remetendo para os objetos de áudio individuais para obter objetos de áudio manipulados. Os objetos de áudio manipulados são misturados utilizando um misturador de objetos para se obter finalmente um sinal de saída de áudio com um ou vários sinais do canal, dependendo de uma configuração de representação específica.

[012] A WO 2008/046531 A1 ([ENG]) descreve um codificador de objetos de áudio para a geração de um objeto codificado utilizando vários objetos de áudio, incluindo um gerador de informação *downmix* para a geração de informação *downmix*, indicando uma distribuição de vários objetos de áudio em pelo menos dois canais *downmix*, um gerador de parâmetros do objeto de áudio para a geração de parâmetros de objeto para os objetos de áudio e uma interface de saída para a geração do sinal de saída de áudio importado utilizando a informação *downmix* e os parâmetros de objeto. Um sintetizador de áudio utiliza a informação *downmix* para a geração de dados de saída utilizáveis para a criação de vários canais de saída da configuração de saída de áudio predefinida.

[013] Seria desejável possuir uma estimativa precisa da percepção sonora média de saída ou da alteração da percepção sonora média sem demora e, quando o programa não altera ou a cena de representação não é alterada, a percepção sonora média também deve permanecer estática.

[014] O objeto da presente invenção consiste em proporcionar conceitos melhorados de codificação, processamento e descodificação do sinal de áudio. O objeto da presente invenção é solucionado com um descodificador de acordo com a reivindicação 1, por um codificador de acordo com a reivindicação 9, com um sistema de acordo com a reivindicação 11, com um método de acordo com a reivindicação 12, com um método de acordo com a reivindicação 13 e com um programa de computador de acordo com a reivindicação 15.

[015] É apresentada uma forma informada de estimativa da percepção sonora da saída em um sistema de codificação áudio baseada em objetos. Os conceitos fornecidos baseiam-se na informação da percepção sonora dos objetos na mistura áudio a ser fornecida ao descodificador. O descodificador utiliza esta informação conjuntamente com a informação de representação para estimativa da percepção sonora do sinal de saída. Assim é possível, por exemplo, estimar a diferença de percepção sonora entre o *downmix* predefinido e a saída representada. É então possível compensar esta diferença de forma a

obter aproximadamente uma percepção sonora constante à saída, independentemente da informação de representação. A estimativa da percepção sonora no decodificador decorre de forma totalmente paramétrica e, em termos informáticos, é muito leve e precisa em comparação com os conceitos de estimativa da percepção sonora baseados em sinais.

[016] São apresentados conceitos de obtenção de informação sobre a percepção sonora da cena de saída específica utilizando conceitos puramente paramétricos, o que permite o processamento da percepção sonora sem estimativa explícita da percepção sonora baseada em sinais no decodificador. Além disso, é descrita a tecnologia específica da Codificação de Objeto de Áudio Espacial (em inglês, Spatial Audio Object Coding - SAOC) normalizada por MPEG [SAOC], mas os conceitos apresentados podem ser utilizados também em conjunto com outras tecnologias de codificação de objetos áudio.

[017] É apresentado um decodificador para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio. O decodificador compreende um interface receptor para receber um sinal de entrada de áudio, que compreende uma pluralidade de sinais de objetos de áudio, para receber informação de percepção sonora nos sinais de objetos de áudio e para receber informação de representação indicadora se um ou mais sinais de objetos de áudio deverão ser amplificados ou atenuados. Além disso, o decodificador compreende um processador de sinal para gerar um ou mais canais de saída de áudio do sinal de saída de áudio. O processador de sinal está configurado para determinar um valor de compensação de percepção sonora dependendo da informação de percepção sonora e dependendo da informação de representação. Além disso, o processador de sinal está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio, dependendo da informação de representação e dependendo do valor de compensação de percepção sonora.

[018] De acordo com uma realização, o processador de sinal pode ser configurado para gerar um ou mais canais de saída de áudio do sinal de saída

de áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora, de tal modo que uma percepção sonora do sinal de saída de áudio é igual a uma percepção sonora do sinal de entrada de áudio ou de tal modo que a percepção sonora do sinal de saída é mais próxima da percepção sonora do sinal de entrada de áudio comparativamente a uma percepção sonora de um sinal de áudio modificado que resultaria da modificação de um sinal de entrada de áudio ao amplificar ou atenuar os sinais de objetos de áudio de acordo com a informação de representação.

[019] De acordo com outra realização, cada sinal de objetos de áudio do sinal de entrada de áudio pode ser atribuído a exatamente um grupo de dois ou mais grupos, em que cada dos dois ou mais grupos pode compreender um ou mais sinais de objetos de áudio do sinal de entrada de áudio, Nesta realização, a interface receptora pode ser configurada para receber um valor de percepção sonora para cada grupo dos dois ou mais grupos da informação de percepção sonora, indicando o referido valor de percepção sonora uma percepção sonora total original de um ou mais sinais de objetos de áudio do referido grupo. Além disso, a interface receptora pode ser configurada para receber a informação de representação indicando quanto a pelo menos um grupo dos dois ou mais grupos se o um ou mais sinais de objetos de áudio do referido grupo serão amplificados ou atenuados através da indicação de uma percepção sonora total modificada do um ou mais sinais de objetos de áudio do referido grupo. Além disso, em uma realização assim, o processador de sinal pode ser configurado para determinar o valor de compensação de percepção sonora dependendo da percepção sonora total modificada de cada um do referido pelo menos um grupo dos dois ou mais grupos e dependendo da percepção sonora total original de cada um dos dois ou mais grupos. Além disso, o processador de sinal pode ser configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da percepção sonora total modificada de cada um do referido pelo menos um

grupo dos dois ou mais grupos e dependendo do valor de compensação da percepção sonora.

[020] Em realizações particulares, pelo menos um grupo dos dois ou mais grupos pode compreender dois ou mais dos sinais de objetos de áudio.

[021] Além disso é apresentado um codificador. O codificador compreende uma unidade de codificação baseada em objetos para codificar uma pluralidade de sinais de objetos de áudio para obter um sinal de áudio codificado que compreende a pluralidade de sinais de objetos de áudio. Além disso, o codificador compreende uma unidade de codificação da percepção sonora do objeto para codificar a informação da percepção sonora nos sinais de objetos de áudio. A informação da percepção sonora compreende um ou mais valores de percepção sonora, em que cada um ou de mais valores de percepção sonora dependem de um ou mais dos sinais de objetos de áudio.

[022] De acordo com uma realização, cada um dos sinais de objetos de áudio do sinal de áudio codificado pode ser atribuído a exatamente um grupo de dois ou mais grupos, em que cada dos dois ou mais grupos compreende um ou mais dos sinais de objetos de áudio do sinal de áudio codificado. A unidade codificadora da percepção sonora do objeto pode ser configurada para determinar o um ou mais valores de percepção sonora da informação de percepção sonora através da determinação de um valor de percepção sonora para cada grupo dos dois ou mais grupos, em que o referido valor de percepção sonora do referido grupo indica uma percepção sonora total original do um ou mais sinais de objetos de áudio do referido grupo.

[023] Além disso é apresentado um sistema. O sistema compreende um codificador de acordo com uma das realizações já descritas para a codificação de uma pluralidade de sinais de objetos de áudio para obter um sinal áudio codificado compreendendo a pluralidade de sinais de objetos de áudio e para codificação da informação de percepção sonora nos sinais de objetos de áudio. Além disso, o sistema compreende um decodificador de acordo com uma das realizações já descritas para gerar um sinal de saída de áudio que compreenda

um ou mais canais de saída de áudio. O decodificador está configurado para receber o sinal de áudio codificado como um sinal de entrada de áudio e a informação da percepção sonora. Além disso, o decodificador está configurado para receber ainda a informação de representação. O decodificador está ainda configurado para determinar um valor de compensação de percepção sonora dependendo da informação de percepção sonora e dependendo da informação de representação. Além disso, o decodificador está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

[024] Além disso, é apresentado método para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio. O método inclui:

[025] - Receber um sinal de entrada de áudio compreendendo uma pluralidade de sinais de objetos de áudio.

[026] - Receber informação da percepção sonora sobre os sinais de objetos de áudio.

[027] - Receber informação de representação indicando se um ou mais dos sinais de objetos de áudios deverão ser amplificados ou atenuados.

[028] - Determinar um valor de compensação de percepção sonora dependendo da informação de percepção sonora e dependendo da informação de representação e:

[029] - Gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio, dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

[030] Além disso, é apresentado um método de decodificação. O método inclui:

[031] - Codificar um sinal de entrada de áudio compreendendo uma pluralidade de sinais de objetos de áudio. e:

[032] - Codificar informação da percepção sonora sobre os sinais de

objetos de áudio, em que a informação de ruído compreende um ou mais valores da percepção sonora, em que cada um ou de mais valores da percepção sonora dependem em um ou mais dos sinais de objetos de áudio.

[033] Além disso, é apresentado um programa de computador para implementar o método descrito acima quando executado em um computador ou processador de sinal.

[034] As realizações preferidas são apresentadas nas reivindicações dependentes.

[035] Seguidamente passa-se à descrição mais detalhada das realizações da presente invenção com referência às figuras, nas quais:

[036] Fig. 1 ilustra um decodificador para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio de acordo com uma realização,

[037] Fig. 2 ilustra um codificador de acordo com uma realização,

[038] Fig. 3 ilustra um sistema de acordo com uma realização,

[039] Fig. 4 ilustra um sistema de Codificação de Objeto Áudio Espacial compreendendo um codificador SAOC e um decodificador SAOC,

[040] Fig. 5 ilustra um decodificador SAOC compreendendo um decodificador de informação lateral, um separador de objetos e um renderizador,

[041] Fig. 6 ilustra um comportamento das estimativas da percepção sonora do sinal de saída em uma alteração da percepção sonora,

[042] Fig. 7 descreve a estimativa informada da percepção sonora de acordo com uma realização, ilustrando os componentes de um codificador e de um decodificador de acordo com uma realização,

[043] Fig. 8 ilustra um codificador de acordo com outra realização,

[044] Fig. 9 ilustra um codificador e um decodificador de acordo com uma realização relacionada com o Melhoramento de Diálogo SAOC, que compreende canais de bypass,

[045] Fig. 10 descreve uma primeira ilustração de uma alteração da

percepção sonora medida e o resultado da utilização dos conceitos fornecidos para estimar a alteração na percepção sonora de forma paramétrica,

[046] Fig. 11 descreve uma segunda ilustração de uma alteração da percepção sonora medida e o resultado da utilização dos conceitos fornecidos para estimar a alteração na percepção sonora de forma paramétrica e

[047] Fig. 12 ilustra outra realização para conduzir a compensação da percepção sonora.

[048] Antes de serem descritas as realizações preferidas de forma detalhada, são descritos a estimativa da percepção sonora, Codificação de Objeto Áudio Espacial (abreviatura em inglês SAOC) e Melhoramento do Diálogo (abreviatura do inglês DE).

[049] Descreve-se primeiro a estimativa da percepção sonora.

[050] Como já foi referido, a recomendação R128 EBU baseia-se no modelo apresentado em ITU-R BS.1770 para a estimativa da percepção sonora. Esta escala será empregue como um exemplo, mas os conceitos descritos abaixo podem também ser aplicados para outras medições da percepção sonora.

[051] A operação da estimativa da percepção sonora de acordo com a norma BS.1770 é relativamente simples e baseia-se nos seguintes passos principais [ITU]:

[052] - o sinal de entrada x_i (ou sinais no caso do sinal multicanal) é filtrado com um filtro K (uma combinação de um filtro *shelving* e um filtro passa alta) para se obter o(s) sinal/sinais y_i .

[053] - é calculada a energia média quadrática z_i do sinal y_i .

[054] - no caso do sinal multicanal, é aplicada a ponderação do canal G_i e são somados os sinais ponderados. A percepção sonora do sinal é então definida como sendo

$$L = c + 10 \log_{10} \sum_i G_i z_i,$$

[055] com o valor constante $c = -0.691$. A saída é então expressa nas unidades de "LKFS" (em inglês, Loudness, K-weighted, relative to Full Scale, isto é Percepção sonora, Ponderação-K, relativa à escala total) que é similar à escala de decibéis.

[056] Na fórmula acima, G_i pode ser, por exemplo, igual a 1 para alguns dos canais, enquanto G_i pode, por exemplo, ser 1,41 para outros canais. Por exemplo, se forem considerados um canal esquerdo, um canal direito, um canal central, um canal *surround* esquerdo e um canal *surround* direito, as respectivas ponderações G_i podem ser, por exemplo, 1 para o canal esquerdo, direito e central e pode ser, por exemplo, 1,41 para o canal *surround* esquerdo e para o canal *surround* direito, ver [ITU].

[057] Como se pode ver, o valor da percepção sonora L está intimamente relacionado com o logaritmo da energia do sinal.

[058] Seguidamente descreve-se a Codificação de Objeto Áudio Espacial.

[059] Os conceitos de codificação de áudio baseada em objetos permitem uma considerável flexibilidade do lado do decodificador da cadeia. Um exemplo de um conceito de codificação de áudio baseada em objetos é a Codificação de Objeto Áudio Espacial (SAOC).

[060] A fig. 4 ilustra um sistema de Codificação de Objeto Áudio Espacial (SAOC) compreendendo um codificador SAOC 410 e um decodificador SAOC 420.

[061] O codificador SAOC 410 recebe N sinais de objetos de áudio S_1, \dots, S_N como entrada. Além disso, o codificador SAOC 410 recebe ainda instruções "informação de mistura \mathbf{D} " sobre como estes objetos devem ser combinados para se obter um sinal *downmix* compreendendo os canais *downmix* M X_1, \dots, X_M . O codificador SAOC 410 extrai alguma informação lateral dos objetos e do processo de *downmix* e esta informação lateral é transmitida e/ou guardada juntamente com os sinais de *downmix*.

[062] Uma propriedade substancial do sistema SAOC consiste no fato de o sinal *downmix* X que compreende os canais *downmix* X_1, \dots, X_M formar um sinal significativo em termos semânticos. Por outras palavras, é possível ouvir o sinal *downmix*. Se, por exemplo, o receptor não possuir a funcionalidade do decodificador SAOC, o receptor pode, ainda assim, fornecer o sinal *downmix* como saída.

[063] A fig. 5 ilustra um decodificador SAOC compreendendo um decodificador de informação lateral 510, um separador de objetos 520 e um renderizador 530. O decodificador SAOC ilustrado pela fig. 5 recebe, por exemplo a partir de um codificador SAOC, o sinal *downmix* e a informação lateral. O sinal *downmix* pode ser considerado como um sinal de entrada de áudio que compreende os sinais de objetos de áudio, já que os sinais de objetos de áudio são misturados com o sinal *downmix* (os sinais de objetos de áudio são misturados dentro de um ou mais canais *downmix* do sinal *downmix*).

[064] O decodificador SAOC pode, por ex. tentar então reconstruir (virtualmente) os objetos originais, por ex. empregando o separador de objetos 520, por ex, utilizando a informação lateral decodificada. Estas reconstruções (virtuais) de objetos $\hat{S}_1, \dots, \hat{S}_N$, por ex. os sinais de objetos de áudio reconstruídos, são então combinadas com base na informação de representação, por ex. uma matriz de representação R para produzir canais de saída de áudio $\mathbf{K} \mathbf{Y}_1, \dots, \mathbf{Y}_K$ de um sinal de saída de áudio Y .

[065] Em SAOC, é frequente os sinais de objetos de áudio serem, por exemplo, reconstruídos, por ex. através do emprego de informação de covariância, por ex. uma matriz de covariância do sinal \mathbf{E} , que é transmitida do codificador SAOC para o decodificador SAOC.

[066] Por exemplo, a fórmula seguinte pode ser empregue para reconstruir os sinais de objetos de áudio no lado do decodificador:

[067] $\hat{\mathbf{S}} = \mathbf{G} \mathbf{X}$ com $\mathbf{G} \approx \mathbf{E} \mathbf{D}^H (\mathbf{D} \mathbf{E} \mathbf{D}^H)^{-1}$

[068] em que

- [069] N número de sinais de objetos de áudio,
- [070] $N_{samples}$ número de amostras consideradas de um sinais de objetos de áudio
- [071] M número de canais *downmix*,
- [072] \mathbf{X} sinal de áudio *downmix*, tamanho $M \times N_{samples}$,
- [073] \mathbf{D} matriz de *downmix*, tamanho $M \times N$
- [074] \mathbf{E} matriz de covariância do sinal, tamanho $N \times N$ definido como $\mathbf{E} = \mathbf{X} \mathbf{X}^H$
- [075] $\hat{\mathbf{S}}$ sinais de objetos de áudio N com reconstrução paramétrica, tamanho $N \times N_{samples}$
- [076] $(\cdot)^H$ operador (hermitiano) auto-adjunto que representa a transposição do conjugado de (\cdot)
- [077] Depois, a matriz de representação \mathbf{R} pode ser aplicada nos sinais de objetos de áudio reconstruídos $\hat{\mathbf{S}}$ para obter os canais de saída de áudio do sinal de saída de áudio \mathbf{Y} , por ex. de acordo com a fórmula:
- [078] $\mathbf{Y} = \mathbf{R} \hat{\mathbf{S}}$
- [079] em que
- [080] K número dos canais de saída de áudio Y_1, \dots, Y_K do sinal de saída de áudio \mathbf{Y} .
- [081] \mathbf{R} matriz de representação de tamanho $K \times N$
- [082] \mathbf{Y} sinal de saída de áudio compreendendo os canais de saída de áudio K ,
- [083] tamanho $K \times N_{samples}$
- [084] Na fig. 5, o processo de reconstrução do objeto, por ex. conduzido pelo separador de objetos 520, é referido com a noção "virtual" ou "opcional", dado que não tem necessariamente de ocorrer, mas a funcionalidade desejada pode ser obtida mediante combinação da reconstrução e as fases de representação no domínio paramétrico (isto é, combinando as equações).
- [085] Por outras palavras, em vez de reconstruir os sinais de objetos de

áudio utilizando a informação de mistura \mathbf{D} e a informação de covariância \mathbf{E} primeiro e depois aplicando a informação de representação \mathbf{R} nos sinais de objetos de áudio reconstruídos para obter os canais de saída de áudio Y_1, \dots, Y_K , ambos os passos podem ser conduzidos em um único passo, de forma que os canais de saída de áudio Y_1, \dots, Y_K são gerados diretamente a partir dos canais *downmix*.

[086] Por exemplo, pode ser empregue a seguinte fórmula:

[087] $\mathbf{Y} = \mathbf{R}\mathbf{G}\mathbf{X}$ com $\mathbf{G} \approx \mathbf{E}\mathbf{D}^H (\mathbf{D}\mathbf{E}\mathbf{D}^H)^{-1}$.

[088] Em princípio, a informação de representação \mathbf{R} pode solicitar qualquer combinação dos sinais de objetos de áudio originais. Na prática, contudo, as reconstruções de objetos podem compreender erros de reconstrução e a cena de saída solicitada não tem necessariamente de ser alcançada. Segundo uma regra geral que abrange muitos casos práticos, quanto mais à cena de saída solicitada difere do sinal de *downmix*, mais serão os erros de reconstrução audíveis.

[089] Segue-se uma descrição do melhoramento do diálogo (abreviatura do inglês DE). A tecnologia SAOC pode, por exemplo, ser empregue para concretizar o cenário. Deve-se ter em atenção que muito embora o nome "Melhoramento de diálogo" sugira uma concentração em sinais orientados para diálogo, o mesmo princípio é aplicado também com outros tipos de sinais.

[090] No cenário DE, os graus de liberdade no sistema são limitados a partir do caso geral.

[091] Por exemplo, os sinais de objetos de áudio $S_1, \dots, S_N = \mathbf{S}$ são agrupados (e possivelmente misturados) em dois meta-objetos de um objeto em primeiro plano (abreviatura do inglês, FGO) \mathbf{S}_{FGO} e um objeto em fundo (abreviatura do inglês, BGO) \mathbf{S}_{BGO} .

[092] Além disso, a cena de saída $Y_1, \dots, Y_K = \mathbf{Y}$ assemelha-se ao sinal *downmix* $X_1, \dots, X_M = \mathbf{X}$. Mais especificamente, ambos os sinais têm as mesmas dimensionalidades, isto é $K = M$ e o utilizador final só pode controlar

os níveis de mistura relativos dos dois meta-objetos FGO e BGO. Para ser mais exato, o sinal *downmix* é obtido através da mistura de FGO e BGO com os mesmos pesos escalares

$$\mathbf{X} = h_{FGO}\mathbf{S}_{FGO} + h_{BGO}\mathbf{S}_{BGO},$$

[093] e a cena de saída é obtida de forma similar com alguma ponderação escalar do FGO e BGO:

$$\mathbf{Y} = g_{FGO}\mathbf{S}_{FGO} + g_{BGO}\mathbf{S}_{BGO}.$$

[094] Dependendo dos valores relativos dos pesos de mistura, o equilíbrio entre FGO e BGO pode variar. Por exemplo, com a configuração

$$\begin{cases} g_{FGO} > h_{FGO} \\ g_{BGO} = h_{BGO} \end{cases}$$

[095] é possível aumentar o nível relativo do FGO na mistura. Se o FGO for o diálogo, esta configuração proporciona uma funcionalidade de melhoramento do diálogo.

[096] Como exemplo de caso de uso, o BGO pode ser os ruídos de um estádio e outros sons de fundo durante um evento desportivo e o FGO é a voz do comentador. A funcionalidade DE permite ao utilizador final amplificar ou atenuar o nível do comentador em relação ao fundo.

[097] As realizações são baseadas na descoberta de que a utilização da tecnologia SAOC (ou similar) em um cenário de emissão permite proporcionar ao utilizador final funcionalidades de manipulação do sinal ampliadas. São fornecidas mais funcionalidades para além da alteração do canal e ajuste do volume de reprodução.

[098] Acima apresenta-se uma descrição breve de uma possibilidade de emprego da tecnologia DE. Se o sinal de emissão, sendo o sinal *downmix* para SAOC, é normalizado em termos de nível, por ex. de acordo com a norma R128, os diferentes programas possuem uma percepção sonora média similar quando não é aplicado o processamento (SAOC-) (ou a descrição de representação for igual à descrição de *downmix*). No entanto, quando é

aplicado algum processamento (SAOC-), o sinal de saída difere do sinal *downmix* predefinido e a percepção sonora do sinal de saída pode ser diferente da percepção sonora do sinal *downmix* predefinido. Do ponto de vista do utilizador final, esta situação pode levar a que a percepção sonora do sinal de saída entre canais ou programas possa apresentar novamente os saltos ou diferenças indesejados. Por outras palavras, as vantagens da normalização aplicadas pela emissora perdem-se em parte.

[099] Este problema não é específico apenas do cenário SAOC ou DE, mas pode ocorrer também com outros conceitos de codificação de áudio que permitem ao utilizador final interagir com o conteúdo. No entanto, em muitos casos, não causa problema algum se o sinal de saída possuir uma percepção sonora diferente do que o *downmix* predefinido.

[100] Como já foi dito, uma percepção sonora total de um programa de sinal de entrada de áudio deve corresponder a um nível especificado com pequenos desvios admissíveis. No entanto, como já foi enfatizado, esta situação levanta problemas significativos quando a representação de áudio é conduzida, dado que a representação tem um efeito significativo na percepção sonora global/total do sinal de entrada de áudio recebido. No entanto, apesar de ser executada a representação da cena, a percepção sonora total do sinal de áudio recebido permanece igual.

[101] Uma abordagem consistiria em estimar a percepção sonora de um sinal enquanto este está a ser reproduzido e com um conceito de integração temporal apropriado, podendo a estimativa convergir para a verdadeira percepção sonora média ao fim de algum tempo. O tempo necessário para a convergência, porém, é problemático do ponto de vista do utilizador final. Quando a estimativa da percepção sonora se altera, mesmo sem que sejam aplicadas alterações ao sinal, a compensação de alteração da percepção sonora deve também reagir e alterar o seu comportamento. Este fato iria conduzir a um sinal de saída com percepção sonora média variável no tempo, o que pode dar uma sensação bastante incômoda.

[102] A fig. 6 ilustra um comportamento das estimativas da percepção sonora do sinal de saída em uma alteração da percepção sonora. Entre outros aspectos, é descrita uma estimativa da percepção sonora do sinal de saída com base no sinal, que ilustra o efeito de uma solução como a descrita. A estimativa aborda a estimativa correta com certa lentidão. Em vez de uma estimativa da percepção sonora do sinal de saída baseada no sinal, seria preferível uma estimativa informada da percepção sonora do sinal de saída que determina de imediato corretamente a percepção sonora do sinal de saída.

[103] Em particular, na fig. 6 o utilizador introduz, por ex. o nível do objeto de diálogo, altera um instante T aumentando o valor. O verdadeiro nível do sinal de saída, e de forma correspondente a percepção sonora, altera no mesmo instante. Quando a estimativa da percepção sonora do sinal de saída é executada a partir do sinal de saída com algum tempo de integração temporal, a estimativa irá alterar-se gradualmente e atingir o valor correto ao fim de certo atraso. Durante este atraso, os valores de estimativa estão em alteração e não podem ser empregues com fiabilidade para continuar o processamento do sinal de saída, por ex. para a correção do nível de percepção sonora.

[104] Como já foi referido, seria desejável possuir uma estimativa precisa da percepção sonora média de saída ou da alteração da percepção sonora média sem demora e, quando o programa não altera ou a cena de representação não é alterada, a percepção sonora média também deve permanecer estática. Por outras palavras, quando é aplicada alguma compensação da alteração da percepção sonora, o parâmetro de compensação só deve alterar-se quando ou o programa se altera ou existe alguma interação com o utilizador.

[105] O comportamento desejado encontra-se ilustrado na ilustração inferior da fig. 6 (estimativa informada da percepção sonora do sinal de saída). A estimativa da percepção sonora do sinal de saída altera-se imediatamente quando é alterada a entrada do utilizador.

[106] A fig. 2 ilustra um codificador de acordo com uma realização.

[107] O codificador compreende uma unidade de codificação 210 baseada em objetos para codificar uma pluralidade de sinais de objetos de áudio para obter um sinal de áudio codificado que compreende a pluralidade de sinais de objetos de áudio.

[108] Além disso, o codificador compreende uma unidade de codificação da percepção sonora do objeto 220 para codificar a informação da percepção sonora nos sinais de objetos de áudio. A informação da percepção sonora compreende um ou mais valores de percepção sonora, em que cada do ou de mais valores de percepção sonora dependem de um ou mais dos sinais de objetos de áudio.

[109] De acordo com uma realização, cada um dos sinais de objetos de áudio do sinal de áudio codificado é atribuído a exatamente um grupo de dois ou mais grupos, em que cada dos dois ou mais grupos compreende um ou mais dos sinais de objetos de áudio do sinal de áudio codificado. A unidade codificadora da percepção sonora do objeto 220 está configurada para determinar o um ou mais valores de percepção sonora da informação de percepção sonora através da determinação de um valor de percepção sonora para cada grupo dos dois ou mais grupos, em que o referido valor de percepção sonora do referido grupo indica uma percepção sonora total original do um ou mais sinais de objetos de áudio do referido grupo.

[110] A fig. 1 ilustra um descodificador para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio de acordo com uma realização.

[111] O descodificador compreende um interface receptor 110 para receber um sinal de entrada de áudio, que compreende uma pluralidade de sinais de objetos de áudio, para receber informação de percepção sonora nos sinais de objetos de áudio e para receber informação de representação indicadora se um ou mais sinais de objetos de áudio deverão ser amplificados ou atenuados.

[112] Além disso, o descodificador compreende um processador de sinal

120 para gerar um ou mais canais de saída de áudio do sinal de saída de áudio. O processador de sinal 120 está configurado para determinar um valor de compensação de percepção sonora dependendo da informação de percepção sonora e dependendo da informação de representação. Além disso, o processador de sinal 120 está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio, dependendo da informação de representação e dependendo do valor de compensação de percepção sonora.

[113] De acordo com uma realização, o processador de sinal 110 está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora, de tal modo que uma percepção sonora do sinal de saída de áudio é igual a uma percepção sonora do sinal de entrada de áudio ou de tal modo que a percepção sonora do sinal de saída é mais próxima da percepção sonora do sinal de entrada de áudio comparativamente a uma percepção sonora de um sinal de áudio modificado que resultaria da modificação de um sinal de entrada de áudio ao amplificar ou atenuar os sinais de objetos de áudio de acordo com a informação de representação.

[114] De acordo com outra realização, cada sinal de objetos de áudio do sinal de entrada de áudio é atribuído a exatamente um grupo de dois ou mais grupos, em que cada dos dois ou mais grupos pode compreender um ou mais sinais de objetos de áudio do sinal de entrada de áudio,

[115] Nesta realização, a interface receptora 110 é configurada para receber um valor de percepção sonora para cada grupo dos dois ou mais grupos da informação de percepção sonora, indicando o referido valor de percepção sonora uma percepção sonora total original de um ou mais sinais de objetos de áudio do referido grupo. Além disso, a interface receptora 110 é configurada para receber a informação de representação indicando quanto a pelo menos um grupo dos dois ou mais grupos se o um ou mais sinais de

objetos de áudio do referido grupo serão amplificados ou atenuados através da indicação de uma percepção sonora total modificada de um ou mais sinais de objetos de áudio do referido grupo. Além disso, em uma realização assim, o processador de sinal 120 é configurado para determinar o valor de compensação de percepção sonora dependendo da percepção sonora total modificada de cada um do referido pelo menos um grupo dos dois ou mais grupos e dependendo da percepção sonora total original de cada um dos dois ou mais grupos. Além disso, o processador de sinal 120 é configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da percepção sonora total modificada de cada um do referido pelo menos um grupo dos dois ou mais grupos e dependendo do valor de compensação da percepção sonora.

[116] Em realizações particulares, pelo menos um grupo dos dois ou mais grupos compreende dois ou mais dos sinais de objetos áudio.

[117] Existe uma relação direta entre a energia e_i de um sinal de objeto de áudio i e a percepção sonora L_i do sinal de objeto de áudio i de acordo com as fórmulas:

$$L_i = c + 10 \log_{10} e_i, \quad e_i = 10^{(L_i - c)/10}$$

[118] em que c é um valor constante.

[119] As realizações baseiam-se nas seguintes descobertas: Sinais de objetos de áudio diferentes do sinal de entrada de áudio podem ter uma percepção sonora diferente e, portanto, uma energia diferente. Se, por ex., um utilizador desejar aumentar a percepção sonora de um dos sinais de objetos de áudio, a informação de representação pode ser ajustada de forma correspondente e o aumento da percepção sonora deste sinal de objeto de áudio aumenta a energia deste objeto de áudio. Deste modo alcançar-se-ia uma percepção sonora aumentada do sinal de saída de áudio. Para manter constante a percepção sonora total, tem de ser executada uma compensação da percepção sonora. Por outras palavras, o sinal de áudio modificado que

resultaria da aplicação da informação de representação no sinal de entrada de áudio teria de ser ajustado. No entanto, o efeito exato da amplificação de um dos sinais de objetos áudio na percepção sonora total do sinal de áudio modificado depende da percepção sonora original do sinal de objeto de áudio amplificado, por ex. do sinal de objeto de áudio, cuja percepção sonora é aumentada. Se a percepção sonora original deste objeto corresponde a uma energia que era consideravelmente baixa, o efeito na percepção sonora total do sinal de entrada de áudio será menor. No entanto, se a percepção sonora original deste objeto corresponde a uma energia que era consideravelmente elevada, o efeito na percepção sonora total do sinal de entrada de áudio será significativo.

[120] Podem ser considerados dois exemplos. Em ambos os exemplos, um sinal de entrada de áudio compreende dois sinais de objetos de áudio e em ambos os exemplos, a aplicação da informação de representação aumenta a energia de um primeiro dos sinais de objetos de áudio em 50 %.

[121] No primeiro exemplo, o primeiro sinal de objeto de áudio contribui 20 % e o segundo sinal de objeto de áudio contribui 80 % para a energia total do sinal de entrada de áudio. No entanto, no segundo exemplo, o primeiro objeto de áudio, o primeiro sinal de objeto de áudio contribui 40 % e o segundo sinal de objeto de áudio contribui 60 % para a energia total do sinal de entrada de áudio. Em ambos os exemplos, estas contribuições podem ser derivadas da informação de percepção sonora nos sinais de objetos de áudio, dado que existe uma relação direta entre a percepção sonora e a energia.

[122] No primeiro exemplo, um aumento de 50 % da energia do primeiro objeto de áudio tem como resultado um sinal de áudio modificado que é gerado através da aplicação da informação de representação no sinal de entrada de áudio com uma energia total de $1,5 \times 20 \% + 80 \% = 110 \%$ da energia do sinal de entrada de áudio.

[123] No segundo exemplo, um aumento de 50 % da energia do primeiro objeto de áudio tem como resultado o sinal de áudio modificado que é gerado

através da aplicação da informação de representação no sinal de entrada de áudio com uma energia total de $1,5 \times 40 \% + 60 \% = 120 \%$ da energia do sinal de entrada de áudio.

[124] Assim, após aplicar a informação de representação no sinal de entrada de áudio, no primeiro exemplo, a energia total do sinal de áudio modificado tem de ser reduzida apenas 9 % (10/110) para obter energia igual tanto no sinal de entrada de áudio e no sinal de saída de áudio, enquanto no segundo exemplo, a energia total do sinal de áudio modificado tem de ser reduzida 17 % (20/120). Para este fim, pode ser calculado o valor de compensação da percepção sonora.

[125] Por exemplo, o valor de compensação da percepção sonora pode ser um valor escalar que é aplicado em todos os canais de saída de áudio do sinal de saída de áudio.

[126] De acordo com uma realização, o processador de sinal é configurado para gerar o sinal de áudio modificado, modificando o sinal de entrada de áudio ao amplificar ou atenuar os sinais de objetos de áudio do sinal de entrada de áudio, de acordo com a informação de representação. Além disso, o processador de sinal é configurado para gerar o sinal de saída de áudio, aplicando o valor de compensação de percepção sonora ao sinal de áudio modificado, de modo a que a percepção sonora do sinal de saída de áudio é igual à percepção sonora do sinal de entrada de áudio ou de modo a que a percepção sonora do sinal de saída de áudio é mais próxima da percepção sonora do sinal de entrada de áudio comparativamente ao sinal de áudio modificado.

[127] Por exemplo, no primeiro exemplo acima, o valor de compensação da percepção sonora lcv , pode, por exemplo, ser definido para um valor $lcv = 10/11$ e um fator de multiplicação de 10/11 pode ser aplicado em todos os canais que resultam da representação dos canais de entrada de áudio, de acordo com a informação de representação.

[128] Assim, por exemplo, no segundo exemplo acima, o valor de

compensação da percepção sonora l_{cv} , pode, por exemplo, ser definido para um valor $l_{cv} = 10/12 = 5/6$, e um fator de multiplicação de $5/6$ pode ser aplicado em todos os canais que resultam da representação dos canais de entrada de áudio, de acordo com a informação de representação.

[129] Em outras realizações, cada um dos sinais de objetos de áudio pode ser atribuído a um de entre vários grupos e um valor de percepção sonora pode ser transmitido para cada um dos grupos, indicando um valor de percepção sonora total dos sinais de objetos de áudio do referido grupo. Se a informação de representação especificar que a energia de um dos grupos é atenuada ou amplificada, por ex. amplificada em 50 %, como acima, pode ser calculado um aumento de energia total e pode ser determinado um valor de compensação de percepção sonora, como anteriormente descrito.

[130] Por exemplo, de acordo com uma realização, cada dos sinais de objetos de áudio do sinal de entrada de áudio ser atribuído a exatamente um grupo de exatamente dois grupos como os dois ou mais grupos. Cada um dos sinais de objetos de áudio do sinal de entrada de áudio ser atribuído a um grupo de objetos de primeiro plano dos dois grupos exatamente ou a um grupo de objetos de segundo plano dos dois grupos exatamente. A interface receptora 110 é configurada para receber a percepção sonora total original de um ou mais sinais de objetos de áudio do grupo de objetos em primeiro plano. Além disso, a interface receptora 110 é configurada para receber a percepção sonora total original de um ou mais sinais de objetos de áudio do grupo de objetos em fundo Além disso, a interface receptora 110 é configurada para receber a informação de representação indicando quanto a pelo menos um grupo dos exatamente dois grupos se o um ou mais sinais de objetos de áudio de cada um do referido pelo menos um grupo serão amplificados ou atenuados através da indicação de uma percepção sonora total modificada do um ou mais sinais de objetos de áudio do referido grupo.

[131] Em uma realização assim, o processador de sinal 120 é configurado para determinar o valor de compensação de percepção sonora dependendo da

percepção sonora total modificada de cada um do referido pelo menos um grupo, dependendo da percepção sonora total original do um ou mais sinais de objetos de áudio do grupo de objetos em primeiro plano e dependendo da percepção sonora total original de um ou mais sinais de objetos áudio do grupo de objetos de fundo. Além disso, o processador de sinal 120 é configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da percepção sonora total modificada de cada um do referido pelo menos um grupo e dependendo do valor de compensação da percepção sonora.

[132] De acordo com algumas realizações, cada um dos sinais de objetos de áudio é atribuído a um de três ou mais grupos e a interface receptora pode ser configurada para receber um valor de percepção sonora para cada um dos três ou mais grupos, indicando a percepção sonora total dos sinais de objetos de áudio do referido grupo.

[133] De acordo com uma realização, para determinar o valor de percepção sonora total de dois ou mais sinais de objetos de áudio, por exemplo, o valor de energia correspondente ao valor da percepção sonora é determinado para cada sinal de objeto de áudio, os valores de energia de todos os valores de percepção sonora são somados para obter uma soma de energia e o valor de percepção sonora correspondente à soma de energia é determinado como o valor de percepção sonora total de dois ou mais sinais de objetos de áudio. Por exemplo, as seguintes fórmulas

$$L_i = c + 10 \log_{10} e_i, \quad e_i = 10^{(L_i - c)/10}$$

[134] podem ser empregues,

[135] Em algumas realizações, os valores da percepção sonora são transmitidos para cada um dos sinais de objetos de áudio ou cada um dos sinais de objetos de áudio é atribuído a um ou dois ou mais grupos, em que para cada um dos grupos, é transmitido um valor da percepção sonora.

[136] No entanto, em algumas realizações, para um ou mais sinais de

objetos de áudio ou para um ou mais dos grupos que compreendem sinais de objetos de áudio, não é transmitido qualquer valor de percepção sonora. Ao invés, o decodificador pode, por exemplo, assumir que estes sinais de objetos de áudio ou grupos de sinais de objetos de áudio, para os quais não é transmitido qualquer valor de percepção sonora, possuem um valor de percepção sonora predefinido. O decodificador, por ex., pode basear todas as outras determinações neste valor da percepção sonora pré-definido.

[137] De acordo com uma realização, a interface receptora 110 está configurada para receber um sinal *downmix* que compreende um ou mais canais *downmix* como o sinal de entrada de áudio, em que um ou mais canais *downmix* compreendem os sinais de objetos de áudio e em que o número dos sinais de objetos de áudio é inferior ao número de um ou mais canais *downmix*. A interface receptora 110 está configurada para receber informação *downmix* indicando como os sinais de objetos de áudio são misturados em um ou mais canais *downmix*, Mais ainda, o processador de sinal 120 está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação *downmix*, dependendo da informação de reconstrução do sinal. e dependendo do valor da percepção sonora. Em uma realização particular, o processador de sinal 120 pode, por exemplo, ser configurado para calcular o valor de compensação da percepção sonora dependendo da informação *downmix*.

[138] Por exemplo, a informação *downmix* pode ser uma matriz *downmix*. Em realizações, o decodificador pode ser um decodificador SAOC. Em realizações como estas, a interface receptora 110 pode, por ex., ser configurada para receber informação de covariância, por ex., uma matriz de covariância tal como acima descrito.

[139] No que respeita à informação de representação que indica se um ou mais sinais de objetos de áudio devem ser amplificados ou atenuados, deverá ter-se em conta, por exemplo, que a informação que indica como um ou mais sinais de objetos de áudio devem ser amplificados ou atenuados, é informação

de representação. Por exemplo, uma matriz de *rendering* **R**, por ex., uma matriz de *rendering* de SAOC, é informação de representação.

[140] A fig. 3 ilustra um sistema de acordo com uma realização.

[141] O sistema compreende um codificador 310 de acordo com uma das realizações acima descritas para codificar uma pluralidade de sinais de objetos de áudio para obter um sinal de áudio codificado que compreenda uma pluralidade de sinais de objetos de áudio.

[142] Mais ainda, o sistema compreende um descodificador 320 de acordo com uma das realizações acima descritas para gerar um sinal de saída de áudio que compreenda um ou mais canais de saída de áudio. O descodificador 320 é configurado para receber o sinal de áudio codificado como um sinal de entrada áudio e para receber a informação da percepção sonora. Mais ainda, o descodificador 320 é configurado para receber ainda informação de representação. Para além disso, o descodificador 320 é configurado para determinar um valor de compensação de percepção sonora, dependendo da informação da percepção sonora e dependendo da informação de representação. Mais ainda, o descodificador 320 é configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

[143] A fig. 7 ilustra a estimativa de percepção sonora informada de acordo com uma realização. À esquerda da corrente de transporte 730, são ilustrados componentes de um codificador de codificação de áudio baseado em objetos. Em particular, uma unidade codificadora baseada em objetos 710 (“codificador áudio baseado em objetos”) e uma unidade codificadora da percepção de sonoridade do objeto 720 é ilustrada (“estimativa da percepção sonora do objeto”).

[144] A própria corrente de transporte 730 compreende informação da percepção sonora **L**, informação *downmix* **D** e a saída do codificador de áudio baseado em objetos 710 **B**.

[145] À direita da corrente de transporte 730, são ilustrados componentes de um processador de sinal de um decodificador de codificação de áudio baseado em objetos. A interface receptora do decodificador não se encontra ilustrada. São representados um estimador da percepção sonora de saída 740 e uma unidade decodificadora de áudio baseada em objetos 750. O estimador da percepção sonora de saída 740 pode ser configurado para determinar o valor de compensação da percepção sonora. A unidade decodificadora de áudio baseada em objetos 750 pode ser configurada para determinar um sinal de áudio modificado a partir de um sinal de áudio, ao ser introduzido no decodificador, aplicando a informação de representação **R**. A aplicação do valor de compensação da percepção sonora no sinal de áudio modificado para compensar uma alteração total da percepção sonora causada pela representação não é apresentada na Fig. 7.

[146] A entrada para o codificador consiste nos objetos de entrada **S** no mínimo. O sistema estima a percepção sonora de cada objeto (ou outra informação relacionada com a percepção sonora, tais como as energias do objeto), por ex., pela unidade codificadora da percepção sonora do objeto 720 e esta informação **L** é transmitida e/ou armazenada. (É igualmente possível que a percepção sonora dos objetos seja facultada como uma entrada para o sistema e a fase de estimativa dentro do sistema pode ser omitida).

[147] Na realização da Fig. 7, o decodificador recebe, pelo menos, a informação da percepção sonora do objeto, por ex., a informação de representação **R** descrevendo a mistura dos objetos para o sinal de saída. Com base nisto, por ex., o estimador da percepção sonora de saída 740 estima a percepção sonora do sinal de saída e facultada esta informação como sua saída.

[148] A informação *downmix* **D** pode ser facultada como a informação de representação, em cujo caso a estimativa da percepção sonora apresenta uma estimativa da percepção sonora do sinal *downmix*. É igualmente possível apresentar a informação *downmix* como uma entrada para a estimativa da percepção sonora do objeto e para transmitir e/ou armazenar a mesma ao

longo da informação da percepção sonora do objeto. A estimativa da percepção sonora do objeto pode então estimar simultaneamente a percepção sonora do sinal *downmix* e a saída de representação e facultar estes dois valores ou suas diferenças como a informação da percepção sonora de saída. O valor de diferença (ou seu inverso) descreve a compensação necessária que deverá ser aplicada no sinal de saída de representação para tornar a sua percepção sonora semelhante à percepção sonora do sinal *downmix*. A informação da percepção sonora do objeto pode adicionalmente incluir informação relativa aos coeficientes de correlação entre vários objetos e esta informação de correlação pode ser utilizada na estimativa da percepção sonora de saída para uma estimativa mais exata.

[149] Seguidamente, é descrita uma realização preferida para aplicação de reforço do diálogo.

[150] Na aplicação de reforço do diálogo, tal como acima descrito, os sinais de objeto de áudio de entrada são agrupados e parcialmente submetidos a *downmix* para formar dois meta-objetos, FGO e BGO, os quais podem ser então trivialmente somados para se obter o sinal *downmix* final.

[151] No seguimento da descrição de SAOC [SAOC], sinais de objeto de entrada N são representados como uma matriz \mathbf{S} do tamanho $N \times N_{\text{Samples}}$, e a informação *downmix* como uma matriz \mathbf{D} do tamanho $M \times N$. Os sinais *downmix* podem então ser obtidos como $\mathbf{X} = \mathbf{D}\mathbf{S}$.

[152] A informação *downmix* \mathbf{D} pode agora ser dividida em duas partes

$$\mathbf{D} = \mathbf{D}_{FGO} + \mathbf{D}_{BGO}$$

[153] para os meta-objetos.

[154] Considerando que cada coluna da matriz \mathbf{D} corresponde a um sinal de objeto de áudio original, as duas matrizes *downmix* componentes podem ser obtidas definindo as colunas, as quais correspondem ao outro meta-objeto em zero (assumindo que não se encontra presente qualquer objeto original em ambos os meta-objetos). Por outras palavras, as colunas que correspondem ao

meta-objeto BGO são definidas para zero \mathbf{D}_{FGO} , e vice-versa.

[155] Estas matrizes *downmixig* descrevem a forma como os dois meta-objetos podem ser obtidos a partir de objetos de entrada, nomeadamente:

$$\mathbf{S}_{FGO} = \mathbf{D}_{FGO}\mathbf{S} \quad \text{e} \quad \mathbf{S}_{BGO} = \mathbf{D}_{BGO}\mathbf{S},$$

[156] e o *downmix* atual é simplificado para

$$\mathbf{X} = \mathbf{S}_{FGO} + \mathbf{S}_{BGO}.$$

[157] Pode igualmente ser considerado que o decodificador do objeto (por ex., SAOC) tente reconstruir os meta-objetos:

$$\tilde{\mathbf{S}}_{FGO} \approx \mathbf{S}_{FGO} \quad \text{e} \quad \tilde{\mathbf{S}}_{BGO} \approx \mathbf{S}_{BGO},$$

[158] e a representação específica de DE pode ser escrita como uma combinação destas duas reconstruções de meta-objetos:

$$\mathbf{Y} = g_{FGO}\mathbf{S}_{FGO} + g_{BGO}\mathbf{S}_{BGO} \approx g_{FGO}\tilde{\mathbf{S}}_{FGO} + g_{BGO}\tilde{\mathbf{S}}_{BGO}.$$

[159] A estimativa da percepção sonora do objeto recebe os dois meta-objetos \mathbf{S}_{FGO} e \mathbf{S}_{BGO} como a entrada e estima a percepção sonora de cada um deles: L_{FGO} sendo a percepção sonora (total/geral) de \mathbf{S}_{FGO} e L_{BGO} sendo a percepção sonora (total/geral) de \mathbf{S}_{BGO} . Estes valores da percepção sonora são transmitidos e/ou armazenados.

[160] Como uma alternativa, utilizando um dos meta-objetos, por ex., o FGO, como referência, é possível calcular a diferença da percepção sonora destes dois objetos, por ex., como

$$\Delta L_{FGO} = L_{BGO} - L_{FGO}.$$

[161] Este valor único é depois transmitido e/ou armazenado.

[162] A fig. 8 ilustra um codificador de acordo com outra realização. O codificador da fig. 8 compreende um *downmixer* de um objeto 811 e um estimador de informação lateral do objeto 812. Para além disso, o codificador da fig. 9 compreende ainda uma unidade codificadora da percepção sonora do objeto 820. Mais ainda, o codificador da fig. 8 compreende um misturador de meta-objeto de áudio 805.

[163] O codificador da fig. 8 utiliza meta-objetos de áudio intermediários como uma entrada para a estimativa da percepção sonora do objeto. Em realizações, o codificador da fig. 8 pode ser configurado para gerar dois meta-objetos de áudio. Em outras realizações, o codificador da fig. 8 pode ser configurado para gerar três ou mais meta-objetos de áudio.

[164] Entre outros aspectos, os conceitos apresentados proporcionam a nova funcionalidade que consiste na possibilidade de o codificador, por ex., estimar a percepção sonora média de todos os objetos de entrada. Os objetos podem, por ex., ser misturados em um sinal *downmix* que é transmitido. Os conceitos apresentados proporcionam ainda a nova funcionalidade que consiste na possibilidade de a percepção sonora do objeto e a informação *downmix* ser, por ex., incluída na informação lateral de codificação do objeto que é transmitida.

[165] O decodificador pode, por ex., utilizar a informação lateral codificadora do objeto para separação (virtual) dos objetos e recombina os objetos utilizando a informação de representação.

[166] Para além disso, os conceitos apresentados proporcionam a nova funcionalidade segundo a qual a informação *downmix* pode ser utilizada para estimar a percepção sonora do sinal *downmix* predefinida, a informação de representação e a percepção sonora do objeto recebido podem ser utilizados para estimar a percepção sonora média do sinal de saída e/ou a alteração da percepção sonora pode ser estimada a partir destes dois valores. Ou, a informação *downmix* e de representação pode ser utilizada para estimar a alteração da percepção sonora a partir do *downmix* predefinido, outra nova funcionalidade dos conceitos apresentados.

[167] Para além disso, os conceitos apresentados proporcionam a nova funcionalidade segundo a qual a saída do decodificador pode ser modificada para compensar a alteração na percepção sonora de modo a que a percepção sonora média do sinal modificado corresponda à percepção sonora média do *downmix* predefinido.

[168] É ilustrado, na Fig. 9, uma realização específica relacionada com SAOC-DE. O sistema recebe os sinais do objeto de áudio de entrada, a informação *downmix* e a informação do agrupamento de objetos para meta-objetos. Com base nestes, o misturador de meta-objeto de áudio 905 forma os dois meta-objetos S_{FGO} e S_{BGO} . É possível que a porção do sinal que é processado com SAOC não constitua o sinal inteiro. Por exemplo, em uma configuração de canal 5.1, SAOC pode ser implantado em um sub-grupo de canais, como no canal posterior (esquerda, direita e centro) enquanto os outros canais (som à esquerda, som à direita e efeitos de baixa frequência) são enviados, passando o SAOC e entregues como tal. Estes canais não processados por SAOC são assinalados com X_{BYPASS} . Os canais de *bypass* possíveis precisam ser fornecidos para o codificador para uma estimativa mais precisa da informação da percepção sonora.

[169] Os canais de *bypass* podem ser manuseados de várias formas.

[170] Por exemplo, os canais de *bypass* podem, por ex., formar um meta-objeto independente. Isto permite definir a representação de modo a que todos os três meta-objetos sejam representados em escala de forma independente.

[171] Ou, por exemplo, os canais de *bypass* podem, por ex., ser combinados com um dos outros dois meta-objetos. As definições de representação desse meta-objeto controlam igualmente a porção do canal de *bypass*. Por exemplo, no cenário de reforço de diálogo, pode ser significativo combinar os canais de *bypass* com o meta-objeto de segundo plano.

$$X_{BGO} = S_{BGO} + X_{BYPASS}.$$

[172] Ou, por exemplo, os canais de *bypass* podem, por ex., ser ignorados.

[173] De acordo com realizações, a unidade de codificação baseada em objetos 210 do codificador é configurada para receber os sinais de objetos de áudio, em que cada um dos sinais de objetos de áudio é atribuído a exatamente um de exatamente dois grupos, em que cada dos exatamente dois

grupos compreende um ou mais dos sinais de objetos de áudio. Mais ainda, em que a unidade de codificação baseada em objetos 210 está configurada para *downmix* os sinais de objetos de áudio, sendo compreendida pelos exatamente dois grupos, para obter um sinal *downmix* que compreenda um ou mais canais de áudio *downmix* como o sinal de áudio codificado, em que o número de um ou mais canais *downmix* é inferior ao número dos sinais de objetos de áudio sendo compreendido pelos exatamente dois grupos. A unidade de codificação da percepção sonora do objeto 220 é designada para receber um ou mais sinais de objetos de áudio de *bypass* adicionais, em que cada um ou mais dos sinais de objetos de áudio de *bypass* adicionais é atribuído a um terceiro grupo, em que cada de um ou mais dos sinais de objetos de áudio de *bypass* adicionais não é compreendido pelo primeiro grupo e não é compreendido pelo segundo grupo, em que a unidade de codificação baseada em objetos 210 é configurada para não *downmix* um ou mais dos sinais de objetos de áudio de *bypass* adicionais com o sinal *downmix*.

[174] Em uma realização, a unidade de codificação da percepção sonora do objeto 220 é configurada para determinar um primeiro valor de percepção sonora, um Segundo valor de percepção sonora e um terceiro valor de percepção Sonora da informação de percepção sonora, o primeiro valor da percepção sonora indicando uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo, o segundo valor da percepção sonora indicando uma percepção sonora total de um ou mais sinais de objetos de áudio do segundo grupo e o terceiro valor da percepção sonora indicando uma percepção sonora total de um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo. Em outra realização, a unidade de codificação da percepção sonora do objeto 220 é configurada para determinar um primeiro valor de percepção sonora e um segundo valor da percepção Sonora da informação da percepção sonora, o primeiro valor da percepção sonora indicando uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo, o segundo valor da percepção sonora indicando uma

percepção sonora total de um ou mais sinais de objetos de áudio do segundo grupo e de um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo.

[175] De acordo com uma realização, a interface receptora 110 do decodificador é configurada para receber o sinal *downmix*. Mais ainda, a interface receptora 110 é configurada para receber um ou mais sinais de objetos de áudio de *bypass* adicionais, em que um ou mais sinais de objetos de áudio de *bypass* adicionais não se encontram misturados com o sinal *downmix*. Para além disso, a interface receptora 110 está configurada para receber a informação de ruído que indica informação sobre a percepção sonora dos sinais de objetos de áudio, os quais se encontram misturados com o sinal *downmix* e indicam informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais que não se encontram misturados com o sinal *downmix*. Mais ainda, o processador de sinal 120 é configurado para determinar o valor de compensação da percepção sonora dependendo da informação sobre a percepção Sonora dos sinais de objetos de áudio, os quais se encontram misturados com o sinal *downmix* e que dependem da informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais que não se encontram misturados com o sinal *downmix*.

[176] A fig. 9 ilustra um codificador e um decodificador de acordo com uma realização relacionada com o SAOC-DE, o qual compreende canais de *bypass*. Entre outros aspectos, o codificador da fig. 9 inclui um codificador SAOC 902.

[177] Na realização da fig. 9, a combinação possível dos canais *bypass* com os outros meta-objetos ocorre nos dois blocos de “inclusão de *bypass*” 913, 914, produzindo os meta-objetos X_{FGO} e X_{BGO} com as partes definidas dos canais de *bypass* incluídas.

[178] A percepção sonora L_{BYPASS} , L_{FGO} , e L_{BGO} de ambos estes meta-objetos são estimadas nas unidades de estimativa de percepção sonora 921,

922, 923. Esta informação da percepção sonora é depois transformada em uma codificação apropriada em um estimador de informação da percepção sonora de um meta-objeto 925 e depois transmitida e/ou armazenada.

[179] O atual codificador e decodificador operam como esperado, extraindo a informação lateral do objeto dos objetos, criando o sinal *downmix* X , e transmitindo e/ou armazenando a informação para o decodificador. Os canais de *bypass* possíveis são transmitidos e/ou armazenados na restante informação para o decodificador.

[180] O decodificador SAOC-DE 945 recebe um valor de ganho "ganho de Diálogo" como uma entrada do utilizador. Com base nesta entrada e informação *downmix* recebida, o decodificador SAOC 945 determina a informação de representação. O decodificador SAOC 945 produz depois a cena de saída de representação como o sinal Y . Adicionalmente, produz um fator ganho (e um valor de atraso) que deveria ser aplicado nos possíveis sinais de *bypass* X_{BYPASS} .

[181] A unidade de "inclusão de *bypass*" 955 recebe esta informação juntamente com a cena de saída de representação e os sinais de *bypass* e cria o sinal de cena de saída completo. O decodificador SAOC 945 produz igualmente um grupo de valores ganhos de meta-objetos, sendo que a quantidades destes depende do agrupamento de meta-objetos e forma de informação da percepção sonora desejada.

[182] Os valores de ganho são fornecidos para o estimador da percepção sonora de mistura 960, o qual recebe igualmente a informação da percepção sonora do meta-objeto a partir do decodificador.

[183] O estimador da percepção sonora de mistura 960 tem depois capacidade para determinar a informação da percepção sonora desejada, a qual pode incluir, mas sem que tal constitua qualquer limitação, a percepção sonora do sinal *downmix*, a percepção sonora da cena de saída de representação e/ou a diferença na percepção sonora entre o sinal *downmix* e a cena de saída de representação.

[184] Em algumas realizações, a própria informação da percepção sonora é suficiente, enquanto que em outras realizações, é desejável processar a saída completa dependendo da informação da percepção sonora determinada. Este processamento pode, por exemplo, ser compensação de qualquer diferença possível na percepção sonora entre o sinal *downmix* e a cena de saída de reprodução. Um processamento deste tipo, por ex., por uma unidade de processamento de percepção sonora 970, faria sentido no cenário de transmissão, uma vez que reduziria as alterações na percepção sonora do sinal detectado independentemente da interação do utilizador (definição do "ganho de diálogo" de entrada).

[185] O processamento relacionado com a percepção sonora nesta realização específica compreende uma pluralidade de novas funcionalidades. Entre outros aspectos, o FGO, BGO e os possíveis canais de *bypass* são pré-misturados na configuração do canal final de modo a que o *downmix* possa ser realizado ao simplesmente se adicionar os dois sinais pré-misturados conjuntamente (por ex., coeficiente da matriz *downmix* de 1), a qual constitui uma nova funcionalidade. Mais ainda, como uma nova funcionalidade adicional, é estimada a percepção sonora média do FGO e BGO e é calculada a diferença. Para além disso, os objetos são misturados em um sinal *downmix* que é transmitido. Mais ainda, como uma nova funcionalidade adicional, a informação da diferença da percepção sonora é incluída na informação lateral que é transmitida. (nova) Para além disso, o descodificador utiliza a informação lateral para separação (virtual) dos objetos e recombina os objetos utilizando a informação de representação, a qual é baseada na informação *downmix* e no ganho de modificação de entrada do utilizador. Mais ainda, como outra nova funcionalidade, o descodificador utiliza o ganho de modificação e a informação da percepção sonora transmitida para estimar a alteração na percepção sonora média da saída do sistema comparativamente ao *downmix* pré-definido.

[186] Seguidamente, é facultada uma descrição formal das realizações.

[187] Assumindo que os valores da percepção sonora dos objetos têm um

comportamento semelhante aos valores de energia quando se somam os objetos, i.e., os valores da percepção sonora devem ser transformados em domínio linear, aí adicionados e finalmente transformados de novo para o domínio logarítmico. A motivação desta situação através da definição da medida de percepção sonora BS.1770 irá ser agora apresentada (para efeitos de simplicidade, o número de canais é definido para um, mas o mesmo princípio pode ser aplicado a sinais multicanal com cálculos apropriados sobre os canais).

[188] A percepção sonora do i^{th} sinal filtrado por K z_i com a energia quadrática média e_i é definida como

$$L_i = c + 10 \log_{10} e_i,$$

[189] em que c é uma constante de deslocamento. Por exemplo, c pode ser $-0,691$. Deste modo, a energia do sinal pode ser determinada a partir da percepção sonora com

$$e_i = 10^{(L_i - c)/10}.$$

[190] A energia da soma de N sinais não correlacionados $z_{SUM} = \sum_{i=1}^N z_i$ é então

$$e_{SUM} = \sum_{i=1}^N e_i = \sum_{i=1}^N 10^{(L_i - c)/10},$$

[191] e a percepção sonora deste sinal de soma é então

$$L_{SUM} = c + 10 \log_{10} e_{SUM} = c + 10 \log_{10} \sum_{i=1}^N 10^{(L_i - c)/10}.$$

[192] Se os sinais não estão não-correlacionados, os coeficientes de correlação $C_{i,j}$ deverão ser considerados quando se aproxima a energia do sinal de soma como

$$e_{SUM} = \sum_{i=1}^N \sum_{j=1}^N e_{i,j},$$

[193] em que a energia cruzada $e_{i,j}$ entre i^{th} e objetos j^{th} é definida como

$$\begin{aligned} e_{i,j} &= C_{i,j} \sqrt{e_i e_j} \\ &= C_{i,j} \sqrt{10^{(L_i-c)/10} 10^{(L_j-c)/10}} , \\ &= C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}} \end{aligned}$$

[194] em que $-1 \leq C_{i,j} \leq 1$ é o coeficiente de correlação entre os dois objetos i e j . Quando dois objetos estão não-correlacionados, o coeficiente de correlação é igual a 0 e quando os dois objetos são idênticos, o coeficiente de correlação é igual a 1.

[195] Alargando mais o modelo com pesos misturados g_i a ser aplicado aos sinais no processo de mistura, i.e., $z_{SUM} = \sum_{i=1}^N g_i z_i$, a energia do sinal de soma será

$$e_{SUM} = \sum_{i=1}^N \sum_{j=1}^N g_i g_j e_{i,j} ,$$

[196] e a percepção do sinal de mistura pode ser obtido a partir daqui, como anteriormente, com

$$L_{SUM} = c + 10 \log_{10} e_{SUM} .$$

[197] A diferença entre a percepção sonora de dois sinais pode ser estimada como

$$\Delta L(i, j) = L_i - L_j .$$

[198] Se a definição de percepção sonora é agora utilizada como anteriormente, tal pode ser escrito como

$$\begin{aligned} \Delta L(i, j) &= L_i - L_j \\ &= (c + 10 \log_{10} e_i) - (c + 10 \log_{10} e_j) , \\ &= 10 \log_{10} \frac{e_i}{e_j} \end{aligned}$$

[199] o que pode ser observado como uma função de energias de sinal.

Deseja-se agora estimar a diferença de percepção sonora entre duas misturas

$$z_A = \sum_{i=1}^N g_i z_i \text{ e } z_B = \sum_{i=1}^N h_i z_i$$

[200] com pesos de mistura possivelmente diferentes g_i e h_i , tal pode ser estimado com

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{e_A}{e_B} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j e_{i,j}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j e_{i,j}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(L_i+L_j-2c)/10}}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(L_i+L_j)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(L_i+L_j)/10}}} \end{aligned}$$

[201] No caso de os objetos serem não-correlacionados ($C_{i,j} = 0, \forall i \neq j$ e $C_{i,j} = 1, \forall i = j$), a estimativa de diferença torna-se

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{(L_i-c)/10}}{\sum_{i=1}^N h_i^2 10^{(L_i-c)/10}} \\ &= 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{L_i/10}}{\sum_{i=1}^N h_i^2 10^{L_i/10}} \end{aligned}$$

[202] Seguidamente, considera-se codificação diferencial.

[203] É possível codificar os valores de percepção sonora por objeto como

diferenças da percepção sonora de um objeto de referência selecionado.

$$K_i = L_i - L_{REF} ,$$

[204] em que L_{REF} é a percepção sonora do objeto de referência. Esta codificação é benéfica se não forem necessários valores absolutos de percepção sonora, porque é agora necessário transmitir um valor a menos e a estimativa da diferença da percepção sonora pode ser escrita como

$$\Delta L(A, B) = 10 \log_{10} \frac{\sum_{i=1}^N \sum_{j=1}^N g_i g_j C_{i,j} \sqrt{10^{(K_i+K_j)/10}}}{\sum_{i=1}^N \sum_{j=1}^N h_i h_j C_{i,j} \sqrt{10^{(K_i+K_j)/10}}} ,$$

[205] ou no caso de objetos não correlacionados

$$\Delta L(A, B) = 10 \log_{10} \frac{\sum_{i=1}^N g_i^2 10^{K_i/10}}{\sum_{i=1}^N h_i^2 10^{K_i/10}} .$$

[206] Seguidamente é considerado um cenário de reforço de diálogo.

[207] Considerando uma vez mais o cenário de aplicação do reforço de diálogo. A liberdade de definir a informação de representação no decodificador é apenas limitada na mudança de níveis dos meta-objetos. Vamos ainda assumir que os dois meta-objetos são não correlacionados, i.e., $C_{FGO,BGO} = 0$. Se os pesos *downmix* dos meta-objetos são h_{FGO} e h_{BGO} , e forem representados com os ganhos f_{FGO} e f_{BGO} , a percepção sonora da saída relativa ao *downmix* pré definido é

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{f_{FGO}^2 10^{(L_{FGO}-c)/10} + f_{BGO}^2 10^{(L_{BGO}-c)/10}}{h_{FGO}^2 10^{(L_{FGO}-c)/10} + h_{BGO}^2 10^{(L_{BGO}-c)/10}} \\ &= 10 \log_{10} \frac{f_{FGO}^2 10^{L_{FGO}/10} + f_{BGO}^2 10^{L_{BGO}/10}}{h_{FGO}^2 10^{L_{FGO}/10} + h_{BGO}^2 10^{L_{BGO}/10}} . \end{aligned}$$

[208] Esta é então a compensação necessária se se deseja ter a mesma percepção sonora na saída como no *downmix* pré definido.

[209] $\Delta L(A, B)$ pode ser considerado como um valor de compensação de percepção sonora que pode ser transmitido pelo processador de sinal 120 do decodificador. $\Delta L(A, B)$ pode igualmente ser denominado como um valor de mudança de percepção sonora e, assim, o valor de compensação atual pode ser um valor inverso. Ou é correto utilizar o nome "fator de compensação de percepção sonora" para esta situação também? Deste modo, o valor de compensação de percepção sonora lcv mencionado anteriormente neste documento corresponderia ao valor g_{Δ} abaixo.

[210] Por exemplo, $g_{\Delta} = 10^{-\Delta L(A, B)/20}$ $1 / \Delta L(A, B)$ pode ser aplicada como um fator de multiplicação em cada canal de um sinal de áudio modificado que resulta da aplicação da informação de representação no sinal de entrada de áudio. Esta equação para g_{Δ} funciona no domínio linear. No domínio logarítmico, a equação seria diferente tal como $1 / \Delta L(A, B)$ e aplicada em concordância.

[211] Se o processo *downmix* é simplificado de modo a que os meta-objetos possam ser misturados com pesos de unidade para se obter o sinal *downmix*, i.e., $h_{FGO} = h_{BGO} = 1$, e neste momento os ganhos de representação para estes dois objetos são assinalados com g_{FGO} e g_{BGO} . Isto simplifica a equação para a alteração de percepção sonora em

$$\begin{aligned} \Delta L(A, B) &= 10 \log_{10} \frac{g_{FGO}^2 10^{(L_{FGO}-c)/10} + g_{BGO}^2 10^{(L_{BGO}-c)/10}}{10^{(L_{FGO}-c)/10} + 10^{(L_{BGO}-c)/10}} \\ &= 10 \log_{10} \frac{g_{FGO}^2 10^{L_{FGO}/10} + g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10} + 10^{L_{BGO}/10}} \end{aligned}$$

[212] Uma vez mais, $\Delta L(A, B)$ pode ser considerado como um valor de compensação de percepção sonora que é determinado pelo processador de sinal 120.

[213] De modo geral, g_{FGO} pode ser considerado como um ganho de representação para o objeto de primeiro plano *FGO* (grupo de objeto de primeiro plano) e g_{BGO} pode ser considerado como um ganho de representação

para o objeto de segundo plano BGO (grupo de objeto de segundo plano).

[214] Tal como anteriormente mencionado, é possível transmitir diferenças de percepção sonora em vez de percepção sonora absoluta. Permitam-nos definir a percepção sonora de referência como a percepção sonora do meta-objeto FGO $L_{REF} = L_{FGO}$, i.e., $K_{FGO} = L_{FGO} - L_{REF} = 0$ e $K_{BGO} = L_{BGO} - L_{REF} = L_{BGO} - L_{FGO}$. Agora, a alteração da percepção sonora é

$$\Delta L(A, B) = 10 \log_{10} \frac{g_{FGO}^2 + g_{BGO}^2 10^{K_{BGO}/10}}{1 + 10^{K_{BGO}/10}}.$$

[215] Também pode ser, tal como no caso do SAOC-DE, que dois meta-objetos não possuam fatores de escala individuais, mas um dos objetos é deixado não modificado enquanto o outro é atenuado para obter o rácio de mistura correta entre os objetos. Nesta definição de representação, a saída será mais baixa em percepção sonora comparativamente à mistura pré definida e a mudança na percepção sonora é

$$\Delta L(A, B) = 10 \log_{10} \frac{\hat{g}_{FGO}^2 + \hat{g}_{BGO}^2 10^{K_{BGO}/10}}{1 + 10^{K_{BGO}/10}},$$

[216] com

$$\hat{g}_{FGO} = \begin{cases} 1 & , \text{ if } g_{FGO} \geq g_{BGO} \\ \frac{g_{FGO}}{g_{BGO}} & , \text{ if } g_{FGO} < g_{BGO} \end{cases}, \text{ e } \hat{g}_{BGO} = \begin{cases} \frac{g_{BGO}}{g_{FGO}} & , \text{ if } g_{BGO} < g_{FGO} \\ 1 & , \text{ if } g_{BGO} \geq g_{FGO} \end{cases}.$$

[217] Esta forma é já mais simples e mais agnóstica relativamente à medida de percepção sonora utilizada. O único requisito real é, que os valores de percepção sonora deverão somar no domínio exponencial. É possível transmitir/armazenar valores de energias de sinal em vez de valores de percepção sonora, uma vez que os dois têm uma relação próxima.

[218] Em cada das fórmulas acima, $\Delta L(A, B)$ pode ser considerado como um valor de compensação de percepção sonora que pode ser transmitido pelo processador de sinal 120 do decodificador.

[219] Seguidamente, serão considerados casos de exemplo. A precisão

dos conceitos facultados é ilustrada através de dois sinais de exemplo. Ambos os sinais têm um *downmix* 5.1 com o *surround* e canais LFE contornados pelo processamento SAOC.

[220] São utilizadas duas abordagens principais: uma ("3-períodos") com três meta-objetos: FGO, BGO e canais *bypass*, por ex.,

$$\mathbf{X} = \mathbf{X}_{FGO} + \mathbf{X}_{BGO} + \mathbf{X}_{BYPASS} ,$$

[221] E outra ("2-períodos") com dois meta-objetos, por ex.:

$$\mathbf{X} = \mathbf{X}_{FGO} + \mathbf{X}_{BGO} .$$

[222] Na abordagem de 2-períodos, os canais *bypass* podem, por ex., ser misturados conjuntamente com o BGO para a estimativa de percepção sonora do meta-objeto. A percepção sonora de ambos (ou todos os três) objetos assim como a percepção sonora do sinal *downmix* são estimadas e os valores são armazenados.

[223] As instruções de representação são da forma

$$\mathbf{Y} = \hat{g}_{FGO} \mathbf{X}_{FGO} + \hat{g}_{BGO} \mathbf{X}_{BGO} + \hat{g}_{BGO} \mathbf{X}_{BYPASS}$$

[224] e

$$\mathbf{Y} = \hat{g}_{FGO} \mathbf{X}_{FGO} + \hat{g}_{BGO} \mathbf{X}_{BGO}$$

[225] para as duas abordagens respectivamente.

[226] Os valores de ganho são, por ex., determinados de acordo com:

$$\hat{g}_{FGO} = \begin{cases} 1 & , \text{ if } g_{FGO} > 1 \\ g_{FGO} & , \text{ otherwise } \end{cases} , \quad \text{ e } \quad \hat{g}_{BGO} = \begin{cases} 1/g_{FGO} & , \text{ if } g_{FGO} > 1 \\ 1 & , \text{ otherwise } \end{cases} ,$$

[227] em que o ganho FGO g_{FGO} é variado entre -24 e +24 dB.

[228] O cenário de saída é representado, a percepção sonora é medida e a atenuação da percepção sonora do sinal *downmix* é calculada.

[229] Este resultado é exibido na fig. 10 e fig. 11 com a linha azul com marcadores circulares. A fig. 10 representa uma primeira ilustração e a fig. 11 representa uma segunda ilustração de uma alteração de percepção sonora medida e o resultado de utilizar os conceitos facultados para estimar a

alteração na percepção sonora em uma maneira puramente paramétrica.

[230] Seguidamente, a atenuação do *downmix* é estimada parametricamente empregando valores de percepção sonora de meta-objetos armazenados e o *downmix* e informação de representação. A estimativa utilizando percepção sonora de três meta-objetos é ilustrada com a linha verde com marcadores quadrados e a estimativa utilizando a percepção sonora de dois meta-objetos é ilustrada com a linha vermelha com marcadores em estrela.

[231] Pode observar-se nas figuras que as abordagens de 2 períodos e 3-períodos apresentam resultados praticamente idênticos e ambas se aproximam bastante do valor medido.

[232] Os conceitos facultados exibem uma pluralidade de vantagens. Por exemplo, os conceitos facultados permitem estimar a percepção sonora de um sinal de mistura a partir da percepção dos sinais de componentes que formam a mistura. O benefício daí resultante é que a percepção sonora do sinal de componente pode ser estimado de uma vez e a estimativa da percepção sonora do sinal de mistura pode ser obtido parametricamente para qualquer mistura sem necessidade da estimativa de percepção sonora baseada no sinal atual. Isto oferece uma melhoria considerável na eficácia computacional de todo o sistema no qual a estimativa da percepção sonora de várias misturas é necessária. Por exemplo, quando o utilizador final altera as definições de representação, a estimativa da percepção sonora da saída fica imediatamente disponível.

[233] Em algumas aplicações, tal como quando da adaptação à recomendação EBU R128, a percepção sonora média de todo o programa é importante. Se a estimativa da percepção sonora no receptor, por ex., em um cenário de transmissão, é realizada com base no sinal recebido, a estimativa converge para a percepção sonora média apenas depois de se receber o programa completo. Consequentemente, qualquer compensação da percepção sonora irá ter erros ou exibir variações temporais. Ao se estimar a percepção

sonora de objetos de componente como proposto e ao se transmitir a informação da percepção sonora, é possível estimar a percepção sonora de mistura média no receptor sem um atraso.

[234] Caso se pretender que a percepção sonora média do sinal de saída permaneça (aproximadamente) constante independentemente das alterações na informação de representação, os conceitos facultados permitem determinar um fator de compensação para esta razão. Os cálculos necessários para tal no decodificador são, na sua complexidade computacional, insignificantes e a funcionalidade é, deste modo, possível de ser adicionada a qualquer decodificador.

[235] Existem casos nos quais o nível de percepção sonora absoluta da saída não é importante, mas a importância assenta em determinar a alteração na percepção sonora a partir de uma cena de referência. Em tais casos, os níveis absolutos dos objetos não são importantes, mas os seus níveis relativos são importantes. Tal permite definir um dos objetos como o objeto de referência e representar a percepção sonora dos restantes objetos em relação à percepção sonora deste objeto de referência. Tal oferece alguns benefícios considerando o transporte e/ou armazenamento da informação de percepção sonora.

[236] Primeiro que tudo, não é necessário transportar o nível de percepção sonora de referência. No caso da aplicação dos dois meta-objetos, a quantidade de dados a transmitir é reduzida a metade. O segundo benefício refere-se à quantização possível e representação dos valores de percepção sonora. Considerando que os níveis absolutos dos objetos podem ser quase nada, os valores de percepção sonora absoluta podem igualmente ser quase nada. Os valores de percepção sonora relativa, por outro lado, são assumidos por ter uma média de 0 e uma distribuição mais agradavelmente formada à volta da média. A diferença entre as representações permite definir a grelha de quantização da representação relativa em uma forma com uma precisão potencialmente maior com o mesmo número de bits utilizado para a

representação quantificada.

[237] A fig. 12 ilustra outra realização para conduzir compensação de percepção sonora. Na fig. 12, compensação de percepção sonora pode ser conduzida, por ex., para compensar a perda em percepção sonora, Para este efeito, por ex., os valores **DE_loudness_diff_dialogue** (= K_{FGO}) e **DE_loudness_diff_background** (= K_{BGO}) de **DE_control_info** podem ser utilizados. Aqui, **DE_control_info** pode especificar informação de controle de "Reforço de Diálogo" (DE) Áudio Limpo

[238] A compensação de percepção sonora é conseguida ao se aplicar um valor de ganho "g" no sinal de saída SAOC-DE e os canais submetidos a *bypass* (no caso de um sinal multicanal).

[239] Na realização da fig. 12, tal é realizado como segue:

[240] Um valor de ganho de modificação de diálogo limitado m_g é utilizado para determinar os ganhos efetivos para o objeto de primeiro plano (FGO, por ex., diálogo) e para o objeto de segundo plano (BGO, por ex., ambiente). Tal é realizado pelo bloco "Mapa de ganhos" 1220, o qual produz os valores de ganho m_{FGO} e m_{BGO} .

[241] O bloco "Estimador de percepção sonora de saída" 1230 utiliza a informação de percepção sonora K_{FGO} e K_{BGO} , e os valores de ganho efetivos m_{FGO} e m_{BGO} para estimar esta alteração possível na percepção sonora comparativamente ao caso *downmix* pré definido. A alteração é depois mapeada para o "fator de compensação de percepção sonora", o qual é aplicado nos canais de saída para produzir os "Sinais de Saída" finais.

[242] Os seguintes passos são aplicados para compensação de percepção sonora:

[243] Receber o valor de ganho limitado m_g do descodificador SAOC-DE (tal como definido em

[244] cláusula 12.8 "Controle da gama de modificação para SAOC-DE"

[DE]), e determinar os ganhos FGO/BGO aplicados:

$$\begin{cases} m_{FGO} = m_G, \text{ and } m_{BGO} = 1 & \text{if } m_G \leq 1 \\ m_{FGO} = 1, \text{ and } m_{BGO} = m_G^{-1} & \text{if } m_G > 1 \end{cases} .$$

[245] Obter a informação de percepção sonora do meta-objeto K_{FGO} e K_{BGO} .

[246] Calcular a alteração na percepção sonora de saída comparativamente ao *downmix* predefinido com

$$\Delta L = 10 \log_{10} \frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}} .$$

[247] Calcular o ganho de compensação da percepção sonora $g_{\Delta} = 10^{-0.05\Delta L}$.

[248] - Calcular os fatores de escala $\mathbf{g} = \begin{bmatrix} g_1 \\ \vdots \\ g_N \end{bmatrix}$, em que

$$g_i = \begin{cases} g_{\Delta} & \text{if channel } i \text{ belongs to SAOC-DE output} \\ m_{BGO} g_{\Delta} & \text{if channel } i \text{ is a by-pass channel} \end{cases} , \text{ e } N \text{ é o número total de}$$

canais de saída. Na Fig. 12, o ajuste do ganho é dividido em duas fases: o ganho dos possíveis "canais *bypass*" é ajustado com m_{BGO} antes de combinar os mesmos com os "canais de saída SAOC-DE" e depois um ganho comum g_{Δ} é então aplicado a todos os canais combinados. Isto é apenas uma possível reordenação das operações de ajuste de ganho, enquanto \mathbf{g} aqui combina ambas as fases de ajuste de ganho em um ajuste de ganho.

[249] Aplicar os valores de escala \mathbf{g} nos canais áudio \mathbf{Y}_{FULL} consistindo nos "canais de saída SAOC-DE" \mathbf{Y}_{SAOC} e os possíveis "canais *bypass*" alinhados no tempo \mathbf{Y}_{BYPASS} : $\mathbf{Y}_{FULL} = \mathbf{Y}_{SAOC} \cup \mathbf{Y}_{BYPASS}$

[250] Aplicar os valores de escala g nos canais áudio Y_{FULL} é conduzido pela unidade de ajuste de ganho 1240.

[251] ΔL como acima calculada pode ser considerada como um valor de compensação de percepção sonora. De modo geral, m_{FGO} indica um ganho de representação para o objeto de primeiro plano FGO (grupo de objeto de primeiro plano) e m_{BGO} indica um ganho de representação para o objeto de segundo plano BGO (grupo de objeto de segundo plano).

[252] Embora tenham sido descritos alguns aspectos no contexto de um aparelho, torna-se claro que estes aspectos representam igualmente uma descrição do método correspondente, em que um bloco ou dispositivo corresponde a uma fase do método ou uma funcionalidade de uma fase do método. Analogamente, aspectos descritos no contexto de uma fase do método representam igualmente uma descrição de um bloco correspondente ou item ou funcionalidade de um aparelho correspondente.

[253] O sinal decomposto inventivo pode ser armazenado em um meio de armazenamento digital ou pode ser transmitido em um meio de transmissão, tal como um meio de transmissão sem fios ou um meio de transmissão com fios, tal como Internet.

[254] Dependendo de determinados requisitos de implementação, as realizações da invenção podem ser implementadas em *hardware* ou em *software*. A implementação pode se realizada utilizando um meio de armazenamento digital, por exemplo uma disquete, um DVD, um CD, ROM, PROM, EPROM, EEPROM ou uma memória FLASH, com sinais de controle eletronicamente legíveis armazenados nos mesmos, os quais cooperam (ou têm capacidade de cooperar) com um sistema informático programável de modo a que o método respectivo seja executado.

[255] Algumas realizações, de acordo com a invenção, compreendem um veículo de dados não transitórios com sinais de controle eletronicamente legíveis, os quais têm capacidade de cooperar com um sistema informático

programável, de modo a que um dos métodos presentemente descrito seja executado.

[256] De modo geral, as realizações da presente invenção podem ser implementadas como um produto de programa informático com um código de programa, sendo o código de programa operativo para realizar um dos métodos quando o produto do programa informático correr em um computador. O código de programa pode, por exemplo, ser armazenado em um veículo legível por máquinas.

[257] Outras realizações compreendem o programa informático para realizar um dos métodos presentemente descritos, armazenado em um veículo legível por máquinas.

[258] Por outras palavras, uma realização do método inventivo é, por conseguinte, um programa informático com um código de programa para realizar um dos métodos presentemente descritos, quando o programa informático correr em um computador.

[259] Outra realização dos métodos inventivos é, por conseguinte, um veículo de dados (ou um meio de armazenamento digital, ou um meio legível por computador) que compreende, gravados no mesmo, o programa informático para realizar um dos métodos presentemente descritos.

[260] Outra realização do método da invenção é, por conseguinte, uma corrente de dados ou uma sequência de sinais que representam o programa informático para realizar um dos métodos presentemente descritos. A corrente de dados ou a sequência de sinais pode, por exemplo, ser configurada para ser transferida via uma conexão de comunicação de dados, por exemplo, através da Internet.

[261] Outra realização compreende um meio de processamento, por exemplo, um computador ou um dispositivo lógico programável, configurado para ou adaptado para realizar um dos métodos presentemente descritos.

[262] Outra realização compreende um computador que tenha instalado o programa informático para realizar um dos métodos presentemente descritos.

[263] Em algumas realizações, um dispositivo lógico programáveis (por exemplo, uma matriz de portas de campo programável) pode ser utilizado para realizar algumas das ou todas as funcionalidades dos métodos presentemente descritos. Em algumas realizações, uma matriz de portas de campo programáveis pode cooperar com um microprocessador de modo a realizar um dos métodos presentemente descritos. De modo geral, os métodos são preferencialmente realizados por qualquer aparelho de *hardware*.

[264] As realizações acima descritas são meramente ilustrativas para os princípios da presente invenção. Compreende-se que modificações e variações das disposições e dos detalhes presentemente descritos serão evidentes para os especialistas na técnica. A intenção é, por conseguinte, limitarmo-nos apenas ao âmbito das reivindicações da patente iminente e não aos detalhes específicos apresentados em forma de descrição e explicação das realizações aqui compreendidas.

REIVINDICAÇÕES

1.Descodificador para gerar um sinal de saída de áudio que compreende um ou mais canais de saída de áudio, **caracterizado por** o decodificador incluir:

um interface receptor (110) para receber um sinal de entrada de áudio que compreende uma pluralidade de sinais de objetos de áudio, para receber informação de percepção sonora nos sinais de objetos de áudio e para receber informação de representação que indica se um ou mais sinais de objetos de áudio deverão ser amplificados ou atenuados, e

um processador de sinal (120) para gerar um ou mais canais de saída de áudio do sinal de saída de áudio,

em que a interface receptora (110) está configurada para receber um sinal *downmix*, que compreende um ou mais canais *downmix* como o sinal de entrada de áudio, em que um ou mais canais *downmix* compreendem os sinais de objetos de áudio e em que o número de um ou mais canais *downmix* é menor que o número dos sinais de objetos de áudio,

em que a interface receptora (110) está configurada para receber informação *downmix* indicando como os sinais de objetos de áudio são misturados em um ou mais canais *downmix*, em que a interface receptora (110) está configurada para receber um ou mais sinais de objetos de áudio de *bypass* adicionais, em que receber um ou mais sinais de objetos de áudio de *bypass* adicionais não são misturados com o sinal *downmix*,

em que a interface receptora (110) está configurada para receber a informação de percepção sonora que indica informação sobre a percepção sonora dos sinais de objetos de áudio, os quais se encontram misturados com o sinal *downmix* e indicam informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais que não se encontram misturados com o sinal *downmix*,

em que o processador de sinal (120) está configurado para determinar um valor de compensação da percepção sonora, dependendo da

informação sobre a percepção sonora dos sinais de objetos de áudio que se encontram misturados com o sinal *downmix* e dependendo da informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais os quais não se encontram misturados com o sinal, e

em que o processador de sinal (120) está configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação *downmix*, dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

2.Descodificador de acordo com a reivindicação 1, **caracterizado por** o processador de sinal (120) estar configurado para gerar um ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora, de tal modo que uma percepção sonora do sinal de saída de áudio é igual a uma percepção sonora do sinal de entrada de áudio ou de tal modo que a percepção sonora do sinal de saída é mais próxima da percepção sonora do sinal de entrada de áudio comparativamente a uma percepção sonora de um sinal de áudio modificado que resultaria da modificação de um sinal de entrada de áudio ao amplificar ou atenuar os sinais de objetos de áudio de acordo com a informação de representação.

3.Descodificador de acordo com a reivindicação 2, **caracterizado por** o processador de sinal (120) estar configurado para gerar o sinal de áudio modificado, modificando o sinal de entrada de áudio ao amplificar ou atenuar os sinais de objetos de áudio do sinal de entrada de áudio, de acordo com a informação de representação, e

por o processador de sinal (120) estar configurado para gerar o sinal de saída de áudio, aplicando o valor de compensação de percepção sonora ao sinal de áudio modificado, de modo a que a percepção sonora do sinal de saída de áudio é igual à percepção sonora do sinal de entrada de áudio ou de

modo a que a percepção sonora do sinal de saída de áudio seja mais próxima da percepção sonora do sinal de entrada de áudio comparativamente ao sinal de áudio modificado.

4.Descodificador de acordo com qualquer uma das reivindicações 1 a 3, **caracterizado por** cada sinal de objetos de áudio do sinal de entrada de áudio ser atribuído a exatamente um grupo de dois ou mais grupos, em que cada dos dois ou mais grupos compreende um ou mais sinais de objetos de áudio do sinal de entrada de áudio,

por a interface receptora (110) estar configurada para receber um valor de percepção sonora para cada grupo dos dois ou mais grupos como informação de percepção sonora,

por o processador de sinal (120) estar configurado para determinar o valor de compensação de percepção sonora dependendo do valor de percepção sonora de cada dos dois ou mais grupos, e

por o processador de sinal (120) estar configurado para gerar o ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo do valor de compensação de percepção sonora.

5.Descodificador de acordo com qualquer uma das reivindicações 1 a 4, **caracterizado por**, pelo menos, um grupo dos dois ou mais grupos compreender dois ou mais dos sinais de objetos áudio.

6.Descodificador de acordo com qualquer uma das reivindicações 1 a 5, **caracterizado por** cada dos sinais de objetos de áudio do sinal de entrada de áudio ser atribuído a exatamente um grupo de exatamente dois grupos como os dois ou mais grupos,

por cada um dos sinais de objetos de áudio do sinal de entrada de áudio ser atribuído a um grupo de objetos de primeiro plano dos dois grupos exatamente ou a um grupo de objetos de segundo plano dos dois grupos exatamente,

por a interface receptora (110) estar configurada para receber o valor da percepção sonora do grupo de objetos de primeiro plano,

por a interface receptora (110) estar configurada para receber o valor da percepção sonora do grupo de objetos de segundo plano,

por o processador de sinal (120) estar configurado para determinar o valor de compensação de percepção sonora dependendo do valor da percepção sonora do grupo de objetos de primeiro plano e dependendo do valor da percepção sonora do grupo de objetos de segundo plano, e

por o processador de sinal (120) estar configurado para gerar o ou mais canais de saída de áudio do sinal de saída de áudio a partir do sinal de entrada de áudio dependendo do valor de compensação de percepção sonora.

7. Descodificador de acordo com a reivindicação 6, **caracterizado por** o processador de sinal (120) estar configurado para determinar um valor de compensação de percepção sonora ΔL de acordo com a fórmula

$$\Delta L = 10 \log_{10} \frac{m_{FGO}^2 10^{K_{FGO}/10} + m_{BGO}^2 10^{K_{BGO}/10}}{10^{K_{FGO}/10} + 10^{K_{BGO}/10}}$$

em que KFGO indica o valor da percepção sonora do grupo de objetos de primeiro plano,

em que KBGO indica o valor da percepção sonora do grupo de objetos de segundo plano,

em que mFGO indica um ganho de representação do grupo de objetos de primeiro plano, e

em que mBGO indica um ganho de representação do grupo de objetos de segundo plano.

8. Descodificador de acordo com a reivindicação 6, **caracterizado por** o processador de sinal (120) estar configurado para determinar um valor de compensação de percepção sonora ΔL de acordo com a fórmula

$$\Delta L(A, B) = 10 \log_{10} \frac{g_{FGO}^2 10^{L_{FGO}/10} + g_{BGO}^2 10^{L_{BGO}/10}}{10^{L_{FGO}/10} + 10^{L_{BGO}/10}}$$

em que LFGO indica o valor da percepção sonora do grupo de objetos de primeiro plano,

em que LBGO indica o valor da percepção sonora do grupo de objetos de segundo plano,

em que gFGO indica um ganho de representação do grupo de objetos de primeiro plano, e

em que gBGO indica um ganho de representação do grupo de objetos de segundo plano.

9.Codificador, **caracterizado por** compreender:

uma unidade de codificação baseada em objetos (210; 710) para codificar uma pluralidade de sinais de objetos de áudio para obter um sinal de áudio codificado que compreende a pluralidade de sinais de objetos de áudio e

uma unidade de codificação da percepção sonora do objeto (220; 720; 820) para codificar informação da percepção sonora nos sinais de objetos de áudio,

em que a informação da percepção sonora compreende um ou mais valores de percepção sonora, em que cada do ou de mais valores de percepção sonora dependem de um ou mais dos sinais de objetos de áudio,

em que unidade de codificação baseada em objetos (210; 710) é configurada para receber os sinais de objetos de áudio, em que cada dos sinais de objetos de áudio é atribuído a exatamente um de dois ou mais grupos, em que cada dos dois ou mais grupos compreende um ou mais sinais de objetos de áudio,

em que a unidade de codificação baseada em objetos (210; 710) está configurada para *downmix* dos sinais de objetos de áudio, sendo compreendida pelos dois ou mais grupos, para obter um sinal *downmix* que compreenda um ou mais canais de áudio *downmix* como o sinal de áudio codificado, em que o número de um ou mais canais *downmix* é inferior ao número de sinais de objetos de áudio que é compreendido por dois ou mais grupos,

em que a unidade de codificação da percepção sonora do objeto (220; 720; 820) é designada para receber um ou mais sinais de objetos de

áudio de *bypass* adicionais, em que cada de um ou mais sinais de objetos de áudio de *bypass* adicionais é atribuído a um terceiro grupo, em que cada de um ou mais sinais de objetos de áudio de *bypass* adicionais não é compreendido pelo primeiro grupo e não é compreendido pelo segundo grupo, em que a unidade de codificação com base no objeto (210; 710) é configurada para não proceder ao *downmix* do um ou mais sinais de objetos de áudio de *bypass* adicionais com o sinal *downmix*, e

em que a unidade de codificação da percepção sonora do objeto (220; 720; 820) está configurada para determinar um primeiro valor da percepção sonora, um segundo valor da percepção sonora e um terceiro valor da percepção sonora da informação da percepção sonora, o primeiro valor da percepção sonora indica uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo, o segundo valor da percepção sonora indica uma percepção sonora total de um ou mais sinais de objetos de áudio do segundo grupo e o terceiro valor da percepção sonora indica uma percepção sonora total de um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo, ou está configurado para determinar um primeiro valor da percepção sonora e um segundo valor da percepção sonora da informação da percepção sonora, o primeiro valor da percepção sonora indicando uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo e o segundo valor da percepção sonora indicando um total da percepção sonora de um ou mais sinais de objetos e áudio do segundo grupo e um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo.

10.Codificador de acordo com a reivindicação 9, **caracterizado por** os dois ou mais grupos serem exatamente dois grupos,

em que cada dos sinais de objetos de áudio é atribuído a exatamente um dos exatamente dois grupos, em que cada dos exatamente dois grupos compreende um ou mais dos sinais de objetos de áudio,

em que a unidade de codificação com base no objeto (210; 710) está configurada para *downmix* os sinais de objetos de áudio, sendo compreendida

pelos exatamente dois grupos, para obter um sinal *downmix* que compreenda um ou mais canais de áudio *downmix* como o sinal de áudio codificado, em que o número de um ou mais canais *downmix* é inferior ao número dos sinais de objetos de áudio sendo compreendido pelos exatamente dois grupos.

11. Sistema **caracterizado por** compreender:

um codificador (310) tal como descrito na reivindicação 9 ou 10 para codificar uma pluralidade de sinais de objetos de áudio para obter um sinal de áudio codificado que compreenda uma pluralidade de sinais de objetos de áudio, e

um descodificador (320) de acordo com uma das reivindicações 1 a 8 para gerar um sinal de saída de áudio que compreenda um ou mais canais de saída de áudio,

em que o descodificador (320) está configurado para receber o sinal de áudio codificado como um sinal de entrada de áudio e para receber a informação da percepção sonora

em que o descodificador (320) está configurado para receber ainda a informação de representação,

em que o descodificador (320) está configurado para determinar um valor de compensação da percepção sonora dependendo da informação da percepção sonora e dependendo da informação de representação, e

em que o descodificador (320) está configurado para gerar um ou mais canais de saída de áudio do sinal de saída e áudio a partir do sinal de entrada de áudio dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

12. Método para gerar um sinal de saída e áudio que **caracterizado por** compreender um ou mais canais de saída de áudio, em que o método inclui:

recepção de um sinal de entrada de áudio compreendendo uma pluralidade de sinais de objetos de áudio,

receber a informação da percepção sonora que indica informação

sobre a percepção sonora dos sinais de objetos de áudio, os quais se encontram misturados com o sinal *downmix* e indicam informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais que não se encontram misturados com o sinal *downmix*, e

receber informação de representação indicando se um ou mais dos sinais de objetos de áudios deverão ser amplificados ou atenuados,

receber um sinal *downmix* que compreende um ou mais canais *downmix* como o sinal de entrada de áudio, em que um ou mais canais *downmix* compreendem os sinais de objetos de áudio e em que o número de um ou mais canais *downmix* é menor que o número dos sinais de objetos de áudio,

receber informação *downmix* indicando como os sinais de objetos de áudio são misturados em um ou mais canais *downmix*,

receber um ou mais sinais de objetos de áudio de *bypass* adicionais, em que um ou mais sinais de objetos de áudio de *bypass* adicionais não são misturados com o sinal *downmix*,

determinar um valor de compensação da percepção sonora dependendo da informação sobre a percepção sonora dos sinais de objetos de áudio os quais se encontram misturados com o sinal *downmix* e dependendo da informação sobre a percepção sonora de um ou mais sinais de objetos de áudio de *bypass* adicionais os quais não se encontram misturados com o sinal, e

gerar um ou mais canais de saída áudio do sinal de saída áudio a partir do sinal de entrada áudio dependendo da informação *downmix*, dependendo da informação de representação e dependendo do valor de compensação da percepção sonora.

13.Método de codificação que **caracterizado por** incluir:

codificação de um sinal de entrada de áudio compreendendo uma pluralidade de sinais de objetos de áudio, e

codificação de informação da percepção sonora sobre os sinais de

objetos de áudio, em que a informação de ruído compreende um ou mais valores da percepção sonora, em que cada do ou de mais valores da percepção sonora dependem em um ou mais dos sinais de objetos de áudio,

em que cada dos sinais de objetos de áudio é atribuído a exatamente um de dois ou mais grupos, em que cada dos dois ou mais grupos compreende um ou mais dos sinais de objetos de áudio,

em que a codificação da informação da percepção sonora sobre os sinais de objetos de áudio é realizada por *downmix* dos sinais de objetos de áudio, sendo compreendida pelos dois ou mais grupos, para obter um sinal *downmix* compreendendo um ou mais canais de áudio *downmix* como o sinal de áudio codificado, em que o número de um ou mais canais *downmix* é inferior ao número dos sinais de objetos de áudio sendo compreendido pelos dois ou mais grupos,

em que cada de um ou mais sinais de objetos de áudio de *bypass* adicionais é atribuído a um terceiro grupo, em que cada de um ou mais sinais de objetos de áudio de *bypass* adicionais não é compreendida pelo primeiro grupo e não é compreendida pelo segundo grupo,

em que a codificação da informação de ruído sobre os sinais de objetos de áudio é realizada por não *downmix* um ou mais sinais de objetos de áudio de *bypass* adicionais com o sinal *downmix*, e

em que a codificação da informação da percepção sonora sobre os sinais de objetos de áudio é realizada, determinando um primeiro valor da percepção sonora, um segundo valor da percepção sonora e um terceiro valor da percepção sonora da informação da percepção sonora, o primeiro valor da percepção sonora indica uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo, o segundo valor da percepção sonora indica um ruído total de um ou mais sinais de objetos de áudio do segundo grupo e o terceiro valor da percepção sonora indica uma percepção sonora total de um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo, ou está configurado para determinar um primeiro valor da percepção

sonora e um segundo valor da percepção sonora da informação da percepção sonora, indicando o primeiro valor da percepção sonora uma percepção sonora total de um ou mais sinais de objetos de áudio do primeiro grupo e o segundo valor de ruído indicando um total de percepção sonora de um ou mais sinais de objetos de áudio do segundo grupo e um ou mais sinais de objetos de áudio de *bypass* adicionais do terceiro grupo.

14. Método de acordo com a reivindicação 13, **caracterizado por** os dois ou mais grupos serem exatamente dois grupos,

em que cada dos sinais de objetos de áudio é atribuído a exatamente um dos exatamente dois grupos, em que cada dos exatamente dois grupos compreende um ou mais dos sinais de objetos de áudio,

em que a codificação da informação da percepção sonora nos sinais de objetos de áudio é realizada ao *downmix* os sinais de objetos de áudio, sendo compreendidos pelos exatamente dois grupos, para obter um sinal *downmix* que compreenda um ou mais canais de áudio *downmix* como o sinal de áudio codificado, em que o número de um ou mais canais *downmix* é inferior ao número dos sinais de objetos de áudio sendo compreendido pelos exatamente dois grupos.

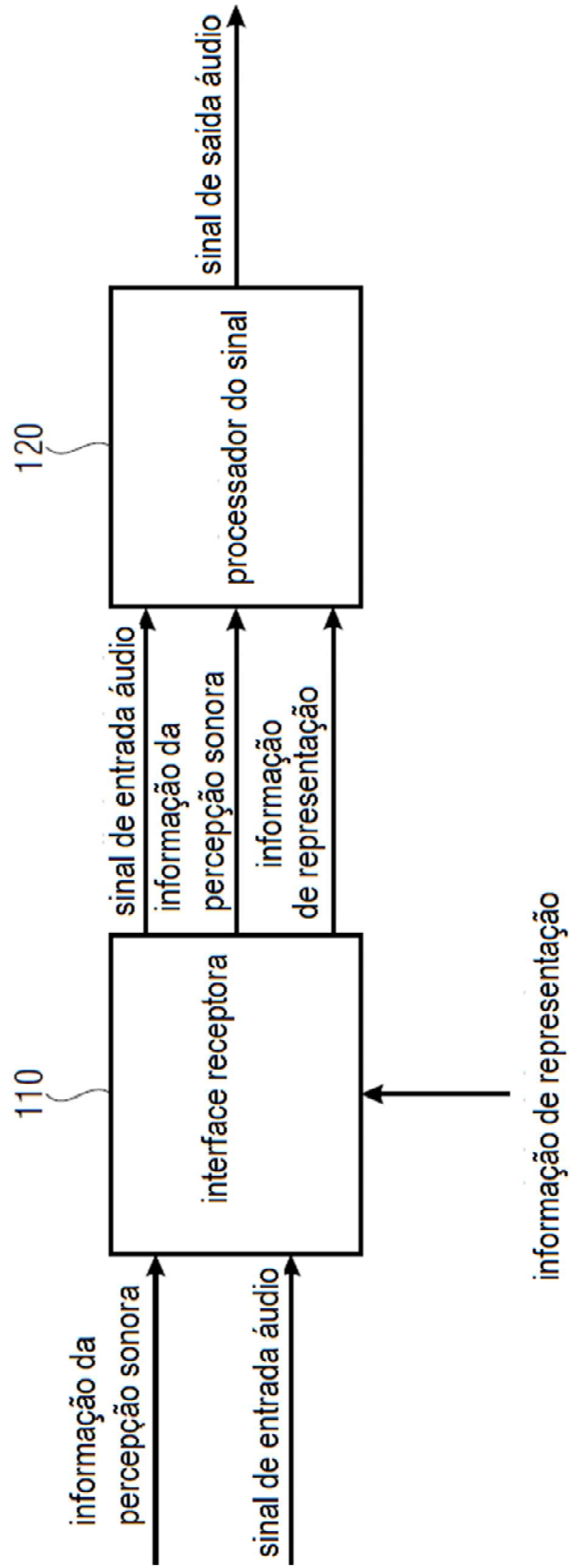


FIG. 1

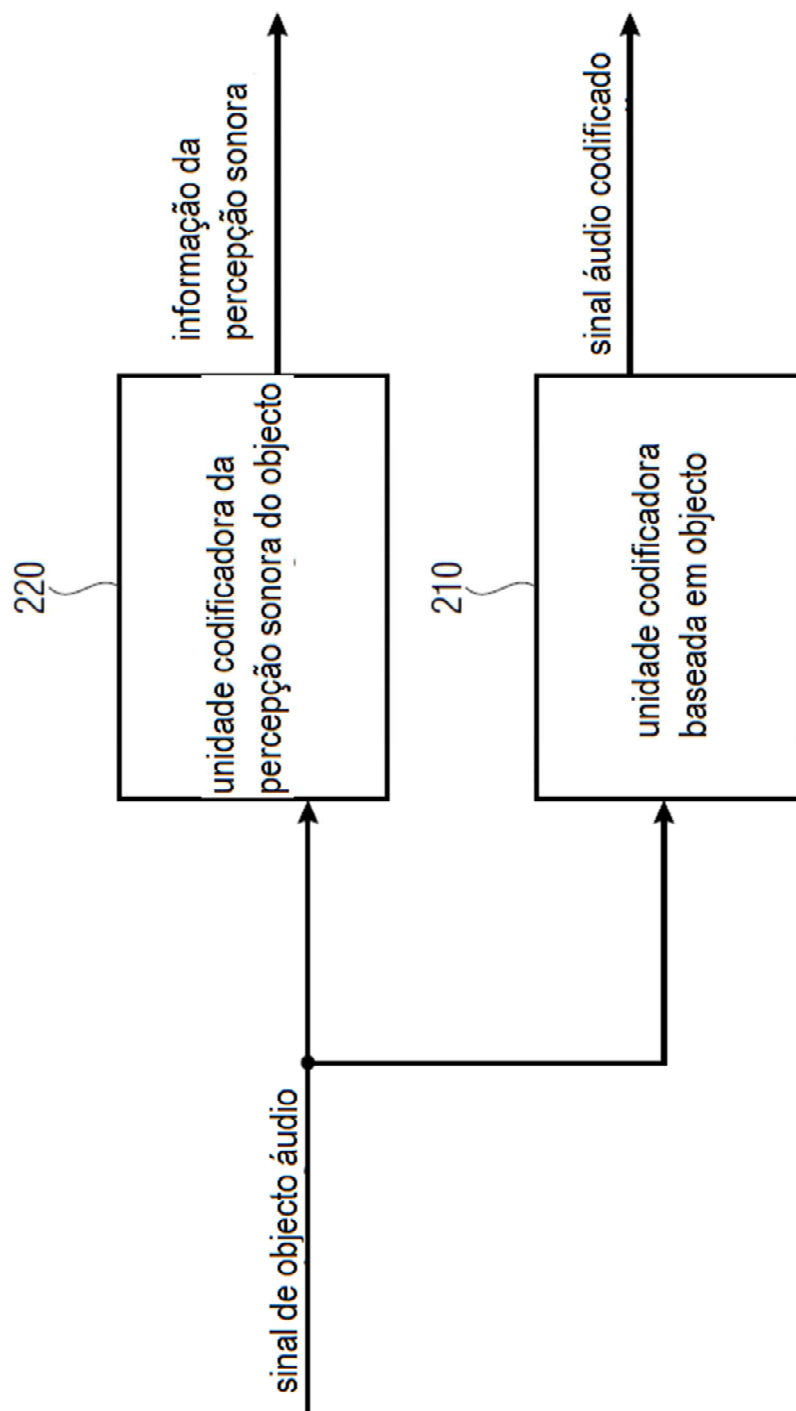


FIG. 2

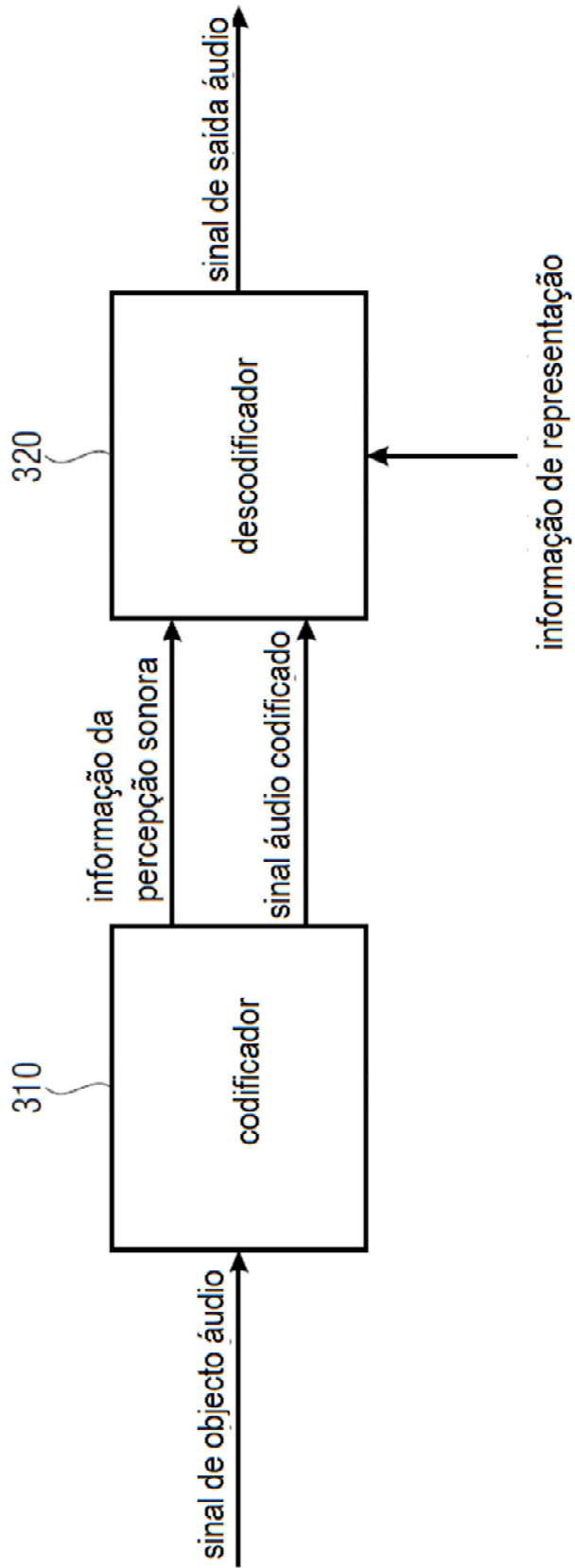


FIG. 3

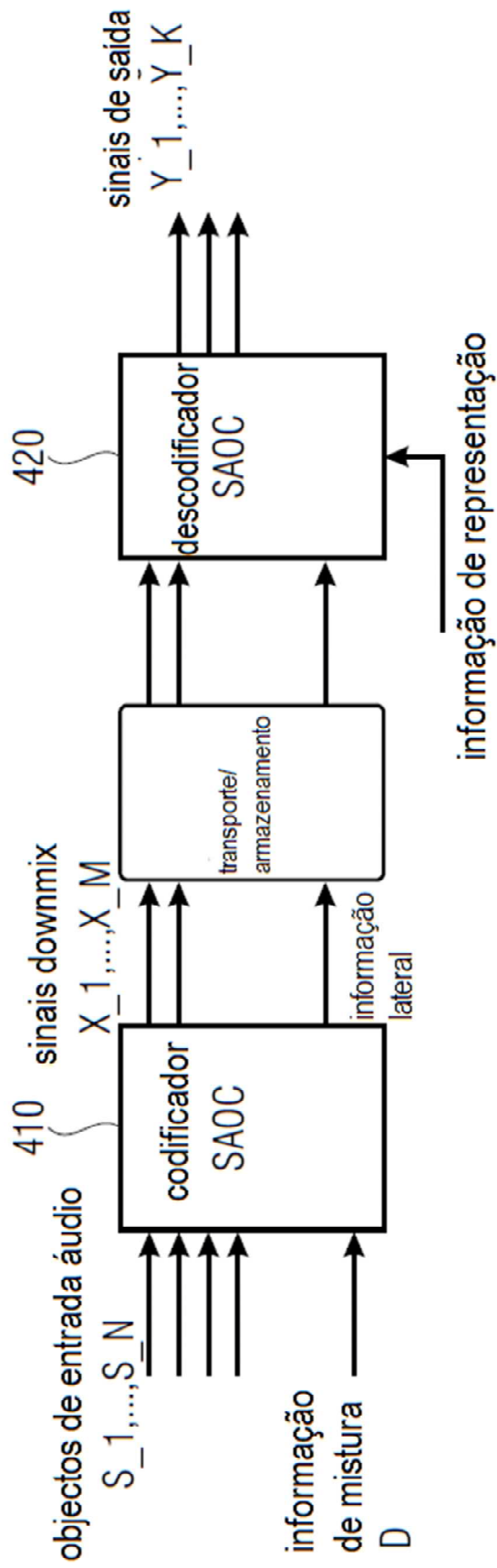


FIG. 4

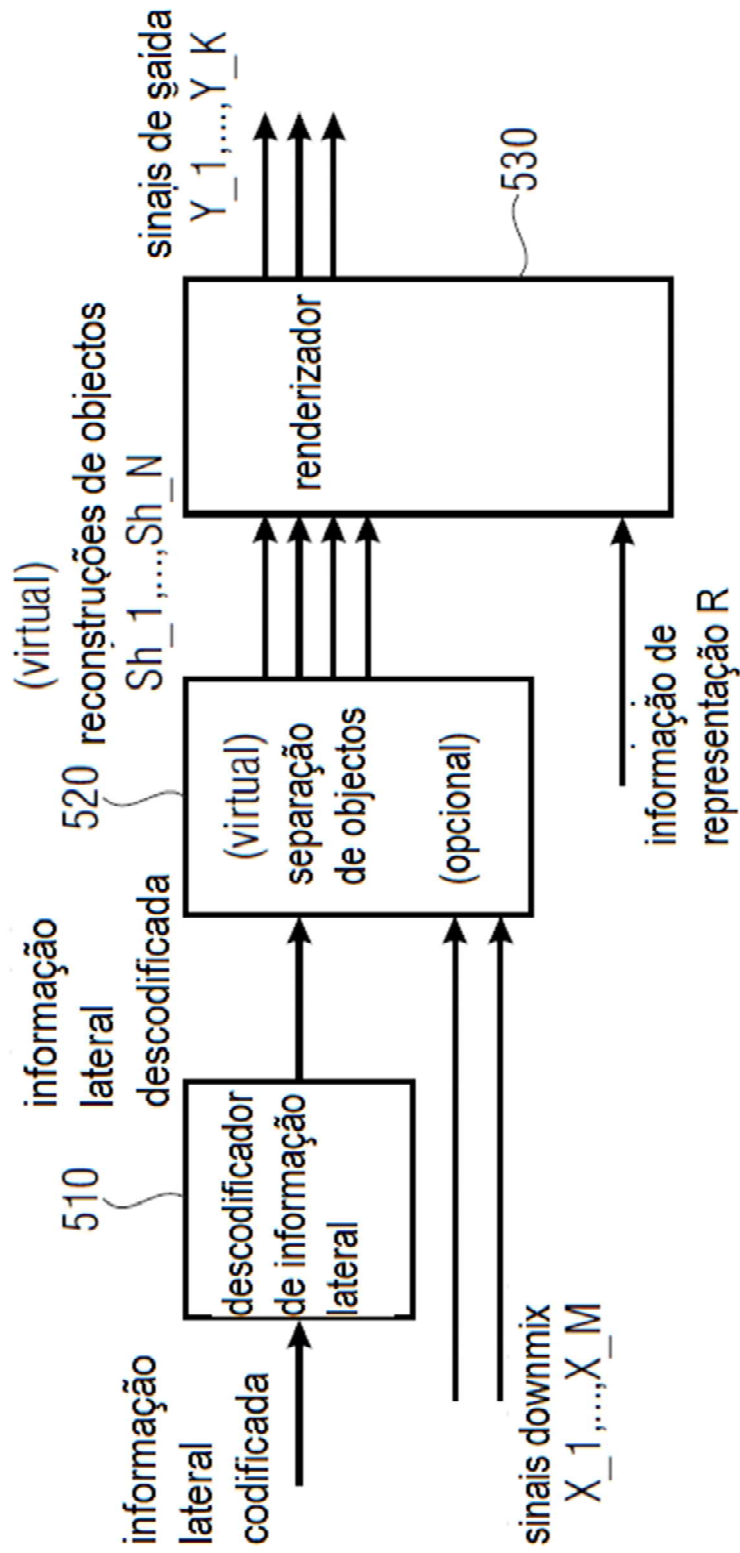


FIG. 5

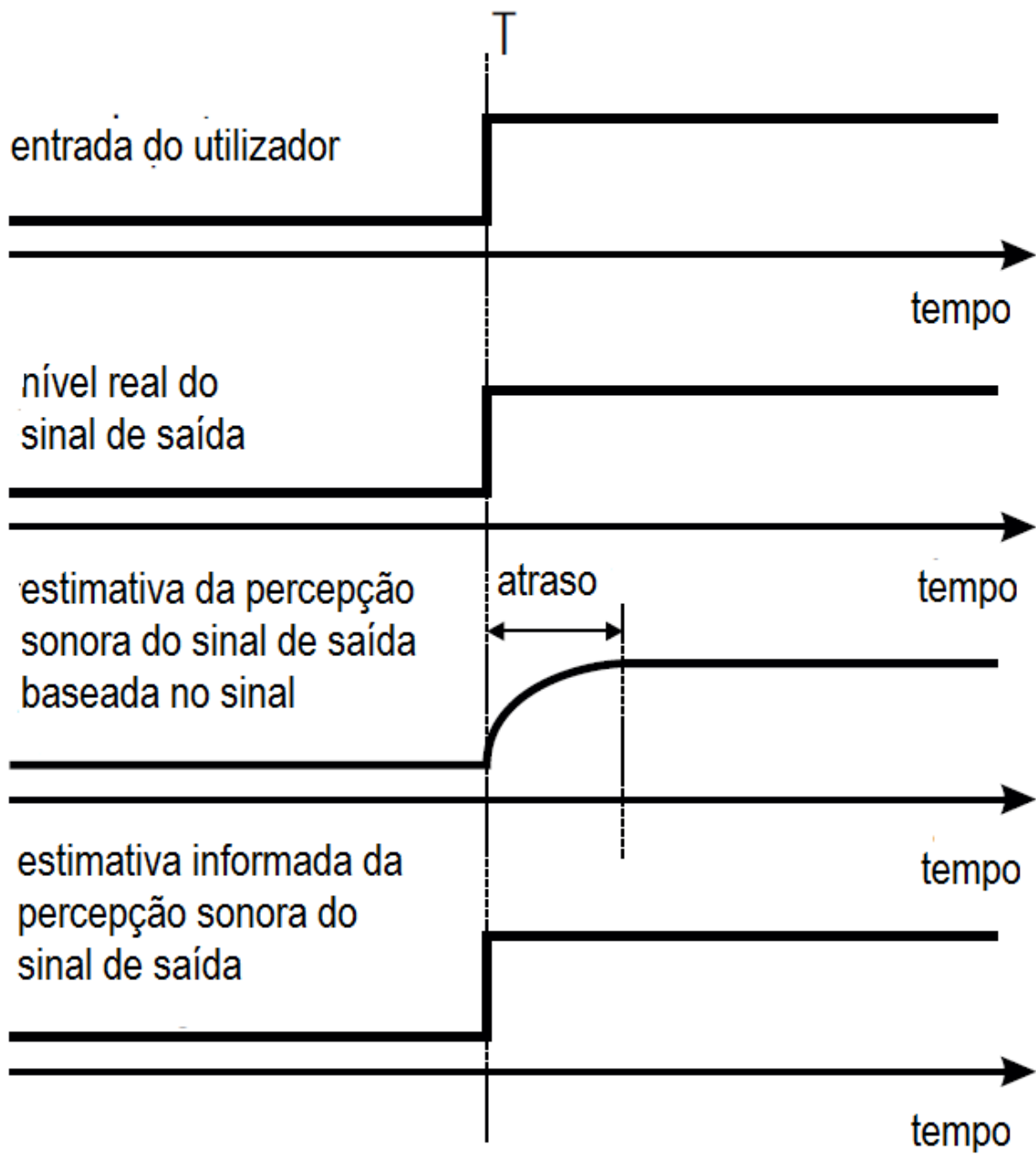


FIG. 6

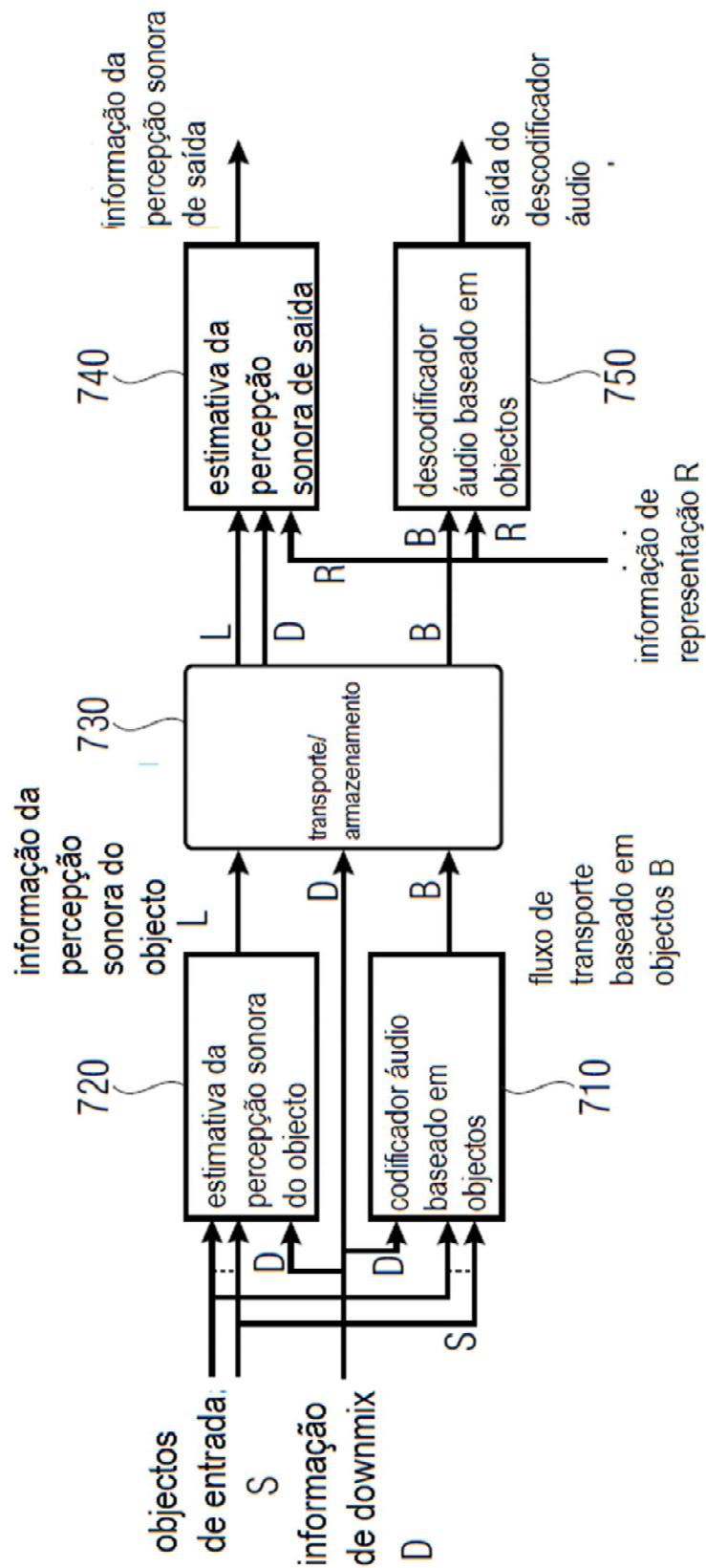


FIG. 7

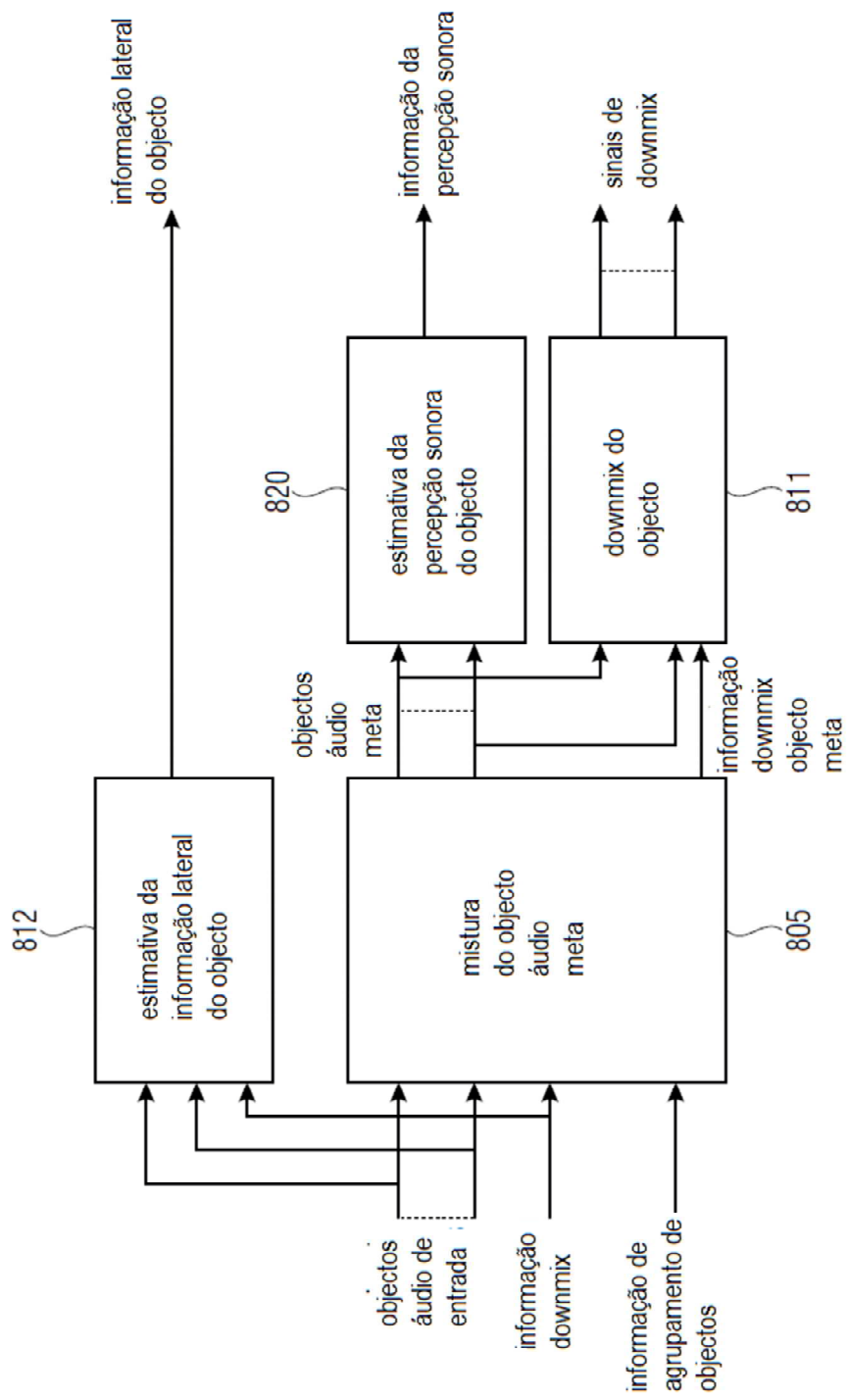


FIG. 8

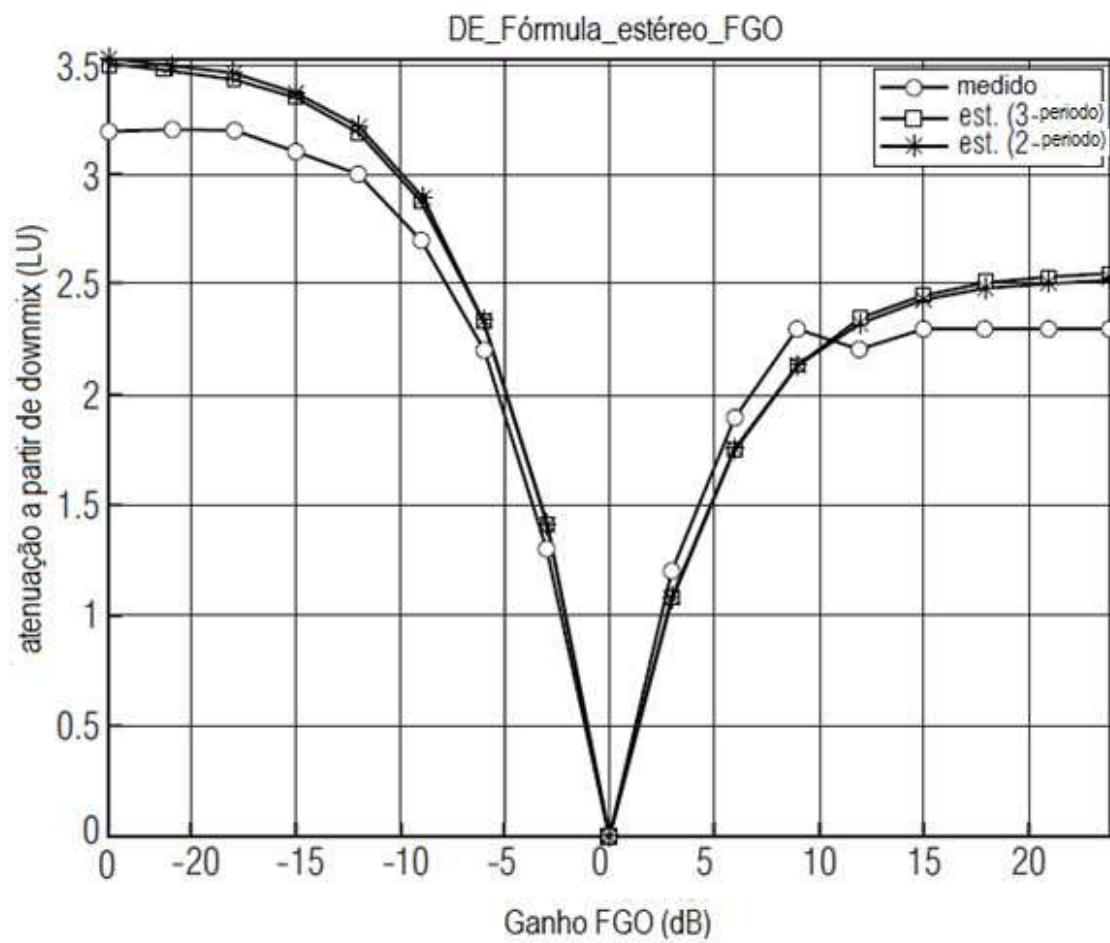


FIG. 10

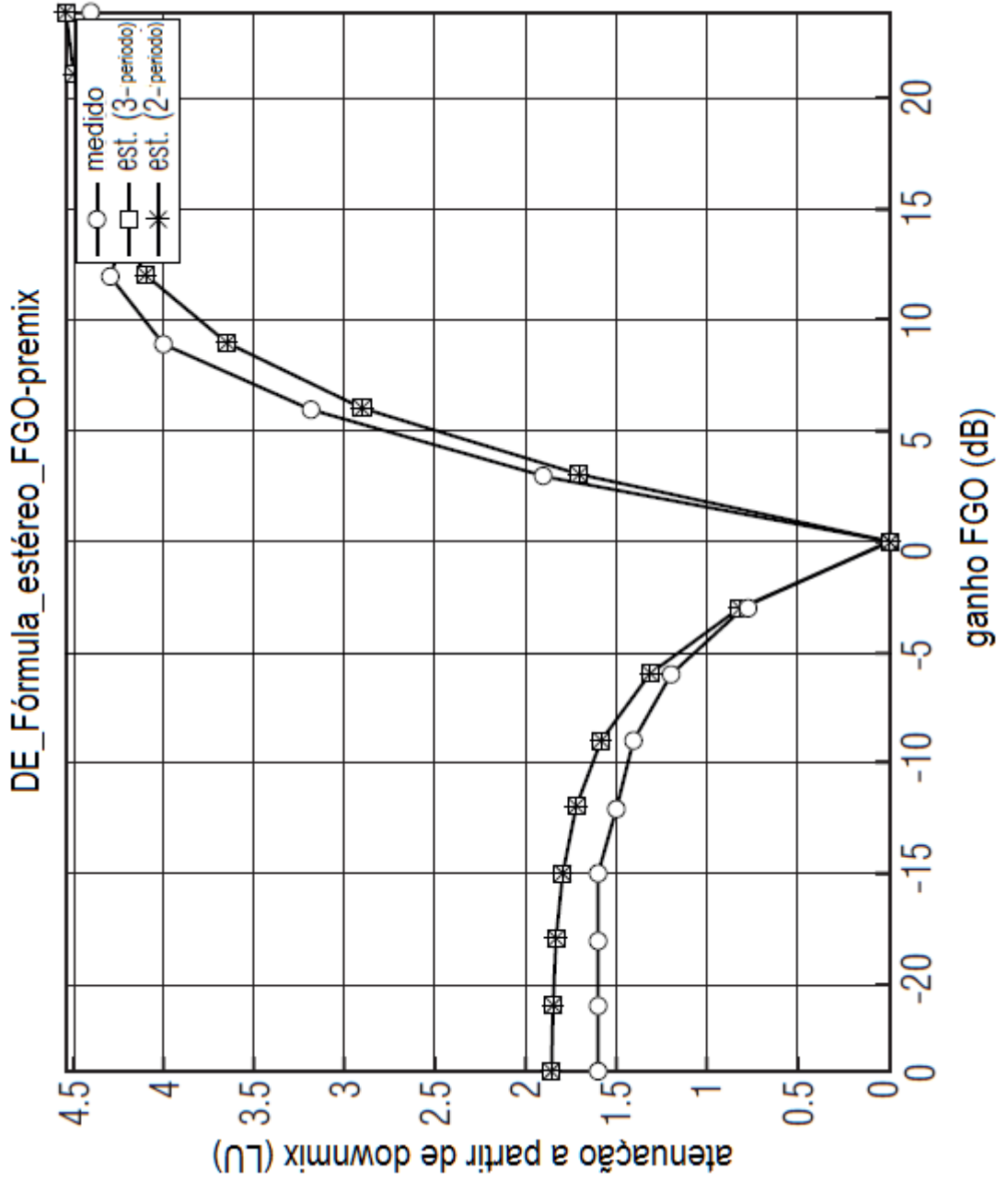


FIG. 11

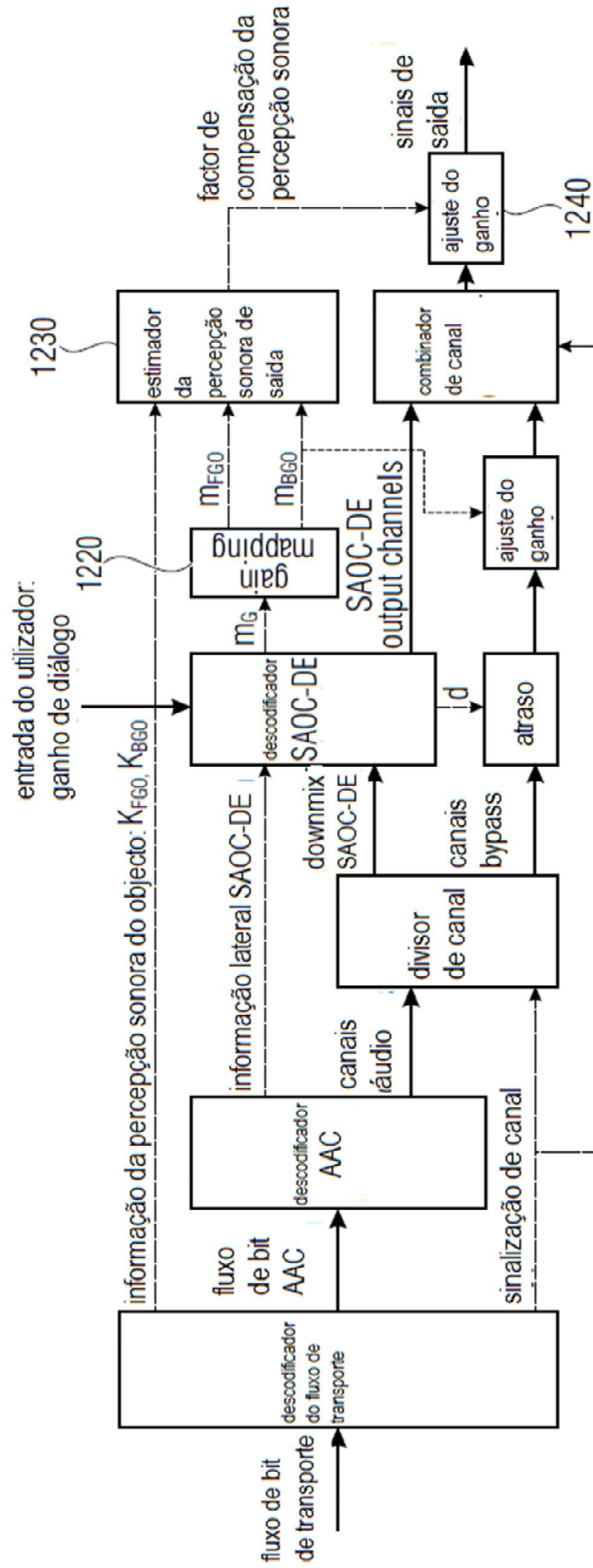


FIG. 12