



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

(11) BR 112016001250-0 B1



(22) Data do Depósito: 17/07/2014

(45) Data de Concessão: 26/07/2022

(54) Título: DECODIFICADOR DE ÁUDIO DE MULTICANAL, CODIFICADOR DE ÁUDIO DE MULTICANAL, MÉTODOS, E REPRESENTAÇÃO DE ÁUDIO CODIFICADA USANDO UMA DESCORRELAÇÃO DOS SINAIS DE ÁUDIO RENDERIZADOS

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

(30) Prioridade Unionista: 18/10/2013 EP 13189345.5; 22/07/2013 EP 13177374.9; 25/03/2014 EP 14161611.0.

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

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

(86) Pedido PCT: PCT EP2014065397 de 17/07/2014

(87) Publicação PCT: WO 2015/011015 de 29/01/2015

(85) Data do Início da Fase Nacional: 19/01/2016

(57) Resumo: DECODIFICADOR DE ÁUDIO DE MULTICANAL, CODIFICADOR DE ÁUDIO DE MULTICANAL, MÉTODOS, PROGRAMA DE COMPUTADOR E REPRESENTAÇÃO DE ÁUDIO CODIFICADA USANDO UMA DESCORRELAÇÃO DOS SINAIS DE ÁUDIO RENDERIZADOS Decodificador de áudio multicanal para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada é configurado para renderizar uma pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, na dependência de um ou mais parâmetros de renderização, para obter uma pluralidade de sinais de áudio renderizados. O decodificador de áudio multicanal é configurado para derivar um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados, e para combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída. Um codificador de áudio de multicanal fornece um parâmetro do método de descorrelação para controlar um decodificador de áudio.

Relatório Descritivo da Patente de Invenção para: **“DECODIFICADOR DE ÁUDIO DE MULTICANAL, CODIFICADOR DE ÁUDIO DE MULTICANAL, MÉTODOS, E REPRESENTAÇÃO DE ÁUDIO CODIFICADA USANDO UMA DESCORRELAÇÃO DOS SINAIS DE ÁUDIO RENDERIZADOS”**.

Descrição

Campo técnico

[001] As modalidades de acordo com a invenção são relacionadas a um decodificador de áudio multicanal para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada.

[002] Modalidades adicionais de acordo com a invenção são relacionadas a um codificador de áudio de multicanal para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada.

[003] Modalidades adicionais de acordo com a invenção são relacionadas a um método para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada.

[004] Modalidades adicionais de acordo com a invenção são relacionadas a um método para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada.

[005] Modalidades adicionais de acordo com a invenção são relacionadas a um programa de computador para realizar um dos referidos métodos.

[006] Modalidades adicionais de acordo com a invenção são relacionadas a uma representação de áudio codificada.

[007] De um modo geral, as modalidades de acordo com a presente invenção são relacionadas a um conceito de descorrelação para sistemas de codificação de objeto de áudio paramétrica de downmix/upmix de multicanal.

Antecedentes da invenção

[008] Nos últimos anos, a demanda por armazenamento e transmissão de conteúdo de áudio tem aumentado constantemente. Além disso, os requisitos de qualidade para o armazenamento e transmissão de conteúdo de áudio também têm aumentado de forma constante. Por conseguinte, os conceitos

para a codificação e decodificação de conteúdo de áudio foram melhorados.

[009] Por exemplo, o assim chamado "Advanced Audio Coding" (AAC) tem sido desenvolvido, o qual é descrito, por exemplo, no padrão internacional ISO/IEC 13818-7: 2003. Além disso, algumas extensões espaciais foram criadas, como por exemplo, o chamado conceito "MPEG Surround", que é descrito, por exemplo, no padrão internacional ISO/IEC 23003-1: 2007. Além disso, as melhorias adicionais para codificação e decodificação de informação espacial de sinais de áudio são descritas na norma internacional ISO / IEC 23003-2: 2010, que se refere à chamada "Spatial Audio Object Coding".

[010] Além disso, um conceito de codificação/decodificação de áudio comutável que oferece a possibilidade de codificar ambos os sinais de áudio gerais e sinais de fala com boa eficiência de codificação e lidar com sinais de áudio multicanal é definido na norma internacional ISO/IEC 23003-3: 2012, que descreve o chamado conceito de "Unified Speech and Audio Coding".

[011] Além disso, outros conceitos convencionais são descritos nas referências, as quais são mencionados no final da presente descrição.

[012] No entanto, existe um desejo de fornecer um conceito ainda mais avançado para uma codificação e decodificação eficiente de cenas de áudio tridimensionais.

Sumário da invenção

[013] Uma modalidade de acordo com a invenção cria um decodificador de áudio multicanal para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada. O decodificador de áudio multicanal é configurado para renderizar uma pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, na dependência de um ou mais parâmetros de renderização, para obter uma pluralidade de sinais de áudio renderizados. O decodificador de áudio multicanal é configurado para derivar um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados. Além disso, o decodificador de áudio multicanal é configurado para combinar os sinais de áudio renderizados, ou uma versão

reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída.

[014] Esta modalidade de acordo com a invenção é baseada na descoberta de que a qualidade de áudio pode ser melhorada em um decodificador de áudio multicanal pela derivação de um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados, que são obtidos com base em uma pluralidade de sinais de áudio decodificados, e combinação dos sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída. Verificou-se que é mais eficiente ajustar as características de correlação, ou as características de covariância, dos sinais de áudio de saída pela adição dos sinais descorrelacionados após a renderização quando comparado à adição de sinais descorrelacionados antes da renderização ou durante a renderização. Verificou-se que este conceito é mais eficiente, em casos gerais, nos quais existem mais sinais de áudio decodificados, que são inseridos para a renderização, do que sinais de áudio renderizados, porque mais descorrelacionadores seriam necessários se a descorrelação fosse realizada antes da renderização ou durante a renderização. Além disso, verificou-se que são frequentemente fornecidos artefatos quando os sinais descorrelacionados são adicionados aos sinais de áudio decodificados antes da renderização, porque a renderização tipicamente traz junto uma combinação de sinais de áudio decodificados. Portanto, o conceito de acordo com a presente modalidade da invenção supera as abordagens convencionais, nas quais sinais descorrelacionados são adicionados antes da renderização. Por exemplo, é possível estimar diretamente as características de correlação ou características de covariância desejadas dos sinais renderizados, e adaptar o fornecimento de sinais de áudio descorrelacionados aos sinais efetivamente fornecidos, o que resulta em uma melhor compensação entre eficiência e qualidade de áudio, e frequentemente ainda resulta em um aumento da eficiência e uma melhor qualidade ao mesmo tempo.

[015] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para obter os sinais de áudio decodificados, que são renderizados para obter a pluralidade de sinais de áudio renderizados, usando uma reconstrução paramétrica. Foi constatado que o conceito de acordo com a presente invenção traz consigo vantagens em combinação com uma reconstrução paramétrica de sinais de áudio, sendo que a reconstrução paramétrica é, por exemplo, baseada em uma informação de lado que descreve sinais de objeto e/ou uma relação entre sinais de objeto (sendo que os sinais de objeto podem constituir os sinais de áudio decodificados). Por exemplo, pode haver um número relativamente grande de sinais de objeto (sinais de áudio decodificados) em tal conceito, e verificou-se que a aplicação do descorrelação com base nos sinais de áudio renderizados é particularmente eficiente e evita artefatos em tal cenário.

[016] Em uma modalidade preferida, os sinais de áudio decodificados são sinais de objeto reconstruídos (por exemplo, sinais de objeto parametricamente reconstruídos) e o decodificador de áudio multicanal é configurado para derivar os sinais de objeto reconstruídos a partir de um ou mais sinais de downmix usando uma informação de lado. Portanto, a combinação dos sinais de áudio renderizados com um ou mais sinais de áudio descorrelacionados, que são baseados nos sinais de áudio renderizados, permite uma reconstrução eficiente das características de correlação ou características de covariância nos sinais de áudio de saída, mesmo se há um número comparativamente grande de sinais de objeto reconstruídos (o qual pode ser maior do que um número de sinais de áudio renderizados ou sinais de áudio de saída).

[017] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para derivar coeficientes não misturados a partir da informação do lado e aplicar os coeficientes não misturados para derivar (parametricamente) os sinais de objeto reconstruídos a partir de um ou mais sinais de downmix usando os coeficientes não misturados. Portanto, os sinais de entrada para a renderização podem ser derivados de uma informação de

lado, que pode, por exemplo, ser uma informação de lado relacionada a objeto (como, por exemplo, uma informação de correlação inter-objeto ou uma informação de diferença de nível de objeto, sendo que o mesmo resultado pode ser obtido usando energias absolutas).

[018] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para, pelo menos, parcialmente conseguir as características de correlação desejadas ou características de covariância dos sinais de áudio de saída. Verificou-se que a combinação dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, que são derivados dos sinais de áudio renderizados, permite um ajuste (ou reconstrução) das características de correlação ou características de covariância desejadas. Além disso, verificou-se que é importante para a impressão aditiva ter as características de correlação ou características de covariância adequadas no sinal de áudio de saída, e que isto pode ser conseguido da melhor maneira pela modificação dos sinais de áudio renderizados usando os sinais de áudio descorrelacionado. Por exemplo, quaisquer degradações, que são causadas nos estágios de processamento anteriores, também podem ser consideradas combinando os sinais de áudio renderizados e os sinais de áudio descorrelacionados com base nos sinais de áudio renderizados.

[019] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para, pelo menos, parcialmente compensar uma perda de energia durante uma reconstrução paramétrica dos sinais de áudio decodificados, que são renderizados para obter a pluralidade de sinais de áudio renderizados. Verificou-se que a aplicação pós-renderização dos sinais de áudio descorrelacionados permite a correção das imperfeições do sinal que são causadas por um processamento antes da renderização, por exemplo, pela reconstrução paramétrica dos sinais

de áudio decodificados. Consequentemente, não é necessário reconstruir as características de correlação ou características de covariância dos sinais de áudio decodificados, que são inseridas na renderização, com alta precisão. Isto simplifica a reconstrução dos sinais de áudio decodificados e, por conseguinte, traz consigo uma alta eficiência.

[020] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para determinar características de correlação desejadas das características de covariância dos sinais de áudio de saída. Além disso, o decodificador de áudio multicanal é configurado para ajustar uma combinação dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída, de modo que as características de correlação ou características de covariância dos sinais de áudio de saída obtidas se aproximam ou são iguais às características de correlação ou características de covariância. O cálculo (ou determinação) das características de correlação ou características de covariância desejadas dos sinais de áudio de saída (que devem ser conseguidas após a combinação dos sinais de áudio renderizados com os sinais de áudio descorrelacionado), é possível para ajustar as características de correlação ou características de covariância em um estágio posterior de processamento, o que, por sua vez, permite uma reconstrução relativamente precisa. Portanto, uma impressão auditiva espacial dos sinais de áudio de saída é bem adaptada a uma impressão auditiva desejada.

[021] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para determinar as características de correlação ou características de covariância na dependência de uma informação de renderização que descreve uma renderização da pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, para obter a pluralidade de sinais de áudio renderizados. Ao considerar o processo de renderização na determinação das características de correlação desejadas ou as características de covariância desejadas, é possível conseguir uma

informação precisa para ajustar a combinação dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, o que traz consigo a possibilidade de ter sinais de áudio de saída que correspondem a uma impressão auditiva desejada.

[022] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para determinar as características de correlação ou características de covariância na dependência de uma informação de correlação de objeto ou uma informação de covariância de objeto que descreve as características de uma pluralidade de objetos de áudio e/ou uma relação entre uma pluralidade de objetos de áudio. Portanto, é possível restaurar as características de correlação ou características de covariância, que são adaptadas aos objetos de áudio, em um estágio de processamento final, ou seja, após a renderização. Portanto, a complexidade para decodificar os objetos de áudio é reduzida. Além disso, ao considerar as características de correlação ou características de covariância dos objetos de áudio após a renderização, um impacto negativo da renderização pode ser evitado e as características de correlação ou características de covariância podem ser reconstruídas com boa precisão.

[023] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para determinar a informação de correlação de objeto ou a informação de covariância de objeto com base em uma informação de lado incluída na representação codificada. Portanto, o conceito pode ser bem adaptado a uma abordagem de codificação de objeto de áudio espacial, que usa a informação de lado.

[024] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para determinar características de correlação ou características de covariância reais dos sinais de áudio renderizados e para ajustar a combinação dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída na dependência das características de correlação ou características de covariância reais dos

sinais de áudio renderizados. Portanto, pode-se compreender que imperfeições nos estágios de processamento anteriores como, por exemplo, uma perda de energia na reconstrução de objetos de áudio, ou imperfeições causadas pela renderização, podem ser consideradas. Assim, a combinação dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados pode ser ajustada de uma maneira muito precisa às necessidades, de modo que a combinação dos sinais de áudio renderizados reais com os sinais de áudio descorrelacionados resulta nas características desejadas.

[025] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, sendo que os sinais de áudio renderizados são ponderados usando uma primeira matriz de mistura **P** e sendo que os um ou mais sinais de áudio descorrelacionados são ponderados usando uma segunda matriz de mistura **M**. Isto permite a derivação simples dos sinais de áudio de saída, sendo que uma operação de combinação linear é realizada, a qual é descrita pela matriz de mistura **P** que é aplicada aos sinais de áudio renderizados e a matriz de mistura **M** que é aplicada aos um ou mais sinais de áudio descorrelacionado.

[026] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para ajustar pelo menos uma dentre a matriz de mistura **P** e a matriz de mistura **M** de modo que as características de correlação ou características de covariância dos sinais de áudio de saída obtidas se aproximam ou são iguais às características de correlação desejadas ou características de covariância desejadas. Assim, há uma maneira de ajustar uma ou mais das matrizes de mistura, o que é normalmente possível com esforço moderado e bons resultados.

[027] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para conjuntamente calcular a matriz de mistura **P** e a matriz de mistura **M**. Portanto, é possível obter as matrizes de mistura de modo que as características de correlação ou características de covariância dos sinais de

áudio de saída obtidas podem ser definidas para se aproximarem ou serem iguais às características de correlação ou características de covariância. Além disso, ao calcular conjuntamente a matriz de mistura **P** e a matriz de mistura **M**, alguns graus de liberdade geralmente estão disponíveis, de modo que é possível encaixar melhor a matriz de mistura **P** e a matriz de mistura **M** às necessidades.

[028] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para obter uma matriz de mistura combinada **F**, que compreende a matriz de mistura **P** e a matriz de mistura **M**, de modo que uma matriz de covariância dos sinais de áudio de saída obtidos é igual a uma matriz de covariância desejada.

[029] Em uma modalidade preferida, a matriz de mistura combinada pode ser calculada de acordo com as equações descritas abaixo.

[030] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para determinar a matriz de mistura combinada **F** usando matrizes, que são determinadas usando a Decomposição de Valor Singular de uma primeira matriz de covariância, que descreve o sinal de áudio renderizado e o sinal de áudio descorrelacionado, e de uma segunda matriz de covariância, que descreve características de covariância desejadas dos sinais de áudio de saída. Usando tal Decomposição de Valor Singular constitui uma solução numericamente eficiente para determinar a matriz de mistura combinada.

[031] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para definir a matriz de mistura **P** para ser uma matriz de identidade, ou um múltiplo da mesma, e para calcular a matriz de mistura **M**. Isso evita uma mistura de diferentes sinais de áudio renderizados, o que ajuda a preservar uma impressão espacial desejada. Além disso, o número de graus de liberdade é reduzido.

[032] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para determinar a matriz de mistura **M** de modo que uma diferença entre a matriz de covariância desejada e a matriz de covariância dos

sinais de áudio renderizados se aproxime ou seja igual a uma covariância dos um ou mais sinais descorrelacionados, após mistura com a matriz de mistura **M**. Desse modo, um conceito computacionalmente simples para obter a matriz de mistura **M** é fornecido.

[033] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para determinar a matriz de mistura **M** usando matrizes que são determinadas usando a Decomposição de Valor Singular da diferença entre a matriz de covariância desejada e a matriz de covariância dos sinais de áudio renderizados e da matriz de covariância dos um ou mais sinais descorrelacionados. Esta é uma abordagem muito eficiente computacionalmente para determinar a matriz de mistura **M**.

[034] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para determinar as matrizes de mistura **P**, **M** sob a restrição de que um dado sinal de áudio renderizado somente é misturado com uma versão descorrelacionada do dado próprio sinal de áudio renderizado. Este conceito se limita a uma pequena modificação (por exemplo, na presença de descorrelacionadores imperfeitos) ou impede uma modificação das características de correlação cruzada ou características de covariância cruzada (por exemplo, no caso de descorrelacionadores ideais) e pode, portanto, ser desejável em alguns casos para evitar uma alteração de uma posição do objeto percebida. Entretanto, na presença de descorrelacionadores não ideais, os valores de autocorrelação (ou valores de autocovariância) são explicitamente modificados, e as mudanças nos termos cruzados são ignoradas.

[035] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados de modo que somente valores de autocorrelação ou valores de autocovariância dos sinais de áudio renderizados são modificados enquanto as características de correlação cruzada ou características de covariância cruzada são deixadas não modificadas ou modificadas com um valor pequeno (por exemplo, na presença de

descorrelacionadores imperfeitos). Novamente, uma degradação de uma posição percebida dos objetos de áudio pode ser evitada. Além disso, a complexidade computacional pode ser reduzida. Entretanto, por exemplo, os valores de variância cruzada são modificados como consequência da modificação das energias (valores de autocorrelação), mas os valores de correlação cruzada permanecem não modificados (eles representam a versão normalizada dos valores de variância cruzada).

[036] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para definir a matriz de mistura **P** para ser uma matriz de identidade, ou um múltiplo da mesma, e para calcular a matriz de mistura **M** sob a restrição de que **M** é uma matriz diagonal. Assim, uma modificação das características de correlação cruzada ou características de covariância cruzada pode ser evitada ou restrita a um valor pequeno (por exemplo, na presença de descorrelacionadores imperfeitos).

[037] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para obter o sinal de áudio de saída, sendo que uma matriz diagonal **M** é aplicada aos um ou mais sinais de áudio descorrelacionados **W**. Neste caso, o decodificador de áudio multicanal é configurado para calcular elementos diagonais da matriz de mistura **M** de modo que elementos diagonais da matriz de covariância dos sinais de áudio de saída são iguais às energias desejadas. Portanto, uma perda de energia, que pode ser obtida pela operação de renderização e/ou pela reconstrução dos objetos de áudio com base em um ou mais sinais de downmix e uma informação de lado espacial, pode ser compensada. Desse modo, uma intensidade adequada dos sinais de áudio de saída pode ser conseguida.

[038] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para calcular os elementos da matriz de mistura **M** na dependência dos elementos diagonais de uma matriz de covariância desejada, elementos diagonais de uma matriz de covariância dos sinais de áudio

renderizados, e elementos diagonais de uma matriz de covariância dos um ou mais sinais descorrelacionados. Elementos não-diagonais da matriz de mistura **M** podem ser definidos para zero, e a matriz de covariância desejada pode ser calculada com base na matriz de renderização usada para a operação de renderização e uma matriz de covariância de objeto. Além disso, um valor limiar pode ser usado para limitar uma quantidade de descorrelação adicionada aos sinais. Este conceito proporciona uma determinação muito eficiente computacionalmente dos elementos da matriz de mistura **M**.

[039] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para considerar características de correlação ou características de covariância dos sinais de áudio descorrelacionados ao determinar como combinar os sinais de áudio renderizados, ou a versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionado. Portanto, imperfeições da descorrelação podem ser consideradas.

[040] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para misturar sinais de áudio renderizados e sinais de áudio descorrelacionados, de modo que um dado sinal de áudio de saída é fornecido com base em dois ou mais sinais de áudio renderizados e pelo menos um sinal de áudio descorrelacionado. Pelo uso deste conceito, as características de correlação cruzada podem ser ajustadas eficientemente sem a necessidade de introduzir grandes quantidades de sinais descorrelacionados (o que pode degradar uma impressão espacial auditiva).

[041] Em uma modalidade preferida, o decodificador de áudio multicanal pode ser configurado para alternar entre modos diferentes, nos quais diferentes restrições são aplicadas para determinar como combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída. Portanto, características de complexidade e processamento podem ser ajustadas aos sinais que são processados.

[042] Em uma modalidade preferida, o decodificador de áudio multicanal

pode ser configurado para alternar entre um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado. Deste modo, ambas as características de complexidade e processamento podem ser ajustadas ao tipo de sinal de áudio que está atualmente sendo renderizado. A modificação das características de autocorrelação ou características de autocovariância e modificação não explicitamente das características de correlação cruzada ou características de covariância cruzada pode, por exemplo, ser útil se uma impressão espacial dos sinais de áudio fosse degradada por tal modificação, enquanto que, no entanto, é desejável ajustar as intensidades dos sinais de áudio de saída. Por outro lado, há casos nos quais é desejável ajustar as características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída. O decodificador de áudio multicanal mencionado aqui permite tal ajuste, sendo que no primeiro modo, é possível combinar sinais de áudio renderizados, de modo que uma quantidade (ou intensidade) dos componentes

de sinal descorrelacionados, que é necessária para ajustar as características de correlação cruzada ou características de covariância cruzada, é comparativamente pequena. Deste modo, componentes de sinal “localizáveis” são usados no primeiro modo para ajustar as características de correlação cruzada ou características de covariância cruzada. Em contraste, no segundo modo, sinais descorrelacionados são usados para ajustar as características de correlação cruzada ou características de covariância cruzada, o que naturalmente trazem junto uma impressão de auditiva diferente. Portanto, pelo fornecimento de três modos diferentes, o decodificador de áudio pode ser bem adaptado ao conteúdo de áudio sendo manipulado.

[043] Em uma modalidade preferida, o decodificador de áudio multicanal é configurado para avaliar um elemento de bitstream da representação codificada que indica quais dos três modos para combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados deve ser usado, e selecionar o modo dependendo do referido elemento de bitstream. Portanto, um codificador de áudio pode sinalizar um modo adequado na dependência de seu conhecimento dos conteúdos de áudio. Deste modo, uma qualidade máxima dos sinais de áudio de saída pode ser conseguida sob qualquer circunstância.

[044] Uma modalidade de acordo com a invenção cria um codificador de áudio de multicanal para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada. O codificador de áudio de multicanal é configurado para fornecer um ou mais sinais de downmix com base pelo menos em dois sinais de áudio de entrada. Além disso, o codificador de áudio de multicanal é configurado para fornecer um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada. Além disso, o codificador de áudio de multicanal é configurado para fornecer um parâmetro do método de descorrelação que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um codificador de áudio. Portanto, o codificador de áudio de

multicanal pode controlar o decodificador de áudio para usar um modo de correlação adequado, que é bem adaptado ao tipo de sinal de áudio que é atualmente codificado. Deste modo, o codificador de áudio de multicanal descrito aqui é bem adaptado para cooperação com o decodificador de áudio multicanal discutido antes.

[045] Em uma modalidade preferida, o codificador de áudio de multicanal é configurado para seletivamente fornecer o parâmetro do método de descorrelação, para sinalizar um dentre os três modos seguintes para a operação de um decodificador de áudio: um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, um segundo modo no qual nenhuma mistura entre diferentes dos sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal de áudio descorrelacionado é combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e um terceiro modo no qual nenhuma mistura entre diferentes dos sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal de áudio descorrelacionado seja combinado com sinais de áudio renderizados diferentes de um sinal de áudio renderizado a partir do que o dado sinal de áudio descorrelacionado é derivado. Deste modo, o codificador de áudio de multicanal pode alternar um decodificador de áudio multicanal através dos três modos discutidos acima na dependência do conteúdo de áudio, sendo que o modo no qual o decodificador de áudio multicanal é operado pode ser bem adaptado pelo codificador de áudio de multicanal ao tipo

do conteúdo de áudio atualmente codificado. Entretanto, em algumas modalidades, somente um ou dois dos três modos mencionados acima para a operação do decodificador de áudio pode ser usado (ou pode estar disponível).

[046] Em uma modalidade preferida, o codificador de áudio de multicanal é configurado para selecionar o parâmetro do método de descorrelação na dependência de se os sinais de áudio de entrada compreendem uma correlação comparativamente elevada ou uma correlação comparativamente menor. Deste modo, uma adaptação da descorrelação, que é usado no decodificador, pode ser feita com base nas características importantes dos sinais de áudio que são atualmente codificados.

[047] Em uma modalidade preferida, o codificador de áudio de multicanal é configurado para selecionar o parâmetro do método de descorrelação para designar o primeiro modo ou o segundo modo se uma correlação ou covariância entre os sinais de áudio de entrada for comparativamente alta, e para selecionar o parâmetro do método de descorrelação para designar o terceiro modo se uma correlação ou covariância entre os sinais de áudio de entrada for comparativamente inferior. Portanto, no caso da correlação comparativamente pequena ou covariância entre os sinais de áudio de entrada, um modo de decodificação é escolhido no qual não há correlação das características de covariância cruzada ou características de correlação cruzada. Verificou-se que esta é uma escolha eficiente para sinais que tem uma correlação comparativamente pequena (ou covariância), já que tais sinais são substancialmente independentes, o que elimina a necessidade de uma adaptação de correlações cruzadas ou covariâncias cruzadas. Em vez disso, um ajuste das correlações cruzadas ou covariâncias cruzadas para sinais de áudio de entrada substancialmente independentes (que tem uma correlação comparativamente pequena ou covariância) degradaria tipicamente uma qualidade de áudio e ao mesmo tempo aumenta uma complexidade de decodificação. Deste modo, este conceito permite uma adaptação razoável do decodificador de áudio multicanal à entrada do sinal no codificador de áudio de

multicanal.

[048] Uma modalidade de acordo com a invenção cria um método para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada. O método compreende renderizar uma pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, na dependência de um ou mais parâmetros de renderização, para obter uma pluralidade de sinais de áudio renderizados. O método também compreende derivar um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados e combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída. Este método baseia-se nas mesmas considerações que o decodificador de áudio de multicanal descrito acima. Além disso, o método pode ser suplementado por qualquer dos recursos e funcionalidades discutidas acima com relação ao decodificador de áudio multicanal.

[049] Outra modalidade de acordo com a invenção cria um método para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada. O método compreende fornecer um ou mais sinais de downmix com base pelo menos em dois sinais de áudio de entrada, fornecer um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada, e fornecer um parâmetro do método de descorrelação que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio. Este método baseia-se nas mesmas considerações que o codificador de áudio de multicanal descrito acima. Além disso, o método pode ser suplementado por qualquer dos recursos e funcionalidades descritas aqui com relação ao codificador de áudio de multicanal.

[050] Outra modalidade de acordo com a invenção cria um programa de computador para realizar um ou mais dos métodos descritos acima.

[051] Outra modalidade de acordo com a invenção cria uma

representação de áudio codificada, compreendendo uma representação codificada de um sinal de downmix, uma representação codificada de um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada, e um parâmetro do método de descorrelação codificado que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio. Esta representação de áudio codificada permite sinalizar um modo de descorrelação adequado e, portanto, ajuda a implementar as vantagens descritas com relação ao codificador de áudio de multicanal e o decodificador de áudio multicanal.

Breve descrição das figuras

[052] As modalidades de acordo com a presente invenção serão descritas posteriormente fazendo referência às figuras anexas nas quais:

[053] Fig. 1 mostra um diagrama em bloco esquemático de um decodificador de áudio multicanal, de acordo com uma modalidade da presente invenção;

[054] Fig. 2 mostra um diagrama em bloco esquemático de um codificador de áudio de multicanal, de acordo com uma modalidade da presente invenção;

[055] Fig. 3 mostra um fluxograma de um método para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada, de acordo com uma modalidade da invenção;

[056] Fig. 4 mostra um fluxograma de um método para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada, de acordo com uma modalidade da presente invenção;

[057] Fig. 5 mostra uma representação esquemática de uma representação de áudio codificada, de acordo com uma modalidade da presente invenção;

[058] Fig. 6 mostra um diagrama em bloco esquemático de um descorrelacionador de multicanal, de acordo com uma modalidade da presente invenção;

[059] Fig. 7 mostra um diagrama em bloco esquemático de um decodificador de áudio multicanal, de acordo com uma modalidade da presente invenção;

[060] Fig. 8 mostra um diagrama em bloco esquemático de um codificador de áudio de multicanal, de acordo com uma modalidade da presente invenção,

[061] Fig. 9 mostra um fluxograma de um método para fornecer a pluralidade de sinais descorrelacionados com base em uma pluralidade de sinais de entrada do descorrelacionador, de acordo com uma modalidade da presente invenção;

[062] Fig. 10 mostra um fluxograma de um método para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada, de acordo com uma modalidade da presente invenção;

[063] Fig. 11 mostra um fluxograma de um método para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada, de acordo com uma modalidade da presente invenção;

[064] Fig. 12 mostra uma representação esquemática de uma representação codificada, de acordo com uma modalidade da presente invenção;

[065] Fig. 13 mostra a representação esquemática que fornece uma visão geral de um conceito de downmix/upmix paramétrico com base em MMSE;

[066] Fig. 14 mostra uma representação esquemática para um princípio de ortogonalidade no espaço tridimensional;

[067] Fig. 15 mostra um diagrama em bloco esquemático de um sistema de reconstrução paramétrico com descorrelação aplicada à saída renderizada, de acordo com uma modalidade da presente invenção;

[068] Fig. 16 mostra um diagrama em bloco esquemático de uma unidade de descorrelação;

[069] Fig. 17 mostra um diagrama em bloco esquemático de uma

unidade de descorrelação de complexidade reduzida, de acordo com uma modalidade da presente invenção;

[070] Fig. 18 mostra uma representação em tabela das posições do alto-falante, de acordo com uma modalidade da presente invenção;

[071] Figs. 19a a 19g mostra representações em tabela dos coeficientes de pré-mistura para $N = 22$ e K entre 5 e 11;

[072] Figs. 20a a 20d mostra representações em tabela dos coeficientes de pré-mistura para $N = 10$ e K entre 2 e 5;

[073] Figs. 21a a 21c mostra representações em tabela dos coeficientes de pré-mistura para $N = 8$ e K entre 2 e 4;

[074] Figs 21d a 21f mostra representações em tabela dos coeficientes de pré-mistura para $N = 7$ e K entre 2 e 4;

[075] Figs. 22a e 22b mostra representações em tabela dos coeficientes de pré-mistura para $N = 5$ e $K = 2$ ou $K = 3$;

[076] Fig. 23 mostra uma representação em tabela dos coeficientes de pré-mistura para $N = 2$ e $K = 1$;

[077] Fig. 24 mostra uma representação em tabela dos grupos de sinais do canal;

[078] Fig. 25 mostra uma representação de sintaxe dos parâmetros adicionais, que podem ser incluídos na sintaxe de `SAOCSpecifigConfig()` ou, equivalentemente, `SAOC3DSpecificConfig()`;

[079] Fig. 26 mostra uma representação em tabela dos diferentes valores da variável do bitstream `bsDecorrelationMethod`;

[080] Fig. 27 mostra uma representação em tabela de um número de descorrelacionadores para diferentes níveis de descorrelação e configurações de saída, indicados pela variável do bitstream `bsDecorrelationLevel`;

[081] Fig. 28 mostra, na forma de um diagrama em bloco esquemático, uma visão geral sobre um codificador de áudio 3D;

[082] Fig. 29 mostra, na forma de um diagrama em bloco esquemático, uma visão geral sobre um decodificador de áudio 3D; e

[083] Fig. 30 mostra um diagrama em bloco esquemático de uma estrutura de um conversor de formato.

[084] Fig. 31 mostra um diagrama em bloco esquemático de um processador de downmix, de acordo com uma modalidade da presente invenção;

[085] Fig. 32 mostra uma tabela que representa os modos de decodificação para número diferente de objetos de downmix SAOC; e

[086] Fig. 33 mostra uma representação de sintaxe de um elemento de bitstream “SAOC3DSpecificConfig”.

[087] Descrição detalhada das modalidades

[088] Decodificador de áudio de multicanal de acordo com a Fig. 1

[089] Fig. 1 mostra um diagrama em bloco esquemático de um decodificador de áudio multicanal 100, de acordo com uma modalidade da presente invenção.

[090] O decodificador de áudio multicanal 100 é configurado para receber uma representação codificada 110 e para fornecer, com base na mesma, ao menos dois sinais de áudio de saída 112, 114.

[091] O decodificador de áudio multicanal 100 de preferência compreende um decodificador 120 que é configurado para fornecer sinais de áudio decodificados 122 com base na representação codificada 110. Além disso, o decodificador de áudio multicanal 100 compreende um renderizador 130, que é configurado para renderizar uma pluralidade de sinais de áudio decodificados 122, que são obtidos com base na representação codificada 110 (por exemplo, pelo decodificador 120) na dependência de um ou mais parâmetros de renderização 132, para obter uma pluralidade de sinais de áudio renderizados 134, 136. Além disso, o decodificador de áudio multicanal 100 compreende um descorrelacionador 140, que é configurado para derivar um ou mais sinais de áudio descorrelacionados 142, 144 a partir dos sinais de áudio renderizados 134, 136. Além disso, o decodificador de áudio multicanal 100 compreende um combinador 150, que é configurado para combinar os sinais de áudio

renderizados 134, 136, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados 142, 144 para obter os sinais de áudio de saída 112, 114.

[092] Entretanto, deve-se observar que uma estrutura de hardware diferente do decodificador de áudio multicanal 100 pode ser possível, contanto que as funcionalidades acima descritas são apresentadas.

[093] Em relação a funcionalidade do decodificador de áudio multicanal 100, deve-se observar que os sinais de áudio descorrelacionados 142, 144 são derivados dos sinais de áudio renderizados 134, 136, e que os sinais de áudio descorrelacionados 142, 144 são combinados com os sinais de áudio renderizados 134, 136 para obter os sinais de áudio de saída 112, 114. Pela derivação dos sinais de áudio descorrelacionados 142, 144 a partir dos sinais de áudio renderizados 134, 136, um processamento particularmente eficiente pode ser conseguido, já que o número de sinais de áudio renderizados 134, 136 é tipicamente independente do número dos sinais de áudio decodificados 122 que são inseridos no renderizador 130. Deste modo, o esforço de descorrelação é tipicamente independente do número de sinais de áudio decodificados 122, o que melhora a eficiência da implementação. Além disso, a aplicação da descorrelação após a renderização evita a introdução de artefatos, o que poderia ser causado pelo renderizador ao combinar vários sinais descorrelacionados no caso em que a descorrelação é aplicada antes da renderização. Além disso, as características dos sinais de áudio renderizados podem ser consideradas na descorrelação realizada pelo descorrelacionador 140, que resulta tipicamente nos sinais de áudio de saída de boa qualidade.

[094] Além disso, deve-se observar que o decodificador de áudio multicanal 100 pode ser suplementado por qualquer dos recursos e funcionalidades descritas aqui. Em particular, deve-se observar que melhorias individuais como as descritas aqui podem ser introduzidas no decodificador de áudio multicanal 100 para, desse modo, melhorar ainda a eficiência do processamento, e/ou a qualidade dos sinais de áudio de saída.

[095] 2. Codificador de áudio de multicanal De acordo com a Fig. 2

[096] Fig. 2 mostra um diagrama em bloco esquemático de um codificador de áudio de multicanal 200, de acordo com uma modalidade da presente invenção. O codificador de áudio de multicanal 200 é configurado para receber dois ou mais sinais de áudio de entrada 210, 212, e para fornecer, com base na mesma, uma representação codificada 214. O codificador de áudio de multicanal compreende um provedor de sinal de downmix 220, que é configurado para fornecer um ou mais sinais de downmix 222 com base pelo menos em dois sinais de áudio de entrada 210, 212. Além disso, o codificador de áudio de multicanal 200 compreende um provedor de parâmetro 230, que é configurado para fornecer um ou mais parâmetros 232 que descrevem uma relação (por exemplo, uma correlação cruzada, uma covariância cruzada, uma diferença de nível ou similares) entre os pelo menos dois sinais de áudio de entrada 210, 212.

[097] Além disso, o codificador de áudio de multicanal 200 também compreende um provedor do parâmetro do método de descorrelação 240, que é configurado para fornecer um parâmetro do método de descorrelação 242 que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio. Os um ou mais sinais de downmix 222, os um ou mais parâmetros 232 e o parâmetro do método de descorrelação 242 são incluídos, por exemplo, em uma forma codificada, na representação codificada 214.

[098] Entretanto, deve-se observar que a estrutura de hardware do codificador de áudio de multicanal 200 pode ser diferente, contanto que as funcionalidades acima descritas sejam satisfeitas. Em outras palavras, a distribuição das funcionalidades do codificador de áudio de multicanal 200 para blocos individuais (por exemplo, para o provedor de sinal de downmix 220, para o provedor de parâmetro 230 e para o provedor do parâmetro do método de descorrelação 240) deve ser considerada somente como um exemplo.

[099] Em relação a funcionalidade do codificador de áudio de multicanal

200, deve-se observar que os um ou mais sinais de downmix 222 e os um ou mais parâmetros 232 são fornecidos de uma maneira convencional, por exemplo como em um codificador de áudio de multicanal SAOC ou em um codificador de áudio de multicanal USAC. Entretanto, o parâmetro do método de descorrelação 242, que também é fornecido pelo codificador de áudio de multicanal 200 e incluído na representação codificada 214, pode ser usado para adaptar um modo de descorrelação aos sinais de áudio de entrada 210, 212 ou a uma qualidade de reprodução desejada. Portanto, o modo de descorrelação pode ser adaptado a diferentes tipos de conteúdo de áudio. Por exemplo, diferentes modos de descorrelação podem ser escolhidos para tipos de conteúdos de áudio nos quais os sinais de áudio de entrada 210, 212 são fortemente correlacionados e para tipos de conteúdo de áudio em que os sinais de áudio de entrada 210, 212 são independentes. Além disso, diferentes modos de descorrelação podem, por exemplo, ser sinalizados pelo parâmetro do modo de descorrelação 242 para tipos de conteúdo de áudio em que uma percepção espacial é particularmente importantes e para tipos de conteúdo de áudio em que uma impressão espacial é mesmo importante ou ainda de importância subordinada (por exemplo, quando comparado a uma reprodução de canais individuais). Portanto, um decodificador de áudio multicanal, que recebe a representação codificada 214, pode ser controlado pelo codificador de áudio de multicanal 200, e pode ser definido para um modo de decodificação que traz consigo um melhor compromisso possível entre complexidade de decodificação e qualidade de reprodução.

[100] Além disso, deve-se observar que o codificador de áudio de multicanal 200 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui. Deve-se observar que os possíveis recursos e melhorias adicionais descritos aqui podem ser adicionados ao codificador de áudio de multicanal 200 individualmente ou em combinação, para assim aperfeiçoar (ou melhorar) o codificador de áudio de multicanal 200.

[101] 3. Método para fornecer ao menos dois sinais de áudio de saída de

acordo com a Fig. 3

[102] A Fig. 3 mostra um fluxograma de um método 300 para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada. O método compreende renderizar 310 uma pluralidade de sinais de áudio decodificados, que são obtidos com base em uma representação codificada 312, na dependência de um ou mais parâmetros de renderização, para obter uma pluralidade de sinais de áudio renderizados. O método 300 também compreende derivar 320 um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados. O método 300 também compreende combinar 330 os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída 332.

[103] Deve-se observar que o método 300 é baseado nas mesmas considerações que o decodificador de áudio multicanal 100 de acordo com a Fig. 1. Além disso, deve-se observar que o método 300 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui (ou individualmente ou em combinação). Por exemplo, o método 300 pode ser suplementado por qualquer dos recursos e funcionalidades descritos com relação aos decodificadores de áudio multicanal descritos aqui.

[104] 4. Método para fornecer uma representação codificada de acordo com a Fig. 4

[105] A Fig. 4 mostra um fluxograma de um método 400 para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada. O método 400 compreende fornecer 410 um ou mais sinais de downmix com base pelo menos em dois sinais de áudio de entrada 412. O método 400 compreende ainda fornecer 420 um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada 412 e fornecer 430 um parâmetro do método de descorrelação que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio. Portanto,

uma representação codificada 432 é fornecida, que de preferência inclui uma representação codificada dos um ou mais sinais de downmix, um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada, e o parâmetro do método de descorrelação.

[106] Deve-se observar que o método 400 é baseado nas mesmas considerações que o codificador de áudio de multicanal 200 de acordo com a Fig. 2, de modo que as explicações acima também se aplicam.

[107] Além disso, deve-se observar que a ordem das etapas 410, 420, 430 pode ser variada flexivelmente, e que as etapas 410, 420, 430 também podem ser realizadas em paralelo até isso ser possível em um ambiente de execução para o método 400. Além disso, deve-se observar que o método 400 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui, ou individualmente ou em combinação. Por exemplo, o método 400 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui com relação aos codificadores de áudio de multicanal. Entretanto, também é possível introduzir recursos e funcionalidades que correspondem aos recursos e funcionalidades dos decodificadores de áudio multicanal descritos aqui, que recebem a representação codificada 432.

[108] 5. Representação de áudio codificada de acordo com a Fig. 5

[109] A Fig. 5 mostra uma representação esquemática de uma representação de áudio codificada 500 de acordo com uma modalidade da presente invenção.

[110] A representação de áudio codificada 500 compreende uma representação codificada 510 de um sinal de downmix, uma representação codificada 520 de um ou mais parâmetros que descrevem uma relação entre ao menos dois sinais de áudio. Além disso, a representação de áudio codificada 500 também compreende um parâmetro do método de descorrelação codificado 530 que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio. Portanto, a representação de áudio codificada permite

sinalizar um modo de descorrelação a partir de um codificador de áudio para um decodificador de áudio. Portanto, é possível obter um modo de descorrelação que é bem adaptado às características do conteúdo de áudio (que é descrito, por exemplo, pela representação codificada 510 de um ou mais sinais de downmix e pela representação codificada 520 de um ou mais parâmetros que descrevem uma relação entre ao menos dois sinais de áudio (por exemplo, os ao menos dois sinais de áudio submisturados na representação codificada 510 de um ou mais sinais de downmix)). Deste modo, a representação de áudio codificada 500 permite uma renderização de um conteúdo de áudio representado pela representação de áudio codificada 500 com uma impressão espacial auditiva particularmente boa e/ou um equilíbrio particularmente bom entre impressão espacial auditiva e complexidade de decodificação.

[111] Além disso, deve-se observar que a representação codificada 500 pode ser suplementada por qualquer dos recursos e funcionalidades descritos com relação aos codificadores de áudio de multicanal e os decodificadores de áudio multicanal, ou individualmente ou em combinação.

[112] 6. Descorrelacionador de multicanal de acordo com a Fig. 6

[113] A Fig. 6 mostra um diagrama em bloco esquemático de um descorrelacionador de multicanal 600, de acordo com uma modalidade da presente invenção.

[114] O descorrelacionador de multicanal 600 é configurado para receber um primeiro conjunto de N sinais de entrada do descorrelacionador 610a a 610n e fornecer, com base no mesmo, um segundo conjunto de N' sinais de saída do descorrelacionador 612a a 612n'. Em outras palavras, o descorrelacionador de multicanal 600 é configurado para fornecer uma pluralidade de (ao menos aproximadamente) sinais descorrelacionados 612a a 612n' com base nos sinais de entrada do descorrelacionador 610a a 610n.

[115] O descorrelacionador de multicanal 600 compreende um pré-misturador 620, que é configurado para pré-misturar o primeiro conjunto de N

sinais de entrada do descorrelacionador 610a a 610n em um segundo conjunto de K sinais de entrada do descorrelacionador 622a a 622k, sendo que K é menor que N (com K e N sendo números inteiros). O descorrelacionador de multicanal 600 também compreende uma descorrelação (ou núcleo do descorrelacionador) 630, que é configurado para fornecer um primeiro conjunto de K' sinais de saída do descorrelacionador 632a a 632k' com base no segundo conjunto de K sinais de entrada do descorrelacionador 622a a 622k. Além disso, o descorrelacionador de multicanal compreende um pós-misturador 640, que é configurado para fazer o upmix do primeiro conjunto de K' sinais de saída do descorrelacionador 632a a 632k' em um segundo conjunto de N' sinais de saída do descorrelacionador 612a a 612n', sendo que N' é maior que K' (com N' e K' sendo números inteiros).

[116] Entretanto, deve-se observar que a dada estrutura do descorrelacionador de multicanal 600 deve ser considerada como um exemplo somente, e que não é necessário subdividir o descorrelacionador de multicanal 600 em blocos funcionais (por exemplo, no pré-misturador 620, a descorrelação ou núcleo do descorrelacionador 630 e o pós-misturador 640) contanto que a funcionalidade descrita aqui seja fornecida.

[117] Em relação a funcionalidade do descorrelacionador de multicanal 600, também deve ser observado que o conceito de realizar uma pré-mistura, para derivar o segundo conjunto de K sinais de entrada do descorrelacionador a partir do primeiro conjunto de N sinais de entrada do descorrelacionador, e de realizar a descorrelação com base no (pré-misturado ou “submisturado”) segundo conjunto de K sinais de entrada do descorrelacionador traz consigo uma redução de uma complexidade quando comparado a um conceito em que a descorrelação efetiva é aplicada, por exemplo, diretamente aos N sinais de entrada do descorrelacionador. Além disso, o segundo conjunto (que sofreu upmix) de N' sinais de saída do descorrelacionador é obtido com base no primeiro conjunto (original) de sinais de saída do descorrelacionador, que são o resultado da descorrelação efetiva, com base em uma pós-mistura, que pode

ser realizada pelo dispositivo de upmix 640. Deste modo, o descorrelacionador de multicanal 600 efetivamente (quando visto a partir de fora) recebe N sinais de entrada do descorrelacionador e fornece, com base na mesma, N' sinais de saída do descorrelacionador, enquanto o núcleo do descorrelacionador efetivo 630 funciona somente em um número menor de sinais (ou seja, K sinais de entrada do descorrelacionador submisturados 622a a 622k do segundo conjunto de K sinais de entrada do descorrelacionador). Deste modo, a complexidade do descorrelacionador de multicanal 600 pode ser substancialmente reduzida, quando comparada aos descorrelacionadores convencionais, pela realização de uma submistura ou “pré-mistura” (que pode de preferência ser uma pré-mistura linear sem qualquer funcionalidade de descorrelação) para um lado de entrada da descorrelação (ou núcleo do descorrelacionador) 630 e pela realização da sobremistura ou “pós-mistura” (por exemplo, uma sobremistura linear sem qualquer funcionalidade de descorrelação adicional) com base nos sinais de saída (originais) 632a a 632k' da descorrelação (núcleo do descorrelacionador) 630.

[118] Além disso, deve-se observar que o descorrelacionador de multicanal 600 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui com relação à descorrelação de multicanal e também com relação aos decodificadores de áudio multicanal. Deve-se observar que os recursos descritos aqui podem ser adicionados ao descorrelacionador de multicanal 600 ou individualmente ou em combinação, para assim aperfeiçoar ou melhorar o descorrelacionador de multicanal 600.

[119] Deve-se observar que um descorrelacionador de multicanal sem redução de complexidade pode ser derivado do descorrelacionador de multicanal descrito acima para $K=N$ (e possivelmente $K'=N'$ ou ainda $K=N=K'=N'$).

[120] 7. Decodificador de áudio de multicanal de acordo com a Fig. 7

[121] A Fig. 7 mostra um diagrama em bloco esquemático de um decodificador de áudio multicanal 700, de acordo com uma modalidade da

invenção.

[122] O decodificador de áudio multicanal 700 é configurado para receber uma representação codificada 710 e para fornecer, com base na mesma, ao menos dois sinais de saída 712, 714. O decodificador de áudio multicanal 700 compreende um descorrelacionador de multicanal 720, que pode ser substancialmente idêntico ao descorrelacionador de multicanal 600 de acordo com a Fig. 6. Além disso, o decodificador de áudio multicanal 700 pode compreender qualquer dos recursos e funcionalidades de um decodificador de áudio multicanal que são conhecidos por uma pessoa versada na técnica ou que são descritos aqui com relação a outros decodificadores de áudio de multicanal.

[123] Além disso, deve-se observar que o decodificador de áudio multicanal 700 compreende uma eficiência particularmente alta quando comparado aos decodificadores de áudio de multicanal convencionais, já que o decodificador de áudio multicanal 700 usa o descorrelacionador de multicanal de alta eficiência 720.

[124] 8. Codificador de áudio de multicanal de acordo com a Fig. 8

[125] A Fig. 8 mostra um diagrama em bloco esquemático de um codificador de áudio de multicanal 800 de acordo com uma modalidade da presente invenção. O codificador de áudio de multicanal 800 é configurado para receber pelo menos dois sinais de áudio de entrada 810, 812 e para fornecer, com base no mesmo, uma representação codificada 814 de um conteúdo de áudio representado pelos sinais de áudio de entrada 810, 812.

[126] O codificador de áudio de multicanal 800 compreende um provedor de sinal de downmix 820, que é configurado para fornecer um ou mais sinais de downmix 822 com base pelo menos em dois sinais de áudio de entrada 810, 812. O codificador de áudio de multicanal 800 também compreende um provedor de parâmetro 830 que é configurado para fornecer um ou mais parâmetros 832 (por exemplo, parâmetros de correlação cruzada ou parâmetros de covariância cruzada, ou parâmetros de correlação inter-objeto

e/ou parâmetros de diferença de nível do objeto) com base nos sinais de áudio de entrada 810,812. Além disso, o codificador de áudio de multicanal 800 compreende um provedor de parâmetro de complexidade da descorrelação 840 que é configurado para fornecer um parâmetro de complexidade de descorrelação 842 que descreve uma complexidade de uma descorrelação a ser usada no lado de um decodificador de áudio (que recebe a representação codificada 814). Os um ou mais sinais de downmix 822, os um ou mais parâmetros 832 e o parâmetro de complexidade de descorrelação 842 são incluídos na representação codificada 814, de preferência em uma forma codificada.

[127] Entretanto, deve-se observar que a estrutura interna do codificador de áudio de multicanal 800 (por exemplo, a presença do provedor de sinal de downmix 820, do provedor de parâmetro 830 e do provedor de parâmetro de complexidade da descorrelação 840) deve ser considerada como um exemplo somente. Diferentes estruturas são possíveis contanto que a funcionalidade descrita aqui seja conseguida.

[128] Em relação a funcionalidade do codificador de áudio de multicanal 800, deve-se observar que o codificador multicanal fornece uma representação codificada 814, sendo que os um ou mais sinais de downmix 822 e os um ou mais parâmetros 832 podem ser similares a, ou iguais aos sinais de downmix e parâmetros fornecidos pelos codificadores de áudio convencionais (como, por exemplo, codificadores de áudio SAOC convencionais ou codificadores de áudio USAC). Entretanto, o codificador de áudio de multicanal 800 é também configurado para fornecer o parâmetro de complexidade de descorrelação 842, que permite determinar uma complexidade de descorrelação que é aplicada no lado de um decodificador de áudio. Portanto, a complexidade de descorrelação pode ser adaptada ao conteúdo de áudio que é atualmente codificado. Por exemplo, é possível sinalizar uma complexidade de descorrelação desejada, que corresponde a uma qualidade de áudio atingível, na dependência de um conhecimento do lado do codificador sobre as características dos sinais de

áudio de entrada. Por exemplo, se for constatado que as características espaciais são importantes para um sinal de áudio, uma complexidade de descorrelação mais alta pode ser sinalizada, usando o parâmetro de complexidade de descorrelação 842, quando comparado a um caso em que as características espaciais não são tão importantes. Alternativamente, o uso de uma alta complexidade de descorrelação pode ser sinalizada usando o parâmetro de complexidade de descorrelação 842, se for constatado que uma passagem do conteúdo de áudio ou todo o conteúdo de áudio é tal que uma descorrelação de alta complexidade é necessária em um lado de um decodificador de áudio por outras razões.

[129] Para resumir, o codificador de áudio de multicanal 800 proporciona a possibilidade de controlar um decodificador de áudio multicanal, para usar uma complexidade de descorrelação que é adaptada para sinalizar características ou características de reprodução desejadas que podem ser definidas pelo codificador de áudio de multicanal 800.

[130] Além disso, deve-se observar que o codificador de áudio de multicanal 800 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui com relação a um codificador de áudio de multicanal, ou individualmente ou em combinação. Por exemplo, alguns ou todos os recursos descritos aqui com relação aos codificadores de áudio de multicanal podem ser adicionados ao codificador de áudio de multicanal 800. Além disso, o codificador de áudio de multicanal 800 pode ser adaptado para cooperação com os decodificadores de áudio multicanal descritos aqui.

[131] 9. Método para fornecer uma pluralidade de sinais descorrelacionados com base em uma pluralidade de sinais de entrada do descorrelacionador, de acordo com a Fig. 9

[132] A Fig. 9 mostra um fluxograma de um método 900 para fornecer uma pluralidade de sinais descorrelacionados com base em uma pluralidade de sinais de entrada do descorrelacionador.

[133] O método 900 compreende a pré-mistura 910 de um primeiro

conjunto de N sinais de entrada do descorrelacionador em um segundo conjunto de K sinais de entrada do descorrelacionador, sendo que K é menor que N . O método 900 também compreende fornecer 920 um primeiro conjunto de K' sinais de saída do descorrelacionador com base no segundo conjunto de K sinais de entrada do descorrelacionador. Por exemplo, o primeiro conjunto de K' sinais de saída do descorrelacionador pode ser fornecido com base no segundo conjunto de K sinais de entrada do descorrelacionador usando a descorrelação, que pode ser realizada, por exemplo, usando um núcleo do descorrelacionador ou usando um algoritmo de descorrelação. O método 900 compreende ainda a pós-mistura 930 do primeiro conjunto de K' sinais de saída do descorrelacionador em um segundo conjunto para N' sinais de saída do descorrelacionador, sendo que N' é maior que K' (com N' e K' sendo números inteiros). Portanto, o segundo conjunto de N' sinais de saída do descorrelacionador, que são a saída do método 900, pode ser fornecido com base no primeiro conjunto de N sinais de entrada do descorrelacionador, que são a entrada para o método 900.

[134] Deve-se observar que o método 900 é baseado nas mesmas considerações que o descorrelacionador de multicanal descrito acima. Além disso, deve-se observar que o método 900 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui com relação ao descorrelacionador de multicanal (e também com relação ao codificador de áudio de multicanal, se aplicável), ou individualmente ou em combinação.

[135] 10. Método para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada, de acordo com a Fig. 10

[136] A Fig. 10 mostra um fluxograma de um método 1000 para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada.

[137] O método 1000 compreende fornecer 1010 ao menos dois sinais de áudio de saída 1014, 1016 com base em uma representação codificada 1012. O método 1000 compreende fornecer 1020 uma pluralidade de sinais

descorrelacionados com base em uma pluralidade de sinais de entrada do descorrelacionador de acordo com o método 900 de acordo com a Fig. 9.

[138] Deve-se observar que o método 1000 é baseado nas mesmas considerações que o decodificador de áudio multicanal 700 de acordo com a Fig. 7.

[139] Também, deve-se observar que o método 1000 pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui com relação aos decodificadores de multicanal, ou individualmente ou em combinação.

[140] 11. Método para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada, de acordo com a Fig. 11

[141] A Fig. 11 mostra um fluxograma de um método 1100 para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada.

[142] O método 1100 compreende fornecer 1110 um ou mais sinais de downmix com base pelo menos em dois sinais de áudio de entrada 1112, 1114. O método 1100 também compreende fornecer 1120 um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada 1112, 1114. Além disso, o método 1100 compreende fornecer 1130 um parâmetro de complexidade de descorrelação que descreve uma complexidade de uma descorrelação a ser usada no lado de um decodificador de áudio. Portanto, uma representação codificada 1132 é fornecida com base pelo menos em dois sinais de áudio de entrada 1112, 1114, sendo que a representação codificada tipicamente compreende os um ou mais sinais de downmix, os um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada e o parâmetro de complexidade de descorrelação em uma forma codificada.

[143] Deve-se observar que as etapas 1110, 1120, 1130 podem ser realizadas em paralelo ou em uma ordem diferente em algumas modalidades de acordo com a invenção. Além disso, deve-se observar que o método 1100 é

baseado nas mesmas considerações que o codificador de áudio de multicanal 800 de acordo com a Fig. 8, e que o método 1100 pode ser suplementado por qualquer dos recursos e funcionalidades descritas aqui com relação ao codificador de áudio de multicanal, ou em combinação ou individualmente. Além disso, deve-se observar que o método 1100 pode ser adaptado para combinar o decodificador de áudio multicanal e o método para fornecer ao menos dois sinais de áudio de saída descritos aqui.

[144] 12. Representação de áudio codificada de acordo com a Fig. 12

[145] A Fig. 12 mostra uma representação esquemática de uma representação de áudio codificada, de acordo com uma modalidade da presente invenção. A representação de áudio codificada 1200 compreende uma representação codificada 1210 de um sinal de downmix, uma representação codificada 1220 de um ou mais parâmetros que descrevem uma relação entre os pelo menos dois sinais de áudio de entrada, e um parâmetro de complexidade de descorrelação decodificada 1230 que descreve uma complexidade de uma descorrelação a ser usada no lado de um decodificador de áudio. Portanto, a representação de áudio codificada 1200 permite ajustar a complexidade de descorrelação usada por um decodificador de áudio multicanal, que traz consigo uma eficiência de decodificação melhorada, e possivelmente uma qualidade de áudio melhorada, ou um equilíbrio melhorado entre eficiência de codificação e qualidade de áudio. Além disso, deve-se observar que a representação de áudio codificada 1200 pode ser fornecida pelo codificador de áudio de multicanal como descrito aqui, e pode ser usada pelo decodificador de áudio multicanal como descrito aqui. Portanto, a representação de áudio codificada 1200 pode ser suplementada por qualquer dos recursos descritos com relação aos codificadores de áudio de multicanal e com relação aos decodificadores de áudio multicanal.

[146] 13. Notação e considerações subjacentes

[147] Recentemente, têm sido propostas técnicas paramétricas para a transmissão/armazenamento de bitrate eficiente de cenas de áudio contendo

vários objetos de áudio no campo de codificação de áudio (ver, por exemplo, as referências [BCC], [JSC], [SAOC], [SAOC1], [SAOC2]) e separação de fonte informada (ver, por exemplo, as referências [ISS1], [ISS2], [ISS3], [ISS4], [ISS5], [ISS6]). Essas técnicas visam a reconstrução de uma cena de áudio de saída desejada ou objeto de fonte de áudio com base em informações de lado adicionais que descrevem a cena de áudio transmitida/armazenada e/ou objetos de origem na cena de áudio. Esta reconstrução ocorre no decodificador utilizando um esquema de separação de fonte informada paramétrico. Além disso, também é feita referência ao chamado conceito "MPEG Surround", que é descrito, por exemplo, no padrão internacional ISO/IEC 23003-1: 2007. Além disso, também é feita referência à chamada "Spatial Audio Object Coding" que é descrita no padrão internacional ISO/IEC 23003-2:2010. Além disso, referência é feita ao chamado conceito de "Unified Speech and Audio Coding", que é descrito no padrão internacional ISO/IEC 23003-3:2012. Os conceitos a partir desses padrões podem ser usados nas modalidades de acordo com a invenção, por exemplo, nos codificadores de áudio de multicanal mencionados aqui e os decodificadores de áudio multicanal mencionados aqui, sendo que algumas adaptações podem ser necessárias.

[148] No seguinte, alguma informação de fundamento será descrita. Em particular, uma visão geral dos esquemas de separação paramétricos será fornecida, usando o exemplo da tecnologia de codificação de objeto de áudio espacial MPEG (SAOC) (ver, por exemplo, a referência [SAOC]). As propriedades matemáticas deste método são consideradas.

[149] 13.1. Notação e Definições

[150] A seguinte notação matemática é aplicada no presente documento:

[151] N_{Objetos} número de sinais de objeto de áudio

[152] N_{DmxCh} número de canais de downmix (processados)

[153] N_{UpmixCh} número de canais de upmix (saída)

[154] N_{Amostras} número de amostras de dados processados

- [155] $\mathbf{D}$ matriz de downmix, tamanho $N_{CDmix} \times N_{Objetos}$
- [156] $\mathbf{x}$ sinal do objeto de áudio de entrada, tamanho $N_{Objetos} \times N_{Amostras}$
- [157] $\mathbf{E}_X$ matriz de covariância do objeto, tamanho $N_{Objetos} \times N_{Objetos}$
- definida como $\mathbf{E}_X = \mathbf{X}\mathbf{X}^H$
- [158] $\mathbf{Y}$ sinal de áudio de downmix, tamanho $N_{CDmix} \times N_{Amostras}$
- definido como $\mathbf{Y} = \mathbf{D}\mathbf{X}$
- [159] $\mathbf{E}_Y$ matriz de covariância dos sinais de downmix, tamanho $N_{CDmix} \times N_{CDmix}$
- definida como $\mathbf{E}_Y = \mathbf{Y}\mathbf{Y}^H$
- [160] $\mathbf{G}$ matriz de estimativa da fonte paramétrica, tamanho $N_{Objetos} \times N_{CDmix}$
- [161] que se aproxima de $\mathbf{E}_X \mathbf{D}^H (\mathbf{D} \mathbf{E}_X \mathbf{D}^H)^{-1}$
- [162] $\hat{\mathbf{X}}$ sinal de objeto parametricamente reconstruído, tamanho $N_{Objetos} \times N_{Amostras}$
- [163] que se aproxima de $\mathbf{x}$ e definido como $\hat{\mathbf{X}} = \mathbf{G}\mathbf{Y}$
- [164] $\mathbf{R}$ matriz de renderização (especificada no lado do decodificador), tamanho $N_{CUpmix} \times N_{Objetos}$
- [165] $\mathbf{Z}$ sinal de cena de saída renderizada ideal, tamanho $N_{CUpmix} \times N_{Amostras}$
- definido como $\mathbf{Z} = \mathbf{R}\mathbf{X}$
- [166] $\hat{\mathbf{Z}}$ saída paramétrica renderizada, tamanho $N_{CUpmix} \times N_{Amostras}$

definida como $\hat{\mathbf{Z}} = \mathbf{R}\hat{\mathbf{X}}$

[167] $\mathbf{C}$ matriz de covariância da saída ideal, tamanho $N_{iCUpmix} \times N_{iCUpmix}$

definido como $\mathbf{C} = \mathbf{R}\mathbf{E}_X\mathbf{R}^H$

[168] $\mathbf{W}$ saídas do descorrelacionador, tamanho $N_{iCUpmix} \times N_{Amostras}$

[169] $\mathbf{S}$ sinal combinado $\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}$, tamanho $2N_{iCUpmix} \times N_{Amostras}$

[170] $\mathbf{E}_S$ sinal combinado matriz de covariância, tamanho $2N_{iCUpmix} \times 2N_{iCUpmix}$

definido como $\mathbf{E}_S = \mathbf{S}\mathbf{S}^H$

[171] $\tilde{\mathbf{Z}}$ saída final, tamanho $N_{iCUpmix} \times N_{Amostras}$

[172] $(\cdot)^H$ operador auto-adjunto (Hermitiano)

[173] que representa o transposto conjugado complexo de $(\cdot)$. A notação $(\cdot)^*$ também pode ser usada.

[174] $F_{decorr}(\cdot)$ função do descorrelacionador

[175] ε é uma constante aditiva ou uma constante de limitação (por exemplo, usada em uma operação “máxima” ou uma operação “max”) para evitar a divisão por zero

[176] $\mathbf{H} = \text{matdiag}(\mathbf{M})$ é uma matriz que contém os elementos da diagonal principal da matriz $\mathbf{M}$ na diagonal principal e valores zero nas posições fora da diagonal.

[177] Sem perda de generalidade, a fim de melhorar a capacidade de leitura das equações, para todas as variáveis introduzidas os índices que denotam dependência de tempo e frequência são omitidos neste documento.

[178] 13.2. Sistemas de separação paramétricos

[179] Sistemas de separação paramétricos gerais visam estimar um

número de fontes de áudio a partir de uma mistura de sinal (downmix), utilizando informações de parâmetro auxiliar (como, por exemplo, valores de correlação inter-canal, valores de diferença de nível inter-canal, valores de correlação inter-objeto e/ou informação de diferença de nível de objeto). Uma solução típica desta tarefa é baseada na aplicação dos algoritmos de estimativa de erro quadrático médio mínimo (MMSE) . A tecnologia SAOC é um exemplo deste tipo de sistemas de codificação/decodificação de áudio paramétricos.

[180] A Fig. 13 mostra o princípio geral da arquitetura de codificador/decodificador SAOC. Em outras palavras, a Fig. 13 mostra, na forma de um diagrama em bloco esquemático, uma visão geral do conceito de downmix/upmix paramétrico com base em MMSE.

[181] Um codificador 1310 recebe uma pluralidade de sinais de objeto 1312a, 1312b a 1312n. Além disso, o codificador 1310 também recebe parâmetros de mistura D, 1314, que podem, por exemplo, ser parâmetros de downmix. O codificador 1310 fornece, com base no mesmo, um ou mais sinais de downmix 1316a, 1316b, e assim por diante. Além disso, o codificador fornece uma informação de lado 1318. Os um ou mais sinais de downmix e a informação de lado pode, por exemplo, ser fornecida em uma forma codificada.

[182] O codificador 1310 compreende um misturador 1320, que é tipicamente configurado para receber os sinais de objeto 1312a a 1312n e para combinar (por exemplo downmix) os sinais de objeto 1312a a 1312n em um ou mais sinais de downmix 1316a, 1316b na dependência dos parâmetros de mistura 1314. Além disso, o codificador compreende um estimador de informação de lado 1330, que é configurado para derivar a informação de lado 1318 a partir dos sinais de objeto 1312a a 1312n. Por exemplo, um estimador de informação de lado 1330 pode ser configurado para derivar a informação de lado 1318 de modo que a informação de lado descreve uma relação entre sinais de objeto, por exemplo, uma correlação cruzada entre sinais de objeto (que pode ser designada como “correlação inter-objeto” IOC) e/ou uma

informação que descreve diferenças de nível entre sinais de objeto (que pode ser designada como uma “informação de diferença de nível de objeto” OLD).

[183] Os um ou mais sinais de downmix 1316a, 1316b e a informação de lado 1318 podem ser armazenados e/ou transmitidos para um decodificador 1350, que é indicado no número de referência 1340.

[184] O decodificador 1350 recebe os um ou mais sinais de downmix 1316a, 1316b e a informação de lado 1318 (por exemplo, em uma forma codificada) e fornece, com base na mesma, uma pluralidade de sinais de áudio de saída 1352a a 1352n. O decodificador 1350 também pode receber uma informação de interação de usuário 1354, que pode compreender um ou mais parâmetros de renderização R (que pode definir uma matriz de renderização). O decodificador 1350 compreende um separador de objeto paramétrico 1360, um processador de informação de lado 1370 e um renderizador 1380. Um processador de informação de lado 1370 recebe a informação de lado 1318 e fornece, com base na mesma, uma informação de controle 1372 para o separador de objeto paramétrico 1360. O separador de objeto paramétrico 1360 fornece uma pluralidade de sinais de objeto 1362a a 1362n com base nos sinais de downmix 1360a, 1360b e a informação de controle 1372, que é derivada a partir da informação do lado 1318 por um processador de informação de lado 1370. Por exemplo, o separador de objeto pode realizar uma decodificação dos sinais de downmix codificados e uma separação de objeto. O renderizador 1380 renderiza os sinais de objeto reconstruídos 1362a a 1362n, para assim obter os sinais de áudio de saída 1352a a 1352n.

[185] No que segue, a funcionalidade do conceito de downmix/upmix do parâmetro com base em MMSE será discutida.

[186] O processamento de downmix/upmix paramétrico geral é realizado de uma maneira seletiva de tempo/frequência e pode ser descrito como uma sequência das etapas a seguir:

[187] O “codificador” 1310 é fornecido com “objetos de áudio” de entrada x e “parâmetros de mistura” D . O “misturador” 1320 faz o downmix dos

“objetos de áudio” x em um número de “sinais de downmix” y usando “parâmetros de mistura” D (ex., ganhos de downmix). O “estimador de info de lado” extrai a informação de lado 1318 que descreve as características dos “objetos de áudio” de entrada x (ex., propriedades de covariância).

[188] Os “sinais de downmix” y e informação de lado são transmitidos ou armazenados. Estes sinais de áudio de downmix podem ser comprimidos adicionalmente usando codificadores de áudio (como MPEG-1/2 Camada II ou III, MPEG-2/4 Advanced Audio Coding (AAC), MPEG Unified Speech e Audio Coding (USAC), etc.). A informação de lado também pode ser representada e codificada eficientemente (ex., como relações codificadas de menos perda das energias do objeto e coeficientes de correlação do objeto).

[189] O “decodificador” 1350 restaura os “objetos de áudio” originais a partir dos “sinais de downmix” decodificados usando a informação de lado transmitida 1318. O “processador de info de lado” 1370 estima os coeficientes não misturados 1372 para serem aplicados aos “sinais de downmix” dentro do “separador de objeto paramétrico” 1360 para obter a reconstrução de objeto paramétrica de x . Os “objetos de áudio” reconstruídos 1362a a 1362n são renderizados para uma cena alvo (multicanal), representada pelos canais de saída $\hat{Z}$, pela aplicação de “parâmetros de renderização” R , 1354.

[190] Além disso, deve-se observar que as funcionalidades descritas com relação ao codificador 1310 e o decodificador 1350 podem ser usadas nos outros codificadores de áudio e decodificadores de áudio descritos aqui também.

[191] 13.3. Princípio de ortogonalidade da Estimativa do erro quadrático médio mínimo

[192] O princípio de ortogonalidade é uma propriedade principal dos estimadores de MMSE. Considere dois espaços de Hilbert W e V , com V gerado por um conjunto de vetores y_i , e um vetor $x \in W$. Se alguém deseja encontrar uma estimativa $\hat{x} \in V$ que irá aproximar x como uma combinação linear de vetores $y_i \in V$, enquanto minimiza o erro quadrático médio, então o

vetor de erro será ortogonal no espaço gerado pelos vetores y_i :

[193] $(x - \hat{x})y^H = 0,$

[194] Como uma consequência, o erro de estimativa e a própria estimativa são ortogonais:

[195] $(x - \hat{x})\hat{x}^H = 0.$

[196] Geometricamente pode-se visualizar isso pelos exemplos mostrados na Fig. 14.

[197] A Fig. 14 mostra uma representação esquemática para o princípio de ortogonalidade no espaço tridimensional. Como pode ser visto, um espaço do vetor é gerado pelos vetores y_1, y_2 . Um vetor x é igual a uma soma de um vetor $\hat{x}$ e um vetor de diferença (ou vetor de erro) e . Como pode ser visto, o vetor de erro e é ortogonal ao espaço do vetor (ou plano) V gerado pelos vetores y_1 e y_2 . Portanto, o vetor $\hat{x}$ pode ser considerado como uma melhor aproximação de x dentro do espaço do vetor V .

[198] 13.4. Erro de reconstrução paramétrico

[199] Na definição de uma matriz compreendendo N sinais: $\mathbf{x}$ e denotando o erro de estimativa com $\mathbf{X}_{Error}$, as identidades a seguir podem ser formuladas. O sinal original pode ser representado como uma soma de uma reconstrução paramétrica $\hat{\mathbf{X}}$ e o erro de reconstrução $\mathbf{X}_{Error}$ como

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{X}_{Error}.$$

[200] Por causa do princípio de ortogonalidade, a matriz de covariância dos sinais originais $\mathbf{E}_X = \mathbf{X}\mathbf{X}^H$ pode ser formulada como uma soma da matriz de covariância dos sinais reconstruídos $\hat{\mathbf{X}}\hat{\mathbf{X}}^H$ e a matriz de covariância dos erros de estimativa $\mathbf{X}_{Error}\mathbf{X}_{Error}^H$ como

$$\begin{aligned}\mathbf{E}_X = \mathbf{X}\mathbf{X}^H &= (\hat{\mathbf{X}} + \mathbf{X}_{Error})(\hat{\mathbf{X}} + \mathbf{X}_{Error})^H = \hat{\mathbf{X}}\hat{\mathbf{X}}^H + \mathbf{X}_{Error}\mathbf{X}_{Error}^H + \hat{\mathbf{X}}\mathbf{X}_{Error}^H + \mathbf{X}_{Error}\hat{\mathbf{X}}^H = \\ &= \hat{\mathbf{X}}\hat{\mathbf{X}}^H + \mathbf{X}_{Error}\mathbf{X}_{Error}^H\end{aligned}$$

[201] Quando os objetos de entrada $\mathbf{x}$ não estão no espaço abrangido

pelos canais de downmix (ex., o número de canais de downmix é menor que o número de sinais de entrada) e os objetos de entrada não podem ser representados como combinações lineares dos canais de downmix, os algoritmos baseados no MMSE introduzem a imprecisão de reconstrução

$$\mathbf{X}_{\text{Erro}} \mathbf{X}_{\text{Erro}}^H$$

[202] 13.5. Correlação inter-objeto

[203] No sistema auditivo, a covariância cruzada (coerência/correlação) é intimamente relacionada com a percepção de envelopamento, de estar cercado pelo som, e à largura percebida de uma fonte de som. Por exemplo nos sistemas baseados em SAOC os parâmetros de Correlação Inter-Objeto (IOC) são usados para caracterização desta propriedade:

$$IOC(i, j) = \frac{\mathbf{E}_X(i, j)}{\sqrt{\mathbf{E}_X(i, i) \mathbf{E}_X(j, j)}}$$

[204] Vamos considerar um exemplo da reprodução de uma fonte de som que usa dois sinais de áudio. Se o valor IOC está próximo de um, o som é percebido como uma fonte de ponto bem localizada. Se o valor IOC está próximo de zero, a largura percebida da fonte de som aumenta e para casos extremos pode ainda ser percebida como duas fontes distintas [Blauert, Chapter 3].

[205] 13.6. Compensação para Imprecisão de Reconstrução

[206] No caso da reconstrução paramétrica imperfeita, o sinal de saída pode exibir uma energia inferior se comparado aos objetos originais. O erro nos elementos diagonais da matriz de covariância pode resultar em diferenças audíveis de nível e erro nos elementos fora da diagonal em uma imagem de som espacial distorcida (em comparação com a saída de referência ideal). O método proposto tem a finalidade de resolver este problema.

[207] No MPEG Surround (MPS), por exemplo, essa questão é tratada somente para alguns cenários de processamento com base em canal específico, ou seja, para downmix mono/estéreo e configurações de saída estática limitadas (ex., mono, estéreo, 5.1, 7.1, etc). Em tecnologias orientadas

por objeto, como SAOC, que também usa o downmix mono/estéreo este problema é tratado pela aplicação da renderização pós-processamento MPS para a configuração de saída 5.1 somente.

[208] As soluções existentes são limitadas às configurações de saída padrão e número fixo de canais de entrada/saída. Ou seja, elas são realizadas como consequente aplicação de vários blocos que implementam apenas métodos de decorrelação de canal “mono-a-estéreo” (ou “estéreo-para-três”).

[209] Portanto, uma solução geral (ex., método de correção de nível de energia e propriedades de correlação) para compensação da imprecisão da reconstrução paramétrica é desejada, que pode ser aplicada para um número flexível de canais de downmix/saída e ajustes de configuração de saída arbitrários.

[210] 13.7. Conclusões

[211] Para concluir, uma visão geral sobre a notação foi fornecida. Além disso, um sistema de separação paramétrico foi descrito no qual as modalidades de acordo com a invenção são baseadas. Além disso, foi descrito que o princípio de ortogonalidade se aplica à estimativa do erro quadrático médio mínimo. Além disso, uma equação para cálculo de uma matriz de covariância $\mathbf{E}_x$ foi fornecida que se aplica na presença de um erro de reconstrução $\mathbf{X}_{\text{Erro}}$. Também, a relação entre os chamados valores de correlação inter-objeto e os elementos de uma matriz de covariância $\mathbf{E}_x$ foi fornecida, que pode ser aplicada, por exemplo, nas modalidades de acordo com a invenção para derivar as características de covariância desejadas (ou características de correlação) a partir dos valores de correlação inter-objeto (que podem ser incluídos na informação de lado paramétrica), e possivelmente formam as diferenças de nível de objeto. Além disso, foi descrito que as características dos sinais de objeto reconstruídos podem diferir das características desejadas por causa de uma reconstrução imperfeita. Além disso, foi descrito que as soluções existentes para lidar com o problema são limitadas a algumas configurações de saída específicas e dependem de uma

combinação específica dos blocos padrão, que torna as soluções convencionais inflexíveis.

[212] 14. Modalidade de acordo com a Fig. 15

[213] 14.1. Visão geral do conceito

[214] As modalidades de acordo com a invenção estendem os métodos de reconstrução paramétricos MMSE usados nos esquemas de separação de áudio paramétricos com uma solução de descorrelação para um número arbitrário de canais de downmix/upmix. As modalidades de acordo com a invenção, como, por exemplo, o aparelho da invenção e o método da invenção, pode compensar a perda de energia durante uma reconstrução paramétrica e restaurar as propriedades de correlação dos objetos estimados.

[215] A Fig. 15 fornece uma visão geral do conceito de downmix/upmix paramétrico com uma trajetória de descorrelação integrada. Em outras palavras, a Fig. 15 mostra, na forma de um diagrama em bloco esquemático, um sistema de reconstrução paramétrico com descorrelação aplicada à saída renderizada.

[216] O sistema de acordo com a Fig. 15 compreende um codificador 1510, que é substancialmente idêntico ao codificador 1310 de acordo com a Fig. 13. O codificador 1510 recebe uma pluralidade de sinais de objeto 1512a a 1512n, e fornece com base no mesmo, um ou mais sinais de downmix 1516a, 1516b, assim como uma informação de lado 1518. Os sinais de downmix 1516a, 1516b podem ser substancialmente idênticos aos sinais de downmix 1316a, 1316b e podem ser designados com **Y**. A informação de lado 1518 pode ser substancialmente idêntica à informação de lado 1318. Entretanto, a informação de lado pode, por exemplo, compreender um parâmetro do modo de descorrelação ou um parâmetro do método de descorrelação, ou um parâmetro de complexidade de descorrelação. Além disso, o codificador 1510 pode receber parâmetros de mistura 1514.

[217] O sistema de reconstrução paramétrico também compreende uma transmissão e/ou armazenamento dos um ou mais sinais de downmix 1516a,

1516b e da informação de lado 1518, sendo que a transmissão e/ou armazenamento é designada com 1540, e sendo que os um ou mais sinais de downmix 1516a, 1516b e a informação de lado 1518 (que pode incluir a informação de lado paramétrica) pode ser codificada.

[218] Além disso, o sistema de reconstrução paramétrico de acordo com a Fig. 15 compreende um decodificador 1550, que é configurado para receber os um ou mais sinais de downmix transmitidos ou armazenados (possivelmente codificados) 1516a, 1516b e a informação de lado transmitida ou armazenada (possivelmente codificada) 1518 e para fornecer, com base na mesma, sinais de áudio de saída 1552a a 1552n. O decodificador 1550 (que pode ser considerado como um decodificador de áudio multicanal) compreende um separador de objeto paramétrico 1560 e um processador de informação de lado 1570. Além disso, o decodificador 1550 compreende um renderizador 1580, um descorrelacionador 1590 e um misturador 1598.

[219] O separador de objeto paramétrico 1560 é configurado para receber os um ou mais sinais de downmix 1516a, 1516b e a informação de controle 1572, que é fornecida pelo processador de informação de lado 1570 com base na informação de lado 1518, e para fornecer, com base na mesma, sinais de objeto 1562a a 1562n, que também são designados com $\hat{X}$, e que podem ser considerados como sinais de áudio decodificados. A informação de controle 1572 pode, por exemplo, compreender coeficientes não misturados para serem aplicados aos sinais de downmix (por exemplo, aos sinais de downmix decodificados derivados dos sinais de downmix codificados 1516a, 1516b) dentro do separador de objeto paramétrico para obter sinais de objeto reconstruídos (por exemplo, os sinais de áudio decodificados 1562a a 1562n). O renderizador 1580 renderiza os sinais de áudio decodificados 1562a a 1562n (que podem ser sinais de objeto reconstruídos, e que podem, por exemplo, corresponder aos sinais de entrada de objeto 1512a a 1512n), para assim obter uma pluralidade de sinais de áudio renderizados 1582a a 1582n. Por exemplo, o renderizador 1580 pode considerar parâmetros de renderização R , que

podem, por exemplo, ser fornecidos pela interação do usuário e que podem, por exemplo, definir uma matriz de renderização. Entretanto, alternativamente, os parâmetros de renderização podem ser retirados da representação codificada (que pode incluir os sinais de downmix codificados 1516a, 1516b e a informação de lado codificada 1518).

[220] O descorrelacionador 1590 é configurado para receber os sinais de áudio renderizados 1582a a 1582n e para fornecer, com base na mesma, sinais de áudio descorrelacionados 1592a a 1592n, que também são designados com **W**. O misturador 1598 recebe os sinais de áudio renderizados 1582a a 1582n e os sinais de áudio descorrelacionados 1592a a 1592n, e combina os sinais de áudio renderizados 1582a a 1582n e os sinais de áudio descorrelacionados 1592a a 1592n, para assim obter os sinais de áudio de saída 1552a a 1552n. O misturador 1598 também pode usar a informação de controle 1574 que é derivada por um processador de informação de lado 1570 a partir da informação de lado codificada 1518, como será descrito abaixo.

[221] 14.2. Função do Descorrelacionador

[222] No que segue, alguns detalhes com relação ao descorrelacionador 1590 serão descritos. Entretanto, deve-se observar que diferentes conceitos de descorrelacionador podem ser usados, alguns dos quais serão descritos abaixo.

[223] Em uma modalidade, a função do descorrelacionador $w = F_{decorr}(\hat{z})$ fornece um sinal de saída w que é ortogonal ao sinal de entrada $\hat{z}$ ($E\{w\hat{z}^H\} = 0$). O sinal de saída w tem propriedades de envelope espectrais e temporais iguais (ao sinal de entrada $\hat{z}$) (ou ao menos propriedades similares). Além disso, o sinal w é percebido da mesma forma e tem a mesma (ou similar) qualidade subjetiva que o sinal de entrada $\hat{z}$ (ver, por exemplo, [SAOC2]).

[224] No caso de vários sinais de entrada, é benéfico se a função de descorrelação produz várias saídas que são mutualmente ortogonais (i.e., $W_i = F_{decorr}(\hat{Z}_i)$, de modo que $W_i \hat{Z}_j^H = 0$ para todos os i e j , e $W_i W_j^H = 0$ para

$i \neq j$).

[225] A especificação exata para a implementação da função do descorrelacionador está fora do escopo desta descrição. Por exemplo, o banco de vários descorrelacionadores baseados no filtro de Resposta de Impulso Infinito (IIR) especificado no Padrão MPEG Surround pode ser usado para fins de descorrelação [MPS].

[226] Assume-se que os descorrelacionadores genéricos descritos nesta descrição são ideais. Isto implica que (além dos requisitos de percepção) a saída de cada descorrelacionador é ortogonal em sua entrada e na saída de todos os outros descorrelacionadores. Portanto, para a dada entrada $\hat{Z}$ com covariância $\mathbf{E}_{\hat{Z}} = \hat{Z}\hat{Z}^H$ e saída $\mathbf{W} = F_{\text{descorr}}(\hat{Z})$ as seguintes propriedades das matrizes de covariância se mantêm:

$$\mathbf{E}_{\mathbf{W}}(i, i) = \mathbf{E}_{\hat{Z}}(i, i), \mathbf{E}_{\mathbf{W}}(i, j) = 0, \text{ para } i \neq j, \hat{Z}\mathbf{W}^H = \mathbf{W}\hat{Z}^H = \mathbf{0}.$$

[227] A partir destas relações, resulta que

$$(\hat{Z} + \mathbf{W})(\hat{Z} + \mathbf{W})^H = \mathbf{E}_{\hat{Z}} + \hat{Z}\mathbf{W}^H + \mathbf{W}\hat{Z}^H + \mathbf{E}_{\mathbf{W}} = \mathbf{E}_{\hat{Z}} + \mathbf{E}_{\mathbf{W}}.$$

[228] A saída do descorrelacionador $\mathbf{W}$ pode ser usada para compensar a imprecisão de previsão em um estimador MMSE (lembrando que o erro de previsão é ortogonal aos sinais previstos) pelo uso dos sinais previstos como as entradas.

[229] Deve-se ainda observar que os erros de previsão não são em um caso geral ortogonais dentre eles mesmos. Deste modo, um objetivo do conceito inventivo (ex., método) é criar uma mistura do sinal “seco” (i.e., entrada do descorrelacionador) (ex., sinais de áudio renderizados 1582a a 1582n) e sinal “molhado” (i.e., saída do descorrelacionador) (ex., sinais de áudio descorrelacionados 1592a a 1592n), de modo que a matriz de covariância da mistura resultante (ex., sinais de áudio de saída 1552a a 1552n) se torna similar à matriz de covariância da saída desejada.

[230] Além disso, deve-se observar que uma redução da complexidade para a unidade de descorrelação pode ser usada, que será descrita em detalhe abaixo, e que pode trazer consigo algumas imperfeições do sinal descorrelacionado, o que pode, entretanto, ser aceitável.

[231] 14.3. Correção da Covariância de Saída usando Sinais descorrelacionados

[232] No que segue, um conceito será descrito para ajustar as características de covariância dos sinais de áudio de saída 1552a a 1552n para obter uma impressão auditiva razoavelmente boa.

[233] O método proposto para a correção do erro de covariância de saída compõe o sinal de saída $\tilde{\mathbf{Z}}$ (ex., os sinais de áudio de saída 1552a a 1552n) como uma soma ponderada do sinal parametricamente reconstruído $\hat{\mathbf{Z}}$ (ex., os sinais de áudio renderizados 1582a a 1582n) e sua parte descorrelacionada $\mathbf{W}$. Esta soma pode ser representada como segue

[234]
$$\tilde{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{Z}} + \mathbf{M}\mathbf{W}.$$

[235] Entretanto, deve-se observar que esta equação pode ser considerada uma formulação mais geral. Uma alteração pode opcionalmente ser aplicada à fórmula acima que é válida (ou que pode ser feita) para todos os “métodos simplificados” descritos aqui.

[236] As matrizes de mistura $\mathbf{P}$ aplicadas ao sinal direto $\hat{\mathbf{Z}}$ e $\mathbf{M}$ aplicadas ao sinal descorrelacionado $\mathbf{W}$ tem a estrutura a seguir (com $N = N_{CUpmix}$,

sendo que N_{CUpmix} designa um número de sinais de áudio renderizados, que pode ser igual a um número de sinais de áudio de saída):

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,N} \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,N} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \cdots & m_{N,N} \end{bmatrix}.$$

[237] Aplicar a notação para a matriz combinada $\mathbf{F} = [\mathbf{P} \quad \mathbf{M}]$ e sinal

$\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}$ rende:

[238] $\tilde{\mathbf{Z}} = \mathbf{F}\mathbf{S}$.

[239] Alternativamente, entretanto, a equação

[240] $\tilde{\mathbf{Z}} = \tilde{\mathbf{F}}\mathbf{S}$

[241] pode ser aplicada, como será descrito em mais detalhe abaixo.

[242] Usando esta representação, a matriz de covariância $\mathbf{E}_{\tilde{\mathbf{Z}}}$ do sinal de saída $\tilde{\mathbf{Z}}$ é definida como

[243] $\mathbf{E}_{\tilde{\mathbf{Z}}} = \mathbf{F}\mathbf{E}_S\mathbf{F}^H$.

[244] A covariância alvo $\mathbf{C}$ da cena de saída renderizada criada idealmente é definida como

[245] $\mathbf{C} = \mathbf{R}\mathbf{E}_X\mathbf{R}^H$.

[246] A matriz de mistura $\mathbf{F}$ é calculada de modo que a matriz de covariância $\mathbf{E}_{\tilde{\mathbf{Z}}}$ da saída final se aproxima, ou se iguala, à covariância alvo $\mathbf{C}$ como

[247] $\mathbf{E}_{\tilde{\mathbf{Z}}} \approx \mathbf{C}$.

[248] A matriz de mistura $\mathbf{F}$ é calculada, por exemplo, como uma função das quantidades conhecidas $\mathbf{F} = \mathbf{F}(\mathbf{E}_S, \mathbf{E}_X, \mathbf{R})$ as

[249] $\mathbf{F} = \left(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H \right) \mathbf{H} \left(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H \right),$

[250] onde as matrizes $\mathbf{U}$, $\mathbf{T}$ e $\mathbf{V}$, $\mathbf{Q}$ podem ser determinadas, por exemplo, usando Decomposição de Valor Singular (SVD) das matrizes de covariância $\mathbf{E}_S$ e $\mathbf{C}$ produzindo

[251] $\mathbf{C} = \mathbf{U}\mathbf{T}\mathbf{U}^H, \mathbf{E}_S = \mathbf{V}\mathbf{Q}\mathbf{V}^H$.

[252] A matriz de protótipo $\mathbf{H}$ pode ser escolhida de acordo com as ponderações desejadas para as trajetórias de sinal direto e descorrelacionado.

[253] Por exemplo, uma possível matriz de protótipo $\mathbf{H}$ pode ser determinada como

[254]
$$\mathbf{H} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & b_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 & 0 & b_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N,N} & 0 & 0 & \dots & b_{N,N} \end{bmatrix}, \quad \text{onde}$$

$a_{i,i}^2 + b_{i,i}^2 = 1.$

[255] No que segue, alguns desvios matemáticos para a estrutura da matriz geral $\mathbf{F}$ serão fornecidos.

[256] Em outras palavras, o desvio da matriz de mistura $\mathbf{F}$ para uma solução geral será descrito a seguir.

[257] As matrizes de covariância $\mathbf{E}_s$ e $\mathbf{C}$ podem ser expressas usando, ex., Decomposição de Valor Singular (SVD) como

[258]
$$\mathbf{E}_s = \mathbf{V}\mathbf{Q}\mathbf{V}^H, \quad \mathbf{C} = \mathbf{U}\mathbf{T}\mathbf{U}^H.$$

[259] com $\mathbf{T}$ e $\mathbf{Q}$ sendo matrizes diagonais com os valores singulares de $\mathbf{C}$ e $\mathbf{E}_s$ respectivamente, e $\mathbf{U}$ e $\mathbf{V}$ sendo matrizes unitárias contendo os vetores singulares correspondentes.

[260] Observe que a aplicação da triangulação de Schur ou decomposição de Autovalor (ao invés de SVD) leva a resultados similares (ou mesmo resultados idênticos se as matrizes diagonais $\mathbf{Q}$ e $\mathbf{T}$ são restritas aos valores positivos).

[261] A aplicação desta decomposição ao requisito $\mathbf{E}_z \approx \mathbf{C}$, rende (ao menos aproximadamente)

[262]
$$\mathbf{C} = \mathbf{F}\mathbf{E}_s\mathbf{F}^H,$$

[263]
$$\mathbf{U}\mathbf{T}\mathbf{U}^H = \mathbf{F}\mathbf{V}\mathbf{Q}\mathbf{V}^H\mathbf{F}^H,$$

$$[264] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = \mathbf{F}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)\mathbf{F}^H,$$

$$[265] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = (\mathbf{F}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H\mathbf{F}^H),$$

$$[266] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)^H = (\mathbf{F}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{F}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)^H.$$

[267] De forma a tomar cuidado com a dimensionalidade das matrizes de covariância, é necessária regularização em alguns casos. Por exemplo, a

matriz de protótipo $\mathbf{H}$ de tamanho $N_{CUpmix} \times 2N_{CUpmix}$, com a propriedade que $\mathbf{H}\mathbf{H}^H = \mathbf{I}_{N_{CUpmix}}$ pode ser aplicada

$$[268] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)\mathbf{H}\mathbf{H}^H(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = \mathbf{F}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)\mathbf{F}^H,$$

$$[269] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)\mathbf{H} = \mathbf{F}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H).$$

[270] Resulta que a matriz de mistura $\mathbf{F}$ pode ser determinada como

$$[271] \quad \mathbf{F} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)\mathbf{H}(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H).$$

[272] A matriz de protótipo $\mathbf{H}$ é escolhida de acordo com as ponderações desejadas para as trajetórias de sinal direto e descorrelacionado. Por exemplo, a possível matriz de protótipo $\mathbf{H}$ pode ser determinada como

$$[273] \quad \mathbf{H} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & b_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 & 0 & b_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N,N} & 0 & 0 & \dots & b_{N,N} \end{bmatrix}, \quad \text{onde}$$

$$a_{i,i}^2 + b_{i,i}^2 = 1.$$

[274] Dependendo da condição da matriz de covariância $\mathbf{E}_S$ dos sinais combinados, a última equação pode precisar incluir alguma regularização, mas, de outra maneira, deve ser numericamente estável.

[275] Para concluir, um conceito foi descrito para derivar os sinais de áudio de saída (representados pela matriz $\tilde{\mathbf{Z}}$, ou equivalentemente, pelo vetor

$\hat{z}$) com base nos sinais de áudio renderizados (representados pela matriz $\hat{\mathbf{Z}}$, ou equivalentemente, vetor $\hat{z}$) e os sinais de áudio descorrelacionados (representados pela matriz $\mathbf{W}$, ou equivalentemente, vetor w). Como pode ser visto, duas matrizes de mistura $\mathbf{P}$ e $\mathbf{M}$ da estrutura de matriz geral são comumente determinadas. Por exemplo, uma matriz combinada $\mathbf{F}$, como definida acima, pode ser determinada, de modo que uma matriz de covariância $\mathbf{E}_z$ dos sinais de áudio de saída 1552a a 1562n se aproxime, ou se iguale à covariância desejada (também designada como covariância alvo) $\mathbf{C}$. A matriz de covariância desejada $\mathbf{C}$ pode, por exemplo, ser derivada com base no conhecimento da matriz de renderização $\mathbf{R}$ (que pode ser fornecido pela interação do usuário, por exemplo) e com base em um conhecimento da matriz de covariância do objeto $\mathbf{E}_x$, que pode, por exemplo, ser derivada com base na informação de lado codificada 1518. Por exemplo, a matriz de covariância do objeto $\mathbf{E}_x$ pode ser derivada usando os valores de correlação inter-objeto IOC, que são descritos acima, e que podem ser incluídos na informação de lado codificada 1518. Deste modo, a matriz de covariância alvo $\mathbf{C}$ pode, por exemplo, ser fornecida por um processador de informação de lado 1570 como a informação 1574, ou como parte da informação 1574.

[276] Entretanto, alternativamente, um processador de informação de lado 1570 também pode fornecer diretamente a matriz de mistura $\mathbf{F}$ como a informação 1574 para o misturador 1598.

[277] Além disso, uma regra de cálculo para a matriz de mistura $\mathbf{F}$ foi descrita, que usa a Decomposição de Valor Singular. Entretanto, deve-se observar que há alguns graus de liberdade, já que as entradas $a_{i,j}$ e $b_{i,j}$ da matriz de protótipo $\mathbf{H}$ podem ser escolhidas. De preferência, as entradas da matriz de protótipo $\mathbf{H}$ são escolhidas por estarem em algum lugar entre 0 e 1. Se os valores $a_{i,j}$ são escolhidos por estarem próximos de um, haverá uma mistura significativa dos sinais de áudio de saída renderizados, enquanto o impacto dos sinais de áudio descorrelacionados é comparativamente pequeno,

o que pode ser desejável em algumas situações. Entretanto, em algumas outras situações pode ser mais desejável ter um impacto comparativamente grande dos sinais de áudio descorrelacionados, enquanto há somente uma mistura fraca entre sinais de áudio renderizados. Neste caso, os valores $b_{i,j}$ são tipicamente escolhidos para serem maiores que $a_{i,j}$. Deste modo, o decodificador 1550 pode ser adaptado às necessidades pela escolha apropriada das entradas da matriz de protótipo $\mathbf{H}$.

[278] 14.4. Métodos simplificados para Correção de covariância de saída

[279] Nesta seção, duas estruturas alternativas para a matriz de mistura $\mathbf{F}$ mencionada acima são descritas juntamente com algoritmos exemplares para determinar seus valores. As duas alternativas são projetadas para diferentes conteúdos de entrada (ex., conteúdo de áudio):

[280] Método de ajuste de covariância para conteúdo altamente correlacionado (ex., entrada com base em canal com alta correlação entre diferentes pares de canal).

[281] Método de compensação de energia para sinais de entrada independentes (ex., entrada com base no objeto, assumido normalmente independente).

[282] 14.4.1. Método de ajuste de covariância (A)

[283] Levando em consideração que o sinal $\hat{\mathbf{Z}}$ (ex., os sinais de áudio renderizados 1582a a 1582n) já é ideal no sentido de MMSE, geralmente não é aconselhável modificar as reconstruções paramétricas $\hat{\mathbf{Z}}$ (ex., os sinais de áudio de saída 1552a a 1552n) para melhorar as propriedades de covariância da saída $\tilde{\mathbf{Z}}$ porque isto pode afetar a qualidade de separação.

[284] Se somente a mistura dos sinais descorrelacionados $\mathbf{W}$ é manipulada, a matriz de mistura $\mathbf{P}$ pode ser reduzida a uma matriz de identidade (ou um múltiplo da mesma). Deste modo, este método simplificado pode ser descrito pela definição de

$$[285] \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,N} \\ m_{2,1} & m_{2,2} & \dots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \dots & m_{N,N} \end{bmatrix}.$$

[286] A saída final do sistema pode ser representada como

$$[287] \quad \tilde{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{M}\mathbf{W}.$$

[288] Consequentemente a covariância da saída final do sistema pode ser representada como:

$$[289] \quad \mathbf{E}_{\tilde{\mathbf{Z}}} = \mathbf{E}_{\hat{\mathbf{Z}}} + \mathbf{M}\mathbf{E}_W\mathbf{M}^H$$

[290] A diferença Δ_E entre a matriz de covariância de saída ideal (ou desejada) $\mathbf{C}$ e a matriz de covariância $\mathbf{E}_{\hat{\mathbf{Z}}}$ da reconstrução paramétrica renderizada (ex., dos sinais de áudio renderizados) é dada por

$$[291] \quad \Delta_E = \mathbf{C} - \mathbf{E}_{\hat{\mathbf{Z}}}.$$

[292] Portanto, a matriz de mistura $\mathbf{M}$ é determinada de modo que

$$[293] \quad \Delta_E \approx \mathbf{M}\mathbf{E}_W\mathbf{M}^H.$$

[294] A matriz de mistura $\mathbf{M}$ é calculada de modo que a matriz de covariância dos sinais descorrelacionados misturados $\mathbf{M}\mathbf{W}$ é igual ou se aproxima da diferença de covariância entre a covariância desejada e a covariância dos sinais secos (ex., dos sinais de áudio renderizados). Consequentemente a covariância da saída final irá se aproximar da covariância alvo $\mathbf{E}_Z \approx \mathbf{C}$:

$$[295] \quad \mathbf{M} = \left(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H \right) \left(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H \right),$$

[296] onde as matrizes $\mathbf{U}$, $\mathbf{T}$ e $\mathbf{V}$, $\mathbf{Q}$ podem ser determinadas, por exemplo, usando Decomposição de Valor Singular (SVD) das matrizes de covariância Δ_E e $\mathbf{E}_W$ produzindo

$$[297] \quad \Delta_E = \mathbf{U}\mathbf{T}\mathbf{U}^H, \quad \mathbf{E}_W = \mathbf{V}\mathbf{Q}\mathbf{V}^H.$$

[298] Esta abordagem garante a boa reconstrução de correlação cruzada maximizando o uso da saída seca (ex., dos sinais de áudio renderizados 1582a a 1582n) e utiliza a liberdade de mistura dos sinais descorrelacionados somente. Em outras palavras, não há mistura entre diferentes sinais de áudio renderizados permitida quando se combina os sinais de áudio renderizados (ou uma versão reduzida dos mesmos) com os um ou mais sinais de áudio descorrelacionado. Entretanto, é permitido que um dado sinal descorrelacionado seja combinado, com um dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída. A combinação é definida, por exemplo, por uma matriz **M** como definido aqui.

[299] No que segue, alguns desvios matemáticos para a estrutura da matriz restrita **F** serão fornecidos.

[300] Em outras palavras, o desvio da matriz de mistura **M** para o método simplificado “A” será explicado.

[301] As matrizes de covariância Δ_E e E_W podem ser expressas usando, ex., Decomposição de Valor Singular (SVD) como

$$\mathbf{[302]} \quad \Delta_E = \mathbf{U} \mathbf{T} \mathbf{U}^H, \quad \mathbf{E}_W = \mathbf{V} \mathbf{Q} \mathbf{V}^H.$$

[303] com **T** e **Q** sendo matrizes diagonais com os valores singulares de Δ_E e E_W respectivamente, e **U** e **V** sendo matrizes unitárias contendo os vetores singulares correspondentes.

[304] Observe, que a aplicação da triangulação de Schur ou decomposição de Autovalor (ao invés de SVD) leva a resultados similares (ou mesmo resultados idênticos se as matrizes diagonais **Q** e **T** são restritas aos valores positivos).

[305] A aplicação desta decomposição ao requisito $E_Z \approx C$, rende (ao menos aproximadamente)

$$\mathbf{[306]} \quad \Delta_E = \mathbf{M} \mathbf{E}_W \mathbf{M}^H,$$

$$[307] \quad \mathbf{U}\mathbf{T}\mathbf{U}^H = \mathbf{M}\mathbf{V}\mathbf{Q}\mathbf{V}^H\mathbf{M}^H,$$

$$[308] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = \mathbf{M}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)\mathbf{M}^H,$$

$$[309] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = (\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H\mathbf{M}^H),$$

$$[310] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)^H = (\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)(\mathbf{M}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H)^H,$$

$$[311] \quad (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) = \mathbf{M}(\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{V}^H).$$

[312] Observando que ambos os lados da equação representam um quadrado de uma matriz, deixamos de lado a quadratura, e resolvemos a matriz completa $\mathbf{M}$.

[313] Resulta que a matriz de mistura $\mathbf{M}$ pode ser determinada como

$$[314] \quad \mathbf{M} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H)(\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H).$$

[315] Este método pode ser derivado do método geral pela definição da matriz de protótipo $\mathbf{H}$ como segue

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

[316] Dependendo da condição da matriz de covariância $\mathbf{E}_w$ dos sinais molhados, a última equação pode precisar incluir alguma regularização, mas, de outra maneira, deve ser numericamente estável.

[317] 14.4.2. Método de compensação de energia (B)

[318] Às vezes (dependendo do cenário de aplicação) não deseja-se permitir a mistura das reconstruções paramétricas (ex., dos sinais de áudio renderizados) ou os sinais descorrelacionados, mas a mistura individualmente de cada sinal parametricamente reconstruído (ex., sinal de áudio renderizado) com seu próprio sinal descorrelacionado somente.

[319] De forma a atingir este requisito, uma restrição adicional deve ser introduzida ao método simplificado “A”. Agora, é preciso que a matriz de mistura $\mathbf{M}$ dos sinais molhados (sinais descorrelacionados) tenha uma forma

diagonal:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \mathbf{M} = \begin{bmatrix} m_{1,1} & 0 & \dots & 0 \\ 0 & m_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_{N,N} \end{bmatrix}.$$

[320] O objetivo principal desta abordagem é usar sinais descorrelacionados para compensar a perda de energia na reconstrução paramétrica (ex., sinal de áudio renderizado), enquanto a modificação fora da diagonal da matriz de covariância do sinal de saída é ignorada, i.e., não há manipulação direta das correlações cruzadas. Portanto, nenhuma dispersão cruzada entre os objetos/canais de saída (ex., entre os sinais de áudio renderizados) é introduzida na aplicação dos sinais descorrelacionados.

[321] Como um resultado, somente a diagonal principal da matriz de covariância alvo (ou matriz de covariância desejada) pode ser alcançada, e as linhas fora da diagonal estão à mercê da precisão da reconstrução paramétrica e dos sinais descorrelacionados adicionados. Este método é mais adequado para aplicações baseadas em objeto somente, nas quais os sinais podem ser considerados como não correlacionados.

[322] A saída final do método (ex., os sinais de áudio de saída) é dada por

$$\tilde{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{M}\mathbf{W} \text{ com uma matriz diagonal } \mathbf{M} \text{ calculada de modo que}$$

as entradas da matriz de covariância correspondentes às energias dos sinais reconstruídos $E_{\tilde{\mathbf{Z}}}(i,i)$ são iguais às energias desejadas

$$E_{\tilde{\mathbf{Z}}}(i,i) = \mathbf{C}(i,i)$$

[323] .

[324] $\mathbf{C}$ pode ser determinada como explicado acima para o caso geral.

[325] Por exemplo, a matriz de mistura $\mathbf{M}$ pode ser derivada diretamente pela divisão das energias desejadas dos sinais de compensação (diferenças entre as energias desejadas (que podem ser descritas pelos elementos

diagonais da matriz de covariância cruzada $\mathbf{C}$) e as energias das reconstruções paramétricas (que podem ser determinadas pelo decodificador de áudio)) com as energias dos sinais descorrelacionados (que podem ser determinadas pelo decodificador de áudio):

$$\mathbf{M}(i, j) = \begin{cases} \sqrt{\min \left(\lambda_{Dec}, \max \left(0, \frac{\mathbf{C}(i, i) - \mathbf{E}_{\hat{z}}(i, i)}{\max(\mathbf{E}_w(i, i), \varepsilon)} \right) \right)} & i = j, \\ 0 & i \neq j. \end{cases}$$

[326] sendo que λ_{Dec} é um limiar não negativo usado para limitar a quantidade de componente descorrelacionado adicionado aos sinais de saída (ex., $\lambda_{Dec} = 4$).

[327] Deve-se observar que as energias podem ser reconstruídas parametricamente (por exemplo, usando OLDs, IOC's e coeficientes de renderização) ou podem ser efetivamente calculadas pelo decodificador (que é tipicamente mais caro computacionalmente).

[328] Este método pode ser derivado do método geral pela definição da matriz de protótipo $\mathbf{H}$ como segue:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

[329] Este método maximiza o uso das saídas renderizadas secas explicitamente. O método é equivalente com a simplificação "A" quando as matrizes de covariância não têm entradas fora da diagonal.

[330] Este método tem uma complexidade computacional reduzida.

[331] Entretanto, deve-se observar que o método de compensação de energia, não implica necessariamente que os termos de correlação cruzada não são modificados. Isso vale somente se nós usamos descorrelacionadores ideais e nenhuma redução da complexidade para a unidade de descorrelação. A ideia do método é recuperar a energia e ignorar as modificações nos termos

cruzados (as mudanças nos termos cruzados não modificarão substancialmente as propriedades de correlação e não afetarão a impressão espacial geral).

[332] 14.5. Requisitos para a Matriz de mistura $\mathbf{F}$

[333] No que segue, será explicado que a matriz de mistura $\mathbf{F}$, um desvio do que foi descrito nas seções 14.3 e 14.4, atende aos requisitos para evitar degradações.

[334] Para evitar degradações na saída, qualquer método para compensar os erros de reconstrução paramétrica deve produzir um resultado com a seguinte propriedade: se a matriz de renderização é igual à matriz de downmix então os canais de saída devem ser iguais (ou ao menos aproximados) aos canais de downmix. O modelo proposto atende esta propriedade. Se a matriz de renderização é igual à matriz de downmix $\mathbf{R} = \mathbf{D}$, a reconstrução paramétrica é dada por

[335] $\hat{\mathbf{Z}} = \mathbf{R}\hat{\mathbf{X}} = \mathbf{D}\hat{\mathbf{X}} = \mathbf{D}\mathbf{G}\mathbf{Y} = \mathbf{D}\mathbf{E}\mathbf{D}^H(\mathbf{D}\mathbf{E}\mathbf{D}^H)^{-1}\mathbf{Y} \approx \mathbf{Y},$

[336] e a matriz de covariância desejada será

[337] $\mathbf{C} = \mathbf{R}\mathbf{E}_X\mathbf{R}^H = \mathbf{D}\mathbf{E}_X\mathbf{D}^H = \mathbf{E}_Y.$

[338] Portanto a equação a ser resolvida para obter a matriz de mistura $\mathbf{F}$ é

[339]
$$\mathbf{E}_Y = \mathbf{F} \begin{bmatrix} \mathbf{E}_Y & \mathbf{0}_{N_{CUpmix}} \\ \mathbf{0}_{N_{CUpmix}} & \mathbf{E}_W \end{bmatrix} \mathbf{F}^H,$$

[340] onde $\mathbf{0}_{N_{UpmixCh}}$ é uma matriz quadrada do tamanho $N_{CUpmix} \times N_{CUpmix}$

de zeros. Resolvendo a equação anterior para $\mathbf{F}$, se obtém:

[341]
$$\mathbf{F} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

[342] Isto significa que os sinais descorrelacionados terão peso zero no somatório, e a saída final será dada pelos sinais secos, que são idênticos aos sinais de downmix

$$[343] \quad \tilde{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{Z}} + \mathbf{M}\mathbf{W} = \hat{\mathbf{Z}} \approx \mathbf{Y}.$$

[344] Como um resultado, o dado requisito para a saída do sistema se igualar ao sinal de downmix neste cenário de renderização é atendido.

[345] 14.6. Estimativa da Matriz de covariância do sinal $\mathbf{E}_S$

[346] Para obter a matriz de mistura $\mathbf{F}$ o conhecimento da matriz de covariância $\mathbf{E}_S$ dos sinais combinados $\mathbf{S}$ é necessário ou ao menos desejável.

[347] Em princípio, é possível estimar a matriz de covariância $\mathbf{E}_S$ diretamente a partir dos sinais disponíveis (ou seja, a partir da reconstrução paramétrica $\hat{\mathbf{Z}}$ e a saída do descorrelacionador $\mathbf{W}$). Embora esta abordagem possa levar a resultados mais precisos, pode não ser prática por causa da complexidade computacional associada. Os métodos propostos usam aproximações paramétricas da matriz de covariância $\mathbf{E}_S$.

[348] A estrutura geral da matriz de covariância $\mathbf{E}_S$ pode ser representada como

$$[349] \quad \mathbf{E}_S = \begin{bmatrix} \mathbf{E}_{\hat{\mathbf{Z}}} & \mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}}^H \\ \mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}} & \mathbf{E}_{\mathbf{W}} \end{bmatrix},$$

[350] onde a matriz $\mathbf{E}_{\hat{\mathbf{Z}}\mathbf{W}}$ é a covariância cruzada entre os sinais diretos $\hat{\mathbf{Z}}$ e descorrelacionados $\mathbf{W}$.

[351] Assumindo que os descorrelacionadores são ideais (i.e., para economia de energia, as saídas sendo ortogonais às entradas, e todas as saídas sendo mutualmente ortogonais), a matriz de covariância $\mathbf{E}_S$ pode ser expressa usando a forma simplificada como

$$[352] \quad \mathbf{E}_S = \begin{bmatrix} \mathbf{E}_{\hat{\mathbf{Z}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{\mathbf{W}} \end{bmatrix}.$$

[353] A matriz de covariância $\mathbf{E}_{\hat{\mathbf{Z}}}$ do sinal parametricamente reconstruído $\hat{\mathbf{Z}}$ pode ser determinada parametricamente como

[354]
$$\mathbf{E}_{\hat{\mathbf{Z}}} = \mathbf{R} \mathbf{E}_{\hat{\mathbf{X}}} \mathbf{R}^H = \mathbf{R} \mathbf{G} \mathbf{D} \mathbf{E}_{\mathbf{X}} \mathbf{D}^H \mathbf{G}^H \mathbf{R}^H .$$

[355] Assume-se que a matriz de covariância $\mathbf{E}_{\mathbf{W}}$ do sinal descorrelacionado $\mathbf{W}$ atende a propriedade de ortogonalidade mútua e contém somente os elementos diagonais de $\mathbf{E}_{\hat{\mathbf{Z}}}$ como segue

[356]
$$\mathbf{E}_{\mathbf{W}}(i, j) = \begin{cases} \mathbf{E}_{\hat{\mathbf{Z}}}(i, i) & \text{para } i = j, \\ 0 & \text{para } i \neq j. \end{cases}$$

[357] Se a suposição da ortogonalidade mútua e/ou economia de energia for violada (ex., no caso quando o número de descorrelacionadores disponível é menor que o número de sinais a serem descorrelacionados), então a matriz de covariância $\mathbf{E}_{\mathbf{W}}$ pode ser estimada como

[358]
$$\mathbf{E}_{\mathbf{W}} = \mathbf{M}_{post} \left[\text{matdiag}(\mathbf{M}_{pre} \mathbf{E}_{\hat{\mathbf{Z}}} \mathbf{M}_{pre}^H) \right] \mathbf{M}_{post}^H .$$

[359] 14.7 Melhoria Opcional: Correção da covariância de saída usando Sinais descorrelacionados e Unidade de ajuste de energia

[360] No que segue, um conceito particularmente vantajoso será descrito, que pode ser combinado com os outros conceitos descritos aqui.

[361] O método proposto para a correção do erro de covariância de saída compõe o sinal de saída como uma soma ponderada de um sinal parametricamente reconstruído $\hat{\mathbf{Z}}$ e sua parte descorrelacionada $\tilde{\mathbf{Z}}$. Esta soma pode ser representada como segue

[362]
$$\tilde{\mathbf{Z}} = \mathbf{P} \hat{\mathbf{Z}} + \mathbf{M} \mathbf{W} . \quad (11)$$

[363] Aplicar a notação para a matriz combinada

[364]
$$\mathbf{F} = [\mathbf{P} \quad \mathbf{M}]$$

[365] e sinal

[366]
$$\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}$$

[367] produz:

[368]
$$\tilde{\mathbf{Z}} = \mathbf{F}\mathbf{S} \quad (I1)$$

[369] Entretanto, deve-se observar que esta equação pode ser considerada uma formulação mais geral. Uma alteração pode opcionalmente ser aplicada à fórmula acima que é válida para todos os “métodos simplificados” descritos aqui.

[370] No que segue, uma funcionalidade será descrita, que pode ser realizada, por exemplo, por uma unidade de ajuste de energia.

[371] Para evitar a introdução de artefatos na saída final, em casos extremos, diferentes restrições podem ser impostas na matriz de mistura $\mathbf{F}$ (ou uma matriz de mistura $\tilde{\mathbf{F}}$). As restrições mencionadas podem ser representadas pelos valores limiares absolutos ou valores limiares relativos com relação à energia e/ou propriedades de correlação dos sinais alvo e/ou reconstruídos parametricamente (ex., sinais de áudio renderizados).

[372] O método descrito nesta seção propõe conseguir isso pela adição de uma etapa de ajuste de energia no bloco de mistura final de saída. A finalidade de tal etapa de processamento é garantir que, após a etapa de mistura com a matriz $\mathbf{F}$ (ou uma matriz de mistura “modificada” $\tilde{\mathbf{F}}$ derivada da mesma), os níveis de energia dos sinais descorrelacionados (molhados) (por exemplo, $\mathbf{A}_{\text{molhado}}\mathbf{M}\mathbf{W}$) e/ou os níveis de energia dos sinais parametricamente reconstruídos (secos) (por exemplo, $\mathbf{A}_{\text{seco}}\mathbf{P}\hat{\mathbf{Z}}$) e/ou os níveis de energia dos sinais de saída finais (por exemplo, $\mathbf{A}_{\text{seco}}\mathbf{P}\hat{\mathbf{Z}} + \mathbf{A}_{\text{molhado}}\mathbf{M}\mathbf{W}$ não excedem determinados valores limiares.

[373] Esta funcionalidade extra pode ser conseguida pela modificação da definição da matriz de mistura combinada $\mathbf{F}$ para ser

[374]
$$\tilde{\mathbf{F}} = \begin{bmatrix} \mathbf{A}_{\text{dry}}\mathbf{P} & \mathbf{A}_{\text{wet}}\mathbf{M} \end{bmatrix}, (I3)$$

[375] sendo que as duas matrizes de ajuste de energia quadradas (ou

diagonais) $\mathbf{A}_{seco}$ e $\mathbf{A}_{molhado}$ (que também podem ser referidas como “matrizes de correção de energia”) são aplicadas nos pesos de mistura (por exemplo, $\mathbf{P}$ e $\mathbf{M}$) dos sinais parametricamente reconstruídos (secos) e descorrelacionados (molhados) respectivamente. Como um resultado, a saída final será

$$\begin{aligned} \hat{\mathbf{Z}} &= \tilde{\mathbf{F}}\mathbf{S} \\ &= \mathbf{A}_{seco}\mathbf{P}\hat{\mathbf{Z}} + \mathbf{A}_{molh}\mathbf{M}\mathbf{W} \end{aligned} \quad (I4)$$

[376]

[377] As matrizes de correção de energia secas e molhadas $\mathbf{A}_{seco}$ e $\mathbf{A}_{molhado}$ são calculadas de modo que a contribuição dos sinais secos e/ou molhados (por exemplo, $\hat{\mathbf{Z}}$ e $\mathbf{W}$) nos níveis dos sinais de saída finais (por exemplo, $\tilde{\mathbf{Z}}$), devido a uma etapa de mistura com a matriz $\tilde{\mathbf{F}}$, não excedem um determinado valor limiar relativo com relação aos sinais parametricamente reconstruídos (por exemplo, $\hat{\mathbf{Z}}$) e/ou sinais descorrelacionados (por exemplo, $\mathbf{W}$) e/ou sinais alvo. Em outras palavras, há, em geral, várias possibilidades para calcular as matrizes de correção.

[378] As matrizes de correção de energia secas e molhadas $\mathbf{A}_{seco}$ e $\mathbf{A}_{molhado}$ podem ser calculadas, por exemplo, como uma função da energia e/ou correlação e/ou propriedades de covariância dos sinais secos (por exemplo, $\hat{\mathbf{Z}}$) e/ou sinais molhados (por exemplo, $\mathbf{W}$) e/ou sinais de saída finais desejados e/ou uma estimativa da matriz de covariância dos sinais de saída secos e/ou molhados e/ou finais após a etapa de mistura. Deve-se observar que as possibilidades mencionadas acima descrevem alguns exemplos de como as matrizes de correção podem ser obtidas.

[379] Uma solução possível é dada pelas seguintes expressões:

$$\mathbf{A}_{seco}(i, j) = \begin{cases} \min \left(1, \sqrt{\max \left(0, \lambda_{seco} \frac{\mathbf{E}_{\hat{\mathbf{Z}}}(i, i)}{\max(\mathbf{C}_{estim}(i, i), \varepsilon)} \right)} \right) & i = j, \\ 0 & i \neq j. \end{cases}$$

[380]

[381] e

$$[382] \quad \mathbf{A}_{\text{molh}}(i, j) = \begin{cases} \min \left(1, \sqrt{\max \left(0, \lambda_{\text{molh}} \frac{\mathbf{E}_{\hat{z}}(i, i)}{\max(\mathbf{C}_{\text{estim}}(i, i), \varepsilon)} \right)} \right) & i = j, \\ 0 & i \neq j. \end{cases}$$

[383] onde λ_{seco} e λ_{molhado} são dois valores limiares que podem ser constantes ou variáveis em tempo/frequência como uma função das propriedades do sinal (ex., energia, correlação, e/ou covariância), ε é uma constante de regularização não negativa pequena (opcional), ex., $\varepsilon = 10^{-9}$, $\mathbf{E}_{\hat{z}}$ representa a informação de covariância e/ou energia dos sinais parametricamente reconstruídos (secos), e $\mathbf{C}_{\text{estim}}$ representa a estimativa da matriz de covariância dos sinais secos ou molhados após a etapa de mistura com matriz $\mathbf{F}$, ou a estimativa da matriz de covariância dos sinais de saída após a etapa de mistura com matriz $\mathbf{F}$, que seria obtida se nenhuma etapa de ajuste de energia como proposto pela presente invenção fosse aplicada (ou redigida de forma diferente, que seria obtida se a unidade de ajuste de energia não fosse usada).

[384] Nas equações acima, a operação “max(.)” no denominador, que fornece o valor máximo de argumentos, $\mathbf{C}_{\text{estim}}(i, i)$ e ε , pode, por exemplo, ser substituída por uma adição de ε ou outro mecanismo para evitar uma divisão por zero.

[385] Por exemplo, $\mathbf{C}_{\text{estim}}$ pode ser dada por:

[386] $\mathbf{C}_{\text{estim}} = \mathbf{M} \mathbf{E}_w \mathbf{M}^H$ - a estimativa da matriz de covariância dos sinais molhados após a etapa de mistura com matriz $\mathbf{M}$.

[387] $\mathbf{C}_{\text{estim}} = \mathbf{P} \mathbf{E}_{\hat{z}} \mathbf{P}^H$ - a estimativa da matriz de covariância dos

sinais secos após a etapa de mistura com matriz $\mathbf{P}$.

[388] $\mathbf{C}_{\text{estim}} = \mathbf{P}\mathbf{E}_{\hat{\mathbf{Z}}}\mathbf{P}^H + \mathbf{M}\mathbf{E}_W\mathbf{M}^H$ - a estimativa da matriz de covariância dos sinais de saída após a etapa de mistura com matriz $\mathbf{F}$.

[389] No que segue, algumas simplificações adicionais serão descritas. Em outras palavras, métodos simplificados para correção de covariância de saída serão descritos.

[390] Levando em consideração que os sinais $\hat{\mathbf{Z}}$ já são ideais no sentido de MMSE, geralmente não é aconselhável modificar as reconstruções paramétricas (sinais secos) $\hat{\mathbf{Z}}$ para melhorar as propriedades de covariância da saída $\tilde{\mathbf{Z}}$ porque isto pode afetar a qualidade de separação.

[391] Se somente a mistura dos sinais descorrelacionados (molhados) $\mathbf{W}$ é manipulada, a matriz de mistura $\mathbf{P}$ pode ser reduzida para uma matriz de identidade. Neste caso, a matriz de ajuste de energia correspondente aos sinais parametricamente reconstruídos (secos) também pode ser reduzida para uma matriz de identidade. Deste modo, este método simplificado pode ser descrito pela definição:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{A}_{\text{seca}} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

[392]

[393] A saída final do sistema pode ser representada como:

$$\tilde{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{A}_{\text{molh}}\mathbf{M}\mathbf{W}$$

[394] 15. Redução da complexidade para a unidade de descorrelação

[395] No que segue, será descrito como a complexidade dos descorrelacionadores usados nas modalidades de acordo com a presente invenção pode ser reduzida.

[396] Deve-se observar que a implementação da função do descorrelacionador é muitas vezes computacionalmente complexa. Em algumas aplicações (ex., soluções de decodificador portátil) limitações sobre o

número de descorrelacionadores podem precisar ser introduzidas devido aos recursos computacionais restritos. Esta seção fornece uma descrição dos meios para redução da complexidade da unidade descorrelacionadora pelo controle do número de descorrelacionadores aplicados (ou descorrelações). A interface da unidade de descorrelação é ilustrada nas Figs. 16 e 17.

[397] A Fig. 16 mostra um diagrama em bloco esquemático de uma unidade de descorrelação simples (convencional). A unidade de descorrelação 1600 de acordo com a Fig. 6 é configurada para receber N sinais de entrada do descorrelacionador 1610a a 1610n, como por exemplo sinais de áudio renderizados $\hat{Z}$. Além disso, a unidade de descorrelação 1600 fornece N sinais de saída do descorrelacionador 1612a a 1612n. A unidade de descorrelação 1600 pode, por exemplo, compreender N descorrelacionadores individuais (ou funções de descorrelação) 1620a a 1620n. Por exemplo, cada um dos descorrelacionadores individuais 1620a a 1620n pode fornecer um dos sinais de saída do descorrelacionador 1612a a 1612n com base em um associado dos sinais de entrada do descorrelacionador 1610a a 1610n. Portanto, N descorrelacionadores individuais, ou funções de descorrelação, 1620a a 1620n podem ser necessários para fornecer os N sinais descorrelacionados 1612a a 1612n com base nos N sinais de entrada do descorrelacionador 1610a a 1610n.

[398] Entretanto, a Fig. 17 mostra um diagrama em bloco esquemático de uma unidade de descorrelação de complexidade reduzida 1700. A unidade de descorrelação de complexidade reduzida 1700 é configurada para receber N sinais de entrada do descorrelacionador 1710a a 1710n e para fornecer, com base na mesma, N sinais de saída do descorrelacionador 1712a a 1712n. Por exemplo, os sinais de entrada do descorrelacionador 1710a a 1710n podem ser sinais de áudio renderizados $\hat{Z}$, e os sinais de saída do descorrelacionador 1712a a 1712n podem ser sinais de áudio descorrelacionados W .

[399] O descorrelacionador 1700 compreende um pré-misturador (ou equivalentemente, uma funcionalidade de pré-mistura) 1720 que é configurado

para receber o primeiro conjunto de N sinais de entrada do descorrelacionador 1710a a 1710n e para fornecer, com base na mesma, um segundo conjunto de K sinais de entrada do descorrelacionador 1722a a 1722k. Por exemplo, o pré-misturador 1720 pode realizar uma chamada “pré-mistura” ou “submistura” para derivar os segundo conjunto de K sinais de entrada do descorrelacionador 1722a a 1722k com base nos primeiro conjunto de N sinais de entrada do descorrelacionador 1710a a 1710n. Por exemplo, os K sinais do segundo conjunto de K sinais de entrada do descorrelacionador 1722a a 1722k podem ser representados usando uma matriz $\hat{\mathbf{Z}}_{mix}$. A unidade de descorrelação (ou, equivalentemente, descorrelacionador de multicanal) 1700 também compreende um núcleo do descorrelacionador 1730, que é configurado para receber os K sinais do segundo conjunto de sinais de entrada do descorrelacionador 1722a a 1722k, e para fornecer, com base na mesma, K sinais de saída do descorrelacionador que constituem um primeiro conjunto de sinais de saída do descorrelacionador 1732a a 1732k. Por exemplo, o núcleo do descorrelacionador 1730 pode compreender K descorrelacionadores individuais (ou funções de descorrelação), sendo que cada um dos descorrelacionadores individuais (ou funções de descorrelação) fornece um dos sinais de saída do descorrelacionador do primeiro conjunto de K sinais de saída do descorrelacionador 1732a a 1732k com base em um descorrelacionador correspondente de sinal de entrada do segundo conjunto de K sinais de entrada do descorrelacionador 1722a a 1722k. Alternativamente, um dado descorrelacionador, ou função de descorrelação, pode ser aplicado K vezes, de modo que cada um dos sinais de saída do descorrelacionador do primeiro conjunto de K sinais de saída do descorrelacionador 1732a a 1732k é baseado em um único dos sinais de entrada do descorrelacionador do segundo conjunto de K sinais de entrada do descorrelacionador 1722a a 1722k.

[400] A unidade de descorrelação 1700 também compreende um pós-misturador 1740, que é configurado para receber os K sinais de saída do descorrelacionador 1732a a 1732k do primeiro conjunto de sinais de saída do

descorrelacionador e para fornecer, com base no mesmo, os N sinais 1712a a 1712n do segundo conjunto de sinais de saída do descorrelacionador (que constituem os sinais de saída “externos” do descorrelacionador).

[401] Deve-se observar que o pré-misturador 1720 pode de preferência realizar uma operação de mistura linear, que pode ser descrita por uma matriz de pré-mistura $\mathbf{M}_{pre}$. Além disso, o pós-misturador 1740 de preferência realiza uma operação de mistura linear (ou sobremistura), que pode ser representada pela pós-matriz de mistura $\mathbf{M}_{pós}$, para derivar os N sinais de saída do descorrelacionador 1712a a 1712n do segundo conjunto de sinais de saída do descorrelacionador a partir do primeiro conjunto de K sinais de saída do descorrelacionador 1732a a 1732k (i.e., a partir dos sinais de saída do núcleo do descorrelacionador 1730).

[402] A ideia principal do método proposto e aparelho é reduzir o número de sinais de entrada para os descorrelacionadores (ou para o núcleo do descorrelacionador) a partir de N a K por:

[403] Pré-mistura dos sinais (ex., os sinais de áudio renderizados) para número inferior de canais com

[404]
$$\hat{\mathbf{Z}}_{mis} = \mathbf{M}_{pre} \hat{\mathbf{Z}}$$

[405] Aplicação da descorrelação usando os K descorrelacionadores disponíveis (ex., do núcleo do descorrelacionador) com

[406]
$$\hat{\mathbf{Z}}_{mis}^{dec} = Decorr(\hat{\mathbf{Z}}_{mis})$$

[407] Sobremistura dos sinais descorrelacionados de volta para os N canais com

[408]
$$\mathbf{W} = \mathbf{M}_{post} \hat{\mathbf{Z}}_{mis}^{dec}$$

[409] A matriz de pré-mistura $\mathbf{M}_{pre}$ pode ser construída com base na informação de downmix/renderização/correlação/etc de modo que o produto da matriz $(\mathbf{M}_{pre} \mathbf{M}_{pre}^H)$ se torne bem condicionado (com relação à operação de

inversão). A matriz de pós-mistura pode ser calculada como

[410]
$$\mathbf{M}_{pós} \approx \mathbf{M}_{pre}^H (\mathbf{M}_{pre} \mathbf{M}_{pre}^H)^{-1}$$

[411] Mesmo que a matriz de covariância dos sinais descorrelacionados

intermediários $\tilde{\mathbf{S}}$ (ou $\hat{\mathbf{Z}}_{mis}^{dec}$) seja diagonal (assumindo descorrelacionadores ideais), a matriz de covariância dos sinais descorrelacionados finais $\mathbf{W}$ muito provavelmente não será mais diagonal quando este tipo de processamento for usado. Portanto, a matriz de covariância pode ter que ser estimada usando as matrizes de mistura como

[412]
$$\mathbf{E}_W = \mathbf{M}_{pós} \left[\text{matdiag}(\mathbf{M}_{pre} \mathbf{E}_{\hat{\mathbf{Z}}} \mathbf{M}_{pre}^H) \right] \mathbf{M}_{pós}^H$$

[413] O número de descorrelacionadores usados (ou descorrelações individuais), K , não é especificado e é dependente da complexidade computacional desejada e descorrelacionadores disponíveis. Seu valor pode ser variado a partir de N (complexidade computacional mais alta) até 1 (complexidade computacional mais baixa).

[414] O número de sinais de entrada para a unidade descorrelacionadora, N , é arbitrário e o método proposto suporta qualquer número de sinais de entrada, independente da configuração de renderização do sistema.

[415] Por exemplo nas aplicações usando conteúdo de áudio 3D, com alto número de canais de saída, dependendo da configuração de saída uma expressão possível para a matriz de pré-mistura $\mathbf{M}_{pre}$ é descrita abaixo.

[416] No que segue, será descrito como a pré-mistura, que é realizada pelo pré-misturador 1720 (e, conseqüentemente, a pós-mistura, que é realizada pelo pós-misturador 1740) é ajustada se a unidade de descorrelação 1700 é usada em um decodificador de áudio multicanal, sendo que os sinais de entrada do descorrelacionador 1710a a 1710n do primeiro conjunto de sinais de entrada do descorrelacionador são associados com diferentes posições espaciais de uma cena de áudio.

[417] Para este fim, a Fig. 18 mostra uma representação em tabela das posições do alto-falante, que são usadas para diferentes formatos de saída.

[418] Na tabela 1800 da Fig. 18, uma primeira coluna 1810 descreve um número do índice de alto-falante. Uma segunda coluna 1820 descreve um rótulo de alto-falante. Uma terceira coluna 1830 descreve uma posição de azimute do respectivo alto-falante, e uma quarta coluna 1832 descreve uma tolerância de azimute da posição do alto-falante. Uma quinta coluna 1840 descreve uma elevação de uma posição do respectivo alto-falante, e uma sexta coluna 1842 descreve uma tolerância de elevação correspondente. Uma sétima coluna 1850 indica que alto-falantes são usados para o formato de saída O-2.0. Uma oitava coluna 1860 mostra que alto-falantes são usados para o formato de saída O-5.1. Uma nona coluna 1864 mostra que alto-falantes são usados para o formato de saída O-7.1. Uma décima coluna 1870 mostra que alto-falantes são usados para o formato de saída O-8.1, uma décima primeira coluna 1880 mostra que alto-falantes são usados para o formato de saída O-10.1, e uma décima segunda coluna 1890 mostra que alto-falantes são usados para o formato de saída O-22.2. Como pode ser visto, dois alto-falantes são usados para o formato de saída O-2.0, seis alto-falantes são usados para o formato de saída O-5.1, oito alto-falantes são usados para o formato de saída O-7.1, nove alto-falantes são usados para o formato de saída O-8.1, 11 alto-falantes são usados para o formato de saída O-10.1, e 24 alto-falantes são usados para o formato de saída O-22.2.

[419] Entretanto, deve-se observar que um alto-falante de efeito de baixa frequência é usado para formatos de saída O-5.1, O-7.1, O-8.1 e O-10.1, e que dois alto-falantes de efeito de baixa frequência (LFE1, LFE2) são usados para o formato de saída O-22.2. Além disso, deve-se observar que, em uma modalidade preferida, um sinal de áudio renderizado (por exemplo, um dos sinais de áudio renderizados 1582a a 1582n) é associado com cada um dos alto-falantes, exceto para os um ou mais alto-falantes de efeito de baixa frequência. Portanto, dois sinais de áudio renderizados são associados com os

dois alto-falantes usados de acordo com o formato O-2.0, cinco sinais de áudio renderizados são associados com os cinco alto-falantes de efeito de frequência não baixa se o formato O-5.1 é usado, sete sinais de áudio renderizados são associados com sete alto-falantes de efeito de frequência não baixa se o formato O-7.1 é usado, oito sinais de áudio renderizados são associados com os oito alto-falantes de efeito de frequência não baixa se o formato O-8.1 é usado, dez sinais de áudio renderizados são associados com os dez alto-falantes de efeito de frequência não baixa se o formato O-10.1 é usado, e 22 sinais de áudio renderizados são associados com os 22 alto-falantes de efeito de frequência não baixa se o formato O-22.2 é usado.

[420] Entretanto, é frequentemente desejável usar um número menor de descorrelacionadores (individuais) (do núcleo do descorrelacionador), como mencionado acima. No que segue, será descrito como o número de descorrelacionadores pode ser reduzido flexivelmente quando o formato de saída O-22.2 é usado por um decodificador de áudio multicanal, de modo que há 22 sinais de áudio renderizados 1582a a 1582n (que podem ser representados pela matriz $\hat{\mathbf{Z}}$, ou por um vetor $\hat{\mathbf{z}}$).

[421] As Figs. 19a a 19g representam diferentes opções para pré-mistura dos sinais de áudio renderizados 1582a a 1582n sob a suposição de que existem $N = 22$ sinais de áudio renderizados. Por exemplo, a Fig. 19a mostra uma representação em tabela das entradas de uma matriz de pré-mistura $\mathbf{M}_{pré}$. As fileiras, rotuladas com 1 a 11 na Fig. 19a, representam as fileiras da matriz de pré-mistura $\mathbf{M}_{pré}$, e as colunas, rotuladas com 1 a 22 são associadas com as colunas da matriz de pré-mistura $\mathbf{M}_{pré}$. Além disso, deve-se observar que cada fileira da matriz de pré-mistura $\mathbf{M}_{pré}$ é associada com um dos K sinais de entrada do descorrelacionador 1722a a 1722k do segundo conjunto de sinais de entrada do descorrelacionador (i.e., com os sinais de entrada do núcleo do descorrelacionador). Além disso, cada coluna da matriz de pré-mistura $\mathbf{M}_{pré}$ é associada com um dos N sinais de entrada do descorrelacionador 1710a a 1710n do primeiro conjunto de sinais de entrada do descorrelacionador, e

consequentemente com um dos sinais de áudio renderizados 1582a a 1582n (já que os sinais de entrada do descorrelacionador 1710a a 1710n do primeiro conjunto de sinais de entrada do descorrelacionador são tipicamente idênticos aos sinais de áudio renderizados 1582 a 1582n em uma modalidade). Portanto, cada coluna da matriz de pré-mistura $\mathbf{M}_{pré}$ é associada com um alto-falante específico e, consequentemente, já que os alto-falantes são associados com posições espaciais, com uma posição espacial específica. Uma fileira 1910 indica à qual alto-falante (e, consequentemente, para essa posição espacial) as colunas da matriz de pré-mistura $\mathbf{M}_{pré}$ são associadas (sendo que os rótulos dos alto-falantes são definidos na coluna 1820 da tabela 1800).

[422] No que segue, a funcionalidade definida pela pré-mistura $\mathbf{M}_{pré}$ da Fig. 19a será descrita em mais detalhe. Como pode ser visto, os sinais de áudio renderizados associados com os alto-falantes (ou, equivalentemente, posições do alto-falante) “CH_M_000” e “CH_L_000” são combinados, para obter um primeiro sinal de entrada do descorrelacionador do segundo conjunto de sinais de entrada do descorrelacionador (i.e., um primeiro sinal de entrada do descorrelacionador submisturados), que é indicado pelos valores “1” na primeira e segunda coluna da primeira fileira da matriz de pré-mistura $\mathbf{M}_{pré}$. Da mesma forma, os sinais de áudio renderizados associados com os alto-falantes (ou, equivalentemente, posições do alto-falante) “CH_U_000” e “CH_T_000” são combinados para obter um segundo sinal de entrada do descorrelacionador submisturados (i.e., um segundo sinal de entrada do descorrelacionador do segundo conjunto de sinais de entrada do descorrelacionador). Além disso, pode ser visto que a matriz de pré-mistura $\mathbf{M}_{pré}$ da Fig. 19a define onze combinações de dois sinais de áudio renderizados cada, de modo que onze sinais de entrada do descorrelacionador submisturados são derivados de 22 sinais de áudio renderizados. Também pode ser visto que quatro sinais centrais são combinados, para obter dois sinais de entrada do descorrelacionador submisturados (conferir as colunas 1 a 4 e fileiras 1 e 2 da matriz de pré-mistura). Além disso, pode ser visto que os

outros sinais de entrada do descorrelacionador submisturados são cada obtidos pela combinação de dois sinais de áudio associados com o mesmo lado da cena de áudio. Por exemplo, um terceiro sinal de entrada do descorrelacionador submisturados, representado pela terceira fileira da matriz de pré-mistura, é obtido pela combinação dos sinais de áudio renderizados associados com uma posição de azimuth de $+135^\circ$ (“CH_M_L135”; “CH_U_L135”). Além disso, pode ser visto que um quarto sinal de entrada do descorrelacionador (representado pela quarta fileira da matriz de pré-mistura) é obtido pela combinação dos sinais de áudio renderizados associados com uma posição de azimuth de -135° (“CH_M_R135”; “CH_U_R135”). Portanto, cada um dos sinais de entrada do descorrelacionador submisturados é obtido pela combinação de dois sinais de áudio renderizados associados com a mesma (ou similar) posição de azimuth (ou, equivalentemente, posição horizontal), sendo que há tipicamente uma combinação de sinais associados com diferente elevação (ou, equivalentemente, posição vertical).

[423] Fazendo referência agora à Fig. 19b, que mostra coeficientes de pré-mistura (entradas da matriz de pré-mistura $\mathbf{M}_{pré}$) para $N = 22$ e $K = 10$. A estrutura da tabela da Fig. 19b é idêntica à estrutura da tabela da Fig. 19a. Entretanto, como pode ser visto, a matriz de pré-mistura $\mathbf{M}_{pré}$ de acordo com a Fig. 19b difere da matriz de pré-mistura $\mathbf{M}_{pré}$ da Fig. 19a pelo fato de que a primeira fileira descreve a combinação de quatro sinais de áudio renderizados que tem IDs de canal (ou posições) “CH_M_000”, “CH_L_000”, “CH_U_000” e “CH_T_000”. Em outras palavras, quatro sinais de áudio renderizados associados com posições verticalmente adjacentes são combinados na pré-mistura para reduzir o número de descorrelacionadores necessários (dez descorrelacionadores ao invés de onze descorrelacionadores para a matriz de acordo com a Fig. 19a).

[424] Fazendo referência agora à Fig. 19c, que mostra coeficientes de pré-mistura (entradas da matriz de pré-mistura $\mathbf{M}_{pré}$) para $N = 22$ e $K = 9$, pode ser visto, que a matriz de pré-mistura $\mathbf{M}_{pré}$ de acordo com a Fig. 19c compreende

somente nove fileiras. Além disso, pode ser visto a partir da segunda fileira da matriz de pré-mistura $\mathbf{M}_{pré}$ da Fig. 19c que os sinais de áudio renderizados associados com os IDs de canal (ou posições) “CH_M_L135”, “CH_U_L135”, “CH_M_R135” e “CH_U_R135” são combinados (em um pré-misturador configurado de acordo com a matriz de pré-mistura da Fig. 19c) para obter um segundo sinal de entrada do descorrelacionador submisturados (sinal de entrada do descorrelacionador do segundo conjunto de sinais de entrada do descorrelacionador). Como pode ser visto, os sinais de áudio renderizados que foram combinados em sinais de entrada separados do descorrelacionador submisturados pelas matrizes de pré-mistura de acordo com as Figs. 19a e 19b submisturados em um sinal de entrada submisturados comum do descorrelacionador de acordo com a Fig. 19c. Além disso, deve-se observar que os sinais de áudio renderizados que tem IDs de canal “CH_M_L135” e “CH_U_L135” são associados com posições horizontais idênticas (ou posições de azimuth) no mesmo lado da cena de áudio e posições verticais espacialmente adjacentes (ou elevações), e que os sinais de áudio renderizados que tem IDs de canal “CH_M_R135” e “CH_U_R135” são associados com posições horizontais idênticas (ou posições de azimuth) em um segundo lado da cena de áudio e posições verticais espacialmente adjacentes (ou elevações). Além disso, pode-se dizer que os sinais de áudio renderizados que tem IDs de canal “CH_M_L135”, “CH_U_L135”, “CH_M_R135” e “CH_U_R135” são associados com um par horizontal (ou mesmo um quádruplo horizontal) de posições espaciais compreendendo uma posição do lado esquerdo e uma posição do lado direito. Em outras palavras, pode ser visto na segunda fileira da matriz de pré-mistura $\mathbf{M}_{pré}$ da Fig. 19c que dois dos quatro sinais de áudio renderizados, que são combinados para serem descorrelacionados usando um único dado descorrelacionador, são associados com posições espaciais em um lado esquerdo de uma cena de áudio, e que dois dos quatro sinais de áudio renderizados que são combinados para serem descorrelacionados usando o mesmo dado descorrelacionador, são associados

com posições espaciais em um lado direito da cena de áudio. Além disso, pode ser visto que os sinais de áudio renderizados do lado esquerdo (dos ditos quatro sinais de áudio renderizados) são associados com posições espaciais que são simétricas, com relação a um plano central da cena de áudio, com as posições espaciais associadas com os sinais de áudio renderizados do lado direito (dos ditos quatro sinais de áudio renderizado), de modo que um quádruplo “simétrico” dos sinais de áudio renderizados são combinados pela pré-mistura a ser descorrelacionada usando um único (individual) descorrelacionador.

[425] Fazendo referência às Figs. 19d, 19e, 19f e 19g, pode ser visto que mais e mais sinais de áudio renderizados são combinados com número decrescente de descorrelacionadores (individual) (i.e. com K decrescente). Como pode ser visto nas Figs. 19a a 19g, tipicamente sinais de áudio renderizados que são submisturados em dois sinais de entrada separados do descorrelacionador submisturados são combinados quando diminui-se o número de descorrelacionadores por 1. Além disso, pode ser visto que tipicamente tais sinais de áudio renderizados são combinados, que são associados com um “quádruplo simétrico” das posições espaciais, sendo que, para um número comparativamente alto de descorrelacionadores, somente sinais de áudio renderizados associados com posições horizontais iguais ou ao menos similares (ou posições de azimuth) são combinados, enquanto para o número comparativamente inferior de descorrelacionadores, os sinais de áudio renderizados associados com posições espaciais nos lados opostos da cena de áudio também são combinados.

[426] Fazendo referência agora às Figs. 20a a 20d, 21a a 21c, 22a a 22b e 23, deve-se observar que conceitos similares também podem ser aplicados para um número diferente de sinais de áudio renderizados.

[427] Por exemplo, as Figs. 20a a 20d descrevem entradas da matriz de pré-mistura $\mathbf{M}_{pré}$ para $N = 10$ e para K entre 2 e 5.

[428] Da mesma forma, as Figs. 21a a 21c descrevem entradas da matriz

de pré-mistura $\mathbf{M}_{pré}$ para $N = 8$ e K entre 2 e 4.

[429] Da mesma forma, as Figs. 21d a 21f descrevem entradas da matriz de pré-mistura $\mathbf{M}_{pré}$ para $N = 7$ e K entre 2 e 4.

[430] As Figs. 22a e 22b mostram entradas da matriz de pré-mistura para $N = 5$ e $K = 2$ e $K = 3$.

[431] Finalmente, a Fig. 23 mostra entradas da matriz de pré-mistura para $N = 2$ e $K = 1$.

[432] Para resumir, as matrizes de pré-mistura de acordo com as Figs. 19 a 23 podem ser usadas, por exemplo, de uma maneira comutável, em um descorrelacionador de multicanal que é parte de um decodificador de áudio multicanal. A comutação entre as matrizes de pré-mistura pode ser realizada, por exemplo, na dependência de uma configuração de saída desejada (que tipicamente determina um número N de sinais de áudio renderizados) e também na dependência da complexidade desejada da descorrelação (que determina o parâmetro K , e que pode ser ajustada, por exemplo, na dependência da informação de complexidade incluída em uma representação codificada de um conteúdo de áudio).

[433] Fazendo referência agora à Fig. 24, a redução da complexidade para o formato de saída 22.2 será descrita em mais detalhe. Como já descrito acima, uma solução possível para construir a matriz de pré-mistura e a matriz de pós-mistura é usar a informação espacial do layout de reprodução para selecionar os canais a serem misturados juntos e calcular os coeficientes de mistura. Com base em sua posição, os alto-falantes geometricamente relacionados (e, por exemplo, os sinais de áudio renderizados associados com eles) são agrupados, tomando pares verticais e horizontais, como descrito na tabela da Fig. 24. Em outras palavras, a Fig. 24 mostra, na forma de uma tabela, um agrupamento de posições do alto-falante, que pode ser associado com sinais de áudio renderizados. Por exemplo, uma primeira fileira 2410 descreve um primeiro grupo de posições de alto-falante, que estão em um centro de uma cena de áudio. Uma segunda fileira 2412 representa um

segundo grupo de posições do alto-falante, que são espacialmente relacionadas. As posições do alto-falante “CH_M_L135” e “CH_U_L135” são associadas com posições de azimuth idênticas (ou equivalentemente posições horizontais) e posições de elevação adjacentes (ou equivalentemente, posições verticalmente adjacentes). Da mesma forma, as posições “CH_M_R135” e “CH_U_R135” compreendem azimuth idêntico (ou, equivalentemente, posição horizontal idêntica) e elevação similar (ou, equivalentemente, posição verticalmente adjacente). Além disso, as posições “CH_M_L135”, “CH_U_L135”, “CH_M_R135” e “CH_U_R135” formam um quádruplo de posições, sendo que as posições “CH_M_L135” e “CH_U_L135” são simétricas às posições “CH_M_R135” e “CH_U_R135” com relação ao um plano central da cena de áudio. Além disso, as posições “CH_M_180” e “CH_U_180” também compreendem posição de azimuth idêntica (ou, equivalentemente, posição horizontal idêntica) e elevação similar (ou, equivalentemente, posição vertical adjacente).

[434] Uma terceira fileira 2414 representa um terceiro grupo de posições. Deve-se observar que as posições “CH_M_L030” e “CH_L_L045” são posições espacialmente adjacentes e compreendem azimuth similar (ou, equivalentemente, posição horizontal similar) e elevação similar (ou, equivalentemente, posição vertical similar). O mesmo vale para as posições “CH_M_R030” e “CH_L_R045”. Além disso, as posições do terceiro grupo de posições formam um quádruplo de posições, sendo que as posições “CH_M_L030” e “CH_L_L045” são espacialmente adjacentes, e simétricas com relação ao um plano central da cena de áudio, para as posições “CH_M_R030” e “CH_L_R045”.

[435] Uma quarta fileira 2416 representa quatro posições adicionais, que tem características similares quando comparadas às das primeiras quatro posições da segunda fileira, e que formam um quádruplo simétrico de posições.

[436] Uma quinta fileira 2418 representa outro quádruplo de posições simétricas “CH_M_L060”, “CH_U_L045”, “CH_M_R060” e “CH_U_R045”.

[437] Além disso, deve-se observar que os sinais de áudio renderizados associados com as posições dos diferentes grupos de posições podem ser combinados mais e mais com número decrescente de descorrelacionadores. Por exemplo, na presença de onze descorrelacionadores individuais em um descorrelacionador de multicanal, os sinais de áudio renderizados associados com posições na primeira e segunda coluna podem ser combinados para cada grupo. Além disso, os sinais de áudio renderizados associados com as posições representadas em uma terceira e uma quarta coluna podem ser combinados para cada grupo. Além disso, os sinais de áudio renderizados associados com as posições mostradas na quinta e sexta coluna podem ser combinadas para o segundo grupo. Portanto, onze sinais de entrada do descorrelacionador de downmix (que são inseridos nos descorrelacionadores individuais) podem ser obtidos. Entretanto, deseja-se ter menos descorrelacionadores individuais, sinais de áudio renderizados associados com as posições mostradas nas colunas 1 a 4 podem ser combinados para um ou mais dos grupos. Também, os sinais de áudio renderizados associados com todas as posições do segundo grupo podem ser combinados, se for desejado reduzir ainda mais um número de descorrelacionadores individuais.

[438] Para resumir, os sinais alimentados ao layout de saída (por exemplo, aos alto-falantes) tem dependências horizontais e verticais, que devem ser preservadas durante o processo de descorrelação. Portanto, os coeficientes de mistura são calculados de modo que os canais correspondentes à diferentes grupos de alto-falante não são misturados juntos.

[439] Dependendo do número de descorrelacionadores disponíveis, ou o nível desejado de descorrelação, em cada grupo primeiro são misturados os pares verticais (entre uma camada média e a camada superior ou entre a camada média e a camada inferior). Em segundo lugar, os pares horizontais (entre esquerda e direita) ou pares verticais restantes são misturados. Por exemplo, em três grupos, primeiro os canais no par vertical esquerdo (“CH_M_L030” e “CH_L_L045”), e no par vertical direito (“CH_M_R030” e

“CH_L_R045”), são misturados, reduzindo assim o número de descorrelacionadores necessários para este grupo a partir de quatro a dois. Se é desejado reduzir ainda mais o número de descorrelacionadores, o par horizontal obtido é submisturado para somente um canal, e o número de descorrelacionadores necessários para este grupo é reduzido de quatro para um.

[440] Com base nas regras de mistura apresentadas, as tabelas mencionadas acima (por exemplo, mostradas nas Figs. 19 a 23) são derivadas para diferentes níveis de descorrelação desejada (ou para diferentes níveis de complexidade de descorrelação desejada).

[441] 16. Compatibilidade com um Renderizador Externo Secundário/Conversor de Formato

[442] No caso quando o decodificador SAOC (ou, mais geralmente, o decodificador de áudio multicanal) é usado junto com um renderizador externo secundário/conversor de formato, as seguintes mudanças ao conceito proposto (método ou aparelho) podem ser usadas:

[443] a matriz de renderização interna $\mathbf{R}$ (ex., do renderizador) é definida para identidade $\mathbf{R} = \mathbf{I}_{N_{Objects}}$ (quando um renderizador externo é usado) ou inicializado com os coeficientes de mistura derivados de uma configuração de renderização intermediária (quando um conversor de formato externo é usado).

[444] o número de descorrelacionadores é reduzido usando os método descritos na seção 15 com a matriz de pré-mistura $\mathbf{M}_{pre}$ calculada com base na informação de retorno recebida do renderizador/conversor de formato (ex., $\mathbf{M}_{pre} = \mathbf{D}_{conteúdo}$ onde $\mathbf{D}_{conteúdo}$ é a matriz de downmix usada dentro do conversor de formato). Os canais que serão misturados fora do decodificador SAOC, são pré-misturados juntos e alimentados ao mesmo descorrelacionador dentro do decodificador SAOC.

[445] Usando um conversor de formato externo, o renderizador interno SAOC irá pré-renderizar para uma configuração intermediária (ex., a

configuração com o número mais alto de alto-falantes).

[446] Para concluir, em algumas modalidades uma informação sobre quais dos sinais de áudio de saída são misturados em um renderizador externo ou conversor de formato são usados para determinar a matriz de pré-mistura $\mathbf{M}_{pré}$, de modo que a matriz de pré-mistura define uma combinação de tais sinais de entrada do descorrelacionador (do primeiro conjunto de sinais de entrada do descorrelacionador) que são efetivamente combinados no renderizador externo. Deste modo, a informação recebida a partir do renderizador externo/conversor de formato (que recebe os sinais de áudio de saída do decodificador de multi-canal) é usado para selecionar ou ajustar a matriz de pré-mistura (por exemplo, quando a matriz de renderização interna do decodificador de áudio multicanal é definida para identidade, ou inicializada com os coeficientes de mistura derivados de uma configuração de renderização intermediária), e o renderizador externo/conversor de formato é conectado para receber os sinais de áudio de saída as mencionada acima com relação ao decodificador de áudio multicanal.

[447] 17. Bitstream

[448] No que segue, será descrito quais informações de sinalização adicionais podem ser usadas em um bitstream (ou, equivalentemente, em uma representação codificada do conteúdo de áudio). Nas modalidades de acordo com a invenção, o método de descorrelação pode ser sinalizado no bitstream para garantir um nível de qualidade desejado. Deste modo, o usuário (ou um codificador de áudio) tem mais flexibilidade para selecionar o método com base no conteúdo. Para este fim, a sintaxe do bitstream MPEG SAOC pode ser, por exemplo, estendida com dois bits para especificar o método de descorrelação usado e/ou dois bits para especificar a configuração (ou complexidade).

[449] A Fig. 25 mostra uma representação de sintaxe dos elementos de bitstream “bsDecorrelationMethod” e “bsDecorrelationLevel”, que podem ser adicionados, por exemplo, a uma porção de bitstream “SAOCSpecifigConfig()” ou “SAOC3DSpecificConfig()”. Como pode ser visto na Fig. 25, dois bits podem

ser usados para o elemento de bitstream “bsDecorrelationMethod”, e dois bits podem ser usados para o elemento de bitstream “bsDecorrelationLevel”.

[450] A Fig. 26 mostra, na forma de uma tabela, uma associação entre valores da variável do bitstream “bsDecorrelationMethod” e os diferentes métodos de decorrelação. Por exemplo, três métodos de decorrelação diferentes podem ser sinalizados por diferentes valores da dita variável do bitstream. Por exemplo, uma correção de covariância de saída usando sinais decorrelacionados, como descrito, por exemplo, na seção 14.3, pode ser sinalizada como uma das opções. Como outra opção, um método de ajuste de covariância, por exemplo, como descrito na seção 14.4.1 pode ser sinalizado. Ainda como outra opção, um método de compensação de energia, por exemplo, como descrito na seção 14.4.2 pode ser sinalizado. Portanto, três métodos diferentes para a reconstrução das características do sinal dos sinais de áudio de saída com base nos sinais de áudio renderizados e os sinais de áudio decorrelacionados podem ser selecionados na dependência de uma variável do bitstream.

[451] O modo de compensação de energia usa o método descrito na seção 14.4.2, o modo de ajuste de covariância limitado usa o método descrito na seção 14.4.1, e o modo de ajuste de covariância geral usa o método descrito na seção 14.3.

[452] Fazendo referência agora à Fig. 27, que mostra, na forma de uma representação de tabela, como diferentes níveis de decorrelação podem ser sinalizados pela variável do bitstream “bsDecorrelationLevel”, um método para selecionar a complexidade de decorrelação será descrito. Em outras palavras, a referida variável pode ser avaliada por um decodificador de áudio multicanal compreendendo o decorrelacionador de multicanal descrito acima para decidir qual complexidade de decorrelação é usada. Por exemplo, o referido parâmetro de bitstream pode sinalizar diferentes “níveis” de decorrelação que podem ser designados com os valores: 0, 1, 2 e 3.

[453] Um exemplo das configurações de decorrelação (que podem, por

exemplo, ser designadas como “níveis” de descorrelação) é dado na tabela da Fig. 27. A Fig. 27 mostra uma representação em tabela de um número de descorrelacionadores para diferentes “níveis” (ex., níveis de descorrelação) e configurações de saída. Em outras palavras, a Fig. 27 mostra o número K de sinais de entrada do descorrelacionador (do segundo conjunto de sinais de entrada do descorrelacionador), que é usado pelo descorrelacionador de multicanal. Como pode ser visto na tabela da Fig. 27, um número de descorrelacionadores (individual) usado no descorrelacionador de multicanal é alternado entre 11, 9, 7 e 5 para uma configuração de saída 22.2, na dependência de qual “nível de descorrelação” é sinalizado pelo parâmetro de bitstream “bsDecorrelationLevel”. Para uma configuração de saída 10.1, uma seleção é feita entre 10, 5, 3 e 2 descorrelacionadores individuais, para uma configuração 8.1, uma seleção é feita entre 8, 4, 3 ou 2 descorrelacionadores individuais, e para uma configuração de saída 7.1, uma seleção é feita entre 7, 4, 3 e 2 descorrelacionadores na dependência do “nível de descorrelação” sinalizado pelo referido parâmetro de bitstream. Na configuração de saída 5.1, há somente três opções válidas para os números de descorrelacionadores individuais, ou seja, 5, 3, ou 2. Para a configuração de saída 2.1, há somente uma escolha entre dois descorrelacionadores individuais (nível de descorrelação 0) e um descorrelacionador individual (nível de descorrelação 1).

[454] Para resumir, o método de descorrelação pode ser determinado no lado do decodificador com base na capacidade computacional e um número de descorrelacionadores disponível. Além disso, a seleção do número de descorrelacionadores pode ser feita no lado do codificador e sinalizada usando um parâmetro de bitstream.

[455] Portanto, ambos o método de como os sinais de áudio descorrelacionados são aplicados, para obter os sinais de áudio de saída, e a complexidade para o fornecimento dos sinais descorrelacionados podem ser controlados a partir do lado de um codificador de áudio usando os parâmetros de bitstream mostrados na Fig. 25 e definidos em mais detalhe nas Figs. 26 e

27.

[456] 18. Campos de aplicação para o processamento da invenção

[457] Deve-se observar que um dos objetivos dos métodos introduzidos é restaurar pistas de áudio, que são de maior importância para a percepção humana de uma cena de áudio. As modalidades de acordo com a invenção melhoram uma precisão de reconstrução do nível de energia e propriedades de correlação e portanto aumenta a qualidade de percepção de áudio do sinal de saída final. As modalidades de acordo com a invenção podem ser aplicadas a um número arbitrário de canais de downmix/upmix. Além disso, os métodos e aparelhos descritos aqui podem ser combinados com algoritmos de separação de fonte paramétrica existentes. As modalidades de acordo com a invenção permitem controlar a complexidade computacional do sistema pelas restrições de definição sobre o número de funções do descorrelacionador aplicadas. As modalidades de acordo com a invenção podem levar a uma simplificação dos algoritmos de construção paramétrica com base no objeto como SAOC pela remoção de uma etapa de transcodificação MPS.

[458] 19. Ambiente de Codificação/Decodificação

[459] No que segue, um ambiente de codificação/decodificação de áudio será descrito no qual os conceitos de acordo com a presente invenção podem ser aplicados.

[460] Um sistema de codec de áudio 3D, nos quais os conceitos de acordo com a presente invenção podem ser usados, é baseado em um codec MPEG-D USAC para codificar os canais e sinais de objeto para aumentar a eficiência para codificar uma grande quantidade de objetos. A tecnologia MPEG-SAOC foi adaptada. Três tipos de renderizadores realizam as tarefas de renderização de objetos para os canais, renderização de canais para fones de ouvido ou renderização de canais para diferentes ajustes de alto-falante. Quando os sinais de objeto são explicitamente transmitidos ou parametricamente codificados usando SAOC, a informação de metadados do objeto correspondente é comprimida e multiplexada para o fluxo de áudio 3D.

[461] As Figs. 28, 29 e 30 mostram os diferentes blocos de algoritmo do sistema de áudio 3D.

[462] A Fig. 28 mostra um diagrama em bloco esquemático de tal codificador de áudio, e a Fig. 29 mostra um diagrama em bloco esquemático de tal decodificador de áudio. Em outras palavras, as Figs. 28 e 29 mostram os diferentes blocos de algoritmo do sistema de áudio 3D.

[463] Fazendo referência agora à Fig. 28, que mostra um diagrama em bloco esquemático de um codificador de áudio 3D 2900, alguns detalhes serão explicados. O codificador 2900 compreende um pré-renderizador/misturador opcional 2910, que recebe um ou mais sinais de canal 2912 e um ou mais sinais de objeto 2914 e fornece, com base nos mesmos, um ou mais sinais de canal 2916 assim como um ou mais sinais de objeto 2918, 2920. O codificador de áudio também compreende um codificador USAC 2930 e opcionalmente um codificador SAOC 2940. O codificador SAOC 2940 é configurado para fornecer um ou mais canais de transporte SAOC 2942 e a informação de lado SAOC 2944 com base em um ou mais objetos 2920 fornecidos para o codificador SAOC. Além disso, o codificador USAC 2930 é configurado para receber os sinais de canal 2916 compreendendo canais e objetos pré-renderizados a partir do pré-renderizador/misturador 2910, para receber um ou mais sinais de objeto 2918 do pré-renderizador/misturador 2910, e para receber um ou mais canais de transporte SAOC 2942 e informação de lado SAOC 2944, e fornece, com base na mesma, uma representação codificada 2932. Além disso, o codificador de áudio 2900 também compreende um codificador de metadados de objeto 2950 que é configurado para receber metadados de objeto 2952 (que podem ser avaliados pelo pré-renderizador/misturador 2910) e para codificar os metadados de objeto para obter metadados de objeto codificados 2954. Os metadados codificados também são recebidos pelo codificador USAC 2930 e usados para fornecer a representação codificada 2932.

[464] Alguns detalhes com relação aos componentes individuais do codificador de áudio 2900 serão descritos abaixo.

[465] Fazendo referência agora à Fig. 29, um decodificador de áudio 3000 será descrito. O decodificador de áudio 3000 é configurado para receber uma representação codificada 3010 e para fornecer, com base na mesma, um sinal de alto-falante multicanal 3012, sinais de fone de ouvido 3014 e/ou sinais de alto-falante 3016 em um formato alternativo (por exemplo, em um formato 5.1). O decodificador de áudio 3000 compreende um decodificador USAC 3020, que fornece um ou mais sinais de canal 3022, um ou mais sinais de objeto pré-renderizados 3024, um ou mais sinais de objeto 3026, um ou mais canais de transporte SAOC 3028, uma informação de lado SAOC 3030 e uma informação de metadados de objeto comprimida 3032 com base na representação codificada 3010. O decodificador de áudio 3000 também compreende um renderizador de objeto 3040, que é configurado para fornecer um ou mais sinais de objeto renderizados 3042 com base nos um ou mais sinais de objeto 3026 e uma informação de metadados de objeto 3044, sendo que a informação de metadados de objeto 3044 é fornecida por um decodificador de metadados de objeto 3050 com base na informação de metadados de objeto comprimida 3032. O decodificador de áudio 3000 também compreende, opcionalmente, um decodificador SAOC 3060, que é configurado para receber o canal de transporte de SAOC 3028 e a informação de lado SAOC 3030, e para fornecer, com base na mesma, um ou mais sinais de objeto renderizados 3062. O decodificador de áudio 3000 também compreende um misturador 3070, que é configurado para receber os sinais de canal 3022, os sinais de objeto pré-renderizados 3024, os sinais de objeto renderizados 3042 e os sinais de objeto renderizados 3062, e para fornecer, com base na mesma, uma pluralidade de sinais de canal mistos 3072, que podem, por exemplo, constituir os sinais de alto-falante multicanal 3012. O decodificador de áudio 3000 pode, por exemplo, compreender também um renderizador binaural 3080, que é configurado para receber os sinais de canal mistos 3072 e para fornecer, com base na mesma, os sinais de fone de ouvido 3014. Além disso, o decodificador de áudio 3000 pode compreender uma conversão de formato 3090, que é configurada para

receber os sinais de canal mistos 3072 e uma informação de layout de reprodução 3092 e para fornecer, com base na mesma, um sinal de alto-falante 3016 para um ajuste de alto-falante alternativo.

[466] No que segue, alguns detalhes com relação aos componentes do codificador de áudio 2900 e do decodificador de áudio 3000 serão descritos.

[467] 19.1. Pré-renderizador/Misturador

[468] O pré-renderizador/misturador 2910 pode ser opcionalmente usado para converter um canal mais cena de entrada do objeto em uma cena de canal antes da decodificação. Funcionalmente, ele pode, por exemplo, ser idêntico ao renderizador de objeto/misturador descrito abaixo.

[469] A pré-renderização dos objetos pode, por exemplo, garantir uma entropia de sinal determinística na entrada do codificador que é basicamente independente do número de sinais de objeto ativos simultaneamente.

[470] Com a pré-renderização dos objetos, nenhum metadado de objeto transmissão é necessário.

[471] Sinais de objeto discretos são renderizados para o layout do canal que o codificador é configurado para usar, os pesos dos objetos para cada canal são obtidos a partir dos metadados de objeto associados (OAM) 1952.

[472] 19.2. Codec Núcleo USAC

[473] O codec núcleo 2930, 3020 para sinais de alto-falante-canal, sinais de objeto discretos, sinais de downmix do objeto e sinais pré-renderizados é baseado na tecnologia USAC MPEG-D. Ele manipula a decodificação dos múltiplos sinais pela criação da informação de mapeamento de canal e objeto com base na informação geométrica e semântica do canal de entrada e atribuição de objeto. Esta informação de mapeamento descreve, como os canais de entrada e objetos são mapeados para os elementos do canal USAC (CPEs, SCEs, LFEs) e a informação correspondente é transmitida para o decodificador.

[474] Todas as cargas úteis adicionais como dados SAOC ou metadados de objeto foram passadas através de elementos de extensão e foram

consideradas no controle de taxa dos codificadores. A decodificação de objetos é possível de diferentes formas, dependentes dos requisitos de taxa/distorção e os requisitos de interatividade para o renderizador. As seguintes variantes de codificação de objeto são possíveis:

[475] Objetos pré-renderizados: sinais de objeto são pré-renderizados e misturados aos sinais de canal 22.2 antes da decodificação. A cadeia de codificação subsequente vê sinais de canal 22.2.

[476] Formas de onda de objeto discretas: objetos como aplicados como formas de onda monofônicas para o codificador. O codificador usa elementos de canal únicos SCEs para transmitir os objetos além dos sinais de canal. Os objetos decodificados são renderizados e misturados no lado receptor. A informação de metadados de objeto comprimida é transmitida para o receptor/renderizador lado a lado.

[477] Formas de onda de objeto paramétricas: propriedades do objeto e sua relação umas com as outras são descritas por meio de parâmetros SAOC. O downmix dos sinais de objeto é codificado com USAC. A informação paramétrica é transmitida lado a lado. O número de canais de downmix é escolhido dependendo do número de objetos e a taxa de dados geral. A informação de metadados de objeto comprimida é transmitida para o renderizador SAOC.

[478] 19.3. SAOC

[479] O codificador SAOC 2940 e o decodificador SAOC 3060 para sinais de objeto são baseados na tecnologia SAOC MPEG. O sistema é capaz de recriar, modificar e renderizar um número de objetos de áudio com base em um número menor de canais transmitidos e dados paramétricos adicionais (diferenças de nível de objeto OLDs, correlações inter-objeto IOCs, ganhos de downmix DMGs). Os dados paramétricos adicionais exibem uma taxa de dados significativamente menor do que necessário para todos os objetos transmitidos individualmente, tornando a decodificação muito eficiente. O codificador SAOC tem como entrada os sinais de objeto/canal como formas de onda monofônicas

e produz a informação paramétrica (que é embalada no bitstream de áudio 3D 2932, 3010) e os canais de transporte SAOC (que são codificados usando elementos de canal únicos e transmitidos). O decodificador SAOC 3000 reconstrói os sinais de objeto/canal a partir dos canais de transporte decodificados SAOC 3028 e informação paramétrica 3030, e gera a cena de áudio de saída com base no layout de reprodução, a informação de metadados de objeto descomprimida e opcionalmente na informação de interação de usuário.

[480] 19.4. Codec de Metadados de objeto

[481] Para cada objeto, os metadados associados que especificam a posição geométrica e volume do objeto no espaço 3D são eficientemente codificados pela quantificação das propriedades do objeto no tempo e espaço. Os metadados de objeto comprimidos cOAM 2954, 3032 são transmitidos para o receptor como informação de lado.

[482] 19.5. Renderizador/Misturador de objeto

[483] O renderizador de objeto usa os metadados de objeto descomprimidos OAM 3044 para gerar formas de onda do objeto de acordo com o dado formato de reprodução. Cada objeto é renderizado para determinados canais de saída de acordo com seus metadados. A saída deste bloco resulta da soma dos resultados parciais.

[484] Se tanto os conteúdos baseados em canal assim como objetos discretos/paramétricos são decodificados, as formas de onda baseadas em canal e as formas de onda do objeto renderizadas são misturadas antes de produzir as formas de onda resultantes (ou antes de alimentá-las a um módulo pós-processador como o renderizador binaural ou o módulo de renderizador de alto-falante).

[485] 19.6. Renderizador binaural

[486] O módulo do renderizador binaural 3080 produz um downmix binaural do material de áudio multicanal, de modo que cada canal de entrada é representado pela fonte de som virtual. O processamento é realizado quadro a

quadro no domínio QMF. A binauralização é baseada nas respostas de impulso do ambiente binaural medidas.

[487] 19.7. Renderizador de Alto-falante/Conversão de formato

[488] O renderizador de alto-falante 3090 converte entre uma configuração de canal transmitida e o formato de reprodução desejado. Ele é, deste modo, chamado “Conversor de Formato” a seguir. O conversor de formato realiza conversões para números inferiores de canais de saída, i.e. cria submisturas. O sistema gera automaticamente matrizes de downmix otimizadas para a dada combinação de formatos de entrada e saída e aplica estas matrizes em um processo de downmix. O conversor de formato permite configurações de alto-falante padrão assim como configurações aleatórias com posições do alto-falante não padrão.

[489] A Fig. 30 mostra um diagrama em bloco esquemático de um conversor de formato. Em outras palavras, a Fig. 30 mostra a estrutura do conversor de formato.

[490] Como pode ser visto, o conversor de formato 3100 recebe sinais de saída do misturador 3110, por exemplo os sinais de canal mistos 3072, e fornece sinais de alto-falante 3112, por exemplo os sinais de alto-falante 3016. O conversor de formato compreende um processo de downmix 3120 no domínio e um configurador de downmix 3130, sendo que o configurador de downmix fornece a informação de configuração para o processo de downmix 3020 com base em uma informação do layout de saída do misturador 3032 e uma informação de layout de reprodução 3034.

[491] 19.8. Observações Gerais

[492] Além disso, deve-se observar que os conceitos descritos aqui, por exemplo, o decodificador de áudio 100, o codificador de áudio 200, o descorrelacionador de multicanal 600, o decodificador de áudio multicanal 700, o codificador de áudio 800 ou o decodificador de áudio 1550 podem ser usados dentro do codificador de áudio 2900 e/ou dentro do decodificador de áudio 3000. Por exemplo, os codificadores/decodificadores de áudio mencionados

acima podem ser usados como parte do codificador SAOC 2940 e/ou como uma parte do decodificador SAOC 3060. Entretanto, os conceitos mencionados acima também podem ser usados em outras posições do decodificador de áudio 3D 3000 e/ou do codificador de áudio 2900.

[493] Naturalmente, os métodos mencionados acima também podem ser usados nos conceitos para codificar ou decodificar a informação de áudio de acordo com as Figs. 28 e 29.

[494] 20. Modalidade adicional

[495] 20.1 Introdução

[496] No que segue, outra modalidade de acordo com a presente invenção será descrita.

[497] A Figura 31 mostra um diagrama em bloco esquemático de um processador de downmix, de acordo com uma modalidade da presente invenção.

[498] O processador de downmix 3100 compreende um desmisturador 3110, um renderizador 3120, um combinador 3130 e um descorrelacionador multicanal 3140. O renderizador fornece sinais de áudio renderizados Y_{seco} para o combinador 3130 e para o descorrelacionador multicanal 3140. O descorrelacionador multicanal compreende um pré-misturador 3150, que recebe os sinais de áudio renderizados (que pode ser considerado como um primeiro conjunto de sinais de entrada do descorrelacionador) e fornece, com base na mesma, um segundo conjunto de sinais de entrada pré-misturado do descorrelacionador para um núcleo do descorrelacionador 3160. O núcleo do descorrelacionador fornece um primeiro conjunto de sinais de saída do descorrelacionador com base no segundo conjunto de sinais de entrada do descorrelacionador para uso por um pós-misturador 3170. O pós-misturador faz a pós-mistura (ou sobremistura (upmix)) os sinais de saída do descorrelacionador fornecida pelo núcleo do descorrelacionador 3160, para obter um segundo conjunto de sinais de saída do descorrelacionador pós-misturado, que é fornecido para o combinador 3130.

[499] O renderizador 3130 pode, por exemplo, aplicar uma matriz $\mathbf{R}$ para a renderização, o pré-misturador pode, por exemplo, aplicar uma matriz $\mathbf{M}_{\text{pré}}$ para a pré-mistura, o pós-misturador pode, por exemplo, aplicar uma matriz $\mathbf{M}_{\text{pós}}$ para a pós-mistura, e o combinador pode, por exemplo, aplicar uma matriz $\mathbf{P}$ para a combinação.

[500] Deve-se observar que o processador de downmix 3100, ou componentes individuais ou funcionalidades dos mesmos, podem ser usados nos decodificadores de áudio descritos aqui. Além disso, deve-se observar que o processador de downmix pode ser suplementado por qualquer dos recursos e funcionalidades descritos aqui.

[501] 20.2 Processamento 3D SAOC

[502] O banco de filtro híbrido descrito em ISO/IEC 23003-1:2007 é aplicado. A desquantificação dos parâmetros de DMG, OLD, IOC segue as mesmas regras como definido em 7.1.2 do ISO/IEC 23003-2:2010.

[503] 20.2.1 Sinais e parâmetros

[504] Os sinais de áudio são definidos para cada partição de tempo n e cada sub-banda híbrida k . Os parâmetros 3D SAOC correspondentes são definidos para cada partição de tempo do parâmetro l e banda de processamento m . O mapeamento subsequente entre um domínio híbrido e de parâmetro é especificado pela Tabela a.31 do ISO/IEC 23003-1:2007. Por isso, todos os cálculos são realizados com relação à determinados índices de tempo/banda e as dimensionalidades correspondentes são implicadas para cada variável introduzida.

[505] Os dados disponíveis no decodificador 3D SAOC consistem nos sinal de downmix multicanal $\mathbf{X}$, a matriz de covariância $\mathbf{E}$, a matriz de renderização $\mathbf{R}$ e matriz de downmix $\mathbf{D}$.

[506] 20.2.1.1 Parâmetros de Objeto

[507] A matriz de covariância $\mathbf{E}$ de tamanho $N \times N$ com elementos $e_{i,j}$ representa uma aproximação da matriz de covariância do sinal original $\mathbf{E} \approx \mathbf{S}\mathbf{S}^*$ e é obtida a partir dos parâmetros OLD e IOC como:

[508]
$$e_{i,j} = \sqrt{OLD_i OLD_j} IOC_{i,j} .$$

[509] Aqui, os parâmetros de objeto desquantificados são obtidos como:

[510]
$$OLD_i = \mathbf{D}_{OLD}(i, l, m) , \quad IOC_{i,j} = \mathbf{D}_{IOC}(i, j, l, m) .$$

[511] 20.2.1.3 Matriz de downmix

[512] A matriz de downmix $\mathbf{D}$ aplicada aos sinais de áudio de entrada $\mathbf{S}$ determina o sinal de downmix como $\mathbf{X} = \mathbf{D}\mathbf{S}$. A matriz de downmix $\mathbf{D}$ de tamanho $N_{dmx} \times N$ é obtida como:

[513]
$$\mathbf{D} = \mathbf{D}_{dmx} \mathbf{D}_{pré-mistura} .$$

[514] A matriz $\mathbf{D}_{dmx}$ e matriz $\mathbf{D}_{pré-mistura}$ tem diferentes tamanhos dependendo do modo de processamento. A matriz $\mathbf{D}_{dmx}$ é obtida a partir dos parâmetros DMG como:

[515]
$$d_{i,j} = \begin{cases} 0 & , \text{ se nenhum dado DMG para } (i,j) \text{ está presente no bitstream} \\ 10^{0.05 DMG_{i,j}} & , \text{ caso contrário} \end{cases}$$

[516] Aqui, os parâmetros de downmix desquantificados são obtidos como:

[517]
$$DMG_{i,j} = \mathbf{D}_{DMG}(i, j, l) .$$

[518] 20.2.1.3.1 Modo Direto

[519] No caso do modo direto, nenhuma pré-mistura é usada. A matriz $\mathbf{D}_{pré-mistura}$ tem o tamanho $N \times N$ e é dada por: $\mathbf{D}_{pré-mistura} = \mathbf{I}$. A matriz $\mathbf{D}_{dmx}$ tem tamanho $N_{dmx} \times N$ e é obtida a partir dos parâmetros DMG de acordo com 20.2.1.3.

[520] 20.2.1.3.2 Modo de Pré-mistura

[521] No caso do modo de pré-mistura a matriz $\mathbf{D}_{pré-mistura}$ tem tamanho $(N_c + N_{pré-mistura}) \times N$ e é dada por:

$$\mathbf{D}_{\text{pré-mistura}} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{A} \end{pmatrix},$$

[522] onde a matriz de pré-mistura $\mathbf{A}$ de tamanho $N_{\text{pré-mistura}} \times N_{\text{obj}}$ é recebida como uma entrada para o decodificador 3D SAOC, a partir do renderizador de objeto.

[523] A matriz $\mathbf{D}_{\text{dmx}}$ tem tamanho $N_{\text{dmx}} \times (N_{\text{ch}} + N_{\text{pré-mistura}})$ e é obtida a partir dos parâmetros DMG de acordo com 20.2.1.3

[524] 20.2.1.4 Matriz de renderização

[525] A matriz de renderização $\mathbf{R}$ aplicada aos sinais de áudio de entrada $\mathbf{S}$ determina a saída alvo renderizada como $\mathbf{Y} = \mathbf{R}\mathbf{S}$. A matriz de renderização $\mathbf{R}$ de tamanho $N_{\text{out}} \times N$ é dada por

[526]
$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{\text{ch}} & \mathbf{R}_{\text{obj}} \end{pmatrix},$$

[527] onde $\mathbf{R}_{\text{ch}}$ do tamanho $N_{\text{out}} \times N_{\text{ch}}$ representa a matriz de renderização associada com os canais de entrada e $\mathbf{R}_{\text{obj}}$ de tamanho $N_{\text{out}} \times N_{\text{obj}}$ representa a matriz de renderização associada com os objetos de entrada.

[528] 20.2.1.4 Matriz de covariância de saída alvo

[529] A matriz de covariância $\mathbf{C}$ de tamanho $N_{\text{out}} \times N_{\text{out}}$ com elementos $c_{i,j}$ representa uma aproximação da matriz de covariância do sinal de saída alvo $\mathbf{C} \approx \mathbf{Y}\mathbf{Y}^*$ e é obtida a partir da matriz de covariância $\mathbf{E}$ e a matriz de renderização $\mathbf{R}$:

[530]
$$\mathbf{C} = \mathbf{R}\mathbf{E}\mathbf{R}^*.$$

[531] 20.2.2 Decodificação

[532] O método para obter um sinal de saída usando parâmetros 3D SAOC e informação de renderização é descrito. O decodificador 3D SAOC pode, por exemplo, e consistir nos processador de parâmetro 3D SAOC e o processador de downmix 3D SAOC.

[533] 20.2.2.1 Processador de downmix

[534] O sinal de saída do processador de downmix (representado no domínio QMF híbrido) é alimentado no banco de filtro de síntese correspondente como descrito em ISO/IEC 23003-1:2007 produzindo a saída final do decodificador 3D SAOC. Uma estrutura detalhada do processador de downmix é ilustrada na Fig. 31

[535] O sinal de saída $\hat{Y}$ é calculado a partir do sinal de downmix multicanal X e o sinal multicanal descorrelacionado X_d como:

$$\hat{Y} = P_{\text{seco}} R U X + P_{\text{molnt}} M_{\text{pos}} X_d,$$

[536] onde U representa a matriz de não mistura paramétrica e é definida em 20.2.2.1.1 e 20.2.2.1.2.

[537] O sinal multicanal descorrelacionado X_d é calculado de acordo com 20.2.3.

$$X_d = \text{decorrFunc}(M_{\text{pre}} Y_{\text{seco}}).$$

[538] A matriz de mistura $P = \begin{pmatrix} P_{\text{seco}} & P_{\text{molnt}} \end{pmatrix}$ é descrita em 20.2.3. As matrizes M_{pre} para configuração de saída diferente são dadas nas Figs. 19 a 23 e as matrizes M_{pos} são obtidas usando a seguinte equação:

$$M_{\text{pos}} = M_{\text{pre}}^* (M_{\text{pre}} M_{\text{pre}}^*)^{-1}.$$

[540] O modo de decodificação é controlado pelo elemento de bitstream `bsNumSaocDmxObjects`, como mostrado na Fig. 32.

[541] 20.2.2.1.1 Modo de Decodificação Combinado

[542] No caso do modo de decodificação combinado a matriz de não mistura paramétrica U é dada por:

$$U = E D^* J.$$

[544] A matriz J de tamanho $N_{\text{dmx}} \times N_{\text{dmx}}$ é dada por $J \approx \Lambda^{-1}$ com $\Lambda = D E D^*.$

[545] 20.2.2.1.2 Modo de Decodificação Independente

[546] No caso do modo de decodificação independente a matriz de não mistura $\mathbf{U}$ é dada por:

$$[547] \quad \mathbf{U} = \begin{pmatrix} \mathbf{U}_{\text{ch}} & 0 \\ 0 & \mathbf{U}_{\text{obj}} \end{pmatrix},$$

[548] onde $\mathbf{U}_{\text{ch}} = \mathbf{E}_{\text{ch}} \mathbf{D}_{\text{ch}}^* \mathbf{J}_{\text{ch}}$ e $\mathbf{U}_{\text{obj}} = \mathbf{E}_{\text{obj}} \mathbf{D}_{\text{obj}}^* \mathbf{J}_{\text{obj}}$.

[549] A matriz de covariância baseada no canal $\mathbf{E}_{\text{ch}}$ de tamanho $N_{\text{ch}} \times N_{\text{ch}}$ e a matriz de covariância baseada no objeto $\mathbf{E}_{\text{obj}}$ de tamanho $N_{\text{obj}} \times N_{\text{obj}}$ são obtidas a partir da matriz de covariância $\mathbf{E}$ pela seleção somente dos blocos diagonais correspondentes:

$$[550] \quad \mathbf{E} = \begin{pmatrix} \mathbf{E}_{\text{ch}} & \mathbf{E}_{\text{ch,obj}} \\ \mathbf{E}_{\text{obj,ch}} & \mathbf{E}_{\text{obj}} \end{pmatrix},$$

[551] onde a matriz $\mathbf{E}_{\text{ch,obj}} = (\mathbf{E}_{\text{obj,ch}})^*$ representa a matriz de covariância cruzada entre os canais de entrada e objetos de entrada e não precisa ser calculada.

[552] A matriz de downmix baseada em canal $\mathbf{D}_{\text{ch}}$ de tamanho $N_{\text{ch}}^{\text{dmx}} \times N_{\text{ch}}$ e a matriz de downmix baseada em objeto $\mathbf{D}_{\text{obj}}$ de tamanho $N_{\text{obj}}^{\text{dmx}} \times N_{\text{obj}}$ são obtidas a partir da matriz de downmix $\mathbf{D}$ pela seleção somente dos blocos diagonais correspondentes:

$$[553] \quad \mathbf{D} = \begin{pmatrix} \mathbf{D}_{\text{ch}} & 0 \\ 0 & \mathbf{D}_{\text{obj}} \end{pmatrix}.$$

[554] A matriz $\mathbf{J}_{\text{ch}} \approx (\mathbf{D}_{\text{ch}} \mathbf{E}_{\text{ch}} \mathbf{D}_{\text{ch}}^*)^{-1}$ de tamanho $N_{\text{ch}}^{\text{dmx}} \times N_{\text{ch}}^{\text{dmx}}$ é derivada portanto para 20.2.2.1.4 para $\Delta = \mathbf{D}_{\text{ch}} \mathbf{E}_{\text{ch}} \mathbf{D}_{\text{ch}}^*$.

[555] A matriz $\mathbf{J}_{\text{obj}} \approx (\mathbf{D}_{\text{obj}} \mathbf{E}_{\text{obj}} \mathbf{D}_{\text{obj}}^*)^{-1}$ de tamanho $N_{\text{obj}}^{\text{dmx}} \times N_{\text{obj}}^{\text{dmx}}$ é derivada portanto para 20.2.2.1.4 para $\Delta = \mathbf{D}_{\text{obj}} \mathbf{E}_{\text{obj}} \mathbf{D}_{\text{obj}}^*$.

[556] 20.2.2.1.4 Cálculo da matriz $\mathbf{J}$

[557] A matriz $\mathbf{J} \approx \mathbf{\Lambda}^{-1}$ é calculada usando a seguinte equação:

[558]
$$\mathbf{J} = \mathbf{V} \mathbf{\Lambda}^{inv} \mathbf{V}^*.$$

[559] Aqui o vetor singular $\mathbf{v}$ da matriz $\mathbf{\Lambda}$ são obtidos usando a seguinte equação de característica:

[560]
$$\mathbf{V} \mathbf{\Lambda} \mathbf{V}^* = \mathbf{\Lambda}.$$

[561] O inverso regularizado $\mathbf{\Lambda}^{inv}$ da matriz de valor singular diagonal $\mathbf{\Lambda}$ é calculado como

$$\lambda_{i,j}^{inv} = \begin{cases} \frac{1}{\lambda_{i,j}}, & \text{se } i = j \text{ e } \lambda_{i,j} \geq T_{reg}^{\Lambda} \\ 0, & \text{caso contrário} \end{cases},$$

[562]

[563] O escalar de regularização relativa T_{reg}^{Λ} é determinado usando o limiar absoluto T_{reg} e valor máximo de $\mathbf{\Lambda}$ como

[564]
$$T_{reg}^{\Lambda} = \max(\lambda_{i,i}) T_{reg}, \quad T_{reg} = 10^{-2}.$$

[565] 20.2.3. Descorrelação

[566] Os sinais descorrelacionados $\mathbf{X}_d$ são criados a partir do descorrelacionador descrito em 6.6.2 do ISO/IEC 23003-1:2007, com `bsDecorConfig == 0` e um índice do descorrelacionador, X , de acordo com as tabelas nas Figs. 19 a 24. Por isso, a $decorrFunc()$ denota o processo de descorrelação:

[567]
$$\mathbf{X}_d = decorrFunc(\mathbf{M}_{pre} \mathbf{Y}_{seco}).$$

[568] 20.2.4. Matriz de mistura $\mathbf{P}$ - Primeira Opção

[569] O cálculo da matriz de mistura $\mathbf{P} = \begin{pmatrix} \mathbf{P}_{seca} & \mathbf{P}_{molin} \end{pmatrix}$ é controlado pelo elemento de bitstream `bsDecorrelationMethod`. A matriz $\mathbf{P}$ tem tamanho $N_{out} \times 2N_{out}$ e a $\mathbf{P}_{seca}$ e $\mathbf{P}_{molin}$ tem ambas o tamanho $N_{out} \times N_{out}$.

[570] 20.2.4.1 O modo de compensação de energia

[571] O modo de compensação de energia usa sinais descorrelacionados

para compensar a perda de energia na reconstrução paramétrica. As matrizes de mistura $\mathbf{P}_{\text{dry}}$ e $\mathbf{P}_{\text{wet}}$ são dadas por:

$$[572] \quad \mathbf{P}_{\text{seca}} = \mathbf{I},$$

$$p_{i,j}^{\text{seca}} = \begin{cases} \min \left(\lambda_{\text{Dec}}, \max \left(0, \frac{\mathbf{C}(i,i) - \mathbf{E}_Y^{\text{seca}}(i,i)}{\max(\varepsilon, \mathbf{E}_Y^{\text{molh}}(i,i))} \right) \right) & i = j, \\ 0 & i \neq j. \end{cases}$$

[573]

[574] onde $\lambda_{\text{Dec}} = 4$ é uma constante usada para limitar a quantidade de componente descorrelacionado adicionado aos sinais de saída.

[575] 20.2.4.2 O modo de ajuste de covariância limitado

[576] O modo de ajuste de covariância limitado garante que a matriz de covariância dos sinais descorrelacionados misturados $\mathbf{P}_{\text{molh}} \mathbf{Y}_{\text{seca}}$ se aproxima da matriz de covariância da diferença $\Delta_E : \mathbf{P}_{\text{molh}} \mathbf{E}_Y^{\text{molh}} \mathbf{P}_{\text{molh}}^* \approx \Delta_E$. As matrizes de mistura $\mathbf{P}_{\text{seca}}$ e $\mathbf{P}_{\text{molhada}}$ são definidas usando as seguintes equações:

$$[577] \quad \mathbf{P}_{\text{seca}} = \mathbf{I},$$

$$[578] \quad \mathbf{P}_{\text{molh}} = \left(\mathbf{V}_1 \sqrt{\mathbf{Q}_1} \mathbf{V}_1^* \right) \left(\mathbf{V}_2 \sqrt{\mathbf{Q}_2^{\text{inv}}} \mathbf{V}_2^* \right),$$

[579] onde o inverso regularizado $\mathbf{Q}_2^{\text{inv}}$ da matriz de valor singular diagonal $\mathbf{Q}_2$ é calculado como

$$[580] \quad \mathbf{Q}_2^{\text{inv}}(i,j) = \begin{cases} \frac{1}{\mathbf{Q}_2(i,j)}, & \text{se } i = j \text{ e } \mathbf{Q}_2(i,j) \geq T_{\text{reg}}^{\Lambda}, \\ 0, & \text{caso contrário} \end{cases}$$

[581] O escalar de regularização relativa T_{reg}^{Λ} é determinado usando o limiar absoluto T_{reg} e valor máximo de $\mathbf{Q}_2^{\text{inv}}$ as

$$[582] \quad T_{\text{reg}}^{\Lambda} = \max(\mathbf{Q}_2^{\text{inv}}(i,i)) T_{\text{reg}}, \quad T_{\text{reg}} = 10^{-2}.$$

[583] A matriz Δ_E é decomposta usando a Decomposição de Valor

Singular as:

[584] $\Delta_E = \mathbf{V}_1 \mathbf{Q}_1 \mathbf{V}_1^*$.

[585] A matriz de covariância dos sinais descorrelacionados $\mathbf{E}_Y^{\text{moln}}$ também é expressa usando a Decomposição de Valor Singular:

[586] $\mathbf{E}_Y^{\text{moln}} = \mathbf{V}_2 \mathbf{Q}_2 \mathbf{V}_2^*$.

[587] 20.2.4.3. Modo de ajuste de covariância geral

[588] O modo de ajuste de covariância geral garante que a matriz de covariância dos sinais de saída finais $\hat{\mathbf{Y}}$ ($\mathbf{E}_{\hat{\mathbf{Y}}} = \hat{\mathbf{Y}} \hat{\mathbf{Y}}^*$) se aproxima da matriz de covariância alvo: $\mathbf{E}_{\hat{\mathbf{Y}}} \approx \mathbf{C}$. A matriz de mistura $\mathbf{P}$ é definida usando a seguinte equação:

[589] $\mathbf{P} = \left(\mathbf{V}_1 \sqrt{\mathbf{Q}_1} \mathbf{V}_1^* \right) \mathbf{H} \left(\mathbf{V}_2 \sqrt{\mathbf{Q}_2^{\text{inv}}} \mathbf{V}_2^* \right),$

[590] onde o inverso regularizado $\mathbf{Q}_2^{\text{inv}}$ da matriz de valor singular diagonal $\mathbf{Q}_2$ é calculado como

[591]
$$\mathbf{Q}_2^{\text{inv}}(i, j) = \begin{cases} \frac{1}{\mathbf{Q}_2(i, j)}, & \text{se } i = j \text{ e } \mathbf{Q}_2(i, j) \geq T_{\text{reg}}^{\Lambda}, \\ 0, & \text{caso contrário} \end{cases}$$

[592] O escalar de regularização relativa T_{reg}^{Λ} é determinado usando limiar absoluto T_{reg} e valor máximo de $\mathbf{Q}_2^{\text{inv}}$ as

[593] $T_{\text{reg}}^{\Lambda} = \max \left(\mathbf{Q}_2^{\text{inv}}(i, i) \right) T_{\text{reg}}, \quad T_{\text{reg}} = 10^{-2}.$

[594] A matriz de covariância alvo $\mathbf{C}$ é decomposta usando a Decomposição de Valor Singular como:

[595] $\mathbf{C} = \mathbf{V}_1 \mathbf{Q}_1 \mathbf{V}_1^*$.

[596] A matriz de covariância dos sinais combinados $\mathbf{E}_Y^{\text{com}}$ também é expressa usando a Decomposição de Valor Singular:

[597] $\mathbf{E}_Y^{\text{com}} = \mathbf{V}_2 \mathbf{Q}_2 \mathbf{V}_2^*$.

[598] A matriz $\mathbf{H}$ representa uma matriz de ponderação de protótipo de tamanho $(N_{\text{out}} \times 2N_{\text{out}})$ e é dada pela seguinte equação:

$$\mathbf{H} = \begin{pmatrix} 1/\sqrt{2} & 0 & \dots & 0 & 1/\sqrt{2} & 0 & \dots & 0 \\ 0 & 1/\sqrt{2} & \dots & 0 & 0 & 1/\sqrt{2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1/\sqrt{2} & 0 & 0 & \dots & 1/\sqrt{2} \end{pmatrix}.$$

[599]

[600] 20.2.4.4 Matrizes de Covariância Introduzidas

[601] A matriz Δ_E representa a diferença entre a matriz de covariância de saída alvo $\mathbf{C}$ e a matriz de covariância $\mathbf{E}_Y^{\text{seco}}$ dos sinais parametricamente reconstruídos e é dada por:

$$\Delta_E = \mathbf{C} - \mathbf{E}_Y^{\text{seco}}.$$

[602] A matriz $\mathbf{E}_Y^{\text{seco}}$ representa a matriz de covariância dos sinais parametricamente estimados $\mathbf{E}_Y^{\text{seco}} \approx \mathbf{Y}_{\text{seco}} \mathbf{Y}_{\text{seco}}^*$ e é definida usando a seguinte equação:

$$[603] \quad \mathbf{E}_Y^{\text{seco}} = \mathbf{R} \mathbf{U} \mathbf{E} \mathbf{U}^* \mathbf{R}^*.$$

[604] A matriz $\mathbf{E}_Y^{\text{molh}}$ representa a matriz de covariância dos sinais descorrelacionados $\mathbf{E}_Y^{\text{molh}} \approx \mathbf{Y}_{\text{molh}} \mathbf{Y}_{\text{molh}}^*$ e é definida usando a seguinte equação:

$$[605] \quad \mathbf{E}_Y^{\text{molh}} = \mathbf{M}_{\text{pós}} : \left[\text{matdiag}(\mathbf{M}_{\text{pré}} \mathbf{E}_Y^{\text{seco}} \mathbf{M}_{\text{pré}}^*) \right] \mathbf{M}_{\text{pós}}^*.$$

[606] Considerando o sinal $\mathbf{Y}_{\text{com}}$ que consiste em uma combinação de sinais paramétricos estimados e descorrelacionados:

$$\mathbf{Y}_{\text{com}} = \begin{pmatrix} \mathbf{Y}_{\text{seco}} \\ \mathbf{Y}_{\text{molh}} \end{pmatrix},$$

[607]

[608] a matriz de covariância de $\mathbf{Y}_{\text{com}}$ é definida pela seguinte equação:

$$\mathbf{E}_Y^{\text{com}} = \begin{pmatrix} \mathbf{E}_Y^{\text{seco}} & 0 \\ 0 & \mathbf{E}_Y^{\text{molh}} \end{pmatrix}.$$

[609]

[610] A matriz $\hat{\mathbf{E}}_Y^{\text{molh}}$ representa, por exemplo, a matriz de covariância estimada dos sinais descorrelacionados após a matriz de mistura $\mathbf{P}_{\text{molh}}$ ter sido aplicada, e é definida usando a seguinte equação:

$$\hat{\mathbf{E}}_Y^{\text{molh}} = \mathbf{P}_{\text{molh}} \mathbf{E}_Y^{\text{molh}} \mathbf{P}_{\text{seca}}^*.$$

[611]

[612] 20.2.5. Matriz de mistura $\mathbf{P}$ - Segunda Opção

[613] O cálculo da matriz de mistura $\mathbf{P} = \begin{bmatrix} \mathbf{P}_{\text{seca}} & \mathbf{A}_{\text{molh}} \mathbf{P}_{\text{molh}} \end{bmatrix}$ é controlado pelo elemento de bitstream **bsDecorrelationMethod**. A matriz $\mathbf{P}$ tem o tamanho $N_{\text{out}} \times 2N_{\text{out}}$ e as matrizes $\mathbf{P}_{\text{seca}}$ e $\mathbf{P}_{\text{molh}}$ tem ambas o tamanho $N_{\text{out}} \times N_{\text{out}}$. A matriz de limitação $\mathbf{A}_{\text{molh}}$ de tamanho $N_{\text{out}} \times N_{\text{out}}$ é dada por:

$$\mathbf{A}_{\text{molh}} = \text{matdiag} \left(\min \left(1, \sqrt{\max \left(0, \lambda_{\text{Dec}} \frac{\mathbf{E}_Y^{\text{seco}}(i,i)}{\max(\varepsilon, \hat{\mathbf{E}}_Y^{\text{molh}}(i,i))} \right)} \right) \right),$$

[614]

[615] onde as matrizes de covariância $\mathbf{E}_Y^{\text{seco}}$, $\mathbf{E}_Y^{\text{molh}}$ e $\hat{\mathbf{E}}_Y^{\text{molh}}$ são dadas, por exemplo, na seção 20.2.4.4 e $\lambda_{\text{Dec}} = 4$ é a constante usada para limitar a quantidade do componente descorrelacionado adicionado aos sinais de saída.

[616] 20.2.5.1 Modo de compensação de energia

[617] O modo de compensação de energia usa sinais descorrelacionados para compensar a perda de energia na reconstrução paramétrica. As matrizes de mistura $\mathbf{P}_{\text{seca}}$ e $\mathbf{P}_{\text{molh}}$ são dadas por:

$$\mathbf{P}_{\text{seca}} = \mathbf{I},$$

[618]

$$p_{i,j}^{\text{molh}} = \begin{cases} \sqrt{\max\left(0, \frac{C(i,i) - E_Y^{\text{seco}}(i,i)}{\max(\varepsilon, E_Y^{\text{molh}}(i,i))}\right)} & i = j, \\ 0 & i \neq j. \end{cases}$$

[619]

[620] 20.2.5.2 Conceitos e detalhes adicionais

[621] Em relação aos conceitos e detalhes adicionais, referência também é feita às seções 20.2.4.2 a 20.2.4.4.

[622] 20.3 Observações relativas à notação

[623] Deve-se observar que diferentes notações são usadas dentro da presente invenção. Entretanto, é claro a partir do contexto que a notação se aplica a uma equação específica.

[624] Por exemplo, a matriz de mistura é designada com F ou $\tilde{F}$ em algumas partes da descrição, enquanto a matriz de mistura é designada com P em outras partes da descrição.

[625] Além disso, um componente da matriz de mistura a ser aplicado a um sinal seco (ou aos sinais secos) é designado com P em algumas partes da descrição e com P_{seco} em outras partes da descrição. Da mesma forma, um componente da matriz de mistura a ser aplicado a um sinal molhado (ou aos sinais molhados) é designado com M em algumas partes da descrição e com P_{molhado} em outras partes da descrição. Além disso, a matriz de covariância E_w dos sinais molhados (antes da etapa de mistura com matriz M) é igual à matriz de covariância E_Y^{molh} dos sinais descorrelacionados.

[626] 21. Alternativas de Implementação

[627] Embora alguns aspectos tenham sido descritos no contexto de um aparelho, é claro que estes aspectos também representam uma descrição do método correspondente, onde um bloco ou dispositivo corresponde a uma etapa do método ou um recurso de uma etapa do método. Analogamente, os aspectos descritos no contexto de uma etapa do método também representam uma descrição de um bloco ou item ou recurso correspondente de um aparelho

correspondente. Algumas ou todas as etapas do método podem ser executadas por (ou usando) um aparelho de hardware, como por exemplo, um microprocessador, um computador programável ou um circuito eletrônico. Em algumas modalidades, algumas uma ou mais das etapas mais importantes do método podem ser executadas por tal aparelho.

[628] O sinal de áudio codificado da invenção pode ser armazenado em um meio de armazenamento digital ou pode ser transmitido em um meio de transmissão como um meio de transmissão remoto ou um meio de transmissão com fio como a Internet.

[629] Dependendo de determinados requisitos de implementação, as modalidades da invenção podem ser implementadas em hardware ou em software. A implementação pode ser realizada usando um meio de armazenamento digital, por exemplo um disquete, um DVD, um Blu-Ray, um CD, uma ROM, uma PROM, uma EPROM, uma EEPROM ou uma memória FLASH, tendo sinais de controle legíveis eletronicamente armazenados nos mesmos, que cooperam (ou são capazes de cooperar) com um sistema de computador programável de modo que o respectivo método é realizado. Portanto, o meio de armazenamento digital pode ser legível por computador.

[630] Algumas modalidades de acordo com a invenção compreendem um portador de dados tendo sinais de controle legíveis eletronicamente, que são capazes de cooperar com um sistema de computador programável, de modo que um dos métodos descritos aqui é realizado.

[631] Geralmente, as modalidades da presente invenção podem ser implementadas como um produto de programa de computador com um código de programa, o código de programa sendo operacional para realizar um dos métodos quando o produto de programa de computador funciona em um computador. O código de programa pode, por exemplo, ser armazenado em um portador legível por máquina.

[632] Outras modalidades compreendem o programa de computador para realizar um dos métodos descritos aqui, armazenado em um portador legível

por máquina.

[633] Em outras palavras, uma modalidade do método da invenção é, portanto, um programa de computador que tem um código de programa para realizar um dos métodos descritos aqui, quando o programa de computador funciona em um computador.

[634] Uma modalidade adicional dos métodos da invenção é, portanto, um portador de dados (ou um meio de armazenamento digital, ou um meio legível por computador) compreendendo, armazenado nele, o programa de computador para realizar um dos métodos descritos aqui. O portador de dados, o meio de armazenamento digital ou o meio gravado são tipicamente tangíveis e/ou não-transitórios.

[635] Uma modalidade adicional do método da invenção é, portanto, um fluxo de dados ou uma sequência de sinais que representam p programa de computador para realizar um dos métodos descritos aqui. O fluxo de dados ou a sequência de sinais pode, por exemplo, ser configurado para ser transferido através de uma conexão de comunicação de dados, por exemplo através da Internet.

[636] Uma modalidade adicional compreende um meio de processamento, por exemplo um computador, ou um dispositivo lógico programável, configurado para ou adaptado para realizar um dos métodos descritos aqui.

[637] Uma modalidade adicional compreende um computador tendo instalado nele o programa de computador para realizar um dos métodos descritos aqui.

[638] Uma modalidade adicional de acordo com a invenção compreende um aparelho ou um sistema configurado para transferir (por exemplo, eletronicamente ou oticamente) um programa de computador para realizar um dos métodos descritos aqui para um receptor. O receptor pode, por exemplo, ser um computador, um dispositivo móvel, um dispositivo de memória ou similares. O aparelho ou sistema pode, por exemplo, compreender um servidor de arquivo para transferir o programa de computador para o receptor .

[639] Em algumas modalidades, um dispositivo lógico programável (por exemplo, um arranjo de porta programável em campo) pode ser usado para realizar algumas ou todas as funcionalidades dos métodos descritos aqui. Em algumas modalidades, um arranjo de porta programável em campo pode cooperar com um microprocessador para realizar um dos métodos descritos aqui. Geralmente, os métodos são de preferência realizados por qualquer aparelho de hardware.

[640] As modalidades descritas acima são meramente ilustrativas para os princípios da presente invenção. Compreende-se que as modificações e variações dos arranjos e os detalhes descritos aqui serão aparentes para outros versados na técnica. A intenção é, portanto, ser limitadas somente pelo escopo das reivindicações de patentes iminentes e não pelos detalhes específicos apresentados por meio de descrição e explicação das modalidades da presente invenção.

[641] Referências

[642] [BCC] C. Faller e F. Baumgarte, “Binaural Cue Coding - Part II: Schemes and applications,” IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003.

[643] [Blauert] J. Blauert, “Spatial Hearing – The Psychophysics of Human Sound Localization”, Revised Edition, The MIT Press, Londres, 1997.

[644] [JSC] C. Faller, “Parametric Joint-Coding of Audio Sources”, 120th AES Convention, Paris, 2006.

[645] [ISS1] M. Parvaix e L. Girin: “Informed Source Separation of underdetermined instantaneous Stereo Mixtures using Source Index Embedding”, IEEE ICASSP, 2010.

[646] [ISS2] M. Parvaix, L. Girin, J.-M. Brossier: “A watermarking-based method for informed source separation of audio signals with a single sensor”, IEEE Transactions on Audio, Speech and Language Processing, 2010.

[647] [ISS3] A. Liutkus e J. Pinel e R. Badeau e L. Girin e G. Richard: “Informed source separation through spectrogram coding and data embedding”,

Signal Processing Journal, 2011.

[648] [ISS4] A. Ozerov, A. Liutkus, R. Badeau, G. Richard: "Informed source separation: source coding meets source separation", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011.

[649] [ISS5] S. Zhang e L. Girin: "An Informed Source Separation System for Speech Signals", INTERSPEECH, 2011.

[650] [ISS6] L. Girin e J. Pinel: "Informed Audio Source Separation from Compressed Linear Stereo Mixtures", AES 42nd International Conference: Semantic Audio, 2011.

[651] [MPS] ISO/IEC, "Information technology – MPEG audio technologies – Part 1: MPEG Surround," ISO/IEC JTC1/SC29/WG11 (MPEG) international Standard 23003-1:2006.

[652] [OCD] J. Vilkamo, T. Bäckström, e A. Kuntz. "Optimized covariance domain framework for time-frequency processing of spatial audio", Journal of the Audio Engineering Society, 2013. em publicação.

[653] [SAOC1] J. Herre, S. Disch, J. Hilpert, O. Hellmuth: "From SAC To SAOC - Recent Developments in Parametric Coding of Spatial Audio", 22nd Regional UK AES Conference, Cambridge, UK, April 2007.

[654] [SAOC2] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers e W. Oomen: "Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding", 124th AES Convention, Amsterdam 2008.

[655] [SAOC] ISO/IEC, "MPEG audio technologies – Part 2: Spatial Audio Object Coding (SAOC)," ISO/IEC JTC1/SC29/WG11 (MPEG) International Standard 23003-2.

[656] Patente Internacional N°. WO/2006/026452, "MULTICHANNEL DECORRELATION IN SPATIAL AUDIO CODING" emitida em 9 de março de 2006.

REIVINDICAÇÕES

1. Decodificador de áudio multicanal (100; 700; 1550; 3000) para fornecer ao menos dois sinais de áudio de saída (112,114; 712,714; 1552a – 1552n; 3012) com base em uma representação codificada (110; 710; 1516a, 1516b,1518),

caracterizado por o decodificador de áudio multicanal ser configurado para renderizar (130; 1580) uma pluralidade de sinais de áudio decodificados (122; 1562a – 1562n, $\hat{x}$), que são obtidos com base na representação codificada, para uma cena alvo de multicanal na dependência de um ou mais parâmetros de renderização (132) que definem uma matriz de renderização, para obter uma pluralidade de sinais de áudio renderizados (134, 136; 1582a-1582n, $\hat{z}$), e

sendo que o decodificador de áudio multicanal é configurado para derivar (140; 1590) um ou mais sinais de áudio descorrelacionados (142,144; 1592a-1592n) a partir dos sinais de áudio renderizados, e

sendo que o decodificador de áudio multicanal é configurado para combinar (150;1598) os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída;

sendo que o decodificador de áudio multicanal é configurado para obter os sinais de áudio decodificados, que são renderizados para obter a pluralidade de sinais de áudio renderizados, usando uma reconstrução paramétrica (120; 1560).

sendo que os sinais de áudio decodificados são sinais de objeto reconstruídos, e

sendo que o decodificador de áudio multicanal é configurado para derivar os sinais de objeto reconstruídos a partir de um ou mais sinais de downmix (1516a, 1516b) usando uma informação de lado (1518).

2. decodificador de áudio multicanal, de acordo com a reivindicação 1, **caracterizado por** o decodificador de áudio multicanal ser configurado para

derivar coeficientes não misturados a partir da informação do lado e aplicar os coeficientes não misturados para derivar os sinais de objeto reconstruídos a partir de um ou mais sinais de downmix usando os coeficientes não misturados.

3. decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 2, **caracterizado por** o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para, pelo menos, parcialmente conseguir as características de correlação desejadas ou características de covariância dos sinais de áudio de saída.

4. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 3, **caracterizado por** o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para, pelo menos, parcialmente compensar uma perda de energia durante uma reconstrução paramétrica (120; 1560) dos sinais de áudio decodificados (122; 1562a a 1562n), que são renderizados para obter a pluralidade de sinais de áudio renderizados.

5. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 4, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar características de correlação desejadas ou características de covariância desejadas dos sinais de áudio de saída, e

sendo que o decodificador de áudio multicanal é configurado para ajustar uma combinação (150; 1598) dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída, de modo que as características de correlação ou características de covariância dos sinais de áudio de saída obtidas se aproximam ou são iguais às características de correlação ou características de covariância (C).

6. Decodificador de áudio multicanal, de acordo com a reivindicação 5, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar as características de correlação ou características de covariância (C) na dependência de uma informação de renderização (R) que descreve uma

renderização (130; 1560) da pluralidade de sinais de áudio decodificados ($\hat{X}$), que são obtidos com base na representação codificada, para obter a pluralidade de sinais de áudio renderizados($\hat{Z}$).

7. Decodificador de áudio multicanal, de acordo com a reivindicação 5 ou 6, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar as características de correlação ou características de covariância (**C**) na dependência de uma informação de correlação de objeto ou uma informação de covariância de objeto (**E_x**) que descreve as características de uma pluralidade de objetos de áudio e/ou uma relação entre uma pluralidade de objetos de áudio.

8. Decodificador de áudio multicanal, de acordo com a reivindicação 7, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar a informação de correlação de objeto ou informação de covariância de objeto (**E_x**) com base em uma informação de lado (1518) incluída na representação codificada.

9. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 5 a 8, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar características de correlação ou características de covariância reais (**E_s**) dos sinais de áudio renderizados e os um ou mais sinais de áudio descorrelacionados, e

para ajustar a combinação (150;1598) dos sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída, na dependência das características de correlação ou características de covariância reais (**E_s**) dos sinais de áudio renderizados e os um ou mais sinais de áudio descorrelacionado.

10. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 9,

caracterizado por o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados $\hat{Z}$ com os um ou mais sinais de áudio descorrelacionados **W**, para obter os sinais de áudio de saída $\tilde{Z}$

de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{Z}} + \mathbf{M}\mathbf{W},$$

sendo que $\mathbf{P}$ é uma matriz de mistura que é aplicada aos sinais de áudio renderizados $\hat{\mathbf{Z}}$, e

sendo que $\mathbf{M}$ é uma matriz de mistura que é aplicada aos um ou mais sinais de áudio descorrelacionados $\mathbf{W}$.

11. Decodificador de áudio multicanal, de acordo com a reivindicação 10,

caracterizado por o decodificador de áudio multicanal ser configurado para ajustar pelo menos uma das matriz de mistura $\mathbf{P}$ e a matriz de mistura $\mathbf{M}$ de modo que as características de correlação ou características de covariância ($E_{\tilde{\mathbf{Z}}}$) dos sinais de áudio de saída obtidos $\tilde{\mathbf{Z}}$ se aproximam ou são iguais às características de correlação ou características de covariância ($\mathbf{C}$).

12. Decodificador de áudio multicanal, de acordo com a reivindicação 10 ou 11,

caracterizado por o decodificador de áudio multicanal ser configurado para conjuntamente calcular a matriz de mistura $\mathbf{P}$ e a matriz de mistura $\mathbf{M}$.

13. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 10 a 12,

caracterizado por o decodificador de áudio multicanal ser configurado para obter uma matriz de mistura combinada $\mathbf{F}$, com

$$\mathbf{F} = [\mathbf{P} \quad \mathbf{M}]$$

de modo que uma matriz de covariância $E_{\tilde{\mathbf{Z}}}$ dos sinais de áudio de saída obtidos $\tilde{\mathbf{Z}}$ se aproxima ou é igual a uma matriz de covariância desejada $\mathbf{C}$.

14. Decodificador de áudio multicanal, de acordo com a reivindicação 13,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar a matriz de mistura combinada $\mathbf{F}$ de modo que a matriz de covariância

$$\mathbf{E}_{\tilde{\mathbf{Z}}} = \mathbf{F} \mathbf{E}_{\mathbf{S}} \mathbf{F}^H$$

é igual à matriz de covariância desejada

$$\mathbf{C} = \mathbf{R} \mathbf{E}_{\mathbf{X}} \mathbf{R}^H,$$

sendo que $\mathbf{E}_{\mathbf{S}}$ é uma matriz de covariância de um sinal $\mathbf{S}$ que combina os sinais de áudio renderizados $\hat{\mathbf{Z}}$ e os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, que é definida como

$$\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}, \text{ e}$$

sendo que $\mathbf{E}_{\mathbf{X}}$ é uma matriz de covariância de objeto.

15. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 9,

caracterizado por o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados $\hat{\mathbf{Z}}$ com os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, para obter os sinais de áudio de saída $\tilde{\mathbf{Z}}$ de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{A}_{seca} \mathbf{P} \hat{\mathbf{Z}} + \mathbf{M} \mathbf{W}$$

ou de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{P} \hat{\mathbf{Z}} + \mathbf{A}_{molhada} \mathbf{M} \mathbf{W}$$

ou de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{A}_{seca} \mathbf{P} \hat{\mathbf{Z}} + \mathbf{A}_{molhada} \mathbf{M} \mathbf{W}$$

sendo que $\mathbf{P}$ é uma matriz de mistura que é aplicada aos sinais de áudio renderizados $\hat{\mathbf{Z}}$, e

sendo que $\mathbf{M}$ é uma matriz de mistura que é aplicada aos um ou mais sinais de áudio descorrelacionados $\mathbf{W}$,

sendo que $\mathbf{A}_{seca}$ é uma primeira matriz de correlação ou uma primeira matriz de ajuste,

sendo que $\mathbf{A}_{molhada}$ é uma segunda matriz de correlação ou a segunda matriz de ajuste.

16. Decodificador de áudio multicanal, de acordo com a reivindicação 15,

caracterizado por o decodificador de áudio multicanal ser configurado para ajustar pelo menos uma das matriz de mistura **P** e a matriz de mistura **M** de modo que as características de correlação ou características de covariância ($E_{\tilde{Z}}$) dos sinais de áudio de saída obtidos $\tilde{Z}$ ou dos sinais de áudio obtidos por uma mistura de $\hat{Z}$ e **W** usando **P** e **M** se aproximam ou são iguais às características de correlação ou características de covariância (**C**).

17. Decodificador de áudio multicanal, de acordo com a reivindicação 15 ou 16,

caracterizado por o decodificador de áudio multicanal ser configurado para conjuntamente calcular a matriz de mistura **P** e a matriz de mistura **M**.

18. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 15 a 17,

caracterizado por o decodificador de áudio multicanal ser configurado para obter uma matriz de mistura combinada **F**, com

$$\mathbf{F} = [\mathbf{P} \quad \mathbf{M}]$$

de modo que uma matriz de covariância $E_{\tilde{Z}}$ dos sinais de áudio de saída obtidos $\tilde{Z}$ ou a matriz de covariância dos sinais de áudio obtidos por uma mistura de $\hat{Z}$ e **W** usando **P** e **M** se aproxima ou é igual a uma matriz de covariância desejada **C**.

19. Decodificador de áudio multicanal, de acordo com a reivindicação 18,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar a matriz de mistura combinada **F** de modo que a matriz de covariância

$$\mathbf{E}_{\tilde{Z}} = \mathbf{F} \mathbf{E}_S \mathbf{F}^H$$

é igual à matriz de covariância desejada

$$\mathbf{C} = \mathbf{R} \mathbf{E}_X \mathbf{R}^H,$$

sendo que **E_S** é uma matriz de covariância de um sinal **S** que combina

os sinais de áudio renderizados $\hat{\mathbf{Z}}$ e os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, que é definida como

$$\mathbf{S} = \begin{bmatrix} \hat{\mathbf{Z}} \\ \mathbf{W} \end{bmatrix}, \text{ e}$$

sendo que $\mathbf{E}_x$ é uma matriz de covariância de objeto.

20. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 15 a 19,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar a primeira matriz de correlação de modo que uma contribuição dos sinais de áudio renderizados sobre os sinais de áudio de saída é limitada, e/ou

sendo que o decodificador de áudio multicanal é configurado para determinar a segunda matriz de correlação de modo que uma contribuição dos sinais de áudio descorrelacionados sobre os sinais de áudio de saída é limitada.

21. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 15 a 20,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar a primeira matriz de correlação na dependência das propriedades dos sinais de áudio renderizados, e/ou na dependência das propriedades dos sinais de áudio descorrelacionados, e/ou na dependência das propriedades dos sinais de áudio de saída desejados, e/ou na dependência das propriedades estimadas dos sinais de áudio renderizados misturados, e/ou na dependência das propriedades estimadas dos sinais de áudio descorrelacionados misturados, de modo que uma contribuição dos sinais de áudio renderizados sobre os sinais de áudio de saída é limitada, e/ou

sendo que o decodificador de áudio multicanal é configurado para determinar a segunda matriz de correlação na dependência das propriedades dos sinais de áudio renderizados, e/ou na dependência das propriedades dos sinais de áudio descorrelacionados, e/ou na dependência das propriedades dos sinais de áudio de saída desejados, e/ou na dependência das propriedades

estimadas dos sinais de áudio renderizados misturados, e/ou na dependência das propriedades estimadas dos sinais de áudio descorrelacionados misturados, de modo que uma contribuição dos sinais de áudio descorrelacionados sobre os sinais de áudio de saída é limitada.

22. Decodificador de áudio multicanal, de acordo com a reivindicação 21, **caracterizado por** as propriedades dos sinais de áudio renderizados, e/ou dos sinais de áudio descorrelacionados, e/ou dos sinais de áudio de saída desejados, e/ou dos sinais de áudio renderizados misturados, e/ ou os sinais de áudio descorrelacionados misturados serem propriedades de energia, ou propriedades de correlação, ou propriedades de covariância.

23. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 22,

caracterizado por o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados $\hat{\mathbf{Z}}$ com os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, para obter os sinais de áudio de saída $\tilde{\mathbf{Z}}$ de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{Z}} + \mathbf{A}_{molhada}\mathbf{M}\mathbf{W},$$

sendo que o decodificador de áudio multicanal é configurado para fornecer a matriz de correlação $\mathbf{A}_{molhada}$ de modo que $\mathbf{A}_{molhada}$ é uma matriz diagonal e de modo que entradas $\mathbf{A}_{molhada}(i,i)$ da matriz de correlação $\mathbf{A}_{molhada}$ são reduzidas se comparadas às entradas diagonais normais, não reduzidas da matriz de correlação $\mathbf{A}_{molhada}$ se uma razão entre uma intensidade ($E_Y^{dry}(i,i)$) de um sinal de áudio renderizado e uma intensidade ($\hat{E}_Y^{wet}(i,i)$) de um sinal de áudio descorrelacionado misturado, com a matriz de mistura $\mathbf{M}$, em um i-ésimo sinal de áudio de saída fosse inferior a um valor limiar.

24. Decodificador de áudio multicanal, de acordo com a reivindicação 23, **caracterizado por** o valor limiar ser um valor limiar constante predeterminado ou sendo que o valor limiar é variável em tempo e/ou variável em frequência na dependência das propriedades do sinal, por exemplo, propriedades de energia, propriedades de correlação e/ou propriedades de

covariância.

25. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 24,

caracterizado por o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados $\hat{\mathbf{Z}}$ com os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, para obter os sinais de áudio de saída $\tilde{\mathbf{Z}}$ de acordo com

$$\tilde{\mathbf{Z}} = \mathbf{P}\hat{\mathbf{Z}} + \mathbf{A}_{molhado}\mathbf{M}\mathbf{W},$$

sendo que $\mathbf{P} = \mathbf{P}_{seca}$,

sendo que $\mathbf{M} = \mathbf{P}_{molhada}$,

sendo que

$$\mathbf{A}_{molh} = \text{matdiag} \left(\min \left(1, \sqrt{\max \left(0, \lambda_{Dec} \frac{\mathbf{E}_Y^{seco}(i, i)}{\max(\varepsilon, \hat{\mathbf{E}}_Y^{molh}(i, i))} \right)} \right) \right),$$

sendo que $\mathbf{E}_Y^{seco}$ é uma matriz de covariância dos sinais de áudio renderizados $\hat{\mathbf{Z}}$, e

sendo que $\hat{\mathbf{E}}_Y^{molhado}$ é uma matriz de covariância estimada dos sinais de áudio descorrelacionados após a matriz $\mathbf{P}_{molhado}$ ter sido aplicada.

26. Decodificador de áudio multicanal, de acordo com a reivindicação 15, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar a matriz de mistura combinada $\mathbf{F}$ de acordo com

$$\mathbf{F} = (\mathbf{U}\sqrt{\mathbf{T}}\mathbf{U}^H) \mathbf{H} (\mathbf{V}\sqrt{\mathbf{Q}^{-1}}\mathbf{V}^H),$$

onde as matrizes $\mathbf{U}$, $\mathbf{T}$, $\mathbf{V}$ e $\mathbf{Q}$ são determinadas usando Decomposição de Valor Singular das matrizes de covariância $\mathbf{E}_s$ e $\mathbf{C}$ produzindo

$$\mathbf{C} = \mathbf{U}\mathbf{T}\mathbf{U}^H,$$

e

$$\mathbf{E}_s = \mathbf{V}\mathbf{Q}\mathbf{V}^H,$$

sendo que a matriz $\mathbf{H}$ é definida como

$$\mathbf{H} = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & b_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 & 0 & b_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{N,N} & 0 & 0 & \dots & b_{N,N} \end{bmatrix}$$

sendo que $a_{i,i}$ e $b_{i,i}$ são escolhidos de modo que

$$a_{i,i}^2 + b_{i,i}^2 = 1.$$

27. Decodificador de áudio multicanal, de acordo com a reivindicação 10 ou 11,

caracterizado por o decodificador de áudio multicanal ser configurado para definir a matriz de mistura $\mathbf{P}$ para ser uma matriz de identidade, ou um múltiplo da mesma, e para calcular a matriz de mistura $\mathbf{M}$.

28. Decodificador de áudio multicanal, de acordo com a reivindicação 27, **caracterizado por** o decodificador de áudio multicanal ser configurado para determinar a matriz de mistura $\mathbf{M}$ de modo que uma diferença Δ_E entre a matriz de covariância desejada $\mathbf{C}$ e a matriz de covariância $\mathbf{E}_{\hat{z}}$, que é definida como

$$\Delta_E = \mathbf{C} - \mathbf{E}_{\hat{z}}$$

é igual a, ou se aproxima, uma covariância

$$\mathbf{M}\mathbf{E}_w\mathbf{M}^H,$$

sendo que a matriz de covariância desejada $\mathbf{C}$ é definida como

$$\mathbf{C} = \mathbf{R}\mathbf{E}_x\mathbf{R}^H,$$

sendo que $\mathbf{R}$ é uma matriz de renderização,

sendo que $\mathbf{E}_x$ é uma matriz de covariância de objeto, e

sendo que $\mathbf{E}_w$ é uma matriz de covariância dos um ou mais sinais descorrelacionados, e

sendo que $\mathbf{E}_{\hat{z}}$ é uma matriz de covariância dos sinais de áudio renderizados.

29. Decodificador de áudio multicanal, de acordo com a reivindicação 28,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar a matriz de mistura **M** de acordo com

$$\mathbf{M} = \left(\mathbf{U} \sqrt{\mathbf{T}} \mathbf{U}^H \right) \left(\mathbf{V} \sqrt{\mathbf{Q}^{-1}} \mathbf{V}^H \right),$$

onde as matrizes **U**, **T**, **V** e **Q** são determinadas usando Decomposição de Valor Singular das matrizes de covariância Δ_E e $\mathbf{E}_W$ produzindo

$$\Delta_E = \mathbf{U} \mathbf{T} \mathbf{U}^H$$

e

$$\mathbf{E}_W = \mathbf{V} \mathbf{Q} \mathbf{V}^H.$$

30. Decodificador de áudio multicanal, de acordo com a reivindicação 10 ou 11,

caracterizado por o decodificador de áudio multicanal ser configurado para determinar as matrizes de mistura **P**, **M** sob a restrição de que um dado sinal de áudio renderizado somente é misturado com uma versão descorrelacionada do dado próprio sinal de áudio renderizado.

31. Decodificador de áudio multicanal, de acordo com a reivindicação 10, 11 ou 30, **caracterizado por** o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados com os um ou mais sinais de áudio descorrelacionados de modo que somente valores de autocorrelação ou valores de autocovariância dos sinais de áudio renderizados são modificados enquanto valores de correlação cruzada ou valores de covariância cruzada são deixados inalterados.

32. Decodificador de áudio multicanal, de acordo com a reivindicação 10, 11, 30 ou 31,

caracterizado por o decodificador de áudio multicanal ser configurado para definir a matriz de mistura **P** para ser uma matriz de identidade, ou um múltiplo da mesma, e para calcular a matriz de mistura **M** sob a restrição de que **M** é uma matriz diagonal.

33. Decodificador de áudio multicanal, de acordo com a reivindicação

30, 31 ou 32, **caracterizado por** o decodificador de áudio multicanal ser configurado para combinar os sinais de áudio renderizados $\hat{\mathbf{Z}}$ com os um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, para obter os sinais de áudio de saída $\tilde{\mathbf{Z}}$ de acordo com

$$\tilde{\mathbf{Z}} = \hat{\mathbf{Z}} + \mathbf{M}\mathbf{W},$$

sendo que $\mathbf{M}$ é uma matriz de mistura diagonal que é aplicada aos um ou mais sinais de áudio descorrelacionados $\mathbf{W}$, e

sendo que o decodificador de áudio multicanal é configurado para calcular elementos diagonais da matriz de mistura $\mathbf{M}$ de modo que elementos diagonais de uma matriz de covariância dos sinais de áudio de saída são iguais às energias desejadas.

34. Decodificador de áudio multicanal, de acordo com a reivindicação 33, **caracterizado por** o decodificador de áudio multicanal ser configurado para calcular os elementos da matriz de mistura $\mathbf{M}$ de acordo com

$$\mathbf{M}(i, j) = \begin{cases} \sqrt{\min\left(\lambda_{Dec}, \max\left(0, \frac{\mathbf{C}(i, i) - \mathbf{E}_{\hat{\mathbf{Z}}}(i, i)}{\max(\mathbf{E}_{\mathbf{W}}(i, i), \varepsilon)}\right)\right)} & i = j, \\ 0 & i \neq j. \end{cases},$$

sendo que a matriz de covariância desejada $\mathbf{C}$ é definida como

$$\mathbf{C} = \mathbf{R}\mathbf{E}_X\mathbf{R}^H,$$

sendo que $\mathbf{R}$ é uma matriz de renderização,

sendo que $\mathbf{E}_X$ é uma matriz de covariância de objeto,

sendo que $\mathbf{E}_W$ é uma matriz de covariância dos um ou mais sinais descorrelacionados, e

sendo que λ_{Dec} é um valor limiar que limita uma quantidade de descorrelação adicionada aos sinais.

35. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 34, **caracterizado por** o decodificador de áudio multicanal ser configurado para considerar características de correlação ou características de covariância dos sinais de áudio descorrelacionados ao determinar como

combinar os sinais de áudio renderizados, ou a versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionado.

36. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 26 ou 35, **caracterizado por** o decodificador de áudio multicanal ser configurado para misturar sinais de áudio renderizados e sinais de áudio descorrelacionados, de modo que um dado sinal de áudio de saída é fornecido com base em dois ou mais sinais de áudio renderizados e pelo menos um sinal de áudio descorrelacionado.

37. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 36, **caracterizado por** o decodificador de áudio multicanal ser configurado para alternar entre modos diferentes, nos quais diferentes restrições são aplicadas para determinar como combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída.

38. Decodificador de áudio multicanal, de acordo com qualquer uma das reivindicações 1 a 37, **caracterizado por** o decodificador de áudio multicanal ser configurado para alternar entre

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de

áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

39. Decodificador de áudio multicanal, de acordo com a reivindicação 37 ou 38, **caracterizado por** o decodificador de áudio de multicanal é configurado para avaliar a elemento de bitstream da representação codificada que indica quais dos três modos para combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados deve ser usado, e selecionar o modo dependendo do referido elemento de bitstream.

40. Codificador de áudio de multicanal (200; 1510; 2900) para fornecer uma representação codificada (214; 1516a, 1516b, 1518; 2932) com base pelo menos em dois sinais de áudio de entrada (210,212; 1512a-1512n; 2912, 2914),

caracterizado por o codificador de áudio de multicanal ser configurado para fornecer (220) um ou mais sinais de downmix (222; 1516a,1516b) com base pelo menos em dois sinais de áudio de entrada, e

sendo que o codificador de áudio de multicanal é configurado para fornecer (230) um ou mais parâmetros (232; 1518) que descrevem uma relação entre a pelo menos dois sinais de áudio de entrada, e

sendo que o codificador de áudio de multicanal é configurado para fornecer (240) a parâmetro do método de descorrelação (242; 1518) que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio;

sendo que o codificador de áudio de multicanal é configurado para seletivamente fornecer o parâmetro do método de descorrelação, para sinalizar um dentre os três modos seguintes para a operação de um decodificador de

áudio:

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

41. Codificador de áudio de multicanal, de acordo com a reivindicação 40, **caracterizado por** o codificador de áudio de multicanal é configurado para selecionar o parâmetro do método de descorrelação na dependência de se os sinais de áudio de entrada compreendem uma correlação comparativamente elevada ou uma correlação comparativamente menor.

42. Codificador de áudio de multicanal, de acordo com a reivindicação 40 ou 41, **caracterizado por** o codificador de áudio de multicanal ser configurado para selecionar o parâmetro do método de descorrelação para designar o primeiro modo ou o segundo modo se uma correlação entre a sinais de áudio de entrada for comparativamente alta, e

sendo que o codificador de áudio de multicanal é configurado para selecionar o parâmetro do método de descorrelação para designar o terceiro modo se uma correlação entre a sinais de áudio de entrada for comparativamente inferior.

43. Método (300) para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada, o método **caracterizado por** compreender:

renderizar (310) uma pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, para uma cena alvo de multicanal na dependência de um ou mais parâmetros de renderização que definem uma matriz de renderização, para obter uma pluralidade de sinais de áudio renderizados,

derivar (320) um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados, e

combinar (330) os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída;

sendo que os sinais de áudio decodificados, que são renderizados para obter a pluralidade de sinais de áudio renderizados, são obtidos usando uma reconstrução paramétrica (120; 1560);

sendo que os sinais de áudio decodificados are sinais de objeto reconstruídos; e

sendo que os sinais de objeto reconstruídos são derivados de um ou mais sinais de downmix (1516a, 1516b) usando uma informação de lado (1518).

44. Método (400) para fornecer uma representação codificada com base pelo menos em dois sinais de áudio de entrada, o método **caracterizado por** compreender:

fornecer (410) um ou mais sinais de downmix com base pelo menos em dois sinais de áudio de entrada,

fornecer (420) um ou mais parâmetros que descrevem uma relação

entre a pelo menos dois sinais de áudio de entrada, e

fornecer (430) a parâmetro do método de descorrelação que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio;

sendo que o método compreende fornecer seletivamente o parâmetro do método de descorrelação, para sinalizar um dentre os três modos seguintes para a operação de um decodificador de áudio:

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

45. Representação de áudio codificada (500), **caracterizada por** compreender:

uma representação codificada (510) de um sinal de downmix;

uma representação codificada (520) de um ou mais parâmetros que

descrevem uma relação entre a pelo menos dois sinais de áudio de entrada, e um parâmetro codificado do método de descorrelação(530) que descreve qual modo de descorrelação dentre uma pluralidade de modos de descorrelação deve ser usado no lado de um decodificador de áudio;

sendo que o parâmetro do método de descorrelação sinaliza um dos três modos de operação a seguir de um decodificador de áudio:

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

46. Decodificador de áudio multicanal (100; 700; 1550; 3000) para fornecer ao menos dois sinais de áudio de saída (112,114; 712,714; 1552a – 1552n; 3012) com base em uma representação codificada (110; 710; 1516a, 1516b,1518),

caracterizado por o decodificador de áudio multicanal ser

configurado para renderizar (130; 1580) uma pluralidade de sinais de áudio decodificados (122; 1562a – 1562n, $\hat{X}$), que são obtidos com base na representação codificada, na dependência de um ou mais parâmetros de renderização (132), para obter uma pluralidade de sinais de áudio renderizados (134, 136; 1582a-1582n, $\hat{Z}$), e

sendo que o decodificador de áudio multicanal é configurado para derivar (140; 1590) um ou mais sinais de áudio descorrelacionados (142, 144; 1592a-1592n) a partir dos sinais de áudio renderizados, e

sendo que o decodificador de áudio multicanal é configurado para combinar (150; 1598) os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída;

sendo que o decodificador de áudio multicanal é configurado para alternar entre

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal

descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

47. Método (300) para fornecer ao menos dois sinais de áudio de saída com base em uma representação codificada, o método **caracterizado por** compreender:

renderizar (310) uma pluralidade de sinais de áudio decodificados, que são obtidos com base na representação codificada, na dependência de um ou mais parâmetros de renderização, para obter uma pluralidade de sinais de áudio renderizados,

derivar (320) um ou mais sinais de áudio descorrelacionados a partir dos sinais de áudio renderizados, e

combinar (330) os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, para obter os sinais de áudio de saída;

sendo que o método compreende alternar entre

um primeiro modo, no qual uma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados,

um segundo modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual é permitido que um dado sinal descorrelacionado seja combinado, com dimensionamento igual ou diferente, com uma pluralidade de sinais de áudio renderizados, ou uma versão reduzida dos mesmos, para ajustar características de correlação cruzada ou características de covariância cruzada dos sinais de áudio de saída, e

um terceiro modo no qual nenhuma mistura entre diferentes sinais de áudio renderizados é permitida ao combinar os sinais de áudio renderizados, ou

uma versão reduzida dos mesmos, com os um ou mais sinais de áudio descorrelacionados, e no qual não é permitido que um dado sinal descorrelacionado seja combinado com sinais de áudio renderizados diferente de um sinal de áudio renderizado a partir do qual o dado sinal descorrelacionado é derivado.

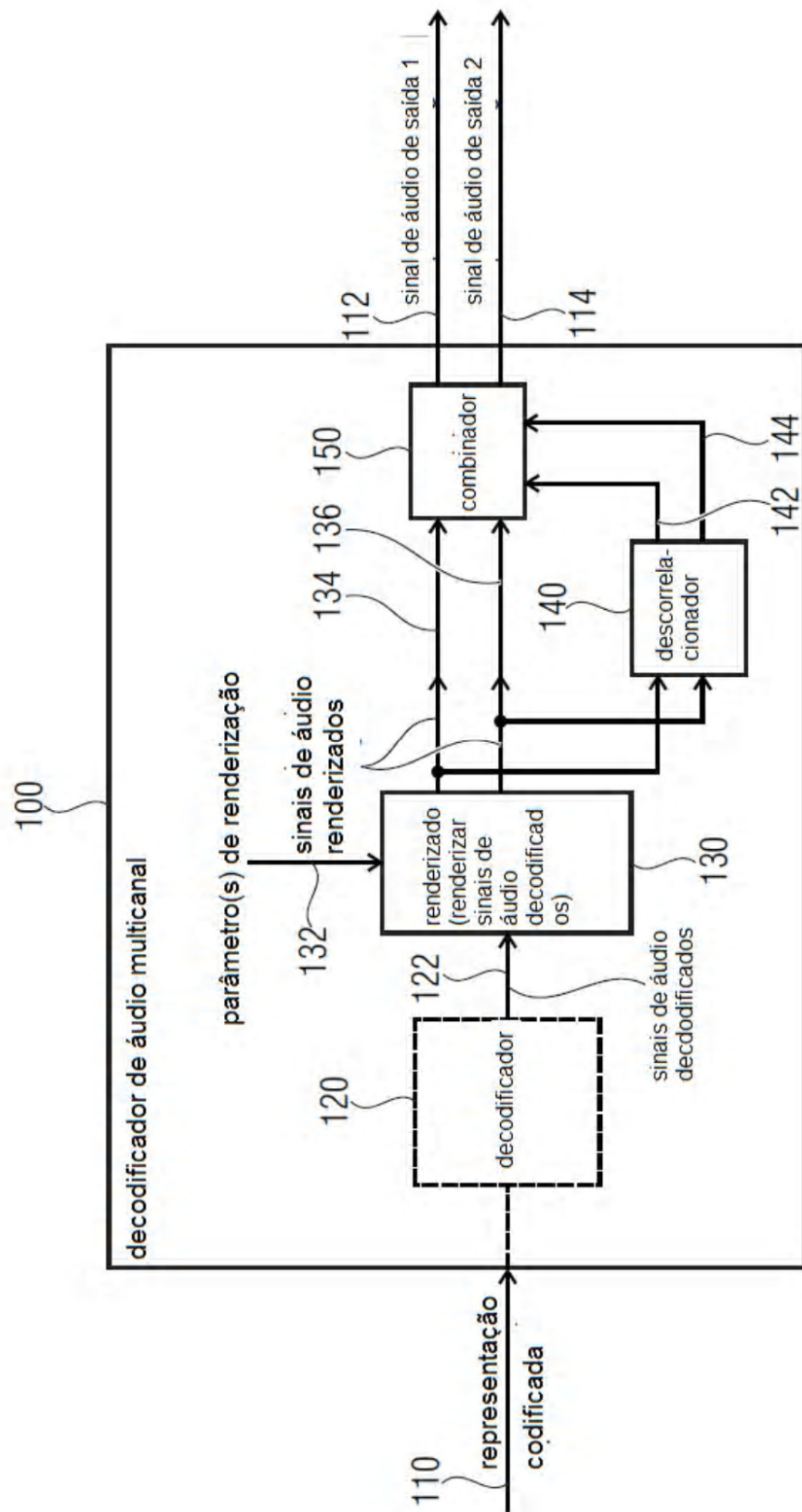


FIG 1

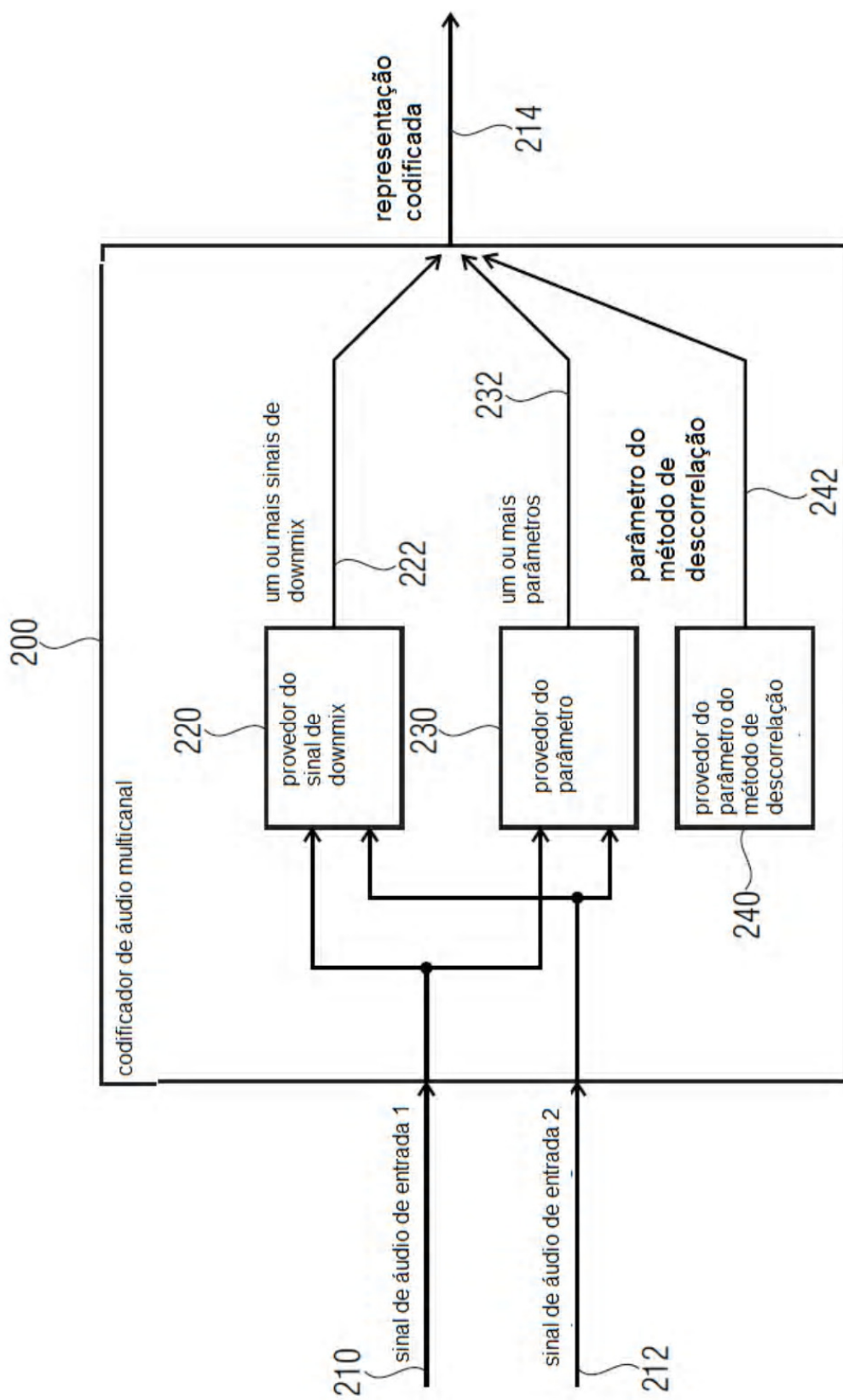


FIG 2

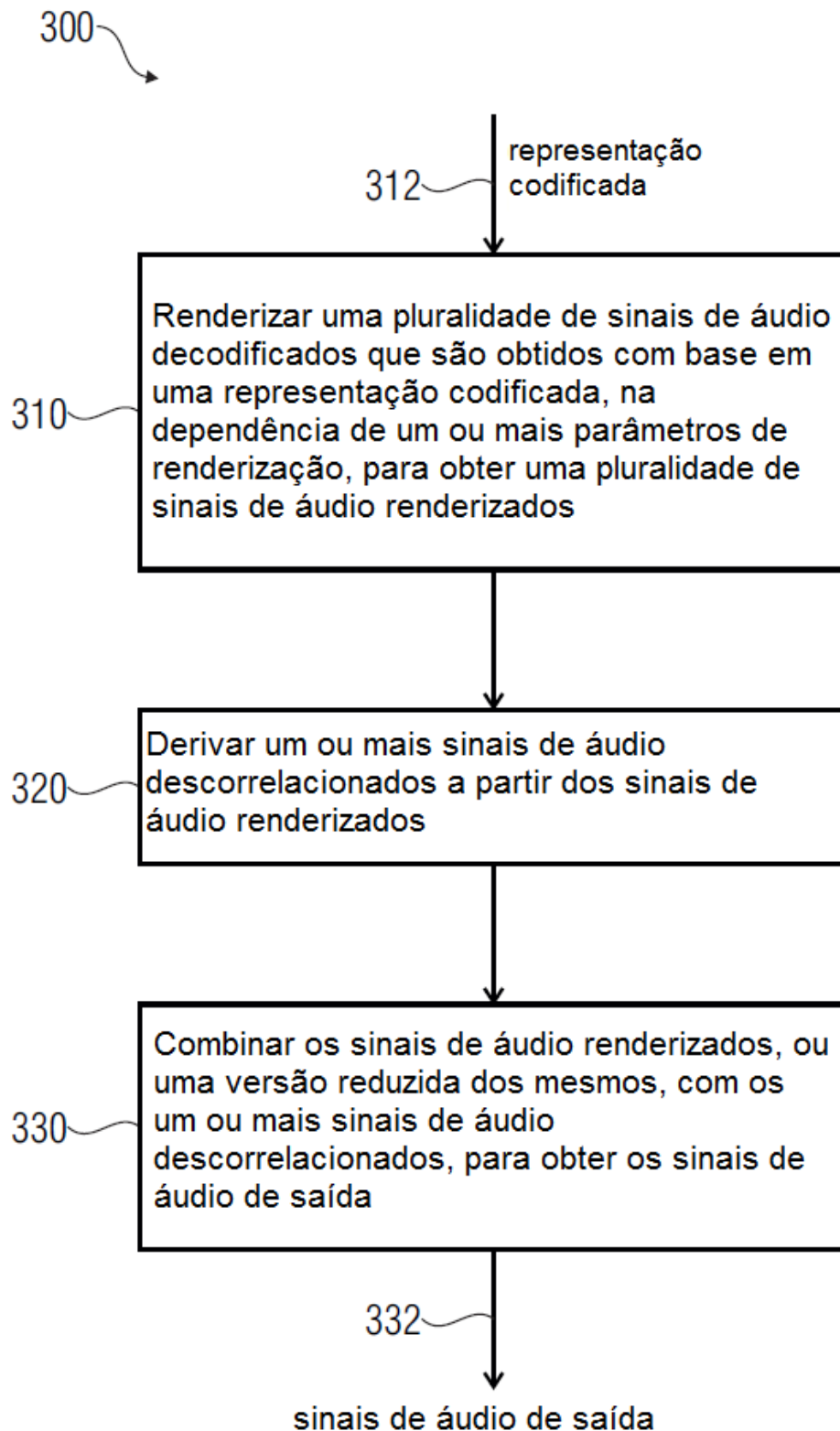


FIG 3

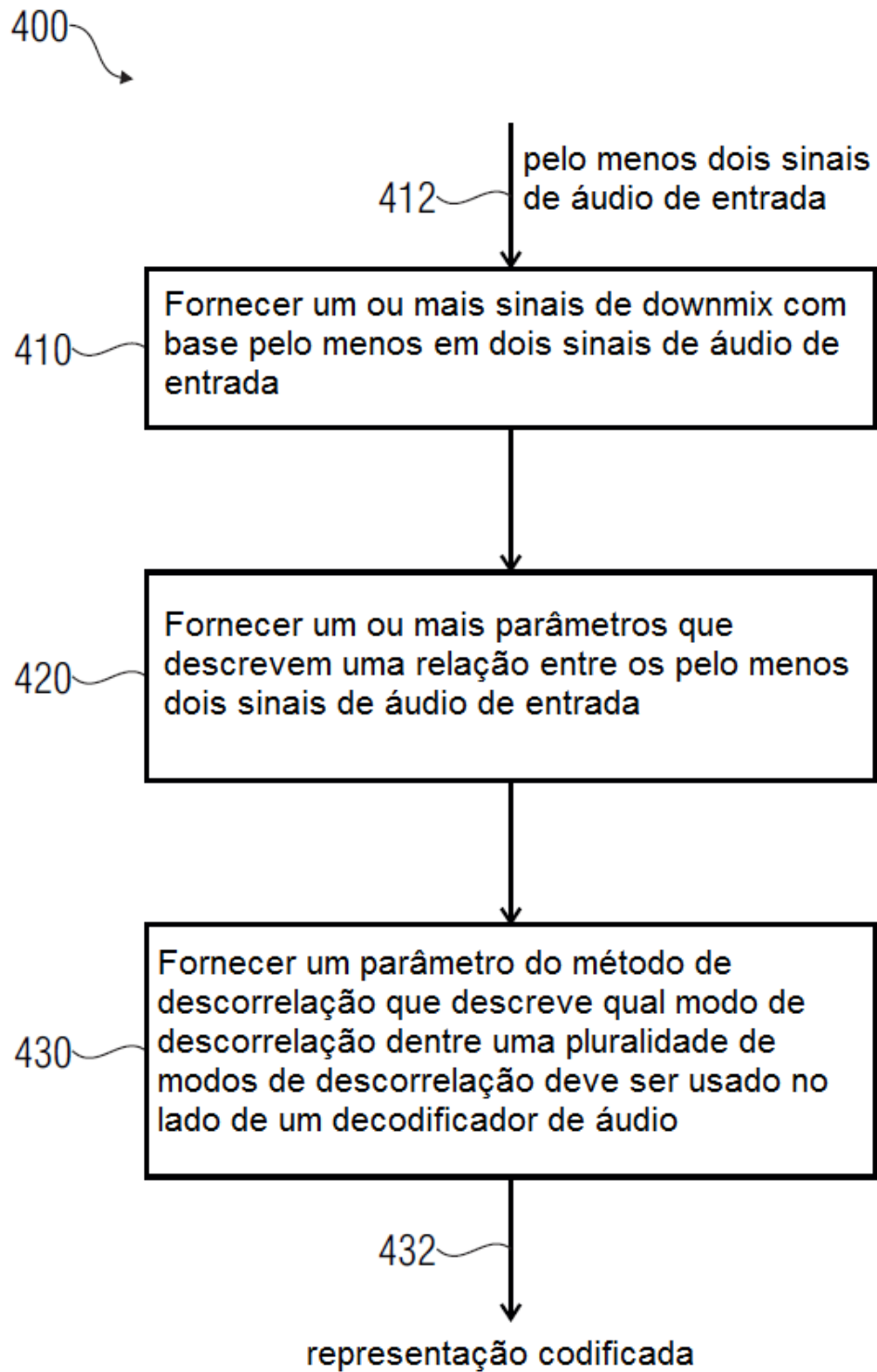


FIG 4

500

representação de áudio codificada

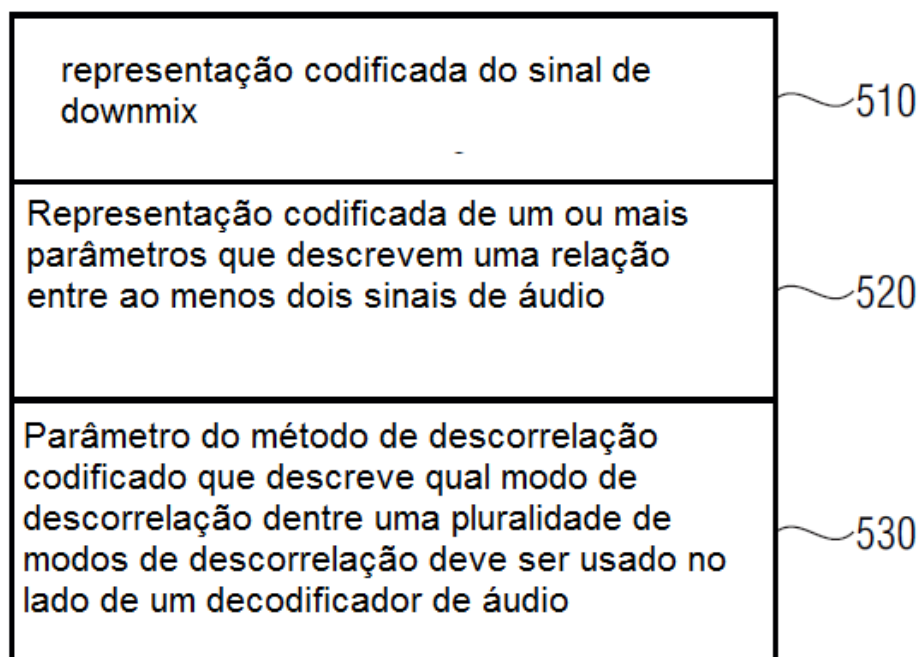


FIG 5

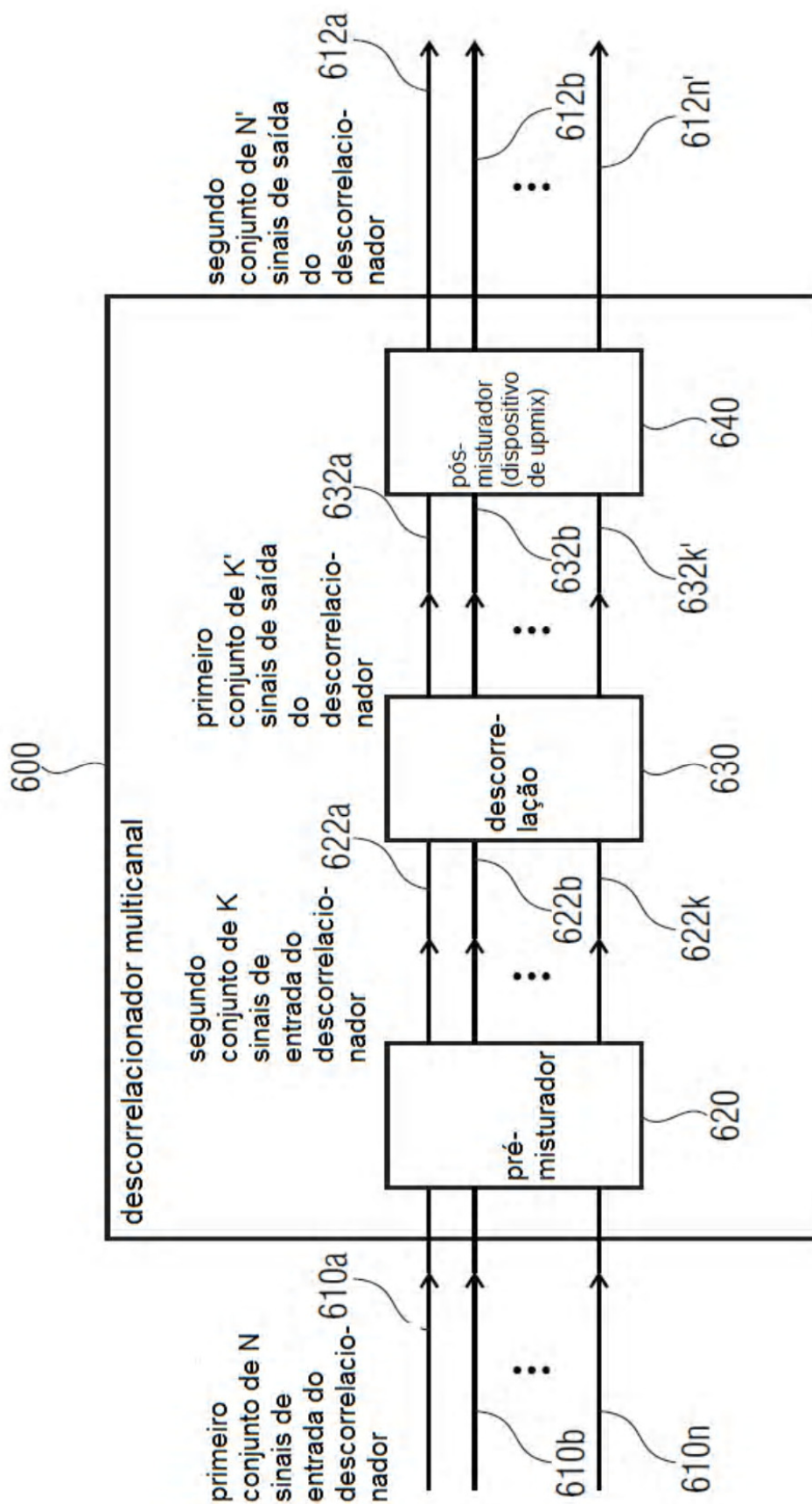


FIG 6

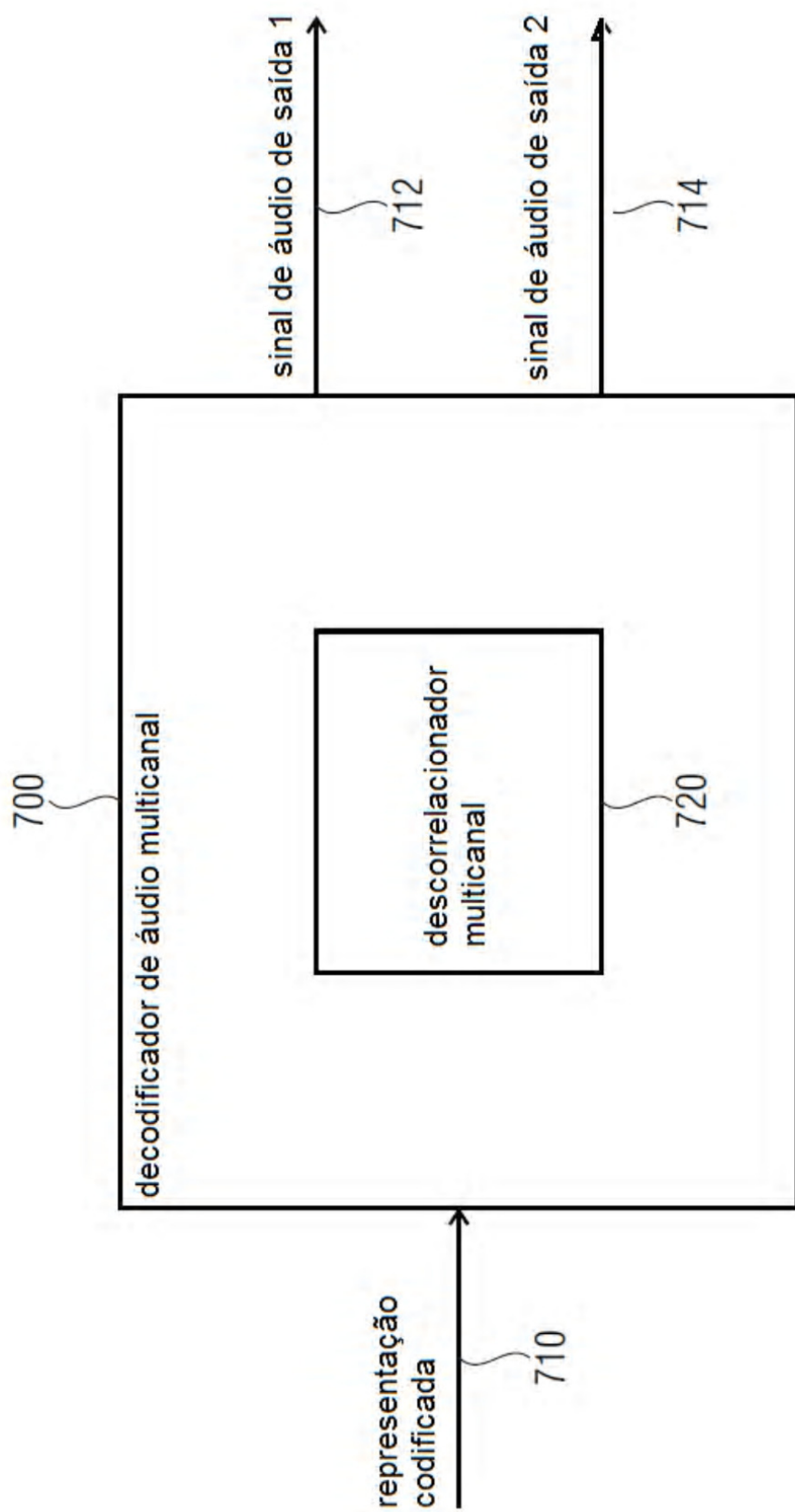


FIG 7

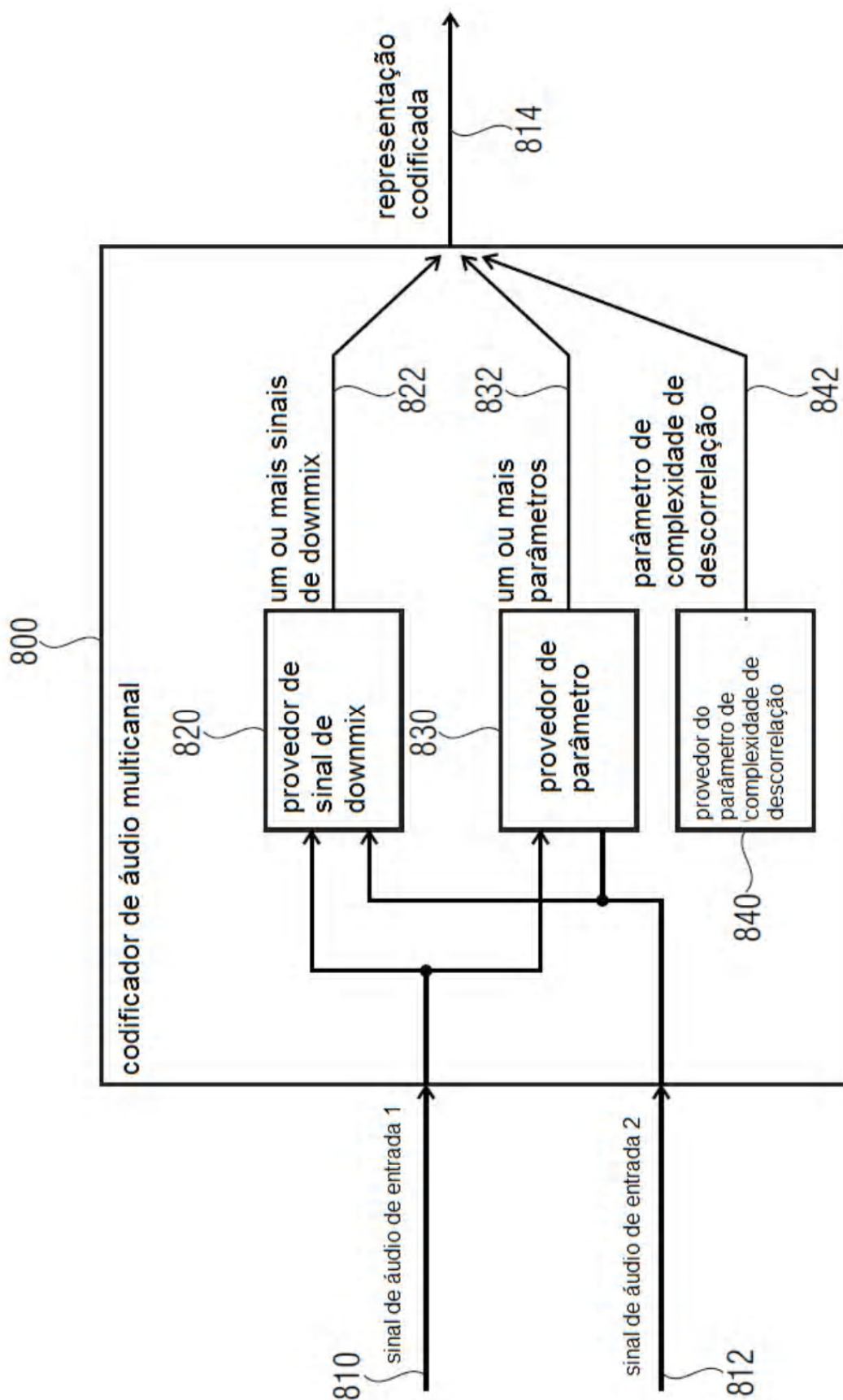


FIG 8

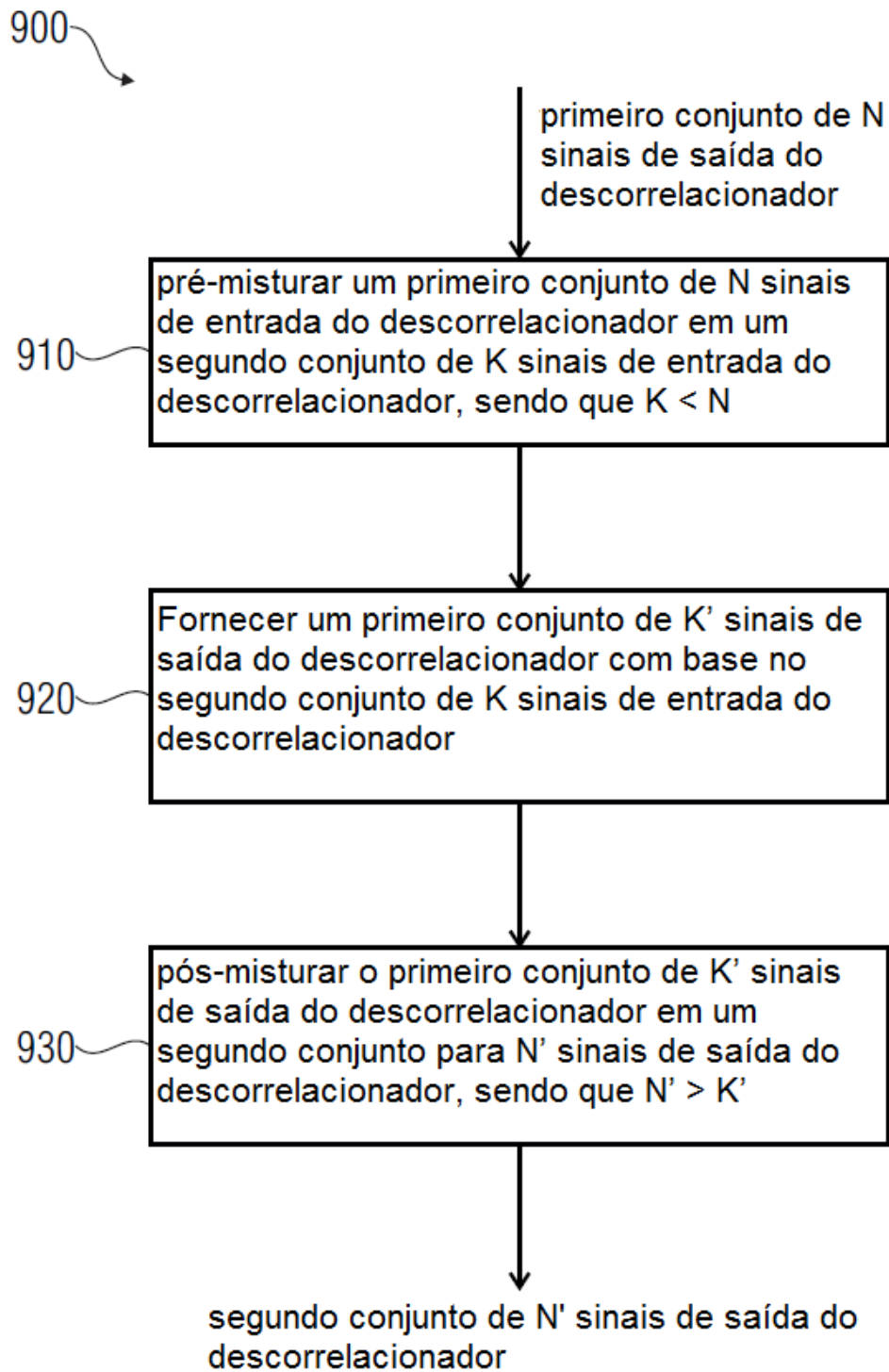


FIG 9

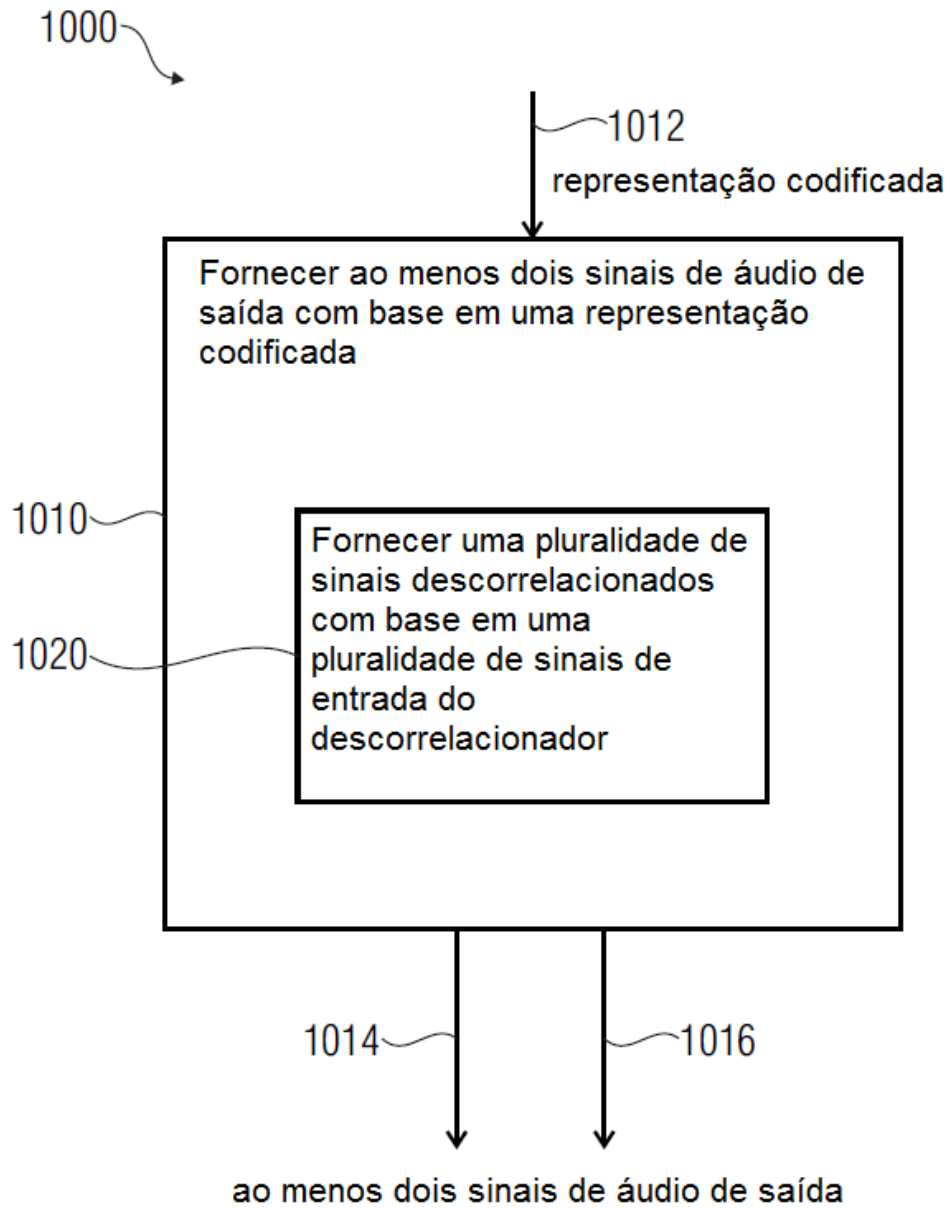


FIG 10

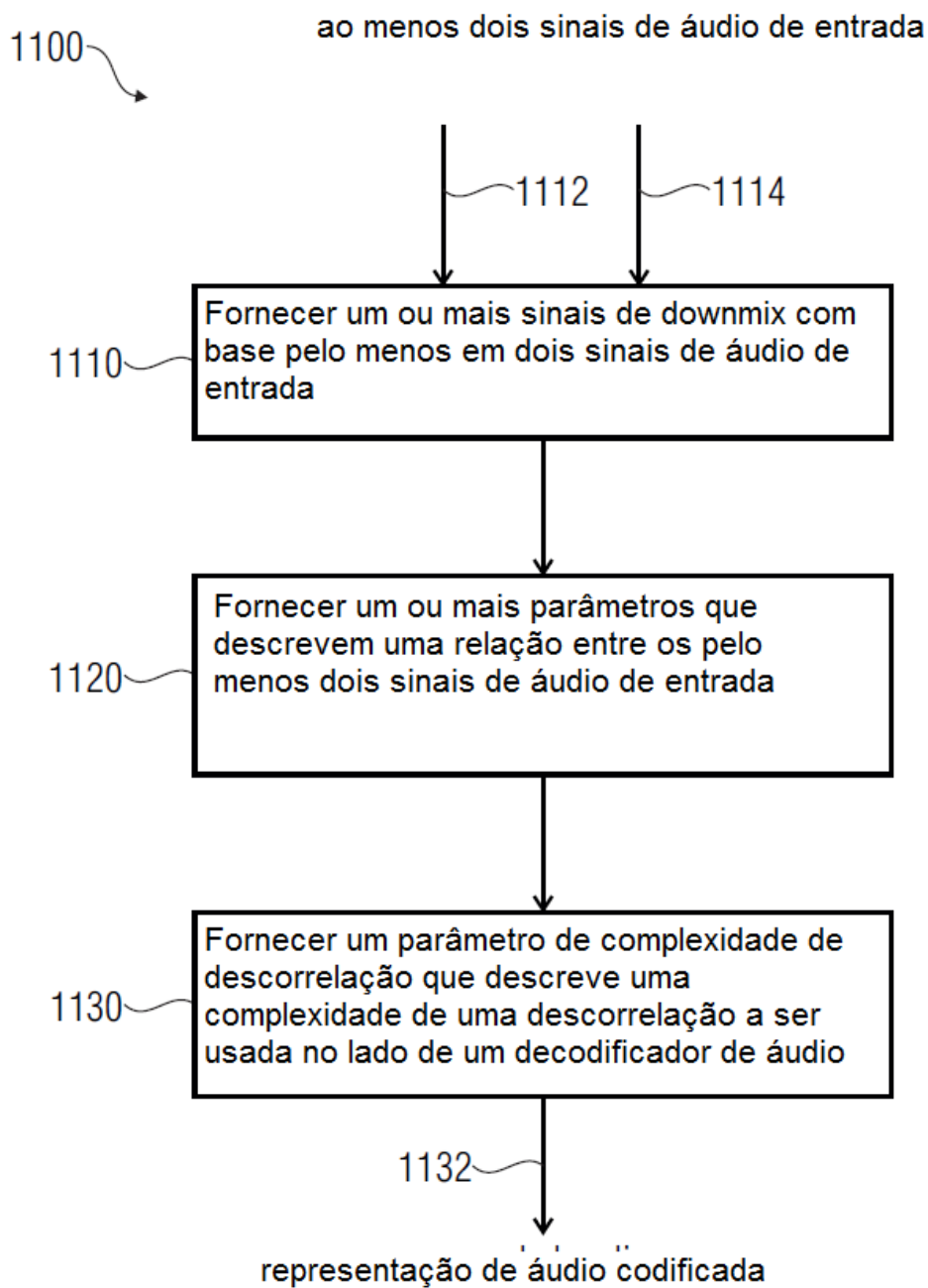


FIG 11

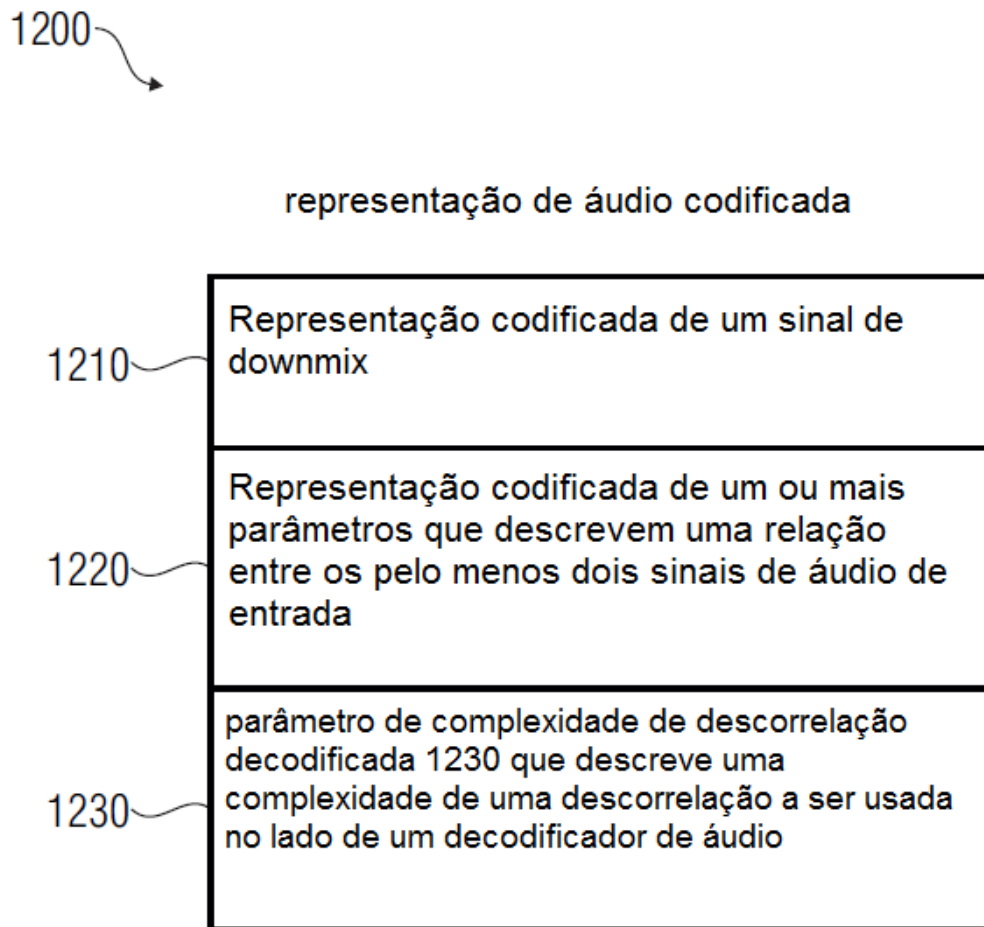


FIG 12

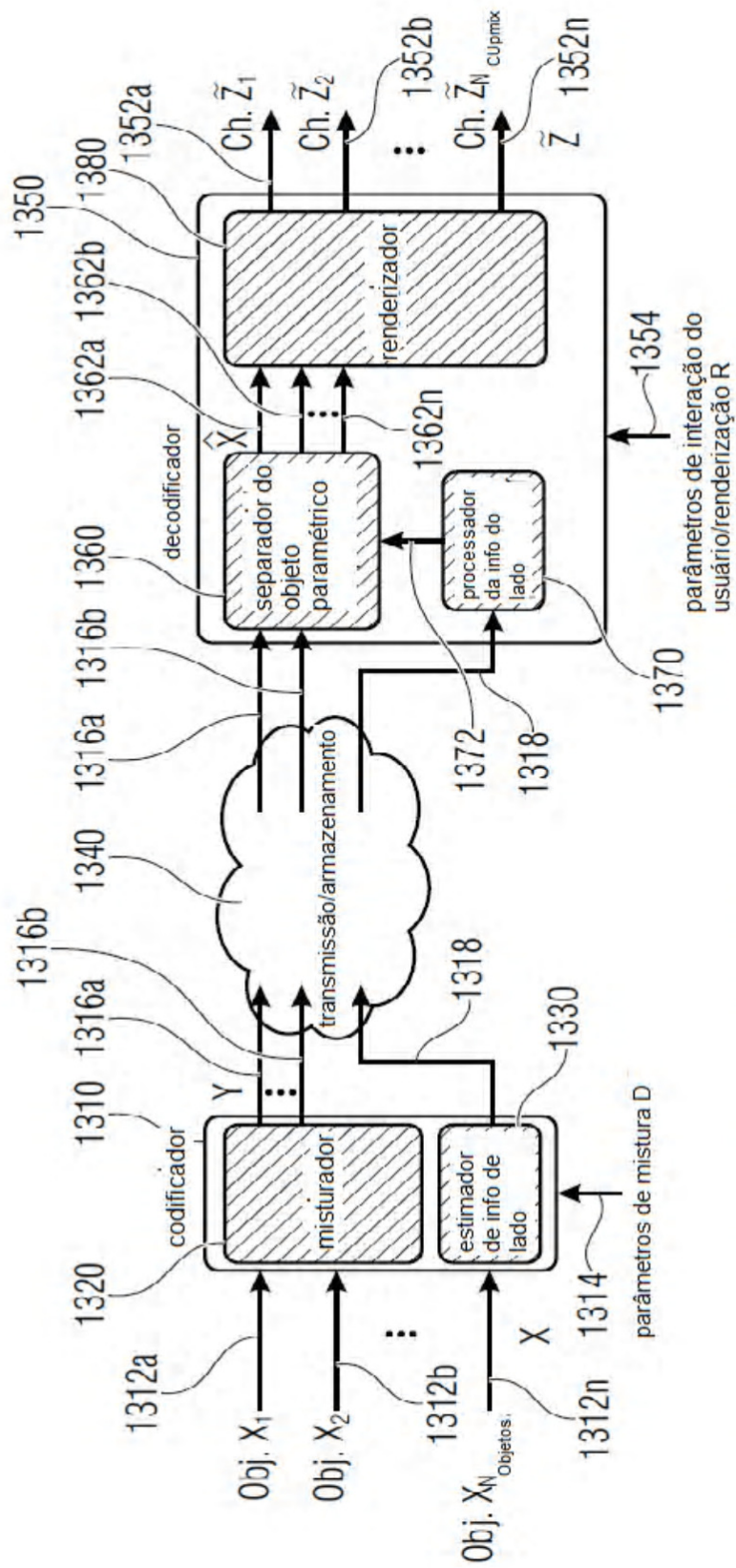
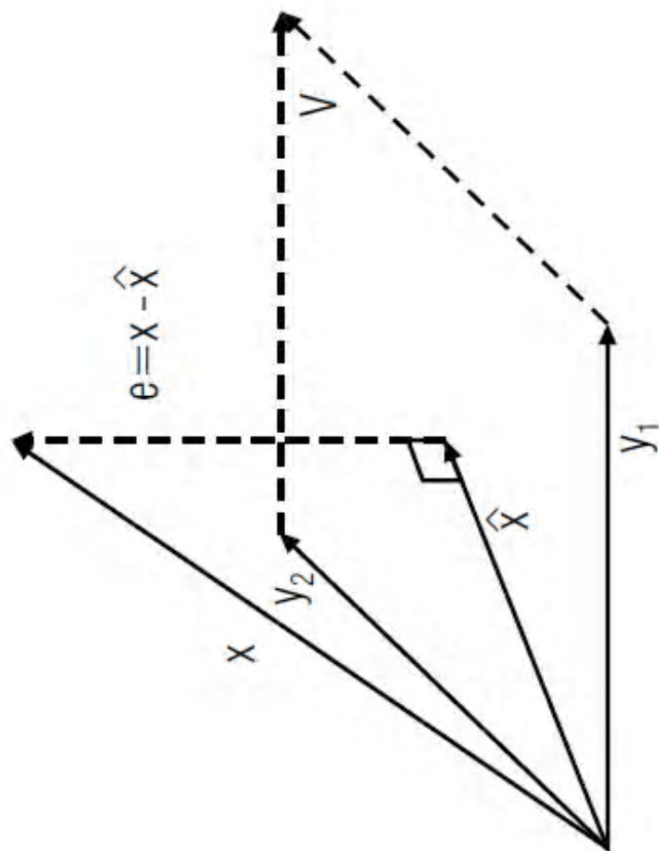


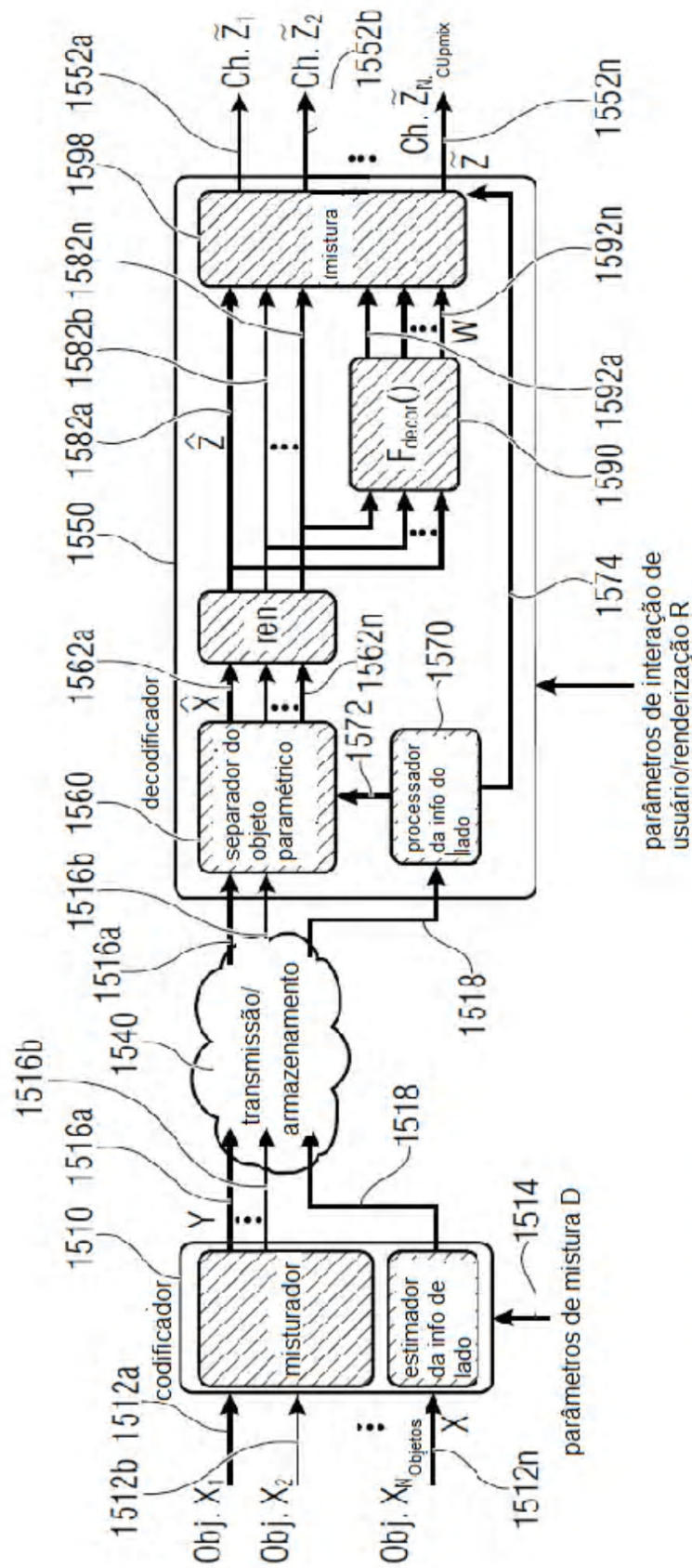
FIG 13

VISÃO GERAL DE UM CONCEITO DE DOWNMIX/UPMIX PARAMÉTRICO COM BASE EM MMSE



Representação geométrica para o princípio de ortogonalidade no espaço tridimensional

FIG 14



SISTEMA DE RECONSTRUÇÃO PARAMÉTRICA COM DESCORRELAÇÃO
APLICADA NA SAÍDA RENDERIZADA

FIG 15

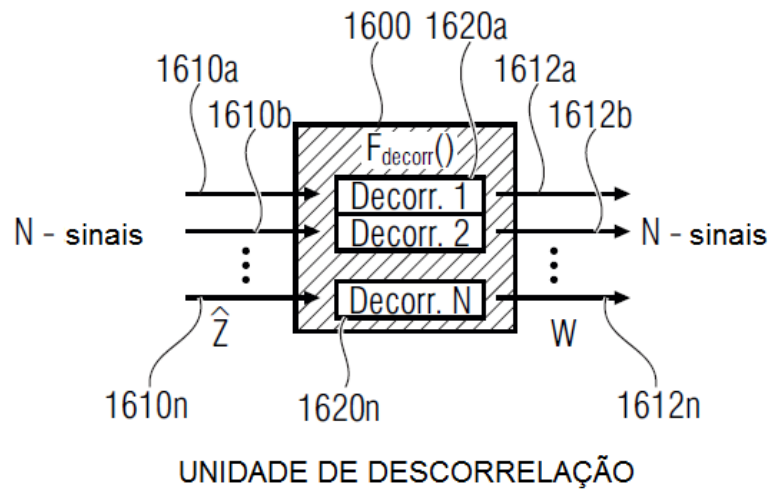


FIG 16

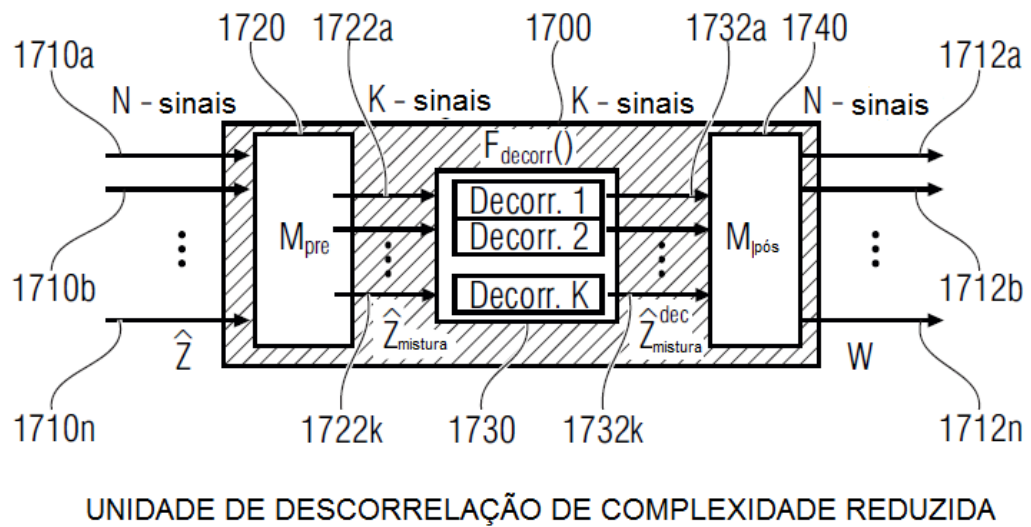


FIG 17

Posições do alto-falante e formatos de saída

1800

Nº	Rótulo LS	Az. °	Az. Tol. °	El. °	El. Tol. °	formatos de saída					
						0-2.0	0-5.1	0-7.1	0-8.1	0-10.1	0-22.2
1	CH_M_000	0	±2	0	±2		3	3		3	3
2	CH_M_L030	30	±2	0	±2	1	1	1	1	1	7
3	CH_M_R030	-30	±2	0	±2	2	2	2	2	2	8
4	CH_M_L060	60	±2	0	±2						1
5	CH_M_R060	-60	±2	0	±2						2
6	CH_M_L090	90	±5	0	±2						11
7	CH_M_R090	-90	±5	0	±2						12
8	CH_M_L110	110	±5	0	±2		5	5	5	5	
9	CH_M_R110	-110	±5	0	±2		6	6	6	6	
10	CH_M_L135	135	±5	0	±2						5
11	CH_M_R135	-135	±5	0	±2						6
12	CH_M_180	180	±5	0	±2						9
13	CH_U_000	0	±2	35	±10				3		15
14	CH_U_L045	45	±5	35	±10						13
15	CH_U_R045	-45	±5	35	±10						14
16	CH_U_L030	30	±5	35	±10			7	7	7	
17	CH_U_R030	-30	±5	35	±10			8	8	8	
18	CH_U_L090	90	±5	35	±10						19
19	CH_U_R090	-90	±5	35	±10						20
20	CH_U_L110	110	±5	35	±10					9	
21	CH_U_R110	-110	±5	35	±10					10	
22	CH_U_L135	135	±5	35	±10						17
23	CH_U_R135	-135	±5	35	±10						18
24	CH_U_180	180	±5	35	±10						21
25	CH_T_000	0	±2	90	±10					11	16
26	CH_L_000	0	±2	-15	+5-25				9		22
27	CH_L_L045	45	±5	-15	+5-25						23
28	CH_L_R045	-45	±5	-15	+5-25						24
29	CH_LFE1	45	±15	-15	±15		4	4	4	4	4
30	CH_LFE2	-45	±15	-15	±15						10

FIG 18

Redução de complexidade para o formato de saída 22.2

- coeficientes de pré-mistura para $N=22$ e $K=$, $\text{cond}(M_{pe}M_{gr}^H)=1$.

Ch. ID	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
M_{gr}^j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

FIG 19A

1910

$\text{cond}(M_{pre}^H M_{pre}^H) = 2.$

Coeficientes de pré-mistura para $N=22$ e $K=10$,

ID. C	M_{pre}^{il}	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

FIG 19B

- Coeficientes de pré-mistura para $N=22$ e $K=9$, $\text{Cond}(M_{pr3}^H M_{pre}^H) = 2$.

ID _C	M_{pre}^H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

FIG 19C

-Coeficientes de pré-mistura para N=22 e K=8
 $\text{cond}(M_{\text{pre}}^H M_{\text{pre}}^H) = 2$.

ID	M_{pre}^H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19D

- Coeficientes de pré-mistura para $N=22$ e $K=7$, $\text{cond}(M_{\text{pre}}^H M_{\text{pre}}^H) = 2$.

ID _C	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
M_{pre}^i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19E

- Coeficientes de pré-mistura para $N=22$ e $K=6$, $\text{cond}(M_{\text{pre}}^H) = 2$.

ID. C	M_{pre}^j	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19F

- Coeficientes de pré-mistura para N=22 e K=5
 $\text{cond}(M_{\text{pre}}^H) = 1.5.$

ID _C	CH_M_000	CH_L_000	CH_U_000	CH_T_000	CH_M_L135	CH_U_L135	CH_M_R135	CH_U_R135	CH_M_180	CH_U_180	CH_M_L030	CH_L_L045	CH_M_R030	CH_L_R045	CH_M_L090	CH_U_L090	CH_M_R090	CH_U_R090	CH_M_L060	CH_U_L045	CH_M_R060	CH_U_R045
M_{j-e}^l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

FIG 19G

Redução de complexidade para o formato 10.1

- Coeficientes de pré-mistura para N=10 e K=5, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

ID. C	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	0	0	1	1	0	0
5	0	0	0	0	0	0	0	0	1	1

FIG 20A

- Coeficientes de pré-mistura para N=10 e K=4, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

ID. C	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_T_000	CH_M_L110	CH_U_L110	CH_M_R110	CH_U_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	0	0	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0	0
3	0	0	0	0	1	1	0	0	0	0
4	0	0	0	0	0	0	1	1	1	1

FIG 20B

- Coeficientes de pré-mistura para $N=10$ e $K=3$ $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

C ID.	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH T 000	CH M L110	CH U L110	CH M R110	CH U R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	0	0	0	0	0	0
2	0	0	0	0	1	1	0	0	0	0
3	0	0	0	0	0	0	1	1	1	1

FIG 20C

- Coeficientes de pré-mistura para $N=10$ e $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1.5$.

C ID.	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH T 000	CH M L110	CH U L110	CH M R110	CH U R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	0	0	0	0
2	0	0	0	0	0	0	1	1	1	1

FIG 20D

Redução de complexidade para o formato de saída 8.1

- Coeficientes de pré-mistura para N=8 e K=4, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

ID. C	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH T 000	CH M L110	CH M R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8
1	1	1	0	0	0	0	0	0
2	0	0	1	1	0	0	0	0
3	0	0	0	0	1	1	0	0
4	0	0	0	0	0	0	1	1

FIG 21A

- Coeficientes de pré-mistura para N=8 e K=3, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

ID. C	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH T 000	CH M L110	CH M R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8
1	1	1	1	1	0	0	0	0
2	0	0	0	0	1	1	0	0
3	0	0	0	0	0	0	1	1

FIG 21B

- Coeficientes de pré-mistura para $N=8$ e $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 3$.

ID. C	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH T 000	CH M L110	CH M R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	0	0
3	0	0	0	0	0	0	1	1

FIG 21C

Redução de Complexidade para o formato de saída 7.1

- Coeficientes de pré-mistura para $N=7$ e $K=4$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1$.

ID. C	CH M L030	CH U L030	CH M R030	CH U R030	CH M 000	CH M L110	CH M R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	0	0	0	0	0
2	0	0	1	1	0	0	0
3	0	0	0	0	1	0	0
4	0	0	0	0	0	1	1

FIG 21D

- Coeficientes de pré-mistura para $N=7$ e $K=3$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

$i \backslash j$ C	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	1	1	0	0	0
2	0	0	0	0	1	0	0
3	0	0	0	0	0	1	1

FIG 21E

Coeficientes de pré-mistura para $N=7$ e $K=2$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2.5$.

$i \backslash j$ C	CH_M_L030	CH_U_L030	CH_M_R030	CH_U_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5	6	7
1	1	1	1	1	1	0	0
2	0	0	0	0	0	1	1

FIG 21F

Redução de complexidade para o formato de saída 5.1

- Coeficientes de pré-mistura para N=5 e K=3, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

ID. C	CH_M_L030	CH_M_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1

FIG 22A

- Coeficientes de pré-mistura para N=5 e K=2, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 1.5$.

ID. C	CH_M_L030	CH_M_R030	CH_M_000	CH_M_L110	CH_M_R110
$M_{\text{pre}}^{i,j}$	1	2	3	4	5
1	1	1	1	0	0
2	0	0	0	1	1

FIG 22B

Redução de complexidade para o formato de saída 2.0

- Coeficientes de pré-mistura para $N=2$ e $K=1$, $\text{cond}(M_{\text{pre}} M_{\text{pre}}^H) = 2$.

ID. C	CH_M_L030	CH_M_R030
$M_{\text{pre}}^{i,j}$	1	2
1	1	1

FIG 23

2410	Grupo 1	CH_M_000 (Ch. ID. 3)	CH_L_000 (Ch. ID. 22)	CH_U_000 (Ch. ID. 15)	CH_T_000 (Ch. ID. 16)		
2412	Grupo 2	CH_M_L135 (Ch. ID. 5)	CH_U_L135 (Ch. ID. 17)	CH_M_R135 (Ch. ID. 6)	CH_U_R135 (Ch. ID. 18)	CH_M_180 (Ch. ID. 9)	CH_U_180 (Ch. ID. 21)
2414	Grupo 3	CH_M_L030 (Ch. ID. 7)	CH_L_L045 (Ch. ID. 23)	CH_M_R030 (Ch. ID. 8)	CH_L_R045 (Ch. ID. 24)		
2416	Grupo 4	CH_M_L090 (Ch. ID. 11)	CH_U_L090 (Ch. ID. 19)	CH_M_R090 (Ch. ID. 12)	CH_U_R090 (Ch. ID. 20)		
2418	Grupo 5	CH_M_L060 (Ch. ID. 1)	CH_U_L045 (Ch. ID. 13)	CH_M_R060 (Ch. ID. 2)	CH_U_R045 (Ch. ID. 14)		

FIG 24

bsDecorrelationMethod;	2	uimsbf
bsDecorrelationLevel;	2	uimsbf

FIG 25

bsDecorrelationMethod

Indica o modo de operação de descorrelação do decodificador de acordo com:

bsDecorrelationMethod

bsDecorrelationMethod	Significado
0	Modo de compensação de energia
1	Modo de ajuste de covariância limitada
2	Modo de ajuste de covariância geral
3	N/A

FIG 26

bsDecorrelationLevel

Define o nível de correlação de acordo com:

bsDecorrelationLevel

bsDecorrelationLevel	22.2	10.1	8.1	7.1	5.1	2.1
0	11	10	8	7	5	2
1	9	5	4	4	3	1
2	7	3	3	3	2	0
3	5	2	2	2	0	0

FIG 27

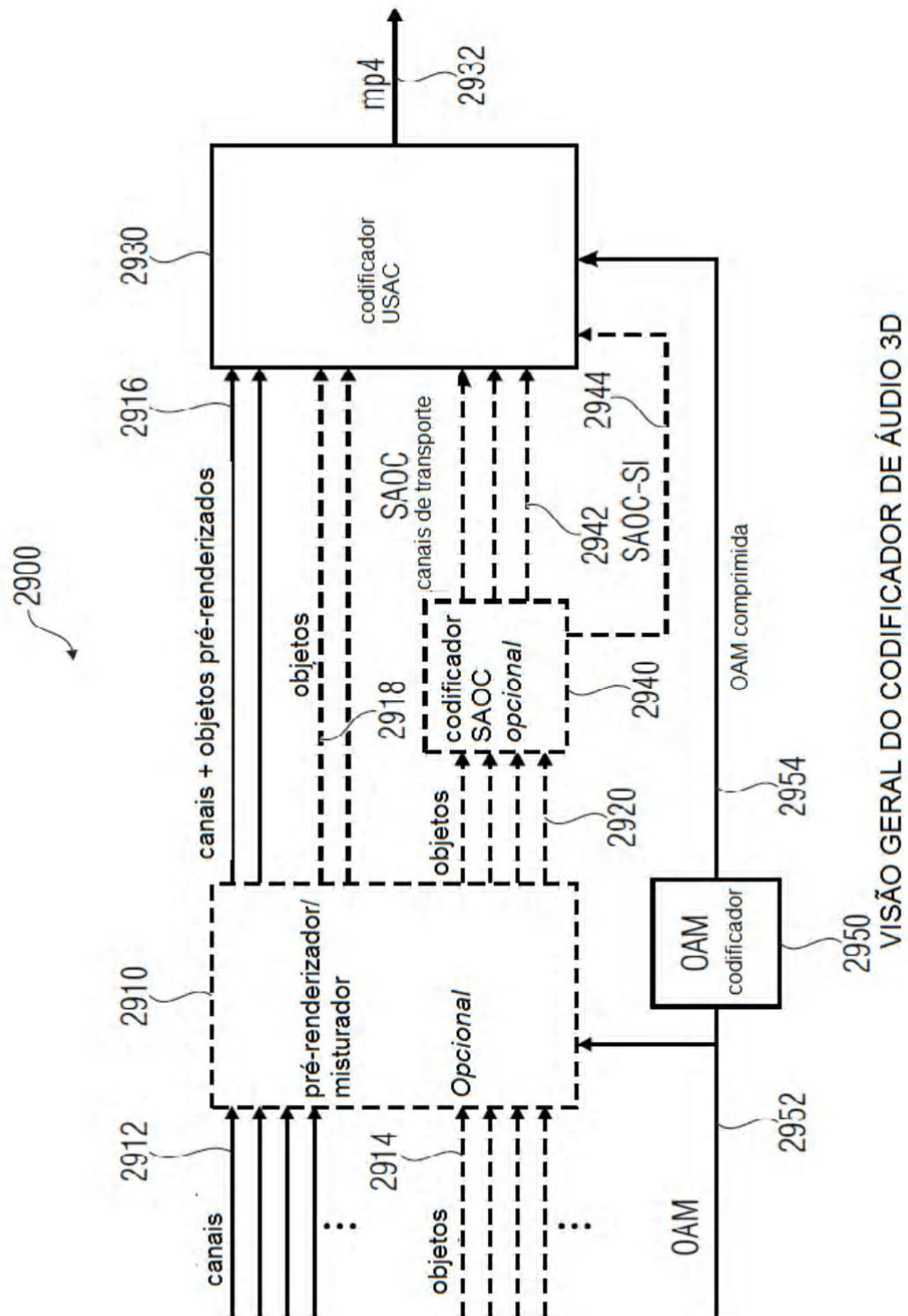
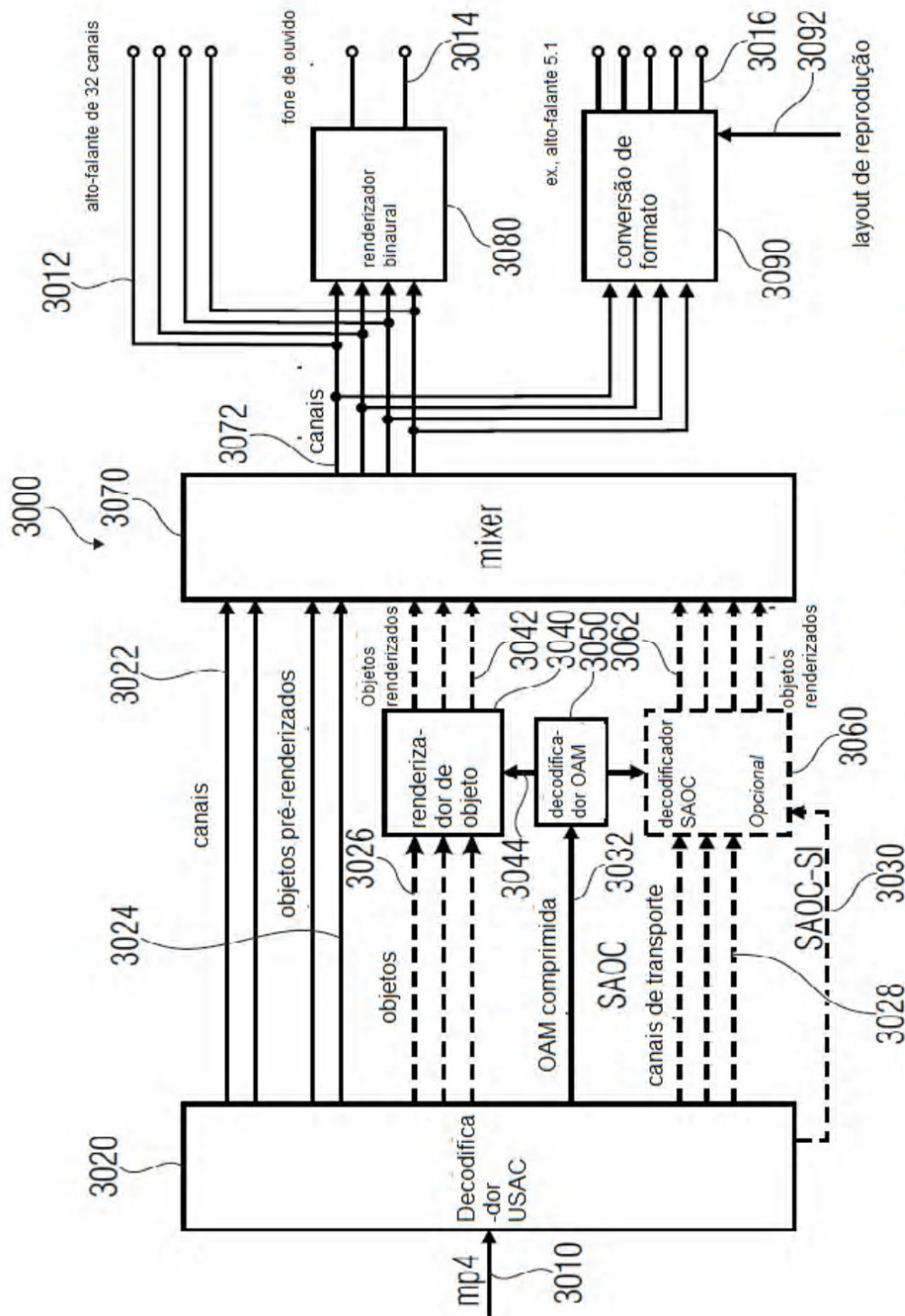
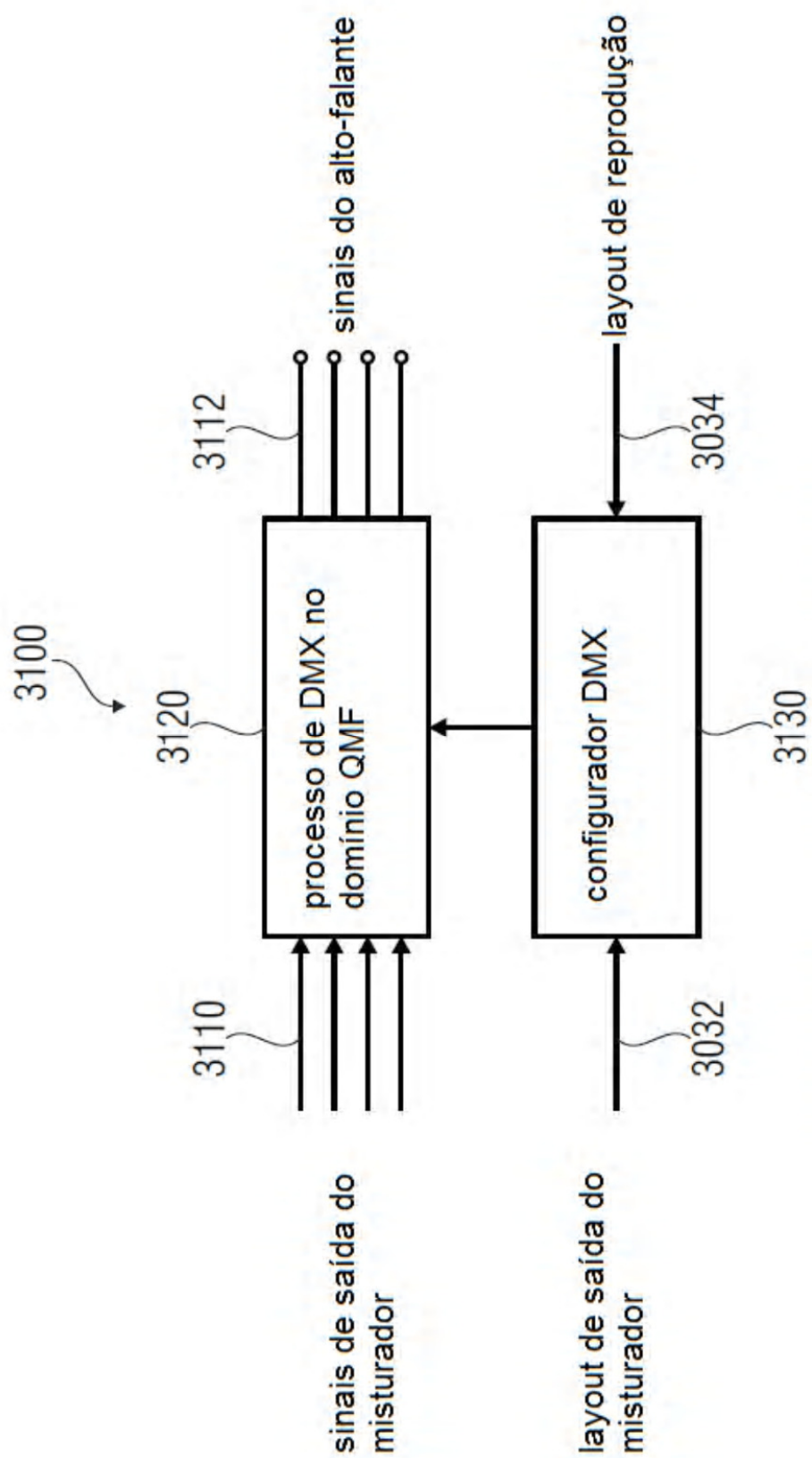


FIG 28



VISÃO GERAL DO DECODIFICADOR DE ÁUDIO 3D

FIG 29



ESTRUTURA DO CONVERSOR DE FORMATO

FIG 30

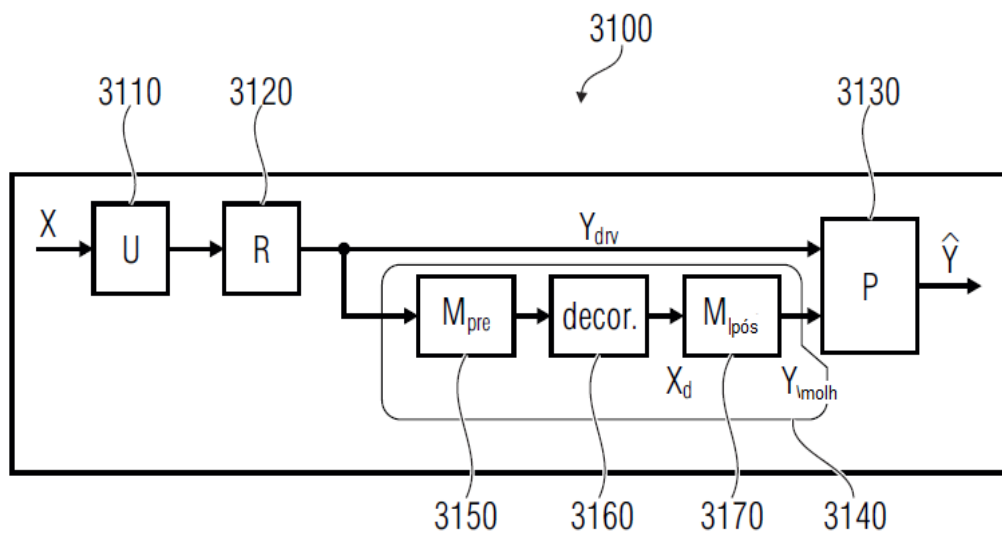


FIG 31

bsNumSaocDmxObjects	módo de decodificação	significado
0	combinado	Os sinais baseados no canal de entrada e sinais baseados no objeto de entrada são submisturados em N_c canais.
≥ 1	independente	Os sinais baseados do canal de entrada são submisturados em N_c canais. Os sinais baseados no objeto de entrada são submisturados em N_{obj} canais.

FIG 32

- Syntaxe de SAOC3DSpecificConfig()

[illegible]

FIG 33A	FIG 33A-1
	FIG 33A-2

FIG 33A-1

```

para ( i=bsNumSaocChannels; i<NumInputSignals; i++ ) {
    para ( j=0; j<bsNumSaocChannels; j++ ) {
        bsRelatedTo[i][j] = 0;
        bsRelatedTo[j][i] = 0;
    }
}
para ( i=bsNumSaocChannels; i<NumInputSignals; i++ ) {
    bsRelatedTo[i][i] = 1;
    para ( j=i+1; j<NumInputSignals; j++ ) {
        bsRelatedTo[i][j];
        bsRelatedTo[j][i] = bsRelatedTo[i][j];
    }
}
bsOneLOC;
bsSaocDmxMethod;
se (bsSaocDmxMethod == 15) {
    bsNumPremixedChannels;
}
bsDualMode;
se (bsDualMode) {
    bsBandsLow;
    bsBandsHigh = numBands;
}também{
    bsBandsLow = numBands;

```

	1	uimsbf
	1	uimsbf
	4	uimsbf
	5	uimsbf
	1	uimsbf
	5	uimsbf

Observação 1

FIG 33A	FIG 33A-1
	FIG 33A-2

FIG 33A-2

}		
bsDcuFlag;	1	uimsbf
se (bsDcuFlag == 1) {		
bsDcuMandatory;	1	uimsbf
bsDcuDynamic;	1	uimsbf
se (bsDcuDynamic == 0) {		
bsDcuMode;	1	uimsbf
bsDcuParam;	4	uimsbf
}		
} ele {		
bsDcuMandatory = 0;		
bsDcuDynamic = 0;		
bsDcuMode = 0;		
bsDcuParam = 0;		
}		
bsSaocReserved;	3	uimsbf
ByteAlign();		
SAOC3DExtensionConfig();		
}		
Observação 1: numBands é definido na Tabela 33 em ISO/IEC 23003-2:2010		

FIG 33B