



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

(11) PI 0924007-1 B1



(22) Data do Depósito: 14/05/2009

(45) Data de Concessão: 03/08/2021

(54) Título: UNIDADE DE SUPRESSÃO DE ECO ACÚSTICO E CONFERÊNCIA DE FRONT-END

(51) Int.Cl.: H04M 9/08; H04R 3/02.

(52) CPC: H04M 9/082; H04R 3/02.

(30) Prioridade Unionista: 15/04/2009 US 61/169,436.

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

(72) Inventor(es): FABIAN KÜCH; OLIVER MOSER; MARKUS KALLINGER; MARKUS SCHMIDT; MERAY ZOURUB; MARCO DIATSCHUK.

(86) Pedido PCT: PCT EP2009003446 de 14/05/2009

(87) Publicação PCT: WO 2010/118763 de 21/10/2010

(85) Data do Início da Fase Nacional: 17/10/2011

(57) Resumo: UNIDADE DE SUPRESSÃO DE ECO ACÚSTICO E CONFERÊNCIA DE FRONT-END Uma unidade de supressão do eco acústico (210) de acordo com uma realização da presente invenção compreendendo uma interface de entrada (230) para extrair um sinal downmix (310) de um sinal de entrada (300) o sinal de entrada compreendendo o sinal downmix (310) e a informação lateral paramétrica (320), em que o downmix e a informação lateral paramétrica juntamente representam um sinal multicanal, uma calculadora (220) para calcular os coeficientes de filtro para um filtro adaptativo (240), em que a calculadora (220) é adaptada para determinar os coeficientes de filtro com base no sinal downmix (310) e no sinal de microfone (340) ou no sinal derivado de um sinal de microfone, e um filtro adaptativo (240) adaptado para filtrar o sinal de microfone (340) ou o sinal derivado do sinal de microfone com base nos coeficientes de filtro para suprimir um eco causado pelo sinal multicanal no sinal de microfone (340) .

**"UNIDADE DE SUPRESSÃO DE ECO ACÚSTICO E
CONFERÊNCIA DE FRONT-END"**

DESCRIÇÃO

Realizações de acordo com a presente invenção referem-se a uma unidade de supressão de eco e um método para suprimir um eco acústico, que pode ser usado, por exemplo, em sistemas de telecomunicação *hands-free* ou outros sistemas acústicos que incluem *pláystack* de alto-falante de multicanal com base em uma representação paramétrica de som espacial.

Ecos acústicos aumentos a partir de um acoplamento acústico ou *feedback* entre os alto-falantes e microfones de dispositivos de telecomunicação. Este fenômeno está especialmente presente em operações *hands-free*. O sinal de *feedback* acústico de um alto-falante é transmitido para o subscritor do *far-end*, que percebe um atraso na sua própria fala. Sinais de eco representam um distúrbio bastante perturbador que pode até mesmo inibir a interação na comunicação entre uma dupla. Adicionalmente, ecos acústicos podem resultar em efeitos uivantes e instabilidade na curva do *feedback* acústico. Em um sistema de comunicação duplo *hands-free*, o controle de eco é deste modo aconselhável a fim de cancelar o acoplamento entre os alto-falantes e os microfones.

A Fig. 9 ilustra o problema de controle de eco na acústica em geral. O sinal *far-end*, emitido por um alto-falante, vai diretamente para o microfone, e através de caminhos refletidos. Assim, o microfone não somente captura a fala do *near-end* local, mas também o eco que é então reproduzido para o usuário no *far-end*.

Um sinal de alto-falante $x(n)$ é fornecido para um alto-falante 100, que transforma o sinal do alto-falante em uma oscilação audível do meio *surrounding* do alto-falante 100. Conforme indicado pela Fig. 9, o microfone 110 pode receber o som emitido pelo alto-falante 100, que é indicado na Fig. 9 por um vetor curvado, em que $y(n)$ denota um sinal de *feedback* do alto-falante 100 para o microfone 110.

Além do sinal de *feedback* $y(n)$, o microfone 110 também grava um sinal sonoro adicional $w(n)$, que pode, por exemplo, representar a fala para um usuário. Ambos os sinais acústicos são gravados pelo microfone 110 e fornecidos, como um sinal de microfone $z(n)$, para uma unidade de remoção de eco 120. A unidade de remoção do eco 120 também recebe o sinal $x(n)$ do alto-falante. Este externa um sinal no qual - idealmente - a contribuição do sinal $x(n)$ do alto-falante é removida do sinal gravado ou o sinal $z(n)$ do microfone.

A partir daí, a Fig. 9 ilustra o início geral do problema do controle acústico do eco. O sinal $x(n)$ do alto-falante é devolvido para o sinal $z(n)$ do microfone. Um processo de remoção de eco remove este eco enquanto - idealmente - deixa passar o sinal $w(n)$ do *near-end* para o local desejado.

O controle do eco acústico é um problema bem conhecido e vários métodos para remover os ecos acústicos têm sido propostos [13]. Abaixo, lembramos brevemente as abordagens para a supressão do eco acústico (AES) como, por exemplo, apresentados em [8, 9], como sendo os mais adequados no contexto considerado de comunicação espacial de áudio.

Quando transmitindo ou retransmitindo sinais de

áudio, os sistemas de multicanais são comumente utilizados. Nestes sistemas alto-falantes múltiplos são usados para retransmitir o som e/ou microfones múltiplos são usados para gravar o som espacial. Tais sistemas de multicanais são, por exemplo, usados em sistemas de teleconferência espacial de áudio que não somente transmitem os sinais de áudio de diferentes partes, como também preserva as informações espaciais do cenário de gravação [12]. Em outros sistemas, as informações espaciais podem ser fornecidas artificialmente ou mudadas interativamente [5].

No caso de o áudio espacial ser aplicado em cenários de telecomunicação, uma representação eficiente dos sinais de áudio multicanal deveriam ser usados, enquanto ainda assegura alta qualidade de áudio. A codificação do áudio espacial paramétrico representa uma abordagem adequada para solucionar este desafio. Abaixo, apresentamos métodos práticos que seguem o paradigma da codificação do áudio espacial paramétrico e são especialmente importantes no contexto da comunicação.

Enquanto sistemas de multicanais como, por exemplo, a codificação de áudio espacial previamente mencionada fornece a oportunidade de transmitir uma pluralidade de sinais de áudio de uma maneira muito eficiente e economizando banda larga, uma implementação simples de um removedor de eco ou processo de supressão do eco em tais sistemas de multicanais necessita uma aplicação para cada um e todos os sinais de microfone com base em cada e em todos os sinais de alto-falante conforme externados pelo sistema de multicanais. Isto, no entanto, pode representar um crescimento exponencial significativo acerca da complexidade

computacional simplesmente devido ao alto número de sinais de microfone e/ou de alto-falante a serem processados. Consequentemente, isto pode causar custos adicionais devido a um alto consumo de energia, a necessidade de uma processabilidade de dados mais alta e, eventualmente, também um pequeno aumento no atraso.

Assim, é um objetivo da presente invenção fornecer uma unidade de supressão do eco acústico e uma conferência de *front-end* que permita uma supressão do eco acústico mais eficiente.

Este objetivo é alcançado por uma unidade de supressão do eco acústico de acordo com a reivindicação 1, um método para suprimir uma acústica de acordo com a reivindicação 8, uma conferência de *front-end* de acordo com a reivindicação 10, um método para fornecer sinais de alto-falante e sinal de microfone de acordo com a reivindicação 14, ou um programa de computador de acordo com a reivindicação 15.

Realizações, de acordo com a presente invenção, têm base na descoberta de que uma supressão de eco acústico mais eficiente é realizável através da extração de um sinal *downmixing* a partir de um sinal de entrada compreendendo o sinal *downmixing* e uma informação lateral paramétrica, que calcula coeficiente de filtro para um filtro adaptativo com base no sinal *downmixing* e um sinal de microfone ou um sinal derivado de um sinal de microfone, e filtrar o sinal de microfone ou sinal derivado a partir do sinal de microfone com base nos coeficientes de filtro calculado. Em outras palavras, no caso de um sistema de multicanais com base em um sinal *downmix* e informações laterais paramétricas formando um

sinal de entrada, em que as informações laterais paramétricas juntamente ao sinal *downmix* representam um sistema de multicanais, a supressão de eco pode ser feita com base no sinal *downmix*.

Enquanto empregando uma realização de acordo com a presente invenção, pode ser, assim, possível evitar a decodificação do sinal de entrada no sinal multicanal e após desempenhar a supressão do eco acústico. Assim pode ser possível reduzir significativamente a complexidade computacional considerando que o número de sinais é drasticamente reduzido em comparação com o sistema de multicanais como previamente descrito. Através do emprego da realização de acordo com a presente invenção é possível desempenhar a supressão do eco acústico na base do sinal *downmix* compreendido no sinal de entrada.

Em outras realizações de acordo com a presente invenção, a supressão do eco pode ser desempenhada com base na referência de energia de espectros, que pode ser determinada com base no sinal *downmix* recebido e no sinal de microfone ou sinal derivado do sinal de microfone. Opcionalmente, a referência de energia de espectros derivada do sinal multicanal pode ser atrasada por um valor de atraso, que pode, por exemplo, ser determinada com base em um valor de correlação.

Portanto, a conferência do *front-end* de acordo com uma realização da presente invenção não compreende somente uma unidade de supressão do eco acústico de acordo com uma realização da presente invenção como também um decodificador de multicanais e pelo menos uma unidade de microfone, em que o decodificador de multicanais é adaptado ao decodificador para decodificar o sinal *downmix* e a informação lateral de parâmetro em

uma pluralidade de sinais de alto-falante. A pelo menos uma unidade de microfone é adaptada ainda para fornecer um sinal de microfone para a unidade de supressão do eco acústico. Em outras realizações da conferência de *front-end* a interface de entrada é adaptada ainda para extrair a informação lateral paramétrica, em que o decodificador de multicanais compreende um *upmixer* e um processador de parâmetro. O processador de parâmetro é então adaptado para receber a informação lateral de parâmetro a partir de uma interface de entrada e para fornecer um sinal de controle de *upmix*. O *upmixer* é adaptado para receber o sinal *downmix* da interface de entrada e o sinal de controle de *upmix* a partir de um processador de parâmetro e é adaptado para fornecer uma pluralidade de sinais de alto-falante com base no sinal *downmix* e do sinal de controle de *upmix*. Por isso, em realizações de acordo com a presente invenção, a interface de entrada da unidade de supressão do eco acústico pode ser aquela do decodificador de multicanais, ou o decodificador de multicanais e a unidade de supressão acústica podem dividir uma interface de entrada em comum.

Além disso, realizações de acordo com a presente invenção podem opcionalmente compreender também um codificador multicanal correspondente adaptado para codificar uma pluralidade de sinais de entrada de áudio em outro sinal *downmix* e outra informação lateral paramétrica juntas representando uma pluralidade de sinais de entrada de áudio, em que o sinal de microfone da pelo menos uma unidade de microfone é um dos sinais de entrada de áudio. Neste caso, a unidade de supressão de eco acústico compreendida na conferência de *front-end* é adaptada para

receber o outro *downmix* como o derivado a partir do sinal do microfone.

Em outras palavras, como será apresentado abaixo, a abordagem de acordo com as realizações da presente invenção
5 permite combinar eficientemente a supressão do eco acústico e a codificação do áudio espacial paramétrico.

Realizações de acordo com a presente invenção serão descritas aqui fazendo referência aos desenhos em anexo.

A Fig. 1 apresenta um diagrama em bloco de uma
10 conferência de *front-end* compreendendo uma unidade de supressão acústica de acordo com uma realização da presente invenção;

A Fig. 2 ilustra uma estrutura geral de um codificador de áudio espacial paramétrico;

A Fig. 3 ilustra uma estrutura geral de um
15 decodificador de áudio paramétrico;

A Fig. 4 ilustra uma cadeia de processamento de sinal usada em decodificador *MPEG surround (MPS)*;

A Fig. 5 ilustra uma estrutura geral do codificador do código do objeto de áudio espacial (SAOC);

20 A Fig. 6a ilustra um transcodificador com base em *monodownmix* para transcodificar dados SAOC para dados MPS;

A Fig. 6b ilustra um transcodificador com base em estéreo *downmix* para transcodificar dados SAOC em dados MPS;

A Fig. 7 mostra uma conferência de *front-end* de
25 acordo com uma realização da presente invenção para ilustrar a abordagem eficiente proposta da supressão do eco acústico com base no *downmix* dos codificadores de áudio espacial paramétrico;

A Fig. 8 ilustra outra realização de acordo com a

presente invenção na forma de uma conferência de *front-end* compreendendo uma unidade de supressão de eco acústico para uma realização da presente invenção;

5 A Fig. 9 ilustra o início geral de um problema de controle do eco acústico.

Com referência nas Figs. 1 a 9 as seguintes realizações diferentes de acordo com a presente invenção e tecnologias subjacentes serão delineadas e descritas em mais detalhes. No entanto, antes de introduzir técnicas de supressão acústica para canal único de supressão do eco acústico, e para 10 gerar um mapa de profundidade 37 (por exemplo, campo de profundidade codificado por cor 72 mostrado na FIG. 7), em uma realização de acordo com a presente invenção, na forma de uma conferência de *front-end* junto a uma unidade de supressão do eco acústico será primeiramente descrita. 15

A Fig. 1 mostra um diagrama em bloco de uma conferência de *front-end* 200 compreendendo, como um componente central, uma unidade de supressão do eco acústico 210 de acordo com uma realização da presente invenção. A unidade de supressão do eco acústico 210 compreende uma calculadora 220, uma interface de 20 entrada 230 e um filtro adaptativo 240. A conferência de *front-end* 200 compreende ainda um decodificador de multicanais 250, que é ponderadamente acoplado a saída de uma pluralidade de alto-falantes 100, os quais são exemplarmente mostrados quatro alto-falantes 100-1,..., 100-4. A conferência de *front-end* compreende 25 ainda um microfone ou uma unidade de microfone 110.

Para ser um pouco mais específico, um sinal de entrada 300 é fornecido, o qual compreende um sinal *downmix* 310 e

uma informação lateral paramétrica 320. A interface de entrada 230 separa ou extrai do sinal de entrada na realização mostrada na Fig. 1, o sinal *downmix* 310 e a informação paramétrica lateral 320. Na realização mostrada na Fig. 1, a interface de entrada 230
5 fornece um sinal *downmix* 310 junto à informação lateral paramétrica 320 para um decodificador de multicanais 250.

O decodificador de multicanais 250 é adaptado para decodificar o sinal *downmix* 310 e a informação lateral paramétrica 320 em uma pluralidade de sinais de alto-falante 330,
10 do qual, somente para o bem da simplicidade, na Fig. 1 somente um é rotulado como tal.

Desde que os alto-falantes 100 são acoplados para saídas apropriadas do decodificador de multicanais 250, o alto-falante 100 recebe os sinais de alto-falante individual 330 e os
15 transforma novamente em sinais acústicos audíveis.

A calculadora 220 é ainda acoplada a uma saída da interface de entrada 230 na qual o canal *downmix* 310 está disponível. Consequentemente, a calculadora 220 é adaptada para receber o sinal *downmix* 310. No entanto, na realização mostrada na
20 Fig. 1, as informações laterais paramétricas 320 do sinal de entrada 300 não são fornecidas para a calculadora 220. Em outras palavras, em realizações de acordo com a presente invenção, a calculadora 220 pode ser usada no sinal *downmix* sozinho em termos dos sinais compreendidos no sinal de entrada.

25 O microfone 110 é ponderadamente acoplado a saída para ambos, à calculadora 220 e o filtro adaptativo 240. Como consequência, a calculadora 220 é também adaptada para receber um sinal de microfone 340 como fornecido pelo microfone 110. Com base

no sinal de microfone 340 e no sinal *downmix* 310, a calculadora 220 é adaptada para determinar coeficientes de filtro para o filtro adaptativo 240 e para fornecer um sinal coeficiente de filtro correspondente 350 para o filtro adaptativo 240 na base do qual o filtro adaptativo 240 filtra o sinal de microfone 340 que entra. O filtro adaptativo 240 fornece na sua saída um sinal de saída, que é uma versão suprimida de um eco do sinal de microfone 340.

Mais detalhes em relação ao modo de operação de uma possível implementação de uma calculadora 220 serão dados abaixo.

Apesar da interface de entrada 230 ser esquematicamente desenhada na Fig. 1 como um componente individual da unidade de supressão do eco acústico 210, a interface 230 pode também ser parte do decodificador 250 ou pode ser dividida pelo decodificador 250 e pela unidade de supressão do eco acústico. Ainda, é possível implementar de acordo com a presente invenção, por exemplo, através da implementação de uma interface de entrada 230 que é capaz de extrair o sinal *downmix* 310 sozinho. Neste caso, o sinal de entrada 300 seria fornecido pelo decodificador de multicanais 250, que por sua vez compreende uma interface apropriada capaz de extrair o sinal *downmix* 310 e as informações laterais paramétricas 320. Em outras palavras, pode ser possível implementar uma unidade de supressão de eco acústico 210 com uma interface de entrada 230 que não é capaz de extrair as informações laterais paramétricas, mas somente o sinal *downmix* 310.

Realizações, de acordo com a presente invenção, representam um método eficiente para a supressão de ecos acústicos

para sistemas de alto-falante multicanal usados em sistemas de comunicação de áudio espacial. O método é aplicável no caso de os sinais de áudio espacial serem representados por um sinal *downmix* e informações laterais paramétricas ou metadados correspondentes.

5 Estes parâmetros capturam a informação que é necessária para computar os sinais de alto-falante no lado de reprodução. A invenção explora o fato de que a supressão do eco pode ser desempenhada diretamente com base no sinal *downmix* recebido em vez de explicitamente computar os sinais de alto-falante antes que
10 eles entrem na supressão acústica do eco (AES). Analogamente, os componentes do eco podem também serem suprimidos no sinal *downmix* do sinal de áudio espacial para ser transmitido ao *far-end*. Esta abordagem tipicamente é também mais eficiente que aplicar a supressão do eco a cada um dos sinais gravados dos microfones
15 usados para capturar o campo de som observado.

A seguir, sinais de referência resumidos serão usados para o objeto que aparece mais de uma vez em uma realização ou figura, mas que são, todavia, iguais ou similares pelo menos em termos de algumas de suas características ou estruturas. Por
20 exemplo, na Fig. 1 os quatro alto-falantes 100-1, ..., 100-4 foram denotados com sinais de referência individuais, no entanto, quando suas características e propriedades básicas como sendo alto-falantes forem discutidas, a referência é feita a "alto-falantes 100".

25 Ainda, para simplificar a descrição, objetos similares ou iguais serão denotados com os mesmos ou iguais sinais de referência. Comparando as Figs. 1 e 9, os alto-falantes foram referenciados com o mesmo sinal de referência 100. Objetos

denotados pelos mesmos sinais de referência ou sinais de referência iguais podem ser implementados idêntica, similar ou diferentemente. Por exemplo, em algumas implementações pode ser aconselhável implementar diferentes tipos de alto-falantes 100 para os diferentes sinais de alto-falantes, enquanto em diferentes aplicações os alto-falantes podem ser implementados identicamente. Assim, o objeto denotado pelo mesmo sinal de referência ou sinal igual pode opcionalmente ser implementado idêntica ou similarmente.

10 Além disso, dever ser notado que quando diversos objetos aparecem mais de uma vez em uma figura, o número retratado de objetos é tipicamente somente para propósito ilustrativo. Desvios do número pode ser feito também para aumentar ou diminuir o número. Por exemplo, a Fig. 1 mostra quatro alto-falantes 100-15 1,..., 100-4. No entanto, em realizações diferentes mais ou menos alto-falantes 100 podem ser igualmente bem implementados. Por exemplo, no caso de um sistema "5.1", 5 alto-falantes com um alto-falante *sub-woof* são tipicamente utilizados.

A seguir, lembramos a abordagem geral da supressão do eco acústico. Desse modo, seguimos basicamente o método conforme descrito em [8, 9].

Como ilustrado na Fig. 9, o sinal de microfone $z(n)$ é composto do sinal de eco acústico $y(n)$ que resulta do *feedback* do sinal de alto-falante $x(n)$ e do sinal de *near-end* $w(n)$. 25 Aqui, assumimos que a resposta do *impulse* de ambiente pode ser expresso como uma combinação de um caminho de propagação direta correspondente para um atraso de amostras d entre o sinal de alto-falante $x(n)$ e do sinal de microfone $z(n)$, e um filtro

linear g_n o qual modela as propriedades acústicas englobadas.

Então, o sinal de microfone $z(n)$ pode ser expresso por $z(n) = g_n * x(n-d) + w(n)$, (1) em que $*$ denota convolução. A representação do domínio de transformação *short-time* 5 *Fourier* (STFT) da equação (1) é dado por

$$Z(k,m) = G(k,m)X_d(k,m) + W(k,m), \quad (2)$$

em que k é um índice de bloco de tempo e m denota um índice de frequência. $X_d(k,m)$ é definido conforme a correspondência do domínio STFT do atraso do sinal de alto- 10 falante. O primeiro termo no lado direito da equação (2) representa os componentes do eco $Y(k,m)$, em que

$$Y(k,m) = G(k,m)X_d(k,m). \quad (3)$$

Deve ser notado que a seguinte discussão da supressão do eco acústico se refere ao STFT como uma representação 15 espectral dos sinais. No entanto, o conceito pode obviamente ser aplicado também a qualquer outra frequência de representação de *subband* adequada alternativa.

A supressão do eco acústico é desempenhada através da modificação da magnitude do STFT do sinal de microfone 20 $Z(k,m)$, enquanto mantém sua etapa inalterada. Isto pode ser expresso por

$$E(k,m) = H(k,m)Z(k,m), \quad (4)$$

em que $H(k,m)$ representa um valor real, fator de atenuação positiva. No seguinte referimo-nos a $H(k,m)$ como um filtro de supressão (ESF).

25 Uma abordagem prática da computação do filtro de supressão do eco $H(k,m)$ é para usar uma abordagem de subtração espectral paramétrica analogamente a [7]:

$$H(k, m) = \left(\frac{|Z(k, m)|^\alpha - \beta |\hat{Y}(k, m)|^\alpha}{|Z(k, m)|^\alpha} \right)^{\frac{1}{\gamma}}, \quad (5)$$

em que α , β e γ representam parâmetros de design para controlar o desempenho da supressão do eco.

Valores típicos para β e γ são valores em torno de 2, enquanto em algumas aplicações α é escolhido para ser o inverso de γ . Em outras palavras, ao escolher valores típicos de $\beta=2$ e $\gamma=2$, α é tipicamente escolhido para ser 0,5 ($=1/2$).

A estimativa do espectro de energia do sinal do eco pode ser obtida por

$$|\hat{Y}(k, m)|^2 = |\hat{G}(k, m)|^2 |X_d(k, m)|^2, \quad (6)$$

Onde $|\hat{G}(k, m)|^2$ representa uma estimativa da função de transferência da força do eco $|\hat{G}(k, m)|^2$. Alternativamente, uma abordagem com base no complexo do espectro de acordo com

$$|\hat{Y}(k, m)|^2 = |\hat{G}(k, m) X_d(k, m)|^2,$$

(7)

pode ser usada.

Note que na prática a função de transferência da força do eco $|\hat{G}(k, m)|^2$ e o atraso d não são conhecidos e devem ser substituídos por estimativas correspondentes, como discutido a seguir. A seguir nos referimos a $|\hat{G}(k, m)|^2$ como filtro estimado do eco (EEF).

Um método possível para estimar o EEF tem sido proposto em [8]. Supondo que o falante *near-end* está em silêncio, a equação (2) implica que o EEF pode ser estimado por

$$|\hat{G}(k, m)|^2 = \frac{\left| E\{Z(k, m) X_d^*(k, m)\} \right|^2}{\left| E\{X_d(k, m) X_d^*(k, m)\} \right|}, \quad (8)$$

em que * denota o operador complexo conjugado, e $E\{\dots\}$ denota o operador de expectativa. O operador de expectativa pode ser aproximado através da média flutuante de seu argumento.

A técnica acima estima efetivamente a função e
5 transferência do caminho do eco e tira a magnitude deste para obter um valor real do EEF. Sempre que a etapa mudar abruptamente, tal como durante a mudança de caminho do eco, flutuação do tempo, etc., esta estimativa do EEF pode ter que reconvergir. Para fazer a equação (8) insensível às variações da etapa, ela pode ser
10 modificada para computar a partir dos espectros de energia mais preferencialmente que a partir do complexo espectra [6]:

$$|\hat{G}(k, m)|^2 = \frac{E\{|X_d(k, m)|^2 |Z(k, m)|^2\}}{E\{|X_d(k, m)|^2 |X_d(k, m)|^2\}}. \quad (9)$$

No [6] é mostrado que a estimativa de acordo com (9) é tendenciosa. Assim, em [6] é proposto usar outra abordagem para estimar o EEF, a saber para estimar $|\hat{G}(k, m)|^2$ com base em
15 flutuações temporais dos espectros de energia dos sinais de alto-falante e do microfone. As flutuações temporais dos espectros de energia podem ser computadas de acordo com

$$\tilde{Z}(k, m) = |Z(k, m)|^2 - E\{|Z(k, m)|^2\} \quad (10)$$

$$\tilde{X}_d(k, m) = |X_d(k, m)|^2 - E\{|X_d(k, m)|^2\}. \quad (11)$$

A estimativa do EEF é então desempenhada analogamente à equação (9), mas com base na flutuação espectral do
20 alto-falante e do microfone:

$$|\hat{G}(k, m)|^2 = \frac{E\{\tilde{X}_d(k, m)\tilde{Z}(k, m)\}}{E\{\tilde{X}_d(k, m)\tilde{X}_d(k, m)\}}. \quad (12)$$

É importante notar que a flutuação da força espectral somente é usada para a estimativa $|\hat{G}(k, m)|^2$. A computação

do filtro de supressão do eco $H(k,m)$ é baseada ainda nos espectros de energia original dos sinal de alto-falante e microfone.

O valor de atraso d pode ser estimado usando a
5 função de coerência ao quadrado em relação os espectros de energia do alto-falante e do microfone de acordo com

$$\Gamma_d(k,m) = \frac{\left(E\{|X(k-d,m)|^2|Z(k,m)|^2\}\right)^2}{E\{|X(k-d,m)|^2|X(k-d,m)|^2\}E\{|Z(k,m)|^2|Z(k,m)|^2\}},$$

(13)

Em geral, o atraso d pode então ser escolhido diferente para cada frequência *bin* m . Aqui, no entanto, consideramos apenas um único atraso para todas as frequências.
10 Portanto, computamos uma previsão de ganho do eco de $\omega_d(k)$ como meio do $\Gamma_d(k,m)$ sobre a frequência

$$\omega_d(k) = \frac{1}{M} \sum_{m=0}^{M-1} \Gamma_d(k,m), \quad (14)$$

em que M denota o número de frequências *bins*. Então d é escolhido de forma que a previsão de ganho do eco seja maximizada, isto é,

15
$$d = \arg \max_d \{\omega_d(k)\} \quad (15)$$

Alternativamente à equação (15), a estimativa do valor do atraso d pode também ser desempenhada em relação a flutuação da espectra, isto é, com base na equação (10), (11).

Note que na prática, a expectativa matemática
20 $E\{\dots\}$, usada nas derivações acima, pode ter que ser substituída por *short-time* correspondentes ou médias flutuantes. Para exemplificar, consideramos

$$\Phi_{AB}(k,m) = E\{A(k)B(k,m)\}, \quad (16)$$

A média *short-time* $\hat{\Phi}_{AB}(k, m)$ correspondente à $\Phi_{AB}(k, m)$ pode, por exemplo, ser obtida por suavização recursiva de acordo com

$$\hat{\Phi}_{AB}(k, m) = (1 - \alpha_{avg}) \hat{\Phi}_{AB}(k - 1, m) + \alpha_{avg} A(k, m) B(k, m) \quad (17)$$

O fator α_{avg} determina o grau de suavização ao longo do tempo e pode ser ajustado para quaisquer requisitos dados.

A seguir discutimos como um único canal AES descrito na seção anterior pode analogamente ser aplicado também a um AES multicanal.

Deixemos que $X_l(k, m)$ denote a representação do domínio STFT do sinal de alto-falante l -th. Um espectro de força conjunta para todos os canais do alto-falante é então computada através da combinação dos espectros de energia dos sinais individuais do alto-falante:

$$|X(k, m)|^2 = \sum_{l=0}^{L-1} |X_l(k, m)|^2, \quad (18)$$

em que L denota o número de canais do alto-falante.

Alternativamente, o espectro de força conjunta dos sinais de alto-falante podem ser obtidos a partir da adição do espectro a cada sinal do alto-falante e então tirar a magnitude quadrada do espectro conjunto:

$$|X(k, m)|^2 = \left| \sum_{l=0}^{L-1} X_l(k, m) \right|^2, \quad (19)$$

Analogamente, um espectro de força conjunta é computado para os canais do microfone de acordo com

$$|Z(k, m)|^2 = \sum_{p=0}^{P-1} |Z_p(k, m)|^2, \quad (20)$$

em que $Z_p(k, m)$ representa o sinal de microfone p -th, e P denota o número de microfones.

No caso dos sinais de alto-falante, o espectro de força do microfone conjunto pode alternativamente ser computado de acordo com

$$|Z(k, m)|^2 = \left| \sum_{p=0}^{P-1} Z_p(k, m) \right|^2, \quad (21)$$

O modelo desejado para o espectro de força do eco é dado analogamente à equação (2), quando supondo independência estatística dos sinais de alto-falante e dos sinais do *near-end*:

$$|Z(k, m)|^2 \approx |G(k, m)|^2 |X_d(k, m)|^2 + |W(k, m)|^2 \quad (22)$$

em que, no caso do multicanal, os espectros de energia $|X(k, m)|^2$ e $|Z(k, m)|^2$ são dados pelas equações (18) e (20), respectivamente.

Para determinar o filtro da estimativa do eco $|G(k, m)|^2$ e o valor de atraso d , respectivamente, podemos também aplicar os diferentes métodos discutidos acima, mas usando os espectros de energia do alto-falante e do microfone conjuntos aqui definido.

A real supressão do eco é então desempenhada para cada sinal de microfone separadamente, mas através do uso dos mesmo filtros removedores de eco para cada canal do microfone:

$$E_p(k, m) = H(k, m) Y_p(k, m) \quad (23)$$

para $p = 0, 1, \dots, P-1$.

Nesta seção reveremos alguns exemplos importantes da representação do áudio espacial paramétrico e codificação do áudio espacial paramétrico. Assim, consideramos as abordagens de Codificação do Áudio Direcional (DirARC) [12], MPEG Surround (MPS)

[1], Codificação do Objeto de Áudio Espacial MPEG (SAOC) [5]. Antes de olhar os detalhes específicos das diferentes abordagens da codificação, consideramos as estruturas básicas codificador/decodificador que são comuns para todos os métodos aqui discutidos.

A estrutura geral do codificador de áudio espacial paramétrico é ilustrada na Fig. 2. A Fig. 2 mostra um multicanal ou um codificador de áudio espacial paramétrico 400. O codificador toma sinais de áudio múltiplos conforme a entrada e gera um sinal *downmix* de um ou mais canais e as informações laterais paramétricas correspondentes. Para ser um pouco mais específico, o codificador multicanal 400 é fornecido com uma pluralidade de sinais de entrada 410-1,...,410-N, que podem, à princípio, ser qualquer sinal de áudio. Com base nos sinais de entrada 410, o codificador 400 fornece um sinal *downmix* 310 e as informações laterais paramétricas 320, que juntas representam a pluralidade de sinais de entrada 410. Em muitos casos e implementações de um codificador multicanal 400 esta representação não é tipicamente sem perdas.

O codificador recebe como entrada múltiplos canais de áudio. Dependendo da real abordagem de codificação, estes canais de entrada de áudio podem representar sinais do microfone [12], sinais de alto-falante [10], ou sinais de entrada correspondentes aos chamados objetos de áudio espacial [5]. A saída do codificador é um sinal *downmix* 310 e as informações laterais 320 correspondentes. O sinal *downmix* compreende um ou mais canais de áudio. A informação lateral inclui metadados paramétricos, representando o campo de som observado, a relação

entre diferentes canais de entrada, ou a relação entre diferentes objetos de áudio. A saída do codificador, isto é, a combinação do sinal *downmix* e da informação lateral, é chamada corrente de áudio espacial ou representação de áudio espacial a seguir.

5 A estrutura geral de um decodificador de áudio espacial paramétrico correspondente é ilustrada na Fig. 3. A Fig. 3 mostra um decodificador 250 (multicanal), que tira um sinal *downmix* 310 e a informação lateral paramétrica 320 correspondente como entrada. O decodificador de multicanais 250 gera uma
10 pluralidade de sinais de saída 420-1,...,420-N, que podem ser, por exemplo, sinais de alto-falante (por exemplo, sinais de alto-falante 330 como mostrados na Fig. 1) correspondendo a uma realização desejada de *playback*. Como pode ser visto, o decodificador tira a corrente de áudio espacial como entrada. Com
15 base no sinal *downmix* e no metadado incluso na informação lateral, o decodificador computa sinais de alto-falante correspondente a uma realização desejada de *playback*. As instalações do alto-falante são, por exemplo, descritas em [1].

Um exemplo de um esquema de codificação de áudio
20 espacial paramétrica é a codificação de áudio direcional, que é também referida como DirAC. O DirAC usa uma representação paramétrica de uma campo de som usando a chegada de direção (DOA) e difusão do som em frequências *subband*s. Por isso, somente tira características de contas são relevantes para audição humana. A
25 abordagem DirAC é baseada na suposição da *interaural time differences* (ITD) e do *interaural level differences* (ILD) serem percebidas corretamente, se a direção de chegada de um campo de som for corretamente reproduzida. Correspondentemente, a coerência

interaural (IC) é supostamente percebida corretamente, se a difusão de um campo de som for reproduzida corretamente. Desta maneira, o lado da reprodução somente precisa dos parâmetros de direção e difusão e um sinal de mono microfone para gerar

5 características que são relevantes para a percepção humana do áudio espacial em uma dada posição de escuta com um conjunto arbitrário de alto-falantes.

Na DirAC, os parâmetros desejados (isto é, DOA $\varphi(k, m)$ do som e a difusão $\square(k, m)$ em cada frequência de banda) são

10 estimadas via uma análise energética do campo de som [12] com base nos sinais do microfone *B-format*. Os sinais do microfone *B-format* tipicamente compreendem um sinal direcional Omni $W(k, m)$, e dois sinais dipolo $(U_x(k, m), U_y(k, m))$ correspondentes as direções x-, y- de um sistema de coordenadas Cartesiana. Os sinais *B-format* podem

15 ser diretamente medidos usando, por exemplo, microfones de campo de som [2]. Alternativamente, uma disposição de microfones direcionais Omni pode ser usada para gerar os sinais *B-format* necessários [11].

No lado de reprodução (decodificador), os

20 diferentes sinais de alto-falante são computados com base no sinal mono *downmix* com os parâmetros de direção e difusão. Os sinais de alto-falante são compostos de componentes do sinal correspondente ao som direto e ao som difuso, respectivamente. O sinal do canal do alto-falante p th pode, por exemplo, ser computado de acordo com

$$X_p(k, m) = g_p(k, m)\sqrt{1 - \Psi(k, m)}W(k, m) + \sqrt{\Psi(k, m)}D_p\{W(k, m)\}, \quad (24)$$

25 em que $\square(k, m)$ denota a difusão na frequência *subband* m e o índice do bloco de tempo k . O ganho de movimento $g_p(k, m)$ depende do DOA do som $\varphi(k, m)$ e da posição do alto-falante p

em relação a posição de escuta desejada. O operador $D_p\{\dots\}$ corresponde a um decorrelator. O decorrelator é aplicado ao sinal *downmix* $W(k,m)$ quando computado o sinal do alto-falante p th.

A partir da discussão acima é seguido que os
5 sinais do microfone (*B-format* ou disposição de microfones direcionais Omni) representam a entrada do codificador DirAC 400. A saída do codificador é dada pelo sinal *downmix* $W(k,m)$ e pela direção $\varphi(k,m)$ e parâmetros da difusão ($\square(k,m)$) como informação lateral.

10 Correspondentemente, o decodificador 250 tira o sinal *downmix* $W(k,m)$ e a informação lateral paramétrica $\varphi(k,m)$ e $\square(k,m)$ como entrada para computar os sinais de alto-falante desejado de acordo com (24).

MPEG Surround (MPS) representa uma abordagem
15 eficiente para codificação do áudio espacial de alta qualidade [10]. Uma especificação completa do MPS pode ser encontrada em [1]. A seguir não olharemos os detalhes do MPS, mas preferencialmente reveremos aquelas partes que são relevantes no contexto das realizações da presente invenção.

20 O MPS explora o fato de que, de um ponto de vista perceptual, sinais de áudio multicanal tipicamente compreendem uma redundância significativa em relação aos diferentes canais de alto-falante. O codificador MPS toma sinais de alto-falante múltiplos como entrada, em que a realização espacial correspondente dos
25 alto-falantes tem que ser conhecida primeiro. Com base nestes sinais de entrada, o codificador MPS 400 computa parâmetros espaciais na frequência *subbands*, como diferenças de níveis de canal (CLD) entre dois canais, a correlação inter canal (ICC) entre

dois canais, e os coeficientes de previsão de canal (CPC) usados para prever um terceiro canal a partir de dois outros canais. A real informação lateral MPS 320 é então derivada a partir destes parâmetros espaciais. Além disso, o codificador 400 computa um
 5 sinal *downmix* que pode compreender um ou mais canais de áudio.

No caso mono, um sinal *downmix* $\mathbf{B}(k,m)$ obviamente compreende somente um canal $\mathbf{B}(k,m)$, considerando que no caso estéreo, o sinal *downmix* pode ser escrito como

$$\mathbf{B}(k, m) = [B_1(k, m), B_2(k, m)], \quad (25)$$

em que, por exemplo, $B_1(k,m)$ corresponde ao canal
 10 do alto-falante esquerdo e $B_2(k,m)$ denota o canal do alto-falante direito de uma configuração de alto-falante estéreo comum.

O decodificador MPS 250 toma o sinal *downmix* e a informação lateral paramétrica como entrada e computa os sinais de alto-falante 330, 420 para uma configuração de alto-falante
 15 desejada. A estrutura geral da cadeia de processamento de sinal usado no decodificador *MPEG surround* é ilustrada na Fig. 4 para o caso estéreo.

A Fig. 4 mostra uma representação esquemática do decodificador *MPEG surround* 250. Para o decodificador 250, o sinal
 20 *downmix* 310 e a informação lateral paramétrica são fornecidos. O sinal *downmix* 310 compreende os canais de sinal *downmix* $B_1(k,m)$ e $B_2(k,m)$, que correspondem aos canais de alto-falante direito e esquerdo de uma configuração de estéreo comum.

Em uma matriz *pre-mixing* 450 (M_1) os dois canais
 25 do sinal *downmix* 310 são transformados em um vetor de sinal intermediário $V(k,m)$. Partes dos componentes do vetor do sinal intermediário $V(k,m)$ são então fornecidos para uma pluralidade de

decorrelatores 460-1,..., 460-P que decorrelatam os respectivos componentes do vetor de sinal intermediário. Os sinais fornecidos pelos decorrelatores 460 e os sinais indecorrelatados ou componentes do sinal do vetor de sinal intermediário $V(k,m)$ formam um segundo vetor de sinal intermediário $R(k,m)$, que por sua vez é fornecido para a matriz *post-mixing* 470 (M_2). A matriz *post-mixing* 470 fornece na sua saída uma pluralidade de sinais de alto-falante 330-1,..., 330-P, que representam os sinais de saída 420 nos termos do decodificador mostrado na Fig. 3.

O decodificador 250 compreende ainda um processador de parâmetro 480 para qual a informação lateral paramétrica 320 é fornecida. O processador de parâmetro 480 é acoplado em ambas, a matriz *pre-mixing* 450 bem como a matriz *post-mixing* 470. O processador de parâmetro 480 é adaptado para receber a informação lateral paramétrica 320 e para gerar os elementos de matriz correspondentes a serem processados pela matriz *pre-mixing* 450 e a matriz *post-mixing* 470. Para facilitar isto, o processador de parâmetro 420 é acoplado a ambas as matrizes *pre-mixing* 450 e *post-mixing* 470.

Como implicado pela Fig. 4, o processos de decodificação pode ser escrito na notação da matriz de acordo com

$$\mathbf{V}(k,m) = \mathbf{M}_1(k,m)\mathbf{B}(k,m), \quad (26)$$

$$\mathbf{X}(k,m) = \mathbf{M}_2(k,m)\mathbf{R}(k,m), \quad (27)$$

Seguindo [1], $\mathbf{M}_1(k,m)$ denota a matriz *pre-mixing* 450 e $\mathbf{M}_2(k,m)$ a matriz *post-mixing* 470. Note que os elementos de $\mathbf{M}_1(k,m)$ e $\mathbf{M}_2(k,m)$ dependem da informação lateral espacial e da configuração do alto-falante usada para o *playback*, a qual pode ser fornecida pelo processador de parâmetro 480.

Como pode ser visto na Fig. 4, a relação entre os vetores de sinal intermediário $\mathbf{V}(k,m)$ e $\mathbf{R}(k,m)$ são dados conforme segue: Uma parte dos elementos do vetor de sinal $\mathbf{V}_p(k,m)$ é mantida inalterada ($\mathbf{R}_p(k,m) = \mathbf{V}_p(k,m)$), enquanto os outros componentes do $\mathbf{R}(k,m)$ são versões decorrelacionadas dos elementos correspondentes do $\mathbf{V}(k,m)$, isto é, $\mathbf{R}_l(k,m) = \mathbf{D}_l\{(k,m)\}$, em que $\mathbf{D}_l\{(k,m)\}$ descreve um operador decorrelator. Os elementos do vetor de sinal $\mathbf{X}(k,m)$ corresponde aos sinais de alto-falante multicanal $\mathbf{X}_p(k,m)$ usado para *playback*.

Deve ser notado que o MPS supõe canais do alto-falante como entrada, considerando que em sistemas de teleconferência a entrada consiste em sinais gravados do microfone. A conversão do sinal de entrada do microfone para canais do alto-falante correspondente pode ser necessária antes do MPS poder ser aplicado para determinar a representação do áudio espacial eficiente desejado do som gravado. Uma abordagem possível é para simplesmente usar microfones direcionais múltiplo que são dispostos de forma que os canais de alto-falante possam ser diretamente computados pela combinação de sinais de entrada do microfone. Alternativamente, uma computação com base DirAC dos canais dos alto-falantes pode ser aplicada, compreendendo uma conexão direta de um codificador DirAC e um decodificador DirAC como descritos nas seções anteriores.

Codificação do objeto de Áudio Espacial (SAOC) é baseada no conceito de representar uma cena de áudio complexa através de número de objetos únicos com uma descrição de cena correspondente. A fim de implementar uma maneira eficiente de alcançar este objetivo, o SAOC aplica técnicas que são bastante

relacionadas ao MPS [5]. Como antes, consideraremos aquelas partes do conceito SAOC que são relevantes no contexto desta invenção. Mais detalhes podem ser encontrados, por exemplo, em [5].

5 A estrutura geral do codificador SAOC é mostrada na Fig. 2, em que os sinais de entrada 410 correspondem a objetos de áudio. A partir destes sinais de entrada 410, o codificador SAOC 400 computa um sinal *downmix* 310 (mono ou estéreo) juntamente com a informação lateral paramétrica 320 correspondente
10 representando a relação de diferentes objetos de áudio na dada cena de áudio. Similar ao MPS, estes parâmetros são computados para cada índice do bloco de tempo e cada frequência *subband*. Estes parâmetros incluem Diferenças do Nível de Objeto (OLD), Coerência Cruzada Inter objeto (IOC), Energias do Objeto (NRG), e
15 outras medidas e parâmetros relacionadas ao sinal *downmix*.

 O decodificador SAOC 250 toma o sinal *downmix* 310 juntamente com a informação lateral 320 correspondente como entrada, e gera os sinais do canal do alto-falante para uma configuração desejada do alto-falante. O decodificador SAOC também
20 usa o motor de renderização MPS para determinar os sinais de alto-falante final. Note que adicionalmente a informação lateral gerada pelo codificador SAOC 400, o decodificador SAOC 250 toma também informação da configuração do alto-falante usado para renderização, ou outra informação interativa em relação ao
25 controle específico dos objetos de áudio, como entrada para computar os sinais de saída final. Isto é ilustrado na Fig. 5.

 A fig. 5 ilustra a estrutura geral de um decodificador SAOC 250. Para o decodificador SAOC 250 um sinal

downmix 310 e a informação lateral paramétrica 320 são fornecidos. Adicionalmente, o decodificador SAOC 250 é também fornecido com informação de interação ou de renderização 490. Como descrito acima, o decodificador SAOC 250 toma o sinal *downmix* 310, a informação lateral paramétrica 320 com um parâmetro de interação/renderização 490 para gerar uma pluralidade de sinais de alto-falante 330-1,..., 330-N. Estes sinais são gerados pelo decodificador SAOC 250.

Vamos considerar agora o decodificador SAOC para o caso de um sinal *downmix* mono e um sinal *downmix* estéreo, respectivamente. Seguindo [5], a estrutura do decodificador SAOC é ilustrada na Fig. 6(a) para o *downmix* mono e a Fig. 6(b) para o caso estéreo.

A Fig. 6a ilustra mais especificamente detalhes em relação a um transcodificador com base em *downmix* mono, que pode ser usado como um transcodificador de SAOC para MPS de acordo com [5]. O sistema mostrado na Fig. 6a compreende um decodificador *surround MPEG* 250, para o qual um sinal *downmix* 310 e uma corrente de *bit surround MPEG* como informação lateral paramétrica 320 é fornecida. O decodificador *surround MPEG* 250 gera na situação mostrada na Fig. 6a pelo menos cinco sinais do alto-falante 330-1,..., 330-5. Opcionalmente, o codificador *surround MPEG* 250 pode também gerar outros sinais de alto-falante, tal como o sinal de alto-falante *sub-woof*. No entanto, um alto-falante *sub-woof* correspondente não é mostrado na Fig. 6a para o bem da simplicidade, enquanto alto-falantes correspondentes 100-1,..., 100-5 para cada um dos alto-falantes 330 são mostrados na Fig. 6a.

Enquanto a corrente de *bit downmix* 310 é

diretamente fornecida para o decodificador *MPEG surround* 250, a
informação lateral paramétrica 320 é fornecida pelo
transcodificador SAOC para MPS 500, o transcodificador 500
compreende uma unidade de análise SAOC 510 para a qual uma
5 corrente de *bit* SAOC é fornecida como um sinal de entrada 520. A
unidade de análise SAOC 510 fornece como um de seus sinais de
saída informação a respeito de um número de objetos 530.

A unidade de análise SAOC 510 é ainda acoplada a
um motor de renderização de cena 540, que processa os dados
10 recebidos da unidade de análise SAOC 510 com base na matriz de
renderização 550 gerada por um gerador de matriz de renderização
560 e correspondendo a informação lateral 320 para o decodificador
surround MPEG 250. Portanto, o motor de renderização de cena 540 e
sua saída na qual a informação lateral 320 é fornecida para o
15 decodificador *MPEG surround* 250 também representa a saída do
transcodificador 500.

O gerador de matriz de renderização 560 é
fornecido com informação a respeito da configuração do *playback*
570, bem como com informação em relação as posições 580 do objeto
20 na base do qual o gerador de matriz de renderização 560 fornece a
matriz de renderização 550.

O decodificador *downmix* mono compreende
transcodificar a informação lateral SAOC em informação lateral MPS
520, com base nas posições 580 do objeto dadas e na realização 570
25 do alto-falante usada para o *playback*. A então determinada
informação lateral 320 MPS é alimentada no decodificador MPS 250
com o sinal *downmix* mono SAOC 310. Desde que o sinal *downmix* 310
permanece inalterado, a computação dos sinais de alto-falante pode

também ser expressa de acordo com as equações (26), (27), em que a matriz *pre-mixing* $M_1(k,m)$ e a matriz *post-mixing* $M_2(k,m)$ são determinadas a partir do transcodificador SAOC para MPS.

A Fig. 6b mostra um transcodificador SAOC para MPS 500 similar comparado ao transcodificador correspondente 500 mostrado na Fig. 6a. Assim, é feita referência às descrições acima. No entanto, o sistema, assim como o transcodificar 500 diferenciam-se principalmente em relação ao sinal *downmix* 310, que está na situação retratada na Fig. 6b um sinal *downmix* estéreo. Portanto, o decodificador *surround* MPEG 250 se difere do decodificador do decodificador *surround* MPEG correspondente na Fig. 6a pelo fato de que o sinal *downmix* 310 compreende dois canais de forma que o decodificador 250 é adaptado para gerar sinais de alto-falante 330 na base da informação lateral 320 e do sinal *downmix* estéreo 310.

O sistema mostrado na Fig. 6b difere-se do sistema mostrado na Fig. 6a em relação a outros detalhes. O transcodificador 500 compreende ainda o transcodificador *downmix* 590 que recebe um sinal *downmix* original 310' assim como a informação controle 600 a partir do motor de renderização de cena 540. O transcoder *downmix* 590 é portanto adaptado para gerar o sinal *downmix* 310 com base na informação controle 600 e o sinal *downmix* original ou de entrada 310'.

No caso estéreo, o sinal *downmix* SAOC 310' pode não representar uma entrada adequada para o decodificador MPS. Um exemplo para tal situação é, se os componentes do sinal do objeto forem inclusos somente no canal esquerdo do SAOC estéreo *downmix* 310', enquanto deveria ser renderizado para o hemisfério direito

durante o processo de *mixing* MPS [5]. Então, como mostrado na Fig. 6(b), o sinal *downmix* SAOC 310' tem que ser processado pelo então chamado transcodificador *downmix* 590 antes de poder usar como entrada para o decodificador MPS 250. As propriedades específicas deste estágio do processamento depende da real informação lateral SAOC 520 e da configuração 570 do *playback*. Obviamente a relação do sinal *downmix* 310 transcodificado e dos canais do alto-falante 330 usados para o *playback* podem então ser expressos também pelas equações (26) e (27).

10 Deve ser notado que o SAOC supõe sinais correspondentes para conjunto de objetos de áudio como entrada, considerando em sistemas de teleconferência, a entrada tipicamente compreende sinais de microfone gravados. Uma conversão do sinal de entrada do microfone para uma representação do objeto de áudio

15 espacial correspondente pode ser útil antes do SAOC poder ser aplicado para determinar a representação de áudio espacial eficiente desejada do som gravado. Uma possível abordagem para determinar diferentes objetos de áudio a partir da entrada da disposição de microfone é dada por técnicas de separação de fonte

20 cega tal como da [3]. Com base nos sinais de entrada do microfone, métodos de separação da fonte cega exploram a independência estatística de diferentes objetos de áudio para estimar sinais de áudio correspondente. No caso da configuração da disposição de microfone ser previamente conhecida, a informação espacial

25 adicional em relação aos objetos de áudio pode ser determinada também [4].

Somente para o bem da simplicidade, deve ser notado que através das informações de descrições e sinais

carregando as respectivas informações tem sido identificadas com o mesmo sinal de referência. Além disso, os sinais e as linhas de dados sobre a qual estes são transportados também foram identificados pelos mesmos sinais de referência. Dependendo da
5 implementação concreta da realização da presente invenção, informações podem ser trocadas entre diferentes unidades ou objetos através de sinais transmitidos diretamente sobre as linhas de sinal ou por meio de uma memória, uma locação de armazenamento ou outro mediador (por exemplo, um trinco) acoplado entre as
10 respectivas unidades ou objetos. Por exemplo, no caso da implementação com base no processador, a informação pode ser, por exemplo, armazenadas em uma memória associada com o respectivo processador. Assim, informações, partes da informação e sinais podem simultaneamente se referirem a este.

15 Com base na discussão da supressão do eco acústico e da codificação do áudio espacial paramétrico apresentados nas seções anteriores, apresentamos agora um método para integrar eficientemente a supressão do eco acústico (AES) em uma estrutura de codificador/decodificador de áudio espacial como
20 usada na telecomunicação de áudio espacial. De acordo com uma realização da presente invenção.

A estrutura geral da abordagem proposta é ilustrada na Fig. 7. A Fig. 7 mostra uma conferência *front-end* 200 de acordo com uma realização da presente invenção, em que a
25 supressão do eco acústico é baseada nos sinais *downmix* dos codificadores de áudio espacial paramétrico.

A conferência *front-end* 200 como mostrada na Fig. 7 compreende uma unidade de supressão do eco acústico 210 de

acordo com uma realização da presente invenção com uma interface de entrada 230, que é acoplada a um removedor de eco ou a uma unidade de supressão do eco 700 de forma que um sinal *downmix* 310 compreendido em um sinal de entrada 300 fornecido para a interface de entrada 230 é fornecida a estes. Na realização mostrada na Fig. 7, a informação lateral paramétrica 320, também separada do sinal de entrada 300 pela interface de entrada 230, não é fornecida para a unidade de supressão do eco 700.

Ambos, o sinal *downmix* 310 bem como a informação lateral paramétrica 320 são fornecidos para um decodificador de multicanais 250, que é ponderadamente acoplado a saída para uma pluralidade de alto-falantes 100-1,..., 100-N. O decodificador 220 fornece para cada alto-falante 100 um sinal do alto-falante correspondente 330-1,..., 330-N.

A conferência *front-end* 200 compreende ainda uma pluralidade de microfones 110-1,...,110-K que fornece sinal de entrada acústica para a conferência *front-end* 200. Em contraste, os alto-falantes 100 fornecem a saída acústica equivalente. Os microfones 110 são acoplados a uma unidade de processamento 710 e ainda a um codificador 400, que é adaptado para gerar outro sinal *downmix* 720 e outra informação lateral paramétrica 730 correspondendo aos sinais do microfone pré-processados recebidos do microfone 110. A unidade de supressão do eco 700 é acoplada a um codificador 400 de forma que a unidade de supressão do eco 700 é capaz de receber o outro sinal *downmix* 720 e a outra informação lateral 730. Em uma saída, a unidade de supressão do eco 700 fornece um sinal *downmix* modificado 740 com a outra informação lateral paramétrica 730 que passa através da unidade de supressão

do eco 700 sem ser alterada.

A unidade de supressão do eco 700 será delineada mais detalhadamente na Fig. 8 e compreende uma calculadora 220 e o filtro adaptativo 240 como mostrado na Fig. 1.

5 Aqui, a aplicação da comunicação do áudio espacial é considerada, em que assumimos que as cenas do áudio espacial no *far-end* e no *near-end* são representadas pelas correntes de áudio espacial que são transmitidas entre os diferentes *subscribers*. Considerando que a operação *hands-free* é
10 frequentemente essencial no caso de *surround playback* com alto-falantes múltiplos, uma unidade AES 210 pode ser útil para remover ecos incômodos na saída do decodificador do *near-end*. Em contraste com os métodos previamente descritos acima, em que o AES é desempenhado com base nos sinais de alto-falante, propomos
15 desempenhar o AES sozinho com base no sinal *downmix* 310 da corrente de áudio espacial 300 recebida a partir do *far-end*. Considerando que o número dos canais *downmix* é em geral bem mais baixo que o número dos sinais de alto-falante usado para o *playback*, o método proposto é significativamente mais eficiente em
20 relação a complexidade. O AES pode ser aplicado para os sinais do microfone no *near-end*, ou, até mesmo mais eficientemente, ao sinal *downmix* do codificador de saída dos *near-ends*, como ilustrado na Fig. 7.

Antes de descrever a unidade de supressão do eco
25 700 em mais detalhes no contexto da Fig. 8, a seguir o processo ou método de acordo com uma realização da presente invenção será descrito em mais detalhes.

Primeiro, uma referência do espectro de força

(RPS) dos sinais do *playback* $P(k, m)$ com base no sinal *downmix* 310 da corrente de áudio espacial recebida é computada. No caso geral de uma sinal de *downmix* de canal N $\mathbf{B}(k, m) = [\mathbf{B}_1(k, m), \mathbf{B}_2(k, m), \dots, \mathbf{B}_N(k, m)]$, isto pode ser desempenhado de acordo com uma combinação

5 linear

$$|P(k, m)|^2 = \sum_{i=0}^{N-1} a_i(k, m) |\mathbf{B}_i(k, m)|^2, \quad (28)$$

Alternativamente, a combinação linear pode ser computada em relação ao espectro complexo nos canais *downmix*

$$|P(k, m)|^2 = \left| \sum_{i=0}^{N-1} a_i(k, m) \mathbf{B}_i(k, m) \right|^2, \quad (29)$$

Os fatores de ponderação $a_i(k, m)$ podem ser usados para controlar a contribuição de diferentes canais de *downmix* para

10 o RPS.

Uma ponderação diferente dos canais pode ser, por exemplo, benéfica no contexto do SAOC. Quando a entrada da AES é determinada antes do transcodificador *downmix* ser aplicado ao SAOC sinal de *downmix* (ver Fig. 6(b)), o comportamento da variação do

15 tempo do transcodificador *downmix* pode não ter que ser modelado pelo filtro de estimativa do eco, mas já é capturado pela computação do espectro de força referência.

Para o caso especial do sinal *downmix* mono, é razoável simplesmente escolher o RPS igual ao espectro de força do

20 sinal *downmix*, isto é, $|P(k, m)|^2 = |\mathbf{B}(k, m)|^2$.

Em outras palavras, os coeficientes de ponderação $a_i(k, m)$ são escolhidos para ser um para o canal *downmix* único compreendido no sinal *downmix* 310.

Analogamente às equações (28) e (29), computamos

25 um RPS $Q(k, m)$ dos sinais gravados com base no sinal *downmix* canal

K $\mathbf{A}(k, m) = [A_1(k, m), A_2(k, m), \dots, A_K(k, m)]$ do codificador do *near-end*:

$$|Q(k, m)|^2 = \sum_{i=0}^{K-1} c_i(k, m) |A_i(k, m)|^2, \quad (30)$$

Alternativamente, a combinação linear pode ser computada em relação ao espectra complexo dos canais *downmix*

$$|Q(k, m)|^2 = \left| \sum_{i=0}^{K-1} c_i(k, m) A_i(k, m) \right|^2, \quad (31)$$

5 Os fatores de ponderação $C_i(k, m)$ podem ser usados para controlar a contribuição dos diferentes canais *downmix* para o RPS. Como antes, podemos simplesmente usar $|Q(k, m)|^2 = |A(k, m)|^2$ no caso de um sinal *downmix* mono ($c_i(k, m) = 1$).

O sinal *downmix* $\mathbf{A}(k, m)$ e, assim também o RPS
10 $|Q(k, m)|^2$, contém tipicamente componentes do eco indesejados resultando a partir do *feedback* dos sinais de alto-falante. Uma estimativa $|\hat{Y}(k, m)|^2$ dos componentes do eco $|Q(k, m)|^2$ é computada com base na versão atrasada do RPS $|P(k, m)|^2$ e na estimativa da função de transferência do poder do eco de acordo com

$$|\hat{Y}(k, m)|^2 = |\hat{G}(k, m)|^2 |P(k-d, m)|^2, \quad (32)$$

15 Analogamente à descrição acima, $|\hat{G}(k, m)|^2$ é chamado filtro de estimativa do eco (EEF) a seguir.

Esta estimativa é então usada para determinar um filtro de supressão do eco (ESF), por exemplo, analogamente à [5]:

$$H(k, m) = \left(\frac{|Q(k, m)|^\alpha - \beta |\hat{Y}(k, m)|^\alpha}{|Q(k, m)|^\alpha} \right)^{\frac{1}{\gamma}} \quad (33)$$

em que α , β e γ representam parâmetros de design
20 para controlar o desempenho da supressão do eco. Valores típicos para α , β e γ foram dados acima.

A remoção dos componentes do eco indesejado é

finalmente obtida através da multiplicação dos canais do sinal *downmix* original do codificador do *near-end* com o ESF

$$E_{A,i}(k, m) = H(k, m)A_i(k, m), \quad (34)$$

A estimativa do EEF pode ser baseada na correlação em relação aos RPSS de acordo com

$$|\hat{G}(k, m)|^2 = \frac{E\{P(k-d, m)|Q(k, m)|^2\}}{E\{P(k-d, m)^2|P(k-d, m)|^2\}}$$

(35)

5 Alternativamente, o filtro EEF pode ser estimado usando flutuações temporais dos RPSS, isto é, analogamente a (12):

$$|\hat{G}(k, m)|^2 = \frac{E\{\tilde{P}(k-d, m)\tilde{Q}(k, m)\}}{E\{\tilde{P}(k-d, m)\tilde{P}(k-d, m)\}},$$

(36)

em que as flutuações temporais dos RPSS são computadas de acordo com

$$\tilde{Q}(k, m) = |Q(k, m)|^2 - E\{|Q(k, m)|^2\}$$

(37)

$$\tilde{P}(k-d, m) = |P(k-d, m)|^2 - E\{|P(k-d, m)|^2\}, \quad (38)$$

10 A estimativa do parâmetro de atraso d pode ser desempenhada analogamente à (13), quando substituindo os sinais de alto-falante e do microfone $X(k, m)$ e $Z(k, m)$, pelo RPS correspondente $P(k, m)$ e $Q(k, m)$, respectivamente.

Deve ser mencionado que tipicamente não há uma relação de etapa significativa entre os sinais *downmix* $A(k, m)$ e $B(k, m)$. Isto porque suas etapas são relacionadas não somente ao longo da resposta de frequência do ambiente, mas também pelo processo altamente variante no tempo de determinação dos sinais de alto-falante a partir do sinal *downmix* e da informação lateral espacial. Assim, as abordagens que usam a informação da etapa para

15

estimar o EEF (ou o atraso), tal como (8), não são adequados quando desempenhando a remoção do eco com base nos sinais *downmix*.

É importante mencionar que as mesmas razões são válidas para o caso do cancelamento do eco usando técnicas de filtragem adaptativa linear devem ser aplicadas em relação aos sinais *downmix*. Tais filtros adaptativos teriam que modelar e rastrear as mudanças altamente variantes no tempo causadas pelo mapeamento do sinal *downmix* para os canais do alto-falante.

A Fig. 8 mostra um diagrama em bloco de uma conferência de *front-end* 200 de acordo com uma realização da presente invenção, que é bastante similar àquela mostrada na Fig. 1. Portanto, a referência é feita para a descrição da Fig. 1.

A conferência de *front-end* 200 também compreende uma unidade de supressão do eco acústico 210 de acordo com uma realização da presente invenção, que por sua vez compreende uma calculadora 220 para desempenhar essencialmente a mesma funcionalidade como descrita no contexto da Fig. 1. No entanto, a seguir será dada uma descrição mais detalhada. A conferência de *front-end* 200 compreende ainda uma interface de entrada 230 e um filtro adaptativo 240. A conferência de *front-end* 200 compreende ainda um decodificador de multicanais 250, que é acoplado a uma pluralidade de alto-falantes 100-1,..., 100-N. A conferência de *front-end* 200 compreende ainda um codificador correspondente ou codificador multicanal 400, que por sua vez é acoplado a uma pluralidade de microfones 110-1,..., 110-K.

Para ser um pouco mais específico, um sinal de entrada 300 é fornecido para a interface de entrada 230 a partir do *far-end* de um sistema de comunicação subjacente ao *front-end*

200. Na realização da Fig. 8, a interface de entrada 230 separa o sinal *downmix* 310 e uma informação lateral paramétrica 320 a partir do sinal de entrada e fornece este como sinais de entrada do decodificador de multicanais 250. Dentro do decodificador de multicanais 250 os dois sinais, o sinal *downmix* 310 e a informação paramétrica lateral 320, são decodificados em uma pluralidade de sinais de alto-falante correspondentes 330, que são então fornecidos para os respectivos alto-falantes 100. Para o bem da simplicidade, somente o primeiro sinal do alto-falante 330-1 é rotulado como tal.

O decodificador 250 compreende, na realização mostrada na Fig. 8, um *upmixer* 705 e um processador de parâmetro 480. O *upmixer* 705 é acoplado a interface de entrada 230 e adaptado para receber um sinal *downmix* 310. Similarmente, o processador de parâmetro 480 é também acoplado a interface de entrada 230, mas adaptado para receber a informação lateral paramétrica 320. O *upmixer* 705 e o processador de parâmetro 480 são interconectados de forma a informação de controle *upmix* 707 derivada a partir da informação lateral paramétrica 320 pode ser transmitida para o *upmixer* 705. O *upmixer* 705 é também acoplado aos alto-falantes 100.

Em relação a sua funcionalidade, o *upmixer* 705 é adaptado para gerar sinais do alto-falante 330 a partir do sinal *downmix* 310 com base na informação de controle do *upmix* 707 derivada a partir da informação lateral paramétrica 320. Para cada alto-falante N (N sendo um número inteiro) 100-1, ..., 100-N, o *upmixer* 705 fornece um sinal do alto-falante 330 individual.

Como discutido antes, o decodificador 250 pode

opcionalmente compreender uma interface que extrai a informação lateral 320 e o *downmix* 310 e fornece estes para o processador de parâmetro 480 e para o *upmixer* 705, respectivamente, no case da interface de entrada 230 não ser dividida pelo decodificador 250 e
5 a unidade de supressão do eco acústico 710.

Como já descrito no contexto da Fig. 1, uma saída da interface de entrada 230 é acoplada a uma calculadora 220 para fornecer o sinal *downmix* 310 para a calculadora 220. Em outras palavras, a calculadora 220 é adaptada para receber o sinal
10 *downmix* 310.

Antes de descrever a estrutura interna da calculadora 220 em mais detalhes, deve ser notado que o microfone 110 fornece um respectivo número K (K sendo um número inteiro) de sinais do microfone 340, dos quais somente o primeiro sinal de
15 microfone 340-1 é rotulado como tal na Fig. 8 para o codificador multicanal 400.

Com base nos sinais do microfone 340 recebidos o codificador multicanal 400 gera outro sinal *downmix* 720 e outra informação lateral paramétrica 730 com base nos sinais do
20 microfone recebidos. Enquanto as outras informações laterais paramétricas 730 são fornecidas para uma saída do sistema de conferência 200, o outro sinal *downmix* 720 é fornecido para ambos, a calculadora 220 e o filtro adaptativo 240. A calculadora 220 também fornece um sinal de coeficiente do filtro 350 para o filtro
25 adaptativo 240 na base do qual o outro sinal *downmix* 720 é filtrado para obter um sinal *downmix* modificado 740 em uma saída do filtro adaptativo 240. O sinal *downmix* modificado 740 representa uma versão eco suprimida do outro sinal *downmix* 720 que

chega. Como consequência, no lado receptor do outro sinal *downmix* 720 e da outra informação lateral paramétrica 730 uma versão eco suprimida do sinal de microfone recebido pelo microfones 110 pode ser reconstruído.

5 Em relação a estrutura interna da calculadora 220, os sinais *downmix* 310 da interface de entrada 330 é fornecido para o primeiro gerador de referência do espectro de força 800 que é adaptado para gerar a previamente descrita referência do espectro de força, por exemplo, de acordo com as equações (28) e
10 (29). Uma saída do primeiro gerador de força de referência 800 é acoplado a um temporizador 810, que é adaptado para atrasar um sinal vindouro através de um valor de atraso d . Uma saída do temporizador 810 é então acoplada a um estimador de eco 820, que pode ser, por exemplo, adaptado para calcular uma estimativa do
15 eco de acordo com a equação (38). Uma saída do estimador de eco 820 é então acoplada a uma entrada do gerador do filtro de supressão do eco 830, que gera ou estima o filtro de supressão do eco de acordo com a equação (33). Uma saída do gerador do filtro de supressão do eco 830 é o sinal de coeficiente do filtro 350
20 compreendendo um coeficiente de filtro, que é fornecido para o filtro adaptativo 240.

O outro sinal *downmix* 720 conforme gerado pelo codificador 400 é fornecido para o gerador do filtro de supressão do eco 830, se este circuito compreender um segundo gerador de
25 referência do espectro de força 840 ou for fornecido para o segundo gerador de referência do espectro de força 840. Para realizar isto, a unidade de supressão acústica do eco 210 pode opcionalmente compreender um interface de entrada adicional para

extrair o outro sinal *downmix* 720, se necessário.

Uma saída do segundo gerador de referência do espectro de força 840 é então acoplada a um gerador de coeficiente do filtro de supressão do eco, que por sua vez é acoplado a um
5 estimador de eco 820 para fornecer coeficientes do filtro de estimativa do eco de acordo com a equação (35) ou (36) para um estimador de eco 820. No caso do gerador de coeficiente do filtro de estimativa do eco 850 operar com base na equação (36), um primeiro e um segundo compensadores opcionais de flutuação
10 temporal 860, 870 são acoplados entre o gerador de coeficiente do filtro de estimativa do eco 850 e uma saída do temporizador 810 e do segundo gerador de referência do espectro de força 840, respectivamente. Os dois compensadores de flutuação temporal 860, 870 podem então serem adaptados para calcular espectro da força de
15 referência com base nas equações (37) e (38), respectivamente. Então, o gerador de coeficiente do filtro de estimativa do eco 850 pode usar o espectro da força de referência modificado para operar com base na equação (36).

Deve ser notado que o temporizador 810 não é
20 necessário, mas frequentemente um componente útil. Uma determinação do valor de atraso d pode ser alcançado com base nas computações de acordo com as equações (13), (14) e (15). Para ser mais preciso, uma realização de acordo com a presente invenção pode desta forma compreender uma calculadora de coerência 880, na
25 qual a entrada ponderada é acoplada a uma saída do primeiro gerador de referência do espectro de força 840 para fornecer a calculadora de coerência 880 com a respectiva referência do espectro de força.

Por exemplo, com base na equação (13) mas com os dois espectros de força de referência conforme fornecido pelos dois geradores de referência do espectro de força 800, 840 a calculadora de coerência 880 pode gerar valores de uma função de coerência de acordo com a equação (13) para uma calculadora de ganho de previsão do eco 890, $\omega_d(k)$ de acordo ou com base na equação (14). Uma saída da calculadora de ganho de previsão do eco é então acoplada em uma entrada do otimizador 900, que pode ser adaptado para otimizar o valor de atraso d de acordo com a equação (15). Para fornecer o valor de atraso d para o temporizador 810, o otimizador 900 é acoplado ao temporizador 810 e o temporizador 810 é adaptado para receber o valor de atraso d . Naturalmente, o temporizador é também neste caso adaptado para atrasar o sinal vindouro (aqui a primeira referência do espectro de força) através do valor de atraso d .

Para o bem da plenitude a unidade de supressão do eco 700 também é mostrada na Fig. 8, que compreende uma calculadora 220 bem como um filtro adaptativo 240 como já delineado no contexto da Fig. 7.

No restante desta seção apresentaremos variações práticas do método acima para o sinal *downmix* com base na supressão do eco.

Podemos obter uma variação da equação (32) de acordo com

$$|\hat{Y}(k, m)|^2 = |\hat{G}(k, m) P(k-d, m)|^2, \quad (39)$$

em que o espectro de referência complexa dos sinais do *playback* $P(k, m)$ é computado em relação ao complexo

espectra dos canais *downmix*, isto é, de acordo com

$$P(k, m) = \sum_{i=0}^{N-1} a_i(k, m) B_i(k, m), \quad (40)$$

A equação (40) resulta da (29) pelo descarte da computação da magnitude.

Outra modificação da abordagem da AES pode ser
 5 obtida pelo desempenho da supressão do eco não nos canais *downmix*,
 como proposto pela (34), mas em relação aos sinais de entrada do
 microfone como alternativa. Em outras palavras, a supressão do eco
 é desempenhada nos sinais do microfone gravados originalmente
 antes de serem usados como entrada do codificado *near-end* ou em
 10 qualquer estágio de pré-processamento, respectivamente.

Muitas realizações de acordo com a presente invenção deste modo dividem as seguintes características:

1. Receber uma primeira representação de áudio espacial paramétrico, consistindo em um sinal *downmix* com a
 15 informação lateral, que é usada para gerar sinais de alto-falantes multicanal.

2. Receber uma segunda representação de áudio espacial paramétrico, consistindo em um sinal *downmix* com a
 20 informação lateral, que tem sido determinada a partir de sinais do microfone gravados.

3. Computar um espectro de força de referência do primeiro e do segundo sinais de *downmix*.

4. Computar um filtro de estimativa do eco para estimar os componentes do eco na referência do espectro de força
 25 do segundo sinal *downmix*.

5. Computar um filtro de remoção do eco a partir

da referência do espectro de força do primeiro sinal *downmix*, a referência do espectro de força do segundo sinal *downmix*, e o filtro de remoção do eco para remover os componentes do eco no sinal *downmix* da segunda representação de áudio espacial.

5 Dependendo de certos requisitos de implementação das realizações dos métodos inventivos, realizações dos métodos inventivos podem ser implementadas em *hardware* ou em *software*. A implementação pode ser desempenhada usando um meio de armazenamento digital, em particular, um disco, um CD ou um DVD
10 tendo um sinal de controle legível eletronicamente instalado nele que coopera com um computador programável ou um processador de forma que uma realização dos métodos inventivos é desempenhada. Geralmente, uma realização da presente invenção é, deste modo, um produto do programa de computador em que o código do programa
15 armazenado em um suporte de leitura ótica, o código do programa sendo operativo para desempenhar uma realização do método inventivo, quando o produto do programa de computador opera no computador do processador. Em outras palavras, realizações do método inventivo são, deste modo, um programa de computador tendo
20 um código do programa para desempenhar pelo menos uma das realizações do método inventivo, quando os programas de computador operar no processador do computador. Um processador pode ser formado por um computador, um cartão de *chip*, um cartão *smart*, um circuito integrado de aplicação específica (ASIC) ou outro
25 circuito integrado.

Realizações de acordo com a presente invenção podem ainda serem implementadas com base em elementos elétricos discretos ou elementos eletrônicos, circuitos integrados ou

combinações destes.

Realizações de acordo com a presente invenção permitem deste modo um controle do eco acústico para reprodução do áudio espacial paramétrico. Como a prévia discussão tem mostrado, as realizações podem representar um método eficiente para a supressão dos ecos acústicos para sistemas de alto-falantes multicanal usados em sistemas de comunicação de áudio espacial. Os métodos são aplicáveis em casos que os sinais de áudio espacial são representados por um sinal *downmix* e informação lateral paramétrica correspondente ou metro dados. As realizações exploram o fato de que a supressão do eco pode ser desempenhada diretamente com base no sinal *downmix* recebido mais preferivelmente que explicitamente computando os sinais de alto-falante antes de eles serem introduzidos em uma supressão de eco acústico. Analogamente, os componentes do eco podem também ser suprimido no sinal *downmix* do sinal de áudio espacial para ser transmitido para o *far-end*.

LISTA DOS SINAIS DE REFERÊNCIA

	100	alto falante
	110	microfone
20	120	unidade de remoção do eco
	200	conferência de <i>front-end</i>
	210	unidade de supressão do eco acústico
	220	calculadora
	230	interface de entrada
25	240	filtro adaptativo
	250	decodificador de multicanais
	300	sinal de entrada
	310	sinal <i>downmix</i>

	320	informação lateral paramétrica
	330	signal de alto-falante
	340	signal de microfone
	350	signal do coeficiente de filtro
5	360	signal de saída
	400	codificador multicanal
	410	signal de entrada
	420	signal de saída
	450	matriz <i>pre-mixing</i>
10	460	decorrelator
	470	matriz <i>post-mixing</i>
	480	processador de parâmetro
	490	informação de renderização/interação
	500	transcodificador
15	510	unidade de análise SAOC
	520	corrente de <i>bit</i> SAOC
	530	número de objetos
	540	motor de renderização de cena
	550	matriz de renderização
20	560	gerador da matriz de renderização
	570	configuração do <i>playback</i>
	580	posição do objeto
	590	transcodificador <i>downmix</i>
	600	informação de controle
25	700	unidade de supressão do eco
	710	unidade de processamento
	720	outro signal <i>downmix</i>
	730	outra informação lateral paramétrica

	740	sinal <i>downmix</i> modificado
	800	gerador da primeira referência do espectro de força
	810	temporizador
5	820	estimador do eco
	830	gerador do filtro de supressão do eco
	840	gerador da segunda referência do espectro de força
10	850	gerador do coeficiente de filtro de estimativa do eco
	860	compensador da primeira flutuação temporal
	870	compensador da segunda flutuação temporal
15	880	calculadora de coerência
	890	calculadora do ganho de previsão do eco
	900	otimizador

REFERÊNCIAS

- [1] ISO/IEC 23003-1:2007. "Information technology
20 - MPEG Audio technologies - Part 1: MPEG Surround". (Tecnologia da Informação - Tecnologias de Áudio MPEG - Parte 1: MPEG Surround). - Organização de Padrões Internacionais, Genebra, Suíça, 2007.

- [2] E. Benjamin and T. Chen. "The native B-format microphone: Part I". (O microfone nativo em *B-format*: Parte I). Na
25 119º Convenção AES, Ensaio 6621, Nova York, Outubro, 2005.

- [3] H. Buchner, R. Aichner, and W. Kellermann. "A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics". (A

generalização de algoritmos de separação de fonte cega para misturas com convolutivas com base em segundas ordens estatísticas) *IEEE trans. em Processamento de Fala e Áudio*, 13(1):120–134, Jan. 2005.

5 [4] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, and W. Kellermann. "Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering". (Localização Simultânea das múltiplas fontes de som usando filtragem MIMO adaptativa cega). Em *Proc. IEEE Int. Conferência em Acústica, Fala e Sinal de Processamento (ICASSP)*, Filadélfia, Março de 2005.

[5] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, and W. Oomen. "Spatial audio object coding (SAOC) – the upcoming MPEG standard on parametric object based audio coding". (Codificação do objeto de audio espacial SAOC) – o próximo padrão MPEG em objeto paramétrico com base na codificação do audio". Na 124ª Convenção AES, Ensaio 7377, Amsterdam, Maio de 2008.

[6] A. Favrot et. al. "Acoustic echo control based on temporal fluctuations of short-time spectra". (Controle do eco acústico). Em *Proc. Intl. Works. on Acoust. Controle do Eco e Ruído (IWAENC)*, Seattle, Setembro de 2008, apresentados.

[7] W. Etter and G. S. Moschytz. "Noise reduction by noise-adaptive spectral magnitude expansion". (Redução de Ruído através da expansão de magnitude espectral adaptativa de ruído). *J. Audio Eng. Soc.*, 42:341–349, Maio de 1994.

[8] C. Faller and C. Toumery. "Estimating the delay and coloration effect of the acoustic echo path for low

complexity echo suppression". (Estimando o atraso e o efeito de coloração do caminho do eco acústico para baixa complexidade para supressão do eco). Em *Proc. Intl. Works. on Acoust. Controle do Eco e de Ruído*, (IWAENC), Setembro de 2005.

5 [9] A. Favrot, C. Faller, M. Kallinger, F. Kuech, and M. Schmidt. "Acoustic echo control based on temporal fluctuations of short-time spectra". (Controle do eco acústico com base em flutuações temporais do espectro *short-time*). Em *Proc. Intl. Works. on Acoust. Controle do Eco e de Ruído*. (IWAENC), Set.
10 2008.

[10] Jürgen Herre, Kristofer Kjørling, Jeroen Breebaart, Christof Faller, Sascha Disch, Heiko Purnhagen, Jeroen Koppens, Johannes Hilpert, Jonas Roden, Werner Oomen, Karsten Linzmeier, and Kok Seng Chong. "MPEG Surround - The ISO / MPEG
15 Standard for efficient and compatible multichannel audio coding". (*MPEG Surround - O ISO/Padrão MPEG para codificação eficiente e compatível do audio multicanal*). *J. Audio Eng. Soc.*, 56(11):932-955, Nov. 2008.

[11] J. Merimaa. "Applications of a 3-D microphone
20 array". (Aplicações de uma disposição de microfone 3-D). Na 112ª Convenção AES, Ensaio 5501, Munique, Maio de 2002.

[12] V. Pulkki. "Spatial sound reproduction with directional audio coding". (Reprodução do Som Espacial com codificação direta do som) *J. Audio Eng. Soc.*, 55(6):503-516,
25 Junho de 2007.

[13] G. Schmidt and E. Hansler. "Acoustic echo and noise control: a practical approach". (Controle do Ecos Acústico e Ruído: uma abordagem prática). Hoboken: Wiley, 2004.

REIVINDICAÇÕES

1. MÉTODO PARA SUPRESSÃO DO ECO ACÚSTICO, caracterizado por compreender:

extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix (310) e as informações laterais paramétricas (320), em que o sinal de downmix (310) e as informações laterais paramétricas (320) juntos representam um sinal multicanal compreendendo pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix(310);

calcular coeficientes de filtro (350) para filtragem adaptativa com base no sinal de downmix (310) e em um sinal do microfone (340) ou um sinal derivado do sinal do microfone;

filtrar de forma adaptativa o sinal do microfone (340) ou o sinal derivado do sinal do microfone com base nos coeficientes de filtro (350) para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340),

em que os coeficientes de filtro (350) são calculados determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal derivado do sinal de microfone (720) , determinando coeficientes de filtro de estimativa de eco com base no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de potência de referência e nos coeficientes de filtro de estimativa de eco e determinando os coeficientes de filtro (350)

com base nos coeficientes de filtro de estimativa de eco e o segundo espectro de potência de referência.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que compreende ainda decodificar o sinal de downmix (310) e a informação lateral paramétrica (320) em uma pluralidade de sinais de alto-falante.

3. INTERFACE DE CONFERÊNCIA (200) caracterizada por uma unidade de supressão de eco acústico (210) que compreende: uma interface de entrada (230) adaptada para extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix (310) e a informação lateral paramétrica (320), em que o sinal de downmix (310) e a informação lateral paramétrica (320) juntos representam um sinal multicanal compreendendo pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix (310);

uma calculadora (220) adaptada para calcular coeficientes de filtro (350) para um filtro adaptativo (240), em que a calculadora (220) está adaptada para receber o sinal de downmix (310), em que a calculadora (220) está ainda adaptada para receber um sinal de microfone (340) ou um sinal derivado do sinal de microfone, em que a calculadora está adaptada para determinar os coeficientes de filtro (350) com base nos sinais recebidos ,

em que os coeficientes de filtro (350) são calculados determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal derivado do sinal de microfone (720) , determinando coeficientes de filtro de estimativa de eco com base

no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de potência de referência e nos coeficientes de filtro de estimativa de eco e determinando os coeficientes de filtro (350) com base nos coeficientes de filtro de estimativa de eco e o segundo espectro de potência de referência ;

um filtro adaptativo (240) adaptado para receber os coeficientes de filtro (350) da calculadora (220) e adaptado para filtrar o sinal do microfone (340) ou o sinal derivado do sinal do microfone (340) com base nos coeficientes do filtro para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340);

um decodificador multicanal (250);

pelo menos uma unidade de microfone (110); em que

o decodificador multicanal (250) é adaptado para decodificar o sinal de downmix (310) e a informação lateral paramétrica (320) para uma pluralidade de sinais de alto-falante; e

a pelo menos uma unidade de microfone (110) está adaptada para fornecer o sinal de microfone (340).

4. INTERFACE DE CONFERÊNCIA (200) de acordo com a reivindicação 3 caracterizada pelo fato de que a interface de entrada (230) é ainda adaptada para extrair a informação lateral paramétrica (320), em que o decodificador multicanal (250) compreende um upmixer (705) e um processador de parâmetro (480), em que o processador de parâmetro (480) está adaptado para receber as informações laterais de parâmetro (320) da entrada interface (230) e está adaptado para proporcionar um upmix sinal de

controle, e em que o upmixer está adaptado para receber a sintetização de sinal a partir da interface de entrada e o upmix sinal de controle do processador de parâmetro (480) e está adaptado para proporcionar a pluralidade de sinais de alto-falantes com base na sintetização do sinal (310) e o sinal de controle do upmixer.

5. INTERFACE DE CONFERÊNCIA (200) de acordo com a reivindicação 3 caracterizada pelo fato de que compreende uma pluralidade de unidades de microfone, em que a pluralidade de unidades de microfone está adaptada para fornecer a pluralidade de sinais de entrada de áudio.

6. Método para fornecer uma pluralidade de sinais de alto-falante e um sinal de microfone (340), caracterizado por compreender:

extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix e as informações laterais paramétricas (320), em que o sinal de downmix (310) e as informações laterais paramétricas (320) juntos representam um sinal multicanal compreendendo pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix (310);

calcular coeficientes de filtro (350) para filtragem adaptativa com base no sinal de downmix (310) e o sinal de microfone (340) ou um sinal derivado do sinal de microfone, em que os coeficientes de filtro (350) são calculados determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal derivado do sinal de microfone (720), determinando coeficientes de

filtro de estimativa de eco com base no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de energia de referência e os coeficientes de filtro de estimativa de eco, e determinando os coeficientes de filtro (350) com base nos coeficientes de filtro de estimativa de eco e o segundo espectro de energia de referência;

filtrar adaptativamente o sinal do microfone (340) ou o sinal derivado do sinal do microfone com base nos coeficientes (350) de filtro para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340);

uma etapa de decodificação multicanal;

uma etapa de receber um sinal de microfone (340), em que, na etapa de decodificação multicanal, o sinal de downmix e a informação lateral paramétrica (320) são decodificados para adquirir uma pluralidade de sinais de alto-falante.

7. Meio não transitório legível por computador, caracterizado por conter um conjunto de instruções que, quando executado em um processador, realizam um método para suprimir um eco acústico, dito método caracterizado pelo fato de que compreende:

extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix (310) e as informações laterais paramétricas (320), em que o sinal de downmix (310) e as informações laterais paramétricas (320) juntos representam um sinal multicanal compreendendo pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix(310);

calcular coeficientes de filtro (350) para filtragem adaptativa com base no sinal de downmix (310) e em um sinal do microfone (340) ou um sinal derivado do sinal do microfone; em que os coeficientes de filtro (350) são calculados determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal derivado do sinal de microfone (720), determinando coeficientes de filtro de estimativa de eco com base no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de potência de referência e nos coeficientes de filtro de estimativa de eco e determinando os coeficientes de filtro (350) com base na estimativa de eco coeficientes de filtro e o segundo espectro de potência de referência;

filtrar adaptativamente o sinal do microfone (340) ou o sinal derivado do sinal do microfone com base nos coeficientes (350) de filtro para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340).

8. Meio não transitório legível por computador, caracterizado por conter um conjunto de instruções que, quando executado em um processador, realizam um método de fornecimento de uma pluralidade de sinais de alto-falante e um sinal de microfone (340), compreendendo:

um método de suprimir um eco acústico, compreendendo:

extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix e as informações

laterais paramétricas (320), em que o sinal de downmix (310) e as informações laterais paramétricas (320) juntos representam um sinal multicanal compreendendo pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix (310);

calcular coeficientes de filtro (350) para filtragem adaptativa com base no sinal de downmix (310) e o sinal de microfone (340) ou um sinal derivado do sinal de microfone, em que os coeficientes de filtro (350) são calculados determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal derivado do sinal de microfone (720), determinando coeficientes de filtro de estimativa de eco com base no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de energia de referência e os coeficientes de filtro de estimativa de eco, e determinando os coeficientes de filtro (350) com base nos coeficientes de filtro de estimativa de eco e o segundo espectro de energia de referência;

filtrar adaptativamente o sinal do microfone (340) ou o sinal derivado do sinal do microfone com base nos coeficientes (350) de filtro para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340);

uma etapa de decodificação multicanal;

uma etapa de receber um sinal de microfone (340), em que, na etapa de decodificação multicanal, o sinal de downmix (310) e a informação lateral paramétrica (320) são decodificados para adquirir uma pluralidade de sinais de alto-falante.

9. Unidade de supressão de eco acústico (210), caracterizada pelo fato de que compreende:

uma interface de entrada (230) para extrair um sinal de downmix (310) de um sinal de entrada (300) que compreende o sinal de downmix (310) e informação lateral paramétrica (320), em que o sinal de downmix (310) e informação lateral paramétrica (320) juntos representam um sinal multicanal com pelo menos mais canais ou um número de canal superior ao número de canais no sinal de downmix (310);

uma calculadora (220) para calcular coeficientes de filtro (350) para um filtro adaptativo (240), em que a calculadora (220) está adaptada para receber o sinal de downmix (310), em que a calculadora (220) está ainda adaptada para receber um microfone sinal (340) ou um sinal derivado do sinal do microfone (720), em que a calculadora (220) é adaptada para determinar os coeficientes de filtro (350) com base nos sinais recebidos;

um filtro adaptativo (240) adaptado para receber os coeficientes de filtro (350) da calculadora (220) e adaptado para filtrar o sinal do microfone (340) ou o sinal derivado do sinal do microfone (720) com base nos coeficientes do filtro (350) para suprimir um eco causado pelo sinal multicanal no sinal do microfone (340),

em que a calculadora (220) está adaptada para determinar os coeficientes de filtro (350) determinando um primeiro espectro de potência de referência com base no sinal de downmix (310), determinando um segundo espectro de potência de referência com base no sinal de microfone (340) ou o sinal

derivado do sinal do microfone (720), determinando coeficientes de filtro de estimativa de eco com base no primeiro e segundo espectros de potência de referência, determinando uma estimativa de eco com base no primeiro espectro de potência de referência e os coeficientes de filtro de estimativa de eco e determinando os coeficientes de filtro (350) com base nos coeficientes de filtro de estimativa de eco e no segundo espectro de potência de referência.

10. UNIDADE DE SUPRESSÃO DE ECO ACÚSTICO (210), de acordo com a reivindicação 9, caracterizada por a calculadora (220) ser adaptada para calcular a primeira referência do espectro de força com base em

$$|P(k, m)|^2 = \sum_{i=0}^{N-1} a_i(k, m) |B_i(k, m)|^2$$

ou

$$|P(k, m)|^2 = \left| \sum_{i=0}^{N-1} a_i(k, m) B_i(k, m) \right|^2,$$

em que $|P(k, m)|^2$ é a primeira referência do espectro de força, $a_i(k, m)$ é um fator de ponderação, $B_i(k, m)$ é um canal i-th do sinal *downmix* (310), em que N é o número dos canais no sinal *downmix* (310), N sendo maior ou igual a 1, em que k é um índice do bloco de tempo e m denota um índice de frequência.

11. UNIDADE DE SUPRESSÃO DE ECO ACÚSTICO (210), de acordo com a reivindicação 9, caracterizada por a calculadora (220) ser adaptada para calcular uma segunda referência do espectro de força com base em

$$|Q(k, m)|^2 = \sum_{i=0}^{K-1} c_i(k, m) |A_i(k, m)|^2$$

ou

$$|Q(k, m)|^2 = \left| \sum_{i=0}^{K-1} c_i(k, m) A_i(k, m) \right|^2,$$

em que $|Q(k, m)|^2$ é a segunda referência do espectro de força, $\mathbf{c}_i(\mathbf{k}, \mathbf{m})$ é um fator de ponderação, $\mathbf{A}_i(\mathbf{k}, \mathbf{m})$ é um canal i-th de um sinal *downmix* (720), em que k é um número de canais no sinal *downmix* (720), k sendo maior ou igual a 1, em que k é um índice de bloco de tempo e m denota um índice de frequência.

12. Unidade de supressão de eco acústico (210), de acordo com a reivindicação 9, caracterizada pelo fato de que a calculadora (220) é ainda adaptada para determinar os coeficientes de filtro de estimativa de eco e a estimativa de eco com base no primeiro espectro de potência de referência em uma versão atrasada, atrasando a primeira potência de referência espectro por um valor de atraso.

13. Unidade de supressão de eco acústico (210), de acordo com a reivindicação 12, caracterizada pelo fato de que a calculadora (220) é ainda adaptada para determinar o valor de atraso determinando um valor de correlação para uma pluralidade de diferentes valores de atraso possíveis, determinando valores de ganho de predição de eco para valores de a pluralidade de diferentes valores de atraso possíveis e determinando o valor da pluralidade de diferentes valores de atraso possíveis como o valor de atraso com um valor máximo dos valores de ganho de predição de eco determinados.

14. Unidade de supressão de eco acústico (210), de acordo com a reivindicação 9, caracterizada pelo fato de que a calculadora (220) é adaptada para determinar um primeiro espectro

de potência modificado com base no primeiro espectro de potência de referência subtraindo um valor médio do primeiro espectro de potência de referência, em que a calculadora (220) é adaptado para determinar um segundo espectro de potência modificado com base no segundo espectro de potência de referência subtraindo um segundo valor médio do segundo espectro de potência de referência, e em que a calculadora (220) está adaptada para determinar os coeficientes de filtro de estimativa de eco com base em o primeiro e o segundo espectros de potência modificados.

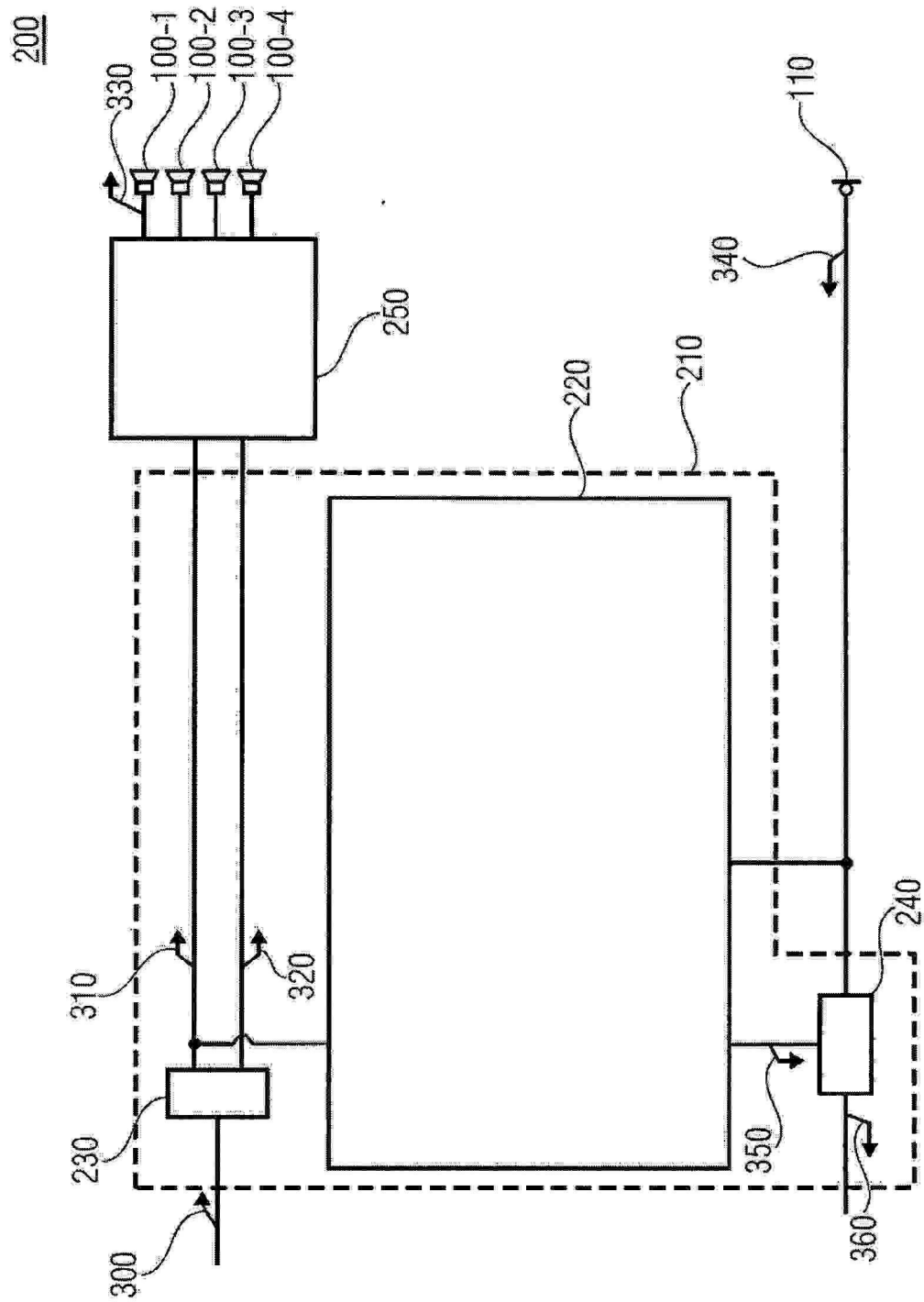


FIG 1

2/9

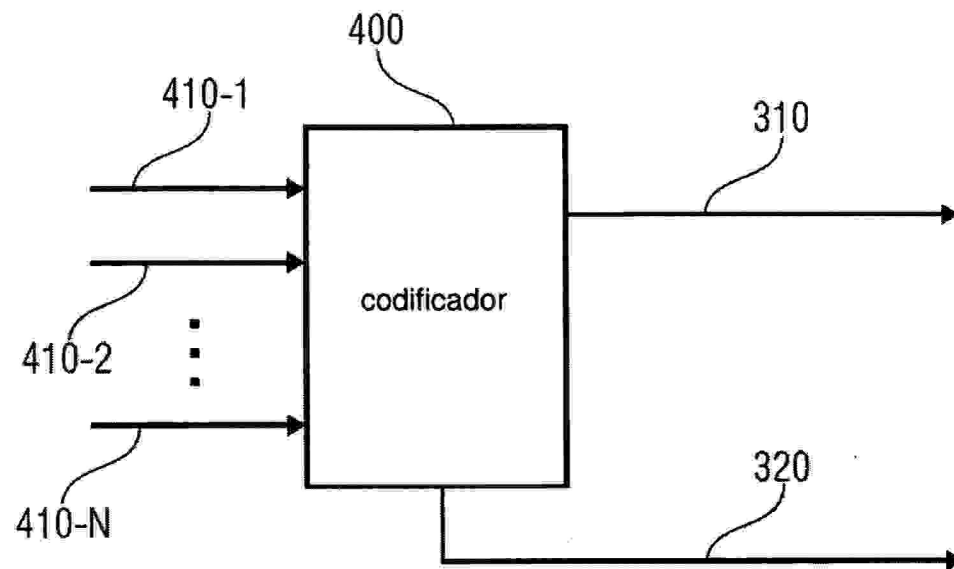


FIG 2

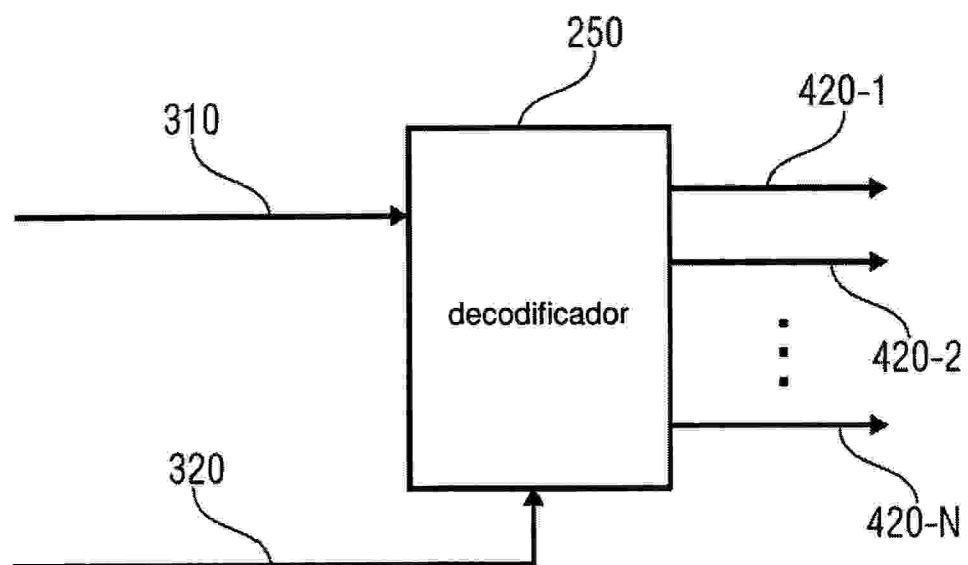


FIG 3

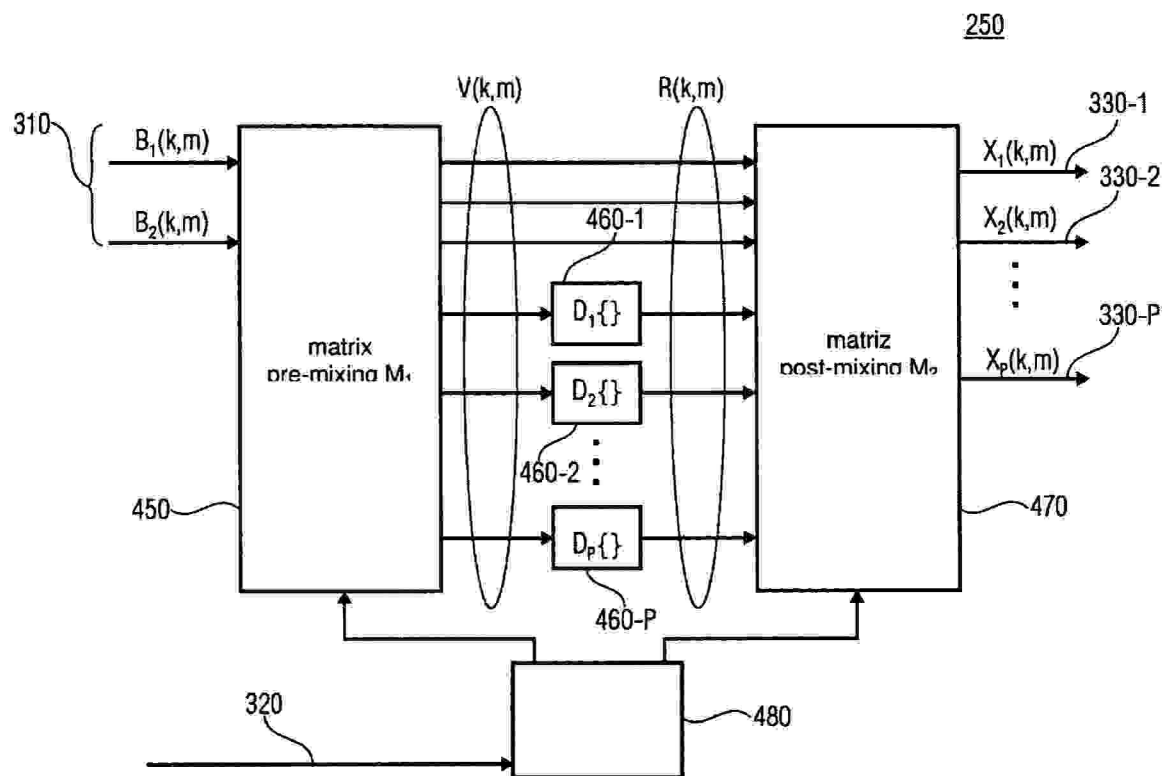


FIG 4

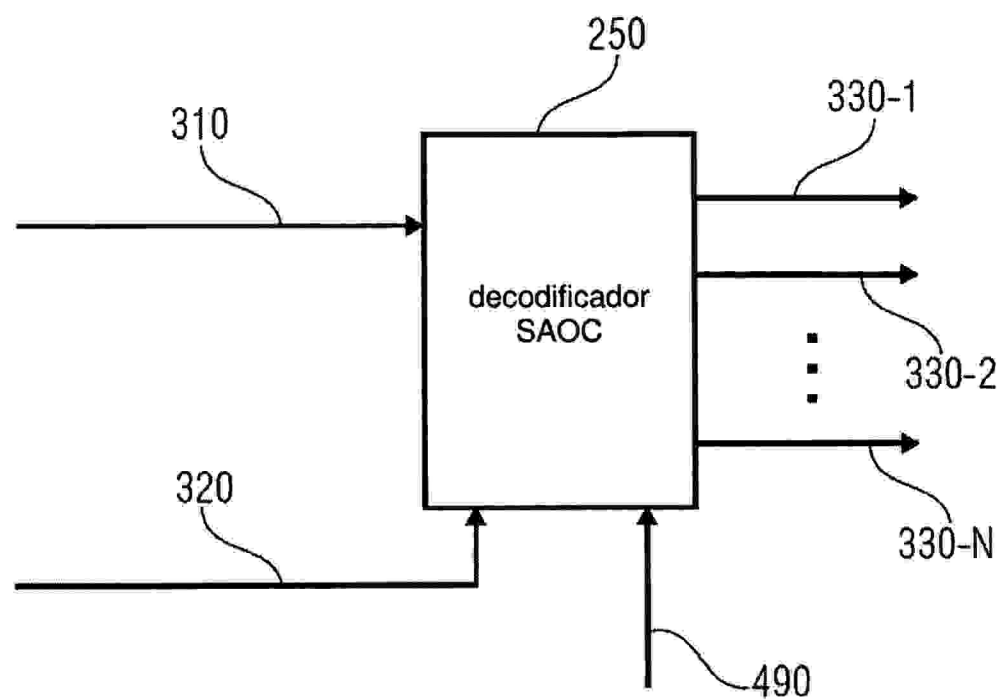


FIG 5

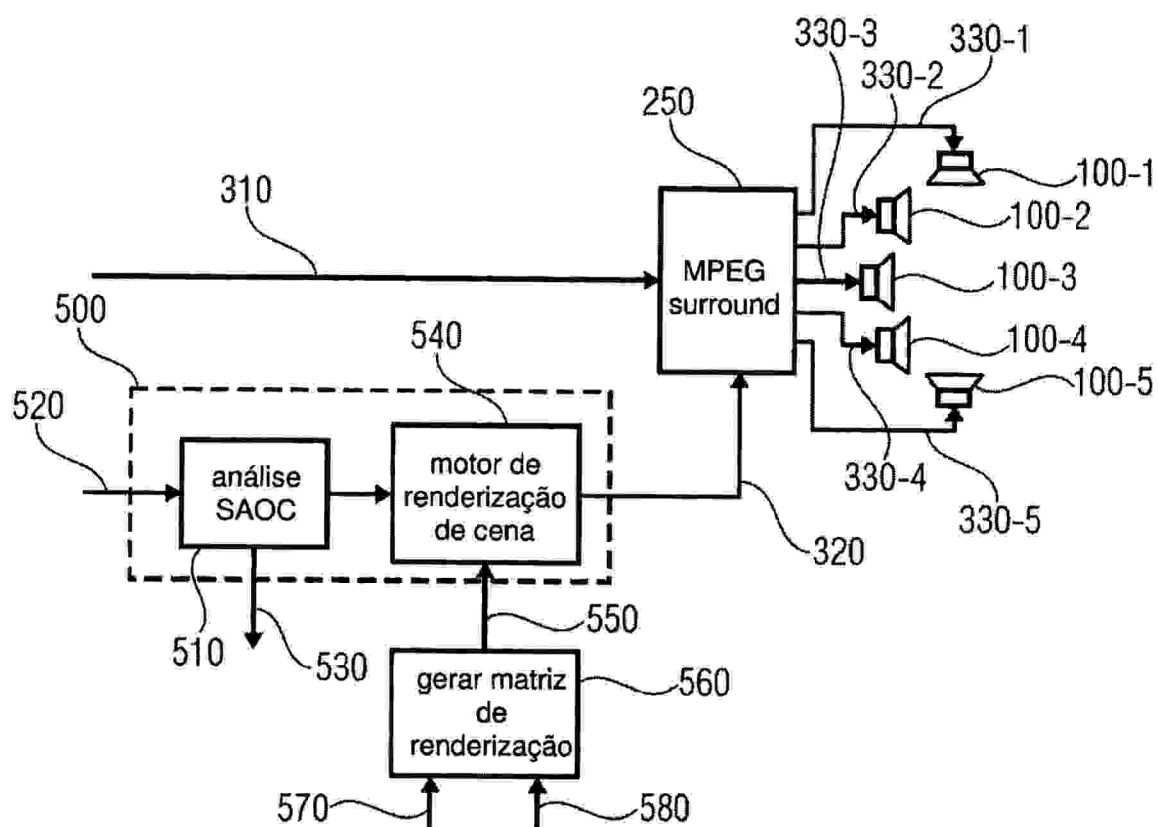


FIG 6A

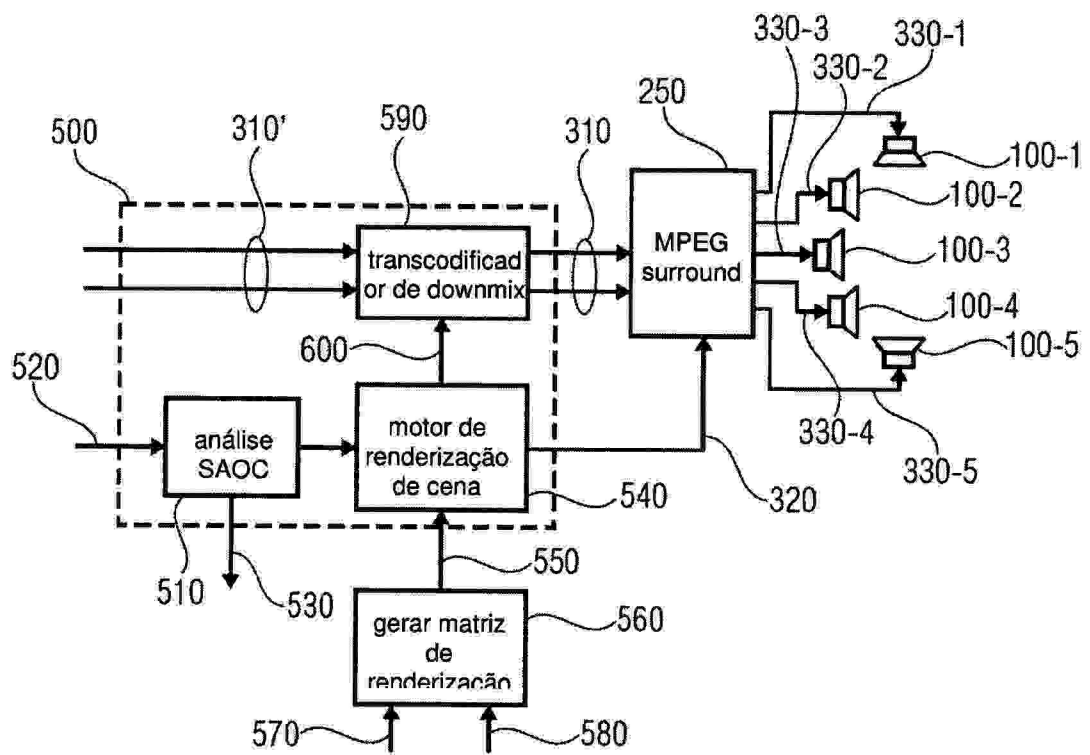


FIG 6B

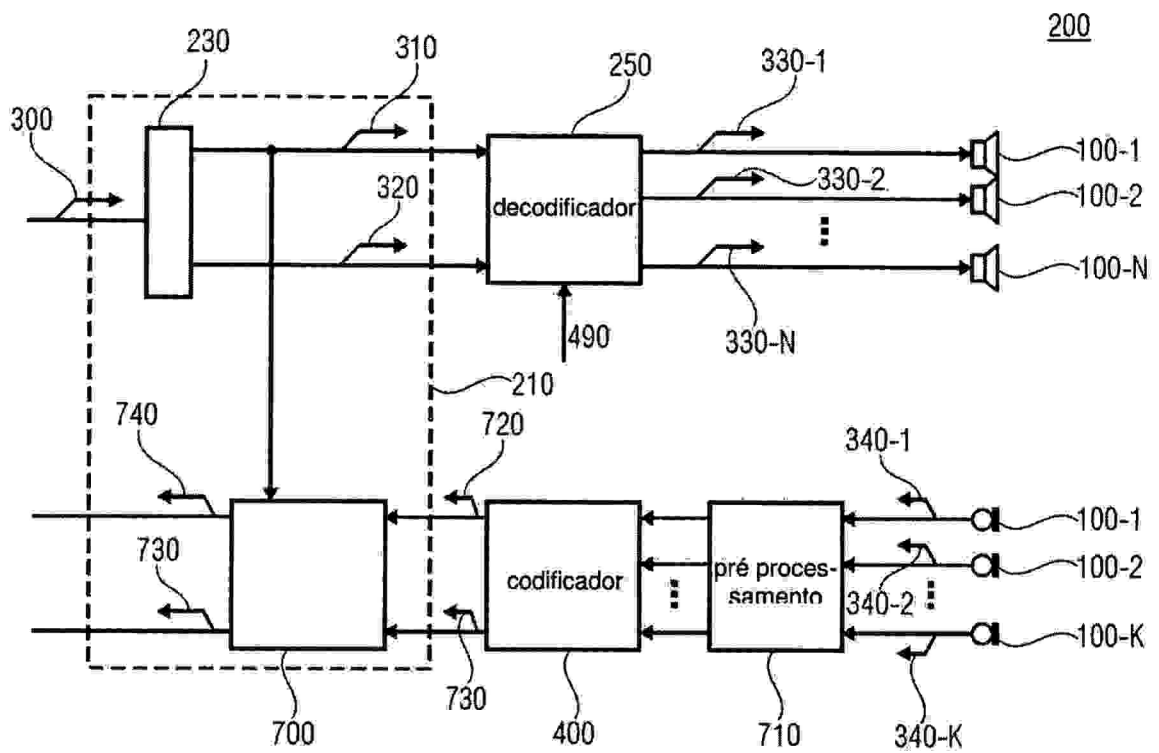


FIG 7

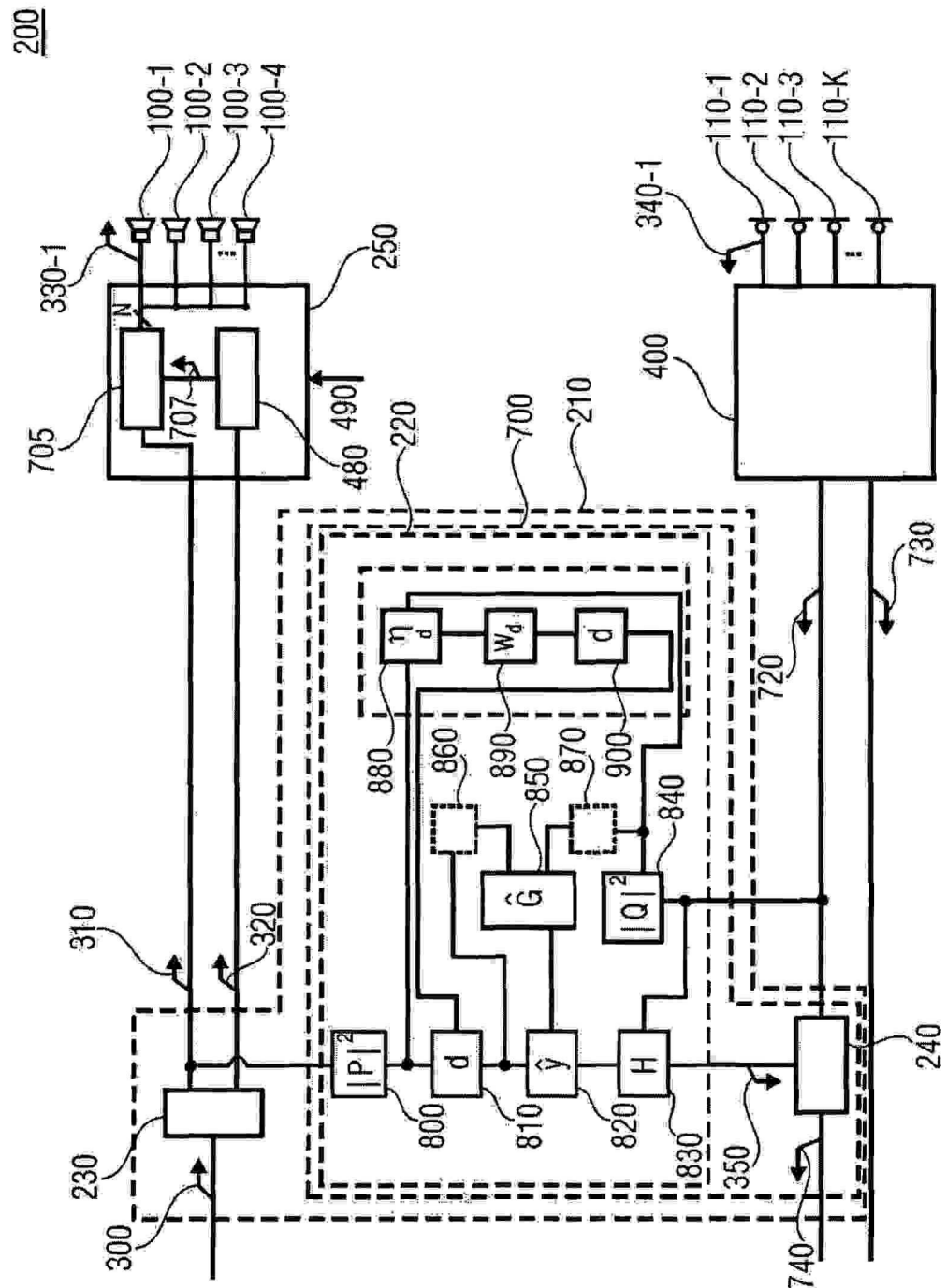


FIG 8

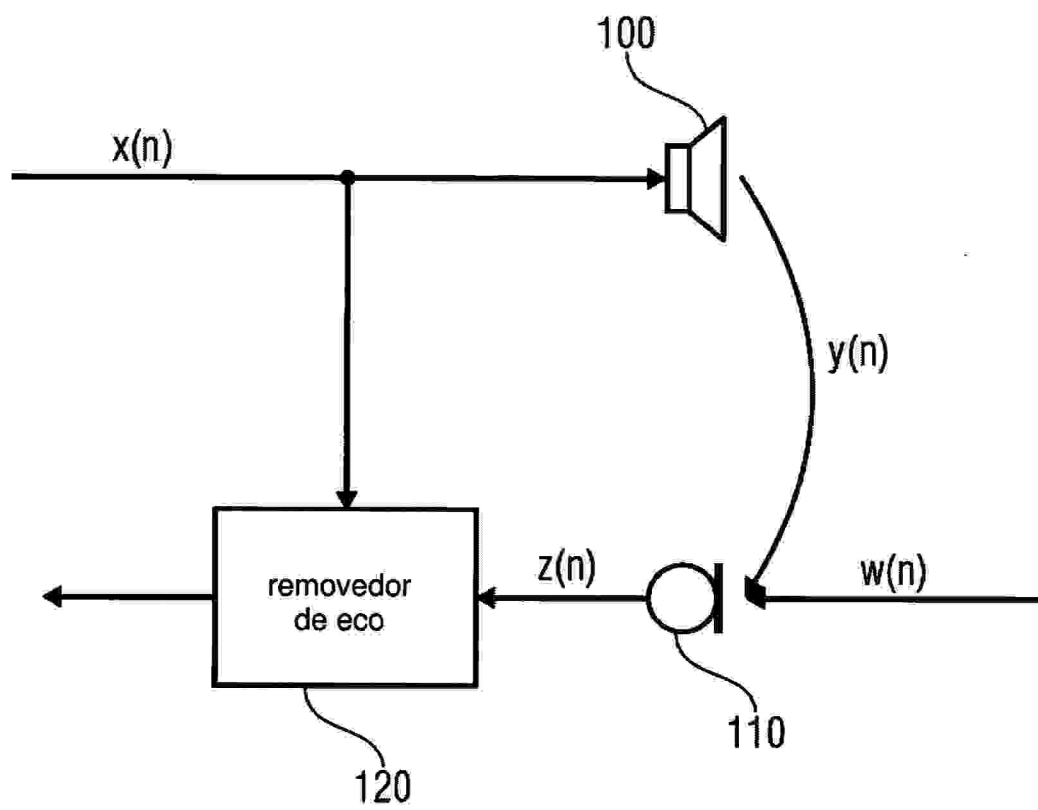


FIG 9