



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

(11) PI 0713236-0 B1



(22) Data do Depósito: 24/04/2007

(45) Data de Concessão: 10/03/2020

(54) Título: CONCEITO PARA COMBINAÇÃO DE MÚLTIPLAS FONTES DE ÁUDIO PARAMETRICAMENTE CODIFICADAS

(51) Int.Cl.: G10L 19/008; H04N 7/15.

(52) CPC: G10L 19/008; H04N 7/15.

(30) Prioridade Unionista: 07/07/2006 US 60/819,419; 24/04/2007 US 11/739,544.

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

(72) Inventor(es): JOHANNES HILPERT; JUERGEN HERRE; KARSTEN LINZMEIER; OLIVER HELLMUTH; THORSTEN KASTNER.

(86) Pedido PCT: PCT EP2007003598 de 24/04/2007

(87) Publicação PCT: WO 2008/003362 de 10/01/2008

(85) Data do Início da Fase Nacional: 07/01/2009

(57) Resumo: CONCEITO PARA COMBINAÇÃO DE MÚLTIPLAS FONTES DE ÁUDIO PARAMETRICAMENTE CODIFICADAS
De acordo com a presente invenção, diversos sinais de áudio parametricamente codificados podem ser eficientemente combinados utilizando um gerador de sinal de áudio (100) que gera um sinal de saída de áudio (120) pela combinação dos canais down-mix (110a,112a) e de parâmetros associados (110b,112b) dos sinais de áudio diretamente dentro do domínio de parâmetro, ou seja, sem reconstruir ou decodificar os sinais individuais de entrada áudio antes da geração do sinal de saída de áudio (120). Isso é realizado pela mixagem direta dos canais down-mix associados (110a,112a) dos sinais de entrada individuais. É uma característica principal da presente invenção que a combinação dos canais down-mix (110a,112a) seja realizada por meio de operações aritméticas simples e baratas do ponto de vista computacional.

"CONCEITO PARA COMBINAÇÃO DE MÚLTIPLAS FONTES DE ÁUDIO PARAMETRICAMENTE CODIFICADAS"

Campo da invenção

A presente invenção refere-se à codificação de áudio multicanais e, em particular, a um conceito de combinação de fluxos de áudio parametricamente codificados de forma flexível e eficiente.

Histórico da invenção e técnica anterior

O recente desenvolvimento na área de codificação de áudio originou diversas técnicas de codificação paramétrica de áudio para codificação conjunta de um sinal de áudio multicanais (por exemplo, 5,1 canais) em um (ou mais) canal down-mix e ainda um fluxo de informações auxiliares. De modo geral, o fluxo de informações auxiliares possui parâmetros relacionados às propriedades dos canais originais do sinal multicanais tanto em relação a outros canais originais do sinal multicanais como em relação ao canal down-mix. A definição particular de parâmetros do canal de referência, aos quais esses parâmetros estão relacionados, depende da implementação específica. Algumas das técnicas conhecidas nesse campo são "binaural cue coding", "spatial áudio coding" e "parametric stereo".

Para mais detalhes sobre essas implementações em particular, faz-se aqui referência a publicações relacionadas. Por exemplo, o *binaural cue coding* é detalhado em:

C. Faller and F. Baumgarte, "Efficient representation of spatial audio using perceptual parametrization," IEEE WASPAA, Mohonk, NY, October 2001; F. Baumgarte and C. Faller, "Estimation of auditory spatial cues for binaural cue coding,"

ICASSP, Orlando, FL, May 2002; C. Faller and F. Baumgarte, "Binaural cue coding: a novel and efficient representation of spatial audio," ICASSP, Orlando, FL, May 2002; C. Faller and F. Baumgarte, "Binaural cue coding applied to audio compression flexible rendering," AES 113th Convention, Los Angeles, Preprint 5686, October 2002; C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," IEEE Trans. on Speech and Audio Proc., vol. 11, no. 6, Nov. 2003, and J. Herre, C. Faller et al., "Spatial Audio Coding: Next-generation efficient and compatible coding of multi-channel audio", Audio Engineering Society Convention Paper, Oct. 28, 2004, San Francisco, CA, USA.

Apesar de o *binaural cue coding* usar múltiplos canais originais, o estéreo paramétrico é uma técnica relacionada para a codificação paramétrica de um sinal estéreo de dois canais resultando em um sinal mono transmitido e informações paramétricas auxiliares, como por exemplo, revistas nas seguintes publicações: J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, "High-Quality Parametric Spatial Audio Coding at Low Bitrates", AES 116th Convention, Berlin, Preprint 6072, May 2004; E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, "Low Complexity Parametric Stereo Coding", AES 116th Convention, Berlin, Preprint 6073, May 2004.

Outras tecnologias se baseiam na multiplexação de números arbitrários de fontes ou objetos de áudio em um único canal de transmissão de áudio. Os esquemas baseados em multiplexação são, por exemplo, apresentados como "apresentação flexível" em publicações relativas ao BCC (binaural cue coding) ou, mais recentemente, por um esquema denominado "joint source

coding" (JSC). As publicações relacionadas são, por exemplo: C. Faller, "Parametric Joint Coding of Audio Sources", Convention Paper 6752, 120th AES Convention, Paris, May 2006. Similar aos esquemas de estéreo paramétrico e *binaural cue coding*, essas técnicas pretendem codificar múltiplos objetos originais de áudio (canais) para transmissão por menos canais down-mix. Pela obtenção adicional de parâmetros baseados em objetos para cada canal de entrada, que podem ser codificados com uma taxa de dados muito baixa e que também são transmitidos para um receptor, esses objetos podem ser separados no lado do receptor e apresentados (mixados) em um determinado número de dispositivos de saída, como por exemplo, fones de ouvido, alto-falantes estéreos de dois canais, ou instalações de alto-falantes multicanais. Esta abordagem permite o ajuste de nível e a redistribuição (*panning*) dos diferentes objetos de áudio em diferentes locais na instalação de reprodução, isto é, no lado do receptor.

Basicamente, essas técnicas operam como transmissor M-k-N, com M sendo o número de objetos de áudio na entrada, k sendo o número de canais down-mix transmitidos, tipicamente $k \leq 2$. N é o número de canais de áudio na saída do apresentador, isto é, por exemplo, o número de alto-falantes. Isto é, $N = 2$ para um apresentador estéreo ou $N = 6$ para uma instalação de falantes multicanais 5.1. Em termos de eficiência de compressão, os valores típicos são, por exemplo, 64 kbps ou menos para um canal down-mix perceptualmente codificado (consistindo de k canais de áudio) e aproximadamente 3 kbps para parâmetros de objeto por objeto de áudio transmitido.

Os cenários de aplicação das técnicas acima são,

por exemplo, a codificação de cenas de áudio espacial relacionadas com produções de filmes e cinema, de maneira a permitir a reprodução espacial do som em um sistema de home theater. Os exemplos comuns são as pistas amplamente conhecidas 5.1 e 7.1 de som surround em mídia de filmes, como DVD e similares. As produções em filmes estão se tornando mais e mais complexas com referência às cenas de áudio, que devem prover uma experiência de audição espacial e, portanto, devem ser mixadas com muito cuidado. Podem ser indicados diferentes engenheiros de som para a mixagem de diferentes fontes surround ou de efeitos sonoros sendo, portanto, desejável a transmissão de cenários multicanais parametricamente codificados entre os engenheiros de som individuais, de maneira a transportar eficientemente os fluxos de áudio dos engenheiros de som individuais.

Outro cenário de aplicação dessa tecnologia é a teleconferência com vários interlocutores em cada ponta de uma conexão ponto a ponto. Para economizar largura de banda, a maior parte das instalações de teleconferência opera com transmissão monofônica. Usando, por exemplo, o *joint source coding* ou uma das demais técnicas de codificação multicanais para transmissão, a redistribuição e o alinhamento de níveis dos diferentes interlocutores na ponta de recepção (cada ponta) podem ser obtidos, e assim, a inteligibilidade e o equilíbrio dos interlocutores são ampliados, gastando-se uma taxa de bits marginalmente aumentada quando comparada com um sistema monofônico. A vantagem da maior inteligibilidade se torna particularmente evidente no caso especial da indicação de cada participante individual da conferência para um único canal (e,

assim, o participante) de uma instalação de falantes multicanais na ponta de recepção. Entretanto, este é um caso especial. No geral, o número de participantes não combina com o número de interlocutores na ponta de recepção. Entretanto, usando a
5 instalação existente de interlocutores, é possível tornar o sinal associado a cada participante, de maneira que pareça estar se originando de qualquer posição desejada. Isto é, o participante individual não somente é reconhecido por sua voz diferente, como também pela localização da fonte de áudio relacionada com o
10 participante interlocutor.

Apesar de as técnicas do estado da arte implementarem conceitos sobre como codificar de maneira eficiente múltiplos canais ou objetos de áudio, todas as técnicas atualmente conhecidas não possuem a possibilidade de combinar dois ou mais
15 desses fluxos de áudio transmitido de maneira eficiente e obterem um fluxo de saída (sinal de saída), que é uma representação de todos os fluxos de entrada de áudio (sinais de entrada de áudio).

O problema surge, por exemplo, quando é considerado um cenário de teleconferência com mais do que duas
20 localizações, cada localização tendo um ou mais interlocutores. Então, é necessária uma instância intermediária para receber os sinais de áudio de entrada das fontes individuais e gerar um sinal de áudio de saída para cada local de teleconferência, tendo somente a informação dos locais restantes da teleconferência. Isto
25 é, a instância intermediária deve gerar um sinal de saída, que é obtido a partir de uma combinação de dois ou mais sinais de áudio de entrada e que permite uma reprodução dos canais individuais de áudio ou objetos de áudio dos dois ou mais sinais de entrada.

Pode ocorrer um cenário similar quando dois engenheiros de áudio de uma produção de filmes - cinema querem combinar seus sinais de áudio espacial de maneira a verificarem a impressão de audição gerada por ambos os sinais. Então, pode ser
5 desejável combinar diretamente dois sinais multicanais codificados para verificar a impressão combinada de audição. Isto é, um sinal combinado deve ser tal que pareça com todos os objetos de áudio (fontes) dos dois engenheiros de áudio.

Entretanto, de acordo com as técnicas anteriores,
10 essa combinação somente é possível por meio da decodificação dos sinais de áudio (fluxos). Então, os sinais decodificados de áudio podem, novamente, ser recodificados pelos codificadores multicanais da técnica anterior, para gerarem um sinal combinado em que todos os canais originais de áudio ou objetos de áudio
15 estão adequadamente representados.

Isto tem a desvantagem de uma alta complexidade computacional, gastando-se então muita energia e tornando, por vezes, impossível a aplicação do conceito, especialmente em cenários de tempo real. Além disso, uma combinação por meio de
20 subsequente decodificação e recodificação de áudio pode provocar um retardo considerável devido às duas etapas de processamento, o que é inaceitável para determinadas aplicações, como em teleconferências/telecomunicações.

Sumário da invenção

25 É objetivo da presente invenção prover um conceito para combinar de forma eficiente múltiplos sinais de áudio codificados parametricamente.

De acordo com um primeiro aspecto da presente

invenção, este objetivo é alcançado por um gerador de sinais de áudio para a geração de um sinal de áudio de saída, o gerador de sinais de áudio compreendendo: um receptor de sinais de áudio para a recepção de um primeiro sinal de áudio compreendendo um primeiro canal down-mix tendo informações sobre dois ou mais primeiros canais originais, e compreendendo um parâmetro original associado a um dos primeiros canais originais, descrevendo uma propriedade de um dos primeiros canais originais com relação a um canal de referência; e um segundo sinal de áudio compreendendo um segundo canal down-mix tendo informações sobre pelo menos um segundo canal original; um combinador de canais para a obtenção de um canal down-mix combinado, combinando o primeiro canal down-mix com o segundo canal down-mix; um calculador de parâmetros para a obtenção de um primeiro parâmetro combinado que descreva as propriedades de um dos primeiros canais originais com relação a um canal comum de referência, e um segundo parâmetro combinado que descreva as propriedades de outro dos primeiros canais originais ou do pelo menos um segundo canal original com relação ao canal comum de referência; e uma interface de saída para enviar o sinal de áudio de saída, compreendendo o canal down-mix combinado, o primeiro e o segundo parâmetros combinados.

De acordo com um segundo aspecto da presente invenção, este objetivo é alcançado por um método de geração de um sinal de áudio de saída, o método compreendendo: receber um primeiro sinal de áudio compreendendo um primeiro canal down-mix tendo informações sobre dois ou mais primeiros canais originais, e compreendendo um parâmetro original associado a um dos primeiros canais originais, descrevendo uma propriedade de um dos primeiros

canais originais com relação a um canal de referência e um segundo sinal de áudio compreendendo um segundo canal down-mix tendo informações sobre pelo menos um segundo canal original; obter um canal down-mix combinado pela combinação do primeiro canal down-mix e do segundo canal down-mix; obter um primeiro parâmetro combinado descrevendo as propriedades de um dos primeiros canais originais com relação a um canal comum de referência e um segundo parâmetro combinado descrevendo as propriedades de outro dos primeiros canais originais ou do pelo menos um segundo canal original com relação a um canal comum de referência; e enviando o sinal de áudio de saída compreendendo o canal down-mix combinado e o primeiro e o segundo parâmetros combinados.

De acordo com um terceiro aspecto da presente invenção, este objetivo é alcançado por uma representação de três ou mais canais de áudio, compreendendo: um canal down-mix combinado, sendo uma combinação de um primeiro canal down-mix tendo informações sobre pelo menos dois primeiros canais originais e um segundo canal down-mix tendo informações sobre pelo menos um segundo canal original; um primeiro parâmetro descrevendo uma propriedade de um dos pelo menos dois primeiros canais originais com relação a um canal de referência; e um segundo parâmetro descrevendo as propriedades de outro canal dos primeiros canais originais ou a propriedade do pelo menos um segundo canal original com relação ao canal de referência.

De acordo com um quarto aspecto da presente invenção, este objetivo é alcançado por um programa de computador que implementa um método para a geração de um sinal de áudio de saída, o método compreendendo: receber um primeiro sinal de áudio

compreendendo um primeiro canal down-mix tendo informações sobre dois ou mais primeiros canais originais, e compreendendo um parâmetro original associado a um dos primeiros canais originais descrevendo uma propriedade de um dos primeiros canais originais com relação a um canal de referência e um segundo sinal de áudio compreendendo um segundo canal down-mix tendo informações sobre pelo menos um segundo canal original; obter um canal down-mix combinado pela combinação do primeiro canal down-mix e do segundo canal down-mix; obter um primeiro parâmetro combinado descrevendo as propriedades de um dos primeiros canais originais com relação a um canal comum de referência e um segundo parâmetro combinado descrevendo as propriedades de outro dos primeiros canais originais ou do pelo menos um segundo canal original com relação a um canal comum de referência; e enviar o sinal de áudio de saída compreendendo o canal down-mix combinado e o primeiro e o segundo parâmetros combinados.

De acordo com um quinto aspecto da presente invenção, este objetivo é alcançado por meio de um sistema de conferência tendo um gerador de sinais de áudio para a geração de um sinal de áudio de saída, compreendendo: um receptor de sinais de áudio para a recepção de um primeiro sinal de áudio compreendendo um primeiro canal down-mix tendo informações sobre dois ou mais primeiros canais originais, e compreendendo um parâmetro original associado a um dos primeiros canais originais descrevendo uma propriedade de um dos primeiros canais originais com relação a um canal de referência; e um segundo sinal de áudio compreendendo um segundo canal down-mix tendo informações sobre pelo menos um segundo canal original; um combinador de

canais para obter um canal down-mix combinado pela combinação do primeiro canal down-mix com o segundo canal down-mix; um calculador de parâmetros para a obtenção de um primeiro parâmetro combinado descrevendo as propriedades de um dos primeiros canais originais com relação a um canal comum de referência, e um segundo parâmetro combinado descrevendo as propriedades de outro dos primeiros canais originais ou do pelo menos um segundo canal original com relação ao canal comum de referência; e uma interface de saída para enviar o sinal de áudio de saída, compreendendo o canal down-mix combinado, o primeiro e o segundo parâmetros combinados.

A presente invenção se baseia no achado que múltiplos sinais de áudio parametricamente codificados podem ser combinados de forma eficiente usando um gerador de sinais de áudio ou combinador de sinais de áudio, que gere um sinal de áudio de saída, combinando os canais down-mix e os parâmetros associados dos sinais de áudio de entrada diretamente dentro do domínio do parâmetro, isto é, sem reconstruir ou decodificar os sinais individuais de áudio de entrada antes da geração do sinal de áudio de saída. Para ser mais específico, isto é obtido pela mixagem direta dos canais down-mix associados dos sinais individuais de entrada, por exemplo, pela soma ou pela formação de uma combinação destes. É uma importante característica da presente invenção que a combinação dos canais down-mix seja obtida por meio de simples operações aritméticas computacionalmente baratas, como por soma.

Isto também é verdade para a combinação dos parâmetros que associam os canais down-mix. Como em geral, pelo menos um subconjunto dos parâmetros associados deverá ser alterado

durante a combinação dos sinais de entrada de áudio, é muito importante que os cálculos realizados para a alteração dos parâmetros sejam simples e, portanto, não precisem de potência computacional significativa, nem que incorram em retardos
5 adicionais, por exemplo, pelo uso de bancos de filtros ou outras operações que envolvam memória.

De acordo com uma configuração da presente invenção, é implementado um gerador de sinais de áudio para a geração de um sinal de áudio de saída para combinar um primeiro e
10 um segundo sinais de áudio, ambos estando parametricamente codificados. Para a geração do sinal de áudio de saída, o gerador de sinais de áudio do invento extrai os canais down-mix dos sinais de entrada de áudio e gera um canal down-mix combinado, pela formação de uma combinação linear de dois canais down-mix. Isto é,
15 os canais individuais são adicionados com pesos adicionais aplicados.

Em uma configuração preferida da presente invenção, os pesos aplicados são obtidos por meio de operações aritméticas extremamente simples, por exemplo, usando o número de
20 canais representado pelo primeiro sinal de áudio e pelo segundo sinal de áudio como base de cálculo.

Em outra configuração preferida, o cálculo de pesos é feito na suposição de que cada canal original de áudio dos sinais de entrada contribui para a energia total de sinais com a
25 mesma quantidade. Isto é, os pesos aplicados são simples relações entre os números de canais dos sinais de entrada e o número total de canais.

Em outra configuração preferida da presente

invenção, os pesos dos canais down-mix individuais são calculados com base na energia contida nos canais down-mix, de maneira a permitir uma reprodução mais autêntica do canal down-mix combinado incluído no sinal de saída de áudio gerado.

5 Em outra configuração preferida da presente invenção, o esforço computacional é ainda reduzido, já que são alterados somente os parâmetros associados a um dos dois sinais de áudio. Isto é, os parâmetros do outro sinal de áudio são transmitidos inalterados, portanto, não precisando de computação e
10 assim, minimizando a carga no gerador de sinais de áudio do invento.

Nos parágrafos a seguir, o conceito do invento será detalhado principalmente para um esquema de codificação que usa o *joint source coding* (JSC). Nesse sentido, a presente
15 invenção estende esta tecnologia para a conexão de múltiplos transceptores monofônicos ou habilitados em JSC para estações remotas, misturando os sinais down-mix JSC e as informações de objeto dentro do domínio do parâmetro. Como mostraram as considerações acima, o conceito do invento não está restrito ao
20 uso da codificação JSC, mas também pode ser implementado com a codificação BCC ou outros esquemas de codificações multicanais, como o *MPEG spatial audio coding* (MPEG Surround) e similares.

Como o conceito do invento será detalhado principalmente usando a codificação JSC, a codificação JSC será
25 revista brevemente nos parágrafos a seguir, de maneira a indicar mais claramente a flexibilidade do conceito do invento e as melhorias que podem ser obtidas em relação à técnica anterior, quando é aplicado o conceito do invento aos esquemas de

codificações de áudio multicanais existentes.

Breve descrição dos desenhos

A Fig.1 mostra um exemplo de um esquema de codificação JSC;

5 A Fig. 2 mostra um exemplo de um apresentador JSC;

A Fig. 3 mostra um cenário de teleconferência com duas localizações;

10 A Fig. 4 mostra um cenário de teleconferência com três localizações;

A Fig. 5 mostra um exemplo de teleconferência usando um gerador de sinais de áudio do invento;

A Fig. 6 mostra outro exemplo de teleconferência usando um gerador de sinais de áudio do invento;

15 A Fig. 6b mostra a compatibilidade para trás do conceito do invento; e

A Fig. 7 mostra um exemplo de um gerador de sinais de áudio do invento.

20 Para a explicação da codificação JSC, será feita referência a seguir às Figuras 1 e 2. Nas figuras seguintes, componentes funcionalmente idênticos compartilham as mesmas marcas de referência, indicando que componentes individuais que proporcionam a mesma funcionalidade podem ser intercambiados entre as configurações individuais da presente invenção sem perderem ou
25 terem funcionalidade restrita, e sem limitar o escopo da presente invenção.

A Fig. 1 mostra um diagrama de blocos do *joint source coding scheme*, um codificador correspondente 2 um

decodificador correspondente 4.

O codificador 2 recebe entradas discretas de áudio $s_i(n)$ 6a, 6b, e 6c e cria um sinal down-mix $s(n)$ 8, por exemplo, somando as formas de ondas.

5 Além disso, um extrator de parâmetros 10 dentro do codificador 2 extrai informações paramétricas auxiliares de cada objeto simples(sinal 6a, 6b, e 6c). Apesar de não ser mostrado na Fig. 1, o sinal down-mix 8 pode ainda ser comprimido por um codificador de fala ou de áudio, sendo transmitido com as
10 informações paramétricas auxiliares adjacentes ao decodificador JSC 4. Um módulo de síntese 12 dentro do decodificador 4 regenera as estimativas 14a, 14b, e 14c ($\hat{s}_i(n)$) dos objetos de entrada (canais 6a, 6b, e 6c).

Para reconstruir as estimativas 14a, 14b, e 14c,
15 sendo perceptualmente similares aos objetos discretos de entrada (canais de entrada) 6a, 6b, e 6c, devem ser extraídas as adequadas informações paramétricas auxiliares de cada canal. Como os canais individuais são somados para a geração do sinal down-mix 8, as taxas de potência entre os canais são essas quantidades adequadas.
20 Portanto, as informações paramétricas dos diferentes objetos ou canais consistem de taxas de potência Δp de cada objeto com relação ao primeiro objeto (objeto de referência).

Essas informações são obtidas no domínio de frequência ou em bandas de frequência não igualmente espaçadas
25 (sub-bandas) correspondentes à resolução crítica de banda da percepção de audição humana. Este é um conceito descrito em mais detalhes, por exemplo, em: J. Blauert, "Spatial Hearing: The Psychophysics of Human Sound Localization", The MIT Press,

Cambridge, MA, revised edition 1997.

Isto é, os canais de entrada de áudio de banda larga são filtrados em várias bandas de frequências com larguras finitas de bandas e para cada uma das bandas individuais de frequência, são feitos os seguintes cálculos. Como já mencionado, a potência de banda do primeiro objeto (objeto de referência ou canal de referência) atua como valor de referência.

$$\Delta p_i(n) = 10 \log_{10} \frac{E\{s_i^2(n)\}}{E\{s_1^2(n)\}}, \quad i = 2 \dots M \quad \text{Equação 1}$$

Para evitar outras introduções de problemas, por exemplo, introduzidos por uma divisão por zero, essas taxas de potência (em representação logarítmica) podem ainda estar limitadas a um máximo de, por exemplo, 24 dB em cada sub-banda. A taxa de potência pode ainda ser quantizada antes da apresentação, para uma maior economia de largura de banda de transmissão.

Não é necessário transmitir explicitamente a potência do primeiro objeto. Ao invés disso, este valor pode ser obtido a partir da suposição de que para objetos estatisticamente independentes, a soma das potências dos sinais sintetizados $\hat{s}_i(n)$ é igual à potência do sinal down-mix $s(n)$. Em termos de expressão matemática, significa:

$$E\{s^2(n)\} = \sum_{i=1}^M E\{\hat{s}_i^2(n)\} \quad \text{Equação 2}$$

Com base nesta suposição e nesta equação, as potências de sub-banda do primeiro objeto (o objeto de referência ou o canal de referência) podem ser reconstruídas, como será mais bem descrito abaixo ao detalhar o conceito do invento.

Para resumir, um sinal de áudio ou fluxo de áudio

de acordo com JSC, compreende um canal down-mix e os parâmetros associados, os parâmetros descrevendo taxas de potência dos canais originais com relação a um canal original de referência. Pode-se notar que este cenário pode facilmente ser alterado quando os
5 outros canais selecionados são o canal de referência. Por exemplo, o próprio canal down-mix pode ser o canal de referência, exigindo a transmissão de outro parâmetro, relacionando a potência do primeiro, canal de referência inicial, com a potência do canal down-mix. Também, o canal de referência pode ser escolhido como
10 variando, já que o canal que tem a maior potência é selecionado como sendo o canal de referência. Assim, como a potência nos canais individuais pode variar com o tempo, o canal de referência também pode variar com o tempo. Também, devido ao fato de que todo o processamento é feito tipicamente na forma seletiva de
15 frequência, o canal de referência pode ser diferente para diferentes bandas de frequência.

A Fig. 2 mostra outro esquema ampliado de codificação JSC, com base no esquema da Fig. 1. As características detalhadas com relação à Fig. 1 estão anexadas na armazenagem ou
20 na caixa de transmissão 20, recebendo os canais de entrada 6 para serem codificados e enviando estimativas 14 dos canais de entrada 6. O esquema da Fig. 2 é ampliado por ainda compreender um mixer 22 que recebe as estimativas. Isto é, os objetos sintetizados 14 não são enviados como simples sinais de áudio
25 diretamente, mas apresentados a N canais de saída no módulo do mixer. Esse mixer pode ser implementado de diferentes formas, por exemplo, recebendo novos parâmetros de mixagem 24 como entrada, para conduzir a mixagem dos objetos sintetizados 14. Somente como

exemplo, pode-se considerar um cenário de teleconferência, em cada um dos canais de saída 26 é atribuído a um participante da conferência. Portanto, um participante na ponta de recepção tem a possibilidade de virtualmente separar os demais participantes
 5 indicando suas vozes a posições individuais. Assim, não somente a voz pode servir como critério para distinguir entre diferentes participantes da conferência telefônica, como também a direção em que o ouvinte recebe a voz de um participante. Além disso, um ouvinte pode dispor do canal de saída de maneira que todos os
 10 participantes do mesmo local de teleconferência estejam agrupados na mesma direção, ampliando ainda mais a experiência perceptual.

Como mostrado na Fig. 2, $s_1(n) \dots s_M(n)$ denota os objetos discretos de áudio na entrada do codificador JSC. Na saída do decodificador JSC $\hat{s}_1(n) \dots \hat{s}_M(n)$ representa os objetos de áudio
 15 ,virtualmente' separados que são enviados ao mixer. Os parâmetros de mixagem 24 podem ser interativamente modificados no lado do receptor para colocar os diferentes objetos em um estágio de som que é reproduzido pelos canais de saída $\hat{x}_1(n) \dots \hat{x}_N(n)$.

A Fig. 3 mostra a aplicação de esquemas de
 20 codificação de áudio multicanais a um cenário de teleconferência básico, ocorrendo entre duas localizações. Aqui, uma primeira localização 40 comunica-se com uma segunda localização 42. A primeira localização pode ter A participantes, isto é, A objetos de áudio, a segunda localização tem B participantes ou objetos de
 25 áudio. Para teleconferência ponto a ponto, a tecnologia descrita da codificação JSC pode ser aplicada para frente de maneira a transmitir sinais de áudio de múltiplos objetos em cada localização para a estação remota correspondente. Isto é, (A-1)

parâmetros a_i e um down-mix associado são transferidos para a localização 42. Na direção oposta, (B-1) parâmetros b_i são transmitidos em conjunto com um down-mix associado para a localização 40.

5 Para teleconferência com mais que dois pontos extremos, a situação é completamente diferente, como ilustrada na Fig. 4.

 A Fig. 4 mostra, além das localizações 40 e 42 uma terceira localização 44. Como pode ser visto na Fig. 4, este
10 cenário precisa de um distribuidor central para os sinais associados de áudio, geralmente chamados de *multi point control unit*, MCU. Cada uma das localizações (sites) 40, 42 e 44 está conectada ao MCU 46. Para cada site 40, 42 e 44, existe um único montante ao MCU que contém o sinal do site. Como cada site
15 individual precisa receber os sinais dos demais sites, a jusante de cada site 40, 42 e 44 existe uma mistura dos sinais dos outros sites, excluindo o próprio sinal do site, que também é denominado como o sinal (N-1). Em geral, para preencher as exigências da instalação e para manter a largura da banda de transmissão
20 razoavelmente baixa, não é viável transmitir N-1 fluxos codificados JSC a partir da MCU para cada site. Esta seria, é claro, uma opção para frente.

 A abordagem do estado da arte para a obtenção dos down-streams individuais é a ressintetização de todos os fluxos de
25 entrada (objetos) dentro da MCU 46 usando um decodificador JSC. Então, os objetos de áudio ressintetizados poderiam ser reagrupados e recodificados, de maneira a fornecer a cada site fluxos de áudio compreendendo os desejados objetos de áudio ou

canais de áudio. Mesmo neste cenário simples, isto significaria três tarefas de decodificação e três de codificação, que devem ser feitas simultaneamente na MCU 46. Apesar das significativas demandas de computação, também podem ser esperados problemas audíveis com este processo paramétrico de "codificação em tandem" (codificação/decodificação repetidas). O aumento do número de sites aumentaria ainda mais o número de fluxos, e assim o número de processos de codificação ou de decodificação necessários, tornando nenhuma das abordagens para frente realizáveis em cenários de tempo real.

De acordo com a presente invenção, portanto, é desenvolvido um esquema para mixar diferentes fluxos parametricamente codificados (fluxos JSC neste exemplo em particular) diretamente dentro do down-mix e do domínio do parâmetro de objeto para este cenário tipo MCU, criando os desejados sinais de saída (fluxos de áudio de saída) com um mínimo de esforço computacional e de perda de qualidade.

Dentro dos parágrafos seguintes, é detalhado o conceito do invento de mixagem direta dos fluxos de áudio multicanais codificados parametricamente dentro do domínio do parâmetro para os fluxos de áudio com codificação JSC.

O conceito do invento é explicado com a combinação de dois sinais originais de áudio (fluxos) em um sinal de saída. Unir três ou mais fluxos em um conjunto pode ser facilmente obtido a partir do caso da combinação de dois fluxos. As seguintes considerações matemáticas são ilustradas na Fig. 5, mostrando um caso em que três canais de áudio do site A devem ser combinados com quatro canais de áudio do site B. Isto é,

claramente, somente um exemplo para a visualização do conceito do invento. Ao usar a codificação JSC, o site 50 (A) tendo três participantes de conferência (interlocutores) 52a a 52c gerando sinais s_{Ax} , transmite um fluxo de áudio ou um sinal de áudio 54. O
 5 sinal de áudio 54 tem um canal down-mix s_A e parâmetros a_2 e a_3 , referentes à potência dos canais 52b e 52c com relação à potência do canal 52a. De forma equivalente, o site 56 (B) transmite um sinal de áudio 58 tendo um canal down-mix s_B e três parâmetros b_2 , b_3 , e b_4 , sendo a representação codificada JSC de quatro
 10 interlocutores 60a a 60d. A MCU 46 combina os sinais de áudio 54 e 58 para obter o sinal de saída 62 tendo um canal down-mix combinado s_Y e 6 parâmetros $y_2, \dots, y_7$.

No lado de recepção, o receptor 64 decodifica o sinal de saída 62 para obter representações dos 7 objetos de áudio
 15 ou canais de áudio dos sites 50 e 56.

Em termos gerais, o objetivo é formar uma única representação combinada 62 de dois fluxos JSC 54 e 58, cada uma representando um número de objetos por um sinal comum down-mix s_Y e um conjunto de parâmetros objeto caracterizando os objetos. De
 20 forma ideal, a representação combinada JSC será idêntica à que seria obtida pela codificação de todo o conjunto de sinais fonte originais subjacente a ambos os fluxos JSC em um único fluxo JSC em uma etapa.

Para manter simples as equações a seguir, supomos
 25 que as taxas de potência relativas da Equação 1 não estão disponíveis no domínio logarítmico, mas somente as taxas de potência. Cada parâmetro objeto $r_i(n)$ de um determinado objeto i pode ser obtido como

$$r_i(n) = \frac{E\{s_i^2(n)\}}{E\{s_i^2(n)\}} \quad \text{Equação 3}$$

A transposição no domínio logarítmico pode ser posteriormente aplicada a cada parâmetro para permitir a quantização usando uma escala de potência logarítmica.

Supõe-se que todos os sinais abaixo sejam decompostos em uma representação de sub-banda e, portanto, cada cálculo é aplicado separadamente a cada sub-banda.

Temos o fluxo A com seu sinal down-mix s_A e os parâmetros (taxas de potência relativas) para U objetos $a_2...a_U$. O fluxo B consiste do sinal down-mix s_B e dos parâmetros para V objetos $b_2...b_V$.

O sinal down-mix combinado s_Y pode ser formado como uma combinação linear de ambos os sinais down-mix s_A e s_B . Para garantir o correto nivelamento de volume das diferentes contribuições de objetos, podem ser aplicados os fatores de ganho g_A e g_B .

$$s_Y = g_A \cdot s_A + g_B \cdot s_B$$

$$\text{com} \quad g_A = \frac{U}{(U+V)}, \quad g_B = \frac{V}{(U+V)}$$

Esse tipo de medição pode ser significativo se fontes sonoras simples de igual potência média tiverem sido somadas e normalizadas em toda a escala do caminho down-mix.

De forma alternativa, pode-se usar uma abordagem de preservação de potência para os fatores de ganho com

$$g_A = \sqrt{\frac{U}{(U+V)}}, \quad g_B = \sqrt{\frac{V}{(U+V)}}$$

Outra possibilidade é escolher o fator de ganho

de maneira que ambos os sinais down-mix contribuam com a mesma energia média para o down-mix combinado, isto é, escolhendo

$$\frac{g_B}{g_A} = \sqrt{\frac{E\{s_A^2(n)\}}{E\{s_B^2(n)\}}}.$$

Os parâmetros de objeto y_i do fluxo combinado s_y representarão todos os objetos $U + V$.

5 Como os parâmetros associados aos canais down-mix são taxas de potência relativas, os parâmetros $a_2, \dots, a_U$ podem ser usados no estado (inalterados) e os parâmetros dos objetos de B podem ser concatenados com os parâmetros $a_2, \dots, a_U$. Uma vez sendo escolhido o objeto do sinal A como sendo o objeto de referência ou
10 canal de referência, os parâmetros originais b_i devem ser transformados para se relacionarem com aquele canal de referência. Pode ser notado que somente os parâmetros de um fluxo devem ser recalculados, reduzindo ainda mais a carga de computação dentro de uma MCU 46.

15 Pode ainda ser notado que não há necessidade de usar o canal de referência de um dos fluxos de áudio originais como o novo canal de referência. O conceito do invento de combinar fluxos de áudio codificados parametricamente dentro do domínio do parâmetro pode também ser muito bem implementado com outros canais
20 de referência, escolhidos a partir do número de canais originais de sites A ou B. Outra possibilidade seria usar o canal down-mix combinado como o novo canal de referência.

 Seguindo esta abordagem de usar o canal de referência original do site A como o novo canal de referência
25 (canal de referência combinado), a energia (potência) do primeiro objeto (canal) de cada sinal A e B deve ser calculada primeiro, já

que estão somente implicitamente disponíveis.

A preservação de potência do sinal down-mix A, supondo-se fontes estatisticamente independentes, fornece:

$$E\{s_A^2(n)\} = \sum_{i=1}^U E\{\hat{s}_i^2(n)\}.$$

As potências de sinal $E\{s_{A_2}^2(n)\} \dots E\{s_{A_U}^2(n)\}$ são
5 definidas por suas potências relativas $a_2 \dots a_U$ a $E\{s_{A_1}^2(n)\}$:

$$E\{s_{A_2}^2(n)\} = a_2 \cdot E\{s_{A_1}^2(n)\}$$

$$E\{s_{A_3}^2(n)\} = a_3 \cdot E\{s_{A_1}^2(n)\}$$

$$E\{s_{A_U}^2(n)\} = a_U \cdot E\{s_{A_1}^2(n)\}$$

Isto leva à potência de s_{A_1} como:

$$E\{s_{A_1}^2(n)\} = \frac{E\{s_A^2(n)\}}{(1 + a_2 + a_3 + \dots + a_U)}$$

Aplicando-a ao sinal down-mix s_B , podemos
calcular a potência do objeto s_{B_1} como:

$$E\{s_{B_1}^2(n)\} = \frac{E\{s_B^2(n)\}}{(1 + b_2 + b_3 + \dots + b_V)}$$

Agora podemos construir o novo conjunto de
10 parâmetros para todos os objetos de sinal s_y :

y_1 : (não transmitido, objeto de referência,
implicitamente disponível)

$$y_2 = a_2$$

$$y_3 = a_3$$

...

$$y_U = a_U$$

$$y_{U+1} = \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

(taxa de potência do primeiro objeto de sinal B com relação ao objeto de referência A1)

$$y_{U+2} = b_2 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

(taxa de potência do segundo objeto de sinal B renormalizado para a potência do objeto de referência A1)

$$y_{U+3} = b_3 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

...

$$y_{U+v} = b_v \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

5 Como já mostrado pelos parágrafos anteriores, o conceito do invento permite a geração de um fluxo combinado de áudio usando somente operações aritméticas simples sendo, portanto, de extrema eficiência computacional. Assim, a combinação de múltiplos fluxos de áudio codificados parametricamente pode ser
10 feita em tempo real.

Para melhor enfatizar a grande flexibilidade do conceito do invento, a Fig. 6 mostra como um sinal monofônico 70, emitido por um único alto-falante no site 56 pode ser combinado, de acordo com o invento, a dois ou mais sinais codificados JSC de
15 alto-falantes no site 50. Isto é, devido à flexibilidade do conceito do invento, sinais monofônicos de sistemas arbitrários de teleconferência podem ser combinados, de acordo com o invento, com fontes (multi-objetos) multicanais codificadas parametricamente para gerarem um sinal de áudio com codificação JSC representando

todos os canais originais de áudio (objetos).

Ampliando a compatibilidade também para estações remotas que não podem transmitir objetos JSC, mas sinais monofônicos tradicionais, esta técnica é também aplicável para
5 inserir um objeto monofônico, por exemplo, a partir de um dispositivo legado de conferência no fluxo baseado no objeto.

O exemplo acima com o fluxo JSC A (down-mix s_A , parâmetros $a_2...a_U$) e um objeto monofônico C (down-mix s_C) leva a um sinal combinado Z com o sinal down-mix

$$s_Z = g_A \cdot s_A + g_C \cdot s_C$$

10 Com os fatores de ganho discutidos anteriormente e seus parâmetros de objetos:

y_1 : não transmitido (canal de referência, implicitamente disponível)

$$y_2 = a_2$$

$$y_3 = a_3$$

...

$$y_U = a_U$$

$$y_{U+1} = \frac{g_C^2 \cdot E\{s_C^2(n)\}}{g_A^2 \cdot E\{s_{A_1}^2(n)\}},$$

(taxa de potência de sinal C com relação ao
15 objeto de referência A1)

O exemplo supramencionado de transcodificação / fusão de dois fluxos JSC depende da representação da potência dos objetos como dados na Equação 1. Não obstante, o mesmo esquema de invento pode ser também aplicado a outras formas de representação
20 dessas informações.

A Fig. 6b enfatiza novamente a grande

flexibilidade do conceito do invento que incorpora uma fonte monofônica de áudio. A Fig. 6b se baseia no cenário multicanais da Fig. 4, mostrando também a facilidade com que um codificador monofônico de áudio da técnica anterior na fonte de áudio C(44) pode ser integrado em uma conferência de áudio multicanais usando a MCU 46 do invento.

Como mencionado anteriormente, o conceito do invento não se restringe à codificação JSC, tendo um canal de referência fixo predeterminado. Portanto, em um exemplo alternativo, a taxa de potência pode ser computada com relação a um canal de referência, que varia com o tempo, o canal de referência sendo o canal de maior energia dentro de um dado intervalo predeterminado de tempo.

Ao invés de normalizar os valores de potência do sinal em banda com a potência da banda correspondente de um canal de referência fixo (objeto) e transpor o resultado para o domínio logarítmico (dB) como definido pela Equação 1, a normalização pode ocorrer relativa à potência máxima em todos os objetos de uma determinada banda de frequência:

$$pnorm_i(n) = \frac{E\{s_i^2(n)\}}{\max_i(E\{s_i^2(n)\})}, \quad i=1...M \quad \text{Equação 4}$$

Esses valores normalizados de potência (que são dados em uma representação linear) não precisam de nenhuma outra limitação até um determinado limite superior já que, de forma inata, podem somente ter valores entre 0 e 1. Essa vantagem traz a desvantagem de ter que transmitir outro parâmetro para o canal de referência não mais conhecido previamente.

O processo de mixagem desse cenário incluiria as

seguintes etapas (que, novamente, devem ser realizadas para cada sub-banda, separadamente):

5 Temos o fluxo A com seu sinal down-mix s_A e dos parâmetros (valores de potência normalizada, Equação 3, Equação 1) para U objetos $a_1...a_U$.

O fluxo B consiste do sinal down-mix s_B e dos parâmetros para V objetos $b_1...b_V$.

Pode ser formado um sinal down-mix combinado de acordo com uma das opções já mostradas:

$$s_Y = g_A \cdot s_A + g_B \cdot s_B$$

10 Todos os valores de potência normalizada da representação combinada y_i devem ser colocados em relação ao objeto com a maior potência entre todos os objetos de sinal Y. Existem dois candidatos para serem este 'objeto máximo' de Y, seja o objeto máximo de A ou o objeto máximo de B, ambos podendo ser
15 identificados como tendo uma taxa de potência normalizada de '1'.

Essa decisão pode ser tomada comparando a potência absoluta de ambos os candidatos. Novamente, podemos usar a relação com a potência dos sinais down-mix (Equação 2) para obter:

$$E\{s_{A_{\max}}^2(n)\} = \frac{E\{s_A^2(n)\}}{\sum_{i=1}^U a_i} \text{ e } E\{s_{B_{\max}}^2(n)\} = \frac{E\{s_B^2(n)\}}{\sum_{i=1}^V b_i}$$

20 Agora, podemos comparar as potências máximas de objeto pesadas com os fatores de ganho do processo down-mix:

$$g_A^2 \cdot E\{s_{A_{\max}}^2(n)\} > g_B^2 \cdot E\{s_{B_{\max}}^2(n)\}?$$

Seja qual for a maior potência de objeto, este objeto servirá como 'objeto máximo' para os parâmetros combinados

y_i .

Como exemplo, seja a_2 o objeto de potência máxima geral $a_{\max}$ de ambos os sinais A e B, então todos os demais parâmetros podem ser combinados como:

$$y_1 = a_1$$

$$y_2 = a_2$$

...

$$y_U = a_U$$

$$y_{U+1} = b_1 \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B\max}^2(n)\}}{E\{s_{A\max}^2(n)\}},$$

5

(taxa de potência do primeiro objeto de sinal B com relação ao 'objeto máximo', aqui a_2)

$$y_{U+2} = b_2 \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B\max}^2(n)\}}{E\{s_{A\max}^2(n)\}}$$

...

$$y_{U+V} = b_V \cdot \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B\max}^2(n)\}}{E\{s_{A\max}^2(n)\}}$$

Para esse exemplo, todos os parâmetros dos objetos de A podem permanecer inalterados, já que o sinal A transporta o objeto máximo geral.

10

Também nesta representação, pode ser feita consequentemente a inserção de um objeto monofônico, por exemplo, supondo $V=1$.

Em geral, o processo de transcodificação é feito de maneira que seu resultado aborde o resultado que seria obtido se todos os objetos originais de ambos os fluxos tivessem inicialmente sido codificados em um único fluxo JSC.

15

A Fig. 7 mostra um exemplo de um gerador de sinais de áudio do invento para a geração de um sinal de áudio de saída, que pode ser usado dentro da MCU 46 para implementar o conceito do invento.

5 O gerador de sinais de áudio 100 compreende um receptor de sinais de áudio 102, um combinador de canais 104, um calculador de parâmetros 106 e uma interface de saída 108.

 O receptor de sinais de áudio 102 recebe um primeiro sinal de áudio 110, compreendendo um primeiro canal down-mix 110a tendo informações sobre dois ou mais primeiros canais originais e compreendendo um parâmetro original 110b associado a um dos primeiros canais originais, descrevendo uma propriedade de um dos primeiros canais originais com relação a um canal de referência. O receptor de sinais de áudio 102 ainda recebe um
10 segundo sinal de áudio 112 compreendendo um segundo canal down-mix 112a tendo informações sobre pelo menos um segundo canal original.

 O receptor de sinais de áudio envia o primeiro canal down-mix 110a e o segundo canal down-mix 112a para uma
20 entrada do combinador de canais 104 e do primeiro canal down-mix 110a, o segundo canal down-mix 112a e o parâmetro original 110b para o calculador de parâmetros 106.

 O combinador de canais 104 recebe um canal down-mix combinado 114 combinando o primeiro canal down-mix 110a com um
25 segundo canal down-mix 112b, isto é, combinando os canais down-mix diretamente sem reconstruir os canais originais de áudio subjacentes.

 O calculador de parâmetros 106 recebe um primeiro

parâmetro combinado 116a descrevendo as propriedades de um dos primeiros canais originais com relação a um canal comum de referência e um segundo parâmetro combinado 116b descrevendo as propriedades de outro dos primeiros canais originais ou do pelo menos um segundo canal original com relação ao mesmo canal comum de referência. O primeiro e o segundo parâmetros combinados são enviados à interface de saída 108, que ainda recebe o canal down-mix combinado 114 do combinador de canais 104. Finalmente, a interface de saída envia um sinal de saída 120 compreendendo o canal down-mix combinado 114 e o primeiro e o segundo parâmetros combinados 116a e 116b.

O sinal de áudio de saída foi assim recebido sem a total reconstrução dos sinais de entrada de áudio e, portanto, sem operações custosas de computador.

Dentro dos parágrafos acima, foi demonstrado o conceito geral de mixagem de dois ou mais sinais, cada qual baseado em uma abordagem paramétrica JSC. Em particular, as equações acima mostram como aplicar essa técnica para um caso, onde as informações paramétricas consistem de taxas relativas de potência. Não obstante, essa técnica não está restrita a uma representação específica de parâmetros de objetos. Portanto, também podem ser usados os parâmetros que descrevem amplitude, medidas ou outras propriedades dos canais individuais de áudio, como correlações. As taxas de potência também podem ser computadas com relação a um canal down-mix combinado, com os custo de transmissão de outro parâmetro. Por um lado, existe o benefício neste cenário alternativo de reduzida complexidade de computação durante a mixagem dos fluxos de áudio, já que a reconstrução da

potência do canal de referência, que não é explicitamente transmitido em JSC "genérico", é obsoleta.

Além disso, a invenção não é limitada a um cenário de teleconferência, podendo ser aplicada sempre que for
5 desejada a multiplexação de objetos paramétricos e um fluxo único. Este pode ser, por exemplo, o caso dentro de esquemas de codificação BCC, *MPEG spatial surround* e outros.

Como demonstrado, o conceito do invento até permite a inclusão sem costuras de estações *legacy* remotas
10 proporcionando um único sinal monofônico no cenário baseado no objeto. Além da combinação de diferentes fluxos de objetos, o conceito do invento também mostra como caminhos diferentes de representação de dados paramétricos podem ser gerados, de maneira que sejam adequados para permitir processos de combinação
15 computacionalmente eficientes. Assim, é uma característica vantajosa de uma sintaxe de fluxo paramétrico de bits do invento expressar as propriedades de objeto de maneira que dois fluxos possam ser combinados por meio de operações muito simples.

Portanto, o conceito do invento também ensina
20 como criar fluxos adequados de bits ou formatos de fluxos de bits para codificar parametricamente múltiplos canais originais de áudio (objetos de áudio), observando-se os seguintes critérios:

- O sinal down-mix combinado é formado simplesmente a partir dos sinais down-mix parciais.
- As informações paramétricas auxiliares
25 combinadas são formadas a partir da combinação de informações paramétricas auxiliares individuais e de algumas simples para a computação das características dos sinais down-mix (por exemplo,

energia).

- Em nenhum caso, é necessária a realização de uma operação complexa, como uma etapa de decodificação / recodificação para os objetos de áudio.

5 Portanto, a representação paramétrica descrevendo os objetos deve ser escolhida de maneira que uma combinação ("adição") de dois ou mais fluxos de objetos seja possível usando somente campos de fluxos de bits disponíveis como parte das informações paramétricas auxiliares, e possivelmente de simples
10 computação da métrica dos sinais down-mix (por exemplo, energia, valor de pico).

 Um exemplo desta representação poderia usar valores de potência normalizada (Equação 4) para cada objeto. Estes podem ser transformados em representação logarítmica (dB) e
15 então quantizados em determinados números de etapas quantizadoras ou seus índices quantizadores representativos. A sintaxe do fluxo de bits deve permitir o fácil aumento (ou redução) do número de parâmetros de objetos em um fluxo, por exemplo, pela simples concatenação, inserção ou remoção de parâmetros.

20 Em resumo, o conceito do invento permite uma combinação de fluxos de áudio codificados parametricamente mais flexível e computacionalmente eficiente. Devido à alta eficiência computacional, o conceito do invento não se restringe a um número máximo de canais a ser combinado. Principalmente, os canais que
25 podem ser combinados em tempo real, podem ser fornecidos a um gerador de sinais de áudio do invento em números arbitrários. Também, não é obrigatória a exata representação paramétrica (JSC) usada para ilustrar o conceito do invento. Além disso, como já

mencionado, outros esquemas de codificação paramétrica, como os conhecidos esquemas surround, podem ser a base para a aplicação e o conceito do invento.

Além disso, as computações necessárias não necessariamente devem ser aplicadas em software. Também podem ser usadas implementações em hardware usando, por exemplo, DSPs, ASICs, e outros circuitos integrados para fazer os cálculos, que aumentarão ainda mais a velocidade do conceito do invento, permitindo a aplicação do conceito do invento em cenários de tempo real.

Devido à flexibilidade do conceito do invento, os fluxos de áudio do invento podem se basear em diferentes representações paramétricas. Os parâmetros a serem transmitidos também podem, por exemplo, ser medidas de amplitude, de diferenças de tempos entre canais originais de áudio, medidas de coerência e outras.

Assim, foi demonstrado o conceito geral de mixagem de dois ou mais sinais, cada qual baseado em uma abordagem paramétrica do estilo JSC.

As equações acima mostram como aplicar essa técnica para um caso, em que as informações paramétricas consistem de taxas relativas de potência. Entretanto, essa técnica não se restringe a uma representação específica de parâmetros de objetos.

Além disso, a invenção não se limita a um cenário de teleconferências, podendo ser aplicada a qualquer caso, em que seja vantajosa a multiplexação de objetos paramétricos em um único fluxo JSC.

Além disso, essa técnica permite a inclusão sem

costuras de estações remotas *legacy* proporcionando um único sinal monofônico no cenário baseado em objeto.

Além do processo real de combinação de diferentes fluxos de objetos, a invenção também mostra como formas diferentes de representação de dados paramétricos são adequadas para habilitar este processo de combinação. Já que nem todas as possíveis representações paramétricas admitem o processo de combinação descrito sem a completa decodificação / recodificação dos objetos, trata-se de uma característica vantajosa de uma sintaxe de fluxo de bits paramétricos expressar as propriedades dos objetos de maneira que dois fluxos possam ser combinados, com a realização de simples operações.

Dependendo de determinadas exigências de implementação dos métodos do invento, estes podem ser implementados em hardware ou em software. A implementação pode ser feita usando um meio de armazenagem digital, em particular um disco, DVD ou um CD dotado de sinais de controle de leitura eletrônica armazenados, que cooperam com um sistema de computador programável, para que os métodos do invento sejam realizados. Portanto, em geral, a presente invenção é um produto de programa de computador com um código de programa armazenado em um veículo com leitura por máquina, o código de programa sendo operacional para a realização dos métodos do invento quando o produto de programa de computador opera em um computador. Em outras palavras, os métodos do invento são, portanto, um programa de computador tendo um código de programa para a realização de pelo menos um dos métodos da invenção quando o programa de computador opera em um computador.

Apesar de o exposto ter sido mostrado e descrito em particular com referência a determinadas configurações, será compreendido pelos técnicos no assunto que várias outras alterações de forma e detalhes podem ser feitas sem abandonar seu
5 escopo. Fica entendido que várias alterações podem ser feitas na adaptação a diferentes configurações, sem abandonar os conceitos mais amplos revelados na presente e contidos pelas reivindicações a seguir.

REIVINDICAÇÕES

1. Gerador de sinal de áudio (100) para gerar um sinal de saída de áudio, caracterizado por compreender:

um receptor de sinal de áudio (102) para receber:

um primeiro sinal de áudio (110) compreendendo um primeiro canal down-mix (110a) tendo informações sobre dois ou mais primeiros canais originais, e compreendendo pelo menos um parâmetro original (110b) associado a um dos primeiros canais originais que descreve uma propriedade de um dos primeiros canais originais em relação a um canal de referência; e

um segundo sinal de áudio (112) compreendendo um segundo canal down-mix (112a) tendo informações sobre pelo menos um segundo canal original;

um combinador de canal (104) para derivar um canal down-mix combinado (114) pela combinação do primeiro canal down-mix (110a) e do segundo canal down-mix (112a) utilizando uma combinação linear do primeiro (110a) e do segundo canal down-mix (112b), onde o coeficientes de combinação linear são dependentes da energia $E(s_A^2(n))$ dentro do primeiro canal down-mix (110a) e da energia $E(s_B^2(n))$ dentro do segundo canal down-mix (112a); e

um calculador de parâmetro (106) para derivar utilizando a energia $E(s_A^2(n))$ do primeiro canal down-mix (110a), a energia $E(s_B^2(n))$ do segundo canal down-mix (112a) e pelo menos ; e pelo menos um parâmetro original (110b),

um primeiro parâmetro combinado (116a) que descreve a propriedade de um dos primeiros canais originais em relação a um canal de referência comum, e

e um segundo parâmetro combinado (116b) que

descreve a propriedade do outro primeiro canal original ou de pelo menos um segundo canal original em relação ao canal de referência comum;

uma interface de saída para emitir o sinal de saída de áudio (120) compreendendo o canal down-mix combinado (114), o primeiro (116a) e o segundo parâmetros combinados (116b).

2. Gerador de sinal de áudio (100), de acordo com a reivindicação 1, caracterizado pelo fato de que o combinador de canal (104) funciona para utilizar uma combinação linear tendo um coeficiente g_A para o primeiro canal down-mix e um coeficiente g_B para o segundo canal down-mix derivado utilizando a seguinte equação:

$$\frac{g_B}{g_A} = \sqrt{\frac{E\{s_A^2(n)\}}{E\{s_B^2(n)\}}}.$$

3. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 2, caracterizado pelo fato de que o combinador de canal (104) funciona para utilizar uma combinação linear tendo coeficientes que dependem do número U dos primeiros canais originais e do número V dos segundos canais originais.

4. Gerador de sinal de áudio (100), de acordo com a reivindicação 3, caracterizado pelo fato de que o combinador de canal (104) funciona para utilizar uma combinação linear tendo um coeficiente g_A do primeiro canal down-mix (110a) e um coeficiente g_B do segundo canal down-mix (112a) derivado de acordo com uma das seguintes equações:

$$g_A = \frac{U}{(U+V)}, \quad g_B = \frac{V}{(U+V)}$$

ou

$$g_A = \sqrt{\frac{U}{(U+V)}}, g_B = \sqrt{\frac{V}{(U+V)}}.$$

5. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 4, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para utilizar um canal predeterminado dos primeiros canais originais ou pelo menos um segundo canal original como canal de referência comum.

6. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 4, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para utilizar o canal de referência do primeiro sinal de áudio (110) como o canal de referência comum.

7. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 4, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para utilizar o canal down-mix combinado (114) como o canal de referência comum.

8. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 4, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para utilizar o canal original como o canal de referência comum que possui a maior energia.

9. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 8, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para calcular a energia $E\{s_{Aref}\}$ do canal de referência pela derivação da energia $E\{s_A^2\}$ do primeiro canal down-mix (110a) e dos parâmetros $a_i\{i = 1, \dots, n\}$ associados a canais que não o canal de referência de

acordo com a equação:

$$E\left\{s_{A_{ref}}\right\} = \frac{E\{s_A^2(n)\}}{1 + \sum_i a_i}.$$

10. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 9, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para utilizar o canal de referência como o canal de referência comum e pelo menos um parâmetro original a_2 como o primeiro parâmetro combinado y_u e para derivar o segundo parâmetro combinado y_{u+1} para pelo menos um segundo canal original em relação ao canal de referência.

11. Gerador de sinal de áudio (100), de acordo com a reivindicação 1, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para ainda utilizar coeficientes g_A associados ao primeiro canal down-mix (110a) e g_B associado ao segundo canal down-mix (112a), os coeficientes utilizados para a combinação linear do primeiro e segundo down-mix utilizado pelo combinador de canal (104).

12. Gerador de sinal de áudio (100), de acordo com a reivindicação 10, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para calcular o segundo parâmetro combinado y_{u+1} para pelo menos um segundo canal original de acordo com a seguinte equação:

$$y_{u+1} = \frac{g_B^2}{g_A^2} \cdot \frac{E\{s_{B_1}^2(n)\}}{E\{s_{A_1}^2(n)\}},$$

onde $E\{s_{A_1}^2(n)\}$ é a energia do canal de referência

derivado utilizando a energia do primeiro canal down-mix $E\{s_A^2(n)\}$ de acordo com a seguinte fórmula:

$$E\{s_A^2(n)\} = \frac{E\{s_A^2(n)\}}{(1+a_2)},$$

onde a_2 é pelo menos um parâmetro original que relaciona um primeiro canal original ao canal de referência.

13. Gerador de sinal de áudio (100), de acordo com qualquer uma das reivindicações de 1 a 12, caracterizado pelo fato de que o calculador de parâmetro (106) funciona para processar porções de frequência do primeiro e segundo canais down-mix associados a intervalos de frequência discretos, de modo que os parâmetros combinados sejam derivados para cada intervalo de frequência discreto.

14. Gerador de sinal de áudio (100), de acordo com a reivindicação 1, caracterizado pelo fato de que o receptor de sinal de áudio funciona para receber sinais de áudio (110, 112) compreendendo canais down-mix (110a, 112a) representados por parâmetros de amostragem amostrados com uma frequência de amostra predeterminada.

15. Método de geração de um sinal de saída de áudio, o método caracterizado por compreender:

recepção de um primeiro sinal de áudio (110) compreendendo um primeiro canal down-mix (110a) tendo informações sobre dois ou mais primeiros canais originais, e compreendendo pelo menos um parâmetro original (110b) associado a um dos primeiros canais originais que descreve uma propriedade de um dos primeiros canais originais em relação a um canal de referência e um segundo sinal de áudio (112) compreendendo um segundo canal down-mix (112a) tendo informações sobre pelo menos um segundo canal original;

derivação de um canal down-mix combinado (114) pela combinação do primeiro canal down-mix (110) e do segundo canal down-mix (112), utilizando uma combinação linear do primeiro (110a) e do segundo canal down-mix (112b), onde o coeficientes de combinação linear são dependentes da energia $E(s_A^2(n))$ dentro do primeiro canal down-mix (110a) e da energia $E(s_B^2(n))$ dentro do segundo canal down-mix (112a); e

derivação, utilizando a energia $E(s_A^2(n))$ do primeiro canal down-mix (110a), a energia $E(s_B^2(n))$ do segundo canal down-mix (112a) e pelo menos ; e pelo menos um parâmetro original (110b),

um primeiro parâmetro combinado (116a) que descreve a propriedade de um dos primeiros canais originais em relação a um canal de referência comum,

e um segundo parâmetro combinado (116b) que descreve também a propriedade do outro primeiro canal original ou de pelo menos um segundo canal original em relação a um canal de referência comum; e

emissão do sinal de saída de áudio (120) compreendendo o canal down-mix combinado (114) e o primeiro (116a) e segundo (116b) parâmetros combinados.

16. Sistema de conferência de acordo com a reivindicação 1, caracterizado por compreender um gerador de sinal de áudio (100) para gerar um sinal de saída de áudio.

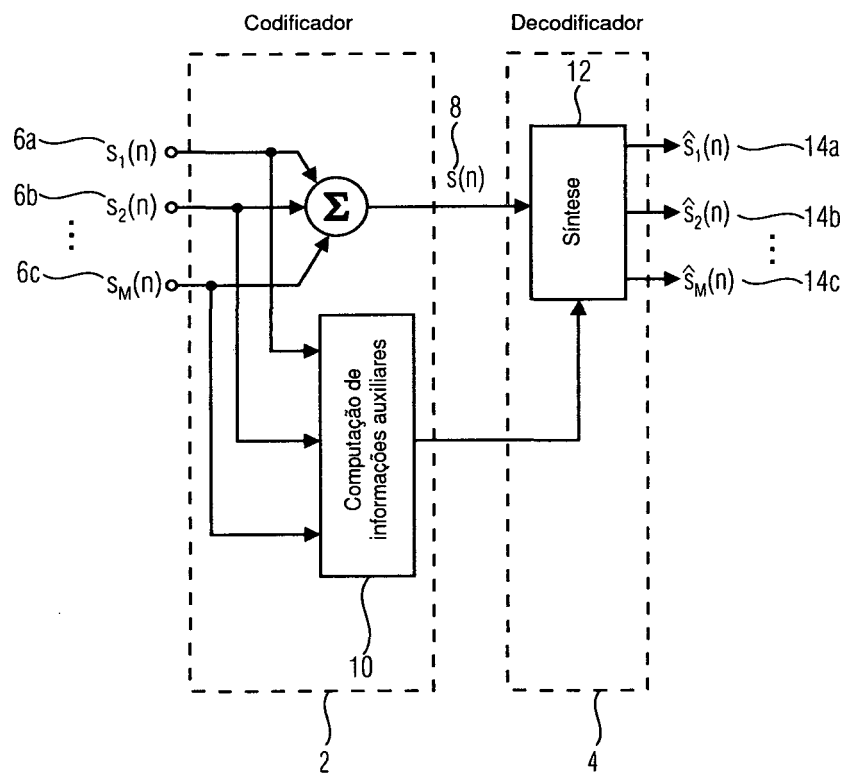


FIG 1

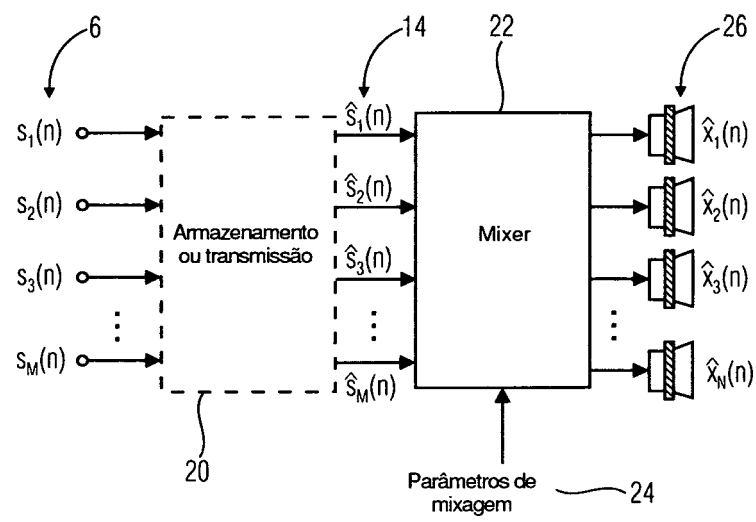


FIG 2

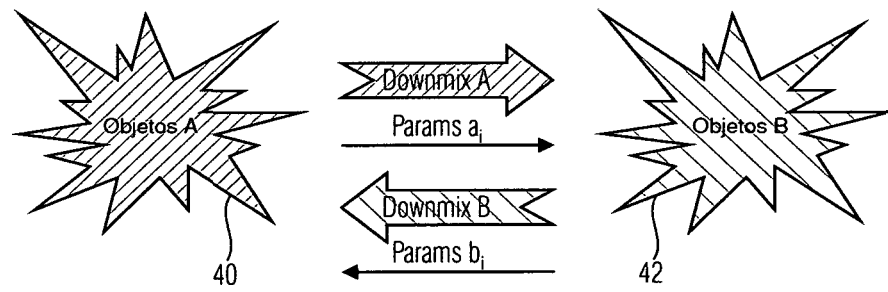


FIG 3

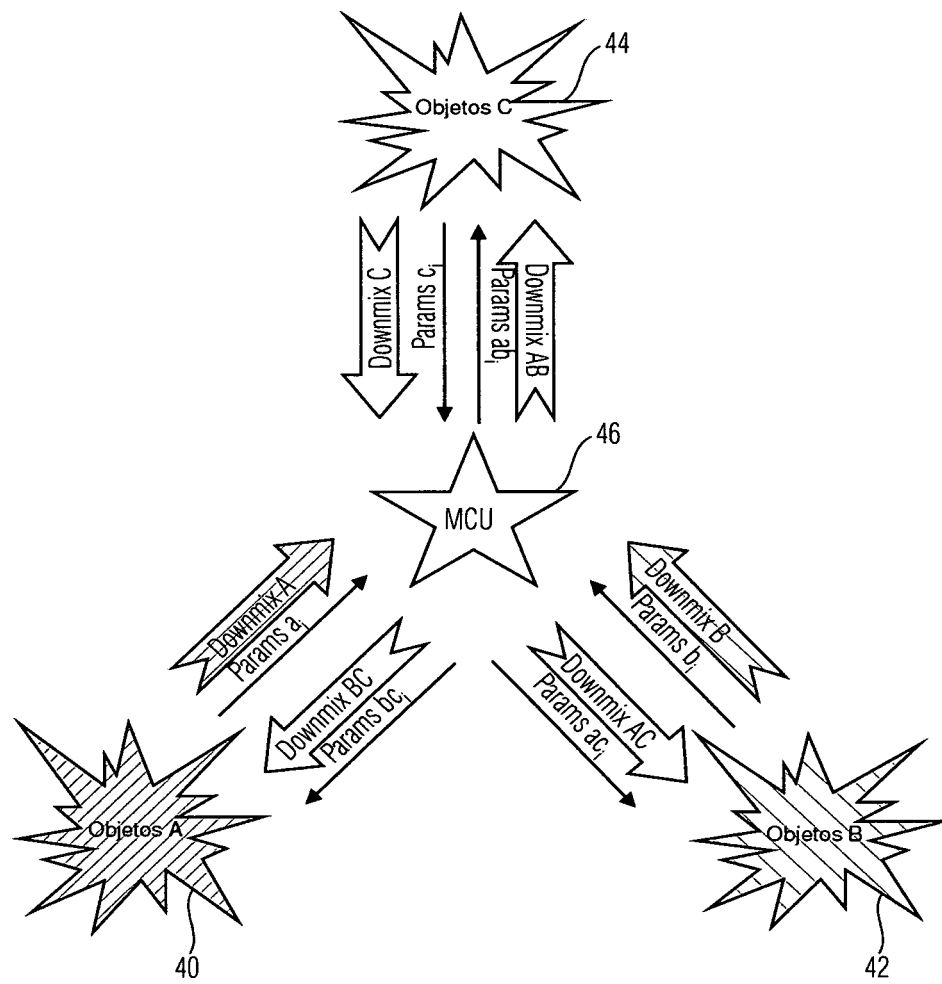


FIG 4

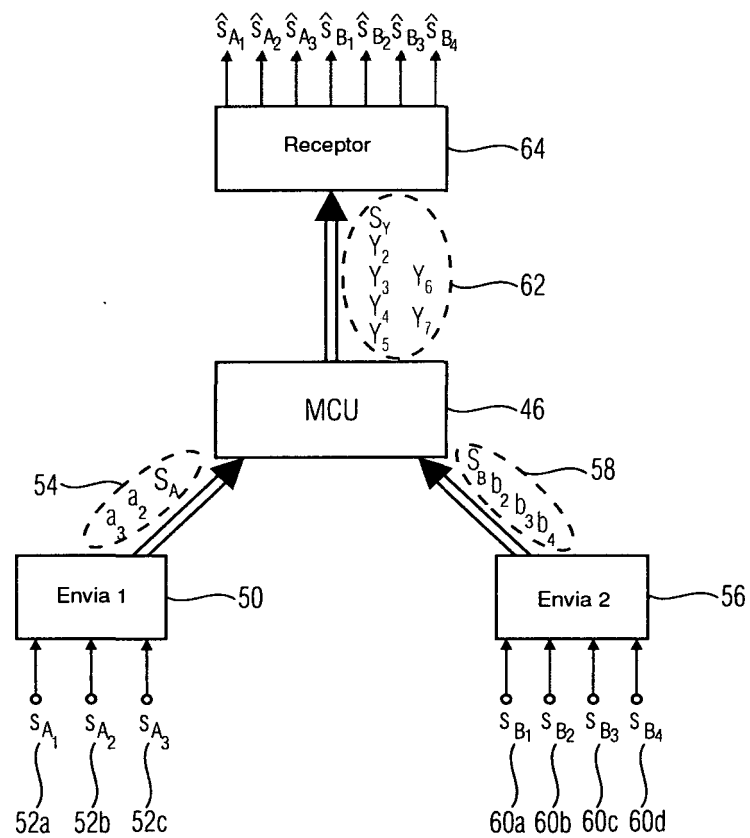


FIG 5

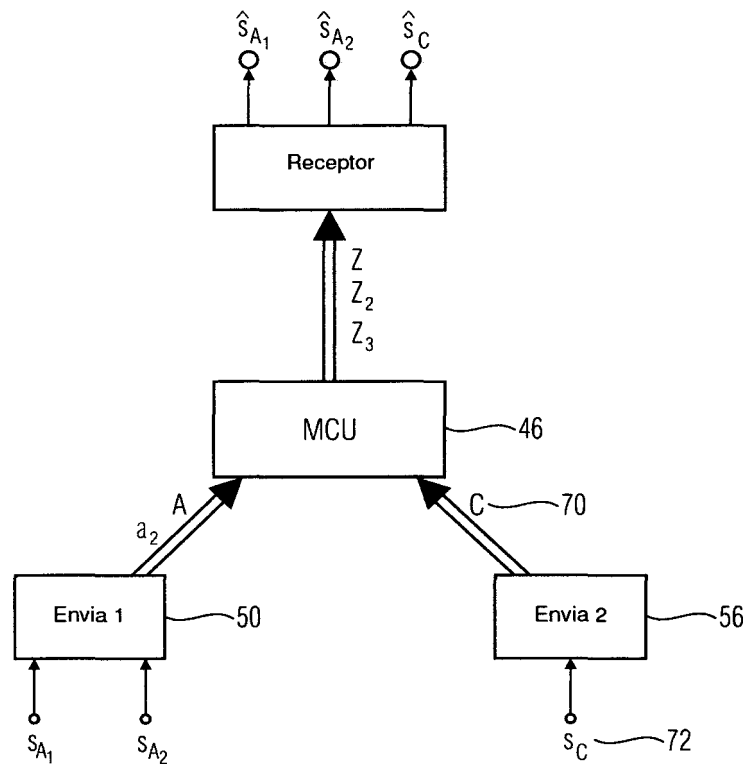


FIG 6

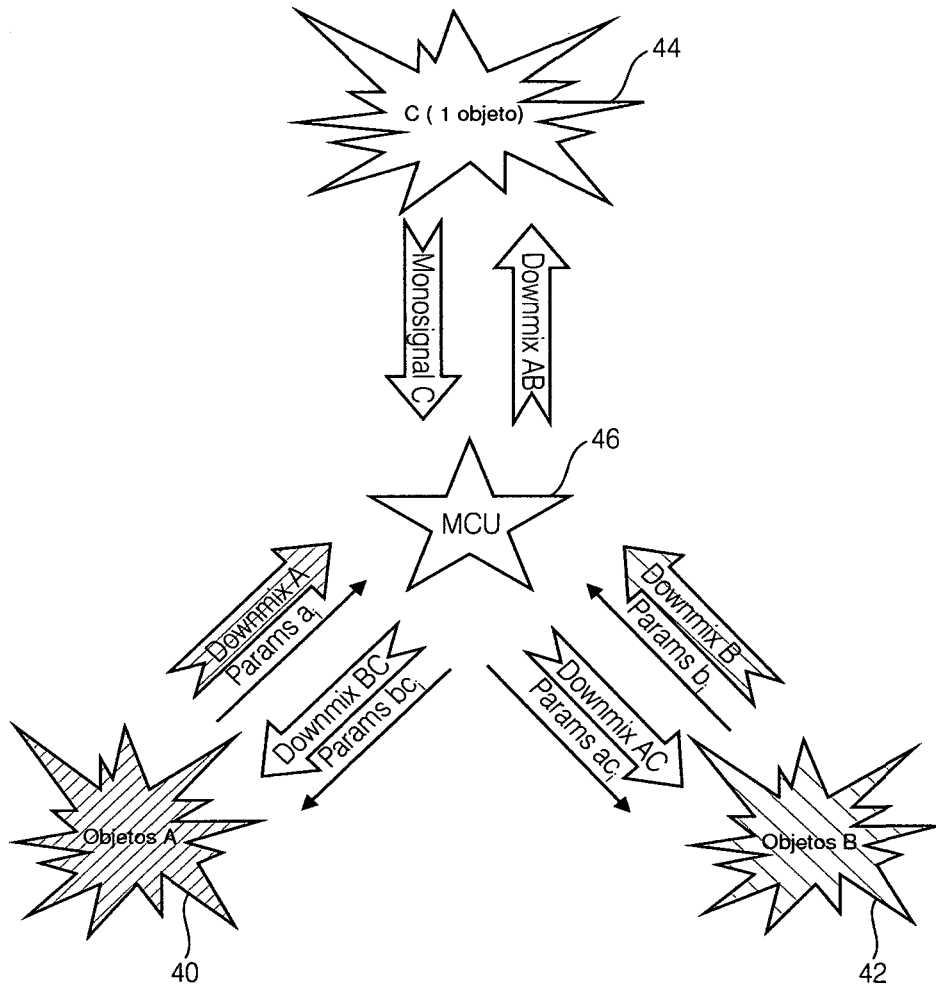


FIG 6B

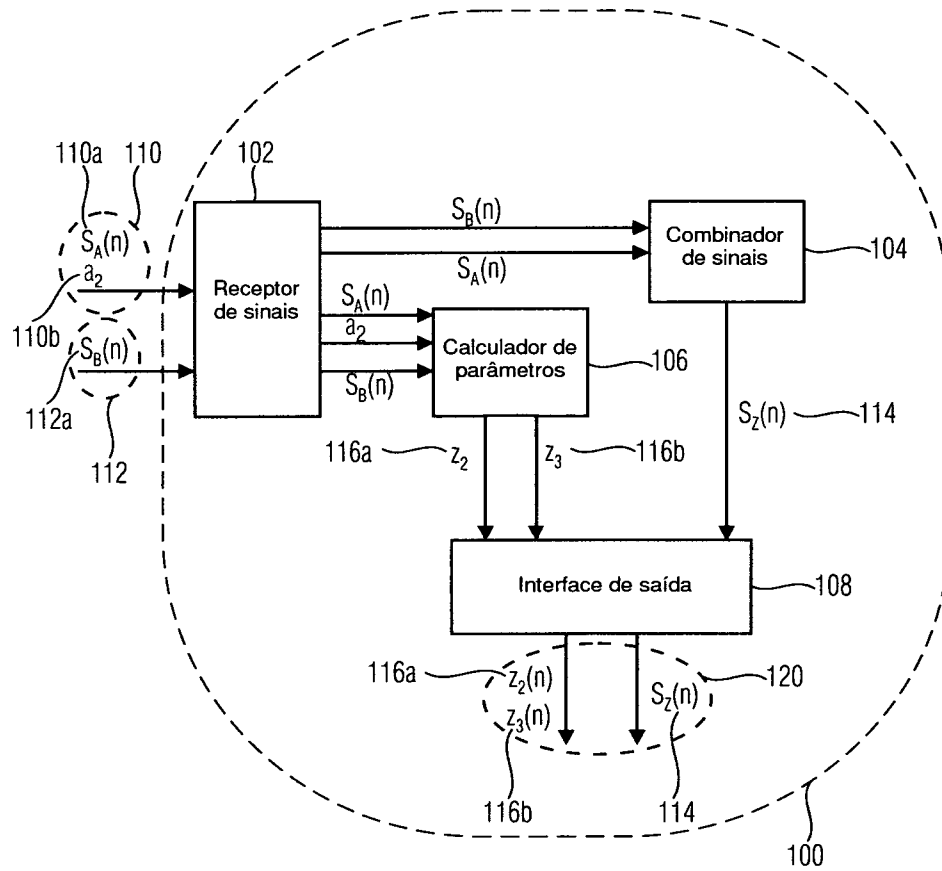


FIG 7