



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

(11) BR 122018072501-0 B1



(22) Data do Depósito: 13/02/2006

(45) Data de Concessão: 09/07/2019

(54) Título: CODIFICAÇÃO PARAMÉTRICA CONJUNTA DE FONTES DE ÁUDIO

(51) Int.Cl.: G10L 19/008; G10L 19/02; H04S 3/00; H04S 7/00.

(52) CPC: G10L 19/008; G10L 19/0204; H04S 3/008; H04S 7/307; H04S 2420/03; (...).

(30) Prioridade Unionista: 14/02/2005 EP 05101055.1.

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

(72) Inventor(es): CHRISTOF FALLER.

(86) Pedido PCT: PCT EP2006050904 de 13/02/2006

(87) Publicação PCT: WO 2006/084916 de 17/08/2006

(85) Data do Início da Fase Nacional: 31/10/2018

(62) Pedido Original do Dividido: PI0607166-0 - 13/02/2006

(57) Resumo: O seguinte cenário de codificação é abordado: diversos sinais de fonte de áudio precisam ser transmitidos ou armazenados para fins de mixagem de síntese de campo de onda, sinais surround de multicanais ou sinais estéreos após a decodificação dos sinais de fonte. A técnica proposta oferece significativo ganho de codificação ao se realizar a codificação conjunta dos sinais de fonte em comparação à codificação separada destes, mesmo quando nenhuma redundância estiver presente entre os sinais de fonte. Isto é possível considerando as propriedades estatísticas dos sinais de fonte, as propriedades das técnicas de mixagem e a audição espacial. A soma dos sinais de fonte é transmitida com as propriedades estatísticas dos sinais de fonte que predominantemente determinam os cues espaciais perceptivelmente importantes dos canais de áudio finais mixados. Os sinais de fonte são recuperados no receptor, de modo que suas propriedades estatísticas se aproximem das propriedades correspondentes dos sinais de fonte originais. Avaliações subjetivas indicam que a alta qualidade de áudio é obtida utilizando-se o esquema proposto.

"CODIFICAÇÃO PARAMÉTRICA CONJUNTA DE FONTES DE ÁUDIO"

Dividido do PI 0607166-0, depositado em 13/02/2006.

INTRODUÇÃO

[001] Em um problema de codificação geral, temos diversos sinais de fonte (mono) $s_i(n)$ ($1 \leq i \leq M$) e um vetor de descrição de cena $\mathbf{S}(n)$, onde n é o índice de tempo. O vetor de descrição de cena contém parâmetros, por exemplo posições de fonte (virtuais), larguras de fonte e parâmetros acústicos, por exemplo, parâmetros de espaço (virtuais). A descrição de cena pode ser invariável no tempo ou pode mudar no decorrer do tempo. Os sinais de fonte e a descrição de cena são codificados e transmitidos para um decodificador. Os sinais de fonte codificados, $\check{s}_i(n)$ são sucessivamente mixados em função da descrição de cena, $\check{S}(n)$, para gerar síntese de campo de onda, sinais de multicanais ou sinais estéreos em função do vetor de descrição de cena. Os sinais de saída do decodificador são denominados $\check{x}_i(n)$ ($0 \leq i \leq N$). Deve ser observado que o vetor de descrição de cena $\mathbf{S}(n)$ pode não ser transmitido, porém pode ser determinado no decodificador. Neste documento, o termo "sinal de áudio estéreo" sempre se refere a sinais de áudio estéreos de dois canais.

[002] ISO/IEC MPEG-4 aborda o cenário de codificação descrito. Define a descrição de cena e utiliza, para cada sinal de fonte ("natural"), um codificador de áudio mono separado, por exemplo, um codificador de áudio AAC. No entanto, quando uma cena complexa com várias fontes precisar ser mixada, a taxa de bits se torna alta, ou seja, a taxa de bits aumenta com o número de fontes. A codificação de um sinal de fonte com alta qualidade requer aproximadamente 60 - 90 kb/s.

[003] Anteriormente, abordamos um caso especial do problema de codificação descrito [1][2] com um esquema denominado *Binaural Cue Coding* (BCC) para *Flexible Rendering*. Ao transmitir somente a soma de determinados sinais de fonte mais as informações colaterais de baixa taxa de bits, a baixa taxa de bits é alcançada. No entanto, os sinais de fonte não podem ser recuperados no decodificador e o esquema foi limitado à geração de sinal *surround* estéreo e de multicanais. Também, foi somente utilizada uma mixagem simplista baseada na amplitude e variação gradual de retardo. Assim, a direção das fontes poderia ser controlada, porém nenhum outro atributo de imagem espacial de audição. Outra limitação desse esquema foi sua limitada qualidade de áudio. Em especial, uma diminuição da qualidade de áudio conforme o número de sinais de fonte aumenta.

[004] O documento [1], (*Binaural Cue Coding*, Estéreo Paramétrico, MP3 *Surround*, MPEG *Surround*) abrange o caso onde N canais de áudio são codificados e N canais de áudio com **cues similares** e então os canais de áudio originais são decodificados. As informações colaterais transmitidas incluem parâmetros de *cue* intercanais relacionados às diferenças entre os canais de entrada.

[005] Os canais de sinais de áudio estéreo e de multicanais contêm mixagens de sinais de fontes de áudio e são assim diferentes, em termos de natureza, dos sinais de fonte de áudio puros. Os sinais de áudio estéreo e de multicanais são mixados de modo que, quando reproduzidos em um sistema de *playback* apropriado, o ouvinte perceberá uma imagem espacial de audição ("estágio de som") conforme é capturado pela configuração de gravação ou designado pelo engenheiro de gravação durante a mixagem. Diversos esquemas para codificação conjunta para os canais de um sinal de

áudio estéreo ou de multicanais foram anteriormente propostos.

SUMÁRIO DA INVENÇÃO

[006] O objetivo da invenção é prover um método para transmitir diversos sinais de fonte enquanto se utiliza uma mínima largura de banda. Na maioria dos métodos conhecidos, o formato de *playback* (por exemplo, estéreo, 5.1) é predefinido e tem uma influência direta sobre o cenário de codificação. O *stream* de áudio no lado do decodificador deve utilizar somente este formato de *playback* predefinido, portanto vinculando o usuário a um cenário de *playback* predefinido (por exemplo, estéreo).

[007] A invenção proposta codifica N sinais de fonte de áudio, tipicamente não canais de um sinal estéreo ou de multicanais, mas sim sinais independentes, por exemplo diferentes sinais de fala ou instrumento. As informações colaterais transmitidas incluem parâmetros estatísticos referentes aos sinais de fonte de áudio de entrada.

[008] A invenção proposta decodifica M canais de áudio com **cues diferentes** dos sinais de fonte de áudio originais. Esses diferentes *cues* são tanto implicitamente sintetizados aplicando-se um mixador ao sinal de soma recebido. O mixador é controlado em função das informações de fonte estatísticas recebidas e dos parâmetros de formato de áudio recebidos (ou localmente determinados) e dos parâmetros de mixagem. Alternativamente, esses diferentes *cues* são explicitamente computados em função das informações de fonte estatísticas recebidas e dos parâmetros de formato de áudio recebidos (ou localmente determinados) e dos parâmetros de mixagem. Esses *cues* computados são utilizados para controlar o decodificador da técnica anterior (*Binaural Cue Coding*, Estéreo Paramétrico, MPEG Surround) para sintetizar os canais de

saída dado o sinal de soma recebido.

[009] O esquema proposto para codificação conjunta de sinais de fonte de áudio é o primeiro de seu gênero. É projetado para codificação conjunta de sinais de fonte de áudio. Os sinais de fonte de áudio são geralmente sinais de áudio mono que não são adequados para *playback* em um sistema de áudio estéreo ou de multicanais. Para facilitar, a seguir, os sinais de fonte de áudio são geralmente denominados sinais de fonte.

[0010] Os sinais de fonte de áudio primeiro precisam ser mixados em sinais de áudio estéreo, de multicanais ou de síntese de campo de onda antes do *playback*. Um sinal de fonte de áudio pode ser um único instrumento ou locutor, ou a soma de diversos instrumentos e locutores. Outro tipo de sinal de fonte de áudio é um sinal de áudio mono capturado com um microfone de ponto durante um concerto. Os sinais de fonte de áudio são geralmente armazenados em gravadores de múltiplas faixas ou em sistemas de gravação em disco rígido.

[0011] O esquema reivindicado de codificação conjunta de sinais de fonte de áudio é baseado somente na transmissão da soma dos sinais de fonte de áudio,

$$s(n) = \sum_{i=1}^M s_i(n) \quad , \quad (1)$$

[0012] ou de uma soma ponderada dos sinais de fonte. Opcionalmente, a soma ponderada pode ser realizada com diferentes pesos em diferentes sub-bandas e os pesos podem ser adaptados ao tempo. A soma com equalização, conforme descrita no Capítulo 3.3.2 em [1], também pode ser aplicada. A seguir, quando nos referimos à soma ou sinal de soma, sempre nos referimos a uma geração de sinal por (1) ou gerado conforme descrito. Além do sinal de soma, são transmitidas informações colaterais. A soma e as informações

colaterais representam o *stream* de áudio emitido. Opcionalmente, o sinal de soma é codificado utilizando um codificador de áudio mono convencional. Esse *stream* pode ser armazenado em um arquivo (CD, DVD, disco rígido) ou transmitido ao receptor. As informações colaterais representam as propriedades estatísticas dos sinais de fonte que são os fatores mais importantes que determinam os *cues* espaciais perceptuais dos sinais de saída do mixador. Será mostrado que essas propriedades estão temporalmente desenvolvendo envelopes espectrais e funções de autocorrelação. Aproximadamente 3 kb/s de informações colaterais são transmitidos por sinal de fonte. No receptor, os sinais de fonte $\tilde{\mathbf{d}}_i(n)$ ($1 \leq i \leq M$) são recuperados com as propriedades estatísticas anteriormente mencionadas aproximando as correspondentes propriedades dos sinais de fonte originais e do sinal de soma.

BREVE DESCRIÇÃO DOS DESENHOS

[0013] A invenção será mais bem compreendida graças às Figuras anexadas nas quais:

[0014] - A figura 1 mostra um esquema no qual a transmissão de cada sinal de fonte é realizada independentemente para futuro processamento,

[0015] - A figura 2 mostra diversas fontes transmitidas como sinal de soma mais informações colaterais,

[0016] - A figura 3 mostra um diagrama de blocos de um esquema de *Binaural Cue Coding* (BCC),

[0017] - A figura 4 mostra um mixador para geração de sinais estéreos com base em vários sinais de fonte,

[0018] - A figura 5 mostra a dependência entre ICTD, ICLD e ICC e o potencial de sub-banda do sinal de fonte,

[0019] - A figura 6 mostra o processo de geração de

informações colaterais,

[0020] - A figura 7 mostra o processo de estimativa dos parâmetros de LPC de cada sinal de fonte,

[0021] - A figura 8 mostra o processo de recriação dos sinais de fonte a partir de um sinal de soma,

[0022] - A figura 9 mostra um esquema alternativo para a geração de cada sinal a partir do sinal de soma,

[0023] - A figura 10 mostra um mixador para a geração de sinais estéreos com base no sinal de soma,

[0024] - A figura 11 mostra um algoritmo de variação gradual de amplitude impedindo que os níveis de fonte dependam dos parâmetros de mixagem,

[0025] - A figura 12 mostra uma matriz de alto-falantes de um sistema de *playback* de síntese de campo de onda,

[0026] - A figura 13 mostra como recuperar uma estimativa dos sinais de fonte no receptor pelo processamento de *downmix* dos canais transmitidos,

[0027] - A figura 14 mostra como recuperar uma estimativa dos sinais de fonte no receptor pelo processamento dos canais transmitidos.

II. DEFINIÇÕES, NOTAÇÃO E VARIÁVEIS

[0028] Seguem as notações e variáveis utilizadas neste documento:

[0029] n índice de tempo;

[0030] i canal de áudio ou índice de fonte;

[0031] d índice de retardo;

[0032] M número de sinais de fonte de entrada do codificador;

[0033] N número de canais de saída do decodificador;

- [0034] $x_i(n)$ sinais de fonte originais mixados;
- [0035] $\ddot{\mathbf{x}}_i(n)$ sinais de saída mixados do decodificador;
- [0036] $s_i(n)$ sinais de fonte de entrada do codificador;
- [0037] $\ddot{\mathbf{x}}_i(n)$ sinais de fonte transmitidos também chamados de pseudo-sinais de fonte;
- [0038] $s(n)$ sinal de soma transmitido;
- [0039] $y_i(n)$ sinal de áudio de canal L ; (sinal de áudio a ser remixado);
- [0040] $\tilde{s}_i(k)$ um sinal de sub-banda de $s_i(n)$ (similarmente definido para outros sinais);
- E $\{\tilde{s}_i^2(n)\}$ estimativa de curto prazo de $\tilde{s}_i^2(n)$ (similarmente definido para outros sinais);
- [0041] ICLD diferença de nível intercanais;
- [0042] ICTD diferença de tempo intercanais;
- [0043] ICC coerência intercanais;
- [0044] $\square L(n)$ ICLD estimada de sub-banda;
- [0045] $\square(n)$ ICTD estimada de sub-banda;
- [0046] $c(n)$ ICC estimada de sub-banda;
- [0047] $\tilde{p}_i(n)$ potencial relativo de sub-banda de fonte;
- [0048] a_i, b_i fatores de escala do mixador;
- [0049] c_i, d_i retardos do mixador;
- [0050] $\square L_i, \square(n)$ nível do mixador e diferença de tempo;
- [0051] G_i ganho de fonte do mixador;

III. CODIFICAÇÃO CONJUNTA DE SINAIS DE FONTE DE ÁUDIO

- [0052] Primeiro, é descrito o *Binaural Cue Coding* (BCC), uma técnica de codificação paramétrica de áudio de multicanais.

Então, é demonstrado que com a mesma percepção na qual o BCC se baseia, pode-se propor um algoritmo para codificação conjunta dos sinais de fonte para um cenário de codificação.

A. *Binaural Cue Coding* (BCC)

[0053] Um esquema BCC [1][2] para codificação de áudio de multicanais é mostrado na figura abaixo. O sinal de áudio de entrada de multicanais é downmixado em um canal único. Em vez da codificação e transmissão de informações sobre todos os formatos de onda de canal, somente o sinal downmixado é codificado (com um codificador convencional de áudio mono) e transmitido. Além disso, as “diferenças de canal de áudio” perceptivelmente motivadas são estimadas entre os canais de áudio originais e também transmitidas ao decodificador. O decodificador gera seus canais de saída de modo que as diferenças de canal de áudio se aproximem das correspondentes diferenças de canal de áudio do sinal de áudio original.

[0054] A localização da soma implica que as diferenças de canal de áudio perceptivelmente relevantes para um par de canais de sinal de alto-falante são a diferença de tempo intercanais (ICTD) e a diferença de nível intercanais (ICLD). A ICTD e a ICLD podem estar relacionadas à direção percebida dos eventos de audição. Outros atributos de imagem espacial de audição, por exemplo, largura de fonte aparente e envolvimento do ouvinte, podem estar relacionados à *coerência interaural* (IC). Para pares de alto-falante na frente ou atrás de um ouvinte, a *coerência interaural* está, de modo geral, diretamente relacionada à coerência intercanais (ICC) que é então considerada como uma terceira medida de diferença de canal de áudio pela BCC. A ICTD, a ICLD e a ICC são estimadas em sub-bandas em função de tempo. Tanto as resoluções espectral como temporal utilizadas são motivadas pela percepção.

B. Codificação Paramétrica Conjunta de Fontes de Áudio

[0055] Um decodificador BCC é capaz de gerar um sinal de áudio de multicanais com qualquer imagem espacial de audição utilizando um sinal mono e sintetizando, em intervalos regulares, um cue específico único de ICTD, ICLD e ICC por sub-banda e par de canais. O bom desempenho dos esquemas BCC para uma ampla faixa de material de áudio [vide 1] implica que a imagem espacial de audição percebida é amplamente determinada pela ICTD, ICLD e ICC. Portanto, em vez de exigir sinais de fonte $s_i(n)$ “limpos” como entrada do mixador na Figura 1, apenas exigimos pseudo-sinais de fonte $\ddot{s}_i(n)$ com a propriedade de resultarem em ICTD, ICLD e ICC similares na saída do mixador como no caso de fornecimento dos reais sinais de fonte ao mixador. Há três objetivos para a geração de $\ddot{s}_i(n)$:

[0056] Se $\ddot{s}_i(n)$ for fornecido a um mixador, os canais de saída do mixador terão aproximadamente os mesmos cues espaciais (ICLD, ICTD, ICC) como se $s_i(n)$ fosse fornecido ao mixador.

[0057] $\ddot{s}_i(n)$ deve ser gerado com a menor quantidade de informação possível sobre os sinais de fonte originais $s(n)$ (pois o objetivo é ter informações colaterais de baixa taxa de bits).

[0058] $\ddot{s}_i(n)$ são gerados a partir do sinal de soma transmitido $s(n)$ de modo que uma quantidade mínima de distorção de sinal seja introduzida.

[0059] Para derivar o esquema proposto, estamos considerando um mixador estéreo ($M = 2$). Outra simplificação do caso geral é que somente a amplitude e variação gradual de retardo são aplicados para mixagem. Se os sinais de fonte discretos estivessem disponíveis no decodificador, um sinal estéreo seria mixado conforme mostrado na Figura 4, ou seja,

$$x_1(n) = \sum_{i=1}^M a_i s_i(n - c_i) \quad x_2(n) = \sum_{i=1}^M b_i s_i(n - d_i) \quad (2)$$

[0060] Nesse caso, o vetor de descrição de cena $\mathbf{s}(n)$ contém apenas direções de fonte que determinam os parâmetros de mixagem,

$$\mathbf{M}(n) = (a_1, a_2, \dots, a_M, b_1, b_2, \dots, b_M, c_1, c_2, \dots, c_M, d_1, d_2, \dots, d_M)^T \quad (3)$$

[0061] onde T é a transposição de um vetor. Deve ser observado que, para os parâmetros de mixagem, ignoramos o índice de tempo para maior conveniência de notação.

[0062] Parâmetros mais convenientes de controle do mixador são a diferença de tempo e nível, τ_i e ΔL_i , que estão relacionados a a_i , b_i , c_i , e d_i por

$$a_i = \frac{10^{G_i/20}}{\sqrt{1+10^{\Delta L_i/10}}} \quad b_i = 10^{(G_i+\Delta L_i)/20} a_i \quad c_i = \max \{-\tau_i, 0\} \\ d_i = \max \{\tau_i, 0\} \quad (4)$$

[0063] onde G_i é um fator de ganho de fonte em dB.

[0064] A seguir, estamos computando ICTD, ICLD e ICC da saída estéreo do mixador em função dos sinais de fonte de entrada $s_i(n)$. As expressões obtidas indicarão quais propriedades de sinal de fonte determinam ICTD, ICLD e ICC (com os parâmetros de mixagem). $\hat{\mathbf{Q}}(n)$ são então gerados de modo que as propriedades de sinal de fonte identificado se aproximem das propriedades correspondentes dos sinais de fonte originais.

B.1 ICTD, ICLD e ICC da saída do mixador

[0065] Os cues são estimados em sub-bandas e em função do tempo. A seguir, assume-se que os sinais de fonte $s_i(n)$ são média zero e mutuamente independentes. Um par de sinais de sub-banda da saída do mixador (2) é denominado $\hat{\mathbf{x}}_1(n)$ e $\hat{\mathbf{x}}_2(n)$. Deve ser observado

que, para maior simplicidade de notação, estamos utilizando o mesmo índice de tempo n para sinais de domínio de tempo e de domínio de sub-banda. Também, nenhum índice de sub-banda é utilizado e a análise/processamento descrita é aplicada a cada sub-banda independentemente. O potencial de sub-banda dos dois sinais de saída do mixador é

$$E \{ \tilde{x}_1^2(n) \} = \sum_{i=1}^M a_i^2 E \{ \tilde{s}_i^2(n) \} \quad E \{ \tilde{x}_2^2(n) \} = \sum_{i=1}^M b_i^2 E \{ \tilde{s}_i^2(n) \} \quad (5)$$

[0066] onde $\tilde{s}_i(n)$ é um sinal de sub-banda de fonte $s_i(n)$ e

$E\{.\}$ denota expectativa de curto prazo, ou seja,

$$E \{ \tilde{s}_2^2(n) \} = \frac{1}{K} \sum_{n=K/2}^{n+K/2-1} \tilde{s}_2^2(n) \quad (6)$$

[0067] onde K determina o comprimento da média de movimento. Deve ser observado que os valores de potencial de sub-banda $E \{ \tilde{s}_2^2(n) \}$ representam, para cada sinal de fonte, o envelope espectral em função do tempo. A ICLD, $\square L(n)$, é

$$\square L(n) = 10 \log_{10} \frac{\sum_{i=1}^M b_i^2 E \{ \tilde{s}_i^2(n) \}}{\sum_{i=1}^M a_i^2 E \{ \tilde{s}_i^2(n) \}} \quad (7)$$

[0068] Para determinar ICTD e ICC, determina-se a função de correlação cruzada normalizada,

$$\square(n, d) = \frac{E \{ \tilde{x}_1(n) \tilde{x}_2(n+d) \}}{\sqrt{E \{ \tilde{x}_1^2(n) \} E \{ \tilde{x}_2^2(n+d) \}}} \quad (8)$$

[0069] A ICC, $c(n)$, é computada de acordo com

$$c(n) = \max_d \square(n, d) \quad (9)$$

[0070] Para a computação da ICTD, $\square(n)$, a localização do maior pico no eixo de retardo é computada,

$$\square(n) = \arg \max_d \square(n, d) \quad (10)$$

[0071] Agora a questão é como a função de correlação

cruzada normalizada pode ser computada em função dos parâmetros de mixagem. Com (2), (8) pode ser escrito como

$$\square(n, d) = \frac{\sum_{i=1}^M E\{a_i b_i \tilde{s}_i(n - c_i) \tilde{s}_i(n - d_i + d)\}}{\sqrt{E\{\sum_{i=1}^M a_i^2 s_i^2(n - c_i)\} E\{\sum_{i=1}^M b_i^2 s_i^2(n - d_i)\}}} \quad (11)$$

[0072] que é equivalente a

$$\square(n, d) = \frac{\sum_{i=1}^M a_i b_i E\{\tilde{s}_i^2(n)\} \Phi_i(n, d_i - T_i)}{\sqrt{(\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}) (\sum_{i=1}^M b_i^2 E\{s_i^2(n)\})}} \quad (12)$$

[0073] onde a função de autocorrelação normalizada $\square(n, e)$ é

$$\square(n, e) = \frac{E\{s_i(n) s_i(n + e)\}}{E\{s_i^2(n)\}} \quad (13)$$

[0074] e $\square_i = d_i - c_i$. Deve ser observado que, para computar (12) dado (11), assumiu-se que os sinais são fixos em amplo sentido dentro da faixa considerada de retardos, ou seja,

$$\begin{aligned} E \{ \tilde{s}_i^2(n) \} &= E \{ \tilde{s}_i^2(n - c_i) \} \\ E \{ \tilde{s}_i^2(n) \} &= E \{ \tilde{s}_i^2(n - d_i) \} \\ \square \{ \tilde{s}_i(n) \tilde{s}_i(n + c_i - d_i + d) \} &= \square \{ \tilde{s}_i(n - c_i) \tilde{s}_i(n - d_i + d) \} \end{aligned}$$

[0075] Um exemplo numérico de dois sinais de fonte, ilustrando a dependência entre ICTD, ICLD e ICC e o potencial de sub-banda de fonte, é mostrado na Figura 5. Os painéis superior, intermediário e inferior da Figura 5 mostram $\square L(n)$, $\square(n)$ e $c(n)$, respectivamente, em função da razão do potencial de sub-banda dos dois sinais de fonte, $a = \square \{ \tilde{s}_1^2(n) \} / (\square \{ \tilde{s}_1^2(n) \} + \square \{ \tilde{s}_2^2(n) \})$, para diferentes parâmetros de mixagem (4) $\square L_1$, $\square L_2$, $\square_1$ e $\square_2$. Deve ser observado que, quando somente uma fonte tiver potencial na sub-banda ($a = 0$ ou $a = 1$), então $\square L(n)$ e $\square(n)$ computados são iguais aos parâmetros de mixagem ($\square L_1$, $\square L_2$, $\square_1$, $\square_2$).

B.2 Informações colaterais necessárias

[0076] A ICLD (7) depende dos parâmetros de mixagem (a_i , b_i , c_i , d_i) e do potencial de sub-banda de curto prazo das fontes, $\{\tilde{s}_i^2(n)\}$ (6). A função de correlação cruzada normalizada de sub-banda $\square(n, d)$ (12), que é necessária para a computação de ICTD (10) e ICC (9), depende de $\{\tilde{s}_i^2(n)\}$ e adicionalmente da função de autocorrelação normalizada de sub-banda, $\square_i(n, e)$ (13), para cada sinal de fonte. O máximo de $\square(n, d)$ fica na faixa $\min_i\{\square_i\} \leq d \leq \max_i\{\square_i\}$. Para a fonte i com o parâmetro $\square_i = d_i - c_i$ do mixador, a faixa correspondente para a qual a propriedade de sub-banda de sinal de fonte $\square_i(n, e)$ (13) é necessária é

$$\min_l \{\square_l\} - \square_i \leq e \leq \max_l \{\square_l\} - \square_i \quad (14)$$

[0077] Uma vez que os cues ICTD, ICLD e ICC dependem das propriedades de sub-banda de sinal de fonte $\{\tilde{s}_i^2(n)\}$ e $\square_i(n, e)$ na faixa (14), a princípio essas propriedades de sub-banda de sinal de fonte precisam ser transmitidas como informações colaterais. Assumimos que qualquer outro tipo de mixador (por exemplo, mixador com efeitos, mixador de síntese de campo de onda/convolutor, etc.) possui propriedades similares e, dessa forma, essas informações colaterais são úteis também quando outros mixadores que não aquele descrito são utilizados. Para reduzir a quantidade de informações colaterais, seria possível armazenar um conjunto de funções de autocorrelação predefinidas no decodificador e somente transmitir índices para escolha daquelas que correspondem mais precisamente às propriedades do sinal de fonte. Uma primeira versão de nosso algoritmo assume que dentro da faixa (14) $\square_i(n, e) = 1$ e assim (12) é computado utilizando somente os valores de potencial de sub-banda (6) como informações colaterais. Os dados mostrados na Figura 5 foram computados assumindo $\square_i(n, e) = 1$.

[0078] Para reduzir a quantidade de informações

colaterais, a faixa dinâmica relativa dos sinais de fonte é limitada. Em cada tempo, para cada sub-banda, o potencial da fonte mais forte é selecionado. Aachamos suficiente reduzir o limite do potencial correspondente de sub-banda de todas as outras fontes a um valor de 24 dB menor que o potencial de sub-banda mais forte. Assim, a faixa dinâmica do quantificador pode ser limitada a 24 dB.

[0079] Assumindo que os sinais de fonte são independentes, o decodificador pode computar a soma do potencial de sub-banda de todas as fontes como $\square \{\tilde{s}^2(n)\}$. Assim, a princípio, é suficiente transmitir ao decodificador somente os valores de potencial de sub-banda de $M - 1$ fontes, enquanto o potencial de sub-banda da fonte remanescente pode ser computado localmente. Dada essa idéia, a taxa de informações colaterais pode ser discretamente reduzida pela transmissão do potencial de sub-banda de fontes com índices $2 \leq i \leq M$ em relação ao potencial da primeira fonte,

$$\square \tilde{p}_i(n) = 10 \log_{10} \frac{E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}_1^2(n)\}}. \quad (15)$$

[0080] Deve ser observado que a limitação de faixa dinâmica conforme descrito anteriormente é realizada antes de (15). Como alternativa, os valores de potencial de sub-banda poderiam ser normalizados em relação ao potencial de sub-banda do sinal de soma, em vez da normalização em relação ao potencial de sub-banda (15) de uma fonte. Para uma frequência de amostragem de 44,1 kHz, utilizamos 20 sub-bandas e transmitimos para cada sub-banda $\square \tilde{p}_i(n)$ ($2 \leq i \leq M$) aproximadamente a cada 12 ms. Vinte sub-bandas correspondem à metade da resolução espectral do sistema auditivo (uma sub-banda tem duas vezes a largura das "larguras de banda críticas"). Experimentos informais indicam que somente discretas melhoras sejam realizadas utilizando mais de 20 sub-bandas, por exemplo, 40 sub-bandas. O

número de sub-bandas e as larguras de banda de sub-banda são escolhidos de acordo com o tempo e a resolução de frequência do sistema auditivo. Uma baixa implementação de qualidade do esquema requer pelo menos três sub-bandas (baixa, média e alta frequências).

[0081] De acordo com uma configuração em particular, as sub-bandas possuem diferentes larguras de banda, sub-bandas em menores frequências possuem menor largura de banda que as sub-bandas em maiores frequências.

[0082] Os valores de potencial relativo são quantificados com um esquema similar ao quantificador ICLD descrito em [2], resultando em uma taxa de bits de aproximadamente $3(M-1)$ kb/s. A Figura 6 ilustra o processo de geração de informações colaterais (corresponde ao bloco de “Geração de Informações Colaterais” na Figura 2).

[0083] A taxa de informações colaterais pode ser adicionalmente reduzida analisando-se a atividade de cada sinal de fonte e transmitindo somente as informações colaterais associadas à fonte se for ativo.

[0084] Em vez da transmissão dos valores de potencial de sub-banda $\square \{ \tilde{s}_i^2(n) \}$ como informações estatísticas, outras informações que representam os envelopes espectrais dos sinais de fonte poderiam ser transmitidas. Por exemplo, os parâmetros de codificação preditiva linear (LPC) poderiam ser transmitidos ou os outros parâmetros correspondentes, por exemplo, parâmetros de filtro de malha ou parâmetros de par espectral de linha (LSP). O processo de estimativa dos parâmetros de LPC de cada sinal de fonte é ilustrado na Figura 7.

B.3 Computação de $\ddot{\mathbf{g}}_i(n)$

[0085] A Figura 8 ilustra o processo que é utilizado para

recriar os sinais de fonte, dado o sinal de soma (1). Esse processo faz parte do bloco de “Síntese” na Figura 2. Os sinais de fonte individuais são recuperados pela classificação de cada sub-banda do sinal de soma com $g_i(n)$ e pela aplicação de um filtro de descorrelação com resposta de impulso $h_i(n)$,

$$\tilde{\mathfrak{s}}_i(n) = h_i(n) * (g_i(n) \tilde{s}(n)) = h_i(n) * \left(\sqrt{\frac{E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}^2(n)\}}} \tilde{s}(n) \right) \quad (16)$$

[0086] onde $*$ é o operador de convolução linear e $\square \{ \tilde{s}_i^2(n) \}$ é computado com as informações colaterais por

$$E\{\tilde{s}_i^2(n)\} = 1/\sqrt{1 + \sum_{i=2}^M 10^{\frac{\Delta\tilde{p}_i(n)}{10}}} \quad \text{para } i = 1 \quad \text{ou} \\ 10^{\frac{\Delta\tilde{p}_i(n)}{10}} E\{\tilde{s}_1^2(n)\} \quad \text{caso contrário} \quad (17)$$

[0087] Como filtros de descorrelação $h_i(n)$, filtros *comb* complementares, filtros do tipo *all-pass*, retardos ou filtros com respostas de impulso aleatório podem ser utilizados. O objetivo do processo de descorrelação é reduzir a correlação entre os sinais enquanto não modifica a maneira que os formatos individuais de onda são percebidos. Diferentes técnicas de descorrelação levam a diferentes artefatos. Filtros *comb* complementares causam coloração. Todas as técnicas descritas estão espalhando a energia dos transientes no decorrer do tempo, causando artefatos como “pré-ecos”. Devido ao seu potencial de artefatos, as técnicas de descorrelação devem ser aplicadas ao mínimo possível. A próxima seção descreve técnicas e estratégias que requerem menor processamento de descorrelação que a simples geração de sinais independentes $\tilde{\mathfrak{s}}(n)$.

[0088] Um esquema alternativo de geração dos sinais $\tilde{\mathfrak{s}}(n)$ é mostrado na Figura 9. Primeiro, o espectro de $s(n)$ é nivelado por

computação do erro de predição linear $e(n)$. A seguir, devido aos filtros de LPC estimados no codificador, f_i , os filtros do tipo *all-pole* correspondentes são computados como a transformada z inversa de

$$\bar{F}_i(z) = \frac{1}{1 - z^{-1}F_i(z)}$$

[0089] Os filtros *all-pole* resultantes, $\bar{f}_i$, representam o envelope espectral dos sinais de fonte. Se outras informações colaterais que não os parâmetros de LPC forem transmitidas, os parâmetros de LPC primeiro precisam ser computados em função das informações colaterais. Assim como no outro esquema, os filtros de descorrelação h_i são utilizados para produzir sinais de fonte independentes.

IV. IMPLEMENTAÇÕES CONSIDERANDO AS RESTRIÇÕES PRÁTICAS

[0090] Na primeira parte desta seção, é apresentado um exemplo de implementação utilizando um esquema de síntese BCC como um mixador estéreo ou de multicanais. Isto é particularmente interessante uma vez que esse esquema de síntese do tipo BCC faz parte de um padrão ISO/IEC MPEG futuro, denominado “codificação de áudio espacial”. Os sinais de fonte $\tilde{\mathbf{q}}(n)$ não são explicitamente computados neste caso, resultando em menor complexidade computacional. Também, esse esquema oferece o potencial de melhor qualidade de áudio, pois é necessária uma descorrelação efetivamente menor que no caso em que os sinais de fonte $\tilde{\mathbf{q}}(n)$ são explicitamente computados.

[0091] A segunda parte dessa seção discute aspectos quando o esquema proposto é aplicado com qualquer mixador e nenhum processamento de descorrelação é aplicado. Esse esquema é de menor complexidade que um esquema com processamento de descorrelação,

porém pode ter outras desvantagens conforme será discutido.

[0092] Idealmente, poderia ser desejado aplicar um processamento de decorrelação de modo que o $\tilde{\mathbf{Q}}_l(n)$ gerado pudesse ser considerado independente. No entanto, uma vez que o processamento de decorrelação é problemático em termos de introdução de artefatos, poderia ser desejado aplicar um processamento de decorrelação ao mínimo possível. A terceira parte dessa seção discute como a quantidade de processamento de decorrelação problemático pode ser reduzida ao mesmo tempo que se obtém benefícios como se o $\tilde{\mathbf{Q}}_l(n)$ gerado fosse independente.

A. Implementação sem computação explícita de $\tilde{\mathbf{Q}}_l(n)$

[0093] A mixagem é diretamente aplicada ao sinal de soma transmitido (1) sem computação explícita de $\tilde{\mathbf{Q}}_l(n)$. Um esquema de síntese BCC é utilizado para esse propósito. A seguir, estamos considerando o caso estéreo, porém todos os princípios descritos também podem ser aplicados para a geração de sinais de áudio de multicanais.

[0094] Um esquema de síntese BCC estéreo (ou um esquema “estéreo paramétrico”), aplicado para processamento do sinal de soma (1), é mostrado na Figura 10. Seria desejado que o esquema de síntese BCC gerasse um sinal que é percebido similarmente como o sinal de saída de um mixador conforme mostrado na Figura 4. Isso ocorre quando a ICTD, ICLD e ICC entre os canais de saída do esquema de síntese BCC são similares aos *cues* correspondentes que aparecem entre os canais de sinal da saída do mixador (4).

[0095] As mesmas informações colaterais que aquelas anteriormente descritas no esquema mais geral são utilizadas, permitindo que o decodificador compute os valores de potencial de sub-banda de curto prazo $\square \{\tilde{s}_i^2(n)\}$ das fontes. Dado $\square \{\tilde{s}_i^2(n)\}$, os

fatores de ganho g_1 e g_2 na Figura 10 são computados como

$$g_1(n) = \sqrt{\frac{\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}^2(n)\}}} \quad g_2(n) = \sqrt{\frac{\sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\}}{E\{\tilde{s}^2(n)\}}} \quad (18)$$

[0096] de modo que o potencial de sub-banda de saída e a ICLD (7) sejam os mesmos que para o mixador na Figura 4. A ICTD $\square(n)$ é computada de acordo com (10), determinando os retardos D_1 e D_2 na Figura 10,

$$D_1(n) = \max\{-\square(n), 0\} \quad D_2(n) = \max\{\square(n), 0\} \quad (19)$$

[0097] A ICC $c(n)$ é computada de acordo com (9) determinando o processamento de descorrelação na Figura 10. O processamento de descorrelação (síntese de ICC) é descrito em [1]. As vantagens da aplicação do processamento de descorrelação aos canais de saída do mixador em comparação à aplicação deste para geração de $\tilde{\mathbf{d}}(n)$ independente são:

[0098] Geralmente, o número de sinais de fonte M é maior que o número de canais de saída de áudio N . Assim, o número de canais de áudio independentes que precisa ser gerado é menor ao se descorrelacionar os N canais de saída em vez da descorrelação dos M sinais de fonte.

[0099] Geralmente, os N canais de saída de áudio estão correlacionados ($ICC > 0$) e menor processamento de descorrelação pode ser aplicado do que seria necessário para gerar M ou N canais independentes.

[00100] Devido ao menor processamento de descorrelação, uma melhor qualidade de áudio é esperada.

[00101] Uma melhor qualidade de áudio é esperada quando os parâmetros do mixador são restringidos de modo que $a_i^2 + b_i^2 = 1$, ou seja, $G_i = 0$ dB. Nesse caso, o potencial de cada fonte no sinal de soma transmitido (1) é o mesmo que o potencial da mesma fonte no

signal de saída mixado do decodificador. O sinal de saída do decodificador (Figura 10) é o mesmo como se o sinal de saída do mixer (Figura 4) fosse codificado e decodificado por um codificador/decodificador BCC nesse caso. Assim, uma qualidade similar também pode ser esperada.

[00102] O decodificador pode não somente determinar a direção na qual cada fonte deve aparecer, mas também o ganho de cada fonte pode ser variado. O ganho é aumentado escolhendo-se $a_i^2 + b_i^2 > 1$ ($G_i > 0$ dB) e reduzido escolhendo-se $a_i^2 + b_i^2 < 1$ ($G_i < 0$ dB).

B. Sem o uso de processamento de descorrelação

[00103] A restrição da técnica anteriormente descrita é que a mixagem é realizada com um esquema de síntese BCC. Poderia-se imaginar a implementação não somente da síntese de ICTD, ICLD e ICC, mas adicionalmente o processamento de efeitos dentro da síntese BCC.

[00104] No entanto, pode ser desejado que os mixadores e processadores de efeitos existentes possam ser utilizados. Isto também inclui os mixadores de síntese de campo de onda (geralmente denominados "convolutores"). Para utilizar os mixadores e processadores de efeitos existentes, os $\ddot{q}_i(n)$ são computados explicitamente e utilizados como se fossem sinais de fonte originais.

[00105] Ao não aplicar nenhum processamento de descorrelação ($h_i(n) = \square(n)$ em (16)), uma boa qualidade de áudio também pode ser alcançada. Este é um compromisso entre os artefatos introduzidos devido ao processamento de descorrelação e artefatos devido ao fato de que os sinais de fonte $\ddot{q}_i(n)$ estão correlacionados. Quando nenhum processamento de descorrelação é utilizado, a imagem espacial de audição resultante pode sofrer instabilidade [1]. Porém o mixer pode auto-introduzir alguma descorrelação quando

reverberadores ou outros efeitos são utilizados e, assim, há menos necessidade de processamento de decorrelação.

[00106] Se $\ddot{\mathbf{q}}_i(n)$ forem gerados sem processamento de decorrelação, o nível das fontes depende da direção para a qual elas são mixadas em relação às outras fontes. Substituindo-se o algoritmo de variação gradual de amplitude nos mixadores existentes por um algoritmo que compensa essa dependência de nível, o efeito negativo de dependência de alto volume sobre os parâmetros de mixagem pode ser evitado. Um algoritmo de amplitude de compensação de nível é mostrado na Figura 11 que tem como objetivo compensar a dependência de nível de fonte dos parâmetros de mixagem. Devido aos fatores de ganho de um algoritmo de variação gradual de amplitude convencional (por exemplo, a Figura 4), a_i e b_i , os pesos na Figura 11, $\bar{a}_i$ e $\bar{b}_i$ são computados por

$$\begin{aligned}\bar{a}_i(n) &= \sqrt{\frac{\sum_{i=1}^M a_i^2 E\{\tilde{s}_i^2(n)\}}{E\{(\sum_{i=1}^M a_i \tilde{s}_i(n))^2\}}} \quad \text{e} \\ \bar{b}_i(n) &= \sqrt{\frac{\sum_{i=1}^M b_i^2 E\{\tilde{s}_i^2(n)\}}{E\{(\sum_{i=1}^M b_i \tilde{s}_i(n))^2\}}} \quad (20)\end{aligned}$$

[00107] Deve ser observado que $\bar{a}_i$ e $\bar{b}_i$ são computados de modo que o potencial de sub-banda de saída seja o mesmo como se $\ddot{\mathbf{q}}_i(n)$ fossem independentes em cada sub-banda.

c. Reduzindo a quantidade de processamento de decorrelação

[00108] Conforme anteriormente mencionado, a geração de $\ddot{\mathbf{q}}_i(n)$ independente é problemática. Aqui, as estratégias são descritas para aplicação de menor processamento de decorrelação, enquanto efetivamente obtém-se um efeito similar como se os $\ddot{\mathbf{q}}_i(n)$ fossem independentes.

[00109] Considere, por exemplo, um sistema de síntese de

campo de onda conforme mostrado na Figura 12. As posições de fonte virtuais desejadas para $s_1, s_2, \dots, s_6$ ($M = 6$) são indicadas. Uma estratégia para computar $\ddot{\mathbf{q}}_i(n)$ (16) sem gerar M sinais totalmente independentes é:

[00110] Gerar grupos de índices de fonte correspondentes às fontes próximas entre si. Por exemplo, na Figura 8, esses grupos poderiam ser: $\{1\}$, $\{2, 5\}$, $\{3\}$ e $\{4, 6\}$.

[00111] Em cada tempo em cada sub-banda, selecionar o índice de fonte da fonte mais forte,

$$i_{\max} = \max_i E\{\bar{s}(n)\} \quad (21)$$

[00112] Não aplicar nenhum processamento de decorrelação para a parte dos índices de fonte do grupo contendo $i_{\max}$, ou seja, $h_i(n) = \bar{\mathbf{q}}(n)$.

[00113] 3. Para cada outro grupo, escolher o mesmo $h_i(n)$ dentro do grupo.

[00114] O algoritmo descrito modifica as componentes de sinal mais forte "least". Além disso, o número de diferentes $h_i(n)$ que são utilizados é reduzido. Isso é uma vantagem, pois a decorrelação é mais fácil quanto menos canais independentes precisarem ser gerados. A técnica descrita também é aplicável quando os sinais de áudio estéreo ou de multicanais são mixados.

V. ESCALABILIDADE EM TERMOS DE QUALIDADE E TAXA DE BITS

[00115] O esquema proposto transmite somente a soma de todos os sinais de fonte, que pode ser codificada com um codificador de áudio mono convencional. Quando nenhuma compatibilidade mono regressiva é necessária e a capacidade está disponível para transmissão/armazenamento de mais que uma forma de onda de áudio, o esquema proposto pode ser classificado para uso com mais de um canal

de transmissão. Isto é implementado pela geração de vários sinais de soma com diferentes subconjuntos dos referidos sinais de fonte, ou seja, para cada subconjunto de sinais de fonte, o esquema de codificação proposto é aplicado individualmente. Espera-se que a qualidade de áudio melhore conforme o número de canais de áudio transmitidos aumenta, pois menos canais independentes precisam ser gerados por descorrelação de cada canal transmitido (em comparação ao caso de um canal transmitido).

VI. COMPATIBILIDADE REGRESSIVA COM OS FORMATOS DE ÁUDIO ESTÉREO E *SURROUND* EXISTENTES

[00116] Considerar o seguinte cenário de disponibilização de áudio. Um cliente obtém um sinal *surround* estéreo ou de multicanais de qualidade máxima (por exemplo, por meio de um CD de áudio, DVD ou loja de música on-line, etc.). O objetivo é disponibilizar opcionalmente ao cliente a flexibilidade de gerar um mix customizado do conteúdo de áudio obtido sem comprometer a qualidade padrão de *playback* estéreo/*surround*.

[00117] Isto é implementado ao se disponibilizar ao cliente (por exemplo, como opção de compra em uma loja de música on-line) um *stream* de bits de informações colaterais que permite a computação de $\tilde{\mathbf{d}}_i(n)$ em função do referido sinal de áudio estéreo ou de multicanais. O algoritmo de mixagem do cliente é então aplicado ao $\tilde{\mathbf{d}}_i(n)$. Na sequência, duas possibilidades de computação de $\tilde{\mathbf{d}}_i(n)$, dados os sinais de áudio estéreo ou de multicanais, são descritas.

A. Estimativa da soma dos sinais de fonte no receptor

[00118] A forma mais direta de utilizar o esquema de codificação proposto com uma transmissão de áudio estéreo ou de multicanais é ilustrada na Figura 13, onde $y_i(n)$ ($1 \leq i \leq L$) são os L

canais do referido sinal de áudio estéreo ou de multicanais. O sinal de soma das fontes é estimado pela downmixagem dos canais transmitidos para um canal único de áudio. A downmixagem é realizada pela computação da soma dos canais $y_i(n)$ ($1 \leq i \leq L$), ou técnicas mais sofisticadas podem ser aplicadas.

[00119] Para melhor desempenho, recomenda-se que o nível dos sinais de fonte seja adaptado antes da estimativa de $\square \{ \tilde{s}_i^2(n) \}$ (6), de modo que a razão de potencial entre os sinais de fonte se aproxime da razão de potencial com a qual as fontes são contidas no referido sinal estéreo ou de multicanais. Nesse caso, o *downmix* dos canais transmitidos é uma estimativa relativamente boa da soma das fontes (1) (ou uma versão classificada desta).

[00120] Um processo automatizado pode ser utilizado para ajustar o nível das entradas do sinal de fonte do codificador $s_i(n)$ antes da computação das informações colaterais. Esse processo, de forma adaptada ao tempo, estima o nível no qual cada sinal de fonte é contido no referido sinal estéreo ou de multicanais. Antes da computação das informações colaterais, o nível de cada sinal de fonte é então, de forma adaptada ao tempo, ajustado de modo que seja igual ao nível no qual a fonte é contida no sinal de áudio estéreo ou de multicanais.

B. Utilizando os canais transmitidos individualmente

[00121] A Figura 14 mostra uma diferente implementação do esquema proposto com transmissão de sinal *surround* estéreo ou de multicanais. Aqui, os canais transmitidos não são downmixados, porém utilizados individualmente para a geração dos $\ddot{q}(n)$. De forma mais geral, os sinais de sub-banda de $\ddot{q}(n)$ são computados por

$$\ddot{\mathfrak{P}}_i(n) = h_i(n) * (g_i(n) \sum_{l=1}^L w_l(n) \tilde{y}_l(n)) \quad (22)$$

[00122] onde $w_l(n)$ são as combinações lineares específicas de determinação de pesos das sub-bandas dos canais transmitidos. As combinações lineares são escolhidas de modo que $\ddot{\mathfrak{P}}_i(n)$ já sejam descorrelacionadas o máximo possível. Assim, nenhum ou somente um baixo processamento de descorrelação precisa ser aplicado, o que é favorável conforme anteriormente discutido.

VII. APLICAÇÕES

[00123] Já mencionamos diversas aplicações dos esquemas de codificação propostos. Aqui, resumimos e mencionamos algumas outras aplicações.

A. Codificação de áudio para mixagem

[00124] Sempre que sinais de fonte de áudio precisarem ser armazenados ou transmitidos antes da mixagem em sinais de áudio estéreos, de multicanais ou de síntese de campo de onda, o esquema proposto pode ser aplicado. Com a técnica anterior, um codificador de áudio mono seria aplicado a cada sinal de fonte independentemente, resultando em uma taxa de bits que se classifica com o número de fontes. O esquema de codificação proposto pode codificar um alto número de sinais de fonte de áudio com um único codificador de áudio mono mais informações colaterais de taxa de bits relativamente baixa. Conforme descrito na Seção V, a qualidade de áudio pode ser melhorada utilizando-se mais que um canal transmitido se houver disponibilidade de memória/capacidade.

B. Remixagem com metadados

[00125] Conforme descrito na Seção VI, os sinais de áudio estéreo e de multicanais existentes podem ser remixados com a ajuda de informações colaterais adicionais (ou seja, “metadados”). Em vez

de somente vender conteúdo de áudio mixado estéreo e multicanais otimizado, os metadados podem ser vendidos permitindo assim que um usuário faça a remixagem de sua música estéreo e de multicanais. Isto pode ser utilizado, por exemplo, para atenuar os vocais em uma canção para karaokê ou para atenuar instrumentos específicos para que um instrumento seja tocado com a música.

[00126] Mesmo que o armazenamento não seja um problema, o esquema descrito seria muito atraente para permitir a mixagem customizada da música, já que é provável que a indústria fonográfica nunca deseje revelar as gravações de múltiplas faixas. O risco de uso abusivo é muito alto. O esquema proposto permite a capacidade de remixagem sem revelar as gravações de múltiplas faixas.

[00127] Além disso, assim que os sinais estéreos ou de multicanais são remixados, ocorre um certo grau de redução da qualidade, tornando a distribuição ilegal de remixes menos atraente.

c. Conversão de estéreo/multicanais em síntese de campo de onda

[00128] Outra aplicação do esquema descrito na Seção VI é descrita a seguir. O áudio estéreo e de multicanais (por exemplo, *surround* 5.1) que acompanha filmes pode ser estendido para a síntese de campo de onda pela adição de informações colaterais. Por exemplo, Dolby AC-3 (áudio em DVD) pode ser estendido para áudio de codificação de compatibilidade regressiva 5.1 para sistemas de síntese de campo de onda, ou seja, som de *playback surround* 5.1 de DVDs em tocadores convencionais e som de síntese de campo de onda em uma nova geração de tocadores que suporta o processamento das informações colaterais.

VIII. AVALIAÇÕES SUBJETIVAS

[00129] Implementamos um decodificador de tempo real dos

algoritmos propostos na Seção IV-A e IV-B. Um banco de filtro STFT baseado em FFT é utilizado. Um FFT de 1024 pontos e uma janela STFT com tamanho de 768 (com *padding* zero) são utilizados. Os coeficientes espectrais são agrupados de modo que cada grupo represente o sinal com uma largura de banda igual a duas vezes a largura de banda retangular equivalente (ERB). A audição informal revelou que a qualidade de áudio não melhorou notavelmente ao se escolher uma maior resolução de frequência. Uma menor resolução de frequência é favorável, pois resulta na transmissão de menos parâmetros.

[00130] Para cada fonte, a amplitude/variação gradual de retardo e o ganho podem ser ajustados individualmente. O algoritmo foi utilizado para codificação de várias gravações de áudio de múltiplas faixas com 12 a 14 faixas.

[00131] O decodificador permite a mixagem *surround* 5.1 utilizando um mixador de variação gradual da amplitude baseada em vetores (VBAP). A direção e o ganho de cada sinal de fonte podem ser ajustados. O software permite a comutação do tipo *on-the-fly* entre a mixagem do sinal de fonte codificado e a mixagem dos sinais de fonte originais discretos.

[00132] A audição casual geralmente revela pouca ou nenhuma diferença entre a mixagem dos sinais codificados ou dos sinais de fonte originais se, para cada fonte, um ganho G_i igual a zero dB for utilizado. Quanto mais os ganhos de fonte são variados, mais artefatos ocorrem. Uma leve amplificação e atenuação das fontes (por exemplo, até ± 6 dB) ainda é satisfatória. Um cenário crítico ocorre quando todas as fontes são mixadas em um lado e somente uma única fonte no lado oposto. Nesse caso, a qualidade de áudio pode ser reduzida, dependendo da mixagem específica e dos sinais de fonte.

IX. CONCLUSÕES

[00133] Foi proposto um esquema de codificação para codificação conjunta de sinais de fonte de áudio, por exemplo, os canais de uma gravação de múltiplas faixas. O objetivo não é codificar os formatos de onda do sinal de fonte com alta qualidade, caso no qual a codificação conjunta proporcionaria mínimo ganho de codificação uma vez que as fontes de áudio são geralmente independentes. O objetivo é que, quando os sinais de fonte codificados são mixados, um sinal de áudio de alta qualidade é obtido. Considerando as propriedades estatísticas dos sinais de fonte, as propriedades dos esquemas de mixagem e a audição espacial, foi demonstrado que uma melhora significativa do ganho de codificação é obtida por meio da codificação conjunta dos sinais de fonte.

[00134] A melhora no ganho de codificação ocorre devido ao fato de que somente uma forma de onda de áudio é transmitida.

[00135] Além disso, são transmitidas informações colaterais que representam as propriedades estatísticas dos sinais de fonte que são os fatores relevantes que determinam a percepção espacial do sinal final mixado.

[00136] A taxa de informações colaterais é de aproximadamente 3 kbs por sinal de fonte. Qualquer mixador pode ser aplicado com os sinais de fonte codificados, ou seja, mixadores estéreos, de multicanais ou de síntese de campo de onda.

[00137] É uma forma direta classificar o esquema proposto para maiores taxas de bits e qualidade por meio da transmissão de mais que um canal de áudio. Além disso, uma variação do esquema foi proposta, a qual permite a remixagem do referido sinal de áudio estéreo ou de multicanais (e mesmo a alteração do formato de áudio,

a saber, estéreo para multicanais ou síntese de campo de onda).

[00138] As aplicações do esquema proposto são múltiplas. Por exemplo, MPEG-4 poderia ser estendido com o esquema proposto para reduzir a taxa de bits quando mais que um “objeto de áudio natural” (sinal de fonte) precisar ser transmitido. Também, o esquema proposto oferece uma representação compacta do conteúdo para sistemas de síntese de campo de onda. Conforme mencionado, os sinais estéreos ou de multicanais existentes poderiam ser complementados com informações colaterais para permitir que o usuário faça a remixagem dos sinais a seu gosto.

REFERÊNCIAS

[00139] [1] C. Faller, Parametric Coding of Spatial Audio, Ph.D. thesis, Swiss Federal Institute of Technology Lausanne (EPFL), 2004, Ph.D. Thesis No. 3062.

[00140] [2] 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.

REIVINDICAÇÕES:

1. Método para sintetizar uma pluralidade de canais de áudio, **caracterizado** por compreender:

recuperação de um fluxo de áudio pelo menos um sinal de soma representando uma soma de sinais de fonte,

recuperação da informação estatística do fluxo de áudio sobre um ou mais sinais de fonte,

recebimento do fluxo de áudio, ou determinar localmente, parâmetros descrevendo um formato de áudio de saída e parâmetros de mistura de fonte,

computação dos parâmetros do mixer de saída da informação estatística recebida, os parâmetros descrevendo um formato de áudio de saída, e os parâmetros de mistura de fonte, e

sintetização da pluralidade de canais de áudio de pelo menos um sinal de soma com base nos parâmetros do mixer de saída computados,

em que pelo menos o sinal de soma é um sinal mono e a pluralidade de canais de áudio é um sinal estéreo.

2. Método, de acordo com a reivindicação 1, **caracterizado** pela informação estatística representar envelopes espectrais dos sinais de fonte, envelopes espectrais de um ou mais sinais de fonte compreendem parâmetros de filtro de malha ou parâmetros espectrais de linha, ou uma potência relativa como uma função de frequência e tempo da pluralidade de sinais de fonte.

3. Método, de acordo com a reivindicação 2, **caracterizado** pela etapa de computação dos parâmetros do mixer de saída compreender as dicas de computação da pluralidade de canais de áudio e computação dos parâmetros do mixer de saída usando as dicas computadas da pluralidade de canais de áudio.

4. Método, de acordo com a reivindicação 1, **caracterizado** pelos canais de áudio serem sintetizados em um domínio de subfaixa de um banco de filtro.

5. Método, de acordo com a reivindicação 4, **caracterizado** por um número e larguras de banda do domínio de subfaixa serem determinados de acordo com uma resolução espectral e temporal de um sistema auditivo humano.

6. Método, de acordo com a reivindicação 4, **caracterizado** pelo número de subfaixas ser entre 3 e 40.

7. Método, de acordo com a reivindicação 4, **caracterizado** pelas subfaixas no domínio de subfaixa ter diferentes larguras de banda, e subfaixas em frequências inferiores ter larguras de banda menores do que as subfaixas em frequências mais altas.

8. Método, de acordo com a reivindicação 4, **caracterizado** por uma transformada de Fourier de tempo curto com base no banco de filtro ser utilizada e coeficientes espectrais serem combinados para formar grupos de coeficientes espectrais de modo que cada grupo de coeficientes espectrais forme uma subfaixa.

9. Método, de acordo com a reivindicação 1,

caracterizado pela informação estatística ainda incluir funções de autocorrelação.

10. Método, de acordo com a reivindicação 2, **caracterizado** pelos envelopes espectrais serem representados como parâmetros de codificação preditiva linear.

11. Método, de acordo com a reivindicação 3, **caracterizado** pelas dicas computadas serem diferença de nível, diferença de tempo, ou coerência para diferentes frequências e instantes de tempo.

12. Aparelho disposto para sintetizar uma pluralidade de canais de áudio, o aparelho **caracterizado** por compreender:

meios para recuperação de um fluxo de áudio pelo menos um sinal de soma representando uma soma de sinais de fonte;

meios para recuperação da informação estatística do fluxo de áudio sobre um ou mais sinais de fonte;

meios para recebimento do fluxo de áudio, ou determinar localmente, parâmetros descrevendo um formato de áudio de saída e misturar parâmetros;

meios para computação parâmetros do mixer de saída da informação estatística recebida, os parâmetros descrevendo um formato de áudio de saída e os parâmetros de mistura de fonte; e

meios para sintetizar a pluralidade de canais de áudio pelo menos do sinal de soma com base nos parâmetros do mixer de saída computados,

em que pelo menos o sinal de soma é um sinal mono
e a pluralidade de canais de áudio é um sinal estéreo.

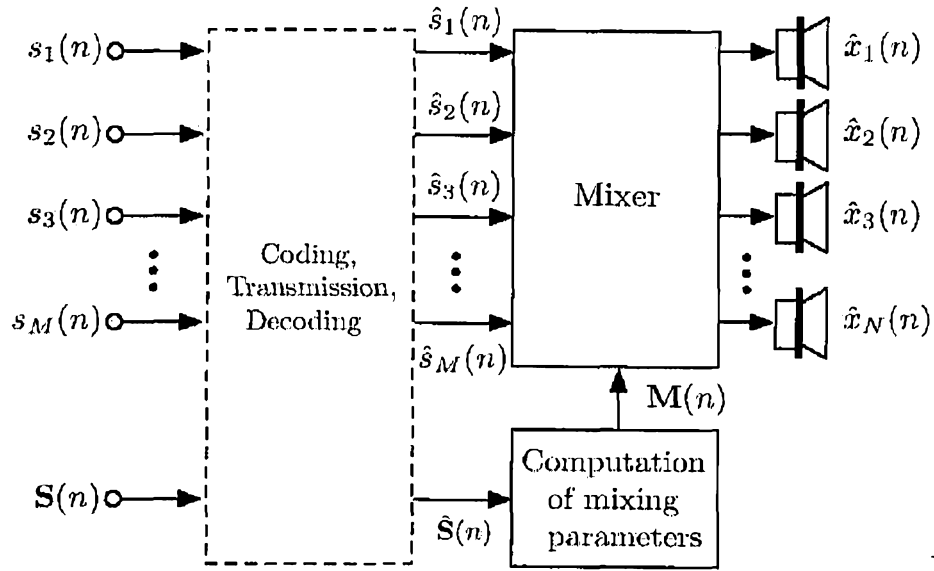


Fig. 1

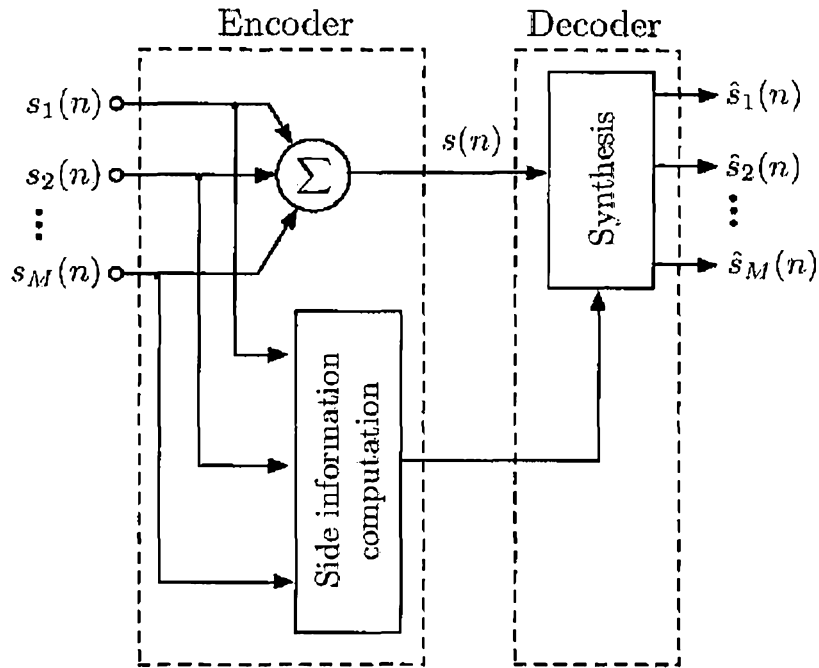


Fig. 2

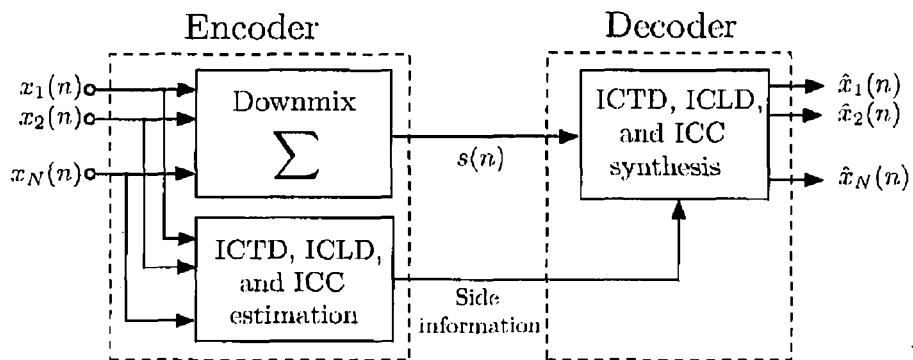


Fig. 3

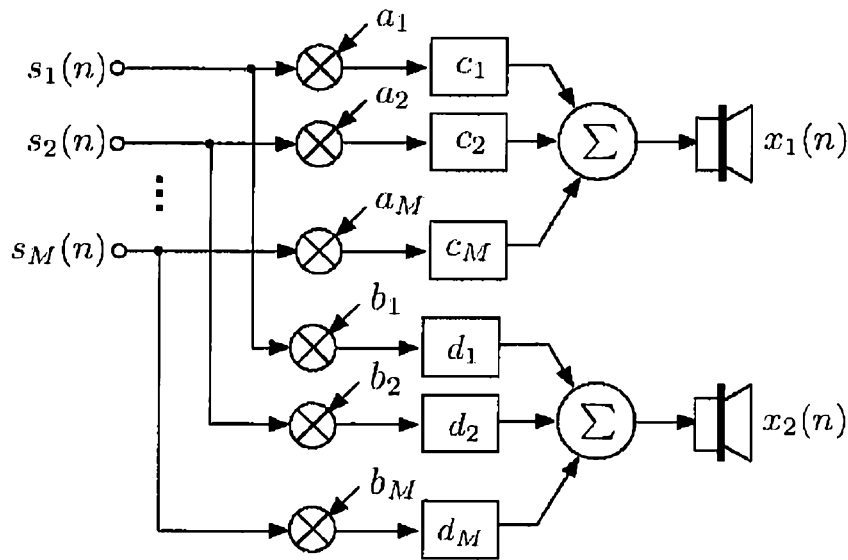


Fig. 4

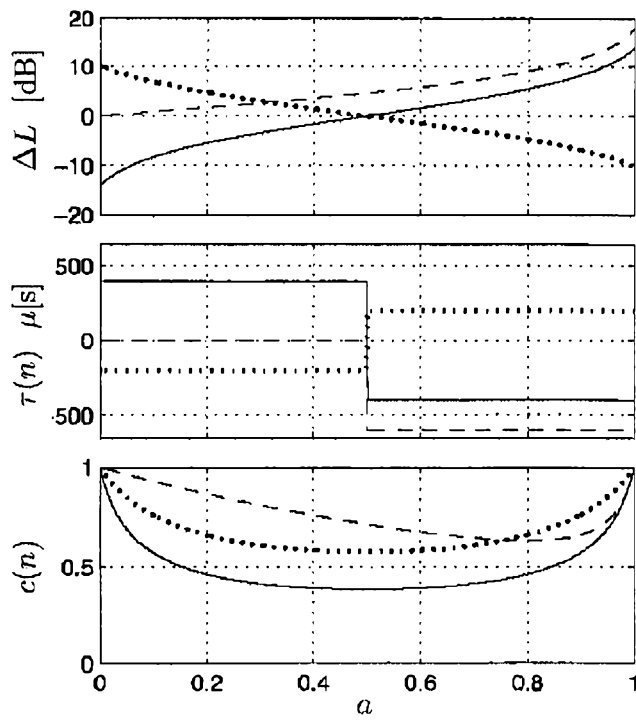


Fig. 5

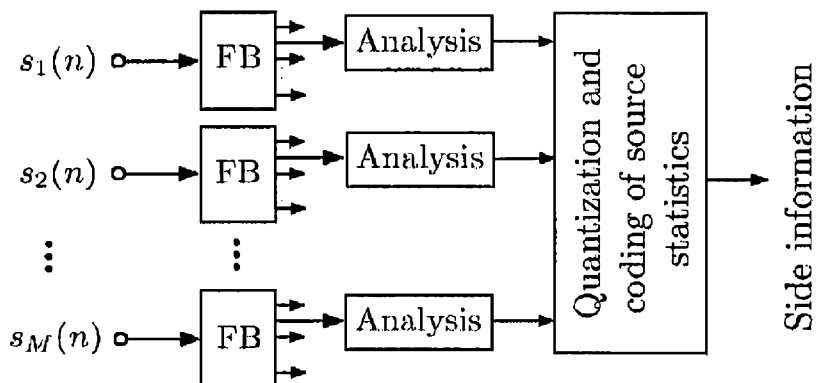


Fig. 6

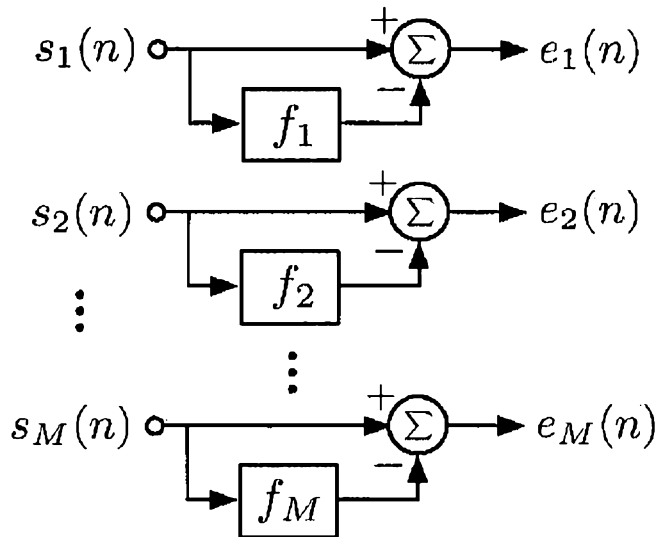


Fig. 7

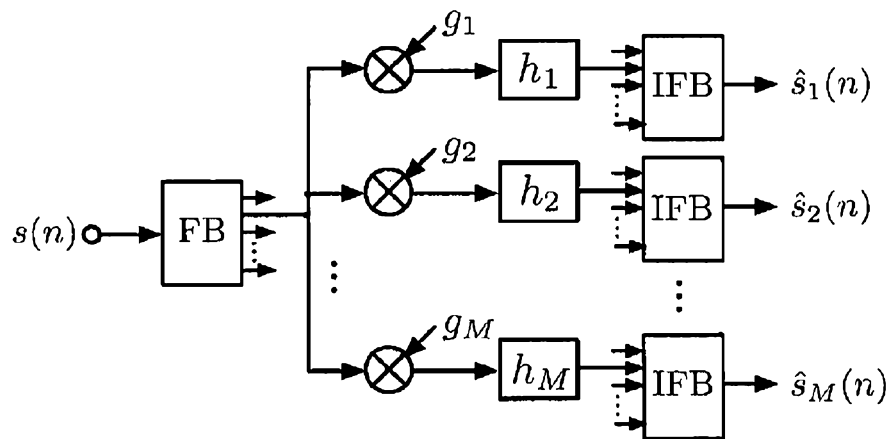


Fig. 8

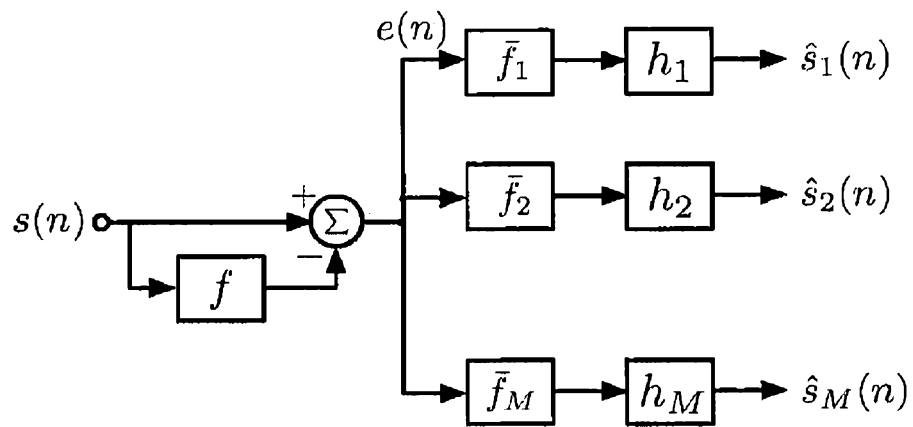
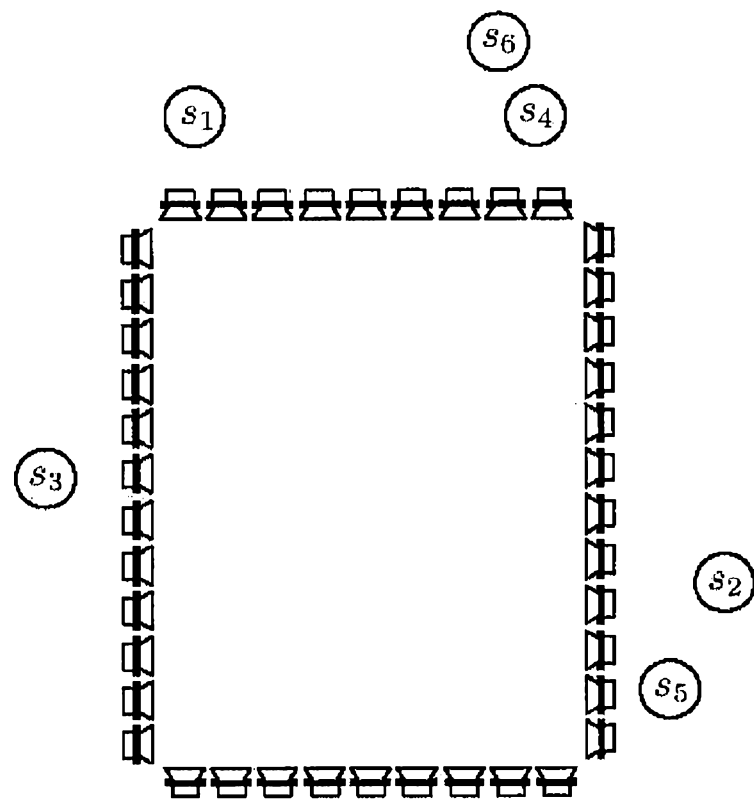
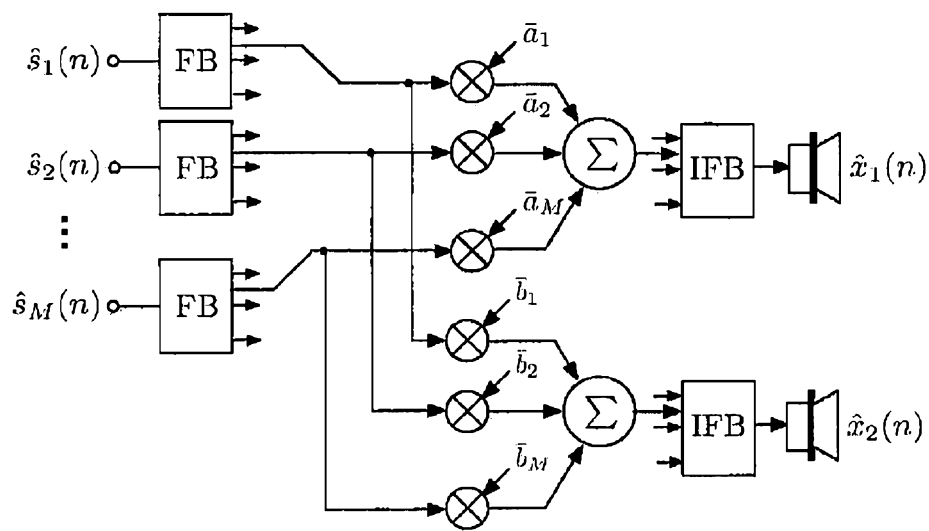
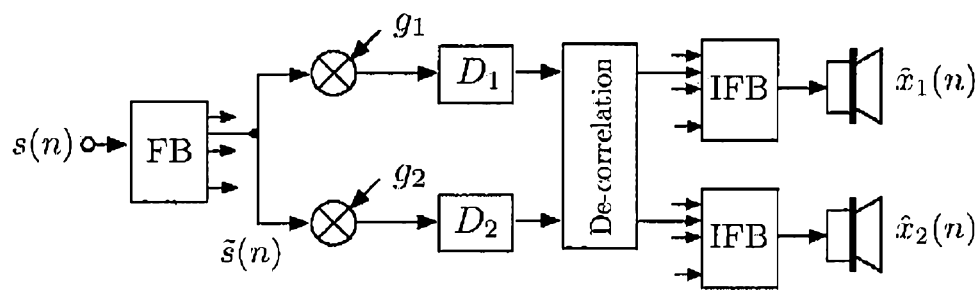


Fig. 9



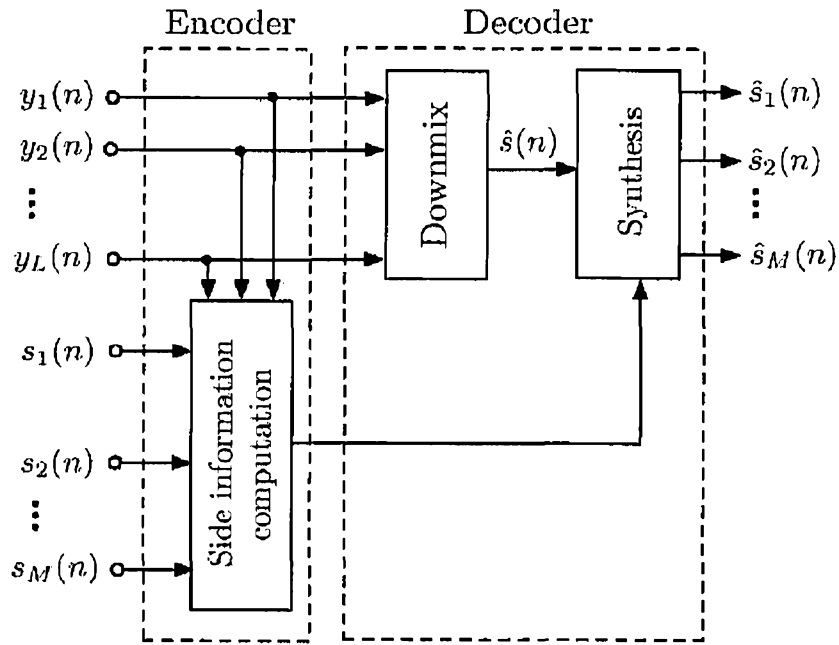


Fig. 13

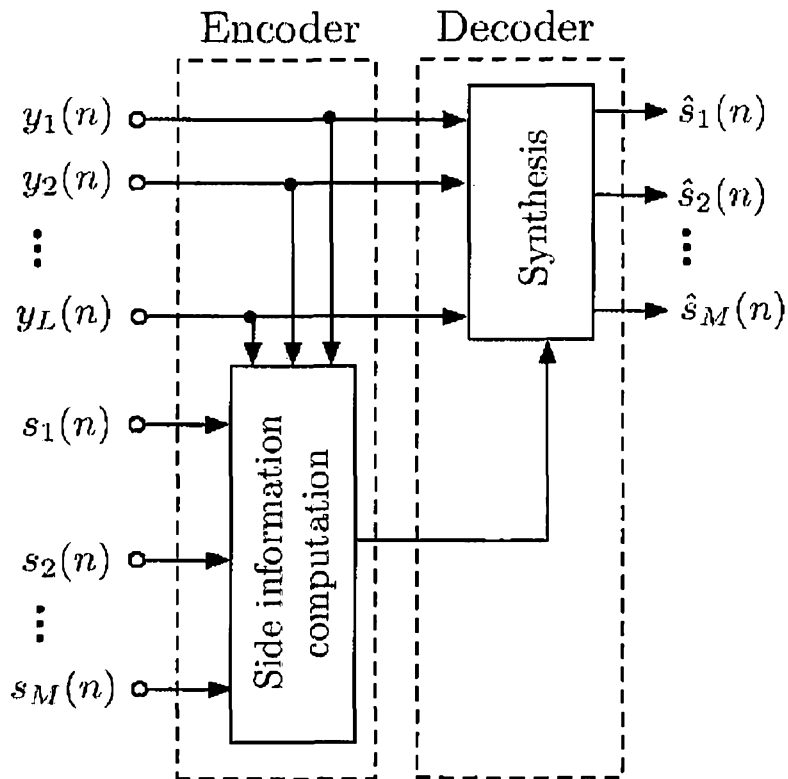


Fig. 14