



**Assinado
Digitalmente**

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA ECONOMIA
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

CARTA PATENTE Nº PI 0910529-8

O INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL concede a presente PATENTE DE INVENÇÃO, que outorga ao seu titular a propriedade da invenção caracterizada neste título, em todo o território nacional, garantindo os direitos dela decorrentes, previstos na legislação em vigor.

(21) Número do Depósito: PI 0910529-8

(22) Data do Depósito: 26/06/2009

(43) Data da Publicação Nacional: 18/08/2020

(51) Classificação Internacional: G10L 19/008; G10L 19/18.

(30) Prioridade Unionista: EP 08017663.9 de 08/10/2008; US 61/079,854 de 11/07/2008; EP 09002271.6 de 18/02/2009.

(54) Título: ESQUEMA DE CODIFICAÇÃO/DECODIFICAÇÃO DE ÁUDIO DE BAIXA TAXA DE BITS QUE APRESENTA COMUTADORES EM CASCATA

(73) Titular: FRAUNHOFER-GELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E. V., Sociedade Alemã. Endereço: HANSASTRASSE 27C, 80686 MÜNCHEN, ALEMANHA(DE); VOICEAGE CORPORATION. Endereço: 750 CHEMIN LUCERNE, SUITE 250, VILLE MONT-ROYAL, QUEBEC H3R 2H6, CANADÁ(CA)

(72) Inventor: GRILL, BERNHARD.; ROCH LEFEBVRE; BESSETTE, BRUNO.; LAPIERRE JIMMY; GOURNAY, PHILIPPE.; REDWAN SALAMI; BAYER, STEFAN.; FUCHS, GUILLAUME; GEYERSBERGER, STEFAN.; GEIGER, RALF.; HILPERT, JOHANNES.; ULRICH KRAEMER; LECOMTE, JÉRÉMIE.; MARKUS MULTRUS; MAX NEUENDORF; HARALD POPP; RETTELACH, NIKOLAUS.

(87) Publicação PCT: WO 2010/003564 de 14/01/2010

Prazo de Validade: 10 (dez) anos contados a partir de 23/02/2021, observadas as condições legais

Expedida em: 23/02/2021

Assinado digitalmente por:

Liane Elizabeth Caldeira Lage

Diretora de Patentes, Programas de Computador e Topografias de Circuitos Integrados

**"ESQUEMA DE CODIFICAÇÃO/DECODIFICAÇÃO DE ÁUDIO DE
BAIXA TAXA DE BITS QUE APRESENTA COMUTADORES EM CASCATA"**

CAMPO DA INVENÇÃO

A presente invenção se refere à codificação de
5 áudio, e particularmente, a esquemas de codificação de áudio de
baixa taxa de bit.

FUNDAMENTOS DA INVENÇÃO

Na técnica, são conhecidos esquemas de
codificação de domínio de frequência como o MP3 ou AAC. Esses
10 codificadores de domínio de frequência têm como base uma conversão
de tempo-domínio/frequência-domínio, um estágio de quantização
subsequente, no qual o erro de quantização é controlado
utilizando-se informações de um módulo psicoacústico, e um estágio
de codificação, no qual os coeficientes de espectro quantizados e
15 as informações secundárias correspondentes são codificados por
entropia com uso de tabelas de códigos.

Por outro lado, existem codificadores que são
bastante adequados ao processamento de fala, como por exemplo, o
AMR-WB+, conforme descrito em 3GPP TS 26.290. Esses esquemas de
20 codificação de fala realizam uma filtragem Linear Preditiva de um
sinal de domínio de tempo. Essa filtragem LP se origina de uma
análise de Previsão Linear do sinal de domínio de tempo de
entrada. Os coeficientes de filtro LP resultantes são então
quantizados/codificados e transmitidos como informação secundária.
25 O processo é conhecido como Codificação de Previsão Linear (LPC).
Na saída do filtro, o sinal residual de previsão ou sinal de erro
de previsão, que também é conhecido como sinal de excitação, é
codificado com uso dos estágios de análise-por-síntese do

codificador ACELP, ou de maneira alternativa, é codificado com uso de um codificador de transformação, que utiliza uma transformação de Fourier com uma sobreposição. A decisão entre a codificação ACELP e a codificação de Excitação de Transformação Codificada, 5 que também é denominada codificação TCX, é feita usando-se um loop fechado ou um algoritmo de loop aberto.

Esquemas de codificação de áudio de domínio de frequência, como o esquema de codificação AAC de alta eficiência, que combina um esquema de codificação AAC e uma técnica de 10 replicação de banda de espectro podem também ser combinados a um estéreo associado ou uma ferramenta de codificação de multicanais que é conhecida pelo termo "MPEG surround".

Por outro lado, codificadores de fala, como o AMR-WB+, também apresentam um estágio de intensificação de alta 15 frequência e uma funcionalidade de estéreo.

Os esquemas de decodificação de domínio de frequência são vantajosos por apresentarem alta qualidade em baixas taxas de bits para sinais musicais. No entanto, a qualidade dos sinais de fala é problemática em baixas taxas de bits.

20 Os esquemas de codificação de fala apresentam alta qualidade para sinais de fala, mesmo em baixas taxas de bits; porém, apresentam má qualidade para sinais musicais em baixas taxas de bits.

SUMÁRIO DA INVENÇÃO

25 O objeto da presente invenção é prover um conceito de codificação/decodificação aperfeiçoado.

Esse objetivo é atingido por um codificador de áudio, de acordo com a reivindicação 1, um método de codificação

de áudio, de acordo com a reivindicação 15, um decodificador, de acordo com a reivindicação 16, um método de decodificação, de acordo com a reivindicação 23, um sinal codificado, de acordo com a reivindicação 24, ou um programa de computador, de acordo com a
5 reivindicação 25.

Um aspecto da presente invenção é um codificador de áudio para codificar um sinal de entrada de áudio, estando o sinal de entrada de áudio em um primeiro domínio, compreendendo: um primeiro ramal de codificação para codificar um sinal de áudio
10 utilizando um primeiro algoritmo de codificação para obter um primeiro sinal codificado; um segundo ramal de codificação para codificar um sinal de áudio utilizando um segundo algoritmo de codificação para obter um segundo sinal codificado, onde o primeiro algoritmo de codificação é diferente do segundo algoritmo
15 de codificação; e um primeiro comutador para transferir entre o primeiro ramal de codificação e o segundo ramal de codificação, de maneira que, para uma parte do sinal de entrada de áudio, seja o primeiro sinal codificado ou o segundo sinal codificado, esteja em um sinal de saída do codificador, onde o segundo ramal de
20 codificação compreende: um conversor para converter o sinal de áudio em um segundo domínio diferente do primeiro domínio, um primeiro ramal de processamento para processar um sinal de áudio no segundo domínio, para obter um primeiro sinal processado; um segundo ramal de processamento para converter um sinal em um
25 terceiro domínio, diferente do primeiro domínio e do segundo domínio, e para processar o sinal no terceiro domínio para obter um segundo sinal processado; e um segundo comutador para transferir entre o primeiro ramal de processamento e o segundo

ramal de processamento, de maneira que, para uma parte da entrada de sinal de áudio no segundo ramal de codificação, ou o primeiro sinal processado ou o segundo sinal processado, esteja no segundo sinal codificado.

5 Um outro aspecto é um decodificador para decodificar um sinal de áudio codificado, sendo que o sinal de áudio codificado compreende um primeiro sinal codificado, um primeiro sinal processado em um segundo domínio, e um segundo sinal processado em um terceiro domínio, onde o primeiro sinal

10 codificado, o primeiro sinal processado, e o segundo sinal processado estão relacionados a diferentes partes de tempo de um sinal de áudio decodificado, e onde um primeiro domínio, o segundo domínio e o terceiro domínio são diferentes uns dos outros, compreendendo: um primeiro ramal de decodificação para decodificar

15 o primeiro sinal codificado com base no primeiro algoritmo de codificação; um segundo ramal de decodificação para decodificar o primeiro sinal processado ou o segundo sinal processado, onde o segundo ramal de decodificação compreende um primeiro ramal de processamento inverso para processar inversamente o primeiro sinal

20 processado para obter um primeiro sinal processado inverso no segundo domínio; um segundo ramal de processamento inverso para processar inversamente o segundo sinal processado para obter um segundo sinal processado inverso no segundo domínio; um primeiro combinador para combinar o primeiro sinal processado inverso e o

25 segundo sinal processado inverso para obter um sinal combinado no segundo domínio; e um conversor para converter o sinal combinado ao primeiro domínio; e um segundo combinador para combinar o sinal convertido no primeiro domínio e a saída do primeiro sinal

convertido pelo primeiro ramal de decodificação para obter um sinal de saída decodificado no primeiro domínio.

Em uma configuração preferida da presente invenção, são providos dois comutadores em ordem sequencial, onde
5 um primeiro comutador decide entre codificação no domínio de espectro usando um codificador de domínio de frequência e codificação no domínio de LPC, isto é, processamento do sinal na saída de um estágio de análise de LPC. O segundo comutador é provido para transferir no domínio de LPC, a fim de codificar o
10. sinal de domínio de LPC, ou no domínio de LPC de maneira a utilizar um codificador de ACELP ou codificando o sinal de domínio de LPC em um domínio de espectro de LPC, que demanda um conversor para converter o sinal de domínio de LPC em um domínio de espectro de LPC, que é diferente de um domínio de espectro, pois o domínio
15 de espectro de LPC apresenta o espectro de um sinal filtrado de LPC e não o espectro do sinal de domínio de tempo.

O primeiro comutador decide entre dois ramais de processamento, onde um ramal é principalmente motivado por um modelo de depósito e/ou um modelo psicoacústico, isto é, por
20 mascaramento auditivo, e o outro é principalmente motivado por um modelo fonte e por cálculos de segmentos SNR. Por exemplo, um ramal tem um codificador de domínio de frequência e o outro ramal tem um codificador à base de LPC, como um codificador de fala. O modelo fonte é em geral o processamento de fala e, portanto, a LPC
25 é usada comumente.

O segundo comutador decide novamente entre dois ramais de processamento; porém, em um domínio diferente do primeiro domínio de ramal "externo". E novamente, um ramal

"interno" é principalmente motivado por um modelo fonte ou por cálculos de SNR, e o outro ramal "interno" pode ser motivado por um modelo de depósito e/ou um modelo psicoacústico, isto é, por meio do mascaramento ou pelo menos incluindo aspectos de código de domínio de frequência/de espectro. Por exemplo, um ramal "interno" tem um conversor de codificador/de espectro de domínio de frequência, e o outro ramal tem uma codificação de codificador no outro domínio, como por exemplo, o domínio de LPC, onde este codificador é, por exemplo, um quantizador/escalonador CELP ou ACELP que processa um sinal de entrada sem conversão de espectro.

Uma outra configuração preferida é um codificador de áudio compreendendo um primeiro ramal de codificação direcionado por depósito de informações, como por exemplo, um ramal de codificação de domínio de espectro, uma segunda fonte de informações ou ramal de codificação direcionado por SNR, como por exemplo, um ramal de codificação de domínio de LPC, e um comutador para transferir entre o primeiro ramal de codificação e o segundo ramal de codificação, onde o segundo ramal de codificação compreende um conversor em um domínio específico, diferente do domínio de tempo, como por exemplo, um estágio de análise de LPC que gera um sinal de excitação, e onde o segundo ramal de codificação compreende ainda um domínio específico, como o ramal de processamento de domínio de LPC, e um domínio de espectro específico, como o ramal de processamento de domínio de espectro de LPC, e um comutador adicional para transferir entre o ramal de codificação de domínio específico e o ramal de codificação de domínio de espectro específico.

Uma outra configuração da invenção é um

decodificador de áudio que inclui um primeiro domínio, como por exemplo, um ramal de decodificação de domínio de espectro, um segundo domínio, como por exemplo, um ramal de decodificação de domínio de LPC para decodificar um sinal, como por exemplo, um
5 sinal de excitação no segundo domínio, e um terceiro domínio, como por exemplo, um ramal de decodificador de espectro de LPC para decodificar um sinal, como um sinal de excitação, em um terceiro domínio, como um domínio de espectro de LPC, onde o terceiro domínio é obtido realizando-se uma conversão de frequência a
10 partir do segundo domínio, onde um primeiro comutador para o sinal do segundo domínio e o sinal do terceiro domínio é provido, e onde um segundo comutador para alternar entre o primeiro decodificador de domínio e o decodificador para o segundo domínio ou terceiro domínio é provido.

15

BREVE DESCRIÇÃO DOS DESENHOS

Configurações preferidas da presente invenção são descritas subsequentemente em relação aos desenhos anexos, nos quais:

A Fig. 1a é um diagrama de bloco de um esquema de
20 codificação, de acordo com um primeiro aspecto da presente invenção;

A Fig. 1b é um diagrama de bloco de um esquema de decodificação, de acordo com o primeiro aspecto da presente invenção;

25

A Fig. 1c é um diagrama de bloco de um esquema de codificação, de acordo com um outro aspecto da presente invenção;

A Fig. 2a é um diagrama de bloco de um esquema de codificação, de acordo com um segundo aspecto da presente

invenção;

A Fig. 2b é um diagrama esquemático de um esquema de decodificação, de acordo com o segundo aspecto da presente invenção;

5 A Fig. 2c é um diagrama de bloco de um esquema de codificação, de acordo com um outro aspecto da presente invenção;

A Fig. 3a ilustra um diagrama de bloco de um esquema de codificação, de acordo com um outro aspecto da presente invenção;

10 A Fig. 3b ilustra um diagrama de bloco de um esquema de decodificação, de acordo com o outro aspecto da presente invenção;

A Fig. 3c ilustra uma representação esquemática do equipamento/método de codificação com comutadores em cascata;

15 A Fig. 3d ilustra um diagrama esquemático de um equipamento ou método para decodificação, no qual são utilizados combinadores em cascata;

A Fig. 3e mostra uma ilustração de um sinal de domínio de tempo e uma representação correspondente do sinal codificado que ilustra regiões de fusão de transmissão curtas que
20 estão incluídas em ambos os sinais codificados;

A Fig. 4a ilustra um diagrama de bloco com um comutador posicionado antes dos ramais de codificação;

A Fig. 4b ilustra um diagrama de bloco de um
25 esquema de codificação com o comutador posicionado subsequente à codificação dos ramais;

A Fig. 4c ilustra um diagrama de bloco para uma configuração preferida de combinador;

A Fig. 5a ilustra uma forma de onda de um segmento de fala de domínio de tempo como um segmento de sinal semelhante ao periódico ou do tipo de impulso;

5 A Fig. 5b ilustra um espectro do segmento da Fig. 5a;

A Fig. 5c ilustra um segmento de fala de domínio de tempo de fala sem voz como exemplo de um segmento do tipo de ruído;

10 A Fig. 5d ilustra um espectro da forma de onda do domínio de tempo da Fig. 5c;

A Fig. 6 ilustra um diagrama de bloco de uma análise por codificador CELP de síntese;

As Figs. 7a a 7d ilustram sinais de excitação com voz/sem voz como exemplo de sinais do tipo de impulso;

15 A Fig. 7e ilustra um estágio de LPC do lado do codificador provendo informações de previsão de curto prazo e o sinal de erro de previsão (excitação);

A Fig. 7f ilustra uma outra configuração de um dispositivo de LPC para gerar um sinal ponderado;

20 A Fig. 7g ilustra uma implementação para transformar um sinal ponderado em um sinal de excitação aplicando-se uma operação de ponderação inversa e uma análise de excitação subsequente, conforme exigido no conversor 537 da Fig. 2b;

25 A Fig. 8 ilustra um diagrama de bloco de um algoritmo de multicanais associado, de acordo com uma configuração da presente invenção;

A Fig. 9 ilustra uma configuração preferida de um algoritmo de extensão de largura de banda;

A Fig. 10a ilustra uma descrição detalhada do comutador ao executar uma decisão de *loop* aberto; e

A Fig. 10b mostra uma ilustração do comutador ao operar em um modo de decisão de *loop* fechado.

5

DESCRIÇÃO DETALHADA DAS CONFIGURAÇÕES PREFERIDAS

A Fig. 1a ilustra uma configuração da invenção com dois comutadores em cascata. Um sinal mono, um sinal estéreo ou um sinal multicanais é inserido em um comutador 200. O comutador 200 é controlado por um estágio de decisão 300. O
10 estágio de decisão recebe, como uma entrada, uma entrada de sinal no bloco 200. Alternativamente, o estágio de decisão 300 pode também receber uma informação secundária que é incluída no sinal mono, no sinal estéreo ou no sinal de multicanais, ou é pelo menos associada a esse sinal, onde existem as informações, que foram,
15 por exemplo, geradas na produção original do sinal mono, do sinal estéreo ou do sinal de multicanais.

O estágio de decisão 300 aciona o comutador 200 para alimentar um sinal em uma parte de codificação de frequência 400 ilustrada em um ramal superior da Fig. 1a ou em uma parte de
20 codificação de domínio de LPC 500 ilustrada em um ramal inferior na Fig. 1a. Um elemento importante do ramal de codificação de domínio de frequência é um bloco de conversão de espectro 410 que é operante para converter um sinal de saída de estágio de pré-processamento comum (conforme será discutido posteriormente) em um
25 domínio de espectro. O bloco de conversão de espectro pode incluir um algoritmo MDCT, um QMF, um algoritmo FFT, uma análise Wavelet ou um banco de filtros, como por exemplo, um banco de filtros criticamente amostrado com um determinado número de canais de

banco de filtros, onde os sinais de sub-banda deste banco de filtros podem ser sinais de valor real ou sinais de valor complexo. A saída deste bloco de conversão de espectro 410 é codificada com o uso de um codificador de áudio de espectro 421, que pode incluir blocos de processamento como eles são conhecidos pelo esquema de codificação de AAC.

Em geral, o processamento no ramal 400 é um processamento em um modelo à base de percepção ou modelo de depósito de informações. Assim, esse modelo representa o sistema auditivo humano recebendo som. Contrariamente a isso, o processamento no ramal 500 serve para gerar um sinal no domínio de excitação, residual ou de LPC. Em geral, o processamento no ramal 500 é um processamento em um modelo de fala ou um modelo de geração de informações. Para sinais de fala, esse modelo é um modelo do sistema de geração de fala/som humano que gera o som. Se, no entanto, um som de uma fonte diferente que demanda um modelo de geração de som diferente for codificado, então o processamento no ramal 500 pode ser diferente.

No ramal de codificação inferior 500, um elemento fundamental é um dispositivo de LPC 510, que produz uma informação de LPC que é usada para controlar as características de um filtro de LPC. Esta informação de LPC é transmitida a um decodificador. O sinal de saída do estágio de LPC 510 é um sinal de domínio de LPC que é composto de um sinal de excitação e/ou um sinal ponderado.

O dispositivo de LPC em geral produz um sinal de domínio de LPC, que pode ser qualquer sinal no domínio de LPC, como o sinal de excitação da Fig. 7e ou um sinal ponderado da Fig. 7f, ou qualquer outro sinal, que tenha sido gerado por meio da

aplicação de coeficientes de filtro LPC a um sinal de áudio. Além disso, um dispositivo de LPC pode também determinar esses coeficientes, e pode ainda quantizar/decodificar esses coeficientes.

5 A decisão no estágio de decisão pode ser adaptativa ao sinal, de maneira que o estágio de decisão execute uma discriminação de música/fala e controle o comutador 200 de tal maneira que sinais de música sejam inseridos no ramal superior 400, e sinais de fala sejam inseridos no ramal inferior 500. Em
10 uma configuração, o estágio de decisão está alimentando suas informações de decisão em um fluxo de bits de saída de maneira que um decodificador possa utilizar essas informações de decisão para executar as operações de decodificação corretas.

Esse decodificador está ilustrado na Fig. 1b. A
15 saída de sinal pelo codificador de áudio de espectro 421 é, após a transmissão, inserida em um decodificador de áudio de espectro 431. A saída do decodificador de áudio de espectro 431 é inserida em um conversor de domínio de tempo 440. Analogamente, a saída do ramal de codificação de domínio de LPC 500 da Fig. 1a recebida no
20 lado do decodificador e processada pelos elementos 531, 533, 534 e 532 para obtenção de um sinal de excitação de LPC. O sinal de excitação de LPC é inserido em um estágio de síntese de LPC 540, que recebe, como uma outra entrada, as informações de LPC geradas pelo estágio de análise de LPC correspondente 510. A saída do
25 conversor de domínio de tempo 440 e/ou a saída do estágio de síntese de LPC 540 são inseridas em um comutador 600. O comutador 600 é controlado por meio de um sinal de controle de comutador que foi, por exemplo, gerado pelo estágio de decisão 300, ou que foi

externamente provido, como por exemplo, por um criador do sinal mono, sinal estéreo ou sinal de multicanais original. A saída do comutador 600 é um sinal mono, sinal estéreo ou sinal de multicanais completo.

5 O sinal de entrada no comutador 200 e o estágio de decisão 300 podem ser um sinal mono, um sinal estéreo, um sinal de multicanais ou em geral um sinal de áudio. Dependendo da decisão que pode ser derivada do sinal de entrada do comutador 200 ou de qualquer fonte externa, como por exemplo um produtor do
10 sinal de áudio original subjacente à entrada de sinal no estágio 200, o comutador transfere entre o ramal de codificação de frequência 400 e o ramal de codificação de LPC 500. O ramal de codificação de frequência 400 inclui um estágio de conversão de espectro 410 e um estágio de quantização/codificação conectado
15 subsequentemente 421. O estágio de quantização/codificação pode incluir quaisquer das funcionalidades conhecidas de codificadores de frequência-domínio modernos, como por exemplo, o codificador AAC. Além disso, a operação de quantização no estágio de quantização/codificação 421 pode ser controlada por meio de um
20 módulo psicoacústico que gera informações psicoacústicas, como limite de mascaramento psicoacústico sobre a frequência, onde essas informações são inseridas no estágio 421.

 No ramal de codificação de LPC, o sinal de saída do comutador é processado através de um estágio de análise de LPC
25 510 que gera informações secundárias de LPC e um sinal de domínio de LPC. O codificador de excitação inventivamente inclui um outro comutador para transferir o processamento adicional do sinal de domínio de LPC entre uma operação de quantização/codificação 522

no domínio de LPC ou um estágio de quantização/codificação 524, que está processando valores no domínio de espectro de LPC. Para essa finalidade, um conversor de espectro 523 é provido na entrada do estágio de quantização/codificação 524. O comutador 521 é controlado em modo de *loop* aberto ou em modo de *loop* fechado, dependendo de configurações específicas, como por exemplo as descritas na especificação técnica do AMR-WB+.

Para o modo de controle de *loop* fechado, o codificador inclui também um quantizador/codificador inverso 531 para o sinal de domínio de LPC, um quantizador/codificador inverso 533 para o sinal de domínio de espectro de LPC e um conversor de espectro inverso 534 para a saída do item 533. Tanto os sinais codificados como os decodificados novamente nos ramos de

processamento do segundo ramal de codificação são inseridos no dispositivo de controle de comutador 525. No dispositivo de controle de comutador 525, esses dois sinais de saída são comparados um ao outro e/ou a uma função-alvo, ou uma função-alvo é calculada, que possa ter como base uma comparação da distorção em ambos os sinais, de maneira que o sinal com a distorção inferior seja usado para decidir qual posição o comutador 521 deve ocupar. Alternativamente, caso ambos os ramos provejam taxas de bits não constantes, o ramal que provê a taxa de bits inferior pode ser selecionado mesmo quando a relação sinal-ruído desse ramal for inferior à relação sinal-ruído do outro ramal.

Alternativamente, a função-alvo poderia utilizar como entrada a relação sinal-ruído de cada sinal e uma taxa de bits de cada sinal, e/ou outros critérios, para encontrar a melhor decisão para um objetivo específico. Se, por exemplo, o objetivo for tal que a

taxa de bits deva ser a mais baixa possível, a função-alvo dependeria muito da taxa de bits da saída de dois sinais pelos elementos 531, 534. No entanto, quando o objetivo principal é ter a melhor qualidade para uma determinada taxa de bits, o controle de comutador 525 pode, por exemplo, descartar cada sinal que esteja acima da taxa de bits permitida, e quando ambos os sinais estiverem abaixo da taxa de bits permitida, o controle de comutador selecionaria o sinal com a melhor relação sinal-ruído, isto é, com as menores distorções de quantização/codificação.

O esquema de decodificação, de acordo com a presente invenção está, conforme já foi informado, ilustrado na Fig. 1b. Para cada um dos três tipos possíveis de sinal de saída,

existe um estágio de decodificação/requantização específico 431, 531 ou 533. Enquanto o estágio 431 produz um espectro de tempo

que é convertido no domínio de tempo usando o conversor de frequência/tempo 440, o estágio 531 produz um sinal de domínio de LPC, e o item 533 produz um espectro de LPC. Para garantir que os sinais de entrada no comutador 532 estejam ambos no domínio de LPC, é provido o espectro de LPC/conversor de LPC. Os dados de

saída do comutador 532 são transformados novamente no domínio de tempo, com uso de um estágio de síntese de LPC 540, que é controlado através de informações de LPC geradas no lado do codificador e transmitidas. Depois, subsequente ao bloco 540,

ambos os ramos têm informações de domínio de tempo que são alternadas de acordo com um sinal de controle de alternância, a fim de finalmente obter um sinal de áudio, como por exemplo, um sinal mono, um sinal estéreo ou um sinal de multicanais, o que depende da entrada de sinal no esquema de codificação da Fig. 1a.

A Fig. 1c ilustra uma outra configuração com uma disposição diferente do comutador 521, semelhante ao princípio da Fig. 4b.

A Fig. 2a ilustra um esquema de codificação preferido, de acordo com um segundo aspecto da invenção. Um esquema de pré-processamento comum conectado à entrada do comutador 200 pode incluir um bloqueio de estéreo surround/associado 101 que gere, como saída, parâmetros de estéreo conjuntos e um sinal mono de saída, que é gerado por meio de downmix do sinal de entrada, que é um sinal com dois ou mais canais. Em geral, o sinal na saída do bloco 101 pode também ser um sinal com mais canais, porém, devido à funcionalidade por meio de downmix o bloco 101, o número de canais na saída do bloco 101 será menor que o número de canais inseridos no bloco 101.

O esquema de pré-processamento comum pode incluir alternativamente ao bloco 101 ou adicionalmente ao bloco 101, um estágio de extensão de largura de banda 102. Na configuração da Fig. 2a, a saída do bloco 101 é inserida no bloco de extensão de largura de banda 102, que, no codificador da Fig. 2a, produz um sinal limitado de banda como o sinal de banda baixa ou o sinal de baixa frequência em sua saída. Preferivelmente, esse sinal é reduzido (por exemplo, por um fator de dois) também. Além disso, para a banda alta da entrada de sinal no bloco 102, os parâmetros de extensão de largura de banda como os parâmetros de envelope de espectro, parâmetros de filtragem inversa, parâmetros de piso de ruído, etc. como os conhecidos pelo perfil HE-AAC de MPEG-4 são gerados e encaminhados a um multiplexor de fluxo de bits 800.

Preferivelmente, o estágio de decisão 300 recebe

a entrada de sinal no bloco 101 ou entrada no bloco 102, para decidir entre, por exemplo, um modo de música ou um modo de fala. No modo de música, o ramal de codificação superior 400 é selecionado, enquanto no modo de fala, o ramal de codificação inferior 500 é selecionado. Preferivelmente, o estágio de decisão controla também o bloco de estéreo conjunto 101 e/ou o bloco de extensão de largura de banda 102 para adaptar a funcionalidade desses blocos ao sinal específico. Assim, quando o estágio de decisão determina que uma determinada parte de tempo do sinal de entrada é do primeiro modo, como por exemplo o modo de música, características específicas do bloco 101 e/ou bloco 102 podem ser controladas pelo estágio de decisão 300. Alternativamente, quando o estágio de decisão 300 determinar que o sinal está em um modo de voz ou, em geral, em um segundo modo de domínio de LPC, características específicas dos blocos 101 e 102 podem ser controladas de acordo com a saída do estágio de decisão.

Preferivelmente, a conversão de espectro do ramal de codificação 400 é feita usando-se uma operação de MDCT, a qual, ainda mais preferivelmente, é a operação de MDCT com distorção temporal, onde a potência ou, em geral a potência de distorção pode ser controlada entre zero e alta potência de distorção. Em uma potência de distorção zero, a operação de MDCT no bloco 411 é uma operação de MDCT direta conhecida na técnica. A potência de distorção de tempo com a informação secundária de distorção de tempo podem ser transmitidas / inseridas no multiplexor de fluxo de bits 800 como informações secundárias.

No ramal de codificação LCP, o codificador de domínio de LPC pode incluir um núcleo de ACELP 526 que calcula um

ganho de altura, um retardo de altura e/ou informações de livro de códigos, como por exemplo, índice e ganho de livro de códigos. O modo TCX conhecido do 3GPP TS 26.290 incorre em um processamento de um sinal perceptualmente ponderado no domínio de transformação.

- 5 Um sinal ponderado de transformação de Fourier é quantizado usando-se uma quantização em treliça de taxa múltipla dividida (VQ algébrico) com quantização de fator de ruído. Uma transformação é calculada em 1024, 512 ou 256 janelas de amostra. O sinal de excitação é recuperado ao filtrar inversamente o sinal ponderado
- 10 quantizado através de um filtro de ponderação inversa. No primeiro ramal de codificação 400, um conversor de espectro preferivelmente inclui uma operação de MDCT adaptada especificamente com determinadas funções de janela seguidas de um estágio de codificação de quantização/entropia, que pode ser composto de um
- 15 ~~único estágio de quantização de vetor,~~ mas preferivelmente é um codificador de quantizador/entropia escalar combinado semelhante ao quantizador/codificador do ramal de codificação de domínio de frequência, isto é, no item 421 da Fig. 2a.

- No segundo ramal de codificação, existe o bloco
- 20 de LPC 510 seguido de um comutador 521, novamente seguido de um bloco de ACELP 526 ou um bloco de TCX 527. O ACELP é descrito no 3GPP TS 26.190 e a TCX é descrito no 3GPP TS 26.290. Em geral, o bloco de ACELP 526 recebe um sinal de excitação de LCP calculado por um procedimento descrito na Fig. 7e. O bloco de TCX 527 recebe
 - 25 um sinal ponderado gerado da maneira que mostra a Fig. 7f.

No TCX, a transformação é aplicada ao sinal ponderado computado filtrando-se o sinal de entrada através de um filtro de ponderação à base de LPC. As configurações preferidas do

filtro de ponderação usado são dadas por $(1-A(z/y))/(1-\mu z^{-1})$.

Assim, o sinal ponderado é um sinal de domínio de LPC e sua transformação é um domínio de espectro de LPC. O sinal processado pelo bloco de ACELP 526 é o sinal de excitação, e é diferente do sinal processado pelo bloco 527, porém, ambos os sinais estão no domínio de LPC.

No lado do decodificador ilustrado na Fig. 2b, depois da transformação de espectro inversa no bloco 537, o inverso do filtro de ponderação é aplicado, isto é, $(1-\mu z^{-1})/(1-A(z/y))$. Depois, o sinal é filtrado por $(1-A(z))$ para ir para o domínio de excitação de LPC. Assim, a conversão para o bloco de domínio de LPC 534 e o bloco TCX^{-1} 537 incluem transformação inversa e depois filtragem por $(1-\mu z^{-1})/(1-A(z/y))$ $(1-A(z))$ para converter do domínio ponderado para o domínio de excitação.

Embora o item 510 das Fig. 1a, 1c, 2a, 2c ilustre um único bloco, o bloco 510 pode produzir diferentes sinais, contanto que esses sinais estejam no domínio de LPC. O modo real do bloco 510, como por exemplo, o modo de sinal de excitação ou o modo de sinal ponderado podem depender da condição real do comutador. Alternativamente, o bloco 510 pode ter dois dispositivos de processamento paralelos, onde um dispositivo é implementado de maneira semelhante à Fig. 7e e o outro dispositivo é implementado como na Fig. 7f. Logo, o domínio de LPC na saída de 510 pode representar o sinal de excitação de LPC ou o sinal ponderado de LPC, ou qualquer outro sinal de domínio de LPC.

No segundo ramal de codificação (ACELP/TCX) da Fig. 2a ou 2c, o sinal é preferivelmente pré-enfatizado através de um filtro $1-0,68z^{-1}$ antes da codificação. No decodificador

ACELP/TCX da Fig. 2b o sinal sintetizado é desenfaturado com o filtro $1/(1-0,68z^{-1})$. A pré-ênfase pode fazer parte do bloco de LPC 510 onde o sinal é pré-enfatizado antes da análise e quantização de LPC. De maneira semelhante, a desenfaturação pode ser parte do

5 bloco de síntese de LPC LPC^{-1} 540.

A Fig. 2c ilustra uma outra configuração para a implementação da Fig. 2a, porém com uma disposição diferente do comutador 521, semelhante ao princípio da Fig. 4b.

Em uma configuração preferida, o primeiro

10 comutador 200 (vide Fig. 1a ou 2a) é controlado através de uma decisão de loop aberto (como na Fig. 4a), e o segundo comutador é controlado através de uma decisão de loop fechado (como na figura 4b).

Por exemplo, a Fig. 2c tem o segundo comutador

- 15 colocado depois dos ramais de ACELP e TCX, como na Fig. 4b. Então, no primeiro ramal de processamento, o primeiro domínio de LPC representa a excitação de LPC, e no segundo ramal de processamento, o segundo domínio de LPC representa o sinal ponderado de LPC. Isto é, o primeiro sinal de domínio de LPC é

20 obtido por filtragem através de $(1-A(z))$ para converter para o domínio residual de LPC, enquanto o segundo sinal de domínio de LPC é obtido filtrando-se pelo filtro $(1-A(z/y))/(1-\mu z^{-1})$ para converter para o domínio ponderado de LPC.

A Fig. 2b ilustra um esquema de decodificação

25 correspondente ao esquema de codificação da Fig. 2a. O fluxo de bits gerado pelo multiplexor de fluxo de bits 800 da Fig. 2a é inserido em um desmultiplexor de fluxo de bits 900. Dependendo de uma informação originada, por exemplo, de um fluxo de dados

através de um bloco de detecção de modo 601, um comutador do lado do decodificador 600 é controlado ou para sinais para frente do ramal superior ou para sinais do ramal mais baixo para o bloco de extensão de largura de banda 701. O bloco de extensão de largura de banda 701 recebe, do desmultiplexor de fluxo de bits 900, informações secundárias e, com base nessas informações secundárias e na saída da decisão de modo 601, reconstroi a banda alta com base na saída de banda baixa pelo comutador 600.

O sinal de banda completo gerado pelo bloco 701 é inserido no estágio de processamento *surround*/estéreo associado 702, que reconstroi dois canais estéreos ou vários multicanais. Em geral, o bloco 702 produzirá mais canais que os que foram inseridos neste bloco. Dependendo da aplicação, a entrada no bloco 702 pode até incluir dois canais, como por exemplo, em um modo estéreo, e pode até incluir mais canais, contanto que a saída por esse bloco tenha mais canais que a entrada nesse bloco.

Foi demonstrado que o comutador 200 alterna entre ambos os ramais, de maneira que somente um ramal receba um sinal para processar e o outro ramal não receba um sinal para processar. Em uma configuração alternativa, no entanto, o comutador pode também ser disposto subsequente, por exemplo, ao codificador de áudio 421 e ao codificador de excitação 522, 523, 524, o que significa que ambos os ramais 400, 500 processam o mesmo sinal em paralelo. Para não dobrar a taxa de bits, no entanto, somente a saída de sinal por um desses ramais de codificação 400 ou 500 é seleccionada para ser gravada no fluxo de bits de saída. O estágio de decisão operará então de maneira que o sinal gravado no fluxo de bits minimize uma determinada função de custo, onde a função de

custo pode ser a taxa de bits gerada ou a distorção perceptual gerada, ou uma função de custo de taxa/distorção combinadas. Portanto, seja neste modo ou no modo ilustrado nas Figuras, o estágio de decisão pode também operar em um modo de *loop* fechado, para garantir que, por fim, somente a saída do ramal de codificação seja gravada no fluxo de bits, o qual possui para uma determinada distorção perceptual a taxa de bits mais baixa ou, para um determinado fluxo de bits, possui a distorção perceptual mais baixa. No modo de *loop* fechado, a entrada de feedback pode ser derivada de saídas dos três blocos quantizadores/escalares 421, 522 e 424, da Fig. 1a.

Na implementação com dois comutadores, isto é, o primeiro comutador 200 e o segundo comutador 521, é preferido que a resolução de tempo para o primeiro comutador seja mais baixa que a resolução de tempo para o segundo comutador. Em outras palavras, os blocos do sinal de entrada no primeiro comutador, que podem ser alternados através de uma operação do comutador, são maiores que os blocos alternados pelo segundo comutador que funciona no domínio de LPC. Por exemplo, o comutador de domínio de frequência/domínio de LPC 200 pode alternar blocos de uma extensão de 1024 amostras, e o segundo comutador 521 pode alternar blocos com 256 amostras cada.

Apesar de algumas das Fig. 1a a 10b estarem ilustradas como diagramas de bloco de um equipamento, essas figuras são simultaneamente uma ilustração de um método, onde as funcionalidades de bloco correspondem às etapas do método.

A Fig. 3a ilustra um codificador de áudio para gerar um sinal de áudio codificado como saída do primeiro ramal de

codificação 400 e um segundo ramal de codificação 500. Além disso, o sinal de áudio codificado preferivelmente inclui informações secundárias, como por exemplo, os parâmetros de pré-processamento pelo estágio de pré-processamento comum ou, conforme foi discutido em relação às Figuras anteriores, informações de controle de comutador.

Preferivelmente, o primeiro ramal de codificação é operacional, para codificar um sinal intermediário de áudio 195 de acordo com um primeiro algoritmo de codificação, onde o primeiro algoritmo de codificação possui um modelo de depósito de informação. O primeiro ramal de codificação 400 gera o primeiro sinal de saída do codificador, que é uma representação de informação de espectro codificada do sinal intermediário de áudio 195.

Além disso, o segundo ramal de codificação 500 é adaptado para codificar o sinal intermediário de áudio 195 de acordo com um segundo algoritmo de codificação, sendo que o segundo algoritmo de codificação possui um modelo de fonte de informação e gera, em um segundo sinal de saída do codificador, parâmetros codificados para o modelo de fonte de informação que representa o sinal de áudio intermediário.

O codificador de áudio inclui ainda o estágio de pré-processamento comum para pré-processar um sinal de entrada de áudio 99, para obter o sinal intermediário de áudio 195. Especificamente, o estágio de pré-processamento comum é operacional para processar o sinal de entrada de áudio 99 de maneira que o sinal intermediário de áudio 195, isto é, a saída do algoritmo de pré-processamento comum, seja uma versão condensada

do sinal de entrada de áudio.

Um método preferido de codificação de áudio para gerar um sinal de áudio codificado inclui uma etapa de codificação 400 de um sinal intermediário de áudio 195, de acordo com um primeiro algoritmo de codificação, sendo que o primeiro algoritmo de codificação possui um modelo de depósito de informação e gera, em um primeiro sinal de saída, informações de espectro codificadas que representam o sinal de áudio; uma etapa da codificação 500 de um sinal intermediário de áudio 195, de acordo com um segundo algoritmo de codificação, sendo que o segundo algoritmo de codificação possui um modelo de fonte de informação e gera, em um segundo sinal de saída, parâmetros codificados para o modelo de fonte de informação que representa o sinal intermediário 195, e uma etapa de pré-processamento 100 comum de um sinal de entrada de áudio 99 para obter o sinal intermediário de áudio 195, onde, na etapa de pré-processamento comum, o sinal de entrada de áudio 99 é processado de maneira que o sinal intermediário de áudio 195 seja uma versão condensada do sinal de entrada de áudio 99, onde o sinal de áudio codificado inclui, para uma determinada parte do sinal de áudio, o primeiro sinal de saída ou o segundo sinal de saída. O método preferivelmente inclui a outra etapa que codifica uma determinada parte do sinal intermediário de áudio, seja usando o primeiro algoritmo de codificação ou usando o segundo algoritmo de codificação, ou codificando o sinal usando-se ambos os algoritmos e produzindo em um sinal codificado o resultado do primeiro algoritmo de codificação ou o resultado do segundo algoritmo de codificação.

Em geral, o algoritmo de codificação de áudio

usado no primeiro ramal de codificação 400 reflete e representa a situação em um depósito de áudio. O depósito de uma informação de áudio é normalmente o ouvido humano. O ouvido humano pode ser representado como um analisador de frequência. Portanto, o primeiro ramal de codificação produz informações de espectro codificadas. Preferivelmente, o primeiro ramal de codificação inclui ainda um modelo psicoacústico para aplicar também um limite de mascaramento psicoacústico. Esse limite de mascaramento psicoacústico é usado ao se quantizar valores de espectro de áudio onde, preferivelmente, a quantização é realizada de maneira que um ruído de quantização seja introduzido quantizando-se os valores de áudio de espectro, que ficam ocultos abaixo do limite de mascaramento psicoacústico.

O segundo ramal de codificação representa um modelo de fonte de informação, que reflete a geração de som de áudio. Portanto, os modelos de fonte de informação podem incluir um modelo de fala que é refletido por um estágio de análise de LPC, isto é, pela transformação de um sinal de domínio de tempo em um domínio de LPC, e pelo subsequente processamento do sinal residual de LPC, isto é, o sinal de excitação. Modelos alternativos de fonte de som, no entanto, são modelos de fonte de som para representar um determinado instrumento ou qualquer outro gerador de som, como por exemplo, uma fonte de som específica que existe no mundo real. Uma seleção entre diferentes modelos de fonte de som pode ser feita quando vários modelos de fonte de som estiverem disponíveis, por exemplo, com base em um cálculo SNR, isto é, com base em um cálculo, qual dos modelos de fonte é o mais adequado para codificar uma determinada parte de tempo e/ou parte

de frequência de um sinal de áudio. Preferivelmente, no entanto, a alternância entre os ramaís de codificação é feita no domínio de tempo, isto é, uma determinada parte de tempo é codificada usando-se um modelo e uma determinada parte de tempo diferente do sinal intermediário é codificada usando-se o outro ramal de codificação.

Os modelos de fonte de informação são representados por determinados parâmetros. Em relação ao modelo de fala, os parâmetros são parâmetros de LPC e parâmetros de excitação codificados, quando um codificador de fala moderno como a AMR-WB+ é considerado. A AMR-WB+ inclui um codificador ACELP e um codificador TCX. Neste caso, os parâmetros de excitação codificados podem ser ganho global, piso de ruído e códigos de extensões variáveis.

A Fig. 3b ilustra um decodificador que corresponde ao codificador ilustrado na Fig. 3a. Em geral, a Fig. 3b ilustra um decodificador de áudio para decodificar um sinal de áudio codificado para obter-se um sinal de áudio decodificado 799. O decodificador inclui o primeiro ramal de decodificação 450 para decodificar um sinal codificado que foi codificado de acordo com um primeiro algoritmo de codificação com um modelo de depósito de informação. O decodificador de áudio inclui ainda um segundo ramal de decodificação 550 para decodificar um sinal de informação codificado que foi codificado de acordo com um segundo algoritmo de codificação com um modelo de fonte de informação. O decodificador de áudio inclui ainda um combinador para combinar sinais de saída do primeiro ramal de decodificação 450 e do segundo ramal de decodificação 550, para obter um sinal combinado. O sinal combinado que está ilustrado na Fig. 3b como sinal

intermediário de áudio decodificado 699 é inserido em um estágio de pós-processamento comum para pós-processar o sinal intermediário de áudio decodificado 699, que é a saída de sinal combinado pelo combinador 600, de maneira que um sinal de saída do estágio de pré-processamento comum seja uma versão expandida do sinal combinado. Assim, o sinal de áudio decodificado 799 possui um teor de informação intensificado, em comparação ao sinal intermediário de áudio decodificado 699. Essa expansão de informação é provida pelo estágio de pós-processamento comum com o auxílio dos parâmetros de pré/pós-processamento que podem ser transmitidos de um codificador para um decodificador, ou que podem ser derivados do próprio sinal intermediário de áudio decodificado. Preferivelmente, no entanto, os parâmetros de pré/pós-processamento são transmitidos de um codificador para um decodificador, pois esse procedimento permite uma melhor qualidade do sinal de áudio decodificado.

A Fig. 3c ilustra um codificador de áudio para codificar um sinal de entrada de áudio 195, que pode ser igual ao sinal de áudio intermediário 195 da Fig. 3a de acordo com a configuração preferida da presente invenção. O sinal de entrada de áudio 195 está presente em um primeiro domínio que pode, por exemplo, ser o domínio de tempo, mas que pode também ser qualquer outro domínio, como por exemplo, um domínio de frequência, um domínio de LPC, um domínio de espectro de LPC ou qualquer outro domínio. Em geral, a conversão de um domínio para o outro domínio é feita por um algoritmo de conversão como qualquer um dos algoritmos de conversão de tempo/frequência ou algoritmos de conversão de frequência/tempo bem conhecidos.

Uma transformação alternativa do domínio de tempo, por exemplo no domínio de LPC, é o resultado de filtragem por LPC de um sinal de domínio de tempo que resulta em um sinal residual de LPC ou sinal de excitação. Qualquer outra operação de

5 filtragem que produza um sinal filtrado que tenha um impacto sobre um número substancial de amostras de sinal antes da transformação pode ser usada como um algoritmo de transformação se for o caso. Portanto, a ponderação de um sinal de áudio com uso de um filtro de ponderação à base de LPC é uma outra transformação, que gera um

10 sinal no domínio de LPC. Em uma transformação de tempo/frequência, a modificação de um único valor de espectro terá um impacto sobre todos os valores de domínio de tempo antes da transformação.

Analogamente, uma modificação de qualquer amostra de domínio de tempo terá um impacto sobre cada amostra de domínio de frequência.

15 De maneira semelhante, ~~uma modificação de uma amostra do sinal de~~ excitação em uma situação de domínio de LPC terá, devido à extensão do filtro de LPC, um impacto em um número substancial de amostras antes da filtração de LPC. De maneira semelhante, uma modificação de uma amostra antes de uma transformação de LPC terá

20 um impacto sobre muitas amostras obtidas por essa transformação de LPC, devido ao efeito de memória inerente do filtro de LPC.

O codificador de áudio da Fig. 3c inclui um primeiro ramal de codificação 400 que gera um primeiro sinal codificado. Esse primeiro sinal codificado pode estar em um quarto

25 domínio que é, na configuração preferida, o domínio de espectro de tempo, isto é, o domínio que é obtido quando um sinal de domínio de tempo é processado através de uma conversão de tempo/frequência.

Portanto, o primeiro ramal de codificação 400 para codificar um sinal de áudio utiliza um primeiro algoritmo de codificação para obter um primeiro sinal codificado, onde o primeiro algoritmo de codificação pode incluir ou não um algoritmo
5 de conversão tempo/frequência.

O codificador de áudio inclui ainda um segundo ramal de codificação 500 para codificar um sinal de áudio. O segundo ramal de codificação 500 utiliza um segundo algoritmo de codificação para obter um segundo sinal codificado, que é
10 diferente do primeiro algoritmo de codificação.

O codificador de áudio inclui, ainda, um primeiro comutador 200 para alternar entre o primeiro ramal de codificação 400 e o segundo ramal de codificação 500, de maneira que para uma parte do sinal de entrada de áudio, ou o primeiro sinal codificado
15 na saída do bloco 400 ou o segundo sinal codificado na saída do segundo ramal de codificação está incluído em um sinal de saída do codificador. Assim, quando para uma determinada parte do sinal de entrada de áudio 195, o primeiro sinal codificado no quarto domínio for incluído no sinal de saída do codificador, o segundo
20 sinal codificado que é ou o primeiro sinal processado no segundo domínio ou o segundo sinal processado no terceiro domínio, não será incluído no sinal de saída do codificador. Isso garante que esse codificador é eficiente em termos de taxa de bits. Em configurações, quaisquer partes de tempo do sinal de áudio que
25 forem incluídas em dois sinais codificados diferentes são pequenas, em comparação a uma extensão de estrutura de um estrutura, como será discutido em relação à Fig. 3e. Essas partes pequenas são úteis para uma fusão de transmissão de um sinal

codificado para outro sinal codificado, no caso de um evento de alternância, para reduzir artefatos que possam ocorrer sem nenhuma fusão de transmissão. Portanto, sem contar a região da fusão de transmissão, cada bloco de domínio de tempo é representado por um
5 sinal codificado de somente um único domínio.

Como ilustra a Fig. 3c, o segundo ramal de codificação 500 inclui um conversor 510 para converter o sinal de áudio do primeiro domínio, isto é, sinal 195 em um segundo domínio. Além disso, o segundo ramal de codificação 500 inclui um
10 primeiro ramal de processamento 522 para processar um sinal de áudio no segundo domínio, para obter um primeiro sinal processado que está, preferivelmente, também no segundo domínio, de maneira que o primeiro ramal de processamento 522 não execute uma mudança de domínio.

15 O segundo ramal de codificação ~~500~~ inclui, ainda, um segundo ramal de processamento 523, 524, que converte o sinal de áudio do segundo domínio em um terceiro domínio, que é diferente do primeiro domínio, e que é também diferente do segundo domínio, e que processa o sinal de áudio no terceiro domínio para
20 obter um segundo sinal processado na saída do segundo ramal de processamento 523, 524.

Além disso, o segundo ramal de codificação inclui um segundo comutador 521 para transferir entre o primeiro ramal de processamento 522 e o segundo ramal de processamento 523, 524, de
25 maneira que, para uma parte da entrada do sinal de áudio no segundo ramal de codificação, ou o primeiro sinal processado no segundo domínio ou o segundo sinal processado no terceiro domínio fica no segundo sinal codificado.

A Fig. 3 ilustra um decodificador correspondente para decodificar um sinal de áudio codificado gerado pelo codificador da Fig. 3c. Em geral, cada bloco do sinal de áudio de primeiro domínio é representado por um sinal de segundo domínio, ou um sinal de terceiro domínio, um sinal codificado de quarto domínio, além de uma região de fusão de transmissão opcional que é preferivelmente curta, em comparação à extensão de uma estrutura, para que se obtenha um sistema que esteja tanto quanto possível no limite crítico de amostragem. O sinal de áudio codificado inclui o primeiro sinal codificado, um segundo sinal codificado em um segundo domínio e um terceiro sinal codificado em um terceiro domínio, onde o primeiro sinal codificado, o segundo sinal codificado e o terceiro sinal codificado estão todos relacionados a diferentes partes de tempo do sinal de áudio decodificado, e onde o segundo domínio, o terceiro domínio e o primeiro domínio de um sinal de áudio decodificado são diferentes uns dos outros.

O decodificador inclui um primeiro ramal de decodificação para decodificar com base no primeiro algoritmo de codificação. O primeiro ramal de decodificação é ilustrado em 431, 440 na Fig. 3d, e inclui preferivelmente um conversor de frequência/tempo. O primeiro sinal codificado está preferivelmente em um quarto domínio e é convertido no primeiro domínio, que é o domínio para o sinal de saída decodificado.

O decodificador da Fig. 3d inclui ainda um segundo ramal de decodificação que inclui vários elementos. Esses elementos são um primeiro ramal de processamento inverso 531 para processar inversamente o segundo sinal codificado, para obter um primeiro sinal processado inverso no segundo domínio, na saída do

bloco 531. O segundo ramal de decodificação inclui ainda um segundo ramal de processamento inverso 533, 534, para processar inversamente um terceiro sinal codificado, para obter um segundo sinal processado inverso no segundo domínio, onde o segundo ramal de processamento inverso inclui um conversor para converter do

5 terceiro domínio no segundo domínio.

O segundo ramal de decodificação inclui ainda um primeiro combinador 532 para combinar o primeiro sinal processado inverso e o segundo sinal processado inverso para obter um sinal

10 no segundo domínio, onde esse sinal combinado é, em um primeiro momento, influenciado somente pelo primeiro sinal processado inverso e é, posteriormente, influenciado somente pelo segundo sinal processado inverso.

O segundo ramal de decodificação inclui ainda um

15 conversor 540 para converter o ~~sinal combinado~~ ao primeiro domínio.

Por fim, o decodificador ilustrado na Fig. 3d inclui um segundo combinador 600 para combinar o primeiro sinal decodificado do bloco 431, 440 e o sinal de saída do conversor

20 540, para obter um sinal de saída decodificado no primeiro domínio. Novamente, o sinal de saída decodificado do primeiro domínio é, em um primeiro momento, influenciado somente pela saída de sinal pelo conversor 540, e é, posteriormente, influenciado somente pela saída do primeiro sinal decodificado pelo bloco 431,

25 440.

Essa situação é ilustrada, de uma perspectiva do codificador, na Fig. 3e. A parte superior da Fig. 3e ilustra, na representação esquemática, um sinal de áudio do primeiro domínio,

como por exemplo, um sinal de áudio do domínio de tempo, onde o índice de tempo aumenta da esquerda para a direita e o item 3 pode ser considerado um fluxo de amostras de áudio que representa o sinal 195 na Fig. 3c. A Fig. 3e ilustra as estruturas 3a, 3b, 3c, 3d, que podem ser geradas alternando-se entre o primeiro sinal codificado e o primeiro sinal processado e o segundo sinal processado, conforme ilustrado no item 4 da Fig. 3e. O primeiro sinal codificado, o primeiro sinal processado e segundo sinal processado estão todos em domínios diferentes, e para garantir que o comutador entre os diferentes domínios não resulte em um artefato no lado do decodificador, as estruturas 3a, 3b do sinal do domínio de tempo apresentam uma faixa de sobreposição que é indicada como uma região de fusão de transmissão, e essa região de fusão de transmissão está presente nas estruturas 3b e 3c. No entanto, não existe essa região de fusão de transmissão entre as estruturas 3d, 3c, o que significa que a estrutura 3d é também representada por um segundo sinal processado, isto é, um sinal no terceiro domínio, e não há nenhuma mudança de domínio entre as estruturas 3c e 3d. Portanto, em geral, é preferido não prover uma região de fusão de transmissão onde não houver mudança de domínio, e prover uma região de fusão de transmissão, isto é, uma parte do sinal de áudio que é codificada por dois sinais subsequentes codificados/processados quando não existir uma mudança de domínio, isto é, uma ação de alternância de nenhum dos dois comutadores. Preferivelmente, as fusões de transmissões são executadas para outras mudanças de domínio.

Na configuração, na qual o primeiro sinal codificado ou o segundo sinal processado tiver sido gerado por um

processamento de MDCT com, por exemplo, sobreposição de 50%, cada amostra de domínio de tempo é incluída em duas estruturas subsequentes. No entanto, devido às características da MDCT, isso não resulta em suspensão, pois a MDCT é um sistema amostrado criticamente. Nesse contexto, amostrado criticamente significa que o número de valores de espectro é o mesmo que o número de valores de domínio de tempo. A MDCT é vantajosa porque o efeito de passagem é provido sem uma região de passagem específica, de maneira que uma passagem de um bloco de MDCT para o próximo bloco de MDCT é provida sem nenhuma suspensão, que violaria a exigência de amostragem crítica.

Preferivelmente, o primeiro algoritmo de codificação do primeiro ramal de codificação baseia-se em um modelo de depósito de informação, e o segundo algoritmo de codificação do segundo ramal de codificação baseia-se em uma fonte de informação ou em um modelo SNR. Um modelo SNR é um modelo que não está especificamente relacionado a um mecanismo de geração de som específico, mas que é um modo de codificação que pode ser selecionado entre uma pluralidade de modos de codificação com base, por exemplo, em uma decisão de *loop* fechado. Assim, um modelo SNR é qualquer modelo de codificação disponível, mas que não necessariamente tem que estar relacionado à constituição física do gerador de som, mas que é qualquer modelo de codificação parametrizada diferente do modelo de depósito de informação, que pode ser selecionado por uma decisão de *loop* fechado e, especificamente, comparando-se diferentes resultados de SNR de diferentes modelos.

Conforme ilustra a Fig. 3c, um controlador 300,

525 é provido. Esse controlador pode incluir as funcionalidades do estágio de decisão 300 da Fig. 1a e adicionalmente, pode incluir a funcionalidade do dispositivo de controle de alternância 525 da Fig. 1a. Em geral, o controlador é para controlar o primeiro comutador e o segundo comutador de maneira adaptativa ao sinal. O controlador é operante para analisar uma entrada de sinal no primeiro comutador ou saída pelo primeiro ou segundo ramal de codificação ou sinais obtidos por codificação e decodificação pelo primeiro e pelo segundo ramal de codificação, em relação a uma função-alvo. Alternativamente, ou adicionalmente, o controlador é operante para analisar a entrada de sinal no segundo comutador ou saída pelo primeiro ramal de processamento ou pelo segundo ramal de processamento, ou obtido por processamento e processamento inverso do primeiro ramal de processamento e do segundo ramal de processamento, novamente em relação a uma função-alvo.

Em uma configuração, o primeiro ramal de codificação ou o segundo ramal de codificação incluem um *aliasing* que introduz um algoritmo de conversão de tempo/frequência como uma MDCT ou um algoritmo de MDCT, que é diferente de uma transformação de FFT direta, a qual não introduz um *aliasing*. Além disso, um ou ambos os ramaís incluem um bloco de codificador quantizador/de entropia. Especificamente, somente o segundo ramal de processamento do segundo ramal de codificação inclui o conversor de tempo/frequência que introduz uma operação de *aliasing*, e o primeiro ramal de processamento do segundo ramal de processamento inclui um quantizador e/ou codificador de entropia, e não introduz nenhum *aliasing*. O conversor de tempo/frequência introdutor do *aliasing* inclui uma formação de janela para aplicar

uma janela de análise e um algoritmo de transformação de MDCT. Especificamente, a formação de janela é operante para aplicar a função de janela a estruturas subsequentes de maneira sobreposta, de modo que uma amostra de um sinal com janela ocorra em pelo
5 menos duas estruturas com janela subsequentes.

Em uma configuração, o primeiro ramal de processamento inclui um codificador ACELP e um segundo ramal de processamento inclui um conversor de espectro de MDCT e o quantizador para quantizar componentes de espectro para obter
10 componentes de espectro quantizados, onde cada componente de espectro quantizado é zero, ou é definido por um índice de quantizador da pluralidade de diferentes índices de quantizador possíveis.

Além disso, é preferido que o primeiro comutador
15 200 opere em modo de *loop* aberto e o segundo comutador opere em modo de *loop* fechado.

Conforme já foi informado, ambos os ramaís de codificação são operantes para codificar o sinal de áudio em blocos, onde o primeiro comutador ou o segundo comutador
20 transferem em blocos, de maneira que ocorra uma ação de transferência, no mínimo, após um bloco de um número pré-definido de amostras de um sinal, sendo que o número pré-definido forma uma extensão de estrutura para o comutador correspondente. Assim, o grânulo para alternar pelo primeiro comutador pode ser, por
25 exemplo, um bloco de 2048 ou 1028 amostras, e a extensão da estrutura, com base no qual o primeiro comutador 200 está alternando, pode ser variável, mas é, preferivelmente, fixa a esse período bem longo.

Contrariamente a isso, a extensão de bloco para o segundo comutador 521, isto é, quando o segundo comutador 521 alterna de um modo para outro, é substancialmente menor que a extensão de bloco para o primeiro comutador. Preferivelmente, 5 ambas as extensões de bloco para os comutadores são selecionadas de maneira que a extensão de bloco mais longa seja um múltiplo integral da extensão de bloco mais curta. Na configuração preferida, a extensão do bloco do primeiro comutador é 2048 ou 1024, e a extensão do bloco do segundo comutador é 1024 ou mais 10 preferivelmente, 512, e ainda mais preferivelmente, 256, e ainda mais preferivelmente, 128 amostras, de maneira que, no máximo, o segundo comutador possa alternar 16 vezes quando o primeiro comutador alternar somente uma única vez. Uma proporção de extensão de bloco máxima preferida, é, no entanto, 4:1.

15 Em uma outra configuração, o controlador 300, 525 é operante para executar uma discriminação de fala/música para o primeiro comutador de maneira que uma decisão para fala seja favorecida em relação a uma decisão para música. Nessa configuração, uma decisão para fala é tomada mesmo quando uma 20 parte de menos de 50% de uma estrutura para o primeiro comutador seja fala e a parte de mais de 50% da estrutura seja música.

Além disso, o controlador é operante para já alternar para o modo de fala, quando uma parte bem pequena do primeiro estrutura for fala e, especificamente, quando uma parte 25 do primeiro estrutura for fala, o que é 50% da extensão do segundo estrutura menor. Assim, uma decisão de alternância de fala/favorecimento preferida já alterna para fala, mesmo quando, por exemplo, somente 6% ou 12% de um bloco correspondente à

extensão do estrutura do primeiro comutador for fala.

Esse procedimento serve preferivelmente para explorar totalmente a capacidade de salvar taxa de bits do primeiro ramal de processamento, o qual possui um núcleo de fala com voz em uma configuração, e para não perder nenhuma qualidade, mesmo para o resto da primeira estrutura grande, que é sem fala, devido ao fato do segundo ramal de processamento incluir um conversor e, assim, ser útil para sinais de áudio que possuem sinais sem fala também. Preferivelmente, esse segundo ramal de processamento inclui uma MDCT sobreposta, que é criticamente amostrada, e que mesmo em tamanhos de janela pequenos, provê uma operação altamente eficiente e operação sem *aliasing*, devido ao processamento de cancelamento de *aliasing* do domínio de tempo, como por exemplo, sobreposição e adição no lado do decodificador.

Além disso, uma extensão de bloco-grande para o primeiro ramal de codificação, que é preferivelmente um ramal de codificação de MDCT do tipo AAC, é útil, pois sinais sem fala são normalmente bem fixos e uma janela de transformação longa provê uma resolução de alta frequência e, portanto, alta qualidade e, além disso, provê uma eficiência de taxa de frequência devida a um módulo de quantização controlado psicoacusticamente, que pode também ser aplicado ao modo de codificação à base de transformação no segundo ramal de processamento do segundo ramal de codificação.

Em relação à ilustração do decodificador da Fig. 3d, é preferido que o sinal transmitido inclua um indicador explícito como informação secundária 4a, conforme ilustra a Fig. 3e. Essa informação secundária 4a é extraída por um divisor de fluxo de bits não ilustrado na Fig. 3d, para encaminhar o primeiro

5 sinal codificado correspondente, primeiro sinal processado ou segundo sinal processado ao processador correto, como por exemplo o primeiro ramal de decodificação, o primeiro ramal de processamento inverso ou o segundo ramal de processamento inverso da Fig. 3d. Portanto, um sinal codificado não só possui os sinais codificados/processados, mas também inclui informações secundárias relacionadas a esses sinais. Em outras configurações, no entanto, pode haver uma sinalização implícita que permita um divisor de fluxo de bits do lado do decodificador, para distinguir entre os
10 sinais determinados. Em relação à Fig. 3e, é estabelecido que o primeiro sinal processado ou o segundo sinal processado é a saída do segundo ramal de codificação e, portanto, o segundo sinal codificado.

15 Preferivelmente, o primeiro ramal de decodificação e/ou o segundo ramal de processamento inverso inclui uma transformação de MDCT para converter do domínio de espectro para o domínio de tempo. Para essa finalidade, é provido um adicionador de sobreposição, para executar uma funcionalidade de cancelamento de *aliasing* de domínio de tempo que, ao mesmo tempo,
20 provê um efeito de fusão de transmissão para evitar artefatos de bloqueio. Em geral, o primeiro ramal de decodificação converte um sinal codificado no quarto domínio no primeiro domínio, enquanto o segundo ramal de processamento inverso realiza uma conversão do terceiro domínio para o segundo domínio, e o conversor
25 subsequentemente conectado ao primeiro combinador provê uma conversão do segundo domínio para o primeiro domínio, de maneira que na entrada do combinador 600, existam somente sinais de primeiro domínio, os quais representam, na configuração da Fig.

3d, o sinal de saída decodificado.

As Figs. 4a e 4b ilustram duas configurações diferentes, que diferem no posicionamento do comutador 200. Na Fig. 4a, o comutador 200 está posicionado entre uma saída do estágio de pré-processamento comum 100 e a entrada dos dois ramaís codificados 400, 500. A configuração da Fig. 4a garante que o sinal de áudio seja inserido em somente um único ramal de codificação, e o outro ramal de codificação, que não está conectado à saída do estágio de pré-processamento comum não funciona e, portanto, é desligado ou fica em modo de suspensão. Essa configuração é preferível porque o ramal de codificação não ativo não consome energia e recursos do computador, o que é útil para aplicações móveis em particular, que sejam movidas a bateria e, portanto, tenham a limitação geral de consumo de energia.

15 ~~Por outro lado, no entanto, a configuração da~~ Fig. 4b pode ser preferível quando o consumo de energia não for problema. Nessa configuração, ambos os ramaís de codificação 400, 500, estão ativos o tempo todo, e somente a saída do ramal de codificação selecionado para uma determinada parte de tempo e/ou uma determinada parte de frequência é encaminhada ao formatador de fluxo de bits, que pode ser implementado como um multiplexor de fluxo de bits 800. Portanto, na configuração da Fig. 4b, ambos os ramaís de codificação ficam ativos o tempo todo, e a saída de um ramal de codificação que é selecionado pelo estágio de decisão 300 25 é inserida no fluxo de bits de saída, enquanto a saída do outro ramal de codificação não selecionado 400 é descartada, isto é, não é inserida no fluxo de bits de saída, isto é, o sinal de áudio codificado.

A Fig. 4c ilustra um outro aspecto de uma implementação de decodificador preferida. Para evitar artefatos audíveis especificamente na situação, na qual o primeiro decodificador é um decodificador que gera *aliasing* de tempo, ou em
 5 resumo, um decodificador de domínio de frequência, e o segundo decodificador é um dispositivo de domínio de tempo, os limites entre os blocos ou estruturas produzidos pelo primeiro decodificador 450 e pelo segundo decodificador 550 não devem ser totalmente contínuos, especificamente em uma situação de
 10 alternância. Assim, quando o primeiro bloco do primeiro decodificador 450 é produzido e, quando para a parte de tempo subsequente, um bloco do segundo decodificador é produzido, é preferido executar uma operação de fusão de transmissão conforme ilustra o bloco de fusão de transmissão 607. Com essa finalidade,
 15 o bloco de fusão de transmissão 607 pode ser implementado conforme ilustra a Fig. 4c em 607a, 607b e 607c. Cada ramal pode ter um ponderador com um fator de ponderação m_1 entre 0 e 1 na escala normalizada, onde o fator de ponderação pode variar conforme indica o plot 609, essa regra de fusão de transmissão garante que
 20 ocorra um fusão de transmissão contínuo e suave, o qual, além disso, garante que o usuário não perceba nenhuma variação de sonoridade. Regras de fusão de transmissão não linear como uma regra de fusão de transmissão de seno² podem ser aplicadas no lugar de uma regra de fusão de transmissão linear.

25 Em determinados casos, o último bloco do primeiro decodificador foi gerado utilizando-se uma janela onde a janela realmente executou um desvanecimento desse bloco. Nesse caso, o fator de ponderação m_1 do bloco 607a é igual a 1 e, na verdade,

não é necessária nenhuma ponderação para esse ramal.

Quando ocorre uma alternância do segundo decodificador para o primeiro decodificador, e quando o segundo decodificador inclui uma janela que realmente desvanece a saída
5 para o final do bloco, então o ponderador indicado por " m_2 " não seria necessário, ou o parâmetro de ponderação pode ser estabelecido para 1 ao longo de toda a região de fusão de transmissão.

Quando o primeiro bloco após um comutador tiver
10 sido gerado com uso de uma operação de formação de janela, e quando essa janela realmente tiver executado um desvanecimento em operação, então o fator de ponderação correspondente pode também ser estabelecido como 1, de maneira que um ponderador não seja realmente necessário. Portanto, quando o último bloco formar uma
15 janela para desvanecimento pelo decodificador e quando o primeiro bloco após o comutador formar uma janela com uso do decodificador para prover uma intensificação, então os ponderadores 607a, 607b não serão necessários, e uma operação de adição pelo adicionador 607c é suficiente.

20 Neste caso, a parte de desvanecimento da última estrutura e a parte de intensificação da próxima estrutura definem a região de fusão de transmissão indicada no bloco 609. Além disso, é preferido nessa situação que o último bloco de um decodificador tenha uma determinada sobreposição de tempo com o
25 primeiro bloco do outro decodificador.

Se uma operação de fusão de transmissão não for necessária ou não for possível ou não for desejada, e se existir somente uma alternância fixa de um decodificador para o outro

decodificador, é preferido executar essa alternância em passagens silenciosas do sinal de áudio, ou pelo menos em passagens do sinal de áudio onde existe energia baixa, isto é, que são percebidas como silenciosas ou quase silenciosas. Preferivelmente, o estágio de decisão 300 garante nessa configuração que o comutador 200 seja
5 ativado somente quando a parte de tempo correspondente que seguir o evento de alternância tiver uma energia que seja, por exemplo, mais baixa que a energia média do sinal de áudio, e seja, preferivelmente, mais baixa que 50% da energia média do sinal de áudio relacionado, por exemplo, a duas ou até mais
10 partes/estruturas de tempo do sinal de áudio.

Preferivelmente, a segunda regra de codificação/regra de decodificação é um algoritmo de codificação à base de LPC. Na codificação de fala à base de LPC, é feita uma
15 diferenciação entre segmentos de sinal ou partes de sinal de excitação semelhantes ao periódico do tipo de impulso e segmentos de sinal ou partes de sinal de excitação do tipo de ruído. Isso é realizado para vocoderes LPC (2,4 kbps) de taxa de bits muito baixa, como na Fig. 7b. No entanto, em codificadores CELP de taxa
20 média, a excitação é obtida para adição de vetores escalonados de um livro de códigos adaptativo e um livro de códigos fixo.

Segmentos de sinal de excitação semelhantes ao periódico e do tipo de impulso, isto é, segmentos de sinal com uma altura específica, são codificados com mecanismos diferentes dos
25 sinais de excitação do tipo de ruído. Enquanto os sinais de excitação semelhantes ao periódico e do tipo de impulso são conectados à fala com voz, os sinais do tipo de ruído estão relacionados à fala sem voz.

Por exemplo, com referência às Fig. 5a a 5d. Aqui, segmentos de sinal ou partes de sinal semelhantes ao periódico e do tipo de impulso e segmentos de sinal ou partes de sinal do tipo de ruído são discutidos como exemplo.

5 Especificamente, uma fala com voz conforme ilustra a Fig. 5a no domínio de tempo e na Fig. 5b no domínio de frequência, é discutida como exemplo para uma parte de sinal semelhante ao periódico e do tipo de impulso, e um segmento de fala sem voz como exemplo de uma parte de sinal do tipo de ruído é discutida em
10 relação às Fig. 5c e 5d. A fala pode ser em geral classificada como com voz, sem voz ou mista. Plots de domínio de tempo-e-frequência para segmentos com voz e sem voz amostrados são mostrados nas Fig. 5a a 5d. A fala com voz é semelhante ao periódico no domínio de tempo e harmonicamente estruturada no
15 domínio de frequência, enquanto a fala sem voz é do tipo aleatória e de banda larga. O espectro de curto prazo da fala com voz é caracterizado por sua estrutura de formante harmônica fina. A estrutura harmônica fina é uma consequência da semelhança à periodicidade da fala, e pode ser atribuída às cordas vocais
20 vibrantes. A estrutura do formante (envelope de espectro) é devida à interação da fonte e dos tratos vocais. Os tratos vocais são compostos da faringe e da cavidade bucal. O formato do envelope de espectro que "se encaixa" no espectro de curto prazo de fala com voz está associado às características de transferência do trato
25 vocal e da inclinação de espectro (6 dB/oitava), devido ao pulso glotal. O envelope de espectro é caracterizado por um conjunto de picos, que são chamados formantes. Os formantes são os modos ressonantes do trato vocal. Para o trato vocal médio existem de

três a cinco formantes abaixo de 5 kHz. As amplitudes e locais dos primeiros três formantes, que em geral ocorrem abaixo de 3 kHz são muito importantes, tanto em síntese como em percepção de fala. Formantes mais altos também são importantes para representações de banda larga e fala sem voz. As propriedades de fala estão relacionadas ao sistema físico de produção de fala, como segue. Excitar o trato vocal com pulsos de ar glotais semelhantes ao periódico gerados pelas cordas vocais em vibração produz fala com voz. A frequência dos pulsos periódicos é denominada frequência fundamental ou altura. Forçar o ar através de uma constrição no trato vocal produz fala sem voz. Sons nasais são devidos ao acoplamento acústico do trato nasal com o trato vocal, e sons plosivos são produzidos liberando-se abruptamente a pressão de ar, que foi acumulada atrás do fechamento do trato.

Assim, uma parte do tipo de ruído do sinal de áudio não demonstra nenhuma estrutura de domínio de tempo do tipo de impulso, nem uma estrutura de domínio de frequência harmônica, conforme ilustrado na Fig. 5c e na Fig. 5d, que é diferente da parte do tipo de impulso semelhante ao periódico, conforme ilustrado, por exemplo, na Fig. 5a e na Fig. 5b. Como será explicado posteriormente, no entanto, a diferenciação entre partes do tipo de ruído e partes do tipo de impulso semelhante ao periódico também podem ser observadas depois de uma LPC para o sinal de excitação. A LPC é um método que modela o trato vocal e extrai do sinal a excitação dos tratos vocais.

Além disso, partes do tipo de impulso semelhantes ao periódico e partes do tipo de ruído podem ocorrer de maneira oportuna, isto é, o que significa que uma parte do sinal de áudio

no tempo é ruidosa e a outra parte do sinal de áudio no tempo é semelhante ao periódico, isto é, tonal. Alternativamente, ou além disso, a característica de um sinal pode ser diferente em diferentes bandas de frequência. Assim, a determinação se o sinal de áudio é ruidoso ou tonal também pode ser feita com seleção de frequência, de maneira que uma determinada banda de frequência ou várias determinadas bandas de frequência sejam consideradas ruidosas e outras bandas de frequência sejam consideradas tonais. Neste caso, uma determinada parte de tempo do sinal de áudio pode incluir componentes tonais e componentes ruidosos.

A Fig. 7a ilustra um modelo linear de um sistema de produção de fala. Esse sistema presume uma excitação de dois estágios, isto é, uma sequência de impulsos para fala com voz conforme indica a Fig. 7c, e um ruído aleatório para fala sem voz conforme indica a Fig. 7d. O trato-vocal é representado como um filtro de todos os pólos 70 que processa os pulsos da Fig. 7c ou Fig. 7d, gerados pelo modelo glotal 72. Por conseguinte, o sistema da Fig. 7a pode ser reduzido a um modelo de filtro de todos os pólos da Fig. 7b com um estágio de ganho 77, uma via de avanço 78, uma via de retorno 79, e um estágio de adição 80. Na via de retorno 79, existe um filtro de previsão 81, e o sistema completo de síntese do modelo de fonte ilustrado na Fig. 7b pode ser representado usando-se funções de domínio z como segue:

$$S(z) = g / (1 - A(z)) \cdot X(z),$$

onde g representa o ganho, A(z) é o filtro de previsão, conforme determinado por uma análise de LP, X(z) é o sinal de excitação, e S(z) é a saída de fala de síntese.

As Figs. 7c e 7d dão uma descrição do domínio de

tempo gráfico da síntese de fala com voz e sem voz, usando o modelo de sistema de fonte linear. Esse sistema e os parâmetros de excitação da equação acima são desconhecidos e devem ser determinados a partir de um conjunto finito de amostras de fala.

5 Os coeficientes de $A(z)$ são obtidos usando-se uma previsão linear do sinal de entrada e uma quantização dos coeficientes de filtro. Em um previsor linear avançado de ordem p , a amostra presente da sequência de fala é prevista a partir de uma combinação linear de amostras passadas por p . Os coeficientes previsores podem ser

10 determinados por algoritmos bem conhecidos, como o algoritmo de Levinson-Durbin, ou em geral um método de autocorrelação ou um método de reflexão.

A Fig. 7e ilustra uma implementação mais detalhada do bloco de análise de LPC 510. O sinal de áudio é

15 inserido em um bloco de determinação de filtro que determina as informações $A(z)$ do filtro. Essas informações são produzidas como as informações de previsão de curto prazo necessárias para um decodificador. As informações de previsão de curto prazo são exigidas pelo filtro de previsão real 85. Em um subtraendo 86, uma

20 amostra atual do sinal de áudio é inserida, e um valor previsto para a amostra atual é subtraído, de maneira que para essa amostra, o sinal de erro de previsão é gerado na linha 84. Uma sequência dessas amostras de sinal de erro de previsão é muito esquematicamente ilustrada na Fig. 7c ou 7d. Portanto, a Fig. 7a,

25 7b pode ser considerada um tipo de sinal do tipo de impulso retificado.

Enquanto a Fig. 7e ilustra uma maneira preferida para calcular o sinal de excitação, a Fig. 7f ilustra uma maneira

preferida para calcular o sinal ponderado. Em comparação à Fig. 7e, o filtro 85 é diferente, quando γ for diferente de 1. Um valor menor que 1 é preferido para γ . Além disso, o bloco 87 está presente, e para μ é preferível um número menor que 1. Em geral, os elementos das Figs. 7e e 7f podem ser implementados como em 3GPP TS 26.190 ou 3GPP TS 26.290.

A Fig. 7g ilustra um processamento inverso, que pode ser aplicado no lado do decodificador, como por exemplo, no elemento 537 da Fig. 2b. Particularmente, o bloco 88 gera um sinal não ponderado a partir do sinal ponderado, e o bloco 89 calcula uma excitação a partir do sinal não ponderado. Em geral, todos os sinais, com exceção do sinal não ponderado da Fig. 7g, estão no domínio de LPC, mas o sinal de excitação e o sinal ponderado são sinais diferentes no mesmo domínio. O bloco 89 produz um sinal de excitação que pode então ser utilizado com a saída do bloco 536. Então, a transformação de LPC inversa comum pode ser executada no bloco 540 da Fig. 2b.

Subsequentemente, um codificador CELP de análise-por-síntese será discutido em relação à Fig. 6, para ilustrar as modificações aplicadas a esse algoritmo. Esse codificador de CELP é discutido em detalhe em "Speech Coding: A tutorial review", Andreas Spanias, Proceedings of IEEE, Vol. 82, No. 10, October 1994, pp. 1541-1582. O codificador de CELP ilustrado na Fig. 6 inclui um componente de previsão de longo prazo 60 e um componente de previsão de curto prazo 62. Além disso, é utilizado um livro de códigos que está indicado em 64. Um filtro de ponderação perceptual $W(z)$ é implementado em 66, e um controlador de minimização de erro é provido em 68. $s(n)$ é o sinal de entrada de

domínio de tempo. Depois de ter sido perceptualmente ponderado, o sinal ponderado é inserido em um subtraendo 69, que calcula o erro entre o sinal de síntese ponderado na saída do bloco 66 e o sinal ponderado real $s_w(n)$. Em geral, os coeficientes de filtro de

5 previsão de curto prazo $A(z)$ é calculada por um estágio de análise de LP, e seus coeficientes são quantizados em $A(z)$ conforme indicado na Fig. 7e. As informações de previsão de longo prazo $A_L(z)$ que incluem o ganho g de previsão de longo prazo e o índice de quantização de vetor, isto é, as referências do livro de

10 códigos são calculadas no sinal de erro de previsão na saída do estágio de análise de LPC mencionada como 10a na Fig. 7e. Os parâmetros de LTP são o retardo e o ganho de altura. Na CELP isso é em geral implementado como um livro de códigos adaptativo que contém o sinal de excitação passado (não o residual). O retardo e

15 ganho de CB adaptativo são encontrados minimizando-se o erro ponderado ao quadrado médio (busca de altura de *loop* fechado).

O algoritmo de CELP codifica, então, o sinal residual obtido após as previsões de curto prazo e longo prazo, utilizando um livro de códigos de, por exemplo, sequências de

20 Gaussian. O algoritmo de ACELP, onde o "A" significa "Algébrico" possui um livro de códigos específico elaborado algebricamente.

Um livro de códigos pode conter mais ou menos vetores, onde cada vetor tem a extensão de algumas amostras. Um fator de ganho g escala o vetor de código e o código ganho é

25 filtrado pelo filtro de síntese de previsão de longo prazo e pelo filtro de síntese de previsão de curto prazo. O vetor de código "ideal" é selecionado de maneira que o erro de quadrado médio perceptualmente ponderado na saída do subtraendo 69 seja

minimizado. O processo de busca na CELP é feito por uma otimização de análise-por-síntese conforme ilustra a Fig. 6.

Para casos específicos, quando uma estrutura for uma mistura de fala sem voz e com voz, ou quando ocorre fala sobre
 5 música, uma codificação TCX pode ser mais adequada para codificar a excitação no domínio de LPC. A codificação TCX processa um sinal ponderado no domínio de frequência sem fazer nenhuma pressuposição de produção de excitação. A TCX é então mais genérica que a codificação CELP, e não está restrita a um modelo de fonte com voz
 10 ou sem voz da excitação. A TCX é ainda uma codificação de modelo registradora de amostra que utiliza um filtro preditor linear para representar os formantes dos sinais do tipo de fala.

Na codificação do tipo do AMR-WB+, ocorre uma seleção entre diferentes modos de TCX e ACELP, conforme é
 15 conhecido pela descrição do AMR-WB+. Os modos de TCX são diferentes porque a extensão da Transformação de Fourier discreta em blocos é diferente para diferentes modos, e o melhor modo pode ser selecionado por uma abordagem de análise por síntese ou por um modo direto de "avanço".

Conforme foi discutido em relação às Figs. 2a e 2b, o estágio de pré-processamento comum 100 inclui preferivelmente um multicanal conjunto (dispositivo surround/estéreo associado) 101 e, além desse, um estágio de extensão de largura de banda 102. De maneira correspondente, o
 25 decodificador inclui um estágio de extensão de largura de banda 701 e um estágio multicanal conjunto conectado subsequentemente 702. Preferivelmente, o estágio multicanal conjunto 101 é, em relação ao codificador, conectado antes do estágio de extensão de

largura de banda 102 e, no lado do decodificador, o estágio de extensão de largura de banda 701 é conectado antes do estágio multicanal conjunto 702 em relação à direção de processamento de sinal. Alternativamente, no entanto, o estágio de pré-
 5 processamento comum pode incluir um estágio multicanal conjunto sem o estágio de extensão de largura de banda conectado subsequentemente ou um estágio de extensão de largura de banda sem um estágio multicanal conjunto conectado.

Um exemplo preferido para um estágio multicanal conjunto no lado do codificador 101a, 101b e no lado do decodificador 702a e 702b está ilustrado no contexto da Fig. 8. Um número de canais de entrada originais E é inserido no *downmixer* 101a de maneira que o *downmixer* gere um número de canais transmitidos K , onde o número K é maior que ou igual a um e é
 10 menor que ou igual a E .

Preferivelmente, os canais de entrada E são inseridos em um analisador de parâmetro multicanal associado 101b que gera informações paramétricas. Essas informações paramétricas são preferivelmente codificadas por entropia, como por exemplo,
 20 por uma codificação de diferença e subsequente codificação de Huffman, ou alternativamente, codificação aritmética subsequente. A saída de informações paramétricas codificada pelo bloco 101b é transmitida a um decodificador de parâmetro 702b que pode fazer parte do item 702 da Fig. 2b. O decodificador de parâmetro 702b
 25 decodifica as informações paramétricas transmitidas e encaminha as informações paramétricas decodificadas ao *upmixer* 702a. O *upmixer* 702a recebe os canais transmitidos K e gera um número de canais de saída L , onde o número de L é maior que ou igual a K e mais baixo

que ou igual a E.

As informações paramétricas podem incluir diferenças de nível entre os canais, diferenças de tempo entre os canais, diferenças de fase entre os canais e/ou medidas de
 5 coerência entre os canais, como é conhecido pela técnica de BBC ou como é conhecido e é descrito em detalhe no padrão *surround* de MPEG. O número de canais transmitidos pode ser um único canal mono para aplicações de taxa de bits ultra-baixas, ou pode incluir uma aplicação estéreo compatível, ou pode incluir um sinal estéreo
 10 compatível, isto é, dois canais. Tipicamente, o número de canais de entrada E pode ser cinco ou talvez ainda mais alto. Alternativamente, o número de canais de entrada E pode também ser objetos de áudio-E, como é conhecido no contexto de codificação de objeto de áudio espacial (SAOC).

15 Em uma implementação, o *downmixer* realiza uma adição ponderada ou não ponderada dos canais de entrada originais E, ou uma adição dos objetos de áudio de entrada E. No caso de objetos de áudio como canais de entrada, o analisador de parâmetro multicanal conjunto 101b calcula parâmetros de objeto de áudio,
 20 como por exemplo, uma matriz de correlação entre os objetos de áudio, preferivelmente para cada parte de tempo, e mais preferivelmente ainda para cada banda de frequência. Com essa finalidade, toda a faixa de frequência pode ser dividida em pelo menos 10 e preferivelmente 32 ou 64 bandas de frequência.

25 A Fig. 9 ilustra uma configuração preferida para a implementação do estágio de extensão de largura de banda 102 na Fig. 2a, e o estágio de extensão de largura de banda correspondente 701 na Fig. 2b. No lado do codificador, o bloco de

extensão de largura de banda 102 preferivelmente inclui um bloco de filtragem de baixa frequência 102b, um bloco de sub-amostrador, que segue o filtro de baixa frequência, ou que faz parte do QMF inverso, que funciona somente na metade das bandas de QMF, e um

5 analisador de banda alta 102a. A entrada de sinal de áudio original no bloco de extensão de largura de banda 102 é filtrada com baixa frequência, gerando o sinal de banda baixa que é então inserido nos ramos de codificação e/ou no comutador. O filtro de

10 baixa frequência apresenta uma frequência de corte que pode estar em uma faixa de 3 kHz a 10 kHz. Além disso, o bloco de extensão de largura de banda 102 inclui ainda um analisador de banda alta para calcular os parâmetros de extensão de largura de banda, como por exemplo uma informação de parâmetro de envelope de espectro, uma

15 informação de parâmetro de piso de ruído, uma informação de parâmetro de filtragem inversa, outras informações paramétricas relacionadas a determinadas linhas harmônicas na banda alta e outros parâmetros discutidos em detalhe no padrão de MPEG-4, no capítulo relacionado à replicação de banda de espectro.

No lado do decodificador, o bloco de extensão de

20 largura de banda 701 inclui um atualizador 701a, um regulador 701b e um combinador 701c. O combinador 701c combina o sinal de banda baixa decodificado e o sinal de banda alta reconstruído e regulado produzido pelo regulador 701b. A entrada no regulador 701b é provida por um atualizador que é operado para derivar o sinal de

25 banda alta do sinal de banda baixa, como por exemplo por replicação de banda de espectro, ou em geral, por extensão de largura de banda. A atualização executada pelo atualizador 701a pode ser uma atualização executada de maneira harmônica ou de

maneira não harmônica. O sinal gerado pelo atualizador 701a é, subsequentemente, regulado pelo regulador 701b com uso da informações paramétricas transmitidas de extensão de largura de banda.

5 Conforme indicado na Fig. 8 e Fig. 9, os blocos descritos podem ter uma entrada de controle de modo em uma configuração preferida. Essa entrada de controle de modo é derivada do sinal de saída do estágio de decisão 300. Nessa configuração preferida, uma característica de um bloco
10 correspondente pode ser adaptada à saída do estágio de decisão, isto é, se, em uma configuração preferida, uma decisão de fala ou uma decisão de música é feita para uma determinada parte de tempo do sinal de áudio ou não. Preferivelmente, o controle de modo relaciona-se somente a uma ou mais das funcionalidades desses
15 blocos, mas não a todas as funcionalidades de blocos. Por exemplo, a decisão pode influenciar somente o atualizador 701a, mas pode não influenciar os outros blocos da Fig. 9, ou pode, por exemplo, influenciar somente o analisador de parâmetro multicanal conjunto 101b da Fig. 8, mas não os outros blocos da Fig. 8. Essa
20 implementação é preferivelmente de modo que uma flexibilidade mais alta e uma qualidade mais alta e sinal de saída de fluxo de bits mais baixo sejam obtidos provendo-se flexibilidade no estágio de pré-processamento comum. Por outro lado, no entanto, o uso de algoritmos no estágio de pré-processamento comum para ambos os
25 tipos de sinais permite implementar um esquema de codificação/decodificação eficiente.

A Fig. 10a e a Fig. 10b ilustram duas implementações diferentes do estágio de decisão 300. Na Fig. 10a,

é indicada uma decisão de loop aberto. Aqui, o analisador de sinal 300a do estágio de decisão tem determinadas regras para decidir se uma determinada parte de tempo ou uma determinada parte de frequência do sinal de entrada tem uma característica que requer
 5 que essa parte de sinal seja codificada pelo primeiro ramal de codificação 400 ou pelo segundo ramal de codificação 500. Com essa finalidade, o analisador de sinal 300a pode analisar o sinal de entrada de áudio no estágio de pré-processamento comum ou pode analisar a saída de sinal de áudio pelo estágio de pré-
 10 processamento comum, isto é, o sinal intermediário de áudio, ou pode analisar um sinal intermediário dentro do estágio de pré-processamento comum, como por exemplo a saída do sinal de *downmix* que pode ser um sinal mono ou que pode ser um sinal com canais *k* indicados na Fig. 8. No lado da saída, o analisador de sinal 300a
 15 gera a decisão de alternância para controlar o comutador 200 no lado do codificador e o comutador correspondente 600 ou o combinador 600 no lado do decodificador.

Apesar de não discutido em detalhe para o segundo comutador 521, deve-se enfatizar que o segundo comutador 521 pode
 20 ser posicionada de maneira semelhante à do primeiro comutador 200, conforme discutido em relação às Figs. 4a e 4b. Assim, uma posição alternativa do comutador 521 na Fig. 3c é na saída de ambos os ramos de processamento 522, 523, 524, de maneira que ambos os ramos de processamento funcionem em paralelo, e somente a saída
 25 de um ramal de processamento seja gravada em um fluxo de bits através de um formador de fluxo de bits que não está ilustrado na Fig. 3c.

Além disso, o segundo combinador 600 pode ter uma

funcionalidade específica de fusão de transmissão discutida na Fig. 4c. Alternativamente ou adicionalmente, o primeiro combinador 532 pode ter a mesma funcionalidade de fusão de transmissão. Além disso, ambos os combinadores podem ter a mesma funcionalidade de fusão de transmissão, ou podem ter diferentes funcionalidades de fusão de transmissão, ou podem não ter nenhuma funcionalidade de fusão de transmissão, de maneira que ambos os combinadores sejam comutadores sem nenhuma funcionalidade de fusão de transmissão adicional.

10 Conforme foi discutido anteriormente, ambas as comutadores podem ser controladas através de uma decisão de *loop* aberto ou de uma decisão de *loop* fechado, conforme discutido em relação à Fig. 10a e à Fig. 10b, onde o controlador 300, 525 da Fig. 3c pode ter funcionalidades iguais ou diferentes para ambas
15 as comutadores.

 Além disso, uma funcionalidade de distorção de tempo que é adaptativa ao sinal pode existir não só no primeiro ramal de codificação ou no primeiro ramal de decodificação, mas pode também existir no segundo ramal de processamento do segundo
20 ramal de codificação do lado do codificador, bem como no lado do decodificador. Dependendo do sinal processado, ambas as funcionalidades de distorção de tempo podem ter a mesma informação de distorção de tempo, de maneira que a mesma distorção de tempo seja aplicada aos sinais no primeiro domínio e no segundo domínio.
25 Isso economiza carga de processamento e pode ser útil em alguns casos, em casos em que blocos subsequentes têm uma característica de tempo de distorção de tempo semelhante. Em configurações alternativas, no entanto, é preferido ter avaliadores de distorção

de tempo independentes para o primeiro ramal de codificação e para o segundo ramal de processamento no segundo ramal de codificação.

O sinal de áudio codificado inventivo pode ser armazenado em um meio de armazenamento digital, ou pode ser transmitido em um meio de transmissão, como por exemplo um meio de transmissão sem fio ou um meio de transmissão com fio, como a Internet.

Em uma configuração diferente, o comutador 200 da Fig. 1a ou 2a alterna entre os dois ramaís de codificação 400, 500. Em uma outra configuração, pode haver outros ramaís de codificação, como por exemplo um terceiro ramal de codificação, ou mesmo um quarto ramal de codificação, ou até mais ramaís de codificação. No lado do decodificador, o comutador 600 da Fig. 1b ou 2b alterna entre os dois ramaís de decodificação 431, 440 e 531, 532, 533, 534, 540. Em uma outra configuração, pode haver outros ramaís de decodificação, como por exemplo um terceiro ramal de decodificação, ou até mesmo um quarto ramal de decodificação, ou até mesmo mais ramaís de decodificação. De maneira semelhante, os outros comutadores 521 ou 532 podem alternar entre mais de dois algoritmos de codificação diferentes, quando esses ramaís de codificação/decodificação adicionais forem providos.

As configurações descritas acima são meramente ilustrativas para os princípios da presente invenção. Compreende-se que modificações e variações das disposições e os detalhes descritos no presente documento serão evidentes para outros técnicos no assunto. Portanto, a intenção é limitar-se somente pelo escopo das reivindicações da patente apresentadas a seguir, e não pelos detalhes específicos apresentados através de descrição e

explicação das configurações do presente documento.

Dependendo de determinados requisitos de implementação dos métodos inventivos, os métodos inventivos podem ser implementados em hardware ou em software. A implementação pode ser realizada utilizando-se um meio de armazenamento digital, em particular, um disco, um DVD ou um CD com sinais de controle eletronicamente legíveis armazenados neles, que cooperem com sistemas de computador programáveis de maneira que o respectivo método seja executado. Em geral, a presente invenção é portanto, um produto de programa de computador com um código de programa armazenado em um portador legível em máquina, sendo o código de programa operado para executar os métodos inventivos quando o produto de programa de computador for executado em um computador. Em outras palavras, os métodos inventivos são, portanto, um programa de computador com um código de programa para executar pelo menos um dos métodos inventivos quando o programa de computador for executado em um computador.

REIVINDICAÇÕES

1. "ESQUEMA DE CODIFICAÇÃO/DECODIFICAÇÃO DE ÁUDIO DE BAIXA TAXA DE BITS QUE APRESENTA COMUTADORES EM CASCATA", onde o Codificador de áudio para codificar um sinal de entrada de áudio (195), estando o sinal de entrada de áudio em um primeiro domínio, é **caracterizado** por compreender:

um primeiro ramo de codificação (400) para codificar um sinal de áudio utilizando um primeiro algoritmo de codificação para obter um primeiro sinal codificado;

um segundo ramo de codificação (500) para codificar um sinal de áudio utilizando um segundo algoritmo de codificação para obter um segundo sinal codificado, onde o primeiro algoritmo de codificação é diferente do segundo algoritmo de codificação; e

um primeiro comutador (200) para alternar entre o primeiro ramo de codificação e o segundo ramo de codificação, de maneira que, para uma parte do sinal de entrada de áudio, o primeiro sinal codificado ou o segundo sinal codificado estejam em um sinal de saída de codificador,

onde o segundo ramo de codificação compreende:

um conversor (510) para converter o sinal de áudio em um segundo domínio diferente do primeiro domínio,

um primeiro ramo de processamento (522) para processar um sinal de áudio no segundo domínio para obter um primeiro sinal processado;

um segundo ramo de processamento (523, 524), para converter um sinal em um terceiro domínio diferente do primeiro domínio e do segundo domínio, e para processar o sinal no terceiro

domínio para obter um segundo sinal processado; e

um segundo comutador (521) para alternar entre o primeiro ramo de processamento (522) e o segundo ramo de processamento (523, 524) de maneira que, para uma parte da entrada de sinal de áudio no segundo ramo de codificação, ou o primeiro sinal processado ou o segundo sinal processado esteja no segundo sinal codificado.

2. Codificador de áudio, de acordo com a reivindicação 1, **caracterizado** pelo fato de que o primeiro algoritmo de codificação do primeiro ramo de codificação (400) se baseia em um modelo de depósito de informação, ou onde o segundo algoritmo de codificação do segundo ramo de codificação (500) se baseia em uma fonte de informações ou em um modelo de relação sinal-ruído (SNR).

3. Codificador de áudio, de acordo com a reivindicação 1 ou 2, **caracterizado** pelo fato de que o primeiro ramo de codificação compreende um conversor (410) para converter o sinal de entrada de áudio em um quarto domínio diferente do primeiro domínio, do segundo domínio e do terceiro domínio.

4. Codificador de áudio, de acordo com uma das reivindicações anteriores, **caracterizado** pelo fato de que o primeiro domínio é o domínio de tempo, o segundo domínio é um domínio de LPC obtido por uma filtração do sinal do primeiro domínio pelo LPC, o terceiro domínio é um domínio de espectro de LPC obtido ao converter um sinal filtrado por LPC em um domínio de espectro, e o quarto domínio é um domínio de espectro obtido pela conversão do domínio de frequência do primeiro sinal de domínio.

5. Codificador de áudio, de acordo com uma das reivindicações anteriores, **caracterizado** por compreender ainda um

controlador (300, 525), para controlar o primeiro comutador (200) ou o segundo comutador (521) de maneira adaptativa de sinal,

onde o controlador é operante para analisar uma entrada de sinal no primeiro comutador (200) ou saída pelo primeiro ramo de codificação ou pelo segundo ramo de codificação, ou um sinal obtido decodificando-se um sinal de saída do primeiro ramo de codificação ou do segundo ramo de codificação em relação a uma função-alvo, ou

onde o controlador (300, 525) é operante para analisar uma entrada de sinal no segundo comutador (521) ou saída pelo primeiro ramo de processamento ou pelo segundo ramo de processamento, ou sinais obtidos por sinais de saída de processamento inverso pelo primeiro ramo de processamento (522) e pelo segundo ramo de processamento (523, 524), em relação a uma função-alvo.

6. Codificador de áudio, de acordo com uma das reivindicações anteriores, **caracterizado** pelo fato de que o primeiro ramo de codificação (400) ou o segundo ramo de processamento (523, 524) do segundo ramo de codificação (500) compreendem um conversor de tempo/frequência introdutor de *aliasing* e um estágio de codificador de quantizador/entropia (421), onde o primeiro ramo de processamento do segundo ramo de codificação inclui um estágio de codificação de quantizador ou entropia (522) sem conversão de introdução de *aliasing*.

7. Codificador de áudio, de acordo com a reivindicação 6, **caracterizado** pelo fato de que o conversor de tempo/frequência introdutor de *aliasing* compreende dispositivo de janelamento para aplicação de uma janela de análise e um algoritmo de transformação de cosseno discreta modificada (MDCT), o disposi-

tivo de janelamento sendo operado para aplicar a função de janela a estruturas subsequentes de maneira sobreposta, de modo que uma amostra de um sinal de entrada no dispositivo de janelamento ocorra em pelo menos duas estruturas subsequentes.

8. Codificador de áudio, de acordo com uma das reivindicações anteriores, **caracterizado** pelo fato de que o primeiro ramo de processamento (522) compreende a codificação de excitação de LPC de um codificador de previsão linear excitada de código algébrico (ACELP), e o segundo ramo de processamento compreende um conversor de espectro de MDCT e um quantizador para quantizar componentes de espectro para obter componentes de espectro quantizados, onde cada componente de espectro quantizado é zero ou é definido por um índice de quantização de uma pluralidade de índices de quantização.

9. Codificador de áudio, de acordo com a reivindicação 5, **caracterizado** pelo fato de que o controlador é operado para controlar o primeiro comutador (200) em modo de *loop* aberto e para controlar o segundo comutador (521) em modo de *loop* fechado.

10. Codificador de áudio, de acordo com uma das reivindicações anteriores, **caracterizado** pelo fato de que o primeiro ramo de codificação e o segundo ramo de codificação são operados para codificar o sinal de áudio em blocos, onde o primeiro comutador ou o segundo comutador alternam em blocos, de maneira que ocorra uma ação de alternância, no mínimo, após um bloco de número pré-definido de amostras de um sinal, sendo que o número pré-definido de amostras foram uma extensão de estrutura para o comutador correspondente (521, 200).

11. Codificador de áudio, de acordo com a reivin-

dicação 10, **caracterizado** pelo fato de que a extensão de estrutura para o primeiro comutador é pelo menos o dobro do tamanho da extensão de estrutura do segundo comutador.

12. Codificador de áudio de acordo com a reivindicação 5, **caracterizado** pelo fato de que o controlador é operado para realizar uma discriminação de fala/música de maneira que uma decisão para fala seja favorecida em relação a uma decisão para música, de maneira que uma decisão para fala seja tomada mesmo quando uma parte de menos de 50% de uma estrutura para o primeiro comutador seja fala e a parte de mais de 50% da estrutura seja música.

13. Codificador de áudio, de acordo com a reivindicação 5 ou 12, **caracterizado** pelo fato de que uma estrutura para o segundo comutador é menor que uma estrutura para o primeiro comutador, e onde o controlador (525, 300) é operado para tomar uma decisão para fala somente quando é estabelecido que uma parte da primeira estrutura que tem uma extensão que é de mais de 50% da extensão da segunda estrutura inclui música.

14. Codificador de áudio, de acordo com a uma das reivindicações anteriores, **caracterizado** pelo fato de que o primeiro ramo de codificação (400) ou o segundo ramo de processamento do segundo ramo de codificação inclui uma funcionalidade de distorção de tempo variável.

15. Método de codificação de um sinal de entrada de áudio (195), estando o sinal de entrada de áudio em um primeiro domínio, **caracterizado** por compreender:

codificação (400) de um sinal de áudio utilizando um primeiro algoritmo de codificação para obter um primeiro sinal

codificado;

codificação (500) de um sinal de áudio utilizando um segundo algoritmo de codificação para obter um segundo sinal codificado, onde o primeiro algoritmo de codificação é diferente do segundo algoritmo de codificação; e

alternância (200) entre a codificação que utiliza o primeiro algoritmo de codificação e a codificação que utiliza o segundo algoritmo de codificação de maneira que, para uma parte do sinal de entrada de áudio, o primeiro sinal codificado ou o segundo sinal codificado esteja em um sinal de saída codificado, ,

onde a codificação (500) que utiliza o segundo algoritmo de codificação compreende:

conversão (510) do sinal de áudio em um segundo domínio diferente do primeiro domínio,

processamento (522) de um sinal de áudio no segundo domínio para obter um primeiro sinal processado;

conversão (523) de um sinal em um terceiro domínio diferente do primeiro domínio e do segundo domínio, e processamento (524) do sinal no terceiro domínio para obter um segundo sinal processado; e

alternância (521) entre processamento (522) do sinal de áudio e conversão (523) e processamento (524) de maneira que, para uma parte do sinal de áudio codificada utilizando o segundo algoritmo de codificação, o primeiro sinal processado ou o segundo sinal processado esteja no segundo sinal codificado.

16. Decodificador para decodificar um sinal de áudio codificado, **caracterizado** pelo sinal de áudio codificado compreender um primeiro sinal codificado, um primeiro sinal pro-

cessado em um segundo domínio, e um segundo sinal processado em um terceiro domínio, onde o primeiro sinal codificado, o primeiro sinal processado e o segundo sinal processado estão relacionados a diferentes partes de tempo de um sinal de áudio decodificado, e onde um primeiro domínio, o segundo domínio e o terceiro domínio são diferentes uns dos outros, compreendendo:

- um primeiro ramo de decodificação (431, 440) para decodificar o primeiro sinal codificado com base no primeiro algoritmo de codificação;

- um segundo ramo de decodificação para decodificar o primeiro sinal processado ou o segundo sinal processado,

 - onde o segundo ramo de decodificação compreende

 - um primeiro ramo de processamento inverso (531) para processar inversamente o primeiro sinal processado para obter um primeiro sinal processado inverso no segundo domínio;

- um segundo ramo de processamento inverso (533, 534) para processar inversamente o segundo sinal processado para obter um segundo sinal processado inverso no segundo domínio;

- um primeiro combinador (532) para combinar o primeiro sinal processado inverso e o segundo sinal processado inverso para obter um sinal combinado no segundo domínio; e

- um conversor (540) para converter o sinal combinado ao primeiro domínio; e

- um segundo combinador (600) para combinar o sinal convertido no primeiro domínio e a saída do primeiro sinal decodificado pelo primeiro ramo de decodificação para obter um sinal de saída decodificado no primeiro domínio.

17. Decodificador, de acordo com a reivindicação

16, **caracterizado** pelo fato de que o primeiro combinador (532) ou o segundo combinador (600) compreende um comutador com uma funcionalidade de fusão de transmissão.

18. Decodificador, de acordo com a reivindicação 16 ou 17, **caracterizado** pelo fato de que o primeiro domínio é um domínio de tempo, o segundo domínio é um domínio de LPC, o terceiro domínio é um domínio de espectro de LPC, ou o primeiro sinal codificado é codificado em um quarto domínio, que é um domínio de espectro de tempo obtido por conversão de tempo/frequência de um sinal do primeiro domínio.

19. Decodificador, de acordo com qualquer das reivindicações 16 a 18, **caracterizado** pelo fato de que o primeiro ramo de decodificação (431, 440) compreende um codificador inverso e um desquantizador, e um conversor de domínio de frequência/domínio de tempo (440), ou

o segundo ramo de decodificação compreende um codificador inverso e um desquantizador no primeiro ramo de processamento inverso ou um codificador inverso e um desquantizador e um domínio de espectro de LPC para conversor de domínio de LPC (534) no segundo ramo de processamento inverso.

20. Decodificador, de acordo com a reivindicação 19, **caracterizado** pelo fato de que o primeiro ramo de decodificação ou o segundo ramo de processamento inverso compreende um adicionador de sobreposição para realizar uma funcionalidade de cancelamento de *aliasing* de domínio de tempo.

21. Decodificador, de acordo com uma das reivindicações de 16 a 20, **caracterizado** pelo fato de que o primeiro ramo de decodificação ou o segundo ramo de processamento inverso

compreende um reversor de distorção controlado por uma característica de distorção incluída no sinal de áudio codificado.

22. Decodificador, de acordo com uma das reivindicações de 16 a 21, **caracterizado** pelo fato de que o sinal codificado inclui, como informações paralelas (4a), uma indicação se um sinal codificado deve ser codificado por um primeiro ramo de codificação ou por um segundo ramo de codificação, ou um primeiro ramo de processamento do segundo ramo de codificação ou um segundo ramo de processamento do segundo ramo de codificação, e

que compreende ainda um analisador para analisar o sinal codificado para determinar, com base nas informações paralelas (4a), se um sinal codificado deve ser processado pelo primeiro ramo de decodificação, ou pelo segundo ramo de decodificação, ou pelo primeiro ramo de processamento inverso do segundo ramo de decodificação ou pelo segundo ramo de processamento inverso do segundo ramo de decodificação.

23. Método de codificação de um sinal de áudio codificado, **caracterizado** pelo o sinal de áudio codificado compreender um primeiro sinal codificado, um primeiro sinal processado em um segundo domínio, e um segundo sinal processado em um terceiro domínio, onde o primeiro sinal codificado, o primeiro sinal processado e o segundo sinal processado estão relacionados a diferentes partes de tempo de um sinal de áudio decodificado, e onde um primeiro domínio, o segundo domínio e o terceiro domínio são diferentes uns dos outros, compreendendo:

decodificação (431, 440) do primeiro sinal codificado com base em um primeiro algoritmo de codificação;

decodificação do primeiro sinal processado ou do

segundo sinal processado,

onde a decodificação do primeiro sinal processado ou do segundo sinal processado compreende:

processamento inverso (531) do primeiro sinal processado para obter um primeiro sinal processado inverso no segundo domínio;

processamento inverso (533, 534) do segundo sinal processado para obter um segundo sinal processado inverso no segundo domínio;

combinação (532) do primeiro sinal processado inverso e do segundo sinal processado inverso para obter um sinal combinado no segundo domínio; e

conversão (540) do sinal combinado ao primeiro domínio; e

combinação (600) do sinal convertido do primeiro domínio e do primeiro sinal decodificado para obter um sinal de saída decodificado no primeiro domínio.

24. Sinal de áudio codificado, **caracterizado** por compreender:

um primeiro sinal codificado ou a ser decodificado utilizando um primeiro algoritmo de codificação,

um primeiro sinal processado em um segundo domínio, e um segundo sinal processado em um terceiro domínio, onde o primeiro sinal processado e o segundo sinal processado são codificados utilizando um segundo algoritmo de codificação,

onde o primeiro sinal codificado, o primeiro sinal processado e o segundo sinal processado estão relacionados a diferentes partes de tempo de um sinal de áudio decodificado,

onde um primeiro domínio, o segundo domínio e o terceiro domínio são diferentes uns dos outros, e

informações paralelas (4a) que indicam se uma parte do sinal codificado é o primeiro sinal codificado, o primeiro sinal processado ou o segundo sinal processado.

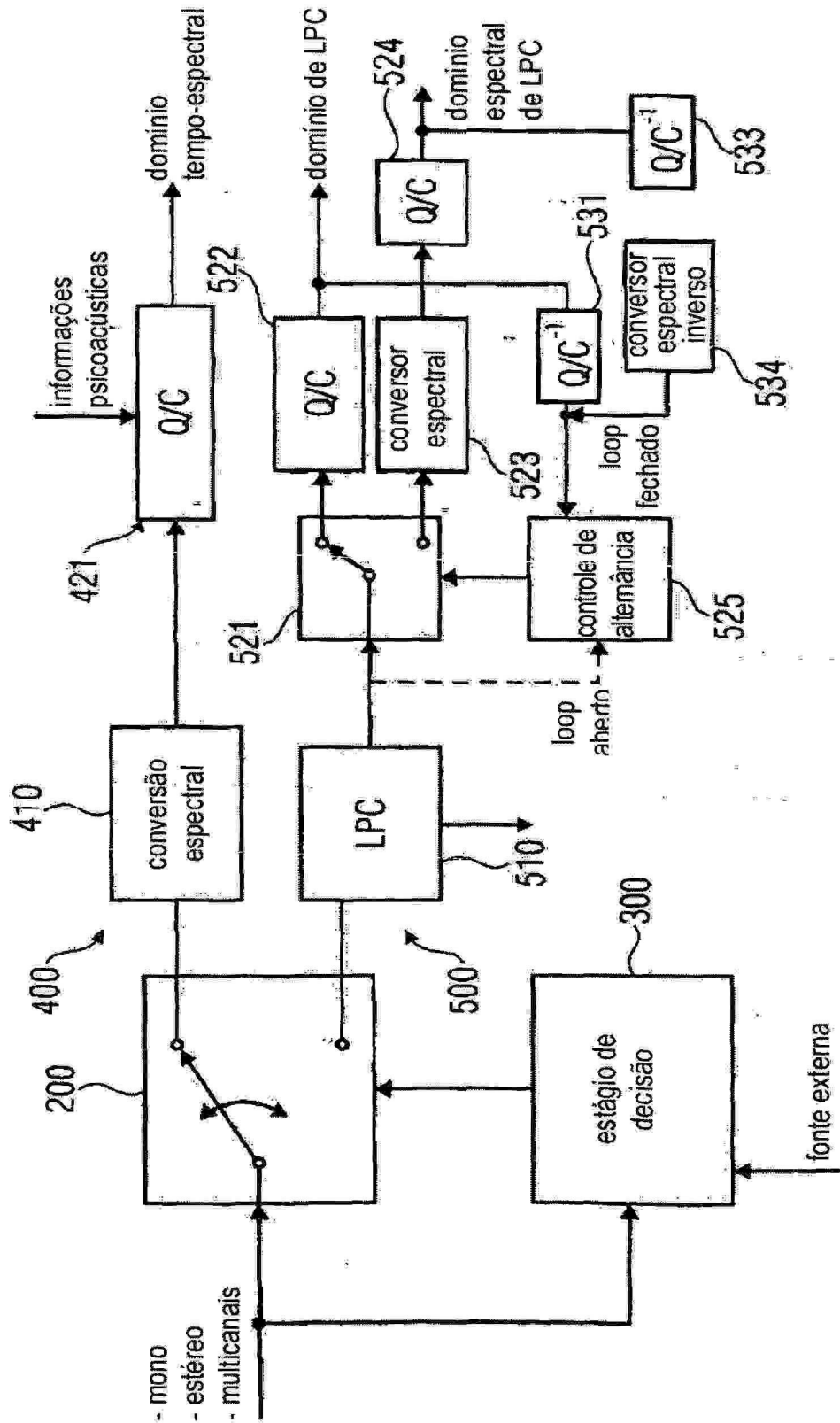


FIG 1A
(Codificador)

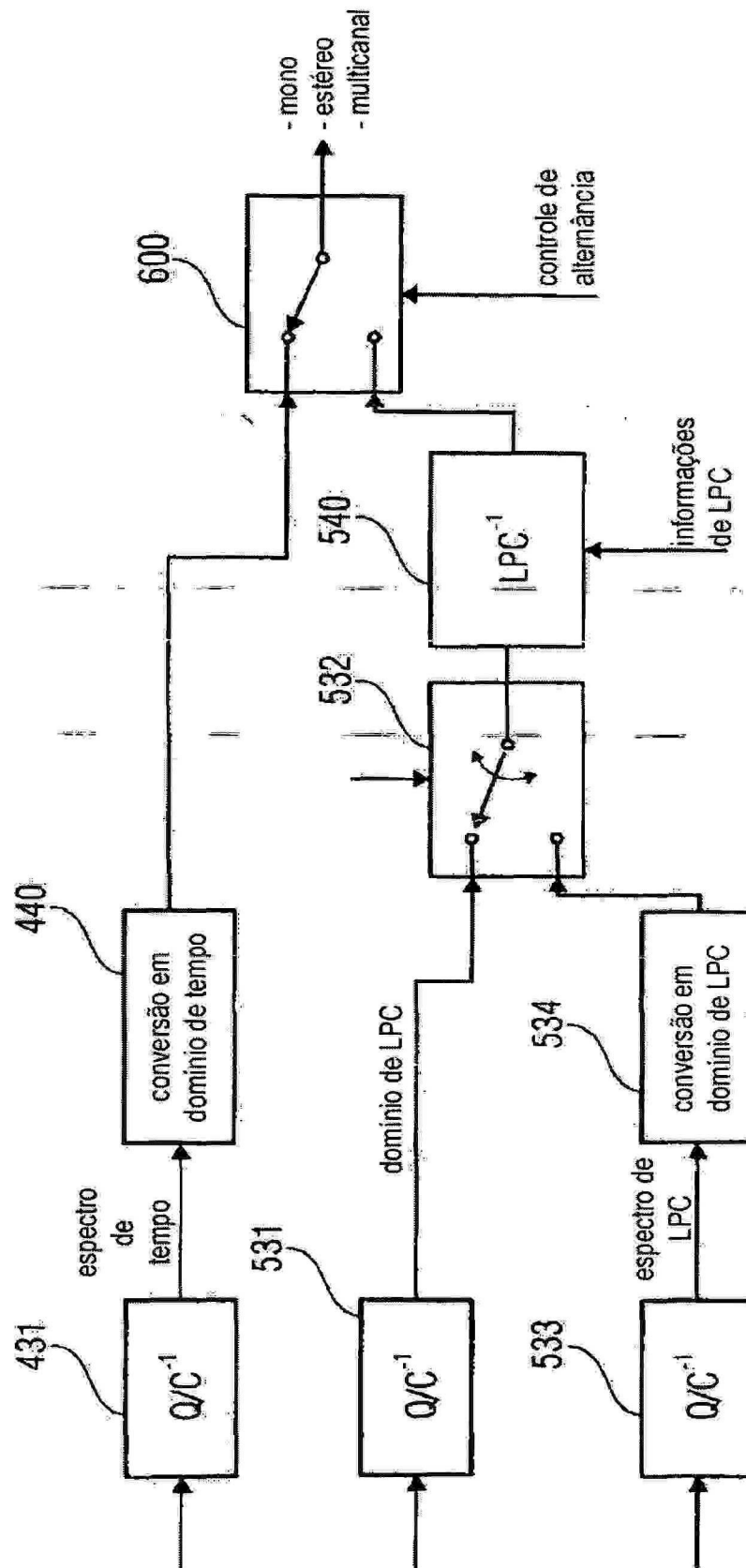


FIG 1B

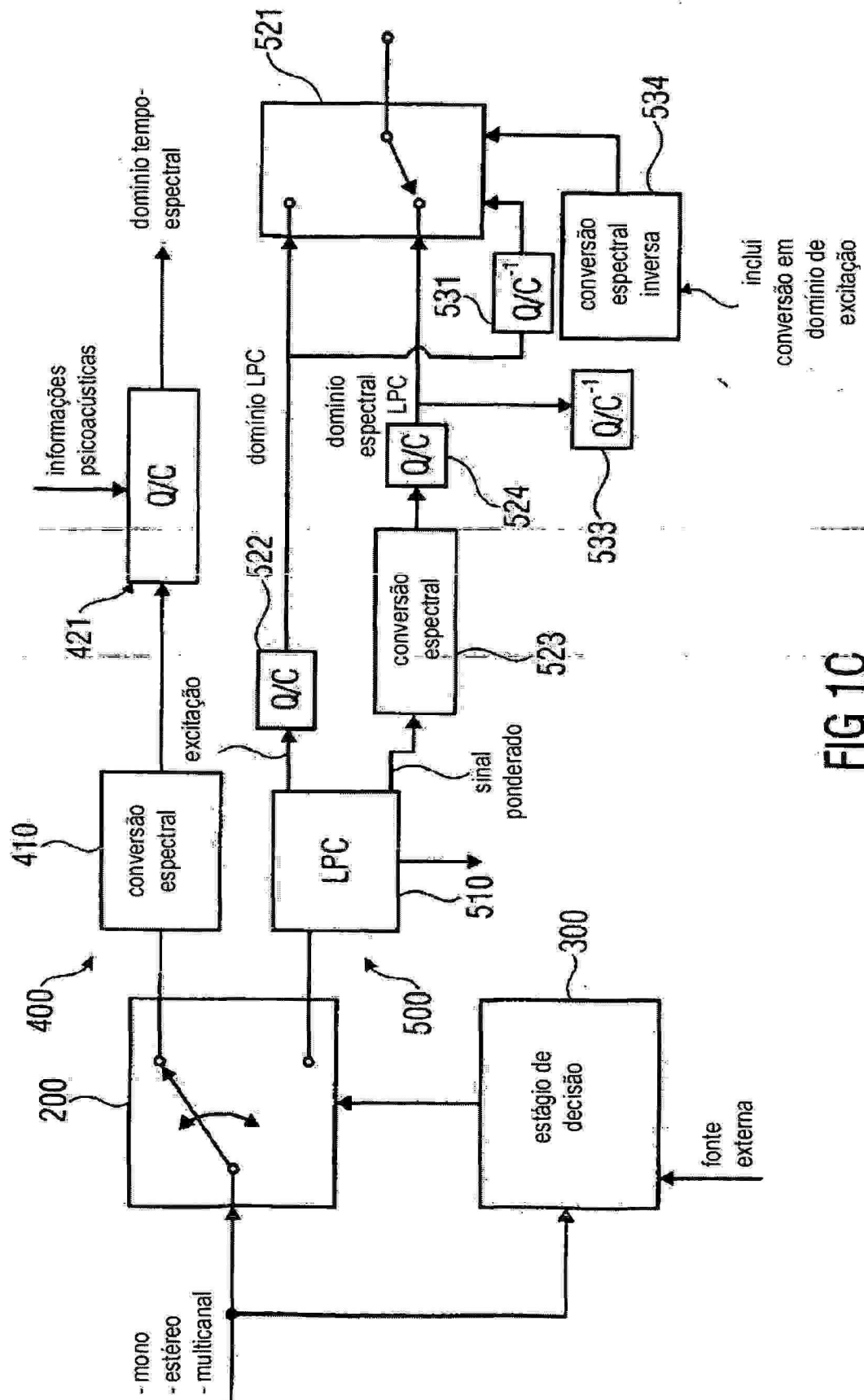


FIG 10

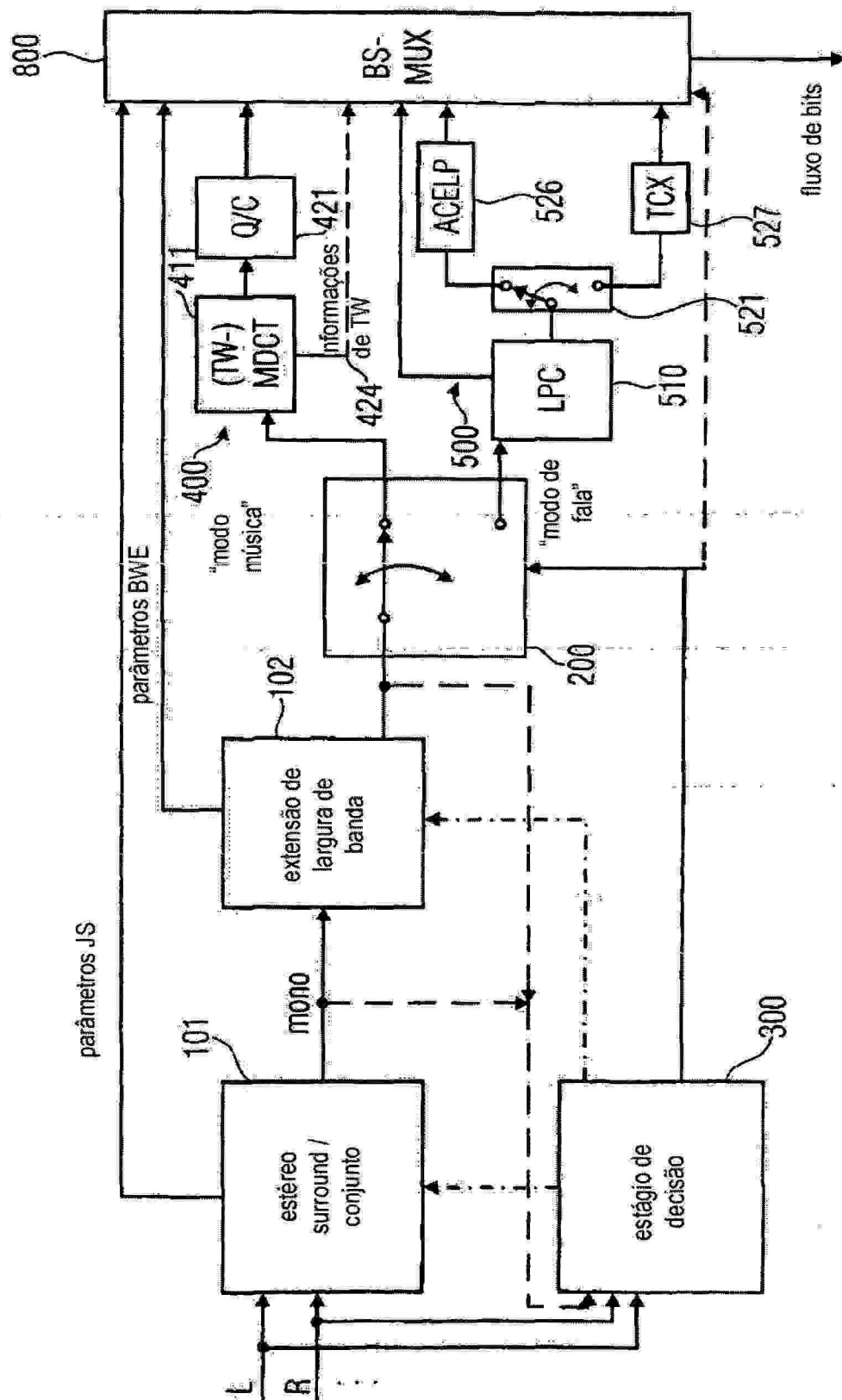


FIG 2A
(Codificador)

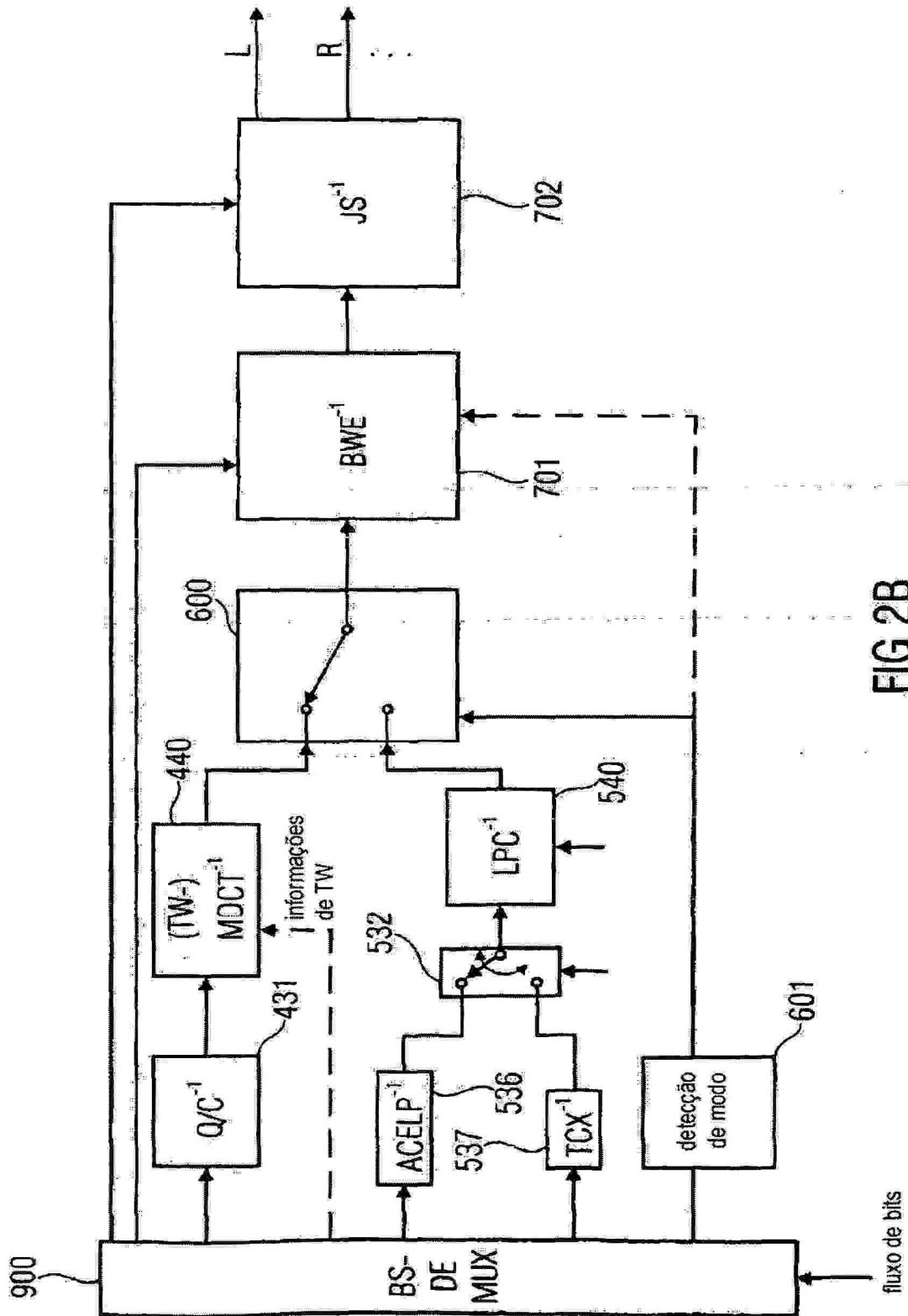


FIG 2B
(Decodificador)

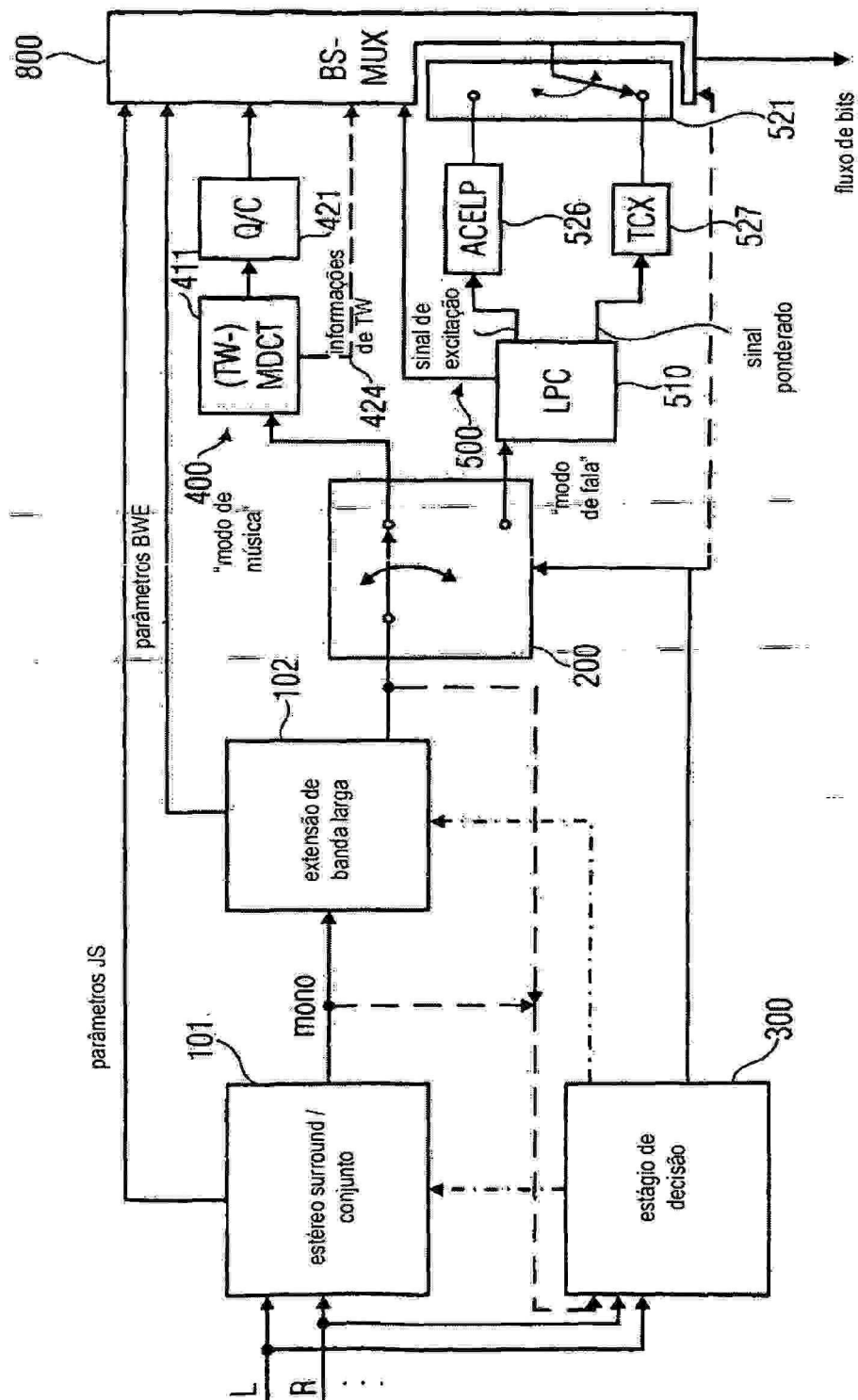


FIG 2C
(Codificador)

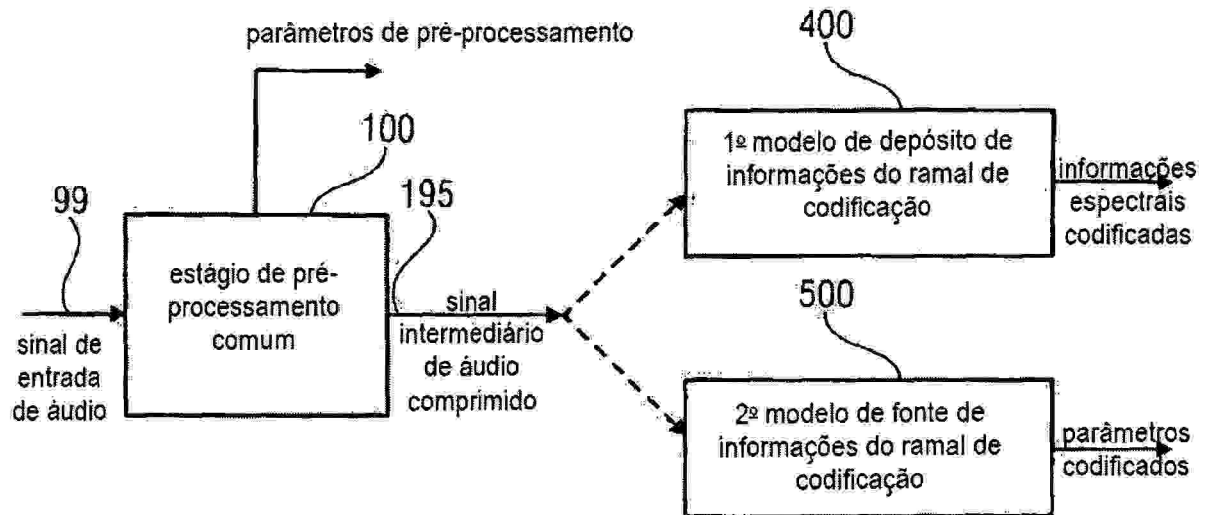


FIG 3A

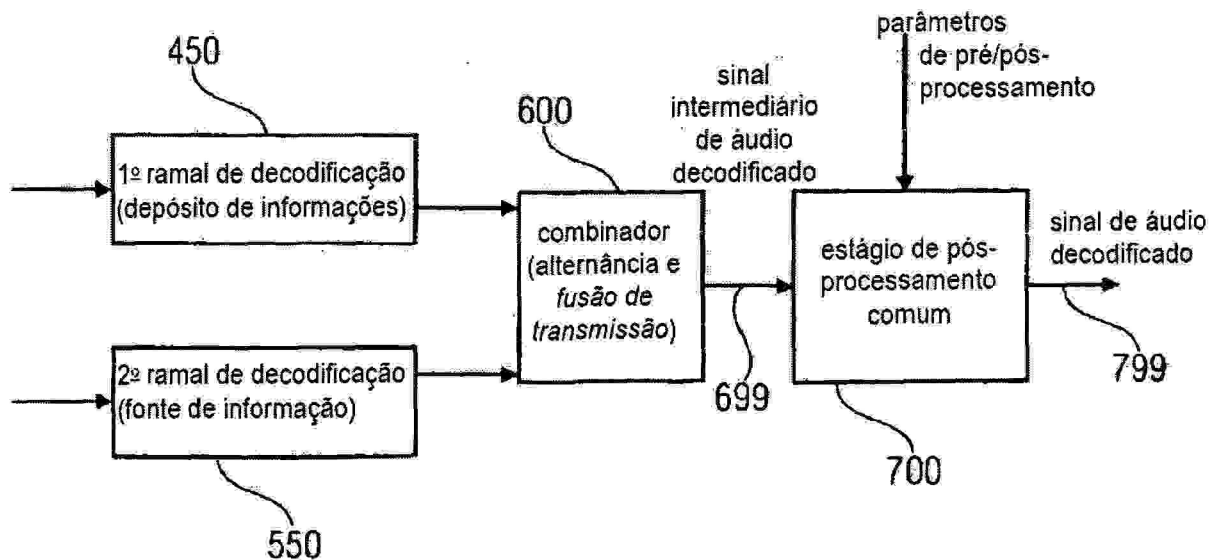
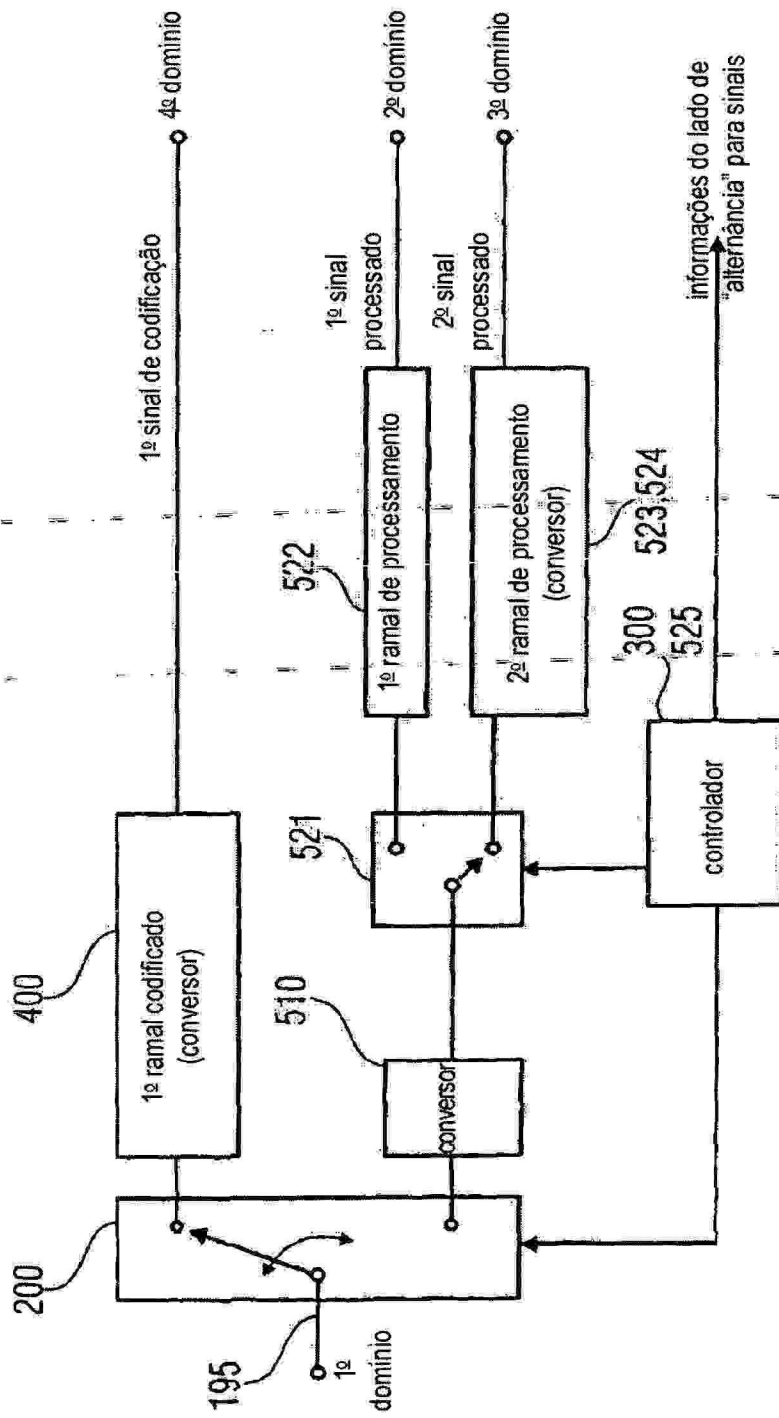


FIG 3B



- cada bloco de sinal de áudio do 1º domínio é representado ou por um sinal codificado de 2º domínio, um 3º domínio, ou 4º domínio, separadamente de uma região de passagem opcional

FIG 3C

9/21

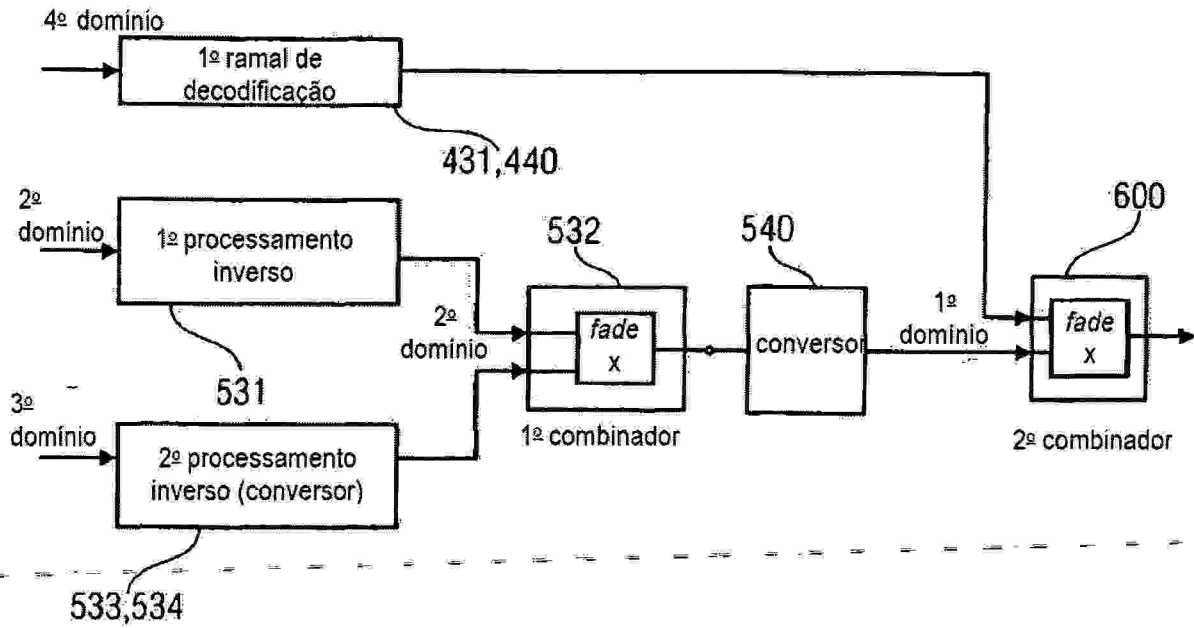


FIG 3D

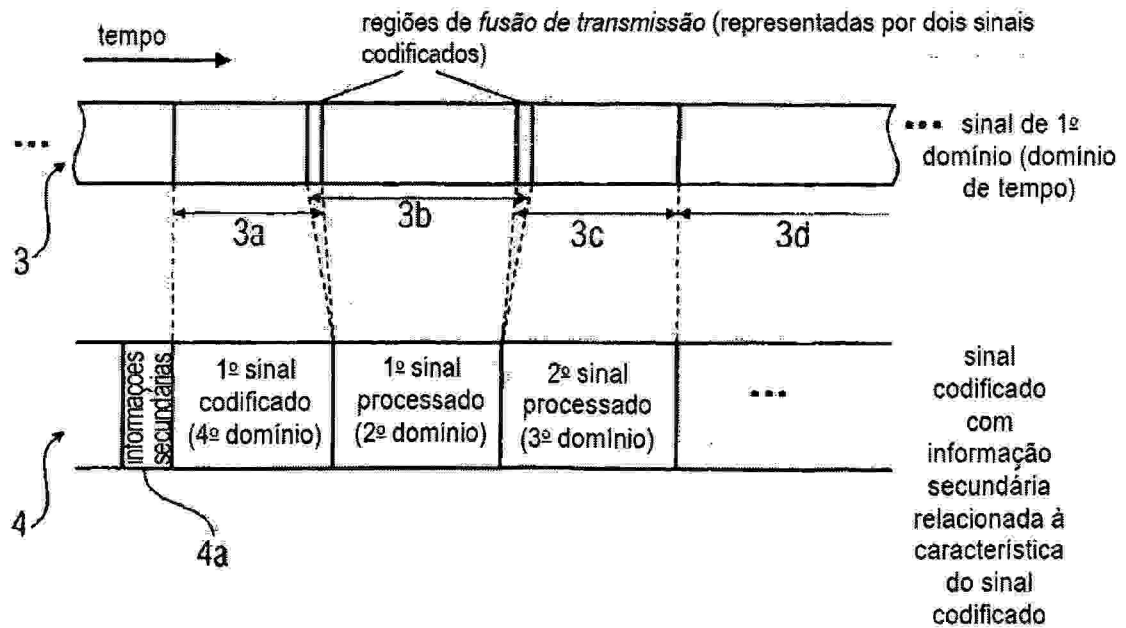


FIG 3E

10/21

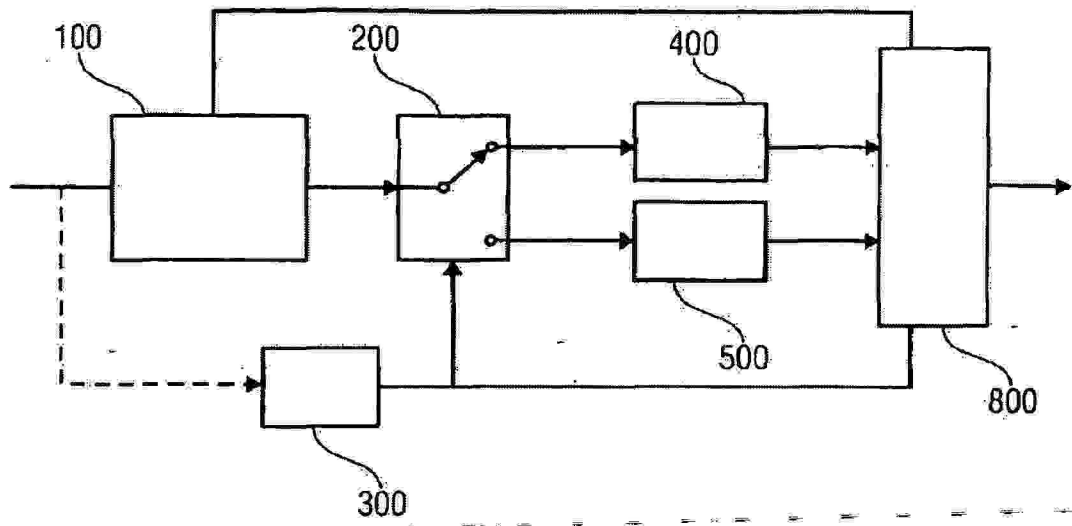


FIG 4A

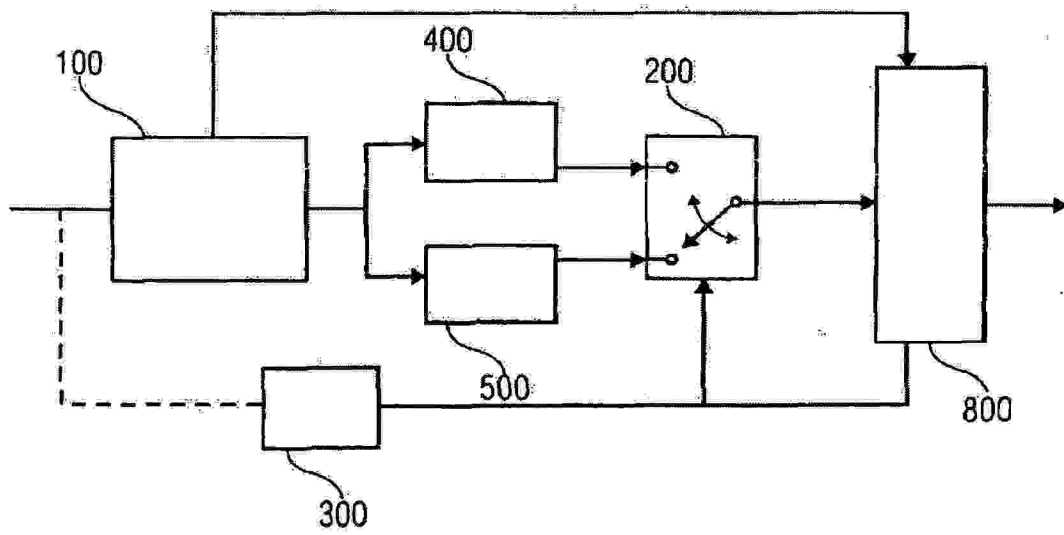


FIG 4B

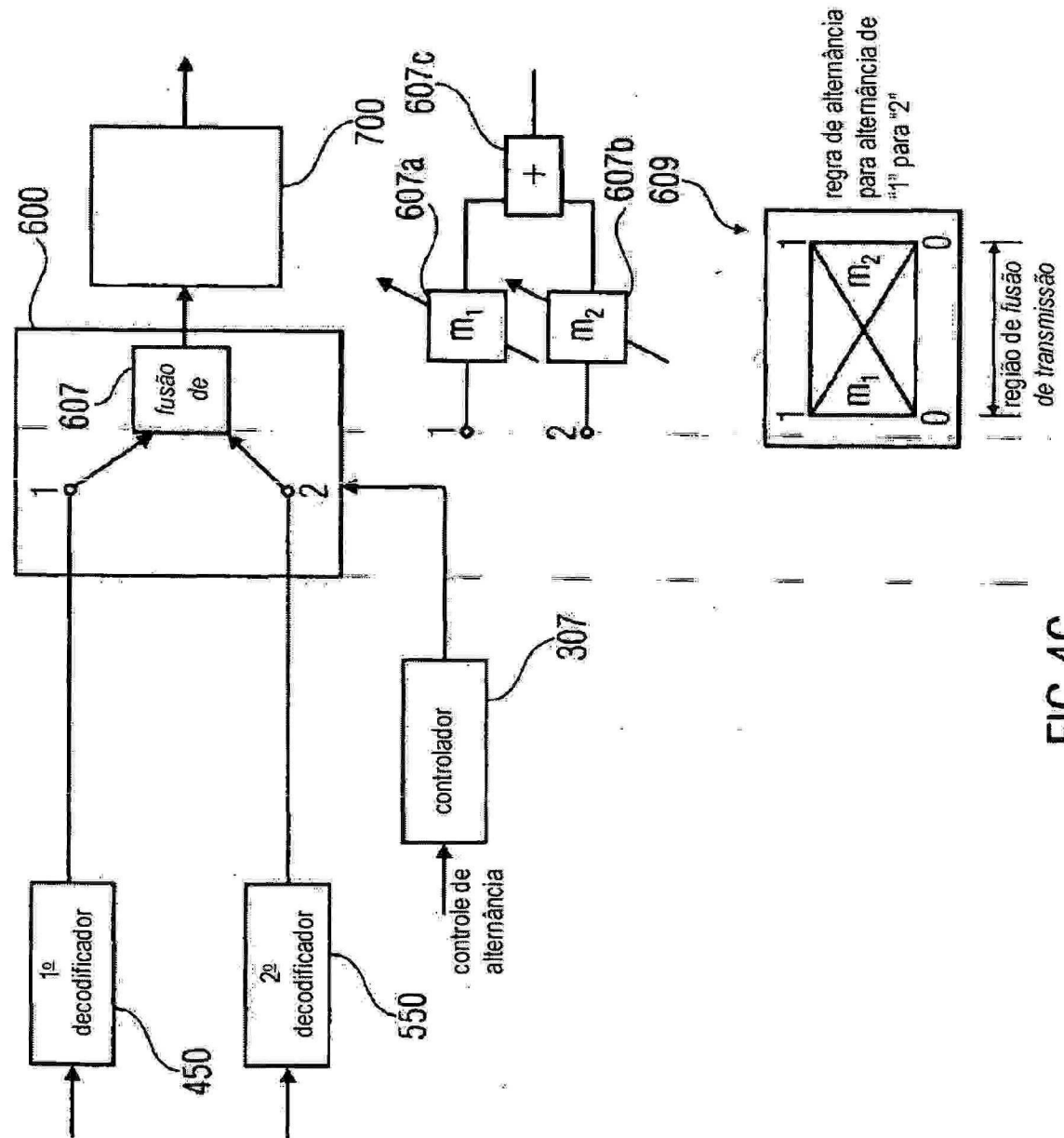


FIG 4C

segmento de sinal do tipo de segmento (por exemplo, fala com voz)

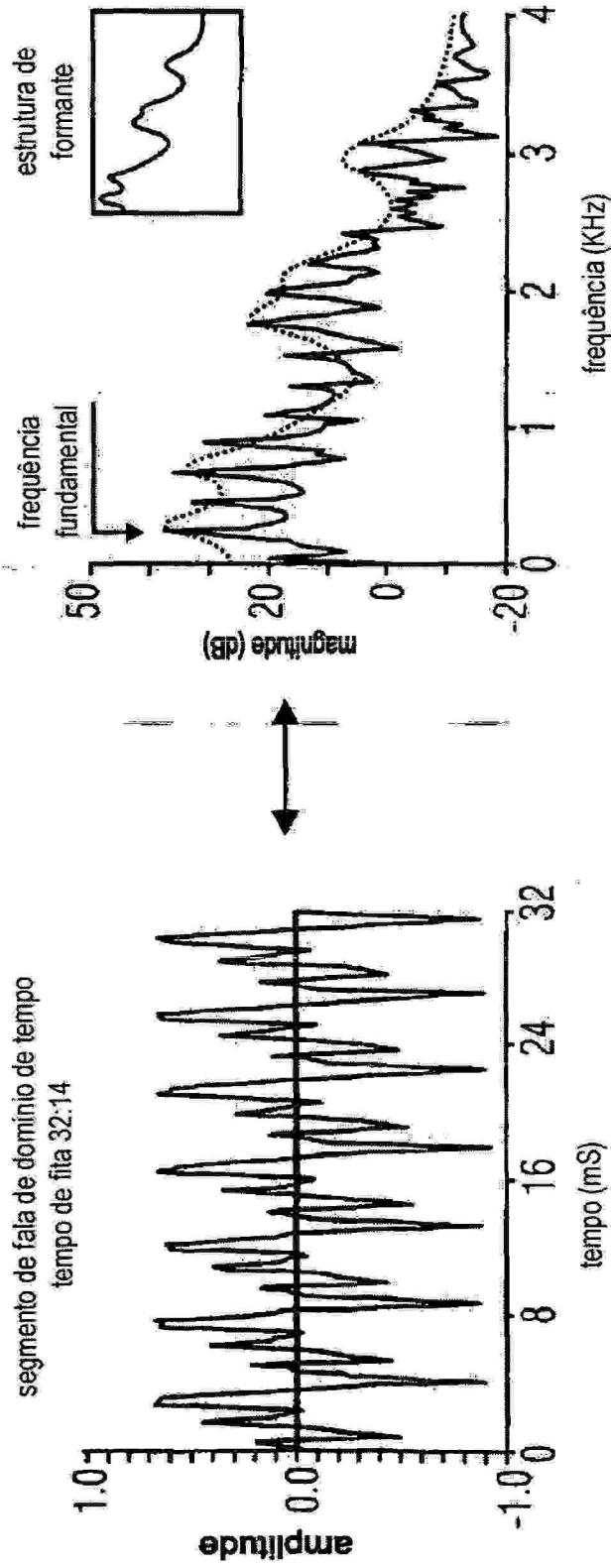


FIG 5A

FIG 5B

segmento fixo (por exemplo, fala sem voz)

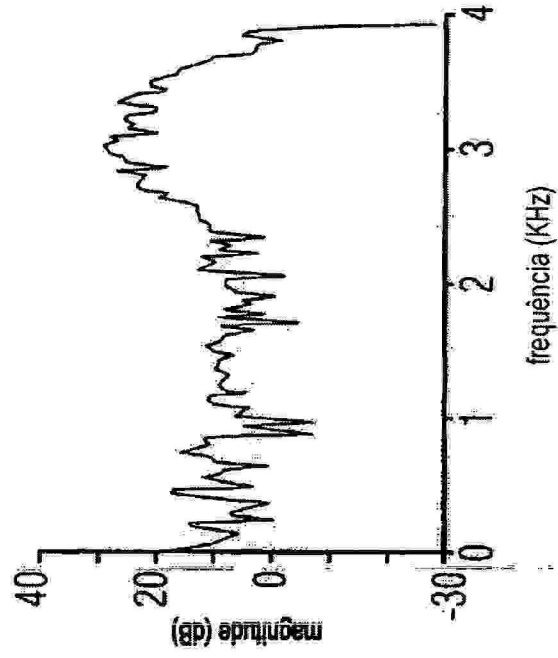
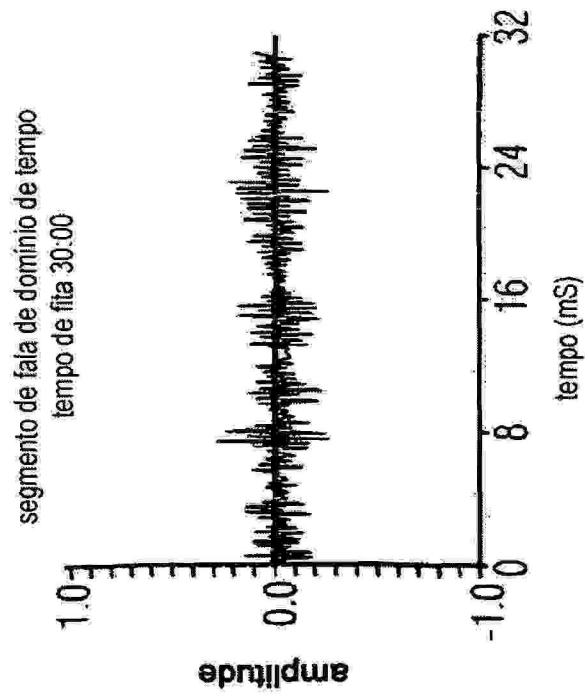
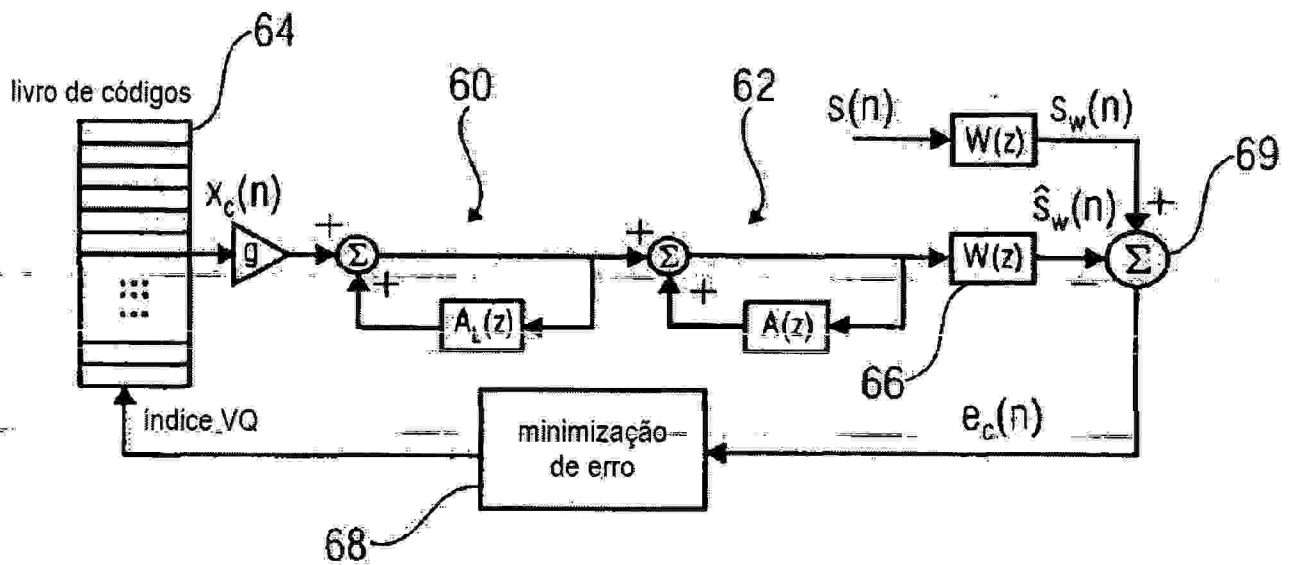


FIG 5C

FIG 5D

CELP de análise-por-síntese



$A_L(z)$: previsão de longo prazo
 $\triangleq$ estrutura (fina) da altura

$A(z)$: previsão de longo prazo
 $\triangleq$ estrutura de formato / envelope espectral

FIG 6

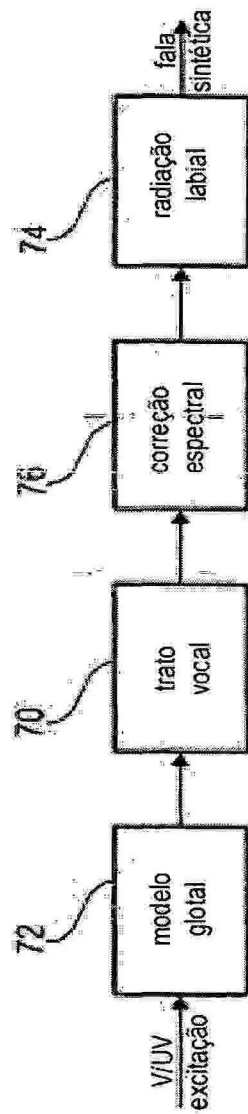


FIG 7A

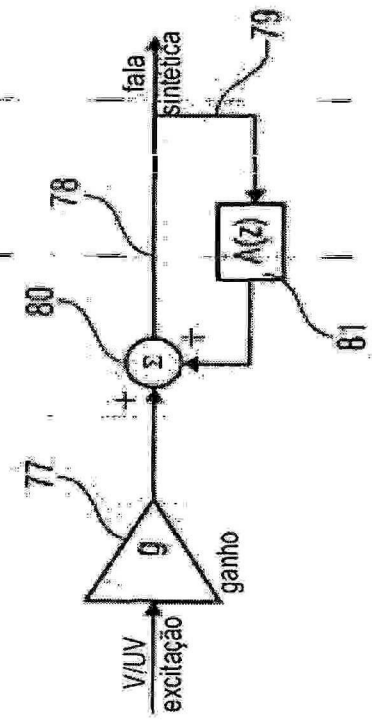


FIG 7B

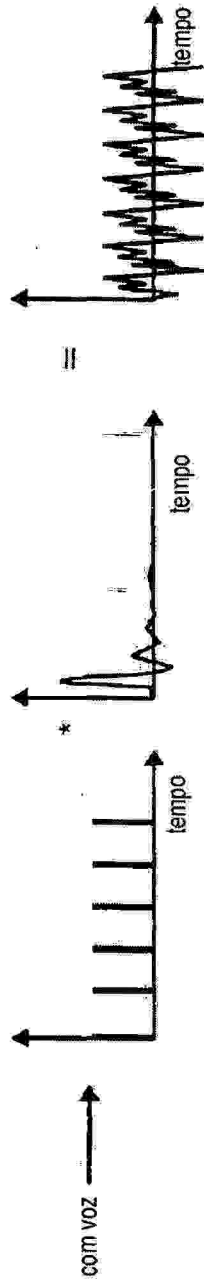


FIG 7C

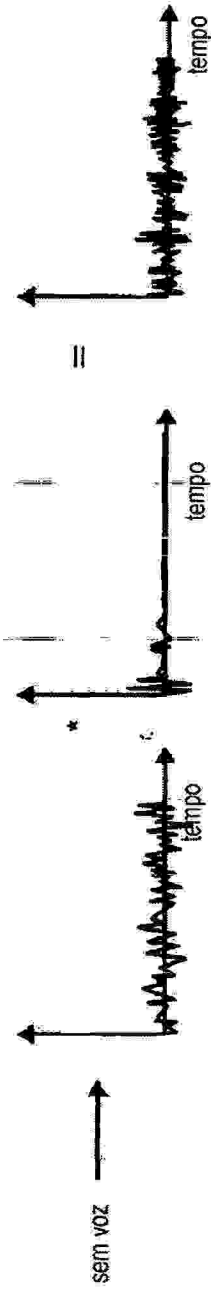


FIG 7D

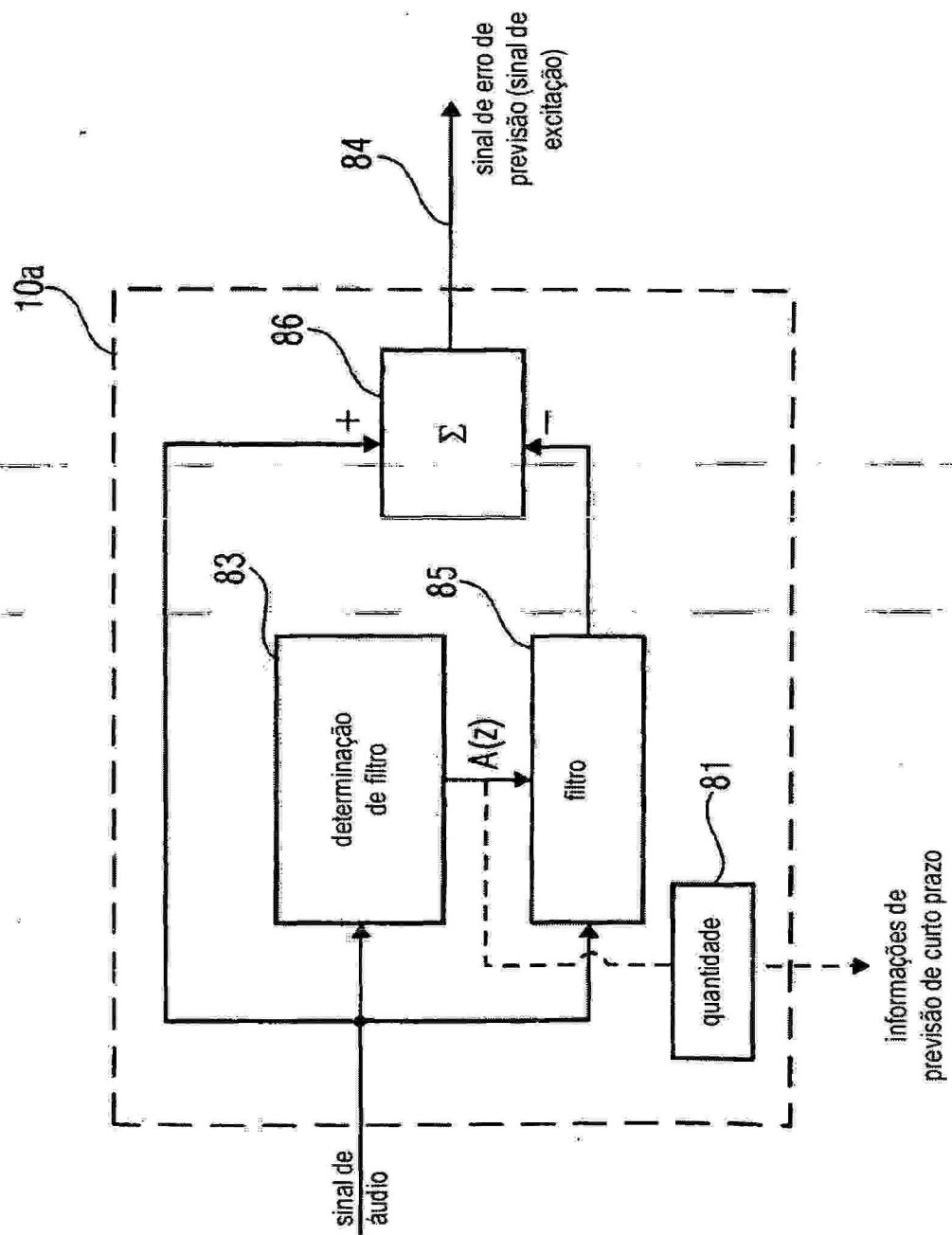


FIG 7E

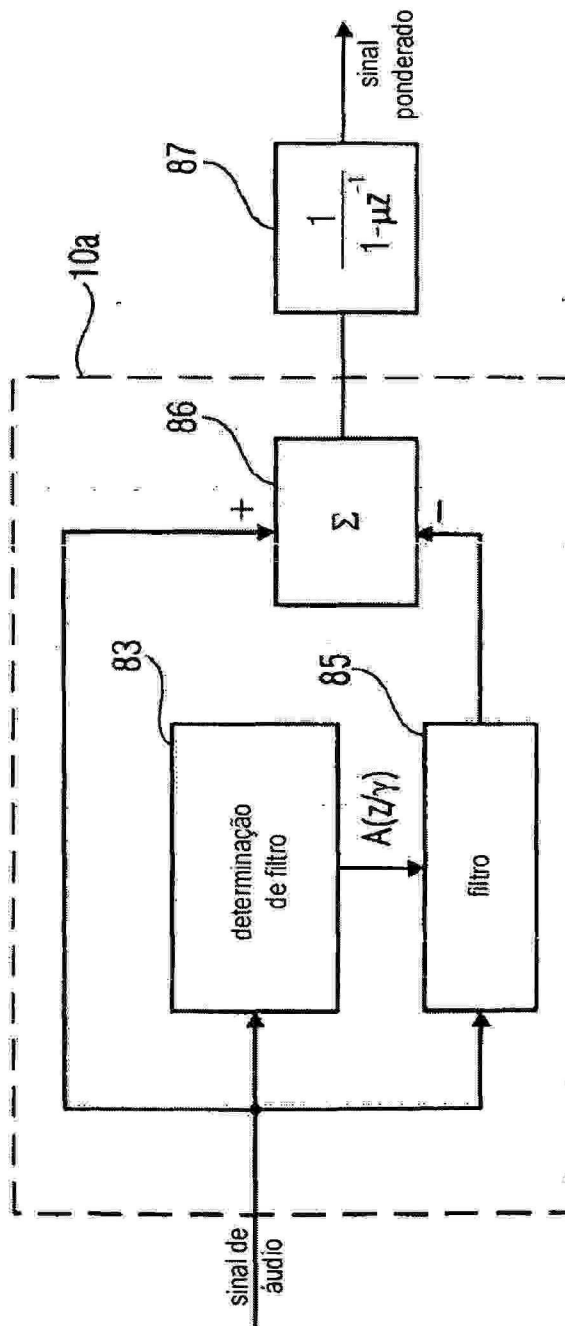


FIG 7F
(lado do codificador)

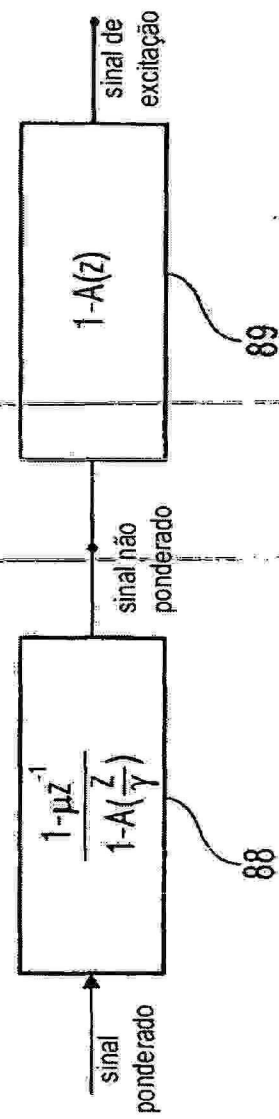


FIG 7G
(lado do decodificador)

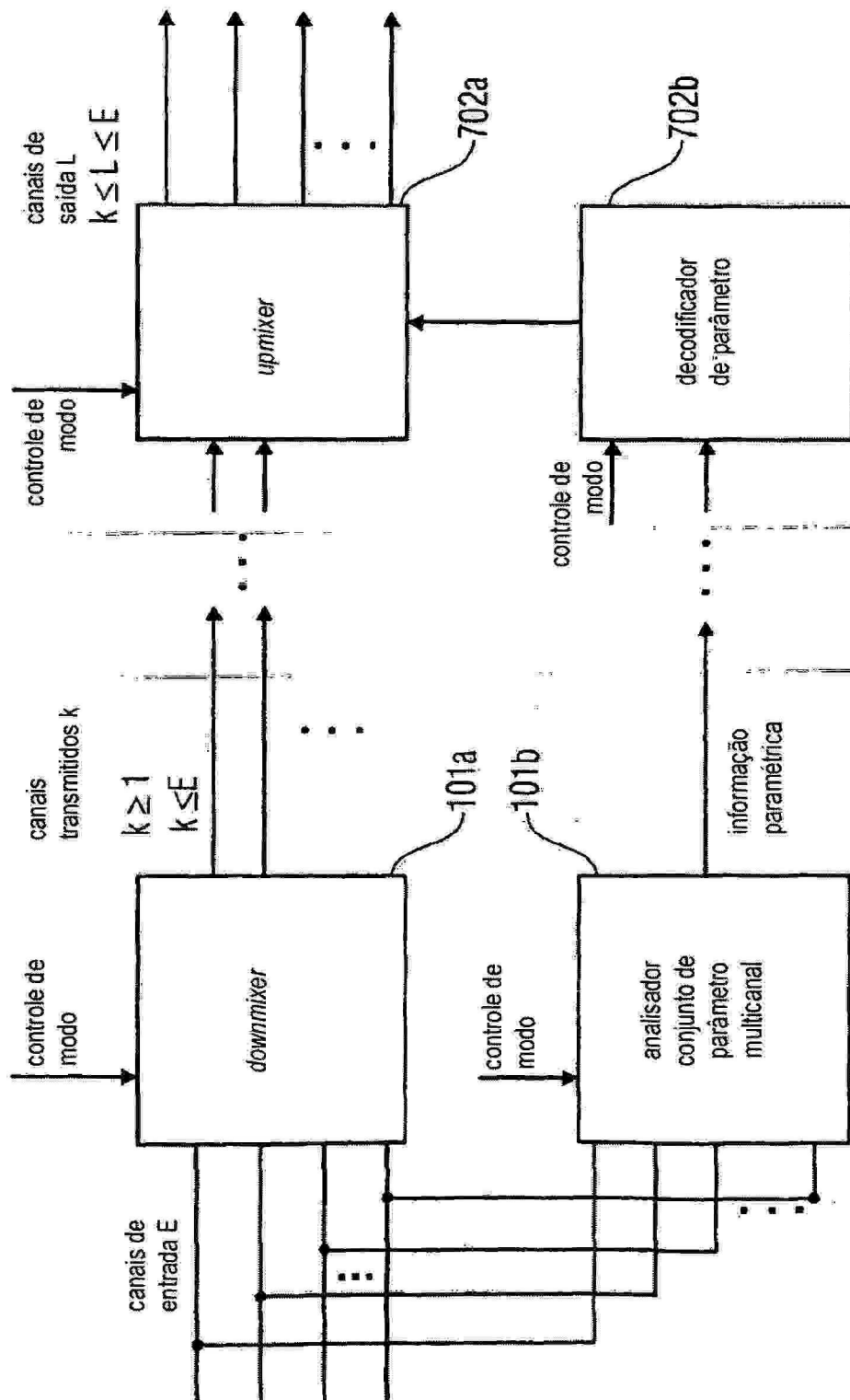


FIG 8

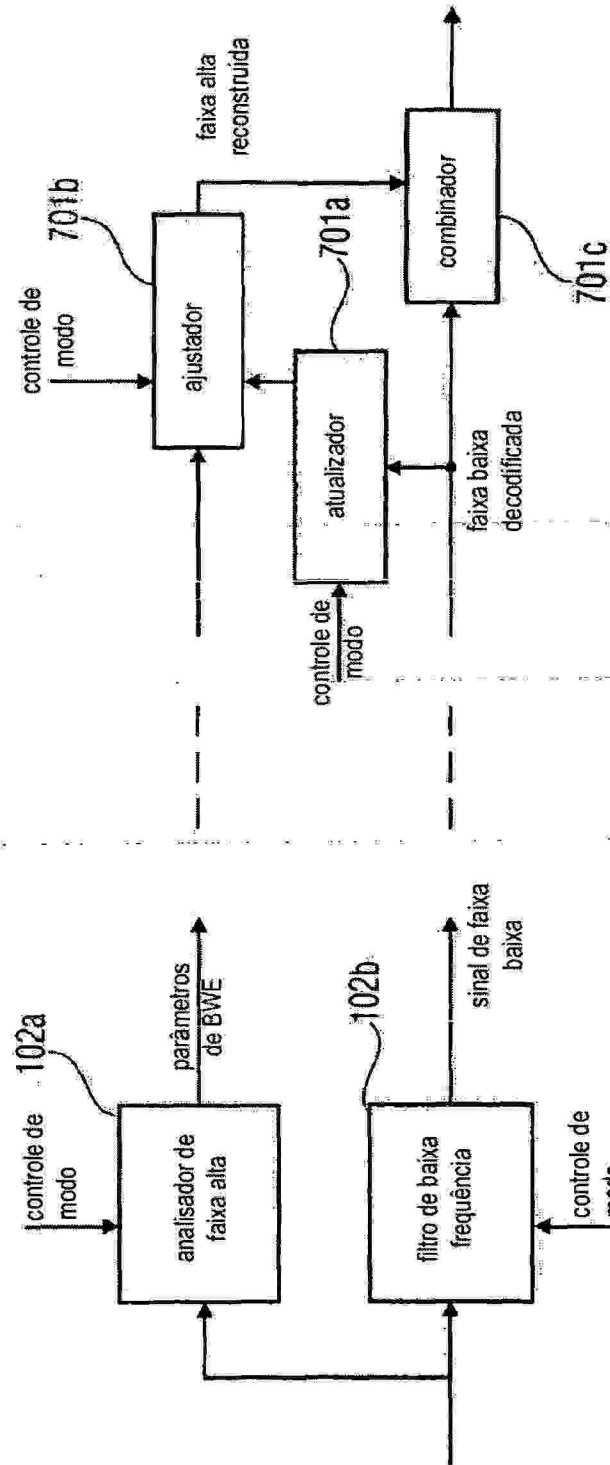
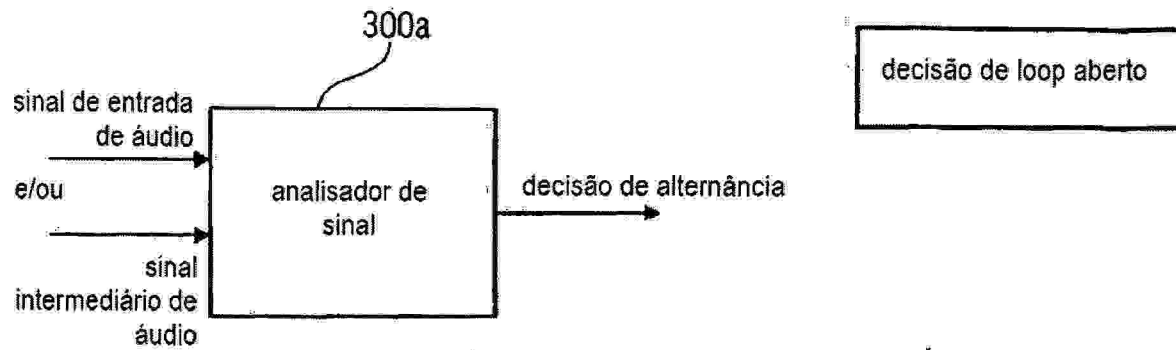


FIG 9

21/21



sinal intermediário de áudio:
- sinal de faixa baixa
- sinal de *downmix*; ou
- parte de faixa baixa do sinal de *downmix*

FIG 10A

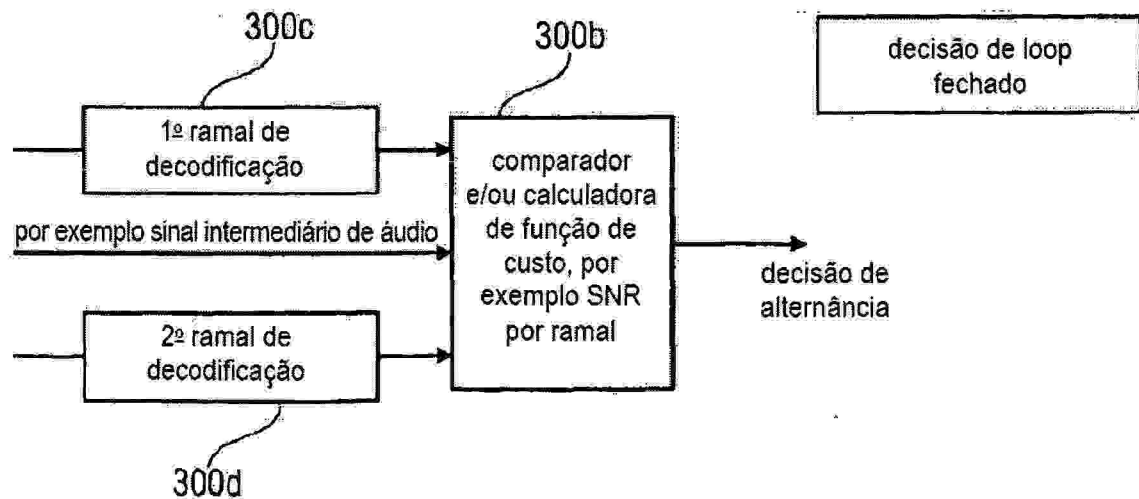


FIG 10B