



INPI
INSTITUTO
NACIONAL
DA PROPRIEDADE
INDUSTRIAL
Assinado
Digitalmente

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

CARTA PATENTE Nº PI 0910527-1

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 0910527-1

(22) Data do Depósito: 04/06/2009

(43) Data da Publicação Nacional: 26/07/2016

(51) Classificação Internacional: G06F 17/14; G10L 19/02; G10L 19/20; G10L 19/08.

(52) Classificação CPC: G10L 19/20; G10L 19/0212; G10L 19/08.

(30) Prioridade Unionista: EP 08017661.3 de 08/10/2008; US 61/103,825 de 08/10/2008; US 61/079,862 de 11/07/2008.

(54) Título: CODIFICADOR E DECODIFICADOR DE ÁUDIO PARA ESTRUTURAS DE CODIFICAÇÃO E DECODIFICAÇÃO DE SINAL DE ÁUDIO TESTADO

(73) Titular: VOICEAGE CORPORATION. Endereço: 750 Lucerne Road, Suite 250 Montreal H3R 2H6, QUEBEC H3R 2H6, CANADÁ(CA); FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V., Sociedade Alemã. Endereço: Hansastrasse 27c, 80686, Munchen, ALEMANHA(DE), Alemanha

(72) Inventor: GEIGER, RALF; GRILL, BERNHARD; BESSETTE, BRUNO.; GOURNAY, PHILIPPE.; FUCHS, GUILLAUME; MULTRUS, MARKUS; NEUENDORF, MAX; SCHULLER, GERALD.

(87) Publicação PCT: WO 2010/003491 de 14/01/2010

Prazo de Validade: 10 (dez) anos contados a partir de 05/11/2019, observadas as condições legais

Expedida em: 05/11/2019

Assinado digitalmente por:

Liane Elizabeth Caldeira Lage

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

**"CODIFICADOR E DECODIFICADOR DE ÁUDIO PARA
ESTRUTURAS DE CODIFICAÇÃO E DECODIFICAÇÃO DE SINAL DE ÁUDIO
TESTADO"**

Especificação

5 A presente invenção refere-se à codificação de fonte e, em especial, à codificação da fonte de áudio, na qual um sinal de áudio é processado por dois codificadores de áudio distintos com algoritmos de codificação diferentes.

 No contexto da tecnologia de áudio com *bitrate*
10 baixo e de codificação de fala, diferentes técnicas de codificação têm sido tradicionalmente empregadas, para alcançar a codificação de *bitrate* baixo desses sinais, com a melhor qualidade subjetiva possível em um determinado *bitrate*. Os codificadores de sinais de música / de som, em geral, têm como objetivo a otimização da
15 qualidade subjetiva, dando forma a um modelo espectral (e temporal) do erro de quantização, de acordo com uma curva de limiar de mascaramento, que é calculada a partir do sinal de entrada, por meio de um modelo perceptual ("codificação de áudio perceptual"). Por outro lado, a codificação da fala, em taxas de
20 *bitrates* muito baixas, tem demonstrado que funciona de forma muito eficiente quando se baseia em um modelo de produção da fala humana, ou seja, empregando a Codificação Preditiva Linear (LPC) para modelar os efeitos de ressonância do trato vocal humano com uma codificação eficiente do sinal de excitação residual.

25 Como consequência dessas duas abordagens diferentes, os codificadores de áudio, em geral, como o MPEG-1 Layer 3 (MPEG = Grupo de Especialistas em Imagens com Movimento), ou o MPEG-2/4 Codificação de Áudio Avançada (AAC), não costumam

realizar, inclusive, para sinais de fala em índices de dados muito baixos, conforme os codificadores de fala baseados em LPC dedicados, em virtude da falta de utilização de um modelo de fonte de fala. Por outro lado, os codificadores de fala baseados em LPC
5 geralmente não atingem resultados convincentes quando aplicados em sinais de música em geral, em virtude de suas incapacidades de modelar flexivelmente o envelope espectral da distorção de codificação, de acordo com uma curva de limiar de mascaramento. A seguir são descritos os conceitos que combinam as vantagens da
10 codificação baseada em LPC e da codificação de áudio perceptual em uma única estrutura e, dessa forma, descrever uma codificação de áudio unificada eficiente tanto para áudio em geral quanto para
sinais de fala.

Tradicionalmente, os codificadores de áudio
15 perceptuais utilizam uma abordagem baseada em banco de filtro para codificar eficientemente os sinais de áudio e modelar a distorção de quantização, de acordo com uma estimativa da curva de mascaramento.

A Fig. 16a mostra o diagrama em blocos básico de
20 um sistema de codificação perceptual monofônico. Um banco de filtro de análise 1600 é utilizado para mapear as amostras no domínio do tempo em componentes espectrais subamostrados. Dependendo da quantidade de componentes espectrais, o sistema também é conhecido como um codificador de sub-bandas (pequena
25 quantidade de sub-bandas, por exemplo, 32) ou um codificador de transformação (grande quantidade de linhas de frequência, por exemplo, 512). Um modelo ("psicoacústico") perceptual 1602 é utilizado para estimar o limiar de mascaramento dependente do

tempo real. Os componentes espectrais ("sub-banda" ou "domínio da frequência") são quantizados e codificados 1604 de tal forma que o ruído de quantização fica oculto sob o sinal real transmitido, além de não ser perceptível após a decodificação. Isso é alcançado
 5 através da variação da granularidade da quantização dos valores espectrais ao longo do tempo e da frequência.

Os coeficientes espectrais quantizados e codificados por entropia ou valores de sub-banda são, além dos metadados, inseridos em um formatador de *bitstream* 1606, que provê
 10 um sinal de áudio codificado, que é apropriado para ser transmitido ou armazenado. O *bitstream* de saída do bloco 1606 pode ser transmitido através da Internet ou pode ser armazenado em qualquer suporte de dados legíveis por máquina.

No lado do decodificador, uma interface de
 15 entrada do decodificador 1610 recebe o *bitstream* codificado. O bloco 1610 separa os valores espectrais/de sub-banda quantizados e codificados por entropia dos metadados. Os valores espectrais codificados são inseridos em um decodificador por entropia, tal como um decodificador Huffman, que é posicionado entre o 1610 e o
 20 1620. As saídas do tal decodificador por entropia são os valores espectrais quantizados. Esses valores espectrais quantizados são inseridos em um requantizador, que realiza uma quantização "inversa", como indicado em 1620, na Fig. 16. A saída do bloco 1620 é inserida em um banco de filtro de síntese 1622, que realiza
 25 uma filtragem de síntese, incluindo uma transformação de frequência/tempo e, tipicamente, uma operação de cancelamento de *aliasing* (efeito que faz com que diferentes sinais se tornem indistinguíveis) no domínio do tempo, tais como sobrepor e

adicionar e/ou uma operação de janelamento na parte da síntese, para, finalmente, obter um sinal de áudio de saída.

Tradicionalmente, a codificação de fala eficiente é baseada na Codificação Preditiva Linear (LPC), para modelar os efeitos de ressonância do trato vocal humano, com uma codificação eficiente do sinal de excitação residual. Ambos os parâmetros de LPC e de excitação são transmitidos a partir do codificador para o decodificador. Esse princípio é ilustrado nas Figs. 17a e 17b.

A Fig. 17a indica a parte do codificador de um sistema de codificação/decodificação baseado em codificação preditiva linear. A entrada da fala é inserida em um analisador LPC 1701, que provê, em sua saída, os coeficientes do filtro LPC. Baseado nesses coeficientes do filtro LPC, um filtro LPC 1703 é ajustado. O filtro LPC emite um sinal de áudio branqueado espectralmente, que também é denominado "sinal de erro de previsão". Esse sinal de áudio branqueado espectralmente é inserido em um codificador de excitação/residual 1705, que gera os parâmetros de excitação. Assim, a entrada de fala é codificada em parâmetros de excitação, por um lado, e os coeficientes LPC, por outro lado.

Na parte do decodificador ilustrada na Fig. 17b, os parâmetros de excitação são inseridos em um decodificador de excitação 1707, que gera um sinal de excitação, que pode ser inserido em um filtro de síntese LPC. O filtro de síntese LPC é ajustado com os coeficientes do filtro LPC transmitidos. Assim, o filtro de síntese LPC 1709 gera um sinal de saída de fala reconstruído ou sintetizado.

Ao longo do tempo, muitos métodos foram propostos

com relação a uma representação convincente eficaz e de forma perceptual do sinal (excitação) residual, como a Excitação Multi-Pulso (MPE), a Excitação de Pulso Regular (RPE) e a Previsão Linear Excitada por Código (CELP).

5 A Codificação Preditiva Linear tenta produzir uma estimativa do valor da amostra atual de uma sequência baseada na observação de certo número de valores passados, como uma combinação linear das observações passadas. A fim de reduzir a redundância no sinal de entrada, o filtro LPC do codificador
10 "branqueia" o sinal de entrada em seu envelope espectral, ou seja, é um modelo do inverso do envelope espectral do sinal. Por outro lado, o filtro de síntese LPC do decodificador é um modelo do envelope espectral do sinal. Especificamente, a análise preditiva linear autorregressiva (AR), bem estabelecida, é conhecida por
15 modelar o envelope espectral do sinal, por meio de uma aproximação em todos os polos.

Normalmente, os codificadores de fala de banda estreita (ou seja, os codificadores de fala com uma taxa de amostragem de 8kHz) utilizam um filtro LPC com uma ordem entre 8 e
20 12. Em virtude da natureza do filtro LPC, a resolução de frequência uniforme é efetiva por toda a faixa de frequência. Isso não corresponde a uma escala de frequência perceptual.

A fim de combinar os pontos fortes do LPC/CELP tradicionais baseados na codificação (melhor qualidade para sinais
25 de fala) e a abordagem de codificação de áudio perceptual tradicional baseada no banco de filtro (ideal para música), uma codificação combinada entre essas arquiteturas foi apresentada. No AMR-WB+ (AMR-WB = Adaptive Multi-Rate WideBand) coder B. Bessette,

R. Lefebvre, R. Salami, "UNIVERSAL SPEECH/AUDIO CODING USING HYBRID ACELP/TCX TECHNIQUES," Proc. IEEE ICASSP 2005, pp. 301-304, 2005, dois kernels de codificação alternados operam em um sinal residual LPC. Um é baseado na ACELP (ACELP = Previsão Linear

5 Excitada por Código Algébrico) e, portanto, é extremamente eficiente para a codificação de sinais de fala. O outro kernel de codificação é baseado em TCX (TCX = Excitação Codificada de Transformação), ou seja, uma abordagem de codificação baseada no banco de filtro parecida com as técnicas tradicionais de

10 codificação de áudio, para obter boa qualidade para sinais de música. Dependendo das características dos sinais do sinal de entrada, um dos dois modos de codificação é selecionado por um curto período de tempo para transmitir o sinal residual LPC. Dessa forma, as estruturas de duração de 80ms podem ser divididas em

15 subestruturas de 40ms ou 20ms, nas quais uma decisão entre os dois modos de codificação é realizada.

O AMR-WB+ (AMR-WB+ = extended Adaptive Multi-Rate WideBand codec), cf. 3GPP (3GPP = Third Generation Partnership Project), número de especificação técnica 26.290, versão 6.3.0,

20 junho de 2005, pode alternar entre os dois modos essencialmente diferentes ACELP e TCX. No modo ACELP, um sinal no domínio do tempo é codificado pela excitação do código algébrico. No modo TCX, uma transformada de Fourier rápida (FFT = transformada de Fourier rápida) é utilizada, e os valores espectrais do sinal

25 ponderado LPC (de onde o sinal de excitação é obtido no decodificador) são codificados com base na quantização vetorial.

A decisão, sobre qual dos modos utilizar, pode ser tomada por tentativa e decodificação de ambas as opções e

comparando as relações sinal-ruído resultantes ($SNR = \text{Relação Sinal-Ruído}$).

Esse caso também é chamado de decisão de loop fechado, pois existe um loop de controle fechado, avaliando o desempenho e/ou a eficiência dos códigos, respectivamente, e, em seguida, escolhendo aquele com a melhor SNR e descartando os demais.

Sabe-se bem que, para aplicações de codificação de fala e de áudio, uma transformação em bloco sem janelamento não é viável. Portanto, para o modo TCX, o sinal recebe a função de janelamento com uma janela de sobreposição baixa, com uma sobreposição de $1/8$. Essa região de sobreposição é necessária, a fim de fazer desaparecer gradualmente um bloco ou estrutura anterior, enquanto faz surgir o próximo, por exemplo, para suprimir artefatos devido ao ruído de quantização não correlacionada em estruturas de áudio consecutivas. Dessa forma, o código extra comparado com a amostragem não-crítica é mantido razoavelmente baixo, e a decodificação necessária para a decisão de loop fechado reconstrói, pelo menos, $7/8$ das amostras da estrutura atual.

O AMR-WB+ introduz $1/8$ do código extra em um modo TCX, ou seja, o número de valores espectrais para ser codificado é $1/8$ superior ao número de amostras de entrada. Isso provê a desvantagem de um código extra de dados aumentados. Além disso, a resposta de frequência dos filtros de passagem de banda correspondente é desvantajosa, em virtude da região de sobreposição excessiva de $1/8$ de estruturas consecutivas.

A fim de elaborar mais sobre o código extra e a

sobreposição das estruturas consecutivas, a Fig. 18 ilustra a definição dos parâmetros de janela. A janela mostrada na Fig. 18 tem uma parte da linha em ascensão, do lado esquerdo, que é indicada por "L" e também chamado de região de sobreposição à esquerda, uma região do centro, que é indicada por "1", que também é chamada uma região de 1 parte do desvio e uma parte da linha em declínio, que é indicada por "R" e também chamada de região de sobreposição à direita. Além disso, a Fig. 18 mostra uma seta indicando a região "PR" de reconstrução perfeita em uma estrutura.

Ademais, a Fig. 18 mostra uma seta indicando o comprimento do núcleo de transformação, que é designado por "T".

A Fig. 19 mostra um gráfico de exibição de uma sequência de janelas AMR-WB+ e, na parte inferior, uma tabela de parâmetro de janela de acordo com a Fig. 18. A sequência de janelas mostradas na parte superior da Fig. 19 é ACELP, TCX20 (para uma estrutura de 20 ms de duração), TCX20, TCX40 (para uma estrutura de 40ms de duração), TCX80 (para uma estrutura de 80ms de duração), TCX20, TCX20, ACELP e ACELP.

A partir da sequência de janelas, as regiões de sobreposição variadas pode ser vistas, que se sobrepõem exatamente 1/8 da parte do centro M. A tabela, na parte inferior da Fig. 19, também mostra que o comprimento de transformação "T" é sempre 1/8 maior do que a região de novas amostras perfeitamente reconstruídas "PR". Além disso, pode ser notado que isso não é apenas o caso de transições ACELP para TCX, mas também para as transições TCXx para TCXx (onde "x" indica as estruturas TCX de comprimento arbitrário). Desse modo, em cada bloco, um código extra de 1/8 é introduzido, ou seja, a amostragem crítica nunca é

atingida.

Ao mudar de TCX para ACELP, as amostras da janela são eliminadas da estrutura FFT-TCX, na região de sobreposição, como, por exemplo, indicado na parte superior da Fig. 19 pela região rotulada com 1900. Ao mudar de ACELP para TCX, a resposta de entrada zero, que recebeu a função de janelamento (ZIR = resposta de entrada zero), que também é indicada pela linha tracejada 1910, na parte superior da Fig. 19, é removida no codificador para janelamento e adição ao decodificador para recuperação. Ao mudar de estruturas TCX para TCX, as amostras que receberam a função de janelamento são usadas para transição. Como as estruturas TCX podem ser quantizadas diferentemente, o erro de quantização ou o ruído de quantização entre estruturas consecutivas pode ser diferente e/ou independente. Com isso, ao mudar de uma estrutura para outra, sem transição, artefatos visíveis podem ocorrer e, portanto, a transição é necessária para atingir certa qualidade.

A partir da tabela, na parte inferior da Fig. 19, pode ser visto que a região de transição cresce com um comprimento crescente da estrutura. A Fig. 20 provê outra tabela com ilustrações das janelas diferentes, para as transições possíveis em AMR-WB+. Quando houver a transição de TCX para ACELP, as amostras de sobreposição podem ser descartadas. Quando houver a transição de ACELP para TCX, a resposta de entrada zero do ACELP é removida no codificador e adicionada no decodificador, para recuperação.

É uma desvantagem significativa do AMR-WB+ que uma sobrecarga código extra de 1/8 seja sempre introduzido.

O objetivo da presente invenção é prover um conceito mais eficiente para a codificação de áudio.

O objetivo é alcançado através de um codificador de áudio, de acordo com a reivindicação 1, um método de
5 codificação de áudio, de acordo com a reivindicação 14, um decodificador de áudio, de acordo com a reivindicação 16, e um método de decodificação de áudio, de acordo com a reivindicação 25.

As configurações da presente invenção são
10 baseadas na constatação de que uma codificação mais eficiente pode ser realizada, se as transformações de introdução de *aliasing* de tempo forem usadas, por exemplo, para a codificação TCX. As transformações de introdução de *aliasing* de tempo podem permitir a
realização de amostragem crítica, enquanto continuam sendo capazes
15 de realizar transição entre estruturas adjacentes. Por exemplo, em uma configuração, a transformação de cosseno discreto modificado (MDCT = Transformação de Cosseno Discreto Modificado) é usada para transformar estruturas de domínio do tempo de sobreposição para o domínio da frequência. Como essa transformação particular produz
20 apenas amostras de domínio de frequência N para amostras no domínio do tempo 2N, a amostragem crítica pode ser mantida mesmo que as estruturas no domínio do tempo possam sobrepor-se em 50%. No decodificador ou no transformador de introdução de *aliasing* de tempo inverso, uma fase de sobreposição e de adição pode ser
25 adaptada para combinar as amostras de domínio de tempo transformadas de volta e a sobreposição de *aliasing* de tempo, de forma que o cancelamento de *aliasing* no domínio do tempo (TDAC = Cancelamento de *Aliasing* no Domínio do Tempo) possa ser executado.

As configurações podem ser utilizadas no contexto de um domínio da frequência mudado e a codificação no domínio do tempo, com janelas de sobreposição baixa, como, por exemplo, o AMR-WB+. As configurações podem usar uma MDCT, em vez de banco de
 5 filtro amostrado não-criticamente. Dessa forma, o código extra, devido à amostragem não-crítica, pode ser vantajosamente reduzido, com base, por exemplo, na propriedade de MDCT de amostragem crítica. Além disso, sobreposições mais extensas são possíveis sem a introdução de código extra. As configurações podem oferecer a
 10 vantagem de que, com base nas sobreposições mais extensas, a transição pode ser realizada de forma mais suave; em outras palavras, a qualidade do som pode ser aumentada no decodificador.

Em uma configuração detalhada da FFT, no AMR-WB+, o modo TCX pode ser substituído por uma MDCT, enquanto mantém as
 15 funcionalidades no AMR-WB+, especialmente a troca entre o modo ACELP e o modo TCX, com base em uma decisão de loop aberto ou fechado. As configurações podem usar a MDCT de forma não-criticamente amostrada para a primeira estrutura TCX após uma estrutura ACELP e, posteriormente, usar a MDCT de forma
 20 criticamente amostrada para todas as estruturas TCX subsequentes. As configurações podem reter o recurso da decisão de loop fechado, utilizando a MDCT com janelas de sobreposição baixa semelhantes ao AMR-WB+ inalterado, mas com sobreposições mais extensas. Isso pode prover a vantagem de uma resposta de frequência melhor em
 25 comparação com as janelas TCX inalteradas.

As configurações da presente invenção serão detalhadas com os números associados, na qual:

A Fig. 1 mostra uma configuração de um

codificador de áudio;

As Figs. 2a-2j mostram equações para uma configuração de um transformador de introdução de *aliasing* no domínio do tempo;

5 A Fig. 3a mostra outra configuração de um codificador de áudio,

A Fig. 3b mostra outra configuração de um codificador de áudio,

10 A Fig. 3c mostra ainda outra configuração de um codificador de áudio,

A Fig. 3d mostra ainda outra configuração de um codificador de áudio;

A Fig. 4a mostra um exemplo do "sinal" de "fala" no domínio do tempo para fala articulada;

15 A Fig. 4b ilustra um espectro de uma amostra do "sinal" de fala articulada;

A Fig. 5a ilustra um sinal no domínio do tempo de uma amostra de uma fala não articulada;

20 A Fig. 5b mostra um espectro de uma amostra de um sinal de fala não articulada;

A Fig. 6 mostra uma configuração de uma análise por síntese CELP;

25 A Fig. 7 ilustra uma fase ACELP da parte do codificador provendo informações de previsão de curto prazo e um sinal de erro de previsão;.

A Fig. 8a mostra uma configuração de um decodificador de áudio;

A Fig. 8b mostra outra configuração de um

decodificador de áudio,

A Fig. 8c mostra outra configuração de um decodificador de áudio,

5 A Fig. 9 mostra uma configuração de uma função de janela;

A Fig. 10 mostra outra configuração de uma função de janela;

A Fig. 11 mostra gráficos de exibição e gráficos de retardo de funções de janela da técnica anterior e uma função
10 de janela de uma configuração;

A Fig. 12 ilustra os parâmetros da janela;

A Fig. 13a mostra uma sequência de funções de janela de acordo com a tabela de parâmetros da janela;

A Fig. 13b mostra transições possíveis para uma
15 configuração baseada em MDCT.

A Fig. 14a mostra uma tabela de transições possíveis em uma configuração;

A Fig. 14b ilustra uma janela de transição de ACELP para TCX80, de acordo com uma configuração;

20 A Fig. 14c mostra a configuração de uma janela de transição de uma estrutura TCXx para uma estrutura TCX20, para uma estrutura TCXx, de acordo com uma configuração;

A Fig. 14d ilustra a configuração de uma janela de transição de ACELP para TCX20, de acordo com uma configuração;

25 A Fig. 14e mostra a configuração de uma janela de transição de ACELP para TCX40, de acordo com uma configuração;

A Fig. 14f ilustra uma configuração da janela de transição para uma transição de uma estrutura TCXx para uma

estrutura TCX80, para uma estrutura TCXx, de acordo com uma configuração;

A Fig. 15 ilustra uma transição de ACELP para TCX80, de acordo com uma configuração;

5 A Fig. 16 ilustra exemplos de codificador e de decodificador convencional;

As Figs. 17a, b ilustram codificação e decodificação LPC;

10 A Fig. 18 ilustra uma janela de transição da técnica anterior;

A Fig. 19 ilustra uma sequência da técnica anterior de janelas AMR-WB+;

A Fig. 20 ilustra janelas utilizadas para a transmissão em AMR-WB+ entre ACELP e TCX.

15 A seguir, as configurações da presente invenção serão descritas em detalhes. Observe que as configurações seguintes não deverão limitar o escopo da invenção; elas deverão ser consideradas realizações ou implementações possíveis entre muitas configurações diferentes.

20 A Fig. 1 mostra um codificador de áudio 10 adaptado para estruturas de codificação de um sinal de áudio testado, para obter estruturas codificadas, caracterizado pelo fato de que uma estrutura composta por um número de amostras de áudio no domínio do tempo, o codificador de áudio 10 compreende
25 uma fase de análise de codificação preditiva 12 para determinar as informações sobre os coeficientes para um filtro de síntese e uma estrutura de domínio de previsão baseada em estruturas de amostras de áudio, por exemplo, a estrutura de domínio de previsão pode ser

baseada em uma estrutura de excitação, a estrutura de domínio de previsão pode incluir amostras ou amostras ponderadas de um sinal de domínio LPC, a partir do qual o sinal de excitação, para o filtro de síntese, pode ser obtido. Em outras palavras, em 5 configurações, uma estrutura de domínio de previsão pode ser baseada em uma estrutura de excitação, incluindo amostras de um sinal de excitação para o filtro de síntese. Em configurações, as estruturas de domínio de previsão podem corresponder a versões filtradas das estruturas de excitação. Por exemplo, a filtragem 10 perceptual pode ser aplicada a uma estrutura de excitação, para obter a estrutura de domínio de previsão. Em outras configurações, a filtragem *high-pass* ou *low-pass* pode ser aplicada nas estruturas de excitação, para obter as estruturas de domínio de previsão. Ainda em outra configuração, as estruturas de domínio de previsão 15 podem corresponder diretamente às estruturas de excitação.

O codificador de áudio 10 compreende ainda um transformador de introdução de *aliasing* de tempo 14 para transformar as estruturas de domínio de previsão de sobreposição com relação ao domínio da frequência, para obter os espectros da 20 estrutura de domínio de previsão, caracterizado pelo fato de que o transformador de introdução de *aliasing* de tempo 14 é adaptado para transformar as estruturas de domínio de previsão de sobreposição, de forma testada criticamente. O codificador de áudio 10 compreende ainda um codificador de redução de redundância 25 16 para a codificação de espectros da estrutura de domínio de previsão, a fim de obter as estruturas codificadas com base nos coeficientes e nos espectros da estrutura de domínio de previsão codificados.

O codificador de redução de redundância 16 pode ser adaptado para o uso de codificação Huffman ou codificação por entropia, para codificar os espectros da estrutura de domínio de previsão e/ou informações sobre os coeficientes.

5 Em configurações, o transformador de introdução de *aliasing* de tempo 14 pode ser adaptado para transformar as estruturas de domínio de previsão de sobreposição de tal forma que um número médio de amostras de um espectro de estrutura de domínio de previsão seja igual ao número médio de amostras em uma

10 estrutura de domínio de previsão, conseguindo, assim, a transformação amostrada criticamente. Além disso, o transformador de introdução de *aliasing* de tempo 14 pode ser adaptado para transformar as estruturas de domínio de previsão de sobreposição, de acordo com uma transformação de cosseno discreto modificado

15 (MDCT = Transformação de Cosseno Discreto Modificado).

A seguir, a MDCT será explicada detalhadamente, com a ajuda das equações ilustradas nas Figs. 2a-2j. A transformação de cosseno discreto modificado (MDCT) é uma transformação relacionada com Fourier baseada na transformação de

20 cosseno discreto tipo IV (DCT-IV = transformação de Cosseno Discreto tipo IV), com a propriedade adicional de ser dobrada, ou seja, ele é projetado para ser executado em blocos consecutivos de um dataset maior, onde blocos subsequentes são sobrepostos, para que, assim, a última metade de um bloco coincida com a primeira

25 metade do bloco seguinte. Essa sobreposição, além das qualidades de compactação de energia da DCT, torna a MDCT especialmente atraente para aplicações de compressão de sinal, uma vez que ajuda a evitar os artefatos decorrentes dos limites do bloco. Assim, uma

MDCT é empregada em MP3 (MP3 = MPEG2/4 layer 3), AC-3 (AC-3 = Audio Codec 3 by Dolby), Ogg Vorbis e AAC (AAC = Codificação de Áudio Avançada) para compressão de áudio, por exemplo.

A MDCT foi proposta por Princen, Johnson, e Bradley, em 1987, na sequência do primeiro trabalho (1986) realizado por Princen e Bradley, para desenvolver princípio subjacente da MDCT de cancelamento de *aliasing* no domínio do tempo (TDAC), descrita abaixo.

Existe também uma transformação análoga, a MDST, baseada na transformação de seno discreto, assim como outro, raramente utilizado, formas de MDCT com base em diferentes tipos de combinações de DCT ou DCT/DST (DST = Transformação de Seno Discreto), que podem ser utilizadas em configurações através do transformador de introdução de *aliasing* de tempo 14.

Em MP3, a MDCT não é aplicada ao sinal de áudio diretamente, mas sim à saída de um banco do filtro de quadratura polifásico de 32 bandas (PQF = Filtro de Quadratura Polifásico). A saída dessa MDCT é pós-processada por uma fórmula de redução de sinal falso, para reduzir o *aliasing* típico do banco do filtro PQF. Essa combinação de um banco de filtro com uma MDCT é chamada de banco de filtro híbrido ou sub-banda MDCT. A AAC, por outro lado, usa normalmente uma MDCT pura; somente a (raramente usada) variante MPEG-4 AAC-SSR (da Sony) usa um banco PQF de quadribanda, seguida por uma MDCT. A ATRAC (ATRAC = Codificação de Áudio de Transformador Adaptativo) usa filtro espelho de quadratura em grande quantidade (QMF), seguida por uma MDCT.

Como uma transformação dobrada, a MDCT é um pouco incomum em comparação com outras transformações relacionadas com

Fourier, na medida em que possui metade tanto das saídas quanto das entradas (em vez de possuir o mesmo número). Em particular, é uma função linear $F: \mathbf{R}^{2N} \rightarrow \mathbf{R}^N$, onde $\mathbf{R}$ representa o conjunto dos números reais. Os números reais $2N$ $x_0, \dots, x_{2N-1}$ são transformados em

5 números reais N $x_0, \dots, x_{N-1}$ de acordo com a fórmula na Fig. 2a.

O coeficiente de normalização na frente dessa transformação, aqui a unidade, é uma convenção arbitrária e difere entre os tratamentos. Apenas o produto das normalizações da MDCT e da IMDCT, abaixo, é limitado.

10 A MDCT inversa é conhecida como a IMDCT. Como existem diferentes números de entradas e saídas, à primeira vista pode parecer que a MDCT não deve ser invertível. No entanto, invertibilidade perfeita é alcançada através da adição das IMDCTs sobrepostas de blocos sobrepostos subsequentes, causando erros

15 para cancelar e os dados originais a serem recuperados; essa técnica é conhecida como cancelamento de *aliasing* no domínio do tempo (TDAC).

A IMDCT transforma números reais N $x_0, \dots, x_{N-1}$ em números reais $2N$ $y_0, \dots, y_{2N-1}$ de acordo com a fórmula na Fig. 2b.

20 Semelhante para DCT-IV, uma transformação ortogonal, o inverso tem a mesma forma que a transformação anterior.

No caso de uma MDCT em janela, com a normalização de janela usual (veja abaixo), o coeficiente de normalização em frente à IMDCT deve ser multiplicado por 2, ou seja, tornando-se

25 $2/N$.

Embora a aplicação direta da fórmula MDCT exigisse operações $O(N^2)$, é possível calcular a mesma coisa com apenas a complexidade $O(N \log N)$, por fatorar recursivamente a

computação, como na transformada de Fourier rápida (FFT). Também pode-se calcular MDCTs através de outras transformações, normalmente uma DFT (FFT) ou uma DCT, combinada com fases pré e pós-processamento $O(N)$. Ainda, como descrito abaixo, qualquer

5 algoritmo para a DCT-IV imediatamente provê um método para calcular a MDCT e a IMDCT do mesmo tamanho.

Em aplicações típicas de compressão de sinais, as propriedades de transformação são melhoradas através de uma função de janela w_n ($n = 0, \dots, 2N-1$), que é multiplicada por x_n e y_n nas

10 fórmulas MDCT e IMDCT, acima, a fim de evitar descontinuidades em $n = 0$ e limites $2N$, fazendo com que a função desça vagarosamente a zero, em tais pontos. Ou seja, os dados recebem a função de janelamento antes da MDCT e após a IMDCT. Em princípio, x e y podem ter funções diferentes de janela, e a função de janela

15 também poderia mudar de um bloco para outro, especialmente para o caso de blocos de dados de tamanhos diferentes serem combinados, mas, pela simplicidade, o caso comum de funções idênticas para janela, para blocos de tamanho semelhante é considerado como primeiro.

20 A transformação permanece invertível, ou seja, TDAC opera para uma janela simétrica $w_n = w_{2N-1-n}$, contanto que w satisfaça a condição Princen-Bradley, de acordo com a Fig. 2c.

Diversas funções de janela diferentes são comuns, um exemplo é dado na Fig. 2d para MP3 e MPEG-2 AAC, e na Fig. 2e

25 para Vorbis. A AC-3 utiliza uma janela derivada Kaiser-Bessel (KBD = derivada Kaiser-Bessel), e MPEG-4 AAC também pode usar uma janela KBD.

Observe que as janelas aplicadas à MDCT são

diferentes das janelas utilizadas para outros tipos de análise de sinais, já que eles devem cumprir a condição Princen-Bradley. Uma das razões para essa diferença é que as janelas MDCT são aplicadas duas vezes, tanto para a MDCT (filtro de análise) e a IMDCT (filtro de síntese).

Como pode ser visto pela inspeção das definições, até mesmo para N , a MDCT é essencialmente equivalente a uma DCT-IV, onde a entrada é desviada por $N/2$ e dois blocos de dados N são transformados imediatamente. Ao examinar com mais cuidado essa equivalência, propriedades importantes como TDAC podem ser facilmente derivadas.

A fim de definir a relação exata da DCT-IV, é preciso perceber que a DCT-IV corresponde a condições de limite par/ímpar alternadas; é par em seu limite à esquerda (em torno de $n = -1/2$), ímpar em seu limite à direita (em torno de $n = N - 1/2$) e assim por diante (ao invés de limites periódicos quanto a DFT). Isso decorre das identidades dadas na Fig. 2f. Assim, se suas entradas são uma array x de comprimento N , imagine estender essa array para $(x, -x_R, -x, x_R, \dots)$, e assim por diante, pode ser imaginado, onde x_R significa x na ordem inversa.

Considere uma MDCT com entradas $2N$ e saídas N , onde as entradas podem ser divididas em quatro blocos (a, b, c, d), cada uma de tamanho $N/2$. Se estes podem ser mudados por $N/2$ (a partir do termo $+N/2$, na definição da MDCT), então (b, c, d) passado ampliado, o fim das entradas N DCT-IV, pois assim elas devem ser "dobradas" de volta, de acordo com as condições de limite acima mencionadas.

Assim, a MDCT de entradas $2N$ (a, b, c, d) é

exatamente equivalente a uma DCT-IV das entradas N : $(c-d, a-bR)$, onde R significa a reversão, como acima. Dessa forma, qualquer algoritmo para calcular a DCT-IV pode ser trivialmente aplicado à MDCT.

5 Da mesma forma, a fórmula IMDCT, como mencionada acima, é precisamente a metade da DCT-IV (que é sua própria inversa), onde a saída é mudada por $N/2$ e ampliada (via condições de limite) para um comprimento $2N$. O inverso DCT-IV devolve simplesmente as entradas $(-c_R-d, a-b_R)$ a partir do anterior.

10 Quando este é mudado e ampliado através de condições de limite, obtém-se o resultado apresentado na Fig. 2g. Metade das saídas IMDCT é, portanto, redundante.

Podemos agora compreender como a TDAC opera. Suponha que haja um cálculo da MDCT do posterior, 50% sobreposto, 15 bloco $2N$ (c, d, e, f) . A IMDCT, então, produzirá análoga ao anterior: $(c-d_R, d-c_R, e+f_R, e_R+f) / 2$. Quando isso é adicionado com o resultado da IMDCT anterior, na metade sobreposta, os termos revertidos cancelam e um obtém simplesmente (c, d) , recuperando os dados originais.

20 A origem do termo "cancelamento de *aliasing* no domínio do tempo" está agora elucidada. O uso de dados de entrada, que se estendem além dos limites da lógica DCT-IV, faz com que os dados a serem *aliased* exatamente da mesma maneira que as frequências além da frequência de Nyquist são *aliased* para baixas 25 frequências, exceto que esse *aliasing* ocorra no domínio do tempo, em vez do domínio da frequência. Por isso, as combinações $c-d_R$ e assim por diante, que têm precisamente os sinais certos para as combinações de cancelar quando forem adicionados.

Para N ímpar (que raramente são usados na prática), $N/2$ não é um número inteiro, portanto para a MDCT não é simplesmente uma permutação de mudança de uma DCT-IV. Nesse caso, a mudança adicional pela metade de uma amostra significa que a MDCT/IMDCT se torna equivalente à DCT-III/II, e a análise é análoga ao descrito acima.

Acima, a propriedade TDAC foi demonstrada para a MDCT ordinária, mostrando que a adição de IMDCTs de blocos, em sua metade sobreposta, recupera os dados originais. A derivação dessa propriedade inversa, para a MDCT em janela, é apenas um pouco mais complicada.

Lembre-se do exposto que, quando (a, b, c, d) e (c, d, e, f) recebem MDCT, e IMDCT e são adicionados em sua metade sobreposta, obtemos os dados originais $(c + d_R, c_R + d) / 2 + (c - d_R, d - c_R) / 2 = (c, d)$.

Agora, considere multiplicar ambas as entradas MDCT e saídas IMDCT por uma função de janela de comprimento $2N$. Como dito acima, assumimos uma função de janela simétrica, que é, portanto, da forma (w, z, z_R, w_R) , onde w e z são vetores de comprimento $N/2$ e R significa reversão, como antes. Então, a condição Princen-Bradley pode ser escrita

$$w^2 + z_R^2 = (1, 1, \dots),$$

com as multiplicações e adições realizadas ao quadrado ("elementwise"), ou de forma equivalente

$$w_R^2 + z^2 = (1, 1, \dots)$$

revertendo w e z .

Portanto, em vez de realizar uma MDCT (a, b, c, d) , a MDCT $(wa, zb, z_R c, w_R d)$ recebe uma MDCT com todas as

multiplicações realizadas elementwise. Quando isso é recebido IMDCT e é multiplicado novamente (elementwise) pela função de janela, a última metade N resulta como mostrado na Fig. 2h.

Observe que a multiplicação por $1/2$ não está mais presente, porque a normalização IMDCT difere por um fator de 2, no caso de janela. Da mesma forma, a MDCT em janelas e a IMDCT de (c, d, e, f) produções, em sua primeira metade N, de acordo com a Fig. 2i. Quando essas duas metades são somadas, os resultados da Fig. 2j são obtidos, recuperando os dados originais.

A Fig. 3a mostra outra configuração do codificador de áudio 10. Na configuração apresentada na Fig. 3a o transformador de introdução de *aliasing* de tempo 14 consiste em um filtro de janelamento 17 para aplicação de uma função de janelamento, para a sobreposição de estruturas de domínio de previsão e um conversor 18 para converter as estruturas de domínio de previsão sobrepostas em janela para os espectros de domínio de previsão. De acordo com o exposto acima, as funções de janela múltiplas são possíveis; algumas das quais serão detalhadas mais adiante.

Outra configuração de um codificador de áudio 10 é mostrada na Fig. 3b. Na configuração apresentada na Fig. 3b o transformador de introdução de *aliasing* de tempo 14 consiste em um processador 19 para a detecção de um evento e para prover uma informação de sequência da janela, se o evento for detectado e caracterizado pelo fato de que o filtro de janelamento 17 está adaptado para a aplicação da função de janelamento, de acordo com as informações de sequência da janela. Por exemplo, o evento pode ocorrer de acordo com certas propriedades de sinal analisadas a

partir das estruturas do sinal de áudio testado. Por exemplo, o comprimento de janela diferente ou de linhas da janela, etc., pode ser aplicado de acordo com o exemplo das propriedades de autocorrelação do sinal, tonalidade transitoriedade, etc. Em 5 outras palavras, diferentes eventos podem ocorrer como parte das propriedades diferentes das estruturas de áudio testado, e o processador 19 pode prover uma sequência de janelas distintas, na dependência de propriedades de estruturas de sinal de áudio. As sequências mais detalhadas e os parâmetros para sequências de 10 janela serão definidos abaixo.

A Fig. 3c mostra outra configuração de um codificador de áudio 10. Na configuração apresentada na Fig. 3d, as estruturas de domínio de previsão não são providas apenas para o transformador de introdução de *aliasing* de tempo 14, mas também 15 para um codificador de *codebook* 13, que é adaptado para codificar as estruturas de domínio de previsão baseadas em um *codebook* predeterminado, para obter uma estrutura codificada do *codebook*. Além disso, a configuração descrita na Fig. 3c consiste em um decisor para decidir se deve usar uma estrutura codificada de 20 *codebook* ou uma estrutura codificada, para obter uma estrutura finalmente codificada, com base na medição de eficiência codificada. A configuração apresentada na Fig. 3c também pode ser chamada de um cenário de loop fechado. Nesse cenário, o decisor 15 tem a possibilidade de obter estruturas codificadas a partir de 25 dois ramais: um ramal com base em transformação e o outro com base em um *codebook*. A fim de determinar uma medida de eficiência de codificação, o decisor pode decodificar as estruturas codificadas de ambos os ramais e, em seguida, determinar a medida de

eficiência de codificação, através da avaliação estatística de erro de diferentes ramais.

Em outras palavras, o decisor 15 pode ser adaptado para reverter o procedimento de codificação, ou seja, realizar a decodificação completa de ambos os ramais. Tendo estruturas totalmente decodificadas, o decisor 15 pode ser adaptado para comparar as amostras decodificadas com as amostras originais, o que é indicado pela seta tracejada na Fig. 3c. Na configuração mostrada na Fig. 3c, o decisor 15 é provido também com as estruturas de domínio de previsão; com isso, ele está habilitado a decodificar estruturas codificadas a partir do codificador de redução da redundância 16 e também decodificar as estruturas do *codebook* codificado 13 e comparar os resultados com as estruturas de domínio de previsão originalmente codificadas.

Com isso, em uma configuração, comparando as diferenças, as medidas de eficiência de codificação, por exemplo, em termos de uma relação sinal-ruído ou de um erro estatístico ou de um erro mínimo, etc., podem ser definidas, em algumas configurações, também, em relação à taxa do código respectivo, ou seja, o número de bits exigido para codificar as estruturas. O decisor 15 pode, então, ser adaptado para selecionar estruturas codificadas, a partir do codificadora de redução da redundância 16 ou das estruturas codificadas do *codebook*, como estruturas finalmente codificadas, com base na medida de eficiência de codificação.

A Fig. 3d mostra outra configuração do codificador de áudio 10. Na configuração mostrada na Fig. 3d há uma chave 20 acoplada ao decisor 15, para mudar as estruturas de domínio de previsão entre o transformador de introdução de

aliasing de tempo 14 e o codificador do *codebook* 13, com base em uma medida de eficiência de codificação. O decisor 15 pode ser adaptado para determinar uma medida de eficiência de codificação, com base em estruturas do sinal de áudio testado, a fim de

5 determinar a posição da chave 20, ou seja, se usar o ramal de codificação baseada em transformação com o transformador de introdução de *aliasing* de tempo 14 e o codificador de redução de redundância 16 ou o ramal de codificação baseado no *codebook* com o codificador de *codebook* 13. Como mencionado anteriormente, a

10 medida de eficiência de codificação pode ser determinada com base nas propriedades das estruturas de sinal de áudio testado, ou seja, as propriedades de áudio propriamente ditas, por exemplo, se a estrutura tem perfil de tom ou de ruído.

O aspecto da configuração mostrado na Fig. 3D

15 também é chamado de configuração de loop aberto, pois o decisor 15 pode decidir, com base nas estruturas de entrada, sem conhecer os resultados do respectivo ramal de codificação. Em ainda outra configuração, o decisor pode decidir com base em estruturas de domínio de previsão, o que é mostrado na Fig. 3d pela seta

20 tracejada. Em outras palavras, em uma configuração, o decisor 15 não pode decidir com base nas estruturas do sinal de áudio testado, mas, sim, sobre as estruturas de domínio de previsão.

Em seguida, o processo de decisão do decisor 15 é elucidado. Geralmente, uma diferenciação entre uma porção de

25 impulso de um sinal de áudio e uma porção estacionária de um sinal estacionário pode ser feita através da aplicação de uma operação de processamento de sinal, onde a característica de impulso e a característica estacionária são medidas. Tais medidas podem, por

exemplo, ser feitas através da análise da forma de onda do sinal de áudio. Para esse efeito, qualquer processamento com base em transformação ou processamento LPC ou qualquer outro processamento pode ser realizado. Uma forma intuitiva para determinar se a

5 porção é de impulso ou não é, por exemplo, olhar para uma forma de onda no domínio do tempo e determinar se esta forma de onda no domínio do tempo tem picos em intervalos regulares ou irregulares, e os picos em intervalos regulares são ainda mais apropriados para o codificador de fala, ou seja, para o codificador do *codebook*.

10 Note que até mesmo as partes de fala articulada ou não articulada podem ser diferenciadas. O codificador do *codebook* 13 pode ser mais eficiente para as partes com sinais de fala ou estruturas de fala, caracterizado pelo fato de que o ramal com base na transformação consiste em um transformador de introdução de

15 *aliasing* de tempo 14 e o codificador de redução de redundância 16 pode ser mais adequado para as estruturas não articuladas. Geralmente, a codificação baseada na transformação também pode ser mais apropriada para sinais estacionários do que os sinais de fala.

20 Como exemplo, é feita referência às Figs. 4a e 4b, 5a e 5b, respectivamente. Os segmentos de sinal de impulso ou as porções de sinal e os segmentos de sinal estacionários ou porções de sinal são exemplarmente discutidos. Geralmente, o decisor 15 pode ser adaptado para decidir com base em diferentes

25 critérios, como, por exemplo, estacionário, transitoriedade, branqueamento de espectros, etc. A seguir, critérios de exemplo são apresentados como parte de uma configuração. Especificamente, uma fala articulada é apresentada na Fig. 4a, no domínio do tempo,

e na Fig. 4b, no domínio da frequência, e é discutida como exemplo de uma porção do sinal de impulso e um segmento de fala não articulada como um exemplo de uma porção do sinal estacionário é discutido com relação às Figs. 5a e 5b.

5 A fala pode ser geralmente classificada como articulada, não articulada ou mista. Os desenhos de domínio de tempo e frequência, para segmentos amostrados articulados e não articulados, são mostrados nas Figs. 4a, 4b, 5a e 5b. A fala articulada é quase periódica no domínio do tempo e harmonicamente
10 estruturada no domínio da frequência, enquanto que a fala não articulada é aleatória e de banda larga. Além disso, a energia dos segmentos articulados geralmente é maior que a energia dos segmentos não articulados. O espectro de curto prazo de fala articulada é caracterizado por sua estrutura fina e formante. A
15 estrutura fina e harmônica é uma consequência da quase periodicidade da fala e pode ser atribuída à vibração das cordas vocais. A estrutura formante, que também é chamada de envelope espectral, ocorre em virtude da interação da fonte e do trato vocal. O trato vocal consiste na faringe e na cavidade bucal. O
20 formato do envelope espectral que "encaixa" o espectro de curto prazo de fala articulada está associado com as características de transferência do trato vocal e com a inclinação espectral (6dB/oitava), devido ao impulso glótico.

 O envelope espectral é caracterizado por um
25 conjunto de picos, que é chamado de formantes. Os formantes são os modos de ressonância do trato vocal. Para o trato vocal médio existem de três a cinco formantes abaixo de 5 kHz. As amplitudes e as posições dos três primeiros formantes, ocorrendo geralmente

abaixo de 3 kHz, são muito importantes, em síntese e percepção de fala. Os formantes mais elevados são também importantes para representações de fala de banda larga e de não articulada. As propriedades da fala estão relacionadas com os sistemas de

5 produção física da fala como se segue. Estimular o trato vocal, com pulsos de ar glóticos quase periódicos gerados pela vibração das cordas vocais, produz a fala articulada. A frequência do pulso periódico é conhecida como a frequência ou altura fundamental. Forçar o ar através de uma constrição no trato vocal produz fala

10 não articulada. Os sons nasais ocorrem devido ao acoplamento acústico do trato nasal ao trato vocal, e os sons plosivos são reduzidos pela redução drástica da pressão do ar, que foi construída por detrás do fechamento no trato.

Assim, uma porção estacionária do sinal de áudio

15 pode ser uma porção estacionária no domínio do tempo, conforme ilustrado na Fig. 5a ou uma porção estacionária no domínio da frequência, que é diferente da porção de impulso, como, por exemplo, ilustrado na Fig. 4a, devido ao fato de que a porção estacionária no domínio do tempo não mostra pulsos de repetição

20 permanentes. Como será descrito mais tarde, porém, a diferenciação entre porções estacionárias e porções de impulso, também pode ser realizada utilizando métodos LPC, que modela o trato vocal e a estimulação do trato vocal. Quando o domínio de frequência do sinal é considerado, os sinais de impulso mostram a aparência

25 proeminente dos formantes individuais, ou seja, picos proeminentes na Fig. 4b, enquanto o espectro estacionário tem um espectro muito amplo, como ilustrado na Fig. 5b, ou, no caso de sinais de harmônicos, um bom patamar de ruído contínuo que possui alguns

picos proeminentes representando tons específicos que ocorrem, por exemplo, em um sinal de música, mas que não têm tal distância regular um do outro como o sinal de impulso, na Fig. 4b.

Além disso, as porções de impulso e as porções
5 estacionárias podem ocorrer de uma maneira oportuna, ou seja, o que significa que uma porção do sinal de áudio, em tempo, é estacionária e outra parte do sinal de áudio, em tempo, é de impulso. Alternativa ou cumulativamente, as características de um sinal podem ser diferentes em diferentes faixas de frequência.
10 Assim, a determinação, se o sinal de áudio é estacionário ou de impulso, também pode ser realizada de frequência-seletiva, para que uma determinada banda de frequência ou várias bandas de frequência sejam consideradas estacionárias e outras bandas de frequências sejam consideradas como de impulso. Nesse caso, uma
15 porção de determinado tempo do sinal de áudio pode incluir uma porção de impulso ou uma porção estacionária.

Voltando para a configuração mostrada na Fig. 3d, o decisor 15 pode analisar as estruturas de áudio, as estruturas de domínio de previsão ou o sinal de excitação, a fim de
20 determinar se eles são certamente de impulso, ou seja, mais adequados para o codificador de *codebook* 13, ou estacionário, ou seja, mais adequados para o ramal de codificação baseado na transformação.

Posteriormente, um codificador CELP de análise
25 por síntese será discutido na Fig. 6. Os detalhes de um codificador CELP também podem ser encontrados em "Speech Coding: A tutorial review", Andreas Spaniers, Proceedings of IEEE, Vol. 84, Nº 10, outubro de 1994, pp 1541-1582. O codificador CELP conforme

apresentado na Fig. 6 inclui um componente de previsão a longo prazo 60 e um componente de previsão a curto prazo 62. Além disso, um *codebook* é utilizado, que é indicado no 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 áudio de entrada. Depois de ter sido perceptualmente ponderado, o sinal de ponderação é inserido em um subtrator 69, que calcula o erro entre o sinal de síntese ponderado (saída do bloco 66) e o sinal de previsão de erro ponderado real $sw(n)$.

10 Geralmente, a previsão a curto prazo $A(z)$ é calculada por uma fase de análise LPC, que será discutida mais adiante. Dependendo dessa informação, a previsão a longo prazo $A(z)$ inclui o retardo de banda de ganho T da previsão a longo prazo (também conhecido como o ganho de altura e retardo de altura). O algoritmo CELP codifica as estruturas de domínio de previsão ou de excitação, usando um *codebook* de sequências de Gaussian, por exemplo. O algoritmo ACELP, onde o "A" significa "algébrico", tem um *codebook* específico projetado algebricamente. O *codebook* pode conter mais ou menos vetores, onde cada vetor tem um comprimento de acordo com a quantidade de amostras. Um fator de ganho g escala o vetor de excitação, e as amostras de excitação são filtradas pelo filtro de síntese de curto prazo e pelo filtro de síntese de longo prazo. O vetor "ideal" é selecionado de modo que o erro de quadrado médio ponderado perceptualmente seja minimizado. O processo de pesquisa no CELP é evidente a partir do esquema de análise por síntese ilustrado na Fig. 6. Note que a Fig. 6 apenas ilustra um exemplo de um CELP de análise por síntese e que as configurações não devem ser limitadas à estrutura

mostrada na Fig. 6.

Na CELP, o predictor de longo prazo é frequentemente implementado como um *codebook* adaptável, contendo o sinal de excitação anterior. O retardo e o ganho de previsão a longo prazo são representados por um índice do *codebook* adaptativo e pelo ganho, que também são selecionados pela minimização do erro de quadrado médio ponderado. Nesse caso, o sinal de excitação consiste na adição de dois vetores de ganho de escala, um de um *codebook* adaptativo e outro de um *codebook* fixo. O filtro de ponderação perceptual, em AMR-WB+, é baseado no filtro LPC, assim, o sinal ponderado perceptualmente é uma forma de um sinal de domínio LPC. No codificador de domínio de transformação, em AMR-WB+, a transformação é aplicada ao sinal ponderado. No decodificador, o sinal de excitação é obtido através da filtragem do sinal decodificado ponderado, através de um filtro que consiste no inverso de filtros de síntese e de ponderação.

Um alvo TCX reconstruído $x(n)$ pode ser filtrado através de um filtro de síntese ponderado inverso de estado zero

$$\frac{\hat{A}(z)(1 - \alpha z^{-1})}{(\hat{A}(z/\lambda))}$$

para encontrar o sinal de excitação que pode ser aplicado ao filtro de síntese. Observe que o filtro LP interpolado, por subestrutura ou estrutura, é usado na filtragem. Uma vez que a excitação é determinada, o sinal pode ser reconstruído através da filtragem da excitação, através de um filtro de síntese $1/\hat{A}(z)$ e, em seguida, desacentuado, por exemplo, por filtração, através do filtro $1/(1 - 0.68z^{-1})$. Observe que a excitação também pode ser usada para atualizar o *codebook*

adaptativo ACELP e permitir mudar de TCX para ACELP, em uma estrutura posterior. Note também que o comprimento da síntese TCX pode ser dado pelo comprimento da estrutura TCX (sem sobreposição): amostras de 256, 512 ou 1024 para mod [] de 1,2 ou 3, respectivamente.

A funcionalidade de uma configuração da fase de análise de codificação preditiva 12 será discutida posteriormente, de acordo com a configuração mostrada na Fig. 7, usando análise LPC e síntese LPC no decisor 15, nas configurações correspondentes.

A Fig. 7 ilustra uma implementação mais detalhada de uma configuração de um bloco de análise LPC 12. O sinal de áudio é inserido no bloco de determinação do filtro, que determina a informação do filtro $A(z)$, ou seja, a informação sobre os coeficientes para o filtro de síntese. Essa informação é quantizada e sai como as informações de previsão de curto prazo necessárias para o decodificador. Em um subtrator 786, uma amostra atual do sinal é inserida e um valor previsto para a amostra atual é subtraída, de modo que para essa amostra o sinal de erro de previsão é gerado na linha 784. Observe que o sinal de erro de previsão também pode ser chamado de sinal de excitação ou estrutura de excitação (geralmente após ser codificado).

Uma configuração de um decodificador de áudio 80, para decodificar as estruturas codificadas, para obter estruturas de um sinal de áudio testado, caracterizada pelo fato de que uma estrutura que consiste em um número de amostras no domínio do tempo, é mostrada na Fig. 8a. O decodificador de áudio 80 consiste em um decodificador de recuperação de redundância 82 para

decodificar as estruturas codificadas, para obter informações sobre os coeficientes para um filtro de síntese e espectros da estrutura de domínio de previsão, ou estruturas de domínio espectral de previsão. O decodificador de áudio 80 inclui ainda um transformador de introdução de *aliasing* de tempo inverso 84, para transformar a estrutura de domínio espectral de previsão, para o domínio do tempo, para a obtenção de estruturas de domínio de previsão sobrepostas, caracterizado pelo fato de que o transformador de introdução de *aliasing* de tempo inverso 84 é adaptado para determinar as estruturas de domínio de previsão sobrepostas a partir de espectros de estrutura de domínio de previsão consecutiva. Além disso, o decodificador de áudio 80 consiste em um combinador de sobreposição/adição 86 para combinar as estruturas de domínio de previsão sobrepostas, para obter uma estrutura de domínio de previsão de uma forma criticamente amostrada. A estrutura de domínio de previsão pode consistir em sinal ponderado baseado na LPC. O combinador de sobreposição/adição 86 também pode incluir um conversor para converter as estruturas de domínio de previsão em estruturas de excitação. O decodificador de áudio 80 inclui ainda uma fase de síntese de previsão 88 para a determinação da estrutura de síntese com base nos coeficientes e na estrutura de excitação.

O combinador de sobreposição/adição 86 pode ser adaptado para a combinação de estruturas de domínio de previsão sobrepostas de tal forma que um número médio de amostras, em uma estrutura de domínio de previsão, é igual a um número médio de amostras do espectro da estrutura de domínio de previsão. Em configurações, o transformador de introdução de *aliasing* de tempo

inverso 84 pode ser adaptado para transformar os espectros da estrutura de domínio de previsão para o domínio no tempo, de acordo com a IMDCT, conforme os detalhes acima.

5 Geralmente, no bloco 86, após o "combinador de sobreposição/adição", pode, opcionalmente, haver nas configurações uma "recuperação de excitação", que é indicada entre parênteses nas Figs. 8a-c. Em configurações, a sobreposição/adição pode ser realizada no domínio ponderado da LPC, então o sinal ponderado pode ser convertido para o sinal de excitação por filtração,
10 através do inverso do filtro da síntese ponderada.

Além disso, em configurações, a fase de síntese de previsão 88 pode ser adaptada para a determinação da estrutura baseada em predição linear, ou seja, LPC. Outra configuração de um decodificador de áudio 80 é mostrada na Fig. 8b. O decodificador
15 de áudio 80, representado na Fig. 8b, mostra os componentes similares ao decodificador de áudio 80, representado na Fig. 8a, entretanto, o transformador de introdução de *aliasing* de tempo inverso 84, na configuração mostrada na Fig. 8b, inclui ainda um conversor 84a para converter os espectros da estrutura de domínio
20 de previsão em estruturas de domínio de previsão de sobreposição e um filtro de janelamento 84b para aplicar uma função de janelamento em estrutura de domínio de previsão, para obter as estruturas de domínio de previsão sobrepostas.

A Fig. 8c mostra outra configuração de um
25 decodificador de áudio 80 com componentes similares ao da configuração mostrada na Fig. 8b. Na configuração mostrada na Fig. 8c, o transformador de introdução de *aliasing* de tempo inverso 84 inclui ainda um processador 84c para a detecção de um evento e

para prover uma informação da sequência da janela, se o evento for detectado o filtro de janelamento 84b, e o filtro de janelamento 84b for adaptado para a aplicação da função de janelamento, de acordo com a informação da sequência de janela. O evento pode ser
5 uma indicação derivada ou provida pelas estruturas codificadas ou por qualquer metadado.

Em configurações de codificadores de áudio 10 e de decodificadores de áudio 80, os filtros de janelamento respectivos 17 e 84 podem ser adaptados para a aplicação de
10 funções de janelamento, de acordo com informações da sequência de janela. A Fig. 9 mostra uma janela geral retangular, na qual as informações da sequência da janela podem incluir uma primeira parte zero, onde a janela mascara as amostras, uma segunda parte
de desvio, onde as amostras de uma estrutura, ou seja, uma
15 estrutura de domínio de previsão ou uma estrutura de domínio de previsão-sobreposta, pode passar através sem ser modificada, e uma terceira parte zero, que, novamente, mascara as amostras no fim da estrutura. Em outras palavras, as funções de janelamento podem ser aplicadas, que suprimem um número de amostras de uma estrutura na
20 primeira parte zero, passam através de amostras em uma segunda parte do desvio e, então, suprimem as amostras no final de uma estrutura, na terceira parte zero. Nesse contexto, a eliminação também pode se referir a acrescentar uma sequência de zeros no início e/ou no final da parte do desvio da janela. A segunda parte
25 do desvio pode ser tal que a função de janelamento simplesmente tem um valor de 1, ou seja, as amostras são passadas sem alteração, ou seja, a função de janelamento muda através das amostras da estrutura.

A Fig. 10 mostra outra configuração de uma sequência de janelamento ou função de janelamento, caracterizada pelo fato de que a sequência de janelas ainda consiste em uma parte da linha em ascensão, entre a primeira parte zero e a
 5 segunda parte do desvio, e uma parte da linha em declínio, entre a segunda parte do desvio e a terceira parte zero. A parte da linha em ascensão também pode ser considerada como uma parte de aparecimento gradual, a parte da linha em declínio pode ser considerada como uma parte de desaparecimento gradual. Nas
 10 configurações, a segunda parte do desvio pode incluir uma sequência de unidades para não modificar, afinal, as amostras da estrutura de domínio LPC.

Em outras palavras, a TCX baseada na MDCT poderá solicitar ao decodificador aritmético uma série de coeficientes
 15 quantizados espectrais, a $1G$, que é determinada pelo $\text{mod}[]$ e valores last_lpd_mode do modo passado. Esses dois valores podem também definir o comprimento e a forma da janela, que serão aplicados na MDCT inversa. A janela pode ser composta de três partes: uma sobreposição à esquerda das amostras L , uma parte do
 20 meio de unidades de amostras M e uma parte de sobreposição à direita de amostras R . Para obter uma janela MDCT de comprimento $2*1g$, zeros ZL podem ser adicionados à esquerda e zeros ZR à direita.

A tabela a seguir ilustra o número de
 25 coeficientes espectrais como uma função do modo lpd passado e $\text{mod}[]$ para algumas configurações:

Valor de last_lpd_mode	Valor de mod [x]	Número lg de coeficientes espectrais	ZL	L	M	R	ZR
0	1	320	160	0	256	128	96
0	2	576	288	0	512	128	224
0	3	1152	512	128	1024	128	512
1..3	1	256	64	128	128	128	64
1..3	2	512	192	128	384	128	192
1..3	3	1024	448	128	896	128	448

A janela MDCT é dada por

$$W(n) = \begin{cases} 0 & \text{for } 0 \leq n < ZL \\ W_{\text{SIN_LEFT},L}(n - ZL) & \text{for } ZL \leq n < ZL + L \\ 1 & \text{for } ZL + L \leq n < ZL + L + M \\ W_{\text{SIN_RIGHT},R}(n - ZL - L - M) & \text{for } ZL + L + M \leq n < ZL + L + M + R \\ 0 & \text{for } ZL + L + M + R \leq n < 2lg \end{cases}$$

As configurações podem prover a vantagem que um
 5 retardo de codificação sistemático da MDCT, IDMCT,
 respectivamente, pode ser reduzido, quando comparado com a MDCT
 original, através da aplicação de funções de janela diferentes. A
 fim de prover mais detalhes sobre essa vantagem, a Fig. 11 mostra
 quatro gráficos de exibição, nos quais o primeiro, na parte
 10 superior, mostra um retardo sistemático, em unidades de tempo T,
 com base nas tradicionais funções de janelamento moldadas em
 triângulos usadas com a MDCT, que são mostradas no segundo gráfico
 de exibição, a partir da parte superior, na Fig. 11.

O retardo sistemático aqui considerado é o
 15 retardo que uma amostra tem experimentado, quando ela atinge a
 fase do decodificador, assumindo que não há retardo para a
 codificação ou a transmissão das amostras. Em outras palavras, o

retardo sistemático mostrado na Fig. 11 considera o retardo de codificação chamado por acumular as amostras de uma estrutura antes de a codificação iniciar. Como explicado acima, a fim de decodificar a amostra em T , as amostras entre 0 e $2T$ têm de ser transformadas. Isso gera um retardo sistemático para a amostra em T de outro T . Porém, essa amostra pode ser decodificada antes de a amostra, todas as amostras da segunda janela, que está centrada em $2T$, têm de estar disponíveis. Portanto, o retardo sistemático salta para $2T$ e cai para T , no centro da segunda janela. O terceiro gráfico de exibição, a partir da parte superior, na Fig. 11, mostra uma sequência de funções de janela, tal como provido por uma configuração. Isso pode ser visto quando comparado com as janelas da técnica anterior, no segundo gráfico de exibição, a partir da parte superior, na Fig. 11 que as áreas de sobreposição da parte de não zero das janelas foram reduzidas em $2T$. Em outras palavras, as funções da janela usadas nas configurações são tão grandes ou largas quanto as janelas da técnica anterior, no entanto, têm uma primeira parte zero e uma terceira parte zero, que se torna previsível.

Em outras palavras, o decodificador já sabe que há uma terceira parte zero e, portanto, a decodificação pode ser iniciada mais cedo, realizando a codificação, respectivamente. Portanto, o retardo sistemático pode ser reduzido por $2T$, como é mostrado na parte inferior da Fig. 11. Em outras palavras, o decodificador não tem que esperar as partes zero, que pode salvar $2T$. É evidente que, naturalmente, após o procedimento de decodificação, todas as amostras têm de ter o mesmo retardo sistemático. Os gráficos de exibição, na Fig. 11, apenas

demonstram o retardo sistemático que uma amostra experiencia até chegar ao decodificador. Em outras palavras, um retardo sistemático global, depois da decodificação seria $2T$ para a abordagem da técnica anterior, e $2T - 2\Delta T$ para as janelas na

5 configuração.

A seguir, uma configuração será considerada onde a MDCT é usada no codec AMR-WB+, substituindo a FFT. Portanto, as janelas serão detalhadas, de acordo com a Fig. 12, que define "L" como área de sobreposição à esquerda ou parte da linha em

10 ascensão; "M" as regiões de unidades ou a segunda parte do desvio; "R" a área de sobreposição à direita ou a parte da linha em declínio. Além disso, as partes do primeiro zero e do terceiro zero são consideradas. Com isso, a região de reconstrução perfeita na estrutura, que é rotulada de "PR", está indicada na Fig. 12

15 pela seta. Ademais, "T" indica a seta do comprimento do núcleo de transformação, que corresponde ao número de amostras no domínio da frequência, isto é, a metade do número de amostras no domínio do tempo, que são compostos pela primeira parte zero, a parte da linha em ascensão "L", a segunda parte do desvio "M", a parte da

20 linha em declínio "R" e a terceira parte zero. Com isso, o número de amostras de frequência pode ser reduzido quando se utiliza a MDCT, onde o número de amostras de frequências para a FFT ou a Transformação de Cosseno Discreto (DCT = Transformação de Cosseno Discreto)

25
$$T = L + M + R$$

em relação ao comprimento do codificador de transformação para a MDCT

$$T = L/2 + M + R/2.$$

A Fig. 13 ilustra o gráfico na parte superior do gráfico de exibição de uma sequência de exemplo de funções de janela para AMR-WB+. Da esquerda para a direita, o gráfico de exibição, na parte superior da Fig. 13a, mostra uma estrutura ACELP, TCX20, TCX20, TCX40, TCX80, TCX20, TCX20, ACELP e ACELP. A linha tracejada mostra a resposta de entrada zero, como já descrito acima.

Na parte inferior da Fig. 13, há uma tabela de parâmetros para as partes da janela diferentes, onde nessa configuração a parte de sobreposição à esquerda ou a parte da linha em ascensão $L=128$, quando qualquer estrutura TCXx segue outra estrutura TCXx. Quando uma estrutura ACELP segue uma estrutura TCXx, as janelas semelhantes são usadas. Se uma estrutura TCX20 ou TCX40 segue uma estrutura ACELP, então, a parte de sobreposição à esquerda pode ser omitida, ou seja, $L=0$. Ao transitar de ACELP para TCX80, uma parte de sobreposição de $L = 128$ pode ser usada. A partir do gráfico de exibição, na tabela na Fig. 13, pode ser visto que o princípio básico é manter-se na amostragem não-crítica, enquanto é suficiente para um código extra na reconstrução perfeita na estrutura, e mudar para amostragem crítica o mais breve possível. Em outras palavras, apenas a primeira estrutura TCX, após uma estrutura ACELP, permanece não criticamente amostrada com a configuração presente.

Na tabela mostrada na parte inferior da Fig. 13a, as diferenças relacionadas com a tabela para o convencional AMR-WB+, conforme ilustrado na Fig. 19, estão em destaque. Os parâmetros em destaque indicam a vantagem de configurações da presente invenção, nos quais a área de sobreposição é ampliada de

tal forma que transição pode ser realizada de forma mais suave e resposta de frequência da janela é melhorada, enquanto mantém criticamente a amostragem.

Na tabela mostrada na parte inferior da Fig. 13a
 5 pode ser visto que apenas para transições de ACELP para TCX, um código extra é introduzido, ou seja, somente para a transição $T > PR$, isto é, a amostragem não crítica é obtida. Para todas as transições de TCXx para TCXx ("x" indica toda duração da estrutura), o comprimento de transformação T é igual ao número de
 10 novas amostras perfeitamente reconstruídas, ou seja, a amostragem crítica é alcançada. A Fig. 13b ilustra a tabela com representações gráficas de todas as janelas para todas as possíveis transições, com a configuração baseada em MDCT do AMR-WB+. Como já indicado na tabela da Fig. 13, a parte esquerda L das
 15 janelas deixa de depender do comprimento de uma estrutura TCX anterior. As representações gráficas na Fig. 14b mostram também que a amostragem crítica pode ser mantida quando se alterna entre diferentes estruturas TCX. Para transições de TCX para ACELP, pode-se ver que um código extra de 128 amostras é produzido. Como
 20 o lado esquerdo da janela não depende do comprimento da estrutura TCX anterior, a tabela mostrada na Fig. 13b pode ser simplificada, tal como mostrado na Fig. 14a. A Fig. 14a mostra novamente uma representação gráfica das janelas para todas as transições possíveis, onde as transições de estruturas TCX podem ser
 25 resumidas em uma linha.

A Fig. 14b ilustra a transição de ACELP para uma janela TCX80, em detalhes. O gráfico de exibição, na Fig. 14b, mostra o número de amostras na abscissa e a função de janela na

ordenada. Considerando a entrada de uma MDCT, a parte zero à esquerda chega a partir de uma amostra 1 para amostra 512. A parte da linha em ascensão está entre a amostra 513 e 640, a segunda parte do desvio entre 641 e 1664, a parte da linha em declínio entre 1665 e 1792, a terceira parte zero entre 1793 e 2304. Sobre a discussão da MDCT acima, na presente configuração 2304, as amostras de domínio do tempo são transformadas em amostras de domínio de frequência 1152. De acordo com a descrição acima, a zona de *aliasing* do domínio do tempo da janela atual está entre as amostras 513 e 640, ou seja, na parte da linha em ascensão, que se estende nas amostras $L=128$. Outra zona de *aliasing* do domínio do tempo se estende entre a amostra 1665 e 1792, ou seja, a parte da linha em declínio de amostras $R=128$. Em virtude da primeira parte zero e da terceira parte zero, há uma zona de *non-aliasing*, onde a reconstrução perfeita é ativada entre amostras 641 e 1664 de tamanho $M=1024$. Na Fig. 14b a estrutura ACELP indicada pela linha tracejada termina na amostra de 640. Diversas opções sobre amostras da parte da linha em ascensão entre 513 e 640 da janela TCX80 surgem. Uma opção é descartar primeiramente as amostras e ficar com a estrutura ACELP. Outra opção é utilizar a saída ACELP, a fim de realizar o cancelamento de *aliasing* no domínio do tempo para a estrutura TCX80.

A Fig. 14c ilustra a transição de qualquer estrutura TCX, indicada por "TCXx" para uma estrutura TCX20 e retorna para qualquer estrutura TCXx. As Figs. 14b até 14f usam a mesma representação de gráfico de exibição como já foi descrito em relação à Fig. 14b. No centro, perto da amostra 256, na Fig. 14c, a janela TCX20 é mostrada. As amostras no domínio do tempo 512 são

transformadas pela MDCT em amostras de domínio de frequência 256. As amostras de domínio no tempo usam 64 amostras para a primeira parte zero, assim como para a terceira parte zero. Com isso, uma zona de não *aliasing* de tamanho $M=128$ estende-se em torno do

5 centro da janela TCX20. A sobreposição à esquerda ou a parte da linha em ascensão, entre as amostras 65 e 192, pode ser combinada para o cancelamento de *aliasing* no domínio do tempo com a parte da linha em declínio de uma janela anterior, como indicado pela linha tracejada. Com isso, uma área de produções de reconstrução

10 perfeita do tamanho $PR=256$. Como todas as partes da linha em ascensão de todas as janelas TCX são $L=128$ e encaixam em todas as partes da linha em declínio $R=128$, a estrutura TCX anterior, assim como as estruturas TCX seguintes, podem ser de qualquer tamanho.

Quando transita de ACELP para TCX20, uma janela diferente pode ser

15 usada como, se indica na Fig. 14d. Como pode ser observado na Fig. 14d, a parte da linha em ascensão foi escolhida para ser $L=0$, ou seja, uma borda retangular.

Com isso, a área de reconstrução perfeita $PR=256$. A Fig. 14e mostra um gráfico de exibição semelhante ao transitar

20 de ACELP para TCX40 e, como outro exemplo, a Fig. 14f ilustra a transição de qualquer janela TCXx para TCX80, para qualquer janela TCXx.

Em resumo, as Figs. 14b até 14f mostram que a região de sobreposição para a janela MDCT é sempre 128 amostras,

25 com exceção para o caso de transitar de ACELP para TCX20, TCX40 ou ACELP.

Ao transitar de TCX para ACELP ou de ACELP para TCX80, várias opções são possíveis. Em uma configuração, a janela

da amostra a partir da estrutura MDCT TCX pode ser descartada na região de sobreposição. Em outra configuração, as amostras em janela podem ser usadas para uma transição e para o cancelamento de um *aliasing* no domínio do tempo, nas amostras MDCT TCX, com

5 base nas amostras ACELP que receberam *aliasing*, na região de sobreposição. Em ainda outra configuração, a transição suave pode ser realizada sem cancelar o *aliasing* no domínio do tempo. Na transição ACELP para TCX, a resposta de entrada zero (ZIR = resposta de entrada zero) pode ser retirada no codificador para

10 janelamento e adicionada ao decodificador para recuperação. Em números, isso é indicado por linhas tracejadas dentro das janelas TCX, na sequência de uma janela ACELP. Na presente configuração, quando transitam de TCX para TCX, as amostras em janela podem ser usadas para transição.

15 Ao transitar de ACELP para TCX80, o comprimento da estrutura é mais longo e pode ser sobreposto com a estrutura ACELP, o cancelamento de *aliasing* no domínio do tempo ou descartar o método pode ser usado.

20 Ao transitar de ACELP para TCX80, a estrutura ACELP anterior pode introduzir um som. O som pode ser reconhecido como uma propagação de erro proveniente da estrutura anterior, devido ao uso de filtragem LPC. O método ZIR utilizado para TCX40 e TCX20 pode contribuir para o som. Uma variante para o TCX80 em configurações é usar o método ZIR com um comprimento de

25 transformação de 1088, ou seja, sem sobreposição com a estrutura ACELP. Em outra configuração o mesmo comprimento de transformação de 1152 pode ser mantido e ocorrer o zeramento da área de sobreposição pouco antes de a ZIR poder ser utilizada, como

mostrado na Fig. 15. A Fig. 15 mostra uma transição de ACELP para TCX80, com redução a zero da área sobreposta e usando o método ZIR. A parte ZIR é novamente indicada pela linha tracejada após o final da janela ACELP.

5 Resumindo, as configurações da presente invenção proveem a vantagem de que a amostragem crítica pode ser realizada para todas as estruturas TCX, quando uma estrutura TCX preceder. Em comparação com a abordagem convencional, uma redução de código extra 1/8 pode ser alcançada. Além disso, as configurações proveem
10 a vantagem de que a área de transição ou de sobreposição, entre estruturas consecutivas, pode ter sempre 128 amostras, ou seja, mais do que para os convencionais AMR-WB+. As áreas de sobreposição melhoradas também proveem uma resposta de frequência aprimorada e uma suave transição. Com isso, uma melhor qualidade
15 de sinal pode ser alcançada com a codificação em geral e o processo de decodificação.

Dependendo dos requisitos de implementação de alguns dos métodos da invenção, estes podem ser implementados em hardware ou em software. A implementação pode ser realizada
20 utilizando um meio de armazenamento digital, em particular, um disco, um DVD, uma memória flash ou um CD com sinais de controle legíveis eletronicamente armazenados, que cooperam com um sistema de computador programável de tal forma que os métodos da invenção sejam executados. Geralmente, a presente invenção é, portanto, um
25 produto de programa de computador com um código de programa armazenado em um suporte legível por máquina, o código do programa que está sendo operado para a realização dos métodos da invenção quando o produto do programa de computador é executado em um

computador. Em outras palavras, os métodos da invenção são, portanto, um programa de computador com um código de programa para realizar, pelo menos, um dos métodos da invenção, quando o programa de computador é executado em um computador.

REIVINDICAÇÕES

1. Codificador de áudio (10) adaptado para codificar quadros de um sinal de áudio amostrado para obter quadros codificados finalmente, em que um quadro compreende inúmeras amostras de áudio de domínio de tempo, **caracterizado** por compreender:

um estágio de análise de codificação preditiva (12) para determinar informações em coeficientes de um filtro de síntese e quadros de excitação com base em quadros de amostras de áudio, um quadro de excitação que compreende amostras de um quadro de excitação para o filtro de síntese;

um transformador de introdução de geração de alias de tempo (14) para transformar quadros de excitação sobrepostos para um domínio de frequência para obter espectros de quadro de excitação, em que o transformador de introdução de geração de alias de tempo (14) é adaptado para transformar os quadros de excitação sobrepostos em uma forma criticamente amostrada;

um codificador de redução redundante (16) para a codificação de espectros de quadro de excitação para obter quadros de excitação com base nos coeficientes e nos espectros de quadro de excitação;

um codificador de livro de códigos (13) para a codificação de quadros de excitação com base em um livro de códigos predeterminado para obter quadros de domínio de excitação codificados de livro de códigos; e

decisor (15) para decidir quanto à

utilização de quadro de domínio de excitação codificado de livro de códigos ou um quadro codificado para obter um quadro codificado finalmente com base em uma medida de eficiência de codificação.

2. Codificador de áudio (10), de acordo com a reivindicação 1, **caracterizado** pelo transformador de introdução de geração de alias de tempo (14) ser adaptado para a transformação de quadros de excitação sobrepostos de modo que uma média de várias amostras de um espectro de quadro de excitação equivalha à média de número de amostras em um quadro de excitação.

3. Codificador de áudio (10), de acordo com as reivindicações 1 e 2, **caracterizado** pelo transformador de introdução de geração de alias de tempo (14) ser adaptado para a transformação de quadros de excitação sobrepostos de acordo com uma transformada discreta do cosseno modificada (MDCT).

4. Codificador de áudio (10), de acordo com as reivindicações 1 a 3, **caracterizado** pelo transformador de introdução de geração de alias de tempo (14) compreender um filtro de exibição parcial (17) para a aplicação de uma função de exibição parcial de quadros de excitação sobrepostos e um conversor (18) para a conversão de quadros de excitação sobrepostos de exibição parcial para os espectros de quadro de excitação.

5. Codificador de áudio (10), de acordo com a reivindicação 4, **caracterizado** pelo transformador de introdução de geração de alias de tempo (14) compreender um

processador (19) para a detecção de um evento e para o fornecimento de informações de uma janela de sequência se o evento for detectado e em que a função de filtro de janela (17) é adaptada para a aplicação de função de exibição parcial de acordo com as informações da janela de sequência.

6. Codificador de áudio (10), de acordo com a reivindicação 5, **caracterizado** pelas informações da janela de sequência compreenderem uma primeira parte zero, uma segunda parte desviada e uma terceira parte zero.

7. Codificador de áudio (10), de acordo com a reivindicação 6, **caracterizado** pelas informações da janela de sequência compreenderem uma parte da borda ascendente entre a primeira parte zero e a segunda parte e a parte descendente entre a segunda parte desviada e a terceira parte zero.

8. Codificador de áudio (10), de acordo com a reivindicação 7, **caracterizado** pela segunda parte desviada compreender uma sequência de uns (1) para não modificar as amostras dos espectros de quadro de excitação.

9. Codificador de áudio, de acordo com qualquer uma das reivindicações 1 a 8, **caracterizado** pela fase de análise de codificação preditiva (12) ser adaptada para a determinação de coeficientes com base em codificação preditiva linear (LPC).

10. Codificador de áudio (10), de acordo com a reivindicação 1, **caracterizado**, ainda, pelo comutador (20) ser acoplado a um decisor (15) para a comutação de

quadros de excitação entre o transformador de introdução de geração de alias de tempo (14) e o codificador de livro de códigos (13) com base em uma medida de eficiência de codificação.

11. Método de codificação de um sinal de áudio amostrado para obter quadros codificados finalmente, em que o quadro compreende inúmeras amostras de áudio de domínio de tempo, **caracterizado** por compreender as etapas de

determinar as informações de coeficientes para um filtro de síntese com base em quadro de amostras de áudio;

determinar os quadros de excitação com base em quadros de amostras de áudio, sendo que um quadro de excitação compreende amostras de um sinal de excitação para o filtro de síntese;

transformar os quadros de excitação sobrepostos para um domínio de frequência para obter espectros de quadro de excitação de modo criticamente amostrado que introduz geração de alias de tempo;

codificar os espectros de quadro de excitação para obter quadros codificados com base nos coeficientes e nos espectros de quadro de excitação;

codificar os quadros de excitação com base em um livro de códigos predeterminado para obter livro de códigos de campo de quadros de excitação codificados; e

decidir entre o uso de livro de códigos de campo de quadro de excitação codificado ou um quadro

codificado para obter um quadro codificado finalmente com base em uma medida de eficiência de codificação.

12. Programa de computador **caracterizado** por conter um código de programação para executar o método, de acordo com reivindicação 11, quando o código do programa executa em um computador ou processador.

13. Decodificador de áudio (80) para decodificar quadros codificados para obter quadros de um sinal de áudio amostrado, em que um quadro compreende inúmeras amostras de áudio de domínio de tempo, **caracterizado** por compreender:

decodificador de recuperação de redundância (82) para decodificar quadros codificados para obter informações de coeficiente para filtro de síntese e espectros de quadro de excitação;

transformador de introdução de geração de alias de tempo invertido (84) para transformar os espectros de quadro de excitação para o campo de tempo para obter quadros de excitação sobrepostos, em que transformador de introdução de geração de alias de tempo invertido (84) é adaptado para determinação de quadros de excitação sobrepostos para espectros de quadro de excitação consecutivo, em que o transformador de introdução de geração de alias de tempo invertido (84) compreende, ainda, um conversor (84a) para conversão dos espectros de quadro de excitação para quadros de excitação sobrepostos convertidos e um filtro de janela (84b) para aplicação de função de exibição parcial para quadros de excitação

sobrepostos convertidos para obter os quadros de excitação sobrepostos, em que o transformador de introdução de geração de alias de tempo invertido(84) processador (84c) para detecção e fornecimento de informações da janela de sequência se o evento é detectado para o filtro de janela (84b) e em que o filtro de janela (84b) é adaptado para a aplicação a função de exibição parcial de acordo com as informações da janela de sequência, e em que as informações da janela de sequência compreende uma primeira parte zero, uma segunda parte desviada e uma terceira parte zero;

combinador de sobreposição/adição (86) para a combinação de quadros de excitação sobrepostos para obter um quadro de excitação em uma forma criticamente amostrada; e

estágio de síntese preditiva (88) para determinação dos quadros de amostras de áudio com base nos coeficientes e no quadro de excitação.

14. Decodificador de áudio (80), de acordo com a reivindicação 13, **caracterizado** pelo combinador de sobreposição/adição (86) ser adaptado para combinar quadros de excitação sobrepostos de modo que a média do número de amostras em um quadro de excitação equivalha à média do número de amostras em um espectro de quadro de excitação.

15. Decodificador de áudio (80), de acordo com qualquer uma das reivindicações 13 ou 14, **caracterizado** pelo transformador de introdução de geração de alias de tempo invertido (84) ser adaptado para transformação dos espectros de quadro de excitação para o campo do tempo de

acordo com uma transformada discreta do cosseno modificada inversa(IMDCT).

16. Decodificador de áudio (80), de acordo com qualquer uma das reivindicações 13 a 15, **caracterizado** pela forma de síntese predita (88) ser adaptada para a determinação do quadro de amostras de áudio baseada na codificação preditiva linear (LPC).

17. Decodificador de áudio (80), de acordo com a reivindicação 13, **caracterizado** pela sequência de janela compreender, ainda, uma parte da borda ascendente entre a primeira parte zero e a segunda parte desviada e pela parte da borda descendente entre a segunda parte desviada e a terceira parte zero.

18. Decodificador de áudio (80), de acordo com a reivindicação 17, **caracterizado** pela segunda parte desviada compreender uma sequência de uns (1) para modificação do quadro de excitação.

19. Método de decodificação de quadros codificados para obter quadros de sinal de áudio amostrado, em que um quadro compreende inúmeras amostras de áudio de domínio de tempo, **caracterizado** por compreender:

decodificar os quadros codificados para obter informações de coeficientes para filtro de síntese e espectros de quadro de excitação;

transformar os espectros de quadro de excitação para o campo de tempo para obter quadros de excitação sobrepostos de espectros de quadro de excitação consecutivo, em que a etapa de transformação compreende

converter os espectros de quadro de excitação para quadros de excitação sobrepostos convertidos,

aplicar uma função de exibição parcial, por um filtro de janela (84b), para os quadros de excitação sobrepostos convertidos para obter os quadros de excitação sobrepostos,

detectar um evento e fornecer informações da janela de sequência se o evento for detectado para o filtro de janela (84b),

em que o filtro de janela (84b) é adaptado para a aplicação da função de exibição parcial de acordo com as informações da janela de sequência, e

em que as informações da janela de sequência compreendem uma primeira parte zero, uma segunda parte desviada e uma terceira parte zero;

combinar quadros de excitação sobrepostos para obter um quadro de excitação em uma forma criticamente amostrada; e

determinar o quadro com base nos coeficientes e no quadro de excitação.

20. Produto de programa de computador **caracterizado** por funcionar para realizar o método, de acordo com a reivindicação 19, quando o programa de computador é executado em um computador ou processador.

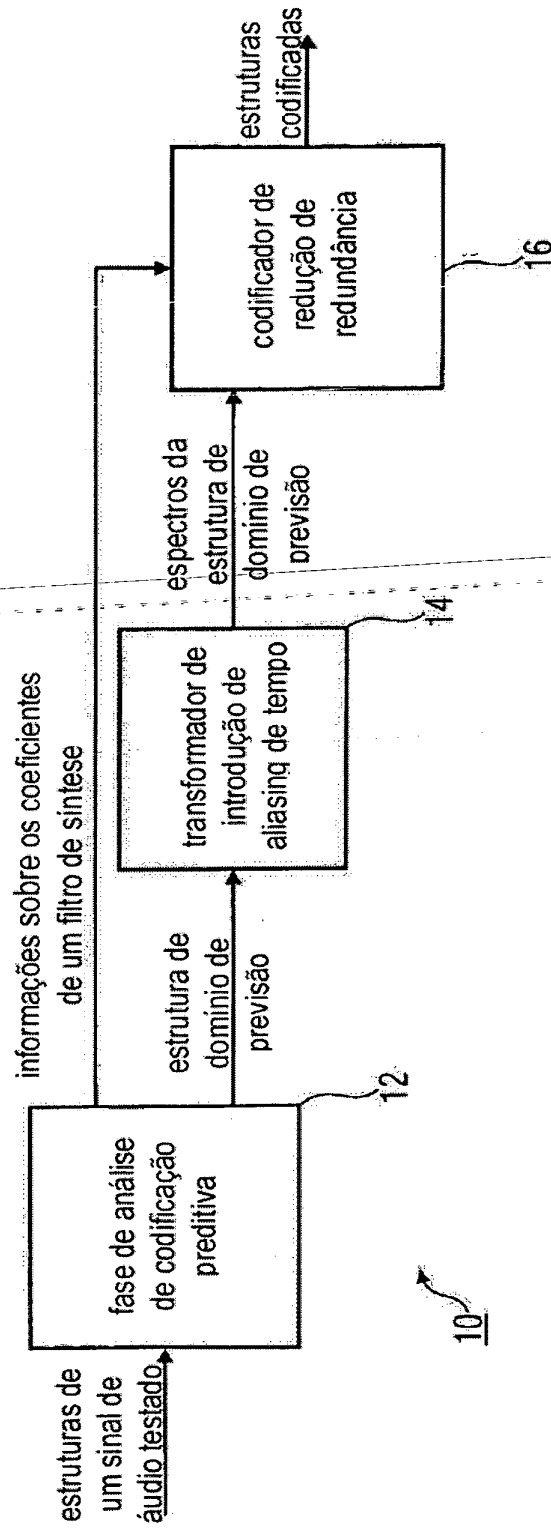


FIG 1

$$a) \quad X_k = \sum_{n=0}^{2N-1} X_n \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

$$b) \quad y_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

$$c) \quad w_n^2 + w_{n+N}^2 = 1$$

$$d) \quad w_n = \sin \left[\frac{\pi}{2N} \left(n + \frac{1}{2} \right) \right]$$

$$e) \quad w_n = \sin \left(\frac{\pi}{2} \sin^2 \left[\frac{\pi}{2N} \left(n + \frac{1}{2} \right) \right] \right)$$

$$f) \quad \cos \left[\frac{\pi}{N} \left(-n-1 + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] = \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] \text{ and } \\ \cos \left[\frac{\pi}{N} \left(2N-n-1 + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right] = -\cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

$$g) \quad \text{IMDCT}(\text{MDCT}(a, b, c, d)) = (a-b_R, b-a_R, c+d_R, c_R+d)/2$$

$$h) \quad (z_R, w_R) \cdot (z_R c + w d_R, z c_R + w_R d) = (z_R^2 c + w z_R d_R, w_R z c_R + w_R^2 d)$$

$$i) \quad (w, z) \cdot (w c - z_R d_R, z d - w_R c_R) = (w^2 c + w z_R d_R, z^2 d - w_R z c_R)$$

$$j) \quad (z_R^2 c + w z_R d_R, w_R z c_R + w_R^2 d) + (w^2 c - w z_R d_R, z^2 d - w_R z c_R) \\ = ([z_R^2 + w^2] c + [w z_R - w z_R] d_R, [w_R^2 + z^2] d + [w_R z - w_R z] c_R) = (c, d)$$

FIG 2

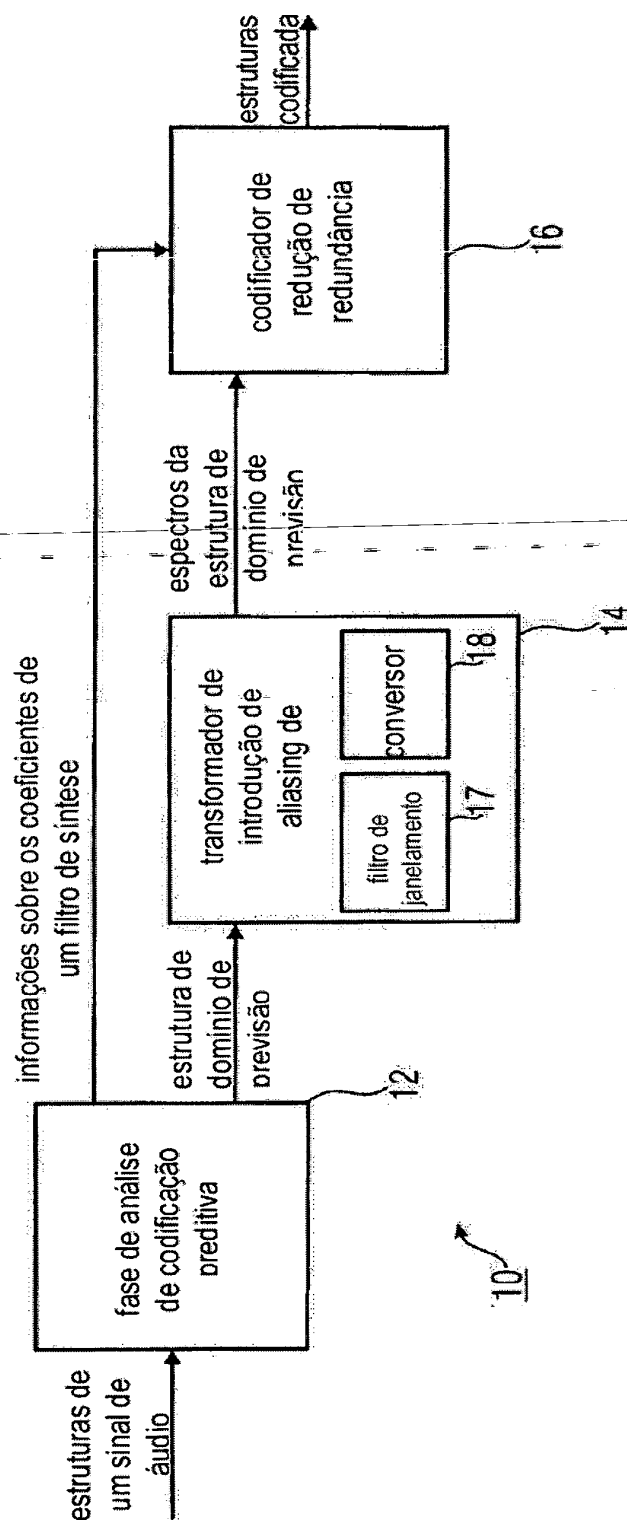


FIG 3A

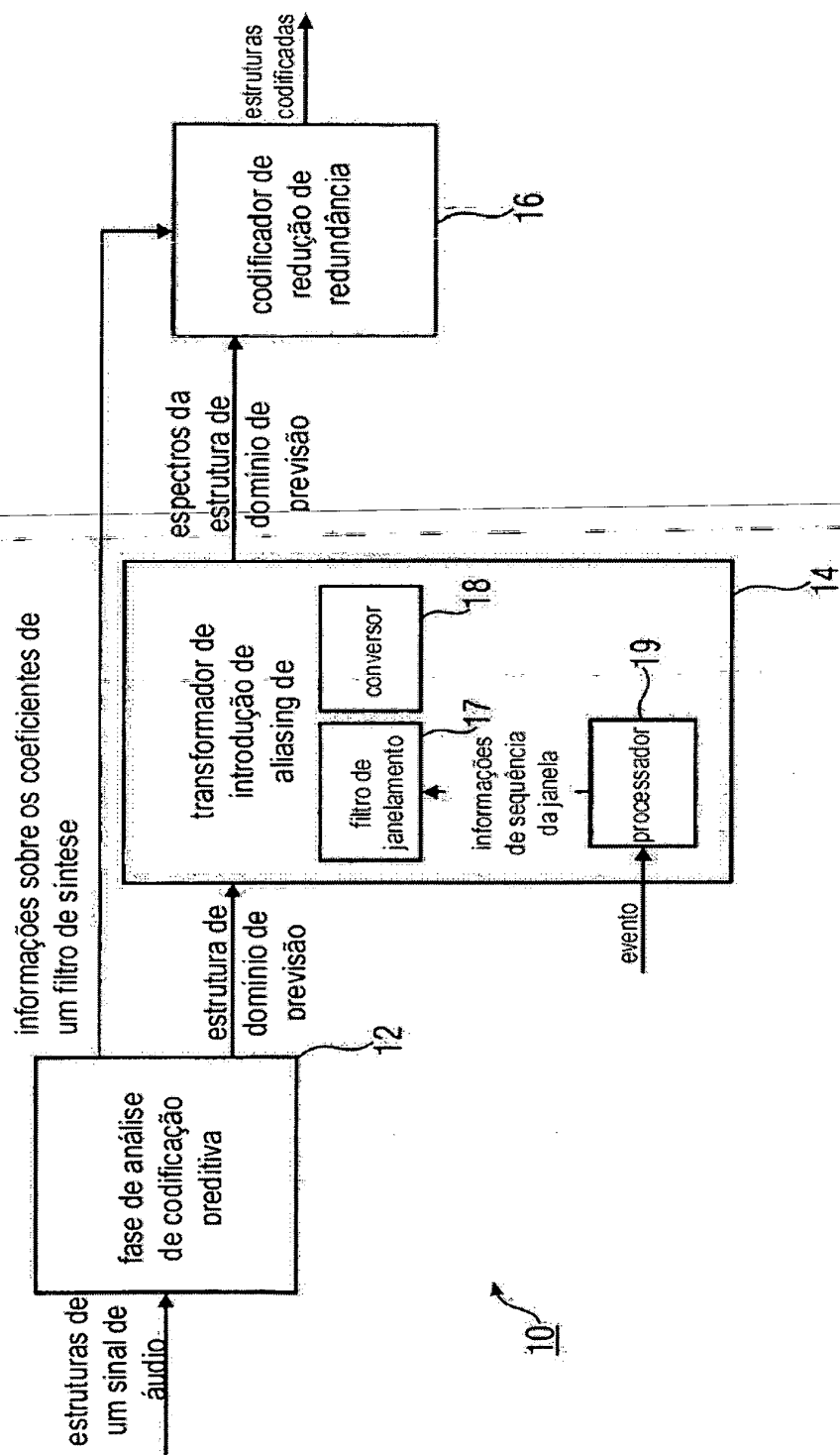


FIG 3B

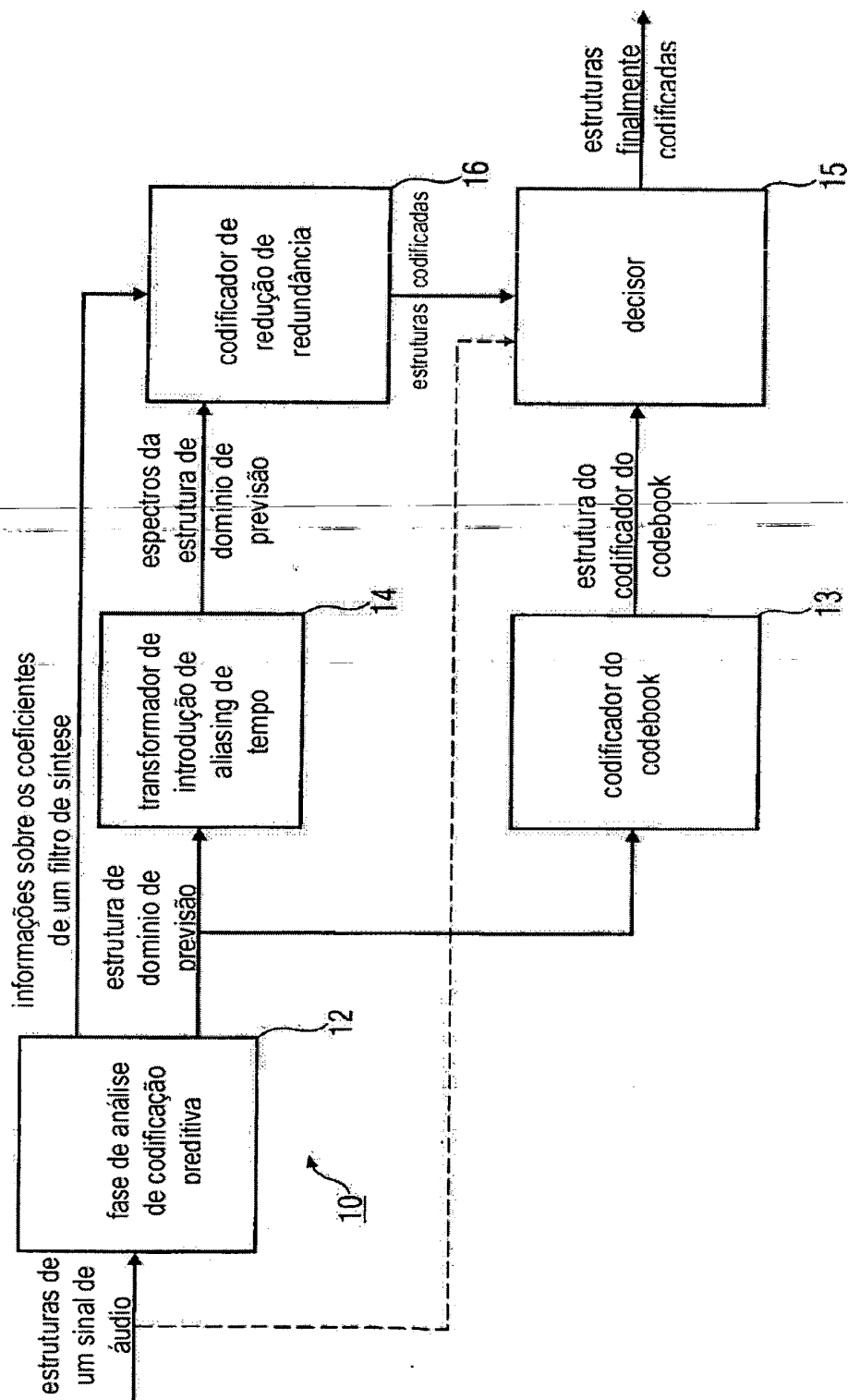


FIG 3C

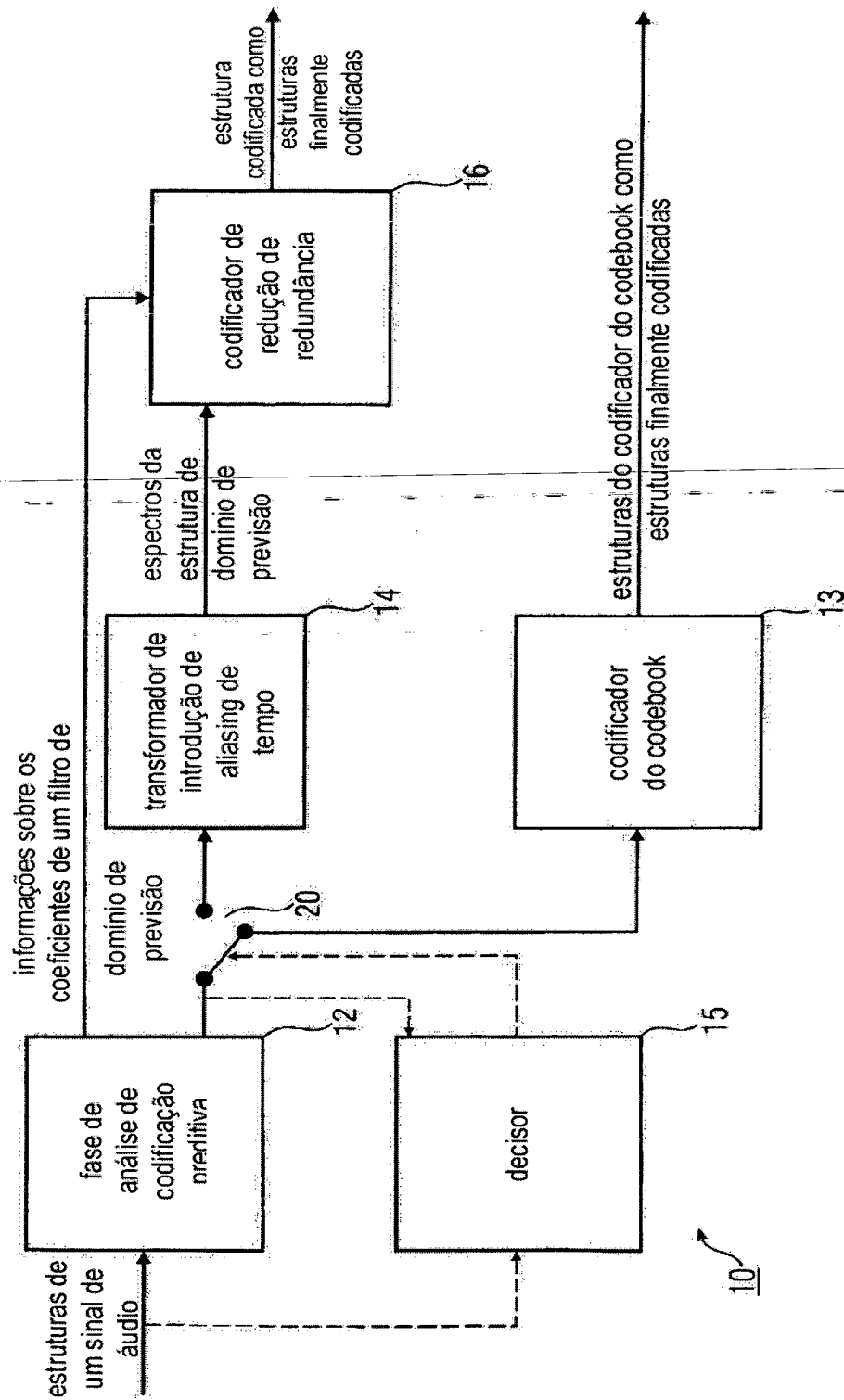


FIG 3D

segmento do sinal impulso (por exemplo, fala articulada)

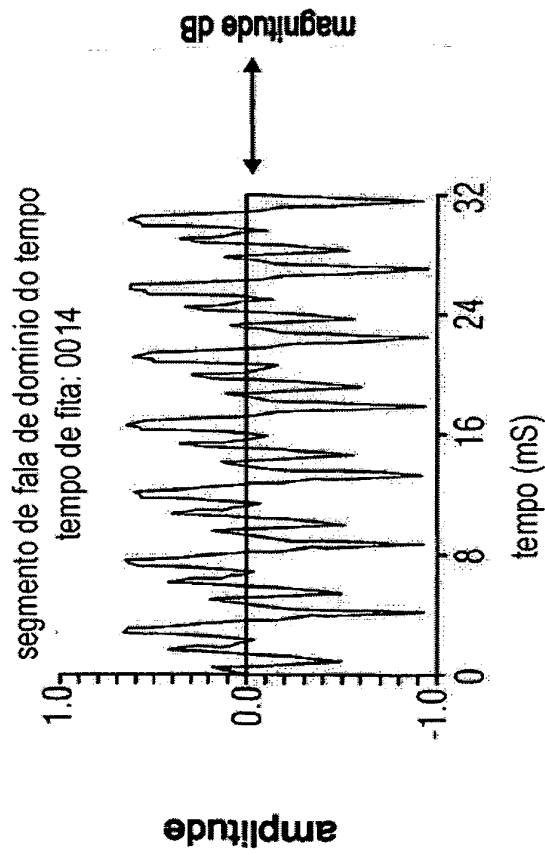


FIG 4A

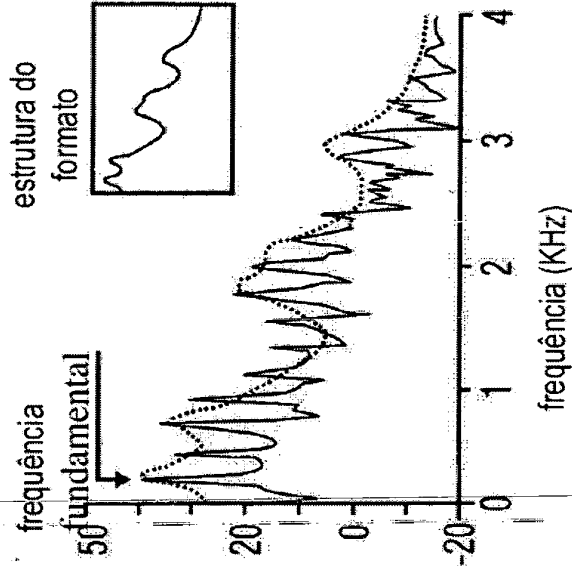


FIG 4B

segmento estacionário (por exemplo, fala não articulada)

magnitude (dB)

segmento de fala de domínio do tempo

amplitude

1.0
0.0
-1.0

tempo (mS)

0 8 16 24 32

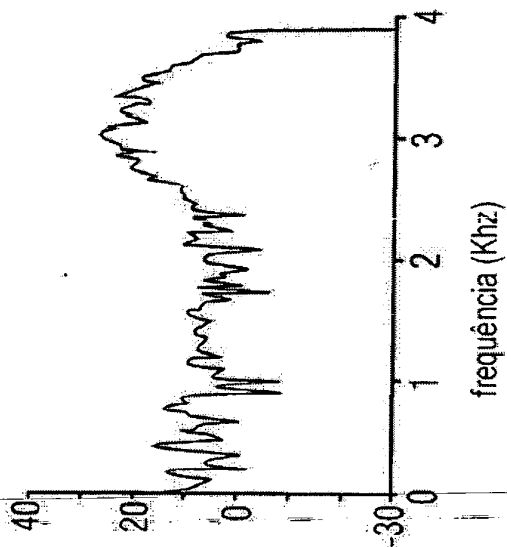
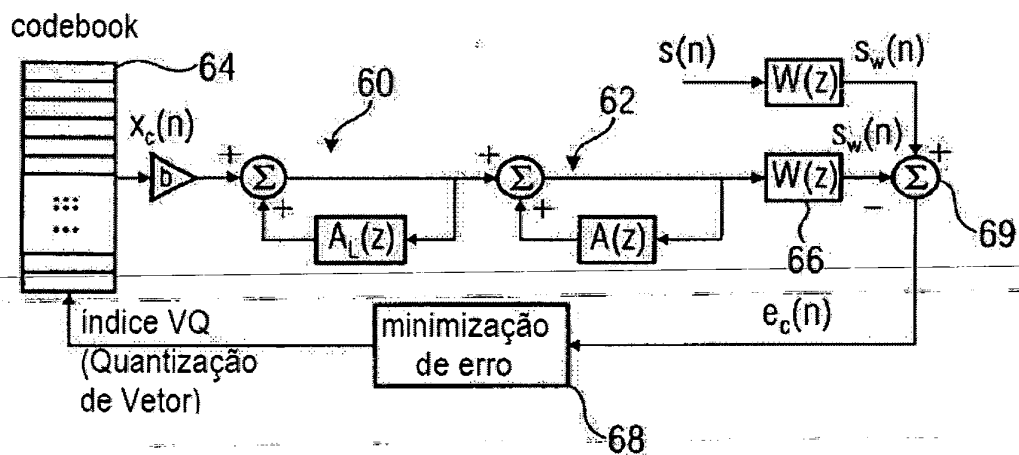


FIG 5A

FIG 5B

análise por síntese CELP



$A_L(z)$: previsão a longo prazo
 $\cong$ estrutura (fina) de passo

$A(z)$: previsão a curto prazo
 $\cong$ estrutura do formato / envelope espectral

FIG 6

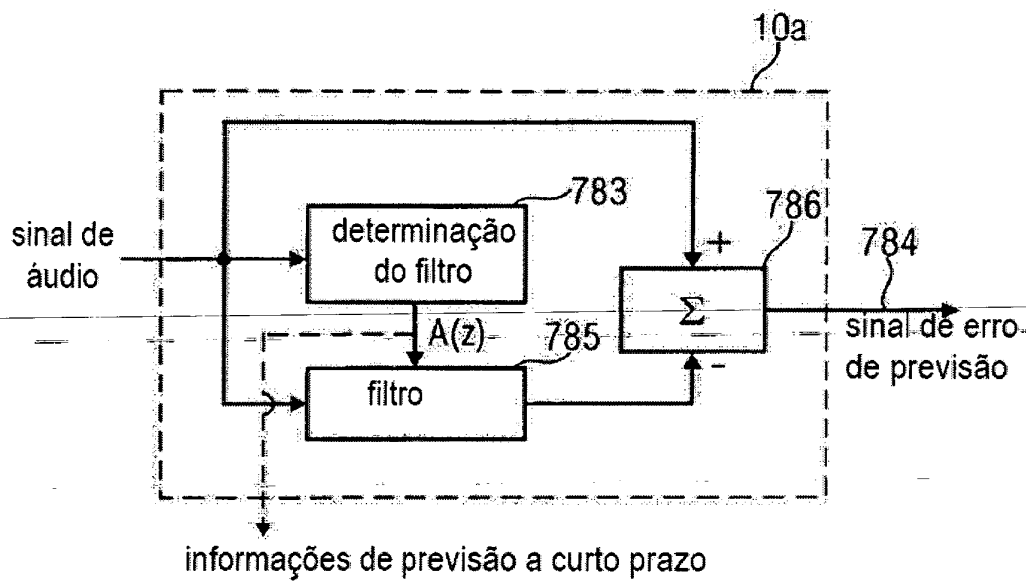


FIG 7

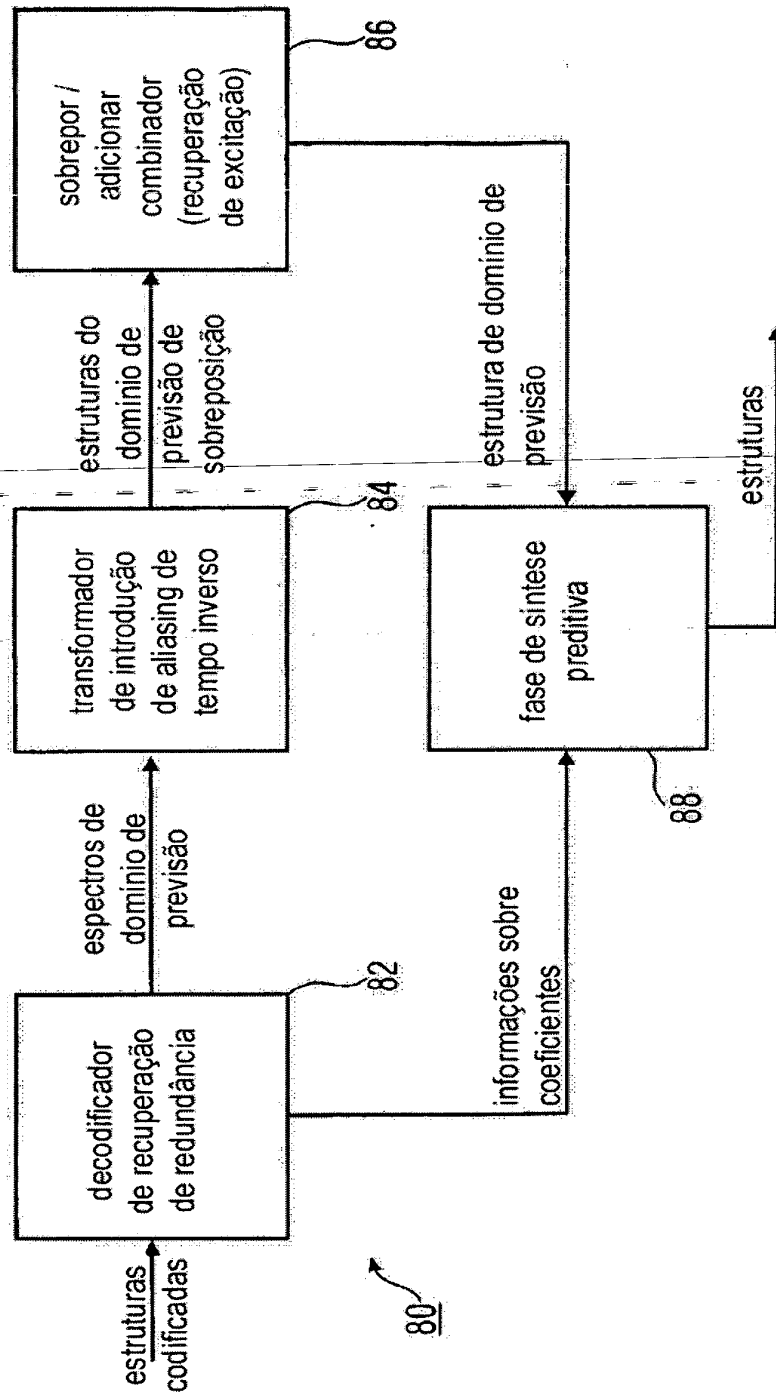


FIG 8A

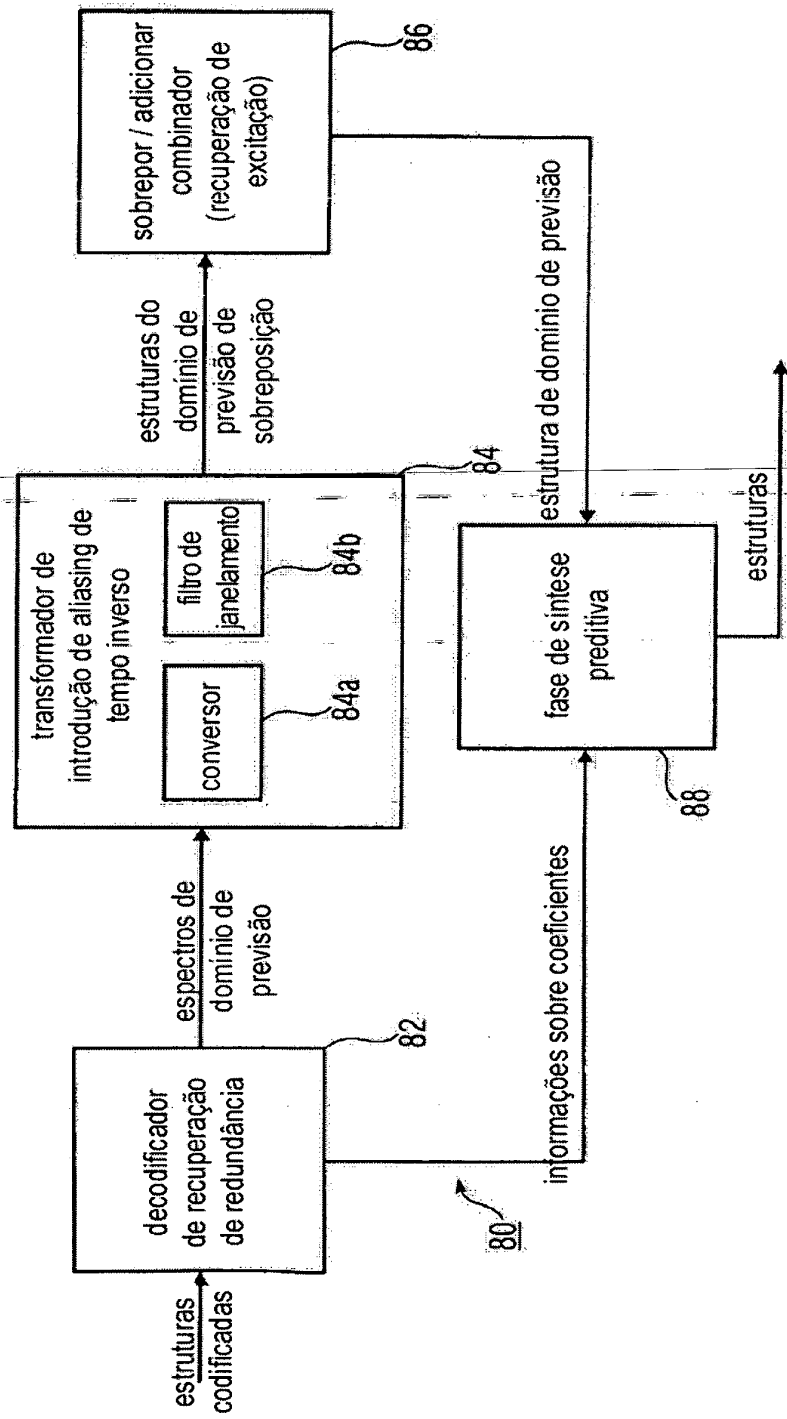


FIG 8B

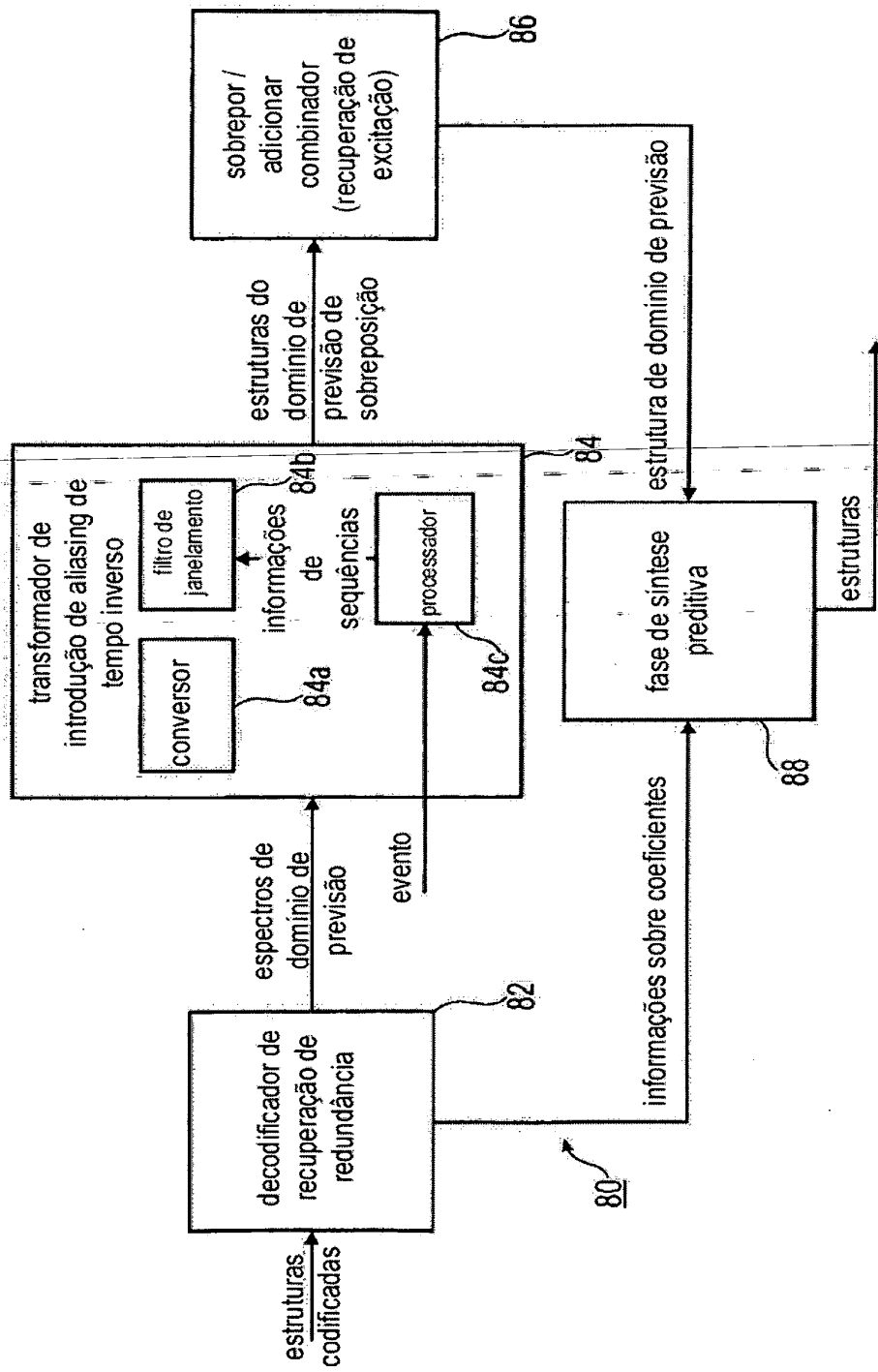


FIG 8C

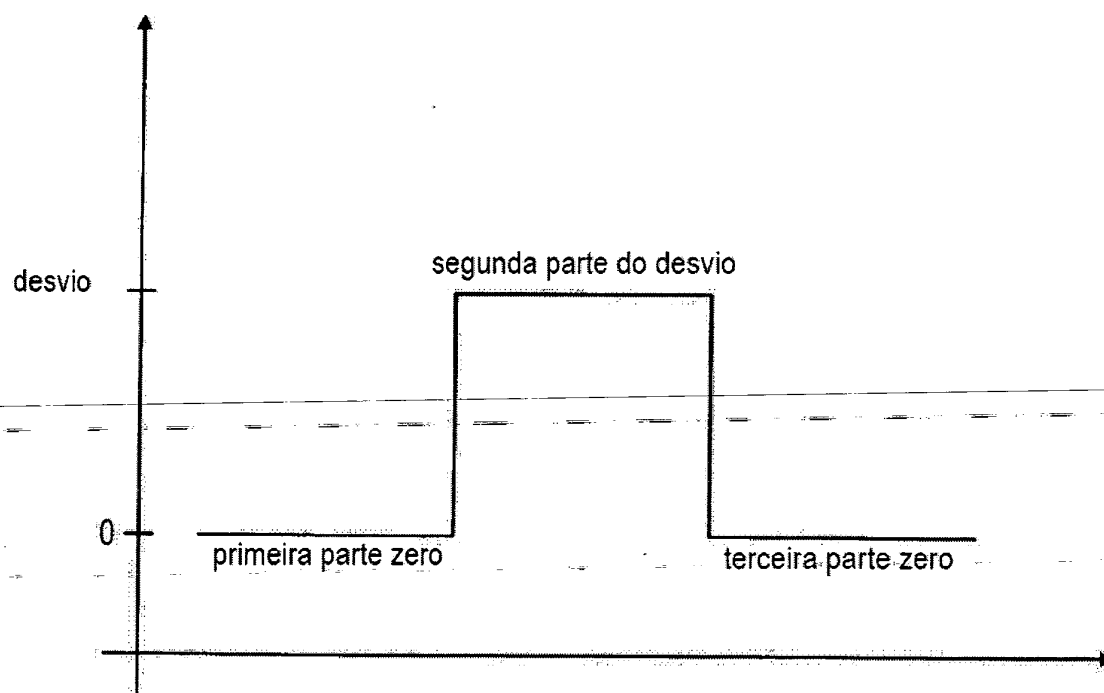


FIG 9

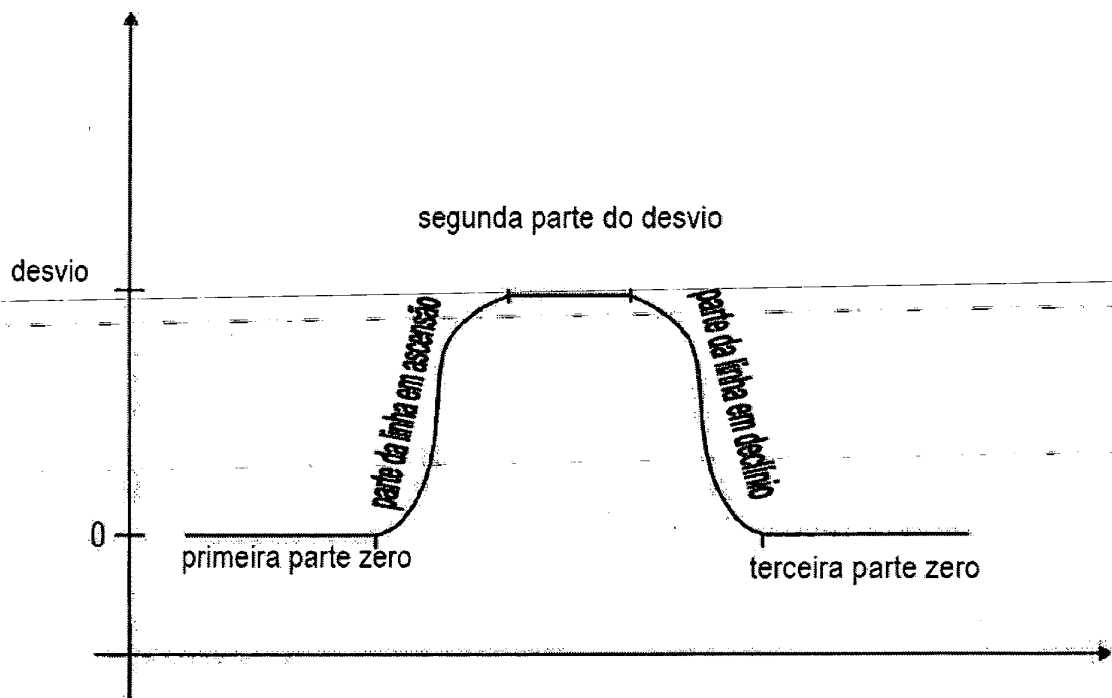
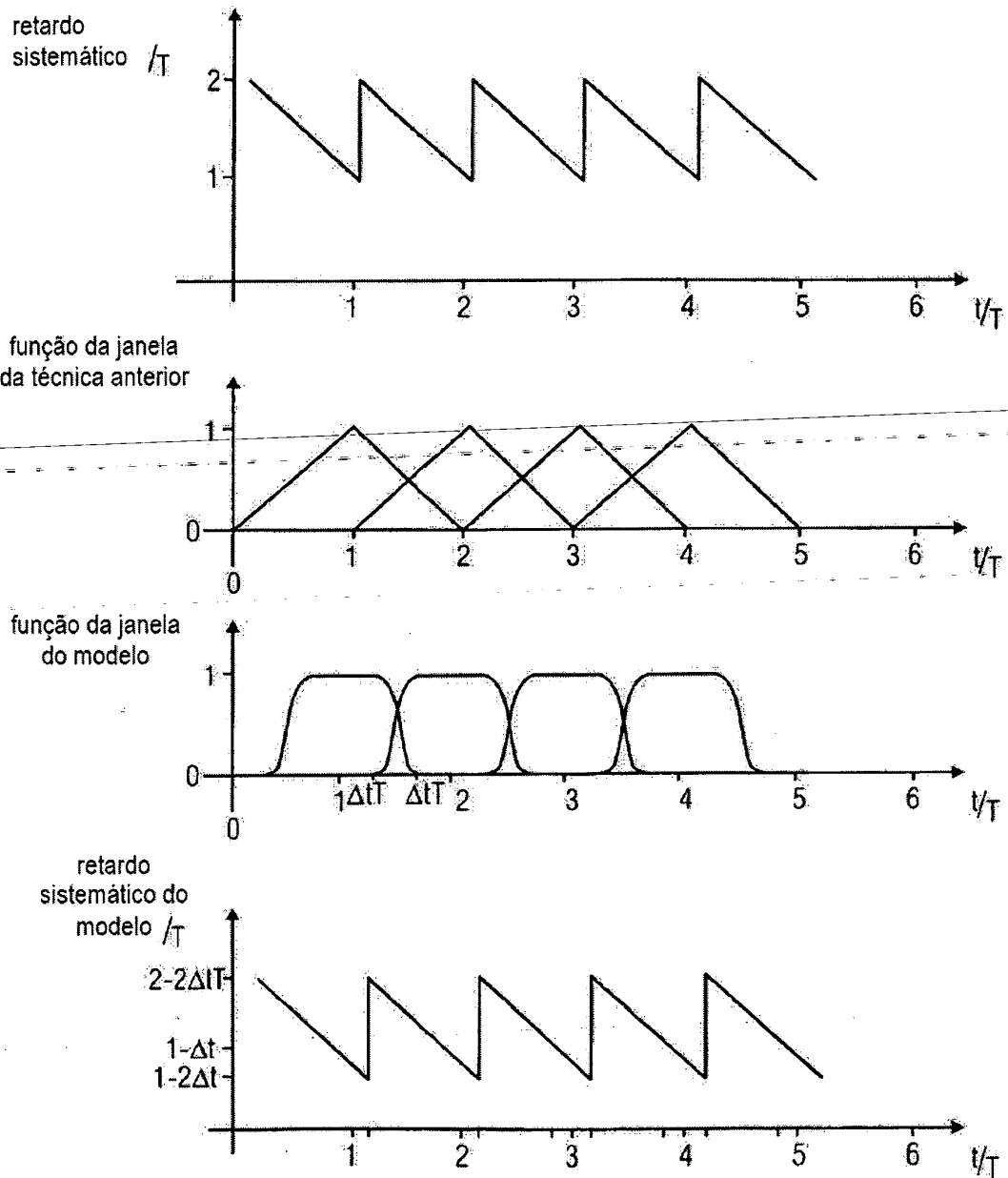


FIG 10



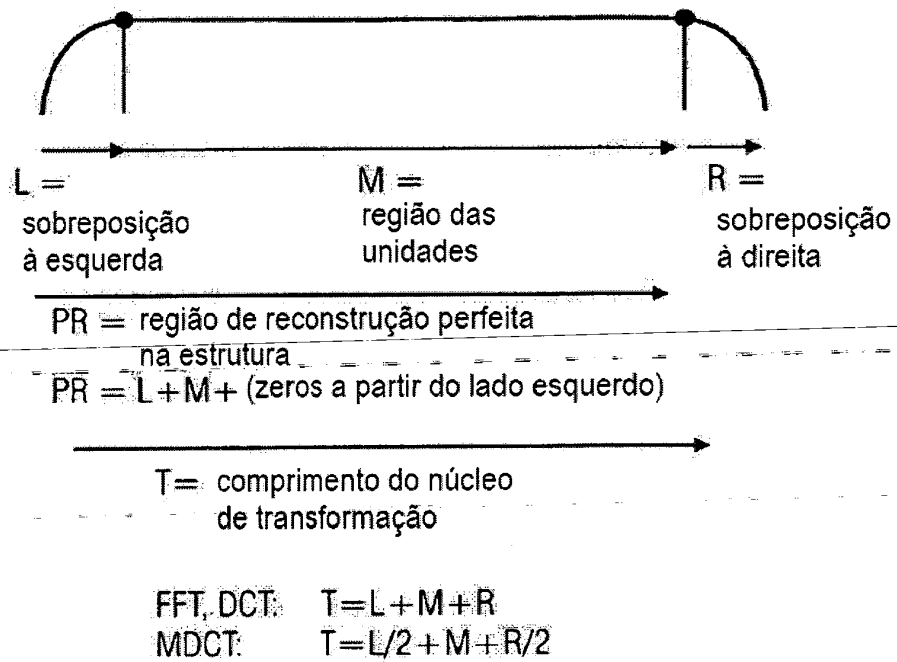
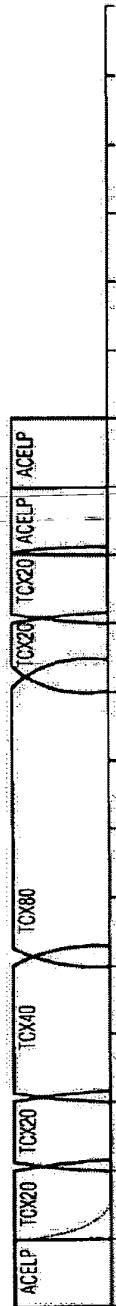


FIG 12



transição para	TCX80	TCX40	TCX20
de			
ACELP	$L=128, M=1024, R=128, PR=1024, T=1152$	$L=0, M=512, R=128, PR=512, T=576$	$L=0, M=256, R=128, PR=256, T=320$
TCX20	$L=128, M=896, R=128, PR=1024, T=1024$	$L=128, M=384, R=128, PR=512, T=512$	$L=128, M=128, R=128, PR=256, T=256$
TCX40	$L=128, M=896, R=128, PR=1024, T=1024$	$L=128, M=384, R=128, PR=512, T=512$	$L=128, M=128, R=128, PR=256, T=256$
TCX80	$L=128, M=896, R=128, PR=1024, T=1024$	$L=128, M=384, R=128, PR=512, T=512$	$L=128, M=128, R=128, PR=256, T=256$

FIG 13A

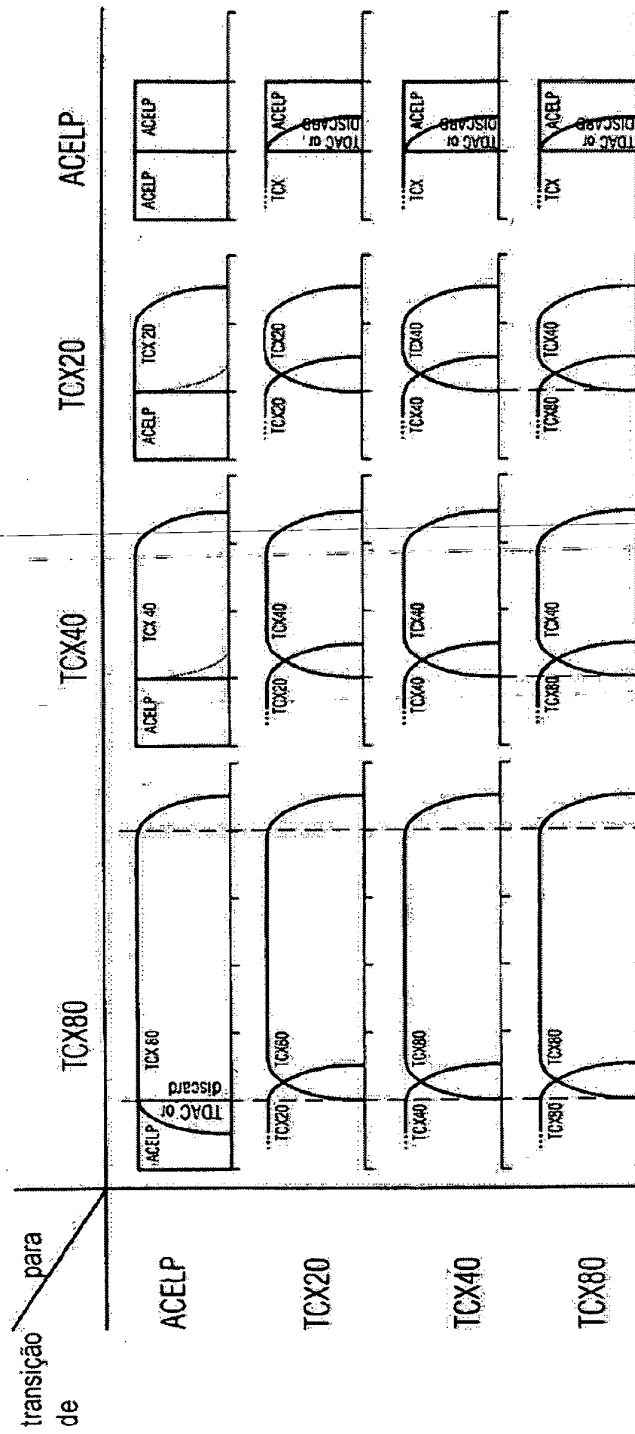


FIG 13B

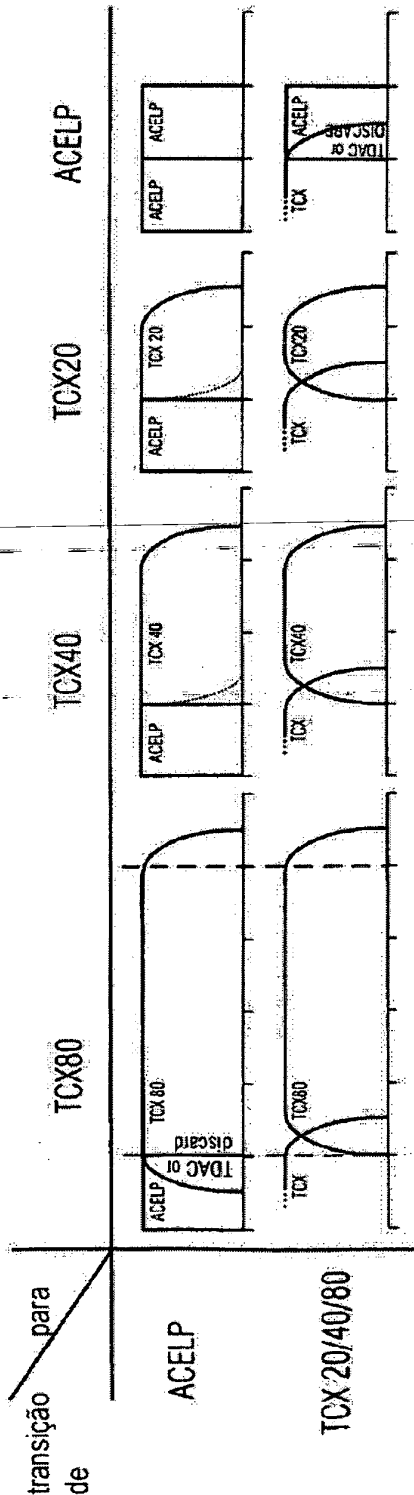


FIG 14A

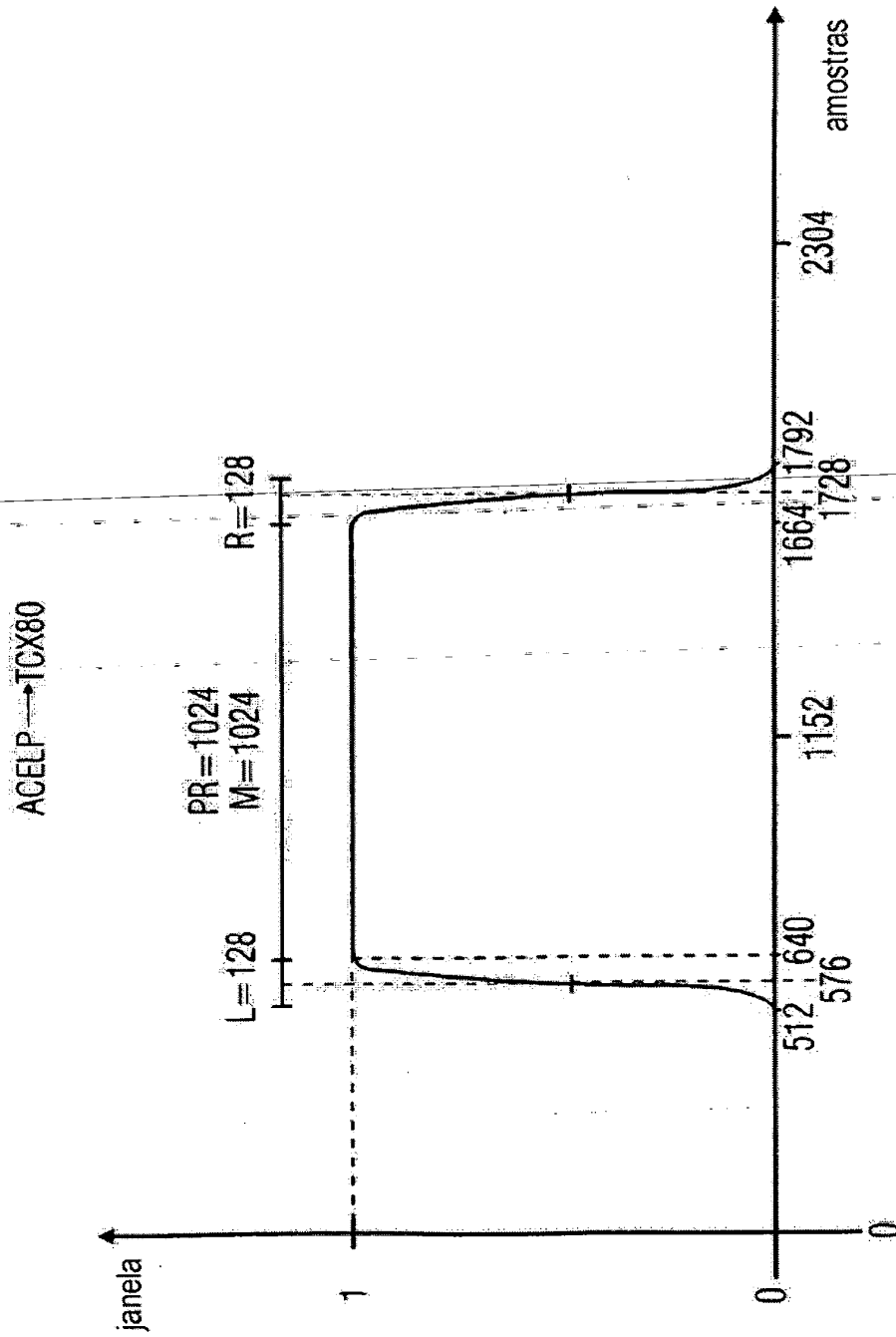


FIG 14B

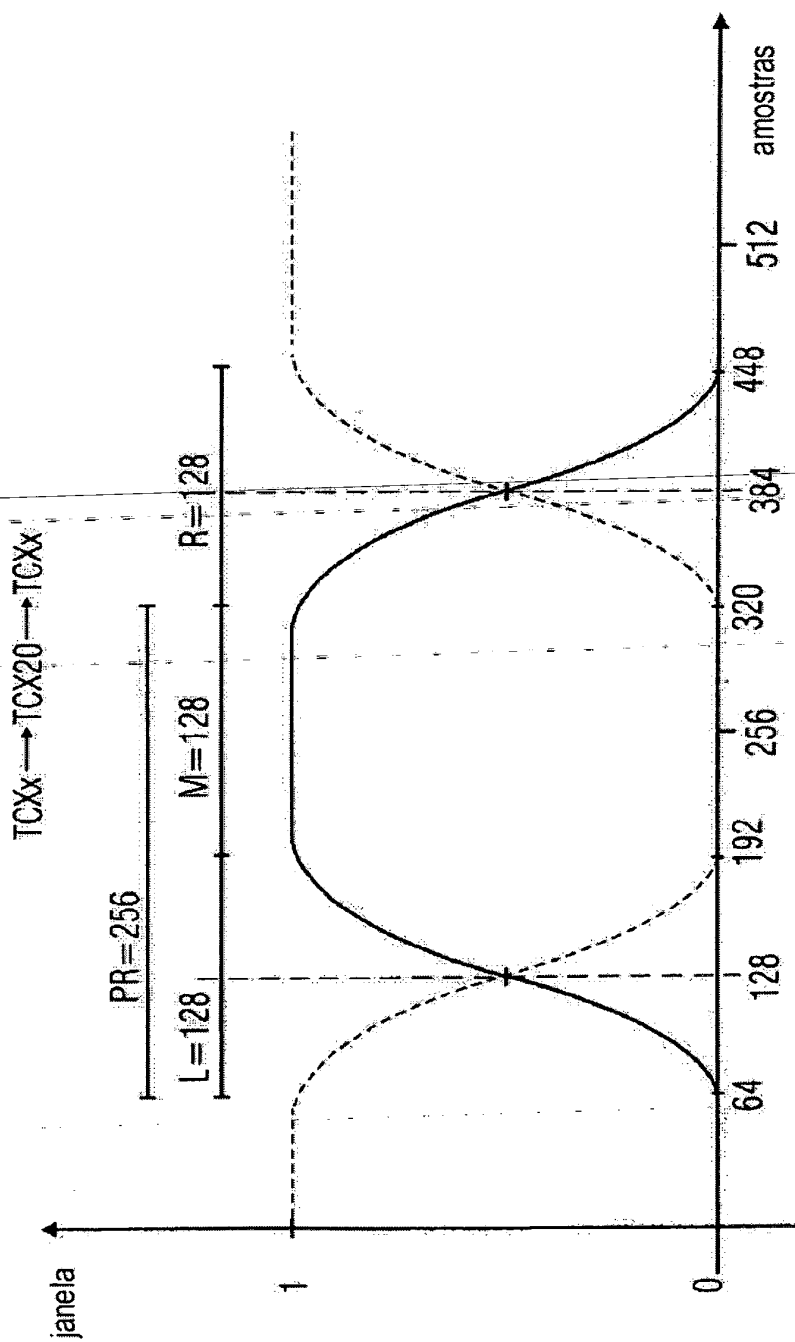


FIG 14C

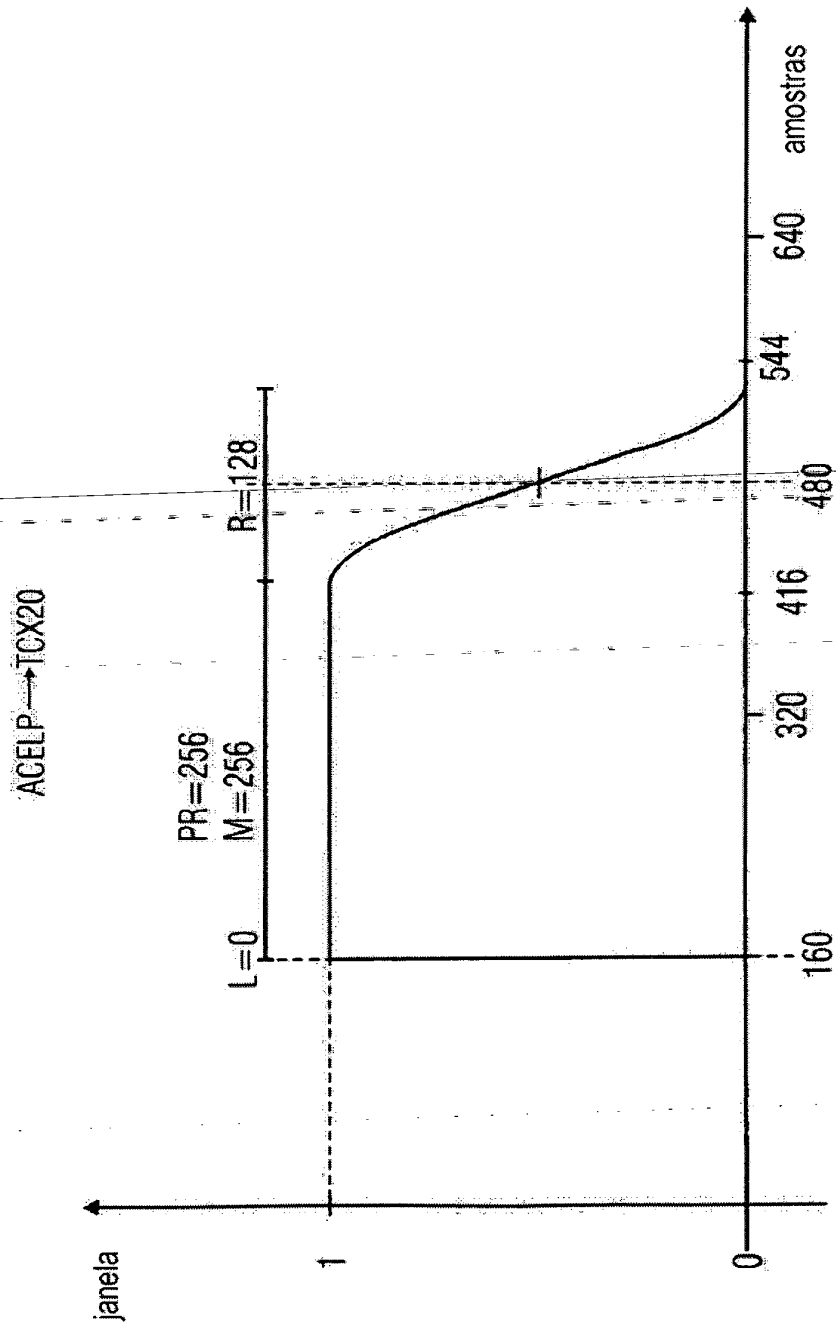


FIG 14D

ACELP → ICX40

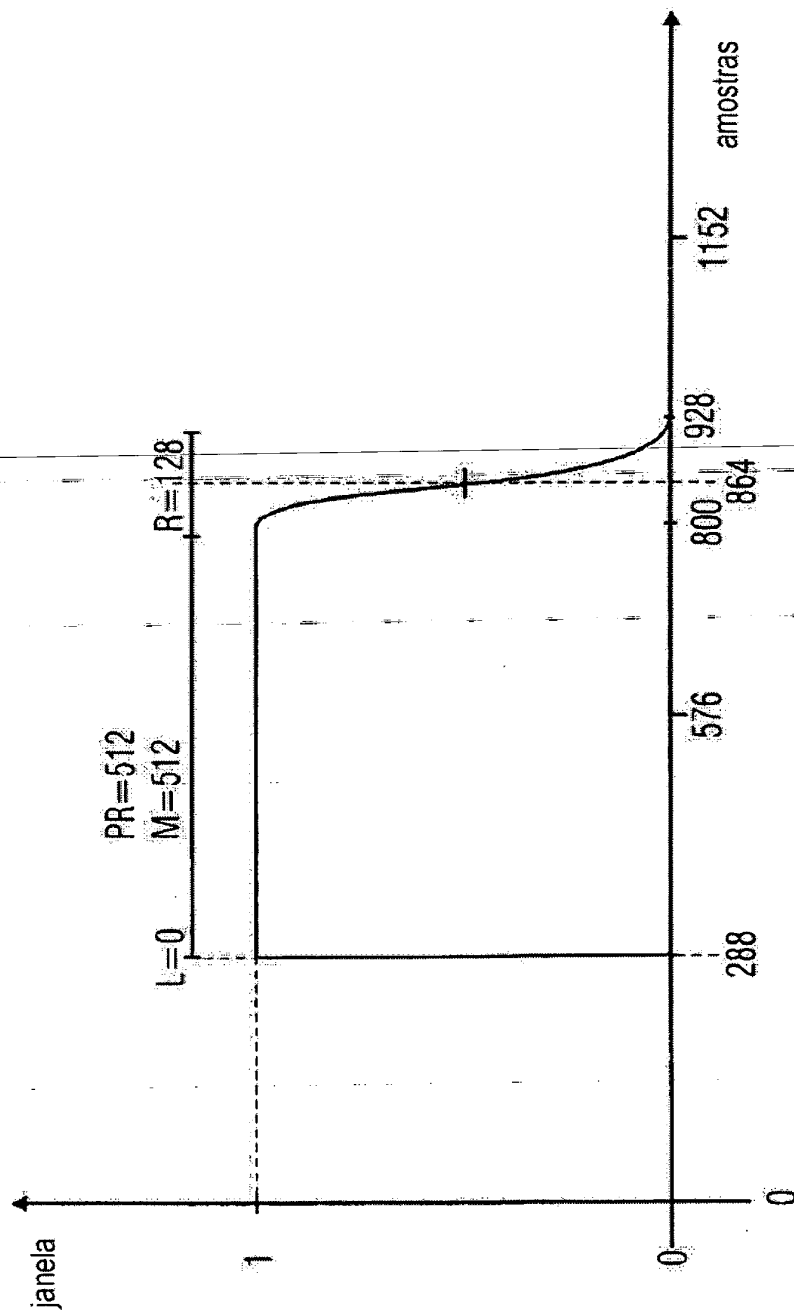


FIG 14E

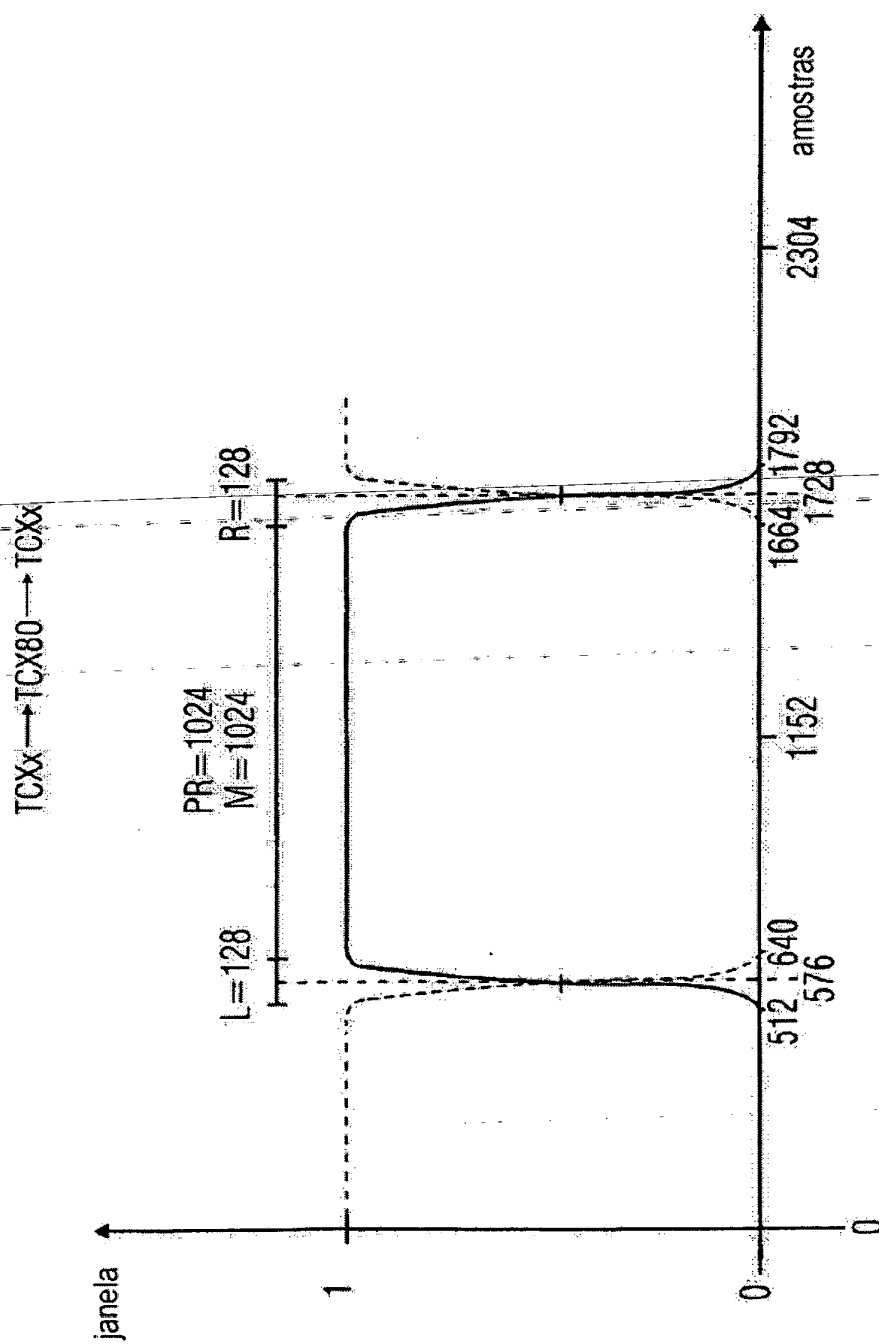


FIG 14F

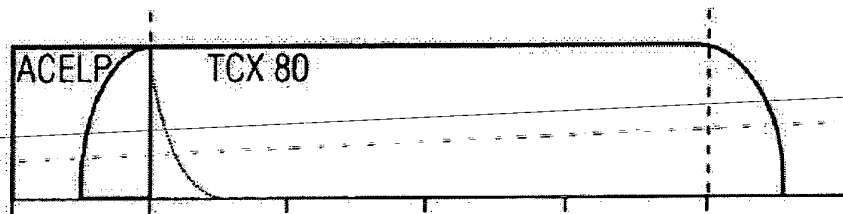


FIG 15

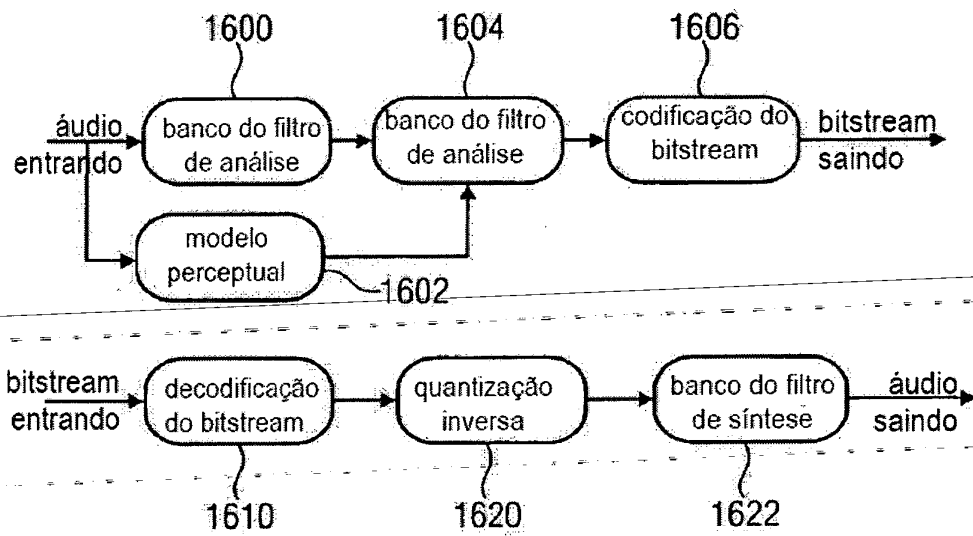


FIG 16
(TÉCNICA ANTERIOR)

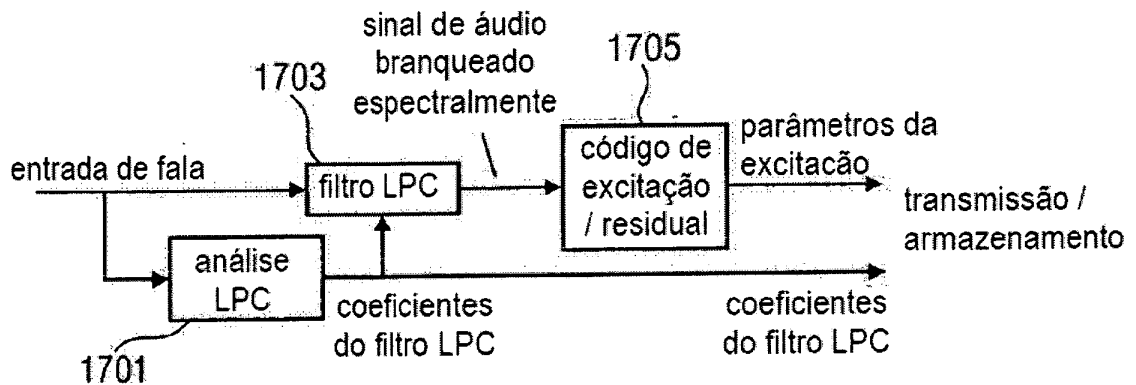


FIG 17A
(ESTADO DA TÉCNICA)

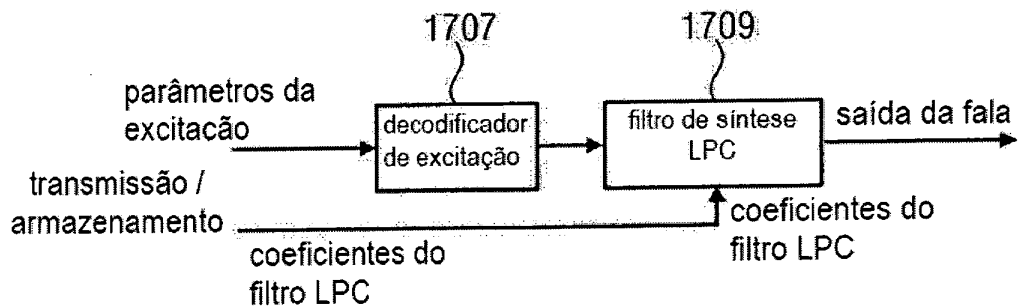


FIG 17B
(ESTADO DA TÉCNICA)

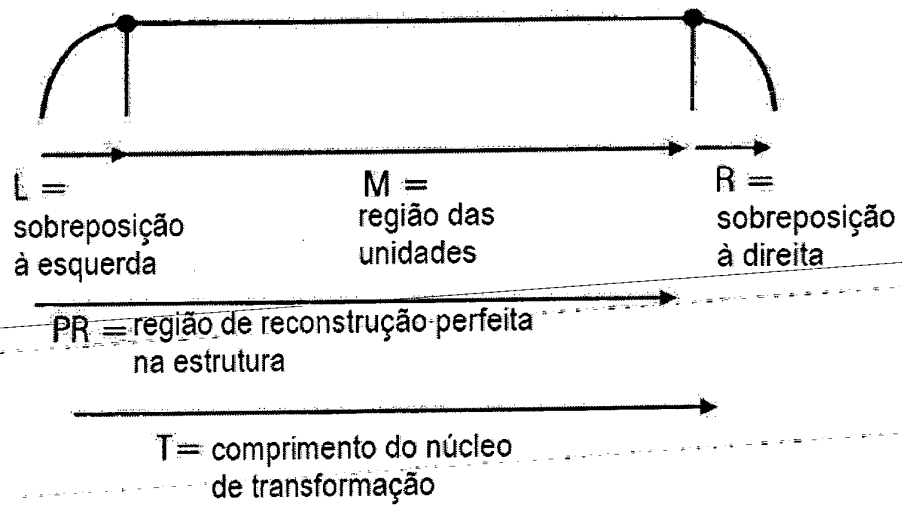


FIG 18
(ESTADO DA TÉCNICA)

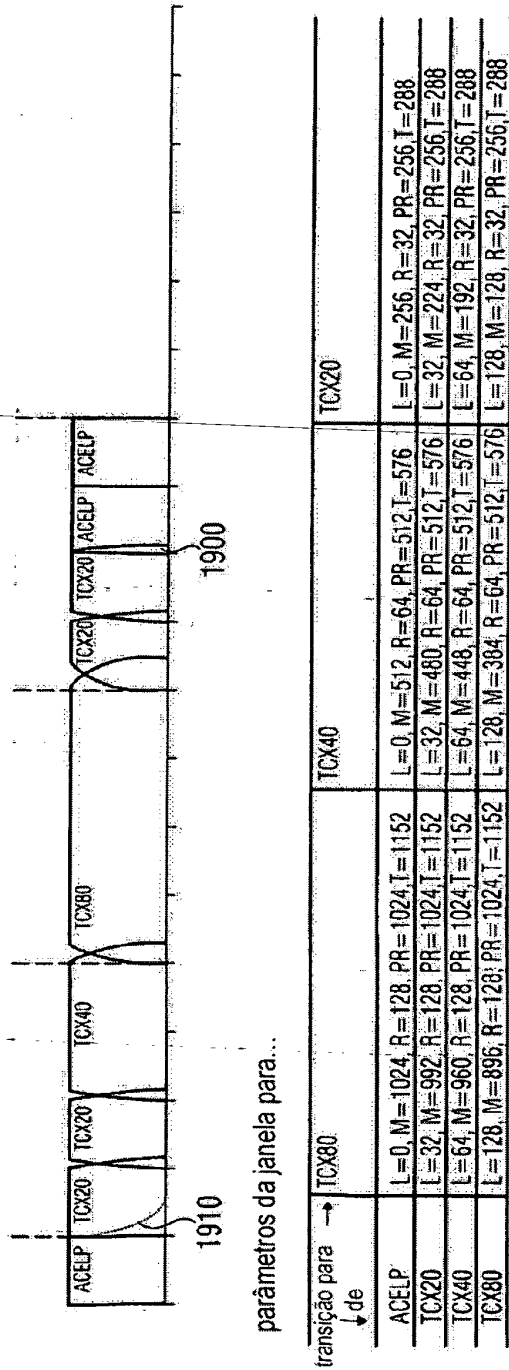


FIG 19
(ESTADO DA TÉCNICA)

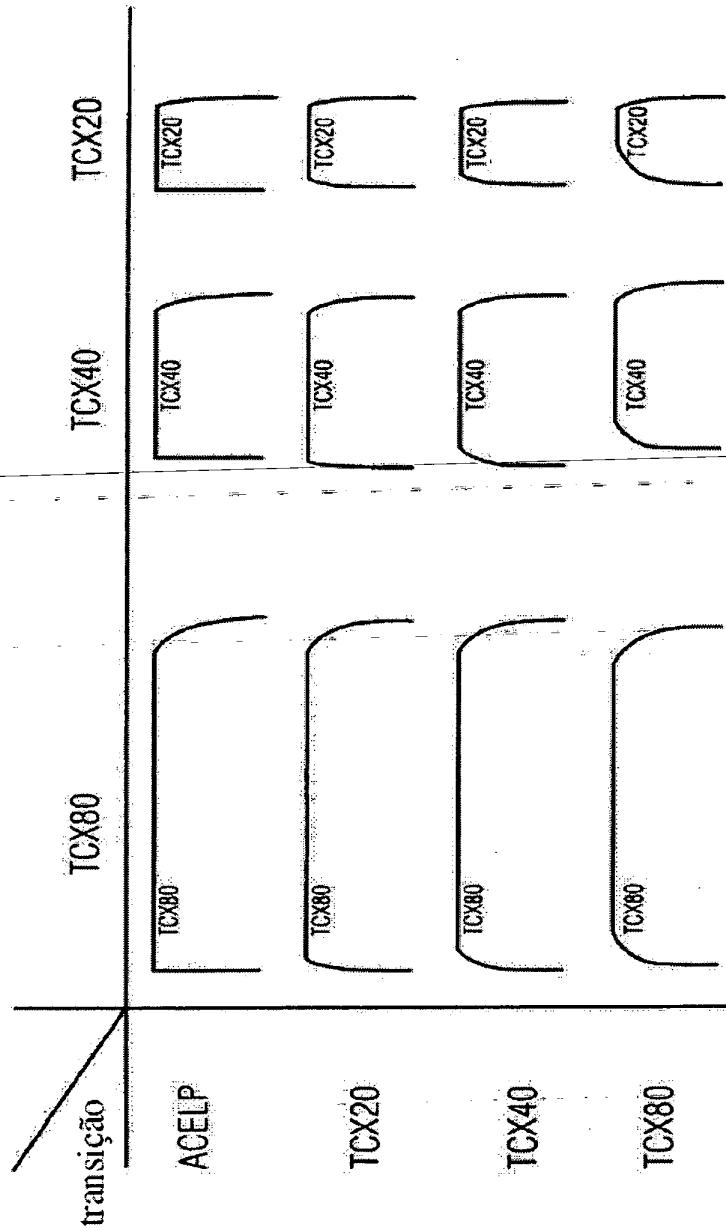


FIG 20
(TÉCNICA ANTERIOR)