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

(11) PI 0910793-2 B1

(22) Data do Depósito: 16/06/2009

(45) Data de Concessão: 24/11/2020

(54) Título: MÉTODO E DISCRIMINADOR PARA A CLASSIFICAÇÃO DE DIFERENTES SEGMENTOS DE UM SINAL

(51) Int.Cl.: G10L 19/20; G10L 19/22; G10L 25/51; G10L 25/78.

(52) CPC: G10L 19/20; G10L 19/22; G10L 25/51; G10L 2025/783.

(30) Prioridade Unionista: 11/07/2008 US 61/079,875.

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

(72) Inventor(es): YOSHIKAZU YOKOTANI; GUILLAUME FUCHS; STEFAN BAYER; JENS HIRSCHFELD; JUERGEN HERRE; JÉRÉMIE LECOMTE; FREDERIK NAGEL; NIKOLAUS RETTELBACH; STEFAN WABNIK.

(86) Pedido PCT: PCT EP2009004339 de 16/06/2009

(87) Publicação PCT: WO 2010/003521 de 14/01/2010

(85) Data do Início da Fase Nacional: 11/01/2011

(57) Resumo: MÉTODO E DISCRIMINADOR PARA A CLASSIFICAÇÃO DE DIFERENTES SEGMENTOS DE UM SINAL. Para classificar os diferentes segmentos de um sinal de que abrange segmentos de pelo menos, um primeiro tipo e um segundo tipo, por exemplo segmentos de áudio e fala, o sinal é classificado como curto prazo (150) com base em pelo menos, um recurso de curto prazo extraído do sinal e um resultado de classificação de curto prazo (152) é entregue. O sinal também é classificado como longo prazo (154) com base em pelo menos, um recurso de curto prazo e em pelo menos, um recurso de longo prazo extraído do sinal e um resultado de classificação de longo prazo (156) é entregue. O resultado de classificação de curto prazo (152) e o resultado de classificação de longo prazo (156) são combinados (158) para fornecer um sinal de saída (160) indicado se um segmento do sinal é de primeiro tipo ou de segundo tipo.

"MÉTODO E DISCRIMINADOR PARA A CLASSIFICAÇÃO DE DIFERENTES SEGMENTOS DE UM SINAL"

HISTÓRICO DA INVENÇÃO

A invenção relata a abordagem para a classificação de diferentes segmentos de um sinal que abrange os segmentos de pelo menos, um primeiro tipo e um segundo tipo. A materialização da invenção refere-se ao campo da codificação de áudio e, particularmente, para a discriminação de fala/música sobre a codificação de um sinal de áudio.

Na arte, o domínio da codificação de esquemas de frequência, tal como o MP3 ou AAC, são conhecidos. Estes codificadores de domínio de frequência são baseados em uma conversão do domínio de tempo/domínio de frequência, um estágio de quantização subsequente, na qual o erro de quantização é controlado usando a informação de um módulo psicoacústico, e um estágio de codificação, no qual o coeficiente coeficientes espectral quantizado e as informações correspondentes são secundárias a codificação entrópica utilizando as tabelas de códigos

Por outro lado existem os codificadores que são muito bem adequados para o processamento da fala como o AMR-WB+ conforme descrito no 3GPP TS 26.290. Tal esquema de codificação de fala realiza uma analise Linear Preditiva. Tal filtragem LP é derivada de uma analise Linear Preditiva do sinal de entrada do domínio de tempo. Os coeficientes resultantes do filtro LP são então codificados e transmitidos como informação secundária. O processo é conhecido como Codificação Linear Preditiva (LPC). Na saída do filtro, o sinal residual preditivo ou o sinal de erro

preditivo que também é conhecido como o sinal de excitação é codificado usando o estágio de análise-por-síntese do codificador ACELP ou, alternativamente, é codificado utilizando um codificador transformado, que usando uma transformada de Fourier com uma sobreposição. A decisão entre a codificação ACELP e a codificação de Excitação da Transformada Codificada que também é chamada de codificação TCX é feita através de um algoritmo de malha fechada ou um algoritmo de malha aberta.

Os esquemas de codificação de áudio de domínio de frequência tal como os esquemas de codificação de alta eficiência-AAC, que combina um esquema de codificação AAC e uma técnica de replicação de largura de faixa espectral pode também ser combinado com um a joint stereo ou uma ferramenta de codificação de multicanal que também é conhecido como o nome de "MPEG surround". Os esquemas de codificação de áudio de domínio são vantajosos na medida em que mostram uma alta qualidade a baixas taxas de bits para os sinais de música. A qualidade dos sinais de voz em baixas taxas de bits, porém é problemática.

Por outro lado, os codificadores de fala como o AMR-WB+ também possuem um estágio de aprimoramento de alta frequência e uma funcionalidade estéreo. Os esquemas de codificação de fala mostram uma alta qualidade para sinais de fala mesmo em baixas taxas de bits, mas mostram uma baixa qualidade para sinais de música em baixas taxas de bits.

Na visualização disponível de um esquema de codificação acima mencionado, alguns dos quais são mais adequados para codificação de fala e outros sendo mais adequados para codificação de música, a segmentação automática e a classificação

de um sinal de áudio a ser codificado é uma importante ferramenta em varias aplicações multimídia e podem ser utilizadas a fim de selecionar um processo apropriado para cada diferente classe que ocorre em um sinal de áudio. O desempenho geral da aplicação é fortemente dependente da confiabilidade da classificação do sinal de áudio. De fato, uma classificação errada gera seleções mal adaptadas e afinações dos seguintes processos.

A Fig. 6 mostra um design convencional de um codificador usado para codificar separadamente a codificação, dependente de fala e música na discriminação de um sinal de áudio. O design do codificador abrange um codificador de seção de fala 100 inclui um codificador de fala apropriado 102, por exemplo, um AMR-WB+ codificador de fala como descrito na "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec", 3GPP TS 26.290 V6.3.0, 2005-06, Especificação Técnica. Além disso, o design do codificador abrange um codificador de seção de música 104 compreendendo de um codificador de música 106, por exemplo um codificador de música AAC como é, por exemplo, descrito na Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding. International Standard 13818-7, ISO/IEC JTC1/SC29/WG11 Moving Pictures Expert Group, 1997.

As saídas dos codificadores 102 e 106 são conectadas a uma entrada de um multiplexador 108. As entradas dos codificadores 102 e 106 são seletivamente conectadas a uma linha de entrada 110 carregando um sinal de áudio de entrada. O sinal de áudio de entrada é aplicado seletivamente para o codificador de fala 102 ou o codificador de música 106 por meio de um comutador 112 mostrado esquematicamente na Fig. 6 e sendo controlado por um

controle de comutação 114. Além disso, o design do codificador abrange um discriminador de fala/música 116 também recebe uma entrada no seu sinal de áudio de entrada e emite um sinal de controle para o controle de comutação 114. O controle de comutação 114 gera uma saída de um sinal indicador do modo em uma linha de 118 que é a entrada em uma segunda entrada do multiplexador 108, para que um sinal indicador de modo possa ser enviado junto com um sinal codificado. O sinal de indicador de modo pode ter somente um bit indicado que o bloco de dados associados com um bit do indicador de modo ou é para a fala codificada ou música codificada de modo que, por exemplo, em um decodificador nenhuma discriminação deve ser feita. Pelo contrário, com base no bit do indicador de modo apresentado junto com os dados codificados para o decodificador secundário de um sinal de comutação apropriado possa ser gerada com base no indicador de modo de encaminhamento dos dados recebidos e codificados em um decodificador apropriado de fala ou de música.

A Fig. 6 e um design tradicional do codificador que é usado para codificar digitalmente os sinais de fala e música aplicada para a linha 110. Normalmente, os codificadores de fala funcionam melhor na fala e os codificadores de áudio funcionam melhor na música. Um esquema de codificação universal pode ser planejado usando um sistema multi-codificador que alterar de um codificador para outro de acordo com a natureza do sinal de entrada. O problema O problema não trivial aqui é planejar um classificador de sinal de entrada bem adequado que conduz o elemento de comutação. O classificador é o discriminador de fala/música 116 mostrado na Fig. 6. Frequentemente uma

classificação confiável de um sinal de áudio introduz um alto atraso, considerando, por outro lado, o atraso é um fator importante nas aplicações em tempo real.

No geral, é desejado que o atraso do algoritmo geral introduzido pelo discriminador de fala/música seja suficientemente baixo para ser capaz de usar os codificadores ligados na aplicação em tempo real.

A Fig. 7 ilustra os atrasos experimentados design do codificador, como mostrado na Fig. 6. Supõe-se que o sinal aplicado na linha de entrada 110 deve ser codificada em uma base de estrutura de 1024 amostras em uma taxa de amostragem de 16 kHz de modo que o discriminador de fala/música deva emitir um resultado em alguma estrutura, ou seja, a cada 64 milissegundos. A transmissão entre dois codificadores é efetuada, por exemplo, da mesma forma como descrita na WO 2008/071353 A2 e o discriminador de fala/música não deve aumentar significantemente o atraso do algoritmo do decodificador comutado que está no total de 1600 amostras sem considerar o atraso necessário para o discriminador de fala/música. É mais desejada fornecer a decisão de fala/música para a mesma estrutura a comutação de bloco AAC é decidido. A situação é descrita na Fig. 7 ilustrando ao longo comutação de bloco AAC tendo um comprimento de 2048 amostras, ou seja, bloco longo 120 abrange duas estruturas de 1024 amostras, um bloco curto AAC 122 de uma estrutura de 1024 amostras, e um AMR-WB+ superestrutura 124 de uma estrutura de 1024 amostras.

Na Fig. 7, a decisão de comutação de bloco AAC e a decisão de fala/música são tomadas nas estruturas 126 e 128 respectivamente de 1024 amostras, que cobre o mesmo período de

tempo. As duas decisões são tomadas nesta posição em particular para fazer a codificação poder utilizar em um momento da janela de transição para ir adequadamente um modo para o outro. Em consequência, um atraso mínimo de 512+64 amostras são introduzidas por duas decisões. Este atraso tem que ser adicionado ao atraso das 1024 amostras geradas por 50% de sobreposição forma a AAC MDCT que resulta um atraso mínimo de 1600 amostras. Em um AAC convencional, somente a comutação de bloco é apresentado e o atraso é exatamente de 1600 amostras. Este atraso é necessário para comutar em um momento de um bloco longo para os blocos curtos quando os transitórios são detectados na estrutura 126. Esta comutação de comprimento de transformação é desejado para evitar o artefato de pré-eco. A estrutura decodificada 130 na Fig. 7 representa a primeira estrutura total que pode ser restituída no decodificador secundário em qualquer caso (blocos longos ou curtos).

Em um codificador comutado utilizando o AAC como codificador de música, a decisão de comutação vindo de um estágio deve evitar adicionar também muito atrasos adicionais ao atraso original do AAC. O atraso adicional vem de uma estrutura lookahead 132 que é necessária para analise de sinal no estágio de decisão. Em uma taxa de amostragem de por exemplo 16kHz, o atraso AAC é de 100 ms quando o discriminador convencional de fala/música usa cerca de 500 ms de lookahead, que resultará em uma estrutura de codificação comutada com um atraso de 600 ms. O atraso total será seis vezes maior do que o atraso do AAC original.

As abordagens convencionais como as descritas acima são desfavoráveis. Como uma classificação confiável de um

sinal de áudio elevado, os atrasos indesejáveis são introduzidos de modo que a necessidade de uma nova abordagem exista para a discriminação de um sinal incluindo segmentos de diferentes tipos, onde um atraso adicional de algoritmo introduzido pelo discriminador seja suficientemente baixa de modo que os codificadores de comutação também possa ser usado para uma aplicação em tempo real.

J. Wang, et. al. "Real-time speech/music classification with a hierarchical oblique decision tree", ICASSP 2008, Conferência Internacional IEEE sobre Acústica, Fala e Processamento de Sinal, 2008, de 31 de março de 2008 a 4 de abril de 2008 descreve uma abordagem para a classificação de fala/música utilizando recursos a curto e longo prazo derivados de um mesmo número de estruturas. Estes recursos a curto e longo prazo são usados para classificar o sinal, mas apenas as propriedades limitadas dos recursos de curto prazo são explorados, por exemplo, a reatividade da classificação não é explorada, embora tenha um papel importante para a maioria das aplicações de codificação de áudio.

RESUMO DA INVENÇÃO

A finalidade da invenção é fornecer uma melhor abordagem para a discriminação em um segmento de sinal de tipo diferente, mantendo qualquer atraso baixo introduzido pela discriminação.

Este finalidade é atingida pelo método da reivindicação 1 e pela discriminação da reivindicação 14.

Uma materialização da invenção fornece um método para classificar diferentes segmentos de um sinal, o sinal

abrangendo os segmentos de pelo menos, um primeiro tipo e um segundo tipo, o método abrange:

classificação de curto prazo do sinal com base em pelo menos, um recurso de curto prazo extraído do sinal e entregando um resultado de classificação de curto prazo;

classificação de longo prazo do sinal com base em pelo menos, um recurso de curto e pelo menos, um recurso de longo prazo extraído do sinal e entregando um resultado da classificação de longo prazo; e

combinando o resultado da classificação de curto prazo e o resultado da classificação de longo prazo para fornecer um sinal de saída indicando se um segmento do sinal é do primeiro tipo ou do segundo tipo.

Outra materialização da invenção proporciona um discriminador, abrangendo:

um classificador de curto prazo configurado para receber um sinal e fornecer um resultado de classificação de curto prazo do sinal com base em pelo menos, um recurso de curto prazo extraído do sinal, o sinal abrange segmentos de pelo menos, um primeiro tipo e de um segundo tipo;

um classificador de longo prazo configurado para receber um sinal e fornecer um resultado de classificação de longo prazo do sinal com base em pelo menos, um recurso de curto prazo do sinal e pelo menos, um recurso de longo prazo extraído do sinal;

um circuito de decisão configurado para combinar o resultado de classificação de curto prazo e o resultado de classificação de longo prazo para fornecer um sinal de saída

indicando se um segmento do sinal é do primeiro tipo ou do segundo tipo.

A materialização de invenção fornece um sinal de saída com base na comparação do resultado da analise de curto prazo para o resultado da analise de longo prazo.

A materialização de invenção relaciona uma abordagem para classificar os diferentes segmentos não-sobreposição de curto espaço de tempo de um sinal de áudio, quer como fala ou como não-fala ou outras classes. A abordagem é baseada na extração de recursos e a analise de suas estatísticas de duas diferentes de análises de comprimentos de janela. A primeira janela é longa e principalmente para o passado. A primeira janela é usada para obter um indício de decisão confiável mas atrasada para a classificação de um sinal. A segunda janela é curta e considera principalmente o processo de segmento no momento presente ou no segmento atual. A segunda janela é usada para obter um indício de decisão instantânea. As duas dicas de decisão são combinadas de modo mais eficiente, preferencialmente por meio de uma decisão de histerese que obtém a informação da memória a partir do indício de decisão atrasada e a informação instantânea a partir da instantânea.

As materializações de uma invenção usam recursos de curto prazo ambos no classificador de curto prazo e no classificador de longo prazo de modo que os dois classificadores explorem estatísticas diferentes do mesmo recurso. O classificador de curto tempo extrai somente a informação instantânea uma vez que ele tem acesso apenas a um conjunto de recursos. Por exemplo, ele pode explorar o meio dos recursos. Por outro lado, o classificador

de longo prazo tem acesso a vários conjuntos de recursos uma vez que ele considera varias estruturas. Como consequência, o classificador de longo prazo pode explorar mais características do sinal ao explorar estatísticas de mais estruturas que o classificador de curto prazo. Por exemplo, o classificador de longo prazo pode explorar a variação do recurso ou a evolução dos recursos todo tempo. Assim, o classificador de longo prazo pode explorar mais informações que o classificador de curto prazo, mas introduz atraso ou latência. Entretanto, os recursos de longo prazo, apesar de introduzir o atraso ou a latência, fará o resultado de classificação de longo prazo mais robusto e confiável. Em algumas materializações os classificadores de curto prazo e de longo prazo podem considerar os mesmos recursos de curto prazo, que podem ser calculados uma vez e utilizados para ambos os classificadores. Assim, em tal materialização o classificador de longo prazo pode receber recursos de curto prazo diretamente a partir do classificador de curto prazo.

A nova abordagem permite, assim, obter uma classificação que é robusta, introduzindo um atraso baixo. Outras abordagens convencionais, a materialização da invenção limita o atraso introduzido pela decisão de fala/música que mantinha uma decisão confiável. Em uma materialização da invenção, o lookahead é limitado a 128 amostras, o que resulta em um atraso de somente 108 ms.

BREVE DESCRIÇÃO DOS DESENHOS

A materialização da invenção será descrita abaixo com a referência acompanhada de desenhos, no qual:

Fig. 1 é um diagrama de bloco de um

discriminador de fala/música de acordo com uma materialização da invenção;

Fig. 2 ilustra a janela de analise usada pelo classificador de longo e curto prazo do discriminador da Fig. 1;

Fig. 3 ilustra a decisão de histerese utilizada no discriminador da Fig. 1;

Fig. 4 é um diagrama de bloco de um esquema exemplar de codificação abrangendo um discriminador de acordo com uma materialização da invenção;

Fig. 5 é um diagrama de bloco de um esquema de decodificação correspondente ao esquema de codificação da Fig. 4;

Fig. 6 mostra um design convencional de codificador usado para codificar separadamente o dependente de fala e música em uma discriminação de um sinal de áudio; e

Fig. 7 ilustra os atrasos experimentado no design do codificador mostrado na Fig. 6.

DESCRIÇÃO DETALHADA

Fig. 1 é um diagrama de bloco de um discriminador de fala/música 116 de acordo com uma materialização da invenção. O discriminador de fala/música 116 abrange um classificador de curto prazo 150 recebe na entrada um sinal de entrada, por exemplo, um sinal de áudio abrangendo os segmentos de fala e música. O classificador de curto prazo 150 emite na linha de saída 152 um resultado de classificação de curto prazo, o indício de decisão instantânea. O discriminador 116 abrange ainda um classificador de longo prazo 154 que também recebe um sinal de entrada e saída em uma linha de saída 156 o resultado de classificação de longo prazo e o indício de decisão atrasada. Além disso, um circuito de

decisão de histerese 158 é fornecido que combina os sinais a partir do classificador de curto prazo 150 e do classificador de longo prazo 154 será descrito de modo mais detalhada abaixo para gerar um sinal decisão de fala/música que é a saída na linha 160 e pode ser usada para controlar o processo posterior de um segmento de uma sinal de saída do modo como está descrito acima com relação a Fig. 6, ou seja o sinal de decisão de fala/música 160 pode ser usado para rotear o segmento do sinal de entra que tem sido classificado para um codificador de fala ou para um codificador de áudio.

Assim, de acordo com uma materialização da invenção dois diferentes classificadores 150 e 154 são usados em paralelo no sinal de entrada aplicado para os respectivos classificadores por meio de uma linha 110. Os dois classificadores são chamados de classificador de longo prazo 154 e classificador de curto prazo 150, onde o em que os dois classificadores diferentes, analisando as estatísticas das características em que a operação sobre as janelas de análise. Os dois classificadores entregam os sinais de saída 152 and 156, nomeados de indício de decisão instantâneo (IDC) e o indício de decisão atrasada (DDC). O classificador de curto prazo 150 gera o IDC com base nos recursos de curto prazo que têm o objetivo de capturar informações instantâneas sobre a natureza do sinal de entrada. Eles estão relacionados com atributos de curto prazo do sinal que podem alterar rapidamente a qualquer momento. Em consequência os recursos de curto prazo deverão ser reativados e não introduzir um atraso longo de todo o processo de discriminação. Por exemplo, desde que a fala é considerado quase estacionária com duração de

5-20ms, os recursos de curto prazo podem ser calculado em cada estrutura de 16 ms em um sinal de amostra de 16 kHz. O classificador de longo prazo 154 gera o DDC com base nos recursos resultantes a partir de longas observações do sinal (recursos de longo prazo) e, portanto, permite alcançar a classificação mais confiável.

A Fig. 2 ilustra a janela de analise usada pelo classificador de longo prazo 154 e pelo classificador de curto prazo 150 mostrado na Fig. 1. Assumindo uma estrutura de 1024 amostras em uma taxa de amostragem de 16 kHz o comprimento da janela do classificador de longo prazo 162 é de 4*1024+128 amostras, ou seja, a janela do classificador de longo prazo 162 transpõe quatro estruturas do sinal de áudio e as 128 amostras adicionais são necessárias pelo classificador de longo prazo 154 para fazer esta analise. Este atraso adicional, que é também referido como um "lookahead", é indicado na Fig. 2 no sinal de referencia 164. A Fig. 2 também mostra a janela do classificador de curto prazo 166 que é 1024+128 amostras, ou seja transpõe uma estrutura do sinal de áudio e o atraso adicional necessário para analisar o segmento atual. O segmento atual é indicado em 128 como o segmento para o qual a decisão de fala/música precisa ser feita.

A janela do classificador de longo prazo indicada na Fig. 2 é suficientemente longa para obter os 4-Hz da modulação de energia da característica da fala. Os 4-Hz da modulação de energia são uma característica relevante e distinta da fala que é tradicionalmente explorada em um robusto discriminador de fala/músicas usadas como por exemplo por Scheirer E. e Slaney M., "Construction and Evaluation of a Robust Multifeature Speech/Music

Discriminator", ICASSP'97, Munich, 1997. Os 4-Hz da modulação de energia são um recurso que pode ser somente extraído pela observação de um sinal em um longo segmento de tempo. O atraso adicional que é introduzido pelo discriminador de fala/música é igual ao lookahead 164 de 128 amostras que é necessário para cada um dos classificadores 150 e 154 fazem a respectiva análise, como uma analise perceptiva linear preditiva como é descrito por H. Hermansky, "Perceptive linear prediction (plp) analysis of speech," Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738-1752, 1990 e H. Hermansky, et al., "Perceptually based linear predictive analysis of speech," ICASSP 5.509-512, 1985. Assim, quando usamos o discriminador da materialização acima em um design de codificador como mostrado na Fig. 6, o atraso total dos codificadores de comutação 102 e 106 serão 1600+128 amostras que é 108 milissegundos que é suficientemente baixo para aplicações em tempo real.

A referência é agora feita para a Fig. 3 descrevendo a combinação do sinal de saída 152 e 156 dos classificadores 150 e 154 do discriminador 116 para obter um sinal de decisão de fala/música 160. O indício de decisão atrasada DDC e o indício de decisão instantânea IDC, de acordo com uma materialização da invenção, é combinado ao usar uma decisão de histerese. Os processos de histerese são amplamente utilizados para divulgar decisões processo a fim de estabilizá-los. A Fig. 3 ilustra uma decisão de dois estados de histerese como uma função do DDC e do IDC para determinar se o sinal decisão de fala/música indicar um segmento atualmente processado do sinal de entrada como sendo um segmento de fala ou de um segmento de música. Os ciclos

de características da histerese é visualizado na Fig. 3 e o IDC e o DDC são normalizados pelos classificadores 150 e 154 de tal forma que os valores estão entre -1 e 1, onde -1 significa que a probabilidade é totalmente semelhante à música, e 1 significa que a probabilidade é totalmente semelhante à fala.

A decisão é baseada nos valores de uma função F(IDC,DDC), esses exemplos que serão descritos abaixo. Na Fig. 3, F1(DDC, IDC) indica um limite que F(IDC,DDC) deve atravessar para ir do estado de música para o estado de fala. A F2(DDC,IDC) indica um limite que F(IDC,DDC) deve atravessar para ir do estado de fala para o estado de música. A decisão final D(n) para um segmento atual ou estrutura atual tendo o índice n, pode então ser calculada com base no seguinte pseudocódigo:

```
%Hysteresis Decision Pseudo Code
If(D(n-1)==music)
    If(F(IDC,DDC)<F1(DDC,IDC))
        D(n)==music
    Else
        D(n)==speech
Else
    If(F(IDC,DDC)>F2(DDC,IDC))
        D(n)==speech
    Else
        D(n)==music
%End Hysteresis Decision Pseudo Code
```

De acordo com uma materialização da invenção a função F(IDC,DDC) e o limite acima mencionado, são definidas a seguir:

F(IDC,DDC)=IDC

F1(IDC,DDC)=0.4-0.4*DDC

F2(IDC,DDC)=-0.4-0.4*DDC

Alternativamente, as seguintes definições podem ser usadas:

F(IDC,DDC)=(2*IDC+DDC)/3

F1(IDC,DDC)=-0.75*DDC

F2(IDC,DDC)=-0.75*DDC

Quando usamos a ultima definição do ciclo de histerese e a decisão é feita somente com base no limite de uma única adaptativa.

A invenção não é limitada pela decisão de histerese descrita acima. Nas materializações seguintes adicionais, será descrito que, combinamos os resultados da análise para a obtenção do sinal de saída.

Um limite simples pode ser usado no lugar da decisão de histerese fazendo de uma forma que o limite explore as características da DDC e IDC. O DDC é considerado como o indício discriminante mais confiável, uma vez que se a trata da observação mais demorada do sinal. Entretanto, o DDC é calculado parcialmente com base em uma observação anterior do sinal. Um classificador convencional que somente compara o valor DDC para o limite 0, e pela classificação do segmento como semelhante à fala quando DDC>0 ou ao contrario, como semelhante à música, temos uma decisão de atraso. Em uma materialização da invenção, podemos adaptar o limite explorando o IDC e tomar a decisão mais reativa. Para este propósito, o limite pode ser adaptado com base no seguinte pseudocódigo:

```
% Pseudo code of adaptive thresholding
If(DDC>-0.5*IDC)
    D(n)==speech
Else
    D(n)==music
%End of adaptive thresholding
```

Em outra materialização, o DDC pode ser usado para tornar o IDC mais confiável. O IDC é conhecido por ser reativo mas não tão confiável quanto o DDC. Além disso, observando a evolução do DDC entre o segmento anterior e o atual pode dar mais uma indicação de como a estrutura 166 na Fig. 2 influencia o DDC calculado no segmento 162. A nota DDC(n) é usada para o valor atual do DDC e DDC(n-1) para o valor. Utilizando ambos os valores, DDC(n) e DDC(n-1), o IDC pode ser mais confiável usando uma árvore de decisão como é descrito a seguir:

```
% Pseudo code of decision tree
If(IDC>0 && DDC(n)>0)
    D(n)=speech
Else if (IDC<0 && DDC(n)<0)
    D(n)=music
Else if (IDC>0 && DDC(n)-DDC(n-1)>0)
    D(n)=speech
Else if (IDC<0 && DDC(n)-DDC(n-1)<0)
D(n)=music
Else if (DDC>0)
    D(n)=speech
Else
D(n)=music
```

%End of decision tree

Na árvore de decisão acima, a decisão é tomada diretamente se ambas as dicas mostrarem o mesmo valor. Se as duas dicas dão indicações contraditórias, observamos para a evolução da DDC. Se a diferença de DDC(n)-DDC(n-1) é positiva, podemos supor que o segmento atual é semelhante à fala. De outra maneira, podemos supor que o segmento atual é semelhante à música. Se esta nova indicação vai na mesma direção do IDC, a decisão final é tomada. Se ambas as tentativas falham ao dar uma decisão clara, a decisão é tomada por considerar somente o atraso no indício DDC desde que a confiabilidade do IDC não possa ser validada.

No seguinte, os respectivos classificadores 150 e 154 de acordo com uma materialização da invenção serão descritos detalhadamente.

Começando pelo primeiro lugar o classificador de longo prazo 154 é o mesmo que se aplica para cada subestrutura de 256 amostras em um conjunto de recursos. O primeiro recurso é o Coeficiente Cepstral de Perceptiva Linear Preditiva (PLPCC) como descrito por H. Hermansky, "Perceptive linear prediction (plp) analysis of speech," Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738-1752, 1990 e H. Hermansky, et al., "Perceptually based linear predictive analysis of speech," ICASSP 5.509-512, 1985. Os PLPCCs são eficientes para classificação de fala ao utilizar a avaliação da percepção auditiva humana. Este recurso pode ser usado para discriminar a fala e a música e, realmente permite as características dos formantes da fala, bem como a modulação silábica da fala de 4 Hz, observando a variação do recurso ao longo do tempo.

Entretanto, para ser mais robusto, os PLPCCs são combinados com outro recurso que é capaz de capturar tom das informações, que é outra característica importante da fala e pode ser critica na codificação. Realmente, a codificação da fala baseia-se na suposição que um sinal de saída é um sinal pseudo mono-periódico. Os esquemas de codificação da fala são eficientes para tal sinal. Por outro lado, as características do tom da fala prejudica muitos a eficiência da codificação dos codificadores de música. A flutuação do atraso de tom suave determina o vibrato natural da fala faz com que a representação de frequência nos codificadores de música sejam incapazes de compactar a energia grande que é necessária para a obtenção de uma alta eficiência de codificação.

Os seguintes recursos das características do tom podem ser determinadas como:

Taxa de Energia dos Pulsos Glótico:

Este recurso calcula a taxa de energia entre os pulsos glóticos e o sinal residual de LPC. Os pulsos glóticos são extraídos do sinal residual de LPC utilizando um algoritmo pick-peaking. Geralmente, o residual de LPC de um segmento sonoro mostra uma grande estrutura semelhante a pulso vindo da vibração glótica. O recurso é alto durante os segmentos sonoros.

Ganho Perceptivo de Longo Prazo:

É o ganho geralmente calculado nos codificadores de fala (ver exemplos "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec", 3GPP TS 26.290 V6.3.0, 2005-06, Especificação Técnica) durante o perceptivo de longo prazo. Este recurso mede a periodicidade do sinal e é baseado no atraso estimativo do tom.

Flutuação do atraso de tom:

Este recurso determina a diferença do atraso estimativo do tom presente quando comparado a ultima sub-estrutura. Para o vozeamento da fala este recurso deve ser baixo mas não zero e evolui suavemente.

Uma vez que o classificador de longo prazo tem extraído o conjunto requerido de recursos, um classificador estático é usado para extrair estes recursos. O classificador é primeiro treinado extraindo os recursos em um conjunto de treinamento de fala e conjunto de treinamento de música. Os recursos extraídos são normalizados para um valor médio de 0 e uma variação de 1 em ambos os conjuntos de treinamento. Para cada conjunto de treinamento, os recursos extraídos e normalizados são reunidos dentro de uma janela do classificador de longo prazo e modelados pelo Gaussians Mixture Model (GMM) usando cinco gaussianos. Ao fim da sequência de treinamento um conjunto de parâmetros de normalização e dois conjuntos de parâmetros GMM são obtidos e salvos.

Para cada estrutura para classificar, os recursos são extraídos primeiros e normalizados com os parâmetros de normalização. A semelhança máxima para a fala (lld_speech) e a t semelhança máxima para a música (lld_music) são calculadas para os recursos extraídos e normalizados usando o GMM de classe de fala e o GMM de classe de música, respectivamente. O indício de decisão atrasada DDC é então calculada pela seguinte:

DDC=(lld_speech-
lld_music)/(abs(lld_music)+abs(lld_speech))

O DDC está vinculado entre -1 e 1, e é positivo

quando a semelhança máxima para a fala seja maior que a semelhança máxima para a música, lld_speech>lld_music.

O classificador de curto prazo utiliza como recurso de curto prazo o PLPCCs. Exceto no classificador de longo prazo, este recurso é somente analisado na janela 128. As estatísticas neste recurso são extraídas neste curto período por um Gaussians Mixture Model (GMM) usando cinco gaussianos. Os dois modelos são treinados, um para música, e outro para fala. Vale a pena notificar, que os dois modelos são diferentes daqueles obtidos pelo classificador de longo prazo. Para cada estrutura para classificar, os PLPCCs são extraídos primeiro e a semelhança máxima para a fala (lld_speech) e a semelhança máxima para a música (lld_music) são calculados usando o GMM de classe de fala e a GMM de classe de música, respectivamente. O indício de decisão instantânea IDC e então calculada a seguir:

IDC=(lld_speech-lld_music)/(abs(lld_music)+abs(lld_speech))

O IDC é variável entre -1 e 1.

Assim, o classificador de curto prazo 150 gera o resultado de classificação de curto prazo do sinal com base no recurso "Coeficiente Cepstral de Perceptiva Linear Preditiva (PLPCC)", e o classificador de longo prazo 154 gera o resultado de classificação de longo prazo do sinal com base no mesmo recurso "Coeficiente Cepstral de Perceptiva Linear Preditiva (PLPCC)" e o(s) recurso(s) adicional(s) acima mencionado(s), por exemplo, o(s) recurso(s) da característica(s) do tom. Além do mais, o classificador de longo prazo pode explorar diferentes características do recurso compartilhado, por exemplo, o PLPCCs,

tem como acesso uma janela de observação mais longa. Assim, a combinação dos resultados de curto e longo prazo, os recursos de curto prazo são considerados suficientemente para a classificação, por exemplo, suas propriedades são suficientemente exploradas.

Abaixo uma materialização para os respectivos classificadores 150 e 154 serão descritos de um modo mais detalhado.

Os recursos de curto prazo analisados pelo classificador de curto prazo de acordo com esta materialização corresponde principalmente ao Coeficiente Cepstral de Perceptiva Linear Preditiva (PLPCCs) mencionado acima. Os PLPCCs são amplamente usados na fala e no reconhecimento da fala assim como os MFCCs (ver acima). Os PLPCCs são retidos uma vez que eles compartilham uma grande parte da funcionalidade da Linear Preditiva (LP) que é usado no mais moderno codificador de fala e assim implementado em um codificador de áudio ligado. O PLPCCs pode extrair a estrutura de formantes da fala como o LP faz, mas levando em conta as considerações perceptivas, o PLPCCs tem mais falantes independentes e portanto, mais relevantes relativos a informação linguística. Uma ordem de 16 é usada na amostra do sinal de entrada de 16 kHz.

Além dos PLPCCs, uma força de vozeamento é calculada como um recurso de curto prazo. A força de vozeamento não é considerado para realmente ser discriminada por si, mas é benéfico na associação com a PLPCCs na dimensão de recursos. A força de vozeamento permite atrair a dimensão de recurso pelo menos, dois grupos correspondentes, respectivamente, para pronuncias de fala de vozeamento e não vozeadas. É baseado em um

calculo de mérito usando diferentes Parâmetros, isto é um Contador de cruzamento por zero, inclinação espectral (tilt), a estabilidade do tom (ps), e a correlação normalizada di tom (nc). Todos os quatro parâmetros são normalizados entre 0 e 1 de maneira que o 0 corresponda ao sinal não sonoro e 1 corresponda a um sinal tipicamente sonoro. Nesta materialização a força de vozeamento é inspirado nos critérios de classificação de fala utilizados no VMR-WB codificador de fala descrito por Milan Jelinek e Redwan Salami, "Wideband speech coding advances in vmr-wb standard," IEEE Trans. on Audio, Speech and Language Processing, vol. 15, no. 4, pp. 1167-1179, maio de 2007. É baseado em um evolução do rastreador de tom baseado na auto-correlação. Para o índice de estrutura k a força de vozeamento u(k) tem a forma abaixo:

$$v(k)=\frac{1}{5}(2*nc(k)+2*ps(k)+tilt(k)+zc(k))$$

A capacidade de discriminação de recursos de curto prazo é avaliada pela Gaussian Mixture Models (GMMS) como um classificador. Dois GMMs, um para a classe de fala e outro para a classe de música, são aplicados. Um número de mesclas são feitas apresentando variações a fim de avaliar o efeito no desempenho. A tabela 1mostra a taxa de precisão pás os diferentes números de mesclas. Uma decisão é calculada para cada segmento de quatro estruturas sucessivas. O atraso total é então igual a 64ms que é adequado para um codificador comutado de áudio. Pode ser observado que o desempenho aumenta com o número de mesclas. O intervalo entre 1-GMMs e 5-GMMs é particularmente importante e pode ser explicado pelo fato de que a representação dos formantes da fala é muito complexa para ser suficientemente definida somente por um

gaussiano.

	1-GMMs	5-GMMs	10-GMMs	20-GMMs
Fala	95.33	96.52	97.02	97.60
Música	92.17	91.97	91.61	91.77
Média	93.75	94.25	94.31	94.68

Tabela 1: % de precisão da classificação de recursos de curto prazo

Retorne para o classificador de longo prazo 154, é observado que vários trabalhos, por exemplo, M. J. Carey, et. al. "A comparison of features for speech and music discrimination," Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICASSP, vol. 12, pp. 149 a 152, março de 1999, considera que as variações dos recursos de estatística são mais exigentes do que os próprios recursos. Como uma regra geral, a música pode ser considerada mais fixo e geralmente exibir uma variação baixa. De modo contrario, a fala pode ser facilmente distinguida pela sua excelente energia de modulação de 4-Hz como o sinal que altera periodicamente entre um segmento sonoro e não sonoro. Além disso a sucessão de diferentes fonemas faz o recurso da fala ser menos constante. Nesta materialização, os dois recursos de longo prazo são considerados, um baseado em um cálculo da variância e o outro baseado um conhecimento priori da entonação da fala. Os recursos de longo prazo são adaptados para o atraso baixo SMD (discriminação de fala/música).

A variação de movimento dos PLPCCs consiste da variação do calculo para cada conjunto de PLPCCs sobre uma janela de analise de sobreposição cobrindo varias estruturas a fim de enfatizar a ultima estrutura. Para limitar a latência introduzida, a janela de analise é assimétrica e considera somente a estrutura

atual e o histórico anterior. Em um primeiro etapa, a média em movimento mam(k) dos PLPCCs é calculada sobre a ultima estrutura N como descrita a seguir:

$$ma_m(k)=\sum_{i=0}^{N-1} PLPC_m(k-i)\cdot w(i)$$

onde o PLPm(k) o coeficiente cepstral mth sobre um total dos coeficientes M vindo da estrutura kth. A variação de movimento mvm(k) é então definida como:

$$mv_m(k)=\sum_{i=0}^{N-1}(PLPC_m(k-i)-ma_m(k))^2\cdot w(i)$$

onde w é a uma janela de comprimento N que esta nesta materialização uma inclinação de rampa definida da seguinte forma:

$$w(i)=(N-i)/N\cdot(N+1)/2$$

A variação de movimento é finalmente calculada sobre a dimensão cepstral:

$$mv(k)=\frac{1}{M}\sum_{m=0}^{M} mv_m(k)$$

O tom da fala possui propriedade excelente e parte deles pode somente ser observados na janela longa de analise. Realmente o tom de voz é suavemente instável durante os segmentos sonoros, mas raramente é constante. De modo contrario, a música exibe muito frequentemente o tom constante durante toda a duração de uma nota e altera repentinamente durantes os transientes. Os recursos de longo prazo abrangem esta característica observando a entonação em um segmento de longo período. Um parâmetro de entonação pc(k) é definido como:

$$pc(k)=\begin{cases}0 & \text{se } |p(k)-p(k-1)|<1\\ 0,5 & \text{se } 1\le|p(k)-p(k-1)|<2\\ 1 & \text{se } 2\le|p(k)-p(k-1)|<20\\ 0,5 & \text{se } 20\le|p(k)-p(k-1)|<25\\ 0 & \text{do contrário}\end{cases}$$

onde p(k) é o atraso de tom calculado no índice da estrutura k na amostra de sinal residual LP em 16Hz. A partir do parâmetro de entonação, um mérito da fala, sm(k), é calculado de modo que é esperado que a fala mostre um atraso de tom suavemente instável durante os segmentos sonoros e uma forte inclinação espectral diante de altas frequências durantes os segmentos não sonoros:

$$sm(k)=\begin{cases}nc(k)\cdot pc(k) & \text{se } v(k)\ge 0.5\\ (1-nc(k))\cdot(1-tilt(k)) & \text{do contrário}\end{cases}$$

onde nc(k), inclinação(k), e v(k) são definidos como acima (ver o classificador de curto prazo). O mérito da fala é medido então pela janela w definida acima e integrada sobre as últimas estruturas N:

$$ams(k)=\sum_{i=0}^{N} m(k-1)w(i)$$

A entonação é também uma indicação importante de que o sinal é adequado para um codificador de fala ou áudio. Realmente os codificadores de fala trabalham principalmente no domínio de tempo e fazem a suposição de que o sinal é harmônico e quasi-estacionários nos segmentos de tempo de cerca de 5ms. Desta forma eles podem modelar eficientemente a flutuação do tom natural da fala. De modo contrario, a mesma flutuação prejudica a eficiência geral dos codificadores de áudio que exploram as transformações lineares na janela longa de analise. A energia

principal do sinal é então espalhada sobre vários coeficientes de transformada.

Tanto os recursos de curto prazo quanto os recursos de longo prazo são avaliados usando um classificador estatístico obtendo assim o resultado de classificação de longo prazo (DDC). Os dois recursos são calculados usando as estruturas N = 25 estruturas, por exemplo, considerando o 400 ms do histórico anterior de um sinal. Uma Analise de Discriminantes Lineares (LDA) é primeiramente aplicado usando 3-GMMs no espaço reduzido unidimensional. A tabela 2 mostra o desempenho medido no treinamento e o conjunto de teste quando os segmentos classificados para as quatro estruturas sucessivas.

	Conjunto de Treinamento	Conjunto de Teste
Fala	97.99	97.84
Música	95.93	95.44
Média	96.96	96.64

Tabela 2: de precisão da classificação de recursos de longo prazo

Os sistemas de classificadores combinados de acordo com a materialização da invenção combina apropriadamente os recursos de curto e longo prazo de modo que eles trazem sua contribuição específica para a decisão final. Para este propósito um estágio decisão final de histerese como descrito acima pode ser usado, onde o efeito de memória é direcionado pelo DDC ou o indício discriminante de longo prazo (LTDC) enquanto a saída imediata vem do IDC ou do indício discriminante de curto prazo (STDC). As duas dicas são saídas dos classificadores de longo e curto prazo como ilustrado na Fig. 1. A decisão é tomada com base

no IDC mas é estabilizada pelo DDC que controla dinamicamente os limites que determinam uma mudança de estado.

O Classificador de longo prazo 154 usa ambos os recursos de longo e curto prazo anteriormente definidos com um LDA seguido por 3-GMMs. O DDC é igual a proporção logarítmica de classificador de longo prazo semelhante a classe de fala e a classe de musica calculada obre a ultima estrutura 4 X K. O numero das estruturas levadas em conta pode variar com o parâmetro K a fim de adicional mais ou menos efeito de memória na decisão final. De modo contrario, o classificador de curto prazo utiliza somente recursos de curto prazo com os 5-GMMs que mostram um bom compromisso entre o desempenho e complexidade. O IDC é igual a proporção logarítmica do classificador de curto prazo semelhante a classe de fala e a classe de musica calculada somente sobre as ultimas 4 estruturas.

A fim de avaliar a abordagem inventiva, especialmente par um codificador comutado de áudio, três diferentes tipos de desempenho foram avaliados. Uma primeira medição de desempenho e a fala convencional contra o desempenho da musica (SvM). É avaliado em mais de um grande conjunto de musicas e itens de fala. Uma segunda medição de desempenho é feita com um grande e único item que possui segmentos de fala e musica alternando a cada 3 segundos. A precisão de discriminação é então chamada de desempenho fala antes/depois da musica (SabM) e reflete principalmente a reatividade do sistema Finalmente, a estabilidade da decisão é avaliada pelo desempenho da classificação em um grande conjunto de musicas e itens de fala. A mescla entre fala e musica é feito em níveis diferentes a partir de um item para

outro. O desempenho da fala sobre a musica (SoM) é então obtido pelo calculo da proporção da comutação de classe de numero que ocorrem sobre o numero total de estruturas.

O classificador de longo e curto prazo são usados como referencias para avaliação da abordagem do classificador simples convencional. O classificador de curto prazo mostra uma boa reatividade quando tem baixa estabilidade e a capacidade de discriminação em geral. Por outro lado, o classificador de longo prazo, especialmente por meio do aumento do número de estruturas 4 X K, pode alcançar uma melhor estabilidade e comportamento discriminatório por comprometer a reatividade da decisão. Quando comparado com a abordagem convencional que acabamos de mencionar, o desempenho do sistema classificador combinado de acordo com a invenção tem várias vantagens. Uma vantagem é que ele mantém uma boa fala pura contra um desempenho de discriminação de música enquanto preserva a reatividade do sistema. Uma outra vantagem é a boa troca entre reatividade e estabilidade.

No seguinte, a referencia é feita para as Figs. 4 e 5 ilustrando os esquemas de codificação e decodificação exemplar que incluem um a discriminador ou estágio de decisão operando de acordo com uma materialização da invenção.

De acordo com os esquemas de codificação exemplar na Fig. 4 um sinal mono, um sinal estéreo ou um sinal multicanal sinal é a entrada em um estágio de pré-processamento comum 200.

O estágio de pré-processamento comum 200 pode ter uma funcionalidade joint stereo, uma funcionalidade surround, e/ou uma funcionalidade de extensão de largura de banda. Na saída de estágio 200 existe um canal mono, um canal estéreo ou canais

múltiplos que é a saída de entrada em um ou mais comutadores 202. O comutador 202 pode ser fornecido para cada saída de estágio 200, quanto o estágio 200 possui duas ou mais saídas, por exemplo, quando as saídas do estágio 200 possuem um sinal estéreo ou um sinal de multicanal. De modo exemplar, o primeiro canal de um sinal estéreo pode ser um canal de fala e o segundo canal de um sinal estéreo pode ser um canal de musica. Neste caso, a decisão em um estágio de decisão 204 pode ser diferente entre os dois canais ao mesmo tempo.

O comutador 202 é controlado pelo estágio de decisão 204. O estágio de decisão é composto com um discriminador de acordo com uma materialização da invenção e recebe, como um sinal de entrada, um sinal dentro do estágio 200 ou um sinal de saída pelo estágio 200. De forma alternativa, o estágio de decisão 204 pode também receber uma informação secundaria que é incluída no sinal mono, no sinal estéreo ou no sinal multicanal ou é pelo menos, associada com tal sinal, onde a informação é existente, que esta, por exemplo, gerada quando inicialmente é produzido o sinal mono, o sinal estéreo ou o sinal multicanal.

Em uma materialização, o estágio de decisão não controla o estágio de pré-processamento 200, e a seta entre o estágio 204 e 200 não existe. Em outra materialização, o processo no estágio 200 é controlado até um certo grau pelo estágio de decisão 204 a fim de definir um ou mais parâmetros no estágio 200 com base na decisão. Isto, porém não influencia o algoritmo geral de 200 estágio de modo que as principais funcionalidades do estágio 200 está ativa, independentemente da decisão no estágio 204.

O estágio de decisão 204 aciona o comutador 202 a fim de alimentar a saída do estágio de pré-processamento comum ou em uma porção de codificação de frequência 206 ilustrada na seção superior da Fig. 4 ou um domínio LPC- codificando a porção 208 ilustrada na seção inferior da Fig. 4.

Em uma materialização, o comutador 202 altera em duas seções codificadas 206, 208. Em outra materialização, pode existir seções codificadas adicionais com uma terceira seções codificadas, ou uma quarta seções codificadas ou até mesmo muitas seções codificadas. Em uma materialização com três seções codificadas, a terceira seções codificadas pode ser idêntica a segunda seções codificadas, mas inclui uma codificação de excitação diferentes para a codificação de excitação 210 na segunda seção 208. Tal como a materialização, a segunda seção abrange O LPC estágio 212 e o codebook é baseado no codificador de excitação 210 tal como no ACELP, e a terceira seção abrange um estágio LPC e um codificador de excitação operando a representação do sinal de saída do estágio.

A frequência de domínio da secção de codificação abrange um bloco de conversão espectral 214 que é operativo para converter o sinal de saída do estágio de pré-processamento comum dentro do domínio do espectro. O bloco de conversão espectral pode incluir um algoritmo MDCT, um QMF e um algoritmo FFT, a analise de Wavelet ou um banco de filtro, tal como os bancos de filtro criticamente amostrados possui um certo numero de canais de banco de filtro, onde o sinal de sub-banda neste banco de filtro pode ser o sinal real valorizado ou o sinal complexo valorizados. A saída do bloco de conversão espectral 214 é codificada usando um

codificador de áudio espectral 216, que pode incluir blocos de processamento tal como é conhecido a partir do esquema de codificação AAC.

A seção codificada baixa 208 é composta de um analisador de modelo de origem como LPC 212, que gera dois tipos de sinais. Um sinal é um sinal de informação LPC, que é usado para controlar a característica do filtro de síntese filtro sintetizador LPC. Esta informação LPC é transmitida por um decodificador. O outro sinal de entrada o do estágio 212 LPC é um sinal de excitação ou um sinal de domínio LPC, que é de entrada em um codificador de excitação 210. O codificador de excitação 210 pode vir de qualquer codificador modelo fonte-filtro como um codificador CELP, um codificador ACELP ou qualquer outro codificador, que processa um sinal de domínio LPC.

Outra implementação do codificador de excitação pode ser uma codificação de transformada do sinal de excitação. Em tal materialização, o sinal de excitação não é codificado usando um mecanismo de codebook ACELP, mas o sinal de excitação é convertido em uma representação espectral e os valores representação espectral tais como sinais de sub-bandas em caso de banco de filtro ou coeficientes de frequência no caso de uma transformação como uma FFT são codificados para obter uma compressão de dados. Uma implementação deste tipo de codificador de excitação é o modo de codificação conhecido como AMR-WB+.

A decisão no estágio de decisão 204 pode ser um sinal adaptativo de modo que o estágio de decisão 204 desenvolve uma discriminação da musica/fala e controla o comutador 202 de tal modo que os sinais de música estão inseridos na seção superior

206, e os sinais de fala são inseridos na seção inferior 208. Em uma materialização, o estágio de 204 abastece suas informações de decisão em um fluxo de bits de saída, de modo que um decodificador pode usar essa informação de decisão, a fim de executar as operações de decodificação correta.

Tais um decodificador é ilustrado na Fig. 5. Após a transmissão, o sinal de saída do codificador de áudio espectral 216 é a entrada em um decodificador espectral de áudio 218. A saída do decodificador de áudio espectral 218 é a entrada em um conversor de domínio do tempo 220. A saída do codificador de excitação 210 da Fig. 4 é de entrada em um decodificador de excitação 222, que gera um sinal de domínio LPC. O sinal de domínio LPC é a entrada em um estágio de síntese LPC 224, que recebe, como uma entrada, as informações LPC geradas pela analise de estágio 212 correspondente. A saída do conversor de domínio do tempo 220 e/ou a saída do estágio de síntese LPC 224 é a entrada em um comutador 226. O comutador 226 é controlado por meio de um sinal de controle do comutador, que foi, por exemplo, gerado pelo estágio de decisão 204, ou que tenham sido fornecidos externamente, como por um criadór do sinal mono original, sinal estéreo ou sinal multicanal.

A saída do comutador 226 é um sinal mono completo que é subsequentemente a entrada em um estágio de pós-processamento de 228, o que pode realizar um processamento joint stereo ou uma extensão da largura de banda, etc. De modo alternativo a saída do comutador também pode ser um sinal estéreo ou um sinal multicanal. É um sinal estéreo, quando o pré-processamento inclui um canal de redução para dois canais. Pode

até ser um sinal de multicanal, quando uma redução de canal para três canais ou nenhuma redução de canal em todos, mas somente uma replicação de faixa espectral é realizada.

Dependendo das funcionalidades específicas do estágio de pós-processamento comum, um sinal mono, um sinal estéreo ou um sinal de multicanal é emitido, que tem, quando o estágio de pós-processamento 228 executa uma operação de extensão de banda larga, uma largura de banda maior do que o sinal de entrada no bloco 228.

Em uma materialização, o comutador 226 alterna entre as duas seções de decodificação 218, 220 e 222, 224. Em outra materialização, pode haver outras seções de decodificação adicionais, como uma terceira seção de decodificação, ou mesmo uma quarta seção de decodificação ou até mesmo mais seções de decodificação. Em uma materialização com três seções de decodificação, a terceira seção de decodificação pode ser semelhante a segunda seção de decodificação, mas inclui um decodificador de excitação diferente do decodificador excitação 222 segunda seção 222, 224. Em tal materialização, segunda seção composta de um estágio LPC 224 e um codebook com base no decodificador de excitação como em um ACELP, e a terceira seção composta de um estádio LPC e um decodificador de excitação operando uma representação espectral do sinal de saída do estágio 224 LPC fase.

Em outra materialização, o estágio de pré-processamento comum composto de um bloco surround/estéreo, que gera, como saída, os parâmetros joint stereo e um sinal de saída mono, que é gerado pelo downmixing do sinal de entrada, que é um

sinal que possui dois ou mais canais. Normalmente, o sinal de saída do bloco pode também pode ser um sinal de que possui mais canais, mas devido à operação downmixing, o número de canais para a saída do bloco será menor do que o número de canais de entrada no bloco. Nesta materialização, a seção de codificação de frequência composta de um estágio de conversão de espectro e um estágio de quantização/codificação subsequentemente conectadas. O estágio de quantização/codificação pode incluir qualquer das funcionalidades como é conhecido desde os modernos codificadores no domínio da frequência, como o codificador AAC. Além disso, a operação do estágio de quantização/codificação pode ser controlada por meio de um módulo de psicoacústica, que gera informações psicoacústicas, como um mascaramento psicoacústico do limite sobre a frequência, onde essa informação é a entrada no estágio. De preferência, a conversão espectral é feita usando uma operação de MDCT que, preferencialmente, é a operação MDCT time-warped, onde a força ou, em geral, a força de deformação pode ser controlada entre zero e uma alta força de deformação. Em uma força deformação zero, a operação de MDCT é uma operação de MDCT direta conhecido na arte. O codificador de domínio LPC pode incluir um núcleo ACELP cálculo de um ganho de tom, com defasagem de tome/ou as informações do codebook como um índice de codebook e um ganho de código.

Embora algumas das figuras ilustrem os blocos de diagramas de um aparelho, é de observado que estas figuras, ao mesmo tempo, ilustrando um método, no qual as funcionalidades do bloco correspondente para os estágios do método.

A materialização da invenção foi descrita acima

com base em um sinal de saída de áudio composto de diferentes segmentos ou estruturas, os diferentes segmentos ou estruturas sendo associados com a informação da fala ou da musica. A invenção não se limita a tais materializações, ao contrário, a abordagem para a classificação de diferentes segmentos de um sinal composto de pelo menos, segmentos de tipo um primeiro e um segundo tipo, também pode ser aplicado a sinais de áudio composto por três ou mais tipos de segmentos diferentes, cada qual se deseja ser codificado por diferentes esquemas de codificação. Os exemplos de tipos de segmento, são:

- Segmentos estacionários/não-estacionários podem ser úteis para o uso de diferente bancos de filtro, janelas ou adaptação de codificação. Por exemplo, uma transitória deve ser codificada com um banco de filtro de resolução de tempo adequada, enquanto uma senóide pura deve ser codificado com um banco de filtro de resolução de frequência adequado.

- Sonoro/não sonoro: os segmentos sonoros são bem tratados pelo codificador CELP, mas para segmentos não sonoros muitos bits são desperdiçados A codificação paramétrica será mais eficiente.

- Silencioso/ativado: o segmento silencioso pode ser codificado com menos bits que o segmento ativado.

- Harmônico/não-harmônico: Será útil para a utilização da codificação segmentos harmônicos usando uma linear preditiva no domínio da frequência.

Além disso, a invenção não se limita ao campo das técnicas de áudio, em vez disso, a abordagem descrita acima para a classificação de um sinal pode ser aplicada a outros tipos de

sinais, como os sinais de vídeo ou dados, onde esses respectivos sinais incluem segmentos de tipos diferentes, que exigem um processamento diferente como, por exemplo:

A presente invenção pode ser adaptada para todas as aplicações em tempo real que precisam de uma segmentação de um sinal de tempo. Por exemplo, a detecção do rosto a partir de uma câmera de vídeo de vigilância pode ser baseado em um classificador que determina para cada pixel de um quadro (aqui um quadro corresponde a uma foto tirada em um tempo n) se ele pertence ao rosto de uma pessoa ou não. A classificação (ou seja, a segmentação do rosto) deve ser feita para cada quadros simples do fluxo de vídeo. No entanto, usando a presente invenção, a segmentação do quadro atual pode levar em conta os sucessivos quadros anteriores para obter uma precisão melhor segmentação tendo a vantagem de que as imagens sucessivas estão fortemente correlacionados. Os dois classificadores podem ser então aplicadas. Um considerando apenas o quadro atual e outro considerando um conjunto de quadros, incluindo o quadro atual e anterior. O último classificador pode integrar o conjunto de quadros e determinar a região de probabilidade para a posição do rosto. A decisão do classificador feito apenas sobre o quadro atual, será então comparada com as regiões de probabilidade. A decisão pode ser validada ou modificada.

A materialização da invenção usa o comutador pra alterar entre as seções de modo que somente uma seção receba um sinal a ser processado e a outra seção não receba o sinal. Em uma materialização alternativa, entretanto, o comutador pode também ser organizado depois do estágio de processamento ou seções, por

exemplo, o codificador de áudio e de fala, de modo que ambas as seções processam o mesmo sinal em paralelo. A entrada de sinal por uma dessas seções é escolhida para ser a saída, por exemplo, a ser escrito em um fluxo contínuo de saída.

Enquanto a materialização da invenção foi descrita com base nos sinais digitais, os segmentos dos quais foram determinados por um número predefinido de amostras obtidos na mesma taxa de amostragem especifica, a invenção não é limitada para tais sinais, especialmente, também é aplicada a sinais analógicos nos quais o segmento deveria então ser determinado por um alcance especifico de frequência ou período de tempo do sinal analógico. Além disso, a materialização da invenção foi descrita em combinação com codificadores incluindo o discriminador. É observado que, basicamente, a abordagem de acordo com uma materialização da invenção para classificação de sinais pode também ser aplicada a decodificadores recebendo um sinal codificado para que diferentes esquemas codificados possam ser classificados, permitindo assim que o sinal codificado para ser fornecido a um decodificador apropriado.

Dependendo dos requisitos de implementação de alguns dos métodos criativos, os métodos inventivos possam ser implementados em hardware ou software. A aplicação pode ser realizada utilizando um meio de armazenamento digital, em particular, um disco, um DVD ou um CD com controlo eletrônico de leitura de sinais nele armazenados, que co-operam com sistemas de computador programáveis de tal forma que os métodos inventivos são executadas. Normalmente, a presente invenção é, portanto, um produto de programa de computador com um código de programa

armazenado em um portador de leitura de máquina, o código do programa que está sendo operado para a realização dos métodos criativos quando o produto de programa de computador é executado em um computador. Em outras palavras, os métodos criativos são, portanto, um programa de computador com um código de programa para realizar pelo menos um dos métodos criativos quando o programa de computador é executado em um computador.

A materialização descrita acima são meramente ilustrativas para os princípios da atual invenção. É entendido que as modificações e variações da disposição e os detalhes descritos neste documento será aparente para os outros qualificados na arte. É a intenção, portanto, ser limitada somente pelo escopo das reivindicações da iminente patente e não com os detalhes específicos, apresentados por meio da descrição e explicação das encarnações neste documento.

Na materialização acima, o sinal é descrito como composto de uma pluralidade estruturas, onde uma estrutura atual é avaliada por uma decisão de comutação. É observado que o segmento atual do sinal que é avaliado por uma decisão de comutação pode ser uma estrutura, entretanto, a invenção não é limitada a tais materializações. Além disso, um segmento do sinal pode ser composto de uma pluralidade, por exemplo, duas ou mais estruturas.

Além disso, na descrição acima a materialização do classificador de curto prazo e do classificador de longo prazo usando o mesmo recurso(s) de curto prazo. Esta abordagem pode ser usada para diferentes motivos, como a necessidade de calcular os recursos de curto prazo somente uma vez, para explorar o mesmo por dois classificadores de formas diferentes o que irá reduzir a

complexidade do sistema, como por exemplo, o recurso de curto prazo pode ser calculado por um dos classificadores curto prazo ou de longo prazo e fornecidos por outro classificador. Também, a comparação entre os resultados do classificador de curto prazo e do longo prazo pode ser mais relevante do que a contribuição para da estrutura atual no resultado de classificação de longo prazo é mais facilmente deduzida pela comparação com o resultado de classificação de curto prazo uma vez que os classificadores compartilham recursos comuns.

A invenção é, entretanto, não é restrita a tal abordagem e o classificador de longo prazo não é restrito para usar o recurso(s) de curto prazo como classificador de curto prazo, por exemplo, tanto o classificador de curto prazo e classificador de longo prazo pode calcular seu respectivo recurso(s) de curto prazo que é diferente para cada um.

Enquanto uma materialização descrita acima mencionou o uso de PLPCCs como recurso de curto prazo, é observado que outros recursos podem ser considerados, por exemplo, a variabilidade do PLPCCs.

REIVINDICAÇÕES

1. **"MÉTODO E DISCRIMINADOR PARA A CLASSIFICAÇÃO DE DIFERENTES SEGMENTOS DE UM SINAL DE AÚDIO"**, **caracterizado** pelo sinal de áudio compreender segmento de fala e segmento de música e o método compreender:

Classificação de curto prazo por um classificador de curto prazo (150), o sinal de áudio usando pelo menos um recurso de curto prazo e pelo menos um recurso de longo prazo extraídos do sinal de áudio e entregam um resultado de classificação de longo prazo (156), e

aplicação de resultado de classificação de curto prazo e do resultado de classificação de longo prazo a um circuito de decisão (158) acoplado a uma saída do classificador de curto prazo (150) e a uma saída do classificador de longo prazo (154), o circuito de decisão (158) combinando o resultado de classificação de curto prazo (152) e a classificação de longo prazo (156) para fornecer um sinal de saída (160), que indica se o segmento atual do sinal de áudio é um segmento de fala ou de um segmento de música.

2. Método, de acordo com a reivindicação 1, **caracterizado** pela etapa de combinação compreender em fornecer o sinal de saída como base em uma comparação do resultado da classificação de curto prazo (152) para o resultado da classificação de longo prazo (156).

3. Método, de acordo com a reivindicação 1 ou 2, **caracterizado** por compreender:

pelo menos, um recurso de curto prazo é obtido através da análise de um segmento atual do sinal de áudio que

deve ser classificado; e

pelo menos, um recurso de longo prazo é obtido através da análise de um segmento atual do sinal de áudio e um ou mais segmentos anteriores do sinal de áudio.

4. Método, de acordo com uma das reivindicações 1 a 3, **caracterizado** por compreender:

pelo menos, um recurso de curto prazo é obtido através da janela de análise (168) de um primeiro comprimento e um método de primeira análise; e

pelo menos, um recurso de longo prazo é obtido através da janela de análise (162) de um segundo comprimento e um método de segunda análise, o primeiro comprimento sendo mais curto que o segundo comprimento, e os métodos da primeira e segunda análise sendo diferentes.

5. Método da reivindicação 4, **caracterizado** pelo primeiro comprimento transpor um segmento atual do sinal de áudio, o segundo comprimento transpõe o segmento atual do sinal de áudio e um ou mais segmentos anteriores do sinal de áudio, e os primeiro e segundo comprimentos abrange um período adicional (164), cobrindo um período de análise.

6. Método, de acordo com uma das reivindicações 1 a 5, **caracterizado** por combinar o resultado da classificação de curto prazo (152) e o resultado da classificação de longo prazo (156), compreendendo uma decisão de histerese com base no resultado combinado, onde o resultado combinado inclui o resultado da classificação de curto prazo (152) e o resultado de classificação de longo prazo (156), cada ponderado por um fator de ponderação predeterminado.

7. Método, de acordo com uma das reivindicações 1 a 6, **caracterizado** pelo sinal de áudio ser um sinal digital e um segmento do sinal de áudio compreende como número predeterminado das amostras obtidas em uma taxa de amostragem específica.

8. Método, de acordo com uma das reivindicações 1 a 7, **caracterizado** por:

pelo menos, um recurso de curto prazo abrange os parâmetros PLPCCs; e

pelo menos, um recurso de longo prazo abrange a informação da característica do tom.

9. Método, de acordo com uma das reivindicações 1 a 8, **caracterizado** pelo recurso de curto prazo ser utilizado pela classificação de curto prazo e o recurso de curto prazo ser utilizado pela classificação de longo prazo são as mesmas ou diferentes.

10. Método para processar um sinal de áudio compreendendo os segmentos de pelo menos de um primeiro e um segundo tipo, o método sendo **caracterizado** por compreender:

classificação (116) de um segmento atual do sinal de áudio de acordo com o método de uma das reivindicações 1 a 9;

dependente do sinal de saída (160) fornecido pela etapa de classificação (116), processamento (102, 206, 106, 208) o segmento atual de acordo com um primeiro ou um segundo processo, e

saída do segmento de processado.

11. Método, de acordo com a reivindicação 10,

caracterizado pelo segmento ser processado por um codificador de voz (102) quando o sinal de saída (160) indicar que o segmento é um segmento de fala; e

o segmento é processado por um codificador de música (106) quando o sinal de saída (160) indicar que o segmento é um segmento de música.

12. Método, de acordo com a reivindicação 11, **caracterizado** por compreender ainda:

combinação (108) do codificador de segmento e informação para o sinal de saída (160) indicando o tipo de segmento.

13. Discriminador, **caracterizado** por compreender:

um classificador de curto prazo (150) configurado para receber um sinal de áudio e fornecer o resultado da classificação de curto prazo (152) do sinal de áudio usando apenas um recurso de curto prazo extraído do sinal de áudio, o sinal de áudio compreendendo segmentos de fala e de música;

um classificador de longo prazo (154) configurado para receber um sinal de áudio e fornecer o resultado da classificação de longo prazo (156) do sinal de áudio usando pelo menos um recurso de curto prazo e pelo menos um recurso de longo prazo extraídos do sinal de áudio; e

um circuito de decisão (158), acoplado a uma saída do classificador de curto prazo (150) e a uma saída do classificador de longo prazo (154), para receber o resultado de classificação de curto prazo (152) e o resultado de

classificação de longo prazo (156), o circuito de decisão (158) sendo configurado para combinar o resultado de classificação de curto prazo (152) e o resultado de classificação de longo prazo (156) para fornecer um sinal de saída (160), que indica se o segmento atual do sinal de áudio é um segmento de fala ou de um segmento de música.

14. Discriminador, de acordo com a reivindicação 13, **caracterizado** pelo circuito de decisão (158) configurado para fornecer o sinal de saída com base em uma comparação do resultado da classificação de curto prazo (152) para o resultado da classificação de longo prazo (156).

15. Aparelho de processamento de sinal de áudio, **caracterizado** por compreender:

uma entrada (110) configurada para receber um sinal de áudio para ser processado, onde o sinal de áudio é composto de segmentos de fala e musica;

um primeiro estágio de processamento (102; 206), configurado para processar os segmentos de fala;

um segundo estágio de processamento (104; 208) configurado para processar os segmentos de musica;

um discriminador (116; 204) da reivindicação 14 ou 15 acoplado à entrada; e

um dispositivo de comutação (112; 202) acoplado entre a entrada (110) e o primeiro e segundo estágios de processamento (102, 104; 206, 208) e configurado para aplicar o sinal de áudio da entrada (110) para um dos primeiro e segundo estágios de processamento (102, 104; 206, 208) dependente no sinal de saída (160) para o discriminador (116).

16. Codificador de áudio, **caracterizado** por compreender um aparelho de processamento de sinal de áudio, de acordo com a reivindicação 15.

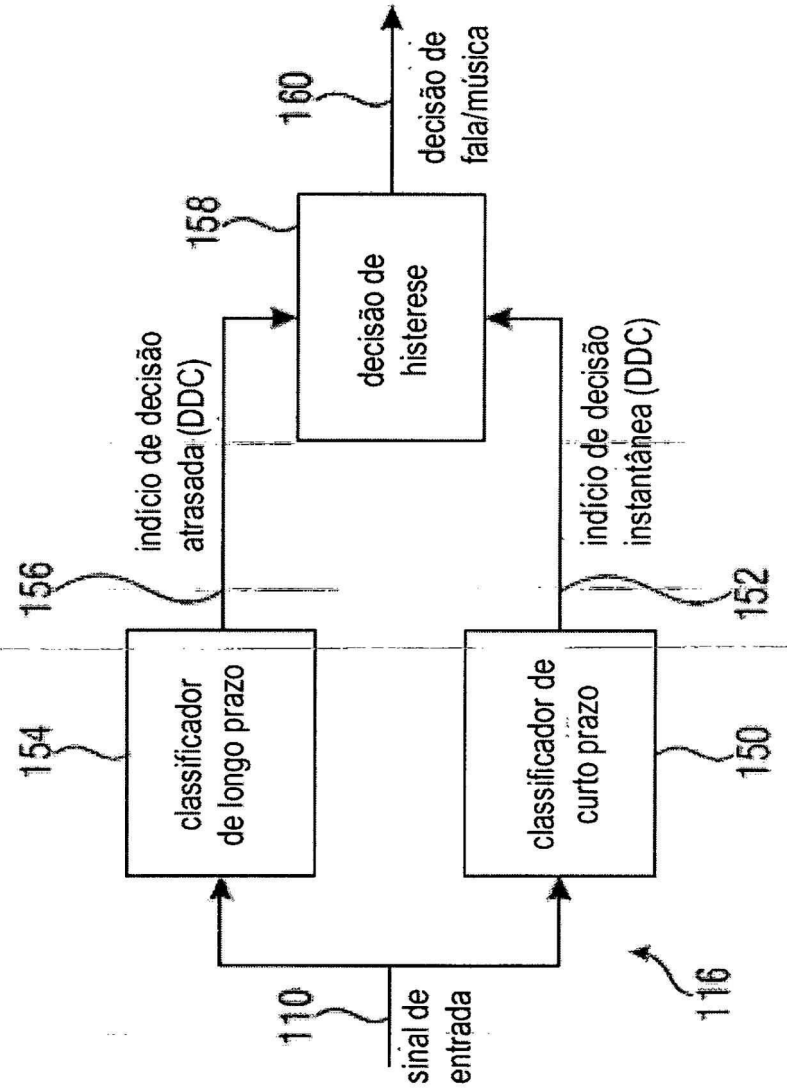
sinal de entrada
110
116
classificador de longo prazo
154
classificador de curto prazo
150
indício de decisão atrasada (DDC)
156
indício de decisão instantânea (DDC)
152
decisão de histerese
158
decisão de fala/música
160

FIGURA 1

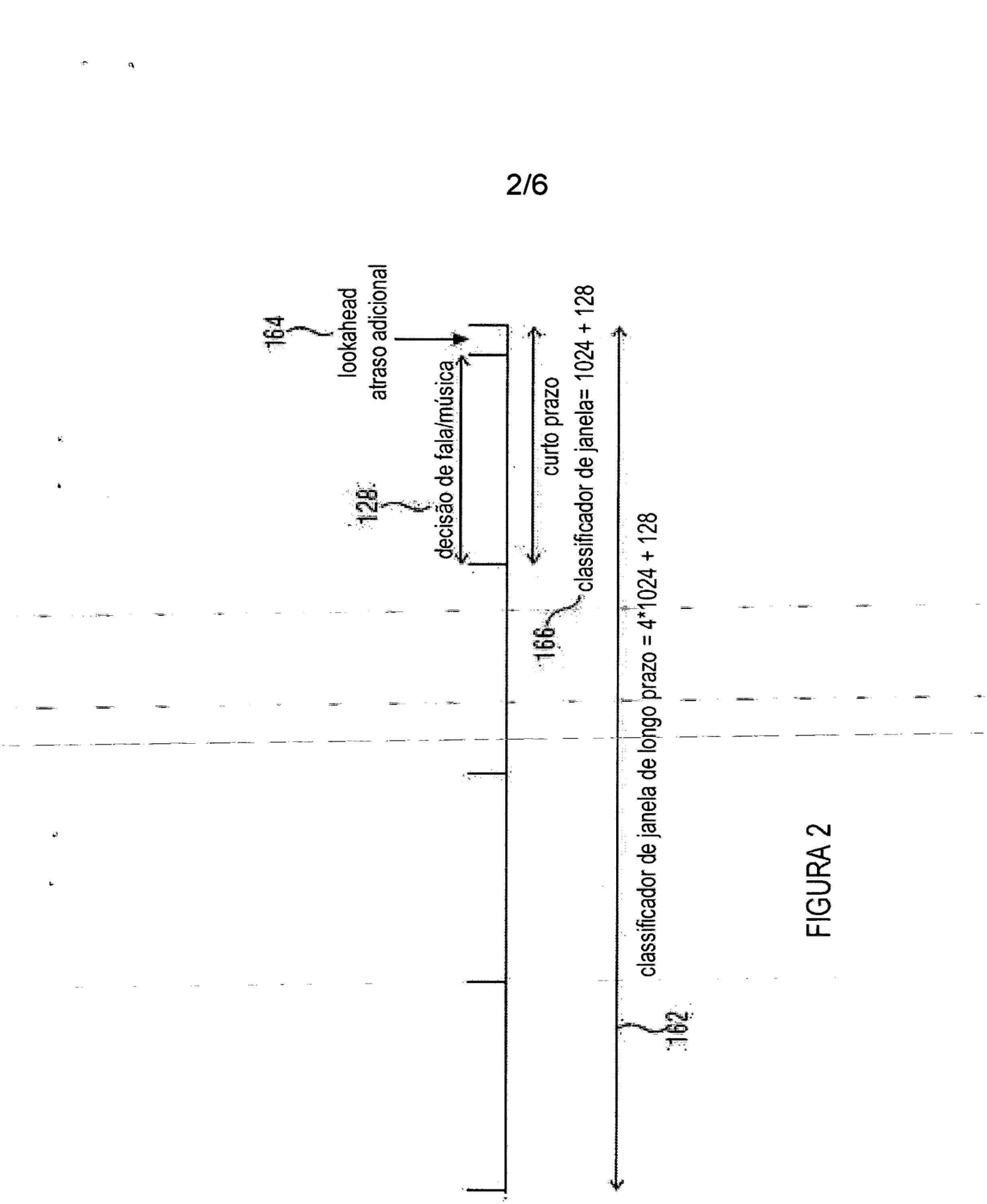
164
lookahead
atraso adicional
128
decisão de fala/música
curto prazo
166
classificador de janela= 1024 + 128
classificador de janela de longo prazo = 4*1024 + 128
162

FIGURA 2

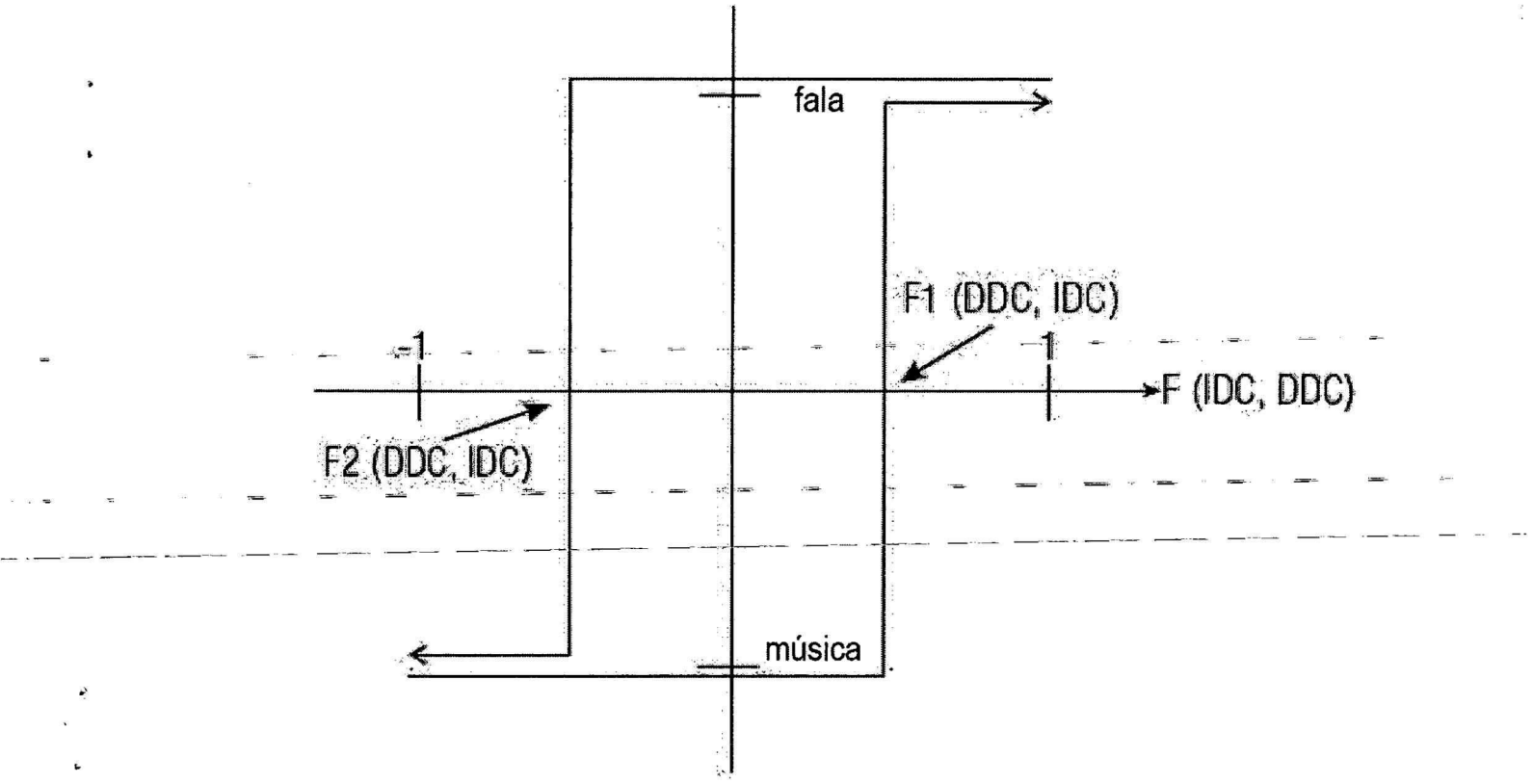

FIGURA 3

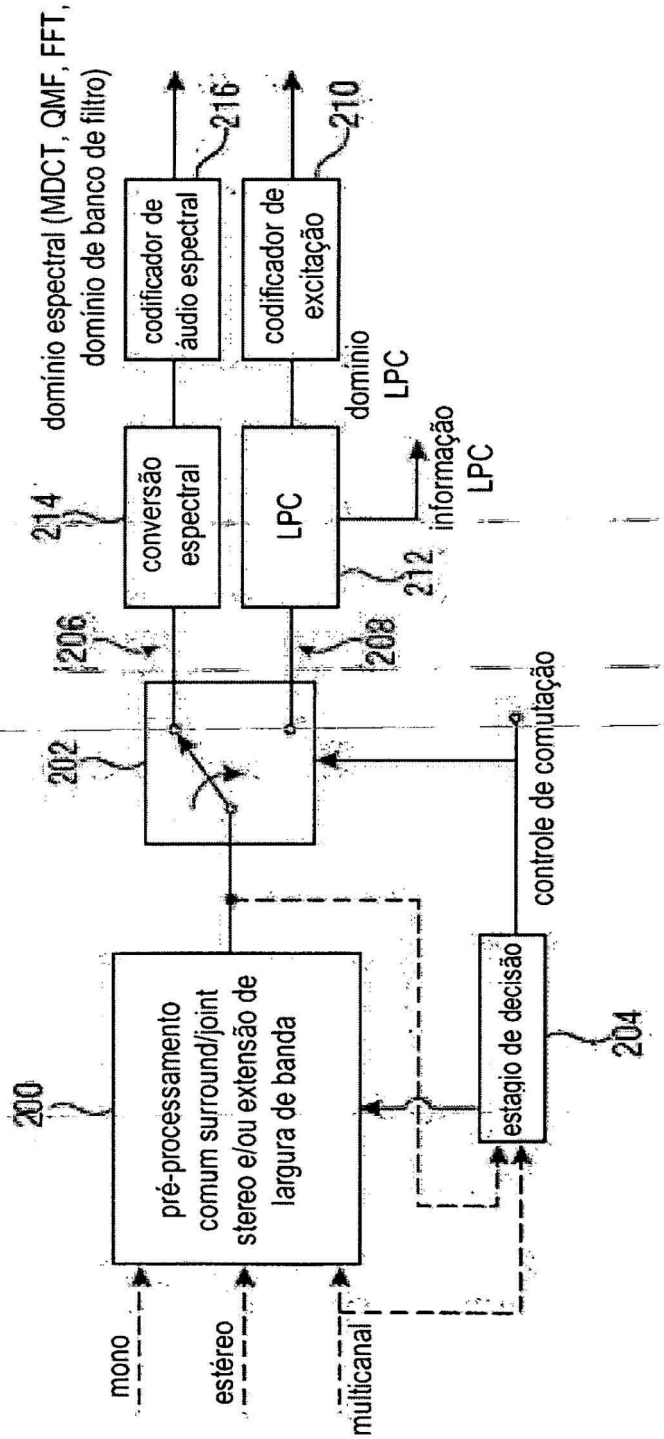

FIGURA 4
(CODIFICADOR)

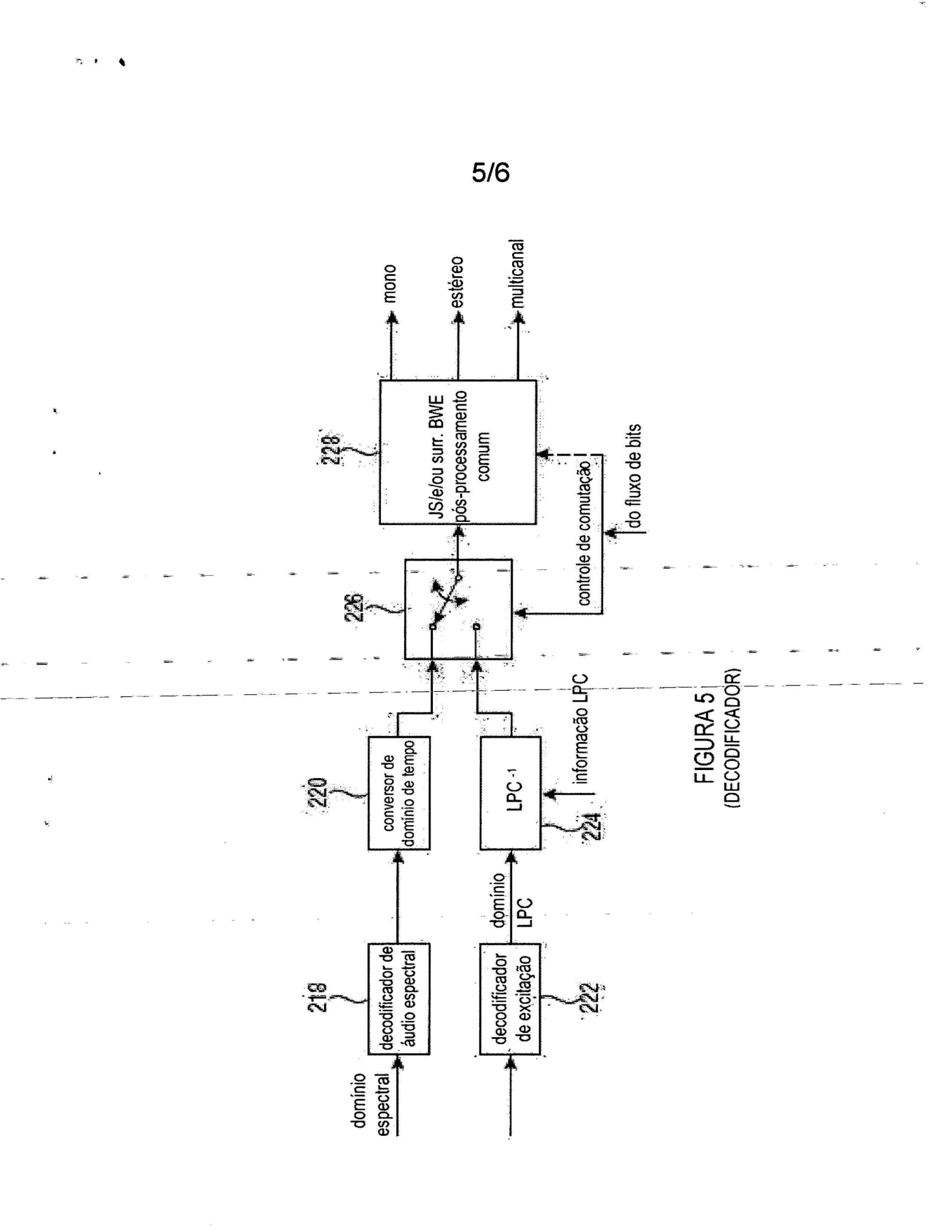

FIGURA 5
(DECODIFICADOR)

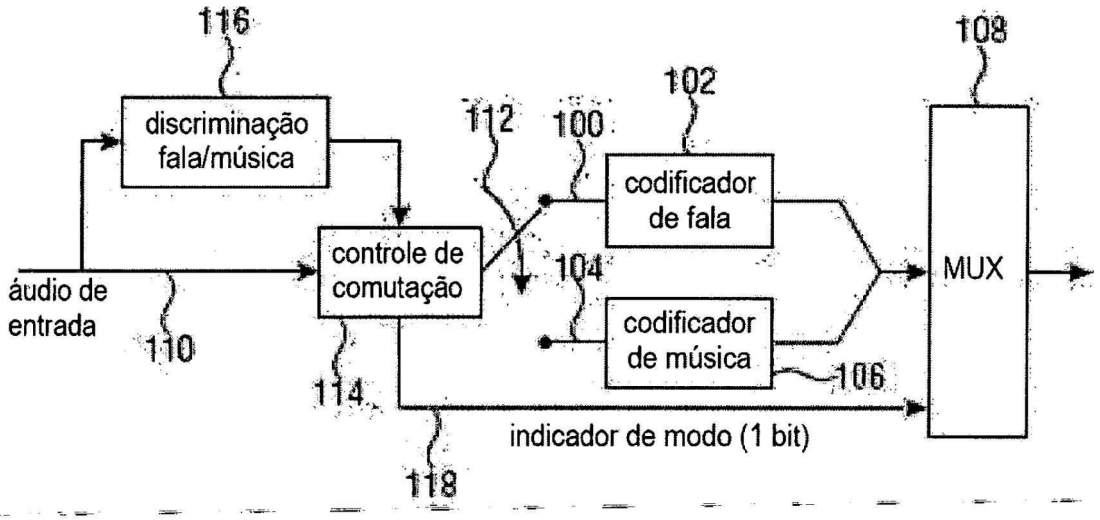

FIGURA 6

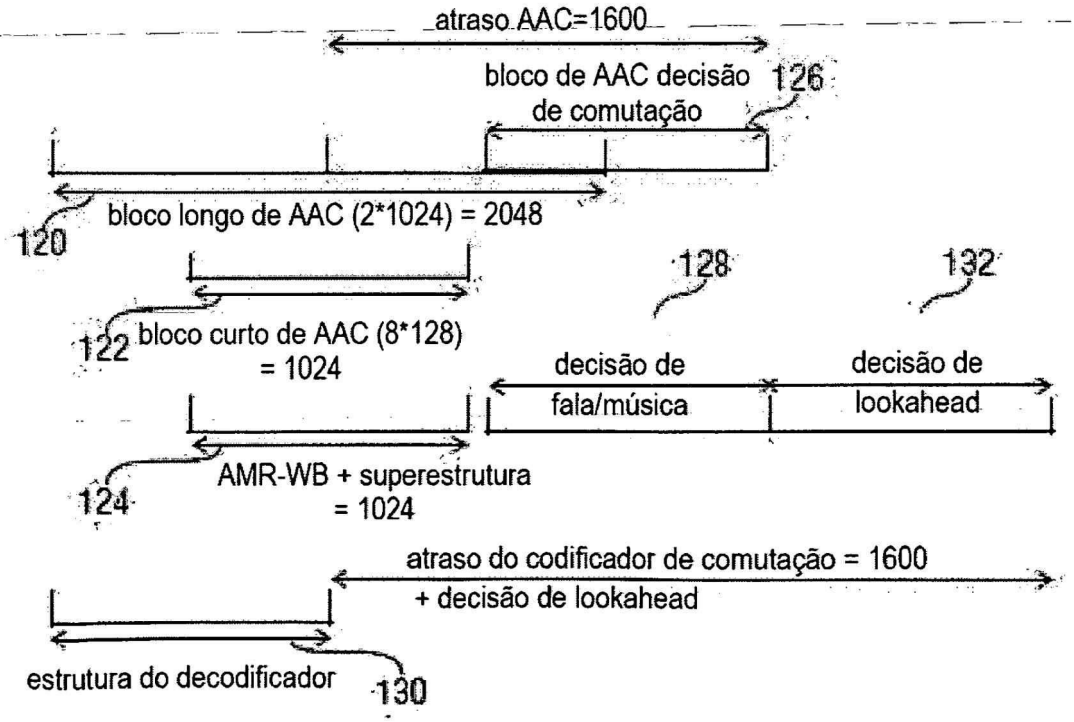

FIGURA 7