



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

(11) PI 0906142-8 B1



(22) Data do Depósito: 17/02/2009

(45) Data de Concessão: 20/10/2020

(54) Título: DISPOSITIVO E MÉTODO PARA A MANIPULAÇÃO DE UM SINAL DE ÁUDIO TENDO UM EVENTO TRANSIENTE

(51) Int.Cl.: G10L 21/04.

(30) Prioridade Unionista: 10/03/2008 US 61/035,317.

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

(72) Inventor(es): DISCH, SASCHA.; NAGEL, FREDERIK.; RETTELBACH, NIKOLAUS.; MULTRUS, MARKUS.; FUCHS, GUILLAUME..

(86) Pedido PCT: PCT EP2009001108 de 17/02/2009

(87) Publicação PCT: WO 2009/112141 de 17/09/2009

(85) Data do Início da Fase Nacional: 10/09/2010

(57) Resumo: DISPOSITIVO E MÉTODO PARA A MANIPULAÇÃO DE UM SINAL DE ÁUDIO TENDO UM EVENTO TRANSIENTE. Um manipulador de sinal para a manipulação de um sinal de áudio tendo um evento transiente pode compreender um removedor de transiente (100), um processador de sinal (110) e um insersor de sinal (120) para a inserção de uma porção de tempo em um sinal de áudio processo em um local de sinal onde o evento transiente foi removido antes do processador pelo referido removedor de transiente, de modo que um sinal de áudio manipulado comporte um evento transiente não influenciado pelo processamento, por meio do qual a coerência vertical do evento transiente é mantida em vez de qualquer processamento realizado no processador de sinal (110), o que destruiria a coerência vertical de um transiente.

**"DISPOSITIVO E MÉTODO PARA A MANIPULAÇÃO DE UM
SINAL DE ÁUDIO TENDO UM EVENTO TRANSIENTE"**

Descrição

A presente invenção refere-se ao processamento de

5 sinal de áudio e, particularmente, à manipulação de sinal de áudio no contexto da aplicação de efeitos de áudio a um sinal contendo eventos transientes..

Sabe-se manipular sinais de áudio de modo que a velocidade de reprodução seja alterada, enquanto o passo é
10 mantido. Métodos conhecidos para esse procedimento são implementados por vocoders de fase ou métodos, tais como de sobreposição-soma (passo síncrono), (P)SOLA, como por exemplo, descrito em J.L. Flanagan and R. M. Golden, *The Bell System Technical Journal*, November 1966, pp. 1394 to 1509; United States
15 Patent 6549884 Laroche, J. & Dolson, M.: *Phase-vocoder pitch-shifting*; Jean Laroche and Mark Dolson, *New Phase-Vocoder Techniques for Pitch-Shifting, Harmonizing And Other Exotic Effects*, Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York, Oct. 17-
20 20, 1999; e Zölzer, U: *DAFX: Digital Effects of Audio*; Wiley & Sons; Edition: 1 (February 26, 2002); pp. 201-298.

Além disso, os sinais de áudio podem ser submetidos a uma transposição utilizando esses métodos, ou seja, vocoders de fase ou (P)SOLA, onde o aspecto especial desse tipo de
25 transposição é que o sinal de áudio transposto tem o mesmo comprimento de reprodução/nova reprodução que o sinal de áudio original antes da transposição, enquanto que o passo é alterado. Isso é obtido por uma reprodução acelerada dos sinais prolongados

onde o fator de aceleração para realizar a reprodução acelerada depende do fator de prolongamento para prolongar o sinal de áudio original no tempo. Quando se tem uma representação de sinal discreto no tempo, esse procedimento corresponde a uma amostragem descendente do sinal prolongado ou decimação do sinal prolongado por um fator igual ao fator de prolongamento onde a frequência de amostragem é mantida.

Um desafio específico nessas manipulações de sinal de áudio são os eventos transientes. Os eventos transientes são eventos em um sinal no qual a energia do sinal em toda a banda ou em uma determinada faixa de frequência está mudando rapidamente, ou seja, aumentando rapidamente ou diminuindo rapidamente. Os aspectos característicos de transientes específicos (eventos transientes) são a distribuição da energia do sinal no espectro. Geralmente, a energia do sinal de áudio durante um evento transiente é distribuída em toda a frequência enquanto que, nas partes de sinal não-transiente, a energia é normalmente concentrada na porção de baixa frequência do sinal de áudio ou em bandas específicas. Isso significa que uma porção de sinal não-transiente, também denominada uma porção de sinal fixa ou tonal, possui um espectro que é não-plano. Em outras palavras, a energia do sinal é incluída em um número comparativamente pequeno de linhas espectrais/bandas espectrais, que são fortemente elevadas em uma base de ruído de um sinal de áudio. Em uma porção transiente, no entanto, a energia do sinal de áudio será distribuída em muitas bandas de frequência diferentes e, especificamente, será distribuída na porção de alta frequência, de modo que um espectro para uma porção transiente do sinal de áudio

será comparativamente plano e, em qualquer evento, será mais plano que um espectro de uma porção tonal do sinal de áudio. Geralmente, um evento transiente é uma forte alteração no tempo, o que significa que o sinal incluirá muitos harmônicos superiores quando 5 uma decomposição de Fourier for realizada. Uma característica importante desses vários harmônicos superiores é que as fases desses harmônicos superiores estão em uma relação mútua muito específica, de modo que uma sobreposição de todas essas ondas de seno resultará em uma rápida alteração da energia do sinal. Em 10 outras palavras, há uma forte correlação através do espectro.

A situação de fase específica entre todas as harmônicas pode também ser denominada como uma "coerência vertical". Esta "coerência vertical" está relacionada a uma representação de spectrograma de tempo/frequência do sinal onde 15 uma direção horizontal corresponde ao desenvolvimento do sinal com o tempo e onde a dimensão vertical descreve a interdependência em relação à frequência dos componentes espectralis (armazenador de frequência de transformada) em um espectro de tempo curto em relação à frequência.

20 Devido às etapas de processamento típicas, que são executadas de modo a prolongar o tempo ou encurtar um sinal de áudio, esta coerência vertical é destruída, o que significa que um transiente é "manchado" com o tempo quando um transiente é submetido a uma operação de prolongamento de tempo ou encurtamento 25 de tempo, como, por exemplo, executada por um vocoder de fase ou qualquer outro método, que executa um processamento dependendo de frequência introduzindo mudanças de fase no sinal de áudio, que são diferentes para coeficientes de frequência diferentes.

Quando a coerência vertical de transientes é destruída por um método de processamento de sinal de áudio, o sinal manipulado será muito similar ao sinal original em porções fixas ou não transientes, mas as porções transientes terão uma 5 qualidade reduzida no sinal manipulado. A manipulação não controlada da coerência vertical de um transiente resulta em dispersão temporal da mesma, visto que muitos componentes harmônicos contribuem para um evento transiente e alterar as fases de todos estes componentes de maneira não controlada 10 inevitavelmente resulta nestes artefatos.

Entretanto, porções transientes são extremamente importantes para a dinâmica de um sinal de áudio, tal como um sinal de música ou um sinal de voz onde alterações repentinas de energia em um tempo específico representam uma grande quantidade 15 da impressão subjetiva do usuário sobre a qualidade do sinal manipulado. Em outras palavras, eventos transientes no sinal de áudio são tipicamente "marcas" muito perceptíveis de um sinal de áudio, as quais têm uma influência desproporcional na impressão subjetiva da qualidade. Transientes manipulados nos quais 20 coerência vertical foi destruída por uma operação de processamento de sinal ou foi degradada com relação à porção transiente do sinal original soarão distorcidas, reverberantes e artificiais ao ouvinte.

Alguns métodos atuais prolongam o tempo ao redor 25 dos transientes para uma extensão mais elevada de modo a ter que executar subsequentemente, durante a duração do transiente, nenhum prolongamento ou apenas prolongamento de tempo muito pequeno. Estas referências e patentes do estado da técnica descrevem

métodos para manipulação de tempo e/ou tom. Referências do Estado da Técnica são: Laroche L., Dolson M.: "Improved phase vocoder timescale modification of audio", IEEE Trans. Speech and Audio Processing, vol. 7, nº 3, página 323 - 332; Emmanuel Ravelli, Mark Sandler e Juan P. Bello: "Fast implementation for non-linear time-scaling of stereo audio"; Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx'05), Madri, Espanha, 20-22 de Setembro de 2005; Duxbury, C. M. Davies, e M. Sandler (Dezembro de 2001). "Separation of transient information in musical audio using multiresolution analysis techniques". No Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Irlanda; e Röbel, A.: "A NEW APPROACH TO TRANSIENT PROCESSING IN THE PHASE VOCODER"; Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), Londres, RU, 8-11 de Setembro de 2003.

Durante o prolongamento de tempo de sinais de áudio por vocoders de fase, porções de sinal transiente são "embaçadas" por dispersão, visto que a coerência vertical do sinal é prejudicada. Métodos usando métodos de adição de sobreposição, como (P)SOLA podem gerar pré e pós-ecos perturbadores de eventos de som transientes. Estes problemas podem realmente serem tratados por prolongamento de tempo aumentado no ambiente de transientes; entretanto, se uma transposição está para ocorrer, o fator de transposição não será mais constante no ambiente dos transientes, isto é o tom de componentes de sinal sobrepostos (possivelmente tonal) será alterado e será percebido como um distúrbio.

É um objetivo da presente invenção prover um conceito de qualidade melhorada para manipulação de sinal de áudio.

Este objetivo é atingido por um equipamento para manipulação de um sinal de áudio de acordo com a reivindicação 1, um equipamento para gerar um sinal de áudio de acordo com a reivindicação 12, um método de manipulação de um sinal de áudio de acordo com a reivindicação 13, um método de geração de um sinal de áudio de acordo com a reivindicação 14, um sinal de áudio tendo uma porção transiente e informação de lado de acordo com a reivindicação 15 ou um programa de computador de acordo com a reivindicação 16.

10 Para tratar de problemas de qualidade ocorrendo em um processamento não controlado de porções transientes, a presente invenção certifica-se que porções transientes não sejam processadas de maneira prejudicial, isto é, sejam removidas antes do processamento e sejam reinseridas após processamento ou os 15 eventos transientes sejam processados, mas sejam removidas do sinal processado e substituídas por eventos transientes não processados.

Preferivelmente, as porções transientes inseridas no sinal processado são cópias de porções transientes 20 correspondentes no sinal de áudio original, de modo que o sinal manipulado consiste de uma porção processada não incluindo uma porção transiente e uma porção não processada ou processada diferentemente incluindo o transiente. Exemplificativamente, o transiente original pode ser submetido a decimação ou qualquer 25 tipo de ponderação ou processamento parametrizado. Alternativamente, entretanto, porções transientes podem ser substituídas por porções transientes sinteticamente criadas, que são sintetizadas de maneira que a porção transiente sintetizada

seja similar à porção transiente original com relação a alguns parâmetros transientes, tais como a quantidade de mudança de energia em um certo tempo ou qualquer outra medição caracterizando um evento transiente. Dessa maneira, seria possível caracterizar 5 ainda uma porção transiente no sinal de áudio original e seria possível remover este transiente antes do processamento ou substituir o transiente processado por um transiente sintetizado, que é sinteticamente criado com base em informação paramétrica transiente. Por questões de eficiência, entretanto, é preferido 10 copiar uma porção do sinal de áudio original antes da manipulação e inserir esta cópia no sinal de áudio processado, visto que este procedimento garante que a porção transiente no sinal processado seja idêntica ao transiente do sinal original. Este procedimento garantirá que a alta influência específica de transientes em uma 15 percepção de sinal de som seja mantida no sinal processado comparado com o sinal original antes do processamento. Dessa maneira, uma qualidade subjetiva ou objetiva com relação aos transientes não é degradada por qualquer tipo de processamento de sinal de áudio para manipulação de um sinal de áudio.

20 Nas configurações preferidas, a presente aplicação provê um método novo para um tratamento perceptual favorável de eventos de som transientes dentro da estrutura deste processamento, que geraria, por outro lado, um "embaçamento" temporal pela dispersão de um sinal. Este método preferido 25 essencialmente comprehende a remoção dos eventos de som transientes antes da manipulação de sinal para o objetivo de prolongamento de tempo e, subsequentemente, adição, enquanto considerando o prolongamento, da porção de sinal transiente não processada ao

sinal modificado (prolongado) de uma maneira precisa.

Configurações preferidas da presente invenção são subsequentemente explicadas com referência aos desenhos em anexo, nos quais:

5 A Figura 1 ilustra uma configuração preferida de um equipamento ou método inventivo para manipular um sinal de áudio tendo um transiente;

A Figura 2 ilustra uma implementação preferida de um removedor de sinal transiente da Figura 1;

10 A Figura 3a ilustra uma implementação preferida de um processador de sinal da Figura 1;

A Figura 3b ilustra uma configuração preferida adicional para implementar o processador de sinal da Figura 1;

A Figura 4 ilustra uma implementação preferida do 15 insensor de sinal da Figura 1;

A Figura 5a ilustra uma visão geral da implementação de um vocoder a ser usado no processador de sinal da Figura 1;

A Figura 5b mostra uma implementação de partes 20 (análises) de um processador de sinal da Figura 1;

A Figura 5c ilustra outras partes (prolongamento) de um processador de sinal da Figura 1;

A Figura 5d ilustra outras partes (síntese) de um processador de sinal da Figura 1;

25 A Figura 6 ilustra uma implementação de transformada de um vocoder de fase a ser usado no processador de sinal da Figura 1;

A Figura 7a ilustra um lado do codificador de um

esquema de processamento de extensão de largura de banda;

A Figura 7b ilustra um lado do decodificador de um esquema de extensão de largura de banda;

A Figura 8a ilustra uma representação de energia 5 de um sinal de entrada de áudio com um evento transiente;

A Figura 8b ilustra o sinal da Figura 8a, mas com um transiente de janela;

A Figura 8c ilustra um sinal sem a porção transiente antes de ser prolongada;

10 A Figura 8d ilustra o sinal da Figura 8c subsequentemente ao seu prolongamento; e

A Figura 8e ilustra o sinal manipulado após a porção correspondente do sinal original ter sido inserida;

A Figura 9 ilustra um equipamento para gerar 15 informação de lado para um sinal de áudio.

A Figura 1 ilustra um equipamento preferido para manipular um sinal de áudio tendo um evento transiente. Preferivelmente, o equipamento compreende um removedor de sinal transiente 100 tendo uma entrada 101 para um sinal de áudio com um 20 evento transiente. A saída 102 do removedor de sinal transiente é conectado a um processador de sinal 110. A saída do processador de sinal 111 é conectada a um insensor de sinal 120. A saída do insensor de sinal 121 na qual um sinal de áudio manipulado com um transiente “natural” não processado ou sintetizado está 25 disponível, pode ser conectada a um dispositivo adicional tal como um condicionador de sinal 130, que pode executar qualquer processamento adicional do sinal manipulado tal como uma amostragem para baixo/decimação para ser requerido para objetivos

de extensão de largura de banda conforme discutido em conexão com as Figuras 7A e 7B.

Entretanto, o condicionador de sinal 130 não pode ser usado se o sinal de áudio manipulado obtido na saída do insensor de sinal 120 for usado como é, isto é, armazenado para processamento adicional, transmitido para um receptor ou transmitido a um conversor digital/analógico que, no final, é conectado a um equipamento de alto-falante para finalmente gerar um sinal de som representando o sinal de áudio manipulado.

No caso de extensão de largura de banda, o sinal na linha 121 poderá ser o sinal de banda alta. Então, o processador de sinal gerou o sinal de banda alta do sinal de banda baixa de entrada, e a porção transiente de banda baixa extraída do sinal de áudio 101 teria que ser colocada na faixa de frequência da banda alta, o que é preferivelmente feito por um processamento de sinal que não perturbe a coerência vertical, tal como uma decimação. Esta decimação seria executada antes do insensor de sinal, de modo que a porção transiente decimada é inserida no sinal de banda alta na saída do bloco 110. Nesta configuração, o condicionador de sinal executaria qualquer processamento adicional do sinal de banda alta tal como formatação de envelope, adição de ruído, filtragem inversa ou adição de harmônica etc., conforme feito, por exemplo, na Replicação de Banda Espectral MPEG 4.

O insensor de sinal 120 preferivelmente recebe informação de lado do removedor 100 por meio da linha 123 de modo a escolher a porção certa do sinal não processado a ser inserida em 111.

Quando a configuração tendo os dispositivos 100,

110, 120, 130 é implementada, uma sequência de sinal conforme discutido em conexão com as Figuras 8a a 8e pode ser obtida. Entretanto, não é necessariamente requerido remover a porção transiente antes de executar a operação de processamento de sinal 5 no processador de sinal 110. Nesta configuração, o removedor de sinal transiente 100 não é requerido e o insensor de sinal 120 determina uma porção de sinal a ser cortada do sinal processado na saída 111 e substituir este sinal cortado por uma porção do sinal original conforme esquematicamente ilustrado pela linha 121 ou por 10 um sinal sintetizado conforme ilustrado pela linha 141 onde este sinal sintetizado pode ser gerado em um gerador de sinal transiente 140. De modo a ser capaz de gerar um transiente adequado, o insensor de sinal 120 é configurado para comunicar parâmetros de descrição transientes para o gerador de sinal 15 transiente. Portanto, a conexão entre os blocos 140 e 120 conforme indicado pelo item 141 é ilustrada como uma conexão de duas vias. Quando um detector de transiente específico é provido no equipamento para manipulação, então a informação sobre o transiente pode ser provida deste detector de transiente (não 20 mostrado na Figura 1) para o gerador de sinal transiente 140. O gerador de sinal transiente pode ser implementado para ter amostras transientes, que podem ser diretamente usadas ou ter amostras transientes pré-armazenadas, que podem ser ponderadas usando parâmetros transientes de modo a realmente gerar/sintetizar 25 um transiente a ser usado pelo insensor de sinal 120.

Em uma configuração, o removedor de sinal transiente 100 é configurado para remover uma primeira porção de tempo do sinal de áudio para obter um sinal de áudio de transiente

reduzido, onde a primeira porção de tempo compreende o evento transiente.

Além do mais, o processador de sinal é preferivelmente configurado para processamento do sinal de áudio de transiente reduzido no qual uma primeira porção de tempo compreendendo o evento transiente é removida ou para processamento do sinal de áudio incluindo o evento transiente para obter o sinal de áudio processado na linha 111.

Preferivelmente, o insensor de sinal 120 é configurado para inserção de uma segunda porção de tempo no sinal de áudio processado em um local de sinal onde a primeira porção de tempo foi removida ou onde o evento transiente está localizado no sinal de áudio, onde a segunda porção de tempo compreende um evento transiente não influenciado pelo processamento executado pelo processador de sinal 110, de modo que o sinal de áudio manipulado na saída 121 seja obtido.

A Figura 2 ilustra uma configuração preferida do removedor de sinal transiente 100. Em uma configuração na qual o sinal de áudio não inclui qualquer informação de lado/meta-informação sobre transientes, o removedor de sinal transiente 100 compreende um detector de transiente 103, um calculador de *fade-out/fade-in* 104 e um primeiro removedor de porção 105. Em uma configuração alternativa na qual informação sobre transientes no sinal de áudio foram coletadas conforme anexado ao sinal de áudio por um dispositivo de codificação, conforme discutido posteriormente com relação à Figura 9, o removedor de sinal transiente 100 compreende um extrator de informação de lado 106, que extrai a informação de lado anexada ao sinal de áudio conforme

indicado pela linha 107. A informação sobre o tempo transiente pode ser provida ao calculador de *fade-out/fade-in* 104 conforme ilustrado pela linha 107. Quando, entretanto, o sinal de áudio inclui, como meta-information, não (apenas) o tempo de transiente, 5 isto é o tempo preciso no qual o evento transiente está ocorrendo, mas o tempo inicial/final da porção a ser excluído do sinal de áudio, isto é, o tempo inicial e o tempo final da "primeira porção" do sinal de áudio, então o calculador de *fade-out/fade-in* 104 não é requerido também, e a informação de tempo inicial/final 10 pode ser diretamente enviada para o primeiro removedor de porção 105 conforme ilustrado pela linha 108. A linha 108 ilustra uma opção e todas as outras linhas, que são indicadas pelas linhas quebradas, são também opcionais.

Na Figura 2, o calculador de *fade-out/fade-in* 104 preferivelmente envia a informação de lado 109. Esta informação de lado 109 é diferente dos tempos inicial/final da primeira porção, visto que a natureza do processamento no processador 110 da Figura 1 é considerada. Além do mais, o sinal de áudio de entrada é preferivelmente alimentado no removedor 105.

20 Preferivelmente, o calculador de *fade-out/fade-in* 104 provê os tempos inicial/final da primeira porção. Estes tempos são calculados com base no tempo transiente, de modo que não apenas o evento transiente, mas também algumas amostras envolvendo o evento transiente são removidas pelo removedor 105 de primeira 25 porção. Além do mais, é preferido não apenas cortar a porção transiente por uma janela retangular de domínio de tempo, mas executar a extração por uma porção de *fade-out* e uma porção de *fade-in*. Para executar uma porção de *fade-out* e/ou *fade-in*,

qualquer tipo de janela tendo uma transição mais suave comparada com um filtro retangular, tal como uma janela de co-seno elevada pode ser aplicada, de modo que a resposta de frequência desta extração não seja tão problemática como seria se uma janela 5 retangular fosse aplicada, embora isto seja também uma opção. Esta operação de janela de domínio de tempo envia o restante da operação de janela, isto é, o sinal de áudio sem a porção de janela.

Qualquer método de supressão de transiente pode 10 ser aplicado neste contexto incluindo estes métodos de supressão de transiente deixando um sinal residual de transiente reduzido ou preferivelmente totalmente não transiente após a remoção do transiente. Comparado com uma remoção completa da porção 15 transiente, na qual o sinal de áudio é definido como zero em relação a certa porção de tempo, a supressão de transiente é vantajosa em situações nas quais um processamento adicional do sinal de áudio sofresse de porções definidas como zero, visto que estas porções definidas como zero são muito artificiais para um sinal de áudio.

Naturalmente, todos os cálculos executados pelo 20 detector de transiente 103 e pelo calculador de *fade-out/fade-in* 104 podem ser aplicados também no lado de codificação conforme discutido em conexão com a Figura 9, contanto que os resultados destes cálculos, tal como tempo transiente e/ou tempo 25 inicial/final da primeira porção sejam transmitidos para um manipulador de sinal, tanto como informação de lado quanto como meta-information juntamente com o sinal de áudio ou separadamente do sinal de áudio, tal como dentro de um sinal de metadados de

áudio separado a ser transmitido via um canal de transmissão separado.

A Figura 3a ilustra uma implementação preferida do processador de sinal 110 da Figura 1. Esta implementação 5 comprehende um analisador seletivo de frequência 112 e um dispositivo de processamento seletivo a frequência subsequentemente conectado 113. O dispositivo de processamento seletivo a frequência 113 é implementado de modo que ele aplica uma influência negativa na coerência vertical do sinal de áudio original. Exemplos para este processamento é o prolongamento de um sinal no tempo ou o encurtamento de um sinal no tempo onde este prolongamento ou encurtamento é aplicado de maneira seletiva a frequência, de modo que, por exemplo, o processamento introduz mudanças de fase no sinal de áudio processado, que são diferentes 10 para bandas de frequência diferentes..

15

Uma maneira preferida de processamento é ilustrada na Figura 3B no contexto de um processamento de *vocoder* de fase. De forma geral, um *vocoder* de fase comprehende um analisador de sub-banda/transformada 114, um processador 20 subsequentemente conectado 115 para executar um processamento seletivo à frequência de uma pluralidade de sinais de saída providos pelo item 114 e, subsequentemente, um combinador de sub-banda/transformada 116, que combina os sinais processados pelo item 115 de modo a finalmente obter um sinal processado no domínio 25 de tempo na saída 117 onde este sinal processado no domínio de tempo, novamente, é um sinal de largura de banda completa ou um sinal filtrado de baixa passagem contanto que a largura de banda do sinal processado 117 seja maior que a largura de banda

representada por uma ramificação única entre o item 115 e 116, visto que o combinador de sub-banda/transformada 116 executa uma combinação de sinais seletivos a frequência...

Detalhes adicionais no vocoder de fase são
5 subsequentemente discutidos em conexão com as Figuras 5A, 5B, 5C e
6.

Subsequentemente, uma implementação preferida do insensor de sinal 120 da Figura 1 é discutida e ilustrada na Figura 4. O insensor de sinal preferivelmente compreende um
10 calculador 122 para calcular o comprimento da segunda porção de tempo. De modo a ser capaz de calcular o comprimento da segunda porção de tempo na configuração na qual a porção transiente foi removida antes do processamento de sinal no processador de sinal 110 na Figura 1, o comprimento da primeira porção removida e o
15 fator de prolongamento de tempo (ou o fator de encurtamento de tempo) são requeridos, de modo que o comprimento da segunda porção de tempo é calculado no item 122. Estes itens de dados podem ser inseridos de fora conforme discutido em conexão com as Figuras 1 e 2. Exemplificativamente, o comprimento da segunda porção de tempo
20 é calculado pela multiplicação do comprimento da primeira porção pelo fator de prolongamento.

O comprimento da segunda porção de tempo é enviado para o calculador 123 para calcular o primeiro ponto limite e o segundo ponto limite da segunda porção de tempo no
25 sinal de áudio. Em especial, o calculador 133 pode ser implementado para executar um processamento de correlação cruzada entre o sinal de áudio processado sem o evento transiente suprido na entrada 124 e o sinal de áudio com o evento transiente, que

provê a segunda porção conforme suprido na entrada 125. Preferivelmente, o calculador 123 é controlado por uma entrada de controle adicional 126 de modo que uma mudança positiva do evento transiente dentro da segunda porção de tempo é preferida em contrapartida com uma mudança negativa do evento transiente conforme discutido posteriormente.

O primeiro ponto limite e o segundo ponto limite da segunda porção de tempo são providos para um extrator 127. Preferivelmente, o extrator 127 corta a porção, isto é, a segunda porção de tempo fora do sinal de áudio original provido na entrada 125. Visto que um *cross-fader* 128 é usado, o corte ocorre usando um filtro retangular. No *cross-fader* 128, a porção inicial da segunda porção de tempo e a porção final da segunda porção de tempo são ponderadas por um peso crescente de 0 a 1 para a porção inicial e/ou peso decrescente de 1 para 0 na porção final de modo que nesta região de *cross-fade*, a porção final do sinal processado juntamente com a porção inicial do sinal extraído, quando adicionadas juntas, resultam em um sinal útil. Um processamento similar é executado no *cross-fader* 128 para o final da segunda porção de tempo e o inicio do sinal de áudio processado após a extração. O *cross-fader* certifica que nenhum artefato de domínio de tempo ocorra que seja, por outro lado, percebido como artefatos de "clique" quando os pontos limites do sinal de áudio processado sem a porção transiente e os pontos limites da segunda porção de tempo não correspondem perfeitamente.

Subsequentemente, é feito referência às Figuras 5a, 5b, 5c e 6 de modo a ilustrar uma implementação preferida do processador de sinal 110 no contexto de um vocoder de fase. A

seguir, com referência às Figuras 5 e 6, implementações preferidas para um vocoder são ilustradas de acordo com a presente invenção. A Figura 5a mostra uma implementação de banco de filtro de um vocoder de fase, onde um sinal de áudio é alimentado em uma 5 entrada 500 e obtido em uma saída 510. Em especial, cada canal do banco de filtro esquemático ilustrado na figura 5a inclui um filtro de passagem de banda 501 e um oscilador a jusante 502. Sinais de saída de todos os osciladores de cada canal são combinados por um combinador, que é, por exemplo, implementado 10 como um adicionador e indicado em 503, de modo a obter o sinal de saída. Cada filtro 501 é implementado de modo que provê um sinal de amplitude por um lado e um sinal de frequência por outro lado. O sinal de amplitude e o sinal de frequência são sinais de tempo 15 ilustrando um desenvolvimento da amplitude em um filtro 501 com o tempo, enquanto o sinal de frequência representa um desenvolvimento da frequência do sinal filtrado por um filtro 501.

Um ajuste esquemático de filtro 501 é ilustrado na Figura 5b. Cada filtro 501 da Figura 5a pode ser definido como na Figura 5b, onde, entretanto, apenas as frequências f_i supridas 20 para dois mixers de entrada 551 e o somador 552 são diferentes de canal para canal. Os sinais de saída do mixer são ambos filtrados de baixa passagem por passagens baixas 553, onde os sinais de baixa passagem são diferentes na medida em que eles foram gerados por frequências de oscilador locais (frequências LO), que estão 25 fora de fase em 90° . O filtro de baixa passagem superior 553 provê um sinal de quadratura 554, enquanto o filtro inferior 553 provê um sinal na fase 555. Estes dois sinais, isto é, I e Q, são supridos para um transformador de coordenada 556 que gera uma

representação de fase de magnitude da representação retangular. O sinal de magnitude ou sinal de amplitude, respectivamente, da Figura 5a com o tempo é enviado em uma saída 557. O sinal de fase é suprido a um desenrolador de fase 558. Na saída do elemento 558,
 5 não existe mais valor de fase presente que esteja sempre entre 0 e 360°, mas um valor de fase que aumenta linearmente. Este valor de fase "desenrolado" é suprido a um conversor de fase/frequência 559 que pode, por exemplo, ser implementado como um formador de diferença de fase simples que subtrai uma fase de um ponto
 10 anterior no tempo de uma fase em um ponto atual no tempo para obter um valor de frequência para o ponto atual no tempo. Este valor de frequência é adicionado ao valor de frequência constante f_i do canal de filtro i para obter um valor de frequência temporariamente variável na saída 560. O valor de frequência na
 15 saída 560 tem um componente direto = f_i e um componente alternativo = o desvio de frequência pelo qual uma frequência atual do sinal no canal de filtro desvia da frequência média f_i .

Dessa maneira, conforme ilustrado nas Figuras 5a e 5b, o vocoder de fase atinge uma separação da informação espectral e informação de tempo. A informação espectral está no canal especial ou na frequência f_i que provê a porção direta da frequência para cada canal, enquanto a informação de tempo está contida no desvio de frequência ou na magnitude em relação ao tempo, respectivamente.

25 A Figura 5c mostra uma manipulação como ela é executada para o aumento de largura de banda de acordo com a invenção, em especial, no vocoder e, em especial, no local do circuito ilustrado plotado em linhas pontilhadas na Figura 5a.

Para escalamento de tempo, por exemplo, os sinais de amplitude $A(t)$ em cada canal ou a frequência dos sinais $f(t)$ em cada sinal pode ser decimado ou interpolado, respectivamente. Por objetivos de transposição, como é útil para a presente invenção, 5 uma interpolação, isto é, uma extensão temporal ou espalhamento dos sinais $A(t)$ e $f(t)$ é executada para obter sinais espalhados $A'(t)$ e $f'(t)$, onde a interpolação é controlada por um fator de espalhamento em um cenário de extensão de largura de banda. Pela interpolação da variação de fase, isto é, o valor antes da adição 10 da frequência constante pelo somador 552, a frequência de cada oscilador individual 502 na Figura 5a não é alterada. A alteração temporal do sinal de áudio geral é diminuída, entretanto, isto é pelo fator 2. O resultado é um tom temporariamente espalhado tendo 15 o tom original, isto é, a onda fundamental original com sua harmônica.

Executando o processamento de sinal ilustrado na Figura 5c, onde este processamento é executado em cada canal de banda de filtro na Figura 5a, e pelo sinal temporal resultante, então, sendo decimado em um decimador, o sinal de áudio é 20 encolhido de volta a sua duração original enquanto todas as frequências são simultaneamente duplicadas. Isto conduz a uma transposição de tom pelo fator 2 onde, entretanto, é obtido um sinal de áudio tendo o mesmo comprimento que o sinal de áudio original, isto é, o mesmo número de amostras.

25 Como uma alternativa à implementação de banco de filtros ilustrada na Figura 5a, uma implementação de transformada de um vocoder de fase pode também ser usada conforme ilustrado na Figura 6. Aqui, o sinal de áudio 100 é alimentado em um

processador FFT, ou mais geralmente, em um Processador de Transformada de Fourier de Curto Tempo 600 como uma sequência de amostras de tempo. O processador FFT 600 é implementado esquematicamente na Figura 6 para executar uma janela de tempo de um sinal de áudio de modo a, então, por meio de um FFT, calcular a magnitude e fase do espectro, onde este cálculo é executado para sucessivos espectros que estão relacionados a blocos do sinal de áudio, que estão fortemente sobrepostos.

Em um caso extremo, para cada nova amostra de sinal de áudio, um novo espectro pode ser calculado, onde um novo espectro pode ser calculado também, por exemplo, apenas para cada vigésima amostra nova. Esta distância a nas amostras entre dois espectros é preferivelmente provido por um controlador 602. O controlador 602 é adicionalmente implementado para alimentar um processador IFFT 604 que é implementado para operar em uma operação de sobreposição. Em especial, o processador IFFT 604 é implementado de modo que ele execute uma Transformação de Fourier inversa de curta duração executando um IFFT por espectro baseado na magnitude e fase de um espectro modificado, de modo a, então, executar uma operação de adição de sobreposição, da qual o sinal de tempo resultante é obtido. A operação de adição de sobreposição elimina os efeitos da janela de análise.

Um espalhamento do sinal de tempo é atingido pela distância b entre dois espetros, conforme eles são processados pelo processador IFFT 604, sendo maior que a distância a entre os espectros na geração dos espectros FFT. A idéia básica é espalhar o sinal de áudio pelos FFTs inversos, simplesmente sendo separados adicionalmente aos FFTs de análise. Como um resultado, alterações

temporais no sinal de áudio sintetizado ocorrem mais lentamente que o sinal de áudio original.

Sem um rescalamento de fase no bloco 606, isto conduziria, entretanto, a artefatos. Quando, por exemplo, um armazenador de frequência único é considerado, para o qual valores de fase sucessivos por 45° são implementados, isto implica que o sinal dentro deste banco de filtro aumenta na fase com uma taxa de $1/8$ de um ciclo, isto é, em 45° por intervalo de tempo, onde o intervalo de tempo aqui é o intervalo de tempo entre FFTs sucessivos. Se, agora, os FFTs inversos estão sendo separados entre si, isto significa que o aumento de fase de 45° ocorre através de um intervalo de tempo mais longo. Isto significa que devido à mudança de fase uma não correspondência no processo de sobreposição-adição subsequente ocorre, conduzindo ao cancelamento de sinal não desejado. Para eliminar este artefato, a fase é rescalada por exatamente o mesmo fator pelo qual o sinal de áudio foi espalhado no tempo. A fase de cada valor espectral FFT é, dessa maneira, aumentada pelo fator b/a , de modo que esta não correspondência é eliminada.

Embora na configuração ilustrada na Figura 5c, o espalhamento por interpolação dos sinais de controle de amplitude/frequência foi atingido para um oscilador de sinal na implementação de banco de filtro da Figura 5a, o espalhamento na Figura 6 é atingido pela distância entre dois espectros IFFT maiores que a distância entre dois espectros FFT, isto é, b sendo maior que a , onde, entretanto, para uma prevenção de artefato, um rescalamento de fase é executado de acordo com b/a .

Com relação a uma descrição detalhada de vocoders

de fase referência é feita aos documentos a seguir:

"The phase Vocoder: A tutorial", Mark Dolson, Computer Music Journal, volume 10, nº 4, pág. 14 - 27, 1986, ou "New phase Vocoder techniques for pitch-shifting, harmonizing and other exotic effects", L. Laroche e M. Dolson, Proceedings 1999 IEEE Workshop on applications of signal processing to audio and acoustics, New Paltz, New York, 17 a 20 de Outubro de 1999, páginas 91 a 94; "New approached to transient processing interphase vocoder", A. Röbel, Proceeding of the 6th international conference on digital audio effects (DAFx-03), Londres, RU, 8 a 11 de Setembro de 2003, páginas DAFx-1 a DAFx-6; "Phase-locked Vocoder", Meller Puckette, Proceedings 1995, IEEE ASSP, Conference on applications of signal processing to audio and acoustics, ou Pedido de Patente Norte-Americana Número 6.549.884.

15 Alternativamente, outros métodos para espalhamento de sinal estão disponíveis, tal como, por exemplo, o método 'Pitch Synchronous Overlap Add'. Adição de Sobreposição Sincronizada de Tom, abreviada como PSOLA, é um método de síntese na qual gravações de sinais de voz estão localizadas no banco de dados. Como estes são sinais periódicos, os mesmos são providos com informação sobre frequência fundamental (tom) e o início de cada período é marcado. Na síntese, estes períodos são cortados com um certo ambiente por meio de uma função de janela, e adicionados ao sinal a ser sintetizado em um local adequado:
20 Dependendo do fato da frequência fundamental desejada ser mais elevada ou mais baixa do que aquela da entrada do banco de dados, elas são combinadas consequentemente mais densa ou menos densa do que no original. Para ajuste da duração do audível, períodos podem

ser omitidos ou enviados em duplicidade. Este método é também denominado de TD-PSOLA, onde TD significa domínio de tempo e enfatiza que os métodos operam no domínio de tempo. Um desenvolvimento adicional é o método de Adição de Sobreposição de Re-síntese de Bandas Múltiplas, abreviado como MBROLA. Aqui, os segmentos no banco de dados são trazidos para uma frequência fundamental uniforme por um pré-processamento e a posição de fase da harmônica é normalizada. Por isto, na síntese de uma transição de um segmento para o próximo, interferências menos perceptivas resultam e a qualidade de voz atingida é mais alta.

Em uma alternativa adicional, o sinal de áudio já foi filtrado por passagem de banda antes do espalhamento, de modo que o sinal após espalhamento e decimação já contém as porções desejadas e a filtragem de passagem de banda subsequente pode ser omitida. Neste caso, o filtro de passagem de banda é definido de modo que a porção do sinal de áudio que teria sido filtrada após extensão de largura de banda está ainda contida no sinal de saída do filtro de passagem de banda. O filtro de passagem de banda, dessa maneira, contém uma faixa de frequência que não está contida no sinal de áudio após espalhamento e decimação. O sinal com esta faixa de frequência é o sinal desejado formando o sinal de alta frequência sintetizado.

O manipulador de sinal conforme ilustrado na Figura 1 pode, adicionalmente, compreender o condicionador de sinal 130 para processamento adicional do sinal de áudio com a transiente "natural" não processada ou sintetizada na linha 121. Este condicionador de sinal pode ser um decimador de sinal dentro de uma aplicação de extensão de largura de banda, que, na sua

saída, gera um sinal de alta banda, que pode, então, ser adicionalmente adaptado para assemelhar-se bastante com as características do sinal de banda elevada original pelo uso de parâmetros de alta frequência (HF) a serem transmitidos juntamente 5 com um fluxo de dados HFR (reconstrução de alta frequência).

As Figuras 7a e 7b ilustram um cenário de extensão de largura de banda, que pode, vantajosamente, usar o sinal de saída do condicionador de sinal dentro do codificador de extensão de largura de banda 720 da Figura 7b. Um sinal de áudio é 10 alimentado em uma combinação de baixa passagem/alta passagem em uma entrada 700. A combinação de baixa passagem/alta passagem por um lado inclui uma passagem baixa (LP), para gerar uma versão filtrada de baixa passagem do sinal de áudio 700, ilustrada em 703 na Figura 7a. Este sinal de áudio filtrado de baixa passagem é 15 codificado com um codificador de áudio 704. O codificador de áudio é, por exemplo, um codificador MP3 (MPEG1 Camada 3) ou um codificador AAC, também conhecido como um codificador MP4 e descrito no Padrão MPEG4. Codificadores de áudio alternativos, provendo uma representação transparente ou vantajosamente 20 transparente perceptualmente do sinal de áudio limitado de banda 703, podem ser usados no codificador 704 para gerar um sinal de áudio 705 codificado completamente ou codificado perceptualmente e preferivelmente codificado perceptualmente de forma transparente, respectivamente.

25 A banda superior do sinal de áudio é enviada em uma saída 706 pela porção de alta passagem do filtro 702, designado por "HP". A porção de alta passagem do sinal de áudio, isto é, a banda superior ou banda HF, também designada como a

porção HF, é suprida para um calculador de parâmetro 707 que é implementado para calcular os diferentes parâmetros. Estes parâmetros são, por exemplo, o envelope espectral da banda superior 706 em uma resolução relativamente grosseira, por 5 exemplo, pela representação de um fator de escala para cada grupo de frequência psicoacústica ou para cada banda Bark na escala Bark, respectivamente. Um parâmetro adicional que pode ser calculado pelo calculador de parâmetro 707 é o piso de ruído na banda superior, cuja energia por banda pode preferivelmente estar 10 relacionada à energia do envelope nesta banda. Parâmetros adicionais que podem ser calculados pelo calculador de parâmetro 707 incluem uma medida de tonalidade para cada banda parcial da banda superior que indica como a energia espectral está distribuída em uma banda, isto é, se a energia espectral na banda 15 está distribuída relativamente de forma uniforme, onde, então, um sinal não tonal existe nesta banda, ou se a energia nesta banda está relativamente concentrada fortemente em uma certa localização na banda, onde, então, apenas um sinal tonal existe para esta banda.

20 Parâmetros adicionais consistem em explicitamente codificar picos relativamente se projetando fortemente na banda superior com relação a sua altura e sua frequência, como o conceito de extensão de largura de banda, na reconstrução sem esta codificação explícita de porções sinusóides proeminentes na banda 25 superior, recuperará apenas os mesmos de forma muito rudimentar, ou não serão recuperados.

Em qualquer caso, o calculador de parâmetro 707 é implementado para gerar apenas parâmetros 708 para a banda

superior que pode ser submetida a etapas de redução de entropia similares, visto que eles podem ser executados no codificador de áudio 704 para valores espectrais quantizados, tal como, por exemplo, codificação diferencial, predição ou codificação de Huffman, etc.. A representação de parâmetro 708 e o sinal de áudio 705 são, então, supridos para um formatador de fluxo de dados 709 que é implementado para prover um fluxo de dados lateral de saída 710 que tipicamente será um fluxo de bits de acordo com um certo formato, como é, por exemplo, padronizado no padrão MPEG4.

O lado do decodificador, como é especialmente adequado para a presente invenção, é ilustrado a seguir com relação à Figura 7b. O fluxo de dados 710 entra em um interpretador de fluxo de dados 711 que é implementado para separar a porção de parâmetro relacionada à extensão de largura de banda 708 da porção de sinal de áudio 705. A porção de parâmetro 708 é decodificada por um decodificador de parâmetro 712 para obter parâmetros decodificados 713. Paralelamente a isto, a porção de sinal de áudio 705 é decodificada por um decodificador de áudio 714 para obter um sinal de áudio...

Dependendo da implementação, o sinal de áudio 100 pode ser enviado por meio de uma primeira saída 715. Na saída 715, um sinal de áudio com uma pequena largura de banda e, dessa maneira, também uma baixa qualidade pode, então, ser obtido. Para melhoramento da qualidade, entretanto, a extensão de largura de banda inventiva 720 é executada para obter o sinal de áudio 712 no lado da saída com uma largura de banda estendida ou alta, respectivamente, e dessa maneira uma alta qualidade.

É conhecido da WO 98/57436 como submeter o sinal

de áudio a uma limitação de banda em uma situação no lado do codificador e a codificar apenas uma banda inferior do sinal de áudio por meio de um codificador de áudio de alta qualidade. A banda superior, entretanto, é apenas muito grosseiramente caracterizada, isto é, por um conjunto de parâmetros que reproduz o envelope espectral da banda superior. No lado do decodificador, a banda superior é, então, sintetizada. Para este objetivo, uma transposição harmônica é proposta, onde a banda inferior do sinal de áudio decodificado é suprida para um banco de filtro. Canais de banco de filtro da banda inferior são conectados a canais de banco de filtro da banda superior, ou são "remendados" e cada sinal de passagem de banda remendado é submetido a um ajuste de envelope. O banco de filtro de síntese pertencendo a um banco de filtro de análise especial aqui recebe sinais de passagem de banda do sinal de áudio na banda inferior e sinais de passagem de banda ajustados a envelope da banda inferior que foram harmonicamente remendados na banda superior. O sinal de saída do banco de filtro de síntese é um sinal de áudio estendido com relação a sua largura de banda, que foi transmitido do lado do codificador para o lado do decodificador com uma taxa de dados muito baixa. Em especial, cálculos de banco de filtro e remendos no domínio de banco de filtro podem se tornar um grande esforço computacional.

O método apresentado aqui resolve os problemas mencionados. A novidade inventiva do método consiste do fato de que, em contraste com os métodos existentes, uma porção de janela, que contém o transiente, é removida do sinal a ser manipulado, e de que do sinal original, uma segunda porção de janela (geralmente diferente da primeira porção) é adicionalmente selecionada, que

pode ser reinserida no sinal manipulado, de modo que o envelope temporal seja preservado tanto quanto possível no ambiente transiente. Esta segunda porção é selecionada de modo que irá precisamente se ajustar no recesso alterado pela operação de 5 prolongamento de tempo. O encaixe preciso é executado pelo cálculo do máximo de correlação cruzada das bordas do recesso resultante com as bordas da porção transiente original.

Dessa maneira, a qualidade de áudio subjetiva do transiente não é mais prejudicada pelo espalhamento e por efeitos 10 de eco.

Determinação precisa da posição do transiente para o objetivo de selecionar uma porção adequada pode ser executada, por exemplo, usando um cálculo centróide em movimento da energia em relação a um período de tempo adequado.

15 Juntamente com o fator de prolongamento de tempo, o tamanho da primeira porção determina o tamanho requerido da segunda porção. Preferivelmente, este tamanho deve ser selecionado de modo que mais que um transiente é acomodado pela segunda porção usada para reinserção apenas se o intervalo de tempo entre os 20 transientes proximamente adjacentes está abaixo do limite para perceptibilidade humana dos eventos temporais individuais.

Ajuste ideal do transiente de acordo com a correlação cruzada máxima pode requerer um ligeiro deslocamento no tempo em relação à posição original do mesmo. Entretanto, devido à 25 existência de efeitos temporais pré- e, especificamente, pós-mascaramento, a posição do transiente reinserido não necessita corresponder precisamente com a posição original. Devido ao período estendido de ação do pós-mascaramento, uma mudança do

transiente na direção de tempo positiva deve ser preferida.

Pela inserção da porção de sinal original, o timbre ou tom da mesma será mudado quando a taxa de amostragem for mudada por uma etapa de decimação subsequente. De forma geral, 5 entretanto, isto é mascarado pelo próprio transiente por meio de mecanismos de mascaramento temporais psicoacústicos. Em especial, se prolongamento por um fator inteiro ocorre, o timbre será mudado apenas ligeiramente, visto que fora do ambiente do transiente, apenas toda enésima ($n=$ fator de prolongamento) onda harmônica 10 será ocupada.

Usando o novo método, artefatos (espalhamento, pré- e pós-ecos) que resultam durante o processamento de transientes por meios de prolongamento de tempo e de métodos de transposição são efetivamente impedidos. Danos potenciais na 15 qualidade de porções de sinal sobrepostas (possível tonal) são evitados.

O método é adequado para quaisquer aplicações de áudio onde as velocidades de reprodução de sinais de áudio ou seus tons devem ser alterados.

20 Subsequentemente, uma configuração preferida no contexto das Figuras 8a a 8e é discutida. A Figura 8a ilustra uma representação do sinal de áudio, mas em contraste com uma sequência de amostra de áudio de domínio de tempo direto, a Figura 8a ilustra uma representação de envelope de energia, que pode, por 25 exemplo, ser obtida quando cada amostra de áudio em uma ilustração de amostra de domínio de tempo é quadrada. Especificamente, a Figura 8a ilustra um sinal de áudio 800 tendo um evento transiente 801 onde o evento transiente é caracterizado por um aumento e

diminuição agudos de energia com o decorrer do tempo. Naturalmente, um transiente seria também um aumento agudo de energia quando esta energia permanece em um certo nível alto ou uma diminuição aguda de energia quando a energia estava em um alto nível para um certo tempo antes da diminuição. Um padrão específico para um transiente é, por exemplo, um bater palmas ou qualquer outro tom gerado por um instrumento de percussão. Adicionalmente, transientes são ataques rápidos de um instrumento, que inicia tocando um tom alto, isto é, que provê energia de som em uma certa banda ou uma pluralidade de bandas acima de certo nível de limite abaixo de certo tempo limite. Naturalmente, outra flutuação de energia tal como a flutuação de energia 802 do sinal de áudio 800 na Figura 8a não é detectada como transientes. Detectores de transiente são conhecidos na técnica e são extensivamente descritos na literatura e se baseiam em muitos algoritmos diferentes, que podem compreender processamento seletivo à frequência e uma comparação de um resultado de um processamento seletivo à frequência com um limite e uma decisão subsequente sobre a existência de um transiente ou não.

A Figura 8b ilustra um transiente de janela. A área delimitada pela linha sólida é subtraída do sinal ponderado pelo formato da janela ilustrada. A área marcada pela linha pontilhada é adicionada novamente após processamento. Especificamente, o transiente ocorrendo em um certo tempo transiente 803 tem que ser cortado do sinal de áudio 800. Para estar no lado seguro, não apenas o transiente, mas também algumas amostras adjacentes/vizinhas devem ser cortadas do sinal original. Portanto, a primeira porção de tempo 804 é determinada, onde a

primeira porção de tempo se estende de um instante de tempo inicial 805 até um instante de tempo final 806. De forma geral, a primeira porção de tempo 804 é selecionada de modo que o tempo transiente 803 está incluído dentro da primeira porção de tempo 804. A Figura 8c ilustra um sinal sem um transiente antes de ser prolongado. Como pode ser visto das bordas degradadas lentamente 807 e 808, a primeira porção de tempo não é apenas cortada por um filtro/janela retangular, mas uma janela é executada para ter bordas ou flancos degradados lentamente do sinal de áudio.

Importante observar que a Figura 8c agora ilustra o sinal de áudio na linha 102 da Figura 1, isto é, subsequente à remoção do sinal transiente. Os flancos de degradação/aumento lentos 807, 808 provêem a região de *fade-in* ou *fade-out* a ser usada pelo *cross-fader* 128 da Figura 4. A Figura 8d ilustra o sinal da Figura 8c, mas em um estado prolongado, isto é, subsequente ao processamento aplicado pelo processador de sinal 110. Dessa maneira, o sinal na Figura 8d é o sinal na linha 111 da Figura 1. Devido à operação de prolongamento, a primeira porção 804 se tornou muito mais longa. Dessa maneira, a primeira porção 804 da Figura 8d foi prolongada até a segunda porção de tempo 809, que tem uma instante inicial da segunda porção de tempo 810 e um instante final da segunda porção de tempo 811. Prolongando o sinal, os flancos 807, 808, foram prolongados também, de modo que o comprimento de tempo dos flancos 807', 808' foi prolongado também. Este prolongamento deve ser considerado quando calculando o comprimento da segunda porção de tempo conforme executado pelo calculador 122 da Figura 4.

Tão logo o comprimento da segunda porção de tempo

é determinado, uma porção correspondente ao comprimento da segunda porção de tempo é cortada do sinal de áudio original ilustrado na Figura 8a, conforme indicado pela linha intermitente na Figura 8b. Com este fim, a segunda porção de tempo 809 foi inserida na Figura 8e. Conforme discutido, o instante de tempo inicial 812, isto é, o primeiro ponto limite da segunda porção de tempo 809 no sinal de áudio original e o instante de tempo final 813 da segunda porção de tempo, isto é, o segundo ponto limite da segunda porção de tempo no sinal de áudio original não têm, necessariamente, que serem simétricos com relação ao tempo de evento transiente 803, 803', de modo que o transiente 801 está localizado exatamente no mesmo instante no tempo como estava no sinal original. Ao invés disso, os instantes de tempo 812, 813 da Figura 8b podem ser ligeiramente variados, de modo que os resultados da correlação cruzada entre um formato de sinal nestes pontos limites no sinal original são, tanto quanto possível, similares a porções correspondentes no sinal prolongado. Dessa maneira, a posição real do transiente 803 pode ser retirada do centro da segunda porção de tempo até certo grau, que é indicado na Figura 8e pelo número de referência 803' indicando certo tempo com relação à segunda porção de tempo, que desvia do tempo 803 correspondente com relação à segunda porção de tempo na Figura 8b. Conforme discutido em conexão com a Figura 4, item 126, uma mudança positiva do transiente para um tempo 803' com relação a um tempo 803 é preferida devido ao efeito de pós-mascaramento, que é mais pronunciado do que o efeito de pré-mascaramento. A Figura 8e adicionalmente ilustra as regiões de cruzamento/transição 813a, 813b nas quais o *cross-fader* 128 provê um *cross-fader* entre o

sinal prolongado sem o transiente e a cópia do sinal original incluindo o transiente.

Conforme ilustrado na Figura 4, o calculador para calcular o comprimento da segunda porção de tempo 122 é configurado para receber o comprimento da primeira porção de tempo e o fator de prolongamento. Alternativamente, o calculador 122 pode também receber uma informação sobre a permissibilidade de transientes vizinhos a serem incluídas dentro de uma e da mesma primeira porção de tempo. Portanto, com base nesta permissibilidade, o calculador pode determinar o comprimento da primeira porção de tempo 804 por si só e, dependendo do fator de prolongamento/encurtamento, calcular então o comprimento da segunda porção de tempo 809.

Conforme discutido acima, a funcionalidade do insensor de sinal é que o insensor de sinal remove uma área adequada para o intervalo na Figura 8e, que é aumentada dentro do sinal prolongado a partir do sinal original e se encaixa nesta área adequada, isto é, a segunda porção de tempo no sinal processado usando um cálculo de correlação cruzada para determinar o instante de tempo 812 e 813 e, preferivelmente, executar uma operação de *cross-fade* nas regiões de *cross-fade* 813a e 813b, também.

A Figura 9 ilustra um equipamento para gerar informação de lado para um sinal de áudio, que pode ser usada no contexto da presente invenção quando a detecção de transiente é executada no lado do codificador e informação de lado referente a esta detecção de transiente é calculada e transmitida para um manipulador de sinal, que então representaria o lado do

decodificador. Para esta finalidade, um detector de transiente similar ao detector de transiente 103 na Figura 2 é aplicado para analisar o sinal de áudio incluindo um evento transiente. O detector de transiente calcula um tempo transiente, isto é, tempo 5 803 na Figura 1 e envia este tempo transiente para um calculador de metadados 104', que pode ser estruturado similarmente para o calculador de *fade-out/fade-in* 104' na Figura 2. De forma geral, o calculador de metadados 104' pode calcular metadados a serem enviados para uma interface de saída de sinal 900 onde estes 10 metadados podem compreender limites para a remoção de transiente, isto é, limites para a primeira porção de tempo, isto é limites 805 e 806 da figura 8b ou limites para a inserção de transiente (segunda porção de tempo) conforme ilustrado em 812, 813 na Figura 8b ou o instante de tempo de evento transiente 803 ou ainda 803'. 15 Mesmo no último caso, o manipulador de sinal estaria em posição para determinar todos os dados requeridos, isto é, os dados da primeira porção de tempo, os dados da segunda porção de tempo, etc., baseado em um instante de tempo de evento transiente 803.

Os metadados conforme gerados pelo item 104' são 20 enviados para a interface de saída de sinal de modo que a interface de saída de sinal gera um sinal, isto é, um sinal de saída para transmissão ou armazenagem. O sinal de saída pode incluir apenas os metadados ou pode incluir os metadados e o sinal de áudio onde, em último caso, os metadados representariam 25 informação de lado para o sinal de áudio. Para esta finalidade, o sinal de áudio pode ser enviado para a interface de saída de sinal 900 por meio da linha 901. O sinal de saída gerado pela interface de saída de sinal 900 pode ser armazenado em qualquer espécie de

meio de armazenagem ou pode ser transmitido por meio de qualquer tipo de canal de transmissão para um manipulador de sinal ou qualquer outro dispositivo requerendo informação transitória.

Deve ser observado que embora a presente invenção 5 tenha sido descrita no contexto de diagramas de bloco onde os blocos representam componentes de hardware real ou lógico, a presente invenção pode também ser implementada por um método implementado por computador. No último caso, os blocos representam etapas de método correspondentes onde estas etapas representam as 10 funcionalidades executadas pelos blocos de hardware lógicos ou físicos correspondentes.

As configurações descritas são meramente ilustrativas dos princípios da presente invenção. É entendido que modificações e variações dos arranjos e dos detalhes descritos 15 aqui ficarão aparentes para aqueles especializados na técnica. É objetivo, portanto, estar limitado apenas pelo escopo das reivindicações de patente independentes e não pelos detalhes específicos apresentados como descrição e explanação das configurações aqui apresentadas.

20 Dependendo de certos requisitos de implementação dos métodos inventivos, os métodos inventivos podem ser implementados em hardware ou em software. A implementação pode ser executada usando um meio de armazenagem digital, em especial, um disco, um DVD ou um CD tendo sinais de controles eletronicamente 25 legíveis armazenados nos mesmos, que cooperam com sistemas de computador programáveis de modo que os métodos inventivos sejam executados. De forma geral, o presente pode, portanto, ser implementado como um produto de programa de computador com um

código de programa armazenado em um veículo legível por máquina, o código de programa sendo operado para executar os métodos inventivos quando o produto de programa de computador opera em um computador. Em outras palavras, os métodos inventivos são, 5 portanto, um programa de computador tendo um código de programa para executar pelo menos um dos métodos inventivos quando o programa de computador opera em um computador. O sinal de metadados inventivo pode ser armazenado em qualquer meio de armazenagem legível por máquina tal como uma mídia de armazenagem 10 digital.

REIVINDICAÇÕES

1. Equipamento para a manipulação de um sinal de áudio tendo um evento transiente (801), **caracterizado por** compreender:

um processador de sinal (110) para o processamento de um sinal de áudio com transiente reduzido, no qual uma primeira porção de tempo (804) compreendendo o evento transiente (801) é removida ou, para o processamento de um sinal de áudio compreendendo o evento transiente (803), para obtenção de um sinal de áudio processado;

um insensor de sinal (120) para a inserção de uma segunda porção de tempo (809) no sinal de áudio processado em um local de sinal, onde a primeira porção foi removida ou onde o evento transiente está localizado no sinal de áudio processado;

pelo fato de que a segunda porção de tempo (809) compreende um evento transiente (801) não influenciado pelo processamento realizado pelo processador de sinal (110);

em que o processador de sinal (110) está configurado para gerar a porção transitória perceptivamente degradada no sinal de áudio processado, esticando ou encurtando o sinal de áudio que compreende o evento transitório (801), de modo que o sinal de áudio processado tenha uma duração maior ou menor que a sinal de áudio, e

que a segunda porção de tempo (809) tem uma duração diferente da duração da primeira porção de tempo (804), em que no caso de alongamento, a segunda porção de tempo (809) é maior que a primeira porção de tempo (804) ou em caso de encurtamento, a segunda porção de tempo (809) é menor que a

primeira porção de tempo (804).

2. Equipamento, de acordo com qualquer uma das reivindicações anteriores, **caracterizado por** compreender um processador de sinal (110) compreende um vocoder, um vocoder de fase ou um processador (P) SOLA.

3. Equipamento, de acordo com qualquer uma das reivindicações anteriores, **caracterizado por** compreender ainda um condicionador de sinal (130) para o condicionamento do sinal de áudio manipulado pela decimação ou interpolação de uma versão discreta no tempo do sinal de áudio manipulado.

4. Equipamento, de acordo com qualquer uma das reivindicações anteriores, no qual o insensor de sinal (120) é caracterizado: para a determinação (122) de um tempo de duração de uma segunda porção de tempo (809) a ser copiada do sinal de áudio tendo o evento transiente (801) em uma primeira porção de tempo (804), para a determinação (123) de um instante de tempo inicial da segunda porção de tempo (809) ou de um instante de tempo final da segunda porção de tempo (809) preferencialmente por meio da descoberta de uma máxima de um cálculo de correlação cruzada, de modo que um ponto limite da segunda porção de tempo (809) corresponda a um ponto limite correspondente do sinal de áudio processado preferencialmente ao máximo possível, em que uma posição no tempo (803') do evento transiente no sinal de áudio manipulado coincide com a posição (803) no tempo do evento transiente (801) no sinal de áudio ou desvia da posição no tempo do evento transiente (801) no sinal de áudio por uma diferença de tempo menor que um grau psico-acusticamente tolerável determinado por um pré-mascaramento ou pós-mascaramento do evento transiente (801).

5. Equipamento, de acordo com qualquer uma das reivindicações anteriores, **caracterizado por** compreender ainda um detector de transiente (103) para a detecção do evento transiente (801) no sinal de áudio, ou compreendendo ainda um extrator de informação lateral (106) para a extração e interpretação de uma informação lateral associada ao sinal de áudio, a informação lateral indicando uma posição no tempo (803) do evento transiente (801) ou indicando um instante de tempo inicial ou um instante de tempo final da primeira porção de tempo (804) ou da segunda porção de tempo (809).

6. Método de manipulação de um sinal de áudio tendo um evento transiente (801) em uma primeira porção de tempo (804) de um sinal de áudio, **caracterizado por** compreender:

processar (110) o sinal de áudio compreendendo o evento transitório (801) na primeira parte do tempo (804) para obter um sinal de áudio processado compreendendo uma porção transitória perceptivamente degradada;

inserir (120) uma segunda porção de tempo (809) no sinal de áudio processado em um local de sinal, em que a porção transitória perceptivamente degradada está localizada no sinal de áudio processado, de modo que um sinal de áudio manipulado seja obtido, em que a segunda porção de tempo (809) compreende o evento transitório (801) não influenciado pela etapa de processamento (110);

em que a etapa de processamento (110) gera a porção transitória perceptivamente degradada no sinal de

áudio processado, esticando ou encurtando o sinal de áudio que compreende o evento transitório (801), de modo que o sinal de áudio processado tenha uma duração maior ou menor que o áudio sinal, e

em que a segunda porção de tempo (809) tem uma duração diferente da duração da primeira porção de tempo (804), em que no caso de alongamento, a segunda porção de tempo (809) é maior que a primeira porção de tempo (804) ou em caso de encurtamento, a segunda porção de tempo (809) é menor que a primeira porção de tempo (804).

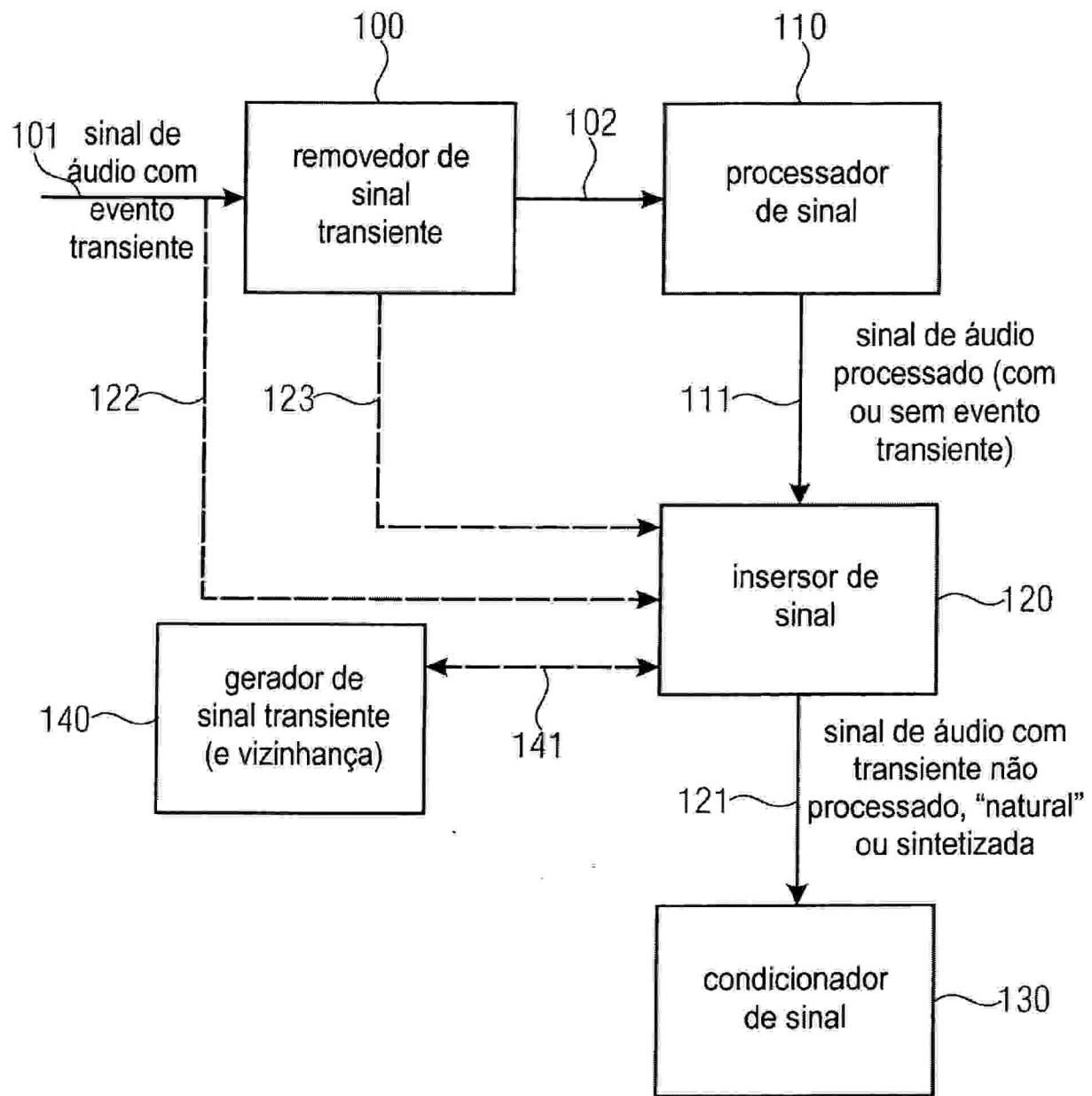


FIGURA 1

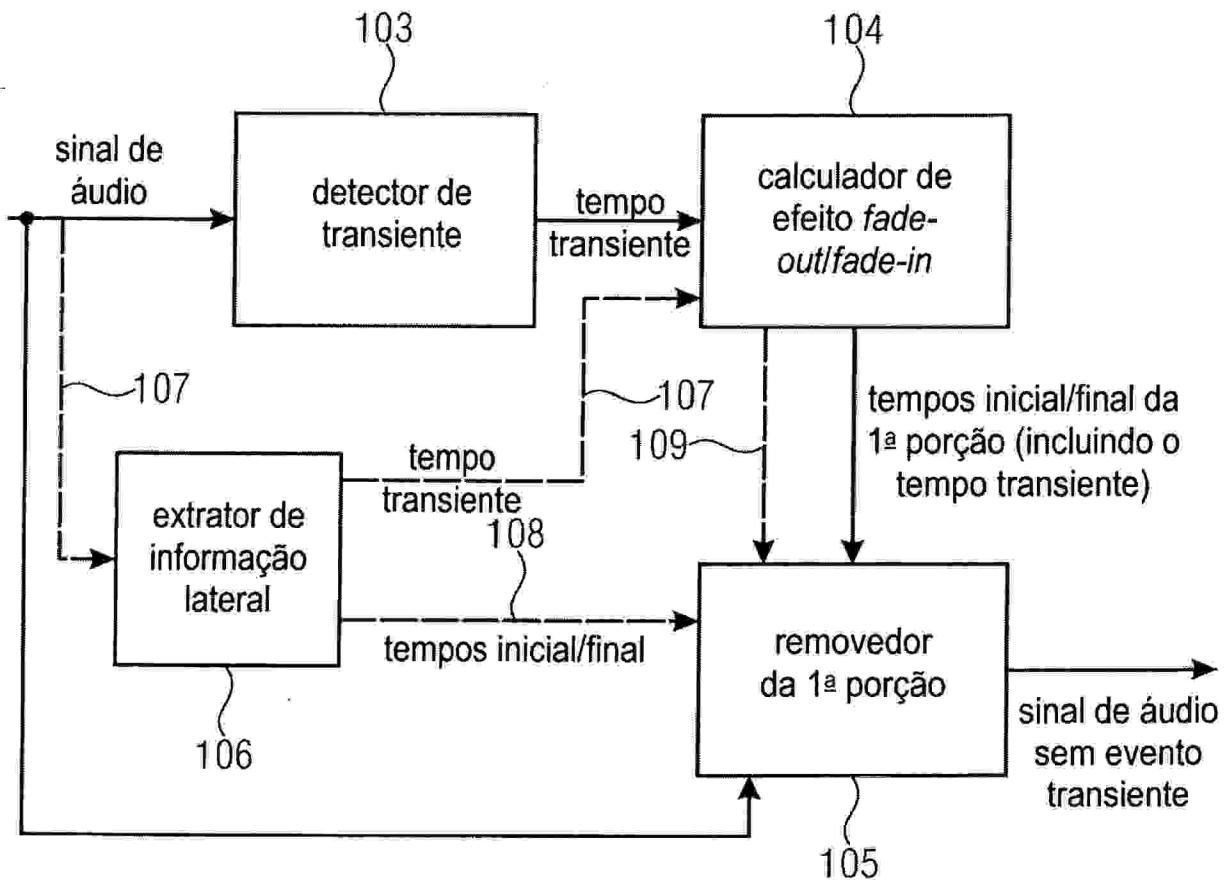


FIGURA 2

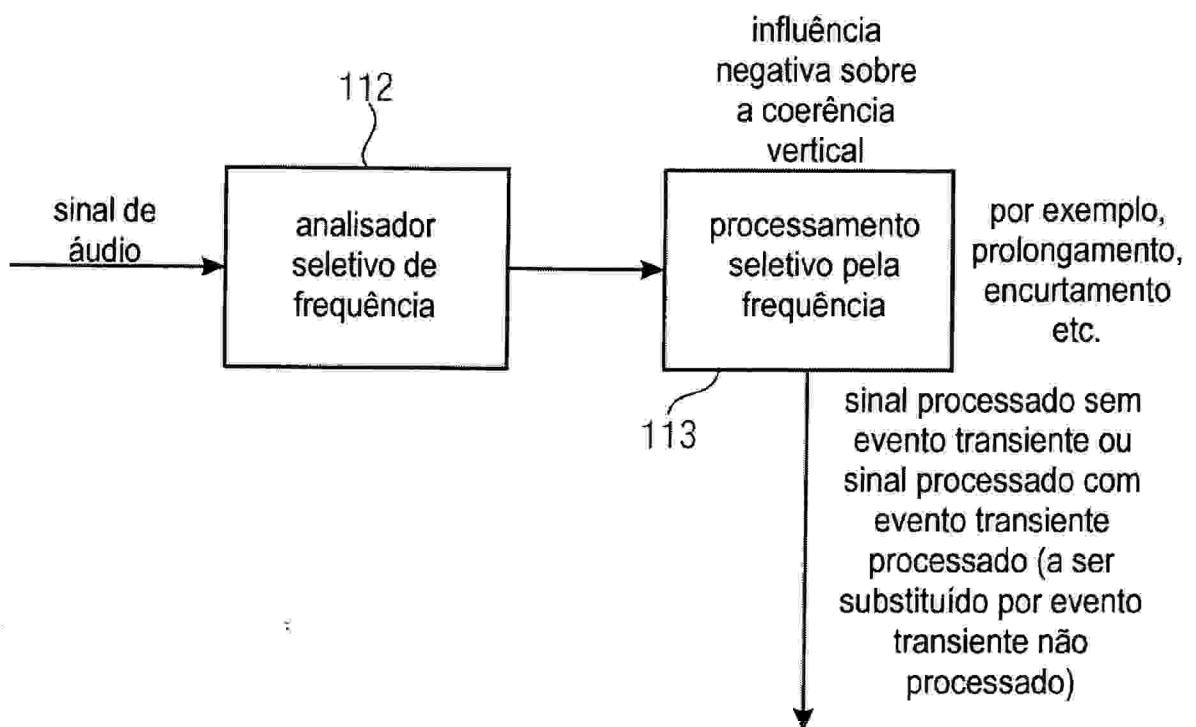


FIGURA 3A

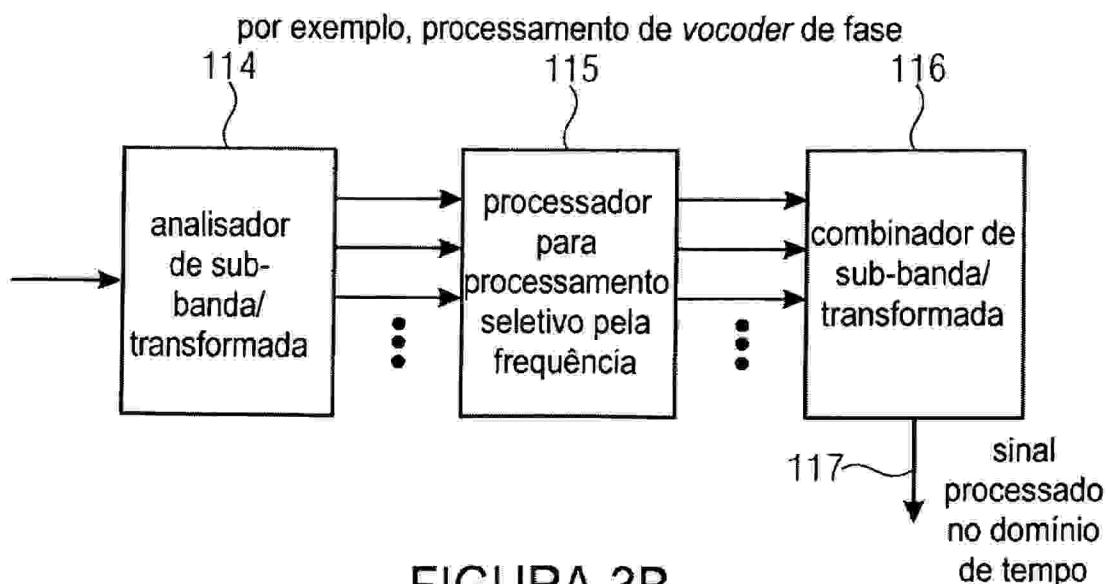


FIGURA 3B

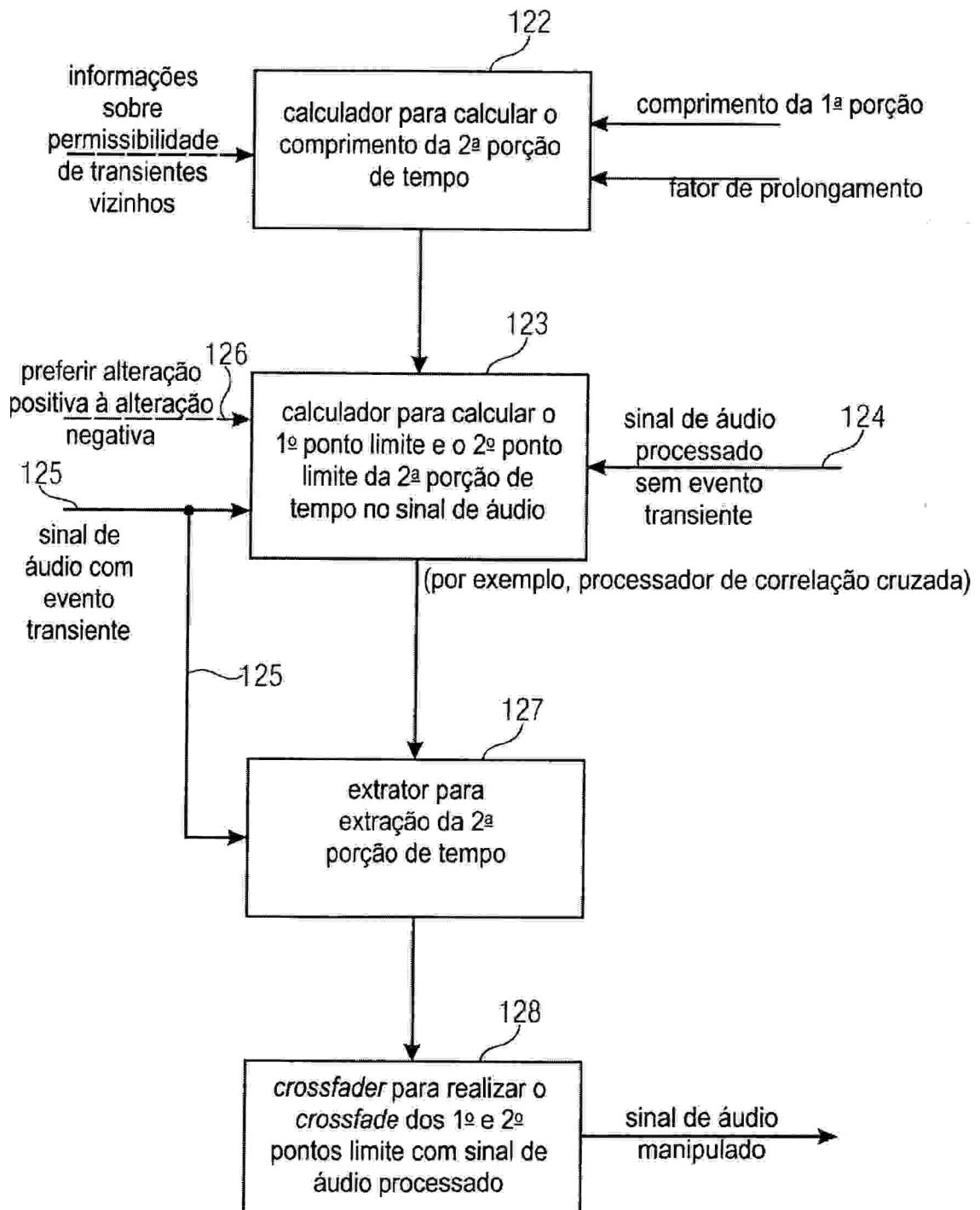


FIGURA 4

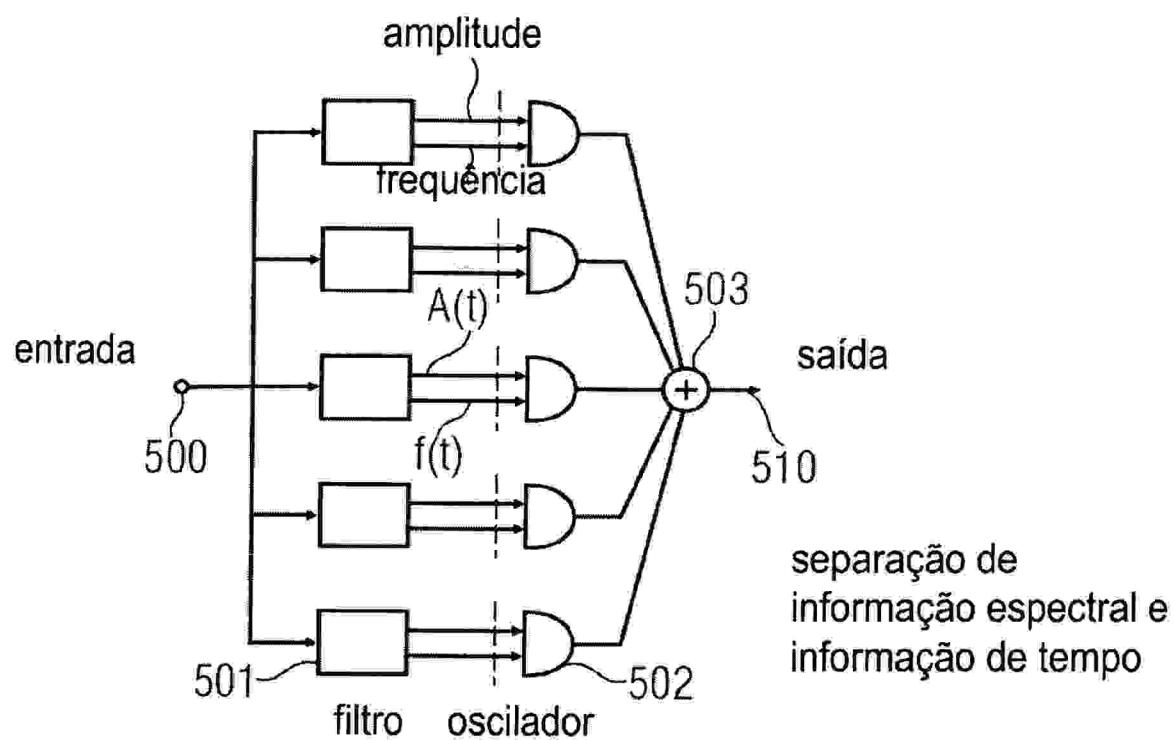


FIGURA 5A
(implementação do banco de filtro)

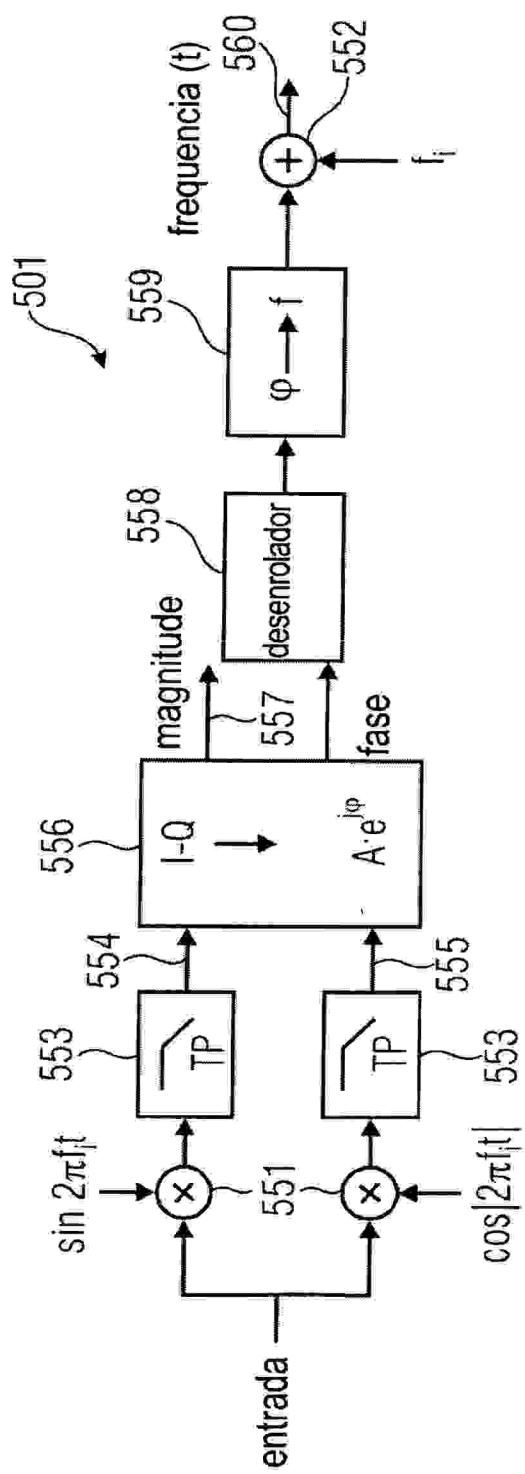


FIGURA 5B

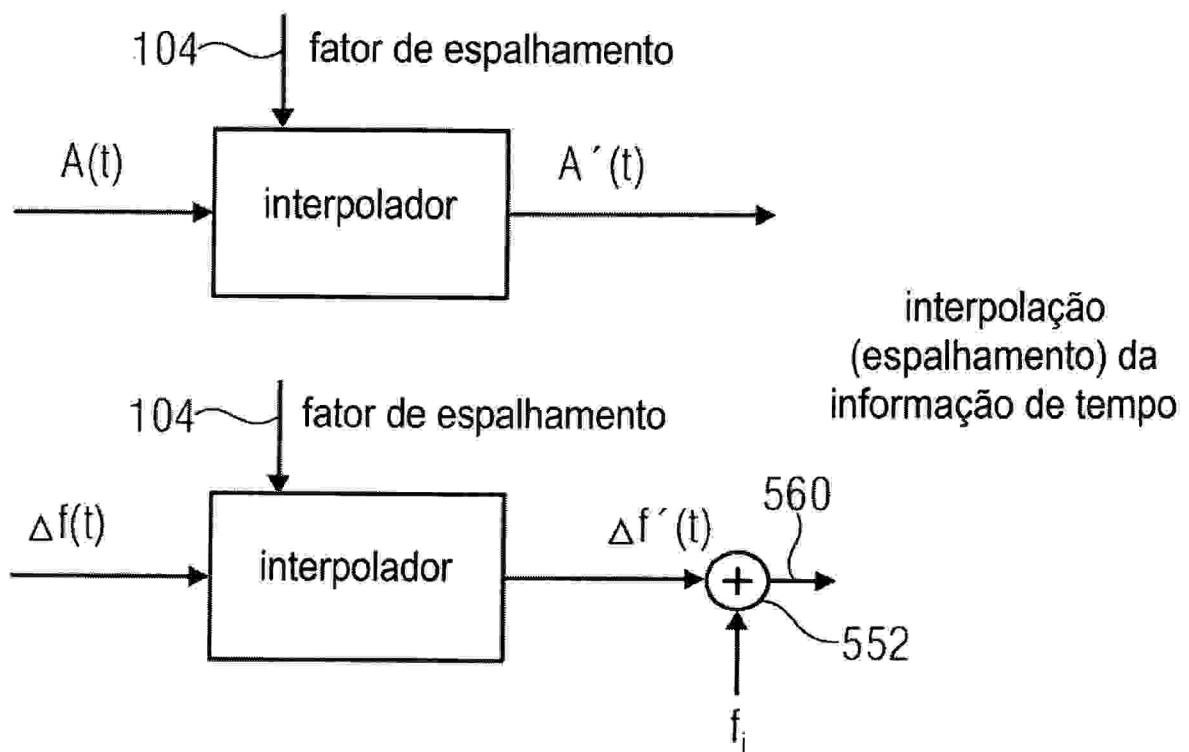


FIGURA 5C

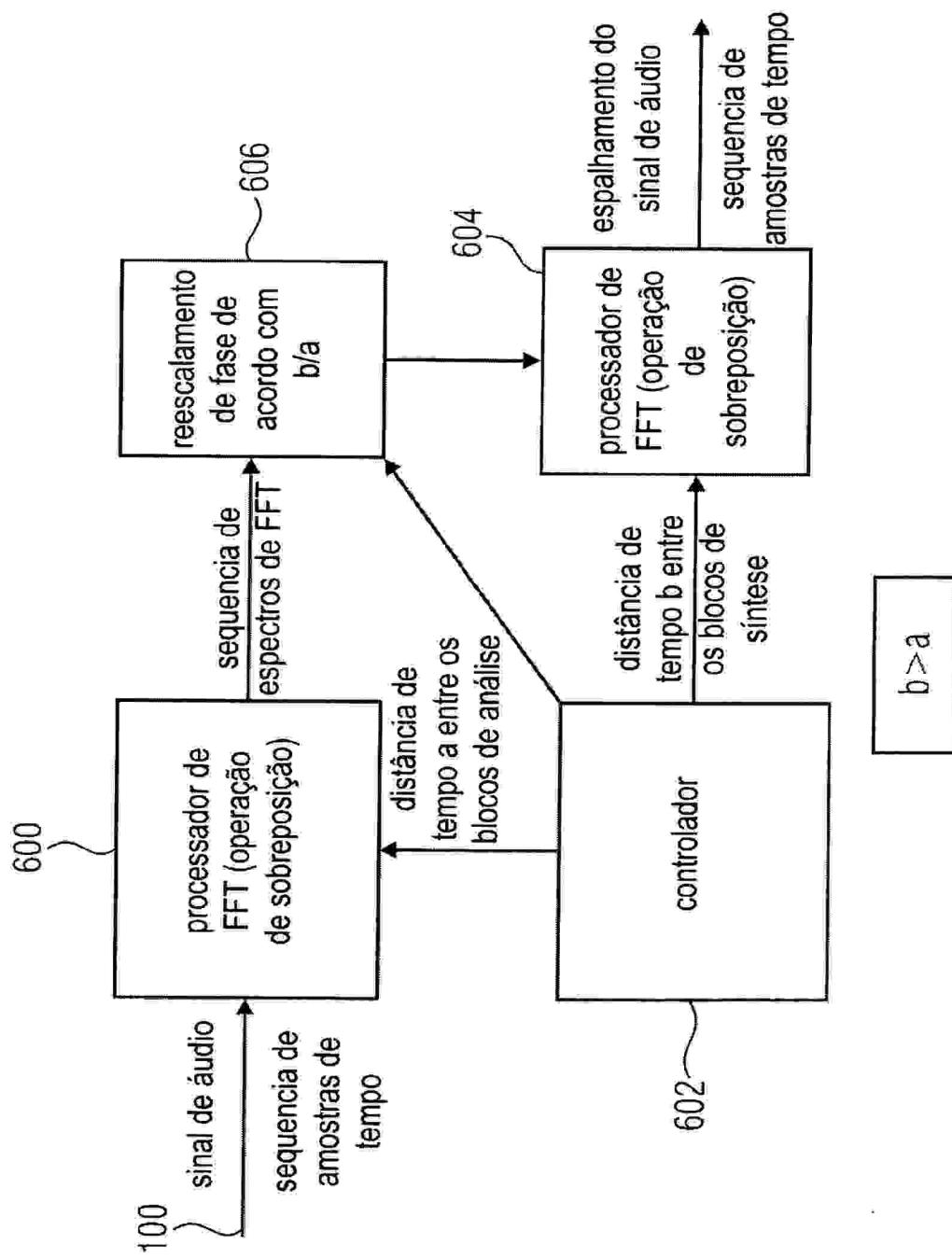


FIGURA 6
(implementação de transformada)

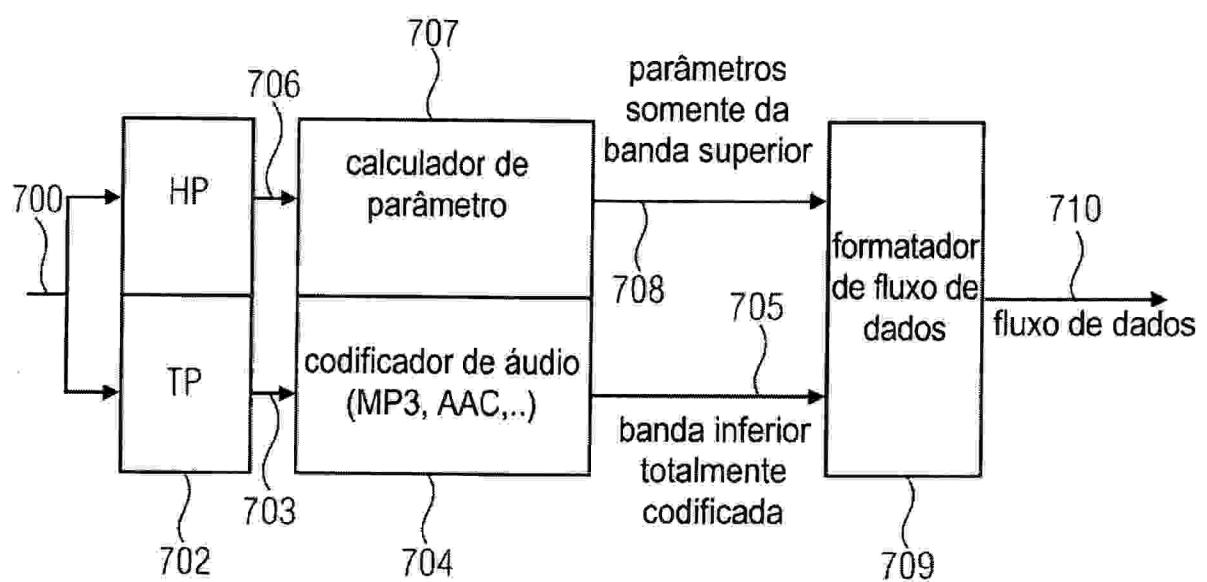


FIGURA 7A
(Lado do Codificador)

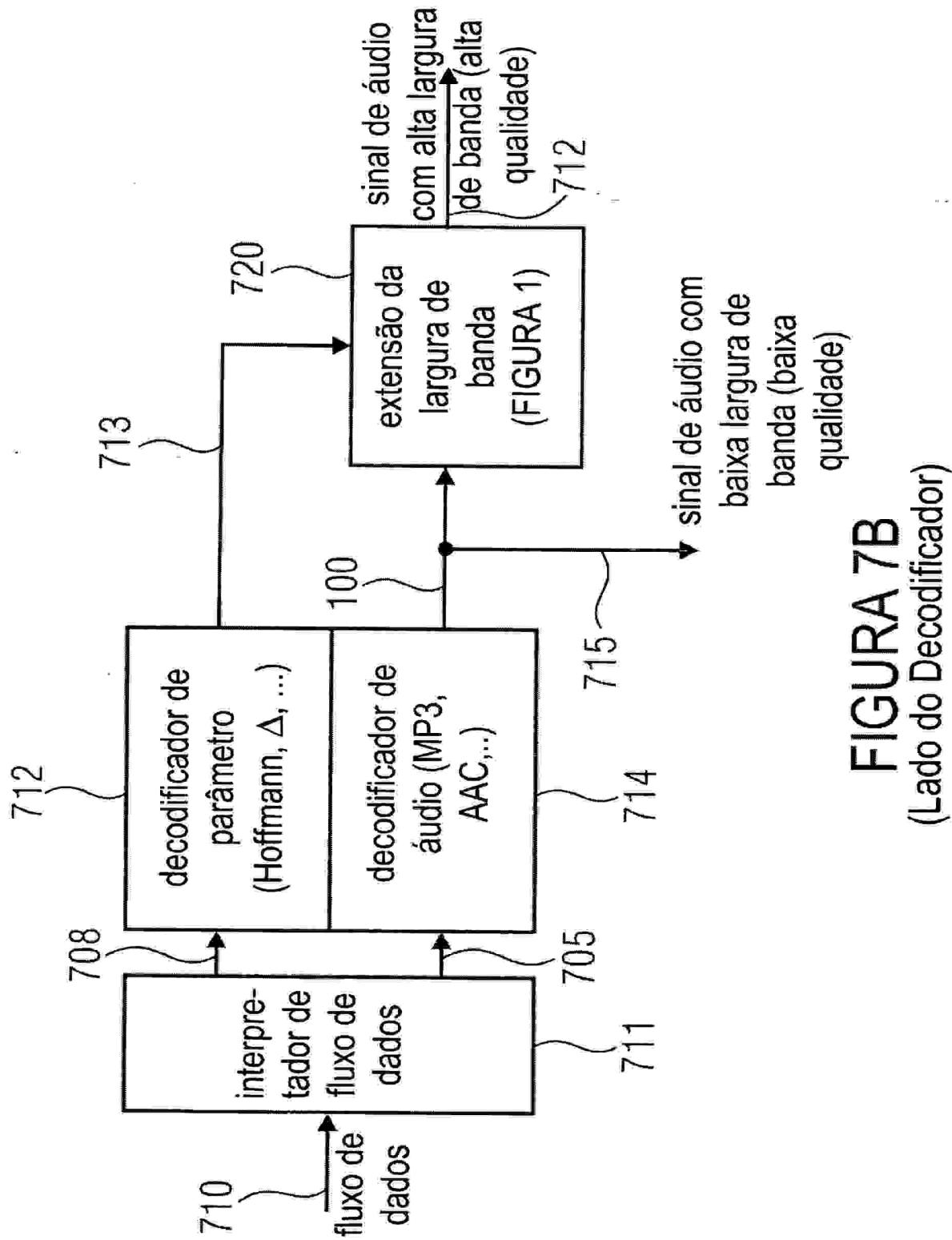


FIGURA 7B
(Lado do Decodificador)

11/13

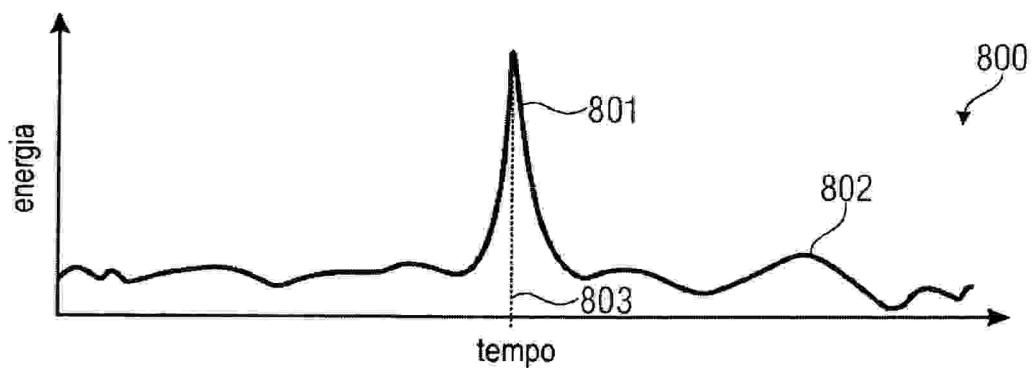


FIGURA 8A

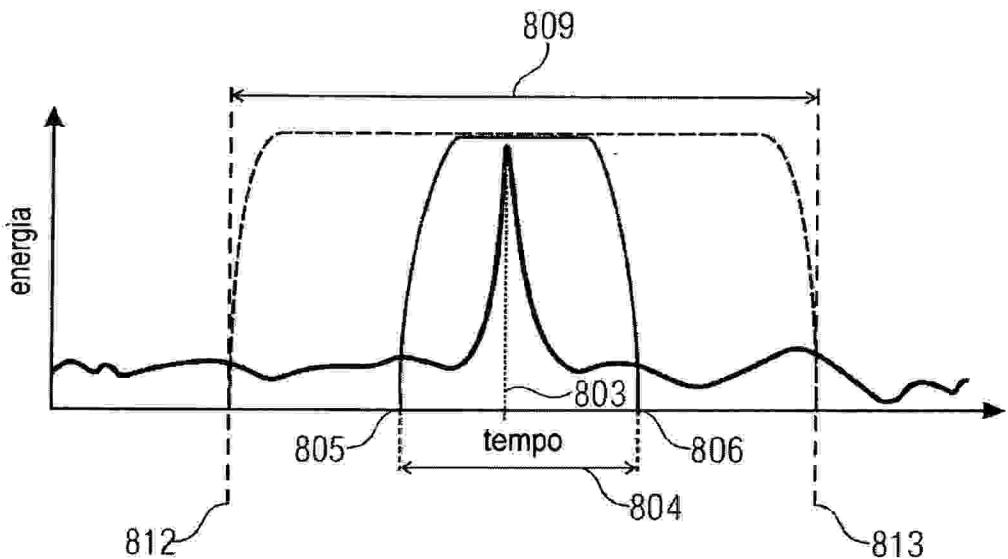


FIGURA 8B

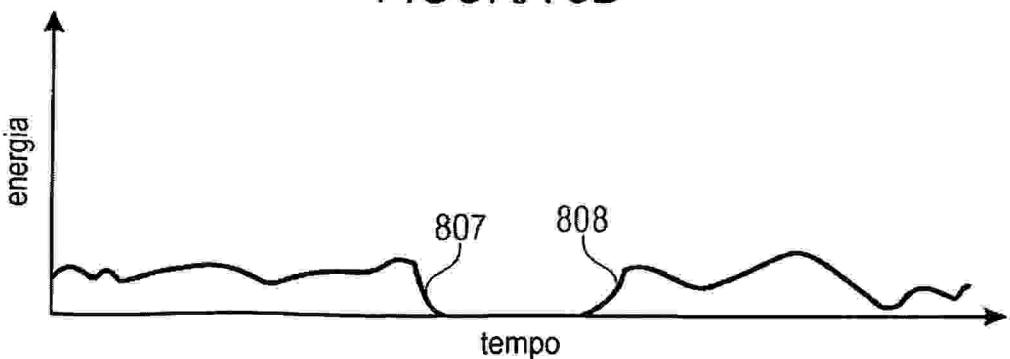


FIGURA 8C

12/13

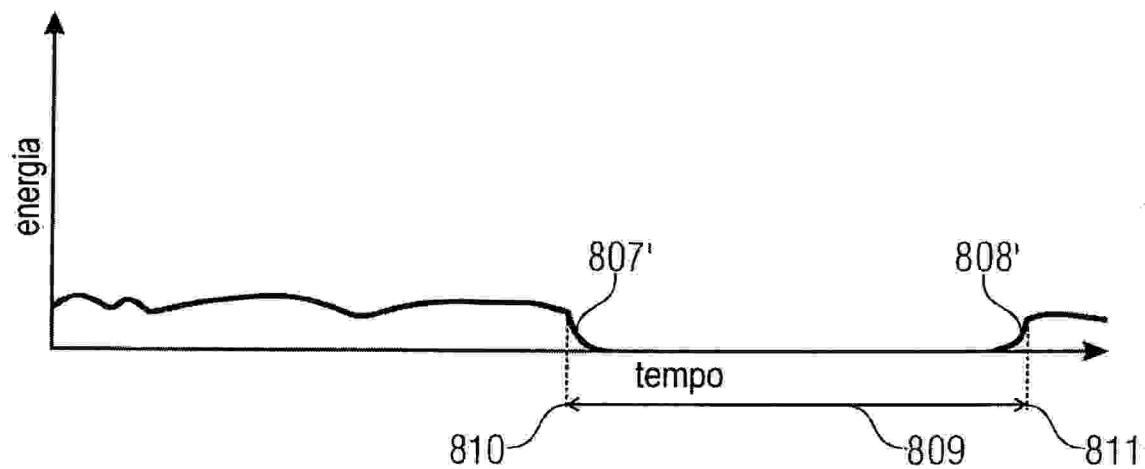


FIGURA 8D

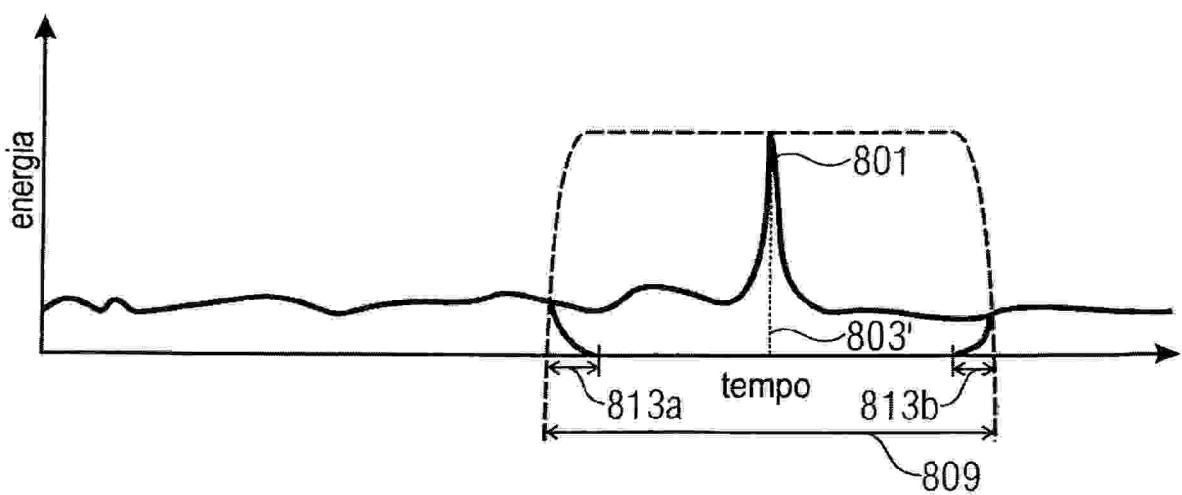


FIGURA 8E

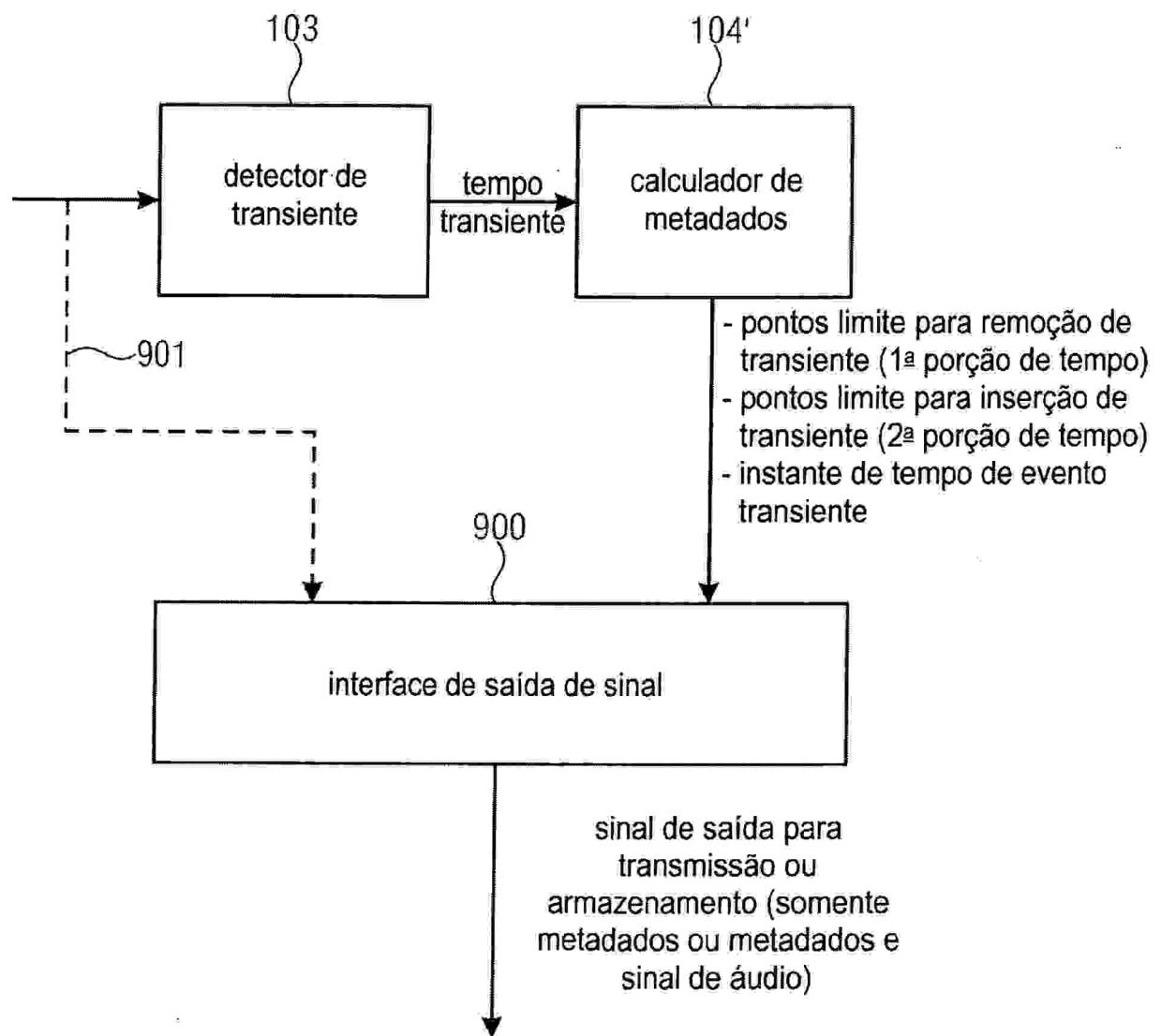


FIGURA 9