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

(11) BR 112016000947-9 B1

(22) Data do Depósito: 15/07/2014

(45) Data de Concessão: 21/06/2022

(54) Título: APARELHO E MÉTODO PARA DECODIFICAÇÃO DE UM SINAL DE ÁUDIO CODIFICADO UTILIZANDO UM FILTRO CRUZADO AO REDOR DE UMA FREQUÊNCIA DE TRANSIÇÃO

(51) Int.Cl.: G10L 21/0388; G10L 19/028; G10L 19/02.

(30) Prioridade Unionista: 18/10/2013 EP 13189389.3; 22/07/2013 EP 13177348.3; 22/07/2013 EP 13177346.7; 22/07/2013 EP 13177350.9; 22/07/2013 EP 13177353.3.

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

(72) Inventor(es): NAGEL FREDERIK; GEIGER RALF; NEUKAM CHRISTIAN; HELMRICH CHRISTIAN; DISCH SASCHA; SCHMIDT KONSTANTIN; FISCHER MICHAEL.

(86) Pedido PCT: PCT EP2014065112 de 15/07/2014

(87) Publicação PCT: WO 2015/010950 de 29/01/2015

(85) Data do Início da Fase Nacional: 15/01/2016

(57) Resumo: APARELHO E MÉTODO PARA DECODIFICAÇÃO DE UM SINAL DE ÁUDIO CODIFICADO UTILIZANDO UM FILTRO CRUZADO AO REDOR DE UMA FREQUÊNCIA DE TRANSIÇÃO. Aparelho para decodificação de um sinal de áudio codificado, caracterizado por compreender um sinal principal codificado (1), compreendendo: um decodificador principal (1400) para decodificar o sinal principal codificado (1401) para obter um sinal principal decodificado; um gerador de porção (1404) para gerar uma ou mais porção(ões) espectral(is), tendo frequências não incluídas no sinal principal decodificado, utilizando uma porção espectral do sinal principal decodificado, e um filtro cruzado (1406) para filtrar de forma cruzada e espectral o sinal principal decodificado e uma primeira porção de frequência, tendo frequências estendendo-se de uma frequência de preenchimento de lacuna (309) a uma frequência de limite superior, ou filtrar de forma cruzada e espectral uma primeira porção de frequência e uma segunda porção de frequência. Figura 14a

"APARELHO E MÉTODO PARA DECODIFICAÇÃO DE UM SINAL DE ÁUDIO CODIFICADO UTILIZANDO UM FILTRO CRUZADO AO REDOR DE UMA FREQUÊNCIA DE TRANSIÇÃO".

RELATÓRIO DESCRITIVO

[0001] A presente invenção refere-se à codificação/decodificação de áudio e, particularmente, à codificação de áudio utilizando Preenchimento de Lacuna Inteligente (IGF | *Intelligent Gap Filling*).

[0002] A codificação de áudio é o domínio de compressão de sinal que lida com a exploração da redundância e irrelevância nos sinais de áudio utilizando o conhecimento psicoacústico. Atualmente, os codecs de áudio tipicamente precisam de aproximadamente 60 kbps/canal para codificar de forma perceptualmente transparente a codificação de quase qualquer tipo de sinal de áudio. Codecs recentes têm o objetivo de reduzir a taxa de bit de codificação explorando as similaridades espectrais no sinal utilizando técnicas como extensão da largura de banda (BWE | *bandwidth extension*). Um esquema de BWE utiliza um parâmetro da taxa de bit baixa definida para representar os componentes de alta frequência (HF | *high frequency*) de um sinal de áudio. O espectro de HF é preenchido com o conteúdo espectral das regiões de baixa frequência (LF | *low frequency*) e a forma espectral, inclinação e continuidade temporal para manter o timbre e cor do sinal original. Tais métodos de BWE permitem que os codecs de áudio retenham a boa qualidade em taxas de bit baixas uniformes de aproximadamente 24 kbps/canal.

[0003] O sistema de codificação de áudio inventivo

codifica eficientemente os sinais de áudio arbitrários em uma ampla faixa de taxas de bit. Enquanto isso, para altas taxas de bit, o sistema inventivo converge em transparência, para distúrbios perceptuais das baixas taxas de bit é reduzido. Assim, o compartilhamento principal da taxa disponível é utilizado para a forma de onda codificar apenas a estrutura perceptualmente mais relevante do sinal no codificador, e as lacunas espectrais resultantes são preenchidas no decodificador com o conteúdo do sinal que aproximadamente aproxima o espectro original. Um montante de bit muito limitado é consumido para controlar o chamado Preenchimento de Lacuna Inteligente (IGF) espectral acionado por parâmetro pela informação adicional dedicada transmitida do codificador ao decodificador.

[0004] O armazenamento ou transmissão dos sinais de áudio é geralmente submetido às restrições de taxa de bit rigorosas. No passado, os codificadores foram forçados a reduzir drasticamente a largura de banda de áudio transmitida quando apenas uma taxa de bit muito baixa estava disponível.

[0005] Codecs de áudio modernos são atualmente capazes de codificar sinais de banda ampla utilizando os métodos de extensão da largura de banda (BWE) [1]. Estes algoritmos dependem de uma representação paramétrica do conteúdo de alta frequência (HF) - que é gerada da parte de baixa frequência (LF) codificada da forma de onda do sinal decodificado por meios de transposição na região espectral HF ("correção" (*patching*)) e aplicação de um pós-processamento acionado por parâmetro. Nos esquemas de BWE, a reconstrução

da região espectral HF acima de chamada frequência cruzada é geralmente com base na correção espectral. Tipicamente, a região HF é composta por várias correções adjacentes e cada uma dessas correções é proveniente das regiões de passa-banda (BP | *band-pass*) do espectro LF abaixo da dada frequência cruzada. Os sistemas do estado da técnica realizam a correção dentro de uma representação do banco de filtro, por exemplo, Banco de Filtro em Espelho de Quadratura (QMF | *Quadrature Mirror Filterbank*), copiando um conjunto de coeficientes de sub-banda adjacentes de uma região de origem para a região alvo.

[0006] Outra técnica encontrada nos codecs de áudio atuais que aumenta a eficiência da compressão e assim permite a largura de banda de áudio estendida em baixas taxas de bit é a substituição sintética acionada por parâmetro de partes adequadas dos espectros de áudio. Por exemplo, partes do sinal do tipo ruído do sinal de áudio original podem ser substituídos sem perda substancial da qualidade subjetiva pelo ruído artificial gerado no decodificador e escalado pelos parâmetros de informação adicional. Ume exemplo é a ferramenta de Substituição do Ruído Perceptual (PNS | *Perceptual Noise Substitution*) contida na Codificação de Áudio Avançada MPEG-4 (AAC | *Advanced Audio Coding*) [5].

[0007] Outra disposição que também permite a largura de banda de áudio estendida em baixas taxas de bit é a técnica de preenchimento de ruído contida na Codificação de Áudio e Discurso Unificado de MPEG-D (USAC | *Unified Speech and Audio Coding*) [7]. As lacunas espectrais (zeros) que são

deduzidas pela zona morta do quantizador devido a uma quantização muito rústica são subsequentemente preenchidas com ruído artificial no decodificador e escaladas por um pós-processamento acionado por parâmetro.

[0008] Outro sistema do estado da técnica é chamado Substituição Espectral Precisa (ASR | *Accurate Spectral Replacement*) [2-4]. Além de um codec da forma de onda, a ASR emprega um estágio de síntese do sinal dedicado que recupera perceptualmente partes sinusoidais importantes do sinal no decodificador. Ainda, um sistema descrito em [5] depende da modelagem sinusoidal na região HF de um codificador da forma de onda para permitir que largura de banda de áudio estendida tendo qualidade perceptual satisfatória em baixas taxas de bit. Todos estes métodos envolvem a transformação dos dados em um segundo domínio longe da transformada de cosseno discreta modificada (MDCT | *modified discrete cosine transform*) e ainda estágios de análise/síntese bastante complexos para a preservação dos componentes sinusoidais HF.

[0009] A figura 13a ilustra um diagrama esquemático de um codificador de áudio para uma tecnologia de extensão de largura de banda como, por exemplo, utilizada na Codificação de Áudio Avançada de Alta Eficiência (HE-AAC | *High Efficiency Advanced Audio Coding*). Um sinal de áudio na linha (1300) é inserido em um sistema de filtro compreendendo um passa-baixa (1302) e um passa alto (1304). O sinal emitido pelo filtro passa-alta (1304) é inserido em um codificador/extrator de parâmetro (1306). O codificador/extrator de parâmetro (1306) é configurado para

calcular e codificar os parâmetros como um parâmetro do envelope espectral, um parâmetro de preenchimento de ruído, um parâmetro ausente harmônico, ou um parâmetro de filtragem inversa, por exemplo. Estes parâmetros extraídos são inseridos em um multiplexador de fluxo contínuo de bits (1308). O sinal de saída passa-baixa é inserido em um processador tipicamente compreendendo a funcionalidade de um redutor da taxa de amostragem (1310) e um codificador principal (1312). O passa-baixa (1302) restringe a largura de banda a ser codificada em uma largura de banda significantemente menor do que ocorre no sinal de áudio de entrada original na linha (1300). Isso fornece um ganho de codificação significante devido ao fato de que todas as funcionalidades que ocorrem no codificador principal devem apenas operar em um sinal com uma largura de banda reduzida. Quando, por exemplo, a largura de banda do sinal de áudio na linha (1300) é 20 kHz e quando o filtro passa-baixa (1302) exemplarmente tem uma largura de banda de 4 kHz, a fim de realizar a o teorema da amostragem, é teoricamente suficiente que o sinal subsequente ao redutor da taxa de amostragem tem uma frequência de amostragem de 8 kHz, que é a redução substancial à taxa de amostragem necessária para o sinal de áudio (1300) que deve ser pelo menos 40 kHz.

[00010] A Figura 13b ilustra um diagrama esquemático de um decodificador da extensão da largura de banda correspondente. O decodificador compreende um multiplexador do fluxo contínuo de bit (1320). O demultiplexador do fluxo contínuo de bit (1320) extrai um sinal de entrada para um

decodificador principal (1322) e um sinal de entrada para um decodificador de parâmetro (1324). Um sinal de saída do decodificador principal tem, no exemplo acima, uma taxa de amostragem de 8 kHz e, assim, uma largura de banda de 4 kHz enquanto, para uma completa reconstrução da largura de banda, o sinal de saída de um reconstrutor de alta frequência (1330) deve ser a 20 kHz exigindo uma taxa de amostragem de, pelo menos, 40 kHz. A fim de tornar isso possível, um processador do decodificador tendo a funcionalidade de um amplificador da taxa de amostragem (1325) e um banco de filtro (1326) é exigido. O reconstrutor de alta frequência (1330), então, recebe o sinal de baixa frequência analisado por frequência emitido pelo banco de filtro (1326) e reconstrói a faixa de frequência definida pelo filtro passa-alta (1304) da Figura 13a utilizando a representação paramétrica da banda de alta frequência. O reconstrutor de alta frequência (1330) tem várias funcionalidades como a regeneração da faixa de frequência superior utilizando a faixa de origem na faixa de frequência baixa, um ajuste do envelope espectral, uma funcionalidade de preenchimento de ruído e uma funcionalidade para introduzir harmonias ausentes na faixa de frequência superior e, se aplicada e calculada no codificador da Figura 13a, uma operação de filtragem inversa a fim de contar o fato que a faixa de frequência mais alta não é tipicamente tão tonal quanto a faixa de frequência inferior. Em HE-AAC, as harmonias ausentes são sintetizadas novamente no lado do decodificador e são colocados exatamente no meio de uma banda de reconstrução. Assim, todas as linhas de harmonia ausente

que foram determinadas em uma determinada banda de reconstrução não são colocadas nos valores de frequência onde eles estavam localizados no sinal original. Ainda, estas linhas de harmonia ausente são colocadas nas frequências no centro da banda principal. Assim, quando uma linha de harmonia ausente no sinal original foi colocada muito próxima ao limite da banda de reconstrução no sinal original, o erro na frequência introduzida colocando esta linha de harmonias ausentes no sinal reconstruído no centro da banda é quase 50% da banda de reconstrução individual, para os parâmetros que foram gerados e transmitidos.

[00011] Além disso, embora os codificadores típicos de áudio centrais operem no domínio espectral, o decodificador principal gera independentemente um sinal de domínio de tempo que é, então, novamente convertido em um domínio espectral pela funcionalidade do banco de filtro 1326. Isso introduz os atrasos de processamento adicional, pode introduzir perturbações devido ao processamento tandem, primeiramente transformando do domínio espectral ao domínio de frequência e novamente transformando em um domínio de frequência tipicamente diferente e, certamente, também exige uma quantidade substancial da complexidade de computação e, assim, energia elétrica que é especificamente uma questão quando a tecnologia de extensão de largura de banda é aplicada em dispositivos móveis como celulares, *tablet* ou *notebooks*, etc.

[00011] Codecs de áudio atuais realizam codificação de áudio de baixa taxa de bit utilizando BWE como uma parte

integral do esquema de codificação. Entretanto, as técnicas de BWE são restritas para substituir apenas o conteúdo de alta frequência. Além disso, elas não permitem que o conteúdo perceptualmente importante acima de uma dada frequência cruzada seja codificado pela forma de onda. Assim, os codecs de áudio temporários tanto perdem o detalhe da HF quanto o timbre quando a BWE é implementada, visto que o alinhamento exato da harmonia tonal do sinal não é considerado na maioria dos sistemas.

[00012] Outra desvantagem dos sistemas de BWE do estado atual da técnica é a necessidade para transformação do sinal de áudio em um novo domínio para implementação da BWE (p.ex., transformação do domínio de MDCT em QMF). Isso leva a complicações de sincronização, complexidade computacional adicional e exigências de memória elevada.

[00013] O armazenamento ou a transmissão de sinais de áudio é geralmente submetido à rigorosas restrições de taxa de bit. No passado, os codificadores eram forçados a reduzir muito a largura de banda do áudio transmitido apenas quando uma taxa de bit muito baixa estava disponível. Os codecs de áudio modernos são, atualmente, capazes de codificar os sinais de banda ampla utilizando os métodos de extensão da largura de banda (BWE) [1-2]. Esses algoritmos dependem de uma representação paramétrica do conteúdo de alta frequência (HF) - que é gerada a partir da parte de baixa frequência codificada por forma de onda (LF) do sinal decodificado por meios da transposição na região espectral de HF (“correção”) e aplicação de um pós-processamento acionado por parâmetro.

[00014] Nos esquemas de BWE, a reconstrução da região espectral de HF acima de uma determinada e denominada frequência cruzada é gerada frequentemente com base na correção espectral. Outros esquemas que são funcionais para preencher as lacunas espectrais, p.ex., o Preenchimento de Lacuna Inteligente (IGF), utilizam as denominadas porções espectrais proximais para regenerar as partes dos espectros HF do sinal de áudio. Tipicamente, a região HF é composta por múltiplas correções ou porções adjacentes e cada uma dessas correções ou porções é proveniente das regiões de passa-banda (BP) do espectro LF abaixo de uma determinada frequência cruzada. Os sistemas do estado da técnica realizam eficientemente a correção ou partição dentro de uma representação do banco de filtro pela cópia de um conjunto de coeficientes de sub-bandas adjacentes de uma região de origem para a região alvo. Ainda, para alguns conteúdos de sinal, a montagem do sinal reconstruído da banda LF e correções adjacentes dentro da banda HF podem levar à pulsação, dissonância e aspereza auditiva.

[00015] Assim, em [19], o conceito de filtragem de banda de guarda da dissonância é apresentado no contexto de um sistema de BWE com base no banco de filtro. Sugere-se aplicar efetivamente um filtro rejeita-faixa de aproximadamente 1 largura de banda na Escala de Bark na frequência cruzada entre LF e HF regenerada por BWE para evitar a possibilidade de dissonância e substituir o conteúdo espectral com zeros ou ruído.

[00016] Entretanto, a solução proposta em [19] tem

algumas desvantagens: Primeira, a restrita substituição do conteúdo espectral por zeros ou ruído pode também prejudicar a qualidade perceptual do sinal. Além disso, o processamento proposto não é adaptativo ao sinal e pode, assim, prejudicar a qualidade perceptual em alguns casos. Por exemplo, se o sinal contém transientes, isso pode levar a pré- e pós-ecos.

[00017] Segunda, as dissonâncias também podem ocorrer nas transições entre as correções HF consecutivas. A solução proposta em [19] é apenas funcional para remediar as dissonâncias que ocorrem na frequência cruzada entre LF e HF regenerada por BWE.

[00018] Finalmente, conforme oposto aos sistemas com base no banco de filtro, como proposto em [19], os sistemas de BWE podem também ser realizados em implementações à base de transformada, como, por exemplo, a Transformada de cosseno discreta modificada (MDCT | *Modified Discrete Cosine Transform*). As transformadas, como a MDCT, são muito propensas à denominada modulação de frequência [20] ou perturbações de toque que ocorrem se as regiões de passa-banda dos coeficientes espectrais são copiadas ou os coeficientes espectrais são definidos a zero, conforme proposto em [19].

[00019] Particularmente, a Patente Norte-Americana 8.412.365 revela o uso, na translação ou dobramento à base de banco de filtro, das denominadas bandas de guarda que são inseridas e feitas de um ou vários canais de sub-banda definidos a zero. Um número de canais de banco de filtro é utilizado como bandas de guarda e uma largura de banda de uma

banda de guarda deveria ser 0.5 na Escala de Bark. Essas bandas de guarda de dissonância são parcialmente reconstruídas utilizando sinais de ruído branco aleatórios, ou seja, as sub-bandas são alimentadas com ruído branco ao invés de ser zero. As bandas de guarda são inseridas independentemente do sinal atual ao processado.

[00020] Os sistemas de extensão de largura de banda são particularmente problemáticos quando são realizados em implementações com base na transformada, como, por exemplo, a Transformada de cosseno discreta modificada (MDCT). As transformadas, como MDCT e outras transformadas, são muito propensas à denominada modulação de frequência, conforme discutido em [3], e perturbações por toque que ocorrem se as regiões de passa-banda dos coeficientes espectrais forem copiadas ou os coeficientes espectrais forem definidos a zero, conforme proposto em [2].

[00021] É o objeto da presente invenção fornecer um aparelho e método melhorados para decodificação de um sinal de áudio codificado.

[00022] Esse objeto é alcançado por um aparelho para decodificação de um sinal de áudio codificado de acordo com a reivindicação 1, um método para decodificação de um sinal de áudio codificado de acordo com a reivindicação 15 ou um programa de computador de acordo com a reivindicação 16.

[00023] De acordo com a presente invenção, um aparelho para decodificação de um sinal de áudio codificado compreende um decodificador principal, um gerador de porção para gerar uma ou mais porção(ões) espectral(is) tendo

frequências não incluídas no sinal principal decodificado utilizando uma porção espectral do sinal principal decodificado e um filtro cruzado para filtragem cruzada espectral do sinal principal decodificado e uma primeira porção de frequência tendo frequências se estendendo de uma frequência de preenchimento de lacuna em uma primeira frequência de parada de porção ou para filtragem cruzada espectral de uma porção e uma porção de frequência adicional, a porção de frequência adicional tendo uma frequência de limite inferior sendo adjacente à frequência em uma frequência de limite superior da porção de frequência.

[00024] Preferivelmente, esse procedimento se destina a ser aplicado dentro de uma extensão de largura de banda com base em uma transformada como a MDCT. Entretanto, a presente invenção é geralmente aplicável e, particularmente em um cenário de extensão de largura de banda que depende de um banco de filtro em espelho de quadratura (QMF), particularmente se o sistema for criticamente testado, por exemplo, quando há uma representação de QMF com valor real como uma conversão de tempo-frequência ou como uma conversão de frequência-tempo.

[00025] A presente invenção é particularmente útil para sinais do tipo transiente, pois para tais sinais do tipo transiente, o toque é uma perturbação audível e irritante. As perturbações de toque de filtro são causadas pela chamada característica de parede de tijolos de um filtro na banda de transição, ou seja, uma transição íngreme de um passa-banda para uma banda de parada em uma frequência de corte. Tais

filtros podem ser eficientemente implementados pelo ajuste de um coeficiente ou grupos de coeficientes para zero em um domínio de frequência de uma transformada de tempo-frequência. Assim, a presente invenção depende de um filtro cruzado em cada frequência de transição entre as correções/porções ou entre uma banda principal e uma primeira correção/porção para reduzir essa perturbação de toque. O filtro cruzado é preferivelmente implementado pela ponderação espectral no domínio da transformada que emprega as funções de ganho adequadas.

[00026] Preferivelmente, o filtro cruzado é adaptativo ao sinal e consiste em dois filtros, um filtro de *fade-out* (conclusão), que é aplicado na região espectral inferior, e um filtro de *fade-in* (início), que é aplicado na região espectral mais alta. Os filtros podem ser simétricos ou assimétricos dependendo da implementação específica.

[00027] Em uma aplicação adicional, uma porção de frequência ou uma correção de frequência não é apenas submetida à filtragem cruzada, mas o gerador de porção preferivelmente realiza, antes de realizar a filtragem cruzada, uma adaptação de correção que compreende um ajuste dos limites de frequência em mínimos espectrais locais e uma remoção ou atenuação das partes tonais que permanecem nas faixas de transição ao redor das frequências de transição.

[00028] Nessa aplicação, uma análise do sinal do lado do decodificador utilizando um analisador é realizada para analisar o sinal principal decodificado antes ou após realizar uma operação de regeneração de frequência para

fornecer um resultado de análise. Então, esse resultado de análise é utilizado por um regenerador de frequência para regenerar partes espectrais não incluídas no sinal principal decodificado.

[00029] Assim, em contraste a um ajuste do decodificador fixo, onde a correção ou partição de frequência é realizado em uma forma fixa, ou seja, onde uma certa faixa de origem é tirada do sinal principal e certos limites de frequência fixos são aplicados a qualquer conjunto de frequência entre a faixa de origem e a faixa de reconstrução ou o limite de frequência entre as duas correções ou porções de frequência adjacentes dentro da faixa de reconstrução, uma correção ou partição dependente de sinal é realizada, na qual, por exemplo, o sinal principal pode ser analisado para localizar os mínimos locais no sinal principal e, então, a faixa principal é selecionada de modo que os limites de frequência da faixa principal coincidem com os mínimos locais no espectro do sinal principal.

[00030] De modo alternativo ou adicional, uma análise de sinal pode ser realizada em um sinal regenerado preliminar ou sinal corrigido ou particionado por frequência preliminar, em que, após o procedimento da regeneração de frequência preliminar, o limite entre a faixa principal e a faixa de reconstrução é analisado a fim de detectar quaisquer partes do sinal que criam perturbação como partes tonais sendo problemáticas em que elas estão bem próximos entre si para gerar uma perturbação de pulsação quando reconstruída. De modo alternativo ou adicional, os limites também podem ser

examinados de modo que um meio-recorte de uma parte tonal é detectado e esse recorte de uma parte tonal também criaria uma perturbação quando reconstruída como é. A fim de evitar esses procedimentos, o limite de frequência da faixa de reconstrução e/ou da faixa de origem e/ou entre as duas porções de frequência ou correções individuais na faixa de reconstrução pode ser modificado por um manipulador de sinal a fim de realizar novamente uma reconstrução com os limites recém definidos.

[00031] De modo adicional ou alternativo, a regeneração de frequência é uma regeneração com base no resultado de análise em que os limites de frequência são deixados como são e uma eliminação ou pelo menos atenuação de partes tonais problemáticas próximas aos limites de frequência entre a faixa de origem e a faixa de reconstrução ou entre as duas porções ou correções de frequência individuais dentro da faixa de reconstrução é feita. Essas partes tonais podem ser tons próximos que resultariam em uma perturbação de pulsação ou poderiam ser partes tonais cortadas.

[00032] Especificamente, quando uma transformada de conservação sem energia é utilizada como uma MDCT, um único tom não mapeia diretamente a uma única linha espectral. Ao invés disso, um único tom mapeará em um grupo de linhas espectrais com certas amplitudes dependendo da fase do tom. Quando uma operação de correção corta essa parte tonal, então isso resultará em uma perturbação após a reconstrução embora uma reconstrução perfeita seja aplicada como em um

reconstrutor de MDCT. Isso é devido ao fato de que o reconstrutor de MDCT exigiria o padrão tonal completo para um tom a gim de finalmente reconstruir completamente esse tom. Devido ao fato de que um recorte ocorreu antes, isso não é possível mais e, assim, uma perturbação de modulação de frequência variável no tempo será criada. Com base na análise de acordo com a presente invenção, o regenerador de frequência evitará essa situação atenuando a parte tonal completa criando uma perturbação ou conforme previamente discutido, mudando as frequências de limite correspondentes ou aplicando ambas as medições ou pela reconstrução uniforme da parte recortada com base em um certo pré-conhecimento em tais padrões tonais.

[00033] A abordagem inventiva destina-se principalmente a ser aplicada dentro de um BWE com base em uma transformada, como a MDCT. Não obstante, os ensinamentos da invenção são geralmente aplicáveis, p.ex., analogamente dentro de um sistema com base no Banco de Filtro em Espelho de Quadratura (QMF), especialmente se o sistema for criticamente testado, por exemplo uma representação de QMF com valor real.

[00034] As aplicações preferidas serão subsequentemente discutidas com relação aos desenhos anexos, nos quais:

a Figura 1a ilustra um aparelho para codificação de um sinal de áudio;

a Figura 1b ilustra um decodificador para decodificação de um sinal de áudio codificado correspondente

ao codificador da Figura 1a;

a Figura 2a ilustra uma implementação preferida do decodificador;

a Figura 2b ilustra uma implementação preferida do codificador;

a Figura 3a ilustra uma representação esquemática de um espectro, conforme gerado pelo decodificador de domínio espectral da Figura 1b;

a Figura 3b ilustra uma tabela, indicando a relação entre os fatores de escala para bandas do fator de escala e energias para bandas de reconstrução e informação de preenchimento de ruído para uma banda de preenchimento de ruído;

a Figura 4a ilustra a funcionalidade do codificador do domínio espectral para aplicar a seleção de partes espectrais no primeiro e segundo conjuntos de partes espectrais;

a Figura 4b ilustra uma implementação da funcionalidade da Figura 4a;

a Figura 5a ilustra uma funcionalidade de um codificador MDCT;

a Figura 5b ilustra uma funcionalidade do decodificador com uma tecnologia MDCT;

a Figura 5c ilustra uma implementação do regenerador de frequência;

a Figura 6a é um aparelho para decodificação de um sinal de áudio codificado, de acordo com uma implementação;

a Figura 6b é uma aplicação adicional de um aparelho para decodificação de um sinal de áudio codificado;

a Figura 7a ilustra uma implementação preferida do regenerador de frequência da Figura 6a ou 6b;

a Figura 7b ilustra uma implementação adicional de uma cooperação entre o analisador e o regenerador de frequência;

a Figura 8 ilustra uma implementação adicional do regenerador de frequência;

a Figura 8b ilustra uma aplicação adicional da invenção;

a Figura 9a ilustra um decodificador com a tecnologia da regeneração de frequência utilizando valores de energia para a faixa de frequência da regeneração;

a Figura 9b ilustra uma implementação mais detalhada do regenerador de frequência da Figura 9a;

a Figura 9c ilustra um esquema que ilustra a funcionalidade da Figura 9b;

a Figura 9d ilustra outra implementação do decodificador da Figura 9a;

a Figura 10a ilustra um diagrama em blocos de um codificador correspondente com o decodificador da Figura 9a;

a Figura 10b ilustra um diagrama em blocos para ilustrar outra funcionalidade da calculadora de parâmetro da Figura 10a;

a Figura 10c ilustra um diagrama em blocos que ilustra outra funcionalidade da calculadora paramétrica

da Figura 10a;

a Figura 10d ilustra um diagrama em blocos que ilustra outra funcionalidade da calculadora paramétrica da Figura 10a;

a Figura 11a ilustra um espectro de um toque do filtro ao redor de um transiente;

a Figura 11b ilustra um espectrograma de um transiente após aplicar a extensão da largura de banda;

a Figura 11c ilustra um espectrograma de um transiente após aplicar a extensão da largura de banda com redução do toque do filtro;

a Figura 12a ilustra um diagrama em blocos de um aparelho para decodificação de um sinal de áudio codificado;

a Figura 12b ilustra os espectros de magnitude (estilizados) de um sinal tonal, uma cópia sem adaptação de correção/porção, uma cópia com limites de frequência alterados e uma eliminação adicional de partes tonais que criam perturbação;

a Figura 12c ilustra uma função exemplar de *cross-fade* (transição suave);

a Figura 13a ilustra um codificador da técnica anterior com extensão da largura de banda; e

a Figura 13b ilustra um decodificador da técnica anterior com a extensão da largura de banda.

a Figura 14a ilustra um aparelho adicional para decodificação de um sinal de áudio codificado, utilizando um filtro cruzado;

a Figura 14b apresenta uma ilustração mais detalhada de um filtro cruzado exemplar.

[00035] A figura 6a ilustra um aparelho para decodificação de um sinal de áudio codificado que compreende um sinal principal codificado e dados paramétricos. O aparelho compreende um decodificador principal (600) para decodificar o sinal principal codificado para obter um sinal principal decodificado, um analisador (602) para analisar o sinal principal decodificado antes ou após realizar uma operação de regeneração de frequência. O analisador (602) é configurado para fornecer um resultado de análise (603). O regenerador de frequência (604) é configurado para regenerar partes espectrais não incluídas no sinal principal decodificado utilizando uma parte espectral do sinal principal decodificado, dados de envelope (605) para as partes espectrais ausentes e o resultado de análise (603). Assim, em contraste às implementações prévias, a regeneração de frequência não é realizada independente do sinal do lado do decodificador, mas é realizada dependente do sinal. Isso tem a vantagem que, quando nenhum problema existe, a regeneração de frequência é realizada como é, mas quando as partes do sinal problemático existem, então isso é detectado pelo resultado de análise (603) e o regenerador de frequência (604) então realiza uma forma adaptada de regeneração de frequência que pode, por exemplo, ser a mudança de um limite de frequência inicial entre a região principal e a banda de reconstrução ou a mudança de um limite de frequência entre as duas porções/correções individuais dentro da banda de

reconstrução. Ao contrário da implementação das bandas de guarda, isso tem a vantagem que os procedimentos específicos são apenas realizados quando necessários e não, como na implementação da banda de guarda, todo o tempo sem qualquer dependência de sinal.

[00036] Preferivelmente, o decodificador principal (600) é implementado como uma entropia (p.ex., decodificador de Huffman ou aritmético) decodificando e dequantizando o estágio (612) conforme ilustrado na Figura 6b. O decodificador principal (600) então emite um espectro do sinal principal e o espectro é analisado pelo analisador espectral (614) que é, bem similar ao analisador (602) na Figura 6a implementado como um analisador espectral ao invés de qualquer analisador arbitrário que poderia, conforme ilustrado na Figura 6a, ainda analisar um sinal de domínio de tempo. Na aplicação da Figura 6b, o analisador espectral é configurado para analisar o sinal espectral de modo que mínimos locais na banda de origem e/ou em uma banda alvo, ou seja, nas correções de frequência ou porções de frequência sejam determinados. Então, o regenerador de frequência (604) realiza, conforme ilustrado em (616), uma regeneração de frequência onde os limites de correção são colocados nos mínimos na banda de origem e/ou na banda alvo.

[00037] Subsequentemente, a Figura 7a é discutida, a fim de descrever uma implementação preferida do regenerador de frequência (604) da Figura 6a. Um regenerador do sinal preliminar (702) recebe, como uma entrada, dados de origem da banda de origem e, adicionalmente, informações preliminares

de correção como frequências limite preliminares. Então, um sinal regenerado preliminar (703) é gerado, que é detectado pelo detector (704) para detectar os componentes tonais dentro do sinal reconstruído preliminar (703). De modo alternativo ou adicional, os dados de origem (705) também podem ser analisados pelo detector correspondente ao analisador (602) da Figura 6a. Então, a etapa de regeneração de sinal preliminar não seria necessária. Quando há um mapeamento bem definido dos dados de origem para os dados de reconstrução, então os mínimos ou partes tonais podem ser detectados mesmo considerando apenas os dados de origem, se há partes tonais próximas ao limite superior da faixa principal ou em um limite de frequência entre duas porções de frequência individualmente geradas conforme será discutido posteriormente com relação à Figura 12b.

[00038] No caso de componentes tonais problemáticos, foram descobertos limites de frequência próximos, um ajustador de frequência de transição (706) realiza um ajuste de uma frequência de transição como uma frequência de transição ou frequência cruzada ou frequência de preenchimento de lacuna inicial entre a banda principal e a banda de reconstrução ou entre partes de frequência individuais geradas por um e os mesmos dados de origem na banda de reconstrução. O sinal de saída do bloco (706) é direcionado a um removedor (708) de componentes tonais nos limites. O removedor é configurado para remover os componentes tonais remanescentes que são ainda subsequentes ao ajuste da frequência de transição pelo bloco (706). O

resultado do removedor (708) é, então, encaminhado a um filtro cruzado (710) a fim de direcionar o problema de toque do filtro e o resultado do filtro cruzado (710) é, então, inserido em um bloco de formação de envelope espectral (712) que realiza uma formação de envelope espectral na banda de reconstrução.

[00039] Conforme discutido no contexto da Figura 7a, a detecção de componentes tonais no bloco (704) pode ser tanto realizada em dados de origem (705) ou um sinal reconstruído preliminar (703). Essa aplicação é ilustrada na Figura 7b, onde um sinal regenerado preliminar é criado conforme mostrado no bloco (718). O sinal correspondente ao sinal (703) da Figura 7a é, então, encaminhado a um detector (720) que detecta os componentes que criam perturbação. Embora o detector (720) possa ser configurado para ser um detector para detectar componentes tonais nos limites de frequência conforme ilustrado em (704) na Figura 7a, o detector também pode ser implementado para detectar outros componentes que criam perturbação. Tais componentes espectrais podem ser ainda outros componentes do que componentes tonais e uma detecção se uma perturbação foi criada pode ser realizada tentando diferentes regenerações e comparando os diferentes resultados de regeneração a fim de observar qual forneceu componentes que criam perturbação.

[00040] O detector (720) agora controla um manipulador (722) para manipular o sinal, ou seja, o sinal regenerado preliminar. Essa manipulação pode ser feita processando realmente o sinal regenerado preliminar pela

linha (723)

ou pela nova realização de uma regeneração, mas agora com, por exemplo, as frequências de transição alteradas conforme ilustrado pela linha (724).

[00041] Uma implementação do procedimento de manipulação é que a frequência de transição é ajustada, conforme ilustrado em (706) na Figura 7a. Uma implementação adicional é ilustrada na Figura 8a, que pode ser realizada ao invés do bloco (706) ou junto com o bloco (706) da Figura 7a. Um detector (802) é fornecido para detectar as frequências inicial e final de uma parte tonal problemática. Então, um interpolador (804) é configurado para interpolação e, preferivelmente interpolação complexa entre o início e o final da parte tonal dentro da faixa espectral. Então, conforme ilustrado na Figura 8a pelo bloco (806), a parte tonal é substituída pelo resultado da interpolação.

[00042] Uma implementação alternativa é ilustrada na Figura 8a pelos blocos (808, 810). Ao invés de realizar uma interpolação, uma geração aleatória de linhas espectrais (808) é realizada entre o início e o fim da parte tonal. Então, um ajuste de energia das linhas espectrais aleatoriamente geradas é realizado conforme ilustrado em (810) e a energia das linhas espectrais aleatoriamente geradas é definida de modo que a energia seja similar às partes espectrais não tonais adjacentes. Então, a parte tonal é substituída pelas linhas espectrais aleatoriamente geradas ajustadas por envelope. As linhas espectrais podem ser aleatórias ou pseudoaleatoriamente geradas, a fim de fornecer

um sinal de substituição que é, tanto quanto possível, livre de perturbação.

[00043] Uma implementação adicional é ilustrada na Figura 8b. Um gerador de porção de frequência localizado dentro do regenerador de frequência (604) da Figura 6a é ilustrado no bloco (820). O gerador de porção de frequência utiliza os limites de frequência predeterminados. Então, o analisador analisa o sinal gerado pelo gerador de porção de frequência e o gerador de porção de frequência (820) é preferivelmente configurado para realizar múltiplas operações de partição para gerar múltiplas porções de frequência. Então, o manipulador (824) na Figura 8b manipula o resultado do gerador de porção de frequência de acordo com o resultado de análise emitido pelo analisador (822). A manipulação pode ser a mudança de limites de frequência ou a atenuação de partes individuais. Então, um ajustador de envelope espectral (826) realiza um ajuste de envelope espectral utilizando as informações paramétricas (605) conforme já discutido no contexto da Figura 6a.

[00044] Então, o sinal espectralmente ajustado emitido pelo bloco (826) é inserido em um conversor de frequência-tempo que, adicionalmente, recebe as primeiras partes espectrais, ou seja, uma representação espectral do sinal de saída do decodificador principal (600). A saída do conversor de frequência-tempo (828) pode, então, ser utilizada para armazenamento ou para transmissão em um alto-falante para interpretação de áudio.

[00045] A presente invenção pode ser aplicada tanto

nos procedimentos conhecidos de regeneração de frequência, como ilustrado nas Figuras 13a, 13b quanto podem ser preferivelmente aplicados dentro do contexto de preenchimento de lacuna inteligente, que é subsequentemente descrito com relação às Figuras de 1a a 5b e de 9a a 10d.

[00046] A Figura 1a ilustra um aparelho para codificar um sinal de áudio (99). O sinal de áudio (99) é inserido em um conversor do espectro de tempo (100) para converter um sinal de áudio tendo uma taxa de amostragem em uma representação espectral (101) emitida pelo conversor do espectro de tempo. O espectro (101) é inserido em um analisador espectral (102) para analisar a representação espectral (101). O analisador espectral (101) é configurado para determinar um primeiro conjunto de partes espectrais primárias (103) a ser codificado com uma primeira resolução espectral e um diferente segundo conjunto de partes espectrais secundárias (105) a ser codificado com uma segunda resolução espectral. A segunda resolução espectral é menor do que a primeira resolução espectral. O segundo conjunto de partes espectrais secundárias (105) é inserido em uma calculadora de parâmetro ou codificador paramétrico (104) para calcular a informação do envelope espectral tendo a segunda resolução espectral. Além disso, um codificador do domínio espectral de áudio (106) é fornecido para gerar uma primeira representação codificada (107) do primeiro conjunto de partes espectrais primárias tendo a primeira resolução espectral. Além disso, a calculadora de parâmetro/codificador paramétrico (104) é configurada para gerar uma segunda

representação codificada (109) do segundo conjunto de partes espectrais secundárias. A primeira representação codificada (107) e a segunda representação codificada (109) são inseridas em um multiplexador de fluxo contínuo de bits ou formador do fluxo contínuo de bits (108) e bloco (108) finalmente emite o sinal de áudio codificado para transmissão ou armazenamento em um dispositivo de armazenamento.

[00047] Tipicamente, uma primeira parte espectral como (306) da Figura 3a será envolvida por duas partes espectrais secundárias como (307a, 307b). Este não é o caso em HE AAC, onde a faixa de frequência do codificador principal é limitada por banda.

[00048] A Figura 1b ilustra um decodificador correspondente com o codificador da Figura 1a. A primeira representação codificada (107) é inserida em um decodificador de áudio de domínio espectral (112) para gerar uma primeira representação decodificada de um primeiro conjunto de partes espectrais primárias, a representação decodificada tendo uma primeira resolução espectral. Além disso, a segunda representação codificada (109) é inserida em um decodificador paramétrico (114) para gerar uma segunda representação decodificada de um segundo conjunto de partes espectrais secundárias tendo uma segunda resolução espectral sendo menor do que a primeira resolução espectral.

[00049] O decodificador compreende, ainda, um regenerador de frequência (116) para gerar novamente uma segunda parte espectral reconstruída, tendo a primeira resolução espectral, utilizando uma primeira parte espectral.

O regenerador de frequência (116) realiza uma operação de adição da porção, ou seja, utiliza uma porção ou parte do primeiro conjunto de partes espectrais primárias e copia este primeiro conjunto de partes espectrais primárias na faixa de reconstrução ou banda de reconstrução tendo a segunda parte espectral e tipicamente realiza a formação do envelope espectral ou outra operação é indicada pela segunda representação decodificada emitida pelo decodificador paramétrico (114), ou seja, utilizando a informação no segundo conjunto de partes espectrais secundárias. O primeiro conjunto decodificado de partes espectrais primárias e o segundo conjunto reconstruído de partes espectrais conforme indicado na saída do regenerador de frequência (116) na linha (117) é inserido em um conversor de espectro-tempo (118) configurado para converter a primeira representação decodificada e a segunda parte espectral reconstruída em uma representação de tempo (119), a representação de tempo tendo uma certa alta taxa de amostragem.

[00050] A Figura 2b ilustra uma implementação do codificador da Figura 1a. Um sinal de entrada de áudio (99) é inserido em um banco de filtro de análise (220) correspondente ao conversor do espectro de tempo (100) da Figura 1a. Então, uma operação do modelador de ruído temporal é realizada no bloco TNS (222). Assim, a entrada no analisador espectral (102) da Figura 1a correspondente a uma máscara tonal do bloco (226) da Figura 2b pode tanto ser valores espectrais completos, quando a operação do modelador de ruído temporal/modelador da porção temporal não é aplicada

ou pode ser valores residuais espectrais, quando a operação de TNS conforme ilustrado na Figura 2b, o bloco (222) é aplicado. Para sinais de dois canais ou sinais multicanais, uma codificação do canal de junção (228) pode adicionalmente ser realizada, de modo que o codificador do domínio espectral (106) da Figura 1a possa compreender o bloco da codificação do canal de junção (228). Além disso, um codificador de entropia (232) para realizar uma compressão de dados sem perda é fornecido sendo também uma parte do codificador do domínio espectral (106) da Figura 1a.

[00051] A máscara do analisador espectral/tonal (226) separa a saída do bloco TNS (222) na banda espectral e os componentes tonais correspondentes ao primeiro conjunto de partes espectrais primárias (103) e os componentes residuais correspondentes ao segundo conjunto de partes espectrais secundárias (105) da Figura 1a. O bloco (224) indicado como codificação de extração do parâmetro IGF corresponde ao codificador paramétrico (104) da Figura 1a e o multiplexador do fluxo contínuo de bit (230) corresponde ao multiplexador do fluxo contínuo de bit (108) da Figura 1a.

[00052] Preferivelmente, o banco de filtro de análise (222) é implementado como uma MDCT (banco de filtro da transformada de cosseno discreta modificada) e a MDCT é utilizada para transformar o sinal (99) em um domínio de tempo-frequência com a transformada de cosseno discreta modificada que age como a ferramenta de análise de frequência.

[00053] O analisador espectral (226) preferivelmente

aplica uma máscara de tonalidade. Este estágio de estimativa da máscara de tonalidade é utilizado para separar os componentes tonais dos componentes do tipo ruído no sinal. Isso permite que o codificador principal (228) codifique todos os componentes tonais com um módulo psicoacústico. O estágio de estimativa da máscara de tonalidade pode ser implementado em várias formas diferentes e é preferivelmente implementado semelhante em sua funcionalidade ao estágio de estimativa de rastreio sinusoidal utilizado na modelagem por seno e ruído para codificação de áudio/fala [8, 9] ou um codificador de áudio com base no modelo HILN descrito em [10]. Preferivelmente, uma implementação é utilizada sendo fácil de implementar sem a necessidade de manter as trajetórias iniciais-finais, mas qualquer outra tonalidade ou detector de ruído também pode ser utilizada.

[00054] O módulo de IGF calcula a similaridade que existe entre uma região de origem e uma região alvo. A região alvo será representada pelo espectro da região de origem. A medição da similaridade entre a região de origem e a região alvo é feita utilizando uma abordagem de correlação. A região alvo é dividida em $nTar$ porções de frequência não sobrepostas. Para cada porção na região alvo, $nSrc$ as porções de origem são criadas de uma frequência inicial fixa. Estas porções de origem sobrepostas por um fator entre 0 e 1, onde 0 significa 0% de sobreposição e 1 significa 100% de sobreposição. Cada uma destas porções de origem está correlacionada com a porção alvo em vários atrasos para encontrar a porção de origem que melhor correspondente com a

porção alvo. O melhor número de porção correspondente é armazenado em $tileNum[idx_tar]$, o atraso no qual melhor correlaciona com o alvo é armazenado em $xcorr_lag[idx_tar][idx_src]$ e o sinal da correlação é armazenado em $xcorr_sign[idx_tar][idx_src]$. No caso em que a correlação é altamente negativa, a porção de origem precisa ser multiplicada por −1 antes do processo de adição da porção no decodificador. O módulo de IGF ainda cuida dos componentes tonais não sobrescritos no espectro visto que os componentes tonais são preservados utilizando a máscara de tonalidade. Um parâmetro da energia de banda ampla é utilizado para armazenar a energia da região alvo que nos permite reconstruir o espectro precisamente.

[00055] Este método tem certas vantagens sobre a SBR clássica [1] em que a grade de harmonia de um sinal de vários tons é preservada pelo codificador principal enquanto apenas as lacunas entre os sinusoides é preenchida com o melhor "ruído modulado" correspondente da região de origem. Outra vantagem deste sistema comparado ao ASR (Substituição Espectral Precisa) [2–4] é a ausência de um estágio da síntese do sinal que cria as partes importantes do sinal no decodificador. Ainda, esta tarefa é assumida pelo codificador principal, permitindo a preservação de importantes componentes do espectro. Outra vantagem do sistema proposto é a escalabilidade contínua que as características oferecem. Apenas utilizando $tileNum[idx_tar]$ e $xcorr_lag = 0$, para cada *porção* é chamado de granularidade bruta correspondente e pode ser utilizado para baixas taxas de bit ao utilizar a variável

xcorr_lag para cada porção nos permite equiparar os espectros de origem e alvo melhor.

[00056] Além disso, uma técnica de estabilização de escolha da porção é proposta removendo as perturbações de domínio de frequência como vibração e ruído musical.

[00057] No caso de pares de canal estéreo, um processamento estéreo conjunto adicional é aplicado. Isso é necessário porque, para uma certa faixa de destino, o sinal pode ser uma de origem de som panorâmico altamente correlacionada. No caso de as regiões de origem escolhidas para esta região particular não estarem bem correlacionadas, embora as energias estejam equipadas para as regiões de destino, a imagem espacial pode sofrer devido às regiões de origem não correlacionadas. O codificador analisa cada banda de energia da região de destino, tipicamente realizando uma correlação dos valores espectrais, e se um determinado limite for excedido, define um identificador conjunto para esta banda de energia. No decodificador, as bandas de energia do canal direita e esquerda são tratadas individualmente se este identificador estéreo conjunto não estiver definido. No caso do identificador estéreo conjunto estar definido, ambas as energias e a correção são realizadas no domínio estéreo conjunto. A informação estéreo conjunta para as regiões de IGF é sinalizada semelhante à informação estéreo conjunta para a codificação principal, incluindo um identificador que indica no caso de previsão se a direção da previsão é de *downmix* a residual ou vice versa.

[00058] As energias podem ser calculadas a partir das

energias transmitidas no domínio L/R.

$$midNrg[k] = leftNrg[k] + rightNrg[k];$$

$$sideNrg[k] = leftNrg[k] - rightNrg[k];$$

com k sendo o índice de frequência no domínio de transformada.

[00059] Outra solução é calcular e transmitir as energias diretamente no domínio estéreo conjunto para bandas onde o estéreo conjunto é ativo, assim, nenhuma transformação de energia adicional é necessária no lado do decodificador.

[00060] As porções de origem são sempre criadas de acordo com a Matriz-Meio/Lateral:

$$midTile\ [k] = 0.5 \cdot (leftTile\ [k] + rightTile\ [k])$$

$$sideTile\ [k] = 0.5 \cdot (leftTile\ [k] - rightTile\ [k])$$

[00061] Ajuste de energia:

$$midTile[k] = midTile[k] * midNrg[k];$$

$$sideTile[k] = sideTile[k] * sideNrg[k];$$

[00062] Estéreo conjunto -> transformação LR:

[00063] Se nenhum parâmetro de previsão adicional estiver codificado:

$$leftTile\ [k] = midTile\ [k] + sideTile\ [k]$$

$$rightTile\ [k] = midTile\ [k] - sideTile\ [k]$$

[00064] Se um parâmetro de previsão adicional estiver codificado e se a direção sinalizada for do meio à lateral:

$$sideTile[k] = sideTile[k] - predictionCoeff \cdot midTile[k]$$
$$leftTile[k] = midTile[k] + sideTile[k]$$
$$rightTile[k] = midTile[k] - sideTile[k]$$

[00065] Se a direção sinalizada for da lateral ao meio:

$$midTile1[k] = midTile[k] - predictionCoeff \cdot sideTile[k]$$
$$leftTile[k] = midTile1[k] - sideTile[k]$$
$$rightTile[k] = midTile1[k] + sideTile[k]$$

[00066] Este processamento garante que a partir das porções utilizadas para regenerar regiões de destino altamente correlacionadas e regiões de destino panorâmicas, os canais esquerdo e direito resultantes ainda representam uma de origem sonora correlacionada e panorâmica mesmo se as regiões de origem não estiverem correlacionadas, preservando a imagem estéreo para tais regiões.

[00067] Em outras palavras, no fluxo contínuo de bits, os conjuntos de identificador estéreo são transmitidos indicando se L/R ou M/S como um exemplo para a codificação do conjunto estéreo geral deve ser utilizado. No decodificador, primeiro, o sinal principal é decodificado conforme indicado pelos conjuntos de identificador estéreo para as bandas espectrais. Segundo, o sinal principal é armazenado tanto na representação L/R quanto na M/S. Para o preenchimento de porção de IGF, a representação da porção de origem é escolhida para ajustar a representação da porção alvo conforme indicado pela informação estéreo conjunta para as bandas de IGF.

[00068] O Modelador de Ruído Temporal (TNS | *Temporal Noise Shaping*) é uma técnica padrão e parte de AAC [11 – 13]. TNS pode ser considerado como uma extensão do esquema básico de um codificador perceptual, inserindo uma etapa de processamento opcional entre o banco de filtro e o estágio de quantização. A principal tarefa do módulo de TNS é ocular o ruído de quantização produzido na região de mascaramento

temporal de sinais do tipo transitórios e assim leva a um esquema de codificação mais eficiente. Primeiro, TNS calcula um conjunto de coeficientes de previsão utilizando “previsão de avanço” no domínio de transformada, por exemplo MDCT. Estes coeficientes são então utilizados para nivelamento do envelope temporal do sinal. Como a quantização afeta o espectro filtrado por TNS, ainda o ruído de quantização é temporariamente plano. Aplicando a filtragem inversa de TNS no lado do decodificador, o ruído de quantização é modulado de acordo com o envelope temporal do filtro de TNS e assim o ruído de quantização fica mascarado pelo transiente.

[00069] O IGF se baseia em uma representação de MDCT. Para codificação eficiente, preferivelmente blocos longos de aproximadamente 20 ms devem ser utilizados. Se o sinal dentro deste bloco longo contém transientes, pré- e pós-ecos audíveis ocorrem nas bandas espectrais de IGF devido à adição de porção. A Figura 7c mostra um efeito típico de pré-eco antes do começo transiente devido à IGF. No lado esquerdo, o espectrograma do sinal original é mostrado e no lado direito o espectrograma da largura de banda sinal estendido sem filtragem de TNS é mostrado.

[00070] Este efeito de pré-eco é reduzido utilizando TNS no contexto de IGF. Aqui, TNS é utilizado como uma ferramenta do modelador de porção temporal (TTS | *temporal tile shaping*) como a regeneração espectral no decodificador é realizado no sinal residual de TNS. Os coeficientes de previsão de TTS necessários são calculados e aplicados utilizando o espectro completo no lado do codificador como de

costume. As frequências de início e parada do TNS/TTS não são afetadas pela frequência inicial de IGF $f_{IGFstart}$ da ferramenta de IGF. Em comparação ao TNS legado, a frequência de parada de TTS é aumentada para a frequência de parada da ferramenta de IGF, que é mais alta do que $f_{IGFstart}$. No lado do decodificador os coeficientes de TNS/TTS são aplicados no espectro completo novamente, ou seja, o espectro principal mais o espectro regenerado mais os componentes tonais do mapa de tonalidade (veja a Figura 7e). A aplicação de TTS é necessária para formar o envelope temporal do espectro regenerado para equiparar o envelope do sinal original novamente. Assim, os pré-ecos mostrados são reduzidos. Além disso, ainda modula o ruído de quantização no sinal abaixo de $f_{IGFstart}$ como de costume com TNS.

[00071] Nos decodificadores de legado, a correção espectral em um sinal de áudio corrompe a correlação espectral nos limites de correção e assim impede o envelope temporal do sinal de áudio introduzindo a dispersão. Assim, outro benefício para realizar o preenchimento de porção de IGF no sinal residual é que, após a aplicação do filtro de modulação, as porções limite são correlacionadas continuamente, resultando em uma reprodução temporal mais fielmente do sinal.

[00072] Em um codificador inventivo, o espectro tendo passado pela filtragem de TNS/TTS, o processamento da máscara de tonalidade e a estimativa do parâmetro de IGF é desprovido de qualquer sinal acima da frequência inicial de IGF exceto para componentes tonais. Este espectro insuficiente é agora

codificado pelo codificador principal utilizando os princípios da codificação aritmética e codificação preditiva. Estes componentes codificados com os bits de sinalização formam os fluxos contínuos de bits do áudio.

[00073] A Figura 2a ilustra a implementação do decodificador correspondente. O fluxo contínuo de bits na Figura 2a correspondente ao sinal de áudio codificado é inserido no demultiplexador/decodificador que seria conectado, com relação à Figura 1b, aos blocos (112 e 114). O demultiplexador do fluxo contínuo de bits separa o sinal de áudio de entrada na primeira representação codificada (107) da Figura 1b e a segunda representação codificada (109) da Figura 1b. A primeira representação codificada tendo o primeiro conjunto de partes espectrais primárias é inserida no bloco de codificação de canal conjunto (204) correspondente ao decodificador de domínio espectral (112) da Figura 1b. A segunda representação codificada é inserida no decodificador paramétrico (114) não ilustrado na Figura 2a e então inserida no bloco IGF (202) correspondente ao regenerador de frequência (116) da Figura 1b. O primeiro conjunto de partes espectrais primárias necessário para a regeneração de frequência é inserido no bloco IGF (202) através da linha (203). Além disso, subsequente à decodificação do canal conjunto (204) a decodificação principal específica é aplicada no bloco da máscara tonal (206) de modo que a saída da máscara tonal (206) corresponda à saída do decodificador de domínio espectral (112). Então, uma combinação pelo combinador (208) é realizada, ou seja,

uma criação de estrutura onde a saída do combinador (208) agora tem o espectro de intervalo completo, mas ainda no domínio filtrado TNS/TTS. Então, no bloco (210), uma operação inversa de TNS/TTS é realizada utilizando a informação do filtro TNS/TTS fornecida através da linha (109), ou seja, a informação adicional de TTS é preferivelmente incluída na primeira representação codificada gerada pelo codificador do domínio espectral (106) que pode, por exemplo, ser um codificador principal direto AAC ou USAC, ou pode ainda ser incluído na segunda representação codificada. Na saída do bloco (210), um espectro completo até a frequência máxima é fornecido que é a frequência de intervalo completo definida pela taxa de amostragem do sinal original de entrada. Então, uma conversão de espectro/tempo é realizada no banco de filtro de síntese (212) para finalmente obter o sinal de saída de áudio.

[00074] A Figura 3a ilustra uma representação esquemática do espectro. O espectro é subdividido em bandas do fator de escala SCB onde há sete bandas do fator de escala SCB1 a SCB7 no exemplo ilustrado da Figura 3a. As bandas do fator de escala podem ser bandas do fator de escala AAC que são definidas no padrão AAC e têm uma largura de banda crescente em frequências superiores conforme ilustrado na Figura 3a esquematicamente. É preferido realizar o Preenchimento de Lacuna Inteligente não do primeiro momento do espectro, ou seja, em baixas frequências, mas para iniciar a operação de IGF em uma frequência inicial de IGF ilustrada em (309). Assim, a banda de frequência principal se estende

da frequência mais baixa à frequência inicial de IGF. Acima da frequência inicial de IGF, a análise do espectro é aplicada para separar os componentes espectrais de alta resolução (304, 305, 306, 307) (o primeiro conjunto de partes espectrais primárias) dos componentes de alta resolução representados pelo segundo conjunto de partes espectrais secundárias. A Figura 3a ilustra um espectro que é exemplarmente inserido no codificador do domínio espectral (106) ou no codificador do canal conjunto (228), ou seja, o codificador principal opera no intervalo completo, mas codifica uma quantidade significativa de valores espectrais zero, ou seja, estes valores espectrais zero são quantizados a zero ou são definidos a zero antes de quantizar ou subsequente à quantização. De qualquer modo, o codificador principal opera no intervalo completo, ou seja, como se o espectro fosse conforme ilustrado, ou seja, o decodificador principal não tem necessariamente que estar ciente de qualquer Preenchimento de Lacuna Inteligente ou codificação do segundo conjunto de partes espectrais secundárias com uma resolução espectral inferior.

[00075] Preferivelmente, a alta resolução é definida por uma codificação em linhas das linhas espectrais como linhas de MDCT, enquanto a segunda resolução ou alta resolução é definida, por exemplo, pelo cálculo apenas de um único valor espectral por banda de fator de escala, onde uma banda de fator de escala abrange as várias linhas de frequência. Assim, a segunda alta resolução é, com relação a sua resolução espectral, muito menor do que a primeira ou

alta resolução definida pela codificação em linhas tipicamente aplicada pelo codificador principal como um codificador principal AAC ou USAC.

[00076] Com relação ao fator de escala ou cálculo de energia, a situação é ilustrada na Figura 3b. Devido ao fato de que o codificador é um codificador principal e devido ao fato de que pode haver, mas não necessariamente deve ser, componentes do primeiro conjunto de partes espectrais em cada banda, o codificador principal calcula um fator de escala para cada banda não apenas na faixa principal abaixo da frequência inicial de IGF (309), mas ainda acima da frequência inicial de IGF até a frequência máxima $f_{IGFstop}$ que é menor ou igual à metade da frequência de amostragem, ou seja, $f_{s/2}$. Assim, as partes tonais codificadas (302, 304, 305, 306, 307) da Figura 3a e, nesta aplicação juntas com os fatores de escala SCB1 a SCB7 correspondem aos dados espectrais de alta resolução. Os dados espectrais de alta resolução são calculados começando da frequência inicial de IGF e correspondem aos valores da informação de energia E_1, E_2, E_3, E_4, que são transmitidos juntos com os fatores de escala SF4 a SF7.

[00077] Particularmente, quando o codificador principal está sob uma condição de baixa taxa de bit, uma operação adicional de preenchimento de ruído na banda espectral, ou seja, menor na frequência do que a frequência inicial de IGF, ou seja, nas bandas do fator de escala SCB1 a SCB3 pode ser aplicada. Além disso, no preenchimento de ruído, há várias linhas espectrais adjacentes que foram

quantizadas a zero. No lado do decodificador, estes valores quantizados em valores espectrais a zero são sintetizados novamente e os valores sintetizados novamente são ajustados em sua magnitude utilizando uma energia de preenchimento de ruído como NF_2 ilustrada em (308) na Figura 3b. A energia de preenchimento de ruído, que pode ser dada em termos absolutos ou nos termos relativos particularmente com relação ao fator de escala como em USAC corresponde à energia do conjunto de valores espectrais quantizados a zero. Estas linhas espectrais do preenchimento de ruído podem ainda ser consideradas como um terceiro conjunto de partes espectrais terciárias que são regeneradas pela síntese direta de preenchimento de ruído sem qualquer operação de IGF dependente da regeneração de frequência utilizando porções de frequência de outras frequências para reconstruir as porções de frequência utilizando valores espectrais de uma faixa de origem e da informação de energia E_1, E_2, E_3, E_4.

[00078] Preferivelmente, as bandas, nas quais a informação de energia é calculada, coincidem com as bandas do fator de escala. Em outras aplicações, um agrupamento do valor da informação de energia é aplicado de modo que, por exemplo, para bandas do fator de escala 4 e 5, apenas um único valor da informação de energia é transmitido, mas ainda nesta aplicação, os limites das bandas de reconstrução agrupadas coincidem com os limites das bandas do fator de escala. Se diferentes separações da banda são aplicadas, então certos recálculos ou cálculos de sincronização podem ser aplicados, e isso pode fazer sentido dependendo da

determinada implementação.

[00079] Preferivelmente, o codificador do domínio espectral (106) da Figura 1a é um codificador psicoacusticamente acionado, conforme ilustrado na Figura 4a. Tipicamente, como, por exemplo, ilustrado no padrão MPEG2/4 AAC ou MPEG1/2, padrão de Camada 3, para ser o sinal de áudio codificado após ter sido transformado na faixa espectral (401 na Figura 4a) é encaminhado a uma calculadora do fator de escala (400). A calculadora do fator de escala é controlada por um modelo psicoacústico adicionalmente que recebe sinal de áudio a ser quantizado ou recebe, como em MPEG1/2 Camada 3 ou padrão MPEG AAC, uma representação espectral complexa do sinal de áudio. O modelo psicoacústico calcula, para cada banda de fator de escala, um fator de escala que representa o limite psicoacústico. Adicionalmente, os fatores de escala são então, por cooperação dos circuitos de iteração interno e externo bem conhecidos ou por qualquer outro procedimento de codificação adequado ajustado de modo que certas condições da taxa de bit sejam realizadas. Então, os valores espectrais a ser quantizados, por um lado, e os fatores de escala calculados, por outro lado, são inseridos em um processador quantizador (404). Na operação do codificador de áudio direto, os valores espectrais a ser quantizados são ponderados pelos fatores de escala e, os valores espectrais ponderados são então inseridos em um quantizador fixo tipicamente tendo uma funcionalidade de compressão em faixas de amplitude superiores. Então, na saída do processador quantizador há índices de quantização que são encaminhados em

um codificador por entropia tipicamente tendo codificação específica e muito eficiente para um conjunto de índices de quantização zero para valores de frequência adjacentes ou, como ainda chamado na técnica, uma “série” de valores zero.

[00080] No codificador de áudio da Figura 1a, entretanto, o processador quantizador tipicamente recebe a informação nas partes espectrais secundárias do analisador espectral. Assim, o processador quantizador (404) certifica-se de que, na saída do processador quantizador (404), as partes espectrais secundárias, conforme identificado pelo analisador espectral (102), são zero ou têm uma representação reconhecida por um codificador ou um decodificador como uma representação zero que pode ser muito eficientemente codificada, especificamente quando existe “séries” de valores zero no espectro.

[00081] A Figura 4b ilustra uma implementação do processador quantizador. Os valores espectrais de MDCT podem ser inseridos em um ajuste ao bloco zero (410). Então, as partes espectrais secundárias já são definidas antes de uma ponderação pelos fatores de escala no bloco (412) são realizadas. Em uma implementação adicional, o bloco (410) não é provido, mas o ajuste na cooperação zero é realizado no bloco (418) subsequente ao bloco de ponderação (412). Em uma implementação ainda uniforme. O ajuste na operação zero pode ainda ser realizado em um ajuste no bloco zero (422) subsequente a uma quantização no bloco quantizador (420). Nesta implementação, os blocos (410 e 418) não estariam presentes. Geralmente, pelo menos um dos blocos (410, 418,

422) é fornecido dependendo da implementação específica.

[00082] Então, na saída do bloco (422), um espectro quantizado é obtido correspondente ao que é ilustrado na Figura 3a. Este espectro quantizado é, então, inserido em um codificador de entropia como (232) na Figura 2b que pode ser um codificador Huffman ou um codificador aritmético como, por exemplo, definido no padrão USAC.

[00083] O ajuste a blocos zero (410, 418, 422), que são fornecidos de modo alternativo entre si ou em paralelos, são controlados pelo analisador espectral (424). O analisador espectral preferivelmente compreende qualquer implementação de um detector de tonalidade bem conhecido ou compreende qualquer diferente tipo de detector operativo para separar um espectro em componentes a ser codificados com uma alta resolução e componentes a ser codificados com uma alta resolução. Outros desses algoritmos implementados no analisador espectral podem ser um detector de atividade de voz, um detector de ruído, um detector de fala ou qualquer outro detector que decide, dependendo da informação espectral ou metadados associados nas exigências de resolução para diferentes partes espectrais.

[00084] A Figura 5a ilustra uma implementação preferida do conversor do espectro de tempo (100) da Figura 1a como, por exemplo, implementado em AAC ou USAC. O conversor do espectro de tempo (100) compreende um janelador (502) controlado por um detector transiente (504). Quando o detector transiente (504) detecta um transiente, então um equipamento de transição de janelas longas a janelas curtas é

sinalizado ao janelador. O janelador (502), então, calcula, para os blocos de sobreposição, estruturas em janela, onde cada estrutura em janela tipicamente tem dois valores N como (2048) valores. Então, uma transformação dentro de um transformador de bloco (506) é realizada, e este transformador de bloco típica e adicionalmente fornece uma dizimação, de modo que uma dizimação/transformada combinada é realizada para obter uma estrutura espectral com N valores como valores espectrais de MDCT. Assim, para uma operação de janela longa, a estrutura na entrada do bloco (506) compreende dois valores N como (2048) valores e uma estrutura espectral então tem (1024) valores. Então, entretanto, uma interrupção é realizada em blocos curtos, quando oito blocos curtos são realizados onde cada bloco curto tem 1/8 valores do domínio de tempo em janela comparados a uma janela longa e cada bloco espectral tem 1/8 valores espectrais comparados a um bloco longo. Assim, quando esta dizimação é combinada com 50% da operação de sobreposição do janelador, o espectro é uma versão criticamente amostrada do sinal de áudio de domínio de tempo (99).

[00085] Subsequentemente, referência é feita à Figura 5b que ilustra uma implementação específica do regenerador de frequência (116) e do conversor de espectro-tempo (118) da Figura 1b, ou da operação combinada de blocos (208, 212) da Figura 2a. Na Figura 5b, uma banda de reconstrução específica é considerada como a banda de fator de escala 6 da Figura 3a. A primeira parte espectral nesta banda de reconstrução, ou seja, a primeira parte espectral (306) da Figura 3a é

inserida em o bloco criador/ajustador de estrutura (510). Além disso, uma segunda parte espectral reconstruída para a banda de fator de escala 6 é inserida no criador/ajustador de estrutura (510) também. Além disso, a informação de energia como E_3 da Figura 3b para uma banda de fator de escala 6 é ainda inserida no bloco (510). A segunda parte espectral reconstruída na banda de reconstrução já foi gerada pelo preenchimento da porção de frequência utilizando uma faixa de origem e a banda de reconstrução então corresponde à faixa alvo. Agora, um ajuste de energia da estrutura é realizado para então finamente obter a estrutura reconstruída completa tendo os N valores como, por exemplo, obtidos na saída do combinador (208) da Figura 2a. Então, no bloco (512), uma transformada/interpolação do bloco inverso é realizada para obter (248) valores de domínio de tempo para, por exemplo, (124) valores espectrais na entrada do bloco (512). Então, uma operação de janelamento de síntese é realizada no bloco (514) que é novamente controlada por uma indicação de janela longa/curta transmitida como informação adicional no sinal de áudio codificado. Então, no bloco (516), uma operação de sobreposição/adição com um período de tempo prévio é realizada. Preferivelmente, a MDCT aplica uma sobreposição de 50% de modo que, para cada novo período de tempo de 2N valores, N valores de domínio de tempo são finalmente emitidos. Uma sobreposição de 50% é dificilmente preferida devido ao fato de que ela fornece amostragem crítica e uma intersecção contínua de uma estrutura para a próxima estrutura devido à operação de sobreposição/adição no bloco

(516).

[00086] Conforme ilustrado em 301 na Figura 3a, uma operação de preenchimento de ruído pode adicionalmente ser aplicada não apenas abaixo da frequência inicial de IGF, mais também acima da frequência inicial de IGF como para a banda de reconstrução contemplada que coincide com a banda de fator de escala 6 da Figura 3a. Então, os valores espectrais de preenchimento de ruído podem ainda ser inseridos no criador/ajustador de estrutura (510) e o ajuste dos valores espectrais de preenchimento de ruído pode ainda ser aplicado dentro deste bloco ou os valores espectrais de preenchimento de ruído podem já ser ajustados utilizando a energia de preenchimento de ruído antes de ser inserida no criador/ajustador de estrutura (510).

[00087] Preferivelmente, uma operação de IGF, ou seja, uma operação de preenchimento da porção de frequência utilizando valores espectrais de outras partes pode ser aplicada no espectro completo. Assim, uma operação de adição da porção espectral não pode ser apenas aplicada na banda alta acima de uma frequência inicial de IGF, mas pode ainda ser aplicada na banda baixa. Além disso, o preenchimento de ruído sem preenchimento da porção de frequência pode ainda ser aplicada não apenas abaixo da frequência inicial de IGF, mas também acima da frequência inicial de IGF. Entretanto, foi observado que codificação de áudio de alta qualidade e alta eficiência pode ser obtida quando a operação de preenchimento de ruído for limitada à faixa de frequência abaixo da frequência inicial de IGF e quando a operação de

preenchimento da porção de frequência é restrita à faixa de frequência acima da frequência inicial de IGF conforme ilustrado na Figura 3a.

[00088] Preferivelmente, as porções alvo (TT | *target tiles*) (tendo frequências maiores do que a frequência inicial de IGF) são ligadas aos limites da banda de fator de escala do codificador de intervalo completo. As porções de origem (ST | *source tiles*), das quais a informação é tirada, ou seja, para frequências menores do que a frequência inicial de IGF, não são ligadas pelos limites da banda de fator de escala. O tamanho da ST deve corresponder ao tamanho da TT associada. Isso é ilustrado utilizando o seguinte exemplo. TT[0] tem um comprimento de 10 posições MDCT. Isso corresponde exatamente ao comprimento de dois SCBs subsequentes (como 4 + 6). Então, toda possível ST que deva ser correlacionada com TT[0], têm um comprimento de 10 posições também. Uma segunda porção alvo TT[1], sendo adjacente a TT[0], tem um comprimento de 15 posições l (com SCB tendo um comprimento de 7 + 8). Então, a ST para isso tem um comprimento de 15 posições ao invés de 10 posições como para TT[0].

[00089] Se surgir o caso onde pode-se encontrar uma TT para uma ST com o comprimento da porção alvo (quando, por exemplo, o comprimento de TT for maior do que a faixa de origem disponível), então uma correlação não é calculada e a faixa de origem é copiada várias vezes nesta TT (a cópia é feita uma após a outra, de modo que uma linha de frequência para a frequência mais baixa da segunda cópia siga

imediatamente – na frequência – a linha de frequência para a frequência mais alta da primeira cópia), até a porção alvo TT estar completamente adicionada.

[00090] Subsequentemente, referência é feita à Figura 5c que ilustra outra aplicação preferida do regenerador de frequência (116) da Figura 1b ou do bloco IGF (202) da Figura 2a. O bloco (522) é um gerador da porção de frequência que recebe, não apenas um ID da banda alvo, mas adicionalmente que recebe um ID da banda de origem. Exemplarmente, foi determinado no lado do codificador que a banda de fator de escala 3 da Figura 3a é muito bem adequada para reconstruir a banda de fator de escala 7. Assim, o ID da banda de origem seria 2 e o ID da banda alvo seria 7. Com base nesta informação, o gerador da porção de frequência (522) aplica uma cópia ou operação de adição da porção harmônica ou qualquer outra operação de adição da porção para gerar a segunda parte não processada dos componentes espectrais (523). A segunda parte não processada dos componentes espectrais tem uma resolução de frequência idêntica à resolução de frequência incluída no primeiro conjunto de partes espectrais primárias.

[00091] Então, a primeira parte espectral da banda de reconstrução como (307) da Figura 3a é inserida em um criador de estrutura (524) e a segunda parte não processada (523) é, ainda, inserida no criador de estrutura (524). Então, a estrutura reconstruída é ajustada pelo ajustador (526) utilizando um fator de ganho para a banda de reconstrução calculada pela calculadora do fator de ganho (528). De forma

importante, entretanto, a primeira parte espectral na estrutura não é influenciada pelo ajustador (526), mas apenas a segunda parte não processada para a reconstrução estrutura é influenciada pelo ajustador (526). Para esta finalidade, a calculadora do fator de ganho (528) analisa a banda de origem ou a segunda parte não processada (523) e adicionalmente analisa a primeira parte espectral na banda de reconstrução para finalmente encontrar o fator de ganho correto (527) de modo que a energia da estrutura ajustada emitida pelo ajustador (526) tenha a energia E_4 quando uma banda de fator de escala 7 é contemplada.

[00092] Neste contexto, é muito importante avaliar a precisão da reconstrução de alta frequência da presente invenção comparada à HE-AAC. Isso é explicado com relação à banda de fator de escala 7 na Figura 3a. Assume-se que um codificador da técnica anterior como ilustrado na Figura 13a detectaria a parte espectral (307) a ser codificada com uma alta resolução como uma "harmonia ausente". Então, a energia deste componente espectral seria transmitida junto com uma informação do envelope espectral para a banda de reconstrução como banda de fator de escala 7 ao decodificador. Então, o decodificador recriaria a harmonia ausente. Entretanto, o valor espectral, no qual a harmonia ausente (307) seria reconstruída pelo decodificador da técnica anterior da Figura 13b estaria no meio da banda 7 em uma frequência indicada pela frequência de reconstrução (390). Assim, a presente invenção evita um erro de frequência (391) que seria introduzido pelo decodificador da técnica anterior da Figura

13d.

[00093] Em uma implementação, o analisador espectral é, ainda, implementado para calcular as similaridades entre as partes espectrais primárias e as partes espectrais secundárias e para determinar, com base nas similaridades calculadas, para uma segunda parte espectral em uma faixa de reconstrução uma primeira parte espectral correspondente com a segunda parte espectral o mais longe possível. Então, nesta implementação da faixa de destino/faixa de origem variável, o codificador paramétrico adicionalmente introduzirá na segunda representação codificada uma informação correspondente indicando para cada faixa de destino uma faixa de origem correspondente. No lado do decodificador, esta informação então seria utilizada por um gerador da porção de frequência (522) da Figura 5c que ilustra uma geração de uma segunda parte não processada (523) com base em um ID da banda de origem e um ID da banda alvo.

[00094] Além disso, conforme ilustrado na Figura 3a, o analisador espectral é configurado para analisar a representação espectral até uma frequência de análise máxima sendo apenas uma pequena quantidade abaixo da metade da frequência de amostragem e preferivelmente sendo pelo menos um quarto da frequência de amostragem ou tipicamente mais alta.

[00095] Conforme ilustrado, o codificador opera sem redução da taxa de amostragem e o decodificador opera sem aumento da taxa de amostragem. Em outras palavras, o codificador do domínio espectral de áudio é configurado para

gerar uma representação espectral, tendo uma Frequência Nyquist definida pela taxa de amostragem do sinal de áudio de entrada original.

[00096] Além disso, conforme ilustrado na Figura 3a, o analisador espectral é configurado para analisar a representação espectral que começa com uma frequência inicial de preenchimento da lacuna e que termina com uma frequência máxima representada por uma frequência máxima incluída na representação espectral, em que uma parte espectral que se estende de uma frequência mínima até a frequência inicial de preenchimento da lacuna pertence ao primeiro conjunto de partes espectrais e em que outra parte espectral como (304, 305, 306, 307) tendo valores de frequência acima da frequência de adição de lacuna adicionalmente é incluída no primeiro conjunto de partes espectrais primárias.

[00097] Conforme descrito, o decodificador de áudio de domínio espectral (112) é configurado, de modo que uma frequência máxima representada por um valor espectral na primeira representação decodificada seja igual a uma frequência máxima incluída na representação de tempo tendo a taxa de amostragem em que o valor espectral para a frequência máxima no primeiro conjunto de partes espectrais primárias é zero ou diferente de zero. De qualquer forma, para esta frequência máxima no primeiro conjunto de componentes espectrais um fator de escala para a banda de fator de escala existe, que é gerado e transmitido independente se todos os valores espectrais nesta banda de fator de escala são definidos a zero ou não conforme discutido no contexto das

Figuras 3a e 3b.

[00098] A invenção é, assim, vantajosa com relação a outras técnicas paramétricas para aumentar a eficiência de compressão, por exemplo, substituição de ruído e preenchimento de ruído (estas técnicas são exclusivamente para representação eficiente do ruído tipo conteúdo do sinal local) a invenção permite uma reprodução de frequência precisa de componentes tonais. Até o momento, nenhuma técnica do estado da técnica direciona a representação paramétrica eficiente do conteúdo do sinal arbitrário pela adição de lacuna espectral sem a restrição de uma divisão prévia anterior na banda baixa (LF) e na banda alta (HF).

[00099] As aplicações do sistema inventivo melhoram as abordagens do estado da técnica e, assim, fornece alta eficiência de compressão, nenhuma ou apenas uma pequena porcentagem de indignação e largura de banda de áudio completo mesmo para baixas taxas de bit.

[000100] O sistema geral consiste em:

- codificação principal da banda completa;
- Preenchimento de Lacuna Inteligente (preenchimento de porção ou preenchimento de ruído);
- Partes tonais insuficientes no núcleo selecionado pela máscara tonal;
- Codificação do par de conjunto estéreo para a banda completa, incluindo o preenchimento de porção;
- TNS na porção;
- Ponderação espectral na faixa de IGF.

[000101] Uma primeira etapa em direção a um sistema

mais eficiente é remover a necessidade de transformar os dados espectrais em um segundo domínio de transformada diferente de um do codificador principal. Como a maioria dos codecs de áudio, como AAC por exemplo, utilizam a MDCT como transforada básica, é útil realizar a BWE no domínio de MDCT também. Uma segunda exigência para o sistema de BWE seria a necessidade de preservar a grade tonal pela qual componentes tonais HF iguais são preservados e a qualidade do áudio do codec é, então, superior aos sistemas existentes. Para cuidar de ambas as exigências mencionadas acima para um esquema de BWE, um novo sistema é proposto chamado Preenchimento de Lacuna Inteligente (IGF). A Figura 2b mostra o diagrama em blocos do sistema proposto no lado do codificador e a Figura 2a mostra o sistema no lado do decodificador.

[000189] A Figura 9a ilustra um aparelho para decodificação de um sinal de áudio codificado, compreendendo uma representação codificada de um primeiro conjunto de partes espectrais primárias e uma representação codificada de dados paramétricos indicando energias espectrais para um segundo conjunto de partes espectrais secundárias. O primeiro conjunto de partes espectrais primárias é indicado em (901a) na Figura 9a, e a representação codificada dos dados paramétricos é indicada em (901b) na Figura 9a. Um decodificador de áudio (900) é fornecido para decodificar a representação codificada (901a) do primeiro conjunto de partes espectrais primárias para obter um primeiro conjunto decodificado de partes espectrais primárias (904) e para decodificar a representação codificada dos dados paramétricos

para obter um dado paramétrico decodificado (902) para o segundo conjunto de partes espectrais secundárias indicando energias individuais para as bandas de reconstrução individuais, onde as partes espectrais secundárias estão localizadas na bandas de reconstrução. Além disso, um regenerador de frequência (906) é fornecido para reconstruir valores espectrais de uma banda de reconstrução compreendendo uma segunda parte espectral. O regenerador de frequência (906) utiliza uma primeira parte espectral do primeiro conjunto de partes espectrais primárias e uma informação de energia individual para a banda de reconstrução, onde a banda de reconstrução compreende uma primeira parte espectral e a segunda parte espectral. O regenerador de frequência (906) compreende uma calculadora (912) para determinar uma informação de energia sobrevivente compreendendo uma energia acumulada da primeira parte espectral tendo frequências na banda de reconstrução. Além disso, o regenerador de frequência (906) compreende uma calculadora (918) para determinar uma informação da energia da porção de outras partes espectrais da banda de reconstrução e para valores de frequência sendo diferentes da primeira parte espectral, onde estes valores de frequência têm frequências na banda de reconstrução, em que outras partes espectrais devem ser geradas pela regeneração de frequência utilizando uma primeira parte espectral diferente da primeira parte espectral na banda de reconstrução.

[000190] O regenerador de frequência (906) compreende, ainda, uma calculadora (914) para uma energia ausente na

banda de reconstrução, e a calculadora (914) opera utilizando a energia individual para a banda de reconstrução e a energia sobrevivente gerada pelo bloco (912). Além disso, o regenerador de frequência (906) compreende um ajustador do envelope espectral (916) para ajustar outras partes espectrais na banda de reconstrução com base na informação da energia ausente e na informação da energia da porção gerada pelo bloco (918).

[000191] Referência é feita à Figura 9c que ilustra uma determinada banda de reconstrução (920). A banda de reconstrução compreende uma primeira parte espectral na banda de reconstrução como a primeira parte espectral 306 na Figura 3a esquematicamente ilustrada em (921). Além disso, o resto dos valores espectrais na banda de reconstrução (920) devem ser gerados utilizando uma região de origem, por exemplo, da banda de fator de escala 1, 2, 3 abaixo da frequência inicial de Preenchimento de Lacuna Inteligente (309) da Figura 3a. O regenerador de frequência (906) é configurado para gerar valores espectrais não processados para as partes espectrais secundárias (922 e 923). Então, um fator de ganho g é calculado conforme ilustrado na Figura 9c a fim de finalmente ajustar os valores espectrais não processados nas faixas de frequência (922, 923) para obter as partes espectrais secundárias reconstruídas e ajustadas na banda de reconstrução (920) que agora têm a mesma resolução espectral, ou seja, a mesma distância da linha que a primeira parte espectral (921). É importante entender que a primeira parte espectral na banda de reconstrução ilustrada em (921) na

Figura 9c é decodificada pelo decodificador de áudio (900) e não é influenciada pelo ajuste do envelope realizado pelo bloco (916) da Figura 9b. Ainda, a primeira parte espectral na banda de reconstrução indicada em (921) é deixada como tal, visto que a primeira parte espectral é emitida pela largura de banda total ou decodificador de áudio de taxa total (900) através da linha (904).

[000192] Subsequentemente, um determinado exemplo com números reais é discutido. A energia sobrevivente restante, conforme calculado pelo bloco (912) é, por exemplo, cinco unidades de energia e esta energia é a energia das quatro linhas espectrais exemplarmente indicadas na primeira parte espectral (921).

[000193] Além disso, o valor de energia E3 para a banda de reconstrução correspondente à banda de fator de escala 6 da Figura 3b ou Figura 3a é igual a 10 unidades. De forma importante, o valor de energia não compreende apenas a energia das partes espectrais (922, 923), mas compreende a energia da banda total de reconstrução (920) conforme calculado no lado do codificador, ou seja, antes de realizar a análise espectral utilizando, por exemplo, a máscara de tonalidade. Assim, as dez unidades de energia abrangem a primeira e a segunda parte espectral na banda de reconstrução. Então, assume-se que os dados da energia da faixa de origem para os blocos (922, 923) ou para os dados do intervalo alvo não processado para o bloco (922, 923) é igual a oito unidades de energia. Assim, uma energia ausente de cinco unidades é calculada.

[000194] Com base na energia ausente dividida pela energia da porção tEk, um fator de ganho de 0,79 é calculado. Então, as linhas espectrais não processadas para as partes espectrais secundárias (922, 923) são multiplicadas pelo fator de ganho calculado. Assim, apenas os valores espectrais para as partes espectrais secundárias (922, 923) são ajustados e as linhas espectrais para a primeira parte espectral (921) não são influenciadas por este ajuste do envelope. Subsequente à multiplicação dos valores espectrais não processados para as partes espectrais secundárias (922, 923), uma banda de reconstrução completa foi calculada consistindo nas partes espectrais primárias na banda de reconstrução, e consistindo nas linhas espectrais nas partes espectrais secundárias (922, 923) na banda de reconstrução (920).

[000195] Preferivelmente, a faixa de origem para gerar os dados espectrais não processados nas bandas (922, 923) é, com relação à frequência, abaixo da frequência inicial de IGF (309) e a banda de reconstrução (920) está acima da frequência inicial de IGF (309).

[000196] Além disso, é preferido que os limites da banda de reconstrução coincidem com os limites da banda de fator de escala. Assim, uma banda de reconstrução tem, em uma aplicação, o tamanho das bandas do fator de escala correspondentes do decodificador de áudio principal ou são dimensionados de modo que, quando o pareamento da energia é aplicado, um valor de energia para uma banda de reconstrução fornece a energia de dois ou um número inteiro mais alto de

bandas do fator de escala. Assim, quando é assumido que o acúmulo de energia é realizado para a banda de fator de escala 4, a banda de fator de escala 5 e a banda de fator de escala 6, então o limite inferior de frequência da banda de reconstrução (920) é igual ao limite inferior da banda de fator de escala 4 e o limite mais alto de frequência da banda de reconstrução (920) coincide com o limite mais alto da banda de fator de escala 6.

[000197] Subsequentemente, a Figura 9d é discutida a fim de mostrar outras funcionalidades do decodificador da Figura 9a. O decodificador de áudio (900) recebe os valores espectrais dequantizados correspondentes às partes espectrais primárias do primeiro conjunto de partes espectrais e, adicionalmente, fatores de escala para bandas do fator de escala como ilustrado na Figura 3b são fornecidos em um bloco de escala inversa (940). O bloco de escala inversa (940) fornece todos os primeiros conjuntos das partes espectrais primárias abaixo da frequência inicial de IGF (309) da Figura 3a e, adicionalmente, as partes espectrais primárias acima da frequência inicial de IGF, ou seja, as partes espectrais primárias (304, 305, 306, 307) da Figura 3a que estão todas localizadas em uma banda de reconstrução conforme ilustrado em (941) na Figura 9d. Além disso, as partes espectrais primárias na banda de origem utilizadas para preenchimento da porção de frequência na banda de reconstrução são fornecidas ao ajustador/calculadora do envelope (942) e este bloco adicionalmente recebe a informação de energia para a banda de reconstrução fornecia como informação adicional paramétrica

ao sinal de áudio codificado conforme ilustrado em (943) na Figura 9d. Então, o ajustador/calculadora do envelope (942) fornece as funcionalidades da Figura 9b e 9c e finalmente emite os valores espectrais ajustados para as partes espectrais secundárias na banda de reconstrução. Estes valores espectrais ajustados (922, 923) para as partes espectrais secundárias na banda de reconstrução e as partes espectrais primárias (921) na banda de reconstrução indicadas na linha (941) na Figura 9d conjuntamente representam a representação espectral da banda de reconstrução completa.

[000198] Subsequentemente, referência é feita às Figuras 10a a 10b para explicar as aplicações preferidas de um codificador de áudio para codificar um sinal de áudio para fornecer ou gerar um sinal de áudio codificado. O codificador compreende um conversor de tempo/espectro (1002) que alimenta um analisador espectral (1004), e o analisador espectral 1004 é conectado a uma calculadora de parâmetro (1006), por um lado e um codificador de áudio (1008), por outro lado. O codificador de áudio (1008) fornece a representação codificada de um primeiro conjunto de partes espectrais primárias e não abrange o segundo conjunto de partes espectrais secundárias. Por outro lado, a calculadora de parâmetro (1006) fornece informação de energia para uma banda de reconstrução que abrange a primeira e a segunda parte espectral. Além disso, o codificador de áudio (1008) é configurado para gerar uma primeira representação codificada do primeiro conjunto de partes espectrais primárias tendo a primeira resolução espectral, onde o codificador de áudio

(1008) fornece fatores de escala para todas as bandas da representação espectral gerada pelo bloco (1002). Adicionalmente, conforme ilustrado na Figura 3b, o codificador fornece a informação de energia pelo menos para bandas de reconstrução localizadas, com relação à frequência, acima da frequência inicial de IGF (309) conforme ilustrado na Figura 3a. Assim, para bandas de reconstrução preferivelmente que coincidem com as bandas do fator de escala ou com grupos de bandas do fator de escala, dois valores são dados, ou seja, o fator de escala correspondente do codificador de áudio (1008) e, adicionalmente, a informação de energia emitida pela calculadora de parâmetro (1006).

[000199] O codificador de áudio preferivelmente tem bandas do fator de escala com diferentes larguras de banda de frequência, ou seja, com um diferente número de valores espectrais. Assim, a calculadora paramétrica compreende um normalizador (1012) para normalizar as energias para a diferente largura de banda com relação à largura de banda da banda de reconstrução específica. Para esta finalidade, o normalizador (1012) recebe, como entradas, uma energia na banda e um número de valores espectrais na banda e o normalizador (1012) então emite uma energia normalizada por reconstrução/banda de fator de escala.

[000200] Além disso, a calculadora paramétrica (1006a) da Figura 10a compreende uma calculadora do valor de energia que recebe a informação de controle do núcleo ou do codificador de áudio (1008) conforme ilustrado pela linha

(1007) na Figura 10a. Esta informação de controle pode compreender a informação em blocos longos/curtos utilizados pelo codificador de áudio e/ou informação de agrupamento. Assim, enquanto a informação sobre os blocos longos/curtos e a informação de agrupamento em janelas curtas referem-se a um agrupamento de "tempo", a informação de agrupamento pode adicionalmente se referir a um agrupamento espectral, ou seja, o agrupamento de duas bandas do fator de escala em uma única banda de reconstrução. Assim, a calculadora do valor de energia (1014) emite um único valor de energia para cada banda agrupada que abrange a primeira e a segunda parte espectral quando apenas as partes espectrais foram agrupadas.

[000201] A Figura 10d ilustra outra aplicação para implementar o agrupamento espectral. Para esta finalidade, o bloco (1016) é configurado para calcular valores de energia para duas bandas adjacentes. Então, no bloco (1018), os valores de energia para as bandas adjacentes são comparados e, quando os valores de energia não são muito diferentes ou menos diferentes do que os definidos, por exemplo, por um limite, então um único valor (normalizado) para ambas as bandas é gerado conforme indicado no bloco (1020). Conforme ilustrado pela linha (1019), o bloco (1018) pode ser ignorado. Além disso, a geração de um único valor para duas ou mais bandas realizadas pelo bloco (1020) pode ser controlada por um controle de taxa de bit do codificador (1024). Assim, quando a taxa de bit deve ser reduzida, o controle da taxa de bit codificada (1024) controla o bloco (1020) para gerar um único valor normalizado para duas ou

mais bandas mesmo embora a comparação no bloco (1018) não seria permitida agrupar os valores da informação de energia.

[000202] No caso em que o codificador de áudio está realizando o agrupamento de duas ou mais janelas curtas, este agrupamento é aplicado para a informação de energia também. Quando o codificador principal realiza um agrupamento de dois ou mais blocos curtos, então, para estes dois ou mais blocos, apenas um único conjunto de fatores de escala é calculado e transmitido. No lado do decodificador, o decodificador de áudio então aplica o mesmo conjunto de fatores de escala para ambas as janelas agrupadas.

[000203] Considerando o cálculo da informação de energia, os valores espectrais na banda de reconstrução são acumulados sobre duas ou mais janelas curtas. Em outras palavras, isso significa que os valores espectrais em determinada banda de reconstrução para um bloco curto e para o bloco curto subsequente são acumulados juntos e apenas um único valor da informação de energia é transmitido por essa banda de reconstrução cobrindo dois blocos curtos. Então, no lado do decodificador, o ajuste do envelope discutido em relação à Figura 9a a 9d não é realizado individualmente para cada bloco curto, mas é realizado junto para o conjunto de janelas curtas.

[000204] A normalização correspondente é, então, novamente aplicada de modo que muito embora qualquer agrupamento em frequência ou agrupamento em tempo foi realizado, a normalização permite facilmente que, para o cálculo de informação de valor de energia no lado do

decodificador, apenas o valor da informação de energia, por um lado e a quantidade de linhas espectrais na banda de reconstrução ou no conjunto de bandas de reconstrução agrupadas tem que ser conhecidas.

[000205] Além disso, enfatiza-se que uma informação sobre as energias espectrais, uma informação sobre as energias individuais ou uma informação de energia individual, uma informação sobre uma energia de sobrevivência ou uma informação de energia de sobrevivência, uma informação da energia de porção ou uma informação da energia de porção, ou uma informação sobre uma energia ausente ou uma informação de energia ausente pode compreender não apenas um valor de energia, mas também um valor de amplitude (p.ex., absoluto), um valor de nível ou qualquer outro valor, do qual um valor de energia final pode ser derivado. Assim, a informação sobre uma energia pode, por exemplo, compreender o próprio valor de energia, e/ou um valor de um nível e/ou de uma amplitude e/ou de uma amplitude absoluta.

[000206] A Figura 12a ilustra uma implementação adicional do aparelho para decodificação. Um fluxo contínuo de bit é recebido por um decodificador principal (1200) que pode, por exemplo, ser um decodificador AAC. O resultado é configurado em um estágio para realizar uma correção ou partição da extensão da largura de banda (1202) correspondente ao regenerador de frequência (604), por exemplo. Então, um procedimento de adaptação de correção/porção e pós-processamento é realizado e, quando uma adaptação de correção foi realizada, o regenerador de

frequência (1202) é controlado para realizar uma regeneração de frequência adicional, mas agora com, por exemplo, limites de frequência ajustados. Além disso, quando um processamento de correção é realizado como pela eliminação ou atenuação de linhas tonais, o resultado é, então, encaminhado ao bloco (1206) realizando a formação do envelope da largura de banda acionada por parâmetro como, por exemplo, também discutido no contexto do bloco (712 ou 826). O resultado é então encaminhado a um bloco de transformada de síntese (1208) para realizar uma transformada no domínio de saída final que é, por exemplo, um domínio de saída PCM, conforme ilustrado na Figura 12a.

[000207] Os recursos principais das aplicações da invenção são conforme segue:

[000208] A aplicação preferida se baseia na MDCT que exibe as perturbações de modulação de frequência referenciadas acima se as áreas espectrais tonais são reduzidas pela escolha infeliz da frequência cruzada e/ou margens de correção, ou componentes tonais serem colocados muito próximos nos limites de correção.

[000209] A Figura 12b mostra como a técnica atualmente proposta reduz as perturbações encontradas nos métodos de BWE do estado da técnica. Na Figura 12, o painel (2), o espectro estilizado de magnitude da saída de um método de BWE contemporâneo é mostrado. Nesse exemplo, o sinal é perceptualmente prejudicado pela pulsação causada por dois tons próximos e também pela divisão de um tom. Ambas as áreas espectrais problemáticas são marcadas com um círculo cada.

[000210] Para superar esses problemas, a nova técnica primeiro detecta a localização espectral dos componentes tonais contidos no sinal. Então, de acordo com um aspecto da invenção, tenta-se ajustar as frequências de transição entre LF e todas as correções por mudanças individuais (dentro dos limites dados) de modo que a divisão ou pulsação de componentes tonais seja reduzido. Para essa finalidade, a frequência de transição preferivelmente tem que combinar um mínimo espectro local. Essa etapa é mostrada na Figura 12b painel (2) e painel (3), onde a frequência de transição f_{x2} é mudada em direção às frequências mais altas, resultando em f'_{x2}.

[000211] De acordo com outro aspecto da invenção, se o conteúdo espectral problemático em regiões de transição permanecer, pelo menos, um dos componentes tonais deslocados é removido para reduzir tanto a perturbação de pulsação nas frequências de transição quanto à modulação de frequência. Isso é feito através da extrapolação ou interpolação / filtragem espectral, conforme mostrado no painel da Figura 2 (3). Um componente tonal é, assim, removido de ponta a ponta, ou seja, do mínimo local esquerdo para seu mínimo local direito. O espectro resultante após a aplicação da tecnologia inventiva é mostrado no painel da Figura 12b (4).

[000212] Em outras palavras, a Figura 12b ilustra, no canto esquerdo superior, ou seja, no painel (1), o sinal original. No canto direito superior, ou seja, no painel (2), um sinal estendido da largura de banda de comparação com áreas problemáticas marcadas por elipses (1220 e 1221) é

mostrado. No canto esquerdo inferior, ou seja, no painel (3), dois recursos de processamento preferidos de correção ou porção de frequência são ilustrados. A divisão das partes tonais foi direcionada pelo aumento do limite de frequência f'_{x2} de modo que um recorte da parte tonal correspondente não está mais lá. Além disso, as funções de ganho (1030) para eliminar a parte tonal (1031 e 1032) são aplicadas ou, de modo alternativo, uma interpolação ilustrada por (1033) é indicada. Finalmente, o canto direito inferior da Figura 12b, ou seja, painel (4) descreve o sinal melhorado resultante de uma combinação de frequência de correção/porção que ajusta por um lado e eliminação ou pelo menos atenuação das partes tonais problemáticas.

[000213] O painel (1) da Figura 12b ilustra, conforme discutido antes, o espectro original e o espectro original tem uma faixa de frequência principal de até o cruzamento ou a frequência inicial do preenchimento de lacuna *fx1*.

[000214] Assim, uma frequência f_{x1} ilustra uma frequência limite (1250) entre a faixa de origem (1252) e a faixa de reconstrução (1254) se estendendo entre a frequência limite (1250) e a frequência máxima que é menor ou igual à frequência de Nyquist $f_{Nyquist}$. No lado do codificador, assume-se que um sinal é limitado pela largura de banda em f_{x1} ou, quando a tecnologia referente ao preenchimento de lacuna inteligente é aplicada, assume-se que f_{x1} corresponde à frequência de preenchimento de lacuna inicial (309) da Figura 3a. Dependendo da tecnologia, a faixa de reconstrução acima de f_{x1} será vazio (no caso da implementação da Figura 13a,

13b) ou compreenderá certas primeiras partes espectrais a ser codificadas com uma alta resolução conforme discutido no contexto da Figura 3a.

[000215] A Figura 12b, painel (2), ilustra um sinal regenerado preliminar, por exemplo, gerado pelo bloco (702) da Figura 7a que tem duas partes problemáticas. Uma parte problemática é ilustrada em (1220). A distância de frequência entre a parte tonal dentro da região principal ilustrada em (1220a) e a parte tonal no início da porção de frequência ilustrada em (1220b) é muito pequena de modo que uma perturbação de pulsação seria criada. O problema adicional é que no limite superior da primeira porção de frequência gerada pela primeira operação de correção ou operação de porção de frequência ilustrada em (1225) é uma parte tonal recortada ao meio ou dividida (1226). Quando essa parte tonal (1226) é comparada a outras partes tonais na Figura 12b, fica claro que a largura é menor que a largura de uma típica parte tonal e isso significa que essa parte tonal foi dividia pela definição do limite de frequência entre a primeira porção de frequência (1225) e a segunda porção de frequência (1227) no local errado na faixa de origem (1252). A fim de direcionar esse problema, a frequência limite f_{x2} foi modificada para se tornar um pouco maior conforme ilustrado no painel (3) na Figura 12b, de modo que um recorte dessa parte tonal não ocorra.

[000216] Por outro lado, esse procedimento, no qual f'_{x2} foi mudado, não aborda efetivamente o problema da pulsação que, assim, é abordado por uma remoção dos

componentes tonais pela filtragem ou interpolação ou qualquer outro procedimento, conforme discutido no contexto do bloco (708) da Figura 7a. Assim, a Figura 12b ilustra uma aplicação sequencial do ajuste da frequência de transição (706) e a remoção dos componentes tonais nos limites ilustrados em (708).

[000217] Outra opção seria definir o limite de transição f_{x1}, de modo que seja pouco mais baixo, a fim de que a parte tonal (1220a) não esteja mais na faixa principal. Então, a parte tonal (1220a) também foi removida ou eliminada pela definição da frequência de transição f_{x1} em um valor inferior.

[000218] Esse procedimento também teria trabalhado para abordar o problema com o componente tonal problemático (1032). Ao definir, mesmo mais alto, f'_{x2}, a parte espectral onde a parte tonal (1032) está localizada poderia ter sido regenerada dentro da primeira operação de correção (1225) e, assim, duas partes tonais adjacentes ou próximas não teriam ocorrido.

[000219] Basicamente, o problema de pulsação depende das amplitudes e da distância na frequência das partes tonais adjacentes. O detector (704, 720) ou declarado de forma mais geral, o analisador (602) é preferivelmente configurado de modo que uma análise da parte espectral inferior localizada na frequência abaixo da frequência de transição como f_{x1}, $f_{x2,}$ f'_{x2} é analisado a fim de localizar qualquer componente tonal. Além disso, a faixa espectral acima da frequência de transição é também analisada a fim de detectar um componente

tonal. Quando a detecção resulta em dois componentes tonais, um à esquerda da frequência de transição com relação à frequência e um à direita (com relação à frequência ascendente), então o removedor de componentes tonais nos limites ilustrados em (708) na Figura 7a é ativado. A detecção dos componentes tonais é realizada em uma certa faixa de detecção que se estende, da frequência de transição, em ambas as direções pelo menos 20% com relação à largura de banda da banda correspondente e preferivelmente apenas se estende até 10% abaixo à esquerda da frequência de transição e acima à direita da frequência de transição com relação à largura de banda correspondente, ou seja, a largura de banda da faixa de origem por um lado e a faixa de reconstrução por outro lado ou, quando a frequência de transição é a frequência de transição entre as duas porções de frequência (1225, 1227), uma quantidade correspondente de 10% da porção de frequência correspondente. Em uma aplicação adicional, a largura de banda de detecção predeterminada é uma Escala de Bark. Deveria ser possível remover as partes tonais dentro de uma faixa de 1 na Escala de Bark ao redor de um limite de correção, de modo que a completa faixa de detecção é 2 na Escala de Bark, ou seja, uma Escala de Bark na banda inferior e uma Escala de Bark na banda mais alta, onde a Escala de Bark 1 na banda inferior é imediatamente adjacente a uma Escala de Bark na banda mais alta.

[000220] De acordo com outro aspecto da invenção, para reduzir a perturbação do toque do filtro, um filtro cruzado no domínio de frequência é aplicado a duas regiões espectrais

consecutivas, ou seja, entre a banda principal e a primeira correção ou entre as duas correções. Preferivelmente, o filtro cruzado é adaptativo ao sinal.

[000221] O filtro cruzado consiste em dois filtros, um filtro de *fade-out* $\boldsymbol{h}_{out}$, que é aplicado à região espectral inferior e um filtro de *fade-in* $\boldsymbol{h}_{in}$, que é aplicado na região espectral mais alta.

[000222] Cada um dos filtros tem comprimento N.

[000223] Além disso, a curva de ambos os filtros é caracterizada por um valor adaptativo ao sinal chamado $Xbias$ que determina a característica de corte do filtro cruzado, com $0 \leq Xbias \leq N$:

[000224] Se $Xbias = 0$, então, a soma de ambos os filtros é igual a 1, ou seja, não há característica de faixa de corte de filtro no filtro resultante.

[000225] Se $Xbias = N$, então, ambos os filtros são completamente zero.

[000226] O desenho básico dos filtros cruzados é restrito às seguintes equações:

$$\boldsymbol{h}_{out}(k) = \boldsymbol{h}_{in}(N-1-k), \forall\, Xbias$$

$$\boldsymbol{h}_{out}(k) + \boldsymbol{h}_{in}(k) = 1, Xbias = 0$$

com $k = 0, 1, \ldots, N-1$ sendo o índice de frequência. A Figura 12c mostra um exemplo de tal filtro cruzado.

[000227] Nesse exemplo, a seguinte equação é utilizada para criar o filtro h_{out}:

$$\boldsymbol{h}_{out}(k) = 0{,}5 + 0{,}5 \cdot cos\left(\frac{k}{N-1-Xbias} \cdot \pi\right), k = 0, 1, \ldots, N-1-Xbias$$

[000228] A seguinte equação descreve como os filtros

h_{in} e h_{out} são, então, aplicados,

$$\boldsymbol{Y}(k_t-(N-1)+k) = \boldsymbol{LF}(k_t-(N-1)+k)\cdot \boldsymbol{h}_{out}(k) + \boldsymbol{HF}(k_t-(N-1)+k)\cdot \boldsymbol{h}_{in}(k), \quad k=0,1,\dots,N-1$$

com $\boldsymbol{Y}$ denotando o espectro montado, k_t sendo a frequência de transição, $\boldsymbol{LF}$ sendo o conteúdo de baixa frequência e $\boldsymbol{HF}$ sendo o conteúdo de alta frequência.

[000229] Em seguida, a evidência dos benefícios dessa técnica será apresentada. O sinal original nos seguintes exemplos é um sinal do tipo transiente, em particular, uma versão de filtro passa baixo, com uma frequência de corte de 22 kHz. Primeiro, esse transiente é a banda limitada a 6 kHz no domínio de transformada. Subsequentemente, a largura de banda do sinal original filtrado por passa baixo é estendido a 24 kHz. A extensão da largura de banda é realizada através da cópia da banda LF três vezes para preencher completamente a faixa de frequência que está disponível acima de 6 kHz dentro da transformada.

[000230] A Figura 11a mostra o espectro desse sinal, que pode ser considerado como um espectro típico de uma perturbação do toque do filtro que espectralmente envolve o transiente devido à dita característica de parede de tijolos da transformada (picos de fala 1100). Pela aplicação da abordagem inventiva, o toque do filtro é reduzido em aproximadamente 20 dB em cada frequência de transição (picos de fala reduzidos).

[000231] O mesmo efeito, mesmo em uma ilustração diferente, é mostrado na Figura 11b, 11c. A Figura 11b mostra o espectrograma do sinal do tipo transiente mencionado com a perturbação do toque do filtro que temporariamente precede e

sucede o transiente após aplicar a técnica de BWE descrita acima sem qualquer redução do toque do filtro. Cada uma das linhas horizontais representa o toque do filtro na frequência de transição entre as correções consecutivas. A Figura 6 mostra o mesmo sinal após aplicar a abordagem inventiva dentro do BWE. Através da aplicação de redução de toque, o toque do filtro é reduzido em aproximadamente 20 dB comparado ao sinal exibido na Figura anterior.

[000232] Subsequentemente, as Figuras 14a, 14b são discutidas, a fim de ilustrar, ainda, o aspecto da invenção do filtro cruzado já discutido no contexto com o recurso do analisador. Entretanto, o filtro cruzado (710) também pode ser implementado independente da invenção discutida no contexto das Figuras 6a–7b.

[000233] A Figura 14a ilustra um aparelho para decodificação de um sinal de áudio codificado que compreende um sinal principal codificado e informação sobre os dados paramétricos. O aparelho compreende um decodificador principal (1400) para decodificar o sinal principal codificado para obter um sinal principal decodificado. O sinal principal decodificado pode ser limitado pela largura de banda no contexto da implementação da Figura 13a, Figura 13b ou o decodificador principal pode ser uma faixa de frequência total ou codificador de taxa total no contexto das Figuras de 1 a 5c ou de 9a a 10d.

[000234] Além disso, um gerador de porção (1404) para regenerar uma ou mais porção(ões) espectral(is), tendo frequências não incluídas no sinal principal decodificado, é

gerado utilizando uma parte espectral do sinal principal decodificado. As porções podem ser segundas partes espectrais reconstruídas dentro de uma banda de reconstrução como, por exemplo, ilustrado no contexto da Figura 3a ou que podem incluir as primeiras partes espectrais a serem reconstruídas com uma alta resolução, mas de modo alternativo, as porções espectrais também podem compreender bandas de frequência completamente vazias quando o codificador realizou uma limitação de banda rígida, conforme ilustrado na Figura 13a.

[000235] Além disso, um filtro cruzado (1406) é fornecido para filtragem cruzada espectral do sinal principal decodificado e a primeira porção de frequência tendo frequências que se estendem de uma frequência de preenchimento de lacuna (309) em uma primeira frequência de parada de porção ou para filtragem cruzada espectral de uma primeira porção de frequência (1225) e uma segunda porção de frequência (1221), a segunda porção de frequência tendo uma frequência de limite inferior sendo adjacente à frequência em uma frequência de limite superior da primeira porção de frequência (1225).

[000236] Em uma implementação adicional, o sinal de saída do filtro cruzado (1406) é inserido em um ajustador de envelope (1408) que aplica a informação do envelope espectral paramétrica incluída em um sinal de áudio codificado como informação do lado paramétrico para finalmente obter um sinal regenerado ajustado pelo envelope. Os elementos (1404, 1406, 1408) podem ser implementados como um regenerador de frequência como, por exemplo, ilustrado na Figura 13b, na

Figura 1b ou na Figura 6a, por exemplo.

[000237] A Figura 14b ilustra uma implementação adicional do filtro cruzado (1406). O filtro cruzado (1406) compreende um subfiltro de *fade-out* que recebe um primeiro sinal de entrada IN1, e um segundo subfiltro de *fade-in* (1422) que recebe uma segunda entrada IN2 e os resultados ou saídas de ambos os filtros (1420 e 1422) são fornecidos a um combinador (1424) que é, por exemplo, um somador. O somador ou combinador (1424) emite os valores espectrais para os índices de frequência. A Figura 12c ilustra uma função exemplar de mixagem que compreende a característica do subfiltro de *fade-out* (1420a) e a característica do subfiltro de *fade-in* (1422a). Ambos os filtros têm uma certa sobreposição de frequência no exemplo na Figura 12c igual a 21, ou seja, N=21. Assim, outros valores de frequência de, por exemplo, a região de origem (1252) não são influenciados. Apenas os índices de frequência mais altos (21) da faixa de origem (1252) são influenciados pela função de *fade-out* (1420a).

[000238] Por outro lado, apenas as linhas de frequência mais baixas (21) da primeira porção de frequência (1225) são influenciadas pela função de *fade-in* (1422a).

[000239] Adicionalmente, fica claro a partir das funções de eliminação de ruído que as linhas de frequência entre 9 e 13 são influenciadas, mas a função de *fade-in* realmente não influencia as linhas de frequência entre 1 e 9 e a função de *fade-out* (1420a) não influencia as linhas de frequência entre 13 e 21. Isso significa que apenas uma

sobreposição seria necessária entre as linhas de frequência 9 e 13 e a frequência cruzada como f_{x1} seria colocada na amostra de frequência ou no índice de frequência 11. Assim, apenas uma sobreposição de dois índices de frequência ou valores de frequência entre a faixa de origem e a primeira porção de frequência seria necessária a fim de implementar a função cruzada ou de eliminação de ruído.

[000240] Dependendo da implementação específica, uma sobreposição mais alta ou mais baixa pode ser aplicada e, adicionalmente, outras funções de eliminação de ruído longe de uma função de cosseno podem ser utilizadas. Além disso, conforme ilustrado na Figura 12c, prefere-se aplicar um certo corte na faixa cruzada. Declarado de forma diferente, a energia nas faixas limite será reduzida devido ao fato que ambas as funções de filtro não concluem como seria o caso em uma função de eliminação de ruído sem corte. Essa perda de energia para os limites da porção de frequência, ou seja, a primeira porção de frequência será atenuada no limite inferior e no limite superior, as energias concentradas mais ao meio das bandas. Devido ao fato, entretanto, que o ajuste de envelope espectral ocorre subsequente ao processamento pelo filtro cruzado, toda a frequência não é tocada, mas é definida pelos dados do envelope espectral como os fatores de escala correspondentes, conforme discutido no contexto da Figura 3a. Em outras palavras, a calculadora (918) da Figura 9b então calcularia “faixa alvo não processada já gerada”, que é a saída do filtro cruzado. Além disso, a perda de energia devido à remoção de uma parte tonal pela interpolação

também seria compensada para devido ao fato de que essa remoção então resultaria em uma energia de porção inferior e o fator de ganho para a completa banda de reconstrução ficar mais alta. Por outro lado, entretanto, a frequência cruzada resulta em uma concentração de energia mais ao meio de uma porção de frequência e isso, no final, efetivamente reduz as perturbações, particularmente causadas pelos transientes, conforme discutido no contexto das Figuras 11a–11c.

[000241] A Figura 14b ilustra diferentes combinações de entrada. Para uma filtragem no limite entre a faixa de frequência de origem e a porção de frequência, a entrada 1 é a parte espectral superior da faixa principal e a entrada 2 é a parte espectral inferior da primeira porção de frequência ou da única single porção de frequência, quando apenas uma única porção de frequência existir. Além disso, a entrada pode ser a primeira porção de frequência e a frequência de transição pode ser o limite de frequência superior da primeira porção e a entrada ao subfiltro (1422) será a parte inferior da segunda porção de frequência. Quando uma terceira porção de frequência adicional existir, então uma frequência de transição adicional será o limite de frequência entre a segunda porção de frequência e a terceira porção de frequência e a entrada ao subfiltro de *fade-out* (1421) será a faixa espectral superior da segunda porção de frequência, conforme determinado pelo parâmetro de filtro, quando a característica da Figura 12c é utilizada e a entrada ao subfiltro de *fade-in* (1422) será a parte inferior da terceira porção de frequência e, no exemplo da Figura 12c, as linhas

espectrais mais baixas 21.

[000242] Conforme ilustrado na Figura 12c, prefere-se ter o parâmetro N igual para o subfiltro de *fade-out* e o subfiltro de *fade-in*. Isso, entretanto, não é necessário. Os valores para *N* podem variar e o resultado então será que o filtro "corte" será assimétrico entre a faixa inferior e a faixa superior. Adicionalmente, as funções *fade-in / fade-out* não têm necessariamente que estar na mesma característica que na Figura 12c. Ainda, as características assimétricas também podem ser utilizadas.

[000243] Além disso, é preferido tornar a característica do filtro cruzado adaptativa ao sinal. Assim, com base em uma análise de sinal, a característica do filtro é adaptada. Devido ao fato que o filtro cruzado é particularmente útil para os sinais transientes, é detectado se os sinais transientes ocorrerem. Quando os sinais transientes ocorrem, então uma característica do filtro como ilustrada na Figura 12c poderia ser utilizada. Quando, entretanto, um sinal não transiente é detectado, é preferido mudar a característica do filtro para reduzir a influência do filtro cruzado. Isso poderia, por exemplo, ser obtido definindo N como zero ou definindo X_{bias} como zero de modo que a soma de ambos os filtros seja igual a 1, ou seja, não há característica de faixa de corte de filtro no filtro resultante. De modo alternativo, o filtro cruzado (1406) poderia ser simplesmente derivado no caso de sinais não transientes. Preferivelmente, entretanto, uma característica de mudança relativamente lenta do filtro mudando os

parâmetros N, X_{bias} é preferida, a fim de evitar as perturbações obtidas pelas características do filtro que mudam rapidamente. Além disso, um filtro passa-baixa é preferido para permitir apenas tais mudanças relativamente pequenas de característica do filtro embora o sinal esteja mudando mais rapidamente conforme detectado por um certo detector transiente/de tonalidade. O detector é ilustrado em (1405) na Figura 14a. Ele pode receber um sinal de entrada em um gerador de porção ou um sinal de saída do gerador de porção (1404) ou pode ainda ser conectado ao decodificador principal (1400) a fim de obter uma informação transiente/não transiente como uma indicação de bloco curto da decodificação AAC, por exemplo. Naturalmente, qualquer outro filtro cruzado diferente daquele mostrado na Figura 12c pode ser utilizado também.

[000244] Então, com base na detecção transiente, ou com base na detecção transiente, ou com base em uma detecção de tonalidade ou com base em qualquer outra detecção de característica de sinal, a característica do filtro cruzado (1406) é mudada conforme discutido.

[000287] Embora alguns aspectos tenham sido descritos no contexto de um aparelho para codificação ou decodificação, é evidente que estes aspectos também representam uma descrição do método correspondente, onde um bloco ou dispositivo corresponde a uma etapa do método ou uma característica de uma etapa do método. De forma análoga, os aspectos descritos no contexto de uma etapa do método também representam uma descrição de um bloco correspondente ou item

ou característica de um aparelho correspondente. Algumas ou todas as etapas do método podem ser executadas por (ou utilizando um) aparelho de hardware, como, por exemplo, um microprocessador, um computador programável ou um circuito eletrônico. Em algumas aplicações, algumas ou a maioria das mais importantes etapas do método podem ser executadas pelo referido aparelho.

[000288] Dependendo de certas exigências de implementação, as aplicações da invenção podem ser implementadas em hardware ou em software. A implementação pode ser realizada utilizando um meio de armazenamento não transitório, tal como um meio de armazenamento digital, por exemplo, um disquete, uma Unidade de Disco Rígido (HDD | *Hard Disk Drive*), um DVD, um Blu-Ray, um CD, uma memória ROM, uma PROM, uma EPROM, uma EEPROM ou uma memória FLASH, tendo sinais de controle eletronicamente legíveis armazenados nele que cooperam (ou são capazes de cooperar) com um sistema de computador programável, de modo que o respectivo método seja realizado. Portanto, o meio de armazenamento digital pode ser legível por computador.

[000289] Algumas aplicações, de acordo com a invenção, compreendem um transportador de dados tendo sinais de controle eletronicamente legíveis que são capazes de cooperar com um sistema de computador programável, de modo que um dos métodos descritos neste documento seja realizado.

[000290] De forma geral, as aplicações da presente invenção podem ser implementadas como um produto do programa de computador com um código do programa, o código do programa

sendo operativo para realizar um dos métodos quando o produto do programa de computador é executado em um computador. O código do programa pode, por exemplo, ser armazenado em um transportador legível por máquina.

[000291] Outras aplicações compreendem o programa de computador para realizar um dos métodos descritos neste documento, armazenados em um transportador legível por máquina.

[000292] Em outras palavras, uma aplicação do método inventivo é, assim, um programa de computador tendo um código do programa para realizar um dos métodos descritos neste documento, quando o programa de computador é executado em um computador.

[000293] Outra aplicação dos métodos inventivos é, portanto, um transportador de dados (ou um meio de armazenamento digital, ou um meio legível por computador) compreendendo, gravado nele, o programa de computador para realizar um dos métodos descritos neste documento. O transportador de dados, o meio de armazenamento digital ou o médio gravado são tipicamente tangíveis e/ou não transitórios.

[000294] Outra aplicação do método inventivo é, portanto, um fluxo de dados ou uma sequência de sinais que representam o programa de computador para realizar um dos métodos descritos neste documento. O fluxo de dados ou a sequência de sinais pode, por exemplo, ser configurado para ser transferido através de uma conexão de comunicação de dados, por exemplo, através da Internet.

[000295] Outra aplicação compreende um meio de processamento, por exemplo, um computador ou um dispositivo lógico programável, configurado para ou adaptado para realizar um dos métodos descritos neste documento.

[000296] Outra aplicação compreende um computador tendo instalado nele o programa de computador para realizar um dos métodos descritos neste documento.

[000297] Uma aplicação adicional, de acordo com a invenção, compreende um aparelho ou sistema configurado para transferir (p.ex., ótica ou eletronicamente) um programa de computador para realizar um dos métodos descritos aqui a um receptor. O receptor pode, por exemplo, ser um computador, um dispositivo móvel, um dispositivo de memória ou similar. O aparelho ou sistema pode, por exemplo, compreender um servidor de arquivos para transferir o programa de computador a um receptor.

[000298] Em algumas aplicações, um dispositivo lógico programável (p.ex., um arranjo de portas de campo programáveis) pode ser utilizado para realizar algumas ou todas as funcionalidades dos métodos descritos neste documento. Em algumas aplicações, um arranjo de portas de campo programáveis pode cooperar com um microprocessador, a fim de realizar um dos métodos descritos neste documento. De forma geral, os métodos são preferivelmente realizados por qualquer aparelho de hardware.

[000299] As aplicações descritas acima são meramente ilustrativas para os princípios da presente invenção. Entende-se que modificações e variações das disposições e os

detalhes descritos no presente documento serão evidentes a outros especialistas na técnica. É intenção da invenção, portanto, ser limitada apenas pelo escopo das reivindicações da patente anexas e não pelos detalhes específicos apresentados em forma de descrição e explicação das aplicações no presente documento.

LISTA DE CITAÇÕES

[000300] [1] Dietz, L. Liljeryd, K. Kjörling and O. Kunz, “Spectral Band Replication, a novel approach in audio coding,” in 112th AES Convention, Munich, May 2002.

[000301] [2] Ferreira, D. Sinha, “Accurate Spectral Replacement”, Audio Engineering Society Convention, Barcelona, Spain 2005.

[000302] [3] D. Sinha, A. Ferreira1 and E. Harinarayanan, “A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)”, Audio Engineering Society Convention, Paris, France 2006.

[000303] [4] R. Annadana, E. Harinarayanan, A. Ferreira and D. Sinha, “New Results in Low Bit Rate Speech Coding and Bandwidth Extension”, Audio Engineering Society Convention, San Francisco, USA 2006.

[000304] [5] T. Żernicki, M. Bartkowiak, “Audio bandwidth extension by frequency scaling of sinusoidal partials”, Audio Engineering Society Convention, San Francisco, USA 2008.

[000305] [6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES Convention, Amsterdam, 1998, Preprint 4720.

[000306] [7] M. Neuendorf, M. Multrus, N. Rettelbach, et al., MPEG Unified Speech and Audio Coding–The ISO/MPEG Standard for High-Efficiency Audio Coding of all Content Types, 132nd AES Convention, Budapest, Hungary, April, 2012.

[000307] [8] McAulay, Robert J., Quatieri, Thomas F. “Speech Analysis/Synthesis Based on a Sinusoidal Representation”. IEEE Transactions on Acoustics, Speech, And Signal Processing, Vol 34(4), August 1986.

[000308] [9] Smith, J.O., Serra, X. “PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation”, Proceedings of the International Computer Music Conference, 1987.

[000309] [10] Purnhagen, H.; Meine, Nikolaus, "HILN-the MPEG-4 parametric audio coding tools," *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on* , vol.3, no., pp.201,204 vol.3, 2000

[000310] [11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio”, Geneva, 1998.

[000311] [12] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Oikawa: "MPEG-2 Advanced Audio Coding", 101st AES Convention, Los Angeles 1996

[000312] [13] J. Herre, “Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction”, 17th AES International Conference on

High Quality Audio Coding, August 1999

[000313] [14] J. Herre, “Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction”, 17th AES International Conference on High Quality Audio Coding, August 1999

[000314] [15] International Standard ISO/IEC 23001-3:2010, Unified speech and audio coding Audio, Geneva, 2010.

[000315] [16] International Standard ISO/IEC 14496-3:2005, Information technology – Coding of audio-visual objects – Part 3: Audio, Geneva, 2005.

[000316] [17] P. Ekstrand, “Bandwidth Extension of Audio Signals by Spectral Band Replication”, in Proceedings of 1st IEEE Benelux Workshop on MPCA, Leuven, November 2002

[000317] [18] F. Nagel, S. Disch, S. Wilde, A continuous modulated single sideband bandwidth extension, ICASSP International Conference on Acoustics, Speech and Signal Processing, Dallas, Texas (USA), April 2010.

[000318] [19] Liljeryd, Lars; Ekstrand, Per; Henn, Fredrik; Kjorling, Kristofer: Spectral translation/folding in the subband domain, United States Patent 8,412,365, April 2, 2013.

[000319] [20] Daudet, L.; Sandler, M.; "MDCT analysis of sinusoids: exact results and applications to coding artifacts reduction," Speech and Audio Processing, IEEE Transactions on , vol.12, no.3, pp. 302- 312, May 2004.

REIVINDICAÇÕES

1. Aparelho para decodificação de um sinal de áudio codificado, caracterizado por compreender um sinal principal codificado (1), compreendendo:

um decodificador principal (1400) configurado para decodificar o sinal principal codificado (1401) para obter um sinal principal decodificado;

um gerador de porção (1404) configurado para gerar uma ou mais porção(ões) espectral(is), tendo frequências não incluídas no sinal principal decodificado, utilizando uma porção espectral do sinal principal decodificado; e

um filtro cruzado (1406) configurado para filtrar de forma cruzada e espectral o sinal principal decodificado e uma primeira porção de frequência, tendo frequências que se estendem de uma frequência de preenchimento de lacuna (309) a uma frequência de limite superior, ou para filtragem cruzada espectral de uma primeira porção de frequência e uma segunda porção de frequência;

em que, o filtro cruzado (1406) é configurado para realizar uma adição ponderada por frequência (1424) do sinal principal decodificado filtrado por um subfiltro de *fade-out* (1420) e, pelo menos, uma parte da primeira porção de frequência filtrada por um filtro de *fade-in* (1422) dentro de uma faixa cruzada estendendo-se sobre, pelo menos, três valores de frequência ou para realizar uma adição ponderada por frequência (1424) de, pelo menos, uma parte de uma primeira porção de frequência filtrada pelo subfiltro de

fade-out (1420) e, pelo menos, uma parte de uma segunda porção de frequência filtrada pelo subfiltro de *fade-in* (1422) dentro de uma faixa cruzada que se estende sobre, pelo menos, três valores de frequência.

2. Aparelho de acordo com a reivindicação 1,

caracterizado por uma porção espectral do sinal principal decodificado, uma porção espectral da primeira porção de frequência ou uma porção espectral da segunda porção de frequência influenciada pelo filtro cruzado (1406) ser menor que 30% da porção espectral coberta por uma banda espectral total da banda de frequência principal decodificada ou uma banda espectral total da primeira ou segunda porção de frequência e ser maior ou igual a uma banda definida por, pelo menos, 5 valores de frequência adjacentes.

3. Aparelho de acordo com qualquer uma das reivindicações anteriores,

caracterizado pelo filtro cruzado (1406) ser configurado para aplicar uma característica de filtro do tipo cosseno para *fading-in* (processo de início) e *fading-out* (processo de conclusão).

4. Aparelho de acordo com qualquer uma das reivindicações anteriores, caracterizado por compreender um ajustador de envelope (1408) configurado para ajuste de envelope de um sinal espectral filtrado cruzado em uma faixa espectral definida pelas faixas espectrais de uma ou mais porção(ões) espectral(is), utilizando a informação de envelope espectral paramétrico (1407) incluída no sinal de áudio codificado.

5. Aparelho de acordo com qualquer uma das reivindicações anteriores,

caracterizado por compreender, ainda, um conversor de frequência-tempo (828) configurado para converter um sinal ajustado por envelope junto com o sinal principal decodificado em uma representação de tempo.

6. Aparelho de acordo com a reivindicação 5, caracterizado pelo conversor de frequência-tempo ser configurado para aplicar uma transformada de cosseno discreta modificada inversa (512, 514, 516), compreendendo um processamento de sobreposição/adição (516) de uma estrutura atual com uma estrutura de tempo anterior.

7. Aparelho de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo filtro cruzado ser um filtro controlável,

em que o aparelho compreende, ainda, um detector de características do sinal (1405), e

em que o detector de características do sinal (1405) é configurado para controlar uma característica de filtro do filtro cruzado (1406), de acordo com um resultado de detecção derivado do sinal principal decodificado.

8. Aparelho de acordo com a reivindicação 7,

caracterizado pelo detector de características do sinal (1405) ser um detector transiente e em que o detector transiente (1405) é configurado para controlar o filtro cruzado de modo que, para uma parte do sinal mais transiente, o filtro cruzado tenha um impacto superior sobre um sinal de entrada do filtro cruzado e que o

filtro cruzado (1406) tenha um impacto inferior sobre o sinal de entrada do filtro cruzado para uma parte do sinal menos transiente.

9. Aparelho de acordo com qualquer uma das reivindicações anteriores,

caracterizado por uma característica do filtro cruzado (1406) ser definida por uma característica de subfiltro de *fade-out* (1420a) e uma característica de subfiltro de *fade-in* (1422a),

em que a característica de subfiltro de *fade-in* $h_{in}(k)$ e a característica de subfiltro de *fade-out* $h_{out}(k)$ são definidas com base nas seguintes equações:

$$\boldsymbol{h}_{out}(k) = \boldsymbol{h}_{in}(N-1-k), \forall\, Xbias$$

$$\boldsymbol{h}_{out}(k) + \boldsymbol{h}_{in}(k) = 1, Xbias = 0$$

$$\boldsymbol{h}_{out}(k) = 0{,}5 + 0{,}5 \cdot cos\left(\frac{k}{N-1-Xbias} \cdot \pi\right), k = 0, 1, \dots, N-1-Xbias,$$

em que X_{bias} é um número inteiro que define uma inclinação de ambos os filtros, estendendo-se entre zero e um número inteiro *N*, em que *k* é um índice de frequência que se estende entre zero e *N-1*, em que *N* é um número inteiro adicional e em que diferentes valores para *N* e X_{bias} resultam em diferentes características de filtro cruzado.

10. Aparelho de acordo com a reivindicação 9,

caracterizado por X_{bias} ser definido entre 2 e 20 e em que *N* é definido entre 10 e 50.

11. Aparelho de acordo com qualquer uma das reivindicações anteriores,

caracterizado pelo gerador de porção (1404)

ser configurado para gerar uma porção de frequência preliminar (703), em que um analisador (702) é configurado para analisar a porção de frequência preliminar, em que o gerador de porção é, ainda, configurado para gerar um sinal regenerado, tendo partes tonais que criam perturbação atenuada ou eliminada com relação à porção de frequência preliminar e em que o gerador de porção é configurado para eliminar ou atenuar os limites de porção de frequência próximos aos componentes tonais (708) para obter um sinal de entrada no filtro cruzado (1406).

12. Aparelho de acordo com a reivindicação 11, caracterizado pelo gerador de porção ser configurado para detectar e remover ou atenuar as porções espectrais tonais dentro de uma faixa de detecção sendo menor que 20% de uma largura de banda de uma porção de frequência ou uma faixa de origem para a regeneração.

13. Aparelho de acordo com qualquer uma das reivindicações anteriores, caracterizado pelo filtro cruzado (1406) ser configurado para filtrar de forma cruzada dentro de uma faixa de sobreposição, a faixa de sobreposição compreendendo uma parte do sinal principal decodificado de frequência superior e uma parte de frequência inferior da primeira porção de frequência, ou

em que o filtro cruzado (1406) é configurado para filtrar de forma cruzada dentro de uma faixa de sobreposição, a faixa de sobreposição compreendendo uma parte de frequência superior de uma primeira porção de frequência e uma parte de frequência inferior de uma segunda porção de

frequência.

14. Método para decodificação de um sinal de áudio codificado, caracterizado por compreender um sinal principal codificado (1), compreendendo:

decodificar (1400) o sinal principal codificado (1401) para obter um sinal principal decodificado;

gerar (1404) uma ou mais porção(ões) espectral(is), tendo frequências não incluídas no sinal principal decodificado, utilizando uma porção espectral do sinal principal decodificado; e

filtrar de forma cruzada e espectral (1406) o sinal principal decodificado e uma primeira porção de frequência, tendo frequências estendendo-se de uma frequência de preenchimento de lacuna (309) a uma frequência de limite superior ou filtrar de forma cruzada e espectral uma primeira porção de frequência e uma segunda porção de frequência.

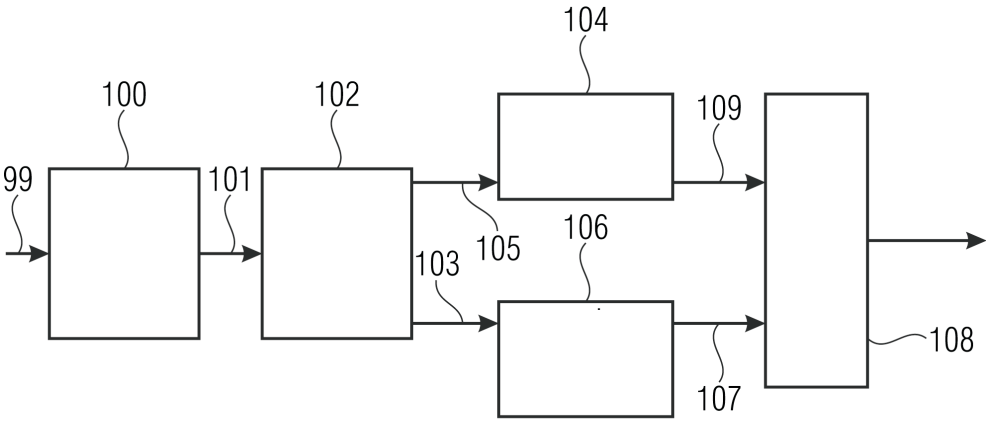

104
100
102
109
99
101
105
106
103
108
107

FIG 1A

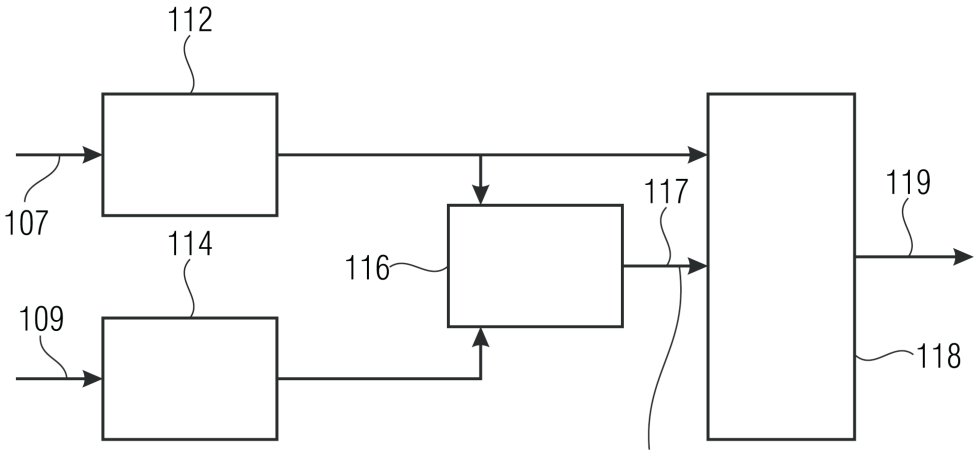

112
107
114
116
117
119
109
118

FIG 1B

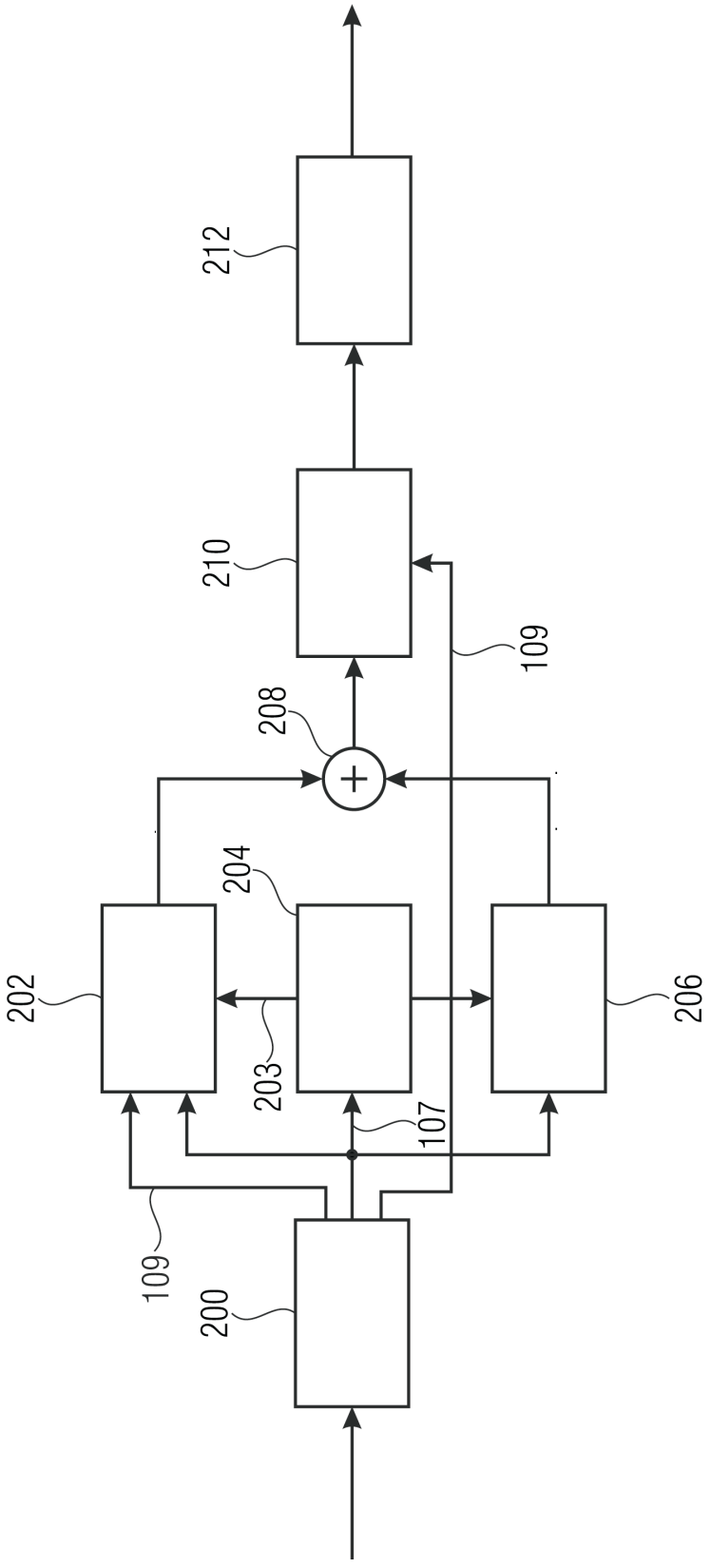


FIG 2A

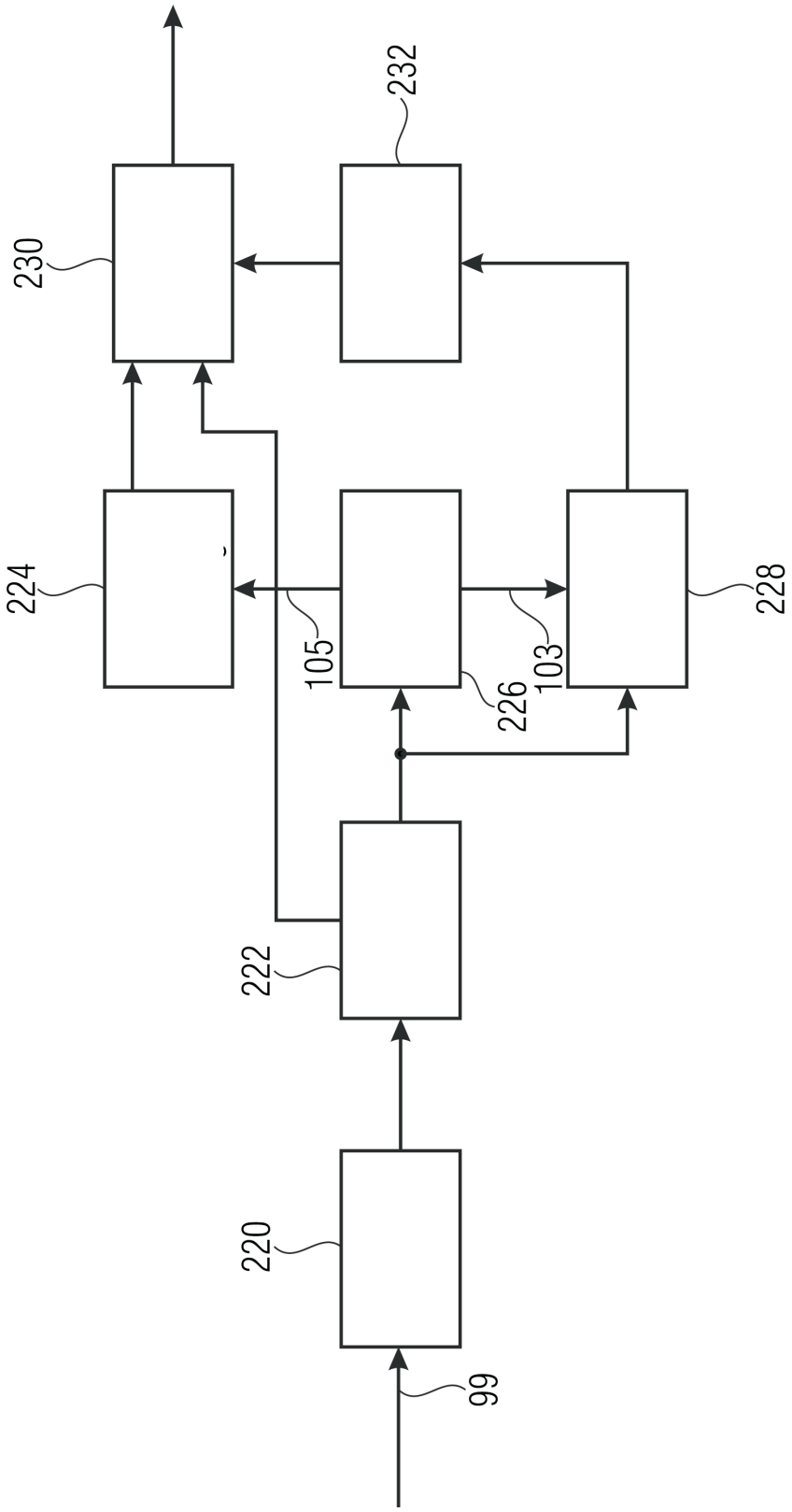


FIG 2B

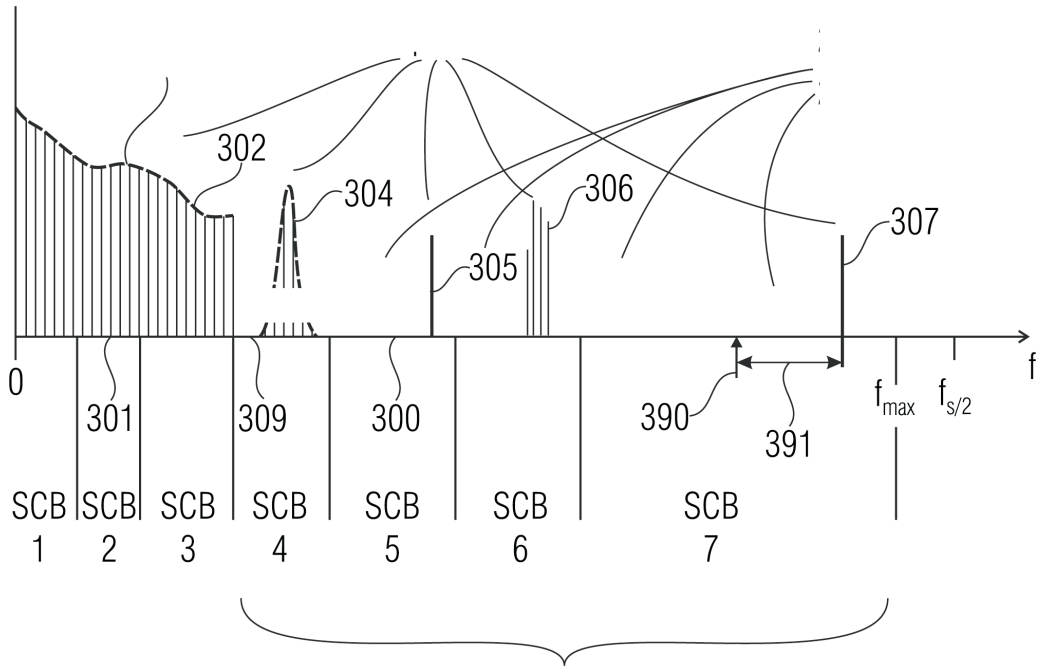

302
304
306
307
305
0
301
309
300
390
391
f
f_{max}
$f_{s/2}$
SCB 1
SCB 2
SCB 3
SCB 4
SCB 5
SCB 6
SCB 7

FIG 3A

SCB1	SCB2	SCB3	SCB4	SCB5	SCB6	SCB7
SF1	SF2	SF3	SF4	SF5	SF6	SF7
			E_1	E_2	E_3	E_4
	NF_2					

308 310 312

FIG 3B

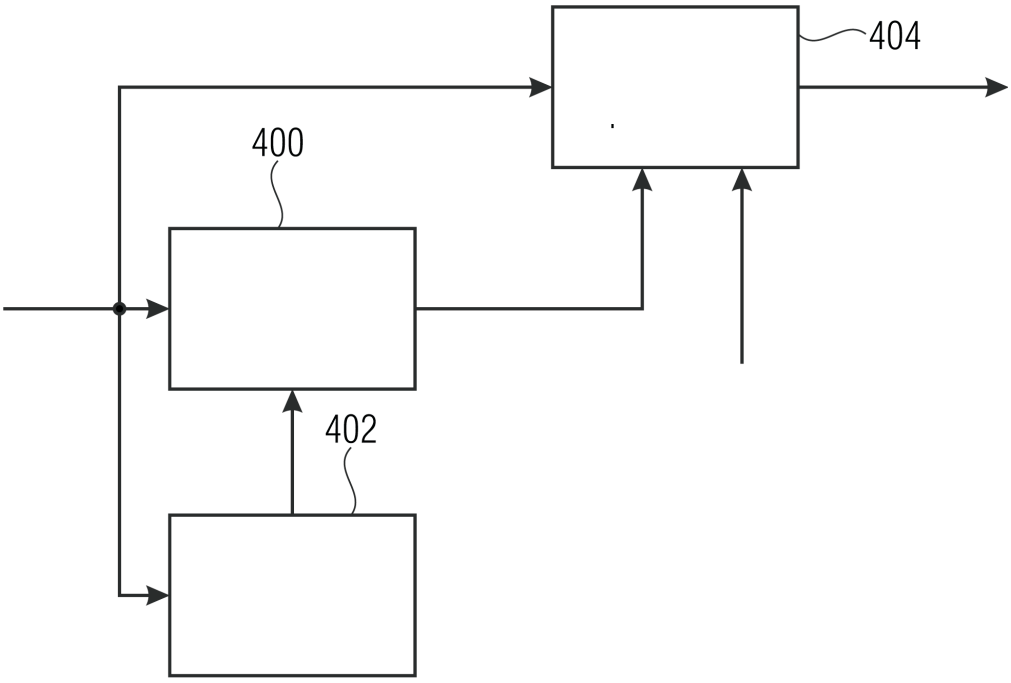


FIG 4A

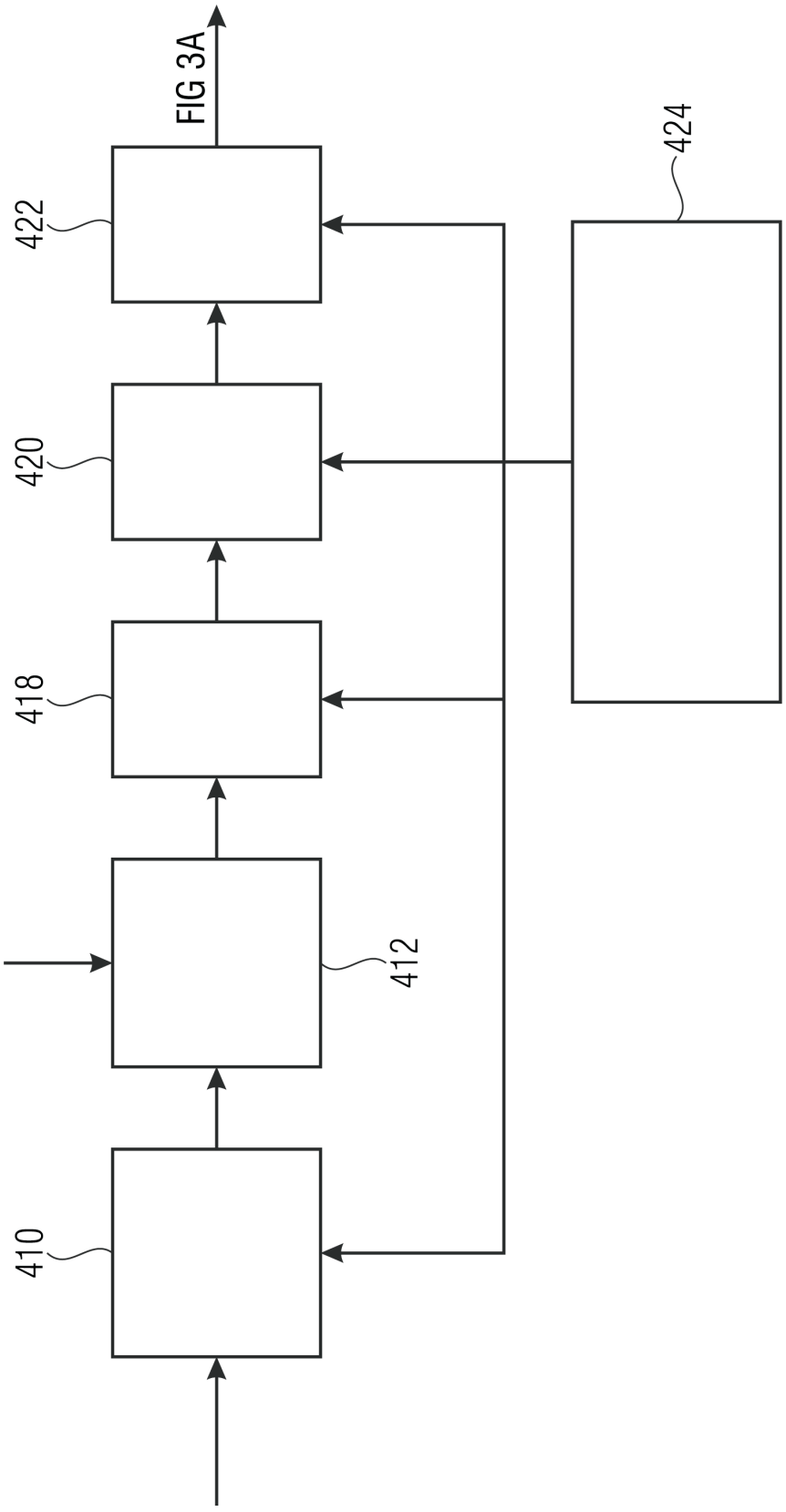


FIG 4B

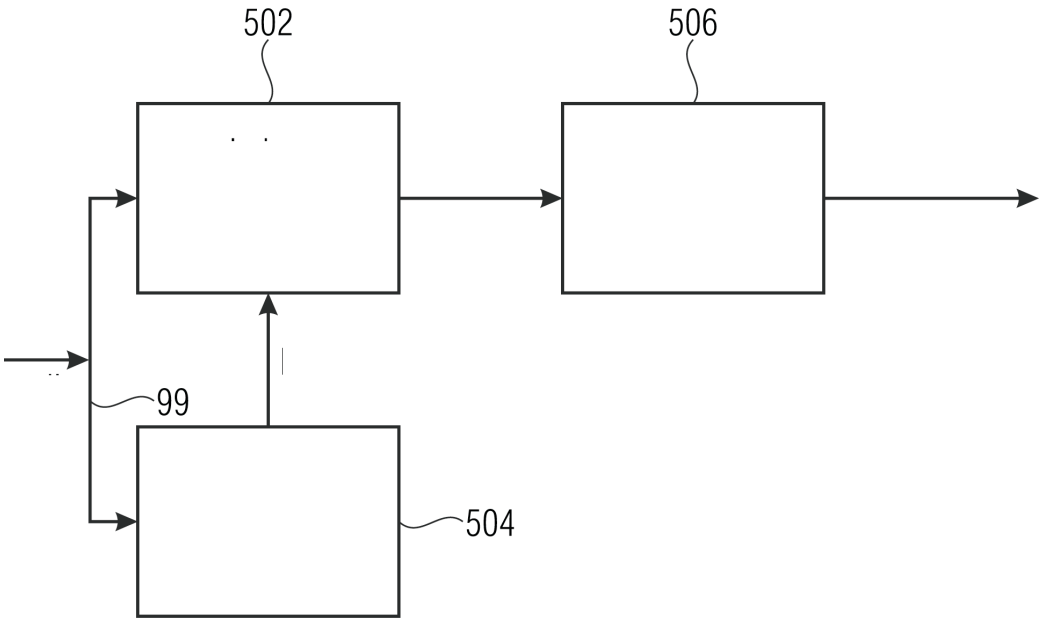


FIG 5A

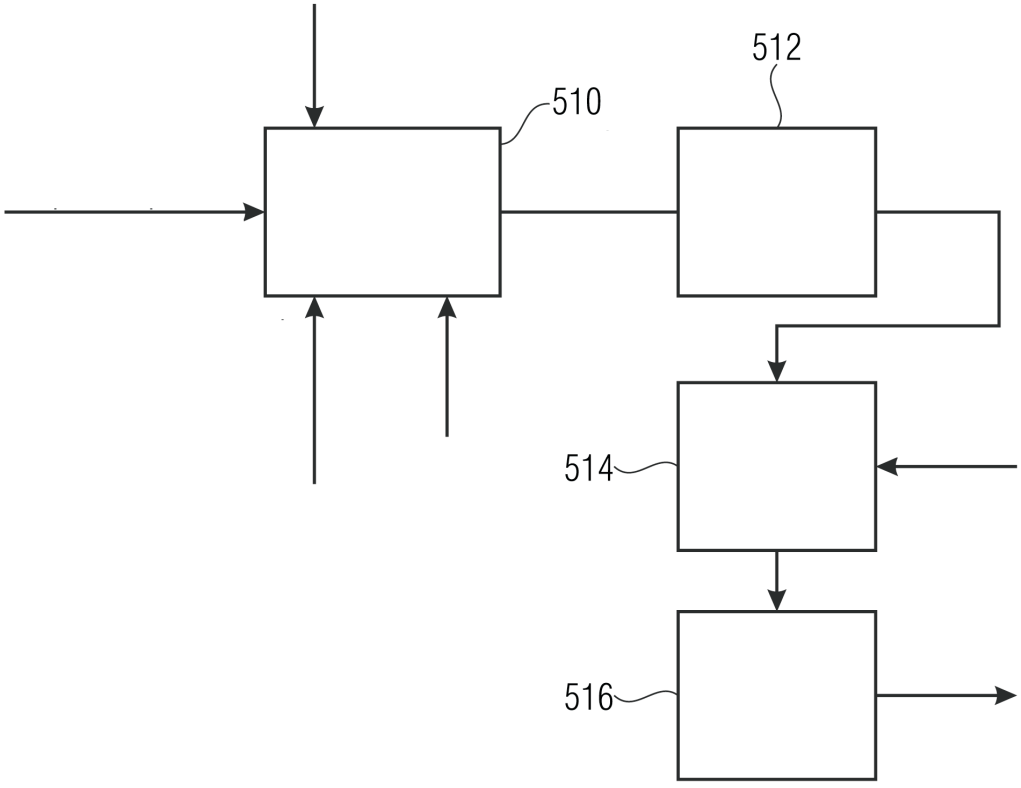


FIG 5B

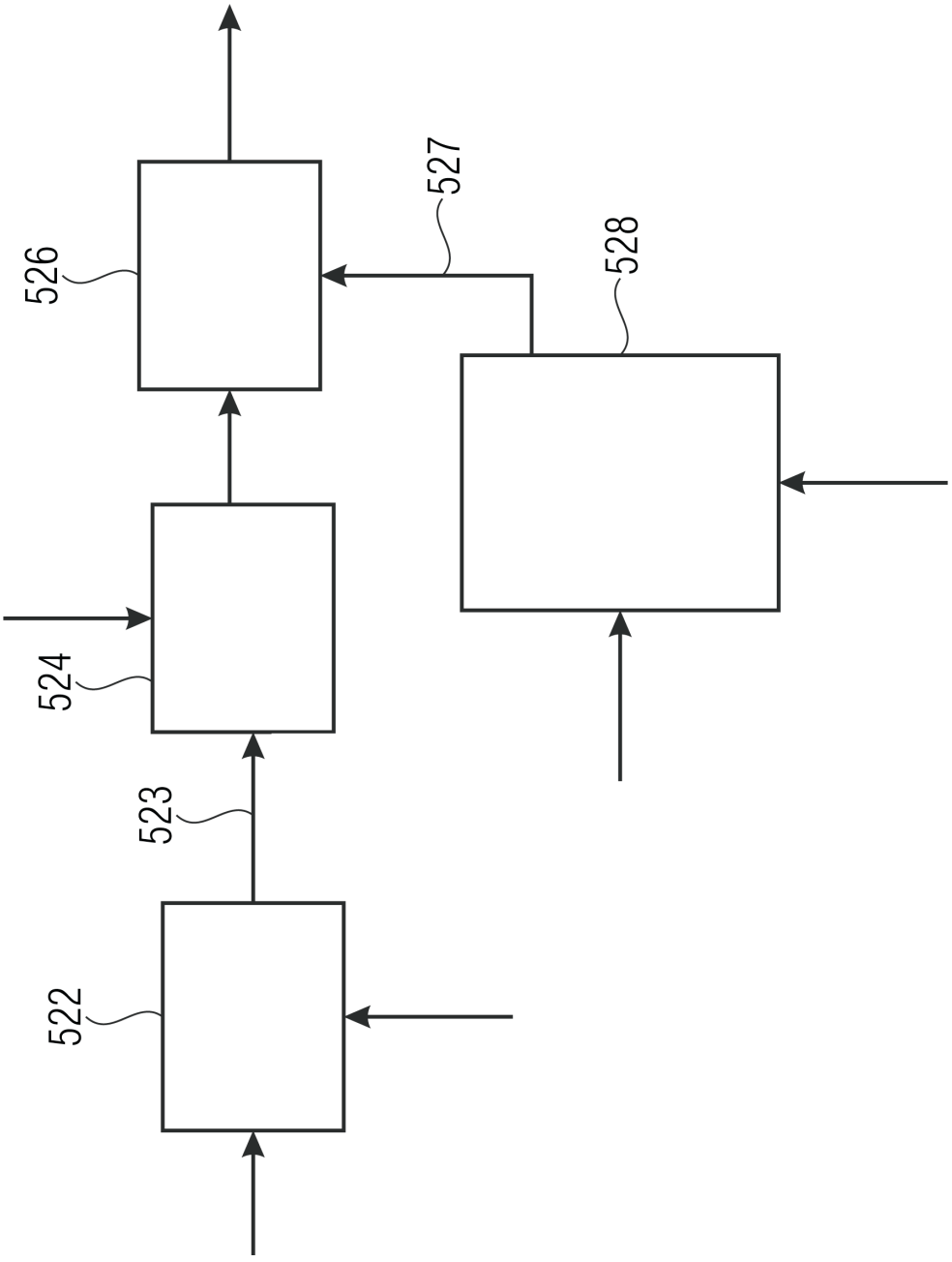


FIG 5C

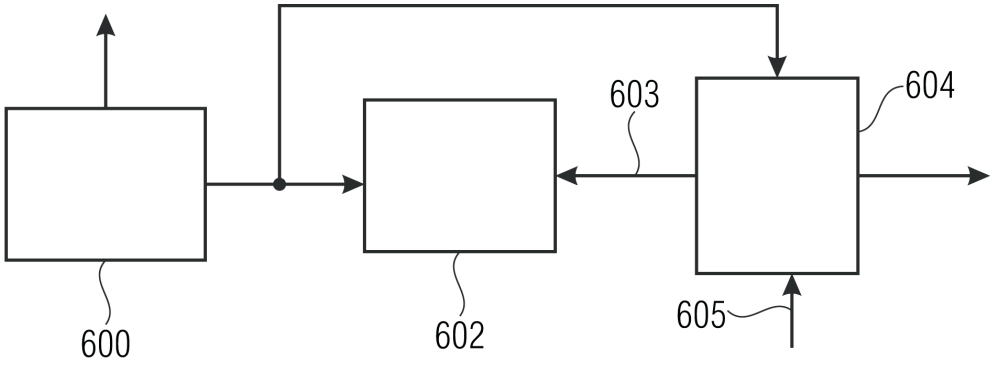


FIG 6A

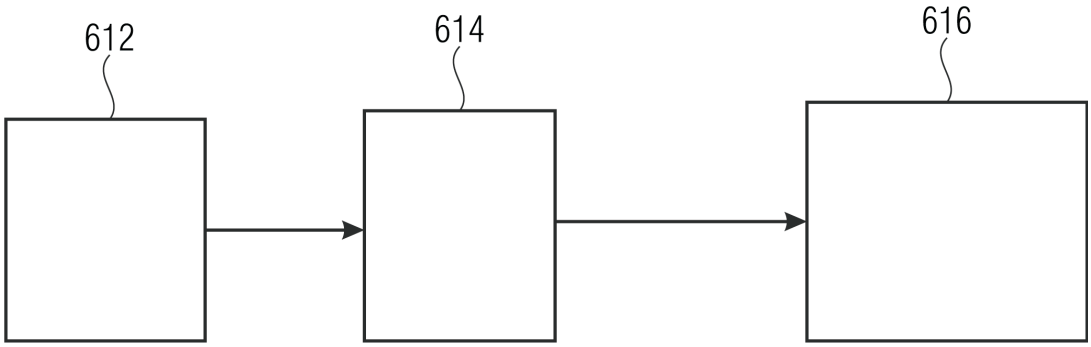


FIG 6B

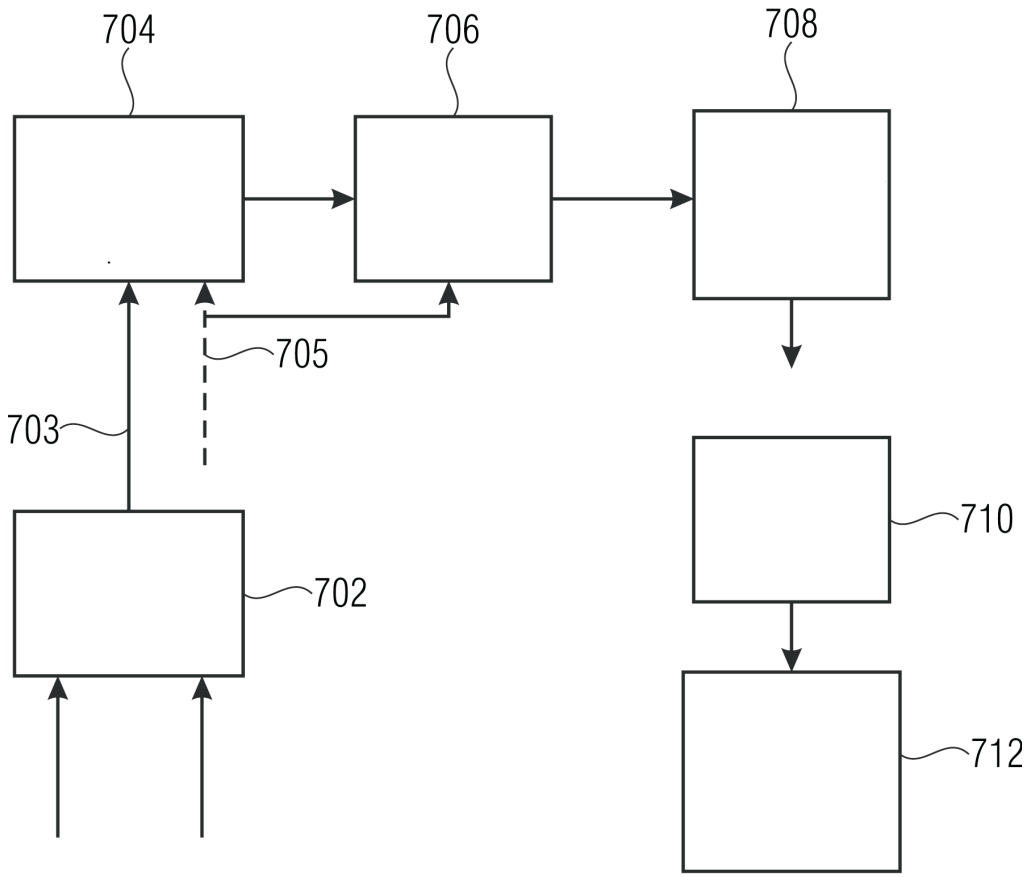


FIG 7A

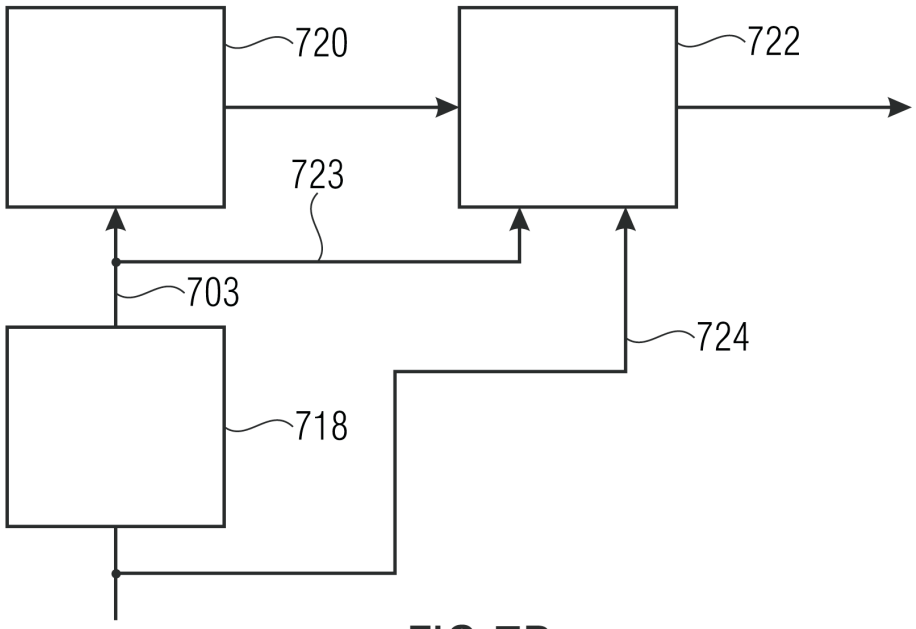


FIG 7B

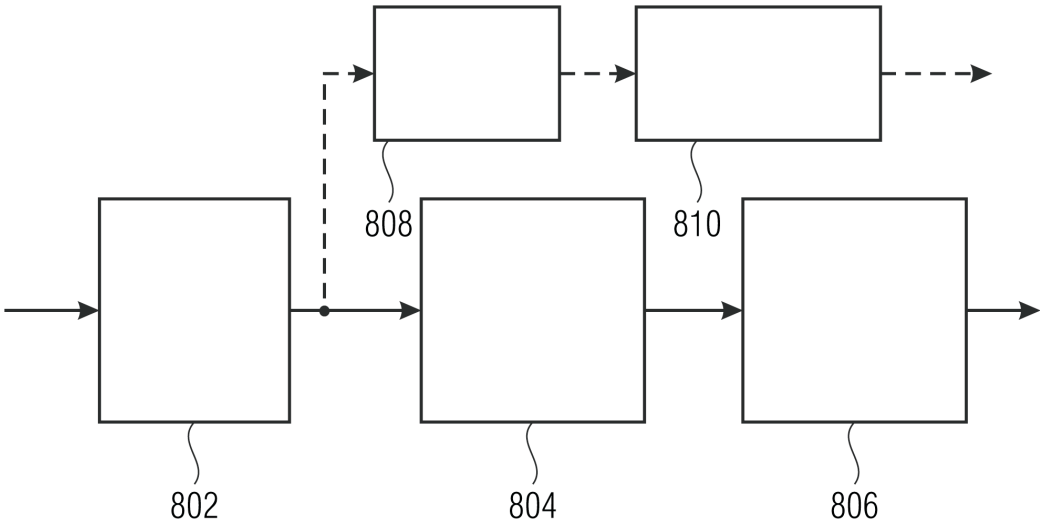


FIG 8A

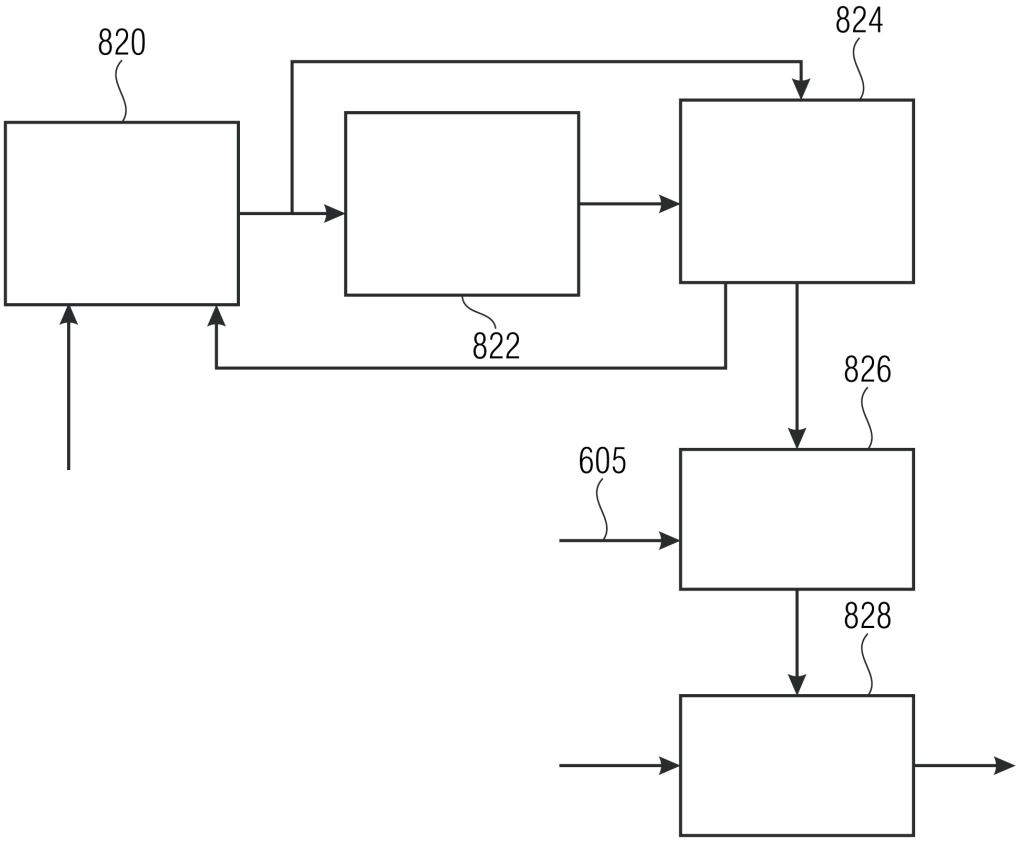


FIG 8B

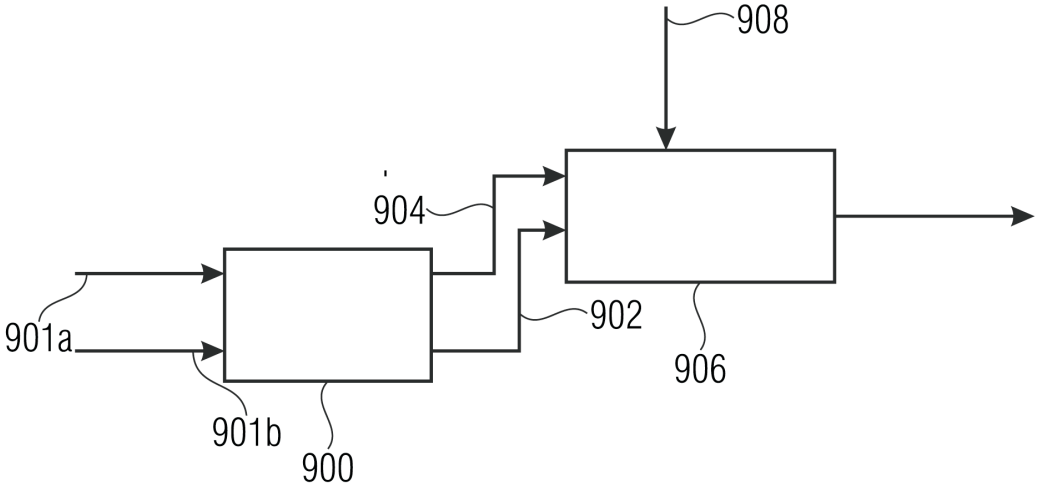


FIG 9A

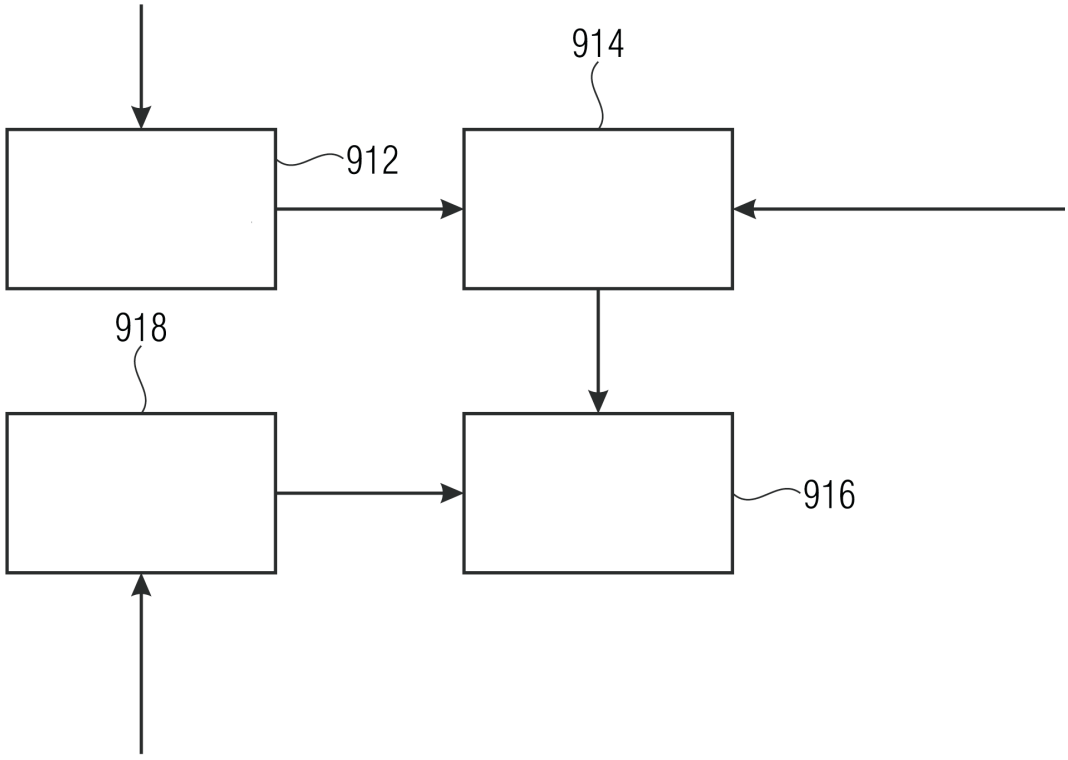


FIG 9B

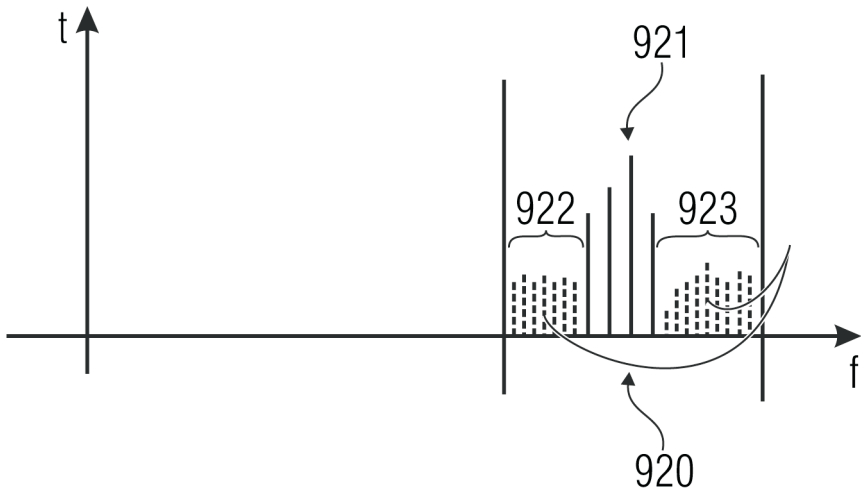


FIG 9C

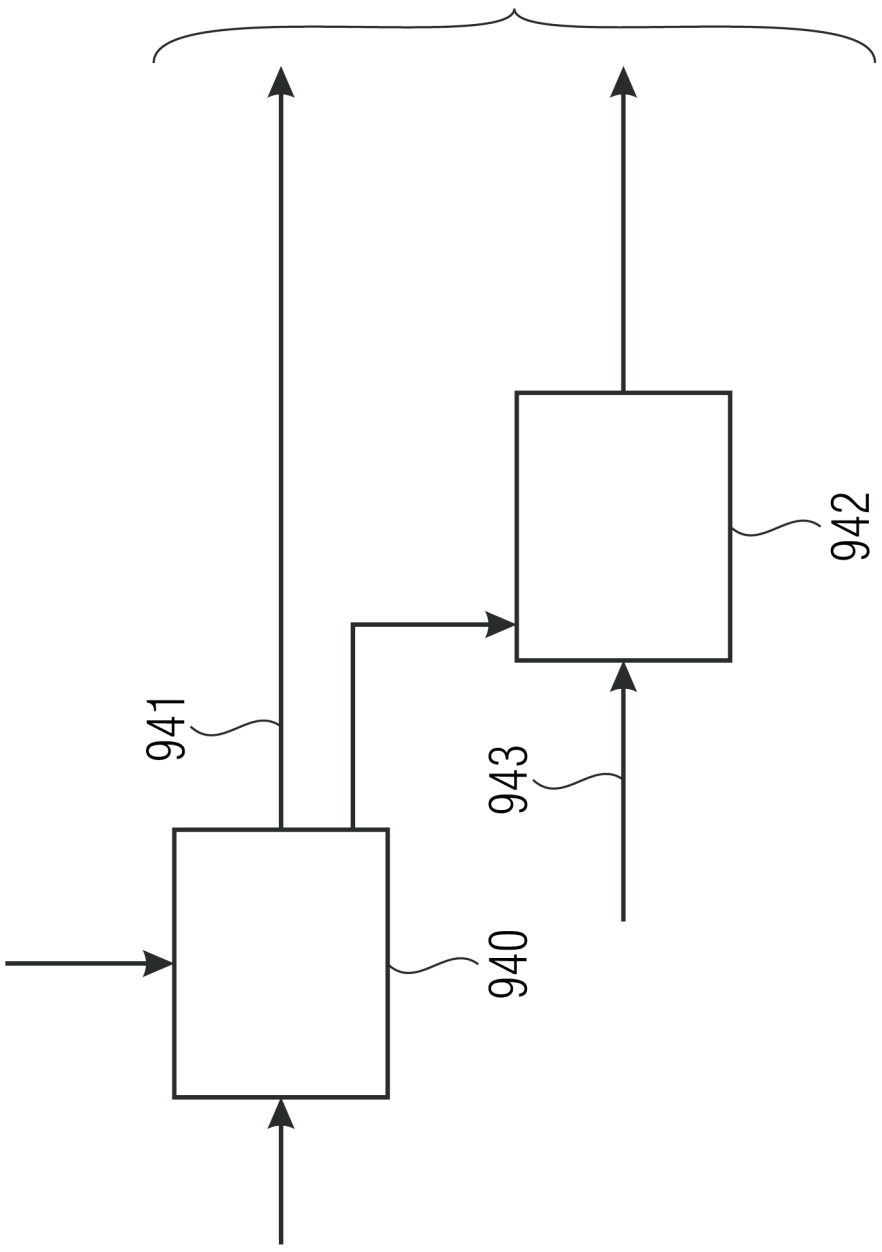


FIG 9D

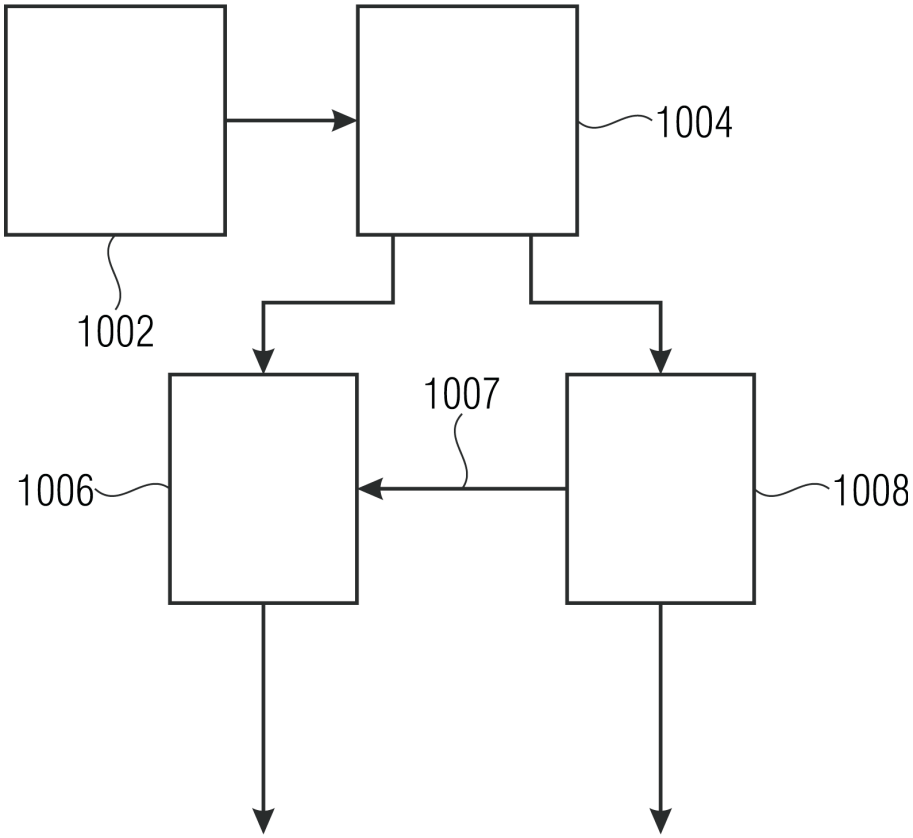


FIG 10A

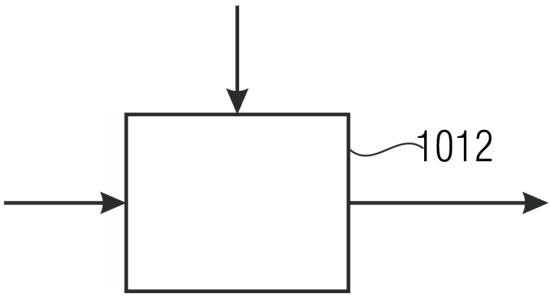


FIG 10B

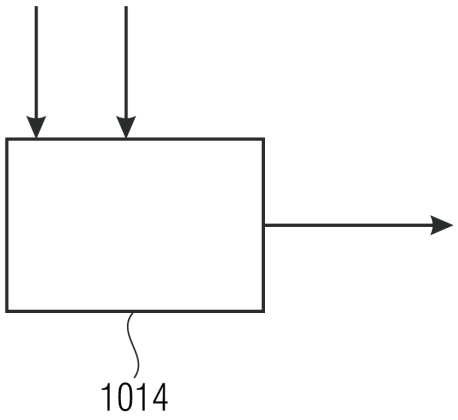


FIG 10C

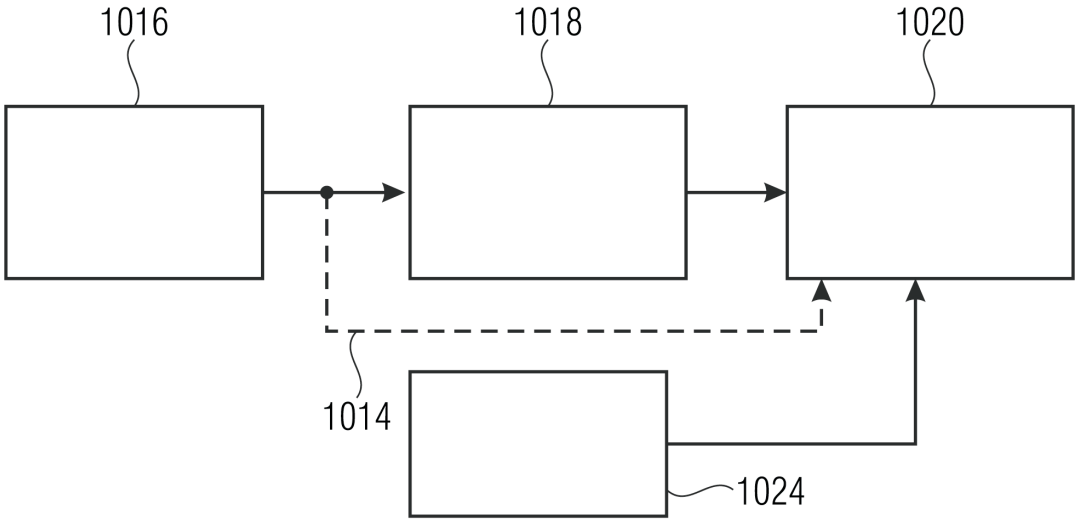


FIG 10D

1100

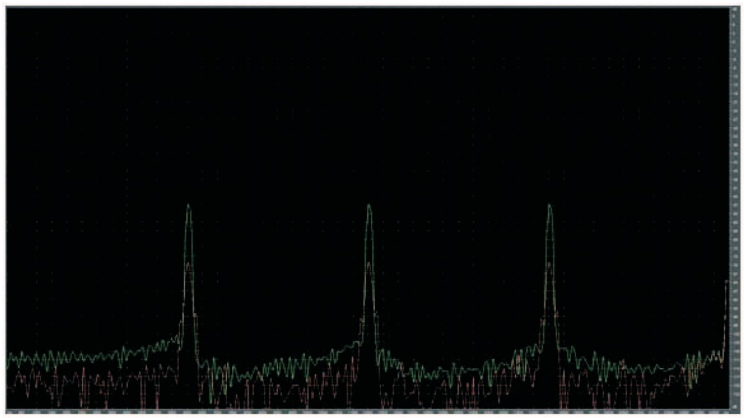


FIG 11A

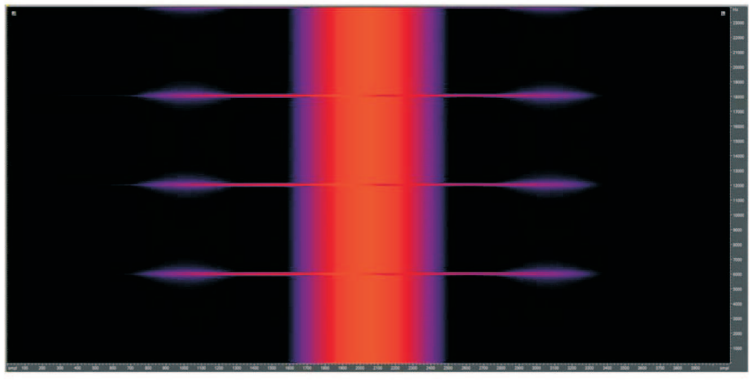


FIG 11B

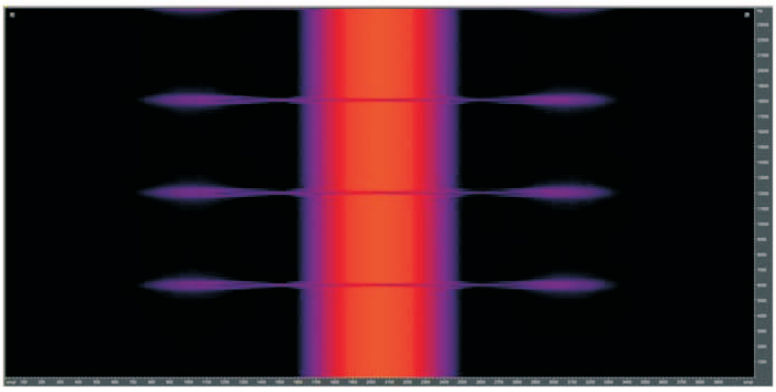


FIG 11C

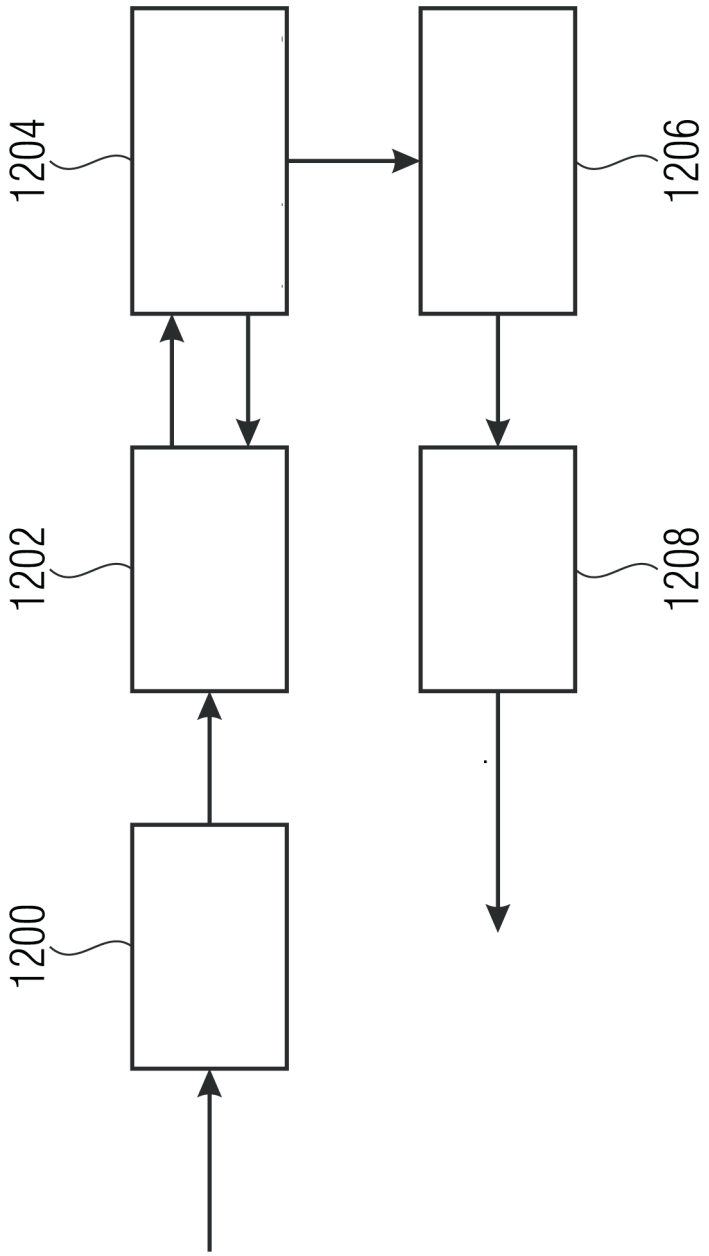


FIG 12A

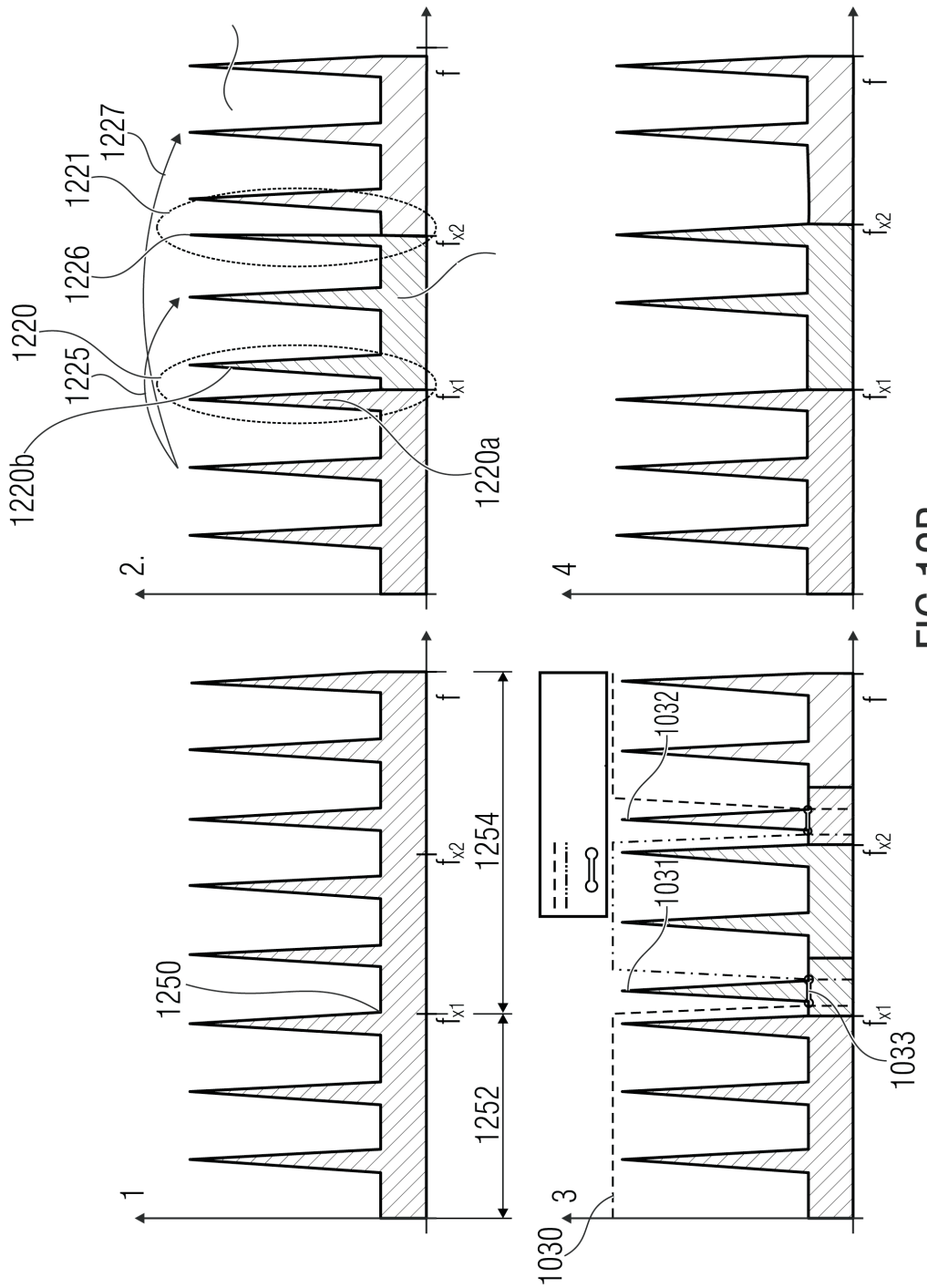

1
2.
3
4
f
f_{x1}
f_{x2}
1250
1252
1254
1220
1220a
1220b
1221
1225
1226
1227
1030
1031
1032
1033

FIG 12B

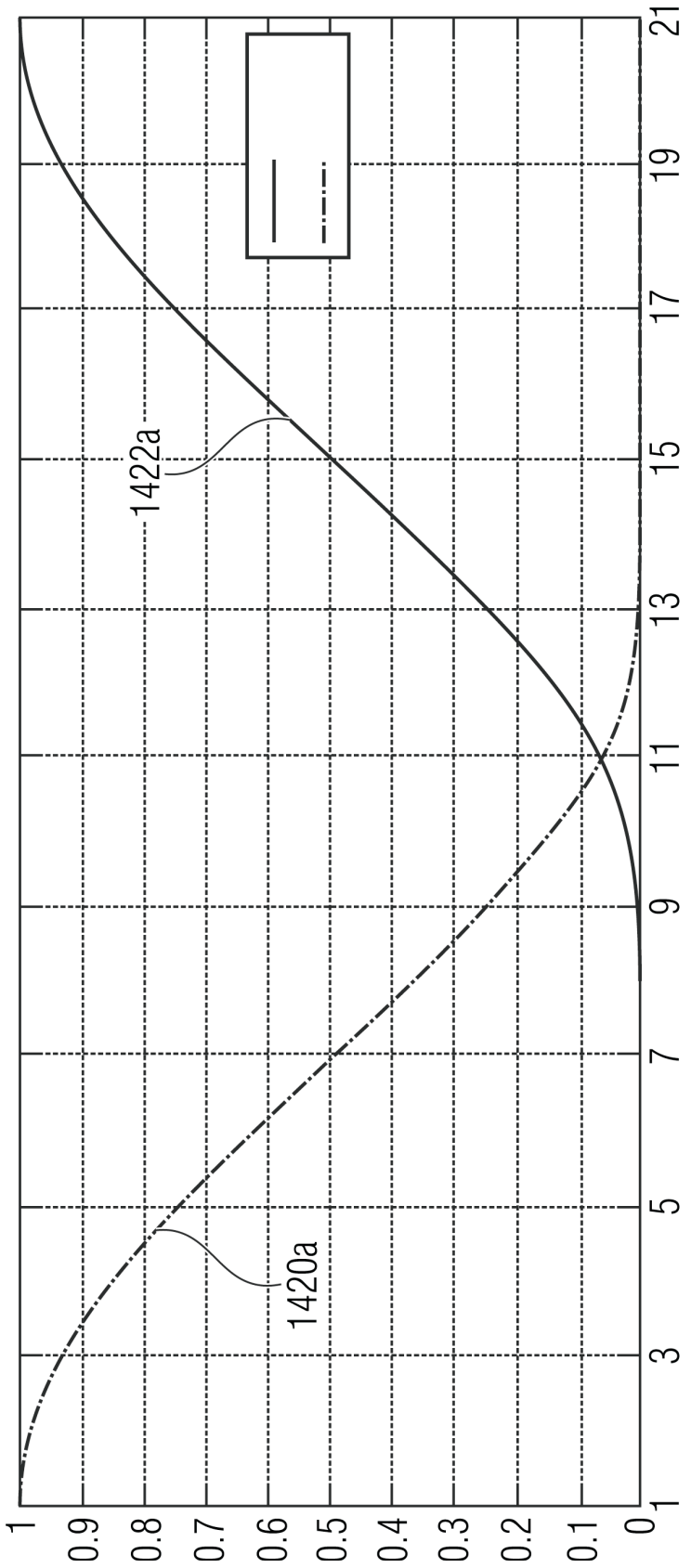


FIG 12C

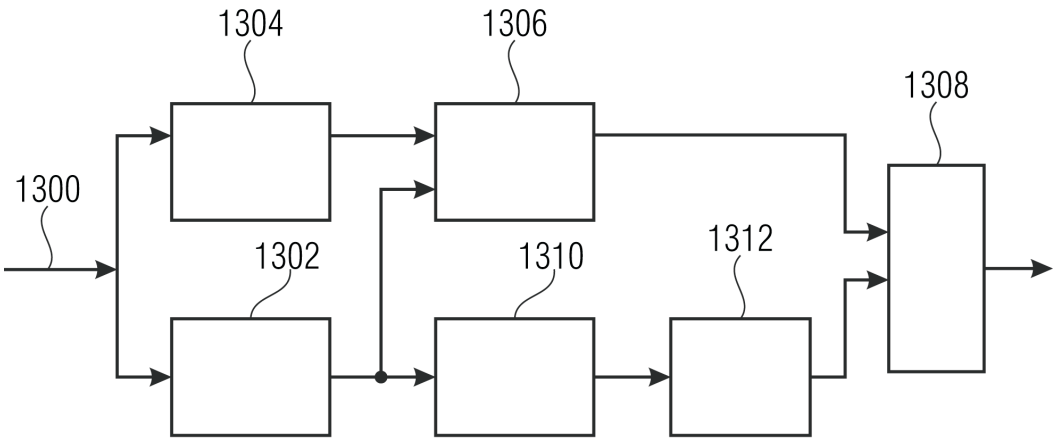


FIG 13A

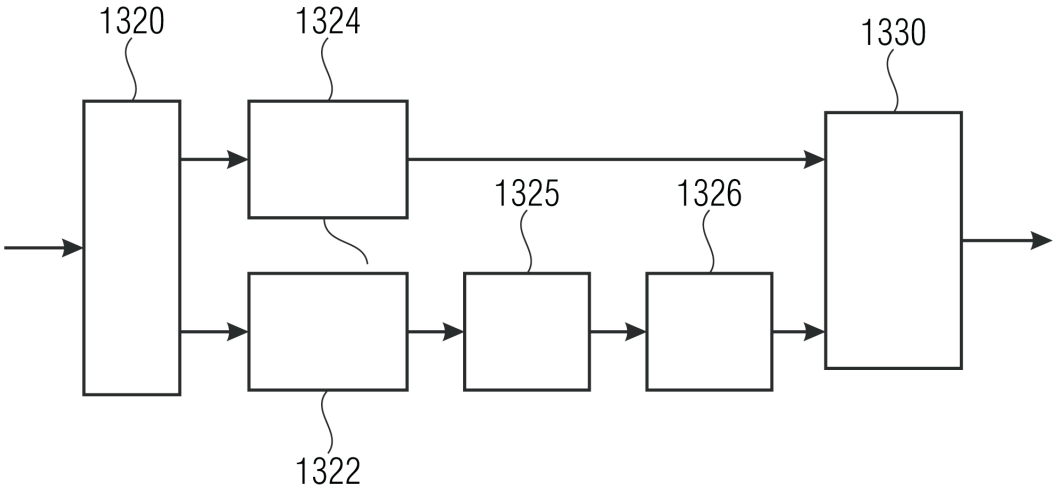


FIG 13B

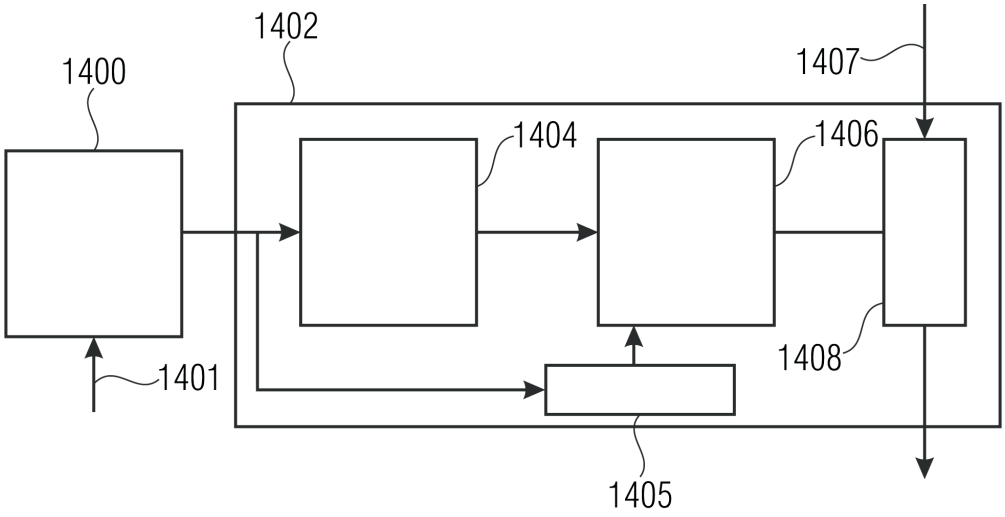


FIG 14A

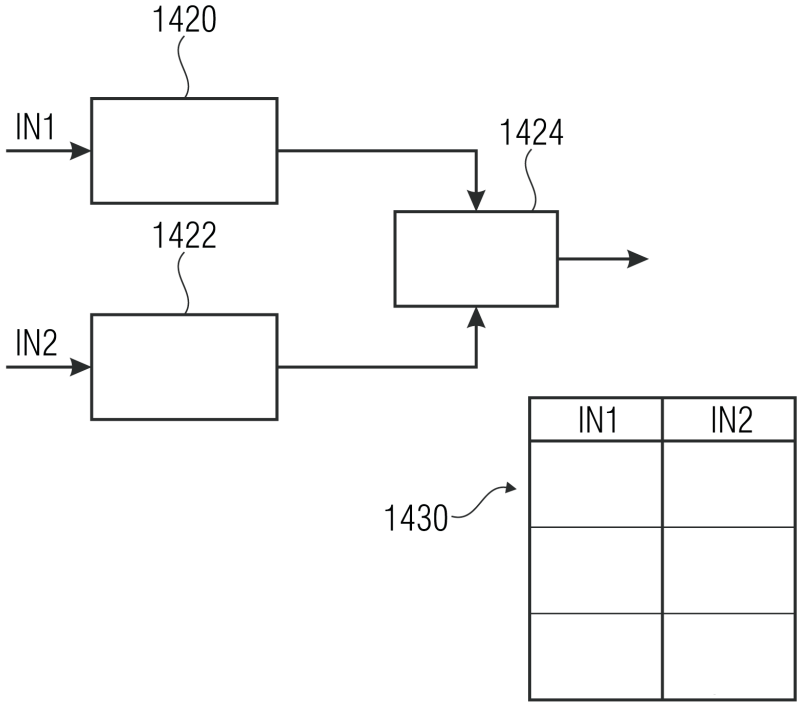


FIG 14B