



República Federativa do Brasil

Ministério do Desenvolvimento, Indústria,
Comércio e Serviços

Instituto Nacional da Propriedade Industrial

(11) BR 122022010965-9 B1

(22) Data do Depósito: 15/07/2014

(45) Data de Concessão: 04/04/2023

(54) Título: APARELHO E MÉTODO PARA CODIFICAR E DECODIFICAR UM SINAL DE ÁUDIO CODIFICADO UTILIZANDO MODULAÇÃO DE CORREÇÃO/RUÍDO TEMPORAL

(51) Int.Cl.: G10L 19/03; G10L 21/0388; G10L 19/02; G10L 19/028.

(30) Prioridade Unionista: 18/10/2013 EP 13189358.8; 22/07/2013 EP 13177353.3; 22/07/2013 EP 13177348.3; 22/07/2013 EP 13177346.7; 22/07/2013 EP 13177350.9.

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

(72) Inventor(es): SASCHA DISCH; FREDERIK NAGEL; RALF GEIGER; BALAJI NAGENDRAN THOSKHAHA; KONSTANTIN SCHMIDT; STEFAN BAYER; CHRISTIAN NEUKAM; BERND EDLER; CHRISTIAN HELMRICH.

(86) Pedido PCT: PCT EP2014065123 de 15/07/2014

(87) Publicação PCT: WO 2015/010954 de 29/01/2015

(85) Data do Início da Fase Nacional: 03/06/2022

(62) Pedido Original do Dividido: BR112015007533-9 - 15/07/2014

(57) Resumo: Um aparelho para decodificar um sinal de áudio codificado compreende: um decodificador de áudio de domínio espectral (602) para gerar uma primeira representação decodificada de um primeiro conjunto de partes espectrais primárias sendo valores residuais de previsão espectrais; um regenerador de frequência (604) para gerar uma segunda parte espectral reconstruída, utilizando uma primeira parte espectral do primeiro conjunto de partes espectrais primárias, caracterizado pela segunda parte espectral reconstruída adicionalmente compreender valores residuais de previsão espectrais; e um filtro de previsão inversa (606) para realizar uma previsão inversa sobre a frequência, utilizando os valores residuais espectrais para o primeiro conjunto de partes espectrais primárias e a segunda parte espectral reconstruída utilizando a informação do filtro de previsão (607) incluída no sinal de áudio codificado.

**"APARELHO E MÉTODO PARA CODIFICAR E
DECODIFICAR UM SINAL DE ÁUDIO CODIFICADO UTILIZANDO MODULAÇÃO
DE CORREÇÃO/RUÍDO TEMPORAL".**

DIVIDIDO DO BR112015007533-9, depositado em 15/07/2014.

DESCRIÇÃO

[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 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.

[0004] 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.

[0005] 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].

[0006] 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.

[0007] 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.

[0008] 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 central 1312. O passa-baixa 1302 restringe a largura de banda a ser codificada em uma largura de banda significativamente menor do que ocorre no sinal de áudio de entrada original na linha 1300. Isso fornece um ganho de codificação significativo devido ao fato de que todas as funcionalidades que ocorrem no codificador central 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.

[0009] 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 central 1322 e um sinal de entrada para um decodificador de parâmetro 1324. Um sinal de saída do decodificador central 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 central. 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.

[00010] Além disso, embora os típicos codificadores de áudio centrais operem no domínio espectral, o decodificador central 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 (por exemplo, transformação do domínio de MDCT em QMF). Isso leva a complicações de sincronização, complexidade

computacional adicional e exigências de memória elevada.

[00013] Particularmente, se um sistema de extensão da largura de banda é implementado em um banco de filtro ou domínio de transformada de tempo-frequência, há apenas uma possibilidade limitada para controlar a forma temporal do sinal da extensão da largura de banda. Tipicamente, a granularidade temporal é limitada pelo tamanho do salto utilizado entre as janelas de transformada adjacentes. Isso pode levar aos pré- ou pós-ecos indesejados na faixa espectral da extensão da largura de banda. A fim de aumentar a granularidade temporal, tamanhos de salto menores ou estruturas mais curtas da extensão da largura de banda podem ser utilizados, mas isso resulta em uma suspensão da taxa de bit devido ao fato de que, por um determinado período de tempo, um número mais alto de parâmetros, tipicamente um certo conjunto de parâmetros para cada período de tempo deve ser transmitido. Caso contrário, se os períodos de tempo individuais se tornam muito grandes, então o pré- e pós-ecos particularmente para as partes transientes de um sinal de áudio são gerados.

[00014] É um objeto da presente invenção fornecer um conceito de codificação/decodificação melhorado.

[00015] Este objeto é obtido por um aparelho para decodificar um sinal de áudio codificado de acordo com a reivindicação 1, um aparelho para codificar um sinal de áudio de acordo com a reivindicação 10, um método para decodificação de acordo com a reivindicação 16, um método para codificação de acordo com a reivindicação 18 ou um

programa de computador de acordo com a reivindicação 19.

[00016] A presente invenção se baseia na constatação que uma qualidade melhorada e taxa de bit reduzida especificamente para sinais compreendendo partes transientes conforme elas ocorrem muito frequente nos sinais de áudio é obtida pela combinação da tecnologia do Modelador do Ruído Temporal (TNS | *Temporal Noise Shaping*) ou do Modelador da Porção Temporal (TTS | *Temporal Tile Shaping*) com reconstrução de alta frequência. O processamento de TNS/TTS no lado do codificador sendo implementado por uma previsão sobre a frequência reconstrói o envelope de tempo do sinal de áudio. Dependendo da implementação, ou seja, quando o filtro do modulador de ruído temporal é determinado dentro de uma faixa de frequência não abrangendo apenas a faixa de frequência alvo, mas também a faixa de frequência alvo a ser reconstruída em um decodificador de regeneração de frequência, o envelope temporal não é apenas aplicado no sinal de áudio central até uma frequência inicial de adição de lacuna, mas o envelope temporal é ainda aplicado nas faixas espectrais das partes espectrais secundárias reconstruídas. Assim, pré-ecos ou pós-ecos que ocorreriam sem modelador de porção temporal são reduzidos ou eliminados. Isso é realizado aplicando uma previsão inversa sobre a frequência não apenas dentro da faixa de frequência central até uma certa frequência inicial de preenchimento da lacuna, mas também dentro de uma faixa de frequência acima da faixa de frequência central. Para esta finalidade, a regeneração de frequência ou a geração de porção de frequência é realizada

no lado do decodificador antes de aplicar uma previsão sobre a frequência. Entretanto, a previsão sobre a frequência pode tanto ser aplicada antes ou subsequentemente à formação do envelope espectral, dependendo se o cálculo da informação de energia foi realizado nos valores residuais espectrais subsequentes à filtragem ou aos valores espectrais (completos) antes do modelador do envelope.

[00017] O processamento TTS sobre uma ou mais porções de frequência adicionalmente estabelece uma continuidade de correlação entre a faixa de origem e a faixa de reconstrução ou em duas faixas de reconstrução adjacentes ou porções de frequência.

[00018] Em uma implementação, é preferido utilizar a filtragem complexa de TNS/TTS. Assim, as perturbações de distorção (temporal) de uma representação real criticamente amostrada, como MDCT, são evitadas. Um filtro TNS complexo pode ser calculado no lado do codificador aplicando não apenas uma transformada de cosseno discreta modificada, mas também uma transformada de seno discreta modificada além de obter uma transformada modificada complexa. Independente, apenas os valores da transformada de cosseno discreta modificada, ou seja, a parte real da transformada complexa é transmitida. No lado do decodificador, entretanto, é possível estimar a parte imaginária da transformada utilizando espectros de MDCT das estruturas precedentes e subsequentes de modo que, no lado do decodificador, o filtro complexo pode ser novamente aplicado na previsão inversa sobre a frequência e, especificamente, a previsão sobre o limite entre a faixa

de origem e a faixa de reconstrução e ainda sobre o limite entre as porções de frequência adjacentes da frequência dentro da faixa de reconstrução.

[00019] Outro aspecto se baseia na constatação que os problemas relacionados à separação da extensão da largura de banda, por um lado, e a codificação central, por outro lado, podem ser direcionados e superados pela realização da extensão da largura de banda no mesmo domínio espectral no qual o decodificador central opera. Assim, um decodificador central da taxa completa é fornecido codificando e decodificando a faixa do sinal de áudio completa. Isso não exige o uso de um redutor da taxa de amostragem no lado do codificador e um amplificador da taxa de amostragem no lado do decodificador. Ao invés disso, todo o processamento é realizado na taxa de amostragem total ou domínio da largura de banda total. A fim de obter um alto ganho de codificação, o sinal de áudio é analisado a fim de encontrar um primeiro conjunto de partes espectrais primárias que deve ser codificado com uma alta resolução, onde este primeiro conjunto de partes espectrais primárias pode incluir, em uma aplicação, as partes tonais do sinal de áudio. Por outro lado, os componentes não-tonais ou com ruído no sinal de áudio que constituem um segundo conjunto de partes espectrais secundárias são parametricamente codificadas com baixa resolução espectral. O sinal de áudio codificado então exige apenas o primeiro conjunto de partes espectrais primárias codificado em uma forma que preserva a forma de onda com uma alta resolução espectral e, adicionalmente, o segundo conjunto de partes

espectrais secundárias codificado parametricamente com uma alta resolução utilizando “porções” de frequência derivadas do primeiro conjunto. No lado do decodificador, o decodificador central, que é um decodificador de banda total, reconstrói o primeiro conjunto de partes espectrais primárias em uma forma que preserva a forma de onda, ou seja, sem qualquer conhecimento que haja qualquer regeneração de frequência adicional. Entretanto, o espectro então gerado tem muitas lacunas espectrais. Estas lacunas são subsequentemente preenchidas com a tecnologia inventiva do Preenchimento de Lacuna Inteligente (IGF), utilizando uma regeneração de frequência que aplica os dados paramétricos, por um lado, e utilizando uma faixa espectral de origem, ou seja, partes espectrais primárias reconstruídas pelo decodificador de áudio de taxa total, por outro lado.

[00020] Em outras aplicações, as partes espectrais, que são reconstruídas apenas pelo preenchimento de ruído ao invés da réplica da largura de banda ou preenchimento da porção de frequência, constituem um terceiro conjunto de partes espectrais terciárias. Devido ao fato de que o conceito de codificação opera em um único domínio para codificação/decodificação central, por um lado, e regeneração de frequência, por outro lado, o IGF não é apenas restrito para preencher uma faixa de frequência mais alta, mas pode preencher as faixas de frequência inferior, tanto pelo preenchimento de ruído sem a regeneração de frequência quanto pela regeneração de frequência utilizando uma porção de frequência em uma diferente faixa de frequência.

[00021] Além disso, é enfatizado que uma informação nas energias espectrais, uma informação sobre as energias individuais ou uma informação de energia individual, uma informação sobre uma energia sobrevivente ou uma informação de energia sobrevivente, uma informação sobre uma energia da porção ou uma informação da energia da porção ou uma informação sobre uma energia ausente ou uma informação da energia ausente pode compreender não apenas um valor de energia, mas ainda um valor de amplitude (por exemplo, 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.

[00022] Outro aspecto se baseia na constatação que a situação de correlação não é apenas importante para a faixa de origem, mas é também importante para a faixa alvo. Além disso, a presente invenção reconhece a situação que diferentes situações de correlação podem ocorrer na faixa de origem e na faixa alvo. Quando, por exemplo, um sinal de fala com ruído de alta frequência é considerado, a situação pode ser que a baixa frequência, compreendendo o sinal de fala com um pequeno número de implicações, é altamente correlacionada no canal esquerdo e no canal direito, quando o alto-falante é colocado no meio. A parte de alta frequência, entretanto, pode ser fortemente não correlacionada devido ao fato de que pode haver um diferente ruído de alta frequência no lado esquerdo comparado a outro ruído de alta frequência ou nenhum

ruído de alta frequência no lado direito. Assim, quando uma operação de adição de lacuna direta seria realizada ignorando esta situação, então a parte de alta frequência também estaria correlacionada, e isso pode gerar sérias perturbações de segregação espacial no sinal reconstruído. A fim de solucionar esta questão, os dados paramétricos para uma banda de reconstrução ou, geralmente, para o segundo conjunto de partes espectrais secundárias que deve ser reconstruído utilizando um primeiro conjunto de partes espectrais primárias é calculado para identificar tanto uma primeira quanto uma segunda representação diferente de dois canais para a segunda parte espectral ou, de forma diferente, para a banda de reconstrução. No lado do codificador, uma identificação de dois canais é, assim, calculada para as partes espectrais secundárias, ou seja, para as partes nas quais, adicionalmente, a informação de energia para as bandas de reconstrução é calculada. Um regenerador de frequência no lado do decodificador, então, regenera uma segunda parte espectral, dependendo de uma primeira parte do primeiro conjunto de partes espectrais primárias, ou seja, a faixa de origem e dados paramétricos para a segunda parte como informação de energia do envelope espectral ou quaisquer outros dados do envelope espectral e, adicionalmente, dependente da identificação de dois canais para a segunda parte, ou seja, para esta banda de reconstrução sob reconsideração.

[00023] A identificação de dois canais é preferivelmente transmitida como um identificador para cada

banda de reconstrução e estes dados são transmitidos de um codificador a um decodificador e o decodificador, então, decodifica o sinal central conforme indicado pelos identificadores preferivelmente calculados para as bandas espectrais. Então, em uma implementação, o sinal central é armazenado em ambas as representações estéreo (por exemplo esquerdo/direito e meio/lateral) e, para o preenchimento da porção de frequência IGF, a representação da porção de origem é escolhida para ajustar a representação da porção alvo conforme indicado pelos identificadores de identificação de dois canais para o Preenchimento de Lacuna Inteligente ou bandas de reconstrução, ou seja, para a faixa alvo.

[00024] É enfatizado que este procedimento não funciona apenas para sinais estéreos, ou seja, para um canal esquerdo e o canal direito, mas também opera para sinais multicanais. No caso de sinais multicanais, vários pares de diferentes canais podem ser processados no modo como um canal esquerdo e um direito como um primeiro par, um canal surround esquerdo e um surround direito como o segundo par e um canal central e um canal LFE como o terceiro par. Outros pareamentos podem ser determinados para formatos do canal de saída mais alta como 7.1, 11.1 e assim por diante.

[00025] Outro aspecto se baseia na constatação que certas imparidades na qualidade de áudio podem ser solucionadas aplicando um esquema de preenchimento da porção de frequência adaptativo do sinal. Para esta finalidade, uma análise no lado do codificador é realizada a fim de encontrar a melhor região de origem correspondente candidata para uma

certa região alvo. Uma informação correspondente que identifica para uma região alvo uma certa região de origem junta com opcionalmente alguma informação adicional é gerada e transmitida como informação adicional ao decodificador. O decodificador, então, aplica uma operação do preenchimento da porção de frequência utilizando a informação correspondente. Para esta finalidade, o decodificador lê a informação correspondente do fluxo de dados transmitido ou arquivo de dados ou acessa a região de origem identificada para uma determinada banda de reconstrução e, se indicado na informação correspondente, adicionalmente realiza algum processamento destes dados da região de origem para gerar dados espectrais não processados para a banda de reconstrução. Então, este resultado da operação do preenchimento da porção de frequência, ou seja, os dados espectrais não processados para a banda de reconstrução, são formados utilizando a informação do envelope espectral para finalmente obter uma banda de reconstrução que compreende também as partes espectrais primárias como partes tonais. Estas partes tonais, entretanto, não são geradas pelo esquema de adição da porção adaptativa, mas estas partes espectrais primárias são emitidas pelo decodificador de áudio ou decodificador central diretamente.

[00026] O esquema de seleção da porção espectral adaptativa pode operar com uma baixa granularidade. Nesta implementação, uma região de origem é subdividida em tipicamente regiões de origem sobrepostas e a região alvo ou as bandas de reconstrução são dadas pelas regiões alvo de

frequência não sobrepostas. Então, as similaridades entre cada região de origem e cada região alvo são determinadas no lado do codificador e o melhor par correspondente de uma região de origem e uma região alvo é identificado pela informação correspondente e, no lado do decodificador, a região de origem identificada na informação correspondente é utilizada para gerar os dados espectrais não processados para a banda de reconstrução.

[00027] Para a finalidade de obter uma granularidade mais alta, cada região de origem é permitida alternar a fim de obter um certo atraso onde as similaridades são máximas. Este atraso pode ser tão fino quanto uma porção de frequência e permite uma correspondência ainda melhor entre a região de origem e a região alvo.

[00028] Ademais, além de apenas identificar um melhor par correspondente, este atraso de correlação pode ainda ser transmitido dentro da informação correspondente e, adicionalmente, ainda um sinal pode ser transmitido. Quando o sinal é determinado como negativo no lado do codificador, então um identificador de sinal correspondente é ainda transmitido dentro da informação correspondente e, no lado do decodificador, os valores espectrais da região de origem são multiplicados por "-1" ou, em uma representação complexa, são "girados" a 180 graus.

[00029] Outra implementação desta invenção aplica uma operação de clareamento da porção. O clareamento de um espectro remove a informação bruta do envelope espectral e enfatiza a estrutura fina espectral que é de maior interesse

avaliar a similaridade da porção. Assim, uma porção de frequência, por um lado, e/ou o sinal de origem, por outro lado, são clareadas antes de calcular uma medição de correlação. Quando apenas a porção é clareada utilizando um procedimento predefinido, um identificador de clareamento é transmitido indicando ao decodificador que o mesmo processo de clareamento predefinido deve ser aplicado na porção de frequência dentro de IGF.

[00030] Com relação à seleção de porção, é preferido utilizar um atraso da correlação para de forma espectral alternar o espectro regenerado por um número inteiro de porções de transformada. Dependendo da transformada subjacente, a mudança espectral pode exigir correções adicionais. No caso de atrasos ímpares, a porção é adicionalmente modulada através da multiplicação por uma sequência temporal alternada de $-1/1$ para compensar a representação reservada por frequência de cada outra banda dentro da MDCT. Além disso, o sinal do resultado de correlação é aplicado ao gerar a porção de frequência.

[00031] Além disso, é preferido utilizar um corte e estabilização da porção a fim de certificar-se de que as perturbações criadas pela rápida mudança de regiões de origem para a mesma região de reconstrução ou região alvo são evitadas. Para esta finalidade, uma análise de similaridade entre as diferentes regiões de origem identificadas é realizada e quando uma porção de origem é semelhante a outras porções de origem com uma similaridade acima do limite, então esta porção de origem pode ser removida do conjunto de

porções de origem potenciais visto que é altamente correlacionada com outras porções de origem. Além disso, como um tipo de estabilização de seleção de porção, é preferido manter a ordem da porção da estrutura anterior se nenhuma das porções de origem na estrutura atual se correlacionarem (melhor do que um dado limite) com as porções alvo na estrutura atual.

[00032] O sistema de codificação de áudio 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.

[00033] As aplicações preferidas da presente invenção são subsequentemente descritas com relação aos desenhos anexos, nos quais:

[00034] A Figura 1a ilustra um aparelho para codificar um sinal de áudio;

[00035] A Figura 1b ilustra um decodificador para

decodificar um sinal de áudio codificado correspondente ao codificador da Figura 1a;

[00036] A Figura 2a ilustra uma implementação preferida do decodificador;

[00037] A Figura 2b ilustra uma implementação preferida do codificador;

[00038] A Figura 3a ilustra uma representação esquemática de um espectro conforme gerado pelo decodificador de domínio espectral da Figura 1b;

- 1ª resolução (alta resolução) para "envelope" do 1º conjunto (codificação de linha)

- 2ª resolução (baixa resolução) para "envelope" do 2º conjunto (fator de escala por SCB);

[00039] 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;

[00040] 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;

[00041] A Figura 4b ilustra uma implementação da funcionalidade da Figura 4a;

[00042] A Figura 5a ilustra uma funcionalidade de um codificador MDCT;

[00043] A Figura 5b ilustra uma funcionalidade do

decodificador com uma tecnologia MDCT;

[00044] A Figura 5c ilustra uma implementação do regenerador de frequência;

[00045] A Figura 6a ilustra um codificador de áudio com funcionalidade do modelador da porção temporal/modelador do ruído temporal;

[00046] A Figura 6b ilustra um decodificador com tecnologia do modelador da porção temporal/modelador do ruído temporal;

[00047] A Figura 6c ilustra outra funcionalidade da funcionalidade do modelador da porção temporal/modelador do ruído temporal com uma ordem diferente do filtro de previsão espectral e do modulador espectral;

[00048] A Figura 7a ilustra uma implementação da funcionalidade do modelador da porção temporal (*TTS | temporal tile shaping*);

[00049] A Figura 7b ilustra uma implementação do decodificador correspondente à implementação do codificador da Figura 7a;

[00050] A Figura 7c ilustra um espectrograma de um sinal original e 7um sinal estendido sem TTS;

[00051] A Figura 7d ilustra uma representação de frequência que ilustra a correspondência entre as frequências de preenchimento da lacuna inteligente e energias do modelador da porção temporal;

[00052] A Figura 7e ilustra um espectrograma de um sinal original e um sinal estendido com TTS;

[00053] A Figura 8a ilustra um decodificador de dois

canais com regeneração de frequência;

[00054] A Figura 8b ilustra uma tabela que ilustra diferentes combinações de representações e faixas de origem/de destino;

[00055] A Figura 8c ilustra o fluxograma que ilustra a funcionalidade do decodificador de dois canais com regeneração de frequência da Figura 8a;

[00056] A Figura 8d ilustra uma implementação mais detalhada do decodificador da Figura 8a;

[00057] A Figura 8e ilustra uma implementação de um codificador para o processamento de dois canais a ser decodificado pelo decodificador da Figura 8a;

[00058] A Figura 9a ilustra um decodificador com a tecnologia da regeneração de frequência utilizando valores de energia para a faixa de frequência da regeneração;

[00059] A Figura 9b ilustra uma implementação mais detalhada do regenerador de frequência da Figura 9a;

[00060] A Figura 9c ilustra um esquema que ilustra a funcionalidade da Figura 9b; por exemplo,

- energia sobrevivente: 5 unidades,
- valor da energia para banda de reconstrução: 10 unidades (abrange 1ª e 2ª partes espectrais na banda de reconstrução),

- energia de dados da faixa fonte ou dados da faixa alvo bruta: 8 unidades,

- energia ausente: 5 unidades,

- Fator de ganho: $g := \sqrt{\frac{mE_k}{pE_k}} = 0,79$

→ apenas valores espectrais para as 2ª partes espectrais são ajustados,

→ 1ª parte espectral não é influenciada pelo ajuste de envelope,

[00061] A Figura 9d ilustra outra implementação do decodificador da Figura 9a;

[00062] A Figura 10a ilustra um diagrama em blocos de um codificador correspondente com o decodificador da Figura 9a;

[00063] A Figura 10b ilustra um diagrama em blocos para ilustrar outra funcionalidade da calculadora de parâmetro da Figura 10a;

[00064] A Figura 10c ilustra um diagrama em blocos que ilustra outra funcionalidade da calculadora paramétrica da Figura 10a;

[00065] A Figura 10d ilustra um diagrama em blocos que ilustra outra funcionalidade da calculadora paramétrica da Figura 10a;

[00066] A Figura 11a ilustra outro decodificador tendo uma identificação da faixa de origem específica para a operação de adição da porção espectral no decodificador;

[00067] A Figura 11b ilustra outra funcionalidade do regenerador de frequência da Figura 11a;

[00068] A Figura 11c ilustra um codificador utilizado para cooperar com o decodificador na Figura 11a;

[00069] A Figura 11d ilustra um diagrama em blocos de uma implementação da calculadora de parâmetro da Figura 11c;

[00070] A Figura 12a e 12b ilustram os rascunhos da

frequência para ilustrar uma faixa de origem e uma faixa alvo;

[00071] A Figura 12c ilustra um gráfico de uma correlação exemplar de dois sinais;

[00072] A Figura 13a ilustra um codificador da técnica anterior com extensão da largura de banda; e

[00073] A Figura 13b ilustra um decodificador da técnica anterior com extensão da largura de banda.

[00074] 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.

[00075] 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 central é limitada por banda.

[00076] 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.

[00077] 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.

[00078] 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.

[00079] 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.

[00080] 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.

[00081] 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 central 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 rastreo 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.

[00082] 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.

[00083] 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 central 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 central, 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.

[00084] Além disso, uma técnica de estabilização de escolha da porção é proposta removendo as perturbações de domínio de frequência como vibração e ruído musical.

[00085] 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 central, incluindo um identificador que indica no caso de previsão se a direção da previsão é de *downmix* a residual ou vice versa.

[00086] As energias podem ser calculadas a partir das energias transmitidas no domínio L/R.

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

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

[00087] com k sendo o índice de frequência no domínio de transformada.

[00088] 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.

[00089] As porções de origem são sempre criadas de acordo com a Matriz-Meio/Lateral:

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

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

[00090] Ajuste de energia:

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

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

[00091] Estéreo conjunto -> transformação LR:

[00092] Se nenhum parâmetro de previsão adicional estiver codificado:

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

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

[00093] Se um parâmetro de previsão adicional estiver codificado e se a direção sinalizada for do meio à lateral:

$$sideTile[k] = sideTile[k] - predictionCoeff \cdot midTile[k]$$

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

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

[00094] Se a direção sinalizada for da lateral ao

meio:

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

$$leftTile[k] = midTile1[k] - sideTile[k]$$

$$rightTile[k] = midTile1[k] + sideTile[k]$$

[00095] 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.

[00096] 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 central é decodificado conforme indicado pelos conjuntos de identificador estéreo para as bandas espectrais. Segundo, o sinal central 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.

[00097] O Modelador de Ruído Temporal (TNS) 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.

[00098] 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.

[00099] 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) 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 central 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.

[000100] 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.

[000101] 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 central 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.

[000102] 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 central 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 central 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.

[000103] 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 central 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 central 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 central opera no intervalo completo, ou seja, como se o espectro fosse conforme ilustrado, ou seja, o decodificador central 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.

[000104] 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 central como um codificador central AAC ou USAC.

[000105] 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 central 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 central calcula um fator de escala para cada banda não apenas na faixa central 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.

[000106] Particularmente, quando o codificador central 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 .

[000107] 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.

[000108] 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.

[000109] 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.

[000110] 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.

[000111] 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.

[000112] O ajuste em 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.

[000113] 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.

[000114] Subsequentemente, a 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 finalmente 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.

[000115] 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.

[000116] 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.

[000117] 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 1 (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].

[000118] 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.

[000119] 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.

[000120] 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.

[000121] 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.

[000122] 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.

[000123] 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.

[000124] Conforme ilustrado, o codificador opera sem subamostragem e o decodificador opera sem sobreamostragem. 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.

[000125] 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.

[000126] 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.

[000127] 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).

[000128] 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.

O sistema geral consiste em:

- codificação central 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.

[000129] 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 central. Como a maioria dos

codecs de áudio, como AAC por exemplo, utilizam a MDCT como transformada 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.

[000130] A Figura 6a ilustra um aparelho para decodificar um sinal de áudio codificado em outra implementação da presente invenção. O aparelho para decodificar compreende um decodificador de áudio de domínio espectral 602 para gerar uma primeira representação decodificada de um primeiro conjunto de partes espectrais e como o regenerador de frequência 604 conectado a jusante do decodificador de áudio de domínio espectral 602 para gerar uma segunda parte espectral reconstruída utilizando uma primeira parte espectral do primeiro conjunto de partes espectrais primárias. Conforme ilustrado em 603, os valores espectrais na primeira parte espectral e na segunda parte espectral são valores residuais de previsão espectrais. A fim de transformar estes valores residuais de previsão espectrais em uma representação espectral completa, um filtro de previsão espectral 606 é fornecido. Este filtro de previsão inversa é configurado para realizar uma previsão inversa

sobre uma frequência utilizando os valores residuais espectrais para o primeiro conjunto de primeira frequência e as partes espectrais secundárias reconstruídas. O filtro de previsão espectral inversa 606 é configurado pela informação do filtro incluída no sinal de áudio codificado. A Figura 6b ilustra uma implementação mais detalhada da Figura 6a aplicação. Os valores residuais de previsão espectrais 603 são inseridos em um gerador da porção de frequência 612 gerando valores espectrais não processados para uma banda de reconstrução ou para uma certa segunda parte da frequência e estes dados não processados tendo agora a mesma resolução que a primeira representação espectral de alta resolução é inserida no modelador espectral 614. O modelador espectral agora modela o espectro utilizando a informação do envelope transmitida no fluxo contínuo de bits e os dados espectralmente modelados são então aplicados ao filtro de previsão espectral 616 finalmente gerando uma estrutura de valores espectrais completos utilizando a informação do filtro 607 transmitida do codificador ao decodificador através do fluxo contínuo de bits.

[000131] Na Figura 6b, assume-se que, no lado do codificador, o cálculo da informação do filtro transmitida através do fluxo contínuo de bits e utilizado através da linha 607 é realizado subsequente ao cálculo da informação do envelope. Assim, em outras palavras, um codificador correspondente com o decodificador da Figura 6b calcularia os valores residuais espectrais primeiro e calcularia, então, a informação do envelope com os valores residuais espectrais

como, por exemplo, ilustrado na Figura 7a. Entretanto, a outra implementação é útil para certas implementações também, onde a informação do envelope é calculada antes de realizar a filtragem de TNS ou TTS no lado do codificador. Então, o filtro de previsão espectral 622 é aplicado antes de realizar a modelação espectral no bloco 624. Assim, em outras palavras, os valores espectrais (completos) são gerados antes da operação da modelação espectral 624 ser aplicada.

[000132] Preferivelmente, um filtro TNS ou TTS com valor complexo é calculado. Isso é ilustrado na Figura 7a. O sinal de áudio original é inserido em um bloco MDCT complexo 702. Então, o cálculo do filtro TTS e do filtro TTS é realizado no domínio complexo. Então, no bloco 706, a informação adicional de IGF é calculada e qualquer outra operação como análise espectral para codificação, etc. é calculada também. Então, o primeiro conjunto de primeira parte espectral gerado pelo bloco 706 é codificado com um codificador acionado pelo modelo psicoacústico ilustrado em 708 para obter o primeiro conjunto de partes espectrais primárias indicada em $X(k)$ na Figura 7a e todos esses dados são encaminhados ao multiplexador do fluxo contínuo de bits 710.

[000133] No lado do decodificador, os dados codificados são inseridos em um demultiplexador 720 para separar a informação adicional de IGF, por um lado, a informação adicional de TTS, por outro lado, e a representação codificada do primeiro conjunto de partes espectrais primárias.

[000134] Então, o bloco 724 é utilizado para calcular um espectro complexo de um ou mais espectros com valor real. Então, tanto os espectros com valor real quanto os espectros complexos são inseridos no bloco 726 para gerar valores de frequência reconstruídos no segundo conjunto de partes espectrais secundárias para uma banda de reconstrução. Então, na estrutura de banda completa adicionada pela porção e obtida completamente, a operação de TTS inversa 728 é realizada e, no lado do decodificador, uma operação de MDCT complexa inversa final é realizada no bloco 730. Assim, a utilização da informação do filtro de TNS complexo permite, quando aplicado não apenas dentro da banda espectral ou dentro das bandas de porção separadas, mas quando aplicado sobre os limites do núcleo/porção ou os limites de porção/porção gera automaticamente um processamento de limite de porção, que, no final, reintroduz uma correlação espectral entre porções. Esta correlação espectral sobre os limites da porção não é obtida apenas pela geração de porções de frequência e realização de um ajuste do envelope espectral nestes dados não processados das porções de frequência.

[000135] A Figura 7c ilustra uma comparação de um sinal original (painel esquerdo) e um sinal estendido sem TTS. Pode ser visto que há perturbações fortes ilustradas pelas partes limitadas na faixa de frequência superior ilustrada em 750. Isso, entretanto, não ocorre na Figura 7e quando a mesma parte espectral em 750 é comparada com o componente relacionado à perturbação 750 da Figura 7c.

[000136] As aplicações ou o sistema de codificação de

áudio inventivo utilizam ao compartilhamento principal de taxa de bit disponível em forma de onda para codificar apenas a estrutura do sinal mais perceptualmente relevante no codificador, e as lacunas espectrais resultantes são adicionadas no decodificador com conteúdo do sinal que aproximadamente de aproxima do espectro original. Um montante de bit muito limitado é consumido para controlar o parâmetro acionado chamado Preenchimento de Lacuna Inteligente (IGF) espectral pela informação adicional dedicada transmitida do codificador ao decodificador.

[000137] O armazenamento ou transmissão dos sinais de áudio é geralmente submetido às restrições de taxa de bit rigorosas. No passado, os codificados foram forçados a drasticamente reduzir a largura de banda de áudio transmitida quando apenas uma taxa de bit muito baixa estava disponível. Os codecs de áudio modernos são atualmente capazes de codificar sinais de banda larga utilizando os métodos de extensão da largura de banda (BWE) como Réplica da Largura de Banda Espectral (SBR) [1]. Estes algoritmos dependem de uma representação paramétrica do conteúdo de alta frequência (HF) - que é gerado da parte de baixa frequência codificada da forma de onda (LF) do sinal decodificado por meios de transposição na região espectral HF ("correção") e aplicação de um pós-processamento acionado por parâmetro. Nos esquemas de BWE, a reconstrução da região espectral HF acima de uma dada 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 derivada

das regiões de passa-banda (BP) do espectro LF abaixo da dada frequência cruzada. Os sistemas do estado da técnica realizam eficientemente a correção dentro de uma representação do banco de filtro pela cópia de um conjunto de coeficientes de sub-banda adjacentes de uma região de origem para uma região alvo.

[000138] Se um sistema de BWE é implementado em um banco de filtro ou domínio de transformada de tempo-frequência, há apenas uma possibilidade limitada de controlar a modulação temporal do sinal da extensão da largura de banda. Tipicamente, a granularidade temporal é limitada pelo tamanho do salto utilizado entre as janelas da transformada adjacente. Isso pode levar ao pré- ou pós-ecos indesejados na faixa espectral de BWE.

[000139] A partir da codificação de áudio perceptual, sabe-se que a modulação do envelope temporal de um sinal de áudio pode ser rearmazenada utilizando as técnicas de filtragem espectral como Modelação do Envelope Temporal (TNS | *Temporal Envelope Shaping*) [14]. Entretanto, o filtro de TNS conhecido do estado da técnica é um filtro com valor real nos espectros com valor real. Este filtro com valor real nos espectros com valor real pode ser seriamente prejudicado pelas perturbações de distorção, especialmente se a transformada real subjacente for uma Transformada de cosseno discreta modificada (MDCT).

[000140] O modelador de porção do envelope temporal aplica a filtragem complexa nos espectros com valor complexo, como obtido de, por exemplo, uma Transformada de cosseno

discreta modificada Complexa (CMDCT | *Complex Modified Discrete Cosine Transform*). Assim, as perturbações de distorção são evitadas.

[000141] O modelador de porção temporal consiste em

- estimativa e aplicação de coeficiente de filtro complexo de um filtro de nivelamento no espectro original do sinal no codificador;
- transmissão dos coeficientes do filtro na informação adicional;
- aplicação de um filtro do modelador no espectro reconstruído adicionado pela porção no decodificador.

[000142] A invenção estende a técnica do estado da técnica conhecida da codificação da transformada de áudio, especificamente Modelador de Ruído Temporal (TNS) pela previsão linear ao longo da direção da frequência, para o uso em uma forma modificada no contexto da extensão da largura de banda.

[000143] Ainda, o algoritmo da extensão da largura de banda inventiva se baseia na Preenchimento de Lacuna Inteligente (IGF), mas emprega uma transformada com valor complexo (CMDCT) sobreamostrada, conforme oposto à configuração padrão de IGF que depende de uma representação de MDCT criticamente amostrada com valor real de um sinal. A CMDCT pode ser vista como a combinação dos coeficientes de MDCT na parte real e os coeficientes de MDST na parte imaginária de cada coeficiente do espectro com valor complexo.

[000144] Embora a nova abordagem seja descrita no contexto de IGF, o processamento inventivo pode ser utilizado em combinação com qualquer método de BWE que se baseia em uma representação do banco de filtro do sinal de áudio.

[000145] Neste novo contexto, a previsão linear com a direção da frequência não é utilizada como modelador de ruído temporal, mas como uma técnica do modelador de porção temporal (TTS). A renomeação é justificada pelo fato de que os componentes do sinal adicionados pela porção são temporariamente modelados por TTS conforme oposto à modelação do ruído de quantização por TNS nos codecs de transformada perceptual do estado da técnica.

[000146] A Figura 7a mostra um diagrama em blocos de um codificador de BWE utilizando IGF e a nova abordagem de TTS.

[000147] Assim, o esquema de codificação básica funciona, conforme segue:

- computar a CMDCT de um sinal de domínio de tempo $x(n)$ para atingir o sinal de domínio de frequência $X(k)$
- calcular o filtro de TTS com valor complexo;
- obter a informação adicional para BWE e remover a informação espectral que deve ser replicada pelo decodificador;
- aplicar a quantização, utilizando o módulo psicoacústico (PAM | *psycho acoustic module*);
- armazenar / transmitir os dados, apenas

os coeficientes de MDCT com valor real são transmitidos.

[000148] A Figura 7b mostra o decodificador correspondente. Ele inverte principalmente as etapas feitas no codificador.

[000149] Aqui, o esquema de decodificação básica funciona, conforme segue:

- estimar os coeficientes de MDST dos valores de MDCT (este processamento adiciona um atraso do bloco decodificador) e combinar a MDCT e coeficientes de MDST nos coeficientes de CMDCT com valor complexo;
- realizar o preenchimento de porção com seu pós-processamento;
- aplicar a filtragem de TTS inversa com os coeficientes do filtro de TTS transmitidos;
- calcular a CMDCT inversa.

[000150] Observe que, de modo alternativo, a ordem da síntese de TTS e pós-processamento de IGF podem ainda ser invertidos no decodificador se a análise de TTS e a estimativa do parâmetro de IGF são consistentemente invertidas no codificador.

[000151] Para a codificação da transformada eficiente, preferivelmente os chamados "blocos longos" de aproximadamente 20 ms devem ser utilizados para atingir o ganho de transformada razoável. Se o sinal dentro deste bloco longo contém transientes, pré- e pós-ecos audíveis ocorrem nas bandas espectrais reconstruídas devido à adição de porção. A Figura 7c mostra típicos efeitos de pré- e pós-eco que prejudicam os transientes devido à IGF. No painel

esquerdo da Figura 7c, o espectrograma do sinal original é mostrado, e no painel direito o espectrograma do sinal adicionado pela porção sem filtragem TTS inventiva é mostrado. Neste exemplo, a frequência inicial de IGF $f_{IGFstart}$ ou f_{split} entre a banda espectral e a banda adicionada por porção é escolhida ser $f_s/4$. No painel direito da Figura 7c, diferentes pré- e pós-ecos são visíveis ao redor dos transientes, especialmente proeminentes na extremidade espectral superior da região de frequência replicada.

[000152] A tarefa principal do módulo TTS é confinar estes componentes do sinal indesejado na proximidade ao redor de um transiente e assim ocultá-los na região temporal governada pelo efeito de mascaramento temporal da percepção humana. Assim, os coeficientes de previsão de TTS necessários são calculados e aplicados utilizando a "previsão de avanço" no domínio de CMDCT.

[000153] Em uma aplicação que combina TTS e IGF em um codec é importante alinhar certos parâmetros de TTS e parâmetros de IGF de modo que uma porção de IGF seja tanto completamente filtrada por um filtro TTS (filtro de nivelamento ou modelação) ou não. Assim, todas as frequências $TTSstart[...]$ ou $TTSstop[...]$ não devem ser compreendidas dentro de uma porção de IGF, mas ainda ser alinhadas às respectivas frequências $f_{IGF...}$. A Figura 7d mostra um exemplo de áreas operacionais de TTS e IGF para um conjunto de três filtros de TTS.

[000154] A frequência de parada de TTS é ajustada na frequência de parada da ferramenta de IGF, que é mais alta do

que $f_{IGFstart}$. If TTS utiliza mais do que um filtro, deve ser garantido que a frequência cruzada entre dois filtros TTS devem corresponder com a frequência de divisão de IGF. Caso contrário, um subfiltro de TTS passará por $f_{IGFstart}$ resultando nas perturbações indesejadas como sobremodelação.

[000155] Na variante da implementação descrita na Figura 7a e Figura 7b, cuidado adicional deve ser tomado em que as energias do decodificador IGF são ajustadas corretamente. Este é especialmente o caso se, no curso de processamento de TTS e IGF, diferentes filtros TTS tendo diferentes ganhos de previsão são aplicados na região de origem (como um filtro de nivelamento) e a região espectral alvo (como um filtro do modelador que não é a contraparte exata do dito filtro de nivelamento) de uma porção de IGF. Neste caso, a razão do ganho de previsão dos dois filtros TTS aplicados não iguala e assim, um ajuste de energia por esta razão deve ser aplicado.

[000156] Na variante da implementação alternativa, a ordem de pós-processamento de IGF e TTS é invertida. No decodificador, isso significa que o ajuste de energia pelo pós-processamento de IGF é calculado subsequente à filtragem de TTS e assim é a etapa de processamento final antes da transformada de síntese. Assim, independente de diferentes ganhos de filtro TTS sendo aplicados a uma porção durante a codificação, a energia final é sempre ajustada corretamente pelo processamento de IGF.

[000157] No lado do decodificador, os coeficientes do

filtro TTS são aplicados no espectro completo novamente, ou seja, o espectro central estendido pelo espectro regenerado. A aplicação de TTS é necessário para formar o envelope temporal do espectro regenerado para combinar o envelope do sinal original novamente. Assim, os pré-ecos mostrados são reduzidos. Além disso, ele ainda modela temporalmente o ruído de quantização no sinal abaixo $f_{IGFstart}$ como com TNS legado.

[000158] Nos codificadores de legado, a correção espectral em um sinal de áudio (por exemplo, SBR) corrompe a correlação espectral nos limites de correção e, assim, prejudica 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 do modelador TTS, os limites da porção são continuamente correlacionados, resultando em uma reprodução temporal mais fiel do sinal.

[000159] O resultado do sinal corretamente processado é mostrado na Figura 7e. Em comparação à versão ultrafiltrada (Figura 7c, painel direito) o sinal filtrado de TTS mostra uma boa redução dos pré- e pós-ecos indesejados (Figura 7e, painel direito).

[000160] Além disso, conforme discutido, a Figura 7a ilustra um codificador correspondente com o decodificador da Figura 7b ou o decodificador da Figura 6a. Basicamente, um aparelho para codificar um sinal de áudio compreende um conversor de espectro de tempo como 702 para converter um sinal de áudio em uma representação espectral. A representação espectral pode ser uma representação espectral

com valor real ou, conforme ilustrado no bloco 702, uma representação espectral com valor complexo. Além disso, um filtro de previsão como 704 para realizar uma previsão sobre uma frequência é fornecido para gerar valores residuais espectrais, em que o filtro de previsão 704 é definido pela informação do filtro de previsão derivada do sinal de áudio e encaminhada a um multiplexador do fluxo contínuo de bits 710, conforme ilustrado em 714 na Figura 7a. Além disso, um codificador de áudio como o codificador psicoacusticamente acionado de áudio 704 é fornecido. O codificador de áudio é configurado para codificar um primeiro conjunto de partes espectrais primárias dos valores residuais espectrais para obter um primeiro conjunto codificado de primeiros valores espectrais. Adicionalmente, um codificador paramétrico como o ilustrado em 706 na Figura 7a é fornecido para codificar um segundo conjunto de partes espectrais secundárias. Preferivelmente, o primeiro conjunto de partes espectrais primárias é codificado com uma resolução espectral mais alta comparada ao segundo conjunto de partes espectrais secundárias.

[000161] Finalmente, conforme ilustrado na Figura 7a, uma interface de saída é fornecida para emitir o sinal codificado, compreendendo o segundo conjunto parametricamente codificado de partes espectrais secundárias, o primeiro conjunto codificado das partes espectrais primárias e a informação do filtro ilustrada como "informação adicional TTS" em 714 na Figura 7a.

[000162] Preferivelmente, o filtro de previsão 704

compreende uma calculadora da informação do filtro configurado para utilizar os valores espectrais da representação espectral para calcular a informação do filtro. Além disso, o filtro de previsão é configurado para calcular os valores residuais espectrais utilizando os mesmos valores espectrais da representação espectral utilizado para calcular a informação do filtro.

[000163] Preferivelmente, o filtro de TTS 704 é configurado da mesma forma que conhecido para os codificadores de áudio da técnica anterior, aplicando a ferramenta de TNS de acordo com o padrão AAC.

[000164] Subsequentemente, outra implementação utilizando decodificação de dois canais é discutida no contexto das Figuras 8a a 8e. Além disso, a referência é feita à descrição dos elementos correspondentes no contexto das Figuras 2a, 2b (codificação do canal conjunto 228 e decodificação do canal conjunto 204).

[000165] A Figura 8a ilustra um decodificador de áudio para gerar um sinal de dois canais decodificados. O decodificador de áudio compreende quatro decodificadores de áudio 802 para decodificar um sinal de dois canais codificados para obter um primeiro conjunto de partes espectrais primárias e adicionalmente um decodificador paramétrico 804 para fornecer dados paramétricos para um segundo conjunto de partes espectrais secundárias e, adicionalmente, uma identificação de dois canais que identificam tanto uma primeira quanto uma segunda

representação diferente de dois canais para as partes espectrais secundárias. Adicionalmente, um regenerador de frequência 806 é fornecido para gerar novamente uma segunda parte espectral, dependendo de uma primeira parte espectral do primeiro conjunto de partes espectrais primárias e dados paramétricos para a segunda parte e a identificação de dois canais para a segunda parte. A Figura 8b ilustra diferentes combinações para representações de dois canais na faixa de origem e na faixa de destino. A faixa de origem pode estar na primeira representação de dois canais e a faixa de destino pode ainda estar na primeira representação de dois canais. De modo alternativo, a faixa de origem pode estar na primeira representação de dois canais e a faixa de destino pode estar na segunda representação de dois canais. Além disso, a faixa de origem pode estar na segunda representação de dois canais e a faixa de destino pode estar na primeira representação de dois canais conforme indicado na terceira coluna da Figura 8b. Finalmente, tanto a faixa de origem quanto a faixa de destino podem estar na segunda representação de dois canais. Em uma aplicação, a primeira representação de dois canais é uma representação separada de dois canais onde os dois canais do sinal de dois canais são individualmente representados. Então, a segunda representação de dois canais é uma representação conjunta onde os dois canais da representação de dois canais são representados conjuntamente, ou seja, onde outro processamento ou transformada da representação é necessário para recalcular uma representação separada de dois canais conforme necessário para emitir aos alto-falantes

correspondentes.

[000166] Em uma implementação, a primeira representação de dois canais pode ser uma representação esquerda/direita (L/R) e a segunda representação de dois canais é uma representação conjunta estéreo. Entretanto, outras representações de dois canais longe da previsão esquerda/direita ou M/S ou estéreo podem ser aplicadas e utilizadas para a presente invenção.

[000167] A Figura 8c ilustra um fluxograma para as operações realizadas pelo decodificador de áudio da Figura 8a. Em uma etapa 812, o decodificador de áudio 802 realiza uma decodificação da faixa de origem. A faixa de origem pode compreender, com relação à Figura 3a, bandas do fator de escala SCB1 a SCB3. Além disso, pode haver uma identificação de dois canais para cada banda de fator de escala e a banda de fator de escala 1 pode, por exemplo, estar na primeira representação (como L/R) e a terceira banda de fator de escala pode estar na segunda representação de dois canais como M/S ou previsão *downmix*/residual. Assim, a etapa 812 pode resultar em diferentes representações para diferentes bandas. Então, na etapa 814, o regenerador de frequência 806 é configurado para selecionar uma faixa de origem para uma regeneração de frequência. Na etapa 816, o regenerador de frequência 806 então verifica a representação da faixa de origem e no bloco 818, o regenerador de frequência 806 compara a representação de dois canais da faixa de origem com a representação de dois canais da faixa alvo. Se ambas as representações são idênticas, o regenerador de frequência 806

fornece uma regeneração de frequência separada para cada canal do sinal de dois canais. Quando, entretanto, ambas as representações como detectadas no bloco 818 não são idênticas, então o fluxo do sinal 824 é considerado e o bloco 822 calcula a outra representação de dois canais da faixa de origem e utiliza esta outra representação de dois canais calculada para a regeneração da faixa alvo. Assim, o decodificador da Figura 8a torna possível regenerar uma faixa de destino indicada como tendo a segunda identificação de dois canais utilizando uma faixa de origem estando na primeira representação de dois canais. Naturalmente, a presente invenção adicionalmente permite regenerar uma faixa alvo utilizando uma faixa de origem tendo a mesma identificação de dois canais. E, adicionalmente, a presente invenção permite regenerar uma faixa alvo tendo uma identificação de dois canais indicando uma representação conjunta de dois canais e para então transformar esta representação em uma representação de canal separado necessária para armazenamento ou transmissão aos alto-falantes correspondentes para o sinal de dois canais.

[000168] É enfatizado que os dois canais da representação de dois canais podem ser dois canais estéreos, como o canal esquerdo e o canal direito. Entretanto, o sinal pode ainda ser um sinal multicanal tendo, por exemplo, cinco canais e um canal *subwoofer* ou tendo ainda mais canais. Então, um processamento de dois canais pareados conforme discutido no contexto da Figura 8a a 8e pode ser realizado onde os pares pode, por exemplo, ser um canal esquerdo e um

canal direito, um canal surround esquerdo e um canal surround direito, e um canal central e um canal LFE (subwoofer). Qualquer outro pareamento pode ser utilizado a fim de representar, por exemplo, seis canais de entrada por três procedimentos de processamento de dois canais.

[000169] A Figura 8d ilustra um diagrama em blocos de um decodificador inventivo correspondente à Figura 8a. Uma faixa de origem ou um decodificador central 830 pode corresponder ao decodificador de áudio 802. Os outros blocos 832, 834, 836, 838, 840, 842 e 846 podem ser partes do regenerador de frequência 806 da Figura 8a. Particularmente, o bloco 832 é um transformador de representação para transformar as representações da faixa de origem nas bandas individuais de modo que, na saída do bloco 832, um conjunto completo da faixa de origem na primeira representação, por um lado, e na segunda representação de dois canais, por outro lado, está presente. Estas duas representações completas da faixa de origem podem ser armazenadas no armazenamento 834 para ambas as representações da faixa de origem.

[000170] Então, o bloco 836 aplica uma geração da porção de frequência utilizando, como na entrada, um ID da faixa de origem e adicionalmente utilizando como uma entrada um ID de dois canais para a faixa alvo. Com base no ID de dois canais para a faixa alvo, o gerador da porção de frequência acessa o armazenamento 834 e recebe a representação de dois canais da faixa de origem correspondentes com o ID de dois canais para a faixa alvo inserida no gerador da porção de frequência em 835. Assim,

quando o ID de dois canais para a faixa alvo indica processamento estéreo conjunto, então o gerador da porção de frequência 836 acesa o armazenamento 834 a fim de obter a representação conjunta estéreo da faixa de origem indicada pelo ID da faixa de origem 833.

[000171] O gerador da porção de frequência 836 realiza esta operação para cada faixa alvo e a saída do gerador da porção de frequência é de modo que cada canal da representação do canal identificada pela identificação de dois canais esteja presente. Então, um ajuste do envelope por um ajustador de envelope 838 é realizado. O ajuste do envelope é realizado no domínio de dois canais identificado pela identificação de dois canais. Para esta finalidade, os parâmetros do ajuste do envelope são necessários e estes parâmetros são tanto transmitidos do codificador ao decodificador na mesma representação de dois canais conforme descrito. Quando, a identificação de dois canais na faixa alvo a ser processada pelo ajustador de envelope tem uma identificação de dois canais indicando uma representação diferente de dois canais do que os dados do envelope para esta faixa alvo, então um transformador de parâmetro 840 transforma os parâmetros do envelope na representação necessária de dois canais. Quando, por exemplo, a identificação de dois canais para uma banda indica codificação conjunta estéreo e quando os parâmetros para esta faixa alvo foram transmitidos como parâmetros do envelope L/R, então o transformador de parâmetro calcula os parâmetros do envelope conjunto estéreo dos parâmetros do envelope L/R

conforme descrito de modo que a representação paramétrica correta seja utilizada para o ajuste do envelope espectral de uma faixa alvo.

[000172] Em outra aplicação preferida os parâmetros do envelope já são transmitidos como parâmetros conjuntos estéreos, quando o conjunto estéreo é utilizado em uma banda alvo.

[000173] Quando assume-se que a entrada no ajustador de envelope 838 é um conjunto de faixas alvo tendo diferentes representações de dois canais, então a saída do ajustador de envelope 838 é um conjunto de faixas alvo em diferentes representações de dois canais também. Quando, uma faixa alvo tem uma representação conjunta como M/S, então esta faixa alvo é processada por um transformador de representação 842 para calcular a representação separada necessária para um armazenamento ou transmissão em alto-falantes. Quando, entretanto, uma faixa alvo já tem uma representação separada, o fluxo do sinal 844 é considerado e o transformador de representação 842 é evitado. Na saída do bloco 842, uma representação espectral de dois canais sendo uma representação separada de dois canais é obtida que pode então ser ainda processada conforme indicado pelo bloco 846, onde este processamento ainda pode, por exemplo, ser uma conversão de frequência/tempo ou qualquer outro processamento necessário.

[000174] Preferivelmente, as partes espectrais secundárias correspondem às faixas de frequência, e a identificação de dois canais é fornecida como uma matriz de

identificadores correspondentes à tabela da Figura 8b, onde um identificador para cada faixa de frequência existe. Então, o decodificador paramétrico é configurado para verificar se o identificador é definido ou não para controlar o regenerador de frequência 106 de acordo com um identificador para utilizar tanto uma primeira representação quanto uma segunda representação da primeira parte espectral.

[000175] Em uma aplicação, apenas a faixa de reconstrução que começa com a frequência inicial de IGF 309 da Figura 3a tem a identificação de dois canais para diferentes bandas de reconstrução. Em outra aplicação, isso ainda é aplicado para a faixa de frequência abaixo da frequência inicial de IGF 309.

[000176] Em outra aplicação, a identificação da banda de origem e a identificação da banda alvo pode ser de forma adaptativa determinada por uma análise de similaridade. Entretanto, o processamento inventivo de dois canais pode ainda ser aplicado quando há uma associação fixa de uma faixa de origem em uma faixa alvo. Uma faixa de origem pode ser utilizada para recriar uma, com relação à frequência, faixa alvo mais ampla tanto pela operação harmônica de preenchimento da porção de frequência ou uma operação de cópia do preenchimento da porção de frequência utilizando duas ou mais operações de preenchimento da porção de frequência semelhantes ao processamento para várias correções conhecidas a partir do processamento AAC de alta eficiência.

[000177] A Figura 8e ilustra um codificador de áudio para codificar um sinal de dois canais de áudio. O

codificador compreende um conversor de espectro de tempo 860 para converter o sinal de dois canais de áudio na representação espectral. Além disso, um analisador espectral 866 para converter o sinal de áudio com canal de áudio de dois canais em uma representação espectral. Além disso, um analisador espectral 866 é fornecido for realizar uma análise a fim de determinar quais partes espectrais devem ser codificadas com uma alta resolução, ou seja, para encontrar o primeiro conjunto de partes espectrais primárias e para adicionalmente encontrar o segundo conjunto de partes espectrais secundárias.

[000178] Além disso, um analisador de dois canais 864 é fornecido para analisar o segundo conjunto de partes espectrais secundárias para determinar uma identificação de dois canais que identifica tanto uma primeira representação de dois canais quanto uma segunda representação de dois canais.

[000179] Dependendo do resultado do analisador de dois canais, uma banda na segunda representação espectral é tanto parametrizada utilizando a primeira representação de dois canais ou a segunda representação de dois canais, e isso é realizado por um codificador de parâmetro 868. A faixa de frequência central, ou seja, a faixa de frequência abaixo da frequência inicial de IGF 309 da Figura 3a é codificada por um codificador central 870. O resultado de blocos 868 e 870 é inserido em uma interface de saída 872. Conforme indicado, o analisador de dois canais fornece uma identificação de dois canais para cada banda tanto acima da frequência inicial de

IGF quanto para toda a faixa de frequência, e esta identificação de dois canais é ainda encaminhada à interface de saída 872 de modo que estes dados ainda estão incluídos em um sinal codificado 873 emitido pela interface de saída 872.

[000180] Além disso, é preferido que o codificador de áudio compreenda um transformador por bandas 862. Com base na decisão do analisador de dois canais 862, o sinal de saída do conversor do espectro de tempo 862 é transformado em uma representação indicada pelo analisador de dois canais e, particularmente, pelo ID de dois canais 835. Assim, uma saída do transformador por bandas 862 é um conjunto de faixas de frequência onde cada faixa de frequência pode estar tanto na primeira representação de dois canais quanto na segunda representação diferente de dois canais. Quando a presente invenção é aplicada na banda completa, ou seja, quando a faixa de origem e a faixa de reconstrução são ambas processadas pelo transformador por bandas, o analisador espectral 860 pode analisar esta representação. De modo alternativo, entretanto, o analisador espectral 860 pode ainda analisar o sinal emitido pelo conversor do espectro de tempo conforme indicado pela linha de controle 861. Assim, o analisador espectral 860 pode tanto aplicar a análise de tonalidade preferida na saída do transformador por bandas 862 quanto a saída do conversor do espectro de tempo 860 antes de ter sido processado pelo transformador por bandas 862. Além disso, o analisador espectral pode aplicar a identificação da melhor faixa de origem correspondente para uma certa faixa alvo tanto no resultado do transformador por bandas 862

quanto no resultado do conversor de espectro de tempo 860.

[000181] Subsequentemente, a referência é feita às Figuras 9a a 9d para ilustrar um cálculo preferido dos valores da informação de energia já discutido no contexto da Figura 3a e Figura 3b.

[000182] O estado moderno dos codificadores de áudio da técnica aplica várias técnicas para reduzir a quantidade de dados que representam um dado sinal de áudio. Codificadores de áudio como USAC [1] aplicam uma transformação de tempo em frequência como a MDCT para obter uma representação espectral de um dado sinal de áudio. Estes coeficientes de MDCT são quantizados utilizando os aspectos psicoacústico do sistema de audição humana. Se a taxa de bit disponível é reduzida a quantização se torna mais grosseira introduzindo grandes números de valores espectrais zero que levam às perturbações audíveis no lado do decodificador. Para melhorar o estado da qualidade perceptual, os decodificadores do estado da técnica adicionam estas partes espectrais com o ruído aleatório. O método de IGF extraem as porções do sinal não zero restante para adicionar estas lacunas. É fundamental para a qualidade perceptual do sinal decodificado de áudio que o envelope espectral e a distribuição de energia de coeficientes espectrais sejam preservados. O método de ajuste de energia apresentado aqui utiliza a informação adicional transmitida para reconstruir o envelope da MDCT espectral do sinal de áudio.

[000183] Dentro de eSBR [15] o sinal de áudio é testado pelo menos por um fator de dois e a parte da

frequência alta do espectro é completamente zerada [1, 17]. Esta parte removida é substituída por técnicas paramétricas, eSBR, no lado do decodificador. eSBR implica o uso de uma transformação adicional, a transformação de QMF que é utilizada para substituir a parte vazia de alta frequência e para testar novamente o sinal de áudio [17]. Isso soma tanto a complexidade computacional quanto o consumo de memória a um codificador de áudio.

[000184] O codificador USAC [15] oferece a possibilidade de adicionar furos espectrais (linhas espectrais zeradas) com ruído aleatório, mas tem as seguintes desvantagens: ruído aleatório não pode preservar a fina estrutura temporal de um sinal transiente e não pode preservar a estrutura harmônica de um sinal tonal.

[000185] A área onde eSBR opera no lado do decodificador foi completamente excluída pelo codificador [1]. Assim eSBR é propenso a remover as linhas tonais na região de alta frequência ou destorcer as harmônicas do sinal original da estrutura. Como a resolução de frequência QMF de eSBR é muito baixa e a inserção repetida de componentes sinusoidais é apenas possível na resolução grosseira do banco de filtra subjacente, a geração repetida de componentes tonais em eSBR na faixa de frequência replicada tem precisão muito baixa.

[000186] eSBR utiliza técnicas para ajustar as energias das áreas corrigidas, o ajuste do envelope espectral [1]. Esta técnica utiliza os valores de energia transmitidos em uma grade de tempo e frequência de QMF para remodelar o

envelope espectral. Esta técnica do estado da técnica não gerencia os espectros parcialmente excluídos e por causa da alta resolução de tempo está propensa a precisar de uma quantidade relativamente grande de bits para transmitir os valores de energia apropriados ou aplicar uma quantização grosseira aos valores de energia.

[000187] O método de IGF não precisa de uma transformação adicional, pois utiliza a transformação de MDCT de legado que é calculada conforme descrito em [15].

[000188] O método de ajuste de energia apresentado aqui utiliza a informação adicional gerada pelo codificador para reconstruir o envelope espectral do sinal de áudio. Esta informação adicional é gerada pelo codificador, conforme descrito abaixo:

a) Aplique uma transformada de MDCT em janelas ao sinal de áudio de entrada [16, seção 4.6], opcionalmente calcule uma MDCT em janelas ou estime uma MDCT em janelas da MDCT calculada;

b) Aplique TNS/TTS nos coeficientes de MDCT [15, seção 7.8];

c) Calcule a energia média para cada banda de fator de escala de MDCT acima da frequência inicial de IGF ($f_{IGFstart}$) até a frequência de parada de IGF ($f_{IGFstop}$);

d) Quantize os valores de energia média.

[000189] $f_{IGFstart}$ e $f_{IGFstop}$ são parâmetros dados pelo usuário.

[000190] Os valores calculados da etapa c) e d) são codificados e transmitidos sem perda como informação

adicional com o fluxo de bit ao decodificador.

[000191] O decodificador recebe os valores transmitidos e utiliza-os para ajustar o envelope espectral.

a) Dequantize os valores de MDCT transmitidos;

b) Aplique preenchimento de ruído USAC de legado se sinalizado;

c) Aplique o preenchimento de porção de IGF;

d) Dequantize os valores de energia transmitidos;

e) Ajuste a banda de fator de escala do envelope espectral;

f) Aplique TNS/TTS se sinalizado;

[000192] Deixe $\hat{x} \in \mathbb{R}^N$ ser a transformada de MDCT, representação espectral com valor real de um sinal de áudio em janelas do comprimento da janela $2N$. Esta transformação é descrita em [16]. O codificador opcionalmente aplica TNS em \hat{x} .

[000193] Em [16, 4.6.2] uma divisão de \hat{x} nas bandas do fator de escala é descrita. As bandas do fator de escala são um conjunto de um conjunto de índices e são denotadas neste texto com *scb*.

[000194] Os limites de cada *scb_k* with $k = 0, 1, 2, \dots, \text{max_sfb}$ são definidos por uma matriz *swb_offset* (16, 4.6.2) , onde *swb_offset[k]* e *swb_offset[k + 1] - 1* define primeiro e último índice para a linha do coeficiente espectral mais alto e mais baixo contida em *scb_k*. Denotamos a banda do fator de escala.

$$scb_k := \{swb_offset[k], 1 + swb_offset[k], 2 + swb_offset[k], \dots, swb_offset[k+1] - 1\}$$

[000195] Se a ferramenta de IGF for utilizada pelo codificador, o usuário define uma frequência inicial de IGF e uma frequência de parada de IGF. Estes dois valores são mapeados para melhor ajustar o índice da banda do fator de escala *igfStartSfb* e *igfStopSfb*. Ambos são sinalizados no fluxo de bit ao decodificador.

[000196] [16] descreve tanto uma transformação do bloco longo quanto do bloco curto. Para os blocos longos apenas um conjunto de coeficientes espectrais juntos com um conjunto de fatores de escala é transmitido ao decodificador. Para os blocos curtos oito janelas curtas com oito diferentes conjuntos de coeficientes espectrais são calculadas. Para salvar a taxa de bit, os fatores de escala destas oito janelas de bloco curto são agrupados pelo codificador.

[000197] No caso de IGF o método apresentado aqui utiliza bandas do fator de escala de legado para agrupar os valores espectrais que são transmitidos ao decodificador:

$$E_k = \sqrt{\frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_i^2}$$

[000198] Onde $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$.

[000199] Para quantizar

$$\hat{E}_k = nINT(4\log_2(E_k))$$

[000200] é calculado. Todos os valores \hat{E}_k são transmitidos ao decodificador.

[000201] Assumimos que o codificador decide agrupar os

conjuntos do fator de escala *num_window_group*. Denotamos com *w* esta divisão por agrupamento do conjunto $\{0,1,2,\dots,7\}$ que são os índices das oito janelas curtas. w_l denota o *l*-th subconjunto de *w*, onde *l* denota o índice do grupo de janela, $0 \leq l < \text{num_window_group}$.

[000202] Para o cálculo do bloco curto o usuário definiu frequência inicial/parada de IGF que é mapeada para as bandas do fator de escala apropriadas. Entretanto, para simplificar um denota para blocos curtos $k = \text{igfStartSfb}, 1 + \text{igfStartSfb}, 2 + \text{igfStartSfb}, \dots, \text{igfEndSfb}$ também.

[000203] O cálculo de energia da IGF utiliza a informação de agrupamento para agrupar os valores $E_{k,l}$:

$$E_{k,l} := \sqrt{\frac{1}{|w_l|} \sum_{j \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_{j,i}^2}$$

[000204] Para quantizar

$$\hat{E}_{k,l} = nINT(4 \log_2(E_{k,l}))$$

[000205] é calculado. Todos os valores $\hat{E}_{k,l}$ são transmitidos ao decodificador.

[000206] As fórmulas de codificação mencionadas acima operam utilizando apenas coeficientes com valor real de MDCT \hat{x} . Para obter uma distribuição de energia mais estável na faixa de IGF, ou seja, para reduzir as flutuações de amplitude temporal, um método alternativo pode ser utilizado para calcular os valores \hat{E}_k :

[000207] Deixe $\hat{x}_r \in \mathbb{R}^N$ ser a transformada de MDCT, a representação espectral com valor real de um sinal de áudio

em janelas de comprimento da janela $2N$, e $\hat{x}_i \in \mathbb{R}^N$ a representação espectral transformada por MDST com valores reais da mesma parte do sinal de áudio. A representação espectral de MDST \hat{x}_i poderia ser tanto calculada exatamente quanto estimada de \hat{x}_r . $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$ denota a representação espectral complexa do sinal de áudio em janelas, tendo \hat{x}_r como sua parte real e \hat{x}_i como sua parte imaginária. O codificador opcionalmente aplica TNS em \hat{x}_r e \hat{x}_i .

[000208] Agora a energia do sinal original na faixa de IGF pode ser medida com

$$E_{ok} = \frac{1}{|\text{scb}_k|} \sum_{i \in \text{scb}_k} \hat{c}_i^2$$

[000209] As energias da banda de reconstrução com valores reais e complexos, ou seja, a porção que deveria ser utilizada no lado do decodificador na reconstrução da faixa de IGF scb_k , é calculada com:

$$E_{tk} = \frac{1}{|\text{scb}_k|} \sum_{i \in \text{tr}_k} \hat{c}_i^2, \quad E_{rk} = \frac{1}{|\text{scb}_k|} \sum_{i \in \text{tr}_k} \hat{x}_{r_i}^2$$

[000210] onde tr_k é um conjunto de índices - a faixa da porção de origem associada, na dependência de scb_k . Nas duas fórmulas acima, ao invés do índice definido scb_k , o $\overline{\text{scb}_k}$ (definido posteriormente neste texto) definido poderia ser utilizado para criar tr_k para atingir os valores mais precisos acima E_t e E_r .

[000211] Calcule

$$f_k = \frac{E_{ok}}{E_{tk}}$$

[000212] se $E_{tk} > 0$, outro $f_k = 0$.

[000213] Com

$$E_k = \sqrt{f_k E_{rk}}$$

[000214] agora, uma versão mais estável de E_k é calculada, visto que um cálculo de E_k com valores de MDCT apenas é evitado pelo fato de que os valores de MDCT não obedecem o teorema de Parseval, e assim eles não refletem a informação de energia completa dos valores espectrais. \hat{E}_k é calculado como acima.

[000215] Conforme observado anteriormente, para blocos curtos assumimos que o codificador decide agrupar conjuntos do fator de escala *num_window_group*. Conforme mencionado acima, w_l denota o l -th subconjunto de w , onde l denota o índice do grupo de janela, $0 \leq l < \text{num_window_group}$.

[000216] Novamente, a versão alternativa descrita acima para calcular uma versão mais estável de $E_{k,l}$ poderia ser calculada. Com isso define $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$, $\hat{x}_r \in \mathbb{R}^N$ sendo a transformada de MDCT e $\hat{x}_i \in \mathbb{R}^N$ sendo o sinal de áudio em janelas transformado por MDST transformado de comprimento $2N$, calcule

$$E_{ok,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{c}_{i,l}^2$$

[000217] Calcule de forma análoga

$$E_{tk,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{c}_{i,l}^2, \quad E_{rk,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{x}_{r,l}^2$$

[000218] e proceda com o fator $f_{k,l}$

$$f_{k,l} = \frac{E_{ok,l}}{E_{tk,l}}$$

[000219] que é utilizado para ajustar $E_{rk,l}$ previamente calculado:

$$E_{k,l} = \sqrt{f_{k,l} E_{rk,l}}$$

$\hat{E}_{k,l}$ é calculado conforme mencionado acima.

[000220] O procedimento de não apenas utilizar a

energia da banda de reconstrução tanto derivada da banda de reconstrução complexa quanto dos valores de MDCT, mais também utilizando uma informação de energia da faixa de origem fornece uma reconstrução de energia melhorada.

[000221] Especificamente, a calculadora de parâmetro 1006 é configurada para calcular a informação de energia para a banda de reconstrução utilizando a informação sobre a energia da banda de reconstrução e adicionalmente utilizando a informação em uma energia de uma faixa de origem para ser utilizada para reconstruir a banda de reconstrução.

[000222] Além disso, a calculadora de parâmetro 1006 é configurada para calcular uma informação de energia (E_{ok}) na banda de reconstrução de um espectro complexo do sinal original, para calcular outra informação de energia (E_{rk}) em uma faixa de origem de uma parte com valor real do espectro complexo do sinal original para ser utilizado para reconstruir a banda de reconstrução, e em que a calculadora de parâmetro é configurada para calcular a informação de energia para a banda de reconstrução utilizando a informação de energia (E_{ok}) e outra informação de energia (E_{rk}).

[000223] Além disso, a calculadora de parâmetro 1006 é configurada para determinar uma primeira informação de energia (E_{ok}) para ser uma banda de fator de escala reconstruída de um espectro complexo do sinal original, para determinar uma segunda informação de energia (E_{tk}) em uma faixa de origem do espectro complexo do sinal original a ser utilizado para reconstruir a banda de fator de escala reconstruída a ser utilizada, para determinar uma terceira

informação de energia (E_{rk}) em uma faixa de origem de uma parte com valor real do espectro complexo do sinal original a ser utilizado para reconstruir a banda de fator de escala a ser reconstruída, para determinar uma informação de ponderação com base em uma relação entre, pelo menos, duas da primeira informação de energia, da segunda informação de energia, e da terceira informação de energia, e para ponderar uma da primeira informação de energia e da terceira informação de energia utilizando a informação de ponderação para obter uma informação ponderada de energia e para utilizar a informação de energia ponderada como a informação de energia para a banda de reconstrução.

[000224] Os exemplos para os cálculos são os seguintes, mas muitos outros podem aparecer aos especialistas na técnica em vista do princípio geral acima:

A)

$$f_k = E_{ok}/E_{tk};$$

$$E_k = \text{sqrt}(f_k * E_{rk});$$

B)

$$f_k = E_{tk}/E_{ok};$$

$$E_k = \text{sqrt}((1/f_k) * E_{rk});$$

C)

$$f_k = E_{rk}/E_{tk};$$

$$E_k = \text{sqrt}(f_k * E_{ok})$$

D)

$$f_k = E_{tk}/E_{rk};$$

$$E_k = \text{sqrt}((1/f_k) * E_{ok})$$

[000225] Todos estes exemplos reconhecem o fato de que

apenas valores reais de MDCT são processados no lado do decodificador, o cálculo real é - devido á sobreposição e adição - do procedimento de cancelamento de distorção de domínio de tempo implicitamente feito utilizando números complexos. Entretanto, particularmente, a determinação 918 da informação da energia da porção de outras partes espectrais 922, 923 da banda de reconstrução 920 para valores de frequência diferentes da primeira parte espectral 921 tendo frequências na banda de reconstrução 920 depende dos valores reais de MDCT. Assim, a informação de energia transmitida ao decodificador será tipicamente menor do que a informação de energia E_{ok} na banda de reconstrução do espectro complexo do sinal original. Por exemplo, para o caso C acima, isso significa que o fator f_k (informação de ponderação) será menor do que 1.

[000226] No lado do decodificador, se a ferramenta de IGF é sinalizada como ON, os valores transmitidos \hat{E}_k são obtidos do fluxo de bit e devem ser dequantizados com

$$E_k = 2^{\frac{1}{4}\hat{E}_k}$$

[000227] para todos os $k = igfStartSfb, 1 + igfStartSfb, 2 + igfStartSfb, \dots, igfEndSfb$.

[000228] Um decodificador dequantiza os valores de MDCT transmitidos a $x \in \mathbb{R}^N$ e calcula a energia sobrevivente remanescente:

$$sE_k := \sum_{i \in sch_k} x_i^2$$

[000229] onde k está na faixa, conforme definido acima.

[000230] Denotamos $\overline{scb_k} = \{i | i \in scb_k \wedge x_i = 0\}$. Este conjunto contém todos os índices da banda do fator de escala scb_k que foram quantizados a zero pelo codificador.

[000231] A IGF que atinge o método de sub-banda (não descrito aqui) é utilizada para adicionar lacunas espectrais resultantes de uma quantização grosseira de valores espectrais de MDCT no lado do codificador utilizando valores não zero da MDCT transmitida. \mathbf{x} adicionalmente irá conter valores que substituem todos os valores zero anteriores. A energia da porção é calculada por:

$$tE_k := \sum_{i \in \overline{scb_k}} x_i^2$$

[000232] onde k está na faixa conforme definido acima.

[000233] A energia ausente na banda de reconstrução é calculada por:

$$mE_k := |scb_k| E_k^2 - sE_k$$

[000234] E o fator de ganho para o ajuste é obtido por:

$$g := \begin{cases} \sqrt{\frac{mE_k}{tE_k}} & \text{if } (mE_k > 0 \wedge tE_k > 0) \\ 0 & \text{else} \end{cases}$$

[000235] Com

$$g' = \min(g, 10)$$

[000236] O ajuste do envelope espectral utilizando o fator de ganho é:

$$x_i := g' x_i$$

[000237] para todos $i \in \overline{scb_k}$ e k está na faixa conforme definido acima.

[000238] Isso remodela o envelope espectral de \mathbf{x} no

modelo do envelope espectral original \hat{x} .

[000239] Com a sequência de janela curta, todos os cálculos, conforme descrito acima, permanecem no princípio do mesmo, mas o agrupamento das bandas do fator de escala é considerado. Denotamos como $E_{k,l}$ os valores de energia agrupados dequantizados obtidos do fluxo de bit. Calcule

$$sE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in scb_{j,k}} x_{j,i}^2$$

e

$$pE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in \overline{scb_{j,k}}} x_{j,i}^2$$

[000240] O índice j descreve o índice da janela da sequência de bloco curto.

[000241] Calcule

$$mE_{k,l} := |scb_k| E_{k,l}^2 - sE_{k,l}$$

[000242] E

$$g := \begin{cases} \sqrt{\frac{mE_{k,l}}{pE_{k,l}}} & \text{if } (mE_{k,l} > 0 \wedge pE_{k,l} > 0) \\ 0 & \text{else} \end{cases}$$

[000243] Com

$$g' = \min(g, 10)$$

[000244] Aplique

$$x_{j,i} := g' x_{j,i}$$

[000245] para todos $i \in \overline{scb_{k,l}}$.

[000246] Para aplicações de baixa taxa de bit um agrupamento em pares dos valores E_k é possível sem perder muita precisão. Este método é aplicado apenas com blocos longos:

$$E_{k \gg 1} = \sqrt{\frac{1}{|\overline{scb_k \cup scb_{k+1}}|} \sum_{i \in \overline{scb_k \cup scb_{k+1}}} \hat{x}_i^2}$$

[000247] onde $k = igfStartSfb, 2 + igfStartSfb, 4 +$
 $igfStartSfb, \dots, igfEndSfb$.

[000248] Novamente, após quantizar todos os valores $E_{k \gg 1}$ são transmitidos ao decodificador.

[000249] A Figura 9a ilustra um aparelho para decodificar 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.

[000250] 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.

[000251] A 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 é, 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.

[000252] 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.

[000253] 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.

[000254] 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.

[000255] 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.

[000256] Além disso, é preferido que os limites da banda de reconstrução coincidam 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 central 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.

[000257] 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 fornecida 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.

[000258] Subsequentemente, a 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.

[000259] 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.

[000260] 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.

[000261] 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.

[000262] 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 central 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.

[000263] 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.

[000264] 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.

[000265] Nos esquemas de BWE do estado da técnica, a reconstrução da região espectral HF acima de uma dada assim chamada frequência cruzada é frequentemente com base na correção espectral. Tipicamente, a região HF é composta de múltiplas correções adjacentes e cada uma dessas correções é proveniente de regiões de passa-banda (BP) do espectro LF abaixo da dada frequência cruzada. Dentro de uma representação do banco de filtro do sinal tais sistemas copiam um conjunto de coeficientes de sub-banda adjacentes para fora do espectro LF na região alvo. Os limites dos conjuntos selecionados são tipicamente dependentes do sistema

e não dependentes do sinal. Para alguns conteúdos do sinal, essa seleção de correções estáticas pode levar a um timbre e coloração desagradável do sinal reconstruído.

[000266] Outras abordagens transferem o sinal LF ao HF através de uma modulação de Banda Lateral Única (SSB | *Single Side Band*) de sinal adaptativo. Tais abordagens são de alta complexidade computacional se comparadas ao [1] visto que elas operam em alta taxa de amostragem em amostras de domínio de tempo. Também, a correção pode se tornar instável, especialmente para sinais não tonais (por exemplo, discurso não vocalizado), e, assim, a correção adaptativa do sinal de última geração pode introduzir imparidades ao sinal.

[000267] A abordagem inventiva é denominada Preenchimento de Lacuna Inteligente (IGF) e, em sua configuração preferida, é aplicada em um sistema BWE com base em uma transformada de frequência temporal, como por exemplo, a Transformada de cosseno discreta modificada (MDCT). Entretanto, os ensinamentos da invenção são geralmente aplicáveis, por exemplo, analogamente dentro de um sistema baseado no Banco de Filtro em Espelho de Quadratura (QMF).

[000268] Uma vantagem da configuração IGF com base na MDCT é a integração contínua na MDCT com base em codificadores de áudio, por exemplo, Codificação de Áudio Avançada MPEG (AAC). Dividir a mesma transformada para codificação de áudio de forma de onda e para BWE reduz a complexidade computacional geral para o codificador de áudio significativamente.

[000269] Além disso, a invenção fornece uma solução

para os problemas de estabilidade inerentes encontrados nos esquemas de correção adaptativos do estado da técnica.

[000270] O sistema proposto se baseia na observação que para alguns sinais, uma seleção de correção não guiada pode levar a alterações de timbre e colocações de sinal. Se um sinal que é tonal na região de origem espectral (SSR | *spectral source region*), mas é com característica de ruído na região alvo espectral (STR | *spectral target region*), corrigindo o sinal com característica de ruído da STR pela estrutura do sinal tonal pode deixa-lo desalinhado ou mesmo destruído pelo processo de correção.

[000271] O sistema IGF proposto realiza uma seleção de porção inteligente utilizando correlação cruzada como uma medição de similaridade entre uma SSR particular e uma STR específica. A correlação cruzada de dois sinais fornece uma medida de similaridade daqueles sinais e ainda o atraso da correlação máxima e seu sinal. Consequentemente, a abordagem de uma correlação com base na seleção de porção pode ainda ser utilizada para ajustar precisamente a compensação espectral do espectro copiado pra torna-lo o mais próximo possível da estrutura espectral original.

[000272] A contribuição fundamental do sistema proposto é a escolha de uma medição de similaridade adequada, e ainda técnicas para estabilizar o processo de seleção de porção. A técnica proposta fornece um balanço ideal entre a adaptação de sinal instantânea e, ao mesmo tempo, a estabilidade temporal. O fornecimento da estabilidade temporal é especialmente importante para sinais que tenham

pouca similaridade de SSR e STR e assim exibem valores baixos da correlação cruzada ou se medições de similaridade que são empregadas são ambíguas. Em tais casos, a estabilização previne o comportamento pseudoaleatório da seleção de porção adaptativa.

[000273] Por exemplo, uma classe de sinais que frequentemente representam problemas para a BWE de última geração é caracterizada por uma concentração distinta de energia para regiões espectrais arbitrárias, como mostrado na Figura 12a (esquerda). Embora haja métodos disponíveis para ajustar o envelope espectral e a tonalidade do espectro reconstruído na região alvo, para alguns sinais esses métodos não estão aptos a preservar o timbre bem como mostrado na Figura 12a (direita). No exemplo mostrado na Figura 12a, a magnitude do espectro na região alvo do sinal original acima de uma assim chamada frequência cruzada f_{xover} (Figura 12a, esquerda) diminui quase linearmente. Em contraste, no espectro reconstruído (Figura 12a, direita), um conjunto distinto de mínimos e máximos está presente, que é como uma perturbação de colorização de timbre.

[000274] Uma etapa importante da nova abordagem é definir um conjunto de porções entre as quais a escolha com base na similaridade subsequente pode acontecer. Primeiro, os limites da porção de ambas, a região de origem e a região alvo, tem que ser definidos em conformidade um com o outro. Assim, a região alvo entre a frequência inicial de IGF do codificador central $f_{IGFstart}$ e a frequência mais alta disponível $f_{IGFstop}$ é dividida em um número inteiro arbitrário

$nTar$ de porções, cada uma dessas tendo um tamanho individual pré-definido. Então, para cada porção alvo $tar[idx_tar]$, um conjunto de porções de origem de tamanhos iguais $src[idx_src]$ é gerado. Dessa forma, o grau básico de liberdade do sistema IGF é determinado. O número total de porções de origem $nSrc$ é determinado pela largura de banda da região de origem,

$$bw_{src} = (f_{IGFstart} - f_{IGFmin})$$

[000275] onde f_{IGFmin} é a frequência mais baixa disponível para a seleção de porção de modo que um número inteiro $nSrc$ de porções de origem encaixe em bw_{src} . O número mínimo de porções de origem é 0.

[000276] Para aumentar mais o grau de liberdade para seleção e ajuste, as porções de origem podem ser definidas para sobrepor uma a outra através de um fator de sobreposição entre 0 e 1, onde 0 significa que não há sobreposição e 1 significa 100% de sobreposição. O caso de sobreposição de 100% implica que apenas uma ou nenhuma porção de origem está disponível.

[000277] A Figura 12b mostra um exemplo de limites de porção de um conjunto de porções. Nesse caso, todas as porções alvo são correlacionadas com cada uma das porções de origem. Nesse exemplo, as porções de origem são 50% sobrepostas.

[000278] Para uma porção alvo, a correlação cruzada é computada com várias porções de origem em atrasos de $xcorr_maxLag$ posições. Para uma dada porção alvo idx_tar e uma porção de origem idx_src , o $xcorr_val[idx_tar][idx_src]$ dá o valor máximo da correlação cruzada absoluta entre as porções,

enquanto `xcorr_lag[idx_tar][idx_src]` dá o atraso no qual o seu máximo ocorre e `xcorr_sign[idx_tar][idx_src]` dá o sinal da correlação cruzada em `xcorr_lag[idx_tar][idx_src]`.

[000279] O parâmetro `xcorr_lag` é utilizado para controlar a proximidade da combinação entre as porções de origem e alvo. Esse parâmetro leva a perturbações reduzidas e ajuda a preservar melhor o timbre e a cor do sinal.

[000280] Em alguns cenários pode acontecer que o tamanho de uma porção alvo específica seja maior que o tamanho das porções de origem disponíveis. Nesse caso, a porção de origem disponível é repetida sempre que necessário para preencher a porção alvo específica completamente. É possível realizar a correlação cruzada entre a porção alvo grande e porção de origem menor para conseguir a melhor posição da porção de origem na porção alvo em termos de atraso da correlação cruzada `xcorr_lag` e sign `xcorr_sign`.

[000281] A correlação cruzada das porções espectrais brutas e o sinal original pode não ser a medição de similaridade mais adequada aplicada aos espectros de áudio com estrutura formante forte. Clarear um espectro remove a informação grosseira do envelope e, assim, enfatiza a estrutura espectral fina, que é de interesse primordial para avaliar a similaridade da porção. Clarear ainda auxílios em um modelador fácil do envelope da STR no decodificador para as regiões processadas pelo IGF. Assim, opcionalmente, a porção e o sinal de origem são clareados antes de calcular a correlação cruzada.

[000282] Em outras configurações, apenas a porção é

clareada utilizando um procedimento predefinido. Um identificador "clareada" transmitido indica ao decodificador que o mesmo procedimento de clareamento predefinido deverá ser aplicado à porção dentro da IGF.

[000283] Para clarear o sinal, primeiro um envelope espectral estimado é calculado. Então, o espectro da MDCT é dividido pelo envelope espectral. O envelope espectral estimado pode ser estimado sobre o espectro da MDCT, as energias do espectro da MDCT, o espectro de energia ou espectro de energia complexo baseado na MDCT. O sinal sobre o qual o envelope é estimado será chamado de *sinal base* a partir de agora.

[000284] Envelopes calculados sobre a MDCT com base no espectro de potência complexo ou estimativas do espectro de potência como sinal base têm a vantagem de não ter flutuação temporal sobre componentes tonais.

[000285] Se o sinal base estiver em um domínio de energia, o espectro da MDCT tem que ser dividido pela raiz quadrada do envelope para clarear o sinal corretamente.

[000286] Há diferentes métodos para cálculo do envelope:

- Transformar o sinal base com transformada de cosseno discreta (DCT), retendo apenas os coeficientes DCT mais baixos (definindo o superior a zero) e então calcular o DCT inverso;
- Calcular um envelope espectral de um conjunto de Coeficientes de Previsão Linear (LPC | *Linear Prediction Coefficients*) calculado na estrutura de áudio de

domínio de tempo;

- Filtrar o sinal base com um filtro passa-baixa.

[000287] Preferivelmente, a última abordagem é escolhida. Para aplicações que exigem baixa complexidade computacional, algumas simplificações podem ser feitas para clarear um espectro da MDCT: Primeiro o envelope é calculado por meio de uma média de movimento. Isso exige apenas dois ciclos de processadores por bin da MDCT. Então para evitar o cálculo da divisão e a raiz quadrada, o envelope espectral é aproximado por 2^n , onde n é o logaritmo de número inteiro do envelope. Nesse domínio a operação de raiz quadrada se torna simplesmente uma operação de comutação e, além disso, a divisão pelo envelope pode ser realizada por outra operação de comutação.

[000288] Após calcular a correlação de cada porção de origem com cada porção alvo, para todas as porções alvo $nTar$ a porção de origem com a correlação mais alta é escolhida para substituí-la. Para combinar melhor com a estrutura espectral original, o atraso da correlação é utilizado para modular o espectro replicado por um número inteiro de porções de transformada. No caso de atrasos estranhos, a porção é adicionalmente modulada através da multiplicação por uma sequência temporal alternada de $-1/1$ para compensar para a representação reservada por frequência de todas as outras bandas dentro da MDCT.

[000289] A Figura 12c mostra um exemplo de uma correlação entre uma porção de origem e uma porção alvo.

Nesse exemplo, o atraso da correlação é 5, então a porção de origem tem que ser modulada a 5 posições em direção às posições de alta frequência no estágio de cópia do algoritmo de BWE. Além disso, o sinal da porção tem que ser invertido visto que o valor de correlação máximo é negativo e uma modulação adicional, conforme descrito acima, conta com o atraso ímpar.

[000290] Então a quantidade total de informação adicional para transmitir do codificador ao decodificador pode consistir dos seguintes dados:

- $tileNum[nTar]$: índice da porção de origem selecionada por porção alvo;
- $tileSign[nTar]$: sinal da porção alvo;
- $tileMod[nTar]$: atraso da correlação por porção alvo.

[000291] A redução e estabilização da porção é uma etapa importante na IGF. Suas necessidades e vantagens são explicadas com um exemplo, assumindo um sinal de áudio tonal adicional, como por exemplo, uma nota de passa estável. A lógica dita que menos perturbações são introduzidas se, para uma dada região alvo, porções de origem são sempre selecionadas a partir da mesma região de origem através das estruturas. Muito embora o sinal seja assumido como estacionário, essa condição não se manteria válida em toda a estrutura visto que a medição de similaridade (por exemplo, correlação) de outra região de origem igualmente similar poderia dominar o resultado de similaridade (por exemplo, correlação cruzada). Isso leva a $tileNum[nTar]$ entre

estruturas adjacentes a vacilar entre duas ou três escolhas muito similares. Isso pode ser a de origem de um barulho musical irritante como perturbação.

[000292] Para eliminar esse tipo de perturbações, o conjunto de porções de origem deverá ser reduzido de modo que os membros restantes do conjunto de origem sejam maximamente diferentes. Isso é alcançado sobre um conjunto de porções de origem

$$S = \{s_1, s_2, \dots, s_n\}$$

[000293] conforme segue. Para qualquer porção de origem s_i , nós a correlacionamos com todas as outras porções de origem, constatando a melhor correlação entre s_i e s_j e a armazenando na matriz S_x . Aqui $S_x[i][j]$ contém o valor absoluto máximo de correlação cruzada entre s_i e s_j . Adicionar a matriz S_x ao longo das colunas, nos dá a soma de correlações cruzadas de uma porção de origem s_i com todas as outras porções de origem T .

$$T[i] = S_x[i][1] + S_x[i][2] \dots + S_x[i][n]$$

[000294] Aqui T representa uma medição de quão bem uma de origem é similar a outras porções de origem. Se para qualquer porção de origem i ,

$$T > \text{limite}$$

[000295] a porção de origem i pode ser diminuída do conjunto de origem em potencial visto que está altamente correlacionada com outras de origem. A porção com a correlação mais baixa a partir do conjunto de porções que satisfaz a condição na equação 1 é escolhida como uma porção representante para esse subconjunto. Dessa forma, nós

garantimos que essas porções de origem são maximamente diferentes umas das outras.

[000296] O método de redução de porção envolve, ainda, uma memória do conjunto de porção reduzido utilizado na estrutura prévia. Porções que eram ativas na estrutura prévia são retidas na próxima estrutura também caso candidatos alternativos para a redução existam.

[000297] Deixe as porções s_3 , s_4 e s_5 serem ativas fora das porções $\{s_1, s_2, \dots, s_5\}$ na estrutura k , então na estrutura $k+1$ mesmo se as porções s_1 , s_3 e s_2 estão competindo para serem reduzidas com s_3 sendo maximamente correlacionadas com as outras, s_3 é retida visto que era uma porção de origem útil na estrutura prévia, e, assim, retê-la no conjunto de porções de origem é benéfico para reforçar a continuidade temporal na seleção de porção. Esse método é preferivelmente aplicado se a correlação cruzada entre a de origem i e alvo j , representada como $T_x[i][j]$ for alta.

[000298] Um método adicional para estabilização de porção é para reter a ordem de porção a partir da estrutura prévia $k-1$ se nenhuma das porções de origem na estrutura atual k se correlacione bem com as porções alvo. Isso pode acontecer se a correlação cruzada entre a porção de origem i e alvo j , representada como $T_x[i][j]$ for muito baixa para todos i, j .

[000299] Por exemplo, se

$$T_x[i][j] < 0,6$$

[000300] um limite provisório sendo utilizado agora, então,

$$\text{tileNum}[\text{nTar}]_k = \text{tileNum}[\text{nTar}]_{k-1}$$

[000301] para todos os nTar dessa estrutura k.

[000302] As duas técnicas acima reduzem amplamente as perturbações que ocorre a partir dos números da porção do conjunto de alteração rápida através das estruturas. Outra vantagem adicionada dessa redução e estabilização de porção é que nenhuma informação extra precisa ser enviada ao decodificador nem uma alteração da arquitetura do decodificador é necessária. Essa redução de porção proposta é uma maneira elegante de reduzir ruídos musicais em potencial, como perturbações, ou ruído excessivo nas regiões espectrais em porção.

[000303] A Figura 11a ilustra um decodificador de áudio para decodificar um sinal de áudio codificado. O decodificador de áudio compreende um decodificador de áudio 1102 (núcleo) para gerar uma primeira representação decodificada de um primeiro conjunto de partes espectrais primárias, a representação decodificada tendo uma primeira resolução espectral.

[000304] Além disso, o decodificador de áudio compreende um decodificador paramétrico 1104 para gerar uma segunda representação decodificada de um segundo conjunto de partes espectrais secundárias tendo uma segunda resolução espectral sendo mais baixa que a primeira resolução espectral. Além disso, um regenerador de frequência 1106 é fornecido que recebe como uma primeira entrada 1101 as partes espectrais primárias decodificadas e como uma segunda entrada

1103 a informação paramétrica incluindo, for cada porção alvo de frequência ou banda alvo de reconstrução uma informação da faixa de origem. O regenerador de frequência 1106 então aplica a regeneração de frequência utilizando valores espectrais a partir da faixa de origem, identificada pela informação correspondente, para gerar os dados espectrais para a faixa alvo. Então, as partes espectrais primárias 1101 e a saída do regenerador de frequência 1107 são ambas inseridas em um conversor de espectro-tempo 1108 para finalmente gerar o sinal decodificado de áudio.

[000305] Preferivelmente, o decodificador de áudio 1102 é um decodificador de áudio de domínio espectral, embora o decodificador de áudio possa ainda ser implementado como qualquer outro decodificador de áudio como um domínio de tempo ou decodificador de áudio paramétrico.

[000306] Conforme indicado na Figura 11b, o regenerador de frequência 1106 pode compreender as funcionalidade de bloco 1120 que ilustra um modulador de porção do seletor de faixa de origem para atrasos estranhos, um filtro de clareamento 1122, quando um identificador de clareamento 1123 é fornecido, e adicionalmente um envelope espectral com funcionalidades de ajuste implementadas ilustradas no bloco 1128, utilizando os dados espectrais não processados gerados pelo bloco 1120 ou bloco 1122 ou pela cooperação de ambos os blocos. De qualquer forma, o regenerador de frequência 1106 pode compreender um interruptor 1124 reativo a um identificador de clareamento 1123 recebido. Quando o identificador de clareamento é

definido, a saída do modulador de porção/seletor da faixa de origem para atrasos estranhos é inserida no filtro de clareamento 1122. Assim, portanto, o identificador de clareamento 1123 não é definido para determinada banda de reconstrução, então, uma linha de derivação 1126 é ativada, de modo que a saída do bloco 1120 é fornecida ao bloco de ajuste do envelope espectral 1128 sem qualquer clareamento.

[000307] Pode haver mais que um nível de clareamento (1123) sinalizado no fluxo contínuo de bits e esses níveis podem ser assinalados por porção. No caso em que há três níveis assinalados por porção, eles deverão ser codificados da seguinte forma:

```
bit = readBit(1);
if(bit == 1) {
    for(tile_index = 0..nT)
        /*same levels as last frame [mesmos níveis que os da
última estrutura]*/
        whitening_level[tile_index] =
whitening_level_prev_frame[tile_index];
} else {
    /*first tile:*/
    tile_index = 0;
    bit = readBit(1);
    if(bit == 1) {
        whitening_level[tile_index] = MID_WHITENING;
    } else {
        bit = readBit(1);
        if(bit == 1) {
```

```

    whitening_level[tile_index] = STRONG_WHITENING;
} else {
    whitening_level[tile_index] = OFF; /*sem clareamento
[no-whitening]*/
}
}

/*remaining tiles:*/
bit = readBit(1);
if(bit == 1) {
    /*níveis de achatamento para as porções restantes iguais
aos da primeira [flattening levels for remaining tiles same
as first].*/

    /*Nenhum bit adicional deverá ser lido [No further bits
have to be read]*/

    for(tile_index = 1..nT)
        whitening_level[tile_index] = whitening_level[0];
    } else {
        /*leia os bits para as porções restantes como para a
primeira porção [read bits for remaining tiles as for first
tile]*/

        for(tile_index = 1..nT) {
            bit = readBit(1);
            if(bit == 1) {
                whitening_level[tile_index] = MID_WHITENING;
            } else {
                bit = readBit(1);
                if(bit == 1) {
                    whitening_level[tile_index] = STRONG_WHITENING;

```



```

    } else {

        whitening_level[tile_index] = OFF; /* sem
clareamento [no-whitening]*/

    }

}

}

}

```

[000308] MID_WHITENING e STRONG_WHITENING referem-se a filtros de clareamento diferentes (1122) que podem diferir na forma que o envelope é calculado (conforme descrito anteriormente).

[000309] O regenerador de frequência adicional do decodificador pode ser controlado por um ID da faixa de origem 1121 quando apenas um esquema de seleção de porção espectral grosseiro é aplicado. Quando, entretanto, um esquema de seleção de porção espectral delineado é aplicado, então, adicionalmente, uma faixa de origem atraso 1119 é fornecida. Além disso, dito que o cálculo de correlação fornece um resultado negativo, então, adicionalmente, um sinal da correlação pode ainda ser aplicado ao bloco 1120 de modo que as linhas espectrais dos dados da página sejam multiplicadas por "-1" para explicar o sinal negativo.

[000310] Assim, a presente invenção, conforme discutido na Figura 11a, 11b certifica que uma qualidade de áudio ideal é obtida devido ao fato de que a melhor faixa de origem ou faixa alvo correspondente para certa finalidade é calculada no lado do codificador e é aplicada no lado do

decodificador.

[000311] A Figura 11c é um determinado codificador de áudio para codificar um sinal de áudio, compreendendo um conversor de espectro de tempo 1130, um analisador espectral 1132 subsequentemente conectado e, adicionalmente, uma calculadora de parâmetro 1134 e um codificador central 1136. O codificador central 1136 extrai faixas de origem codificadas e a calculadora de parâmetro 1134 extrai informação correspondente para faixas alvo.

[000312] As faixas de origem codificadas são transmitidas para um decodificador juntamente com a informação correspondente para as faixas alvo de modo que o decodificador ilustrado na Figura 11a esteja na posição para realizar uma regeneração de frequência.

[000313] A calculadora de parâmetro 1134 é configurada para calcular similaridades entre as partes espectrais primárias e as partes espectrais secundárias e para determinar, com base nas similaridades calculadas, para uma segunda parte espectral uma combinação da primeira parte espectral correspondente com a segunda parte espectral. Preferivelmente, resultados de adequação para faixas de origem e faixas alvos diferentes, conforme ilustrado nas Figuras 12a, 12b para determinar um par de combinações selecionado compreendendo a segunda parte espectral, e a calculadora de parâmetro é configurada para fornecer essa informação correspondente identificando o par de combinações em um sinal de áudio codificado. Preferivelmente, essa calculadora de parâmetro 1134 é configurada para utilizar

regiões alvo pré-definidas no segundo conjunto de partes espectrais secundárias ou regiões de origem pré-definidas no primeiro conjunto de partes espectrais primárias, conforme ilustrado, por exemplo, na Figura 12b. Preferivelmente, as regiões alvo pré-definidas são de não sobreposição ou as regiões de origem pré-definidas são de sobreposição. Quando as regiões de origem pré-definidas são um subconjunto do primeiro conjunto de partes espectrais primárias, abaixo uma frequência inicial de preenchimento da lacuna 309 da Figura 3a e preferivelmente a região alvo pré-definida cobrindo uma região espectral mais baixa coincidem com o seu limite de frequência mais baixo com a frequência inicial de preenchimento da lacuna, de modo que quaisquer faixas alvo sejam localizadas acima da frequência inicial de preenchimento da lacuna e faixas de origem sejam localizadas abaixo da frequência inicial de preenchimento da lacuna.

[000314] Conforme discutido, uma granularidade fina é obtida, comparando uma região alvo com uma região de origem sem qualquer atraso para a região de origem e a mesma região de origem, mas com certo atraso. Esses atrasos são aplicados na calculadora de correlação cruzada 1140 da Figura 11d e a seleção do par de combinações é finalmente realizada pelo seletor de porção 1144.

[000315] Além disso, é preferido realizar um clareamento de faixas alvo e/ou de origem ilustrado no bloco 1142. Este bloco 1142 então fornece um identificador de clareamento ao fluxo contínuo de bits que é utilizado para controlar o interruptor do lado do decodificador 1123 da

Figura 11b. Além disso, se a calculadora de correlação cruzada 1140 fornece um resultado negativo, então este resultado negativo é ainda sinalizado a um decodificador. Assim, em uma aplicação preferida, o seletor de porção emite um ID da faixa de origem para uma faixa alvo, um atraso, um sinal e bloco 1142 adicionalmente fornece um identificador de clareamento.

[000316] Além disso, a calculadora de parâmetro 1134 é configurada para realizar um corte da porção de origem 1146, reduzindo o número de faixas de origem potenciais em que uma correção de origem é derrubada de um conjunto de porções de origem potenciais com base em um limite de similaridade. Assim, quando duas porções de origem são semelhantes a ou iguais a um limite de similaridade, então, uma dessas duas porções de origem é removida do conjunto de origem potenciais e a porção de origem removida não é utilizada mais para processar e, especificamente, não pode ser selecionada pelo seletor de porção 1144 ou não é utilizado para o cálculo de correlação cruzada entre as diferentes faixas de origem e faixas alvo conforme realizado no bloco 1140.

[000317] Diferentes implementações foram descritas com relação às diferentes figuras. As Figuras 1a-5c referem-se a um esquema do codificador/decodificador da taxa completa ou largura de banda total. As Figuras 6a-7e referem-se a um esquema do codificador/decodificador com processamento TNS ou TTS. As Figuras 8a-8e referem-se a um esquema do codificador/decodificador com processamento específico de dois canais. As Figuras 9a-10d referem-se a um cálculo

específico da informação de energia e aplicação, e as Figuras 11a-12c refere-se a uma forma específica da seleção de porção.

[000318] Todos estes diferentes aspectos podem ser de uso inventivo independentes entre eles, mas adicionalmente podem, ainda, ser aplicados juntos como basicamente ilustrado nas Figuras 2a e 2b. Entretanto, o processamento específico de dois canais pode ser aplicado a um esquema do codificador/decodificador ilustrado na Figura 13 também, e o mesmo é verdadeiro para o processamento de TNS/TTS, o cálculo da informação de energia do envelope e aplicação na banda de reconstrução ou a identificação da faixa de origem adaptativa e aplicação correspondente no lado do decodificador. Por outro lado, o aspecto da taxa completa pode ser aplicado com ou sem o processamento de TNS/TTS, com ou sem processamento de dois canais, com ou sem uma identificação da faixa de origem adaptativa ou com outros tipos de cálculo de energias para a representação espectral do envelope. Assim, é claro que as características de um desses aspectos individuais podem ser aplicadas também em outros aspectos.

[000319] 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.

[000320] 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.

[000321] 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.

[000322] 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.

[000323] Outras aplicações compreendem o programa de computador para realizar um dos métodos descritos neste documento, armazenados em um transportador legível por máquina.

[000324] 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.

[000325] 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.

[000326] 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.

[000327] 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.

[000328] Outra aplicação compreende um computador tendo instalado nele o programa de computador para realizar um dos métodos descritos neste documento.

[000329] Uma aplicação adicional, de acordo com a invenção, compreende um aparelho ou sistema configurado para transferir (por exemplo, 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.

[000330] Em algumas aplicações, um dispositivo lógico programável (por exemplo, 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.

[000331] 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

- [000332] [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.
- [000333] [2] Ferreira, D. Sinha, "Accurate Spectral Replacement", Audio Engineering Society Convention, Barcelona, Spain 2005.
- [000334] [3] D. Sinha, A. Ferreira and E. Harinarayanan, "A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)", Audio Engineering Society Convention, Paris, France 2006.
- [000335] [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.
- [000336] [5] T. Żernicki, M. Bartkowiak, "Audio bandwidth extension by frequency scaling of sinusoidal partials", Audio Engineering Society Convention, San Francisco, USA 2008.
- [000337] [6] J. Herre, D. Schulz, Extending the MPEG-4 AAC Codec by Perceptual Noise Substitution, 104th AES Convention, Amsterdam, 1998, Preprint 4720.

[000338] [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.

[000339] [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.

[000340] [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.

[000341] [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

[000342] [11] International Standard ISO/IEC 13818-3, Generic Coding of Moving Pictures and Associated Audio: Audio", Geneva, 1998.

[000343] [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

[000344] [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

[000345] [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

[000346] [15] International Standard ISO/IEC 23001-3:2010, Unified speech and audio coding Audio, Geneva, 2010.

[000347] [16] International Standard ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, Geneva, 2005.

[000348] [17] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication", in Proceedings of 1st IEEE Benelux Workshop on MPCA, Leuven, November 2002

[000349] [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.

REIVINDICAÇÕES

1. Aparelho para decodificar um sinal de áudio codificado, compreendendo:

um decodificador de áudio de domínio espectral (602) para gerar uma primeira representação decodificada de um primeiro conjunto de partes espectrais primárias sendo valores residuais de previsão espectrais;

um regenerador de frequência (604) para gerar uma segunda parte espectral reconstruída, utilizando uma primeira parte espectral do primeiro conjunto de partes espectrais primárias, caracterizado pela segunda parte espectral reconstruída e o primeiro conjunto de partes espectrais primárias compreenderem valores residuais de previsão espectrais; e

um filtro de previsão inversa (606, 616, 622) para realizar uma previsão inversa sobre a frequência utilizando os valores residuais de previsão espectrais para o primeiro conjunto de partes espectrais primárias e a segunda parte espectral reconstruída utilizando a informação do filtro de previsão (607) incluída no sinal de áudio codificado,

em que o regenerador de frequência (604) é configurado para

correção espectral ou transposição dos valores residuais de previsão espectral na primeira parte espectral de áudio para obter a segunda porção espectral de áudio reconstruída, ou

usar os valores residuais de previsão

espectral na primeira parte espectral de áudio dentro de uma operação de preenchimento de lacuna inteligente para obter a segunda parte espectral de áudio reconstruída, ou usar os valores residuais de previsão espectral na primeira parte espectral de áudio dentro de uma operação de replicação de banda espectral para obter a segunda parte espectral de áudio reconstruída.

2. Aparelho de acordo com a reivindicação 1, caracterizado pelo filtro de previsão inversa (606) ser um filtro complexo definido pela informação do filtro de previsão (607).

3. Aparelho de acordo com qualquer uma das reivindicações 1 ou 2 caracterizado pela primeira representação decodificada do primeiro conjunto de partes espectrais primárias compreender valores espectrais reais, em que o aparelho compreende, ainda, um estimador (724) para estimar os valores imaginários para o primeiro conjunto de partes espectrais primárias do primeiro conjunto com valor real de partes espectrais primárias, e em que o filtro de previsão inversa (728) é um filtro de previsão inversa complexa definido pela informação do filtro de previsão com valor complexo (714) e em que o aparelho compreende, ainda, um conversor de frequência-tempo (730) configurado para realizar uma conversão de um espectro com valor complexo em um sinal de áudio de domínio de tempo.

4. Aparelho de acordo com qualquer uma das reivindicações de 1 a 3, caracterizado pelo filtro de

previsão inversa (606, 728) ser configurado para aplicar uma pluralidade de subfiltros, em que um limite de frequência de cada subfiltro coincide com um limite de frequência de uma banda de reconstrução que coincide com uma porção de frequência.

5. Método para decodificar um sinal de áudio codificado, compreendendo:

gerar (602) uma primeira representação decodificada de um primeiro conjunto de partes espectrais primárias sendo valores residuais de previsão espectrais;

regenerar (604) uma segunda parte espectral reconstruída, utilizando uma primeira parte espectral do primeiro conjunto de partes espectrais primárias, caracterizado pela segunda parte espectral reconstruída e pelo primeiro conjunto de partes espectrais primárias compreenderem valores residuais de previsão espectrais; e

realizar (606, 616, 622) por um preenchedor de previsão inversa, uma previsão inversa sobre a frequência, utilizando os valores residuais de previsão espectrais para o primeiro conjunto de partes espectrais primárias e a segunda parte espectral reconstruída utilizando informação do filtro de previsão (607) incluída no sinal de áudio codificado,

em que o regenerador (604) compreende

corrigir espectralmente ou transposicionalmente os valores residuais de previsão espectral na primeira parte espectral de áudio para obter a segunda porção espectral de áudio reconstruída, ou

usar os valores residuais de previsão espectral na primeira parte espectral de áudio dentro de uma operação de preenchimento de lacuna inteligente para obter a segunda parte espectral de áudio reconstruída, ou

usar os valores residuais de previsão espectral na primeira parte espectral de áudio dentro de uma operação de replicação de banda espectral para obter a segunda parte espectral de áudio reconstruída.

6. Mídia de armazenamento não transitória tendo gravadas instruções lidas por um processador, caracterizada por compreender instruções que, quando executadas, realizam o método da reivindicação 5.

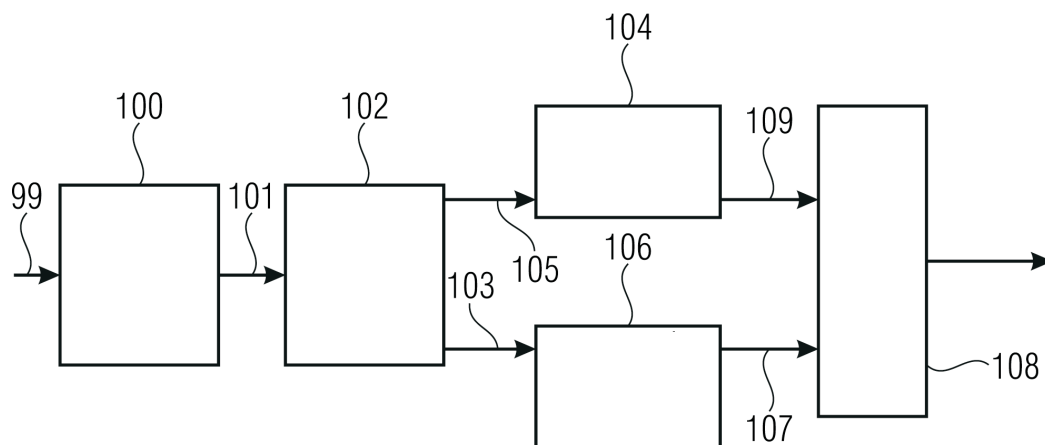


FIG 1A

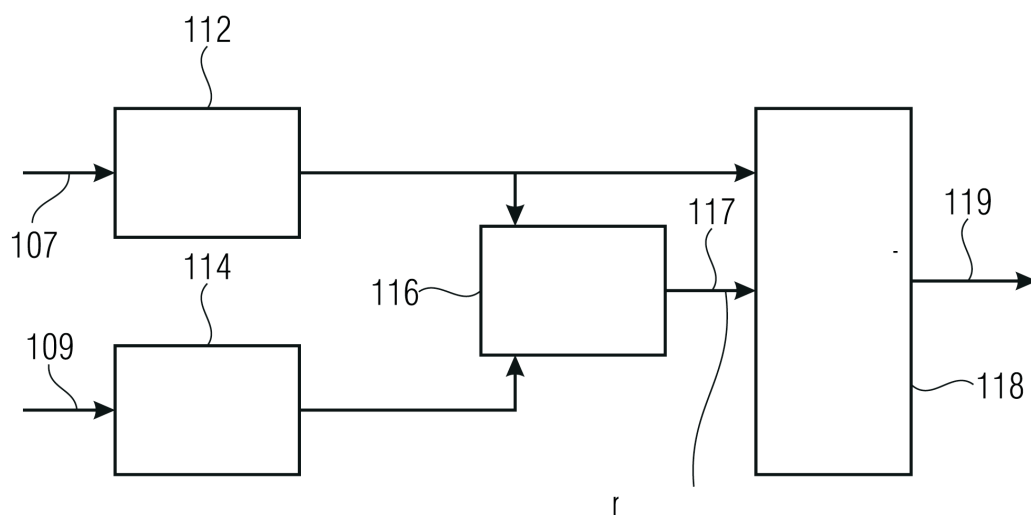


FIG 1B

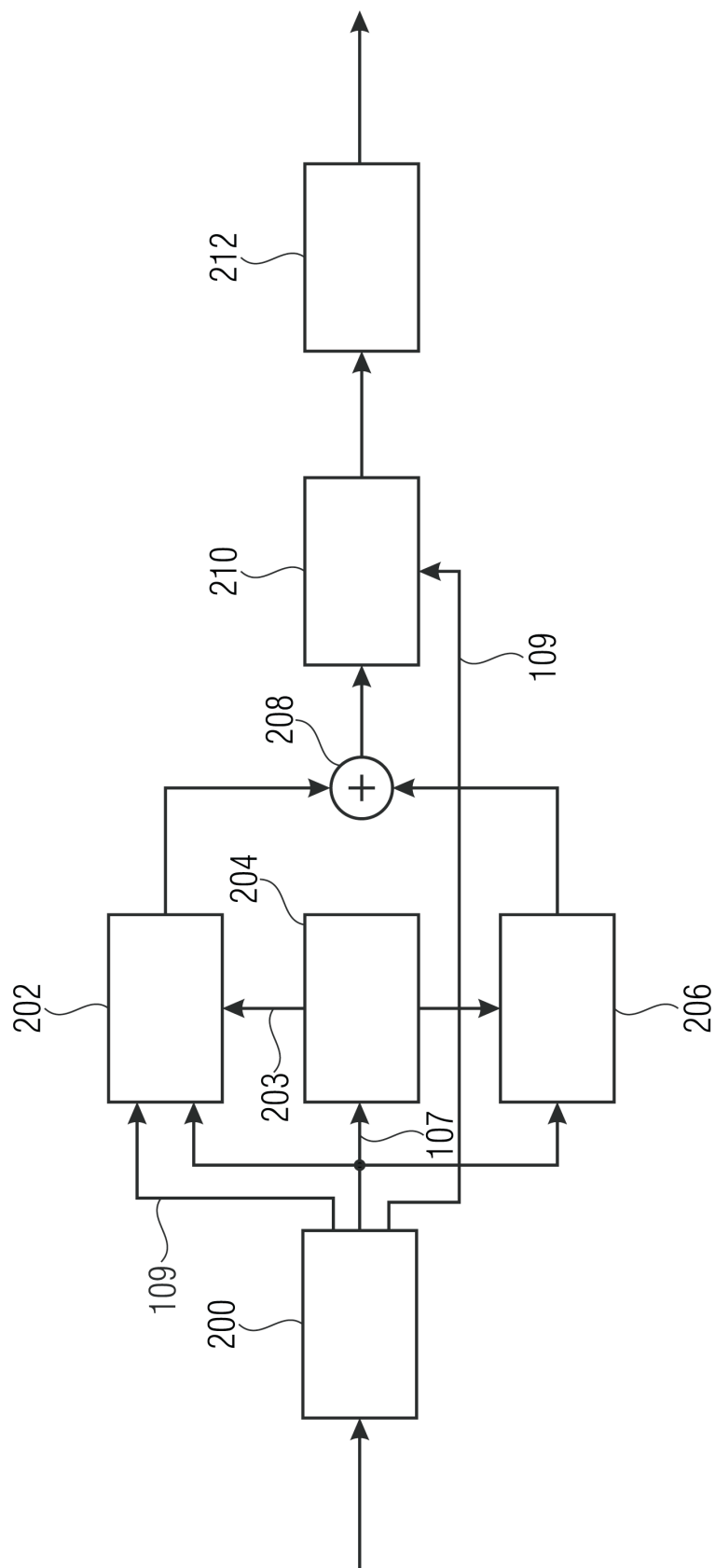


FIG 2A

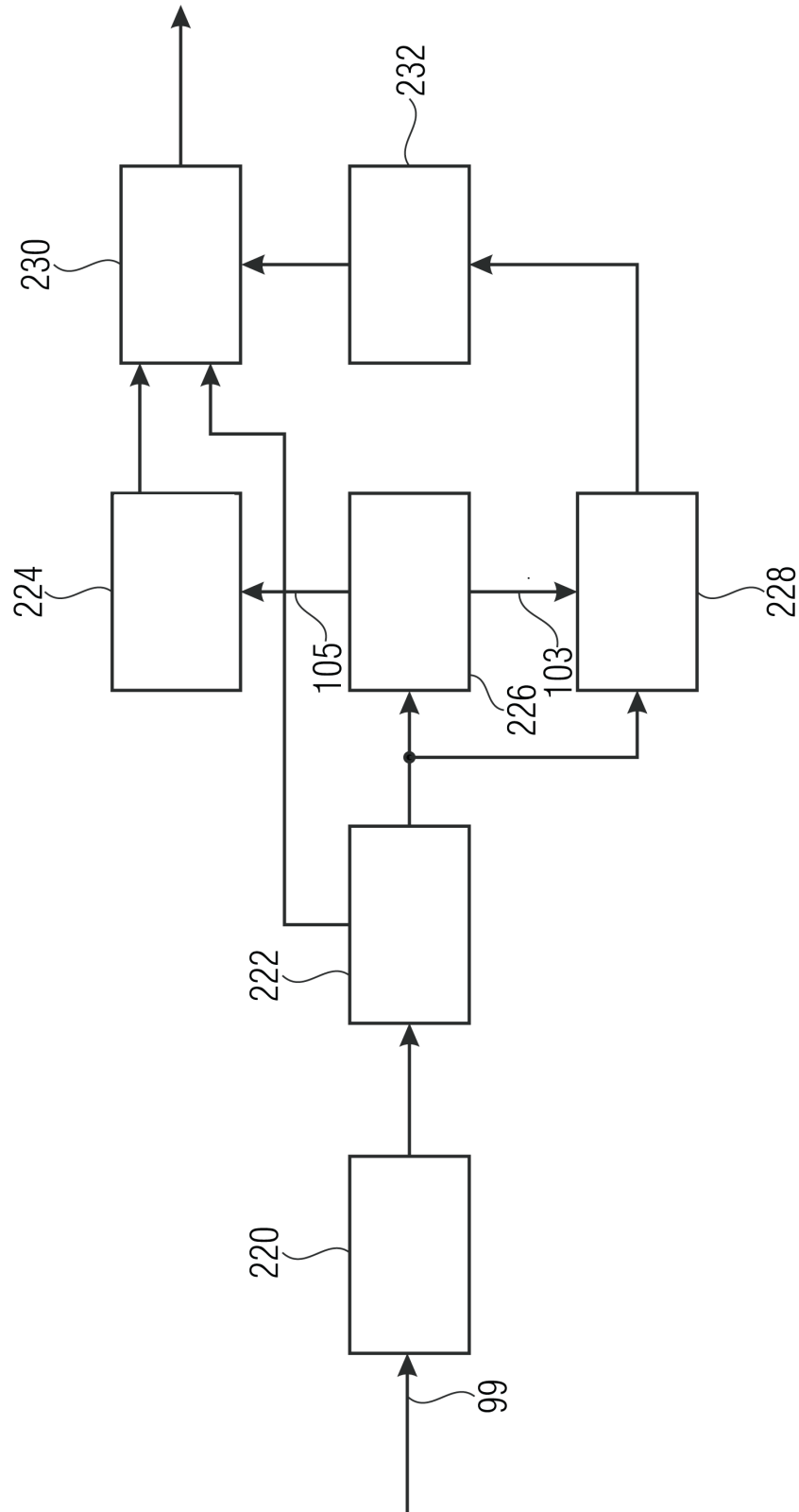


FIG 2B

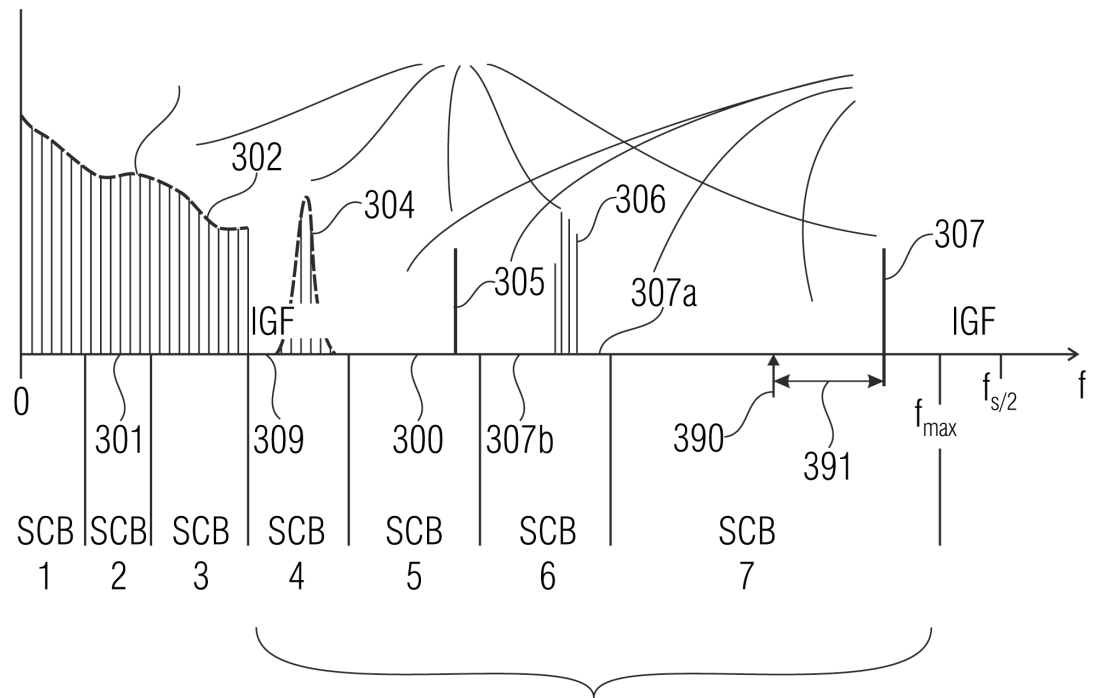


FIG 3A

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

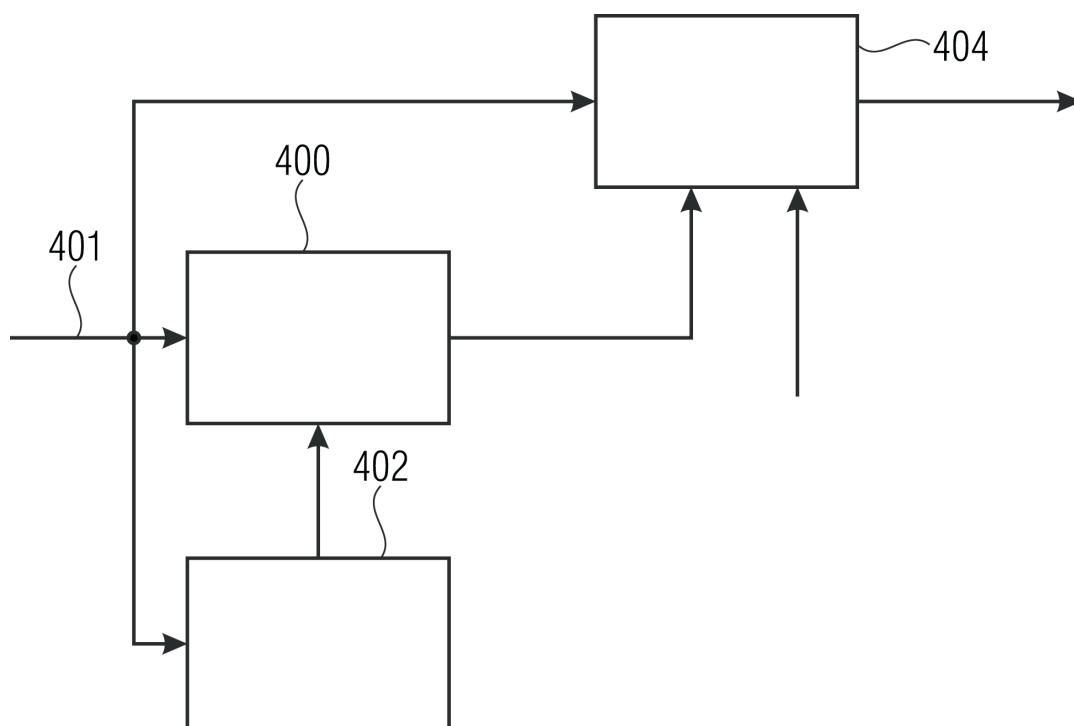


FIG 4A

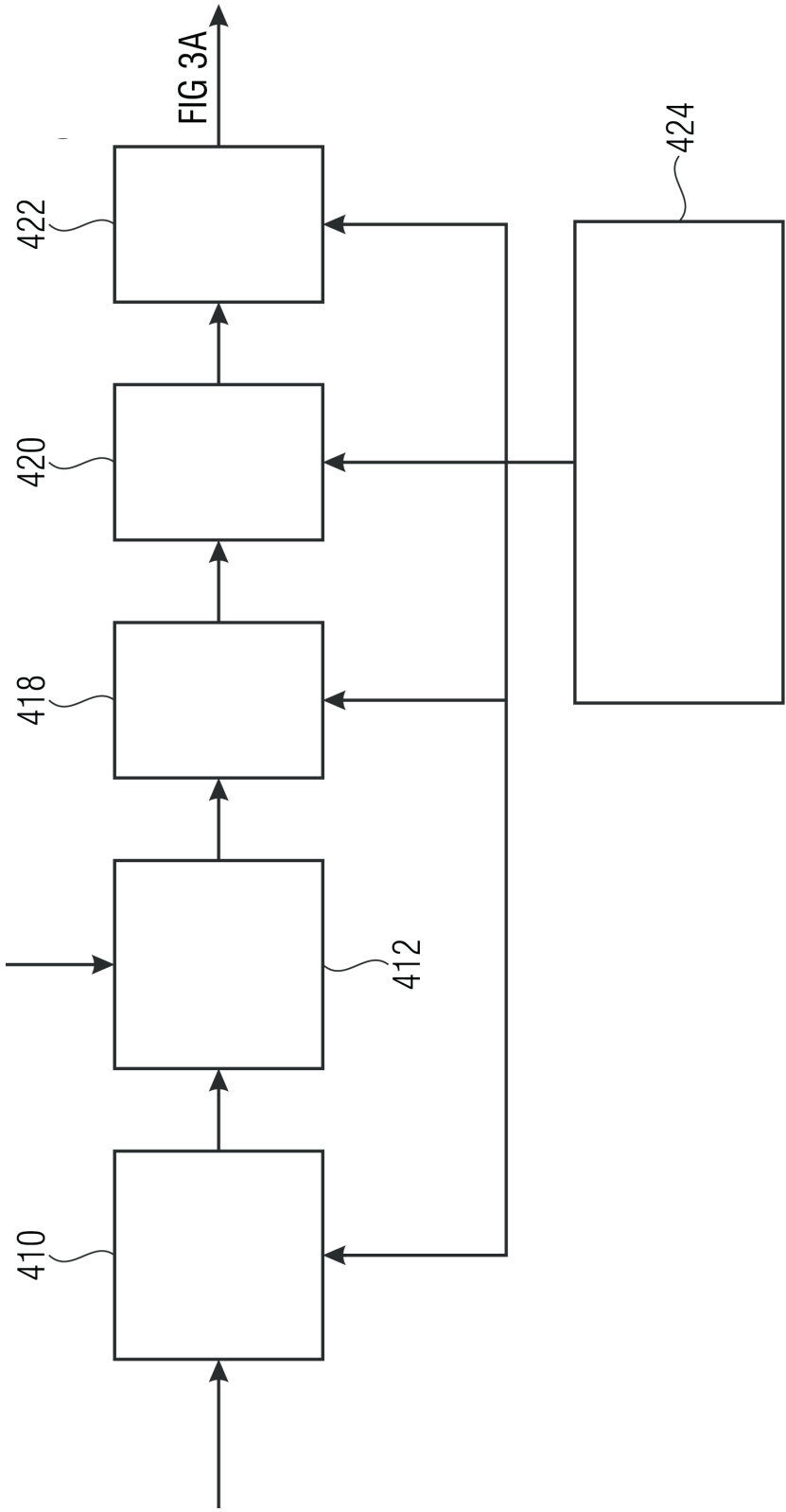


FIG 4B

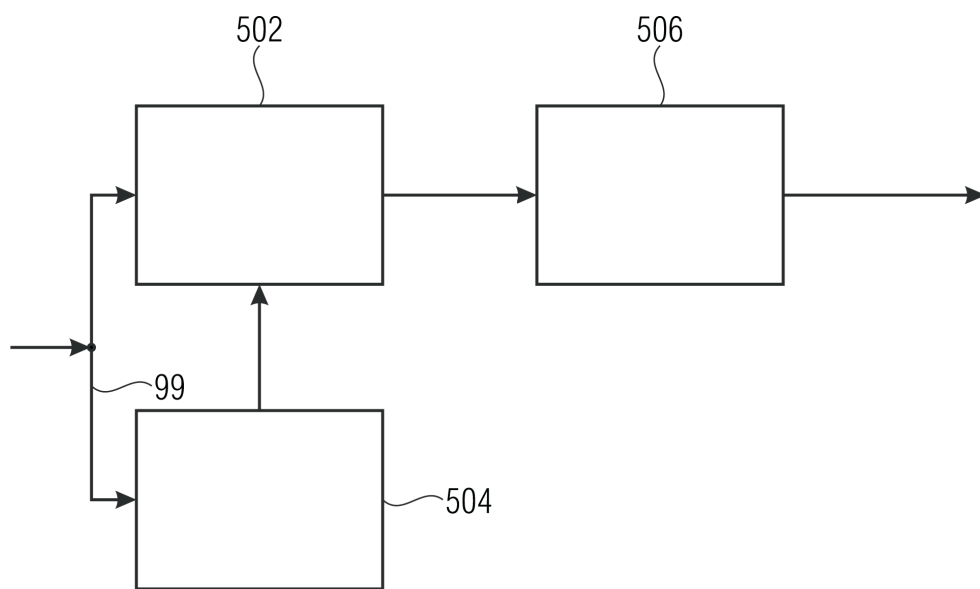


FIG 5A

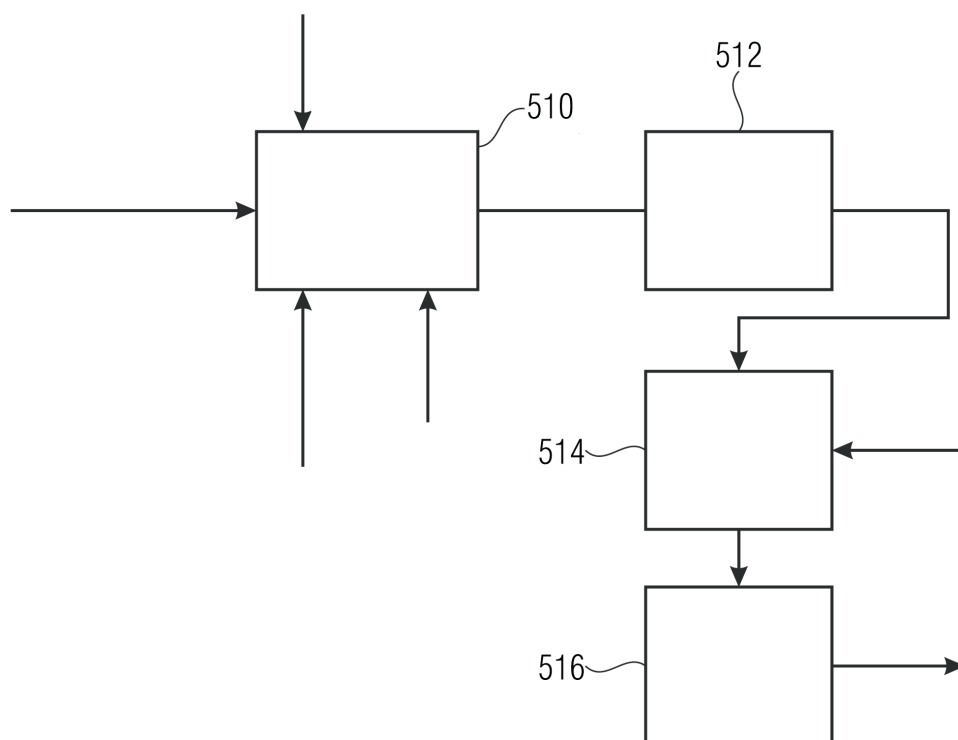


FIG 5B

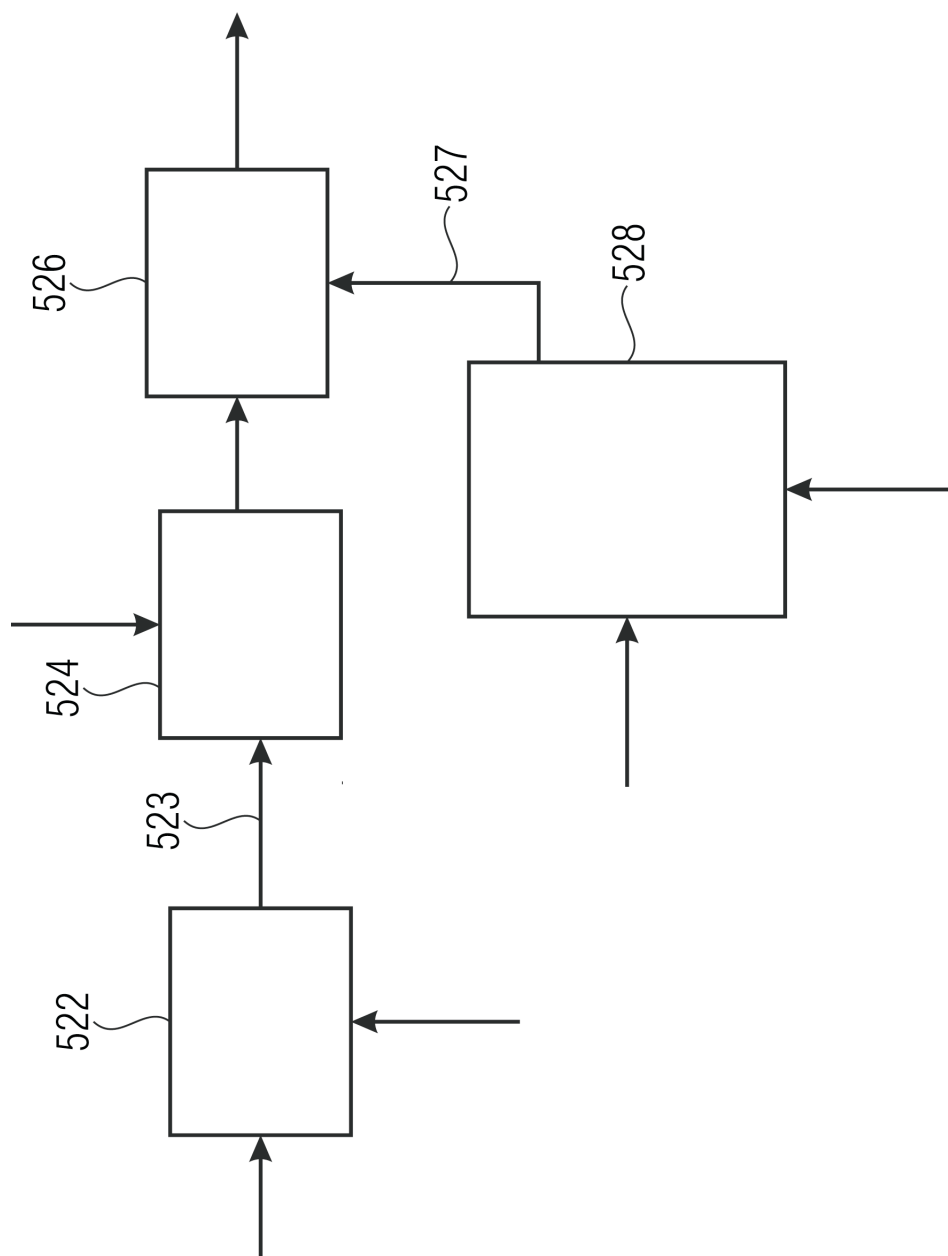


FIG 5C

10/29

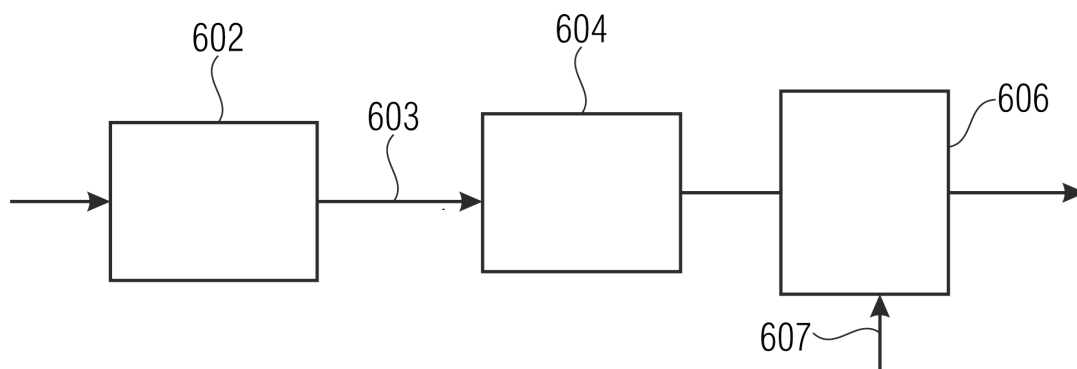


FIG 6A

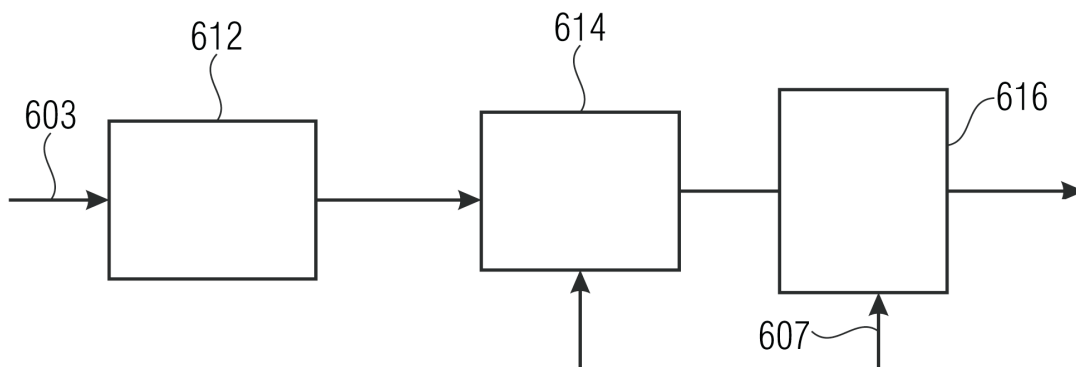


FIG 6B

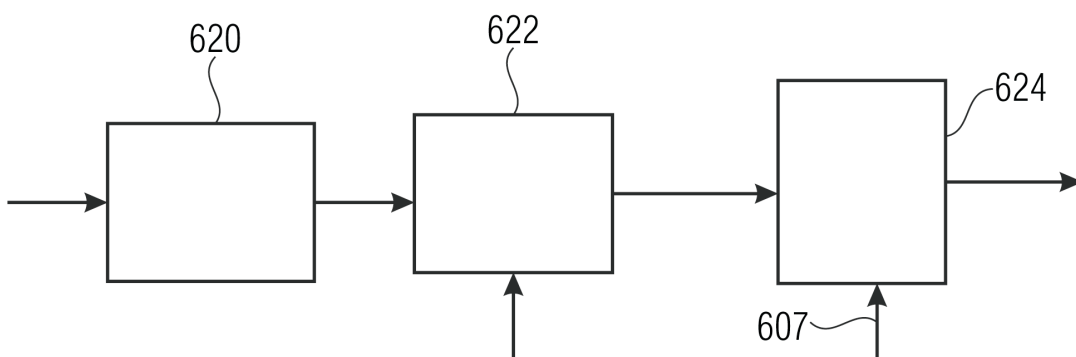


FIG 6C

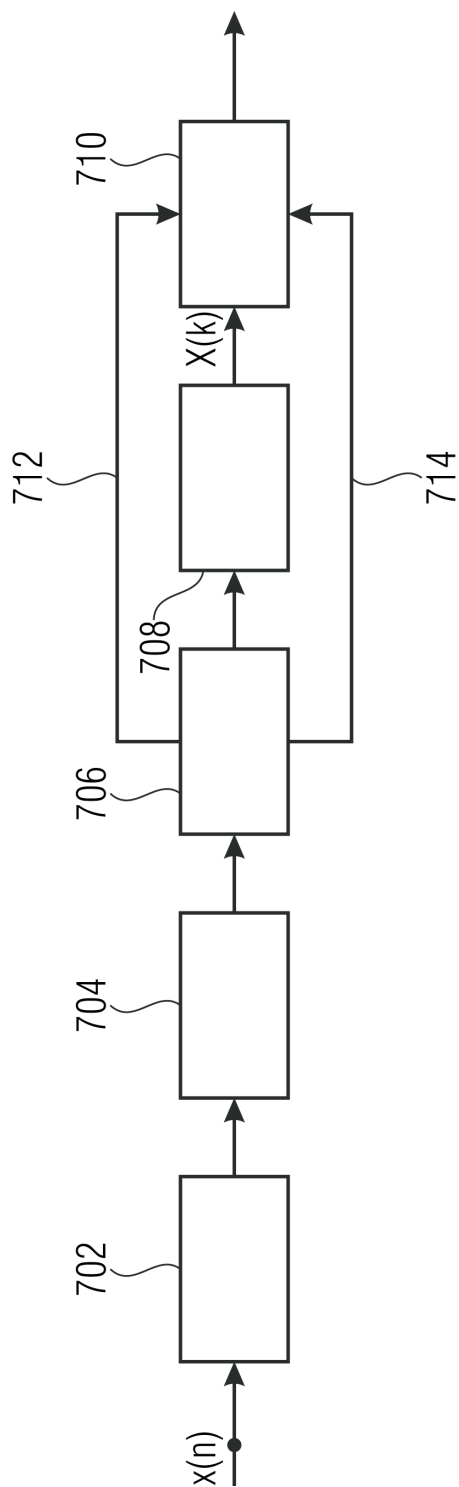


FIG 7A

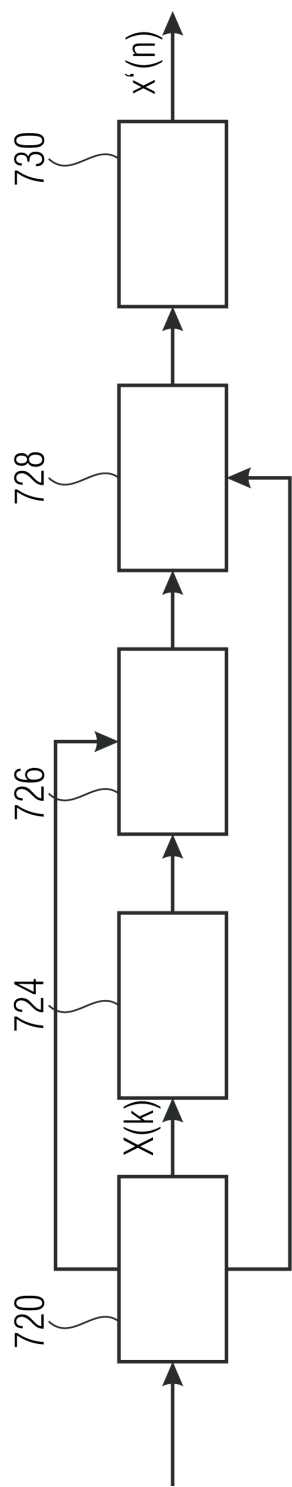


FIG 7B

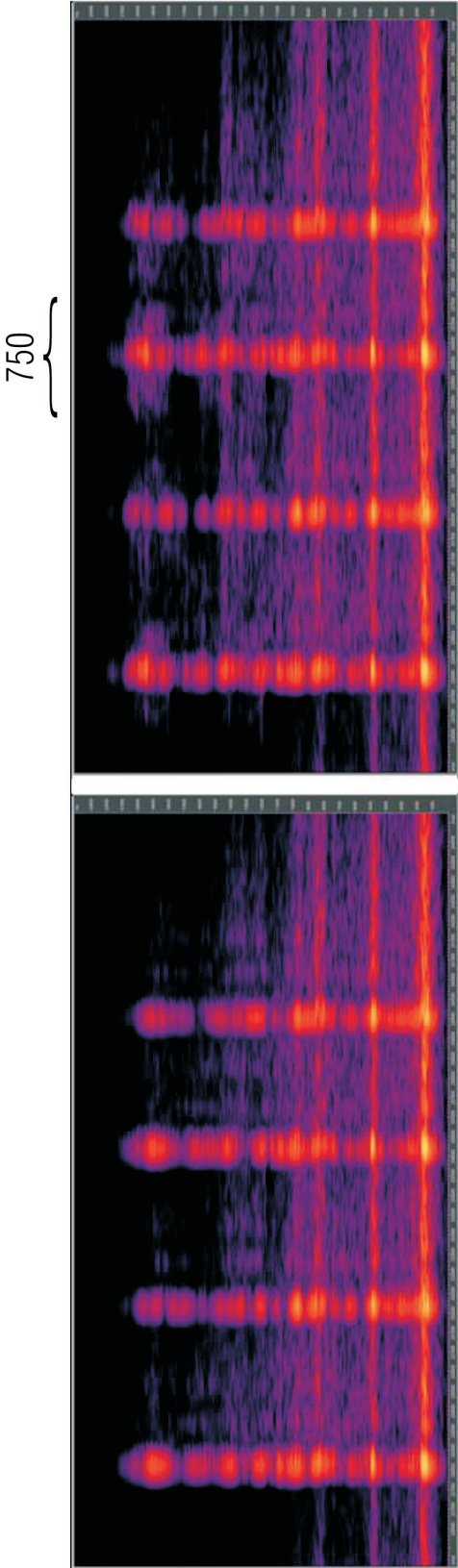


FIG 7C

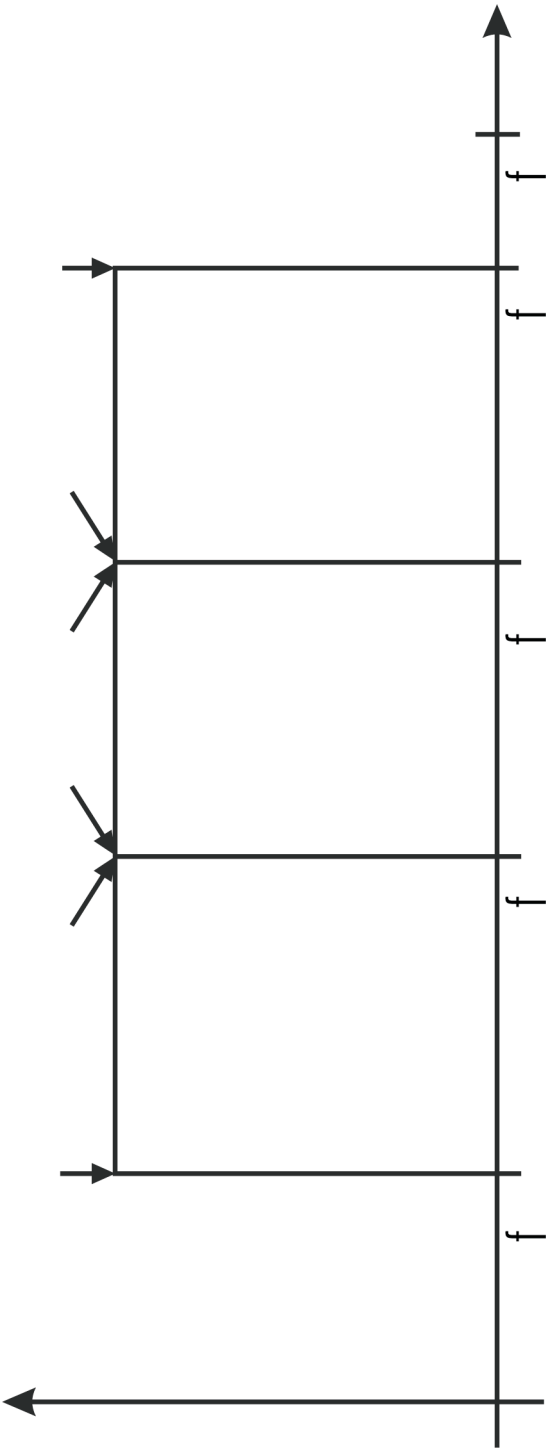


FIG 7D

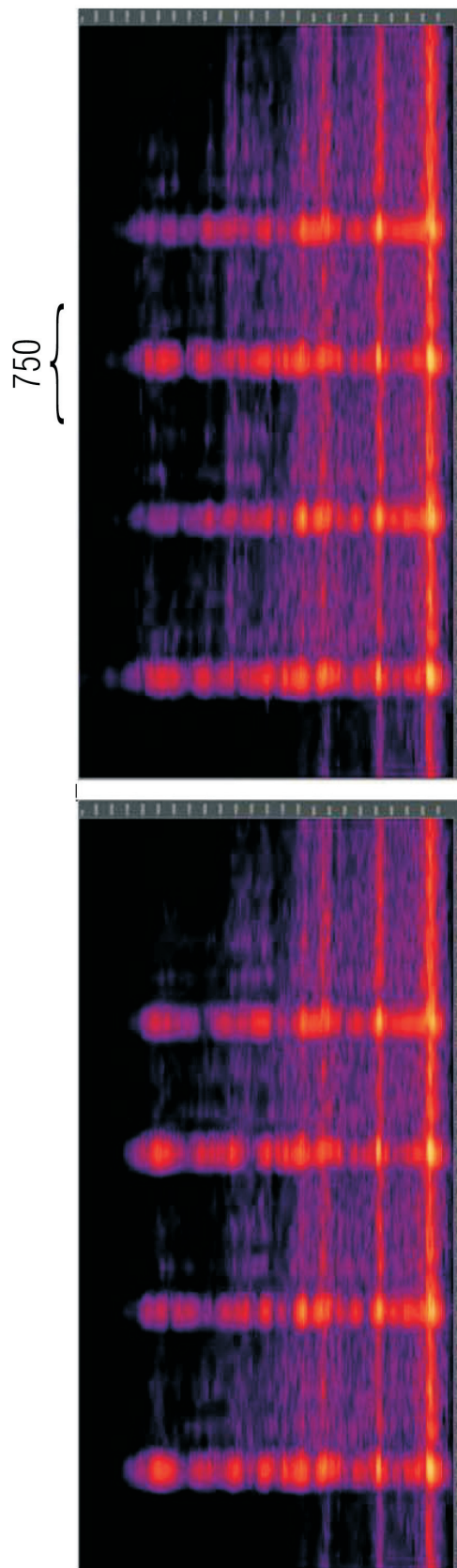


FIG 7E

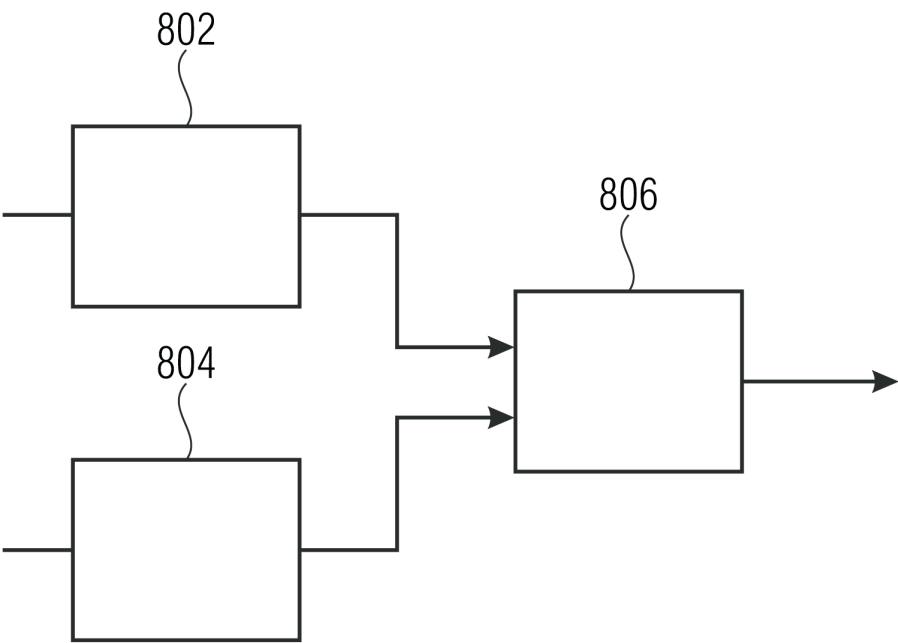


FIG 8A

	1	1	2	2	1:
	1	2	1	2	2:

FIG 8B

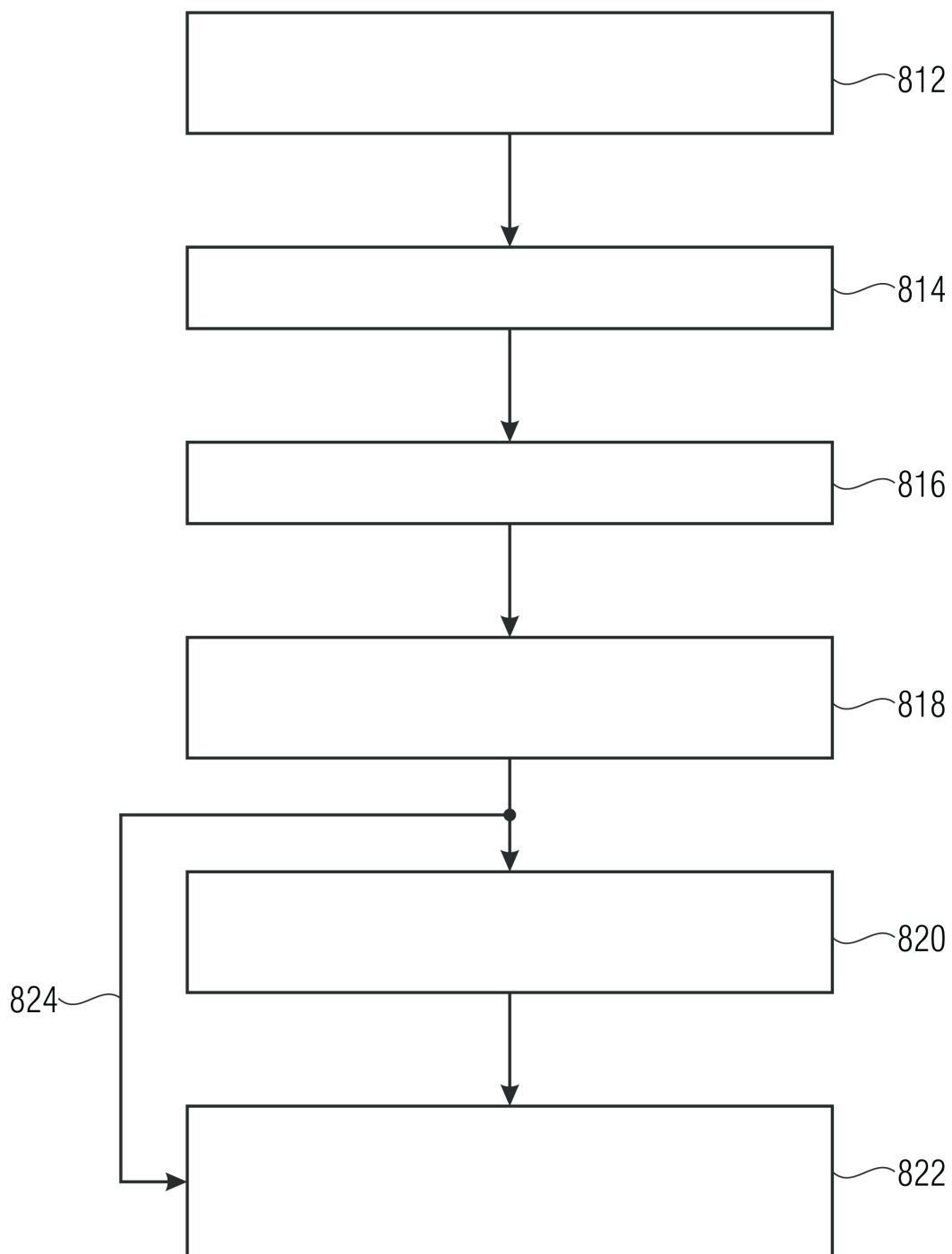


FIG 8C

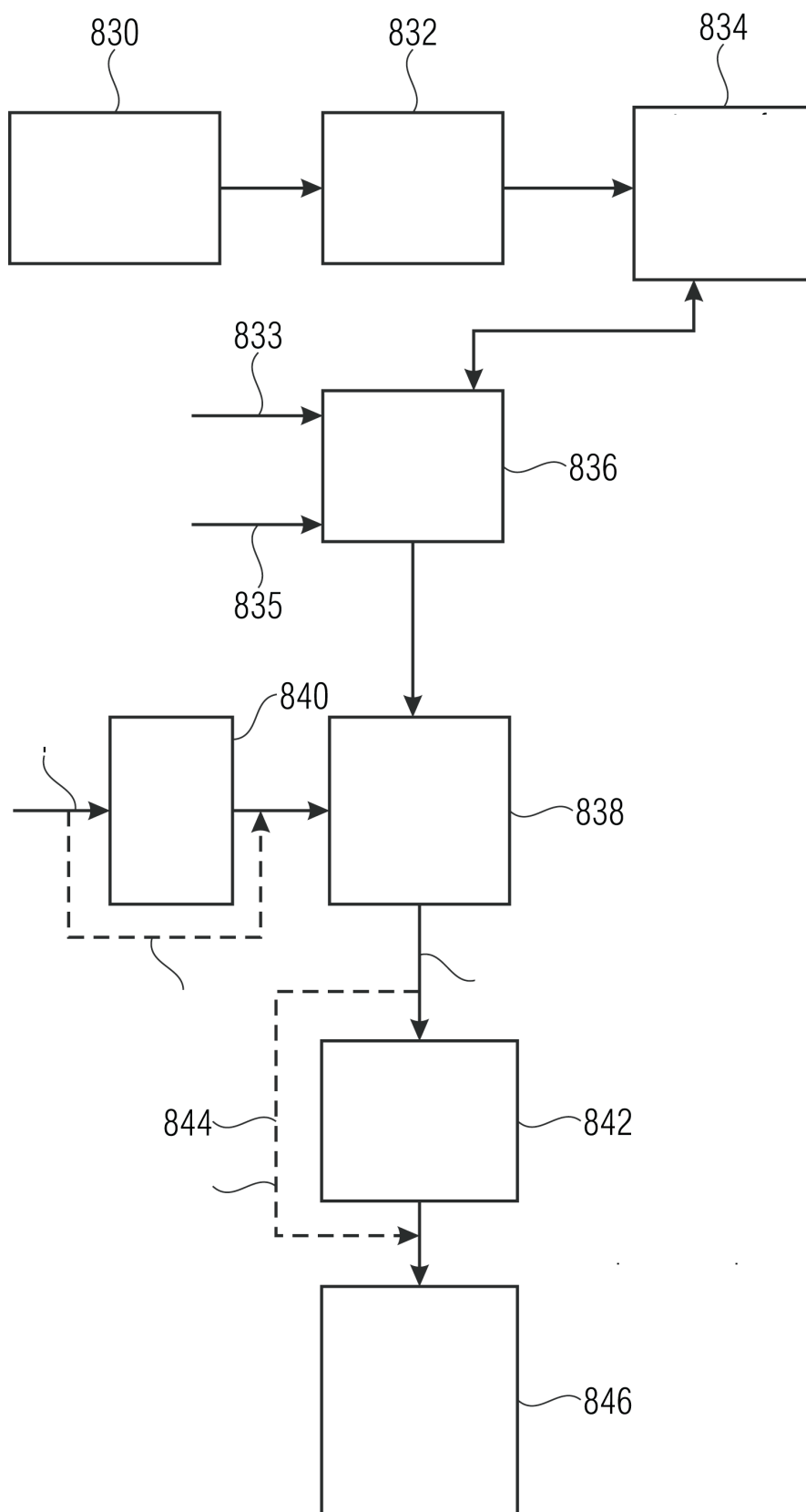


FIG 8D

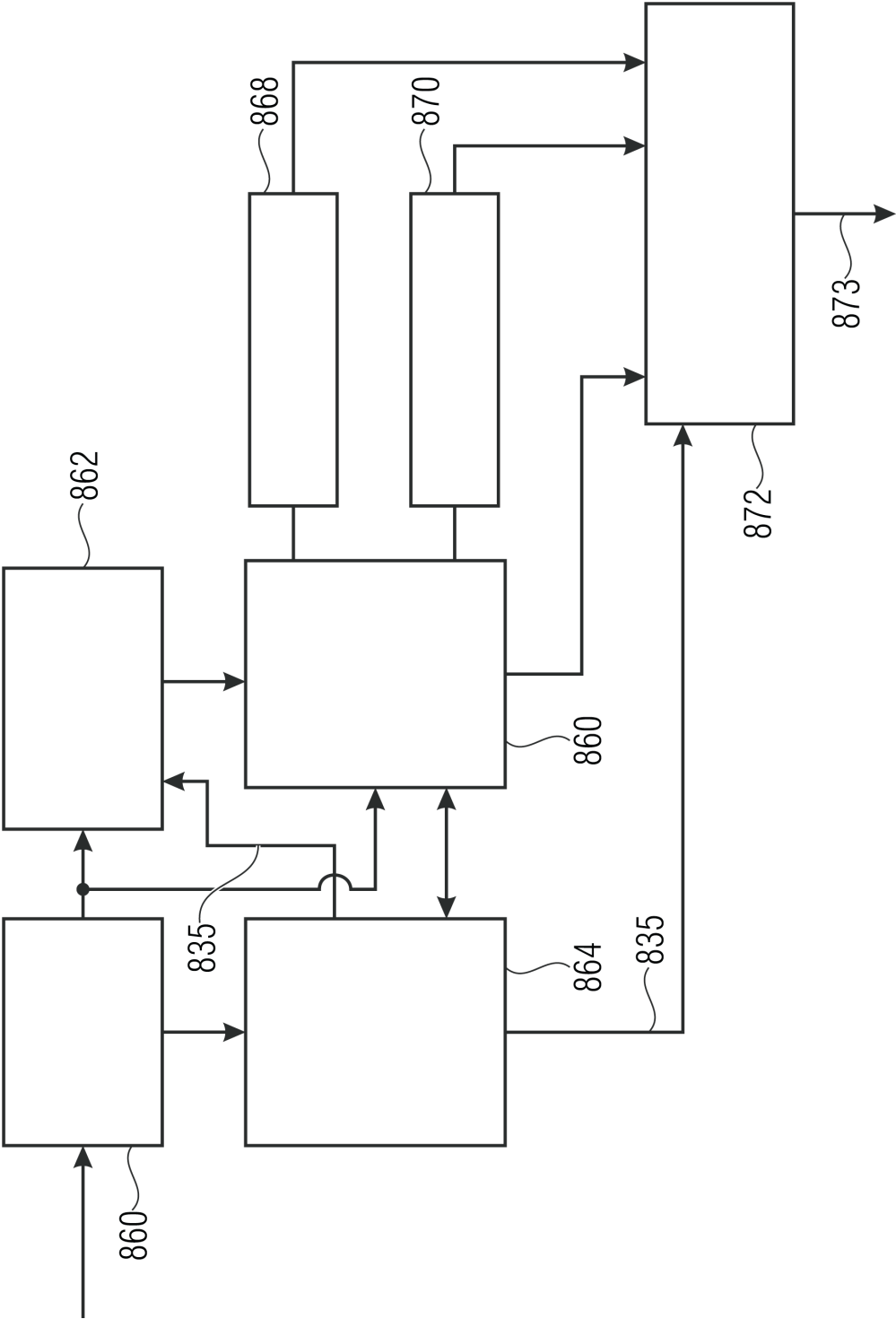


FIG 8E

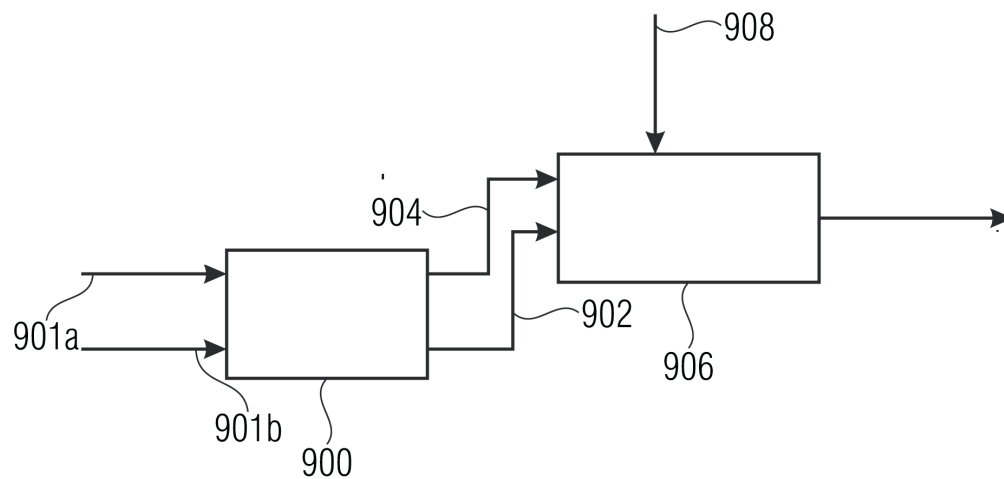


FIG 9A

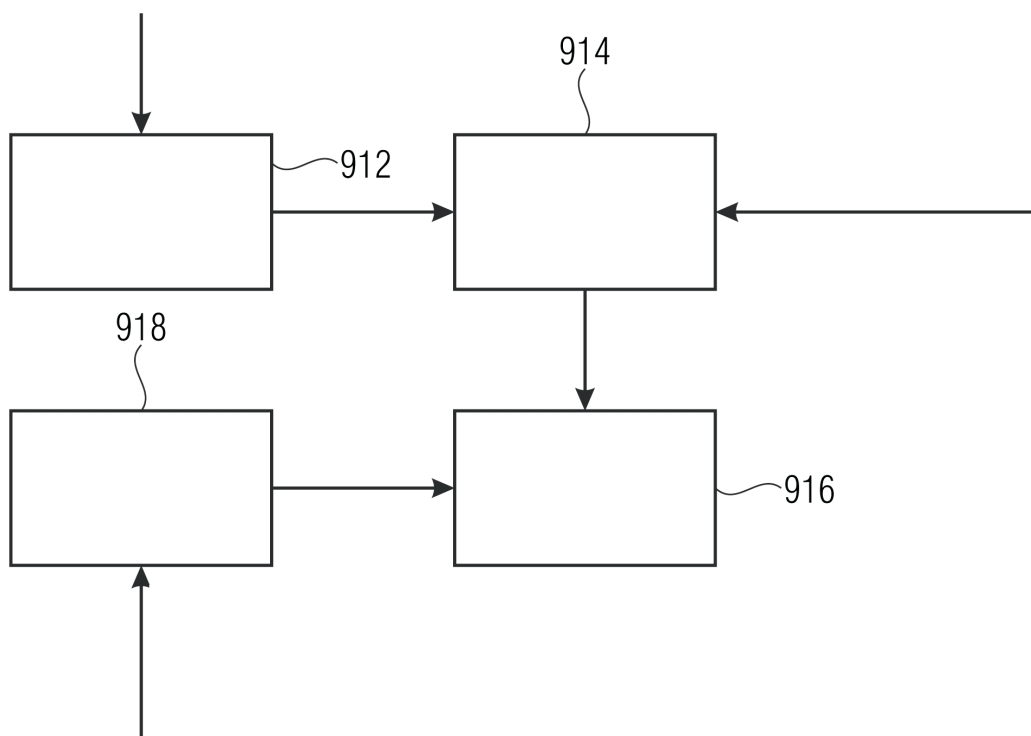


FIG 9B

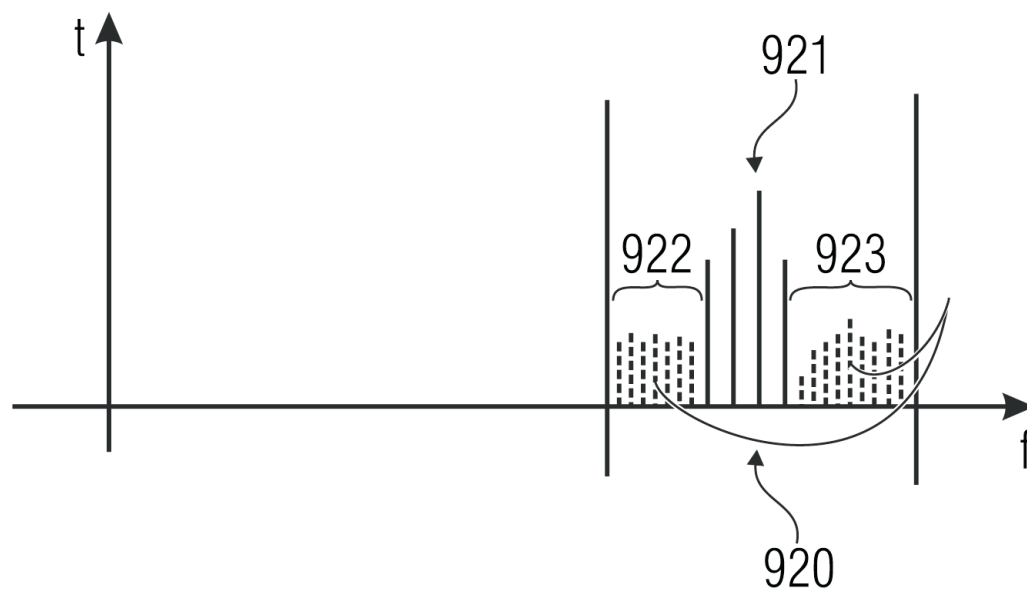


FIG 9C

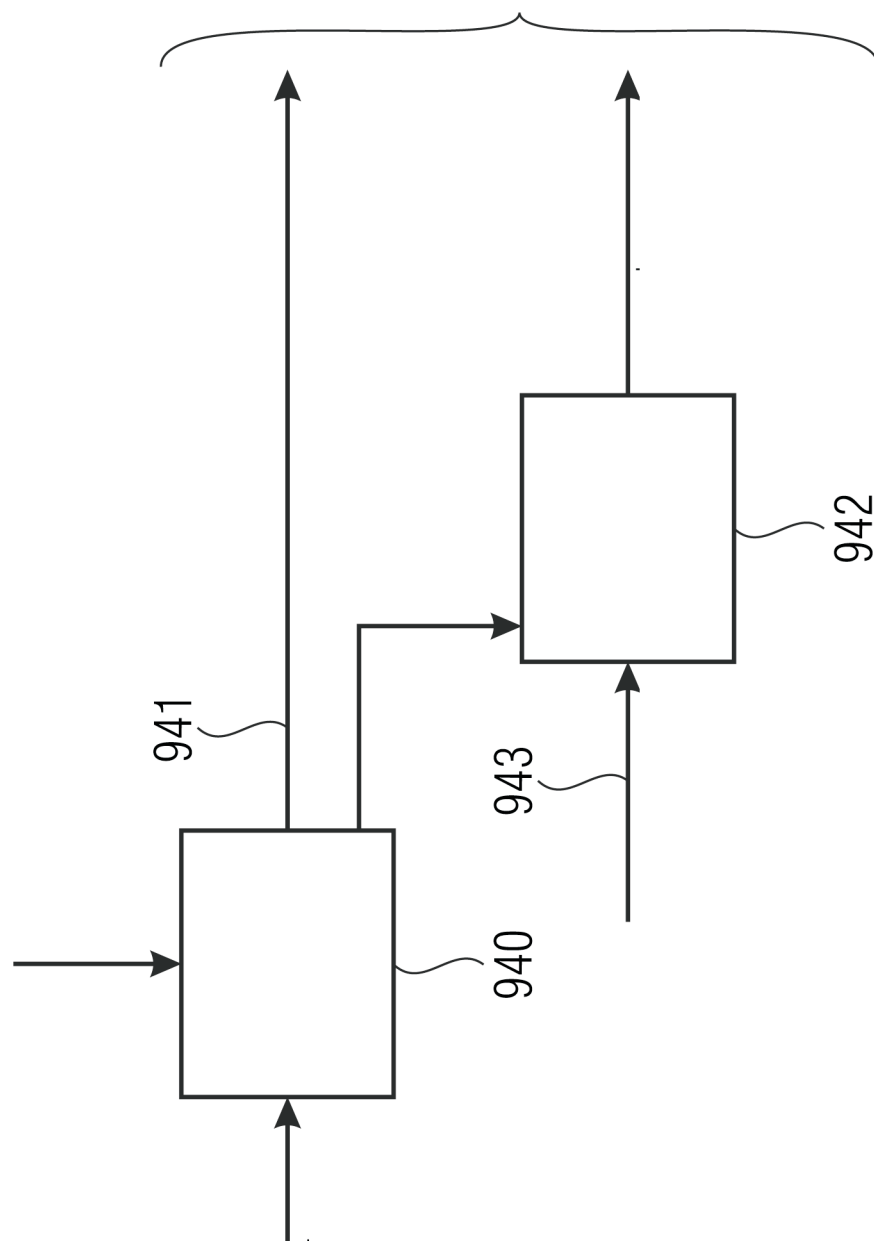


FIG 9D

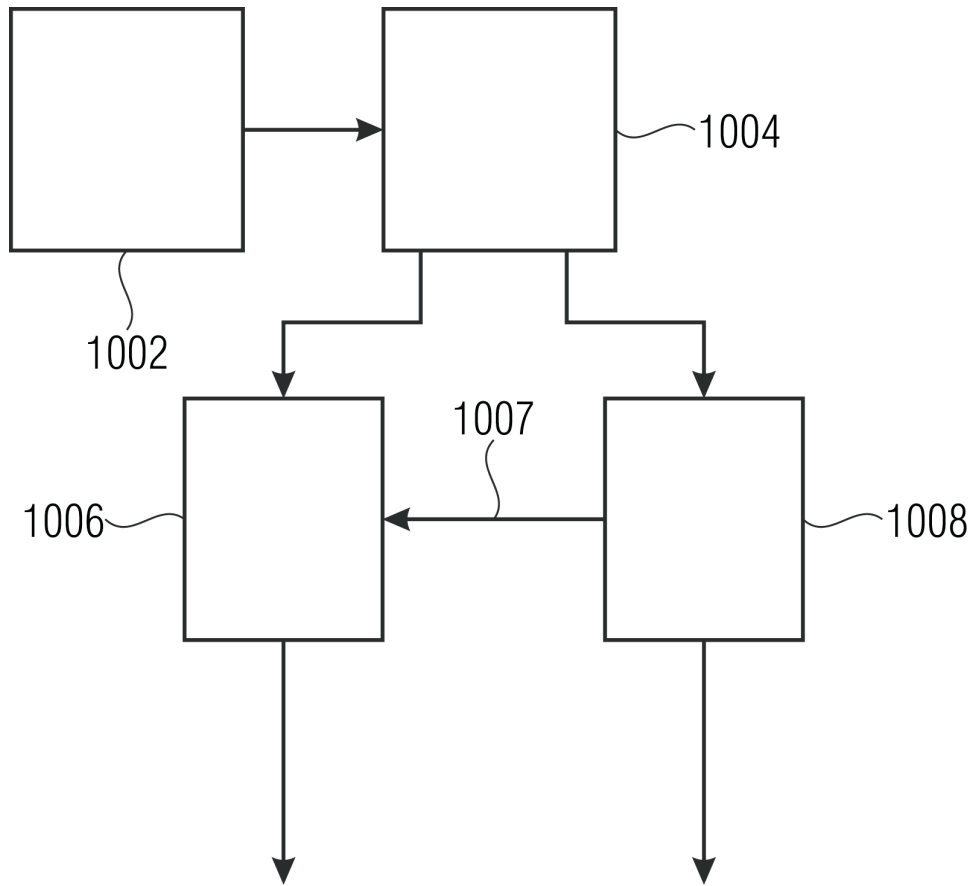


FIG 10A

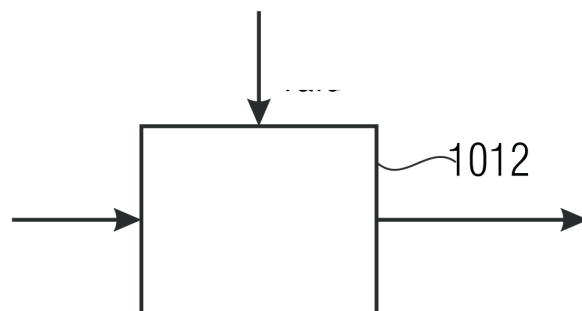


FIG 10B

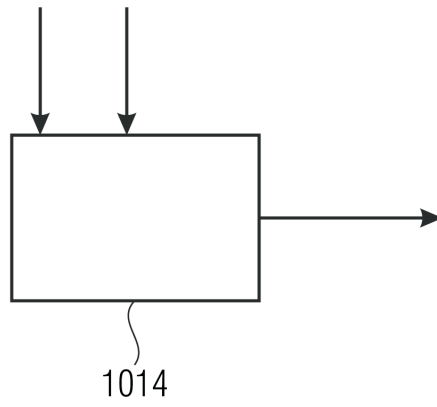


FIG 10C

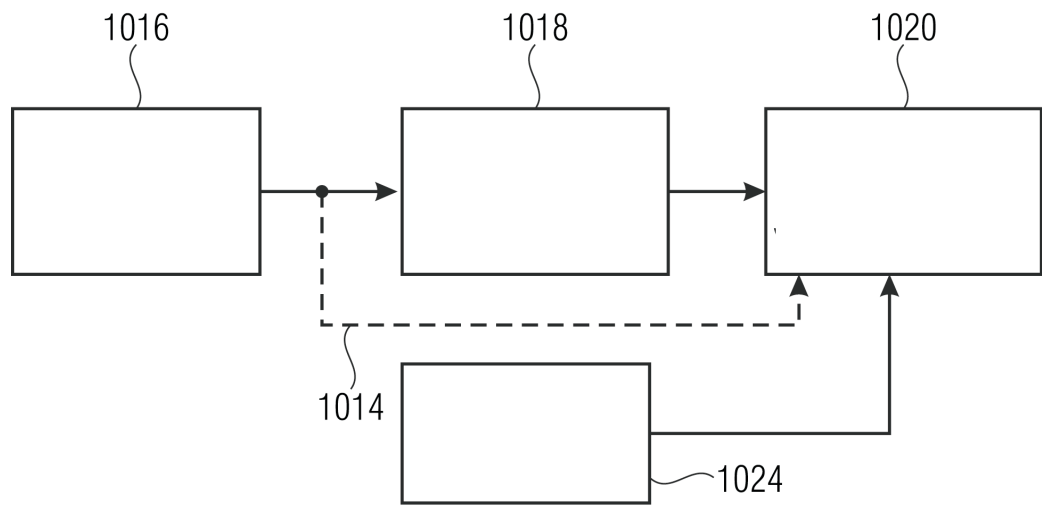


FIG 10D

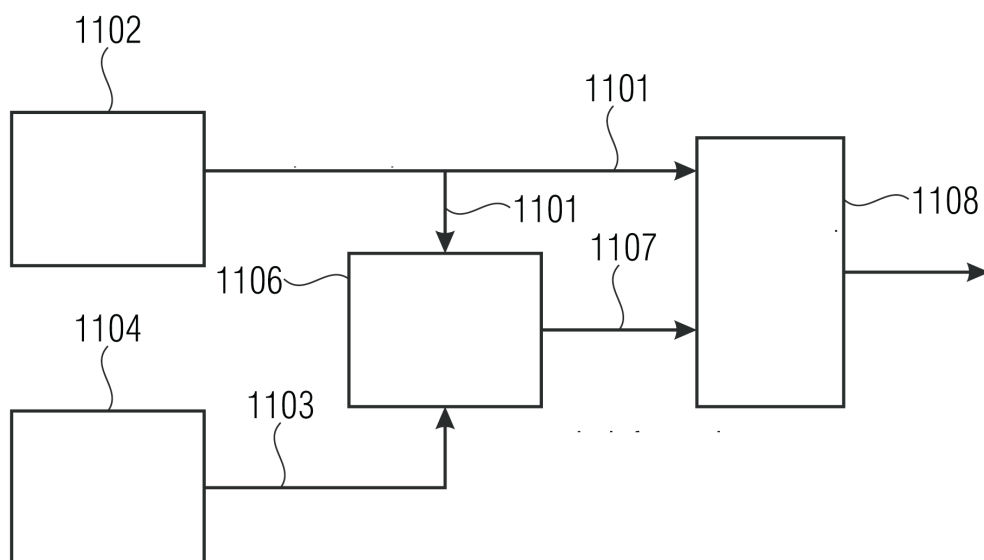


FIG 11A

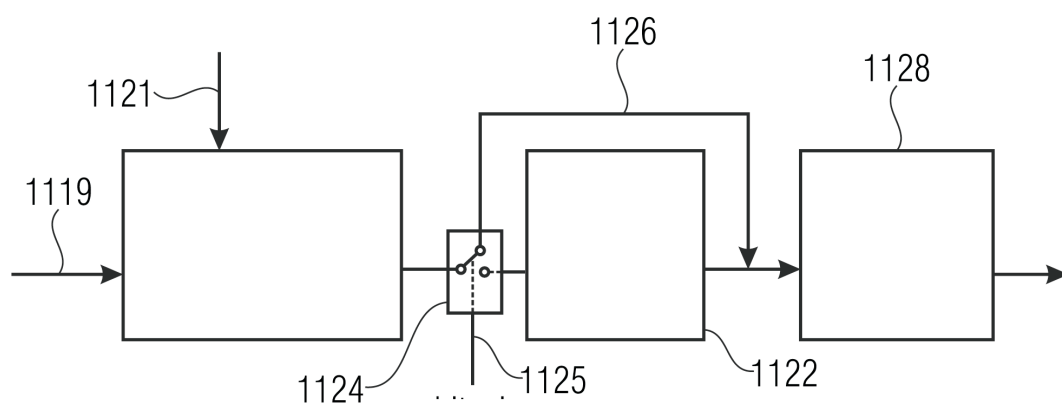


FIG 11B

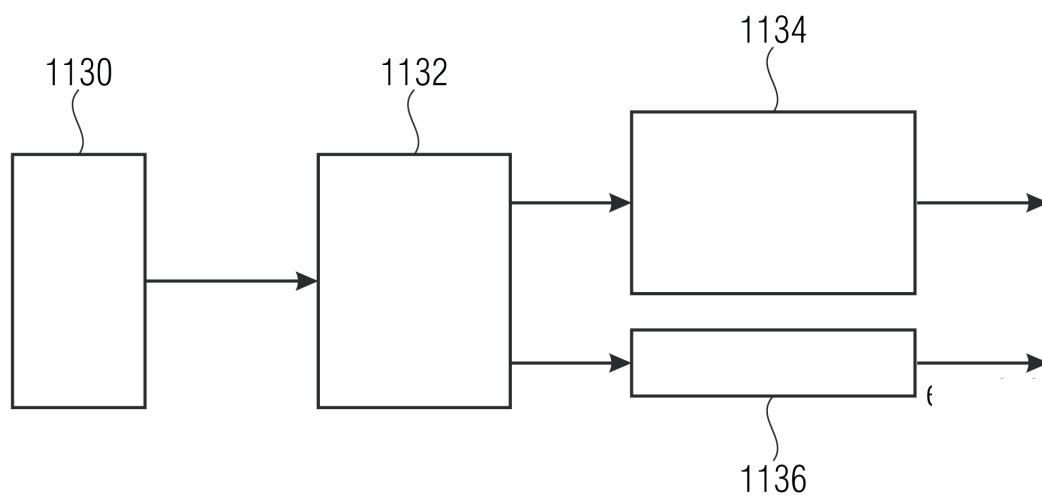


FIG 11C

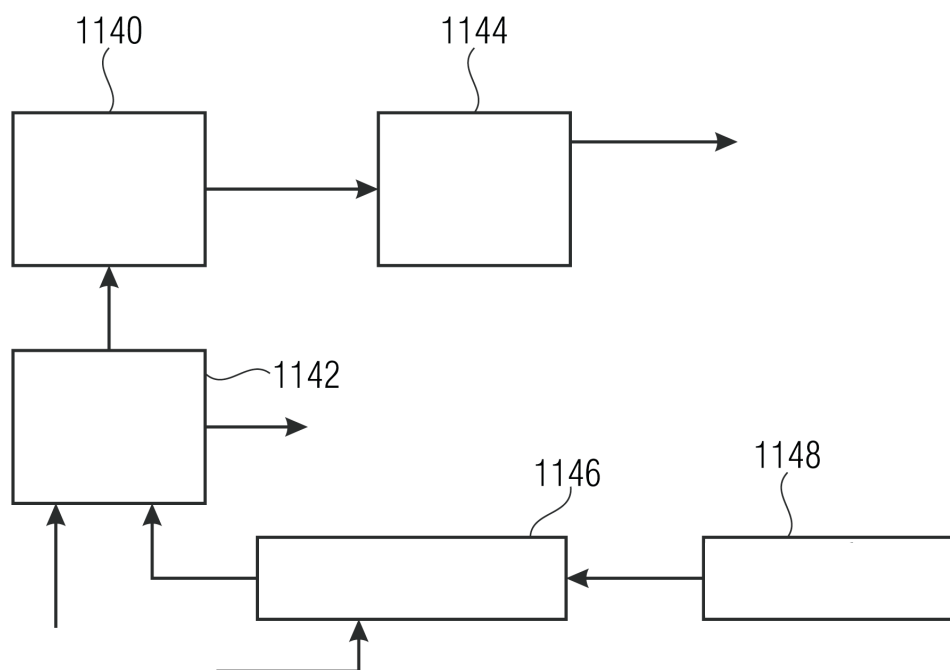


FIG 11D

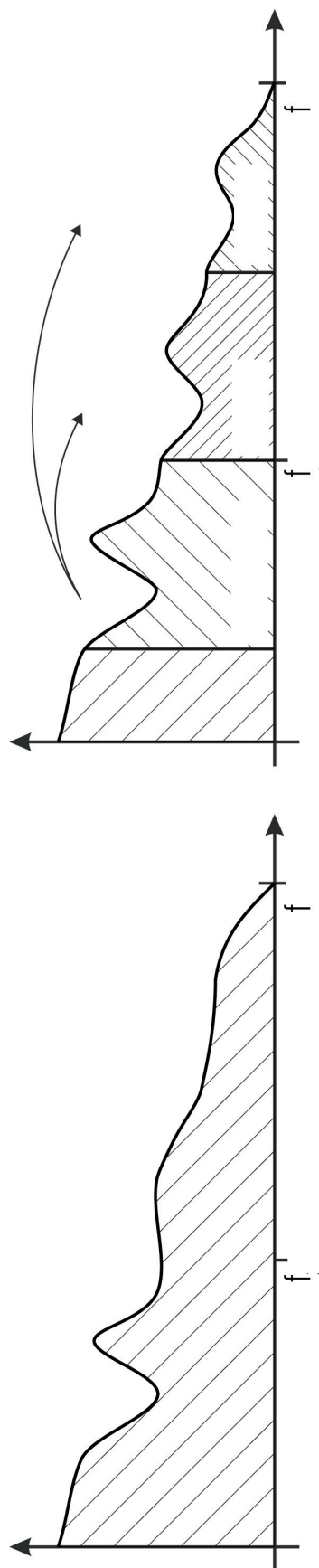


FIG 12A

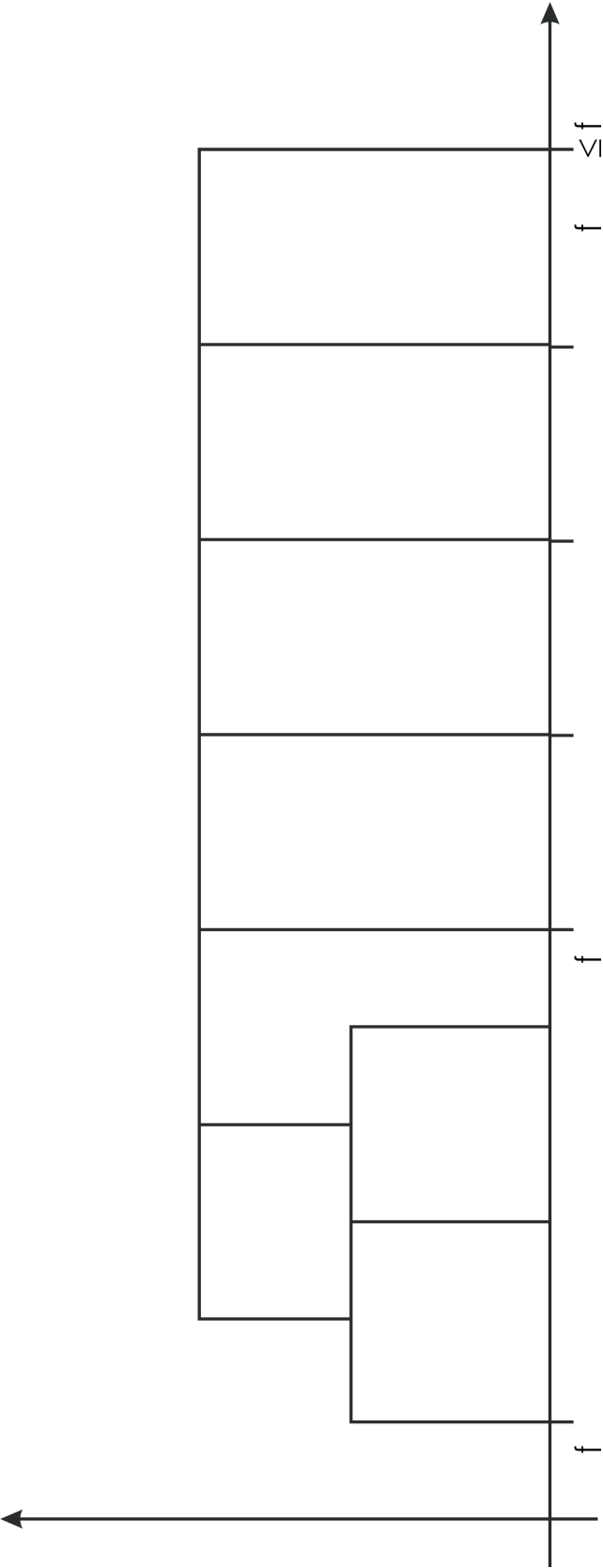


FIG 12B

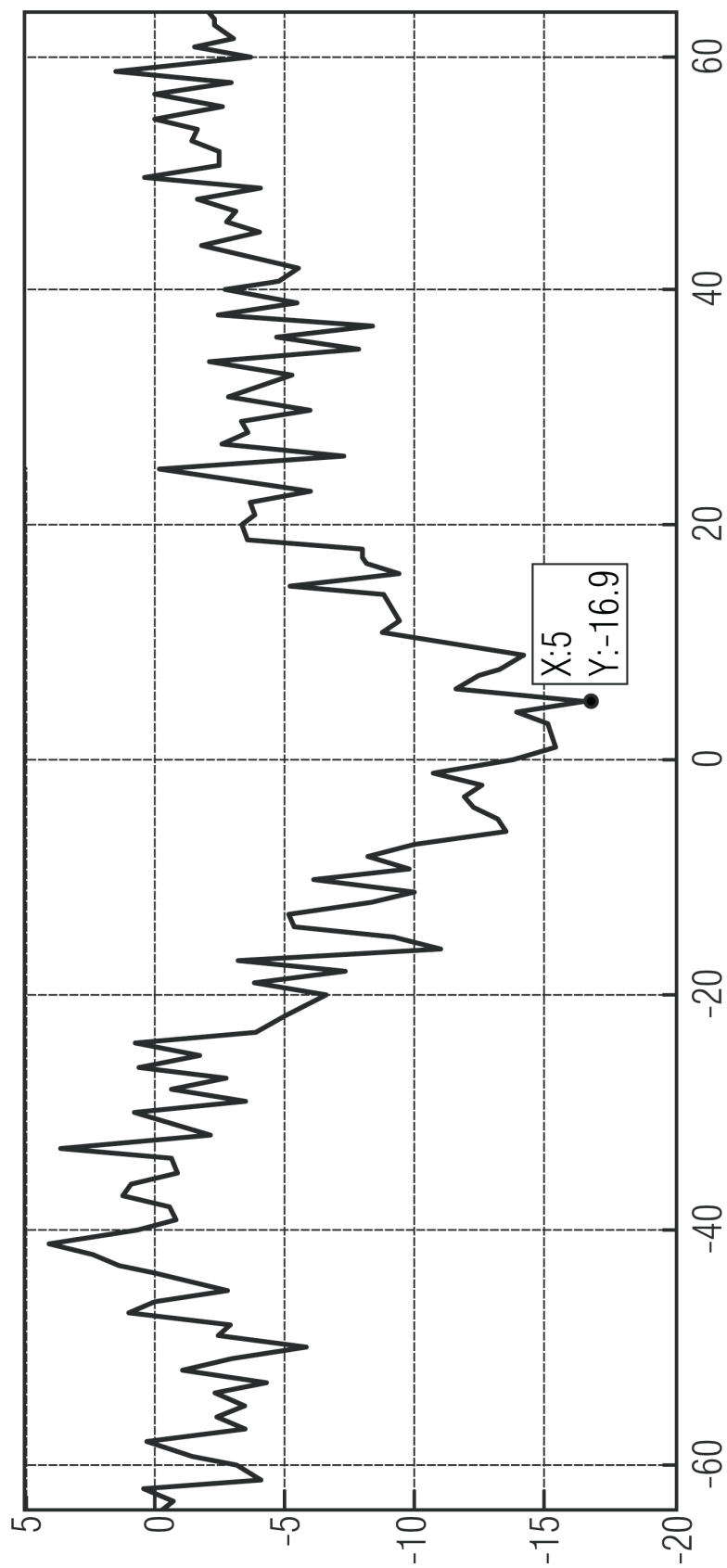


FIG 12C

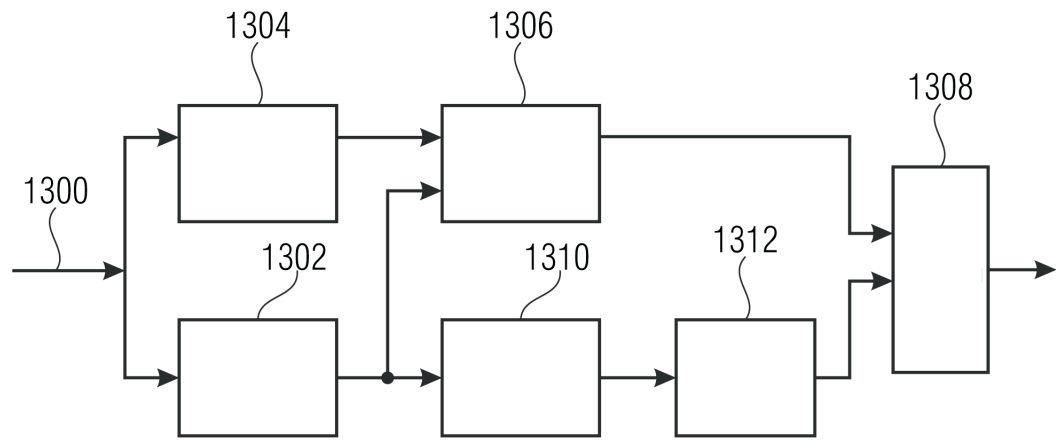


FIG 13A

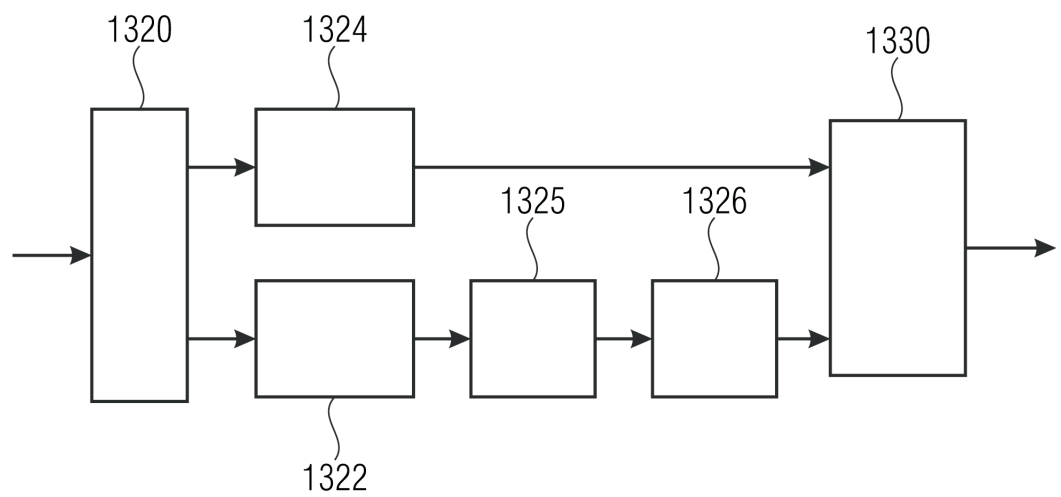


FIG 13B