



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

**(11) PI 0906079-0 B1**



**(22) Data do Depósito: 04/03/2009**

**(45) Data de Concessão: 29/12/2020**

**(54) Título:** MIXAGEM DE FLUXOS DE DADOS DE ENTRADA E GERAÇÃO DE UM FLUXO DE DADOS DE SAÍDA A PARTIR DOS MESMOS

**(51) Int.Cl.:** G10L 21/038; G10L 19/008; G10L 19/26.

**(52) CPC:** G10L 21/038; G10L 19/008; G10L 19/265.

**(30) Prioridade Unionista:** 04/03/2008 US 61/033,590.

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

**(72) Inventor(es):** SCHNELL, MARKUS; LUTZKY, MANFRED.; MULTRUS, MARKUS.

**(86) Pedido PCT:** PCT EP2009001534 de 04/03/2009

**(87) Publicação PCT:** WO 2009/109374 de 11/09/2009

**(85) Data do Início da Fase Nacional:** 03/09/2010

**(57) Resumo:** MIXAGEM DE FLUXOS DE DADOS DE ENTRADA E GERAÇÃO DE UM FLUXO DE DADOS DE SAÍDA A PARTIR DOS MESMOS. Equipamento (500) para mixagem de uma pluralidade 5 de fluxos de dados de entrada (510) é descrito, onde os fluxos de dados de entrada (510), cada um, compreende uma estrutura (540) de dados de áudio no domínio espectral, uma estrutura (540) de um fluxo de dados de entrada (510) compreendendo informações espectrais para uma pluralidade de componentes espectrais. O equipamento compreende 10 uma unidade de processamento (520) adaptada para comparar as estruturas (540) da pluralidade de fluxos de dados de entrada (510). A unidade de processamento (520) é, além disso, adaptada para determinar, corri base na comparação, para um componente espectral de uma estrutura de saída (550) de um fluxo de dados de saída (530), 15 exatamente um fluxo de dados de entrada (510) de uma pluralidade de fluxos de dados de entrada (510) . A unidade de processamento (520) é, além disso, adaptada para gerar o fluxo de dados de saída (530), ao copiar pelo menos uma parte de uma informação de um componente espectral correspondente da estrutura do fluxo de dados determinado 20 (510) de modo a descrever o componente espectral da estrutura de saída (550) do fluxo de dados de saída (530). Além disso, ou alternativamente, o valor de controle das (...).

**"MIXAGEM DE FLUXOS DE DADOS DE ENTRADA E  
GERAÇÃO DE UM FLUXO DE DADOS DE SAÍDA A PARTIR DOS MESMOS"**

**[0001] Descrição**

[0002] As configurações de acordo com a presente invenção se referem a uma mixagem de uma pluralidade de fluxos de dados de entrada para se obter um fluxo de dados de saída e gerar um fluxo de dados de saída por meio da mixagem do primeiro e do segundo fluxo de dados de entrada, respectivamente. O fluxo de dados de saída pode, por exemplo, ser usado no campo de sistemas de conferência, incluindo sistemas de vídeo conferência e sistemas de teleconferência.

[0003] Em muitas aplicações, mais de um sinal de áudio deve ser processado de tal forma que, a partir do número de sinais de áudio, um sinal, ou pelo menos um número reduzido de sinais seja gerado, o que muitas vezes é mencionado como "mixagem". O processo de mixagem de sinais de áudio, conseqüentemente, pode ser citado como agrupamento de diversos sinais individuais de áudio em um sinal resultante. Este processo é usado, por exemplo, ao se criar composições musicais para um disco compacto ("mistura de sons"). Neste caso, diferentes sinais de áudio de diferentes instrumentos junto com um ou mais sinais de áudio compreendendo performances vocais (cantos), são normalmente mixadas em uma canção.

[0004] Outros campos de aplicação, onde a mixagem representa um papel importante, são sistemas de videoconferência

e sistemas de teleconferência. Tais sistemas são normalmente capazes de conectar diversos participantes distribuídos no espaço em uma conferência por meio do uso de um servidor central, que adequadamente mixa os dados de entrada de vídeo e áudio dos participantes registrados e envia a cada um dos participantes um sinal resultante, como retorno. Este sinal resultante, ou sinal de saída, compreende os sinais de áudio de todos os outros participantes da conferência.

[0005] Nos modernos sistemas digitais de conferência, um número de metas e aspectos parcialmente contraditórios, competem entre si. A qualidade de um sinal de áudio reconstruído, assim como a aplicabilidade e a utilidade de algumas técnicas de codificação e decodificação para diferentes tipos de sinais de áudio (por exemplo, sinais de fala, comparados com sinais gerais de áudio e sinais musicais), devem ser levados em consideração. Outros aspectos que também precisam ser considerados ao projetar e implementar sistemas de conferência são a largura da banda disponível e questões de atraso.

[0006] Por exemplo, ao comparar a qualidade por um lado, e a largura de banda por outro lado, um compromisso, na maioria das vezes, é inevitável. No entanto, melhorias relacionadas à qualidade podem ser obtidas pela implementação de modernas técnicas de codificação e decodificação, tais como, a técnica AAC-ELD (AAC = Codec Avançado de Áudio; ELD = Baixo

Atraso Otimizado). No entanto, a qualidade alcançável pode ser negativamente afetada em sistemas usando tais técnicas modernas em função de mais problemas e aspectos fundamentais.

[0007] Para citar apenas um desafio a ser vencido, todas as transmissões de sinais digitais enfrentam o problema de uma quantização necessária, que pode, pelo menos a princípio, ser evitada sob circunstâncias ideais em um sistema análogo livre de ruídos. Devido ao processo de quantização, inevitavelmente uma determinada quantidade de ruído de quantização é introduzido no sinal a ser processado. Para contrabalançar possíveis e audíveis distorções, podemos ficar tentados a aumentar o número de níveis de quantização e, conseqüentemente, aumentar a resolução da quantização de acordo. Isto, no entanto, leva a um maior número de valores de sinal a serem transmitidos, e, conseqüentemente, a um aumento da quantidade de dados a serem transmitidos. Em outras palavras, a melhoria da qualidade por meio da redução de possíveis distorções introduzidas pelos ruídos de quantização, podem, em determinadas circunstâncias, aumentar a quantidade de dados a serem transmitidos e pode eventualmente violar restrições de largura de banda impostas a um sistema de transmissão.

[0008] No caso dos sistemas de conferência, os desafios em melhorar o *tradeoff* entre qualidade, largura de banda disponível e outros parâmetros pode ser ainda mais complicado pelo fato de que normalmente mais de um sinal de

entrada de áudio deve ser processado. Conseqüentemente, condições limítrofes impostas por mais de um sinal de áudio podem precisar ser levadas em consideração ao gerar o sinal de saída ou resultar o sinal produzido pelo sistema de conferência.

[0009] Principalmente em vista dos desafios adicionais ao implementar sistemas de conferência com um atraso suficientemente baixo para permitir uma comunicação direta entre os participantes de uma conferência sem a introdução de atrasos substanciais que possam ser considerados inaceitáveis pelos participantes, isto aumenta ainda mais o desafio.

[00010] Nas implementações de baixo atraso em sistemas de conferência, fontes de atraso são normalmente restritas em termos de número, o que por outro lado pode levar ao desafio do processamento de dados fora do domínio de tempo, no qual a mixagem de sinais de áudio pode ser obtida por superposição ou adição de respectivos sinais.

[00011] Genericamente falando, é favorável escolher um *tradeoff* entre qualidade, largura de banda disponível e outros parâmetros adequados para sistemas de conferência, de forma cuidadosa, no sentido de competir com a codificação extra de processamento para mixagem em tempo real, diminuir a quantidade de hardware necessária, e manter razoáveis os custos em termos de hardware e codecs extras para transmissão, sem comprometer a qualidade de áudio.

[00012] Para reduzir uma quantidade de dados

transmitidos, os modernos codecs de áudio muitas vezes utilizam ferramentas altamente sofisticadas para descrever informações espectrais em relação a componentes espectrais de um respectivo sinal de áudio. Utilizando tais ferramentas que são baseadas em fenômenos psico-acústicos e resultados de exames, um *tradeoff* melhorado entre parâmetros parcialmente contraditórios e condições limítrofes, tais como, a qualidade do sinal de áudio reconstruído a partir dos dados transmitidos, complexidade computacional, taxa de bits, e outros parâmetros, podem ser obtidos.

[00013] Exemplos de tais ferramentas são, por exemplo, substituição de ruídos perceptíveis (PNS), ajuste de ruídos temporais (TNS), e replicação de banda espectral (SBR), para citar apenas alguns. Todas estas técnicas são baseadas na descrição de pelo menos uma parte de informações espectrais com um número reduzido de bits, de forma que, em comparação a um fluxo de dados baseado em não usar estas ferramentas, mais bits podem ser alocados para partes espectrais importantes do espectro. Como consequência, enquanto é mantida a taxa de bits, um perceptível nível de qualidade pode ser melhorado por meio do uso de tais ferramentas. Naturalmente, um *tradeoff* diferente pode ser selecionado, ou seja, para reduzir o número de bits transmitido por estrutura de dados de áudio contendo e impressão global de áudio. Diferentes *tradeoffs* localizados entre estes dois extremos, também podem ser igualmente bem obtidos.

[00014] Estas ferramentas também podem ser usadas em aplicativos de telecomunicações. No entanto, quando mais de dois participantes em tal situação de comunicação estão presentes, poderá ser bastante vantajoso usar um sistema de conferência para mixar dois ou mais fluxos de bits de mais de dois participantes. Situações como estas podem ocorrer em situações puramente baseadas em áudio ou em situações de teleconferência, assim como, em situações de vídeo-conferências.

[00015] Um sistema de conferência operando em um domínio de frequência é, por exemplo, descrito em US 2008/0097764 A1 que realiza a mixagem atual no domínio de frequência, desta forma, omitindo a re-transformação dos sinais de áudio de volta ao domínio de tempo.

[00016] No entanto, o sistema de conferência desta forma descrito não leva em consideração as possibilidades de ferramentas conforme descrito acima, que permitem uma descrição de informações espectrais de pelo menos um componente espectral de uma forma mais condensada. Como resultado, tal sistema de conferência exige etapas adicionais de transformação para reconstruir os sinais de áudio fornecidos ao sistema de conferência, pelo menos até tal nível em que os respectivos sinais de áudio estejam presentes no domínio de frequência. Além disso, o sinal de áudio mixado resultante também precisa ser re-transformado com base nas ferramentas adicionais citadas acima. Estas etapas de re-transformação e transformação exigem, no

entanto, o uso de algoritmos complexos, que podem levar a uma maior complexidade computacional, e, por exemplo, de aplicações portáteis, energeticamente críticas, a um consumo maior de energia e, conseqüentemente a uma limitação no tempo de operação.

[00017] É, portanto, um problema a ser resolvido por meio de configurações de acordo com a presente invenção, para permitir um melhor *tradeoff* entre qualidade, largura de banda disponível e outros parâmetros adequados para sistemas de conferência, ou possibilitar a redução da complexidade computacional exigida em um sistema de conferência, conforme descrito acima.

[00018] Este objetivo é alcançado por meio do uso de um equipamento de acordo com a reivindicação 1 ou 12, um método para mixagem de uma pluralidade de fluxos de dados de entrada de acordo com a reivindicação 10 ou 26, ou um programa de computador de acordo com a reivindicação 11 ou 27.

[00019] De acordo com um primeiro aspecto, configurações de acordo com a presente invenção são baseadas na revelação de que, ao se mixar uma pluralidade de fluxos de dados de entrada, é possível obter um *tradeoff* melhor entre os parâmetros acima citados e as metas, determinando um fluxo de dados de entrada baseado em uma comparação, e copiando pelo menos parcialmente, informações espectrais do determinado fluxo de dados de entrada, ao fluxo de dados de saída. Copiando

informações espectrais pelo menos parcialmente do fluxo de dados de entrada, uma re-quantização pode ser omitida e, conseqüentemente, a re-quantização de ruído. No caso de informações espectrais para as quais não é possível determinar fluxos de entrada dominantes, a mixagem das informações espectrais correspondentes no domínio de frequências pode ser obtido por meio de uma configuração de acordo com a presente invenção.

[00020] A comparação pode, por exemplo, ser baseada em um modelo psico-acústico. A comparação pode, além disso, ser relacionada a informações espectrais correspondentes a um componente espectral em comum (por exemplo, uma frequência ou uma banda de frequência) de pelo menos dois fluxos de dados de entrada diferentes. Pode ser, portanto, uma comparação intra-canal. Caso a comparação seja baseada em um modelo psico-acústico, a comparação pode, conseqüentemente, ser descrita como considerando uma máscara intra-canal.

[00021] De acordo com um segundo aspecto, configurações de acordo com a presente invenção são baseadas na revelação de que uma complexidade de operações efetuadas durante a mixagem de um primeiro fluxo de dados de entrada e de um segundo fluxo de dados de entrada para gerar um fluxo de saída de dados, pode ser reduzida levando em consideração valores de controle associados a dados de carga útil do respectivo fluxo de dados de entrada, onde os valores de controle indicam uma forma onde os dados de

carga útil representam pelo menos uma parte das informações espectrais correspondentes ou o domínio espectral dos respectivos sinais de áudio. Caso os valores de controle dos dois fluxos de dados de entrada sejam iguais, uma nova decisão sobre a forma em que o domínio espectral na respectiva estrutura do fluxo de dados de saída possa ser omitido e, em vez disso, a geração do fluxo de saída poderá contar com a decisão já tomada e de comum acordo determinada pelos codificadores dos fluxos de dados de entrada, por exemplo, adotando o valor de controle disto. Dependendo da forma indicada pelos valores de controle, poderá até mesmo ser possível e preferível evitar a retransformação dos respectivos dados de carga útil de volta para outra forma de representação do domínio espectral, como a forma normal ou plena com um valor espectral por tempo/amostra espectral. No último caso, um processamento direto dos dados de carga útil para obter os dados de carga útil correspondentes do fluxo de saída de dados e o valor de controle sendo igual aos valores de controle do primeiro e do segundo fluxo de dados de entrada, pode ser gerado com o sentido de "direcionamento" "sem alterar a forma em que o domínio espectral é representado", como por meio de PNS ou recursos de áudio semelhantes, descritos abaixo em mais detalhes.

[00022] Em configurações de acordo com uma configuração da presente invenção, os valores de controle se relacionam a pelo menos um componente espectral somente. Além disso, em

configurações de acordo com a presente invenção, tais operações podem ser realizadas quando as estruturas do primeiro fluxo de dados de entrada e do segundo fluxo de dados de entrada, corresponderem ao índice comum de tempo com relação a uma seqüência adequada de estruturas dos dois fluxos de dados de entrada.

[00023] Caso os valores de controle do primeiro e do segundo fluxo de dados não forem iguais, configurações de acordo com a presente invenção podem realizar a etapa de transformação dos dados de carga útil de uma estrutura de um dos primeiro e segundo fluxos de dados de entrada para se obter uma representação dos dados de carga útil de uma estrutura de outro fluxo de dados de entrada. Os dados de carga útil do fluxo de dados de saída podem então ser gerados com base nos dados de transformação de carga útil e os dados de carga útil dos outros dois fluxos. Em alguns casos, configurações de acordo com a presente invenção transformando os dados de carga útil da estrutura de um fluxo de dados de entrada para a representação dos dados de carga útil da estrutura do outro fluxo de dados de entrada, podem ser diretamente realizadas sem a transformação dos respectivos sinais de áudio de volta para o domínio de frequência plena.

[00024] Configurações de acordo com a presente invenção serão descritas a seguir, fazendo referência às seguintes figuras.

[00025] A Fig. 1 representa um diagrama em bloco de um sistema de conferência;

[00026] A Fig. 2 representa um diagrama em bloco do sistema de conferência baseado em um codec geral de áudio;

[00027] Fig. 3 representa um diagrama em bloco de um sistema de conferência em um domínio de frequência usando a tecnologia de mixagem de fluxo de bits;

[00028] Fig. 4 representa um desenho esquemático de fluxo de dados compreendendo uma pluralidade de estruturas;

[00029] Fig. 5 ilustra diferentes formas de componentes espectrais e dados ou informações espectrais;

[00030] Fig. 6 ilustra um equipamento para mixagem de uma pluralidade de fluxo de dados de entrada de acordo com uma configuração da presente invenção em mais detalhes;

[00031] Fig. 7 ilustra um modo de operação do equipamento da fig. 6 de acordo com uma configuração da presente invenção;

[00032] Fig. 8 representa um diagrama em bloco de um equipamento para mixagem de uma pluralidade de fluxos de dados de entrada de acordo com outra configuração da presente invenção no contexto de um sistema de conferência;

[00033] Fig. 9 representa um diagrama em bloco simplificado de um equipamento para geração de um fluxo de dados de saída de acordo com uma configuração da presente invenção;

[00034] Fig. 10 representa um diagrama em bloco mais detalhado de um equipamento para geração de um fluxo de dados de saída de

acordo com uma configuração da presente invenção;

[00035] Fig. 11 representa um diagrama em bloco de um equipamento para geração de um fluxo de dados de saída de uma pluralidade de fluxos de dados de entrada de acordo com outra configuração da presente invenção no contexto de um sistema de conferência;

[00036] Fig. 12a ilustra uma operação de um equipamento de geração de fluxo de dados de acordo com uma configuração da presente invenção para uma implementação PNS;

[00037] Fig. 12b ilustra uma operação de um equipamento de geração de fluxo de dados de acordo com uma configuração da presente invenção para uma implementação SBR; e

[00038] Fig. 12c ilustra uma operação de um equipamento de geração de fluxo de dados de acordo com uma configuração da presente invenção para uma implementação M/S.

[00039] Com relação às fig. 4 a 12C, diferentes configurações de acordo com a presente invenção serão descritas em maiores detalhes. No entanto, antes de descrever estas configurações em maiores detalhes, primeiramente será dada uma breve introdução com relação às fig. 1 a 3 em face dos desafios e exigências que possam vir a se tornar importantes na estrutura dos sistemas de conferência.

[00040] A Fig. 1 representa um diagrama em bloco de um sistema de conferência 100, que também poderá ser citado como unidade de controle multiponto (MCU). Como poderá ser visto na

descrição em relação à sua funcionalidade, o sistema de conferência 100, conforme representado na fig. 1 é um sistema operando no domínio de tempo.

[00041] O sistema de conferência 100, conforme representado na fig. 1 é adaptado para receber uma pluralidade de fluxos de dados de entrada por meio de um número adequado de entradas 110-2, 110-3, ..., dos quais, na fig. 1 apenas três são representados. Cada uma das entradas 110 é acoplada a um respectivo decodificador 120. Para ser mais exato, a entrada 110-1 para o primeiro fluxo de dados de entrada é acoplada a um primeiro decodificador 120-1, enquanto a segunda entrada 120-2 é acoplada a um segundo decodificador 120-2, e a terceira entrada 110-2 é acoplada a um terceiro decodificador 120-3.

[00042] O sistema de conferência 100, além disso, compreende um número adequado de somadores 130-1, 130-2, 130-3, ... dos quais, mais uma vez, três são representados na fig. 1. Cada um dos somadores é associado a uma das entradas 110 do sistema de conferência 100. Por exemplo, o primeiro somador 130-2 é associado à primeira entrada 110-1 e ao decodificador correspondente 120-1.

[00043] Cada um dos somadores 130 é acoplado às saídas de todos os decodificadores 120 à parte do decodificador 120 ao qual a entrada 110 é acoplada. Em outras palavras, o primeiro somador 130-1 é acoplado a todos os decodificadores 120, à parte do primeiro decodificador 120-1. Conseqüentemente, o segundo

somador 130-2 é acoplado a todos os decodificadores 120, à parte do segundo decodificador 120-2.

[00044] Cada um dos somadores 130, além disso, compreende uma saída que é acoplada a um codificador 140, cada. Conseqüentemente, o primeiro somador 130-1 é acoplado em direção à saída, ao primeiro codificador 140-1. Conseqüentemente, o segundo e o terceiro somadores 130-2, 130-3, são também acoplados ao segundo e ao terceiro codificadores 140-2, 140-3, respectivamente.

[00045] Em contrapartida, cada um dos codificadores 140 é acoplado à respectiva saída 150. Em outras palavras, o primeiro codificador é, por exemplo, acoplado a uma primeira saída 150-1. O segundo e o terceiro codificadores 140-2, 140-3, são também acoplados à segunda e terceira saídas 150-2, 150-3, respectivamente.

[00046] No sentido de possibilitar a descrição da operação de um sistema de conferência 100, conforme descrito na fig. 1 em maiores detalhes, a fig. 1 também representa um terminal de conferência 160 de um primeiro participante. O terminal de conferência 160 pode, por exemplo, ser um telefone digital (por exemplo, um telefone ISDN = rede digital de serviço integrado), um sistema compreendendo uma infra-estrutura de fala sobre IP, ou terminal semelhante.

[00047] O terminal de conferência 160 compreende um codificador 170 que é acoplado à primeira saída 110-1 do sistema

de conferência 100. O terminal de conferência 160 também compreende um codificador 180 que é acoplado à primeira saída 150-1 do sistema de conferência 100.

[00048] Terminais de conferência 160 semelhantes também podem estar presentes em sites de outros participantes. Estes terminais de conferência não estão representados na fig. 1, apenas por questões de segurança. Também deve ser observado que o sistema de conferência 100 e os terminais de conferência 160, de forma alguma, precisam estar fisicamente presentes em proximidade um ao outro. Os terminais de conferência 160 e o sistema de conferência 100 podem estar dispostos em sites diferentes, que podem, por exemplo, estar conectados somente por meios de técnicas WAN (WAN = redes de área ampla).

[00049] Os terminais de conferência 160 podem ainda compreender ou estar conectados a componentes adicionais, tais como, microfones, amplificadores e auto-falantes ou fones de ouvido para possibilitar uma troca de sinais de áudio com um usuário humano de uma forma mais compreensível. Estes não são representados na fig. 1 somente por questões de facilidade.

[00050] Conforme citado anteriormente, o sistema de conferência 100 representado na fig. 1 é um sistema operando no domínio de tempo. Quando, por exemplo, o primeiro participante fala ao microfone (não representado na fig. 1), o codificador 170 do terminal de conferência 160 codifica o respectivo sinal de áudio em um fluxo de bits correspondente e transmite o fluxo

de bits à primeira entrada 110-1 do sistema de conferência 100.

[00051] Dentro do sistema de conferência 100, o fluxo de bits é decodificado pelo primeiro decodificador 120-1 e transformado de volta ao primeiro domínio. Uma vez que o primeiro decodificador 120-2 é acoplado ao segundo e ao terceiro mixer 130-1, 130-3, o sinal de áudio, conforme gerado pelo primeiro participante, pode ser mixado no domínio de tempo pela simples adição do sinal de áudio reconstruído com sinais de áudio também reconstruídos do segundo e do terceiro participantes, respectivamente.

[00052] Isto também se aplica para os sinais de áudio transmitidos pelo segundo e pelo terceiro participante, recebidos pela segunda e pela terceira entrada 110-2, 110-3 e processados pelo segundo e pelo terceiro decodificadores 120-2, 120-3, respectivamente. Estes sinais de áudio reconstruídos do segundo e do terceiro participantes são então transmitidos ao primeiro mixer 130-1, que por sua vez, transmite o sinal de áudio adicionado no domínio de tempo, ao primeiro codificador 140-1. O codificador 140-1 recodifica o sinal de áudio adicionado para formar um fluxo de bits e transmite o mesmo na primeira saída 150-1 para o primeiro terminal de participantes de conferência 160.

[00053] Semelhantemente, também o segundo e o terceiro codificadores 140-2, 140-3 codificam os sinais de áudio adicionados no domínio de tempo recebido do segundo e do

terceiro somadores 130-2, 130-3, respectivamente, e transmitem os dados codificados de volta aos respectivos participantes por meio da segunda e da terceira saídas 150-2, 150-3, respectivamente.

[00054] Para realizar a atual mixagem, os sinais de áudio são completamente decodificados e adicionados em forma não comprimida. Depois disso, opcionalmente, pode ser realizado um ajuste de níveis comprimindo os respectivos sinais de saída para evitar efeitos de cortes (por exemplo, ultrapassar um limite permitido de valores). Cortes podem aparecer quando valores únicos de amostras excedem ou caem abaixo de uma faixa permitida de valores de forma que os valores correspondentes sofram cortes (são cortados). No caso de uma quantização de 16 bits, como é, por exemplo, usada no caso de CDs, uma faixa inteira de valores entre -32768 e 32767 por valor de amostra é disponibilizada.

[00055] Para contrabalançar um possível direcionamento do sinal para cima ou para baixo, são usados algoritmos de compressão. Estes algoritmos limitam o desenvolvimento acima ou abaixo de um determinado valor limiar para manter os valores de amostra dentro de uma faixa permitida de valores.

[00056] Ao codificar dados de áudio em sistemas de conferência, tais como sistemas de conferência 100, conforme representado na fig. 1, algumas reduções são aceitas no sentido de realizar uma mixagem no estado não codificado de uma maneira mais facilmente executável. Além disso, as taxas de dados dos

sinais de áudio codificados são adicionalmente limitadas a uma faixa menor de frequências transmitidas, uma vez que uma menor largura de banda permite uma menor frequência de amostragem e, conseqüentemente, menos dados, de acordo com o teorema de amostragem Nyquist-Shannon. O teorema de amostragem Nyquist-Shannon afirma que a frequência de amostragem depende da largura da banda do sinal de amostra e precisa ser (no mínimo) duas vezes maior que a largura da banda.

[00057] A União Internacional de Telecomunicações (ITU) e seu setor de padronização de telecomunicações (ITU-T) desenvolveram diversos padrões para sistemas de conferências multimídia. O H.320 é o protocolo padrão de conferências para ISDN. O H.323 define o sistema padrão de conferências para uma rede baseada em pacotes (TCP/IP). O H.324 define sistemas de conferências para redes de telefones análogos e sistemas de rádio telecomunicações.

[00058] Nestes padrões, não apenas a transmissão de sinais, mas também a codificação e o processamento de dados de áudio são definidos. A gestão de uma conferência é assumida por um ou mais servidores, as assim chamadas unidades de controle multiponto (MCU) de acordo com o padrão H.231. As unidades de controle multiponto são também responsáveis pelo processamento e distribuição de dados de vídeo e de áudio dos diversos participantes.

[00059] Para atingir isso, a unidade de controle

multiponto envia a cada participante uma saída mixada ou sinal resultante compreendendo os dados de áudio de todos os outros participantes e transmite o sinal aos respectivos participantes. A Fig. 1 não apenas representa um diagrama de bloco de um sistema de conferência 100, mas também, um fluxo de sinal em tal situação de conferência.

[00060] Na estrutura dos padrões H.323 e H. 320, codecs de áudio da classe G.7xx são definidos para operação nos respectivos sistemas de conferência. O padrão G.711 é usado para transmissões ISDN em sistemas de telefones por cabo. Em uma frequência de amostragem de 8 kHz, o padrão G.711 cobre uma largura de banda de áudio entre 300 e 3.400 Hz, exigindo uma taxa de bits de 64 kbit/s em uma (quantização) profundidade de 8 bits. A codificação é formada por uma codificação logarítmica única chamada  $\mu$ Law ou A-Law que cria um atraso muito pequeno de apenas 0,125 ms.

[00061] O padrão G.722 codifica uma largura de banda de áudio maior de 50 a 7.000 Hz a uma frequência de amostragem de 16 kHz. Como consequência, o codec alcança uma melhor qualidade comparado os codecs de menor largura de banda G.7xx em taxas de bits de 48, 56, ou 64 Kbit/s, a um atraso de 1,5 ms. Além disso, dois desenvolvimentos adicionais, o G.722.1 e o G.722.2 existem, o que fornece qualidade de fala a menores taxas de bits. O G722.2 permite uma escolha de taxas de bits entre 6.6 kbit/s e 23.85 kbit/s a um atraso de 25 ms.

[00062] O padrão G.729 é normalmente usado em caso de telecomunicações de telefone IP, que também são chamadas de comunicações de fala sobre IP (VoIP). O codec é otimizado para sinais de fala e transmite um jogo de parâmetros de falas analisadas para uma posterior síntese junto com um sinal de erro. Como resultado, o G.729 alcança uma codificação significativamente melhor de aproximadamente 8 kbit/s a uma taxa de amostragem e largura de banda comparáveis, quando comparadas com o padrão do G.711. O algoritmo mais complexo, no entanto, cria um atraso de aproximadamente 15 ms.

[00063] Como uma desvantagem, os codecs G.7.xx são otimizados para codificação de falas e apresentam, à parte de uma largura de banda de baixa frequência, significativos problemas ao codificar músicas junto com a fala, ou apenas música.

[00064] Conseqüentemente, embora o sistema de conferência 100, conforme representado na fig. 1 possa ser usado para uma qualidade aceitável ao transmitir e processar sinais de fala, sinais gerais de áudio não são satisfatoriamente processados com o uso de codecs de baixo atraso, otimizados para falas.

[00065] Em outras palavras, o uso de codecs para codificar e decodificar sinais de fala para processar sinais gerais de áudio, inclusive, por exemplo, sinais de áudio com música, não leva a resultados satisfatórios em relação à

qualidade. Com o uso de codecs para codificação e decodificação de sinais gerais de áudio na estrutura do sistema de conferência 100, conforme representado na fig. 1, a qualidade é improvável. No entanto, conforme será mostrado em detalhes no contexto da fig. 2 em maiores detalhes, o emprego de codecs gerais de áudio em tal sistema de conferência, pode levar a outros efeitos indesejáveis, tais como, um atraso maior, para citar apenas um.

[00066] No entanto, antes de descrever a fig. 2 em maiores detalhes, é necessário citar que na presente descrição, objetos são simbolizados com os mesmos sinais ou sinais de referência similares, quando os respectivos objetos aparecem mais de uma vez em uma configuração ou figura, ou aparecem em diversas configurações ou figuras. A menos que explicitamente ou implicitamente simbolizados de outra forma, objetos simbolizados pelos mesmos sinais ou por sinais de referência semelhantes podem ser implementados de uma forma similar ou igual, por exemplo, quanto a seus circuitos, programação, características, ou outros parâmetros. Conseqüentemente, objetos que aparecem em diversas configurações de figuras e sendo simbolizados com os mesmos sinais de referência ou sinais de referência similares, podem ser implementados possuindo as mesmas especificações, parâmetros, e características. Naturalmente, variações e adaptações também podem ser implementadas, por exemplo, quando condições limite ou outros parâmetros mudam de figura para figura, ou de configuração para configuração.

[00067] Além disso, na sumarização a seguir, sinais serão usados para simbolizar um grupo ou classe de objetos, antes de objetos individuais. Na estrutura da Fig. 1, isto já foi feito, por exemplo, ao simbolizar a primeira entrada como entrada 110-1, a segunda entrada como entrada 110-2 e a terceira entrada como entrada 110-3, sendo que as entradas foram discutidas somente em termos do sinal de referência de sumarização 110. Em outras palavras, a menos que de outra forma citado, partes da descrição se referindo a objetos simbolizados com sinais de referência sumarizados, podem também se relacionar a outros objetos exibindo sinais de referência individuais correspondentes.

[00068] Uma vez que isto também se aplica a objetos simbolizados com sinais de referência iguais ou similares, ambas as medidas ajudam a encurtar a descrição e a descrever as configurações reveladas nelas de uma forma mais clara e concisa.

[00069] Fig. 2 representa um diagrama em bloco de outro sistema de conferência 100 junto com um sistema de conferência 160, que são ambos similares aos representados na fig. 1. O sistema de conferência 100 representado na fig. 2 também compreende entradas 110, decodificadores 120, somadores 130, codificadores 140, e saídas 150, são igualmente interconectadas ao sistema de conferência 100 representado na fig. 1. O terminal de conferência 160 representado na fig. 2 também compreende novamente um codificador 170 e um decodificador 180. Portanto,

referência é feita à descrição do sistema de conferência 100 representado na fig. 1.

[00070] No entanto, o sistema de conferência representado na fig. 2, assim como, um sistema de conferência 160 representado na fig. 2, são adaptados para usar um sistema geral de áudio (Codificador - Decodificador). Conseqüentemente, cada um dos codificadores 140, 170, compreende uma conexão em série de um conversor 190 de tempo/frequência, acoplado á frente de um quantizador/codificador 200. O conversor de tempo/frequência 190 também é ilustrado na fig. 2 como "T/F", enquanto os quantizadores/codificadores 200 são indicados na fig.2 com "Q/C".

[00071] Os decodificadores 120, 180, cada um, compreendem uma decodificador/quantizador 210, que são citados na fig. 2 como "Q/C"<sup>-1</sup>, conectados em série com um conversor 220 de frequência/tempo, que é citado na fig. 2 como "T/F"<sup>-1</sup>". Por uma questão de simplicidade somente, o conversor de tempo/frequência 190, o quantizador/codificador 200 e o decodificador/dequantizador 210, assim como, o conversor de frequência/tempo 220 são indicados desta forma somente no caso do codificador 140-3 e do decodificador 120-3. No entanto, a seguinte descrição também se refere aos outros destes elementos.

[00072] Começando com um codificador como os codificadores 140, e o codificador 170, o sinal de áudio transmitido ao conversor 190 de tempo/frequência é convertido do

domínio de tempo para um domínio de frequência ou um domínio relacionado a frequência por meio do conversor 190. Depois disso, os dados de áudio convertidos são, em uma representação espectral, gerados pelo conversor de tempo/frequência 190, quantizados e codificados para formar uma fluxo de bits, que são então transmitidos, por exemplo, para as saídas 150 do sistema de conferência 100, no caso de um codificador 140.

[00073] Em termos dos decodificadores como dos decodificadores 120 ou do decodificador 180, o fluxo de bits transmitido aos decodificadores é primeiramente decodificado e re-quantizado para formar a representação espectral de pelo menos uma parte de um sinal de áudio, que é então convertido de volta ao domínio de tempo pelos conversores de frequência/tempo 220.

[00074] Os conversores de tempo/frequência 190, assim como os elementos inversos, os conversores de tempo/frequência 220, são desta forma adaptados para gerar uma representação espectral de pelo menos uma peça de sinal de áudio transmitida a eles e para retransformar o representante espectral para dentro das respectivas partes do sinal de áudio no domínio de tempo, respectivamente.

[00075] No processo de conversão de um sinal de áudio do domínio de tempo para o domínio de frequência, e de volta do domínio de frequência para o domínio de tempo, poderão ocorrer alterações de forma que o sinal de áudio restabelecido,

reconstruído ou decodificado, pode divergir do sinal de áudio original ou de fonte. Outros artefatos podem ser acrescentados pelas etapas adicionais de quantização e dequantização realizadas na estrutura do decodificador de quantização 200 e do recodificador 210. Em outras palavras, o sinal de áudio original, assim como, os sinais de áudio restabelecidos podem divergir um do outro.

[00076] Os conversores 190 de tempo/frequência, assim como os conversores de frequência/tempo 220, podem, por exemplo, ser implementados com base em MDCT (transformada discreta de modificação de cosseno), uma MDST (transformada discreta de modificação de seno), um conversor baseado em FFT (FFT = transformada rápida de Fourier) ou outro conversor baseado em Fourier. A quantização e re-quantização na estrutura do quantizador/codificador 200 e do decodificador/dequantizador 210 pode, por exemplo, ser implementada com base em uma quantização linear, uma quantização logarítmica, ou outro algoritmo de quantização mais complexo, levando em consideração, mais especificamente, as características auditivas dos humanos. As partes de codificador e decodificador do quantizador/codificador 200 e do decodificador/dequantizador 210, podem, por exemplo, trabalhar por meio do uso de um esquema de codificação ou decodificação Huffman.

[00077] No entanto, outros conversores 190, 220 de tempo/frequência e frequência/tempo mais complexos, assim como,

quantizadores/codificadores e decodificadores/dequantizadores 200, 210 mais complexos podem ser usados em diferentes configurações e sistemas conforme aqui descritos, fazendo parte da formação, por exemplo, de um codificador AAC-ELD como codificadores 140, 170, e um decodificador AAC-ELD como decodificadores 120, 180.

[00078] É desnecessário dizer que pode ser apropriado implementar codificadores 170, 140 idênticos, e codificadores 180, 120 ou pelo menos decodificadores 180, 120 compatíveis na estrutura do sistema de conferência 100 e dos terminais de conferência 160.

[00079] O sistema de conferência 100, conforme representado na fig. 2, baseado em esquema geral de codificação e decodificação de sinal de áudio, também realiza a atual mixagem dos sinais de áudio no domínio de tempo. Os somadores 130 são providos com os sinais de áudio reconstruídos no domínio de tempo para realizar uma superposição e de transmitir os sinais mixados no domínio de tempo aos conversores 190 de tempo/frequência dos seguintes codificadores 140. Conseqüentemente, o sistema de conferência mais uma vez compreende uma conexão em série de decodificadores 120 e codificadores 140, o que é a razão pela qual o sistema de conferência 100, conforme representado nas fig. 1 e 2 é normalmente citado como o "sistema de codificação tandem".

[00080] Sistemas de codificação tandem muitas vezes

mostram a desvantagem de uma alta complexidade. A complexidade de mixagem em muito depende da complexidade dos decodificadores e codificadores usados, e pode se multiplicar significativamente no caso de diversos sinais de áudio de entrada e de saída. Além disso, devido ao fato de que a maioria dos esquemas de codificação e decodificação não serem livres de perdas, o esquema de codificação tandem, conforme usado nos sistemas de conferência 100 representados nas fig. 1 e 2, normalmente leva a uma influência negativa sobre a qualidade.

[00081] Como outra desvantagem, as etapas repetidas de decodificação e codificação também aumentam o atraso total entre as entradas 110 e as saídas 150 do sistema de conferência 100, o que também é citado como atraso ponta a ponta. Dependendo de um atraso inicial, dos decodificadores e codificadores usados, o sistema de conferência 100 em si, pode aumentar o atraso até um nível que torna o uso na estrutura do sistema de conferência, sem atrativos, ou mesmo, inoportuno, ou até, impossível. Muitas vezes, um atraso de aproximadamente 50 ms é considerado como sendo o atraso máximo que os participantes aceitam em conversações.

[00082] Como principais fontes de atraso, os conversores 190 de tempo/frequência, assim como, os conversores 220 de frequência/tempo, são responsáveis pelo atraso ponta a ponta do sistema de conferência 100, e o atraso adicional imposto pelos terminais de conferência 160. O atraso causado

pelos outros elementos, principalmente os quantizadores/codificadores 200 e os decodificadores/dequantizadores 210 é de menor importância, uma vez que estes componentes podem ser operados a uma frequência bem maior que em comparação com os conversores de tempo/frequência e os conversores frequência/tempo 190, 220. A maioria dos conversores de tempo/frequência e frequência/tempo 190, 220 é operada em bloco ou estruturada, o que significa que em muitos casos, um mínimo de atraso como quantidade de tempo precisa ser levado em consideração, o que equivale ao tempo necessário para encher um buffer ou uma memória possuindo o comprimento de estrutura de um bloco. Este tempo é, no entanto, significativamente influenciado pela frequência de amostragem, o que é normalmente no âmbito de poucos kHz até alguns 10 kHz, enquanto a velocidade operacional dos quantizadores/codificadores 200, assim como, dos decodificadores/dequantizadores 210 é principalmente determinada pela frequência de relógio do sistema subordinado. Isto é normalmente pelo menos 2, 3, 4 ou mais ordens de magnitude, maior.

[00083] Conseqüentemente, em sistemas de conferência usando codecs gerais de sinais de áudio, a assim chamada tecnologia de mixagem de fluxos de bits foi introduzida. O método de mixagem de fluxos de bits pode, por exemplo, ser implementada com base no codec MPEG-4 AAC-ELD, que oferece a

possibilidade de evitar pelo menos algumas das desvantagens mencionadas acima e introduzidas pela codificação tandem.

[00084] Deve-se, no entanto, observar que em princípio o sistema de conferência 100 conforme representado na fig. 2, pode também ser implementado com base no codec MPEG-4 AAC-ELD com uma taxa de bits similar e uma largura de banda significativamente maior, comparado aos códigos baseados em fala previamente citados da família codec G.7xx. Isto imediatamente também implica que uma qualidade significativamente melhor de qualidade de áudio para todos os tipos de sinais pode ser alcançada a um custo de uma taxa de bits significativamente maior. Embora o MPEG-4 AAC-ELD ofereça um atraso que fica no âmbito do codec G.7xx, a implementação do mesmo na estrutura de um sistema de conferência, conforme representado na fig. 2, poderá não levar a um sistema de conferência 100 prático. A seguir, com relação à fig. 3, um sistema mais prático baseado na anteriormente citada e assim chamada mixagem de fluxos de bits, será detalhada.

[00085] Deve ser observado que somente por uma questão de simplicidade, o foco estará voltado principalmente ao codec MPEG-4 AAC-ELD e seus fluxos de dados e fluxos de bits. No entanto, também outros codificadores e decodificadores podem ser usados no ambiente de um sistema de conferência 100, conforme ilustrado e representado na fig. 3.

[00086] Fig. 3 representa um diagrama em bloco de um

sistema de conferência 100, trabalhando de acordo com o princípio de mixagem de fluxos de bits juntamente com um terminal de conferência 160, conforme descrito no contexto da fig. 2. O sistema de conferência 100 em si é uma versão simplificada do sistema de conferência 100 representado na fig. 2. Para ser mais preciso os decodificadores 120 do sistema de conferência 100 na fig. 2 foram substituídos por decodificadores/dequantizadores 220-1, 220-2, 210-3,... conforme representado na fig. 3. Em outras palavras, os conversores de frequência/tempo 120 dos decodificadores 120 foram removidos na comparação do sistema de conferência 100 representado nas fig. 2 e 3. De forma semelhante, os codificadores 140 do sistema de conferência 100 da fig.2 foram substituídos por quantizadores/codificadores 200-1, 200-2, 200-3. Conseqüentemente, os conversores de tempo/frequência 190 dos codificadores 140 foram removidos na comparação do sistema de conferência 100 representado nas fig. 2 e 3.

[00087] Como resultado, os somadores 130 não operam mais no domínio de tempo, mas, devido à falta dos conversores de frequência/tempo 220 e dos conversores de tempo/frequência 190, na frequência ou em um domínio relacionado à frequência.

[00088] Por exemplo, no caso dos codecs MPEG-4 AAC-ELD, o conversor de tempo/frequência 190 e o conversor de frequência/tempo 220, que somente estão presentes no sistema de conferência 160, são baseados em uma transformação MDCT. Desta

forma, dentro do sistema de conferência 100, os mixers 130 operam diretamente nas contribuições dos sinais de áudio na representação de frequência MDCT.

[00089] Uma vez que os conversores 190, 220 representam a principal fonte de atraso no caso do sistema de conferência 100 representado na fig. 2, o atraso é significativamente reduzido pela remoção destes conversores 190, 220. Além disso, a complexidade introduzida pelos dois conversores 190, 200 dentro do sistema de conferência 100 também é significativamente reduzida. Por exemplo, no caso de um decodificador MPEG-2 AAC, a transformação MDCT inversa realizada na estrutura do conversor de frequência/tempo 220, é responsável por aproximadamente 20% do total da complexidade. Uma vez que o conversor MPEG-4 também é baseado em uma transformação similar, uma contribuição não relevante sobre a complexidade total pode ser removida por meio da remoção apenas do conversor de frequência/tempo 220 do sistema de conferência 100.

[00090] A mixagem de sinais de áudio no domínio MDCT, ou outro domínio de frequência é possível, uma vez que, no caso de uma transformação MDCT ou no caso de uma transformação similar baseada em Fourier, estas transformações são transformações lineares. As transformações, portanto, possuem a propriedade de adição matemática, isto é,

$$[00091] \quad f(x + y) = f(x) + f(y), \quad (1)$$

[00092] e a de homogeneidade matemática, isto é,

[00093]  $f(a \cdot x) = a \cdot f(x),$  (2)

[00094] onde  $f(x)$  é uma função de transformação,  $x$  e  $y$  argumentos apropriados da mesma e  $a$ , uma constante de valor real ou de valor complexo.

[00095] Ambas as características da transformação MDCT ou outra transformação baseada em Fourier permitem uma mixagem no respectivo domínio de frequência, similar à mixagem no domínio de tempo. Conseqüentemente, todos os cálculos podem ser igualmente bem realizados com base nos valores espectrais. A transformação dos dados no domínio de tempo não é exigida.

[00096] Sob algumas circunstâncias, outra condição pode precisar ser atendida. Todos os dados espectrais relevantes devem ser iguais com relação a seus índices de tempo durante o processo de mixagem para todos os componentes espectrais relevantes. Isto eventualmente poderá não ser o caso quando, durante a transformação a assim chamada técnica de comutação em bloco é usada, de forma que o codificador dos terminais de conferência 160 possa ser livremente comutado entre diferentes comprimentos de blocos, dependendo de determinadas condições. A comutação de blocos pode colocar em risco a possibilidade de unicamente atribuir valores espectrais individuais a amostras no domínio de tempo devido à comutação entre diferentes comprimentos de blocos e comprimentos de janelas MDCT correspondentes, a menos que os dados a serem mixados tenham sido processados com a mesma janela. Uma vez que em um sistema

geral com terminais distribuídos de conferência 160, isto eventualmente não possa ser garantido, podem ser necessárias interpolações complexas, que, por sua vez, pode criar atrasos adicionais e complexidade. Como consequência, eventualmente poderá ser indicada a implantação de um processo de mixagem de fluxo de bits, baseado em comprimentos de blocos de comutação.

[00097] Em contraste, o codec AAC-ELD é baseado em um comprimento único de bloco, e, portanto, é capaz de garantir mais facilmente a atribuição previamente descrita de dados de frequência, de forma que uma mixagem pode ser mais facilmente realizada. O sistema de conferência 100 representado na fig. 3 é, em outras palavras, um sistema que é capaz de realizar a mixagem no domínio de transformação ou no domínio de frequência.

[00098] Conforme citado anteriormente, no sentido de eliminar o atraso adicional introduzido pelos conversores 190, 200 no sistema de conferência 100 representado na fig. 2, os codecs usados nos terminais de conferência 160 usam uma janela de comprimento e forma fixa. Isto permite a implementação do processo de mixagem descrito diretamente sem a transformação do fluxo de áudio de volta ao domínio de tempo. Este caminho é capaz de limitar a quantidade de atraso algorítmico adicionalmente introduzido. Além disso, a complexidade é diminuída devido à ausência das etapas inversas de transformação no decodificador e das etapas de transformação para frente, no codificador.

[00099] No entanto, também na estrutura de um sistema de conferência 100 conforme representado na fig.3, poderá ser necessário re-quantizar os dados de áudio após a mixagem pelos somadores 130, o que poderá introduzir ruído adicional de quantização. O ruído adicional de quantização poderá, por exemplo, ser criado devido a diferentes etapas de quantização de diferentes sinais de áudio transmitidos pelo sistema de conferência 100. Como resultado, por exemplo, no caso de transmissões de taxas de bits muito baixas, em que um número de etapas de quantização já é limitado, o processo de mixagem de dois sinais de áudio do domínio de frequência ou no domínio de transformação poderá resultar em uma quantidade não desejada de ruídos adicionais ou outras distorções no sinal gerado.

[000100] Antes de descrever uma primeira configuração de acordo com a presente invenção na forma de um equipamento para mixar uma pluralidade de fluxos de dados de entrada com relação à fig. 4, um fluxo de dados ou fluxo de bits, junto com dados ali compreendidos, será brevemente descrito.

[000101] Fig. 4 esquematicamente representa um fluxo de bits ou fluxo de dados 250 que compreende no mínimo uma ou, muitas vezes, mais de uma estrutura 260 de dados de áudio em um domínio espectral. Mais precisamente, a fig.4 representa três estruturas 260-1, 260-2, e 260-3 de dados de áudio em um domínio espectral. Além disso, o fluxo de dados 250 pode também compreender informações adicionais ou blocos de informações 270,

tais como, valores de controle indicando, por exemplo, uma forma pela qual os dados de áudio são codificados, outros valores de controle ou informações relacionadas a índices de tempo ou outros dados relevantes. Naturalmente, o fluxo de dados 250 conforme representado na fig. 4 pode também compreender estruturas adicionais ou uma estrutura 260 pode compreender dados de áudio de mais de um canal. Por exemplo, no caso de um sinal de áudio estéreo, cada uma das estruturas 260 pode, por exemplo, compreender dados de áudio de um canal esquerdo, um canal direito, dados de áudio derivados de ambos os canais, direito e esquerdo, ou qualquer combinação dos dados anteriormente mencionados.

[000102] Conseqüentemente, a fig. 4 mostra que um fluxo de dados 250 pode não apenas compreender uma estrutura de dados de áudio em um domínio espectral, mas também informações adicionais de controle, valores de controle, valores de status, informações de status, valores relacionados a protocolos, (por exemplo, somas de checagem), ou similares.

[000103] Dependendo das concretas implementações do sistema de conferência conforme descrito no contexto das fig. 1 a 3, ou dependendo da concreta implementação de um equipamento de acordo com uma configuração da presente invenção, conforme será descrito a seguir, em especial, de acordo com o descrito com relação às fig. 9 a 12C, os valores de controle que indiquem dados de carga útil associados a uma forma associada à

estrutura, representam no mínimo uma parte do domínio espectral ou informações espectrais de um sinal de áudio, podem igualmente bem estar compreendidos nas estruturas 260 em si, ou no bloco associado 270 de informações adicionais. No caso de valores relacionados a componentes espectrais, os valores de controle podem ser codificados dentro das estruturas 260 em si. Se, no entanto, um valor de controle estiver relacionado a uma estrutura inteira, ele pode igualmente bem estar compreendido nos blocos 270 de informações adicionais. No entanto, os locais anteriormente citados para inclusão dos valores de controle não estão, como descrito acima, de forma alguma compreendidos necessariamente nas estruturas 260, ou no bloco 270 dos blocos adicionais. Caso um valor de controle se relacionar somente a um único ou a poucos componentes espectrais, ele poderá também estar perfeitamente compreendido no bloco 270. Por outro lado, um valor de controle relacionado a uma estrutura completa 260, também pode estar compreendido nas estruturas 260.

[000104] Fig. 5 esquematicamente ilustra informações (espectrais) relacionadas a componentes espectrais como, por exemplo, compreendidas na estrutura 260 do fluxo de dados 250. Para ser mais preciso, a fig. 5 representa um diagrama simplificado de informações em um domínio espectral de um canal único de uma estrutura 260. No domínio espectral, uma estrutura de dados de áudio pode, por exemplo, ser descrita em termos de seus valores de intensidade  $I$  como uma função da frequência  $f$ .

Em sistemas discretos, tais como, por exemplo, sistemas digitais, a resolução de frequência também é discreta, de forma que as informações espectrais normalmente só estão presentes para determinados componentes espectrais, tais como, frequências individuais ou bandas estreitas ou sub-bandas. Frequências individuais ou bandas estreitas, assim como, sub-bandas, são citadas como componentes espectrais.

[000105] Fig. 5 esquematicamente representa uma distribuição de intensidade para seis frequências individuais 300-1, ..., 300-6, assim como uma banda de frequência ou sub-banda 310 compreendendo, no caso ilustrado na fig. 5, quatro frequências individuais. Ambas, frequências individuais ou bandas estreitas correspondentes 300, assim como a sub-banda ou banda de frequência 310 de componentes espectrais com relação às quais a estrutura compreende informações de dados de áudio no domínio espectral.

[000106] As informações relativas à sub-banda 310 podem, por exemplo, ser uma intensidade ampla ou um valor de intensidade médio. À parte da intensidade ou outros valores relacionados à energia, tais como a amplitude, a energia do respectivo componente espectral em si, ou outro valor derivado da energia ou da amplitude, informações de fase e outras informações, também podem ser compreendidas na estrutura e, desta forma, ser consideradas como informações relacionadas a componentes espectrais.

[000107] Após haver descrito alguns dos problemas envolvidos e algumas experiências sobre sistemas de conferência, configurações de acordo com um primeiro aspecto da presente invenção são descritas de acordo com as quais um fluxo de dados de entrada é determinado com base em uma comparação no sentido de copiar pelo menos parcialmente, informações espectrais do fluxo de dados de saída determinado, desta forma possibilitando a omissão de uma re-quantização e, conseqüentemente, ruídos de re-quantização associados a isto.

[000108] Fig. 6 representa um diagrama em bloco de um equipamento 500 para mixagem de uma pluralidade de fluxos de dados de entrada 510, dos quais, dois são mostrados, 510-1, 510-2. O equipamento 500 compreende uma unidade de processamento 520 que é adaptada para receber os fluxos de dados 510 e para gerar um fluxo de dados de saída 530. Cada um dos fluxos de dados de entrada 510-1, 510-2, compreende uma estrutura 540-1, 540-2, respectivamente, o que, de forma similar à estrutura 260 mostrada na fig.4, no contexto com a fig. 5, compreende um dado de áudio em um domínio espectral. Isto é novamente ilustrado por um sistema de coordenadas na fig. 6 na abscissa, da qual a frequência  $f$  e na ordenada da qual, a intensidade  $I$  é representada. O fluxo de dados de saída 530 também compreende uma estrutura de saída 550 que compreende dados de áudio em um domínio espectral, e também é mostrado pelo sistema de coordenadas correspondente.

[000109] A unidade de processamento 520 é adaptada para comparar as estruturas 540-1, 540-2, de uma pluralidade de fluxos de dados de entrada 510. Conforme será detalhado mais abaixo, esta comparação pode, por exemplo, ser baseada em um modelo psico-acústico levando efeitos máscara e outras propriedades das características da audição humana, em consideração. Baseado neste resultado comparativo, a unidade de processamento 520 é também adaptada para determinar pelo menos para um componente espectral, por exemplo, os componentes espectrais 560 mostrados na fig. 6, que está presente em ambas as estruturas 540-1, 540-2, exatamente um fluxo de dados de uma pluralidade de fluxos de dados 510. Então, a unidade de processamento 520, pode ser adaptada para gerar o fluxo de dados de saída 530, compreendendo a estrutura de saída 550, de forma que uma informação compreendendo o componente espectral 560 é copiada da estrutura determinada 540 do respectivo fluxo de dados de entrada 510.

[000110] Para ser mais preciso, a unidade de processamento 520 é adaptada de forma que comparando a estrutura 540 da pluralidade de fluxos de dados de entrada 510 seja baseada em pelo menos duas peças de informação - os valores de intensidade são valores relacionados a energia - correspondentes ao mesmo componente espectral 560 de estruturas 540 de dois diferentes fluxos de dados de entrada 510.

[000111] Para ilustrar isto mais detalhadamente, a fig.

7 esquematicamente representa o caso em que as peças de informação (a intensidade  $I$ ), correspondentes aos componentes espectrais 560, o que é aqui assumido ser uma frequência ou uma banda de frequência estreita da estrutura 540-1 de um primeiro fluxo de dados de entrada 510-1. Isto é comparado com o valor de intensidade  $I$  correspondente, sendo a peça de informação relativa ao componente espectral 560 da estrutura 540-2 do segundo fluxo de dados de entrada 510-2. A comparação pode, por exemplo, ser feita com base na avaliação de uma taxa de energia entre o sinal mixado onde somente alguns fluxos de entrada são inclusos e um completo sinal mixado. Isto pode, por exemplo ser alcançado de acordo com

$$[000112] \quad E_c = \sum_{n=1}^N E_n \quad (3)$$

[000113] e

$$[000114] \quad E_{f^{(n)}} = \sum_{\substack{n=1 \\ n \neq 1}}^N E_i \quad (4)$$

[000115] e calculando a taxa  $r(n)$  de acordo com

$$[000116] \quad r(n) = 20 \cdot \log \frac{E_{f^{(n)}}}{E_c}, \quad (5)$$

[000117] Onde  $n$  é um índice de um fluxo de dados de entrada e  $N$  é o número de todos, ou dos fluxos de dados de entrada relevantes. Se a razão  $r(n)$  for suficientemente alta, os canais menos dominantes ou as estruturas menos dominantes dos fluxos de dados de entrada 510 podem ser vistos como mascarados

pelos mais dominantes. Desta forma, uma redução de irrelevância pode ser processada, o que significa que somente tais componentes espectrais de um fluxo são incluídos quando forem perceptíveis, enquanto os outros fluxos são descartados.

[000118] Os valores de energia que devem ser considerados na estrutura de equações (3) a (5) podem, por exemplo, ser derivados dos valores de intensidade conforme representado na fig. 6, por meio do cálculo do quadrado dos respectivos valores de intensidade. Caso informações relativas a componentes espectrais compreendam outros valores, um cálculo semelhante pode ser feito dependendo da forma das informações compreendidas na estrutura 510. Por exemplo, no caso de informações de valores complexos, pode ser necessário calcular os módulos dos componentes reais e imaginários dos valores individuais atualizando as informações referentes aos componentes espectrais.

[000119] À parte de frequências individuais, para a aplicação do módulo psico-acústico de acordo com as equações (3) a (5), as somas das equações (3) e (4) podem compreender mais de uma frequência. Em outras palavras, nas equações (3) e (4) os respectivos valores de energia  $E_n$  podem ser substituídos por um valor geral de energia correspondente a uma pluralidade de frequências individuais, uma energia de uma banda de frequência, ou para colocar em termos gerais, por uma peça única de informações espectrais ou uma pluralidade de informações

espectrais relativas a um ou mais componentes espectrais.

[000120] Por exemplo, uma vez que o AAC-ELD opera em linhas espectrais em forma de banda, similar a grupos de frequências onde o sistema de audição humano trata ao mesmo tempo, a estimativa de irrelevância ou o modelo psico-acústico podem ser realizados de uma mesma maneira. Por meio do uso do modelo psico-acústico desta forma, é possível remover ou substituir parte de um sinal de apenas uma banda única de frequência, se necessário.

[000121] Como os exames psico-acústicos têm mostrado a dissimulação de um sinal por outro sinal, depende dos respectivos tipos de sinal. Como um limiar mínimo para uma determinação de irrelevância, um cenário de pior caso pode ser usados. Por exemplo, para a dissimulação de ruídos por um sinusóide ou outro som distinto e bem definido, uma diferença de 21 a 28 db é normalmente exigida. Testes têm mostrado que um valor limítrofe de aproximadamente 28,5 db leva a bons resultados substitutos. Este valor pode eventualmente ser melhorado, levando em consideração também as bandas de frequências atuais.

[000122] Conseqüentemente, valores  $r(n)$  de acordo com a equação (5) maiores que 28,5 db podem ser considerados irrelevantes em termos e uma avaliação psico-acústica ou avaliação de irrelevância com base no componente espectral ou nos componentes espectrais considerados. Para diferentes

componentes espectrais, diferentes valores podem ser usados. Desta forma, usando limiares como indicadores para uma irrelevância psico-acústica de um fluxo de dados de entrada em termos de uma estrutura sob consideração de 10 dB a 40 dB, 20 dB a 30 dB, ou 25 dB a 30 dB podem ser considerados úteis.

[000123] Na situação descrita na fig. 7, isto significa que com relação ao componente espectral 560, o primeiro fluxo de dados de entrada 510-1 é determinado, enquanto o segundo fluxo de dados de entrada 510-2 é descartado com relação ao componente espectral 560. Como resultado, a peça de informações relativa ao componente espectral 560 é pelo menos parcialmente copiada da estrutura 540-1 do primeiro fluxo de dados de entrada 510-1 para o fluxo de saída 550 do fluxo de dados de saída 530. Isto é ilustrado na fig. 7 pela seta 570. Ao mesmo tempo, as peças de informação referentes aos componentes espectrais 560 da estrutura 540 dos outros fluxos de dados de entrada 510 (por exemplo, na fig. 7, a estrutura 540-2 do fluxo de dados de entrada 510-2) são desconsideradas conforme ilustrado pela linha interrompida 580.

[000124] Ainda, em outras palavras, o equipamento 500, que pode, por exemplo, ser usado como um MCU ou um sistema de conferência 100, é adaptado de forma que o fluxo de dados de saída 530, junto com sua estrutura de saída 550 seja gerado, de forma que as informações do componente espectral correspondente, seja copiado somente da estrutura 540-1 de determinado fluxo de

dados de entrada 510-1 descrevendo o componente espectral 560 da estrutura de saída 550 do fluxo de dados de saída 530. Normalmente, o equipamento 500 pode também ser adaptado de forma que as informações relativas a mais de um componente espectral sejam copiadas de um fluxo de dados de entrada desconsiderando os outros fluxos de dados de entrada, pelo menos, com relação a estes componentes espectrais. Além disso, é possível, que um equipamento 500, ou sua unidade de processamento 520, seja adaptado de forma que, para diferentes componentes espectrais, diferentes fluxos de dados de entrada 510 sejam determinados. O mesmo fluxo de saída 550 do fluxo de dados de saída 530 pode compreender informações espectrais copiadas relacionadas a diferentes componentes espectrais de diferentes fluxos de dados de entrada 510.

[000125] Normalmente, é aconselhável implementar o equipamento 500 de forma que em caso de uma seqüência de estruturas 540 em um fluxo de dados de entrada 510, somente as estruturas 540 sejam consideradas durante a comparação e determinação, o que corresponde a um índice de tempo similar ou igual.

[000126] Em outras palavras, a fig. 7 ilustra os princípios operacionais de um equipamento para mixagem de uma pluralidade de fluxos de dados de entrada conforme descritos acima, de acordo com uma configuração. Conforme citado anteriormente, a mixagem não é feita de uma forma direta no

sentido de que todos os fluxos recebidos sejam decodificados, o que inclui uma transformação inversa para o domínio de tempo, mixagem, e novamente a codificação dos sinais.

[000127] As configurações das fig. 6 a 8 são baseadas em mixagens feitas do domínio de frequência do respectivo codec. Um possível codec pode ser um codec AAC-ELD, ou qualquer outro codec com uma janela uniforme de transformação. Neste caso, não é necessária qualquer transformação de tempo/frequência para que seja possível mixar os respectivos dados. Configurações de acordo com uma configuração da presente invenção fazem uso do fato de que o acesso a todos os parâmetros de fluxo de bits, tais como, o tamanho da etapa de quantização e outros parâmetros seja possível e que estes parâmetros possam ser usados para gerar um fluxo mixado de saída de bits.

[000128] As configurações das fig. 6 a 8 usam o fato de que a mixagem de linhas espectrais ou informações espectrais relacionadas a componentes espectrais pode ser feita por uma soma qualificada de linhas espectrais de fonte ou informações espectrais. Fatores de qualificação podem ser zero ou um, ou, em princípio, qualquer valor entre eles. Um valor de zero significa que as fontes são tratadas como irrelevantes e não serão usadas. Grupos de linhas, tais como bandas ou bandas de fatores de escala podem usar o mesmo fator de pesagem. No entanto, conforme ilustrado anteriormente, os fatores de qualificação (por exemplo, uma distribuição de zeros e um) podem ser variados para

os componentes espectrais de uma estrutura única 540 de um único fluxo de dados de entrada 510. Além disso, não é necessário usar exclusivamente os fatores de peso zero ou um ao mixar informações espectrais. Pode existir o caso em que sob determinadas circunstâncias, não para um único um, uma pluralidade de informações espectrais totais de uma estrutura 540 de um fluxo de dados de entrada 510, os respectivos fatores de qualificação possam ser diferentes de zero ou um.

[000129] Um caso em especial é quando todas as bandas ou componentes espectrais de uma fonte (fluxo de dados de entrada 510) são ajustados a um fator de um e todos os fatores das outras fontes são ajustados para zero. Neste caso, o fluxo completo de bits de entrada do participante é identicamente copiado como um fluxo final de bits mixados. Os fatores de qualificação podem ser calculados em uma base estrutura para estrutura, mas também podem ser calculados ou determinados com base em grupos mais longos ou seqüência de estruturas. Normalmente, mesmo dentro de uma destas seqüências de estruturas ou dentro de estruturas únicas, os fatores de qualificação podem diferir para diferentes componentes espectrais, conforme frisado anteriormente. Os fatores de qualificação podem ser calculados de acordo com resultados do modelo psico-acústico.

[000130] Um exemplo de modelo psico-acústico já foi descrito acima no contexto das equações (3), (4), e (5). O modelo psico-acústico ou um respectivo modelo calcula a taxa de

energia  $r(n)$  entre um sinal mixado, onde apenas alguns fluxos de entrada são inclusos, levando a um valor de energia  $E_f$  e o sinal mixado completo tendo um valor de energia de  $E_c$ . A razão de energia  $r(n)$  é então calculada de acordo com a equação (5) como 20 vezes o logaritmo de  $E_f$  dividido por  $E_c$ .

[000131] Se a taxa for suficientemente alta, os canais menos dominantes podem ser considerados como mascarados pelos dominantes. Desta forma, uma redução de irrelevância é processada, significando que apenas tais fluxos que não sejam notáveis, sejam inclusos, aos quais um fator de qualificação é atribuído, enquanto todos os outros fluxos - pelo menos uma informação espectral de um componente espectral - são descartados. Em outras palavras, a estes, um fator de qualificação zero, é atribuído.

[000132] A vantagem é que menos ou nenhuma codificação tandem ocorre devido ao número reduzido de etapas de requantização, que podem ser introduzidas. Uma vez que cada etapa de quantização revela um risco significativo de redução de ruído adicional de quantização, a qualidade total do sinal de áudio pode ser melhorada pelo emprego de uma das configurações acima mencionadas para a mixagem de uma pluralidade de fluxos de dados de entrada. Isto pode ser o caso quando a unidade de processamento 500, como, por exemplo, representado na fig. 6, é adaptada de forma que o fluxo de dados de saída 530 seja gerado de forma que uma distribuição de níveis de quantização,

comparada a uma distribuição de níveis de quantização na estrutura do determinado fluxo de entrada, ou partes dele, seja mantido. Em outras palavras, copiando, e, conseqüentemente, reusando os respectivos dados sem re-codificar as informações espectrais, uma introdução de ruído adicional de quantificação, pode ser omitido.

[000133] Além disso, o sistema de conferência, por exemplo, um sistema de tele/vídeo conferência com mais de dois participantes usando qualquer das configurações acima descritas com relação às fig. 6 a 8, pode oferecer a vantagem de uma menor complexidade em comparação com uma mixagem de domínio de tempo, uma vez que, etapas de transformação de tempo/freqüência e etapas de re-codificação podem ser omitidas. Além disso, nenhum outro atraso é causado por estes componentes em comparação à mixagem no domínio de tempo devido à ausência do atraso do banco de filtros.

[000134] Resumindo, as configurações acima descritas podem, por exemplo, ser adaptadas de forma que bandas ou informações espectrais correspondentes a componentes espectrais, que são retiradas completamente de uma fonte, não são re-quantizadas. Desta forma, apenas bandas ou informações espectrais que foram mixadas, serão quantizadas, o que reduz ruídos adicionais de quantização.

[000135] No entanto, as configurações acima descritas também podem ser usadas em diferentes aplicações, como

substituição de ruído perceptual (PNS), ajuste de ruído temporal (TNS), replicação espectral de banda (SBR), e modos de codificação estéreo. Antes de descrever a operação de um equipamento capaz de processar ao menos um dos parâmetros PNS, parâmetros TNS, parâmetros SBR, ou parâmetros de codificação estéreo, uma configuração será descrita em maiores detalhes com relação à fig. 8.

[000136] Fig. 8 representa um diagrama em bloco esquematizado de um equipamento 500 para mixagem de uma pluralidade de fluxos de dados de entrada compreendendo uma unidade de processamento 520. Para ser mais preciso, a fig. 8 representa um equipamento 500 altamente flexível, sendo capaz de processar sinais de áudio altamente diferentes, codificados em fluxos de dados de entrada (fluxos de bits). Alguns dos componentes que serão descritos a seguir são, portanto, componentes opcionais que não são exigidos a serem implementados sob todas as circunstâncias.

[000137] A unidade de processamento 520 compreende um decodificador de fluxo de bits 700 para cada um dos fluxos de dados de entrada ou fluxos de bits de áudio codificados, a serem processados pela unidade de processamento 520. Por questões de simplicidade apenas, a fig. 8 representa apenas dois decodificadores de fluxos de bits 700-1, 700-2. Naturalmente, dependendo do número de fluxos de dados de entrada a serem processados, um maior número de decodificadores de fluxos de

bits 700, ou um menor número, pode ser implementado, se, por exemplo, um decodificador de fluxos de bits 700 for capaz de seqüencialmente processar mais de um fluxo de dados de entrada.

[000138] O decodificador de fluxo de bits 700-1, assim como o outro decodificador de fluxo de bits 700-2, ... cada um compreende um leitor de fluxo de bits 710 que é adaptado para receber e processar os sinais recebidos, e isolar e extrair dados compreendidos no fluxo de bits. Por exemplo, o leitor de fluxo de bits 710 pode ser adaptado para sincronizar os dados recebidos com um relógio interno e pode, além disso, ser adaptado para separar o fluxo de bits de entrada nas estruturas apropriadas.

[000139] O decodificador de fluxo de bits 700 também compreende um decodificador Huffman 720 acoplado à saída do leitor de fluxo de bits 710 para receber os dados isolados do leitor de fluxo de bits 710. Uma saída do decodificador Huffman 720 é acoplada a um dequantizador 730, que é também mencionado como um quantificador inverso. O dequantizador 730 sendo acoplado atrás do decodificador Huffman 720 é seguido de um escalonador 740. O decodificador Huffman 720, o dequantizador 730 e o escalonador 740 de uma primeira unidade 750 na saída da qual pelo menos uma parte do sinal de áudio do respectivo fluxo de dados de entrada está disponível no domínio de frequência ou no domínio relacionado à frequência, no qual o codificador do participante (não mostrado na fig. 8) opera.

[000140] O decodificador de fluxo de bits 700 também compreende uma segunda unidade 760 que é acoplada no sentido de dados à primeira unidade 750. A segunda unidade 760 compreende um decodificador estéreo 770 (módulo M/S) atrás do qual é acoplado um decodificador PNS. O decodificador PNS 780 é seguido no sentido de dados por um decodificador TNS 790, que, junto com o decodificador PNS 780 no decodificador estéreo 770, forma a segunda unidade 760.

[000141] À parte do fluxo de dados de áudio descrito, o decodificador de fluxo de bits 700 também compreende uma pluralidade de conexões entre diferentes módulos referentes a dados de controle. Para ser mais preciso o leitor de fluxo de bits 710 também é acoplado ao decodificador Huffman 720 para receber dados de controle apropriados. Além disso, o decodificador Huffman 720 é diretamente acoplado ao escalonador 740 para transmitir informações de escalonamento ao escalonador 740. O decodificador 770, o decodificador PNS 780, e o decodificador TNS 790 são também, cada qual, acoplados ao leitor de fluxo de bits 710 para receber os dados de controle apropriados.

[000142] A unidade de processamento 520 também compreende uma unidade de mixagem 800 que por sua vez compreende um mixer espectral 810 que é acoplado no sentido da entrada aos decodificadores de fluxo de bits 700. O mixer espectral 810 pode, por exemplo, compreender um ou mais somadores para

realizar a atual mixagem do domínio de frequência. Além disso, o mixer espectral 810 pode também compreender multiplicadores para permitir uma combinação linear arbitrária das informações espectrais transmitidas pelos decodificadores de fluxo de bits 700.

[000143] A unidade de mixagem 800 também compreende um módulo de otimização 820 que é acoplado na direção de dados a uma saída do mixer espectral 810. O módulo otimizador 820, no entanto, é também acoplado ao mixer espectral 810 para prover o mixer espectral 810 com informações de controle. Na direção de dados, o módulo de otimização 820 representa uma saída da unidade de mixagem 800.

[000144] A unidade de mixagem 800 também compreende um mixer SBR 830 que é diretamente acoplado a uma saída do leitor de fluxo de bits 710 dos diferentes decodificadores de fluxo de bits 700. Uma saída do mixer SBR 830 forma outra saída da unidade de mixagem 800.

[000145] A unidade de processamento 520 também compreende um codificador de fluxo de bits 850 que é acoplado à unidade de mixagem 800. O codificador de fluxo de bits 850 compreende uma terceira unidade 860 compreendendo um codificador TNS 870, codificador PNS 880, e um codificador estéreo 890, que são acoplados em série na ordem descrita. A terceira unidade 860, conseqüentemente, forma uma unidade inversa da primeira unidade 750 do decodificador de fluxo de bits 700.

[000146] O codificador de fluxo de bits 850, além disso, compreende uma quarta unidade 900 que compreende um escalonador 910, um quantizador 920, e um codificador Huffman 930 formando uma conexão em série entre uma saída da quarta unidade e uma saída dela. A quarta unidade 900, conseqüentemente, forma um módulo inverso da primeira unidade 750. Desta forma, o escalonador 910 também é diretamente acoplado ao codificador Huffman 930 para prover o codificador Huffman 930 com os respectivos dados de controle.

[000147] O codificador de fluxo de bits 60 também compreende um escritor de fluxo de bits 940 que é acoplado à saída do codificador Huffman 930. Além disso, o escritor de fluxo de bits 940 também é acoplado ao decodificador TNS 870, ao codificador PNS 880, ao codificador estéreo 890, ao codificador Huffman 930 para receber dados de controle e informações destes módulos. Uma saída do escritor de fluxo de bits 940 forma uma saída da unidade de processamento 520 e do equipamento 500.

[000148] O codificador de fluxo de bits 850 também compreende um módulo psico-acústico 950, que também é acoplado à saída da unidade de mixagem 800. O codificador de fluxo de bits 850 é adaptado para prover os módulos da terceira unidade 860 com informações adequadas de controle, indicando, por exemplo, qual pode ser usada para codificar a saída de sinal de áudio pela unidade de mixagem 800 na estrutura das unidades da terceira unidade 860.

[000149] Em princípio, nas saídas da segunda unidade 760 até a entrada da terceira unidade 860, um processamento do sinal de áudio no domínio espectral, conforme definido pelo codificador usado no lado do emissor, portanto, é possível. No entanto, conforme citado anteriormente, uma completa decodificação, dequantização, descalcificação, e futuras etapas de processamento poderá eventualmente não ser necessária se, por exemplo, informações espectrais da estrutura de um dos fluxos de dados de entrada forem dominantes. Pelo menos uma parte das informações espectrais dos respectivos componentes espectrais, é então copiada para o componente espectral da respectiva estrutura do fluxo de dados de saída.

[000150] Para viabilizar tal procedimento, o equipamento 500 e a unidade de processamento 520 compreendem outras linhas de sinais para uma troca de dados otimizada. Para viabilizar tal processamento na configuração representada na fig. 8, uma saída do decodificador Huffman 720, assim como, as saídas do escalonador 740, do decodificador estéreo 770, e do codificador PNS 780 são, junto com os respectivos componentes dos outros leitores de fluxo de bits 710, acoplados ao módulo de otimização 820 da unidade de mixagem 800 para um respectivo processamento.

[000151] Para facilitar, após um respectivo processamento, um fluxo de dados dentro do codificador de fluxo de bits 850, linhas de dados correspondentes para um fluxo de dados otimizado, também são implementadas. Para ser mais

preciso, uma saída do módulo de otimização 820 é acoplada a uma saída do codificador PNS 780, ao codificador estéreo 890, a uma saída da quarta unidade 900 e ao escalonador 910, assim como, a uma entrada no codificador Huffman 930. Além disso, a saída do módulo de otimização 820 também é diretamente acoplada ao escritor de fluxo de bits 940.

[000152] Conforme citado anteriormente, quase todos os módulos conforme descritos acima são módulos adicionais, que não precisam ser implementados. Por exemplo, no caso dos fluxos de dados de áudio compreendendo um único canal, as unidades de codificação e decodificação estéreo 770, 890, podem ser omitidas. Desta forma, no caso em que nenhum sinal baseado em PNS precise ser processado, o decodificador PNS correspondente e o codificador PNS 780, 880, também podem ser omitidos. Os módulos TNS 790, 870 também podem ser omitidos caso o sinal a ser processado e o sinal a ser enviado não seja baseado em dados TNS. Dentro da primeira e da quarta unidades 750, 900 o quantizador inverso 730, o escalonador 740, o quantizador 920, assim como, o escalonador 910, podem eventualmente também ser omitidos. O decodificador Huffman 720 e o codificador Huffman 930 podem ser implementados de forma diferente, usando outro algoritmo, ou serem completamente omitidos.

[000153] O mixer SBR 830 também pode ser eventualmente omitido se, por exemplo, não houver parâmetros SBR de dados presentes. Além disso, o mixer espectral 810 pode ser

implementado diferentemente, por exemplo, como módulo de otimização 820 e o módulo psico-acústico 860. Desta forma, também estes módulos devem ser considerados componentes opcionais.

[000154] Com relação ao modo de operação do equipamento 500 junto com a unidade de processamento 520 nele compreendido, um fluxo de dados de entrada chegando, é inicialmente lido e separado em peças adequadas de informações por meio do leitor de fluxo de bits 710. Após a decodificação Huffman, as informações espectrais resultantes podem eventualmente se re-quantizadas pelo dequantizador 730 e escalonadas apropriadamente pelo de-escalador 740.

[000155] Depois disso, dependendo das informações de controle compreendidas no fluxo de dados de entrada, o sinal de áudio codificado no fluxo de dados de entrada pode ser decomposto em sinais de áudio para dois ou mais canais na estrutura do decodificador estéreo 770. Se, por exemplo, o sinal de áudio compreender um canal intermediário (M) e um canal lateral (S), os dados do canal direito e do canal esquerdo podem ser obtidos pela adição e subtração dos dados do canal intermediário e do canal lateral, um do outro. Em muitas implementações, o canal intermediário é proporcional à soma dos dados de áudio do canal esquerdo e do canal direito, enquanto o canal lateral é proporcional a uma diferença entre o canal esquerdo (L) e o canal direito (R). Dependendo da implementação,

os canais acima citados podem ser somados e/ou subtraídos considerando um fator  $\frac{1}{2}$  para evitar efeitos de cortes. Falando de forma genérica, os diferentes canais podem ser processados por combinações lineares para produzir os canais correspondentes.

[000156] Em outras palavras, depois que o decodificador estéreo 770, os dados de áudio podem, se adequado, ser decompostos em dois canais individuais. Naturalmente, também uma decodificação inversa pode ser realizada pelo decodificador estéreo 770. Se, por exemplo, o sinal de áudio, conforme recebido pelo leitor de fluxo de bits 710 compreender um canal esquerdo e um canal direito, o decodificador estéreo 770 pode igualmente calcular ou determinar dados apropriados de canal intermediário e de canal lateral.

[000157] Dependendo da implementação, não apenas do equipamento 500, mas também, dependendo da implementação do codificador do participante transmitindo o respectivo fluxo dados de entrada, o respectivo fluxo de dados pode compreender parâmetros PNS (PNS = substituição de ruídos perceptuais). O PNS é baseado no fato de que o ouvido humano quase sempre não é capaz de distinguir sons parecidos com ruídos em uma faixa de frequência limitada ou componente espectral, como uma banda ou uma frequência individual, de um ruído gerado sinteticamente. O PNS desta forma substitui a contribuição atual parecida com ruídos do sinal de áudio com um valor de energia indicando um

nível de ruído a ser sinteticamente introduzido no respectivo componente espectral e negligencia o sinal de áudio atual. Em outras palavras, o decodificador PNS 780 pode regenerar em um ou mais componentes espectrais a contribuição atual de sinal de áudio semelhante a ruídos em um parâmetro PNS compreendido no fluxo de dados de entrada.

[000158] Em termos do decodificador TNS 790 e do codificador TNS 870, respectivos sinais de áudio podem precisar ser re-transformados em uma versão não modificada com relação a um módulo TNS operando no lado do emissor. Ajuste temporal de ruído (TNS) é uma forma de reduzir artefatos de pré-eco causados pelos ruídos de quantização, que podem estar presentes no caso de um sinal do tipo transiente em uma estrutura do sinal de áudio. Para compensar este transiente, pelo menos um filtro preditivo ajustável é usado nas informações espectrais a partir do lado baixo do espectro, o lado alto do espectro, ou ambos os lados do espectro. Os comprimentos dos filtros preditivos podem ser adaptados, assim como, as faixas de frequências às quais os respectivos filtros são aplicados.

[000159] Em outras palavras, a operação do módulo TNS é baseada na comutação de um ou mais filtros ajustáveis IIR (IIR = resposta infinita de impulso) e pela codificação e transmissão de um sinal de erro descrevendo a diferença entre o sinal de áudio predito e do atual, junto com os coeficientes de filtro dos filtros preditivos. Como consequência, pode ser possível

aumentar a qualidade de áudio mantendo a taxa de bits do fluxo de dados de transmissão por meio da cópia com os sinais parecidos com transientes, aplicando um filtro de predição no domínio de frequência para reduzir a amplitude do sinal de erro remanescente, que pode então ser codificado usando menos etapas de quantização comparando com a codificação direta do sinal de áudio parecido com transiente, com um ruído de quantização similar.

[000160] Em termos de uma aplicação TNS, pode ser aconselhável sob determinadas circunstâncias, usar a função do decodificador TNS 760 para decodificar a parte TNS do fluxo de dados de entrada para chegar a uma "pura" representação no domínio espectral determinado pelo codec usado. Esta aplicação da funcionalidade dos decodificadores TNS 790 pode ser útil se uma estimativa do modelo psico-acústico (por exemplo, usado no módulo psico-acústico 950) não puder ser ainda estimado com base nos coeficientes dos filtros compreendidos nos parâmetros TNS. Isto pode ser especialmente importante no caso em que pelo menos um dos fluxos de dados de entrada usar TNS, enquanto outros não usar.

[000161] Quando a unidade de processamento determinar, com base na comparação das estruturas de fluxos de dados de entrada, que as informações espectrais de uma estrutura de um fluxo de dados de entrada usando TNS estiverem para serem usados, os parâmetros TNS podem ser usados para a estrutura dos

dados de saída. Se, por exemplo, por motivos de incompatibilidade o receptor do fluxo de dados de saída não estiver capacitado a decodificar dados TNS, pode ser útil não copiar os respectivos dados espectrais do sinal de erro e dos outros parâmetros TNS, mas processar os dados reconstruídos a partir dos dados relacionados a TNS para obter as informações no domínio espectral, e não usar o codificador TNS 870. Isto, mais uma vez mostra que partes dos componentes ou módulos representados na fig. 8 não precisam ser implementados mas, podem, opcionalmente, serem deixados de lado.

[000162] No caso de pelo menos um fluxo de entrada de áudio comparando dados PNS, uma estratégia similar pode ser usada. Se na comparação das estruturas para um componente espectral dos fluxos de dados de entrada revelar que um fluxo de dados de entrada é dominante em termos de sua estrutura presente e do respectivo componente espectral ou dos componentes espectrais, os respectivos parâmetros PNS (por exemplo, os respectivos valores de energia) podem também ser copiados diretamente para o respectivo componente espectral das estruturas de saída. Se, no entanto, o receptor não estiver capacitado para aceitar os parâmetros PNS, as informações espectrais podem ser reconstruídas a partir dos parâmetros PNS para os respectivos componentes espectrais, gerando ruídos com o nível de energia adequado conforme indicado pelo respectivo valor de energia. Então, os dados de ruído podem ser

adequadamente processados no domínio espectral.

[000163] Conforme citado anteriormente, os dados transmitidos podem também compreender dados SBR que podem ser processados no mixer SBR 830. Replicação de banda espectral (SBR) é uma técnica para replicar uma parte de um espectro de um sinal de áudio com base em contribuições e a parte de baixo dos mesmos espectros. Como consequência, a parte superior de espectro não precisa ser transmitida, à parte de parâmetros SBR que descrevem valores de energia em uma maneira dependente de frequência e dependente de tempo usando uma grade adequada de tempo/frequência. Como consequência, a parte superior de espectro não precisa ser transmitida. Para estar apto a melhorar ainda mais a qualidade do sinal reconstruído, contribuições adicionais de ruídos e contribuições senoidais podem ser adicionadas na parte superior do espectro.

[000164] Para ser um pouco mais específico, para frequências acima de uma frequência de interseção  $f_x$ , o sinal de áudio é analisado em termos de um banco de filtros QMF (QMF = filtro de espelho de quadratura) que cria um número específico de sinais de sub-bandas (por exemplo, 32 sinais de sub-bandas) possuindo uma resolução de tempo que é reduzida por um fator igual a, ou proporcional ao número de sub-bandas do filtro de bandas QMF (por exemplo, 32 ou 64). Como consequência, uma grade de tempo/frequência pode ser determinada compreendendo no eixo de tempo, dois ou mais assim chamados envelopes e, para cada

envelope, normalmente 7 a 16 valores de energia descrevendo a respectiva parte superior do espectro.

[000165] Além disso, os parâmetros SBR podem compreender informações relacionadas a ruídos adicionais e senóides que são então atenuados ou determinados com relação a sua força pela grade de tempo/frequência anteriormente citada.

[000166] No caso de um fluxo de dados de entrada baseado em SBR sendo o fluxo de dados de entrada dominante com relação à presente estrutura, a cópia dos respectivos parâmetros SBR junto com os componentes espectrais, pode ser realizada. Se mais uma vez, o receptor não for capaz de decodificar sinais baseados em SBR, uma respectiva reconstrução dentro do domínio de frequência poderá ser feita codificando o sinal reconstruído de acordo com exigências do receptor.

[000167] Uma vez que o SBR permite dois canais estéreo de codificação, codificando o canal esquerdo e o canal direito separadamente, assim como, codificando os mesmos em termos de um canal de união (C), de acordo com uma configuração da presente invenção, copiando os respectivos parâmetros SBR ou pelo menos parte deles, pode compreender a cópia de elementos C dos parâmetros SBR aos dois, elemento esquerdo e elemento direito dos parâmetros SBR a serem determinados e transmitidos, ou vice-versa, dependendo dos resultados da comparação e do resultado da determinação.

[000168] Além disso, uma vez que em diferentes

configurações da presente invenção fluxos de dados de entrada podem compreender ambos, sinais de áudio mono e estéreo, compreendendo um ou dois canais individuais, respectivamente, uma mixagem para cima de mono para estéreo ou uma mixagem para baixo de estéreo para mono, pode ser adicionalmente feita na estrutura de cópia de pelo menos partes das informações quando gerarem pelo menos parte das informações de um componente espectral correspondente da estrutura do fluxo de dados de saída.

[000169] Como a descrição anterior demonstrou o grau de cópia de informações espectrais e/ou respectivos parâmetros relacionados a componentes espectrais e informações espectrais (Por exemplo, parâmetros TNS, parâmetros SBR, parâmetros PNS) podem ser baseadas em diferentes números de dados a serem copiados e podem determinar se as informações espectrais subordinadas ou peças das mesmas também precisam ser copiadas. Por exemplo, no caso de copiar dados SBR, pode ser aconselhável copiar toda a estrutura do respectivo fluxo de dados para evitar uma complicada mixagem de informações espectrais para diferentes componentes espectrais. A mixagem deles pode exigir uma re-quantização que pode de fato reduzir o ruído de quantização.

[000170] Em termos de parâmetros TNS também pode ser aconselhável copiar os respectivos parâmetros TNS junto com as informações espectrais de toda a estrutura do fluxo de dados de entrada dominante para o fluxo de dados de saída, para evitar

uma re-quantização.

[000171] Em caso de informações espectrais baseadas em PNS, a cópia de valores de energia individuais sem copiar os componentes espectrais de apoio, pode ser uma forma viável. Adicionalmente, neste caso, copiando apenas os respectivos parâmetros PNS do componente espectral dominante da estrutura de uma pluralidade de fluxos de dados de entrada para o componente espectral correspondente da estrutura de saída do fluxo de dados de saída, ocorre sem a introdução de ruído de quantização adicional. Deve ser mencionado que mesmo com a re-quantização de um valor de energia em forma de um parâmetro PNS, ruído de quantização adicional pode ser introduzido.

[000172] Conforme citado anteriormente, a configuração citada acima pode também ser realizada simplesmente copiando uma informação espectral relativa ao componente espectral após comparar as estruturas da pluralidade de fluxos de dados e após determinar, com base na comparação, um componente espectral de uma estrutura de saída do fluxo de dados de saída, precisamente um fluxo de dados para ser a fonte das informações espectrais.

[000173] O algoritmo de reposição na estrutura do módulo psico-acústico 950 examina cada uma das informações espectrais relativas aos componentes espectrais relacionados (por exemplo, bandas de frequência) do sinal resultante para identificar componentes espectrais com apenas um único componente ativo. Para estas bandas, os valores quantizados dos respectivos fluxos

de dados de entrada podem ser copiados do codificador sem recodificar ou re-quantizar os respectivos dados espectrais para o componente espectral específico. Sob certas circunstâncias todos os dados quantizados podem ser obtidos a partir de um único sinal ativo de entrada para formar o fluxo de saída de bits ou fluxo de dados de saída, de forma que - em termos de um equipamento 500 - uma codificação sem perdas do fluxo de dados de entrada seja alcançado.

[000174] Além disso, pode ser possível omitir etapas de processamento, tais como, a análise psico-acústica dentro do codificador. Isto permite encurtar o processo de decodificação e, desta forma, reduzir a complexidade computacional, uma vez que, em princípio, apenas a cópia de dados do fluxo de bits para outro fluxo de bits precisa ser realizada sob certas circunstâncias.

[000175] Por exemplo, no caso de PNS, uma reposição pode ser realizada uma vez que, fatores de ruído da banda codificada PNS podem ser copiados de um dos fluxos de dados de saída para o fluxo de dados de saída. A reposição de componentes espectrais individuais com parâmetros PNS é possível, uma vez que os parâmetros PNS são específicos a componentes espectrais, ou, em outras palavras, a uma aproximação independente muito boa, um do outro.

[000176] No entanto, pode ocorrer que uma aplicação muito agressiva do algoritmo descrito leve a uma experiência

auditiva degradada ou a uma redução não desejada em qualidade. Pode, conseqüentemente, ser aconselhável limitar a reposição de estruturas individuais, preferivelmente a informações espectrais, relacionadas a componentes espectrais. Em tal modo de operação a estimativa de irrelevância ou determinação de irrelevância, assim como, a análise de reposição podem ser realizadas sem alterações. No entanto, uma reposição pode neste modo de operação, somente ser realizada quando todos, ou pelo menos um número significativo de componentes espectrais dentro da estrutura ativa, forem substituíveis.

[000177] Embora isto possa levar a um número menor de reposições, uma força interna das informações espectrais pode, em algumas situações, ser melhorada, levando a uma qualidade até mesmo levemente melhor.

[000178] A seguir, configurações de acordo com um segundo aspecto da presente invenção são descritas de acordo com as quais os valores de controle associados aos dados de carga útil dos respectivos fluxos de dados de entrada são levados em consideração. Os valores de controle indicando uma forma em que os dados de carga útil representam pelo menos uma parte das informações espectrais correspondentes ou domínio espectral dos respectivos sinais de áudio, onde, no caso dos valores de controle dos dois fluxos de dados de entrada são iguais, uma nova decisão sobre a forma em que o domínio espectral da respectiva estrutura do fluxo de dados de saída é evitado e, ao

invés disso a geração do fluxo de saída se baseia na decisão já determinada pelos codificadores no fluxo de dados de entrada. De acordo com algumas configurações descritas abaixo, a re-transformação dos respectivos dados de carga útil de volta a outra forma de representação do domínio espectral, como a forma normal ou plena com um valor espectral por amostra de tempo/espectro, é evitada.

[000179] Conforme citado anteriormente, configurações de acordo com a presente invenção são baseadas na realização de uma mixagem, que não é feita de uma forma direta no sentido de que todos os fluxos de entrada são decodificados, o que inclui uma transformação inversa para o domínio de tempo, mixando e novamente re-codificando os sinais. Configurações de acordo com a presente invenção são baseadas em mixagens feitas do domínio de frequência do respectivo codec. Um possível codec pode ser o codec AAC-ELD, ou qualquer outro codec com uma janela de transformação uniforme. Em tal caso, não é necessária uma transformação de tempo/frequência para que seja possível uma mixagem dos respectivos dados. Além disso, um acesso a todos os parâmetros de fluxos de bits, tais como, tamanhos de etapas de quantização e outros parâmetros é possível, e estes parâmetros podem ser usados para gerar um fluxo mixado de bits de saída.

[000180] Além disso, a mixagem de linhas espectrais ou informações espectrais relacionadas a componentes espectrais pode ser feita por uma soma qualificada das linhas espectrais de

fonte ou informações espectrais. Fatores de qualificação podem ser zero ou um, ou, em princípio, qualquer valor entre eles. Um valor de zero significa que as fontes são tratadas como irrelevantes e não serão usadas. Grupos de linhas, tais como bandas ou bandas de fator de escala podem usar o mesmo fator de qualificação. Os fatores de qualificação (por exemplo, uma distribuição de zeros e um) podem ser variados para os componentes espectrais de uma única estrutura de um único fluxo de dados de entrada. As configurações descritas abaixo não são de forma alguma exigidas exclusivamente a usar os fatores de qualificação de zero e um ao mixar informações espectrais. Pode acontecer, em determinadas circunstâncias que, não para uma única, uma pluralidade de informações espectrais gerais de uma estrutura de um fluxo de dados de entrada, os respectivos fatores de qualificação sejam diferentes de zero ou um.

[000181] Um caso em especial se dá quando todas as bandas ou componentes espectrais de uma fonte (fluxo de dados de entrada) são ajustadas para um fator de um e todos os fatores das outras fontes são ajustados para zero. Neste caso, o fluxo total de bits de entrada de um participante pode igualmente ser copiado como fluxo de bits mixados final. Os fatores de qualificação podem ser calculados em uma base de estrutura para estrutura, mas podem também ser calculados ou determinados em grupos mais longos ou seqüências de estruturas. Normalmente, mesmo dentro de tal seqüência de estruturas ou dentro de

estruturas únicas, os fatores de qualificação podem diferir para diferentes componentes espectrais, conforme citado acima. Os fatores de qualificação podem, em algumas configurações, ser calculados ou determinados de acordo com resultados do modelo psico-acústico.

[000182] Tal comparação pode, por exemplo, ser feita baseada na avaliação de uma taxa de energia entre o sinal mixado onde apenas alguns fluxos de entrada são incluídos, e um completo sinal mixado. Isto pode, por exemplo, ser alcançado conforme descrito acima com relação às equações (3) a (5). Em outras palavras, o modelo psico-acústico pode calcular a razão de energia  $r(n)$  entre o sinal mixado onde apenas alguns fluxos de entrada são incluídos levando a um valor de energia  $E_f$  e o completo sinal mixado possuindo um valor de energia  $E_c$ . A razão de energia  $r(n)$  é então calculada de acordo com a equação (5) como 20 vezes o logaritmo de  $E_f$  dividido por  $E_c$ .

[000183] De acordo com isso, de forma similar à descrição acima das configurações com relação às fig. 6 a 8, se a razão for suficientemente grande, os canais menos dominantes podem ser considerados mascarados pelos dominantes. Desta forma, uma redução de irrelevância é processada, significando que apenas tais fluxos são incluídos, que não forem de forma alguma perceptíveis, aos quais, o valor de qualificação de um é atribuído, enquanto todos os outros fluxos - pelo menos uma informação espectral deste componente espectral - são

descartados. Em outras palavras, a estes, um fator de qualificação de zero, é atribuído.

[000184] Isto pode levar a uma vantagem adicional que menos ou nenhum efeito de codificação tandem ocorre devido a um número reduzido de etapas de re-quantização. Uma vez que cada etapa de quantização revela um risco significativo de redução de ruído de quantização adicional, a qualidade geral do sinal de áudio pode, conseqüentemente, ser melhorada.

[000185] De forma semelhante às configurações acima descritas das fig. 6 a 8, as configurações descritas abaixo podem ser usadas com um sistema de conferência que pode, por exemplo, ser um sistema de tele/vídeo conferência com mais de dois participantes, e pode oferecer a vantagem de uma menor complexidade em comparação a uma mixagem de domínio de tempo, uma vez que as etapas de transformação de tempo-freqüência e etapas de recodificação podem ser omitidas. Além disso, nenhum outro atraso é causado por estes componentes em comparação à mixagem no domínio de tempo-freqüência devido à ausência do atraso de banco de filtros.

[000186] Fig. 9 representa um diagrama em bloco simplificado de um equipamento 500 para mixagem de fluxos de dados de entrada de acordo com uma configuração da presente invenção. A maioria dos sinais de referência foi adotada das configurações das fig. 6 a 8, no sentido de facilitar o entendimento e evitar descrições em duplicata. Outros sinais de

referência foram aumentados por 1000 no sentido de simbolizar que a funcionalidade dos mesmos é definida de forma diferente em comparação às configurações acima das fig. 6 a 8 - seja em funcionalidades adicionais ou funcionalidade alternativa, mas com as funções gerais do respectivo elemento, sendo comparáveis.

[000187] Com base no primeiro fluxo de dados de entrada 510-1, e um segundo fluxo de dados de entrada 510-2, uma unidade de processamento 1520 compreendida no equipamento 1500 é adaptada para gerar um fluxo de dados de saída 1530. O primeiro e o segundo fluxo de dados de entrada 510, cada um, compreende uma estrutura 540-1, 540-2, respectivamente, que, cada uma, compreende um valor de controle 1545-1, 1545-2, respectivamente, que indicam uma forma em que os dados de carga útil da estrutura 540 representam pelo menos uma parte do domínio espectral ou informações espectrais de um sinal de áudio.

[000188] O fluxo de dados de saída 530 também compreende uma estrutura de saída 1550 com um valor de controle 555, indicando de uma maneira parecida, uma forma em que os dados de carga útil da estrutura de saída 550 representam informações espectrais no domínio espectral do sinal de áudio codificado no fluxo de dados de saída 530.

[000189] A unidade de processamento 1520 do equipamento 1500 é adaptada para comparar os valores de controle 1545-1 da estrutura 540-1 do primeiro fluxo de dados de entrada 510-1 e o valor de controle 1545-2 de uma estrutura 540-2 do segundo fluxo

de dados de entrada 510-2 para chegar a um resultado comparativo. Com base neste resultado comparativo, a unidade de processamento 1520 é, além disso, adaptada para gerar o fluxo de dados de entrada 530 compreendendo a estrutura de saída 550, de forma que, quando o resultado comparativo indicar que os valores de controle 1545 das estruturas 540 do primeiro e do segundo fluxo de dados de entrada 510 forem idênticos ou iguais, a estrutura 550 compreenda como o valor de controle 1550, um valor igual ao dos valores de controle 1545 das estruturas 540 dos dois fluxos de dados de entrada 510. Os dados de carga útil compreendidos na estrutura de saída 550 são derivados dos dados de carga útil correspondentes das estruturas 540 com relação aos valores de controle idênticos 1545 das estruturas 540, por processamento no domínio espectral, por exemplo, sem visitar o domínio de tempo.

[000190] Se, por exemplo, os valores de controle 1545 indicarem uma codificação especializada de informações espectrais de um ou mais componentes espectrais (por exemplo, dados PNS), e os respectivos valores de controle 1545 dos dois fluxos de dados de entrada forem idênticos, então as informações espectrais correspondentes da estrutura de saída 550, correspondente ao mesmo componente espectral ou componentes espectrais, podem ser obtidas pelo processamento dos dados de carga útil correspondentes no domínio espectral, mesmo diretamente, que não esteja deixando o tipo de representação do

domínio espectral. Conforme será delineado a seguir, no caso de uma representação baseada em PNS, isto pode ser alcançado pela soma dos respectivos dados PNS, opcionalmente acompanhados por um processo de normalização. Isto é, os dados PNS de nenhum dos fluxos de dados de entrada é convertido de volta para a plena representação com um valor por amostra espectral.

[000191] Fig. 10 representa um diagrama mais detalhado de um equipamento 1500 que difere da fig. 9 principalmente com relação a uma estrutura interna da unidade de processamento 1520. Para ser mais específico a unidade de processamento 1520 compreende um comparador 1560, que é acoplado a entradas apropriadas para o primeiro e o segundo fluxos de dados de entrada 510 e que é adaptado para comparar os valores de controle 1545 de suas respectivas estruturas 540. Os fluxos de dados de entrada são, além disso, transmitidos a um transformador opcional 1570-1, 1570-2, para cada um dos dois fluxos de dados de entrada 510. O comparador 1560 é também acoplado aos transformadores opcionais 1570 para prover os mesmos com os resultados comparativos.

[000192] A unidade de processamento 1520 compreende ainda um mixer 1580, que é acoplado no sentido da entrada aos transformadores opcionais 1570 - ou, caso, um ou mais dos transformadores 1570 não sejam implementados - às entradas correspondentes para os fluxos de dados de entrada 510. O mixer 1580 é acoplado com uma saída a um normalizador opcional 1590,

que por sua vez é acoplado, se implementado, com uma saída da unidade de processamento 1520 e a do equipamento 1500 para prover o fluxo de dados de saída 530.

[000193] Conforme citado anteriormente, o comparador 1560 é adaptado para comparar os valores de controle das estruturas 1540 dos dois fluxos de dados de entrada 510. O comparador 1560 provê, quando implementado, os transformadores 1570 com um sinal indicando se os valores de controle 1540 das respectivas estruturas 540 são idênticos, ou não. Se o sinal representando o resultado comparativo indicar que os dois valores de controle 1545 são, pelo menos com relação a um componente espectral, idênticos ou iguais, os transformadores 1570 não transformam os respectivos dados de carga útil conforme compreendido nas estruturas 540.

[000194] Os dados de carga útil compreendidos nas estruturas 540 dos fluxos de dados de entrada 510 são então mixados pelo mixer 1580 e enviados ao normalizador 1590, se implementado, para realizar uma etapa de normalização no sentido de assegurar que os valores resultantes não excedam ou restrinjam o âmbito dos valores. Exemplos de mixagem de dados de carga útil serão resumidos em mais detalhes a seguir no contexto com as fig. 12a a 12c.

[000195] O normalizador 1590 pode ser implementado como um quantizador adaptado para re-quantizar os dados de carga útil de acordo com seus respectivos valores, alternativamente, o

normalizador 1590 pode também ser adaptado para apenas alterar um fator de escala indicando a distribuição de etapas de quantização ou um valor absoluto de um nível mínimo ou máximo de quantização, dependendo da concreta implementação do mesmo.

[000196] Caso o comparador 1560 indicar que os valores de controle 1545 estão pelo menos com relação a um ou mais componentes espectrais diferentes, o comparador 1560 pode prover um ou os dois transformadores 1570 com um respectivo sinal de controle indicando os respectivos transformadores 1570 para transformar os dados de carga útil de pelo menos um dos fluxos de dados de entrada 510 para os do outro fluxo de dados de entrada. Neste caso, o transformador pode ser adaptado para simultaneamente mudar o valor de controle da estrutura transformada de forma que o mixer 1580 seja capaz de gerar a estrutura de saída 550 do fluxo de dados de saída 530 com um valor de controle 1555 sendo igual ao da estrutura 540 dos dois fluxos de dados de entrada que não seja transformado ou que seja um valor comum de dados de carga útil de ambas as estruturas 540.

[000197] Exemplos mais detalhados serão descritos a seguir no contexto das fig. 12a a 12c para diferentes aplicações, tais como, implementações PNS, implementações SBR, e implementações M/S, respectivamente.

[000198] Deve ser mencionado que as configurações das fig. 9 a 12C não estão de forma alguma limitadas a dois fluxos

de dados de entrada 1510-1, 1510-2, conforme representado nas fig. 9, 10 e na fig. 11 a seguir. Mais propriamente, a mesma deve ser adaptada para processar uma pluralidade de fluxo de dados de entrada compreendendo mais de dois fluxos de dados de entrada 510. Neste caso, o comparador 560 pode, por exemplo, ser adaptado para comparar um número adequado de fluxos de dados de entrada 510 e as estruturas 540 ali compreendidas. Além disso, dependendo da implementação concreta, um número adequado de transformadores 1570 também pode ser implementado. O mixer 1580 junto com o normalizador opcional 1590 pode eventualmente ser adaptado para o número incrementado de fluxos de dados a serem processados.

[000199] No caso de mais de apenas dois fluxos de dados de entrada 510, o comparador 1560 pode ser adaptado para comparar todos os valores de controle relevantes 1545 do fluxo de dados de entrada 510 para decidir se uma etapa de transformação deve ser realizada por um ou mais transformadores opcionais implementados 1570. Alternativamente, ou adicionalmente, o comparador 1560 pode também ser adaptado para determinar um conjunto de fluxos de dados de entrada a serem transformados pelo transformador 1570, quando o resultado da comparação indicar que uma transformação para uma maneira comum de representação dos dados de carga útil pode ser alcançada. Por exemplo, a menos que a representação diferente de dados de carga útil envolvida exija uma determinada representação, o comparador

1560 pode, por exemplo, ser adaptado para ativar os transformadores 1570 de tal forma a minimizar a complexidade total. Isto pode, por exemplo, ser alcançado com base em estimativas predeterminadas de valores de complexidade armazenados dentro do comparador 1560 ou disponíveis ao comparador 1560 de forma diferente.

[000200] Além disso, deve ser mencionado que o transformador 1570 pode eventualmente ser omitido quando, por exemplo, uma transformação para o domínio de frequência possa opcionalmente ser realizada pelo mixer 1580 a pedidos. Alternativamente, ou adicionalmente, a funcionalidade dos transformadores 1570 também pode ser incorporada no mixer 1580.

[000201] Além disso, deve ser mencionado que as estruturas 540 podem compreender mais de um valor de controle, tal como substituição perceptual de ruídos (PNS), ajuste temporal de ruídos (TNS) e modos de codificação estéreo. Antes de descrever a operação de um equipamento capaz de processar pelo menos um dos parâmetros PNS, parâmetros TNS ou parâmetros de codificação estéreo, referência é feita à fig. 11 que equivale à Fig. 8 com, no entanto, os sinais de referência 1500 e 1520 sendo usados ao invés dos 500 e 520, respectivamente, no sentido de mostrar que a fig. 8 já representa uma configuração para gerar um fluxo de dados de saída de um primeiro e de um segundo fluxo de dados de entrada nos quais a unidade de processamento 520 e 1520, respectivamente, pode também ser

adaptada para realizar a funcionalidade descrita com relação às fig. 9 e 10. Particularmente, dentro da unidade de processamento 1520, a unidade de mixagem 800 compreendendo o mixer espectral 810, o módulo de otimização 820, e o mixer SBR 830 realiza as funções anteriormente descritas apresentadas com relação às fig. 9 e 10. Conforme anteriormente citado, os valores de controle compreendidos nas estruturas dos fluxos de dados de entrada podem igualmente ser parâmetros PNS, parâmetros SBR, ou dados de controle relacionados à codificação estéreo, em outras palavras, parâmetros M/S. Caso os respectivos valores de controle sejam iguais ou equivalentes, a unidade de mixagem 800 pode processar os dados de carga útil para gerar dados de carga útil correspondentes a serem posteriormente processados para serem compreendidos na estrutura de saída do fluxo de dados de saída. Com relação a isto, conforme já citado acima, uma vez SBR permite que dois canais de codificação estéreo, codifiquem o canal esquerdo e o canal direito separadamente, assim como, codificando os mesmos por meio de um canal de acoplagem (C), de acordo com uma configuração da presente invenção, processando os respectivos parâmetros SBR ou pelo menos partes dos mesmos, pode compreender o processamento de elementos C dos parâmetros SBR para obter ambos, elementos esquerdos e direitos do parâmetro SBR, ou vice-versa, dependendo dos resultados da comparação e dos resultados da determinação. De maneira semelhante, o grau de processamento de informações espectrais e/ou respectivos

parâmetros relacionados a componentes espectrais e informações espectrais (por exemplo, parâmetros TNS, parâmetros SBR, parâmetros PNS) podem ser baseadas em diferentes números de dados a serem processados e podem determinar se as informações espectrais subordinadas ou peças delas também precisam ser decodificadas. Por exemplo, no caso da cópia de dados SBR, pode ser aconselhável processar toda a estrutura do respectivo fluxo de dados para evitar a complicada mixagem de informações espectrais para diferentes componentes espectrais. A mixagem destes pode exigir uma re-quantização que pode de fato reduzir o ruído de quantização. Em termos de parâmetros TNS pode também ser aconselhável decompor os respectivos parâmetros TNS junto com as informações espectrais de toda estrutura do fluxo dominante de dados de entrada para o fluxo de dados de saída para evitar a re-quantização. No caso de informações espectrais baseadas em PNS, o processamento de valores individuais de energia sem copiar os componentes espectrais subordinados pode ser uma forma viável. Além disso, neste caso, por meio do processamento apenas dos respectivos parâmetros PNS do componente espectral dominante das estruturas da pluralidade de fluxos de dados de entrada para o componente espectral correspondente da estrutura de saída do fluxo de dados de saída, ocorre sem a introdução de ruídos adicionais de quantização. É necessário frisar que também por meio da quantização de um valor de energia em forma de um parâmetro PNS, um ruído adicional de

quantização pode ser introduzido.

[000202] Com relação às fig. 12a a 12C, três diferentes modos de mixagem dos dados de carga útil com base em uma comparação dos respectivos valores de controle serão descritos em mais detalhes. A fig. 12a representa um exemplo de uma implementação baseada em PNS de um equipamento 500 de acordo com uma configuração da presente invenção, onde a fig. 12b representa uma implementação SBR semelhante e a fig. 12c mostra uma implementação M/S da mesma.

[000203] Fig. 12a representa um exemplo com um primeiro e um segundo fluxo de dados de entrada 510-1, 510-2, respectivamente, com estruturas de entrada apropriadas 540-1, 540-2, e respectivos valores de controle 545-1, 545-2. Conforme mostrado pelas setas na fig. 11a, os valores de controle 1545 das estruturas 540 dos fluxos de dados de entrada 510 indicam que um componente espectral não está descrito em termos de informações espectrais indiretamente, mas em termos de um valor de energia de uma fonte de ruídos, ou, em outras palavras, por meio de um parâmetro PNS adequado. Mais especificamente, a fig. 12a mostra um primeiro parâmetro PNS 2000-1 e a estrutura 540-2 do segundo fluxo de dados de entrada 510-2 compreende um parâmetro PNS 2000-2.

[000204] Uma vez que, conforme assumido com relação à fig. 12a, os valores de controle 1545 das duas estruturas 540 dos dois fluxos de dados de entrada 510, indicam que o

componente espectral específico deve ser substituído por seu respectivo parâmetro PNS 2000, a unidade de processamento 1500, conforme anteriormente descrita, é capaz de mixar os dois parâmetros PNS 2000-1, 2000-2 para chegarem a um parâmetro 2000-3 da estrutura de saída 550 a ser incluída dentro do fluxo de dados de saída 530. O respectivo valor de controle 1555 da estrutura de saída 550 essencialmente também indica que o respectivo componente espectral será substituído pelo parâmetro PNS 2000-3 mixado. Este processo de mixagem é mostrado na fig. 12a pela representação do parâmetro PNS 2000-3 como sendo a combinação dos parâmetros 2000-1, 2000-2 das respectivas estruturas 540-1, 540-2.

[000205] No entanto, a determinação dos parâmetros PNS 2000-3, o que também é citado como parâmetro de saída PNS, também pode ser feito com base em uma combinação linear de aqodo com

$$[000206] \quad PNS = \sum_{i=1}^N a_i \cdot PNS(i), \quad (6)$$

[000207] onde  $PNS(i)$  é o respectivo parâmetro PNSS do fluxo de dados de entrada  $I$ ,  $N$  é o número de fluxos de dados de entrada a serem mixados e  $a_i$  é um fator apropriado de qualificação. Dependendo da concreta implementação, os fatores de qualificação  $a_i$  podem ser escolhidos para serem iguais a

$$[000208] \quad a_1 = \dots = a_N \quad . \quad (7)$$

[000209] Uma implementação direta, que é mostrada na

fig. 12a, pode ser aquela, quando todos os parâmetros de qualificação  $a_i$  são iguais a 1, em outras palavras,

$$[000210] \quad a_1 = \dots = a_N = 1 \quad . \quad (8)$$

[000211] Caso um normalizador 1590 conforme mostrado na fig. 10 deva ser omitido, os fatores de qualificação podem igualmente ser definidos para serem iguais a  $1/N$ , de forma que a equação

$$[000212] \quad a_1 = \dots = a_N = \frac{1}{N} \quad (9)$$

[000213] se conserve.

[000214] O parâmetro  $N$  aqui é o número de fluxos de dados de entrada a serem mixados, e o número de fluxos de dados de entrada transmitido ao equipamento 500, é um número semelhante. Por questões de simplicidade, deve ser notado que também diferentes normalizações em termos dos fatores de qualificação  $a_1$ , podem ser implementados.

[000215] Em outras palavras, no caso de uma ferramenta PNS ativada no lado do participante, o fator de energia de ruído substitui um fator de escala apropriado junto com os dados quantizados em um componente espectral (por exemplo, banda espectral). À parte deste fator, outros dados não serão fornecidos ao fluxo de dados de entrada pela ferramenta PNS. No caso da mixagem de componentes espectrais, podem resultar dois casos distintos.

[000216] Conforme descrito acima, quando os respectivos

componentes espectrais de todas as estruturas 540 dos fluxos relevantes de dados de entrada são, cada qual, expressos em termos de parâmetros PNS. Uma vez que os dados de frequência de uma descrição relacionada a PNS de um componente de frequência (por exemplo, banda de frequência) são diretamente derivados do fator de energia de ruído (parâmetro PNS), os fatores apropriados podem ser mixados pela simples adição dos respectivos valores. O parâmetro PNS mixado gerará então dentro do decodificador PNS no lado do receptor, uma resolução de frequência equivalente a ser mixada com os puros valores espectrais de outros componentes espectrais. Caso seja usado um processo de normalização durante a mixagem, poderá ser interessante implementar um fator de normalização semelhante em termos do fator de qualificação  $a_1$ . Por exemplo, ao normalizar com um fator proporcional a  $1/N$ , os fatores de qualificação  $a_1$  podem ser escolhidos de acordo com a equação (9).

[000217] Caso os valores de controle 1545 de pelo menos um fluxo de dados de entrada 510 forem diferentes em relação a uma componente espectral, e se os respectivos fluxos de dados de entrada não tiverem de ser descartados devido a um baixo nível de energia, pode ser aconselhável que o decodificador PNS conforme mostrado na fig. 11 gere as informações espectrais ou dados espectrais, com base nos parâmetros PNS, e mixe os respectivos dados da estrutura do mixer espectral 810 da unidade de mixagem ao invés de mixar os parâmetros PNS na estrutura do

módulo de otimização 820.

[000218] Devido à independência dos componentes espectrais PNS um em relação ao outro, e com relação a parâmetros globalmente definidos do fluxo de dados de saída, assim como, dos fluxos de dados de entrada, uma seleção do método de mixagem pode ser adaptada em uma base no sentido da banda. Caso tal mixagem baseada em PNS não seja possível, pode ser aconselhável considerar uma recodificação do respectivo componente espectral pelo codificador PNS 1880 após a mixagem no domínio espectral.

[000219] Fig. 12b representa outro exemplo de um princípio operacional de uma configuração de acordo com uma configuração da presente invenção. Para ser mais preciso, a fig. 12b mostra o caso de dois fluxos de dados de entrada 510-1, 510-2 com as estruturas adequadas 540-1, 540-2 e seus valores de controle 1545-1, 1545-2. As estruturas 540 compreendem dados SBR para componentes espectrais acima, uma assim chamada frequência cruzada  $f_x$ . O valor de controle 1545 compreende informações sobre o fato de parâmetros SBR serem usados, e informações relacionadas à grade de estrutura atual ou grade de tempo/frequência.

[000220] Conforme caracterizado acima, a ferramenta SBR replica em uma banda espectral superior acima das partes das de frequências cruzadas  $f_x$  do espectro replicando uma parte inferior de um espectro que é codificado de forma diferente. A

ferramenta SBR determina um número de conectores para cada estrutura SBR que é igual às estruturas 540 do fluxo de dados de entrada 510 compreendendo também outras informações espectrais. Os conectores de tempo separam o âmbito de frequência da ferramenta SBR em pequenas bandas de frequência igualmente espaçadas ou componentes espectrais. O número destas bandas de frequência em uma estrutura SBR será determinado pelo emissor ou pela ferramenta SBR antes da codificação. Em caso de um MPEG-4 AAC-ELD, o número de conectores de tempo é fixado para ser 16.

[000221] Os conectores de tempo não estão incluídos nos assim chamados envelopes de forma que tais envelopes compreendem pelo menos dois ou mais conectores de tempo formando um respectivo grupo. Cada envelope é atribuído a um número de dados de frequências SBR. Na grade de estruturas ou grade de tempo/frequência, o número e o comprimento em unidade de conectores de tempo dos envelopes individuais, é armazenado.

[000222] A resolução de frequência do envelopes individuais determina quantos dados de energia SBR são calculados para um envelope e armazenados com relação ao mesmo. A ferramenta SBR só diferencia entre uma resolução baixa e alta, onde o envelope compreendendo uma alta resolução compreende o dobro dos valores de um envelope com baixa resolução. O número de valores de frequência ou componentes espectrais para envelopes compreendendo uma alta ou uma baixa resolução depende de outros parâmetros do codificador, tais como, taxas de bits,

freqüência de amostragem e assim por diante.

[000223] No contexto de MPEG-4 AAC ELD a ferramenta SBR muitas vezes utiliza 16 a 14 valores com relação ao envelope que possui uma alta resolução.

[000224] Devido à divisão dinâmica da estrutura 540 com um número adequado de valores de energia com relação à freqüência, uma forma transiente foi considerada. No caso em que um transiente está presente em uma estrutura, o codificador SBR divide a respectiva estrutura em números apropriados de envelopes. Esta distribuição é padronizada no caso de uma ferramenta SBR usada com um codec AAC ELD e depende da posição de transposição em unidades do conversor de tempo. Em muitos casos, a estrutura de grade resultante, ou grade de tempo/freqüência compreende três envelopes quando um transiente está presente. Um primeiro envelope, o envelope inicial, compreende o início de uma estrutura até o conector de tempo recebendo o transiente possuindo os índices de conectores de tempo zero para o inversor-1. O segundo envelope compreende um comprimento de dois conectores de tempo envolvendo o transiente do índice do inversor de tempo para o inversor +2. O terceiro envelope compreende todos os conectores de tempo restantes com os índices do inversor +3 a 16.

[000225] No entanto, o comprimento mínimo de um envelope é de dois conectores de tempo. Conseqüentemente, estruturas compreendendo um transiente perto das bordas da estruturas podem

eventualmente compreender apenas dois envelopes. Caso não houver transiente presente na estrutura, os conectores de tempo são distribuídos sobre envelopes do mesmo comprimento.

[000226] Fig. 12b representa uma grade de tempo/frequência ou grade de estruturas dentro da estrutura 540. Caso os valores de controle 1545 indiquem que as mesmas grades de tempo SBR ou grades de tempo/frequência estejam presentes nas duas estruturas 540-1, 540-2, os respectivos dados SBR podem ser copiados de forma semelhante ao método descrito no contexto com as equações (6) a (9) acima. Em outras palavras, neste caso a ferramenta de mixagem SBR ou o mixer SBR 830, conforme mostrado na fig. 11 pode copiar a grade de tempo/frequência ou a grade de estrutura das respectivas estruturas de entrada para a estrutura de saída 550, e calcular os respectivos valores de energia semelhantes às equações (6) a (9). Mais uma vez, em outras palavras, os dados de energia SBR da grade de estrutura podem ser mixados pela simples soma dos respectivos dados e, opcionalmente, por normalização dos respectivos dados.

[000227] Fig. 12c representa outro exemplo de um modo de operação de uma configuração de acordo com a presente invenção. Para ser mais preciso a fig. 12c representa uma implementação M/S. Mais uma vez, a fig. 12c representa dois fluxos de dados de entrada 510 junto com duas estruturas 540 e valores de controle associados 545 indicando uma forma em que a estrutura de dados de carga útil é representada, pelo menos com relação à no mínimo

um componente espectral delas.

[000228] As estruturas 540, cada, compreendem dados de áudio ou de informações espectrais de dois canais, um primeiro canal 2020, e um segundo canal 2030. Dependendo do valor de controle 1545 da respectiva estrutura 540, o primeiro canal 2020 pode ser, por exemplo, um canal esquerdo ou um canal intermediário, enquanto que o segundo canal 2030 pode ser um canal direito de um sinal estéreo, ou um canal lateral. O primeiro dos modos de codificação muitas vezes é citado como sendo um modo LR, enquanto que o segundo modo é muitas vezes citado como o modo M/S.

[000229] No modo M/S, que algumas vezes também é citado como um estéreo conjunto, o canal intermediário (M) deve ser definido como sendo proporcional à soma de canal esquerdo (L) e do canal direito (R). Muitas vezes, um fator adicional de  $\frac{1}{2}$  é incluído na definição, de forma que o canal intermediário compreende, em ambos, o domínio de tempo e o domínio de frequência, um valor médio dos dois canais estéreo.

[000230] O canal lateral é normalmente definido como sendo proporcional a uma diferença dos dois canais estéreo, citadamente, sendo proporcional a uma diferença do primeiro canal esquerdo (L) e do canal direito (R). Algumas vezes também um fator adicional de  $\frac{1}{2}$  é incluso, de forma que o canal lateral efetivamente representa a metade dos valores de variação entre os dois canais do sinal estéreo, ou a variação do canal

intermediário. Desta forma, a canal esquerdo pode ser reconstruído pela soma do canal intermediário e do canal lateral, enquanto o canal direito pode ser obtido pela subtração do canal lateral do canal intermediário.

[000231] Caso, para as estruturas 540-1 e 540-2 seja usada a mesma codificação estéreo (L/R ou M/S), uma re-transformação dos canais compreendidos na estrutura pode ser omitida, permitindo uma mixagem direta no respectivo domínio codificado L/R ou M/S.

[000232] Neste caso, a mixagem pode mais uma vez ser feita diretamente do domínio de frequência levando a uma estrutura 550 compreendida em um fluxo de dados de saída 530 possuindo o respectivo valor de controle 1555 com um valor igual aos valores de controle 1545-1, 1545-2 das duas estruturas 540. A estrutura de saída 550 compreende, de forma correspondente, dois canais 2020-3, 2030-3, derivados do primeiro e do segundo canal das estruturas do fluxo de dados de entrada.

[000233] Caso os valores de controle 1545-1, 1545-2 das duas estruturas 540 não forem iguais, pode ser aconselhável transformar uma das estruturas na outra representação com base no processo descrito acima. O valor de controle 1555 da estrutura de saída 550 pode ser ajustado de acordo ao valor indicativo da estrutura transformada.

[000234] De acordo com configurações da presente invenção, pode ser possível que os valores de controle 1545,

1555 indiquem uma representação de toda a estrutura 540, 550, respectivamente, ou os respectivos valores de controle podem ser específicos a componentes de frequência. Enquanto, no primeiro caso, os canais 2020, 2030 são codificados sobre toda a estrutura por um dos métodos específicos, no segundo caso, em princípio, cada uma das informações espectrais com relação a um componente espectral pode ser codificada de forma diferente. Naturalmente, também subgrupos de componentes espectrais podem ser descritos por uma dos valores de controle 1545.

[000235] Além disso, um algoritmo de reposição pode ser realizado na estrutura do módulo psico-acústico 950 para examinar cada uma das peças das informações espectrais relativas aos componentes espectrais subordinados (por exemplo, bandas de frequência) do sinal resultante para identificar componentes espectrais com apenas um componente único ativado. Para estas bandas, os valores quantizados do respectivo fluxo de dados de entrada de fluxos de bits de entrada podem ser copiados de um codificador sem recodificar ou re-quantizar os respectivos dados espectrais para o componente espectral específico. Sob determinadas circunstâncias todos os dados quantizados podem ser obtidos de um único sinal de entrada ativo para formar o fluxo de bits de saída ou fluxo de dados de saída, de forma que - em termos do equipamento 1500 - uma codificação sem perdas do fluxo de dados de entrada seja atingido.

[000236] Além disso, pode ser possível omitir etapas de

processamento, tais como, a análise psico-acústica dentro do codificador. Isto permite encurtar o processo de codificação e, desta forma, reduzir a complexidade computacional, uma vez que, em princípio, apenas a cópia dos dados de um fluxo de bits para outro fluxo de bits precisa ser realizada sob determinadas circunstâncias.

[000237] Por exemplo, no caso de PNS, uma reposição pode ser realizada uma vez que fatores de ruído da banda codificada PNS podem ser copiados de um dos fluxos de dados de saída para o fluxo de dados de saída. A reposição de componentes espectrais individuais com parâmetros PNS apropriados é possível, uma vez que, os parâmetros PNS são componentes espectrais específicos independentes um do outro.

[000238] No entanto, pode acontecer que uma aplicação demasiadamente agressiva do algoritmo descrito leve a uma experiência de audição degradada ou a uma indesejada redução na qualidade. Pode, conseqüentemente, ser aconselhável limitar a reposição a estruturas individuais do que a informações espectrais relacionadas a componentes espectrais. Em tal modo de operação a estimativa de relevância ou determinação de relevância, assim como, a análise de reposição podem ser realizadas sem mudanças. No entanto, uma reposição pode, neste modo de operação, somente ser realizada quando todos, ou pelo menos um número significativo de componentes espectrais dentro da estrutura ativa, forem substituíveis.

[000239] Embora isto possa levar a um número menor de reposições, uma força interna das informações espectrais pode, em alguns casos, ser melhorada, levando mesmo a uma qualidade levemente melhorada.

[000240] As configurações citadas acima, naturalmente, diferem em relação a suas implementações. Embora nas configurações anteriores uma decodificação e codificação Huffman tenham sido descritas como um esquema único de codificação de medida termodinâmica, outros esquemas de decodificação de medidas termodinâmicas podem ser usados. Além disso, a implementação de um codificador de medidas termodinâmicas ou um decodificador de medidas termodinâmicas não é de forma alguma exigida. Desta forma, embora a descrição das configurações anteriores tenha focado principalmente os codecs ACC-ELD, outros codecs também podem ser usados para transmitir os fluxos de dados de entrada para decodificação dos fluxos de dados de saída no lado do participante. Por exemplo, qualquer codec sendo baseado em, por exemplo, uma única janela sem alteração de comprimento de bloco, pode ser usado.

[000241] Como a descrição anterior das configurações mostradas nas fig. 8 e 11, por exemplo, também já demonstrou, os módulos descritos aqui não são mandatórios. Por exemplo, um equipamento de acordo com uma configuração da presente invenção pode simplesmente ser realizada pela operação nas informações espectrais das estruturas.

[000242] Deve ser mencionado que as configurações descritas acima com relação às fig. 6 a 12C podem ser concebidos de diferentes formas. Por exemplo, um equipamento 500/1500 para mixagem de uma pluralidade de fluxos de dados de entrada e sua unidade de processamento 520/1520 pode ser concebida com base em diferentes dispositivos elétricos e eletrônicos, tais como, resistores, transistores, indutores, e outros. Além disso, configurações de acordo com a presente invenção podem também ser concebidos com base somente em circuitos integrados, por exemplo, na forma de SOCs (SOC = sistema em chip), processadores tais como CPUs (CPU = unidade central de processamento), e outros circuitos integrados (IC) tais como circuitos integrados para aplicações específicas (ASIC).

[000243] Também deve ser mencionado que dispositivos elétricos fazendo parte de uma implementação discreta ou fazendo parte de um circuito integrado podem ser usados para diferentes propósitos e diferentes funções ao longo da implementação de um equipamento de acordo com uma configuração da presente invenção. Naturalmente, também uma combinação de circuitos baseada em circuitos integrados e circuitos discretos podem ser usada para implementar uma configuração de acordo com a presente invenção.

[000244] Com base em um processador, configurações de acordo com a presente invenção podem também ser implementadas com base em um programa de computador, um programa de software, ou um programa que seja executado em um processador.

[000245] Em outras palavras, dependendo de exigências de determinadas implementações de métodos inventivos, as configurações dos métodos inventivos podem ser implementadas em hardware ou em software. A implementação pode ser realizada com o uso de uma mídia de armazenagem digital, especialmente um disco, um CD ou DVD possuindo sinais legíveis de forma eletrônica armazenados neles, o que colabora com um computador ou processador programável, de tal forma que uma configuração do método inventivo seja permitida. Normalmente, uma configuração da presente invenção é, portanto, um produto de programa de computador com um código de programa armazenado em um portador com leitura mecânica, o código de programa sendo operacional para realizar uma configuração do método inventivo quando o produto do computador roda em um computador ou processador. Ainda, em outras palavras, configurações dos métodos inventivos são desta forma, um programa de computador possuindo um código de programa para realizar no mínimo uma das configurações do método inventivo, quando o programa de computador rodar em um computador ou processador. Um processador pode ser formado por um computador, um cartão de chip, um cartão inteligente, um circuito integrado específico à aplicação, um sistema em chip (SOC), ou um circuito integrado (IC).

[000246] Lista de sinais de referência  
[000247] 100 Sistema de conferências  
[000248] 110 Entrada

|          |     |                                  |
|----------|-----|----------------------------------|
| [000249] | 120 | Decodificador                    |
| [000250] | 130 | Somador                          |
| [000251] | 140 | Codificador                      |
| [000252] | 150 | Saída                            |
| [000253] | 160 | Terminal de Conferência          |
| [000254] | 170 | Codificador                      |
| [000255] | 180 | Decodificador                    |
| [000256] | 190 | Conversor de tempo/frequência    |
| [000257] | 200 | Quantizador/codificador          |
| [000258] | 210 | Decodificador/desquantizador     |
| [000259] | 220 | Conversor de frequência/tempo    |
| [000260] | 250 | Fluxo de dados                   |
| [000261] | 260 | Estrutura                        |
| [000262] | 270 | Blocos de informações adicionais |
| [000263] | 300 | Frequência                       |
| [000264] | 310 | Banda de frequência              |
| [000265] | 500 | Equipamento                      |
| [000266] | 510 | Fluxo de dados de entrada        |
| [000267] | 520 | Unidade de processamento         |
| [000268] | 530 | Fluxo de dados de saída          |
| [000269] | 540 | Estrutura                        |
| [000270] | 550 | Estrutura de saída               |
| [000271] | 560 | Componente espectral             |
| [000272] | 570 | Seta                             |
| [000273] | 580 | Linha interrompida               |

|          |     |                                |
|----------|-----|--------------------------------|
| [000274] | 700 | Decodificador de fluxo de bits |
| [000275] | 710 | Leitor de fluxo de bits        |
| [000276] | 720 | Codificador de Huffman         |
| [000277] | 730 | Desquantizador                 |
| [000278] | 740 | Escalonador                    |
| [000279] | 750 | Primeira unidade               |
| [000280] | 760 | Segunda unidade                |
| [000281] | 770 | Decodificador estéreo          |
| [000282] | 780 | Decodificador PNS              |
| [000283] | 790 | Decodificador TNS              |
| [000284] | 800 | Unidade de mixagem             |
| [000285] | 810 | Mixer espectral                |
| [000286] | 820 | Módulo otimizador              |
| [000287] | 830 | Mixer SBR                      |
| [000288] | 850 | Codificador de fluxo de bits   |
| [000289] | 860 | Terceira unidade               |
| [000290] | 870 | Codificador TNS                |
| [000291] | 880 | Codificador PNS                |
| [000292] | 890 | Codificador estéreo            |
| [000293] | 900 | Quarta unidade                 |
| [000294] | 910 | Escalonador                    |
| [000295] | 920 | Quantizador                    |
| [000296] | 930 | Codificador de Huffman         |
| [000297] | 940 | Escritor de fluxo de bits      |
| [000298] | 950 | Módulo psico-acústico          |

|          |                               |
|----------|-------------------------------|
| [000299] | 1500 Equipamento              |
| [000300] | 1520 Unidade de processamento |
| [000301] | 1545 Valor de controle        |
| [000302] | 1550 Estrutura de saída       |
| [000303] | 1555 Valor de controle.       |

**REIVINDICAÇÕES**

1. Equipamento para misturar uma pluralidade de fluxos de dados de entrada, em que os fluxos de dados de entrada compreendem cada um quadro de dados de áudio em um domínio espectral, um quadro de um fluxo de dados de entrada compreendendo informações espectrais para uma pluralidade de componentes espectrais,

o aparelho **caracterizado** pelo fato de que compreende:

uma unidade de processamento adaptada para comparar os quadros da pluralidade de fluxos de dados de entrada,

em que a unidade de processamento é ainda adaptada para determinar, com base na comparação, para um componente espectral de um quadro de saída de um fluxo de dados de saída, exatamente um fluxo de dados de entrada da pluralidade de fluxos de dados de entrada; e

em que a unidade de processamento é ainda adaptada para gerar o fluxo de dados de saída copiando pelo menos uma parte da informação de um componente espectral correspondente do quadro do fluxo de dados de entrada determinado para descrever o componente espectral do quadro de saída do fluxo de dados de saída.

2. Equipamento, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a unidade de processamento é adaptada para comparar os quadros da pluralidade de fluxos de dados de entrada com base em um modelo psicoacústico.

3. Equipamento, de acordo com a reivindicação 1, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que a comparação dos quadros da pluralidade de fluxos de

dados de entrada se baseie em pelo menos duas informações espectrais correspondentes ao mesmo componente espectral dos quadros de dois fluxos de dados de entrada diferentes.

4. Equipamento, de acordo com qualquer uma das reivindicações anteriores, **caracterizado** pelo aparelho ser adaptado de modo que um componente espectral de uma pluralidade de componentes espectrais corresponda a uma frequência ou uma banda de frequência..

5. Equipamento, de acordo com qualquer uma das reivindicações anteriores, **caracterizado** pela unidade de processamento ser adaptada de modo que a geração do fluxo de dados de saída compreenda a cópia de pelo menos parte da informação do componente espectral correspondente apenas do quadro do fluxo de dados de entrada determinado para descrever o componente espectral do quadro de saída dos dados de saída corrente.

6. Equipamento, **caracterizado** pelo fato de que a unidade de processamento é adaptada para que a geração do fluxo de dados de saída compreenda a cópia de dados de áudio no domínio espectral correspondente ao componente espectral do quadro do fluxo de dados de entrada determinado.

7. Equipamento, de acordo com a reivindicação 6, **caracterizado** pelo fato de que os fluxos de dados de entrada da pluralidade de fluxos de dados de entrada compreendem, em relação ao tempo, cada uma sequência de quadros de dados de áudio no domínio espectral e em que a unidade de processamento é adaptada de modo a comparar o quadros é baseado em quadros que correspondem apenas a um índice de tempo comum da sequência de quadros.

8. Equipamento, de acordo com a reivindicação 1,

**caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que a geração do fluxo de dados de saída mantenha uma distribuição dos níveis de quantização em comparação com uma distribuição dos níveis de quantização da pelo menos parte da informação do componente espectral correspondente da estrutura de o fluxo de entrada determinado..

9. Equipamento, de acordo com a reivindicação 1, **caracterizado** pelo fato de que pelo menos parte das informações do componente espectral correspondente compreende informações sobre níveis de quantização, um parâmetro de substituição de ruído perceptivo, um parâmetro de substituição de ruído temporal ou um parâmetro de replicação de banda espectral.

10. Método para misturar uma pluralidade de fluxos de dados de entrada, em que os fluxos de dados de entrada compreendem um quadro de dados de áudio em um domínio espectral, um quadro de um fluxo de dados de entrada compreendendo uma pluralidade de componentes espectrais,

o método **caracterizado** por:

comparar os quadros da pluralidade de fluxos de dados de entrada;

determinar, com base na comparação, para um componente espectral de um quadro de saída de um fluxo de dados de saída exatamente um fluxo de dados de entrada da pluralidade de fluxos de dados de entrada; e

gerar o fluxo de dados de saída copiando pelo menos uma parte de uma informação de um componente espectral correspondente do quadro do fluxo de dados de entrada determinado para descrever o componente espectral do quadro do fluxo de dados de

saída..

11. Método para executar, em um processador, um método para misturar uma pluralidade de fluxos de dados de entrada, em que os fluxos de dados de entrada compreendem um quadro de dados de áudio em um domínio espectral, um quadro de um fluxo de dados de entrada compreendendo uma pluralidade de componentes espectrais,

o método **caracterizado** por:

comparar os quadros da pluralidade de fluxos de dados de entrada;

determinar, com base na comparação, para um componente espectral de um quadro de saída de um fluxo de dados de saída exatamente um fluxo de dados de entrada da pluralidade de fluxos de dados de entrada; e

gerar o fluxo de dados de saída copiando pelo menos uma parte de uma informação de um componente espectral correspondente do quadro do fluxo de dados de entrada determinado para descrever o componente espectral do quadro do fluxo de dados de saída..

12. Equipamento para gerar um fluxo de dados de saída a partir de um primeiro fluxo de dados de entrada e um segundo fluxo de dados de entrada, em que o primeiro e o segundo fluxos de dados de entrada compreendem um quadro, em que os quadros compreendem um valor de controle e dados de carga útil associados, o valor de controle indicando uma maneira pela qual os dados da carga útil representam pelo menos uma parte de um domínio espectral de um sinal de áudio,

**Caracterizado** por:

uma unidade de processador adaptada para comparar o

valor de controle do quadro do primeiro fluxo de dados de entrada e o valor de controle do quadro do segundo fluxo de dados de entrada para produzir um resultado de comparação,

em que a unidade processadora é adaptada ainda mais, se o resultado da comparação indicar que os valores de controle dos quadros dos primeiro e segundo fluxos de dados de entrada são idênticos, geram o fluxo de dados de saída compreendendo um quadro de saída, de modo que o quadro de saída compreenda um valor de controle igual ao quadro dos primeiro e segundo fluxos de dados de entrada e dados de carga útil derivados dos dados de carga útil dos quadros dos primeiro e segundo fluxos de dados de entrada, processando os dados de áudio no domínio espectral.

13. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que o valor de controle do quadro da primeira ou segunda corrente de dados de entrada se relacione com pelo menos apenas um componente espectral e em que os dados de carga útil associados ao valor de controle representem uma descrição de o sinal de áudio em relação a pelo menos um componente espectral.

14. Equipamento, de acordo com a reivindicação 13, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que o valor de controle do quadro do primeiro fluxo de dados de entrada e o valor de controle do quadro do segundo fluxo de dados de entrada e os dados de carga útil associados dos quadros do o primeiro e o segundo fluxos de dados de entrada se relacionam com o mesmo componente espectral.

15. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade de processamento é adaptada

de modo que o primeiro fluxo de dados de entrada e o segundo fluxo de dados de entrada compreendam cada uma sequência de quadros em relação ao tempo, e em que a unidade processadora é adaptada para comparar os valores de controle de os quadros do primeiro e do segundo fluxos de dados de entrada para quadros associados a um índice de tempo comum dos quadros em relação à sequência de quadros.

16. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade processadora é ainda adaptada para transformar os dados de carga útil da estrutura de um dos primeiro e segundo fluxos de dados de entrada em uma representação dos dados de carga útil da estrutura da outra da primeira e segundos fluxos de dados de entrada, quando o resultado da comparação indica que os valores de controle dos primeiro e segundo fluxos de dados de entrada não são idênticos, antes de gerar o quadro de saída compreendendo um valor de controle igual ao do quadro do outro do primeiro e do segundo fluxos de dados de entrada e dados de carga útil derivados dos dados de carga útil dos quadros de um fluxo de dados de entrada e a representação transformada do outro fluxo de dados de entrada processando os dados de áudio no domínio espectral.

17. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade processadora é adaptada para gerar o quadro de saída de modo que uma distribuição dos níveis de quantização seja mantida em relação a pelo menos uma parte de pelo menos um dos quadros dos primeiro e segundo fluxos de dados de entrada.

18. Equipamento, de acordo com a reivindicação 17, **caracterizado** pelo fato de que a parte do pelo menos um quadro

corresponde apenas a um componente espectral, ao qual o valor de controle e os dados de carga útil associados ao valor de controle se referem.

19. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que os dados de carga útil do quadro de um primeiro fluxo de dados de entrada e os dados de carga útil do quadro de um segundo fluxo de dados de entrada compreendam cada um uma representação do primeiro canal de áudio e um segundo canal de áudio do sinal de áudio no domínio espectral, e em que o valor de controle do quadro do primeiro fluxo de dados de entrada e o valor de controle do quadro do segundo fluxo de dados de entrada indicam se o primeiro canal é um canal esquerdo (Canal L) e o segundo canal é o canal direito (canal R) do sinal de áudio ou se o primeiro canal é um canal intermediário (canal M) e o segundo canal é um canal lateral (canal S) de o sinal de áudio.

20. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que os valores de controle dos quadros dos primeiro e segundo fluxos de dados de entrada indiquem se os dados de carga útil associados aos respectivos valores de controle compreendem um valor relacionado à energia de uma fonte de ruído.

21. Equipamento, de acordo com a reivindicação 20, **caracterizado** pelo fato de que o valor relacionado à energia é um parâmetro de substituição de ruído perceptivo (parâmetro PNS).

22. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade de processamento é adaptada de modo que o valor de controle do quadro do primeiro fluxo de dados

de entrada e o valor de controle do quadro do segundo fluxo de dados de entrada compreenda informações sobre um envelope de dados SBR compreendidos em os dados de carga útil associados ao referido valor de controle e em que a unidade processadora está adaptada para gerar o fluxo de dados de saída em um domínio espectral SBR, quando o resultado da comparação indica envelopes idênticos.

23. Equipamento de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade processadora é ainda adaptada para comparar os quadros do primeiro e segundo fluxo de dados de entrada, em que a unidade processadora é ainda adaptada para determinar, com base na comparação dos quadros, exatamente um fluxo de dados de entrada dos primeiro e segundo fluxos de dados de entrada e em que a unidade processadora é ainda adaptada para gerar o fluxo de dados de saída copiando os dados de carga útil e o valor de controle do quadro do fluxo de entrada determinado.

24. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que o aparelho está adaptado para processar uma pluralidade de fluxos de dados de entrada compreendendo mais de dois fluxos de dados de entrada, a pluralidade de fluxos de dados de entrada compreendendo os primeiro e segundo fluxos de dados de entrada.

25. Equipamento, de acordo com a reivindicação 12, **caracterizado** pelo fato de que a unidade processadora é ainda adaptada para gerar o fluxo de dados de saída derivando os dados de carga útil do fluxo de dados de saída a partir dos dados de carga útil dos quadros dos primeiro e segundo fluxos de dados de entrada permanecendo no caminho de representação do domínio espectral, conforme indicado pelos valores de controle..

26. Método para gerar um fluxo de dados de saída a partir de um primeiro fluxo de dados de entrada e um segundo fluxo de dados de entrada, em que o primeiro e o segundo fluxos de dados de entrada compreendem um quadro, em que o quadro compreende o valor de controle e os dados de carga útil associados, indicando o valor de controle uma maneira como os dados da carga útil representam pelo menos uma parte de um domínio espectral de um sinal de áudio,

**Caracterizado por:**

comparar o valor de controle do quadro do primeiro fluxo de dados de entrada e o valor de controle do quadro do segundo fluxo de dados de entrada para produzir um resultado de comparação;  
e

se o resultado da comparação indicar que os valores de controle dos quadros dos primeiro e segundo fluxos de dados de entrada são idênticos, gerando o fluxo de dados de saída compreendendo um quadro de saída, de modo que o quadro de saída compreenda um valor de controle igual ao do quadro do primeiro e segundo fluxos de dados de entrada e dados de carga útil derivados dos dados de carga útil dos quadros dos primeiro e segundo fluxos de dados de entrada processando os dados de áudio no domínio espectral.

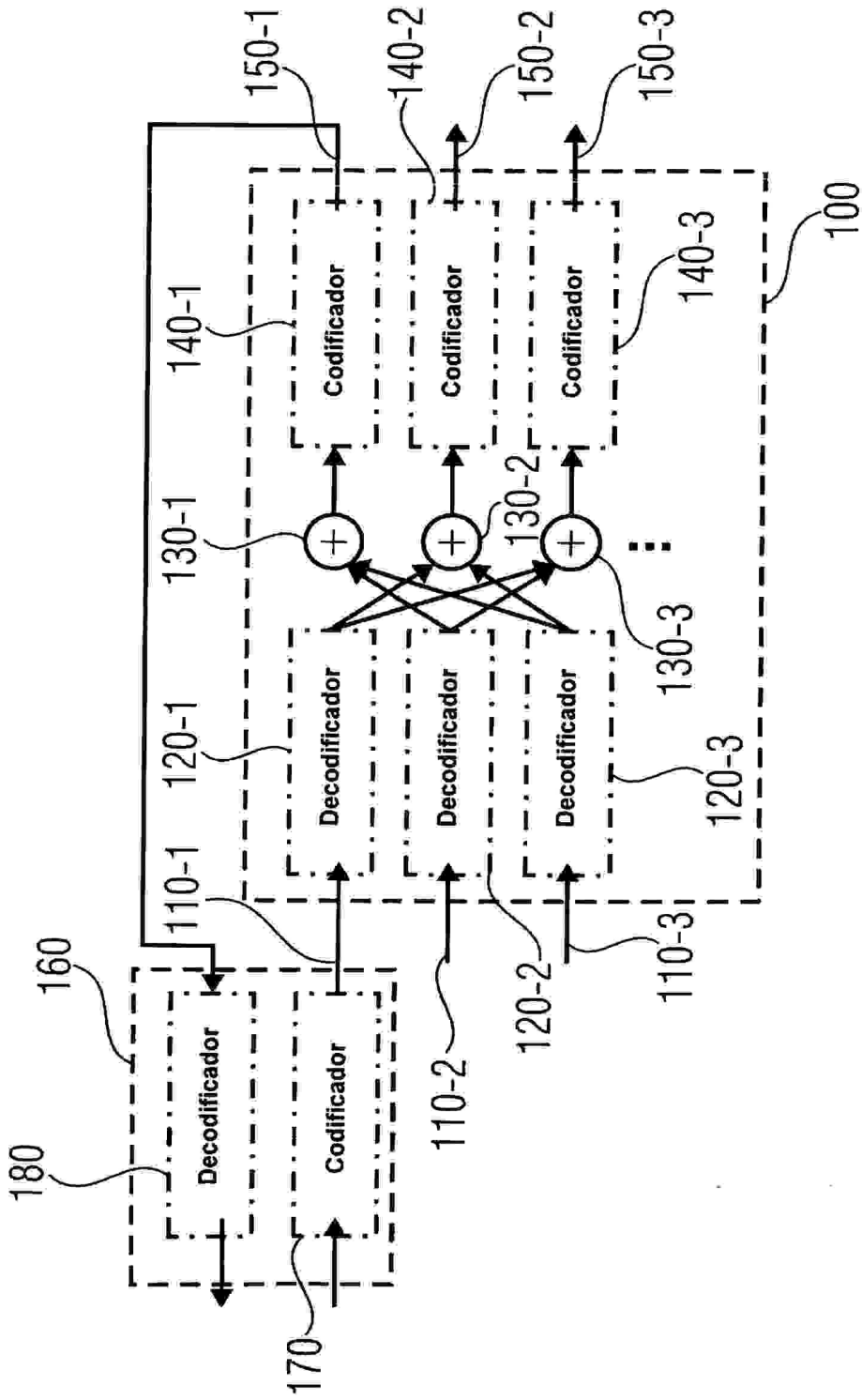


FIG 1

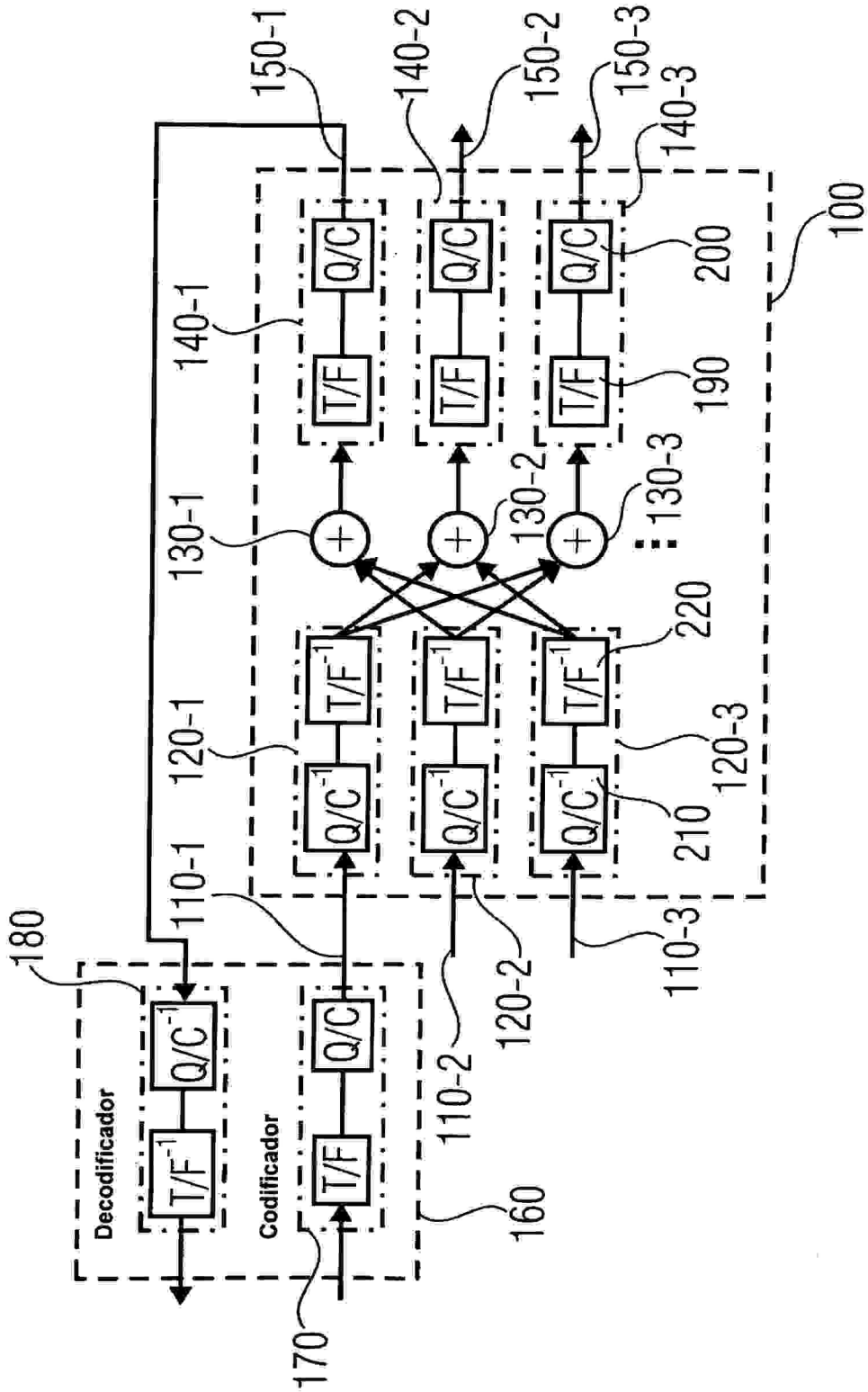


FIG 2

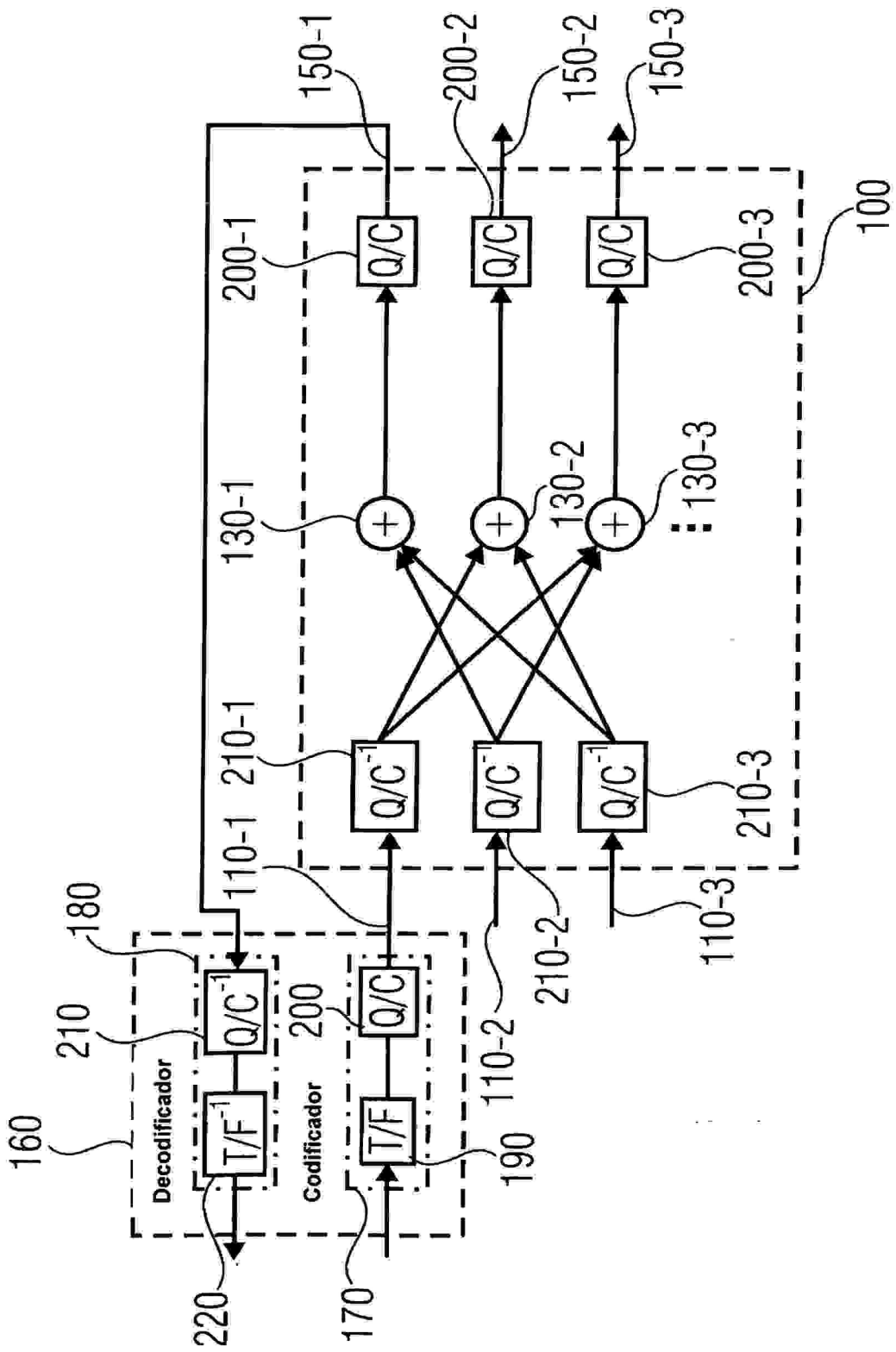


FIG 3

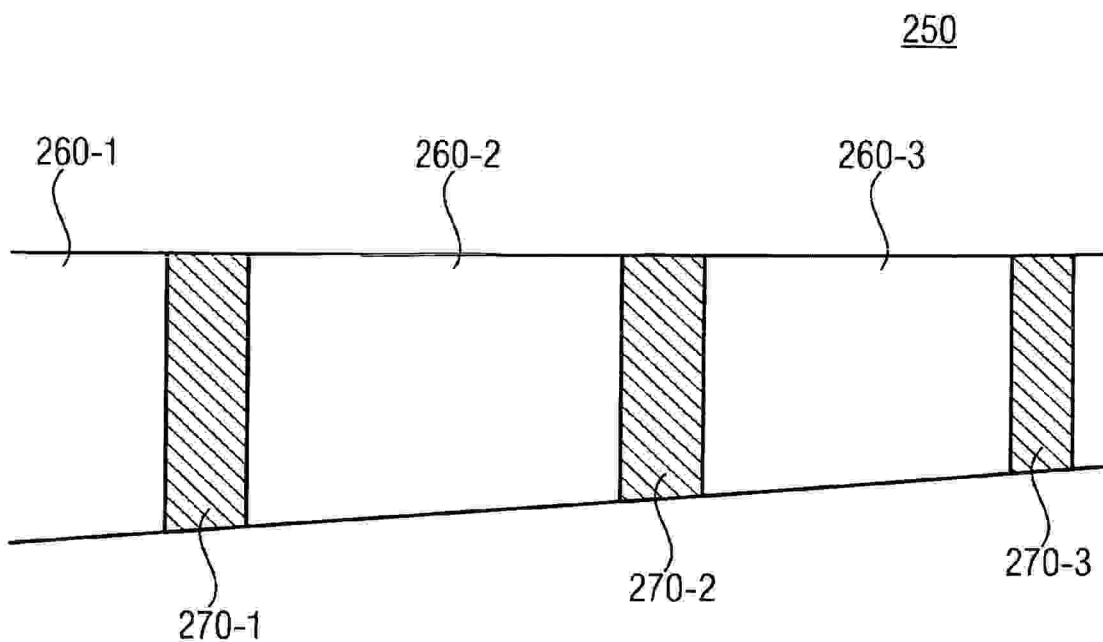


FIG 4

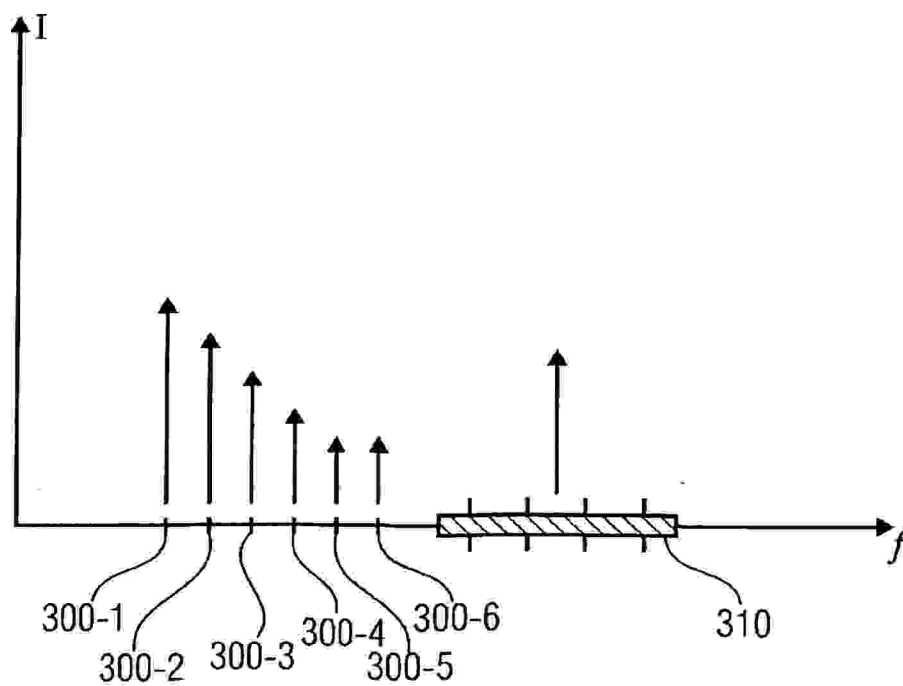


FIG 5

500

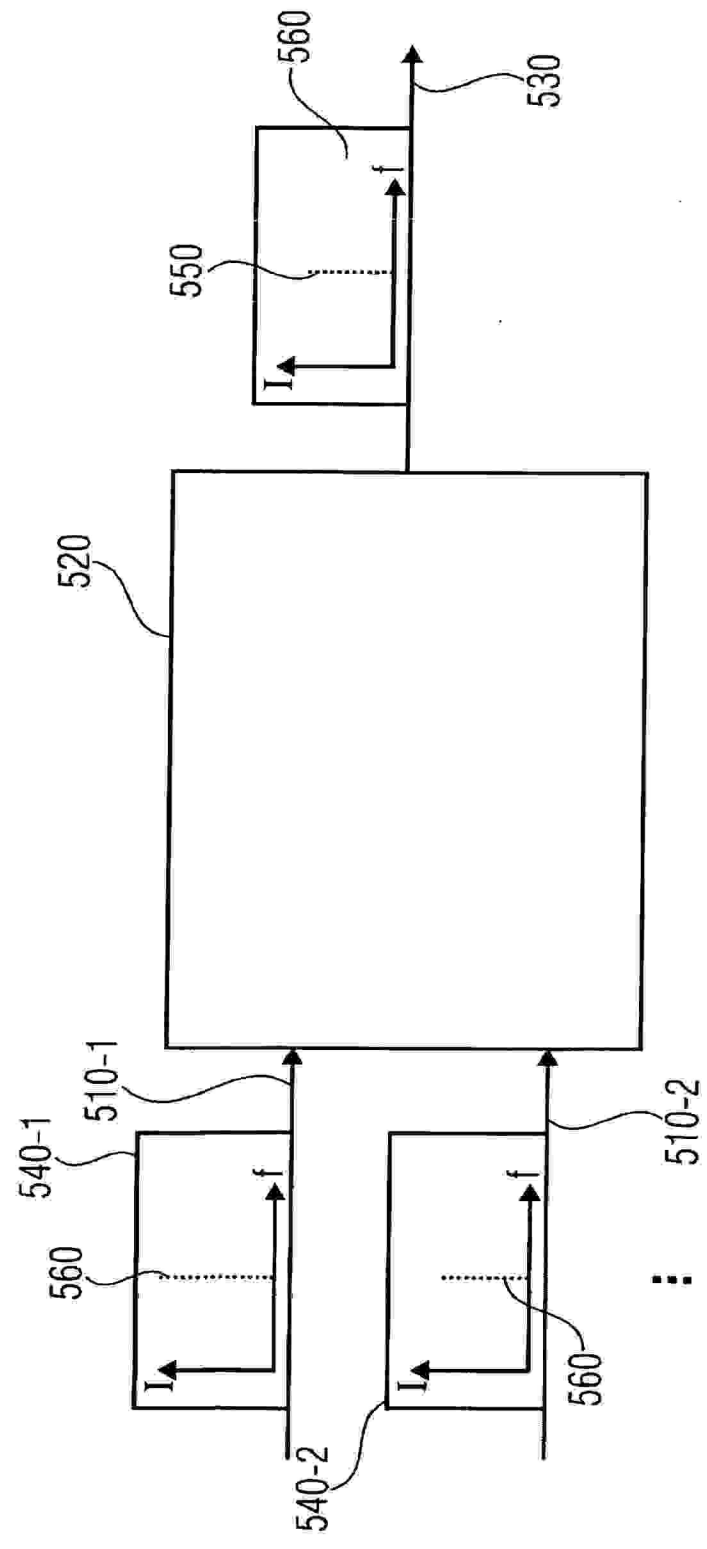


FIG 6

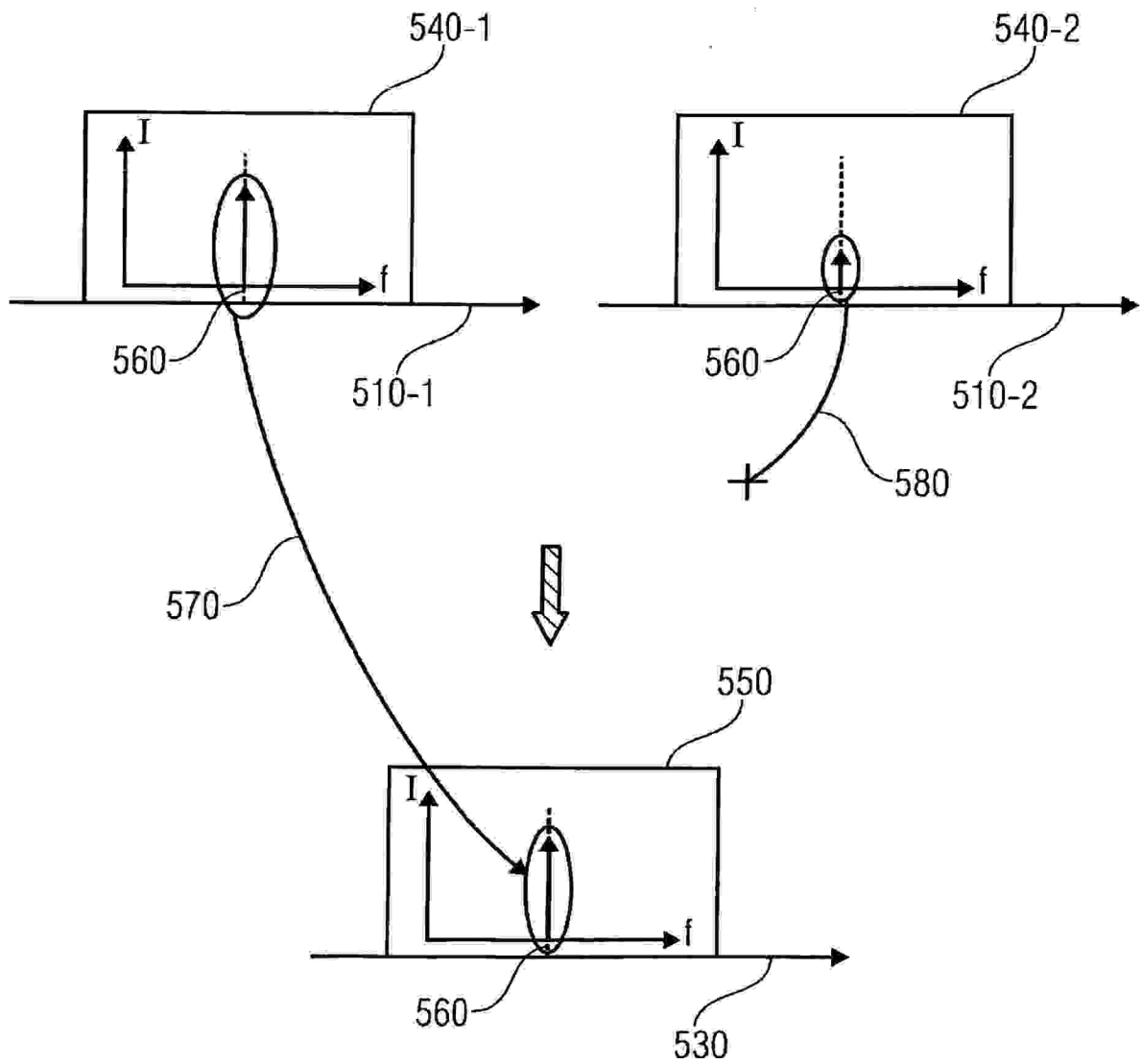


FIG 7

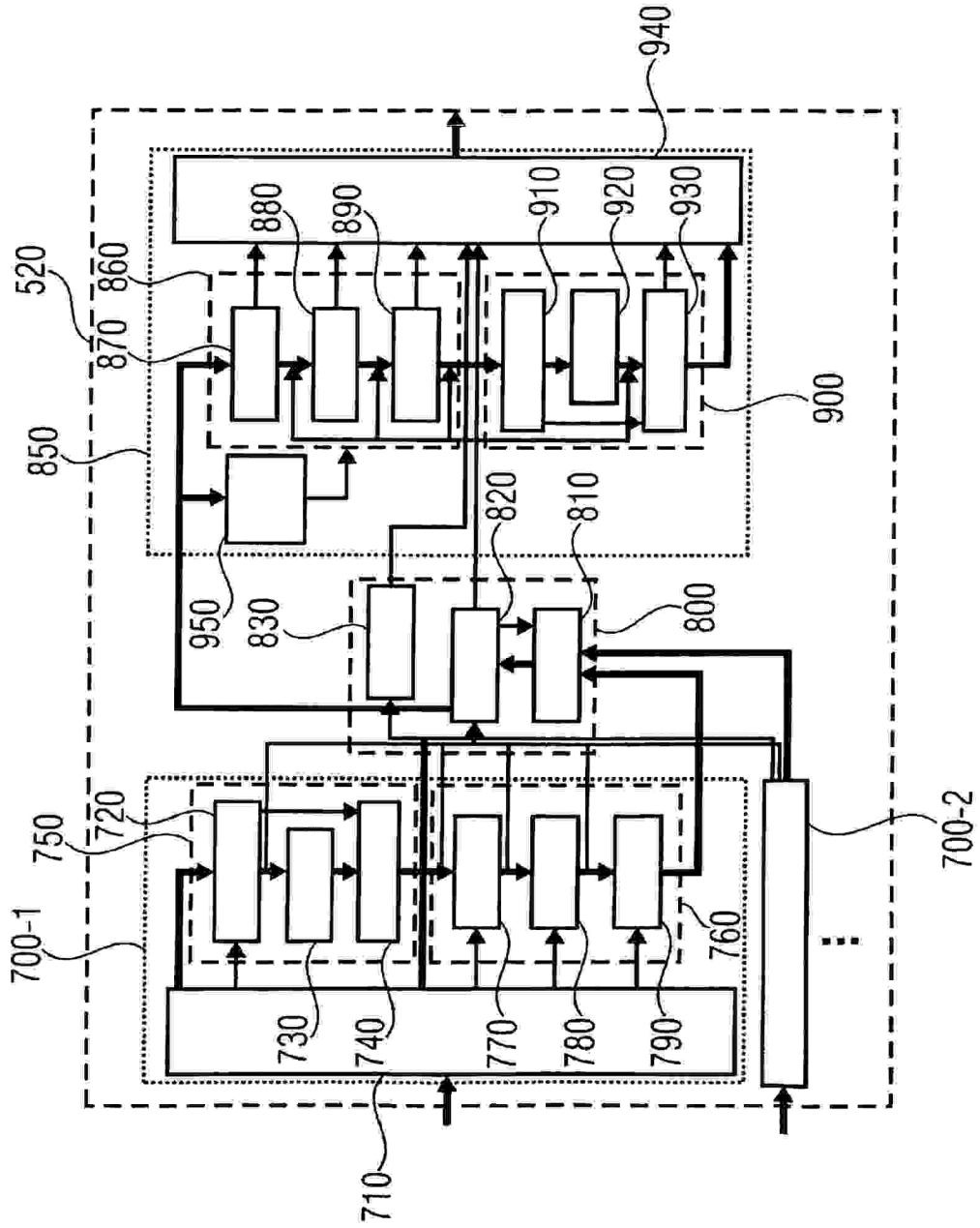


FIG 8

1500

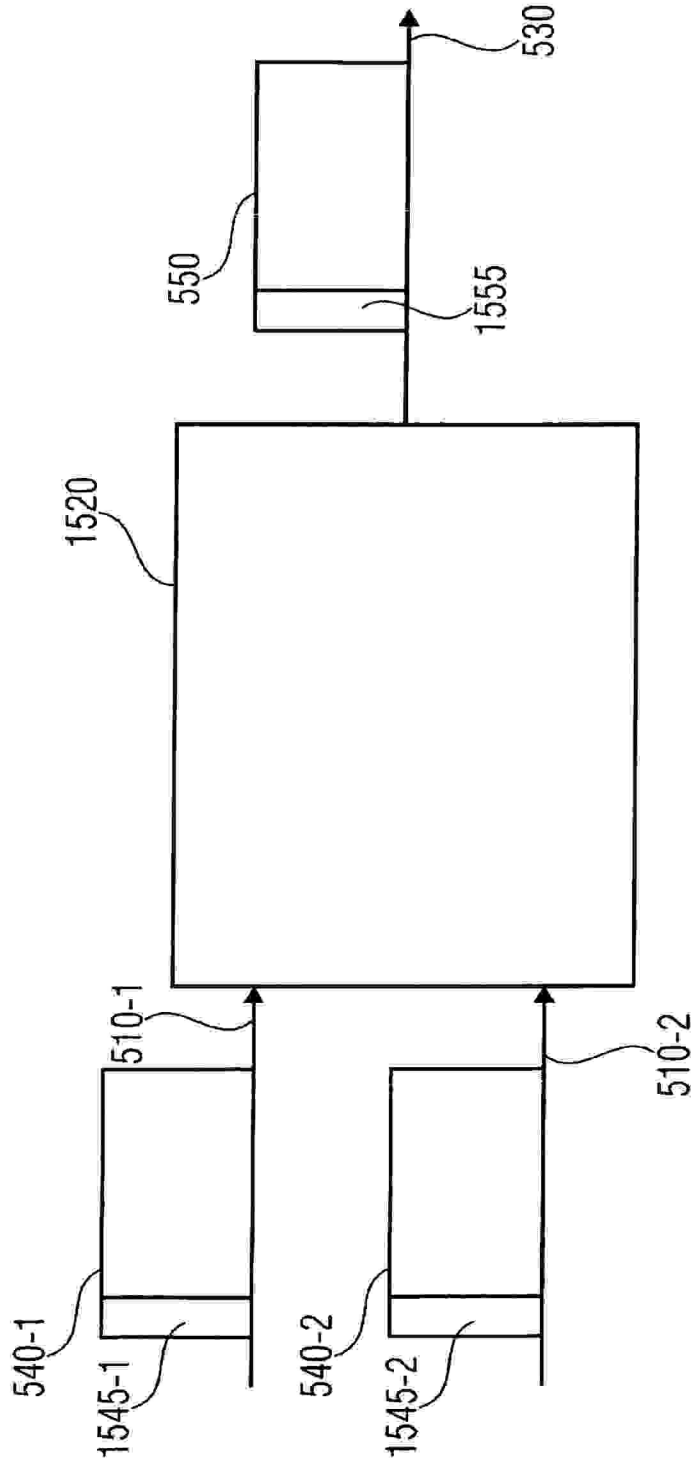


FIG 9

1500

9/12

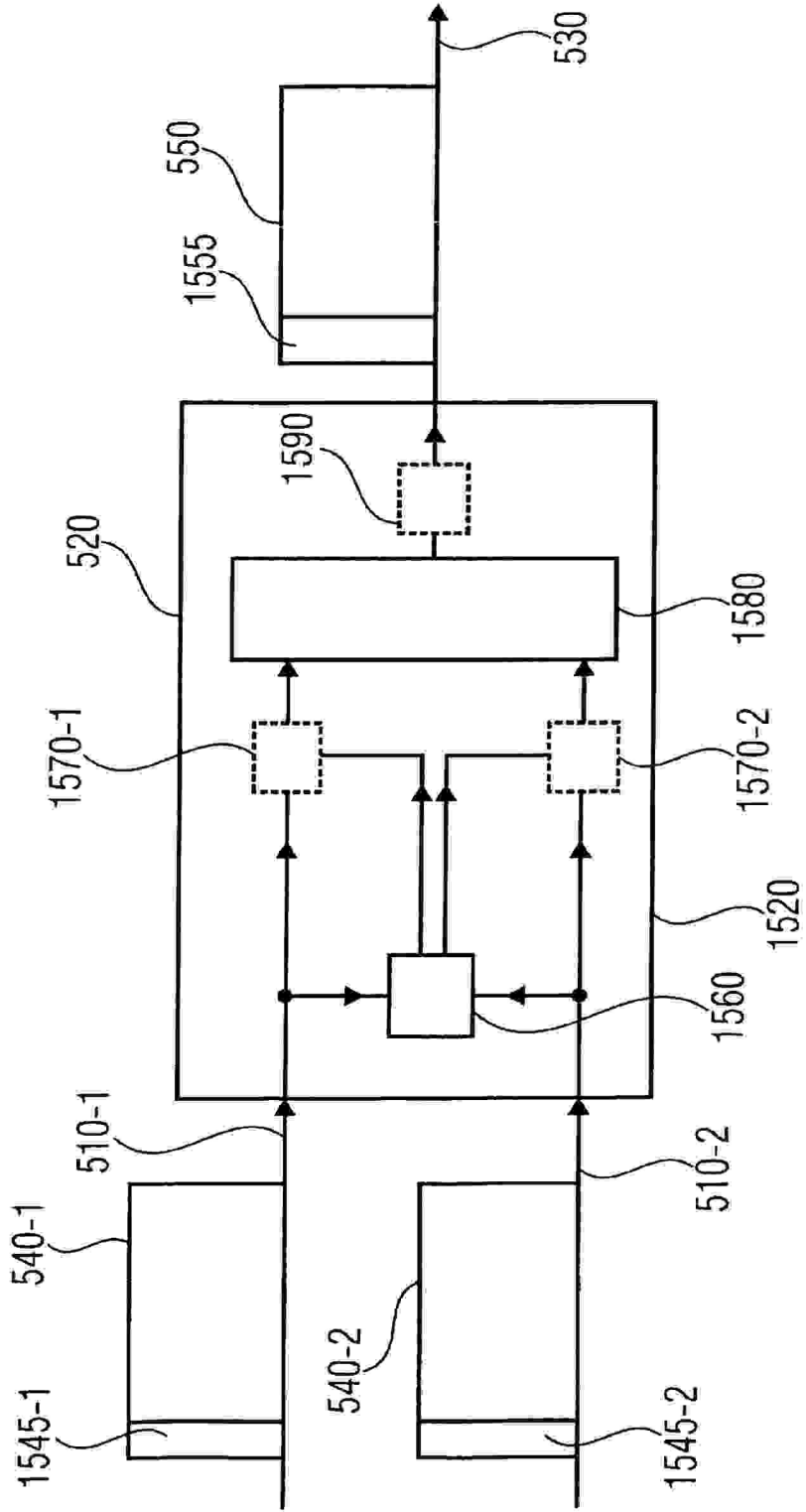


FIG 10

1500

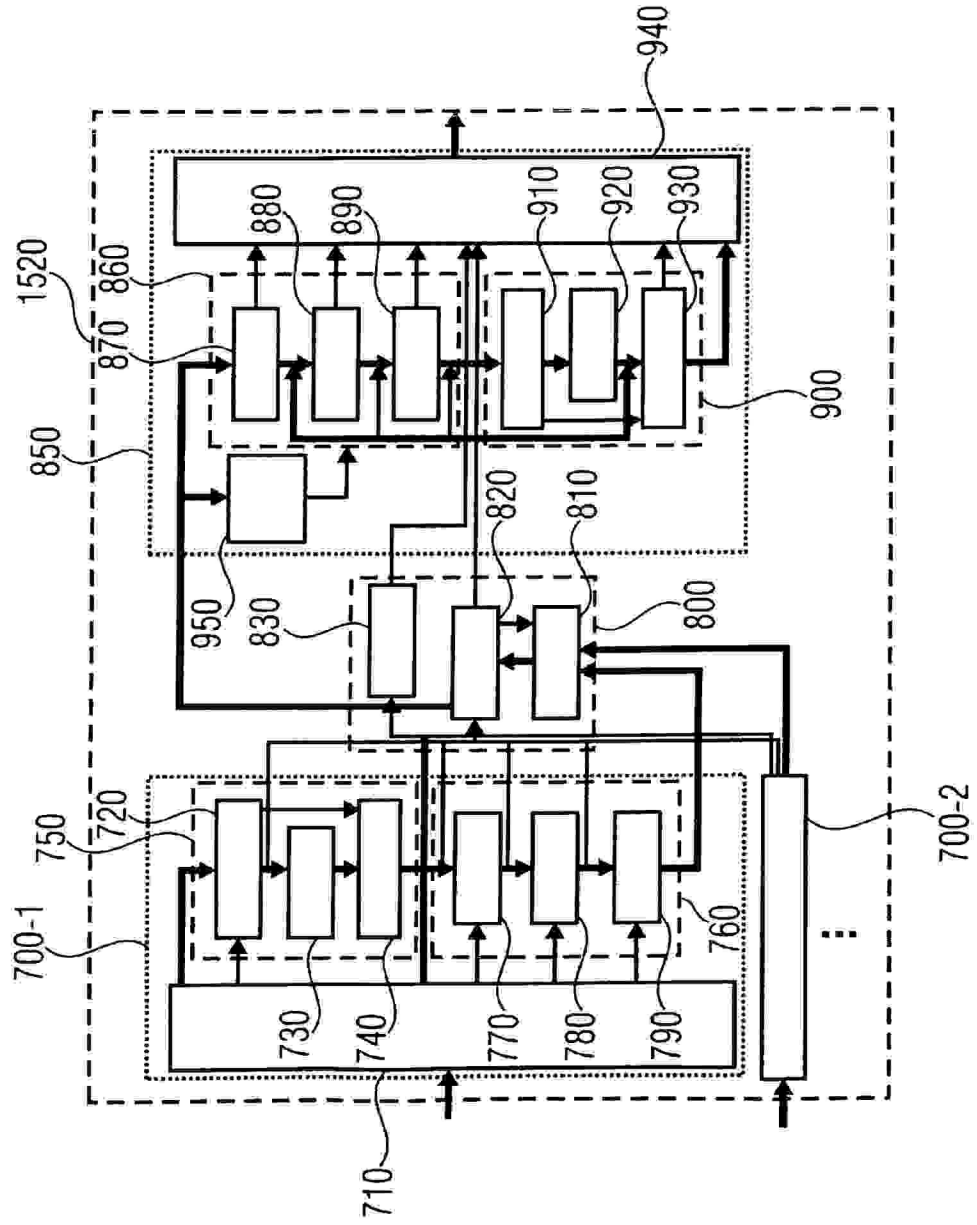


FIG 11

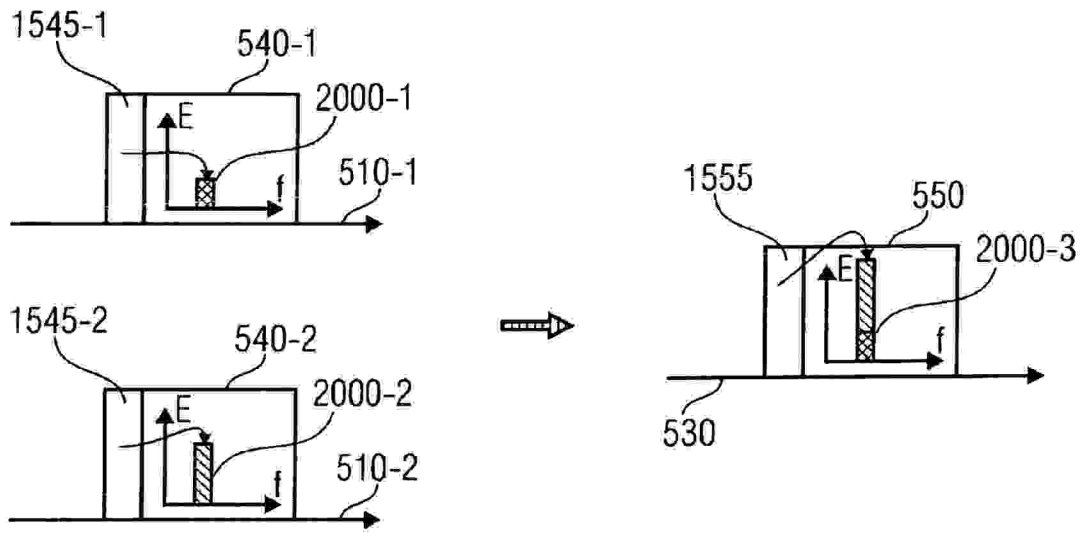


FIG 12A

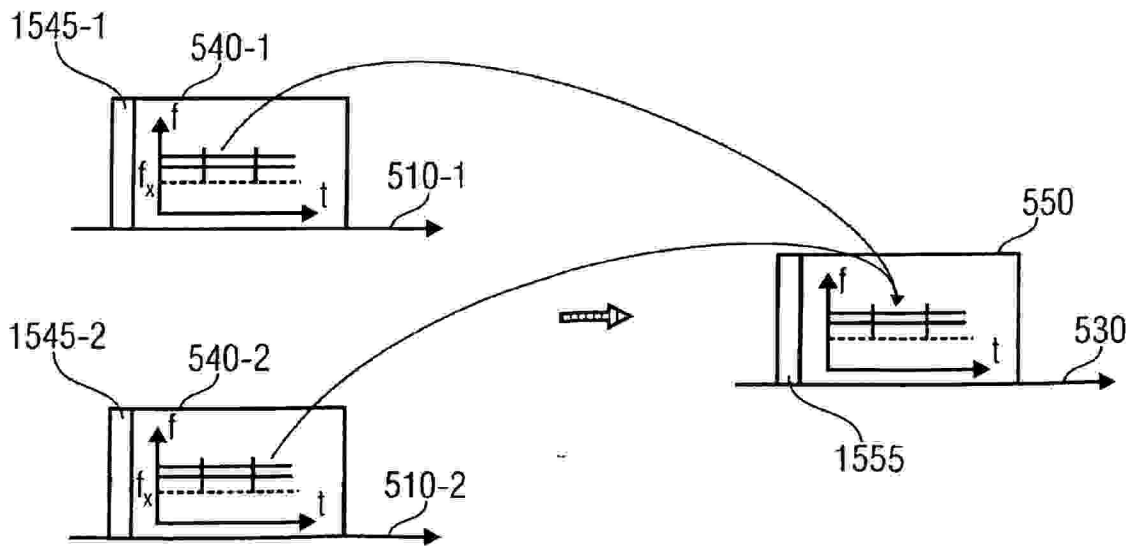


FIG 12B

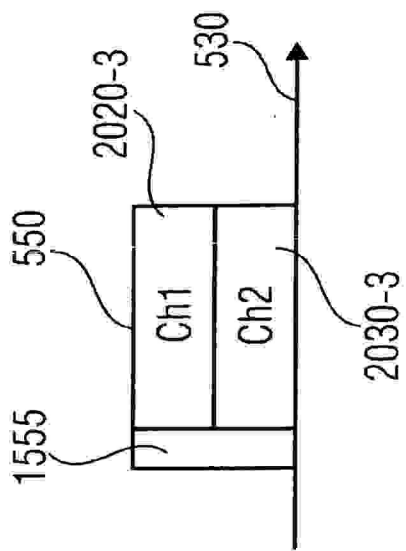
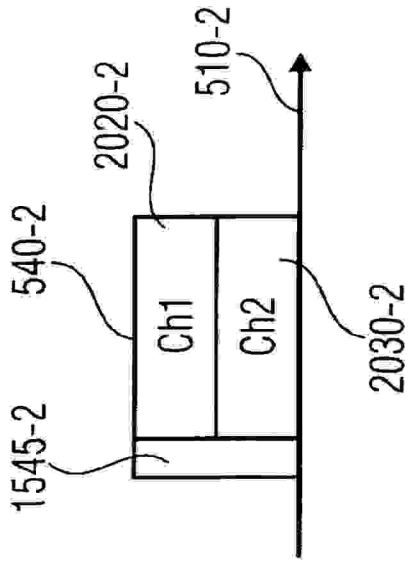
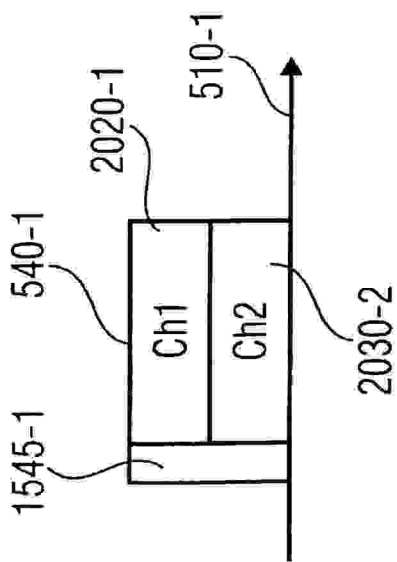


FIG 12C