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

(11) BR 122020017207-0 B1

(22) Data do Depósito: 27/08/2014

(45) Data de Concessão: 06/12/2022

(54) Título: MÉTODO, SISTEMA DE PROCESSAMENTO DE MÍDIA, APARELHO E MEIO DE ARMAZENAMENTO LEGÍVEL POR COMPUTADOR NÃO TRANSITÓRIO

(51) Int.Cl.: G10L 19/008; G10L 19/20; G10L 19/22; G10L 21/02; G10L 21/0324; (...).

(30) Prioridade Unionista: 25/10/2013 US 61/895,959; 25/11/2013 US 61/908,664; 28/08/2013 US 61/870,933.

(73) Titular(es): DOLBY INTERNATIONAL AB; DOLBY LABORATORIES LICENSING CORPORATION.

(72) Inventor(es): JEROEN KOPPENS; HANNES MUESCH.

(86) Pedido PCT: PCT US2014052962 de 27/08/2014

(87) Publicação PCT: WO 2015/031505 de 05/03/2015

(85) Data do Início da Fase Nacional: 24/08/2020

(62) Pedido Original do Dividido: BR112016004299-9 - 27/08/2014

(57) Resumo: A presente invenção refere-se a um método para a melhora da fala híbrida que emprega melhora codificada paramétrica (ou mistura de melhora codificada paramétrica e codificada por forma de onda), sob algumas condições de sinal e melhora codificada por forma de onda (ou uma mistura diferente de melhora codificada paramétrica e codificada por forma de onda), sob outras condições de sinal. Outros aspectos são métodos para a geração de um fluxo de bits indicativo de um programa de áudio, incluindo a fala e outros conteúdos, tal que a melhora da fala híbrida pode ser realizada com o programa, um decodificador incluindo um buffer, que armazena, pelo menos, um segmento de um fluxo de bits de áudio codificado gerado por qualquer modalidade do método da invenção, e um sistema ou dispositivo (por exemplo, um codificador ou decodificador) configurado (por exemplo, programado) para realizar qualquer modalidade do método da invenção. Pelo menos algumas das operações de melhora da fala são realizadas por um decodificador de áudio receptor com metadados para melhora de fala Médio/Lateral gerados por um codificador de áudio à montante.

"MÉTODO, SISTEMA DE PROCESSAMENTO DE MÍDIA, APARELHO E MEIO DE ARMAZENAMENTO LEGÍVEL POR COMPUTADOR NÃO TRANSITÓRIO"

Dividido do BR112016004299-9 depositado em 27 de agosto de 2014.

REFERÊNCIA CRUZADA A PEDIDOS RELACIONADOS

[001] Este pedido reivindica prioridade do Pedido de Patente Provisório US 61/870.933, depositado em 28 de agosto de 2013, Pedido de Patente Provisório US 61/895.959, depositado em 25 de Outubro de 2013 e Pedido de Patente Provisório US 61/908.664, depositado em 25 de Novembro de 2013, cada uma das quais é aqui incorporado por referência na sua totalidade.

TECNOLOGIA

[002] A invenção refere-se ao processamento de sinal de áudio, e mais particularmente, a melhoria do conteúdo da fala de um programa de áudio em relação ao outro conteúdo do programa, em que a melhora da fala é "híbrida", no sentido de que inclui melhora codificada em forma de onda (ou relativamente mais melhoria codificada em forma de onda) sob algumas condições de sinal e melhora codificada paramétrica (ou relativamente mais melhora codificada paramétrica) sob outras condições de sinal. Outros aspectos são codificação, decodificação, e prestação de programas de áudio que incluem dados suficientes para permitir tal melhora de fala híbrida.

ANTECEDENTES

[003] No cinema e na televisão, diálogo e narrativa são frequentemente apresentados em conjunto com outros, áudio não fala, como música, efeitos, ou ambiente de eventos desportivos. Em muitos casos, os sons de fala e não fala são capturados separadamente e misturados em conjunto sob o controlo de um engenheiro de som. O engenheiro de som seleciona o nível da fala em relação ao nível da não fala de uma forma que é apropriada para a maioria dos ouvintes. No entanto, alguns ouvintes, por exemplo, aqueles com uma deficiência auditiva, sentem

dificuldades na compreensão do conteúdo da fala dos programas de áudio (tendo proporções de mistura de fala para não fala determinada pelo engenheiro) e preferem se a fala deve ser misturada em um nível relativamente mais elevado.

[004] Existe um problema a ser resolvido ao permitir que estes ouvintes aumentem a audibilidade do conteúdo de fala do programa fala de áudio em relação ao do conteúdo de áudio não fala.

[005] Uma abordagem atual é a de proporcionar aos ouvintes dois fluxos de áudio de alta qualidade. Um fluxo transporta áudio de conteúdo principal (principalmente da fala) e o outro transporta áudio de conteúdo secundário (o restante do programa de áudio, que exclui a fala) e ao usuário é dado o controle sobre o processo de mistura. Infelizmente, este esquema é impraticável porque não se constrói sobre a prática atual de transmissão de um programa de áudio completamente misturados. Além disso, requer cerca de duas vezes a largura de banda de transmissão atual prática porque dois fluxos independentes de áudio, cada um de qualidade de transmissão, devem ser fornecidos ao usuário.

[006] Um outro método de melhora da fala (a ser referido aqui como melhora "Codificada por forma de onda") é descrito na Publicação do Pedido de Patente US 2010/0106507 A1, publicada em 29 de abril de 2010, atribuída a Dolby Laboratories, Inc. e nomeando Hannes Muesch como inventor. Em melhora codificada por forma de onda, a proporção de fala de fundo (não fala) de uma mistura de áudio original de fala e de conteúdo não fala (por vezes referido como uma mistura principal) é aumentada pela adição à principal mistura uma versão reduzida de qualidade (cópia de baixa qualidade) do sinal de fala limpa, que foi enviada para o receptor ao lado da mistura principal. Para reduzir a sobrecarga de largura de banda, a cópia de baixa qualidade é normalmente codificada a uma taxa de bits muito baixa. Devido à baixa codificação de taxa de bit, artefatos de codificação estão associados com a cópia de baixa qualidade, e os artefatos de codificação são claramente audíveis quando a cópia de

baixa qualidade é processada e ouvida de forma isolada. Assim, a cópia de baixa qualidade tem qualidade questionável quando ouvida de forma isolada. Melhora codificada por forma de onda tenta esconder esses artefatos de codificação adicionando à cópia de baixa qualidade para a mistura principal apenas durante momentos em que o nível dos componentes não fala é elevado para que os artefatos de codificação sejam mascarados pelos componentes de não fala. Como será detalhado mais tarde, as limitações dessa abordagem incluem o seguinte: a quantidade de melhoria de fala normalmente não pode ser constante ao longo do tempo, e artefatos de áudio podem se tornar audíveis quando os componentes de fundo (não fala) da mistura principal são fracos ou a seus espectros de amplitude de frequência diferem drasticamente daqueles do ruído de codificação.

[007] De acordo com a melhora codificada por forma de onda, um programa de áudio (para entrega a um decodificador para decodificar e entrega subsequente) é codificado como um fluxo de bits que inclui a cópia da fala de baixa qualidade (ou uma versão codificada da mesma) como um fluxo lateral da mistura principal. O fluxo de bits pode incluir metadados indicativos de um parâmetro de escala que determina a quantidade de melhora de fala codificada por forma de onda a ser executada (ou seja, o parâmetro de escalonamento determina um fator de escalonamento a ser aplicado à cópia de baixa da fala de qualidade antes da cópia de fala ser escalonada, a cópia de baixa qualidade é combinada com a mistura principal, ou um valor máximo de um tal fator de escalonamento que vai garantir mascaramento de artefatos de codificação). Quando o valor atual do fator de escalonamento é zero, o decodificador não executa a melhora de fala sobre o segmento correspondente da mistura principal. O valor atual do parâmetro de escalonamento (ou o valor máximo atual que este pode atingir) é tipicamente determinado no codificador (uma vez que é tipicamente gerado por um modelo psico-acústico computacionalmente intensivo), mas pode ser gerado no decodificador. Neste último caso, não haveria necessidade de nenhum metadado

indicativo do parâmetro de escalonamento para ser enviado a partir do codificador para o decodificador, e o decodificador, ao contrário, poderia determinar a partir da mistura principal uma proporção de potência do conteúdo de fala da mistura para a potência da mistura e aplicar um modelo para determinar o valor atual do parâmetro de escalonamento em resposta ao valor atual da proporção de potência.

[008] Um outro método (a ser referido aqui como melhora "codificada paramétrica") para aumentar a inteligibilidade da fala na presença de áudio concorrente (fundo) é para segmentar o programa de áudio original (normalmente uma trilha sonora) em ladrilhos de tempo/frequência e impulsionar as ladrilhos de acordo com a proporção da potência (ou nível) do seu conteúdo de fala e de fundo, para alcançar um aumento da componente de fala em relação ao fundo. A ideia subjacente desta abordagem é semelhante à da supressão de ruído de subtração espectral guiada. Num exemplo extremo desta abordagem, no qual todos os ladrilhos com SNR (isto é, a proporção de potência, ou nível, do componente de fala para aquela do conteúdo de som concorrente) abaixo de um limiar predeterminado são completamente suprimidas, demonstrou fornecer robustas melhorias de inteligibilidade da fala. Na aplicação deste método à radiodifusão, a proporção de fala para fundo (SNR) pode ser inferida através da comparação da mistura de áudio original (conteúdo de fala e não fala) para o componente de fala da mistura. A SNR inferida pode então ser transformada em um conjunto adequado de parâmetros de melhora, que são transmitidos juntamente com a mistura de áudio original. No receptor, estes parâmetros podem (opcionalmente) ser aplicados à mistura de áudio original para derivar um sinal indicativo de fala melhorada. Como será descrito mais tarde, funções de melhoria codificadas paramétricas são melhores quando o sinal de fala (o componente de fala da mistura) domina o sinal de fundo (o componente não fala da mistura).

[009] A melhora codificada por forma de onda exige que uma cópia de baixa

qualidade do componente de fala de um programa de áudio disponível esteja disponível ao receptor. Para limitar os dados aéreos incorridos na transmissão que copiam ao longo da mistura principal de áudio, esta cópia é codificada em uma taxa de bits muito baixa e apresenta distorções de codificação. Estas distorções de codificação são susceptíveis de serem mascaradas pelo áudio original, quando o nível dos componentes não fala é alto. Quando as distorções de codificação são mascaradas a qualidade resultante do áudio melhorado é muito boa.

[010] A melhora codificada paramétrica baseia-se na análise do sinal de mistura de áudio principal em ladrilhos de tempo/frequência e a aplicação de ganhos/atenuações adequados a cada um destes ladrilhos. A taxa de dados necessária para transmitir esses ganhos para o receptor é baixa quando comparada com a de melhora codificada por forma de onda. No entanto, devido à limitada resolução temporal-espectral dos parâmetros, a fala, quando misturada com o áudio não fala, não pode ser manipulada sem afetar também o áudio não fala. A melhora codificada paramétrica do conteúdo da fala de uma mistura de áudio introduz, assim, a modulação no conteúdo de não fala da mistura, e esta modulação ("modulação de fundo") pode tornar-se desagradável durante a reprodução da mistura melhorada de fala. As modulações de fundo são mais propensas a serem desagradáveis quando a proporção de fala para fundo é muito baixa.

[011] As abordagens descritas nesta seção são abordagens que poderiam ser buscadas, mas não necessariamente abordagens que tenham sido previamente concebidas ou buscadas. Portanto, a menos que indicado de outra forma, não deve considerar-se que qualquer uma das abordagens descritas nesta seção qualificam como técnica anterior meramente em virtude da sua inclusão na presente seção. Da mesma forma, os problemas identificados no que diz respeito a uma ou mais abordagens não devem presumir ter sido reconhecido em qualquer técnica anterior com base na presente seção, salvo indicação em contrário.

BREVE DESCRIÇÃO DOS DESENHOS

[012] A presente invenção é ilustrada a título de exemplo, e não como forma de limitação, nas figuras dos desenhos em anexo e nos quais números de referência semelhantes se referem a elementos semelhantes e nos quais:

a FIGURA 1 é um diagrama de blocos de um sistema configurado para gerar parâmetros de predição para reconstituir o conteúdo de fala de um sinal de conteúdo misturado de canal único (tendo conteúdo fala e não fala).

a FIGURA 2 é um diagrama de blocos de um sistema configurado para gerar parâmetros de predição para reconstituir o conteúdo de fala de um sinal de conteúdo misturado de vários canais (tendo conteúdo de fala e não fala).

a FIGURA 3 é um diagrama de blocos de um sistema que inclui um codificador configurado para executar uma modalidade do método de codificação da invenção para gerar um fluxo de bits de áudio codificado indicativo de um programa de áudio, e um decodificador configurado para decodificar e executar a melhora de fala (de acordo com uma modalidade do método da invenção) no fluxo de bits de áudio codificado.

a FIGURA 4 é um diagrama de blocos de um sistema configurado para processar um sinal de áudio do conteúdo misturado de vários canais, incluindo através da realização de melhora da fala convencional na mesma.

a FIGURA 5 é um diagrama de blocos de um sistema configurado para processar um sinal de áudio do conteúdo misturado de vários canais, incluindo através da realização de melhora de fala convencional codificada paramétrica na mesma.

a FIGURA 6 e FIGURA 6A são diagramas de blocos de sistemas configurados para processar um sinal de áudio do conteúdo misturado de vários canais, incluindo através da realização de uma modalidade do método de melhora de fala inventivo na mesma.

a FIGURA 7 é um diagrama de blocos de um sistema para a execução e modalidade do método de codificação da invenção usando um modelo de

mascaramento auditivo;

a FIGURA 8A e FIGURA 8B ilustram os fluxos exemplo de processo; e

a FIGURA 9 ilustra uma plataforma de hardware exemplar em que um computador ou um dispositivo computacional, tal como aqui descrito podem ser implementados.

DESCRIÇÃO DE MODALIDADES EXEMPLARES

[013] As modalidades exemplares que se relacionam com melhora fala codificada por forma de onda híbrida e codificada paramétrica, são aqui descritas. Na descrição seguinte, para os fins de explicação, numerosos detalhes específicos são apresentados a fim de proporcionar um entendimento completo da presente invenção. Será evidente, no entanto, que a presente invenção pode ser praticada sem estes detalhes específicos. Em outros casos, estruturas e bem conhecidas e dispositivos não estão descritos em detalhes exaustivos, a fim de evitar desnecessariamente oclusão, obscurecimento, ou ofuscamento da presente invenção.

[014] As modalidades exemplares são aqui descritas de acordo com o esquema seguinte:

1. VISÃO GERAL
2. NOTAÇÃO E NOMENCLATURA
3. GERAÇÃO DE PARÂMETROS DE PREDIÇÃO
4. OPERAÇÕES DE MELHORA DE FALA
5. PROCESSAMENTO DA FALA
6. REPRESENTAÇÃO MÉDIAS/LATERAIS
7. FLUXOGRAMAS EXEMPLARES
8. MECANISMOS DE IMPLEMENTAÇÃO - VISÃO GERAL DE HARDWARE
9. EQUIVALENTES, EXTENSÕES, ALTERNATIVAS E MISCELÂNEA

1. VISAO GERAL

[015] Esta visão geral apresenta uma descrição básica de alguns aspectos

de uma modalidade da presente invenção. Deve notar-se que esta visão geral não é exaustiva ou extensiva ao sumário dos aspectos da modalidade. Além disso, deve notar-se que esta visão geral não se destina a ser entendida como identificação de quaisquer aspectos ou elementos particularmente significativos da modalidade, nem como delineando qualquer escopo da modalidade em particular, nem a invenção em geral. Esta visão geral apenas apresenta alguns conceitos que se relacionam com a modalidade exemplar em um formato condensado e o, e deve ser entendida como apenas um prelúdio conceitual para uma descrição mais detalhada de modalidades exemplares que se segue abaixo. Note-se que, embora modalidades separadas sejam aqui discutidas, qualquer combinação das modalidades e/ou modalidades parciais aqui discutidas podem ser combinadas para formar outras modalidades.

[016] Os inventores reconheceram que os pontos fortes e fracos de melhora codificada paramétrica e melhora codificada por forma de onda podem compensar uma à outra, e que a melhora da fala convencional pode ser substancialmente melhorada por um método de melhora híbrida que emprega melhora codificada paramétrica (ou uma mistura de melhora codificada paramétrica e melhora codificada por forma de onda), sob algumas condições de sinal e melhora codificada por forma de onda (ou uma mistura diferente de melhora codificada paramétrica e melhora codificada por forma de onda), sob outras condições de sinal. As modalidades típicas do método de melhora híbrida da invenção proporcionam uma melhoria da qualidade de fala mais consistente e melhor do que pode ser alcançada por qualquer uma melhoria codificada paramétrica ou codificada por forma de onda sozinha.

[017] Numa classe de modalidades, o método da invenção inclui as etapas de: (a) receber um fluxo de bits indicativo de um programa de áudio incluindo fala tendo uma forma de onda não melhorada e outro conteúdo de áudio, em que o fluxo de bits inclui: dados de áudio indicativos da fala e o outro conteúdo de áudio, dados de forma de onda indicativos de uma versão reduzida de qualidade da fala (onde os

dados de áudio foram gerados por mistura de dados de fala com dados não fala, os dados da forma de onda compreendem, tipicamente, menos bits do que os dados de fala), em que a versão reduzida de qualidade tem uma segunda forma de onda semelhante (por exemplo, pelo menos, substancialmente semelhante) para a forma de onda não melhorada, e a versão de qualidade reduzida teria qualidade questionável se ouvida isoladamente, e os dados paramétricos, em que os dados paramétricos com os dados de áudio determinam fala parametricamente construída, e a fala parametricamente construída é uma versão reconstruída parametricamente da fala, que corresponde a pelo menos substancialmente (por exemplo, é uma boa aproximação de) a fala; e (b) realização de melhora da fala sobre o fluxo de bits em resposta a um indicador de mistura, gerando assim dados indicativos de um programa de áudio de fala melhorada, incluindo através da combinação dos dados de áudio com uma combinação de dados de fala de baixa qualidade determinados a partir dos dados de forma de onda, e os dados de fala reconstruídos, em que a combinação é determinada pelo indicador de mistura (por exemplo, a combinação tem uma sequência de estados determinada por uma sequência de valores atuais do indicador de mistura), os dados de fala reconstruídos são gerados em resposta a pelo menos alguns dos dados paramétricos e, pelo menos, alguns dos dados de áudio, e o programa de áudio com fala melhorada com menos artefatos audíveis de melhora da fala (por exemplo, artefatos de melhora da fala que são mais bem mascarados e, portanto, menos audíveis quando o programa de áudio de fala melhorada é processado e ouvido) do que seria ou um programa de áudio de fala melhorada puramente codificado por forma de onda determinado pela combinação apenas dos dados de fala de baixa qualidade (o que é indicativo da versão reduzida da qualidade da fala) com os dados de áudio ou um programa de áudio de fala melhorada puramente codificada paramétrica determinado a partir dos dados paramétricos e os dados de áudio.

[018] Aqui, "artefato de melhora de fala" (ou "artefato de codificação de melhora de fala") indica uma distorção (tipicamente uma distorção mensurável) de um sinal de áudio (indicativo de um sinal de fala e um sinal de áudio não fala) causado por uma representação do sinal de fala (por exemplo, sinal de fala codificada por forma de onda, ou dados paramétricos, em conjunto com o sinal de conteúdo misturado).

[019] Em algumas modalidades, o indicador de mistura (que pode ter uma sequência de valores, por exemplo, um para cada uma sequência dos segmentos de fluxo de bits) está incluído no fluxo de bits recebido na etapa (a). Algumas modalidades incluem uma etapa de gerar o indicador de mistura (por exemplo, em um receptor que recebe e decodifica o fluxo de bits), em resposta ao fluxo de bits recebido no etapa (a).

[020] Deve ser entendido que a expressão "indicador de mistura" não pretende exigir que o indicador de mistura seja um parâmetro ou valor único (ou uma sequência de parâmetros ou valores únicos) para cada segmento do fluxo de bits. Em vez disso, contempla-se que em algumas modalidades, um indicador de mistura (para um segmento do fluxo de bits) pode ser um conjunto de dois ou mais parâmetros ou valores (por exemplo, para cada segmento, um parâmetro de controle de melhora codificada paramétrica, e um parâmetro de controle de melhora codificada por forma de onda) ou uma sequência de conjuntos de parâmetros ou valores.

[021] Em algumas modalidades, o indicador de mistura para cada segmento pode ser uma sequência de valores indicando a mistura por banda de frequência do segmento.

[022] Os dados da forma de onda e os dados paramétricos não precisam ser fornecidos para (por exemplo, incluídos em) cada segmento do fluxo de bits, e ambos os dados da forma de onda e os dados paramétricos não precisam ser usados para executar a melhora de fala em cada segmento do fluxo de bits. Por exemplo, em alguns casos, pelo menos, um segmento pode incluir apenas os dados de forma de

onda (e a combinação determinada pelo indicador de mistura para cada dito segmento pode consistir em apenas dados da forma de onda) e pelo menos um outro segmento pode incluir apenas os dados paramétricos (e a combinação determinada pelo indicador de mistura para cada dito segmento pode consistir somente em dados de fala reconstruídos).

[023] É contemplado que, tipicamente, um codificador gera o fluxo de bits incluindo por codificação (por exemplo, comprimindo) os dados de áudio, mas não através da aplicação da mesma codificação aos dados de forma de onda ou os dados paramétricos. Assim, quando o fluxo de bits é enviado para um receptor, o receptor tipicamente analisaria o fluxo de bits para extrair os dados de áudio, os dados de forma de onda, e os dados paramétricos (e o indicador de mistura, se for entregue no fluxo de bits), mas apenas descodificaria os dados de áudio. O receptor tipicamente executaria a melhora da fala nos dados de áudio descodificados (utilizando os dados de forma de onda e/ou dados paramétricos) sem aplicar aos dados da forma de onda ou aos dados paramétricos o mesmo processo de descodificação que é aplicado aos dados de áudio.

[024] Normalmente, a combinação (indicada pelo indicador de mistura) dos dados da forma de onda e os dados de fala reconstruídos mudam ao longo do tempo, com cada estado da combinação relativo à fala e outros conteúdos de áudio de um segmento correspondente do fluxo de bits. O indicador de mistura é gerado de tal modo que o estado atual da combinação (de dados da forma de onda e os dados de fala reconstruídos) é pelo menos parcialmente determinado pelas propriedades de sinal da fala e outros conteúdos de áudio (por exemplo, uma proporção entre a potência de conteúdo de fala e a potência de outro conteúdo de áudio) no segmento correspondente do fluxo de bits. Em algumas modalidades, o indicador de mistura é gerado de tal modo que o estado atual da combinação é determinado por propriedades do sinal da fala e o outro conteúdo de áudio no segmento correspondente do fluxo de

bits. Em algumas modalidades, o indicador de mistura é gerado de tal modo que o estado atual da combinação é determinado ambos por propriedades do sinal de fala e o outro conteúdo de áudio no segmento correspondente do fluxo de bits e uma quantidade de artefatos de codificação nos dados da forma de onda.

[025] A etapa (b) pode incluir um etapa de realização de melhora da fala codificada por forma de onda pela combinação (por exemplo, mistura ou mistura), pelo menos, de alguns da dos dados da fala de baixa qualidade com os dados de áudio de, pelo menos, um segmento do fluxo de bits, e realizando melhoria da fala codificada paramétrica, combinando os dados de fala reconstruídos com os dados de áudio de pelo menos um segmento do fluxo de bits. Uma combinação de melhora de fala codificada por forma de onda e melhora de fala codificada paramétrica é realizada em pelo menos um segmento do fluxo de bits por mistura de ambos os dados de fala de baixa qualidade e fala parametricamente construída para o segmento com os dados de áudio do segmento. Sob algumas condições de sinal, apenas um (mas não ambos) de melhora da fala codificada por forma de onda e melhora da fala codificada paramétrica é realizado (em resposta ao indicador de mistura) em um segmento (ou em cada um de mais do que um segmento) de fluxo de bits.

[026] Aqui, a expressão "SNR" (relação sinal-ruído) será utilizada para designar a proporção de potência (ou diferença de nível) do conteúdo da fala de um segmento de um programa de áudio (ou de todo o programa) daquele conteúdo de não fala do segmento ou programa, ou do conteúdo da fala de um segmento do programa (ou de todo o programa) para aquele todo conteúdo (fala e não fala) do segmento ou programa.

[027] Numa classe de modalidades, o método da invenção implementa comutação baseada em SNR temporal "cega" entre melhora codificada paramétrica e melhora codificada por de forma de onda de segmentos de um programa de áudio. Neste contexto, "cega" indica que a comutação não é perceptivamente guiada por um

modelo de mascaramento auditivo complexo (por exemplo, de um tipo a ser descrito aqui), mas é guiada por uma sequência de valores de SNR (indicadores de mistura) correspondendo aos segmentos do programa. Numa modalidade desta classe, a melhora de fala codificada híbrida é conseguido por comutação temporal entre melhora codificada paramétrica e melhora codificada por forma de onda, de modo que qualquer melhora codificada paramétrica ou melhora codificada por forma de onda (mas não ambas melhora codificada paramétrica e melhora codificada por forma de onda) é executada em cada segmento de um programa de áudio em que a melhora de fala é executada. Reconhecendo que melhora codificada por forma de onda tem um melhor desempenho sob a condição de baixa SNR (em segmentos tendo baixos valores de SNR) e melhora codificada paramétrica executa melhor em SNRs favoráveis (em segmentos tendo elevados valores de SNR), a decisão de comutação é tipicamente baseada na proporção de fala (diálogo) para o restante de áudio em uma mistura de áudio original.

[028] Modalidades que implementam comutação baseada em SNR temporal "cega" normalmente incluem etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias de tempo consecutivas (segmentos), e determinar para cada segmento a SNR entre o conteúdo da fala e outros conteúdos de áudio (ou entre o conteúdo da fala e conteúdo de áudio total) do segmento; e para cada segmento, comparar a SNR a um limiar e proporcionar um parâmetro de controle de melhora codificada paramétrica para o segmento (isto é, o indicador de mistura para o segmento indica que a melhora codificada paramétrica deve ser realizada) quando a SNR é maior do que o limiar ou fornecer um parâmetro de controle de melhora codificada por forma de onda para o segmento (isto é, o indicador de mistura para o segmento indica que a melhora codificada por forma de onda deve ser realizada) quando a SNR não é maior do que o limiar. Tipicamente, o sinal de áudio é entregue não melhorado (por exemplo, transmitido) com os parâmetros de controle incluídos

como metadados para um receptor, e o receptor executa (em cada segmento) o tipo de melhora de fala indicada pelo parâmetro de controle para o segmento. Assim, o receptor realiza melhora codificada paramétrica em cada segmento para o qual o parâmetro de controle é um parâmetro de controle de melhora codificada paramétrica, e melhora codificada por forma de onda em cada segmento para o qual o parâmetro de controle é um parâmetro de controle de melhora codificada por forma de onda.

[029] Se alguém está disposto a arcar com os custos de transmissão (com cada segmento de uma mistura de áudio original) ambos os dados (para a implementação de melhora da fala codificada por forma de onda) e parâmetros de melhora codificada paramétrica com uma mistura original (Não melhorada), um maior grau de melhora de fala pode ser conseguido através da aplicação de ambos melhora codificada por forma de onda e melhora codificada paramétrica aos segmentos individuais da mistura. Assim, numa classe de modalidades, o método da invenção implementa mistura baseada em SNR "cega" temporal entre melhora codificada paramétrica e melhora codificada por forma de onda de segmentos de um programa de áudio. Neste contexto, também, "cega" indica que a comutação não é perceptivamente guiada por um modelo complexo de mascaramento auditivo (por exemplo, de um tipo a ser descrito aqui), mas é guiada por uma sequência de valores de SNR correspondentes aos segmentos do programa.

[030] Modalidades que implementam mistura baseada em SNR temporal "cega" normalmente incluem etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias de tempo consecutivas (segmentos), e determinar para cada segmento da SNR entre o conteúdo da fala e o outro conteúdo de áudio (ou entre o conteúdo da fala e conteúdo de áudio total) do segmento; e para cada segmento, fornecer um indicador de controle da mistura, onde o valor do indicador de controle é determinado por (é uma função de) SNR para o segmento.

[031] Em algumas modalidades, o método inclui uma etapa de determinar

(por exemplo, receber um pedido para) uma quantidade total ("T") de melhora de fala, e o indicador de controle da mistura é um parâmetro, α, para cada segmento de modo que T = αPw + (1-α)Pp, onde Pw é melhora codificada por forma de onda para o segmento que produziria o montante total predeterminado de melhora, T, se aplicado ao conteúdo de áudio não melhorado do segmento utilizando os dados da forma de onda fornecidos para o segmento (onde o conteúdo da fala do segmento tem uma forma de onda não melhorada, os dados da forma de onda para o segmento são indicativos de uma versão reduzida da qualidade do conteúdo da fala do segmento, a versão reduzida da qualidade tem uma forma de onda semelhante (por exemplo, pelo menos, substancialmente semelhante) para a forma de onda não melhorada, e a versão reduzida da qualidade do conteúdo da fala é de qualidade questionável quando processada e percebida de forma isolada), e Pp é a melhora codificada paramétrica que produziria o montante total predeterminado de melhora, T, se aplicada ao conteúdo de áudio não melhorado do segmento usando dados paramétricos fornecidos para o segmento (onde os dados paramétricos para o segmento, com o conteúdo de áudio não melhorado do segmento, determinam uma versão reconstruída parametricamente do conteúdo da fala do segmento). Em algumas modalidades, o indicador de controle de mistura para cada um dos segmentos é um conjunto desses parâmetros, incluindo um parâmetro para cada banda de frequência do segmento relevante.

[032] Quando o sinal áudio não melhorado é fornecido (por exemplo, transmitido) com os parâmetros de controle como metadados para um receptor, o receptor pode executar (em cada segmento) a melhora de fala híbrida indicada pelos parâmetros de controle para o segmento. Alternativamente, o receptor gera os parâmetros de controle a partir do sinal de áudio não melhorado.

[033] Em algumas modalidades, o receptor executa (em cada segmento do sinal de áudio não melhorado) uma combinação de melhora codificada paramétrica

(numa quantidade determinada pela melhora Pp escalonada pelo parâmetro α para o segmento) e melhora codificada por forma de onda (numa quantidade determinada pela melhora Pw escalonada pelo valor (1 - α) para o segmento), de tal modo que a combinação de melhora codificada paramétrica e melhora codificada por forma de onda gera o valor total predeterminado de melhora:

$$T = \alpha Pw + (1-\alpha)Pp \quad (1)$$

[034] Numa outra classe de modalidades, a combinação de melhora codificada por forma de onda e codificada paramétrica para ser executada em cada segmento de um sinal de áudio é determinada por um modelo de mascaramento auditivo. Em algumas modalidades nesta classe, a proporção de mistura ideal para uma mistura de melhora codificada por forma de onda e codificada paramétrica para ser executada em um segmento de um programa de áudio utiliza a maior quantidade de melhora codificada por forma de onda que apenas mantém o ruído de codificação de se tornar audível. Deve ser entendido que a disponibilidade de ruído de codificação num decodificador está sempre na forma de uma estimativa estatística, e não pode ser exatamente determinada.

[035] Em algumas modalidades desta classe, o indicador de mistura para cada segmento de dados de áudio é indicativo de uma combinação de melhora codificada por forma de onda e codificada paramétrica para ser executada no segmento, e a combinação é pelo menos substancialmente igual a uma combinação de maximização codificada por forma de onda determinada para o segmento pelo modelo de mascaramento auditivo, em que a combinação maximizando codificada por forma de onda especifica uma maior quantidade relativa de melhora codificada por forma de onda que garante que o ruído de codificação (devido à melhora codificada por forma de onda) no segmento correspondente do programa de áudio melhorado com a fala não é desagradavalmente audível (por exemplo, não é audível). Em algumas modalidades, quanto maior o valor relativo de melhoria codificada por forma

de onda que garante que o ruído de codificação em um segmento do programa de áudio de fala melhorada não é desagradavelmente audível maior é o valor relativo, que assegura que a combinação de melhora codificada por forma de onda e codificada paramétrica para ser executada (em um segmento correspondente de dados de áudio) gera um valor total predeterminado de melhora de fala para o segmento, e/ou (onde artefatos do melhora codificada paramétrica são incluídos na avaliação executada pelo modelo de mascaramento auditivo) ele pode permitir codificação de artefatos (devido à melhora codificada por forma de onda) para ser audível (quando este é favorável) sobre artefatos da melhora codificada paramétrica (por exemplo, quando os artefatos de codificação audíveis (devido a melhora codificada por forma de onda) são menos desagradáveis do que os artefatos audíveis da melhora codificada paramétrica).

[036] A contribuição de melhoria codificada por forma de onda no esquema de codificação híbrido inventivo pode ser aumentada, assegurando que o ruído de codificação não se torne desagradavelmente audível (por exemplo, não se torna audível) usando um modelo de mascaramento auditivo para prever com maior precisão a forma como a codificação de ruído na cópia da fala de qualidade reduzida (para ser usada para implementar a melhora codificada por forma de onda) está sendo mascarada pela mistura de áudio do programa principal e para selecionar a proporção de mistura, por conseguinte.

[037] Algumas modalidades que utilizam um modelo de mascaramento auditivo incluem etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias de tempo consecutivas (segmentos), e proporcionando uma cópia da fala de qualidade reduzida em cada segmento (para uso na melhora codificada por forma de onda) e os parâmetros de melhora codificada paramétrica (para uso em melhora codificada paramétrica) para cada segmento; para cada um dos segmentos, utilizando o modelo de mascaramento auditivo para determinar uma

quantidade máxima de melhora codificada por forma de onda que pode ser aplicada sem codificação artefatos se tornam desagradavelmente audíveis; e gerar um indicador (para cada segmento do sinal de áudio não melhorado) de uma combinação de melhora codificada por forma de onda (em uma quantidade que não exceda a quantidade máxima de melhora codificada por forma de onda determinada utilizando o modelo de mascaramento auditivo para o segmento, e quais pelo menos, correspondem substancialmente à quantidade máxima de melhora codificada pro forma de onda determinada utilizando o modelo de mascaramento auditivo para o segmento) e melhora codificada paramétrica, de tal modo que a combinação de melhora codificada por forma de onda e melhora codificada paramétrica gera uma quantidade total predeterminada de melhora da fala para o segmento.

[038] Em algumas modalidades, cada indicador é incluído (por exemplo, por um codificador) em um fluxo de bits, que também inclui os dados de áudio codificados indicativos do sinal de áudio não melhorado.

[039] Em algumas modalidades, o sinal de áudio não melhorado é segmentado em fatias de tempo consecutivas e cada fatia de tempo é segmentada em bandas de frequência, para cada uma das bandas de cada uma das fatias de tempo de frequência, o modelo de mascaramento auditivo é usado para determinar uma quantidade máxima de melhora codificada por forma de onda que pode ser aplicada sem artefatos de codificação se tornando desagradavelmente audível, e um indicador é gerado para cada banda de frequência de cada fatia de tempo do sinal de áudio não melhorado.

[040] Opcionalmente, o método também inclui uma etapa de realizar (em cada um dos segmentos do sinal de áudio não melhorado) em resposta ao indicador para cada segmento, a combinação de melhora codificada por forma de onda e melhora codificada paramétrica determinada pelo indicador, de tal modo que a combinação de melhora codificada por forma de onda e melhora codificada

paramétrica gera a quantidade total predeterminada de melhora de fala para o segmento.

[041] Em algumas modalidades, o conteúdo de áudio é codificado em um sinal de áudio codificado para uma configuração de canal de áudio de referência (ou representação), como uma configuração de som surround, uma configuração de alto-falante 5.1, uma configuração de alto-falante 7.1, uma configuração de alto-falante 7.2, etc. A configuração de referência pode incluir canais de áudio, como canais estéreo, canal frontal esquerdo e direito, canais surround, canais de alto-falantes, canais objetos, etc. Um ou mais dos canais que transportam o conteúdo da fala podem não ser canais de uma representação canal de áudio Mid/Side (M/S). Tal como aqui utilizado, uma representação de canal de áudio M/S (ou simplesmente representação M/S) compreende, pelo menos, um canal médio e um canal lateral. Em um exemplo de modalidade, o canal médio representa uma soma dos canais esquerdo e direito (por exemplo, igualmente ponderado, etc.), enquanto o canal lateral representa uma diferença dos canais esquerdo e direito, em que os canais esquerdo e direito podem ser afirmados qualquer combinação de dois canais, por exemplo, canais dianteiros centrais e dianteiros esquerdos.

[042] Em algumas modalidades, o conteúdo de fala de um programa pode ser misturado com o conteúdo não fala e podem ser distribuídos ao longo de dois ou mais canais não M/S, como os canais esquerdo e direito, os canais esquerdo e direito dianteiros, etc., na configuração do canal de áudio de referência. O conteúdo da fala pode, mas não é obrigado a, ser representado em um centro fantasma no conteúdo estéreo em que o conteúda de fala é igualmente alto em dois canais não M/S, como os canais esquerdo e direito, etc. O conteúdo estéreo pode conter conteúdo não fala que não é, necessariamente, igualmente alto ou que está ainda presente em ambos os dois canais.

[043] Em algumas abordagens, conjuntos múltiplos de dados de controle não

M/S, parâmetros de controle, etc., para a melhora da fala correspondentes aos canais múltiplos de áudio não M/S sobre os quais o conteúdo da fala é distribuído são transmitidos como parte de metadados de áudio gerais a partir de um codificador de áudio para decodificadores de áudio à jusante. Cada um dos conjuntos múltiplos de dados de controle não M/S, os parâmetros de controle, etc., para a melhora de fala correspondem a um canal de áudio específico dos canais múltiplos de áudio não M/S durante o qual o conteúdo da fala é distribuído e pode ser utilizado por um decodificador de áudio à jusante para controlar as operações de melhora da fala relacionadas com o canal de áudio específico. Tal como aqui utilizado, um conjunto de dados de controle não M/S, os parâmetros de controle, etc., referem-se aos dados de controle, aos parâmetros de controle, etc., para as operações de melhora de fala em um canal de áudio de uma representação não M/S, tal como a configuração de referência em que um sinal de áudio, tal como aqui descrito está codificado.

[044] Em algumas modalidades, metadados para melhora a fala M/S são transmitidos - em complemento ou em substituição de um ou mais conjuntos de dados de controle não M/S, os parâmetros de controle, etc. - como uma parte de metadados de áudio a partir de um codificador de áudio para decodificadores de áudio à jusante. Os metadados para melhora de fala M/S podem compreender um ou mais conjuntos de dados de controle M/S, os parâmetros de controle, etc., para a melhora da fala. Tal como aqui utilizado, um conjunto de dados de controle M/S, os parâmetros de controle, etc., referem-se aos dados de controle, os parâmetros de controle, etc., para as operações de melhora de fala em um canal de áudio da representação M/S. Em algumas modalidades, metadados para melhora de fala M/S para a melhora da fala são transmitidos por um codificador de áudio para descodificadores de áudio à jusante com o conteúdo misturado codificado na configuração do canal de áudio de referência. Em algumas modalidades, o número de conjuntos de dados de controle M/S, os parâmetros de controle, etc., para a melhora da fala nos metadados para melhora de

fala M/S podem ser menores do que o número de canais múltiplos de áudio não M/S, na representação de canal de áudio de referência durante o qual o conteúdo da fala no conteúdo misturado é distribuído. Em algumas modalidades, mesmo quando o conteúdo da fala no conteúdo misturado é distribuído ao longo de dois ou mais canais de áudio não M/S, como os canais esquerdo e direito, etc., na configuração do canal de áudio de referência, apenas um conjunto de dados de controle M/S, os parâmetros de controle, etc., para a melhora da fala - por exemplo, correspondentes ao canal médio da representação M/S – são enviados como os metadados para melhora de fala M/S por um codificador de áudio para descodificadores à jusante. O conjunto único de dados de controle M/S, parâmetros de controle, etc., para a melhora da fala pode ser usado para realizar operações de melhora da fala para todos os dois ou mais canais de áudio não M/S, como os canais esquerdo e direito, etc. Em algumas modalidades, as matrizes de transformação entre a configuração de referência e a representação M/S podem ser usadas para aplicar as operações de melhora de fala baseadas no dados de controle M/S, parâmetros de controle, etc., para a melhora da fala como aqui descrito.

[045] Técnicas como aqui descritas podem ser usadas em cenários em que o conteúdo da fala é movimentada no centro fantasma dos canais esquerdo e direito, o conteúdo da fala não é completamente deslocado no centro (por exemplo, não igualmente alto em ambos os canais esquerdo e direito, etc.), etc. Em um exemplo, estas técnicas podem ser utilizadas em situações em que uma grande percentagem (por exemplo, 70+%, 80+%, 90+%, etc.) da energia do conteúdo da fala está no sinal médio ou canal médio da representação M/S. Em outro exemplo, (por exemplo, espacial, etc.), as transformações como movimentação, rotação, etc., podem ser usadas para transformar o conteúdo de fala inigualável na configuração de referência para ser igual ou substancialmente igual na configuração M/S. Os vetores de geração, matrizes de transformação, etc., que representam movimentação, rotação, etc.,

podem ser utilizados em como uma parte de, ou em conjunto com, as operações de melhora de voz.

[046] Em algumas modalidades (por exemplo, um modo híbrido, etc.), uma versão (por exemplo, uma versão reduzida, etc.) do conteúdo da fala é enviada para um decodificador de áudio à jusante como apenas um sinal de canal médio ou ambos canal médio e os sinais de canal lateral na representação M/S, juntamente com o conteúdo misturado enviado na configuração do canal de áudio de referência, possivelmente, com uma representação não M/S. Em algumas modalidades, quando a versão do conteúdo da fala é enviada para um decodificador de áudio à jusante uma vez que apenas um sinal do canal médio na representação M/S, um vetor de geração correspondente que opera (por exemplo, executa a transformação, etc.) no sinal de canal médio para gerar porções de sinais em um ou mais canais não M/S de uma configuração de canal de áudio não M/S (por exemplo, a configuração de referência, etc.) com base no sinal do canal médio é também enviado para o decodificador de áudio à jusante.

[047] Em algumas modalidades, um algoritmo de melhora de diálogo/fala (por exemplo, em um decodificador de áudio à jusante, etc.) que implementa comutação baseada em SNR temporal "cega" entre melhora codificada paramétrica (por exemplo, a predição de diálogo independente de canal, predição de diálogo multicanal, etc.) e melhora codificada por forma de onda de segmentos de um programa de áudio opera pelo menos em parte, na representação M/S.

[048] As técnicas como aqui descritas que implementam as operações de melhora de fala, pelo menos, parcialmente na representação M/S podem ser utilizadas com a predição independente de canal (por exemplo, no canal médio, etc.), a predição de multicanais (por exemplo, em canal médio e canal lateral, etc.), etc. Estas técnicas também podem ser usadas para suportar melhora de fala para um, dois ou mais diálogos ao mesmo tempo. Zero, um ou mais conjuntos adicionais de parâmetros de

controle, dados de controle, etc., como parâmetros de predição, ganhos, vetores de processamento, etc., pode ser fornecidos no sinal de áudio codificado como uma parte de metadados para melhora de fala M/S para suportar diálogos adicionais.

[049] Em algumas modalidades, a sintaxe do sinal de áudio codificado (por exemplo, a saída do codificador, etc.) suporta a transmissão de um sinalizador M/S a partir de um codificador de áudio à montante para descodificadores de áudio à jusante. O sinalizador M/S está presente/ajustado quando as operações de melhora de fala estão sendo realizadas pelo menos em parte com os dados de controle M/S, parâmetros de controle, etc., que são transmitidos com o sinalizador M/S. Por exemplo, quando o sinalizador M/S é definido, um sinal estéreo (por exemplo, a partir dos canais esquerdo e direito, etc.) em canais não M/S podem ser primeiro transformados por um decodificador de áudio de receptor para o do canal médio e o canal lateral da representação M/S antes de aplicar as operações de melhora de fala M/S, com os dados de controle M/S, parâmetros de controle, etc., tal como recebidos com o sinalizador M/S, de acordo com um ou mais dos algoritmos de melhora de fala (por exemplo, a predição de diálogo independente do canal, a predição de diálogo multicanal, baseado em forma de onda, híbrido paramétrico-forma de onda, etc.). Após as operações de melhoria da fala M/S serem realizadas, os sinais melhorados de fala na representação M/S podem ser transformados de volta para os canais não M/S.

[050] Em algumas modalidades, o programa de áudio cujo conteúdo de fala deve ser aumentado em conformidade com a invenção inclui canais de alto-falantes, mas não qualquer canal objeto. Em outras modalidades, o programa de áudio de fala cujo conteúdo deve ser melhorado de acordo com a invenção é um programa de áudio baseado em objeto (tipicamente um programa de áudio baseado em objeto multicanal) que compreende pelo menos um canal objeto e, opcionalmente, também, pelo menos, um canal de alto-falante.

[051] Um outro aspecto da invenção é um sistema que inclui um codificador configurado (por exemplo, programado) para realizar qualquer modalidade do método de codificação da invenção para gerar um fluxo de bits incluindo dados de áudio codificados, dados de forma de onda, e os dados paramétricos (e opcionalmente também um indicador de mistura (por exemplo, dados indicando a mistura) para cada segmento de dados de áudio), em resposta aos dados de áudio indicativos de um programa incluindo o conteúdo de fala e não fala, e um decodificador configurado para interpretar o fluxo de bits para recuperar os dados de áudio codificados (e, opcionalmente, ainda, cada indicador de mistura) e para decodificar os dados de áudio codificados para recuperar os dados de áudio. Alternativamente, o decodificador está configurado para gerar um indicador de mistura para cada segmento de dados de áudio, em resposta aos dados de áudio recuperados. O decodificador está configurado para executar a melhora de fala híbrida com os dados de áudio recuperados em resposta a cada indicador de mistura.

[052] Um outro aspecto da invenção é um decodificador configurado para realizar qualquer modalidade do método da invenção. Numa outra classe de modalidades, a invenção é um decodificador incluindo uma memória buffer (buffer) que armazena (por exemplo, de forma não transitória), pelo menos um segmento (por exemplo, quadro) de um fluxo de bits de áudio codificado que foi gerado por qualquer modalidade do método da invenção.

[053] Outros aspectos da invenção incluem um sistema ou dispositivo (por exemplo, um codificador, um decodificador, ou um processador) configurado (por exemplo, programado) para realizar qualquer modalidade do método da invenção, e um meio legível por computador (por exemplo, um disco) que armazena o código para a implementação de qualquer modalidade do método da invenção ou suas etapas. Por exemplo, o sistema da invenção pode ser ou incluir um processador programável de uso geral, o processador de sinal digital, ou microprocessador, programados com

software ou firmware e/ou de outra forma configurados para executar qualquer uma de uma variedade de operações sobre os dados, incluindo uma modalidade do método inventivo ou etapas dos mesmos. Um tal processador de uso geral pode ser ou incluir um sistema de computador, incluindo um dispositivo de entrada, uma memória e circuitos de processamento programados (e/ou de outra forma configurados) para executar uma modalidade do método da invenção (ou etapas do mesmo) em resposta aos dados afirmados ao mesmo.

[054] Em algumas modalidades, os mecanismos como aqui descritos formam uma parte de um sistema de processamento de mídia, incluindo mas não limitados a: um dispositivo audiovisual, uma TV de tela plana, um dispositivo portátil, máquinas de jogos, televisão, sistema de home-theater, tablets, dispositivo móvel, computador portátil, computador netbook, telefone celular, leitor de livro eletrônico, ponto de terminal de venda, computador desktop, estação de trabalho do computador, quiosque de computador, vários outros tipos de terminais e de unidades de processamento de mídia, etc.

[055] Várias modificações às modalidades preferenciais e os princípios e características genéricos aqui descritos serão prontamente evidentes para os especialistas na técnica. Assim, a descrição não se destina a ser limitada às modalidades mostradas, mas deve ser aplicado o mais amplo escopo consistente com os princípios e características aqui descritos.

2. NOTAÇÃO E NOMENCLATURA

[056] Ao longo desta descrição, incluindo nas concretizações, os termos "diálogo" e "fala" são utilizados alternadamente como sinônimos para designar o conteúdo do sinal de áudio percebido como uma forma de comunicação por um ser humano (ou personagem em um mundo virtual).

[057] Ao longo desta descrição, incluindo nas concretizações, a expressão executar uma operação "em" um sinal ou dados (por exemplo, a filtragem,

escalonamento, transformação, ou a aplicação de ganho para o sinal ou dados) é utilizada em um sentido amplo para denotar a execução da operação diretamente no sinal ou dados, ou em uma versão processada do sinal ou dados (por exemplo, em uma versão do sinal que foi submetido a filtração preliminar ou pré-processamento antes da realização da operação no mesmo).

[058] Ao longo desta descrição, incluindo nas concretizações, a expressão "sistema" é utilizada em um sentido amplo para designar um dispositivo, sistema ou subsistema. Por exemplo, um subsistema que implementa um decodificador pode ser referido como um sistema de decodificador, e um sistema que inclui um tal subsistema (por exemplo, um sistema que gera sinais de saída X, em resposta a várias entradas, em que o subsistema gera M das entradas e as outras entradas X - M são recebidas a partir de uma fonte externa) pode também ser referido como um sistema de decodificador.

[059] Ao longo desta descrição, incluindo nas concretizações, o termo "processador" é usado em um sentido amplo para designar um sistema ou dispositivo programável ou de outra forma configurável (por exemplo, com um software ou firmware) para executar operações sobre os dados (por exemplo, áudio, ou vídeo ou outros dados de imagem). Exemplos de processadores incluem uma matriz de gate programável por campo (ou outro circuito integrado configurável ou conjunto de chip), um processador de sinal digital programado e/ou configurado para executar o processamento em pipeline em dados de áudio ou outros dados de som, um processador de uso geral programável ou computador, e um chip de microprocessador programável ou conjunto de chip.

[060] Ao longo desta descrição, incluindo nas concretizações, as expressões "processador de áudio" e "unidade de processamento de áudio" são utilizadas indiferentemente, e em um sentido amplo, para designar um sistema configurado para processar dados de áudio. Exemplos de unidades de processamento de áudio

incluem, mas não estão limitados a, codificadores (por exemplo, transcodificadores), descodificadores, codecs, sistemas de pré-processamento, sistemas de pós-processamento, e sistemas de processamento de fluxo de bits (por vezes referidos como ferramentas de processamento de fluxo de bits).

[061] Ao longo desta descrição, incluindo nas concretizações, a expressão "metadados" refere-se a dados separados e diferentes dos dados de áudio correspondentes (conteúdo de áudio de um fluxo de bits que também inclui metadados). Metadados está associado com dados de áudio, e indica, pelo menos, uma característica ou características dos dados de áudio (por exemplo, o(s) tipo(s) de processamento já tenham sido executados, ou devem ser executados, nos dados de áudio, ou a trajetória de um objeto indicado pelos dados de áudio). A associação dos metadados com os dados de áudio é de tempo síncrono. Assim, metadados presentes (mais recentemente recebidos ou atualizados) podem indicar que os dados de áudio correspondentes simultaneamente têm uma característica indicada e/ou compreendem os resultados de um tipo indicado de processamento de dados de áudio.

[062] Ao longo desta descrição, incluindo nas concretizações, o termo "acopla" ou "acoplado" é usado para significar uma ligação direta ou indireta. Assim, se um primeiro dispositivo acopla a um segundo dispositivo, esta conexão pode ser por meio de ligação direta, ou através de uma ligação indireta através de outros dispositivos e conexões.

[063] Ao longo desta descrição, incluindo nas concretizações, as seguintes expressões têm as seguintes definições:

- falante e alto-falante são usados como sinônimos para designar qualquer transdutor de emissão de som. Esta definição inclui alto-falantes implementados como transdutores múltiplos (por exemplo, woofer e tweeter);

- alimentação de alto-falante: um sinal de áudio para ser aplicado diretamente

a um alto-falante ou um sinal de áudio que deve ser aplicado a um amplificador e alto-falante em série;

- canal (ou "canal de áudio"): um sinal de áudio monofônico. Dito um sinal pode tipicamente ser processado de tal modo que seja equivalente à aplicação do sinal diretamente a um alto-falante a uma posição desejada ou nominal. A posição desejada pode ser estática, como é normalmente o caso com alto-falantes físicos, ou dinâmica;

- programa de áudio: um conjunto de um ou mais canais de áudio (pelo menos um canal de alto-falante e/ou pelo menos um canal objeto) e opcionalmente também metadados associados (por exemplo, os metadados que descrevem uma apresentação de áudio espacial desejada);

- canal de alto-falante (ou "canal de alimentação de alto-falante"): um canal de áudio que está associado com um alto-falante nomeado (a uma posição desejada ou nominal), ou com uma zona de alto-falante nomeada dentro de uma configuração de alto-falante definida. Um canal de alto-falante é processado de tal modo a ser equivalente à aplicação do sinal de áudio diretamente para o alto-falante nomeado (na posição desejada ou nominal) ou para um alto-falante na zona do alto-falante nomeado;

- canal objeto: um canal de áudio indicativo de som emitido por uma fonte de áudio (por vezes referido como um áudio "objeto"). Normalmente, um canal objeto determina uma descrição de fonte de áudio paramétrica (por exemplo, metadados indicativos da descrição de fonte de áudio paramétrica estão incluídos ou fornecidos com o canal objeto). A descrição de fonte pode determinar o som emitido pela fonte (como uma função do tempo), a posição aparente (por exemplo, coordenadas espaciais 3D) da fonte como uma função do tempo, e, opcionalmente, pelo menos, um parâmetro adicional (por exemplo, o tamanho aparente da fonte ou largura) caracterizando a fonte;

- programa de áudio baseado em objeto: um programa de áudio que compreende um conjunto de um ou mais canais objetos (e, opcionalmente, compreendendo também pelo menos um canal de alto-falante) e opcionalmente metadados também associados (por exemplo, metadados indicativos de uma trajetória de um áudio objeto que emite som indicado por um canal objeto, ou metadados de outra forma indicativos de uma apresentação de áudio espacial desejada de som indicada por um canal objeto, ou metadados indicativos de uma identificação de, pelo menos, um áudio objeto que é uma fonte de som indicada por um canal objeto) ; e

- processar: o processo de converter um programa de áudio em uma ou mais alimentações de alto-falante, ou o processo de converter um programa de áudio em uma ou mais alimentações de alto-falante e converter as alimentações do alto-falante para soar usando um ou mais alto-falantes (neste último caso, o processamento é por vezes aqui referido como processando "pelo(s)" alto-falante(s)). Um canal de áudio pode ser trivialmente processado ("em" uma posição desejada) através da aplicação do sinal diretamente a um alto-falante física na posição desejada, ou um ou mais canais de áudio podem ser processados utilizando um de uma variedade de técnicas de virtualização concebidas para serem substancialmente equivalentes (para o ouvinte) a tal processamento trivial. Neste último caso, cada canal de áudio pode ser convertido para uma ou mais alimentações de alto-falantes sendo aplicadas para alto-falante(s) em locais conhecidos, que são, em geral, diferentes da posição desejada, de tal modo que som emitido pelo(s) alto-falante(s) em resposta às alimentações será percebido como emitindo a partir da posição desejada. Exemplos de tais técnicas de virtualização incluem processamento binaural via fones de ouvido (por exemplo, usando o processamento Dolby Headphone, que simula até 7.1 canais de som surround para o usuário dos fones de ouvido) e síntese de campo de onda.

[064] As modalidades da invenção de codificação, descodificação, e

métodos de melhora de fala, e sistemas configurados para implementar os métodos serão descritas com referência à FIGURA 3, a FIGURA 6, e FIGURA 7.

3 . GERAÇÃO DE PARÂMETROS DE PREDIÇÃO

[065] A fim de executar a melhora de fala (incluindo melhora de fala híbrida de acordo com modalidades da invenção), é necessário ter acesso ao sinal de fala sendo reforçado. Se o sinal de fala não está disponível (separadamente a partir de uma mistura do conteúdo da fala e da não fala do sinal misturado a ser melhorado) no momento em que a melhora da fala é executada, técnicas paramétricas podem ser usadas para criar uma reconstrução da fala da mistura disponível.

[066] Um método para a reconstrução paramétrica do conteúdo da fala de um sinal de conteúdo misturado (indicativo de uma mistura de conteúdo de fala e não fala) baseia-se em reconstruir a potência da fala em cada ladrilho da frequência de tempo do sinal, e gera parâmetros de acordo com:

$$p_{n,b} = \sqrt{\sum_{s \in n, f \in b} \frac{D_{s,f}^2}{M_{s,f}^2}} \qquad (2)$$

em que $p_{n,b}$ é o parâmetro (valor de melhora de fala codificada paramétrica) para o ladrilho tendo índice temporal *n* e índice de banda de frequência *b*, o valor $D_{s,f}$ representa o sinal de fala em ranhura de tempo *s* e bin de frequência *f* do ladrilho, o valor $M_{s,f}$ representa o sinal do conteúdo misturado na mesma ranhura de tempo e bin de frequência do ladrilho, e o somatório é sobre todos os valores de *s* e *f* em todos os ladrilhos. Os parâmetros $p_{n,b}$ podem ser entregues (como metadados) com o próprio sinal de conteúdo misturado, para permitir que um receptor reconstitua o conteúdo da fala de cada segmento do sinal de conteúdo misturado.

[067] Como representado na FIGURA 1, cada parâmetro $p_{n,b}$ pode ser determinado através da realização de uma transformação de domínio de tempo para o domínio de frequência no sinal de conteúdo misturado ("áudio misturado"), cujo conteúdo da fala deve ser melhorado, realizando um domínio tempo para a transformação de domínio de frequência em fala sinal (o conteúdo da fala do sinal de

conteúdo misturado), integrando a energia (de cada ladrilho de frequência de tempo com índice temporal *n* e índice de bandas de frequência *b* do sinal de fala) ao longo de todos os slots tempo e depósitos de frequências em ladrilho, e integrando a energia da ladrilho de frequência de tempo correspondente do sinal de conteúdo misturado sobre todos os slots de tempo e depósitos de frequência na ladrilho, e dividindo o resultado da primeira integração pelo resultado da segunda integração para gerar o parâmetro $p_{n,b}$ sobre a ladrilho.

[068] Quando cada ladrilho de frequência no tempo do sinal de conteúdo misturado é multiplicado pelo parâmetro $p_{n,b}$ para o ladrilho, o sinal resultante tem envelopes espectrais e temporais semelhantes como o conteúdo da fala do sinal de conteúdo misturado.

[069] Programas de áudio típicos, por exemplo, programas de áudio estéreo ou 5.1 canais, incluem múltiplos canais de alto-falante. Tipicamente, cada canal (ou cada um de um subconjunto de canais) é indicativo do conteúdo da fala e não fala, e um sinal de conteúdo misturado determina cada canal. O método de reconstrução da fala paramétrico descrito pode ser aplicado de forma independente para cada canal para reconstruir o componente de fala de todos os canais. Os sinais de fala reconstruídos (um para cada um dos canais) podem ser adicionados sinais de canal de conteúdo misturado correspondentes, com um ganho adequado para cada canal, para alcançar um aumento desejado do conteúdo da fala.

[070] Os sinais de conteúdo misturado (canais) de um programa multicanal podem ser representados como um conjunto de vetores de sinal, em que cada elemento do vetor é uma coleção de ladrilhos de frequência no tempo que corresponde a um conjunto de parâmetros específicos, ou seja, todo os depósitos de frequência (f) na banda de parâmetro (b) e slots de tempo (s) no quadro (n). Um exemplo de um tal conjunto de vetores, para um sinal de conteúdo misturado de três canais é:

$$M_{n,b} = \begin{pmatrix} M_{c_1,n,b} \\ M_{c_2,n,b} \\ M_{c_3,n,b} \end{pmatrix} \quad (3)$$

em que c_i um indica o canal. O exemplo assume três canais, mas o número de canais é uma quantidade arbitrária.

[071] Do mesmo modo o conteúdo da fala de um programa de multicanais pode ser representado como um conjunto de matrizes 1 x 1 (onde o conteúdo da fala consiste em apenas um canal), $D_{n,b}$. A multiplicação de cada elemento da matriz do sinal de conteúdo misturado com um valor escalar resulta em uma multiplicação de cada sub-elemento com o valor escalar. Um valor de fala reconstruído para cada ladrilho é assim obtido através do cálculo

$$D_{r,n,b} = diag(P) \cdot M_{n,b} \quad (4)$$

para cada *n* e *b*, onde *P* é uma matriz cujos elementos são parâmetros de predição. A fala reconstruída (para todos os ladrilhos) também pode ser denotada por:

$$D_r = diag(P) \cdot M \quad (5)$$

[072] O conteúdo nos multicanais de um sinal de conteúdo misturado multicanal faz com que as correlações entre os canais que podem ser utilizados para fazer uma melhor predição do sinal de fala. Ao empregar um preditor de erro mínimo quadrado médio (MMSE) (por exemplo, de um tipo convencional), os canais podem ser combinados com os parâmetros de predição de modo a reconstruir o conteúdo da fala com um erro mínimo de acordo com o critério de erro quadrado médio (MSE). Como mostrado na FIGURA 2, assumindo um sinal de entrada de conteúdo misturado de três canais, como um preditor MMSE (operando no domínio da frequência) iterativamente gera um conjunto de parâmetros de predição p_i (em que o índice *i* é 1, 2, ou 3), em resposta ao sinal de entrada de conteúdo misturado e um único sinal de fala de entrada indicativo do conteúdo da fala do sinal de entrada de conteúdo misto.

[073] Um valor da fala reconstruído a partir de um ladrilho de cada canal do sinal de entrada de conteúdo misturado (cada ladrilho tendo os mesmos índices *n* e *b*) é uma combinação linear do conteúdo ($M_{ci,\ n,b}$) de cada canal (*i* = 1, 2, ou 3) do sinal

de conteúdo misturado controlado por um parâmetro de peso para cada canal. Estes parâmetros de peso são os parâmetros de predição, p_i, para os ladrilhos tendo os mesmos índices *n* e *b*. Assim, a fala reconstruída a partir de todos os ladrilhos de todos os canais do sinal de conteúdo misturado é:

$$D_r = p_1 \cdot M_{c1} + p_2 \cdot M_{c2} + p_3 \cdot M_{c3} \qquad (6)$$

ou em forma de matriz de sinais:

$$D_r = PM \qquad (7)$$

[074] Por exemplo, quando a fala é coerentemente presente em múltiplos canais do sinal de conteúdo misturado enquanto sons de fundo (não fala) são incoerentes entre os canais, uma combinação aditiva de canais irá favorecer a energia da fala. Para dois canais isto resulta numa melhor separação da fala 3 dB em relação à reconstrução independente de canal. Como outro exemplo, quando a fala está presente em um canal e sons de fundo estão coerentemente presentes em múltiplos canais, uma combinação subtrativa de canais irá (parcialmente) eliminar os sons de fundo enquanto a fala é preservada.

[075] Numa classe de modalidades, o método da invenção inclui as etapas de: (a) receber um fluxo de bits indicativo de um programa de áudio, incluindo fala tendo uma forma de onda não melhorada e outros conteúdos de áudio, em que o fluxo de bits inclui: dados de áudio não melhorados indicativos da fala e o outro conteúdo de áudio, dados de forma de onda indicativos de uma versão reduzida de qualidade da fala, em que a versão reduzida de qualidade da fala tem uma segunda forma de onda semelhante (por exemplo, pelo menos, substancialmente semelhante) para a forma de onda não melhorada, e a versão reduzida de qualidade teria qualidade questionável se ouvida isoladamente, e os dados paramétricos, em que os dados paramétricos com os dados de áudio não melhorados determinam fala parametricamente construída, e a fala parametricamente construída é uma versão parametricamente reconstruída da fala, que corresponde a pelo menos

substancialmente (por exemplo, é um boa aproximação de) a fala; e (b) realização de melhora da fala sobre o fluxo de bits em resposta a um indicador de mistura, gerando assim dados indicativos de um programa de áudio de fala melhorada, incluindo através da combinação dos dados de áudio não melhorados com uma combinação de dados de fala de baixa qualidade determinados a partir dos dados de forma de onda, e dados de fala reconstruídos, em que a combinação é determinada pelo indicador de mistura (por exemplo, a combinação tem uma sequência de estados determinada por uma sequência de valores atuais do indicador de mistura), os dados de fala reconstruídos são gerados em resposta a, pelo menos, alguns dos dados paramétricos e, pelo menos, alguns dos dados áudio não melhorados, e o programa de áudio de fala melhorada tem menos artefatos de codificação de melhora da fala menos audíveis (por exemplo, artefatos de codificação de melhora da fala que são mais bem mascarados) seriam ou simplesmente um programa de áudio de fala melhorada codificada por forma de onda determinado pela combinação apenas dos dados de fala de baixa qualidade com os dados de áudio não melhorados ou simplesmente um programa de áudio de fala melhorada codificada paramétrica determinado a partir dos dados paramétricos e os dados de áudio não melhorados.

[076] Em algumas modalidades, o indicador de mistura (que pode ter uma sequência de valores, por exemplo, um para cada um dos segmentos de uma sequência de fluxo de bits) está incluído no fluxo de bits recebido na etapa (a). Em outras modalidades, o indicador de mistura é gerado (por exemplo, em um receptor que recebe e decodifica o fluxo de bits), em resposta ao fluxo de bits.

[077] Deve ser entendido que a expressão "indicador de mistura" não se destina a indicar um único parâmetro ou valor (ou uma sequência de parâmetros individuais ou valores) para cada segmento do fluxo de bits. Em vez disso, contempla-se que em algumas modalidades, um indicador de mistura (para um segmento do fluxo de bits) pode ser um conjunto de dois ou mais parâmetros ou valores (por exemplo,

para cada segmento, um parâmetro de controle de melhora codificada paramétrica e um parâmetro de controle de melhora codificada por forma de onda). Em algumas modalidades, o indicador de mistura para cada segmento pode ser uma sequência de valores indicando a mistura por banda de frequência do segmento.

[078] Os dados da forma de onda e os dados paramétricos não precisam ser fornecidos para (por exemplo, incluídos em) cada segmento do fluxo de bits, ou usados para executar a melhora de fala em cada segmento do fluxo de bits. Por exemplo, em alguns casos, pelo menos, um segmento pode incluir apenas os dados de forma de onda (e a combinação determinada pelo indicador de mistura para cada tal segmento pode consistir em apenas os dados em forma de onda) e pelo menos um outro segmento pode incluir apenas os dados paramétricos (e a combinação determinada pelo indicador de mistura para cada tal segmento pode consistir somente em dados de fala reconstruídos).

[079] É contemplado que, em algumas modalidades, um codificador gera o fluxo de bits incluindo por codificação (por exemplo, compressão) os dados de áudio não melhorados, mas não os dados da forma de onda ou os dados paramétricos. Assim, quando o fluxo de bits é enviado para um receptor, o receptor pode analisar o fluxo de bits para extrair os dados de áudio não melhoradas, os dados de forma de onda, e os dados paramétricos (e o indicador de mistura, se for entregue no fluxo de bits), mas apenas decodificar os dados de áudio não melhorados. O receptor poderia executar a melhora da fala nos dados de áudio descodificados não melhorados (com base nos dados da forma de onda e/ou dados paramétricos) sem aplicar aos dados de forma de onda ou aos dados paramétricos o mesmo processo de descodificação que é aplicado aos dados de áudio.

[080] Normalmente, a combinação (indicada pelo indicador de mistura) dos dados da forma de onda e as alterações de dados de fala reconstruídos ao longo do tempo, com cada estado da combinação relativa à fala e outros conteúdos de áudio

de um segmento correspondente do fluxo de bits. O indicador de mistura é gerado de tal modo que o estado atual da combinação (de dados da forma de onda e os dados de fala reconstruídos) é determinado por propriedades do sinal da fala e outros conteúdos de áudio (por exemplo, uma proporção entre a potência de conteúdo de fala e o poder de outro conteúdo de áudio) no segmento correspondente do fluxo de bits.

[081] A etapa (b) pode incluir uma etapa de realização de melhora da fala codificada por forma de onda pela combinação (por exemplo, mistura ou mistura), pelo menos, de alguns dos dados de fala de baixa qualidade com os dados de áudio não melhorados de, pelo menos, um segmento do fluxo de bits, e realizando melhora de fala codificada paramétrica pela combinação de dados de fala reconstruídos com os dados de áudio não melhorados de pelo menos um segmento do fluxo de bits. Uma combinação de melhora de fala codificada por forma de onda e melhora de fala codificada paramétrica é realizada em pelo menos um segmento do fluxo de bits por mistura de ambos os dados de fala e dados de baixa qualidade de fala reconstruídos para o segmento com os dados de áudio não melhorados do segmento. Sob algumas condições de sinal, apenas um (mas não ambos) da melhora de fala codificada por forma de onda fala e melhora de fala codificada paramétrica é realizado (em resposta ao indicador de mistura) em um segmento (ou em cada um de mais do que um segmento) do fluxo de bits.

4. OPERAÇÕES DE MELHORA DE FALA

[082] Aqui, "SNR" (relação sinal para ruído) é utilizada para designar a proporção da potência (ou nível) do componente de fala (isto é, conteúdo da fala) de um segmento de um programa de áudio (ou de todo o programa) ao do componente não fala (isto é, o conteúdo não fala) do segmento ou programa ou para todo o conteúdo (fala e não fala) do segmento ou programa. Em algumas modalidades, a SNR é derivada de um sinal de áudio (para submeter-se à melhoria da fala) e um sinal

separado indicativo de conteúdo da fala do sinal de áudio (por exemplo, uma cópia de baixa qualidade do conteúdo de fala que foi gerado para o uso na melhora codificada por forma de onda). Em algumas modalidades, a SNR é derivada de um sinal de áudio (para submeter-se à melhoria da fala) e a partir de dados paramétricos (que foram gerados para o uso na melhora codificada paramétrica do sinal de áudio).

[083] Em uma classe de modalidades, o método da invenção implementa comutação baseada em SNR temporal "cega" entre melhora codificada paramétrica e melhora codificada por forma de onda de segmentos de um programa de áudio. Neste contexto, "cega" denota que a comutação não é perceptivelmente guiada por um modelo de mascaramento auditório (por exemplo, de um tipo a ser descrito aqui), mas é guiado por uma sequência de valores SNR (indicadores de mistura) correspondendo aos segmentos do programa. Em uma modalidade nesta classe, melhora de fala codificada híbrida é alcançada por comutação temporal entre melhora codificada paramétrica e melhora codificada por forma de onda (em resposta a um indicador de mistura, por exemplo, um indicador de mistura gerado em subsistema 29 do codificador da FIGURA 3, que indica que apenas melhora codificada paramétrica ou melhora codificada por forma de onda deve ser realizada nos dados de áudio correspondentes), de modo que melhora codificada paramétrica ou melhora codificada por forma de onda (mas não ambas melhora codificada paramétrica e melhora codificada por forma de onda) é realizada em cada segmento de um programa de áudio em que a melhora de fala é realizada. Reconhecer que melhora codificada por forma de onda realizada melhor sob a condição de baixa SNR (em segmentos tendo valores baixos de SNR) e realiza melhora codificada paramétrica melhor em SNRs favoráveis (em segmentos tendo valores altos de SNR), a decisão de comutação é tipicamente baseada na proporção de fala (diálogo) para o áudio restante em uma mistura de áudio original.

[084] As modalidades que implementam comutação temporal baseada em

SNR "cega" tipicamente incluem as etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias de tempo consecutivas (segmentos), e determinar para cada segmento a SNR entre o conteúdo de fala e o outros conteúdos de áudio (ou entre o conteúdo da fala e conteúdo de áudio total) do segmento; e para cada segmento, comparar o SNR a um limiar e proporcionar um parâmetro de controle de melhora codificada paramétrica para o segmento (isto é, o indicador de mistura para o segmento indica que a melhora codificada paramétrica deve ser realizada) quando a SNR é maior do que o limiar ou o fornecimento de um parâmetro de controle de melhora codificada por forma de onda para o segmento (isto é, o indicador de mistura para o segmento indica que a melhora codificada por forma de onda deve ser realizada) quando a SNR não é maior do que o limiar.

[085] Quando o sinal áudio não melhorado é fornecido (por exemplo, transmitido) com os parâmetros de controle incluídos como metadados para um receptor, o receptor pode executar (em cada segmento) do tipo de melhora de fala indicada pelo parâmetro de controle para o segmento. Assim, o receptor realiza melhora codificada paramétrica em cada segmento para o qual o parâmetro de controle é um parâmetro de controle melhora codificada paramétrica, e melhora codificada por forma de onda em cada segmento para o qual o parâmetro de controle é um parâmetro de controle de melhora codificada por forma de onda.

[086] Se alguém está disposto a arcar com os custos de transmissão (com cada segmento de uma mistura de áudio original) ambos os dados de forma de onda (para a implementação de melhora da fala codificada por forma de onda) e parâmetros de melhora codificada paramétrica com uma mistura original (não melhorada), um maior grau de melhora de fala pode ser conseguido através da aplicação de ambos melhora codificada por forma de onda e melhora codificada paramétrica aos segmentos individuais da mistura. Assim, numa classe de modalidades, o método da invenção implementa mistura baseada em SNR temporal "cega" entre melhora

codificada paramétrica e melhora codificada por forma de onda de segmentos de um programa de áudio. Neste contexto, também, "cega" indica que a comutação não é perceptivamente guiada por um modelo de mascaramento auditivo complexo (por exemplo, de um tipo a ser descrito aqui), mas é guiada por uma sequência de valores de SNR correspondentes aos segmentos de programa.

[087] As modalidades que implementam mistura à base de SNR temporal "cega" tipicamente incluem as etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias de tempo consecutivas (segmentos), e determinação para cada segmento a SNR entre o conteúdo de fala e o outros conteúdos de áudio (ou entre o conteúdo da fala e conteúdo de áudio total) do segmento; determinação (por exemplo, receber um pedido para) uma quantidade total ("T") de melhora da fala; e para cada segmento, proporcionando um parâmetro de controle de mistura, onde o valor do parâmetro de controle de mistura é determinado por (é uma função de) SNR para o segmento.

[088] Por exemplo, o indicador de mistura para um segmento de um programa de áudio pode ser um parâmetro indicador de mistura (ou conjunto de parâmetros) gerado no subsistema 29 do codificador da FIGURA 3 para o segmento.

[089] O indicador de controle de mistura pode ser um parâmetro, α, para cada segmento, tais que $T = \alpha Pw + (1-\alpha)Pp$, em que Pw é a melhora codificada por forma de onda para o segmento que produziria o montante total predeterminado de melhora, T, se aplicado ao conteúdo de áudio não melhorado do segmento utilizando os dados da forma de onda previstos para o segmento (onde o conteúdo da fala do segmento tem uma forma de onda não melhorada, os dados de forma de onda para o segmento são indicativos de uma versão reduzida de qualidade do conteúdo da fala do segmento, a versão reduzida de qualidade tem uma forma de onda semelhante (por exemplo, pelo menos substancialmente semelhante) para a forma de onda não melhorada, e a versão reduzida de qualidade do conteúdo da fala é de qualidade

questionável quando processada e percebida de forma isolada), e Pp é a melhora codificada paramétrica que produziria o montante total predeterminado de melhora, T, se aplicado ao conteúdo de áudio não melhorado do segmento utilizando os dados paramétricos fornecidos para o segmento (em que os dados paramétricos para o segmento, com o conteúdo de áudio não melhorado do segmento, determinam uma versão parametricamente reconstruída do conteúdo da fala do segmento).

[090] Quando o sinal áudio é fornecido não melhorado (por exemplo, transmitido) com os parâmetros de controle como metadados para um receptor, o receptor pode executar (em cada segmento) a melhora de fala híbrida indicada pelos parâmetros de controle para o segmento. Alternativamente, o receptor gera os parâmetros de controle a partir do sinal de áudio não melhorado.

[091] Em algumas modalidades, o receptor executa (em cada segmento do sinal de áudio não melhorado) uma combinação de melhora codificada paramétrica Pp (escalonado pelo parâmetro α para o segmento) e melhora codificada por forma de onda Pw (escalonado pelo valor (1 - α) para o segmento), tal que a combinação de melhora codificada paramétrica dimensionada e melhora codificada por forma de onda dimensionada gera o valor total predeterminado de melhora, como na expressão (1) ($T = \alpha Pw + (1-\alpha)Pp$).

[092] Um exemplo da relação entre α e SNR para um segmento é como se segue: α é uma função não decrescente de SNR, o intervalo de α é de 0 a 1, α tem o valor de 0, quando a SNR para o segmento é inferior ou igual a um valor de limiar ("SNR_poor"), e α tem o valor 1 quando a SNR é maior do que ou igual a um maior valor de limiar ("SNR_high"). Quando a SNR é favorável, α é alta, resultando em uma grande proporção de melhora codificada paramétrica. Quando a SNR é baixa, α é baixo, o que resulta em uma grande proporção de melhora codificada por forma de onda. A localização dos pontos de saturação (SNR_poor e SNR_high) devem ser selecionadas para acomodar as implementações específicas de ambos os algoritmos

da melhora codificada por forma de onda e codificada paramétrica.

[093] Numa outra classe de modalidades, a combinação de melhora codificada por forma de onda e codificada paramétrica para ser executada em cada segmento de um sinal de áudio é determinada por um modelo de mascaramento auditivo. Em algumas modalidades desta classe, a proporção de mistura ideal para uma mistura de melhora codificada por forma de onda e codificada paramétrica para ser executada em um segmento de um programa de áudio utiliza a maior quantidade de intensificação codificada por forma de onda que apenas mantém o ruído de codificação de se tornar audível.

[094] Nas modalidades de mistura à base de SNR cegas acima descritas, a proporção de mistura para um segmento é derivada a partir da SNR, e a SNR é assumida como sendo indicativa da capacidade da mistura de áudio para mascarar o ruído de codificação na versão reduzida de qualidade (cópia) da fala a ser utilizada para a melhora codificada por forma de onda. Vantagens da abordagem baseada na SNR cega são a simplicidade na implementação e baixa carga computacional no codificador. No entanto, a SNR é um indicador pouco fiável de quão bem o ruído de codificação será mascarado e uma grande margem de segurança deve ser aplicada para garantir que o ruído de codificação permanecerá mascarado em todos os momentos. Isto significa que, pelo menos, uma parte do tempo o nível de qualidade da cópia de fala reduzida que é misturada é inferior ao que poderia ser, ou, se a margem é definida de forma mais agressiva, o ruído de codificação torna-se audível durante algum tempo. A contribuição de melhora codificada por forma de onda no esquema de codificação híbrido da invenção pode ser aumentado, assegurando que o ruído de codificação não se torna audível usando um modelo de mascaramento auditivo para prever com mais precisão como o ruído de codificação na cópia da fala reduzida de qualidade está sendo mascarada pela mistura de áudio do programa principal e para selecionar a proporção de mistura, por conseguinte.

[095] As modalidades típicas que empregam um modelo de mascaramento auditivo incluem etapas de: segmentar o sinal de áudio não melhorado (mistura de áudio original) em fatias consecutiva (segmentos), e proporcionando uma cópia da fala de qualidade reduzida em cada segmento (para utilização na melhora codificada por forma de onda) e os parâmetros do melhora codificada paramétrica (para utilização na melhora codificada paramétrica) para cada segmento; para cada um dos segmentos, utilizando o modelo de mascaramento auditivo para determinar uma quantidade máxima de melhora codificada por forma de onda que pode ser aplicada sem artefatos se tornando audíveis; e gerar um indicador de mistura (para cada segmento do sinal de áudio não melhorado) de uma combinação de melhora codificada por forma de onda (em uma quantidade que não exceda a quantidade máxima de melhora codificada por forma de onda determinada utilizando o modelo de mascaramento auditivo para o segmento, e que, de preferência, pelo menos, corresponde substancialmente a quantidade máxima de melhora codificada por forma de onda determinada utilizando o modelo de mascaramento auditivo para o segmento) e melhora codificada paramétrica, de tal modo que a combinação de melhora codificada paramétrica e melhora codificada por forma de onda gera uma quantidade total predeterminada de melhora de fala para o segmento.

[096] Em algumas modalidades, cada dito indicador de mistura é incluído (por exemplo, por um codificador) em um fluxo de bits, que também inclui os dados de áudio codificados indicativos do sinal de áudio não melhorado. Por exemplo, o subsistema 29 de codificador 20 da FIGURA 3 pode ser configurado para gerar tais indicadores de mistura, e o subsistema 28 de codificador 20 pode ser configurado para incluir os indicadores de mistura no fluxo de bits a ser emitido a partir do codificador 20. Para outro exemplo, os indicadores de mistura podem ser gerados (por exemplo, no subsistema 13 do codificador da FIGURA 7) a partir dos parâmetros g_{max}(t) gerados pelo subsistema 14 do codificador da FIGURA 7, e subsistema 13 do codificador da

FIGURA 7 pode ser configurado para incluir os indicadores de mistura no fluxo de bits a ser produzida a partir do codificador da FIGURA 7 (ou subsistema 13 pode incluir, no fluxo de bits a ser emitido a partir do codificador da FIGURA 7, os parâmetros g_{max}(t) gerados pelo subsistema 14, e um receptor que recebe e analisa a fluxo de bits pode ser configurado para gerar os indicadores da mistura em resposta para os parâmetros g_{max}(t).

[097] Opcionalmente, o método também inclui uma etapa de realizar (em cada um dos segmentos do sinal de áudio não melhorado) em resposta ao indicador de mistura para cada segmento, a combinação de melhora codificada por forma de onda e melhora codificada paramétrica determinada pelo indicador de mistura, de tal modo que a combinação de melhora codificada por forma de onda e melhora codificada paramétrica gera o valor total predeterminado de melhora de fala para o segmento.

[098] Um exemplo de uma modalidade do método da invenção que emprega um modelo de mascaramento auditivo será descrito com referência à FIGURA 7. Neste exemplo, uma mistura de fala e áudio de fundo, A(t) (a mistura de áudio não melhorada) é determinada (no elemento 10 da FIGURA 7) e passada para o modelo de mascaramento auditivo (implementado pelo elemento 11 da FIGURA 7) no qual se prevê um limiar de mascaramento Θ(f,t) para cada segmento da mistura de áudio não melhorado. A mistura de áudio não melhorada A(t) é também fornecida ao elemento de codificação 13 para a codificação para a transmissão.

[099] O limiar de mascaramento gerado pelo modelo indica como uma função de frequência e tempo a excitação auditiva que qualquer sinal deve exceder, a fim de ser audível. Tais modelos de mascaramento são bem conhecidos na técnica. O componente de fala, s(t), de cada segmento da mistura de áudio não melhorado, A(t), é codificado (em baixa taxa de bits do codificador de áudio 15) para gerar uma cópia de qualidade reduzida, s’(t), do conteúdo de fala do segmento. A cópia de qualidade

reduzida, s'(t) (que compreende menos bits do que a fala original, s(t)), pode ser conceitualizada como a soma da fala original, s(t), e ruído de codificação, n(t). Esta codificação de ruído pode ser separada a partir da cópia de qualidade reduzida para análise através de subtração (no elemento 16) do sinal de fala alinhado no tempo, s(t), a partir da cópia de qualidade reduzida. Alternativamente, o ruído de codificação pode ser diretamente a partir do codificador de áudio.

[0100] O ruído de codificação, n, é multiplicado no elemento 17 por um fator de escalonamento, g(t), e o ruído de codificação escalonado é passado para um modelo auditivo (implementado pelo elemento 18) que prevê a excitação auditiva, N(f,t), gerada pelo ruído de codificação escalonado. Tais modelos de excitação são conhecidos na técnica. Em um etapa final, a excitação auditiva N(f,t) é comparada com o limiar de mascaramento previsto Θ(f,t) e o maior fator de escalonamento, g_{max}(t), que garante que o ruído de codificação é mascarado, isto é, quanto maior valor de g(t) que assegura que a N(f,t) < Θ(f,t), é encontrado (no elemento 14). Se o modelo auditivo é não linear, este pode ter de ser feito de forma iterativa (como indicado na Fig 2) por iteração do valor de g(t) aplicado ao ruído de codificação, n(t) no elemento 17; se o modelo auditivo é linear isto pode ser feito em um etapa para a frente simples de alimentação. O fator de escalonamento g_{max}(t) resultante é o maior fator de escalonamento que pode ser aplicado para a cópia de fala de qualidade reduzida, s’(t), antes de ser adicionado ao segmento correspondente da mistura de áudio não melhorada, A(t), sem os artefatos de codificação na cópia de fala de qualidade reduzida escalonada tornando-se audível na mistura da cópia de fala de qualidade reduzida escalonada, g_{max}(t)* s’(t), e a mistura de áudio não melhorada, A(t).

[0101] O sistema da FIGURA 7 também inclui elemento 12, o qual está configurado para gerar (em resposta à mistura de áudio não melhorada, A(t) e a fala, s(t)) parâmetros de melhora codificada paramétrica, p(t), para a realização de melhora de fala codificada paramétrica sobre cada segmento do mistura de áudio não

melhorada.

[0102] Os parâmetros da melhora codificada paramétrica, p(t), bem como a cópia de fala de qualidade reduzida, s'(t), gerados no codificador 15, e o fator, g_{max}(t), gerado no elemento 14, para cada segmento de programa de áudio, são também afirmados que codificam para elemento 13. Elemento 13 gera um fluxo de bits de áudio codificado indicativo da mistura de áudio não melhorada, A(t), os parâmetros de melhora codificada paramétrica,p (t), cópia de fala de reduzida qualidade, s'(t), e o fator, g_{max}(t), para cada segmento de programa de áudio, e este fluxo de bits codificado de áudio pode ser transmitido ou de outra forma entregues a um receptor.

[0103] No exemplo, a melhora da fala é executada (por exemplo, em um receptor ao qual a saída codificada do elemento 13 foi entregue) como se segue em cada segmento da mistura de áudio não melhorada, A(t), para aplicar um (por exemplo, solicitado) valor total predeterminado de melhoria, T, utilizando o fator de escalonamento g_{max}(t) para o segmento. O programa de áudio codificado é descodificado para extrair a mistura de áudio não melhorada, A(t), os parâmetros de melhora codificada paramétrica, p(t), a cópia de fala de qualidade reduzida, s'(t), e o fator g_{max}(t) para cada segmento do programa de áudio. Para cada segmento, a melhora codificada por forma de onda, Pw, é determinada para ser a melhora codificada por forma de onda que produziria o valor total predeterminado da melhora, T, se aplicado ao conteúdo de áudio não melhorado do segmento usando a cópia de fala de qualidade reduzida, s'(t), para o segmento, e melhora codificada paramétrica, Pp, é determinado sendo a melhora codificada paramétrica que produziria o valor total predeterminado de melhora, T, se aplicado ao conteúdo de áudio não melhorado do segmento utilizando os dados paramétricos fornecidos para o segmento (em que os dados paramétricos para o segmento, com o conteúdo de áudio não melhorado do segmento, determinam uma versão reconstruída parametricamente do conteúdo da fala do segmento). Para cada segmento, uma combinação de melhora codificada

paramétrica (numa quantidade escalonada por um parâmetro α_2 para o segmento) e melhora codificada por forma de onda (numa quantidade determinada pelo valor α_1 para o segmento) é realizada, de tal modo que a combinação da melhora codificada paramétrica e melhora codificada por forma de onda gera o valor total predeterminado de melhora utilizando o maior valor da melhora codificada por forma de onda permitida pelo modelo: $T = (\alpha_1(Pw) + \alpha_2(Pp))$, onde, fator α_1 é o valor máximo que não excede $g_{max}(t)$ para o segmento e permite a realização da igualdade indicada ($T = (\alpha_1(Pw) + \alpha_2(Pp))$, e o parâmetro α_2 é o valor mínimo não negativo, que permite a realização da igualdade indicada ($T = (\alpha_1(Pw) + \alpha_2(Pp))$.

[0104] Em uma modalidade alternativa, os artefatos da melhora codificada paramétrica são incluídos na avaliação (executada pelo modelo de mascaramento auditivo), de modo a permitir que os artefatos de codificação (devido à melhora codificada por forma de onda) para se tornar audível quando esta está favorável ao longo dos artefatos da melhora codificada paramétrica.

[0105] Em variações na modalidade da FIGURA 7 (e modalidades semelhante às da FIGURA 7, que utilizam um modelo de mascaramento auditivo), por vezes referidas como modalidades separadoras de multibanda guiadas por modelo auditivo, a relação entre o ruído de codificação de melhora codificada por forma de onda, N(f,t), na cópia da fala de qualidade reduzida e o limiar de mascaramento Θ(f,t) pode não ser uniforme em todas as bandas de frequência. Por exemplo, as características espectrais do ruído de codificação com melhora codificada por forma de onda podem ser de tal modo que em uma primeira região de frequência o ruído de mascaramento é de cerca para exceder o limiar de mascaramento, enquanto numa segunda região de frequência o ruído de mascaramento é bem abaixo do limiar mascarado. Na modalidade da FIGURA 7, a contribuição máxima de melhora codificada por forma de onda seria determinada pelo ruído de codificação na primeira região de frequência e o fator de escalonamento máximo, g, que pode ser aplicado para a cópia de fala de

qualidade reduzida é determinada pelo ruído de codificação e as propriedades de mascaramento na primeira região da frequência. Ele é menor do que o máximo fator de escalonamento, g, que pode ser aplicado, se a determinação do fator de escalonamento máxima foi baseada somente na segunda região de frequência. O desempenho global poderia ser melhorado se os princípios da mistura temporal foram aplicados separadamente nas duas regiões de frequência.

[0106] Numa implementação de divisão multi-banda guiada por modelo auditivo, o sinal de áudio não melhorado é dividido em M bandas de frequência contíguas não sobrepostas e os princípios da mistura temporal (isto é, a melhora da fala híbrida com uma mistura de melhora codificada por forma de onda e codificada paramétrica, de acordo com uma modalidade da invenção) são aplicados de forma independente em cada uma das M bandas. Uma alternativa de implementação de partições do espectro em uma banda baixa abaixo de uma frequência de corte, fc, e uma banda elevada acima da frequência de corte, fc. A banda baixa é sempre melhorada com melhora codificada por forma de onda e a banda superior é sempre melhorada com melhora codificada paramétrica. A frequência de corte é variada ao longo do tempo e sempre selecionada para ser tão elevado quanto possível de acordo com a restrição de que a melhora codificada por forma de onda codificando ruído a um valor total predeterminado de melhora de fala, T, está abaixo do limiar de mascaramento. Em outras palavras, a frequência máxima de corte, em qualquer momento é:

$$\max(fc \mid T*N(f<fc,t) < \Theta(f,t)) \qquad (8)$$

[0107] As modalidades descritas acima têm assumido que os meios disponíveis para manter melhora codificada por forma de onda codificando artefatos de se tornar audível é para ajustar a proporção de mistura (melhora codificada por forma de onda para codificada paramétrica) ou para reduzir a quantidade total de melhora. Uma alternativa é a de controlar a quantidade de ruído de codificação de

melhora codificada por forma de onda através de uma alocação variável de taxa de bits para gerar a cópia reduzida de qualidade da fala. Em um exemplo desta modalidade alternativa, uma quantidade base constante de melhora codificada paramétrica é aplicada, e a melhora codificada por forma de onda adicional é aplicada para alcançar a quantidade desejada (predeterminada) de melhora total. A cópia da fala de qualidade reduzida é codificada com uma taxa de bits variável, e essa taxa de bits é selecionada como a mais baixa taxa de bits que mantém ruído de codificação de melhora codificada por forma de onda abaixo do limiar mascarado de áudio principal melhorado codificado paramétrico.

[0108] Em algumas modalidades, o programa de áudio cujo conteúdo fala deve ser aumentado em conformidade com a invenção inclui canais de alto-falantes, mas não qualquer canal objeto. Em outras modalidades, o programa de áudio de fala cujo conteúdo deve ser melhorado de acordo com a invenção é um programa de áudio com base em objeto (tipicamente um programa de áudio com base em objeto multicanal) que compreende pelo menos um canal objeto e, opcionalmente, também, pelo menos, um canal de alto-falante.

[0109] Outros aspectos da invenção incluem um codificador configurado para realizar qualquer modalidade do método de codificação da invenção para gerar um sinal de áudio codificado em resposta a um sinal de entrada de áudio (por exemplo, em resposta a dados de áudio indicativos de um sinal de entrada de áudio de multicanais), um decodificador configurado para decodificar um tal sinal codificado e executar a melhora da fala sobre o conteúdo de áudio descodificado, e um sistema que inclui um dito codificador e dito decodificador. O sistema da FIGURA 3 é um exemplo de um dito sistema.

[0110] O sistema da FIGURA 3 inclui o codificador 20, o qual está configurado (por exemplo, programado) para executar uma modalidade do método de codificação da invenção para gerar um sinal de áudio codificado em resposta a dados de áudio

indicativos de um programa de áudio. Normalmente, o programa é um programa de áudio multicanal. Em algumas modalidades, o programa de áudio multicanal compreende apenas os canais de alto-falante. Em outras modalidades, o programa de áudio de múltiplos canais é um programa de áudio baseado no objeto que compreende pelo menos um canal objeto e, opcionalmente, também, pelo menos, um canal de alto-falante.

[0111] Os dados de áudio incluem dados (identificados como dados de "áudio misturados" na FIGURA 3) indicativos do conteúdo de áudio misturado (uma mistura de conteúdo de fala e não fala) e de dados (identificados como dados de "fala" na FIGURA 3) indicativos do conteúdo da fala do conteúdo de áudio misturado.

[0112] Os dados de fala passam por uma transformação de domínio do tempo para domínio de frequência (QMF) no estágio 21, e os componentes do QMF resultantes são afirmados um elemento de geração de parâmetro de melhora 23. Os dados de áudio misturados passam por uma transformação de domínio de tempo para domínio de frequência (QMF) no estágio 22, e os componentes do QMF resultantes são afirmados para o elemento 23 e para o subsistema de codificação 27.

[0113] Os dados de fala são também afirmados para o subsistema 25 que está configurado para gerar dados de forma de onda (por vezes aqui referidos como uma "qualidade reduzida" ou "baixa qualidade" da cópia da fala) indicativos de uma cópia de baixa qualidade dos dados de fala, para usar em melhora de fala codificada por forma de onda do conteúdo misturado (fala e não fala) determinado pelos dados de áudio misturados. A cópia de fala de baixa qualidade compreende menos bits do que os dados de fala original, é de qualidade desagradável quando processada e percebida de forma isolada, e quando processada é indicativo de fala tendo uma forma de onda semelhante (por exemplo, pelo menos, substancialmente semelhante) para a forma de onda da fala indicada pelos dados de fala originais. Métodos de execução de subsistema 25 são conhecidos na técnica. Exemplos são de codificadores de fala

de predição linear excitados por código (CELP) como AMR e G729.1 ou codificadores de mistura modernos como MPEG Unified Speech and Audio Coding (USAC), normalmente operados em uma taxa de bits baixa (por exemplo, 20 kbps). Alternativamente, podem ser utilizados codificadores de domínio da frequência, os exemplos incluem Siren (G722.1), MPEG 2 Layer II/III, MPEG AAC.

[0114] A melhora de fala híbrida realizada (por exemplo, no subsistema 43 de decodificador 40) de acordo com modalidades típicas da invenção inclui uma etapa de realização (nos dados da forma de onda) o inverso da codificação realizada (por exemplo, no subsistema 25 de codificador 20) para gerar os dados de forma de onda, para recuperar uma cópia de baixa qualidade do conteúdo da fala do sinal de áudio misturado a ser melhorado. A cópia de baixa qualidade recuperada da fala é então utilizada (com os dados paramétricos, e dados indicativos do sinal de áudio misturado) para realizar as etapas restantes da melhora da fala.

[0115] Elemento 23 é configurado para gerar dados paramétricos em resposta à saída de dados a partir de estágios 21 e 22. Os dados paramétricos, com os dados de áudio misturados originais, determinam fala parametricamente construída, que é uma versão reconstruída parametricamente da fala indicada pelos dados de fala originais (isto é, o conteúdo da fala dos dados de áudio misturados). A versão parametricamente reconstruída da fala corresponde a pelo menos substancialmente (por exemplo, é uma boa aproximação de) à fala indicada pelos dados de fala originais. Os dados paramétricos determinam um conjunto de parâmetros de melhora codificada paramétrica, $p(t)$, para a realização de melhora da fala codificada paramétrica em cada segmento do conteúdo misturado não melhorado determinado pelos dados de áudio misturados.

[0116] Elemento de geração de indicador de mistura 29 é configurado para gerar um indicador de mistura ("BI"), em resposta à saída de dados a partir de estágios 21 e 22. Considera-se que o programa de áudio indicado pela saída de fluxo de bits

do codificador 20 passará por melhora híbrida de fala (por exemplo, decodificador 40) para determinar um programa de áudio de fala melhorada, incluindo através da combinação dos dados de áudio não melhorados do programa original com uma combinação de dados de fala de baixa qualidade (determinados a partir dos dados da forma de onda), e os dados paramétricos. O indicador de mistura determina tal combinação (por exemplo, a combinação tem uma sequência de estados determinada por uma sequência de valores atuais do indicador de mistura), de modo que o programa de áudio de fala melhorada tem artefatos de codificação de melhora de fala menos audível (por exemplo, os artefatos de codificação de melhora da fala que são mais bem mascarados) seria ou um programa de áudio de fala melhorada codificado por forma de onda puramente determinado pela combinação apenas de dados de fala de baixa qualidade com dados de áudio não melhorados ou um programa de áudio de fala melhorada codificado paramétrico puramente determinado pela combinação apenas da fala parametricamente construída com os dados de áudio não melhorados.

[0117] Em variações na modalidade da FIGURA 3, o indicador de mistura utilizado para a melhora de fala híbrida da invenção não é gerado no codificador da invenção (e não está incluído na saída do fluxo de bits do codificador), mas em vez é gerado (por exemplo, em uma variação no receptor 40), em resposta à saída de fluxo de bits a partir do codificador (cujo fluxo de bits inclui dados em forma de onda e os dados paramétricos).

[0118] Deve ser entendido que a expressão "indicador de mistura" não se destina a indicar um único parâmetro ou valor (ou uma sequência de parâmetros individuais ou valores) para cada segmento do fluxo de bits. Em vez disso, contempla-se que em algumas modalidades, um indicador de mistura (para um segmento do fluxo de bits) pode ser um conjunto de dois ou mais parâmetros ou valores (por exemplo, para cada segmento, um parâmetro de controle de melhora codificada paramétrica, e um parâmetro de controle de melhora codificada por forma de onda).

[0119] Subsistema de codificação 27 gera dados de áudio codificados indicativos do conteúdo de áudio dos dados de áudio misturados (normalmente, uma versão comprimida dos dados de áudio misturados). O subsistema de codificação 27 tipicamente implementa um inverso da transformação executada no estágio 22, bem como outras operações de codificação.

[0120] O estágio de formatação 28 é configurado para montar a saída de dados paramétrico do elemento 23, a saída de dados da forma de onda do elemento 25, o indicador de mistura gerado no elemento 29, e a saída de dados do áudio codificado a partir do subsistema 27 para um fluxo de bits codificado indicativo do programa de áudio. O fluxo de bits (os quais podem ter formato de E-AC-3 ou AC-3, em algumas implementações) inclui os dados paramétricos não codificados, dados de forma de onda, e indicador de mistura.

[0121] O fluxo de bits de áudio codificado (um sinal de áudio codificado) emitido do codificador 20 é fornecido ao subsistema de entrega 30. Subsistema de entrega 30 está configurado para armazenar o sinal de áudio codificado (por exemplo, para armazenar dados indicativos do sinal de áudio codificado) gerado pelo codificador 20 e/ou para transmitir o sinal de áudio codificado.

[0122] O decodificador 40 é acoplado e configurado (por exemplo, programado) para receber o sinal de áudio codificado a partir do subsistema 30 (por exemplo, pela leitura ou recuperação de dados indicativos do sinal de áudio codificado a partir do armazenamento no subsistema 30 ou receber o sinal de áudio codificado que foi transmitido pelo subsistema 30), e para decodificar os dados indicativos do conteúdo áudio misturado (fala e não fala) do sinal de áudio codificado, e para executar a melhora de fala híbrida no conteúdo de áudio misturado descodificado. Decodificador 40 é tipicamente configurado para gerar e emitir (por exemplo, para um sistema de processamento, não mostrado na FIGURA 3) uma melhora de fala, sinal de áudio descodificados indicativo de uma versão melhorada da fala da entrada de

conteúdo de áudio misturado ao codificador 20. Alternativamente, este inclui um tal sistema de processamento que está acoplado para receber a emissão do subsistema 43.

[0123] Buffer de 44 (uma memória buffer) do decodificador 40 armazena (por exemplo, de forma não transitória), pelo menos um segmento (por exemplo, quadros) do sinal de áudio codificado (fluxo de bits) recebido pelo decodificador 40. Em típica operação, uma sequência de segmentos do fluxo de bits de áudio codificado é fornecida ao buffer 44 e considerada a partir do buffer 44 para o estágio de desformatação 41.

[0124] O estágio de desformatação (análise) 41 do decodificador 40 é configurado para analisar o fluxo de bits codificado do subsistema de entrega 30, para extrair do mesmo os dados paramétricos (gerados pelo elemento 23 do codificador 20), os dados em forma de onda (gerados pelo elemento 25 do codificador 20), o indicador de mistura (gerado no elemento 29 do codificador 20), e os dados de áudio misturados codificados (fala e não fala) (gerados no subsistema de codificação 27 do codificador 20).

[0125] Os dados de áudio misturados codificados são descodificados no subsistema de descodificação 42 do decodificador 40, e os dados de áudio resultantes decodificados misturados (fala e não fala) são afirmados a partir do subsistema de melhora de fala híbrido 43 (e são emitidos opcionalmente a partir do decodificador 40 sem sofrer melhora da fala).

[0126] Em resposta aos dados de controle (incluindo o indicador de mistura) extraídos através do estágio 41 a partir do fluxo de bits (ou gerados no estágio 41 em resposta aos metadados incluídos no fluxo de bits), e em resposta aos dados paramétricos e os dados de forma de onda extraídos pelo estágio 41, subsistema de melhora de fala 43 realiza melhora de fala híbrida nos dados de áudio misturados descodificados (fala e não fala) a partir do subsistema de descodificação 42, em

conformidade com uma modalidade da invenção. A emissão de sinal de áudio fala melhorada de subsistema 43 é indicativo de uma versão melhorada da fala da entrada de conteúdo de áudio misturado ao codificador 20.

[0127] Em várias implementações do codificador 20 da FIGURA 3, o subsistema 23 pode gerar qualquer um dos exemplos descritos de parâmetros de predição, p_i, para cada ladrilho de cada canal do sinal de entrada de áudio misturado, para ser utilizado (por exemplo, decodificador 40) para reconstrução do componente de fala de um sinal de áudio misturado descodificado.

[0128] Com um sinal de fala indicativo do conteúdo da fala do sinal de áudio misturado descodificado (por exemplo, a cópia de baixa qualidade da fala gerada pelo subsistema 25 de codificador 20, ou uma reconstrução do conteúdo de fala gerado usando parâmetros de predição, p_i, gerado pelo subsistema 23 de codificador 20), a melhora da fala pode ser realizada (por exemplo, no subsistema 43 de decodificador 40 da FIGURA 3) por mistura do sinal de fala com o sinal de áudio misturado descodificado. Ao aplicar um ganho para a fala a ser adicionado (misturados), é possível controlar a quantidade de melhora de fala. Para um aumento de 6 dB, a fala pode ser adicionada com um ganho de 0 dB (desde que a fala na mistura de fala melhorada tenha o mesmo nível que o sinal de fala transmitido ou reconstruído). O sinal de fala melhorada é:

$$M_e = M + g \cdot D_r \qquad (9)$$

[0129] Em algumas modalidades, para atingir um ganho de melhora de fala, *G,* o seguinte ganho de mistura é aplicado:

$$g = 10^{G/20} - 1 \qquad (10)$$

[0130] No caso de reconstrução de fala independente de canal, a mistura melhorada de fala, M_e, é obtida na forma de:

$$M_e = M \cdot (1 + diag(P) \cdot g) \qquad (11)$$

[0131] No exemplo acima descrito, a contribuição da fala em cada canal do

sinal de áudio misturado é reconstruída com a mesma energia. Quando a fala foi transmitida como um sinal lateral (por exemplo, como uma cópia de baixa qualidade do conteúdo da fala de um sinal de áudio misturado), ou quando a fala é reconstruída utilizando múltiplos canais (como com um preditor MMSE), a mistura de melhora da fala requer informação processamento de modo a misturar a fala com a mesma distribuição nos diferentes canais, enquanto o componente de fala já presente no sinal de áudio misturado é melhorado.

[0132] Esta informação de processamento pode ser fornecida por um parâmetro de processamento r_i para cada canal, que pode ser representada como um vetor de processamento R tem forma

$$R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \quad (12)$$

quando existem três canais. A mistura de melhora da fala é:

$$M_e = M + R \cdot g \cdot D_r \quad (13)$$

[0133] No caso em que existem múltiplos canais, e a fala (para ser misturada com cada canal de um sinal de áudio misturado) é reconstruída utilizando p parâmetros de predição p_i, a equação anterior pode ser escrita como:

$$M_e = M + R \cdot g \cdot P \cdot M = (I + R \cdot g \cdot P) \cdot M \quad (14)$$

[0134] Onde I é a matriz de identidade.

5. PROCESSAMENTO DA FALA

[0135] A FIGURA 4 é um diagrama de blocos de um sistema de processamento de fala que implementa mistura de melhora de mistura fala convencional de forma:

$$M_e = M + R \cdot g \cdot D_r \quad (15)$$

[0136] Na FIGURA 4, o sinal de áudio misturado em três canais a ser melhorado está no (ou se transforma em) domínio da frequência. Os componentes de frequência fo canal esquerdo são afirmados como uma entrada do elemento de mistura 52, os componentes de frequência do canal central são afirmados a uma

entrada do elemento de mistura 53, e os componentes de frequência do canal direito são afirmados a uma entrada do elemento de mistura 54.

[0137] O sinal de fala a ser misturado com o sinal de áudio misturado (para aumentar o último sinal) pode ter sido transmitido como um sinal lateral (por exemplo, como uma cópia de baixa qualidade do conteúdo da fala do sinal de áudio misturado), ou pode ter sido reconstruído a partir de parâmetros de predição, p_i, transmitido com o sinal de áudio misturado. O sinal de fala é indicado pelos dados de domínio de frequência (por exemplo, estes compreendem componentes de frequência gerados pela transformação de um sinal de domínio de tempo para o domínio da frequência), e estes componentes de frequência são afirmados a uma entrada do elemento de mistura 51, em que são multiplicados pelo parâmetro de ganho, g.

[0138] A saída do elemento 51 é afirmada ao subsistema de processamento 50. Também afirmados à entrega do subsistema de processamento 50 são parâmetros CLD (diferença de nível de canal), CLD_1 e CLD_2, que foram transmitidos com o sinal de áudio misturado. Os parâmetros CLD (para cada segmento do sinal de áudio misturado) descreve a forma como o sinal de fala é misturada com os canais do referido segmento do conteúdo do sinal de áudio misturado. CLD_1 indica um coeficiente de panning para um par de canais de alto-falantes (por exemplo, que define panning da fala entre os canais esquerdo e central), e CLD_2 indica um coeficiente de panning para o outro par dos canais de alto-falantes (por exemplo, que define panning da fala entre canais central e direito). Assim, subsistema de processamento 50 afirma (para elemento 52) dados indicativos de $R \cdot g \cdot D_r$ para o canal esquerdo (o conteúdo da fala, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal esquerdo), e estes dados são somados com o canal esquerdo do sinal de áudio misturado no elemento 52. O subsistema de processamento 50 afirma (para o elemento 53) dados indicativos de $R \cdot g \cdot D_r$ para o canal central (o conteúdo da fala, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal

central), e estes dados são somados com o canal central do sinal de áudio misturado no elemento 53. O subsistema de processamento 50 afirma (para o elemento 54) dados indicativos de $R \cdot g \cdot D_r$ para o canal direito (o conteúdo da fala, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal direito) e esses dados são somados com o canal direito do sinal de áudio misturado no elemento 54.

[0139] As saídas dos elementos 52, 53 e 54 são utilizadas, respectivamente, para dirigir alto-falante da esquerda L, alto-falante central C, e alto-falante direito "Right".

[0140] A FIGURA 5 é um diagrama de blocos de um sistema de processamento de fala que implementa convencional mistura de melhora de fala da forma:

$$M_e = M + R \cdot g \cdot P \cdot M = (I + R \cdot g \cdot P) \cdot M \qquad (16)$$

[0141] Na FIGURA 5, o sinal de áudio misturado em três canais a ser melhorado está no (ou se transforma em) domínio da frequência. Os componentes de frequência do canal esquerdo são afirmados a uma entrada do elemento de mistura 52, os componentes de frequência do canal central são afirmados a uma entrada do elemento de mistura 53, e os componentes de frequência do canal direito são afirmados a uma entrada do elemento de mistura 54.

[0142] O sinal de fala a ser misturado com o sinal de áudio misturado é reconstruído (como indicado) a partir de parâmetros de predição, p_i, transmitidos com o sinal de áudio misturado. Parâmetro de predição p_1 é empregado para reconstruir a fala a partir do primeiro canal (à esquerda) do sinal de áudio misturado, o parâmetro de predição p_2 é empregado para reconstruir a fala a partir do segundo canal (central) do sinal de áudio misturado, e parâmetro de predição p_3 é empregado para reconstruir a fala a partir do terceiro (direita) canal do sinal de áudio misturado. O sinal de fala é indicado pelos dados de domínio de frequência, e estes componentes de frequência são afirmados a uma entrada do elemento de mistura 51, no qual eles são

multiplicados pelo parâmetro de ganho, *g*.

[0143] A saída do elemento 51 é afirmada ao subsistema de processamento 55. Também afirmou ao subsistema de processamento estão parâmetros CLD (diferença de nível de canal), CLD_1 e CLD_2, que foram transmitidos com o sinal de áudio misturado. Os parâmetros CLD (para cada segmento do sinal de áudio misturado) descrevem a forma como o sinal de fala é misturado aos canais do referido segmento do conteúdo do sinal de áudio misturado. CLD_1 indica um coeficiente de panning para um par de canais de alto-falantes (por exemplo, que define panning da fala entre os canais esquerdo e central), e CLD_2 indica um coeficiente de panning para o outro par dos canais de alto-falantes (por exemplo, que define panning da fala entre os canais central e da direita). Assim, subsistema de processamento 55 afirma (para elemento 52) dados indicativos de $R \cdot g \cdot P \cdot M$ para o canal esquerdo (o conteúdo de fala reconstruído misturado com o canal esquerdo do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal esquerdo, misturado com o canal esquerdo do conteúdo de áudio misturado) e esses dados são somados com o canal esquerdo do sinal de áudio misturado no elemento 52. O subsistema de processamento 55 afirma (para o elemento 53) dados dos elementos indicativos de $R \cdot g \cdot P \cdot M$ para o canal central (o conteúdo de fala reconstruída misturado com o canal central do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal central), e esses dados são somados com o canal central do sinal de áudio misturado no elemento 53. O subsistema de processamento 55 afirma (para o elemento 54) dados indicativos de $R \cdot g \cdot P \cdot M$ para o canal direito (o conteúdo da fala reconstruído misturado com o canal direito do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal direito) e esses dados são somados com o canal direito do sinal de áudio misturado no elemento 54.

[0144] As saídas dos elementos 52, 53 e 54 são utilizadas, respectivamente,

para direcionar o alto-falante L, alto-falante central C, e alto-falante direito "Right".

[0145] Os parâmetros CLD (diferença de nível de canal) são convencionalmente transmitidos com sinais de canal de alto-falante (por exemplo, para determinar relações entre os níveis em que diferentes canais devem ser prestados). Estes são utilizados de uma nova maneira em algumas modalidades da invenção (por exemplo, para fala melhorada pan, entre os canais de alto-falante de um programa de áudio de fala melhorada).

[0146] Em modalidades típicas, os parâmetros de processamento r_i são (ou são indicativos de) coeficientes mistura ascendente da fala, que descrevem a forma como o sinal de fala é misturado com os canais do sinal de áudio misturado a ser melhorado. Esses coeficientes podem ser eficientemente transmitidos para o melhorador de fala utilizando parâmetros de diferença de nível do canal (CLD). Um CLD indica coeficientes de panning para dois alto-falantes. Por exemplo,

$$\beta_1 = \sqrt{\frac{1}{1 + 10^{\frac{CLD}{10}}}} \qquad (17)$$

$$\beta_2 = \sqrt{\frac{10^{\frac{CLD}{10}}}{1 + 10^{\frac{CLD}{10}}}} \qquad (18)$$

onde β_1 indica ganho para a alimentação de alto-falante para o primeiro alto-falante e β_2 indica ganho para a alimentação de alto-falante para o segundo alto-falante em um instante durante o pan. Com CLD = 0, o panning está totalmente no primeiro alto-falante, enquanto que com CLD se aproximando do infinito, o panning está totalmente no segundo alto-falante. Com CLDs definidos no domínio de dB, um número limitado de níveis de quantização pode ser suficiente para descrever o panning.

[0147] Com dois CLDs, panning ao longo de três alto-falantes pode ser definidos. Os CLDs podem ser derivados da seguinte forma a partir dos coeficientes de processamento:

$$CLD_1 = 10 \cdot \log_{10}\left(\frac{\bar{r}_2^2}{\bar{r}_1^2}\right) \qquad (19)$$

$$CLD_2 = 10 \cdot \log_{10}\left(\frac{\overline{r}_3^2}{\overline{r}_1^2+\overline{r}_2^2}\right) \qquad (20)$$

em que $\overline{r}_x^2 = \frac{r_x^2}{\sum_i r_i^2}$ são os coeficientes de processamento de tal modo que

$$\overline{r}_1^2 + \overline{r}_2^2 + \overline{r}_3^2 = 1 \qquad (21)$$

[0148] Os coeficientes de processamento podem então ser reconstruídos a partir de CLDs por:

$$R = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{1}{\left(1+10^{\frac{CLD_1}{10}}\right)\left(1+10^{\frac{CLD_2}{10}}\right)}} \\ \sqrt{\frac{10^{\frac{CLD_1}{10}}}{\left(1+10^{\frac{CLD_1}{10}}\right)\cdot\left(1+10^{\frac{CLD_2}{10}}\right)}} \\ \sqrt{\frac{10^{\frac{CLD_2}{10}}}{1+10^{\frac{CLD_2}{10}}}} \end{pmatrix} \qquad (22)$$

[0149] Como observado em outros lugares aqui, a melhora da fala codificada por forma de onda utiliza uma cópia de baixa qualidade do conteúdo da fala do sinal de conteúdo misturado a ser melhorada. A cópia de baixa qualidade é geralmente codificada a uma taxa de bits baixa e transmitida como um sinal lateral, com o sinal de conteúdo misturado, e, portanto, a cópia de baixa qualidade, tipicamente, contém artefatos de codificação significativos. Assim, a melhora da fala codificada por forma de onda proporciona um bom desempenha de melhora da fala em situações com uma SNR baixa (isto é, baixa proporção entre a fala e todos os outros sons indicados pelo sinal de conteúdo misto), e proporciona, tipicamente, um fraco desempenho (isto é, resulta em artefatos de codificação indesejáveis audíveis) em situações com alta SNR.

[0150] Por outro lado, quando o conteúdo da fala (de um sinal de conteúdo misturado é melhorado) é escolhido fora (por exemplo, é fornecido como o único conteúdo de um canal central de um sinal de conteúdo misturado multicanal) ou o sinal de conteúdo misturado, de outro modo, SNR alta, melhora da fala codificada paramétrica fornece um bom desempenho de melhora de fala.

[0151] Assim, a melhora da fala codificada por forma de onda e melhora da fala codificada paramétrica têm desempenho complementar. Com base nas propriedades do sinal cujo conteúdo de fala de ser melhorado, uma classe de

modalidades da presente invenção combina os dois métodos para aproveitar os seus desempenhos.

[0152] A FIGURA 6 é um diagrama de blocos de um sistema de processamento de fala nesta classe de modalidades que está configurado para executar a melhora de fala híbrida. Numa implementação, o subsistema 43 de decodificador 40 da FIGURA 3 incorpora o sistema da FIGURA 6 (exceto para os três alto-falantes mostradas na FIGURA 6). A melhora de fala híbrida (mistura) pode ser descrita por

$$M_e = R \cdot g_1 \cdot D_r + (I + R \cdot g_2 \cdot P) \cdot M \qquad (23)$$

onde $R \cdot g_1 \cdot D_r$ é a melhora da fala codificada por forma de onda do tipo implementado pelo sistema convencional da FIGURA 4, $R \cdot g_2 \cdot P$ é melhora de fala codificada paramétrica do tipo implementado pelo sistema convencional da FIGURA 5, e os parâmetros de g_1 e g_2 controlam o ganho de melhora global e o trade-off entre os dois métodos de melhora de fala. Um exemplo de uma definição dos parâmetros g_1 e g_2 é:

$$g_1 = \alpha_c \cdot (10^{G/20} - 1) \qquad (24)$$

$$g_2 = (1 - \alpha_c) \cdot (10^{G/20} - 1) \qquad (25)$$

onde o parâmetro de α_c define o trade-off entre a melhora de fala codificada paramétrica e métodos de melhora de fala codificada paramétrica. Com um valor de $\alpha_c = 1$, apenas a cópia da fala de baixa qualidade é usada para a melhora da fala codificada por forma de onda. O modo de melhora codificada paramétrica está contribuindo plenamente para a melhora quando $\alpha_c = 0$. Valores de α_c entre 0 e 1 misturam os dois métodos. Em algumas implementações, α_c é um parâmetro de banda larga (aplicável a todas as bandas de frequência dos dados de áudio). Os mesmos princípios podem ser aplicados dentro das faixas de frequências individuais, de tal modo que a mistura é otimizada de um modo dependente de frequência utilizando um valor diferente do parâmetro α_c para cada banda de frequência.

[0153] Na FIGURA 6, o sinal de áudio misturado em três canais a ser melhorado está no (ou se transforma em) domínio da frequência. Os componentes de frequência do canal esquerdo são afirmados a uma entrada do elemento de mistura 65, os componentes de frequência do canal central são afirmados a uma entrada do elemento de mistura 66, e os componentes de frequência do canal direito são afirmados a uma entrada do elemento de mistura 67.

[0154] O sinal de fala a ser misturado com o sinal de áudio misturado (para aumentar o último sinal) inclui uma cópia de baixa qualidade (identificada como "Fala" na FIGURA 6) do conteúdo da fala do sinal de áudio misturado que tenha sido gerado a partir de dados em forma de onda transmitidos (em conformidade com a melhora de fala codificada por forma de onda) com o sinal de áudio misturado (por exemplo, como um sinal lateral), e um sinal de fala reconstruído (emitido a partir do elemento de reconstrução de fala codificada paramétrica 68 da FIGURA 6), que é reconstruído a partir do sinal de áudio misturado e parâmetros de predição, p_i, transmitidos (em conformidade com a melhora de fala codificada paramétrica) com o sinal de áudio misturado. O sinal de fala é indicado pelos dados de domínio de frequência (por exemplo, este compreende componentes de frequência gerados pela transformação de um sinal de domínio de tempo para o domínio da frequência). Os componentes de frequência da cópia de baixa qualidade de fala são afirmados a uma entrada do elemento de mistura 61, no qual eles são multiplicados pelo parâmetro de ganho, g_2. Os componentes de frequência do sinal de fala parametricamente reconstruído são afirmados a partir da saída do elemento 68 a uma entrada do elemento 62, no qual eles são multiplicados pelo parâmetro de ganho, g_1. Em modalidades alternativas, a mistura realizada para aplicar a melhora de fala é realizada no domínio do tempo, em vez de no domínio da frequência como na modalidade da FIGURA 6.

[0155] A saída dos elementos 61 e 62 é somada pelo elemento de soma 63 para gerar o sinal de fala a ser misturado com o sinal de áudio misturado, e este sinal

de fala é afirmado partir da saída do elemento 63 para subsistema de processamento 64. Também afirmado para subsistema de processamento 64 são parâmetros CLD (diferença de nível de canal), CLD_1 e CLD_2, que foram transmitidos com o sinal de áudio misturado. Os parâmetros CLD (para cada segmento do sinal de áudio misturado) descrevem a forma como o sinal de fala é misturado aos canais do referido segmento do conteúdo do sinal de áudio misturado. CLD_1 indica um coeficiente de panning para um par de canais de alto-falantes (por exemplo, que define panning da fala entre os canais esquerdo e central), e CLD_2 indica um coeficiente de panning para o outro par dos canais de alto-falantes (por exemplo, que define panning da fala entre os canais centro direito e central). Assim, subsistema de processamento 64 afirma (para elemento 52) dados indicativos de $R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$ para o canal esquerdo (o conteúdo de fala reconstruído misturado com o canal esquerdo do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal esquerdo, misturado com o canal esquerdo do conteúdo de áudio misturado) e esses dados são somados com o canal esquerdo do sinal de áudio misturado no elemento 52. O subsistema de processamento 64 afirma (para o elemento 53) dados indicativos de $R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$ para o canal central (o conteúdo de fala reconstruído misturado com o canal central do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal central), e estes dados são somados com o canal central do sinal de áudio misturado no elemento 53. O subsistema de processamento 64 afirma (para o elemento 54) dados indicativos de $R \cdot g_1 \cdot D_r + (R \cdot g_2 \cdot P) \cdot M$ para o canal direito (o conteúdo da fala reconstruído misturado com o canal direito do conteúdo de áudio misturado, escalonado pelo parâmetro de ganho e o parâmetro de processamento para o canal direito) e esses dados são somados com o canal direito do sinal de áudio misturado no elemento 54.

[0156] As saídas dos elementos 52, 53 e 54 são utilizadas, respectivamente,

para direcionar alto-falante L, alto-falante central C, e alto-falante direito "Direito".

[0157] O sistema da FIGURA 6 pode implementar comutação baseada em SNR temporal, quando o parâmetro de α_c é constrangido para ter o valor $\alpha_c = 0$ ou o valor de $\alpha_c = 1$. Uma tal aplicação é especialmente útil em situações de taxa de bits fortemente restringidas em que ou os dados da cópia de fala de baixa qualidade podem ser enviados ou os dados paramétricos podem ser enviados, mas não ambos. Por exemplo, em uma tal aplicação, a cópia de fala de baixa qualidade é transmitida com o sinal de áudio misturado (por exemplo, como um sinal lateral) apenas em segmentos para os quais $\alpha_c = 1$, e os parâmetros de predição, p_i, são transmitidos com o sinal de áudio misturado (por exemplo, como um sinal lateral) apenas em segmentos para os quais $\alpha_c = 0$.

[0158] O interruptor (implementado por elementos 61 e 62 da presente implementação da FIGURA 6) determina se melhora codificada por forma de onda ou melhora codificada paramétrica deve ser executada em cada segmento, com base na proporção (SNR) entre a fala e todos outros conteúdos de áudio no segmento (esta relação por sua vez, determina o valor de α_C). Essa implementação pode usar um valor limiar da SNR para decidir qual método escolher:

$$\alpha_c = \begin{cases} \mathbf{0} & if\ SNR > \tau \\ \mathbf{1} & if\ SNR \leq \tau \end{cases} \qquad (26)$$

onde τ é um valor de limiar (por exemplo, τ pode ser igual a 0).

[0159] Algumas implementações da FIGURA 6 empregam histerese para evitar a comutação de alternância rápida entre a melhora codificada por forma de onda e modos de melhora codificada paramétrica quando a SNR está em torno do valor de limiar para vários quadros.

[0160] O sistema da FIGURA 6 pode implementar mistura à base de SNR temporal, quando o parâmetro α_C é deixado ter qualquer valor real na faixa de 0 a 1, inclusive.

[0161] Uma implementação do sistema da FIGURA 6 utiliza dois valores alvos,

τ_1 e τ_2 (da SNR de um segmento do sinal de áudio misturado para ser melhorado) além do qual um método (ou melhora codificada por forma de onda ou melhora codificada paramétrica) é sempre considerado para fornecer o melhor desempenho. Entre esses alvos, a interpolação é utilizada para determinar o valor do parâmetro α_C para o segmento. Por exemplo, a interpolação linear pode ser empregada para determinar o valor do parâmetro de α_C para o segmento:

$$\alpha_c = \begin{cases} \mathbf{0} & if\ SNR > \tau_2 \\ 1 - \dfrac{SNR - \tau_1}{\tau_2 - \tau_1} & if\ \tau_1 < SNR \leq \tau_2 \\ \mathbf{1} & if\ SNR \leq \tau_1 \end{cases} \quad (27)$$

[0162] Em alternativa, outros esquemas de interpolação apropriados podem ser utilizados. Quando a SNR não estiver disponível, os parâmetros de predição em muitas implementações podem ser utilizados para proporcionar uma aproximação da SNR.

[0163] Numa outra classe de modalidades, a combinação de melhora codificada por forma de onda e codificada paramétrica para ser executada em cada segmento de um sinal de áudio é determinada por um modelo de mascaramento auditivo. Em modalidades típicas desta classe, a proporção de mistura ideal para uma mistura de melhora codificada por forma de onda e codificada paramétrica para ser executada em um segmento de um programa de áudio utiliza a maior quantidade de melhora codificada por forma de onda que apenas mantém o ruído de codificação de se tornar audível. Um exemplo de uma modalidade do método da invenção, que emprega um modelo de mascaramento auditivo é aqui descrito com referência à FIGURA 7.

[0164] De modo mais geral, as considerações seguintes referem-se às modalidades em que um modelo de mascaramento auditivo é usado para determinar uma combinação (por exemplo, mistura) de melhora codificada por forma de onda e codificada paramétrica para ser executada em cada segmento de um sinal de áudio. Em tais modalidades, os dados indicativos de uma mistura de fala e de áudio de fundo,

A(t), a ser referida como uma mistura de áudio não melhorada, são proporcionados e processados de acordo com o modelo de mascaramento auditivo (por exemplo, o modelo implementado pelo elemento 11 da FIGURA 7). O modelo prevê um limiar de mascaramento Θ(f,t) para cada segmento da mistura de áudio não melhorada. O limiar de mascaramento de cada ladrilho da da frequência de tempo da mistura de áudio não melhorado, com índice temporal *n* e de índice de banda de frequência *b*, pode ser denotado como $\Theta_{n,b}$.

[0165] O limiar de mascaramento $\Theta_{n,b}$ indica para o quadro *n* e banda *b* quanta distorção pode ser adicionada sem ser audível. Deixar $\varepsilon_{D,n,b}$ ser o erro de codificação (ou seja, o ruído de quantização) da cópia de fala de baixa qualidade (para ser utilizada para a melhora codificada por forma de onda), e $\varepsilon_{P,n,b}$ ser o erro de predição paramétrico.

[0166] Algumas modalidades desta classe implementam um interruptor físico ao método (melhora codificada por forma de onda ou codificada paramétrica) que é mais bem mascarado pelo conteúdo mistura de áudio não melhorado:

$$\alpha_c = \begin{cases} \mathbf{0} & if \sum_{n,b} \Theta_{n,b} - \varepsilon_{P,n,b} > \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \\ \mathbf{1} & if \sum_{n,b} \Theta_{n,b} - \varepsilon_{P,n,b} \leq \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \end{cases} \quad (28)$$

[0167] Em muitas situações práticas, o erro de predição exata paramétrica $\varepsilon_{P,n,b}$ pode não estar disponível no momento da geração dos parâmetros de melhora de fala, uma vez que estes podem ser gerados antes da mistura misturada não melhorado ser codificada. Os esquemas de codificação especialmente paramétricos podem ter um efeito significativo sobre o erro de uma reconstrução paramétrica da fala a partir dos canais de conteúdo misturado.

[0168] Assim, algumas modalidades alternativas misturadas na melhora da fala codificada paramétrica (com melhora codificada por forma de onda) quando os artefatos de codificação na cópia da fala de baixa qualidade (para ser utilizada para a melhora codificada por forma de onda) não são mascaradas pelo conteúdo misturado:

$$\alpha_c = \begin{cases} \mathbf{1} & if \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} \geq \mathbf{0} \\ 1 - \frac{\sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b}}{\tau_a} & if \ -\tau_a \leq \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} < \mathbf{0} \\ \mathbf{0} & if \sum_{n,b} \Theta_{n,b} - \varepsilon_{D,n,b} < -\tau_a \end{cases} \quad (29)$$

em que τ_a é um limiar de distorção além do qual apenas a melhora codificada paramétrica é aplicada. Esta solução é iniciada misturando a melhora codificada por forma de onda e codificada paramétrica quando a distorção global é maior do que o potencial geral de mascaramento. Na prática, isso significa que as distorções já eram audíveis. Por conseguinte, um segundo limiar pode ser utilizado com um valor maior do que 0. Em alternativa, pode-se utilizar as condições que, em vez de focarem sobre os ladrilhos de frequência de tempo não mascarados em vez do comportamento médio.

[0169] Da mesma forma, esta abordagem pode ser combinada com uma regra de mistura guiada por SNR quando as distorções (artefatos de codificação) na cópia da fala de baixa qualidade (para ser utilizada para a melhora codificada por forma de onda) são muito elevadas. Uma vantagem desta abordagem é que, nos casos de muito baixa SNR o modo de melhora codificada paramétrica não é utilizado uma vez que produz ruído mais audível do que as distorções da cópia de fala de baixa qualidade.

[0170] Em outra modalidade, o tipo de melhora de fala realizada para alguns ladrilhos de frequência no tempo se desvia daquela determinada pelos esquemas exemplares descritos acima (ou esquemas semelhantes), quando um buraco espectral é detectado em cada um desses ladrilho de frequência de tempo. Buracos espectrais podem ser detectados, por exemplo, através da avaliação da energia no ladrilho correspondente na reconstrução paramétrica enquanto que a energia é 0 na cópia da fala de baixa qualidade (para ser utilizada para a melhora codificada por forma de onda). Se esta energia excede um limiar, isto pode ser considerado como

áudio relevante. Nestes casos, o parâmetro α_C para o ladrilho pode ser ajustado para 0 (ou, dependendo da SNR o parâmetro α_C para o ladrilho pode ser direcionado para 0).

[0171] Em algumas modalidades, o codificador da invenção é operável em qualquer um selecionado dos seguintes modos:

[0172] 1. Canal paramétrico independente - Neste modo, um conjunto de parâmetros é transmitido para cada canal que contém fala. Usando estes parâmetros, um decodificador que recebe o programa de áudio codificado pode executar a melhora de fala codificada paramétrica sobre o programa para impulsionar a fala nestes canais por uma quantidade arbitrária. Um exemplo de taxa de bits para a transmissão do conjunto de parâmetros é de 0,75 - 2,25 kbps.

[0173] 2. Predição de fala multicanal - Neste modo múltiplos canais do conteúdo misturado são combinados numa combinação linear para prever o sinal de fala. Um conjunto de parâmetros é transmitido para cada canal. Usando estes parâmetros, um decodificador que recebe o programa de áudio codificado pode executar a melhora de fala codificada paramétrica sobre o programa. Os dados de posicionamento adicionais são transmitidos com o programa de áudio codificado para permitir o processamento da fala impulsionada de volta para a mistura. Um exemplo de taxa de bits para a transmissão do conjunto de parâmetros e os dados de posição são de 1,5 - 6,75 kbps por diálogo.

[0174] 3. Fala codificada por forma de onda - Neste modo, uma cópia de baixa qualidade do conteúdo da fala do programa de áudio é transmitida em separado, por qualquer meio adequado, em paralelo com o conteúdo de áudio normal (por exemplo, como um subfluxo separado). Um decodificador que recebe o programa de áudio codificado pode executar a melhora da fala codificada por forma de onda sobre o programa através de mistura na cópia de baixa qualidade em separado do conteúdo da fala com a mistura principal. Misturar a cópia de baixa qualidade da fala com um

ganho de 0 dB normalmente irá impulsionar a fala em 6 dB, como a amplitude é dobrada. Para este modo os dados de posicionamento também são transmitidos de modo que o sinal de fala é distribuído corretamente sobre os canais relevantes. Um exemplo de taxa de bits para a transmissão da cópia de baixa qualidade da fala e dados de posição é mais do que 20 kbps por diálogo.

[0175] 4. Híbrido paramétrico de forma de onda - Neste modo, uma cópia de baixa qualidade do conteúdo da fala do programa áudio (para uso na realização de melhora de fala codificada por forma de onda no programa), e um conjunto de parâmetros para cada canal contendo fala (para uso na realização de melhora de fala codificada paramétrica sobre o programa) são transmitidos em paralelo com a mistura não melhorada (fala e não fala) do conteúdo de áudio do programa. Quando a taxa de bits para a cópia de baixa qualidade da fala é reduzida, mais artefatos de codificação se tornam audíveis nsinal e a largura de banda necessária para a transmissão é reduzida. Também é transmitido um indicador mistura que determina uma combinação de melhora da fala codificada por forma de onda e melhora da fala codificada paramétrica a serem executadas em cada segmento do programa usando a cópia da fala de baixa qualidade e o conjunto de parâmetros. Em um receptor, melhora da fala híbrida é realizada no programa, incluindo através da realização de uma combinação de melhora de fala codificada por forma de onda e melhora de fala codificada paramétrica determinado pelo indicador de mistura, gerando assim dados indicativos de um programa de áudio de fala melhorada. Mais uma vez, os dados de posição também são transmitidos com o conteúdo de áudio misturado não melhorado do programa para indicar onde processar o sinal de fala. Uma vantagem desta abordagem é que a complexidade do receptor/decodificador necessária pode ser reduzida se o receptor/decodificador descarta a cópia de baixa qualidade da fala e aplica-se apenas o conjunto de parâmetros para executar a melhora codificada paramétrica. Um exemplo de taxa de bits para a transmissão da cópia de baixa

qualidade da fala, conjunto de parâmetros, indicador de mistura, e os dados de posição é de 8 - 24 kbps por diálogo.

[0176] Por motivos práticos, o ganho de melhora de fala pode ser limitado ao intervalo de 0 - 12 dB. Um codificador pode ser implementado para ser capaz de reduzir ainda mais o limite superior deste intervalo ainda mais por meio de um campo de fluxo de bits. Em algumas modalidades, a sintaxe do programa codificado (saída do codificador) iria suportar múltiplos diálogos melhoráveis simultâneos (em adição ao conteúdo não fala do programa), de tal modo que cada diálogo pode ser reconstruído e processado separadamente. Nestas modalidades, nos últimos modos, melhoras de fala para diálogos simultâneos (a partir de várias fontes em diferentes posições espaciais) seriam processados em uma única posição.

[0177] Em algumas modalidades em que o programa de áudio codificado é um programa de áudio baseado em objetos, um ou mais (do número total máximo de) aglomerados objetos podem ser selecionados para a melhora da fala. Os pares de valores CLD podem ser incluídos no programa codificado para utilização pela melhora de fala e sistema de processamento para pan a fala melhorada entre os aglomerados objetos. Do mesmo modo, em algumas modalidades em que o programa de áudio codificado inclui canais de alto-falantes em um formato convencional de 5,1, um ou mais dos canais de alto-falantes frontais podem ser selecionados para a melhora da fala.

[0178] Um outro aspecto da invenção é um método (por exemplo, um método realizado pelo decodificador 40 da FIGURA 3) para a descodificação e execução de melhora de fala híbrida em um sinal de áudio codificado que foi gerado de acordo com uma modalidade da codificação do método da invenção.

[0179] A invenção pode ser implementada em hardware, firmware ou software, ou uma combinação de ambos (por exemplo, como uma matriz lógica programável). A menos que especificado em contrário, os algoritmos ou processos incluídos como

parte da invenção, não são intrinsecamente relacionados com qualquer computador particular ou outro aparelho. Em particular, várias de máquinas de uso geral podem ser usadas com programas escritos de acordo com os ensinamentos aqui descritos, ou pode ser mais conveniente construir aparelhos mais especializados (por exemplo, circuitos integrados) para executar as etapas do método necessárias. Assim, a invenção pode ser implementada em um ou mais programas de computadores que executam em um ou mais sistemas programáveis de computador (por exemplo, um sistema de computador que implementa o codificador 20 da FIGURA 3, ou o codificador da FIGURA 7, ou decodificador 40 da FIGURA 3), cada um compreendendo, pelo menos, um processador, pelo menos um sistema de armazenamento de dados (incluindo a memória volátil e não volátil e/ou elementos de armazenamento), pelo menos um dispositivo de entrada ou de porta, e, pelo menos, um dispositivo de saída ou porta. O código de programa é aplicado para introduzir dados para executar as funções aqui descritas e gerar informação de saída. A informação de saída é aplicada a um ou mais dispositivos de saída, de uma forma conhecida.

[0180] Cada dito programa pode ser implementado em qualquer linguagem de computador desejada (incluindo máquina, montagem, ou elevado nível processual, lógico, orientado por objeto, ou linguagens de programação) para se comunicar com um sistema de computador. Em qualquer caso, a linguagem pode ser uma linguagem compilada ou interpretada.

[0181] Por exemplo, quando implementado por sequências de instruções de software de computador, as várias funções e etapas de modalidades da invenção podem ser implementadas por sequências de instruções de software de vários segmentos que funcionam no hardware de processamento de sinal digital adequado, caso em que os diferentes dispositivos, etapas e funções das modalidades podem corresponder a porções das instruções de software.

[0182] Cada um desses programa de computador é de preferência armazenado ou transferido para uma mídia de armazenamento ou dispositivo (por exemplo, a memória de estado sólido ou mídia ou mídia magnética ou óptica) pode ser lido por um computador programável de uso geral ou especial, para configurar e operar o computador quando o meio de armazenamento ou dispositivo é lido pelo sistema de computador para executar os procedimentos aqui descritos. O sistema da invenção pode também ser implementado como um meio de armazenamento legível por computador, configurado com (isto é, armazenamento) um programa de computador, em que o meio de armazenamento assim configurado faz com que um sistema de computador opere de uma maneira específica e predefinida para executar as funções aqui descritas.

[0183] Um número de modalidades da invenção foi descrito. No entanto, será entendido que várias modificações podem ser feitas sem se afastar do espírito e escopo da invenção. Numerosas modificações e variações da presente invenção são possíveis à luz dos ensinamentos anteriores. Deve ser entendido que dentro do escopo das concretizações, a invenção pode ser praticada de modo diferente da especificamente descrita aqui.

6. REPRESENTAÇÃO MÉDIA/LATERAL

[0184] As operações de melhora da fala, tal como aqui descritas podem ser realizadas por um decodificador de áudio baseado, pelo menos em parte, em dados de controle, parâmetros de controle, etc., na representação M/S. Os dados de controle, os parâmetros de controle, etc., na representação M/S podem ser gerados por um codificador de áudio à montante e extraídos pelo decodificador de áudio a partir de um sinal de áudio codificado gerado pelo codificador de áudio à montante.

[0185] Em um modo de melhora codificada paramétrica em que o conteúdo de fala (por exemplo, um ou mais diálogos, etc.) é predito a partir do conteúdo misturado, as operações de melhora de fala podem ser geralmente representadas

com uma única matriz, *H*, como mostrado na seguinte expressão:

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = H \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \qquad (30)$$

onde o lado esquerdo (LHS) representa um sinal de conteúdo misturado de fala melhorada gerado pelas operações de melhora de fala como representado pela matriz *H* operando em um sinal de conteúdo misturado original no lado direito (RHS).

[0186] Com o objetivo de ilustração, cada um do sinal de conteúdo misturado de fala melhorada (por exemplo, a LHS da expressão (30), etc.) e o sinal de conteúdo misturado original (por exemplo, o sinal de conteúdo misturado original é operado por *H* na expressão (30), etc.) compreende dois sinais de componente tenda conteúdo de fala melhorada e original misturado em dois canais, c_1 e c_2, respectivamente. Os dois canais c_1 e c_2 podem ser canais de áudio não M/S (por exemplo, canal esquerdo frontal, canal direito frontal, etc.) com base em uma representação não M/S. Deve notar-se que em várias modalidades, cada um de sinal de conteúdo misturado de fala melhorada e o sinal de conteúdo misturado original pode compreender ainda os sinais de componente tendo um conteúdo não fala em canais (por exemplo, canais surround, um canal de efeito de frequência de baixo, etc.) além dos dois canais não M/S c_1 e c_2. Deve ainda notar-se que em várias modalidades, cada sinal de conteúdo misturado de fala melhorada e o sinal de conteúdo misturado original podem possivelmente compreender sinais de componente tendo conteúdo da fala em um, dois, como ilustrado na expressão (30), ou mais do que dois canais. Conteúdo de fala, tal como aqui descrito pode compreender um, dois ou mais diálogos.

[0187] Em algumas modalidades, as operações de melhora de fala como representadas por *H* na expressão (30) podem ser usadas (por exemplo, como indicado por uma regra de mistura guiada por SNR, etc.) para fatias de tempo (segmentos) do conteúdo misturado com valores SNR relativamente altos entre o conteúdo da fala e outros (por exemplo, não fala, etc.) conteúdos no conteúdo misturado.

[0188] A matriz *H* pode ser reescrita/expandida como um produto de uma matriz, H_{MS} representando operações de melhora na representação M/S, multiplicada na direita, com uma matriz de transformação direta a partir da representação não M/S para a representação M/S e multiplicada à esquerda com uma inversa (que compreende um fator de 1/2) da matriz de transformação para a frente, como se mostra na seguinte expressão:

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_{MS} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \qquad (31)$$

em que a matriz de transformação exemplar da direita da matriz H_{MS} define o sinal de conteúdo misturado do canal médio na representação M/S como a soma dos dois sinais de conteúdo misturado nos dois canais c_1 e c_2, e define o sinal de conteúdo misturado no canal lateral na representação M/S como a diferença dos dois sinais de conteúdo misturados nos dois canais c_1 e c_2, com base na matriz de transformação direta. Deve notar-se que em várias modalidades, outras matrizes de transformação (por exemplo, a atribuição de diferentes pesos de diferentes canais não M/S, etc.), além das matrizes de transformação exemplares mostradas na expressão (31) podem também ser utilizadas para transformar os sinais de conteúdo misturados a partir de uma representação para uma representação diferente. Por exemplo, para a melhora de diálogo com o diálogo não processado no centro fantasma, mas panned entre os dois sinais com pesos desiguais λ_1 e λ_2. As matrizes de transformação M/S podem ser modificadas para minimizar a energia do componente de diálogo no sinal lateral, como mostrado na seguinte expressão:

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \frac{1}{2} \cdot \lambda_1 \cdot \lambda_2 \cdot \begin{pmatrix} \frac{1}{\lambda_2} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & -\frac{1}{\lambda_1} \end{pmatrix} \cdot H_{MS} \cdot \begin{pmatrix} \frac{1}{\lambda_1} & \frac{1}{\lambda_2} \\ \frac{1}{\lambda_1} & -\frac{1}{\lambda_2} \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \qquad (31A)$$

[0189] Em um exemplo de modalidade, a matriz H_{MS} representando operações de melhora na representação M/S pode ser definida como uma matriz diagonalizada (por exemplo, Hermitiana, etc.), como mostrado na seguinte expressão:

$$H_{MS} = \begin{pmatrix} g \cdot p_1 + 1 & 0 \\ 0 & g \cdot p_2 + 1 \end{pmatrix} \quad (32)$$

onde p_1 e p_2 representam parâmetros de predição do canal médio e de canal lateral, respectivamente. Cada um dos parâmetros de predição p_1 e p_2 pode compreender um conjunto de parâmetro de predição variável com o tempo ladrilhos de um sinal de conteúdo misturado correspondente na representação M/S a ser utilizada para a reconstrução do conteúdo de fala a partir do sinal de conteúdo misturado. O parâmetro de ganho *g* corresponde a um ganho de melhora de fala, *G*, por exemplo, como mostrado na expressão (10).

[0190] Em algumas modalidades, as operações de melhora de fala na representação M/S são realizadas no modo de melhora de canal independente paramétrico. Em algumas modalidades, as operações de melhora de fala na representação M/S são realizadas com o conteúdo da fala previsto, em ambos o sinal de canal médio e o sinal de canal lateral, ou com o conteúdo da fala predito no sinal de canal médio apenas. Para fins de ilustração, as operações de melhora de fala na representação M/S são realizadas com o sinal de conteúdo misturado apenas no canal médio, como se mostra na seguinte expressão:

$$H_{MS} = \begin{pmatrix} g \cdot p_1 + 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (33)$$

onde o parâmetro de predição p_1 compreende um único conjunto de parâmetros de predição para ladrilhos de frequência no tempo do sinal de conteúdo misturado no canal médio da representação M/S a ser utilizada para a reconstrução do conteúdo da fala a partir do sinal de conteúdo misturado no único canal médio.

[0191] Com base na matriz diagonalizada H_{MS} apresentada na expressão (33), as operações de melhora de fala no modo de enriquecimento paramétrico, como representada pela expressão (31), pode ser ainda reduzida com a seguinte expressão, que fornece um exemplo explícito da matriz H na expressão (30):

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 2 + g \cdot p_1 & g \cdot p_1 \\ g \cdot p_1 & 2 + g \cdot p_1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \quad (34)$$

[0192] Em um modo de melhora de híbrida paramétrica de forma de onda, as

operações de melhora de fala podem ser representadas na representação M/S com os seguintes exemplos de expressões:

$$M_e = g_1 \cdot \begin{pmatrix} d_{c,1} \\ 0 \end{pmatrix} + \begin{pmatrix} g_2 \cdot p_1 + 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \qquad (35)$$

$$= H_d \cdot D_c + H_p \cdot M$$

em que m_1 e m_2 denotam o sinal de conteúdo misturado do canal médio (por exemplo, a soma dos sinais de conteúdo misturado em canais não M/S, como os canais frontais esquerdo e direito, etc.) e o sinal de conteúdo misturado de canal lateral (por exemplo, a diferença dos sinais de conteúdo misturado em canais não M/S, como os canais frontal esquerdo e direito, etc.), respectivamente, em um vetor de sinal de conteúdo misturado *M*. Um sinal, $d_{c,1}$ denota o sinal em forma de onda do diálogo do canal médio (por exemplo, formas de onda codificadas representando uma versão reduzida de um diálogo no conteúdo misturado, etc.) em um vetor de sinal de diálogo D_C da representação M/S. Uma matriz, H_d, representa as operações de melhora de fala na representação M/S com base no sinal do diálogo $d_{C,1}$ no do canal médio da representação M/S e pode compreender apenas um elemento de matriz em linha 1 e coluna 1 (1x1). Uma matriz, H_P, representa as operações de melhora de fala na representação M/S com base em um diálogo reconstruído utilizando o parâmetro de predição p_1 para o canal médio da representação M/S. Em algumas modalidades, os parâmetros de ganho g_1 e g_2 em conjunto (por exemplo, depois de ter sido aplicado, respectivamente, ao sinal de forma de onda de diálogo e o diálogo reconstruído, etc.) correspondem a um ganho de melhora de fala, *G*, por exemplo, como representado nas expressões (23) e (24). Especificamente, o parâmetro g_1 é aplicado em operações de melhora de fala codificada por forma de onda relacionadas com o diálogo de sinal $d_{C,1}$ no canal médio da representação M/S, enquanto que o parâmetro g_2 é aplicado em operações de melhora de fala codificada paramétrica relativas aos sinais de conteúdo misturado m_1 e m_2, no do canal médio e o canal lateral da representação M/S. Parâmetros g_1 e g_2 controlam o ganho de melhora global e o trade-off entre os

dois métodos de melhora de fala.

[0193] Na representação não M/S, as operações de melhora da fala correspondentes aos representados com a expressão (35) podem ser representadas com as seguintes expressões:

$$\begin{pmatrix} M_{e,c_1} \\ M_{e,c_2} \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_d \cdot D_c + \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot H_p \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \quad (36)$$
$$= \frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \left(H_d \cdot D_c + H_p \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} M_{c_1} \\ M_{c_2} \end{pmatrix} \right)$$

em que os sinais de conteúdo misturado m_1 e m_2 na representação M/S como se mostra na expressão (35) são substituídos com sinais de conteúdo misturado M_{c1} e M_{c2} nos canais não M/S da esquerda multiplicado com a matriz de transformação direta entre a representação não M/S e a representação M/S. A matriz de transformação inversa (com um fator de ½) na expressão (36) converte os sinais de conteúdo misturado da melhorada de fala na representação M/S, como mostra a expressão (35), de volta para os sinais de conteúdo misturado da fala melhorada na representação não M/S (por exemplo, canais esquerdo e direito frontais, etc.).

[0194] Além disso, opcionalmente, ou, em alternativa, em algumas modalidades em que nenhum processamento adicional com base em QMF é feito depois de operações de melhora de fala, algumas ou todas as operações de melhora de fala (por exemplo, tal como representado pelas transformações H_d, H_P, etc.) que combinam conteúdo de melhora da fala com base no sinal de diálogo $d_{c,1}$ e conteúdo misturado de melhora de fala baseado no diálogo reconstruído por meio de predição pode ser realizada depois de um banco de filtros de síntese QMF no domínio do tempo, por razões de eficiência.

[0195] Um parâmetro de predição utilizado para construir/prever conteúdo da fala a partir de um sinal de conteúdo misturado em um ou ambos do canal médio e o canal lateral da representação M/S pode ser gerado com base em um de um ou mais métodos de geração de parâmetros de predição, incluindo, mas não limitado apenas a, qualquer um de: métodos de predição de diálogo independente de canal tal como

representado na FIGURA 1, os métodos de predição de diálogo multicanal, como representado na FIGURA 2, etc. Em algumas modalidades, pelo menos um dos métodos de geração de parâmetros de predição pode ser baseado em MMSE, gradiente descendente, um ou mais outros métodos de otimização, etc.

[0196] Em algumas modalidades, uma comutação baseada em SNR temporal "cega" como discutido anteriormente pode ser usada entre os dados da melhora codificada paramétrica (por exemplo, relacionados com conteúdos de melhora de fala com base no sinal de diálogo $d_{c,1}$ etc.) e melhora codificada por forma de onda (por exemplo, relacionada com o conteúdo misturado de melhora de fala baseado no diálogo reconstruído por meio de predição, etc.) de segmentos de um programa de áudio na representação M/S.

[0197] Em algumas modalidades, uma combinação (por exemplo, indicada por um indicador de mistura anteriormente discutido, uma combinação de g_1 e g_2 na expressão (35), etc.) de dados de forma de onda (por exemplo, relativos ao conteúdo melhorado de fala baseado no sinal de diálogo $d_{c,1}$, etc.) e os dados de fala reconstruídos (por exemplo, relacionados com o conteúdo misturado de melhora de fala baseado no diálogo reconstruído por meio de predição, etc.) nas alterações de representação M/S ao longo do tempo, com cada estado da combinação pertencente ao conteúdo de fala e outros conteúdos de áudio de um segmento correspondente ao fluxo de bits que transporta os dados em forma de onda e o conteúdo misturado utilizado na reconstrução dos dados da fala. O indicador de mistura é gerado de tal modo que o estado atual da combinação (de dados da forma de onda e os dados de fala reconstruídos) é determinado por propriedades do sinal da fala e outros conteúdos de áudio (por exemplo, uma proporção entre a potência de conteúdo de fala e a potência de outro conteúdo de áudio, uma SNR, etc.) no segmento correspondente do programa. O indicador de mistura para um segmento de um programa de áudio pode ser um parâmetro indicador de mistura (ou conjunto de parâmetros) gerado no

subsistema 29 do codificador da FIGURA 3 para o segmento. Um modelo de mascaramento auditivo, como discutido anteriormente pode ser usado para prever com mais precisão como ruídos de codificação na cópia de fala de qualidade reduzida no vetor de sinal de diálogo *Dc* estão sendo mascarados pela mistura de áudio do programa principal e para selecionar a proporção de mistura, por conseguinte.

[0198] Subsistema 28 do codificador 20 da FIGURA 3 pode ser configurado para incluir indicadores de mistura relativos às operações de melhora da fala M/S no fluxo de bits como parte dos metadados para melhora de fala M/S a serem emitidos a partir do codificador 20. Os indicadores de mistura relativos às operações de melhora da fala M/S podem ser gerados (por exemplo, no subsistema 13 do codificador da FIGURA 7) a partir de fatores de escala g_{max}(t) relativos aos artefatos de codificação do sinal de diálogo *Dc*, etc. Os fatores de escala g_{max}(t) podem ser gerados pelo subsistema 14 do codificador da FIGURA 7. Subsistema 13 do codificador da FIGURA 7 pode ser configurado para incluir os indicadores de mistura no fluxo de bits a ser produzido a partir do codificador da FIGURA 7. Além disso, opcionalmente, ou, alternativamente, o subsistema 13 pode incluir, no fluxo de bits a ser emitido a partir do codificador da FIGURA 7, os fatores de escala g_{max}(t) gerados pelo subsistema 14.

[0199] Em algumas modalidades, a mistura de áudio não melhorada, A(t), gerada pela operação 10 da FIGURA 7 representa (por exemplo, segmentos de tempo de, etc.) um vetor de sinal de conteúdo misturado na configuração do canal de áudio de referência. Os parâmetros de melhora codificada paramétrica, p(t), gerados pelo elemento 12 da FIGURA 7 representam, pelo menos, uma parte de metadados da melhora de fala M/S para a realização de melhora da fala codificada paramétrica na representação M/S em relação a cada segmento do vetor de sinal de conteúdo misturado. Em algumas modalidades, a cópia de fala de qualidade reduzida, s’(t), gerada pelo codificador 15 da FIGURA 7 representa um vetor de sinal de diálogo na representação M/S (por exemplo, com o sinal de diálogo do canal médio, o sinal de

diálogo de canal lateral, etc.).

[0200] Em algumas modalidades, o elemento 14 da FIGURA 7 gera os fatores de escala, $g_{max}(t)$, e fornece-os ao elemento codificador 13. Em algumas modalidades, um elemento 13 gera um fluxo de bits de áudio codificado indicativo do vetor de sinal de conteúdo misturado (por exemplo, não melhorado, etc.) na configuração do canal de áudio de referência, os metadados para melhora de fala M/S, o sinal de vetor de diálogo narepresentação M/S se for o caso, e os fatores de escala $g_{max}(t)$ se for o caso, para cada segmento de programa de áudio, e este fluxo de bits de áudio codificado pode ser transmitido ou de outra forma entregue a um receptor.

[0201] Quando o sinal de áudio não melhorado em uma representação não M/S é entregue (por exemplo, transmitido) com metadados para melhora de fala M/S para um receptor, o receptor pode transformar cada segmento do sinal de áudio não melhorado na representação M/S e executar operações de melhora da fala M/S indicados pelos metadados para melhora da fala M/S para o segmento. O vetor de sinal de diálogo na representação M/S para um segmento de programa pode ser fornecida com o vetor de sinal de conteúdo misturado não melhorado em representação não M/S se as operações de melhora de fala para o segmento estão sendo executadas no modo de melhora de fala híbrido, ou no modo de enriquecimento codificado por forma de onda. Se for o caso, um receptor que recebe e analisa o fluxo de bits pode ser configurado para gerar os indicadores de mistura em resposta a fatores de escala $g_{max}(t)$ e determinar os parâmetros de ganho g_1 e g_2 na expressão (35).

[0202] Em algumas modalidades, as operações de melhora da fala são realizadas, pelo menos parcialmente na representação M/S durante um receptor ao qual a saída codificada do elemento 13 foi entregue. Em um exemplo, em cada segmento do sinal de conteúdo misturado não melhorado, os parâmetros de ganho g_1 e g_2 na expressão (35) correspondentes a uma quantidade total predeterminada (por

exemplo, solicitada) de melhora pode ser aplicada baseado, pelo menos em parte, nos indicadores de mistura analisados a partir do fluxo de bits recebido pelo receptor. Em outro exemplo, em cada segmento do sinal de conteúdo misturado não melhorado, os parâmetros de ganho g_1 e g_2 na expressão (35) correspondentes a um valor total predeterminado (por exemplo, solicitado,) melhora podem ser aplicados com base, pelo menos em parte, em indicadores de mistura como determinado a partir de fatores de escala $g_{max}(t)$ para o segmento analisado a partir do fluxo de bits recebido pelo receptor.

[0203] Em algumas modalidades, elemento 23 de codificador 20 da FIGURA 3 está configurado para gerar dados paramétricos incluindo metadados para melhora de fala M/S (por exemplo, parâmetros de predição para reconstruir o conteúdo de diálogo/fala do conteúdo misturado no canal médio e/ou no canal lateral, etc.) em resposta à saída de dados de estágios 21 e 22. Em algumas modalidades, elemento de geração de indicador de mistura 29 do codificador 20 da FIGURA 3 é configurado para gerar um indicador de mistura ("BI") para a determinação de uma combinação de conteúdo melhorado da fala parametricamente (por exemplo, com o parâmetro de ganho g_1, etc.) e do conteúdo melhorado da fala baseado em forma de onda (por exemplo, com o parâmetro de ganho g_1, etc.) em resposta à saída de dados a partir dos estágios 21 e 22.

[0204] Em variações na modalidade da FIGURA 3, o indicador de mistura empregado para a melhora da fala híbrida M/S não é gerado no codificador (e não está incluído na saída do fluxo de bits do codificador), mas é gerado em vez (por exemplo, em uma variação no receptor 40), em resposta à saída do fluxo de bits do codificador (cujo fluxo de dados inclui dados de forma de onda nos canais M/S e metadados para melhora de fala M/S).

[0205] O decodificador 40 é acoplado e configurado (por exemplo, programado) para receber o sinal de áudio codificado a partir do subsistema 30 (por

exemplo, pela leitura ou recuperação de dados indicativos do sinal de áudio codificado a partir do armazenamento no subsistema 30 ou recebendo o sinal de áudio codificado que foi transmitido pelo subsistema 30), e para decodificar os dados indicativos de vetor de sinal de conteúdo misturado (fala e não fala) na configuração do canal de áudio de referência a partir do sinal de áudio codificado, e para realizar operações de melhora de fala, pelo menos em parte, na representação M/S sobre o conteúdo misturado decodificado na configuração do canal de áudio de referência. Decodificador 40 pode ser configurado para gerar e emitir (por exemplo, para um sistema de processamento, etc.) uma fala melhorada, sinal de áudio indicativo descodificado do conteúdo misturado de melhora da fala.

[0206] Em algumas modalidades, alguns ou todos os sistemas de processamento representados na FIGURA 4 até a FIGURA 6 podem ser configurados para processar conteúdo misturado de melhora de fala gerado por operações de melhora de fala M/S, pelo menos, algumas das quais são operações realizadas na representação M/S. FIGURA 6A ilustra um sistema de processamento exemplar configurado para executar as operações de melhora de fala tal como representadas na expressão (35).

[0207] O sistema de processamento de FIGURA 6A pode ser configurado para realizar operações de melhora de fala paramétricas, em resposta à determinação de que, pelo menos, um parâmetro de ganho (por exemplo, g_2 na expressão (35), etc.) utilizado nas operações de melhora de fala paramétricas diferentes de zero (por exemplo, no modo híbrido de melhora, no modo de melhora paramétrica, etc.). Por exemplo, mediante uma dita determinação, o subsistema 68A da Figura6A pode ser configurado para executar uma transformação em um vetor de sinal de conteúdo misturado ("áudio misturado (T/F)") que é distribuído através de canais não M/S para gerar um vetor de sinal de conteúdo misturado correspondente que é distribuído ao longo dos canais M/S. Esta transformação pode utilizar uma matriz de transformação

direta, conforme apropriado. Parâmetros de predição (por exemplo, p_1 e p_2, etc.), os parâmetros de ganho (por exemplo, g_2 na expressão (35), etc.) para operações de melhora paramétrica podem ser aplicados para predizer o conteúdo de fala a partir do vetor de sinal de conteúdo misturado dos canais M/S e melhorar o conteúdo da fala previsto.

[0208] O sistema de processamento de FIGURA 6A pode ser configurado para executar operações de melhora de fala codificada por forma de onda em resposta à determinação de que, pelo menos, um parâmetro de ganho (por exemplo, g_1 na expressão (35), etc.) utilizado nas operações de melhora de fala codificada por forma de onda não é zero (por exemplo, no modo de melhora híbrida, no modo de melhora codificada por forma de onda, etc.). Por exemplo, em dita uma determinação, o sistema de processamento da FIGURA 6A pode ser configurado para receber/extrair, a partir do sinal de áudio recebido codificado, um vetor de sinal de diálogo (por exemplo, com uma versão reduzida do conteúdo de fala presente no vetor de sinal de conteúdo misturado) que é distribuído através de canais M/S. Os parâmetros de ganho (por exemplo, g_1 na expressão (35), etc.) para operações de melhora codificada por forma de onda podem ser aplicados para melhorar o conteúdo de fala representado pelo vetor de sinal de diálogo de canais M/S. Um ganho de melhora definido pelo usuário (G) pode ser usado para derivar parâmetros de ganho g1 e g2 usando um parâmetro de mistura, que podem ou não estar presentes no fluxo de bits. Em algumas modalidades, o parâmetro de mistura a ser usado com o ganho de melhora definidos pelo usuário (G) para obter parâmetros de ganho g1 e g2 pode ser extraído a partir de metadados no sinal de áudio codificado recebido. Em algumas outras modalidades, um dito parâmetro de mistura não pode ser extraído a partir de metadados no sinal de áudio codificado recebido, mas pode ser derivado por um codificador de receptor com base no conteúdo de áudio no sinal de áudio codificado recebido.

[0209] Em algumas modalidades, uma combinação do conteúdo de melhora

de fala paramétrica e o conteúdo de melhora de fala codificada por forma de onda na representação M/S é afirmado ou introduzido no subsistema 64A da FIGURA 6A. Subsistema 64A da FIGURA 6 pode ser configurado para executar uma transformação na combinação do conteúdo da fala melhorada que é distribuído através de canais M/S para gerar um vetor de sinal de conteúdo de fala melhorada que é distribuído através de canais não M/S. Esta transformação pode utilizar uma matriz de transformação inversa, conforme apropriado. O vetor de sinal de conteúdo de melhora da fala dos canais não M/S pode ser combinado com o vetor do sinal de conteúdo misturado ("áudio misturado (T/F)") que é distribuído ao longo dos canais não M/S para gerar um vetor de conteúdo de misturado de melhora da fala.

[0210] Em algumas modalidades, a sintaxe do sinal de áudio codificado (por exemplo, a saída do codificador 20 da FIGURA 3, etc.) suporta a transmissão de um sinalizador M/S a partir de um codificador de áudio à montante (por exemplo, o codificador 20 da FIGURA 3, etc.) para descodificadores de áudio à jusante (por exemplo, decodificador 40 da FIGURA 3, etc.). O sinalizador M/S está presente/definido pelo codificador de áudio (por exemplo, elemento 23 no codificador 20 da FIGURA 3, etc.) quando as operações de melhora de fala devem ser executadas por um decodificador de receptor de áudio (por exemplo, decodificador 40 da FIGURA 3, etc.), pelo menos em parte, com dados de controle M/S, parâmetros de controle, etc., que são transmitidos com o sinalizador M/S. Por exemplo, quando o sinalizador M/S é definido, um sinal estéreo (por exemplo, a partir dos canais esquerdo e direito, etc.) em canais não M/S podem ser primeiro transformados pelo decodificador de áudio receptor (por exemplo, decodificador 40 da FIGURA 3, etc.) para o canal médio e o canal lateral da representação M/S antes de aplicar as operações de melhora de fala M/S com os dados de controle M/S, parâmetros de controle, etc., tal como recebidos com o sinalizador M/S, de acordo com um ou mais dos algoritmos de fala de melhora (por exemplo, a predição de diálogo independente de canal, a predição de

diálogo multicanal, com base em forma de onda, híbrida paramétrica em forma de onda, etc.). No decodificador de áudio receptor (por exemplo, decodificador 40 da FIGURA 3, etc.), depois das operações de melhora de fala M/S serem realizadas, os sinais de melhora de fala na representação M/S podem ser transformados de volta para os canais não M/S.

[0211] Em algumas modalidades, os metadados para melhora de fala gerados por um codificador de áudio (por exemplo, codificador 20 da FIGURA 3, elemento 23 do codificador 20 da FIGURA 3, etc.), como aqui descrito, podem conter um ou mais sinalizadores específicos para indicar a presença de um ou mais conjuntos de dados de controle de melhora da fala, parâmetros de controle, etc., para um ou mais tipos diferentes de operações de melhora de fala. Os um ou mais conjuntos de dados de controle de melhora da fala, parâmetros de controle, etc., para os um ou mais tipos diferentes de operações de melhora de fala podem, mas não estão limitados a apenas, incluir um conjunto de dados de controle M/S, parâmetros de controle, etc., como metadados para melhora de fala M/S. Os metadados para melhora de fala podem também incluir um sinalizador de preferência para indicar que tipo de operações de melhora de fala (por exemplo, operações de melhora de fala M/S, as operações de melhora de fala não M/S, etc.) é o preferido para o conteúdo de áudio para ser fala melhorada. Os metadados para melhora de fala podem ser entregues para um decodificador à jusante (por exemplo, decodificador 40 da FIGURA 3, etc.) como uma parte de metadados entregues em um sinal de áudio codificado que inclui o conteúdo de áudio misturado codificado para uma configuração de canal de áudio não M/S de referência. Em algumas modalidades, apenas metadados para melhora de fala M/S, mas não metadados para melhora não M/S estão incluídos no sinal de áudio codificado.

[0212] Além disso, opcionalmente, ou, alternativamente, um decodificador de áudio (por exemplo, 40 da FIGURA 3, etc.) pode ser configurado para determinar e

executar um tipo específico (por exemplo, melhora de fala M/S, melhora de fala não M/S, etc.) de operações de melhora de fala com base em um ou mais fatores. Esses fatores podem incluir, mas não se limitam a: uma ou mais entradas do usuário que especificam uma preferência para um tipo selecionado pelo usuário específico de operação de melhora de fala, entrada de usuário que especifica uma preferência para um tipo selecionado pelo sistema de operações de melhora de fala, capacidades da configuração do canal de áudio específico operado pelo decodificador de áudio, a disponibilidade de metadados para melhora de fala para o tipo específico de operação de melhora de fala, qualquer sinalizador de preferência gerado pelo codificador de um tipo de operação de melhora de fala, etc. Em algumas modalidades, o decodificador de áudio pode implementar uma ou mais regras de precedência, poderá solicitar nova entrada do usuário, etc., para determinar um tipo específico de operação de melhora fala se esses fatores conflitarem entre si.

7. FLUXOS DE PROCESSO EXEMPLARES

[0213] As FIGURA 8A e FIGURA 8B ilustram os fluxos de processo exemplares. Em algumas modalidades, um ou mais dispositivos de computador ou unidades de um sistema de processamento de mídia podem realizar este fluxo de processo.

[0214] FIGURA 8A ilustra um fluxo de processo de exemplo que pode ser implementado por um codificador de áudio (por exemplo, o codificador 20 da FIGURA 3), tal como aqui descrito. No bloco 802 da FIGURA 8A, o codificador de áudio recebe conteúdo de áudio misturado, tendo uma mistura de conteúdo de fala e conteúdo de áudio não fala, em uma representação de canal de áudio de referência, que é distribuído ao longo de uma pluralidade de canais de áudio da representação de canal de áudio de referência.

[0215] No bloco 804, o codificador de áudio transforma uma ou mais partes do conteúdo de áudio misturado que são distribuídos ao longo de um ou mais canais não

Médio/Lateral (M/S) na pluralidade de canais de áudio da representação de canal de áudio de referência em uma ou mais partes de conteúdo de áudio misturado transformado em uma representação de canal de áudio M/S que são distribuídos ao longo de um ou mais canais M/S da representação do canal de áudio M/S.

[0216] No bloco 806, o codificador de áudio determina metadados para melhora de fala M/S para uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S.

[0217] No bloco 808, o codificador de áudio gera um sinal de áudio que compreende o conteúdo de áudio misturado na representação de canal de áudio de referência e os metadados para melhora de fala M/S para uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S.

[0218] Numa modalidade, o codificador de áudio é ainda configurado para executar: a geração de uma versão do conteúdo da fala, na representação de canal de áudio M/S, separado do conteúdo de áudio misturado; e emitir o sinal de áudio codificado com a versão do conteúdo da fala na representação de canal de áudio M/S.

[0219] Numa modalidade, o codificador de áudio é ainda configurado para executar: gerar mistura indicando dados que permitem um decodificador de áudio receptor para aplicar melhora de fala para o conteúdo de áudio misturado com uma combinação quantitativa específica de melhora de fala codificada por forma de onda com base na versão do conteúdo da fala na representação de canal de áudio M/S e melhora de fala paramétrica com base numa versão reconstruída do conteúdo da fala na representação de canal de áudio M/S; e emitir o sinal de áudio codificado com a mistura indicando dados.

[0220] Numa modalidade, o codificador de áudio é ainda configurado para a codificação de uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S, como uma parte do sinal de áudio.

[0221] A FIGURA 8B ilustra um fluxo de processo exemplar que pode ser

implementado por um decodificador de áudio (por exemplo, decodificador 40 da FIGURA 3), tal como aqui descrito. No bloco 822 da FIGURA 8B, o decodificador de áudio recebe um sinal de áudio que inclui conteúdo de áudio misturado em uma representação de canal de áudio de referência e metadados para melhora de fala Médio/Lateral (M/S).

[0222] No bloco 824 da FIGURA 8B, o decodificador de áudio transforma uma ou mais partes do conteúdo de áudio misturado que são distribuídas ao longo de um, dois ou mais canais não M/S, numa pluralidade de canais de áudio da representação de canal de áudio de referência em uma ou mais partes de conteúdo de áudio misturado transformado numa representação de canal de áudio M/S que são distribuídas ao longo de um ou mais canais M/S da representação do canal de áudio M/S.

[0223] No bloco 826 da FIGURA 8B, o decodificador de áudio executa uma ou mais operações de melhora de fala M/S, com base nos metadados para melhora de fala M/S, em que uma ou mais partes de conteúdo de áudio misturado transformado na representação canal de áudio M/S para gerar uma ou mais partes de conteúdo de melhora de fala na representação M/S.

[0224] No bloco 828 da FIGURA 8B, o decodificador de áudio combina uma ou mais partes de conteúdo de áudio misturado transformado na representação canal de áudio M/S com um ou mais de conteúdo de melhora de fala na representação M/S para gerar uma ou mais partes de conteúdo de áudio misturado de melhora de fala na representação M/S.

[0225] Numa modalidade, o decodificador de áudio é ainda configurado para transformar inversamente as uma ou mais partes de conteúdo de áudio misturado de melhora de fala na representação M/S para uma ou mais partes de conteúdo de áudio misturado de melhora de fala na representação de canal de áudio de referência.

[0226] Numa modalidade, o decodificador de áudio é ainda configurado para

executar: extrair uma versão do conteúdo da fala, na representação de canal de áudio M/S, separado do conteúdo de áudio misturado a partir do sinal de áudio; e executar uma ou mais operações de melhora de fala, com base nos metadados para melhora fala M/S, em uma ou mais partes da versão do conteúdo da fala na representação de canal de áudio M/S para gerar uma ou mais segundas porções do conteúdo de melhora de fala na representação canal de áudio M/S.

[0227] Numa modalidade, o decodificador de áudio é ainda configurado para executar: determinar dados indicadores de mistura para a melhora da fala; e gerar, a partir de dados indicadores de mistura para a melhora de fala, uma combinação quantitativa específica de melhora de fala codificada por forma de onda baseada na versão do conteúdo da fala na representação de canal de áudio M/S e a melhora da fala paramétrica baseada em uma versão reconstruída do o conteúdo da fala na representação de canal de áudio M/S.

[0228] Numa modalidade, os dados indicando mistura são gerados com base, pelo menos em parte, de um ou mais valores de SNR para uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S. Os um ou mais valores de SNR representam uma ou mais da proporção do poder do conteúdo de fala e o conteúdo de áudio não fala de uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S, ou proporções de alimentação de conteúdos de fala e o conteúdo de áudio total de uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S.

[0229] Numa modalidade, a combinação quantitativa específica de melhora de fala codificada por forma de onda baseada na versão do conteúdo da fala na representação de canal de áudio M/S e a melhora da fala paramétrica baseada em uma versão reconstruída do conteúdo da fala da representação de canal de áudio M/S é determinada com um modelo de mascaramento auditivo em que a melhora da fala

codificada por forma de onda com base na versão do conteúdo da fala na representação de canal de áudio M/S representa uma maior quantidade relativa de melhora de fala numa pluralidade de combinações de melhora de fala codificada por forma de onda e a melhora da fala paramétrica que assegura que a codificação de ruído em um programa de áudio de melhora de fala de saída não é desagradavelmente audível.

[0230] Numa modalidade pelo menos uma porção dos metadados para melhora fala M/S permite que um decodificador de áudio receptor para reconstruir uma versão do conteúdo da fala na representação M/S a partir do conteúdo de áudio misturado na representação de canal de áudio de referência.

[0231] Numa modalidade, os metadados para melhora de fala M/S compreendem metadados relativos a uma ou mais das operações de melhora de fala codificada por forma de onda na representação de canal de áudio M/S, ou operações de melhora de fala paramétricas no canal de áudio M/S.

[0232] Numa modalidade, a representação de canal de áudio de referência compreende canais de áudio, relativamente ao circundar alto-falantes. Numa modalidade, os um ou mais canais não M/S da representação do canal de áudio de referência compreendem um ou mais de um canal central, um canal esquerdo, ou um canal direito, enquanto que os um ou mais canais M/S da representação de canal de áudio M/S compreendem um ou mais de um canal médio ou um canal lateral.

[0233] Numa modalidade, os metadados para melhora de fala M/S compreendem um único conjunto de metadados para melhora de fala relativa a um do canal médio da representação de canal de áudio M/S. Numa modalidade, metadados para melhora de fala M/S representam uma parte de metadados gerais de áudio codificados no sinal de áudio. Numa modalidade, os metadados de áudio codificados no sinal de áudio compreendem um campo de dados para indicar um presença dos metadados para melhora de fala M/S. Numa modalidade, o sinal de áudio é uma parte

de um sinal audiovisual.

[0234] Numa modalidade, um aparelho que compreende um processador é configurado para realizar qualquer um dos métodos tal como aqui descrito.

[0235] Numa modalidade, um meio de armazenamento legível por computador não transitório, que compreende instruções de software, que quando executadas por um ou mais processadores causam o desempenho de qualquer um dos métodos tal como aqui descrito. Note-se que, embora modalidades separadas sejam aqui discutidas, qualquer combinação das modalidades e/ou modalidades aqui discutidas parciais podem ser combinadas para formar outras modalidades.

8. MECANISMOS DE IMPLEMENTAÇÃO - VISÃO GERAL DO HARDWARE

[0236] De acordo com uma modalidade, as técnicas aqui descritas são executadas por um ou mais dispositivos de computador para fins especiais. Os dispositivos de computação para fins especiais podem ser com fios para executar as técnicas, ou podem incluir dispositivos eletrônicos digitais, como um ou mais circuitos integrados específicos por aplicação (ASICs) ou matrizes de portas programáveis de campo (FPGAs) que são persistentemente programadas para executar as técnicas, ou podem incluir um ou mais processadores de hardware gerais para fins programados para executar as técnicas de acordo com a instruções de programa em firmware, memória, outro armazenamento, ou uma combinação. Tais dispositivos de finalidade específica de computação também podem combinar lógica personalizada de fio rígido, ASICs ou FPGAs com programação personalizada para realizar as técnicas. Os dispositivos de computação de propósito especial podem ser sistemas de computadores desktop, sistemas de computadores portáteis, dispositivos portáteis, dispositivos de rede ou qualquer outro dispositivo que incorpora lógica de hardware com fio e/ou programa para implementar as técnicas.

[0237] Por exemplo, a FIGURA 9 é um diagrama de blocos que ilustra um sistema de computador 900 no qual uma modalidade da invenção pode ser

implementada. O sistema de computador 900 inclui um barramento 902 ou outro mecanismo de comunicação para comunicar informações, e um processador de hardware 904 juntamente com o barramento 902 para o processamento de informações. Processador hardware 904 pode ser, por exemplo, um microprocessador de uso geral.

[0238] O sistema de computador 900 inclui ainda uma memória principal 906, tal como uma memória de acesso aleatório (RAM) ou outro dispositivo de armazenamento dinâmico, acoplado ao barramento 902 para armazenar informações e instruções para serem executadas pelo processador 904. A memória principal 906 também pode ser usada para armazenar variáveis temporárias ou outras informações intermediárias durante a execução de instruções a serem executadas pelo processador 904. Tais instruções, quando armazenadas em meios de armazenamento não transitórios acessíveis ao processador 904, processam sistema de computador 900 em uma máquina de propósito especial que é específica de dispositivo para realizar as operações especificadas no manual de instruções.

[0239] O sistema de computador 900 inclui ainda uma memória apenas de leitura (ROM) 908 ou outro dispositivo de armazenamento estático acoplado ao barramento 902 para armazenar informações estáticas e instruções para o processador 904. Um dispositivo de armazenamento 910, tal como um disco magnético ou um disco óptico, é fornecido e acoplado ao barramento 902 para armazenar informações e instruções.

[0240] O sistema de computador 900 pode ser acoplado através do barramento 902 a um visor 912, tal como uma tela de cristais líquidos (LCD), para exibir informações a um usuário do computador. Um dispositivo de entrada 914, incluindo alfanumérico e outras teclas, é acoplado ao barramento 902 para a comunicação de informação e de seleções de comando para o processador 904. Um outro tipo de dispositivo de entrada do usuário é o controle de cursor 916, como um

mouse, um trackball ou as teclas de direção do cursor para comunicar a informação de direção e de seleções de comando a um processador 904 e para controlar o movimento do cursor no mostrador 912. Este dispositivo de entrada tipicamente tem dois graus de liberdade em dois eixos, um primeiro eixo (por exemplo, x) e um segundo eixo (por exemplo, y), isto permite que o dispositivo especifique posições em um plano.

[0241] O sistema de computador 900 pode implementar as técnicas aqui descritas usando a lógica específica do dispositivo hard-wired, um ou mais ASICs ou FPGAs, firmware e/ou lógica do programa, que em combinação com o sistema de computador causa ou programa o sistema de computador 900 para ser uma máquina para fins especiais. De acordo com uma modalidade, as técnicas aqui são realizadas pelo sistema de computador 900 em resposta ao processador 904 de executar uma ou mais sequências de uma ou mais instruções contidas na memória principal 906. Tais instruções podem ser lidas na memória principal 906 a partir de outro meio de armazenamento, tal como o dispositivo de armazenamento 910. A execução das sequências de instruções contidas na memória principal 906 faz com que o processador 904 execute as etapas do processo aqui descritas. Em modalidades alternativas, circuitos de hard-wired podem ser utilizados em lugar de ou em combinação com instruções de software.

[0242] O termo "meios de armazenamento", tal como aqui utilizado refere-se a quaisquer meios não transitórios que armazenam dados e/ou instruções que fazem com que uma máquina opere de uma forma específica. Tais meios de armazenamento podem compreender meios voláteis e/ou meios não voláteis. Os meios não voláteis incluem, por exemplo, discos ópticos ou magnéticos, como o dispositivo de armazenamento 910. Os meios voláteis incluem a memória dinâmica, tal como a memória principal 906. As formas mais comuns de meios de armazenamento incluem, por exemplo, um disquete, um disco flexível, disco rígido, drive de estado sólido, fita

magnética ou qualquer outro meio de armazenamento de dados magnéticos, um CD-ROM, qualquer outro meio de armazenamento de dados ópticos, qualquer meio físico com padrões de furos, uma RAM, uma PROM e EPROM, uma FLASH- EPROM, NVRAM, qualquer outro chip de memória ou cartucho.

[0243] Os meios de armazenamento são distintos de, mas podem ser utilizados em conjunto com meios de transmissão. Os meios de transmissão participam na transferência de informação entre os meios de armazenamento. Por exemplo, os meios de transmissão incluem os cabos coaxiais, fios de cobre e as fibras ópticas, incluindo os fios que compõem barramento 902. Os meios de transmissão também podem assumir a forma de ondas de luz ou acústicas, como as geradas durante as comunicações de ondas de rádio e dados de infravermelhos.

[0244] As várias formas de meios podem estar envolvidas na realização de uma ou mais sequências de uma ou mais instruções para o processador 904 para execução. Por exemplo, as instruções podem inicialmente ser transportadas em um disco magnético ou uma unidade de estado sólido de um computador remoto. O computador remoto pode carregar as instruções para a memória dinâmica e enviar as instruções através de uma linha telefônica usando um modem. Um modem local para o sistema de computador 900 pode receber os dados na linha telefônica e usar o transmissor de infravermelhos para converter os dados para um sinal infravermelho. Um detector de infravermelho pode receber os dados transportados no sinal infravermelho e circuitos apropriados podem colocar os dados no barramento 902. O barramento 902 transporta os dados para a memória principal 906, a partir da qual o processador 904 recupera e executa as instruções. As instruções recebidas pela memória principal 906 podem, opcionalmente, ser armazenadas no dispositivo de armazenamento 910 antes ou após a execução pelo processador 904.

[0245] O sistema de computador 900 também inclui uma interface de comunicação 918 acoplada ao barramento 902. A interface de comunicação 918

proporciona um acoplamento de duas vias de comunicação de dados para uma ligação de rede 920 que está ligada a uma rede local 922. Por exemplo, a interface de comunicação 918 pode ser um cartão de rede digital de serviços integrados (RDIS), modem de cabo, modem via satélite, ou um modem para fornecer uma conexão de comunicação de dados a um correspondente tipo de linha telefônica. Como outro exemplo, a interface de comunicação 918 pode ser um cartão de rede de área local (LAN), para proporcionar uma ligação de comunicação de dados a uma rede LAN compatível. As ligações sem fios podem ser implementadas. Em qualquer aplicação, interface de comunicação 918 envia e recebe sinais elétricos, eletromagnéticos ou ópticos que transportam fluxos de dados digitais que representam vários tipos de informação.

[0246] A ligação de rede 920 normalmente fornece comunicação de dados através de uma ou mais redes para outros dispositivos de dados. Por exemplo, ligação de rede 920 pode fornecer uma conexão através de rede local 922 a um computador hospedeiro 924 ou equipamento de dados operado por um provedor de serviço de Internet (ISP) 926. ISP 926, por sua vez fornece serviços de comunicação de dados através da rede de comunicação de dados de pacote de largura mundial agora comumente referida como o "Internet" 928. A rede local 922 e Internet 928 ambas usam sinais elétricos, eletromagnéticos ou ópticos que transportam fluxos de dados digitais. Os sinais através das várias redes e os sinais na ligação de rede 920 e através da interface de comunicação 918, que transportam os dados digitais de e para o sistema de computador 900, são exemplos de formas de mídia de transmissão.

[0247] O sistema de computador 900 pode enviar mensagens e receber dados, incluindo o código do programa, através das redes, ligação de rede 920 e interface de comunicação 918. No exemplo Internet, um servidor 930 pode transmitir um código solicitado para um programa aplicativo através da Internet 928, ISP 926, rede local 922 e interface de comunicação 918.

[0248] O código recebido pode ser executado pelo processador 904, uma vez que é recebido e/ou armazenado no dispositivo de armazenamento 910, ou outro tipo de armazenamento não volátil para execução posterior.

9. EQUIVALENTES, EXTENSÕES, ALTERNATIVAS E DIVERSOS

[0249] Na descrição anterior, as modalidades da invenção foram descritas com referência a numerosos detalhes específicos que podem variar de implementação para implementação. Assim, o indicador único e exclusivo do que é a invenção, e os requerentes pretendem que seja a invenção, é o conjunto de concretizações que expedirem deste pedido na forma específica na qual tais concretizações expedem, incluindo qualquer correção posterior. Quaisquer definições expressamente estabelecidas aqui para termos contidos em tais concretizações regerão o significado desses termos utilizados nas concretizações. Por isso, nenhuma limitação, elemento, característica, característica, vantagem ou atributo que não seja expressamente mencionado em uma concretização deve limitar o escopo de tal concretização de qualquer forma. O relatório descritivo e os desenhos são, por conseguinte, para serem considerados em um sentido ilustrativo em vez de restritivo.

REIVINDICAÇÕES

1. Método **CARACTERIZADO** pelo fato de que compreende:

receber conteúdo de áudio misturado, em uma representação de canal de áudio de referência, que é distribuído através de uma pluralidade de canais de áudio da representação de canal de áudio de referência, o conteúdo de áudio misturado possuindo uma mistura de conteúdo de fala e conteúdo de áudio não fala;

transformar uma ou mais partes do conteúdo de áudio misturado que são distribuídas ao longo de dois ou mais canais não Médio/Lateral (não M/S) na pluralidade de canais de áudio da representação de canal de áudio de referência em uma ou mais partes de conteúdo de áudio misturado transformado em uma representação de canal de áudio M/S que são distribuídos ao longo de um ou mais canais da representação de canal de áudio M/S, em que a representação de canal de áudio M/S compreende pelo menos um sinal de canal médio e um sinal de canal lateral, em que o sinal de canal médio representa uma soma ponderada ou não ponderada de dois canais não M/S da representação de canal de áudio de referência, e em que o sinal de canal lateral representa uma diferença ponderada ou não ponderada de dois canais não M/S da representação de canal de áudio de referência;

determinar metadados para melhora de fala para a uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S;

decodificar, através de um decodificador de áudio, o sinal de canal médio e o sinal de canal lateral em um sinal de canal esquerdo e um sinal de canal direito, em que a decodificação inclui decodificar com base nos metadados para melhora de fala, em que os metadados para melhora de fala incluem um sinalizador de preferência que indica pelo menos um tipo de operação para melhora de fala a ser executado no sinal de canal médio e no sinal de canal lateral durante decodificação, e em que os metadados para melhora de fala indicam ainda um primeiro tipo de melhora de fala para o sinal de canal médio e um segundo tipo de melhora de fala do sinal de canal

médio; e

gerar um sinal de áudio que compreende o conteúdo de áudio misturado e os metadados para melhora de fala para as uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S;

em que o método é realizado por um ou mais dispositivos de computação.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o conteúdo de áudio misturado está em uma representação de canal de áudio não M/S.

3. Método, de acordo com a reivindicação 1 ou 2, **CARACTERIZADO** pelo fato de que compreende ainda

gerar uma versão do conteúdo de fala, na representação de canal de áudio M/S, separado do conteúdo de áudio misturado; e

emitir o sinal de áudio codificado com a versão do conteúdo de fala na representação de canal de áudio M/S.

4. Método, de acordo com a reivindicação 3, **CARACTERIZADO** pelo fato de que compreende ainda:

gerar dados indicando mistura indicando uma combinação quantitativa específica de primeiro e segundo tipos de melhora de fala a ser gerada por um decodificador de áudio receptor, em que o primeiro tipo de melhora de fala é melhora de fala codificada em forma de onda com base na versão do conteúdo de fala na representação de canal de áudio M/S e em que o segundo tipo de melhora da fala é melhora de fala paramétrica com base em uma versão reconstruída do conteúdo de fala na representação de canal de áudio M/S; e

emitir o sinal de áudio codificado com os dados indicando mistura.

5. Método, de acordo com a reivindicação 4, **CARACTERIZADO** pelo fato de que pelo menos parte dos metadados para melhora de fala permite que um decodificador de áudio reconstrua a versão reconstruída de melhora de fala na

representação M/S do conteúdo de áudio misturado na representação de canal de áudio de referência.

6. Método, de acordo com a reivindicação 4 ou 5, **CARACTERIZADO** pelo fato de que os dados indicando mistura são gerados com base, pelo menos em parte, em um ou mais valores SNR para as uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S, em que os um ou mais valores SNR representam uma ou mais dentre proporções de potência de conteúdo de fala e conteúdo de áudio não fala das uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S, ou proporções de potência de conteúdo de fala e conteúdo de áudio total das uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S.

7. Método, de acordo com qualquer uma das reivindicações 4 a 6, **CARACTERIZADO** pelo fato de que a combinação quantitativa específica de primeiro e segundo tipos de melhora de fala é determinada com um modelo de mascaramento auditivo em que o primeiro tipo de melhora de fala representa uma maior quantidade relativa de melhora de fala em uma pluralidade de combinações dos primeiro e segundo tipos de melhora de fala que garante que ruído de codificação em um programa de áudio de fala melhorada emitido não é desagradavelmente audível.

8. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que pelo menos uma parte de metadados para melhora de fala permite que um decodificador de áudio receptor reconstrua uma versão do conteúdo de fala na representação M/S a partir do conteúdo de áudio misturado na representação de canal de áudio de referência.

9. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que os metadados para melhora de fala compreendem metadados relativos a uma ou mais dentre operações de melhora de

fala na representação de canal de áudio M/S, baseado na versão do conteúdo de fala, ou operações de melhora de fala paramétricas na representação de canal de áudio M/S.

10. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a representação de canal de áudio de referência compreende canais de áudio relacionados a alto-falantes *surround*.

11. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que os metadados para melhora de fala compreendem um único conjunto de metadados para melhora de fala relativo a um canal médio da representação de canal de áudio M/S.

12. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que compreende ainda impedir codificação das uma ou mais partes de conteúdo de áudio misturado transformado na representação de canal de áudio M/S como uma parte do sinal de áudio.

13. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que os metadados para melhora de fala representam uma parte de metadados de áudio geral codificados no sinal de áudio.

14. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que metadados de áudio codificados no sinal de áudio compreendem um campo de dados para indicar uma presença dos metadados para melhora de fala.

15. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que o sinal de áudio é uma parte de um sinal audiovisual.

16. Sistema de processamento de mídia **CARACTERIZADO** pelo fato de que é configurado para realizar qualquer um dos métodos definidos nas reivindicações 1 a 15.

17. Aparelho **CARACTERIZADO** pelo fato de que compreende um processador e é configurado para realizar qualquer um dos métodos definidos nas reivindicações 1 a 15.

18. Meio de armazenamento legível por computador não transitório **CARACTERIZADO** pelo fato de que compreende um conjunto de instruções que, quando executadas por um ou mais processadores, geram execução de qualquer um dos métodos definidos nas reivindicações 1 a 15.

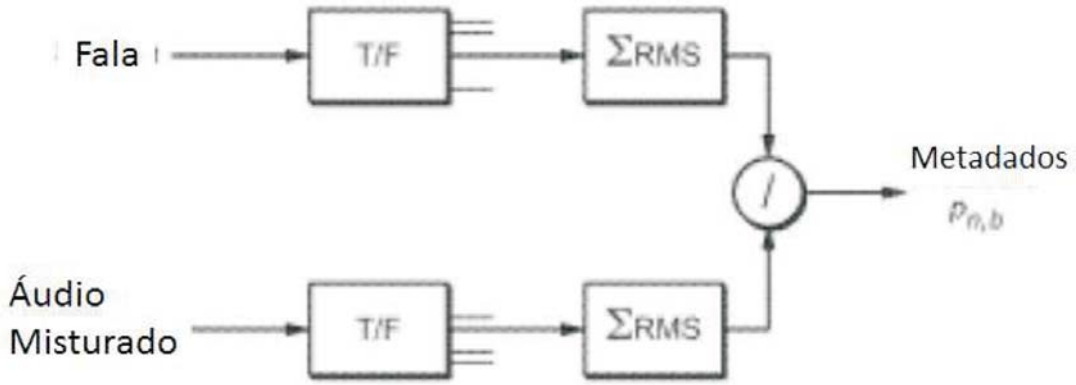

Fala
T/F
ΣRMS
Metadados
$p_{n,b}$
Áudio
Misturado
T/F
ΣRMS

FIG. 1

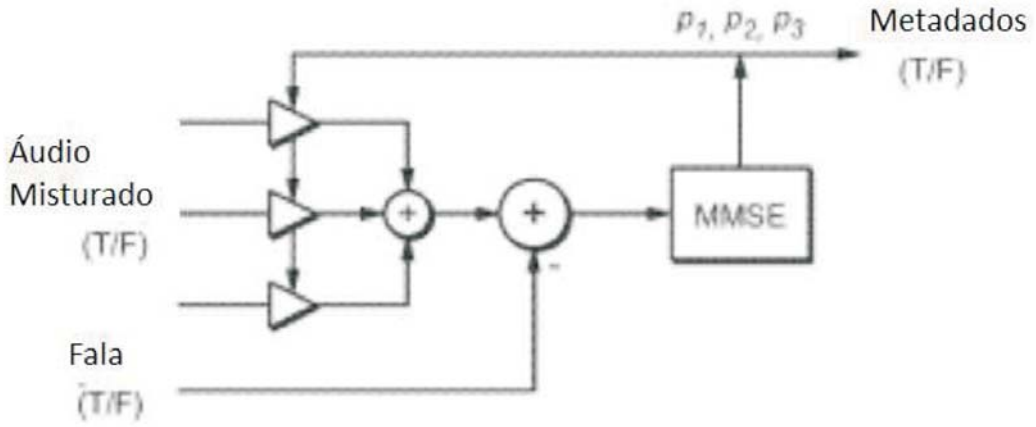

p_1, p_2, p_3
Metadados
(T/F)
Áudio
Misturado
(T/F)
MMSE
Fala
(T/F)

FIG. 2

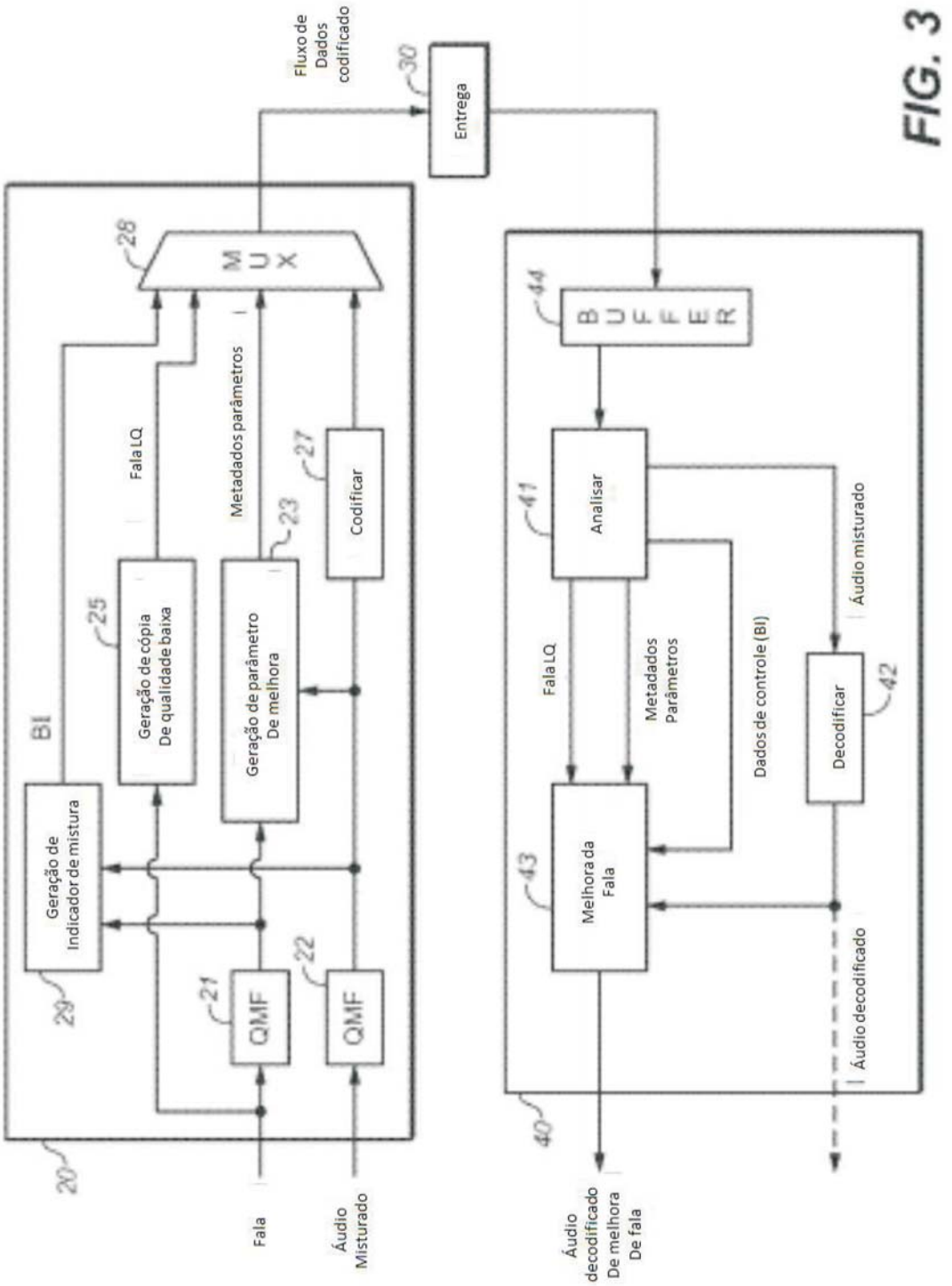

Fluxo de Dados codificado
30
Entrega
28
MUX
BUFFER
44
Fala LQ
Metadados parâmetros
27
Codificar
23
Analisar
41
25
Geração de cópia De qualidade baixa
Geração de parâmetro De melhora
BI
Geração de Indicador de mistura
29
21
QMF
22
QMF
Fala LQ
Metadados Parâmetros
Dados de controle (BI)
Áudio misturado
Decodificar
42
43
Melhora da Fala
20
40
Áudio decodificado
Áudio Misturado
Fala
Áudio decodificado De melhora De fala

FIG. 3

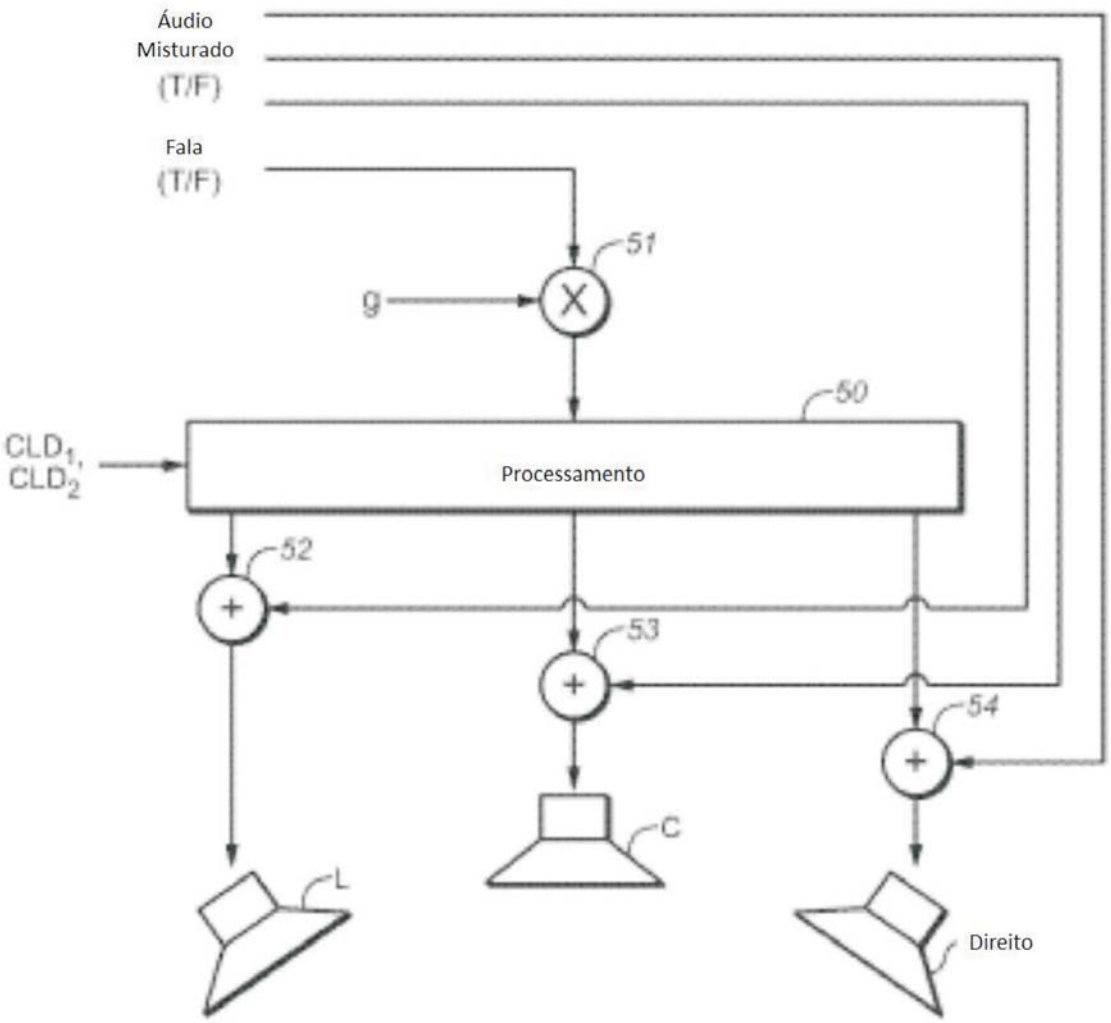

Áudio
Misturado
(T/F)
Fala
(T/F)
51
g
50
CLD1,
CLD2
Processamento
52
53
54
C
L
Direito

FIG. 4

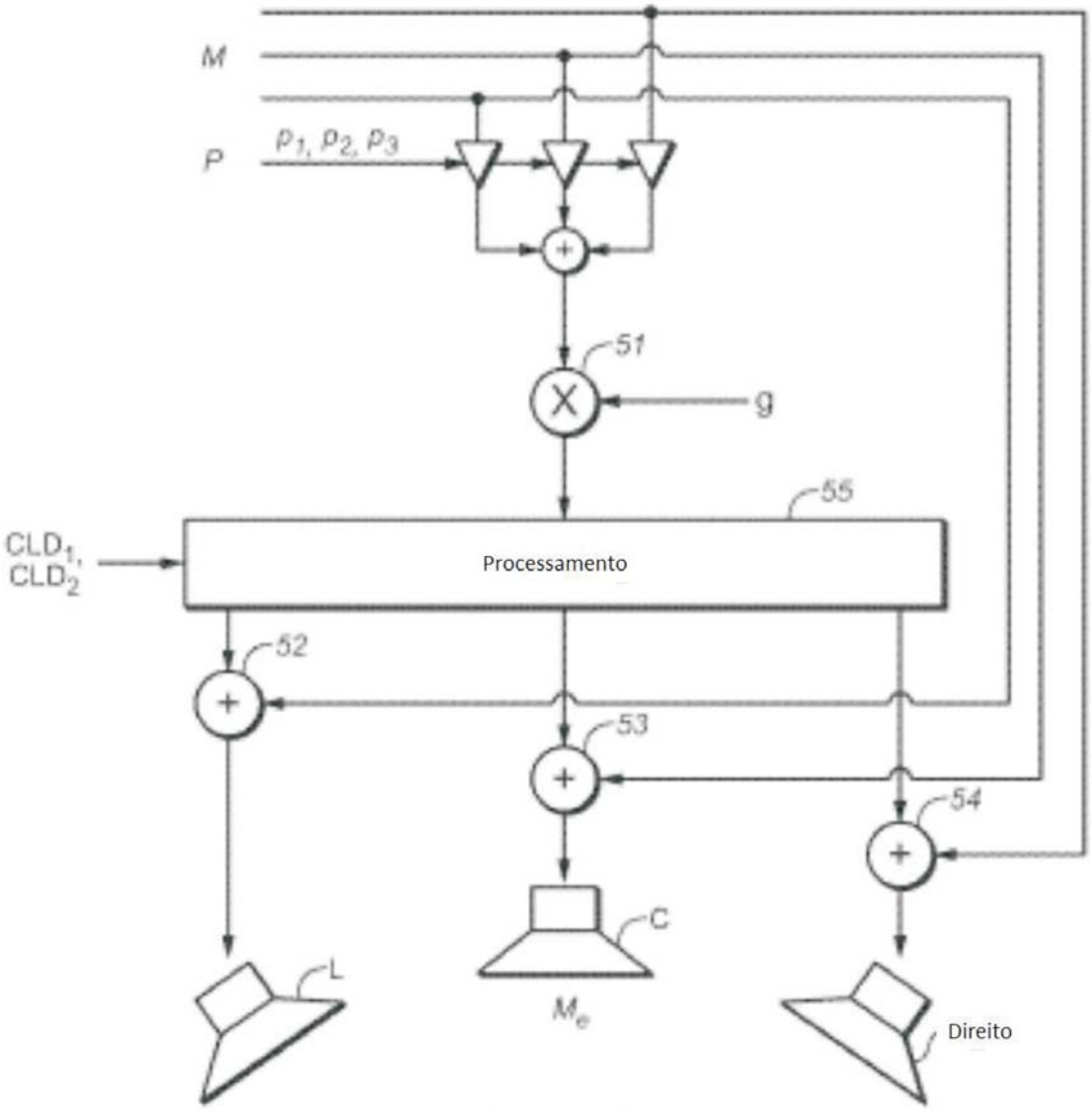

M
P
p_1, p_2, p_3
51
g
55
CLD_1, CLD_2
Processamento
52
53
54
C
L
M_e
Direito

FIG. 5

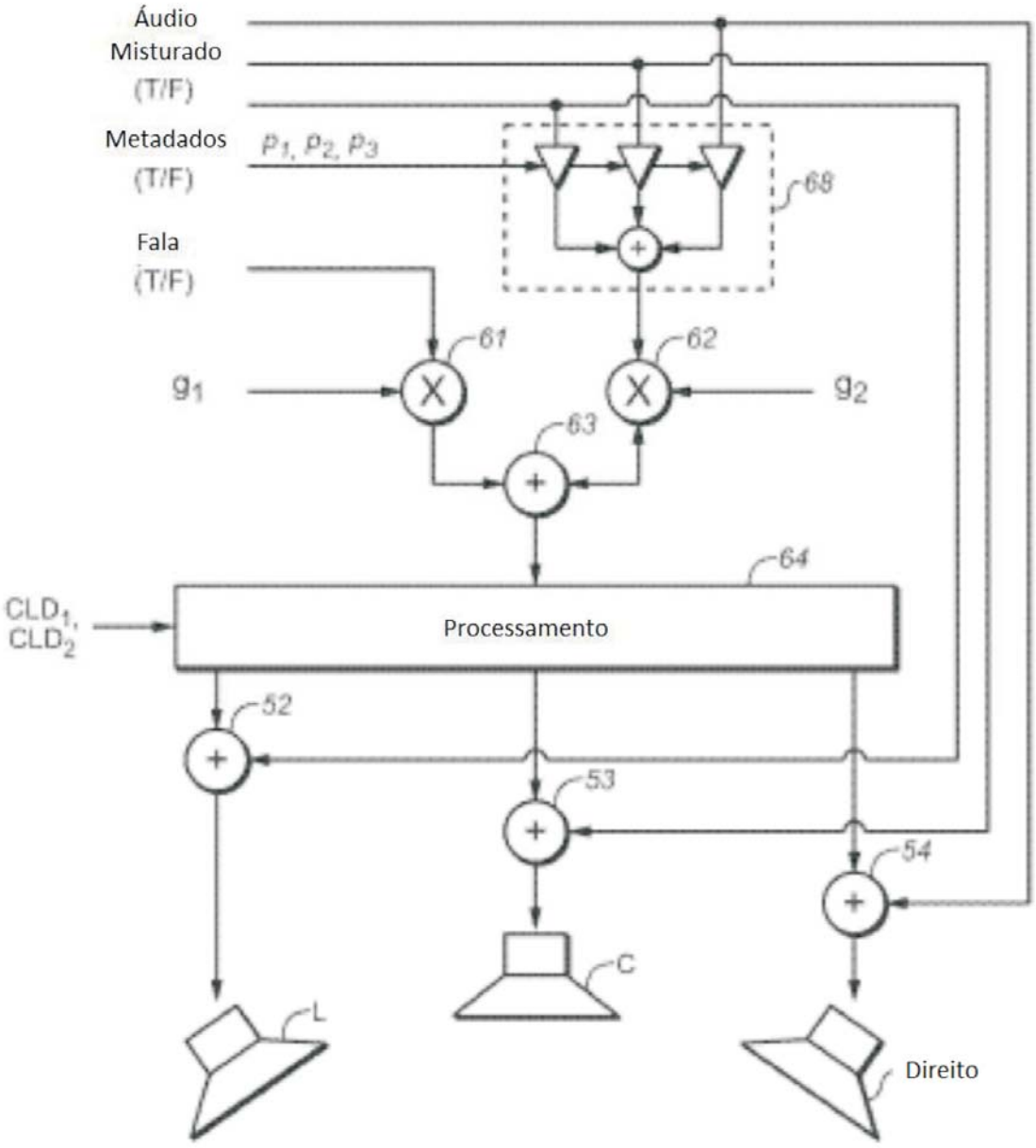

Áudio
Misturado
(T/F)
Metadados
(T/F)
P_1, P_2, P_3
68
Fala
(T/F)
61
62
g_1
g_2
63
64
CLD_1,
CLD_2
Processamento
52
53
54
C
L
Direito

FIG. 6

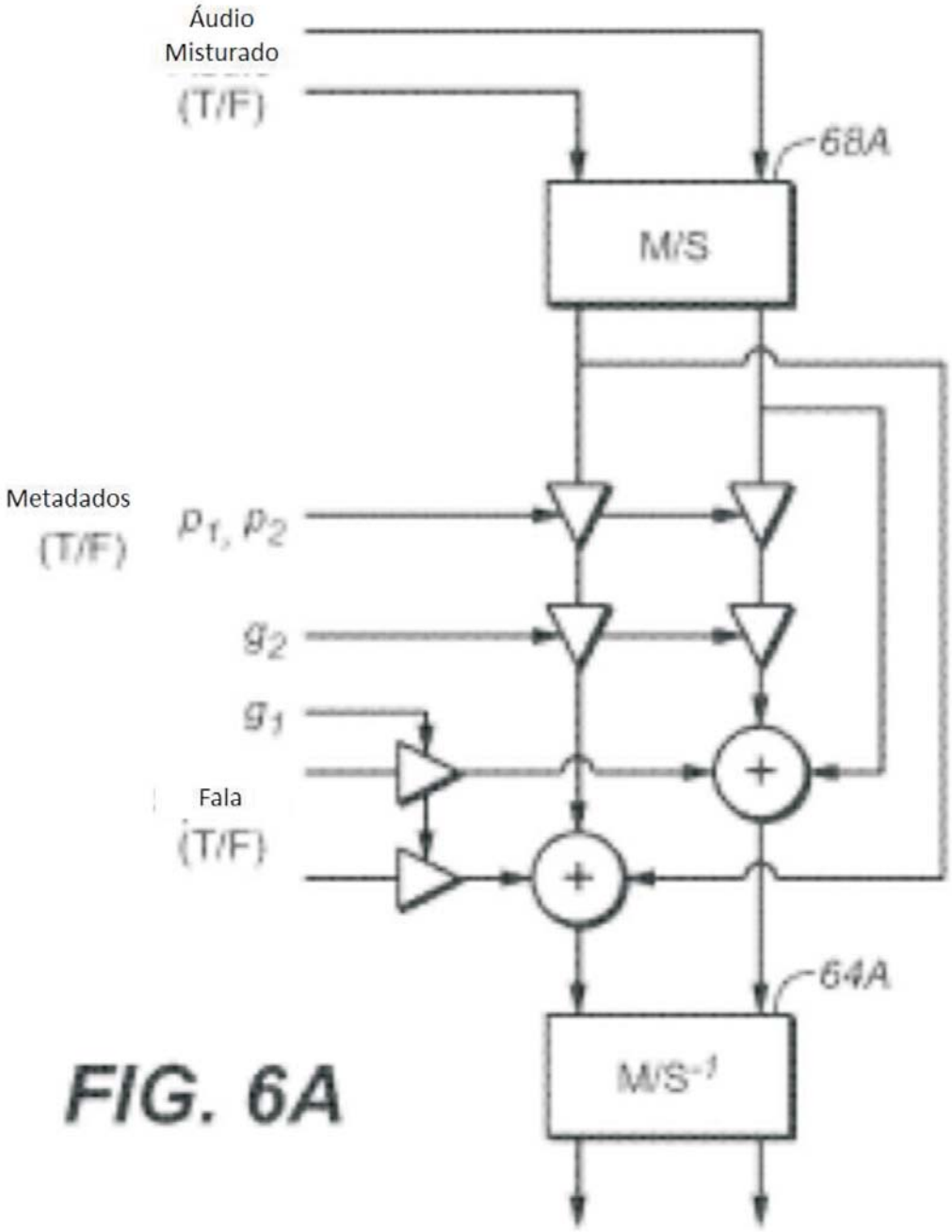

Áudio
Misturado
(T/F)
68A
M/S
Metadados
(T/F)
p_1, p_2
g_2
g_1
Fala
(T/F)
64A
M/S^{-1}

FIG. 6A

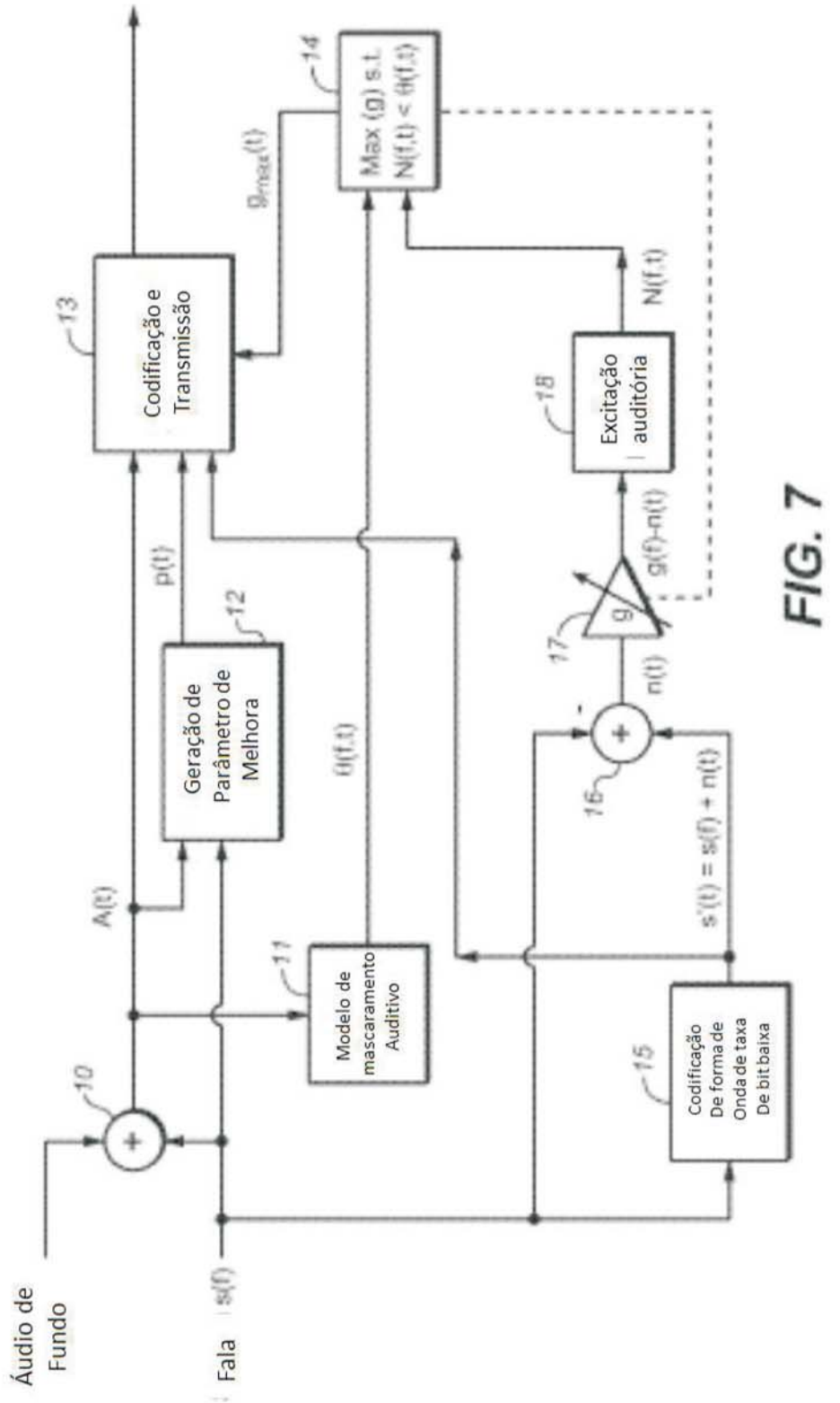

Áudio de Fundo
Fala
s(f)
10
A(t)
11
Modelo de mascaramento Auditivo
12
Geração de Parâmetro de Melhora
p(t)
θ(f,t)
13
Codificação e Transmissão
$g_{max}(t)$
14
Max (g) s.t. N(f,t) < θ(f,t)
15
Codificação De forma de Onda de taxa De bit baixa
s'(t) = s(f) + n(t)
16
n(t)
17
g
g(f)·n(t)
18
Excitação auditória
N(f,t)

FIG. 7

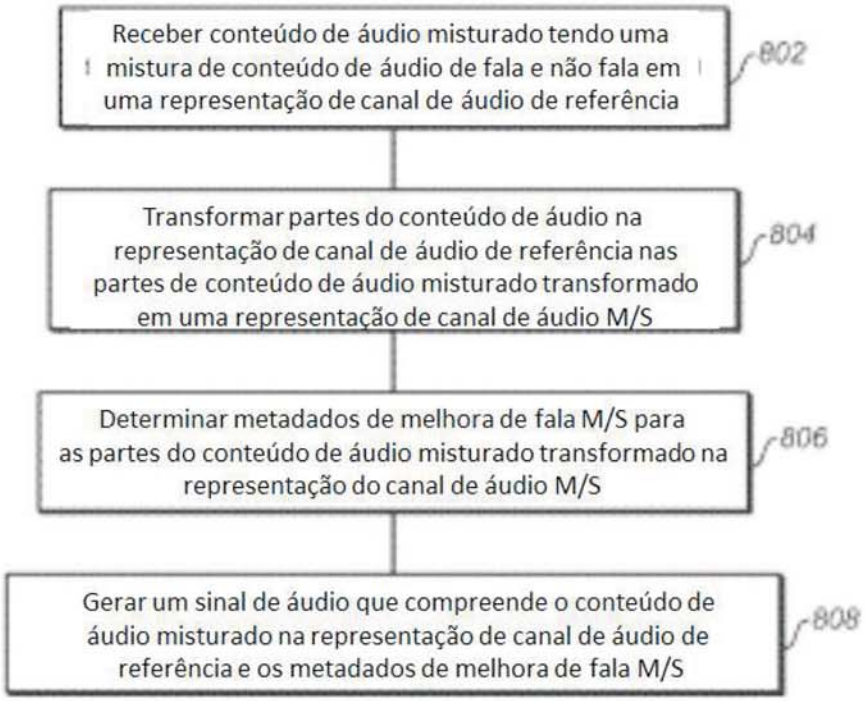

Receber conteúdo de áudio misturado tendo uma mistura de conteúdo de áudio de fala e não fala em uma representação de canal de áudio de referência
802
Transformar partes do conteúdo de áudio na representação de canal de áudio de referência nas partes de conteúdo de áudio misturado transformado em uma representação de canal de áudio M/S
804
Determinar metadados de melhora de fala M/S para as partes do conteúdo de áudio misturado transformado na representação do canal de áudio M/S
806
Gerar um sinal de áudio que compreende o conteúdo de áudio misturado na representação de canal de áudio de referência e os metadados de melhora de fala M/S
808

FIG. 8A

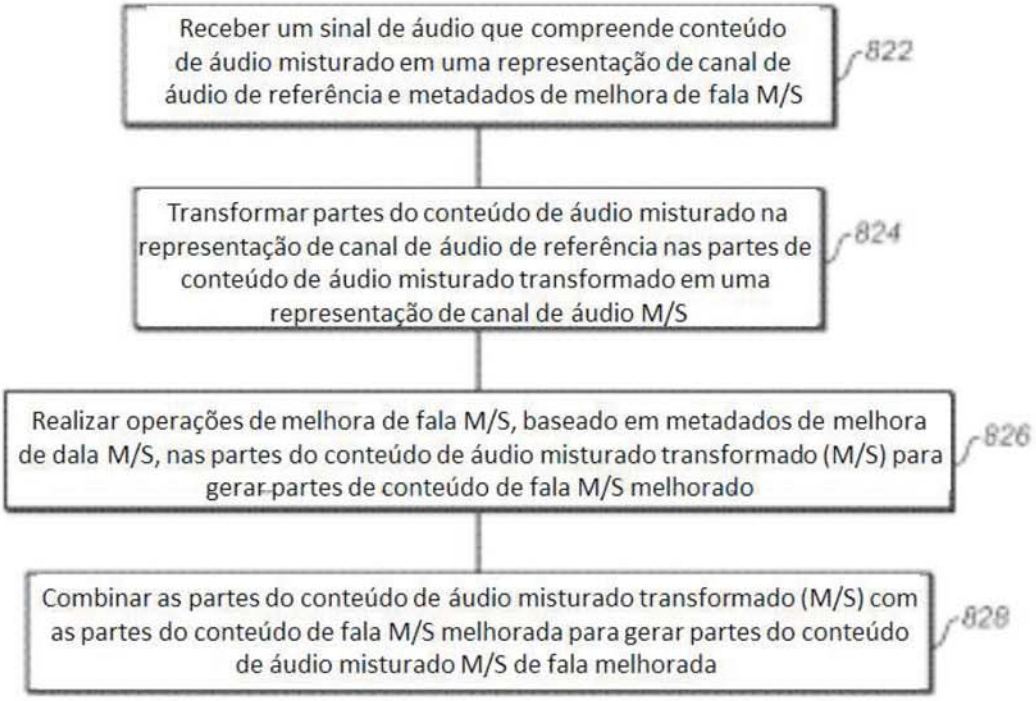

Receber um sinal de áudio que compreende conteúdo de áudio misturado em uma representação de canal de áudio de referência e metadados de melhora de fala M/S
822
Transformar partes do conteúdo de áudio misturado na representação de canal de áudio de referência nas partes de conteúdo de áudio misturado transformado em uma representação de canal de áudio M/S
824
Realizar operações de melhora de fala M/S, baseado em metadados de melhora de dala M/S, nas partes do conteúdo de áudio misturado transformado (M/S) para gerar partes de conteúdo de fala M/S melhorado
826
Combinar as partes do conteúdo de áudio misturado transformado (M/S) com as partes do conteúdo de fala M/S melhorada para gerar partes do conteúdo de áudio misturado M/S de fala melhorada
828

FIG. 8B

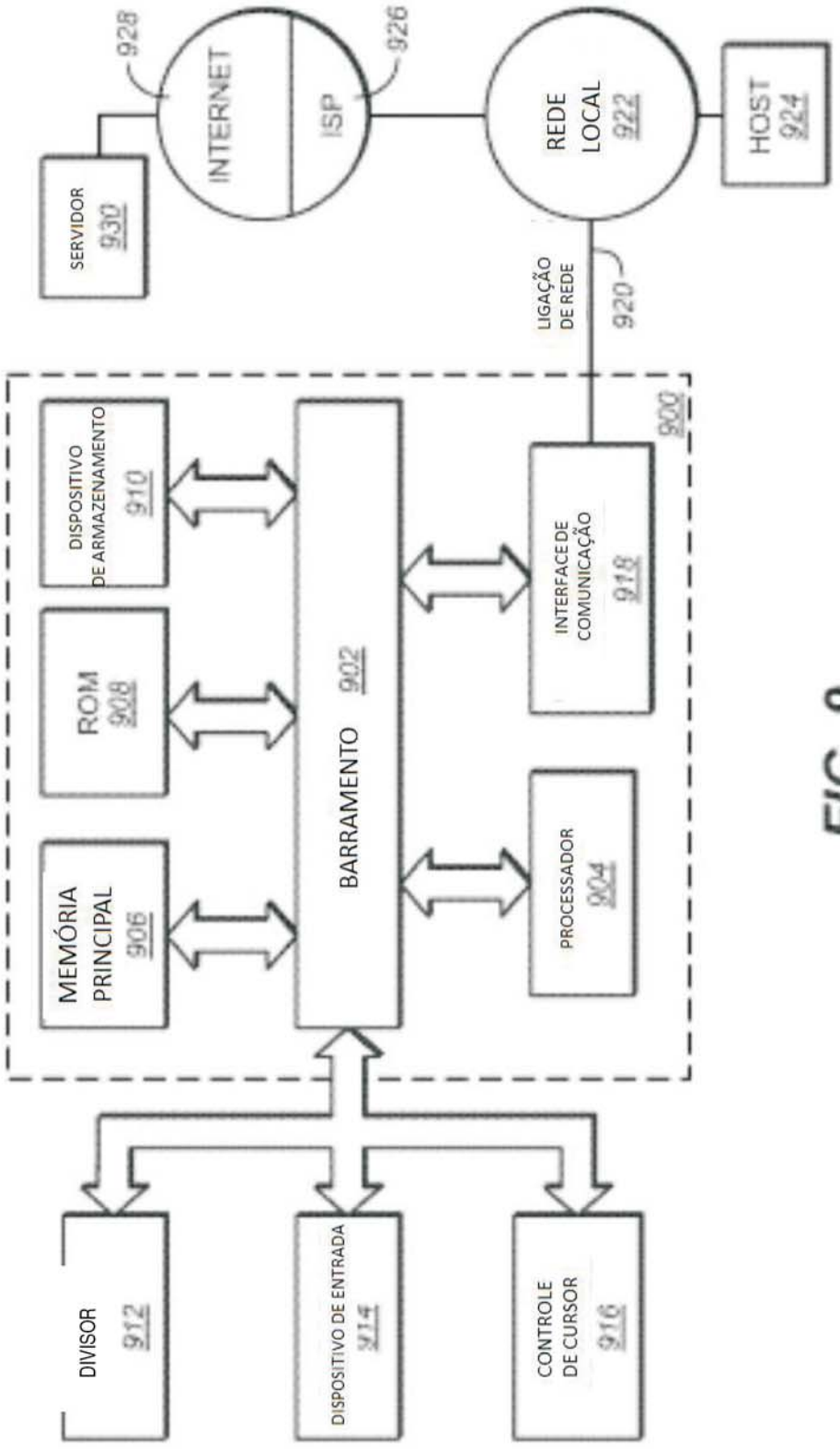

928
INTERNET
ISP
926
REDE LOCAL 922
HOST 924
SERVIDOR 930
LIGAÇÃO DE REDE 920
900
DISPOSITIVO DE ARMAZENAMENTO 910
ROM 908
MEMÓRIA PRINCIPAL 906
BARRAMENTO 902
INTERFACE DE COMUNICAÇÃO 918
PROCESSADOR 904
DIVISOR 912
DISPOSITIVO DE ENTRADA 914
CONTROLE DE CURSOR 916

FIG. 9