



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

(11) BR 112015017064-1 B1



(22) Data do Depósito: 15/01/2014

(45) Data de Concessão: 22/03/2022

(54) Título: MÉTODO, MEIO LEGÍVEL EM COMPUTADOR E APARELHO PARA OTIMIZAR O NÍVEL DE INTENSIDADE DO SOM E A FAIXA DINÂMICA ATRAVÉS DE DISPOSITIVOS DE REPRODUÇÃO DIFERENTES

(51) Int.Cl.: H03G 9/00.

(30) Prioridade Unionista: 05/04/2013 US 61/809,250; 16/05/2013 US 61/824,010; 21/01/2013 US 61/754,882.

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

(72) Inventor(es): JEFFREY RIEDMILLER; SCOTT GREGORY NORCROSS; KARL JONAS ROEDEN.

(86) Pedido PCT: PCT US2014011681 de 15/01/2014

(87) Publicação PCT: WO 2014/113471 de 24/07/2014

(85) Data do Início da Fase Nacional: 16/07/2015

(57) Resumo: SISTEMA E MÉTODO PARA OTIMIZAR O NÍVEL DE INTENSIDADE DO SOM E A FAIXA DINÂMICA ATRAVÉS DE DISPOSITIVOS DE REPRODUÇÃO DIFERENTES. A presente invenção refere-se a um método e um sistema para receber, em um fluxo de bits, metadados associados com dados de áudio, e analisar os metadados para determinar se um parâmetro de nível de intensidade do som para um primeiro grupo de dispositivos de áudio de reprodução está disponível no fluxo de bits. Em resposta à determinação que os parâmetros estão presentes para o primeiro grupo, o sistema usa os parâmetros e os dados de áudio para definir o áudio. Em resposta à determinação que os parâmetros de nível de intensidade do som não estão presentes para o primeiro grupo, o sistema analisa uma ou mais características do primeiro grupo, e determina o parâmetro com base em uma ou mais características.

Relatório Descritivo da Patente de Invenção para
**"MÉTODO, MEIO LEGÍVEL EM COMPUTADOR E APARELHO PARA
OTIMIZAR O NÍVEL DE INTENSIDADE DO SOM E A FAIXA
DINÂMICA ATRAVÉS DE DISPOSITIVOS DE REPRODUÇÃO
DIFERENTES".**

REFERÊNCIA CRUZADA A PEDIDOS DE PATENTE
RELACIONADOS

[001] O presente pedido de patente reivindica a prioridade para o Pedido de Patente Provisório U.S. nº. 61/754.882 depositado em 21 de janeiro de 2013; Pedido de Patente Provisório U.S. nº. 61/809.250 depositado em 05 de abril de 2013; e Pedido de Patente Provisório U.S. nº. 61/824.010 depositado em 16 de maio de 2013, todos incorporados a título de referência no presente documento.

CAMPO DA INVENÇÃO

[002] Uma ou mais modalidades referem-se de modo geral ao processamento de sinal de áudio, e mais especificamente ao processamento de fluxos de bits de dados de áudio com metadados indicativos das características do nível de intensidade do som e da faixa dinâmica do teor de áudio com base em ambientes e dispositivos da reprodução.

ANTECEDENTES

[003] O objeto discutido na seção Antecedentes não deve ser suposto como sendo a técnica anterior meramente em consequência de sua menção na seção Antecedente. Similarmente, um problema mencionado na seção Antecedentes ou associado com o objeto da seção Antecedentes não deve ser suposto como sendo reconhecido previamente na técnica anterior. O objeto na seção Antecedentes representa meramente as abordagens diferentes, nas quais e em que as mesmas também podem ser invenções.

[004] A faixa dinâmica de um sinal de áudio é de modo geral a

relação entre os maiores e menores valores possíveis do som embutidos no sinal, e é medida normalmente como um valor em decibel (base 10). Em muitos sistemas de processamento de áudio, o controle da faixa dinâmica (ou a compressão da faixa dinâmica, DRC) é usado para reduzir o nível de sons altos e/ou amplificar o nível de sons baixos para encaixar o teor largo da fonte de faixa dinâmica em uma faixa dinâmica gravada mais estreita que possa ser mais facilmente armazenada e reproduzida ao usar um equipamento eletrônico. Para o teor de áudio/visual (AV), um nível de referência de diálogo pode ser usado para definir o ponto "nulo" para a compressão através do mecanismo de DRC. A DRC age para intensificar o teor abaixo do nível de referência de diálogo e para cortar o teor acima do nível de referência.

[005] Em um sistema de codificação de áudio conhecido, os metadados associados com o sinal de áudio são usados para ajustar o nível de DRC com base no tipo e no uso pretendido do conteúdo. O modo de DRC ajusta a quantidade de compressão aplicada ao sinal de áudio e define o nível de referência de saída do decodificador. Tais sistemas podem ser limitados a dois ajustes do nível de DRC que são programados no codificador e selecionados pelo usuário. Por exemplo, um valor de dialnorm (normalização de diálogo) de -31 dB (Linha) é usado tradicionalmente para o conteúdo que é reproduzido em um AVR ou em dispositivos com capacidade de faixa dinâmica completa, e um valor de dialnorm de -20 dB (RF) é usado para o conteúdo reproduzido em aparelhos de televisão ou em dispositivos similares. Este tipo de sistema permite que um único fluxo de bits de áudio seja usado em dois cenários comuns, mas muito diferentes de reprodução através do uso de dois conjuntos diferentes de metadados de DRC. Tais sistemas, no entanto, são limitados aos valores de dialnorm pré-ajustados e não são otimizados para a reprodução na ampla variedade de diferentes

dispositivos de reprodução e ambientes de audição que são agora possíveis através do advento de mídia digital e na tecnologia de streaming baseada na Internet.

[006] Nos sistemas de codificação de áudio baseados em metadados atuais, um fluxo de dados de áudio pode incluir ambos o conteúdo de áudio (por exemplo, um ou mais canais de conteúdo de áudio) e os metadados indicativos de pelo menos uma característica do conteúdo de áudio. Por exemplo, em um fluxo de bits AC-3 há diversos parâmetros de metadados de áudio que são destinados especificamente para o uso na mudança do som do programa entregue a um ambiente de audição. Um dos parâmetros de metadados é o parâmetro de dialnorm, que indica o nível de intensidade do som médio do diálogo (ou o nível de intensidade do som médio do conteúdo) que ocorre em um programa de áudio, e é usado para determinar o nível de áudio do sinal de reprodução.

[007] Durante a reprodução de um fluxo de bits que compreende uma sequência de segmentos de programa de áudio diferentes (cada um dos quais tem um parâmetro de dialnorm diferente), um decodificador AC-3 usa o parâmetro de dialnorm de cada segmento para executar um tipo de processamento de nível de intensidade do som que modifica o nível de reprodução do segmento ou o nível de intensidade do som de maneira tal que o nível de intensidade do som percebido do diálogo do segmento encontra-se em um nível consistente. Cada segmento de áudio codificado (item) em uma sequência de itens de áudio codificados (em geral) tem um parâmetro diferente de dialnorm, e o decodificador deve escalar o nível de cada um dos itens de maneira tal que o nível de reprodução ou o nível de intensidade do som do diálogo para cada item seja o mesmo ou muito similar, embora isto possa requerer a aplicação de quantidades diferentes de ganho a itens diferentes durante a reprodução.

[008] Em algumas modalidades, o parâmetro de dialnorm é ajustado por um usuário, e não gerado automaticamente, embora haja um valor de dialnorm padrão caso nenhum valor seja ajustado pelo usuário. Por exemplo, um criador de conteúdo pode fazer medições do nível de intensidade do som com um dispositivo externo a um codificador AC-3 e então transferir o resultado (indicativo do nível de intensidade do som do diálogo falado de um programa de áudio) ao codificador para ajustar o valor de dialnorm. Desse modo, é preciso se basear no criador de conteúdo para ajustar corretamente o parâmetro de dialnorm.

[009] Há diversas razões diferentes pelas quais o parâmetro de dialnorm em um fluxo de bits AC-3 pode estar incorreto. Em primeiro lugar, cada codificador AC-3 tem um valor de dialnorm padrão que é usado durante a geração do fluxo de bits se um valor de dialnorm não for ajustado pelo criador de conteúdo. Este valor de defeito pode ser substancialmente diferente do que o nível de intensidade do som do diálogo real do áudio. Em segundo lugar, mesmo se um criador de conteúdo medir o nível de intensidade do som e ajustar o valor de dialnorm de modo correspondente, pode ter sido usado um algoritmo de medição do nível de intensidade do som ou medidor que não se conforma ao método recomendado da medição do nível de intensidade do som, o que resulta em um valor incorreto de dialnorm. Em terceiro lugar, mesmo se um fluxo de bits AC-3 for criado com o valor de dialnorm medido e ajustado corretamente pelo criador de conteúdo, ele pode ter sido mudado para um valor incorreto por um módulo intermediário durante a transmissão e/ou o armazenamento do fluxo de bits. Por exemplo, não é incomum em aplicações de transmissão de televisão que os fluxos de bits AC-3 sejam decodificados, modificados e então recodificados ao usar informações de metadados de dialnorm incorretas. Desse modo, um valor de dialnorm incluído em um fluxo de

bits AC-3 pode ser incorreto ou impreciso e, portanto, pode ter um impacto negativo na qualidade da experiência de audição.

[0010] Além disso, o parâmetro de dialnorm não indica o estado de processamento do nível de intensidade do som dos dados de áudio correspondentes (por exemplo, qual(is) tipo(s) de processamento de nível de intensidade do som foi executado nos dados de áudio). Além disso, sistemas de nível de intensidade de som e DRC presentemente empregados, tais como sistemas em sistemas Dolby Digital (DD) e Dolby Digital Plus (DD+), foram projetados para tornar o conteúdo de AV em uma sala de estar de um consumidor ou em cinema. Para adaptar tal conteúdo para a reprodução em outros ambientes e no equipamento de audição (por exemplo, um dispositivo móvel), o pós-processamento deve ser aplicado 'às cegas' no dispositivo de reprodução para adaptar o conteúdo de AV para esse ambiente de audição. Em outras palavras, um pós-processador (ou um decodificador) supõe que o nível de intensidade do som do conteúdo recebido encontra-se em um nível particular (por exemplo, -31 ou -20 dB) e o pós-processador ajusta o nível a um nível alvo fixo predeterminado apropriado para um dispositivo particular. Se o nível de intensidade do som suposto ou o nível alvo predeterminado estiver incorreto, o pós-processamento pode ter o oposto de seu efeito pretendido; isto é, o pós-processamento pode tornar o áudio de saída menos desejável para um usuário.

[0011] As modalidades divulgadas não são limitadas ao uso com um fluxo de bits AC-3, um fluxo de bits E-AC-3 ou um fluxo de bits Dolby E, no entanto, para fins de conveniência tais fluxos de bits serão discutidos conjuntamente com um sistema que inclui os metadados do estado de processamento do nível de intensidade do som. Dolby, Dolby Digital, Dolby Digital Plus e Dolby E são marcas registradas da Dolby Laboratories Licensing Corporation. A Dolby Laboratories fornece

implementações proprietárias de AC-3 e E-AC3 conhecidos como Dolby Digital e Dolby Digital Plus, respectivamente.

BREVE SUMÁRIO DAS MODALIDADES

[0012] As modalidades são dirigidas a um método para decodificar dados de áudio, mediante a recepção de um fluxo de bits que contém os metadados associados com os dados de áudio, e a análise dos metadados no fluxo de bits para determinar se um parâmetro do nível de intensidade do som para um primeiro grupo de dispositivos de reprodução de áudio está disponível no fluxo de bits. Em resposta à determinação que os parâmetros estão presentes para o primeiro grupo, um componente de processamento usa os parâmetros e os dados de áudio para definir o áudio. Em resposta à determinação que os parâmetros do nível de intensidade do som não estão presentes para o primeiro grupo, o componente de processamento analisa uma ou mais características do primeiro grupo, e determina o parâmetro com base em uma ou mais características. O método também pode usar os parâmetros e os dados de áudio para definir o áudio mediante a transmissão do parâmetro e dos dados de áudio a um módulo a jusante que define o áudio para reprodução. O parâmetro e os dados de áudio também podem ser usados para obter o áudio mediante a definição dos dados de áudio com base no parâmetro e nos dados de áudio.

[0013] Em uma modalidade, o método também compreende a determinação de um dispositivo de saída que deve definir o fluxo de áudio recebido, e a determinação se ou o dispositivo de saída pertence ou não ao primeiro grupo de dispositivos de reprodução de áudio; em que a etapa de análise dos metadados no fluxo para determinar se o parâmetro do nível de intensidade do som para o primeiro grupo de dispositivos de reprodução de áudio está disponível é executada depois da etapa de determinação se o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio. Em uma modalidade, a

etapa de determinação se o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio compreende: a recepção de uma indicação de um módulo conectado ao dispositivo de saída que indica uma identidade do dispositivo de saída ou que indica uma identidade de um grupo de dispositivos que inclui o dispositivo de saída, e a determinação se o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio com base na indicação recebida.

[0014] As modalidades também são dirigidas a um aparelho ou um sistema que inclui componentes de processamento que executam as ações descritas nas modalidades do método de codificação acima.

[0015] As modalidades também são ainda dirigidas a um método de decodificação de dados de áudio mediante a recepção dos dados e de áudio e dos metadados associados com os dados de áudio, a análise dos metadados no fluxo de bits para determinar se a informação do nível de intensidade do som associada com os parâmetros do nível de intensidade do som para um primeiro grupo de dispositivos de áudio está disponível no fluxo e, em resposta à determinação que a informação do nível de intensidade do som está presente para o primeiro grupo, a determinação da informação do nível de intensidade do som a partir do fluxo, e a transmissão dos dados de áudio e da informação do nível de intensidade do som para o uso na definição do áudio, ou, se a informação do nível de intensidade do som não estiver presente para o primeiro grupo, a determinação da informação do nível de intensidade do som associada com um perfil de saída, e a transmissão da informação do nível de intensidade do som determinada para o perfil de saída para o uso na definição do áudio. Em uma modalidade, a etapa de determinação da informação do nível de intensidade do som associada com um perfil de saída também pode incluir a análise das características do perfil de saída, a determinação dos parâmetros com base nas características, e a transmissão da

informação do nível de intensidade do som determinada compreende a transmissão dos parâmetros determinados. A informação do nível de intensidade do som pode incluir parâmetros do nível de intensidade do som para ou as características de um perfil de saída. Em uma modalidade, o método também pode compreender a determinação de um fluxo codificado com taxa de bits baixa a ser transmitido, em que a informação do nível de intensidade do som compreende características para um ou mais perfis de saída.

[0016] As modalidades também são dirigidas a um aparelho ou um sistema que inclui componentes de processamento que executam as ações descritas nas modalidades do método de decodificação acima.

BREVE DESCRIÇÃO DOS DESENHOS

[0017] Nos desenhos a seguir os mesmos números de referência são usados para indicar os mesmos elementos. Embora as seguintes figuras ilustrem vários exemplos, as implementações descritas no presente documento não são limitadas aos exemplos ilustrados nas figuras.

[0018] A FIG. 1 é um diagrama de blocos de uma modalidade de um sistema de processamento de áudio configurado para executar a otimização do nível de intensidade do som e da faixa dinâmica, sob algumas modalidades.

[0019] A FIG. 2 é um diagrama de blocos de um codificador para o uso no sistema da FIG. 1, sob algumas modalidades.

[0020] A FIG. 3 é um diagrama de blocos de um decodificador para o uso no sistema da FIG. 1, sob algumas modalidades.

[0021] A FIG. 4 é um diagrama de um quadro AC-3, incluindo os segmentos nos quais é dividido.

[0022] A FIG. 5 é um diagrama do segmento de Informação da Sincronização (SI) de um quadro AC-3, incluindo os segmentos nos quais é dividido.

[0023] A FIG. 6 é um diagrama do segmento de Informação de Fluxo de Bits (BSI) de um quadro AC-3, incluindo os segmentos nos quais é dividido.

[0024] A FIG. 7 é um diagrama de um quadro E-AC-3, incluindo os segmentos nos quais é dividido.

[0025] A FIG. 8 é uma tabela que ilustra determinados quadros de um fluxo de bits codificados e o formato dos metadados, sob algumas modalidades.

[0026] A FIG. 9 é uma tabela que ilustra um formato dos metadados do estado de processamento do nível de intensidade do som, sob algumas modalidades.

[0027] A FIG. 10 é um diagrama de blocos mais detalhado do sistema de processamento de áudio da FIG. 1 que pode ser configurado para executar a otimização do nível de intensidade do som e da faixa dinâmica, sob algumas modalidades.

[0028] A FIG. 11 é uma tabela que ilustra requisitos de faixas dinâmicas diferentes para uma variedade de dispositivos de reprodução e de ambientes de audição de fundo em um caso de uso exemplificador.

[0029] A FIG. 12 é um diagrama de blocos de um sistema de otimização da faixa dinâmica, sob uma modalidade.

[0030] A FIG. 13 é um diagrama de blocos que ilustra uma interface entre perfis diferentes para uma variedade de classes de dispositivo de reprodução diferentes, sob uma modalidade.

[0031] A FIG. 14 é uma tabela que ilustra a correlação entre o nível de intensidade do som de longa duração e a faixa dinâmica de curta duração para uma pluralidade de perfis definidos, sob uma modalidade.

[0032] A FIG. 15 ilustra exemplos de perfis do nível de intensidade do som para tipos diferentes de conteúdo de áudio, sob uma modalidade.

[0033] A FIG. 16 é um fluxograma que ilustra um método de

otimização do nível de intensidade do som e da faixa dinâmica através de dispositivos e aplicações de reprodução, sob uma modalidade.

DESCRIÇÃO DETALHADA

Definições e Nomenclatura

[0034] Por toda esta divulgação, inclusive nas concretizações, a expressão que executa uma operação "em" um sinal ou dados (por exemplo, filtragem, escalação, transformação, ou aplicação de ganho a, de sinal ou dados) é usada em um sentido amplo para denotar a execução da operação diretamente no sinal ou nos dados, ou em uma versão processada do sinal ou dos dados (por exemplo, em uma versão do sinal que foi submetido a uma filtragem preliminar ou pré-processamento antes da execução da operação no mesmo). A expressão "sistema" é usada em um sentido amplo para denotar um dispositivo, um sistema, ou um subsistema. Por exemplo, um subsistema que implementa um decodificador pode ser indicado como um sistema de decodificador, e um sistema que inclui tal subsistema (por exemplo, um sistema que gera M sinais de saída em resposta a múltiplas entradas, em que o subsistema gera M das entradas e as outras $X - M$ entradas são recebidas de uma fonte externa) também pode ser indicado como um sistema de decodificador. O termo "processador" é usado em um sentido amplo para denotar um sistema ou um dispositivo programável ou então configurável (por exemplo, com software ou firmware) para executar operações em dados (por exemplo, dados de áudio, ou vídeo ou outros dados de imagem). Os exemplos dos processadores incluem uma disposição de porta programável em campo (ou outro circuito integrado ou conjunto de chips configurável), um processador de sinal digital programado e/ou então configurado para executar um processamento canalizado em dados de áudio ou outros dados de som, um processador ou computador de finalidade geral programável, e um chip ou conjunto de chips de microprocessador

programável.

[0035] As expressões "processador de áudio" e "unidade de processamento de áudio" são usadas intercambiavelmente e, em um sentido amplo, para denotar um sistema configurado para processar dados de áudio. Os exemplos de unidades de processamento de áudio incluem, mas sem ficar a eles limitadas, codificadores (por exemplo, transcodificadores), decodificadores, codecs, sistemas de pré-processamento, sistemas de pós-processamento, e sistemas de processamento de fluxo de bits (indicados às vezes como ferramentas de processamento de fluxo de bits). A expressão "metadados de estado de processamento" (por exemplo, tal como na expressão "metadados de estado de processamento de nível de intensidade do som") refere-se à separação e dados diferentes dos dados de áudio correspondentes (o conteúdo de áudio de um fluxo de dados de áudio que também inclui os metadados de estado de processamento). Os metadados de estado de processamento são associados com os dados de áudio, indicam o estado de processamento do nível de intensidade do som dos dados de áudio correspondentes (por exemplo, qual(is) tipo(s) de processamento já foi executado nos dados de áudio), e opcionalmente também indica pelo menos uma peculiaridade ou característica dos dados de áudio. Em alguma modalidade, a associação dos metadados de estado de processamento com os dados de áudio é temporalmente síncrona. Desse modo, os presentes metadados de estado de processamento (recebidos ou atualizados mais recentemente) indicam que os dados de áudio correspondentes compreendem contemporaneamente os resultados do(s) tipo(s) indicado(s) de processamento de dados de áudio. Em alguns casos, o metadados de estado de processamento podem incluir o histórico de processamento e/ou alguns ou todos os parâmetros que são usados em e/ou derivados dos tipos de processamento indicados. Além disso, os metadados de estado de

processamento podem incluir pelo menos uma peculiaridade ou característica dos dados de áudio correspondentes, que foram computados ou extraídos dos dados de áudio. Os metadados de estado de processamento também podem incluir outros metadados que não estejam relacionados nem sejam derivados de qualquer processamento dos dados de áudio correspondentes. Por exemplo, dados de terceiros, informações de rastreamento, identificadores, informações proprietárias ou padrão, dados de anotação do usuário, dados de preferência do usuário, etc., podem ser adicionados por uma unidade de processamento de áudio particular para serem designados para outras unidades de processamento de áudio.

[0036] A expressão "metadados de estado de processamento de nível de intensidade do som" (ou "LPSM") denota os metadados de estado de processamento indicativos do estado de processamento do nível de intensidade do som dos dados de áudio correspondentes (por exemplo, qual(is) tipo(s) de processamento de nível de intensidade do som foi(foram) executado(s) nos dados de áudio) e opcionalmente também pelo menos uma peculiaridade ou característica (por exemplo, o nível de intensidade do som) dos dados de áudio correspondentes. Os metadados de estado de processamento do nível de intensidade do som podem incluir dados (por exemplo, outros metadados) que não são (isto é, quando considerados sozinhos) metadados de estado de processamento do nível de intensidade do som. O termo "acopla" ou "acoplado" é usado para se referir a uma conexão direta ou indireta.

[0037] Os sistemas e os métodos são descritos para um codificador/decodificador de áudio que normalize de maneira não destrutiva o nível de intensidade do som e a faixa dinâmica do áudio através de vários dispositivos que requerem ou usam valores diferentes do nível de intensidade do som alvo e têm capacidades de faixa dinâmica diferentes. Os métodos e os componentes funcionais de

acordo com algumas modalidades enviam a informação sobre o conteúdo de áudio do codificador ao decodificador para um ou mais perfis do dispositivo. Um perfil de dispositivo especifica o nível de intensidade do som alvo desejado e a faixa dinâmica para um ou mais dispositivos. O sistema é extensível, de maneira tal que novos perfis do dispositivo com alvos do nível de intensidade do som "nominais" diferentes podem ser suportados.

[0038] Em uma modalidade, o sistema gera os ganhos apropriados com base em requisitos de controle de nível de intensidade do som e faixa dinâmica no codificador ou gera os ganhos no decodificador, sob o controle do codificador através da parameterização dos ganhos originais para reduzir a taxa de dados. O sistema de faixa dinâmica inclui dois mecanismos para implementar o controle do nível de intensidade do som: um perfil de faixa dinâmica artístico que propicia aos criadores de conteúdo um controle sobre como o áudio será reproduzido, e um mecanismo separado de proteção para assegurar que não ocorra sobrecarga para os vários perfis de reprodução. O sistema também é configurado para permitir que outros parâmetros (internos ou externos) dos metadados sejam usados para controlar corretamente o nível de intensidade do som e os ganhos e/ou os perfis de faixa dinâmica. O decodificador é configurado para suportar uma entrada auxiliar de n canais que vai nivelar os arranjos/processamento do nível de intensidade do som e da faixa dinâmica no lado do decodificador.

[0039] Em algumas modalidades, os metadados de estado de processamento do nível de intensidade do som (LPSM) são embutidos em um ou mais campos reservados (ou entalhes) de segmentos de metadados de um fluxo de bits de áudio que também inclui dados de áudio em outros segmentos (segmentos de dados de áudio). Por exemplo, pelo menos um segmento de cada quadro do fluxo de bits inclui LPSM, e pelo menos um outro segmento do quadro inclui dados

de áudio correspondentes (isto é, os dados de áudio cujo estado de processamento do nível de intensidade do som e nível de intensidade do som é indicado pelo LPSM). Em algumas modalidades, o volume dos dados de LPSM pode ser suficientemente pequeno para ser carregado sem afetar a taxa de bits alocada para carregar os dados de áudio.

[0040] A comunicação dos metadados do estado de processamento do nível de intensidade do som em uma cadeia de processamento de dados de áudio é particularmente útil quando duas ou mais unidades de processamento de áudio precisam trabalhar em tandem uma com a outra através da cadeia de processamento (ou ciclo de vida do conteúdo). Sem a inclusão dos metadados do estado de processamento do nível de intensidade do som em um fluxo de bits de áudio, podem ocorrer problemas de processamento de mídia tais como a qualidade, o nível e as degradações espaciais, por exemplo, quando dois ou mais codecs de áudio são utilizados na cadeia e a nivelção de volume de uma só extremidade é aplicada mais de uma vez durante o curso do fluxo de bits para um dispositivo de com sumo de mídia (ou um ponto de definição do conteúdo de áudio do fluxo de bits).

Sistema de processamento de metadados do nível de intensidade do som e da faixa dinâmica

[0041] A FIG. 1 é um diagrama de blocos de uma modalidade de um sistema de processamento de áudio que pode ser configurado para executar a otimização do nível de intensidade do som e da faixa dinâmica, sob algumas modalidades ao usar alguns componentes de processamento de metadados (por exemplo, pré-processamento e pós-processamento) componentes. A FIG. 1 ilustra uma cadeia de processamento de áudio exemplificadora (um sistema de processamento de dados de áudio), em que um ou mais dos elementos do sistema podem ser configurados de acordo com uma modalidade da presente invenção. O sistema 10 da FIG. 1 inclui os seguintes

elementos, acoplados uns aos outros, tal como mostrado: uma unidade de pré-processamento 12, um codificador 14, uma unidade de análise de sinal e correção de metadados 16, um transcodificador 18, um decodificador 20, e uma unidade de pós-processamento 24. Nas variações no sistema mostrado, um ou mais dos elementos são omitidos, ou unidades de processamento de dados de áudio adicionais são incluídas. Por exemplo, em uma modalidade, a unidade de pós-processamento 22 faz parte do decodificador 20 em vez de ser uma unidade separada.

[0042] Em algumas implementações, a unidade de pré-processamento da FIG. 1 é configurada para aceitar amostras de PCM (domínio do tempo) que compreendem o conteúdo de áudio como entrada 11, e para emitir amostras de PCM processadas. O codificador 14 pode ser configurado para aceitar as amostras de PCM como entrada e para emitir um fluxo de bits de áudio (por exemplo, comprimido) codificado indicativo do conteúdo de áudio. Os dados do fluxo de bits que são indicativos do conteúdo de áudio são indicados às vezes no presente documento como "dados de áudio". Em uma modalidade, a saída de áudio do fluxo de bits do codificador inclui os metadados de estado de processamento do nível de intensidade do som (e opcionalmente também outros metadados), bem como dados de áudio.

[0043] A unidade de análise de sinal e correção de metadados 16 pode aceitar um ou mais fluxos de bits de áudio codificados como entrada e determinar (por exemplo, validar) se o metadados de estado de processamento em cada fluxo de bits de áudio codificado está correto, ao executar a análise do sinal. Em algumas modalidades, a validação pode ser executada por um componente validador de estado, tal como o elemento 102 mostrado na FIG. 2, e tal técnica de validação é descrita a seguir no contexto de validador de estado 102. Em algumas modalidades, a unidade 16 é incluída no codificador e a validação é feita

pela unidade 16 ou pelo validador 102. Se a unidade de análise de sinal e correção de metadados verificar que os metadados incluídos são inválidos, a unidade de correção de metadados 16 executa a análise do sinal para determinar o(s) valor(es) correto(s) e substitui o(s) valor(es) incorreto pelo(s) valor(s) correto(s) determinado(s). Desse modo, cada fluxo de bits de áudio codificado emitido da unidade de análise de sinal e correção de metadados pode incluir os metadados de estado de processamento, bem como os dados de áudio codificados. A unidade de análise de sinal e correção de metadados 16 pode fazer parte da unidade de pré-processamento 12, do codificador 14, do transcodificador 18, do decodificador 20, ou da unidade de pós-processamento 22. Alternativamente, a unidade de análise de sinal e correção de metadados 16 pode ser uma unidade separada ou parte de uma outra unidade na cadeia de processamento de áudio.

[0044] O transcodificador 18 pode aceitar fluxos de bits de áudio codificados como entrada, e emitir fluxos de bits de áudio modificados (por exemplo, codificados distintamente) em resposta (por exemplo, ao decodificar um fluxo de entrada e ao recodificar o fluxo decodificado em um formato de codificação diferente). A saída de áudio do fluxo de bits do transcodificador inclui os metadados de estado de processamento do nível de intensidade do som (e opcionalmente também outros metadados), bem como dados de áudio codificados. Os metadados podem ter sido incluídos no fluxo de bits.

[0045] O decodificador 20 da FIG. 1 pode aceitar fluxos de bits de áudio codificados (por exemplo, comprimidos) como entrada, e emitir (em resposta) fluxos de amostras de áudio PCM decodificadas. Em uma modalidade, a saída do decodificador é ou inclui qualquer um dos seguintes: um fluxo de amostras de áudio, e um fluxo de metadados do estado de processamento do nível de intensidade do som correspondente (e opcionalmente também outros metadados) extraídos

de um fluxo de bits codificado de entrada; um fluxo de amostras de áudio, e um fluxo correspondente de bits de controle determinado a partir dos metadados de estado de processamento do nível de intensidade do som (e opcionalmente também outros metadados) extraídos de um fluxo de bits codificado de entrada; ou um fluxo de amostras de áudio, sem um fluxo correspondente de metadados de estado de processamento ou bits de controle determinados a partir dos metadados de estado de processamento. Neste último caso, o decodificador pode extrair os metadados de estado de processamento do nível de intensidade do som (e/ou outros metadados) do fluxo de bits codificado de entrada e executar pelo menos uma operação nos metadados extraídos (por exemplo, validação), mesmo que não emita os metadados extraídos ou os bits de controle determinados a partir dos mesmos.

[0046] Com a configuração da unidade de pós-processamento da FIG. 1 de acordo com uma modalidade da presente invenção, a unidade de pós-processamento 22 é configurada para aceitar um fluxo de amostras de áudio PCM decodificadas, e para executar o pós-processamento na mesma (por exemplo, a nivelção do volume do conteúdo de áudio) ao usar os metadados de estado de processamento do nível de intensidade do som (e opcionalmente também outros metadados) recebidos com as amostras, ou os bits de controle (determinados pelo decodificador dos metadados de estado de processamento do nível de intensidade do som e opcionalmente também outros metadados) recebidos com as amostras. A unidade de pós-processamento 22 também é opcionalmente configurada para definir o conteúdo de áudio pós-processado para a reprodução por um ou mais alto-falantes. Estes alto-falantes podem ser incorporados em qualquer um de uma variedade de dispositivos ou itens de audição diferentes do equipamento da reprodução, tais como computadores,

televisores, sistemas estereofônicos (em casa ou no cinema), telefones móveis, e outros dispositivos de reprodução portáteis. Os alto-falantes podem ser de qualquer tamanho e taxa de potência apropriados, e podem ser providos na forma de drivers autônomos, invólucros de alto-falantes, sistemas de som do tipo surround, colunas de som, fones de ouvido, fones auriculares, e assim por diante.

[0047] Algumas modalidades provêm uma cadeia de processamento de áudio realçada em que as unidades de processamento de áudio (por exemplo, codificadores, decodificadores, transcodificadores, e unidades de pré- e pós-processamento) adaptam seu processamento respectivo a ser aplicado aos dados de áudio de acordo com um estado contemporâneo dos dados de mídia tal como indicado pelos metadados de estado de processamento do nível de intensidade do som recebidos respectivamente pelas unidades de processamento de áudio. A entrada de dados de áudio 11 para qualquer unidade de processamento de áudio do sistema 100 (por exemplo, o codificador ou o transcodificador da FIG. 1) pode incluir os metadados de estado de processamento do nível de intensidade do som (e opcionalmente também outros metadados), bem como dados de áudio (por exemplo, dados de áudio codificados). Estes metadados podem ter sido incluídos no áudio de entrada por um outro elemento ou por uma outra fonte de acordo com algumas modalidades. A unidade de processamento que recebe o áudio de entrada (com metadados) pode ser configurada para executar pelo menos uma operação nos metadados (por exemplo, validação) ou em resposta aos metadados (por exemplo, processamento adaptável do áudio de entrada), e também para incluir opcionalmente em seu áudio da saída os metadados, uma versão processada dos metadados, ou os bits de controle determinados a partir dos metadados.

[0048] Uma modalidade da unidade de processamento de áudio (ou

do processador de áudio) é configurada para executar o processamento adaptável dos dados de áudio com base no estado dos dados de áudio tal como indicado pelos metadados de estado de processamento do nível de intensidade do som que corresponde aos dados de áudio. Em algumas modalidades, o processamento adaptável é (ou inclui) o processamento de nível de intensidade do som (se os metadados indicam que o processamento de nível de intensidade do som, ou um processamento similar a este, já não foi executado nos dados de áudio, mas não é (e não inclui) o processamento de nível de intensidade do som (se os metadados indicam que tal processamento de nível de intensidade do som, ou processamento similar a este, já foi executado nos dados de áudio). Em algumas modalidades, o processamento adaptável deve ou inclui a validação dos metadados (por exemplo, executados em uma subunidade de validação de metadados) para assegurar que a unidade de processamento de áudio execute outro processamento adaptável dos dados de áudio com base no estado dos dados de áudio tal como indicado pelos metadados de estado de processamento do nível de intensidade do som. Em algumas modalidades, a validação determina a confiabilidade dos metadados de estado de processamento do nível de intensidade do som associados com (por exemplo, incluídos em um fluxo de bits com) os dados de áudio. Por exemplo, se os metadados forem validados de modo a serem confiáveis, então os resultados de um tipo de processamento de áudio previamente executado podem ser reutilizados e o desempenho adicional do mesmo tipo de processamento de áudio pode ser evitado. Por outro lado, se for verificado que os metadados foram alterados (ou então não são confiáveis), então o tipo de processamento de mídia previamente executado propositadamente (tal como indicado pelos metadados não confiáveis) pode ser repetido pela unidade de processamento de áudio, e/ou outro processamento pode ser

executado pela unidade de processamento de áudio nos metadados e/ou nos dados de áudio. A unidade de processamento de áudio também pode ser configurada para sinalizar a outras unidades de processamento de áudio a jusante em uma cadeia de processamento de mídia realçada que os metadados de estado de processamento do nível de intensidade do som (por exemplo, presentes em um fluxo de bits dos meios) são válidos, se a unidade determinar que os metadados de estado de processamento são válidos (por exemplo, com base em uma combinação de um valor criptográfico extraído e um valor criptográfico de referência).

[0049] Para a modalidade da FIG. 1, o componente de pré-processamento 12 pode fazer parte do codificador 14, e o componente de pós-processamento 22 pode fazer parte do decodificador 22. Alternativamente, o componente de pré-processamento 12 pode ser incorporado em um componente funcional que seja separado do codificador 14. Similarmente, o componente de pós-processamento 22 pode ser incorporado em um componente funcional que seja separado do decodificador 20.

[0050] A FIG. 2 é um diagrama de blocos de um codificador 100 que pode ser usado conjuntamente com o sistema 10 da FIG. 1. Qualquer um dos componentes ou elementos do codificador 100 pode ser implementado como um ou mais processos e/ou um ou mais circuitos (por exemplo, ASICs, FPGAs, ou outros circuitos integrados), em hardware, software, ou em uma combinação de hardware e software. O codificador 100 compreende o armazenamento temporário de quadros 110, o analisador gramatical 111, o decodificador 101, o validador de estado de áudio 102, o estágio de processamento de nível de intensidade do som 103, o estágio de seleção de fluxo de áudio 104, o codificador 105, o estágio de estufador/formatador, o estágio de geração de metadados 106, o subsistema de medição de nível de intensidade

do som de diálogo 108, e o armazenamento temporário de quadro 109, conectados tal como mostrado. Ainda opcionalmente, o codificador 100 inclui outros elementos de processamento (não mostrados). O codificador 100 (que é um transcodificador) é configurado para converter um fluxo de bits de áudio de entrada (que, por exemplo, pode ser um dentre um fluxo de bits AC-3, um fluxo de bits E-AC-3, ou um fluxo de bits Dolby E) em um fluxo de bits de áudio codificado de saída (que, por exemplo, pode ser outro dentre um fluxo de bits AC-3, um fluxo de bits E-AC-3, ou um fluxo de bits Dolby E) inclusive pela execução do processamento do de nível de intensidade do som adaptável e automatizado ao usar os metadados de estado de processamento do nível de intensidade do som incluídos no fluxo de bits de entrada. Por exemplo, o codificador 100 pode ser configurado para converter um fluxo de bits Dolby E de entrada (um formato usado tipicamente nas instalações de produção e de transmissão, mas não nos dispositivos do consumidor que recebem programas de áudio que foram transmitidos aos mesmos) em um fluxo de bits de áudio codificado de saída (apropriado para a transmissão aos dispositivos de consumidor) no formato AC-3 ou E-AC-3.

[0051] O sistema da FIG. 2 também inclui o subsistema de entrega de áudio codificado 150 (que armazena e/ou entrega os fluxos de bits codificados emitidos do codificador 100) e o decodificador 152. Um fluxo de bits de áudio codificado emitido do codificador 100 pode ser armazenado pelo subsistema 150 (por exemplo, na forma de um DVD ou um disco de BluRay), ou transmitido pelo subsistema 150 (que pode implementar um link ou uma rede de transmissão), ou pode ser armazenado e transmitido pelo subsistema 150. O decodificador 152 é configurado para decodificar um fluxo de bits de áudio codificado (gerado pelo codificador 100) que ele recebe através do subsistema 150, inclusive pela extração dos metadados de estado de

processamento do nível de intensidade do som (LPSM) de cada quadro do fluxo de bits, e pela geração de dados de áudio decodificados. Em uma modalidade, o decodificador 152 é configurado para executar o processamento de nível de intensidade do som adaptável nos dados de áudio decodificados ao usar LPSM, e/ou para encaminhar os dados de áudio decodificados e o LPSM a um pós-processador configurado para executar o processamento de nível de intensidade do som adaptável nos dados de áudio decodificados ao usar o LPSM. Opcionalmente, o decodificador 152 inclui um armazenamento temporário, o qual armazena (por exemplo, de uma maneira não transitória) o fluxo de bits de áudio codificado recebido do subsistema 150.

[0052] Várias implementações do codificador 100 e do decodificador 152 são configuradas para executar modalidades diferentes descritas no presente documento. O armazenamento temporário de quadro 110 é uma memória de armazenamento temporário acoplada para receber um fluxo de bits de áudio codificado de entrada. Em operação, o armazenamento temporário 110 armazena (por exemplo, de uma maneira não transitória) pelo menos um quadro do fluxo de bits de áudio codificado, e uma sequência de quadros do fluxo de bits de áudio codificado é designada do armazenamento temporário 110 ao analisador gramatical 111. O analisador gramatical 111 é acoplado e configurado para extrair os metadados de estado de processamento do nível de intensidade do som (LPSM) e outros metadados de cada quadro de áudio codificado de entrada, para passar pelo menos o LPSM ao validador de estado de áudio 102 ao estágio de processamento de nível de intensidade do som 103, ao estágio 106 e ao subsistema 108, para extrair dados de áudio do áudio de entrada codificado, e para passar os dados de áudio ao decodificador 101. O decodificador 101 do codificador 100 é configurado para decodificar os dados de áudio para gerar dados de áudio decodificados, e para passar

os dados de áudio decodificados ao estágio de processamento de nível de intensidade do som 103, ao estágio de seleção de fluxo de áudio 104, ao subsistema 108, e opcionalmente também ao validador de estado 102.

[0053] O validador de estado 102 é configurado para autenticar e validar o LPSM (e opcionalmente outros metadados) designados a este. Em algumas modalidades, o LPSM é (ou é incluído em) um bloco de dados que é incluído no fluxo de bits de entrada (por exemplo, de acordo com uma modalidade da presente invenção). O bloco pode compreender um indicador criptográfico (um código de autenticação da mensagem de mensagem à base de indicador ou um "HMAC") para processar o LPSM (e opcionalmente também outros metadados) e/ou os dados de áudio subjacentes (fornecido pelo decodificador 101 ao validador 102). O bloco de dados pode ser digitalmente assinado nestas modalidades, de modo que uma unidade de processamento de áudio a jusante possa autenticar e validar de maneira relativamente fácil os metadados de estado de processamento.

[0054] Por exemplo, o HMAC é usado para gerar uma compilação, e o(s) valor(es) de proteção incluídos no fluxo de bits da invenção pode(m) incluir a compilação. A compilação pode ser gerada tal como segue para um quadro de AC-3: (1) Depois que os dados de AC-3 e LPSM são codificados, os bytes de dados de quadro (concatenated frame_data #1 e frame_data #2) e os bytes de dados de LPSM são usados como entrada para o HMAC de função de indicação. Outros dados, que podem estar presentes dentro de um campo auxdata, não são levados em consideração para calcular a compilação. Tais outros dados podem ser bytes que não pertencem aos dados de AC-3 nem aos dados de LPSM. Os bits de proteção incluídos em LPSM podem não ser considerados para calcular a compilação de HMAC. (2) Depois que a compilação é calculada, ela é gravada no fluxo de bits em um campo

reservado para bits de proteção. (3) A última etapa da geração de quadro de AC-3 completo é o cálculo da verificação de CRC. Este é gravado bem na extremidade do quadro e todos os dados que pertencem a este quadro são levados em consideração, incluindo os bits de LPSM.

[0055] Outros métodos criptográficos que incluem, mas sem ficar a eles limitados, qualquer um ou mais métodos criptográficos que não de HMAC podem ser usados para a validação de LPSM (por exemplo, no validador 102) para assegurar a transmissão e a recepção seguras de LPSM e/ou dos dados de áudio subjacentes. Por exemplo, a validação (ao usar tal método criptográfico) pode ser executada em cada unidade de processamento de áudio que recebe uma modalidade do fluxo de bits de áudio para determinar se os metadados de estado de processamento do nível de intensidade do som e os dados de áudio correspondentes incluídos no fluxo de bits foram submetidos (e/ou resultaram de) ao processamento de nível de intensidade do som específico (tal como indicado pelos metadados) e não foram modificados após a execução de tal processamento de nível de intensidade do som específico.

[0056] O validador de estado 102 designa dados de controle ao estágio de seleção de fluxo de áudio 104, ao gerador de metadados 106 e ao subsistema de medição do nível de intensidade do som de diálogo 108, para indicar os resultados da operação de validação. Em resposta aos dados de controle, o estágio 104 pode selecionar (e para repassar ao codificador 105) tanto: (1) a saída de maneira adaptável processada do estágio de processamento do nível de intensidade do som 103 (por exemplo, quando os LPSMs indicarem que os dados de áudio provenientes do decodificador 101 não foram submetidos a um tipo específico de processamento de nível de intensidade do som, e os bits de controle do validador 102 indicar que os LPSMs são válidos); quanto

(2) os dados de áudio provenientes do decodificador 101 (por exemplo, quando os LPSMs indicarem que os dados de áudio provenientes do decodificador 101 já foram submetidos ao tipo específico de processamento do nível de intensidade do som que deve ser executado pelo estágio 103, e os bits de controle do validador 102 indicarem que os LPSMs são válidos). Em uma modalidade, o estágio de processamento do nível de intensidade do som 103 corrige o nível de intensidade do som para o alvo específico e a faixa do nível de intensidade do som.

[0057] O estágio 103 do codificador 100 é configurado para executar o processamento de nível de intensidade do som adaptável nos dados de áudio decodificados provenientes do decodificador 101, com base em uma ou mais características de áudio dos dados indicadas por LPSM extraído pelo decodificador 101. O estágio 103 pode ser um processador de controle do nível de intensidade do som em tempo real do domínio de transformação adaptável e da faixa dinâmica. O estágio 103 pode receber a entrada do usuário (por exemplo, valores de nível de intensidade do som/faixa dinâmica alvo do usuário ou valores de dialnorm), ou outros metadados de entrada (por exemplo, um ou mais tipos de dados de terceiros, informações de rastreamento, identificadores, informação proprietária ou padrão, dados de anotação do usuário, dados da preferência do usuário, etc.) e/ou outras entradas (por exemplo, de um processo de formação de impressão digital), e usa tal entrada para processar os dados de áudio decodificada proveniente do decodificador 101.

[0058] O subsistema de medição do nível de intensidade do som de diálogo 108 pode operar para determinar o nível de intensidade do som dos segmentos de áudio decodificado (do decodificador 101) que são indicativos de diálogo (ou uma outra fala), por exemplo, ao usar o LPSM (e/ou outros metadados) extraídos pelo decodificador 101, quando os

bits de controle do validador 102 indicam que os LPSM são inválidos. A operação do subsistema de medição do nível de intensidade do som de diálogo 108 pode ser desabilitada quando os LPSMs indicam um nível de intensidade do som previamente determinado de segmentos de diálogo (ou outra fala) do áudio decodificado (do decodificador 101) quando os bits do controle do validador 102 indicam que os LPSMs são válidos.

[0059] Existem ferramentas úteis (por exemplo, o medidor do nível de intensidade do som Dolby LM100) para medir o nível de diálogo no conteúdo de áudio de maneira conveniente e fácil. Algumas modalidades de APU (por exemplo, o estágio 108 do codificador 100) são implementados para incluir (ou para desempenhar as funções de) tal ferramenta para medir o nível de intensidade do som de diálogo médio do conteúdo de áudio de um fluxo de bits de áudio (por exemplo, um fluxo de bits AC-3 decodificado designado para o estágio 108 do decodificador 101 ao codificador 100). Se o estágio 108 for implementado para medir o nível de intensidade do som de diálogo médio confiável de dados de áudio, a medição pode incluir uma etapa de isolamento dos segmentos do conteúdo de áudio que contêm predominantemente a fala. Os segmentos de áudio que são predominantemente uma fala são processados então de acordo com um algoritmo de medição do nível de intensidade do som. Para os dados de áudio decodificados de um fluxo de bits AC-3, este algoritmo pode ser uma medida do nível de intensidade do som K-ponderada padrão (de acordo com a norma internacional ITU-R BS.1770). Alternativamente, outras medidas do nível de intensidade do som podem ser usadas (por exemplo, aquelas baseados em modelos psicoacústico do nível de intensidade do som).

[0060] O isolamento de segmentos de fala não é essencial para medir o nível de intensidade do som de diálogo médio de dados de

áudio. No entanto, ele melhora a exatidão da medida e propicia resultados mais satisfatórios da perspectiva de um ouvinte. Devido ao fato que nem todo o conteúdo de áudio contém diálogo (fala), a medida do nível de intensidade do som de todo o conteúdo de áudio pode fornecer uma aproximação suficiente do nível de diálogo do áudio, caso a fala tenha estado presente.

[0061] O gerador de metadados 106 gera os metadados a serem incluídos pelo estágio 107 no fluxo de bits codificado a ser emitido do codificador 100. O gerador de metadados 106 pode passar através do estágio 107 o LPSM (e/ou outros metadados) extraídos pelo codificador 101 (por exemplo, quando os bits de controle do validador 102 indicarem que os LPSMs e/ou outros metadados são válidos), ou gerar um novo LPSM (e/ou outros metadados) e designar os novos metadados ao estágio 107 (por exemplo, quando os bits de controle do validador 102 indicarem que os LPSM e/ou outros metadados extraídos pelo decodificador 101 são inválidos, ou pode designar ao estágio 107 uma combinação de metadados extraída pelo decodificador 101 e os metadados recentemente gerados. O gerador de metadados 106 pode incluir os dados do nível de intensidade do som gerados pelo subsistema 108, e pelo menos um valor indicativo do tipo de processamento de nível de intensidade do som executado pelo subsistema 108, no LPSM designa ao estágio 107 para a inclusão no fluxo de bits codificado a ser emitido do codificador 100. O gerador de metadados 106 pode gerar os bits de proteção (que podem consistir em ou incluir um código de autenticação de mensagem à base de indicador ou "HMAC") úteis para pelo menos uma dentre a descritografia, a autenticação ou a validação do LPSM (e opcionalmente também de outros metadados) a ser incluído no fluxo de bits codificado e/ou nos dados de áudio subjacentes a ser incluídos no fluxo de bits codificado. O gerador de metadados 106 pode fornecer tais bits de proteção ao

estágio 107 para a inclusão no fluxo de bits codificado.

[0062] Em uma modalidade, o subsistema de medição do nível de intensidade do som de diálogo 108 processa a saída de dados de áudio do decodificador 101 para gerar em resposta a isso os valores do nível de intensidade do som (por exemplo, valores do nível de intensidade do som de diálogo conectados e não conectados) e valores da faixa dinâmica. Em resposta a estes valores, o gerador de metadados 106 pode gerar os metadados do estado de processamento do nível de intensidade do som (LPSM) para a inclusão (pelo estufador/formatador 107) no fluxo de bits codificado a ser emitido do codificador 100. Em uma modalidade, o nível de intensidade do som pode ser calculado com base nas técnicas especificadas pelas normas ITU-R BS.1770-1 e ITU-R BS.1770-2, ou outras normas de medição do nível de intensidade do som similares. O nível de intensidade do som conectado pode ser um nível de intensidade do som conectado de diálogo ou um nível de intensidade do som conectado relativo, ou uma combinação destes tipos de nível de intensidade do som conectados, e o sistema pode empregar blocos de conexão apropriados dependendo dos requisitos da aplicação e das restrições do sistema.

[0063] Além disso, opcionalmente, ou então alternativamente, os subsistemas de 106 e/ou 108 do codificador 100 podem executar a análise adicional dos dados de áudio para gerar os metadados indicativos de pelo menos uma característica dos dados de áudio para a inclusão no fluxo de bits codificado a ser emitido do estágio 107. O codificador 105 codifica (por exemplo, ao executar a compressão no mesmo) a saída de dados de áudio do estágio de seleção 104, e designa o áudio codificado ao estágio 107 para a inclusão no fluxo de bits codificado a ser emitido do estágio 107.

[0064] O estágio 107 multiplexa o áudio codificado do codificador 105 e os metadados (incluindo LPSM) do gerador 106 para gerar o fluxo

de bits codificado a ser emitido do estágio 107, de modo que o fluxo de bits codificado tenha o formato tal como especificado por uma modalidade. O armazenamento temporário de quadro 109 é uma memória de armazenamento temporário que armazena (por exemplo, de uma maneira não transitória) pelo menos um quadro do fluxo de bits de áudio codificado proveniente do estágio 107, e uma sequência de quadros de fluxo de bits de áudio codificado é designada então do armazenamento temporário 109 como saída do codificador 100 ao sistema de entrega 150.

[0065] O LPSM gerado pelo gerador de metadados 106 e incluído no fluxo de bits codificado pelo estágio 107 é indicativo do estado de processamento do nível de intensidade do som dos dados de áudio correspondentes (por exemplo, qual(is) tipo(s) de processamento do nível de intensidade do som foi(foram) executado(s) nos dados de áudio) e do nível de intensidade do som (por exemplo, o nível de intensidade do som de diálogo medido, o nível de intensidade do som conectado e/ou não conectado, e/ou a faixa dinâmica) dos dados de áudio correspondentes. No presente documento, a "conexão" do nível de intensidade do som e/ou as medições de nível executadas nos dados de áudio refere-se a um nível específico ou limite do nível de intensidade do som onde o(s) valor(es) computado(s) que excedem o limite é(são) incluído(s) na medição final (por exemplo, ignorando os valores do nível de intensidade do som de curta duração abaixo de -60 dBFS nos valores medidos finais). A conexão em um valor absoluto refere-se a um nível fixo ou nível de intensidade do som, ao passo que a conexão em um valor relativo refere-se a um valor que é dependente de um valor de medição "não conectado" atual.

[0066] Em algumas implementações do codificador 100, o fluxo de bits codificado armazenado na memória 109 (e emitido ao sistema de entrega 150) é um fluxo de bits AC-3 ou um fluxo de bits E-AC-3, e

compreende os segmentos de dados de áudio (por exemplo, os segmentos AB0-AB5 do quadro mostrado na FIG. 4) e os segmentos de metadados, onde os segmentos de dados de áudio são indicativos de dados de áudio, e cada um de pelo menos alguns dos segmentos de metadados incluem os metadados do estado de processamento do nível de intensidade do som (LPSM). O estágio 107 insere LPSM no fluxo de bits no formato seguinte. Cada um dos segmentos de metadados que inclui LPSM é incluído em um campo "addbsi" do segmento de Informação de Fluxo de Bits ("BSI") de um quadro de fluxo de bits, ou em um campo auxdata (por exemplo, o segmento AUX mostrado na Fig. 4) na extremidade de um quadro do fluxo de bits.

[0067] Um quadro do fluxo de bits pode incluir um ou dois segmentos de metadados, cada um dos quais inclui LPSM e, se o quadro incluir dois segmentos de metadados, um deles está presente no campo addbsi do quadro e o outro no campo AUX do quadro. Cada segmento de metadados que inclui LPSM inclui um segmento de carga útil de LPSM (ou de recipiente) que tem o seguinte formato: um cabeçalho (por exemplo, incluindo uma palavra de sincronismo que identifica o começo da carga útil de LPSM, seguida por pelo menos um valor de identificação, por exemplo, a versão do formato de LPSM, o comprimento, o período, a contagem e os valores da associação de subfluxo indicados na Tabela 2 a seguir); e, depois do cabeçalho, pelo menos um valor da indicação do diálogo (por exemplo, o parâmetro "Canal(is) de diálogo" da Tabela 2) que indica se os dados de áudio correspondentes indicam o diálogo ou não indicam o diálogo (por exemplo, quais os canais de dados de áudio correspondentes que indicam o diálogo); pelo menos um valor de conformidade de regulação do nível de intensidade do som (por exemplo, o parâmetro "Tipo de Regulação de Nível de Intensidade do Som" da Tabela 2) que indica se os dados de áudio correspondentes estão em conformidade com um

conjunto indicado de regulações do nível de intensidade do som; pelo menos um valor de processamento do nível de intensidade do som (por exemplo, um ou mais parâmetros "sinalizador de Correção do Nível de Intensidade do Som conectado de Diálogo", "Tipo de Correção do Nível de Intensidade do Som", da Tabela 2) que indica pelo menos um tipo de processamento do nível de intensidade do som que foi executado nos dados de áudio correspondentes; e pelo menos um valor do nível de intensidade do som (por exemplo, um ou mais dos parâmetros "Nível de Intensidade do Som Conectado Relativo ITU", "Nível de Intensidade do Som Conectado de Fala ITU", "Nível de Intensidade do Som 3s de Curta duração ITU (EBU 3341)" e "Pico real" da Tabela 2) que indica pelo menos uma característica do nível de intensidade do som (por exemplo, o nível de intensidade do som de pico ou médio) dos dados de áudio correspondentes.

[0068] Em algumas implementações, cada um dos segmentos de metadados inseridos pelo estágio 107 em um campo "addbsi" ou em um campo auxdata de um quadro de fluxo de bits tem o seguinte formato: um cabeçalho de núcleo (por exemplo, incluindo uma palavra de sincronismo que identifica o começo do segmento de metadados, seguida por valores de identificação, por exemplo, a versão do elemento do núcleo, o comprimento, e o período, a contagem estendida do elemento, e os valores de associação de subfluxo indicados na Tabela 1 a seguir); e, depois do cabeçalho do núcleo, pelo menos um valor de proteção (por exemplo, a compilação de HMAC e os valores de Impressão Digital de áudio da Tabela 1) útil para pelo menos uma dentre a descritografia, a autenticação ou a validação de pelo menos um dos metadados do estado de processamento do nível de intensidade do som ou dos dados de áudio correspondentes); e, também depois do cabeçalho do núcleo, se o segmento de metadados incluir LPSM, a identificação de carga útil de LPSM ("ID") e os valores de tamanho de

carga útil de LPSM que identificam os metadados seguintes como uma carga útil de LPSM e indicam o tamanho da carga útil de LPSM.

[0069] O segmento de carga útil de LPSM (ou recipiente) (por exemplo, que tem o formato especificado acima) segue a ID de carga útil ID de LPSM e os valores do tamanho de carga útil de LPSM.

[0070] Em algumas modalidades, cada um dos segmentos de metadados no campo auxdata (ou no campo do "addbsi") de um quadro tem três níveis de estrutura: uma estrutura de alto nível, incluindo um sinalizador que indica se o campo auxdata (ou addbsi) inclui os metadados, pelo menos um valor de ID que indica qual(is) tipo(s) de metadados está(estão) presente(s), e opcionalmente também um valor que indica quantos bits de metadados (por exemplo, de cada tipo) estão presentes (se os metadados estiverem presentes). Um tipo de metadados que poderia estar presente é LSPM, e um outro tipo de metadados que poderia estar presente consiste nos metadados de pesquisa de mídia (por exemplo, metadados de pesquisa de mídia de Nielsen); uma estrutura de nível intermediário, a qual compreende um elemento de núcleo para cada tipo identificado de metadados (por exemplo, cabeçalho de núcleo, valores de proteção, ID de carga útil de LPSM e valores de tamanho de carga útil de LPSM, tal como mencionado acima, para cada tipo identificado de metadados); e uma estrutura de nível baixo, a qual compreende cada carga útil para um elemento de núcleo (por exemplo, uma carga útil de LPSM, se uma for identificada pelo elemento de núcleo como estando presente, e/ou uma carga útil de metadados de um outro tipo, se uma for identificada pelo elemento de núcleo como estando presente).

[0071] Os valores dos dados em uma estrutura de três níveis podem ser aninhados. Por exemplo, o(s) valor(es) de proteção para uma carga útil de LPSM e/ou uma outra carga útil de metadados identificados por um elemento do núcleo podem ser incluídos depois de cada carga útil

identificada pelo elemento de núcleo (e desse modo depois do cabeçalho de núcleo do elemento de núcleo). Em um exemplo, um cabeçalho de núcleo pode identificar uma carga útil de LPSM e uma outra carga útil de metadados, a ID de carga útil e os valores de tamanho de carga útil para a primeira carga útil (por exemplo, a carga útil de LPSM) podem seguir o cabeçalho de núcleo, a própria primeira carga útil pode seguir a ID e os valores de tamanho, a ID de carga útil e o valor de tamanho de carga útil para a segunda carga útil podem seguir a primeira carga útil, a própria segunda carga útil pode seguir essa ID e valores de tamanho, e os bits de proteção para ambas as cargas úteis (ou para valores do elemento de núcleo e ambas as cargas úteis) podem seguir a última carga útil.

[0072] Em algumas modalidades, se o decodificador 101 receber um fluxo de bits de áudio gerado de acordo com uma modalidade da invenção com um indicador criptográfico, o decodificador é configurado para analisar gramaticalmente e recuperar o indicador criptográfico de um bloco de dados determinado a partir do fluxo de bits, em que o dito bloco compreende os metadados do estado de processamento do nível de intensidade do som (LPSM). O validador 102 pode usar o indicador criptográfico para validar o fluxo de bits recebido e/ou os metadados associados. Por exemplo, o validador 102 verifica que o LPSM é válido com base em uma combinação entre um indicador criptográfico de referência e o indicador criptográfico recuperado do bloco dos dados, e então pode desabilitar a operação do processador 103 nos dados de áudio correspondentes e fazer com que o estágio de seleção 104 passe através (inalterado) dos dados de áudio. Além disso, opcionalmente, ou então alternativamente, outros tipos de técnicas criptográficas podem ser usados no lugar de um método baseado em um indicador criptográfico.

[0073] O codificador 100 da FIG. 2 pode determinar (em resposta

ao LPSM extraído pelo decodificador 101) que uma unidade de pós/pré-processamento executou um tipo de processamento do nível de intensidade do som nos dados de áudio a serem codificados (nos elementos 105, 106, e 107) e desse modo pode criar (no gerador 106) os metadados do estado de processamento do nível de intensidade do som que incluem os parâmetros específicos usados em e/ou derivados do processamento do nível de intensidade do som previamente executado. Em algumas implementações, o codificador 100 pode criar (e incluir no fluxo de bits codificado emitido do mesmo) os metadados do estado de processamento indicativo do histórico de processamento no conteúdo de áudio contanto que o codificador esteja ciente dos tipos de processamento que foram executados no conteúdo de áudio.

[0074] A FIG. 3 é um diagrama de blocos de um decodificador que pode ser usado conjuntamente com o sistema 10 da FIG. 1. Qualquer um dos componentes ou elementos do decodificador 200 e do pós-processador 300 pode ser implementado como um ou mais processos e/ou um ou mais circuitos (por exemplo, ASICs, FPGAs, ou outros circuitos integrados), em hardware, software ou uma combinação de hardware e software. O decodificador 200 compreende o armazenamento temporário de quadro 201, o analisador gramatical 205, o decodificador de áudio 202, o estágio de validação de estado de áudio (validador) 203, e o estágio de geração de bits de controle 204, conectados tal como mostrado. O decodificador 200 pode incluir outros elementos de processamento (não mostrados). O armazenamento temporário de quadro 201 (uma memória de armazenamento temporário) armazena (por exemplo, de uma maneira não transitória) pelo menos um quadro do fluxo de bits de áudio codificado recebido pelo decodificador 200. Uma sequência de quadros do fluxo de bits de áudio codificado é designada do armazenamento temporário 201 ao analisador gramatical 205. O analisador gramatical 205 é acoplado e

configurado para extrair os metadados do estado de processamento do nível de intensidade do som (LPSM) e outros metadados de cada quadro de áudio de entrada codificado, para designar pelo menos o LPSM ao validador de estado de áudio 203 e ao estágio 204, para designar o LPSM como saída (por exemplo, para o pós-processador 300), para extrair dados de áudio do áudio de entrada codificado, e para designar os dados de áudio extraídos ao decodificador 202. O fluxo de bits de áudio codificado inserido no decodificador 200 pode ser um dentre um fluxo de bits AC-3, um fluxo de bits E-AC-3 ou um fluxo de bits Dolby E.

[0075] O sistema da FIG. 3 também inclui o pós-processador 300. O pós-processador 300 compreende o armazenamento temporário de quadro 301 e outros elementos de processamento (não mostrados) que incluem pelo menos um elemento de processamento acoplado ao armazenamento temporário 301. O armazenamento temporário de quadro 301 armazena (por exemplo, de uma maneira não transitória) pelo menos um quadro de fluxo de bits de áudio decodificado recebido do decodificador 200 pelo pós-processador 300. Os elementos de processamento do pós-processador 300 são acoplados e configurados para receber e processar de maneira adaptável uma sequência de quadros do fluxo de bits de áudio decodificado proveniente do armazenamento temporário 301, ao usar metadados (incluindo valores de LPSM) provenientes do decodificador 202 e/ou bits de controle provenientes do estágio 204 do decodificador 200. Em uma modalidade, o pós-processador 300 é configurado para executar o processamento de nível de intensidade do som adaptável nos dados de áudio decodificados ao usar os valores de LPSM (por exemplo, com base no estado de processamento do nível de intensidade do som, e/ou uma ou mais características de áudio dos dados, indicadas por LPSM). Várias implementações do decodificador 200 e do pós-processador 300 são

configuradas para executar modalidades diferentes dos métodos de acordo com as modalidades descritas no presente documento.

[0076] O decodificador de áudio 202 do decodificador 200 é configurado para decodificar os dados de áudio extraídos pelo analisador gramatical 205 para gerar dados de áudio decodificados, e para designar os dados de áudio decodificados como saída (por exemplo, ao pós-processador 300). O validador de estado 203 é configurado para autenticar e validar o LPSM (e opcionalmente outros metadados) designados ao mesmo. Em algumas modalidades, o LPSM é (ou é incluído em) um bloco de dados que é incluído no fluxo de bits de entrada (por exemplo, de acordo com uma modalidade da presente invenção). O bloco pode compreender um indicador criptográfico (um código de autenticação de mensagem à base de indicador ou "HMAC") para processar o LPSM (e opcionalmente também outros metadados) e/ou os dados de áudio subjacentes (fornecidos pelo analisador gramatical 205 e/ou decodificador 202 ao validador 203). O bloco de dados pode ser digitalmente assinado nestas modalidades, de modo que uma unidade de processamento de áudio a jusante possa autenticar e validar de maneira relativamente fácil os metadados do estado de processamento.

[0077] Outros métodos criptográficos que incluem, mas sem qualquer limitação, qualquer um dentre um ou mais métodos criptográficos que não de HMAC podem ser usados para a validação de LPSM (por exemplo, no validador 203) para assegurar a transmissão e a recepção seguras do LPSM e/ou dos dados de áudio subjacentes. Por exemplo, a validação (ao usar tal método criptográfico) pode ser executada em cada unidade de processamento de áudio que recebe uma modalidade do fluxo de bits de áudio da invenção para determinar se os metadados do estado de processamento do nível de intensidade do som e os dados de áudio correspondentes incluídos no fluxo de bits

foram submetidos (e/ou resultam de) ao processamento de nível de intensidade do som específico (tal como indicado pelos metadados) e não foram modificados após a execução de tal processamento de nível de intensidade do som específico.

[0078] O validador de estado 203 designa dados de controle para controlar o gerador de bits 204, e/ou designa os dados de controle como saída (por exemplo, ao pós-processador 300), para indicar os resultados da operação de validação. Em resposta aos dados de controle (e opcionalmente também a outros metadados extraídos do fluxo de bits de entrada), o estágio 204 pode gerar (e designar ao pós-processador 300) qualquer um de: bits de controle que indicam que os dados de áudio decodificada provenientes do decodificador 202 foram submetidos a um tipo específico de processamento de nível de intensidade do som (quando os LPSM indicarem que os dados de áudio provenientes do decodificador 202 foram submetidos ao tipo específico de processamento de nível de intensidade do som, e os bits de controle de validador 203 indicarem que os LPSM são válidos); ou os bits de controle que indicam que os dados de áudio decodificados provenientes do decodificador 202 devem ser submetidos a um tipo específico de processamento de nível de intensidade do som (por exemplo, quando os LPSM indicarem que os dados de áudio provenientes do decodificador 202 não foram submetidos ao tipo específico de processamento de nível de intensidade do som, ou quando os LPSM indicarem que os dados de áudio provenientes do decodificador 202 foram submetidos ao tipo específico de processamento de nível de intensidade do som, mas os bits de controle de validador 203 indicarem que os LPSM são inválidos).

[0079] Alternativamente, o decodificador 200 designa o LPSM (e quaisquer outros metadados) extraído pelo decodificador 202 do fluxo de bits de entrada ao pós-processador 300, e o pós-processador 300

executa o processamento de nível de intensidade do som nos dados de áudio decodificados ao usar o LPSM, ou executa a validação do LPSM e executa então o processamento de nível de intensidade do som nos dados de áudio decodificados ao usar o LPSM se a validação indicar que os LPSM são válidos.

[0080] Em algumas modalidades, se o decodificador 201 receber um fluxo de bits de áudio gerado de acordo com uma modalidade da invenção com indicador criptográfico, o decodificador é configurado para analisar gramaticalmente e recuperar o indicador criptográfico de um bloco de dados determinado a partir do fluxo de bits, em que o dito bloco compreende os metadados do estado de processamento do nível de intensidade do som (LPSM). O validador 203 pode usar o indicador criptográfico para validar o fluxo de bits recebido e/ou os metadados associados. Por exemplo, se o validador 203 verificar que o LPSM é válido com base em uma combinação entre um indicador criptográfico de referência e o indicador criptográfico recuperado do bloco de dados, então pode sinalizar a uma unidade de processamento de áudio a jusante (por exemplo, o pós-processador 300, que pode ser ou incluir uma unidade de nivelção de volume) para passar através (inalterado) dos dados de áudio do fluxo de bits. Além disso, opcionalmente, ou então alternativamente, outros tipos de técnicas criptográficas podem ser usados no lugar de um método baseado em um indicador criptográfico.

[0081] Em algumas implementações do decodificador 100, o fluxo de bits codificado recebido (e armazenado na memória 201) é um fluxo de bits AC-3 ou um fluxo de bits E-AC-3, e compreende os segmentos de dados de áudio (por exemplo, os segmentos AB0-AB5 do quadro mostrado na Fig. 4) e os segmentos dos metadados, onde os segmentos de dados de áudio são indicativos de dados de áudio, e cada um de pelo menos alguns dos segmentos de metadados inclui os

metadados do estado de processamento do nível de intensidade do som (LPSM). O estágio do decodificador 202 é configurado para extrair do fluxo de bits LPSM que tem o seguinte formato. Cada um dos segmentos de metadados que inclui LPSM é incluído em um campo "addbsi" do segmento de Informação de Fluxo de Bits ("BSI") de um quadro do fluxo de bits, ou em um campo auxdata (por exemplo, o segmento AUX mostrado na Fig. 4) na extremidade de um quadro do fluxo de bits. Um quadro do fluxo de bits pode incluir um ou dois segmentos de metadados, cada um dos quais inclui LPSM e, se o quadro incluir dois segmentos de metadados, um deles está presente no campo addbsi do quadro e o outro no campo AUX do quadro. Cada segmento de metadados que inclui LPSM inclui um segmento de carga útil de LPSM (ou recipiente) que tem o seguinte formato: um cabeçalho (por exemplo, incluindo uma palavra de sincronismo que identifica o começo da carga útil de LPSM, seguida por valores da identificação, por exemplo, a versão do formato de LPSM, o comprimento, o período, a contagem, e os valores da associação do subfluxo indicados na Tabela 2 abaixo); e, depois do cabeçalho, pelo menos um valor de indicação de diálogo (por exemplo, o parâmetro "Canal(is) de diálogo" da Tabela 2) que indica se os dados de áudio correspondentes indicam o diálogo ou não indicam o diálogo (por exemplo, quais os canais de dados de áudio correspondentes que indicam o diálogo); pelo menos um valor de conformidade de regulação do nível de intensidade do som (por exemplo, o parâmetro "Tipo de Regulação do Nível de Intensidade do Som" da Tabela 2) que indica se os dados de áudio correspondentes se conformam com um conjunto indicado de regulamentos do nível de intensidade do som; pelo menos um valor de processamento do nível de intensidade do som (por exemplo, um ou mais dos parâmetros "sinalizador de Correção do Nível de Intensidade do Som conectado de Diálogo", "Tipo de Correção de Nível de Intensidade do Som", da Tabela

2) que indica pelo menos um tipo de processamento do nível de intensidade do som que foi executado nos dados de áudio correspondentes; e pelo menos um valor do nível de intensidade do som (por exemplo, um ou mais dos parâmetros "Nível de Intensidade do Som Conectado Relativo ITU", "Nível de Intensidade do Som Conectado de Fala ITU", "Nível de Intensidade do Som 3s de Curta duração ITU (EBU 3341)" e "Pico real" da Tabela 2) que indica pelo menos uma característica do nível de intensidade do som (por exemplo, nível de intensidade do som de pico ou médio) dos dados de áudio correspondentes.

[0082] Em algumas implementações, o estágio do decodificador 202 é configurado para extrair, do campo "addbsi" ou de um campo auxdata de um quadro do fluxo de bits, cada segmento de metadados que tem o seguinte formato: um cabeçalho de núcleo (por exemplo, incluindo uma palavra de sincronismo que identifica o começo do segmento de metadados, seguida por pelo menos um valor de identificação, por exemplo, a versão do elemento do núcleo, o comprimento, e o período, a contagem estendida do elemento, e os valores da associação do subfluxo indicados na Tabela 1 a seguir); e, depois do cabeçalho do núcleo, pelo menos um valor de proteção (por exemplo, a compilação de HMAC e os valores de áudio de impressão digital da Tabela 1) útil para pelo menos uma dentre a descryptografia, a autenticação, ou a validação de pelo menos um dentre os metadados do estado de processamento do nível de intensidade do som ou os dados de áudio correspondentes); e também, depois do cabeçalho de núcleo, se o segmento de metadados incluir a identificação da carga útil de LPSM ("ID") e os valores de tamanho da carga útil de LPSM que identificam os metadados seguintes como uma carga útil de LPSM e indicam o tamanho da carga útil de LPSM. O segmento de carga útil de LPSM (ou recipiente) (por exemplo, que tem o formato especificado

acima) segue a ID da carga útil de LPSM e os valores do tamanho da carga útil ID de LPSM.

[0083] De modo mais geral, o fluxo de bits de áudio codificado gerado por uma modalidade tem uma estrutura que provê um mecanismo para etiquetar os elementos e os subelementos como de núcleo (compulsório) ou expandidos (elementos opcionais). Isto permite que a taxa de dados do fluxo de bits (incluindo seus metadados) seja escalonada através de numerosas aplicações. Os elementos de núcleo (compulsórios) da sintaxe do fluxo de bits também devem poder sinalizar que os elementos expandidos (opcionais) associados com o conteúdo de áudio estão presentes (na faixa) e/ou em um local remoto (fora da faixa).

[0084] Em alguma modalidade, o(s) elemento(s) de núcleo precisa(m) estar presente(s) em cada quadro do fluxo de bits. Alguns subelementos de elementos de núcleo são opcionais e podem estar presentes em qualquer combinação. Os elementos expandidos não precisam estar presentes em cada quadro (para limitar a taxa de bits de baixo desempenho). Desse modo, os elementos expandidos podem estar presentes em alguns quadros e não em outros. Alguns subelementos de um elemento expandido são opcionais e podem estar presentes em qualquer combinação, ao passo que alguns subelementos de um elemento expandido podem ser compulsórios (isto é, se o elemento expandido estiver presente em um quadro do fluxo de bits).

[0085] Em algumas modalidades, um fluxo de bits de áudio codificado que compreende uma sequência de segmentos de dados de áudio e de segmentos de metadados é gerado (por exemplo, por uma unidade de processamento de áudio que incorpore a invenção). Os segmentos de dados de áudio são indicativos de dados de áudio, cada um de pelo menos alguns dos segmentos de metadados inclui os

metadados do estado de processamento do nível de intensidade do som (LPSM), e os segmentos de dados de áudio são multiplexados com divisão de tempo com os segmentos de metadados. Em algumas modalidades nesta classe, cada um dos segmentos de metadados tem um formato a ser descrito no presente documento. Em um formato, o fluxo de bits codificado é um fluxo de bits AC-3 ou um fluxo de bits E-AC-3, e cada um dos segmentos de metadados que inclui LPSM é incluído (por exemplo, pelo estágio 107 do codificador 100) como informação do fluxo de bits adicional no campo "addbsi" (mostrado na FIG. 6) do segmento da Informação de Fluxo de Bits ("BSI") de um quadro do fluxo de bits, ou em um campo auxdata de um quadro do fluxo de bits. Cada um dos quadros inclui um elemento de núcleo no campo addbsi do quadro que tem o formato mostrado na Tabela 1 da FIG. 8.

[0086] Em um formato, cada um dos campos addbsi (ou auxdata) que contém LPSM contém um cabeçalho de núcleo (e opcionalmente também elementos de núcleo adicionais), e, depois do cabeçalho de núcleo (ou cabeçalho de núcleo e outros elementos de núcleo), os seguintes valores de LPSM (parâmetros): uma ID de carga útil (que identifica os metadados como LPSM) depois dos valores do elemento de núcleo (por exemplo, tal como especificado na Tabela 1); um tamanho da carga útil (que indica o tamanho da carga útil de LPSM) que segue a ID da carga útil; e dados de LPSM (seguindo a ID da carga útil e o valor do tamanho da carga útil) que têm o formato tal como indicado na Tabela 2 da FIG. 9.

[0087] Em um segundo formato de um fluxo de bits codificado, o fluxo de bits é um fluxo de bits AC-3 ou um fluxo de bits E-AC-3, e cada um dos segmentos de metadados que inclui LPSM é incluído (por exemplo, pelo estágio 107 do codificador 100) em qualquer um de: o campo "addbsi" (mostrado na Fig. 6) do segmento de Informação de

Fluxo de Bits ("BSI") de um quadro do fluxo de bits; ou um campo auxdata (por exemplo, o segmento AUX mostrado na Fig. 4) na extremidade de um quadro do fluxo de bits. Um quadro pode incluir um ou dois segmentos de metadados, cada um dos quais inclui LPSM e, se o quadro incluir dois segmentos de metadados, um está presente no campo addbsi do quadro e o outro no campo AUX do quadro. Cada segmento de metadados que inclui LPSM tem o formato especificado acima com referência às Tabelas 1 e 2 acima (isto é, inclui os elementos de núcleo especificados na Tabela 1, seguidos pela ID da carga útil (que identifica os metadados como LPSM) e pelos valores do tamanho da carga útil especificados acima, seguidos pela carga útil (os dados de LPSM que têm o formato tal como indicado na Tabela 2).

[0088] Em outro formato, o fluxo de bits codificado é um fluxo de bits Dolby E, e cada um dos segmentos de metadados que inclui LPSM consiste nos primeiros locais de N amostras do intervalo da faixa de proteção Dolby E. Um fluxo de bits Dolby E que inclui tal segmento de metadados, que inclui LPSM, por exemplo, inclui um valor indicativo do comprimento da carga útil de LPSM sinalizado na palavra Pd do preâmbulo SMPTE 337M (a taxa de repetição da palavra SMPTE 337M Pa pode continuar idêntica à taxa de quadros de vídeo associada).

[0089] Em um formato em que o fluxo de bits codificado é um fluxo de bits E-AC-3, cada um dos segmentos de metadados que inclui LPSM é incluído (por exemplo, pelo estágio 107 do codificador 100) como informação de fluxo de bits adicional no campo "addbsi" do segmento de Informação de Fluxo de Bits ("BSI") de um quadro do fluxo de bits. Os aspectos adicionais da codificação de um fluxo de bits E-AC-3 com LPSM neste formato são descritos tal como segue: (1) durante a geração de um fluxo de bits E-AC-3, quando o codificador E-AC-3 (que insere os valores de LPSM no fluxo de bits) estiver "ativo", para cada quadro (quadro de sincronismo) gerado, o fluxo de bits deve incluir um

bloco de metadados (incluindo LPSM) carregado no campo addbsi do quadro. Os bits requeridos para carregar o bloco de metadados não devem aumentar a taxa de bits do codificador (comprimento do quadro);

(2) cada bloco de metadados (que contém LPSM) deve conter as seguintes informações: loudness_correction_type_flag [sinalizador_de tipo_de correção do_nível de intensidade do som]: onde '1' indica que o nível de intensidade do som dos dados de áudio correspondentes foi corrigido a montante do codificador, e '0' indica que o nível de intensidade do som foi corrigido por um corretor do nível de intensidade do som embutido no codificador (por exemplo, o processador do nível de intensidade do som 103 do codificador 100 da Fig. 2); speech_channel [canal_de fala]: indica que o(s) canal(is) de origem contém(em) a fala (no 0,5 segundo precedente). Se nenhuma fala for detectada, isto será então indicado tal como; speech_loudness [nível de intensidade do som_da fala]: indica o nível de intensidade do som integrado da fala de cada canal de áudio correspondente que contém a fala (no 0,5 segundo precedente); ITU_loudness [nível de intensidade do som_ITU]: indica o nível de intensidade do som de ITU BS.1770-2 integrado de cada canal de áudio correspondente; ganho: o(s) ganho(s) composto do nível de intensidade do som para a reversão em um decodificador (para demonstrar a reversibilidade).

[0090] Embora o codificador de E-AC-3 (que insere os valores de LPSM no fluxo de bits) esteja "ativo" e esteja recebendo um quadro de AC-3 com um sinalizador 'confiável', o controlador do nível de intensidade do som no codificador (por exemplo, o processador de nível de intensidade do som 103 do codificador 100 da Fig. 2) é desviado. Os valores de dialnorm e DRC de origem 'confiáveis' são passados através (por exemplo, pelo gerador 106 do codificador 100) ao componente de codificador E-AC-3 (por exemplo, estágio 107 do codificador 100). A geração do bloco de LPSM continua e o loudness_correction_type_flag

é ajustado em '1'. A sequência de desvio do controlador do nível de intensidade do som é sincronizada ao começo do quadro AC-3 decodificado onde o sinalizador 'confiável'. A sequência de desvio do controlador do nível de intensidade do som é implementada tal como segue: o controle de `leveler_amount` é decrescido de um valor de 9 para um valor igual a 0 por 10 períodos de bloco de áudio (isto é, 53,3 ms) e o controle de `leveler_back_end_meter` é colocado no modo de desvio (esta operação deve resultar em uma transição sem emenda). O termo desvio "confiável" do nivelador implica que o valor de `dialnorm` do fluxo de bits de origem também é reutilizado na saída do codificador (por exemplo, se o fluxo de bits de origem 'confiável' tiver um valor de `dialnorm` de -30, então a saída do codificador deve utilizar -30 para o valor de `dialnorm` de saída).

[0091] Embora o codificador de E-AC-3 (que insere os valores de LPSM no fluxo de bits) esteja "ativo" e esteja recebendo um quadro de AC-3 sem o sinalizador 'confiável', o controlador do nível de intensidade do som embutido no codificador (por exemplo, o processador de nível de intensidade do som 103 do codificador 100 da Fig. 2) é ativo. A geração de bloco de LPSM continua e o `loudness_correction_type_flag` é ajustado em '0'. A sequência de ativação do controlador do nível de intensidade do som é sincronizada ao começo do quadro de AC-3 decodificado onde o sinalizador 'confiável' desaparece. A sequência de ativação do controlador do nível de intensidade do som é implementado tal como segue: o controle de `leveler_amount` é incrementado de um valor de 0 para um valor de 9 por 1 período de bloco de áudio (isto é, 5,3 ms) e o controle de `leveler_back_end_meter` é colocado no modo 'ativo' (esta operação resulta em uma transição sem emenda e inclui uma integração de `back_end_meter` restaurada); e, durante a codificação, uma interface gráfica do usuário (GUI) indicou a um usuário os seguintes parâmetros: "Inserir Programa de Áudio: [Confiável/Não

confiável]" - o estado deste parâmetro é baseado na presença do sinalizador "confiável" dentro do sinal de entrada; e "Correção do Nível de Intensidade do Som em Tempo Real: [Habilitada/Desabilitada]" - o estado deste parâmetro é baseado no fato se este controlador do nível de intensidade do som embutido no codificador está ativo.

[0092] Quando da decodificação de um fluxo de bits de AC-3 ou E-AC-3 que tem um LPSM (no formato descrito) incluído no campo "addbsi" do segmento da Informação de Fluxo de Bits ("BSI") de cada quadro do fluxo de bits, o decodificador analisa gramaticalmente os dados do bloco de LPSM (no campo addbsi) e passa os valores extraídos de LPSM a uma interface gráfica do usuário (GUI). O conjunto de valores extraídos de LPSM é restaurado a cada quadro.

[0093] Em um outro formato ainda, o fluxo de bits codificado é um fluxo de bits de AC-3 ou um fluxo de bits E-AC-3, e cada um dos segmentos de metadados que inclui o LPSM é incluído (por exemplo, pelo estágio 107 do codificador 100) como informação de fluxo de bits adicional no campo "addbsi" (mostrado na FIG. 6) do segmento de Informação de Fluxo de Bits ("BSI") (ou no segmento Aux) de um quadro do fluxo de bits. Neste formato (que é uma variação do formato descrito acima com referências às Tabelas 1 e 2), cada um dos campos addbsi (ou Aux) que contém LPSM contém os seguintes valores de LPSM: os elementos do núcleo especificados na Tabela 1, seguidos pela ID da carga útil ID (que identifica os metadados como LPSM) e os valores do tamanho da carga útil, seguidos pela carga útil (dados de LPSM) que tem o seguinte formato (similar aos elementos indicados na Tabela 2 acima): versão da carga útil de LPSM: um campo de 2 bits que indica a versão da carga útil de LPSM; dialchan: um campo de 3 bits que indica se os canais da esquerda, da direita e/ou do centro de dados de áudio correspondentes contêm um diálogo falado. A alocação de bits do campo dialchan pode ser tal como segue: bit 0, que indica a presença

de diálogo no canal da esquerda, é armazenado no bit mais significativo do campo dialchan; e bit 2, que indica a presença de diálogo no canal do centro, é armazenado no bit menos significativo do campo dialchan. Cada bit do campo dialchan é ajustado em '1' se o canal correspondente contiver o diálogo falado durante 0,5 segundo precedente do programa; loudregtyp: um campo de 3 bits que indica com qual padrão de regulação do nível de intensidade do som o nível de intensidade do som do programa é compatível. O ajuste do campo "loudregtyp" em '000' indica que o LPSM não indica a conformidade do regulamento do nível de intensidade do som. Por exemplo, um valor deste campo (por exemplo, 000) pode indicar que a conformidade com um padrão de regulação do nível de intensidade do som não está indicada, um outro valor deste campo (por exemplo, 001) pode indicar que os dados de áudio do programa estão em conformidade com a norma ATSC A/85, e um outro valor deste campo (por exemplo, 010) pode indicar que os dados de áudio do programa estão em conformidade com a norma EBU R128. No exemplo, se o campo for ajustado em qualquer valor com exceção de '000', os campos loudcorrdialgat e loudcorrtyp devem seguir na carga útil; loudcorrdialgat: um campo de um bit que indica se a correção do nível de intensidade do som conectada de diálogo foi aplicada. Se o nível de intensidade do som do programa tiver sido corrigido ao usar a conexão de diálogo, o valor do campo loudcorrdialgat é ajustado em '1'. Ou então é ajustado em '0'; loudcorrtyp: um campo de um bit que indica o tipo de correção do nível de intensidade do som aplicado ao programa. Se o nível de intensidade do som do programa for corrigido com um processo de correção do nível de intensidade do som de previsão infinito (baseado em arquivo), o valor do campo loudcorrtyp é ajustado em '0'. Se o nível de intensidade do som do programa for corrigido ao usar uma combinação da medição do nível de intensidade do som em tempo real e do controle da faixa dinâmica, o

valor deste campo é ajustado em '1'; loudrelgate: um campo de um bit que indica se os dados do nível de intensidade do som conectados relativos (ITU) existem. Se o campo loudrelgate for ajustado em '1', um campo ituloudrelgat de 7 bits deve seguir na carga útil; loudrelgat: um campo de 7 bits que indica o nível de intensidade do som de programa conectado relativo (ITU). Este campo indica o nível de intensidade do som integrado do programa de áudio, medido de acordo com ITU-R BS.1770-2 sem nenhum ajuste de ganho devido à aplicação da compressão de dialnorm e da faixa dinâmica. Os valores de 0 a 127 são interpretados como de -58 LKFS a +5,5 LKFS, em etapas de 0,5 de LKFS; loudspchgate: um campo de um bit que indica se os dados do nível de intensidade do som conectados de fala (ITU) existem. Se o campo loudspchgate for ajustado em '1', um campo loudspchgat de 7 bits deve seguir na carga útil; loudspchgat: um campo de 7 bits que indica o nível de intensidade do som de programa conectado de fala. Este campo indica o nível de intensidade do som integrado de todo o programa de áudio correspondente, medido de acordo com a fórmula (2) de ITU-R BS.1770-3 e sem nenhum ajuste de ganho devido à aplicação da compressão de dialnorm e da faixa dinâmica. Os valores de 0 a 127 são interpretados como de -58 a +5,5 LKFS, em etapas de 0,5 LKFS; loudstrm3se: um campo de um bit que indica se os dados do nível de intensidade do som de curta duração (3 segundos) existem. Se o campo for ajustado em '1', um campo de 7 bits loudstrm3s deve seguir na carga útil; loudstrm3s: um campo de 7 bits que indica o nível de intensidade do som não conectado de 3 segundos precedentes do programa de áudio correspondente, medidos de acordo com ITU-R BS.1771-1 e sem nenhum ajuste de ganho devido à aplicação da compressão de dialnorm e da faixa dinâmica. Os valores de 0 a 256 são interpretados como de -116 LKFS a +11.5 LKFS em etapas de 0,5 LKFS; truepke: um campo de um bit que indica se os dados do nível de

intensidade do som de pico confiáveis existem. Se o campo truepk for ajustado em '1', um campo de 8 bits truepk deve seguir na carga útil; e truepk: um campo de 8 bits que indica o valor da amostra do programa de pico real, medido de acordo com o Anexo 2 da norma ITU-R BS.1770-3 e sem nenhum ajuste de ganho devido à aplicação da compressão de dialnorm e da faixa dinâmica. Os valores de 0 a 256 são interpretados como de -116 LKFS a +11.5 LKFS em etapas de 0,5 LKFS.

[0094] Em algumas modalidades, o elemento de núcleo de um segmento de metadados em um campo auxdata (ou no campo "addbsi") de um quadro de um fluxo de bits AC-3 ou um fluxo de bits E-AC-3 compreende um cabeçalho de núcleo (opcionalmente incluindo valores de identificação, por exemplo, a versão do elemento de núcleo) e, depois do cabeçalho do núcleo: os valores indicativos do fato se os dados da impressão digital (ou outros valores de proteção) estão incluídos para os metadados do segmento de metadados, os valores indicativos se os dados externos (relacionados aos dados de áudio que correspondem aos metadados do segmento de metadados) existem, a ID da carga útil e os valores do tamanho da carga útil para cada tipo de metadados (por exemplo, LPSM, e/ou metadados de um tipo com exceção de LPSM) identificados pelo elemento de núcleo, e os valores da proteção para pelo menos um tipo de metadados identificados pelo elemento de núcleo. A(s) carga(s) útil(eis) dos metadados do segmento de metadados segue o cabeçalho de núcleo, e (em alguns casos) é aninhado dentro dos valores do elemento de núcleo.

Sistema do Nível de Intensidade do Som e da Faixa Dinâmica Otimizado

[0095] O esquema de codificação e transporte de metadados seguro descrito acima é usado conjuntamente com um sistema escalonável e extensível para otimizar o nível de intensidade do som e a faixa dinâmica através de diferentes dispositivos de reprodução,

aplicações, e ambientes de audição, tal como ilustrado na FIG. 1. Em uma modalidade, o sistema 10 é configurado para normalizar os níveis do nível de intensidade do som e da faixa dinâmica do áudio de entrada 11 através dos vários dispositivos que requerem valores diferentes do nível de intensidade do som alvo e têm diferentes capacidades de faixa dinâmica. Para normalizar os níveis do nível de intensidade do som e da faixa dinâmica, o sistema 10 inclui perfis diferentes do dispositivo com o conteúdo de áudio e a normalização é feita com base nestes perfis. Os perfis podem ser incluídos por uma das unidades de processamento de áudio nas cadeias de processamento de áudio e os perfis incluídos podem ser usados por uma unidade de processamento a jusante na cadeia de processamento de áudio para determinar o nível de intensidade do som desejado alvo e a faixa dinâmica para um dispositivo alvo. Componentes de processamento adicionais podem fornecer ou processar as informações para o gerenciamento do perfil do dispositivo (incluindo (mas sem nenhuma limitação a) os seguintes parâmetros incluindo: a faixa de banda nula, o limite de pico real, a faixa do nível de intensidade do som, a constante de tempo rápido/lento (coeficientes) e o impulso máximo), o controle de ganho e as funções de geração de ganho de banda larga/múltiplas bandas.

[0096] A FIG. 10 ilustra um diagrama mais detalhado do sistema da FIG. 1 para um sistema que propicia um controle otimizado do nível de intensidade do som e da faixa dinâmica, sob algumas modalidades. Para o sistema 321 da FIG. 10, o estágio de codificador compreende um componente 304 do codificador de núcleo que codifica a entrada de áudio 303 em um formato digital apropriado para a transmissão ao decodificador 312. O áudio é processado de modo que possa ser reproduzido em uma variedade de ambientes de audição diferentes, cada um dos quais pode requerer ajustes alvo do nível de intensidade do som e/ou da faixa dinâmica diferentes. Desse modo, tal como

mostrado na FIG. 10, o decodificador emite um sinal digital que é convertido no formato analógico por um conversor digital em analógico 316 para a reprodução através de uma variedade de tipos diferentes de driver incluindo os alto-falantes 320 de faixa completa, os alto-falantes em miniatura 322, e os fones de ouvido 324. Estes drivers ilustram apenas alguns exemplos de drivers de reprodução possíveis, e quaisquer transdutores ou drivers de qualquer tamanho e tipo apropriados podem ser usados. Além disso, os drivers/transdutores 320 a 324 da FIG. 10 podem ser incorporados em qualquer dispositivo de reprodução apropriado para o uso em qualquer ambiente de audição correspondente. Os tipos de dispositivo podem incluir, por exemplo, AVRs, televisores, equipamento estereofônico, computadores, telefones móveis, computadores do tipo tablet, reprodutores de MP3, e assim por diante; e os ambientes de audição podem incluir, por exemplo, auditórios, lares, carros, cabines de audição, e assim por diante.

[0097] Uma vez que a faixa de ambientes de reprodução e os tipos de drivers podem variar de contextos privados muito pequenos a locais públicos muito grandes, a extensão das configurações de nível de intensidade do som e de faixa dinâmica de reprodução possíveis e ideais pode variar bastante dependendo do tipo de conteúdo, dos níveis de ruído de fundo, e outros ainda. Por exemplo, em um ambiente de cinema em casa, o amplo conteúdo da faixa dinâmica pode ser reproduzido através de um equipamento de som do tipo surround e um conteúdo de faixa dinâmica mais estreita pode ser reproduzido através de um sistema de televisão normal (tal como um tipo de LED/LCD de painel plano), ao passo que um modo de faixa dinâmica muito estreita pode ser usado para determinadas condições de audição (por exemplo, à noite ou em um dispositivo com grandes limitações de potência de saída acústica, por exemplo, alto-falantes de telefone móvel/tablet ou

saída de fone de ouvido) quando grandes variações de nível não são desejadas. Em contextos de audição portáteis ou móveis tais como ao usar alto-falantes de computador ou de doca, ou fones de ouvido/fones auriculares, a faixa dinâmica ideal da reprodução pode variar dependendo do ambiente. Por exemplo, em um ambiente quieto a faixa dinâmica ideal pode ser maior em comparação a um ambiente barulhento. As modalidades do sistema de processamento de áudio adaptável da FIG. 10 irão variar a faixa dinâmica para tornar o conteúdo de áudio mais inteligível dependendo de parâmetros, tais como o ambiente do dispositivo de audição e o tipo de dispositivo de reprodução.

[0098] A FIG. 11 é uma tabela que ilustra requisitos de faixa dinâmica diferentes para uma variedade de dispositivos de reprodução e de ambientes de audição de fundo em um caso de uso exemplificador. Requisitos similares podem ser derivados para o nível de intensidade do som. Os requisitos de faixa dinâmica e do nível de intensidade do som diferentes geram perfis diferentes que são usados pelo sistema de otimização 321. O sistema 321 inclui um componente de medição do nível de intensidade do som e da faixa dinâmica 302 que analisa e mede o nível de intensidade do som e a faixa dinâmica do áudio de entrada. Em uma modalidade, o sistema analisa o conteúdo total do programa para determinar o parâmetro do nível de intensidade do som total. Neste contexto, o nível de intensidade do som refere-se ao nível de intensidade do som de programa de longa duração ou ao nível de intensidade do som médio de um programa, em que um programa é uma única unidade de conteúdo de áudio, tal como um filme, um show de televisão, um comercial, ou um conteúdo de programa similar. O nível de intensidade do som é usado para fornecer uma indicação do perfil artístico da faixa dinâmica que é usado por criadores de conteúdo para controlar a maneira na qual o áudio será reproduzido. O nível de

intensidade do som está relacionado ao valor dos metadados de dialnorm em que dialnorm representa o nível de intensidade do som de diálogo médio de um único programa (por exemplo, filme, show, comercial, etc.). A faixa dinâmica de curta duração quantifica as variações nos sinais por um período de tempo muito mais curto do que o nível de intensidade do som do programa. Por exemplo, a faixa dinâmica de curta duração pode ser medida na ordem de segundos, ao passo que o nível de intensidade do som do programa pode ser medido por uma extensão de minutos ou até mesmo de horas. A faixa dinâmica de curta duração fornece um mecanismo de proteção que é independente do nível de intensidade do som do programa para assegurar que uma sobrecarga não ocorra para vários perfis de reprodução e tipos de dispositivo. Em uma modalidade, o alvo do nível de intensidade do som (nível de intensidade do som de programa de longa duração) é baseado no nível de intensidade do som do diálogo e a faixa dinâmica de curta duração é baseada no nível de intensidade do som conectado e/ou não conectado relativo. Neste caso, determinados componentes de DRC e do nível de intensidade do som no sistema estão inseridos no contexto no que diz respeito ao tipo de conteúdo e/ou aos tipos e às características de dispositivo alvo. Como parte desta capacidade inserida no contexto, o sistema é configurado para analisar uma ou mais características do dispositivo de saída para determinar se o dispositivo é um membro de grupos particulares de dispositivos que são otimizados para determinadas condições de reprodução de DRC e nível de intensidade do som, tais como os dispositivos do tipo AVR, televisores, computadores, dispositivos portáteis, e assim por diante.

[0099] Um componente de pré-processamento analisa o conteúdo do programa para determinar o nível de intensidade do som, os picos, os picos reais, e os períodos quietos para criar metadados singulares para cada perfil de uma pluralidade de perfis diferentes. Em uma

modalidade, o nível de intensidade do som pode ser um nível de intensidade do som conectado de diálogo e/ou um de nível de intensidade do som conectado relativo. Os perfis diferentes definem vários modos de DRC (Controle de Faixa Dinâmica) e de nível de intensidade do som alvo em que valores de ganho diferentes são gerados no codificador dependendo das características do conteúdo de áudio da fonte, do nível de intensidade do som desejado alvo e do tipo e/ou do ambiente do dispositivo da reprodução. O decodificador pode oferecer modos diferentes e DRC e do nível de intensidade do som alvo (habilitados pelos perfis mencionados acima) e pode incluir DRC e nível de intensidade do som alvo desligado/desabilitado que permite uma listagem da faixa dinâmica completa sem nenhuma compressão do sinal de áudio e nenhuma normalização do nível de intensidade do som, normalização de DRC desligado/desabilitado e nível de intensidade do som com um alvo de modo de linha de -31 LKFS para reprodução nos sistemas de cinema em casa, o que fornece uma compressão da faixa dinâmica através de valores ganhos gerados (especificamente para este modo de reprodução e/ou perfil do dispositivo) no codificador com a normalização do nível de intensidade do som com um alvo de -31 LKFS; modo de RF para reprodução através de alto-falantes da tevê, o que fornece uma grande quantidade de compressão da faixa dinâmica com a normalização do nível de intensidade do som com um alvo de -24, -23 ou -20 LKFS, de um modo intermediário para reprodução em computadores ou dispositivos similares, o que fornece a compressão com a normalização do nível de intensidade do som a um alvo de -14 LKFS, e o modo portátil, que fornece uma compressão muito grande da faixa dinâmica com um alvo da normalização do nível de intensidade do som de -11 LKFS. Os valores do nível de intensidade do som alvo de -31, -23/-20, -14 e -11 LKFS se prestam como exemplos de perfis diferentes de reprodução/dispositivo que podem ser definidos para o

sistema sob algumas modalidades, e quaisquer outros valores apropriados do nível de intensidade do som alvo podem ser usados, e o sistema gera valores apropriados de ganho especificamente para estes modos de reprodução e/ou perfil do dispositivo. Além disso, o sistema é extensível e adaptável de modo que os dispositivos de reprodução e ambientes de audição diferentes possam ser acomodados através da definição de um novo perfil no codificador ou em outra parte e carregados no codificador. Desta maneira, perfis novos e singulares de reprodução/dispositivo podem ser gerados para suportar dispositivos de reprodução melhorados ou diferentes para aplicações futuras.

[00100] Em uma modalidade, os valores de ganho podem ser calculados em qualquer componente de processamento apropriado do sistema 321, tal como no codificador 304, no decodificador 312, ou no transcodificador 308, ou qualquer componente de pré-processamento associado com o codificador ou qualquer componente pós-processamento associado com o decodificador.

[00101] A FIG. 13 é um diagrama de blocos que ilustra uma interface entre perfis diferentes para uma variedade de classes de dispositivos de reprodução diferentes, sob uma modalidade. Tal como mostrado na FIG. 13, um codificador 502 recebe a entrada de áudio 501 e um de vários perfis possíveis diferentes 506. O codificador combina os dados de áudio com o perfil selecionado para gerar um arquivo de fluxo de bits de saída que é processado nos componentes do decodificador presentes ou associados com o dispositivo de reprodução alvo. Para o exemplo da FIG. 13, os dispositivos de reprodução diferentes podem ser um computador 510, um telefone móvel 512, um AVR 514, e um televisor 516, embora muitos outros dispositivos de saída também sejam possíveis. Cada um dos dispositivos 510 a 516 inclui ou é acoplado aos alto-falantes (incluindo drivers e/ou transdutores) tais como os drivers 320 a 324. A combinação do processamento,

avaliações de potência e tamanhos dos dispositivos de reprodução e os alto-falantes associados ditam de modo geral qual perfil é o mais ideal para esse alvo particular. Desse modo, os perfis 506 podem ser especificamente definidos para a reprodução através de AVRs, televisores, alto-falantes móveis, fones de ouvido móveis, e assim por diante. Eles também podem ser definidos para modos ou condições operacionais específicos, tais como o modo quieto, modo noturno, ao ar livre, dentro de casa, e assim por diante. Os perfis mostrados na FIG. 13 são modos exemplificadores apenas e qualquer perfil apropriado pode ser definido, incluindo perfis customizados para alvos e ambientes específicos.

[00102] Embora a FIG. 13 ilustre uma modalidade em que um codificador 502 recebe os perfis 506 e gera os parâmetros apropriados para o nível de intensidade do som e o processamento de DRC, deve ser observado que os parâmetros gerados com base em um perfil e no conteúdo de áudio podem ser executados em qualquer unidade de processamento de áudio apropriada tal como um codificador, decodificador, transcodificador, pré-processador, pós-processador, e assim por diante. Por exemplo, cada dispositivo de saída 510 a 516 da FIG. 13 tem ou é acoplado a um componente de decodificador que processa os metadados no fluxo de bits no arquivo 504 enviado do codificador 502 para permitir a adaptação do nível de intensidade do som e da faixa dinâmica para combinar o dispositivo ou o tipo de dispositivo do dispositivo de saída alvo.

[00103] Em uma modalidade, a faixa dinâmica e o nível de intensidade do som do conteúdo de áudio são otimizados para cada dispositivo de reprodução possível. Isto é obtido ao manter o nível de intensidade do som de longa duração no alvo e ao controlar a faixa dinâmica de curta duração para otimizar a experiência de áudio (pelo controle da dinâmica do sinal, pelos picos da amostra e/ou por picos

reais) para cada um dos modos de reprodução alvo. Elementos de metadados diferentes são definidos para o nível de intensidade do som de longa duração e a faixa dinâmica de curta duração. Tal como mostrado na FIG. 10, o componente 302 analisa todo o sinal de áudio de entrada (ou partes do mesmo, tal como o componente de fala, se for aplicável) para derivar as características relevantes para ambos estes componentes de DR separados. Isto permite que valores de ganho diferentes sejam definidos para ganhos artísticos contra valores de ganho de clips (proteção contra sobrecarga).

[00104] Estes valores de ganho para o nível de intensidade do som de longa duração e a faixa dinâmica de curta duração são então mapeados a um perfil 305 para gerar os parâmetros que descrevem os valores do ganho do controle do nível de intensidade do som e da faixa dinâmica. Estes parâmetros são combinados com o sinal de áudio codificado do codificador 304 em um multiplexador 306, ou um componente similar para a criação de um fluxo de bits que é transmitido através do transcodificador 308 a um estágio do decodificador. O fluxo de bits inserido no estágio do decodificador é desmultiplexado no desmultiplexador 310. Ele é então decodificado no decodificador 312. O componente de ganho 314 aplica os ganhos que correspondem ao perfil apropriado para gerar os dados de áudio digitais que são então processados através da unidade DACS para reprodução 416 através dos dispositivos de reprodução e drivers ou transdutores 320 a 324 apropriados.

[00105] A FIG. 14 é uma tabela que ilustra a correlação entre o nível de intensidade do som de longa duração e a faixa dinâmica de curta duração para uma pluralidade de perfis definidos, sob uma modalidade. Tal como mostrado na Tabela 4 da FIG. 14, cada perfil compreende um conjunto de valores de ganho que ditam a quantidade de compressão da faixa dinâmica (DRC) aplicada no decodificador do sistema ou em

cada dispositivo alvo. Cada um dos N perfis, denotados como Perfis 1 - N, ajustam parâmetros do nível de intensidade do som de longa duração particulares (por exemplo, dialnorm) e parâmetros de compressão de sobrecarga ao ditar valores de ganho correspondentes aplicados no estágio do decodificador. Os valores de ganho de DRC para os perfis podem ser definidos por uma fonte externa que seja aceita pelo codificador, ou podem ser gerados internamente dentro do codificador como valores de ganho padrão se valores externos não forem fornecidos.

[00106] Em uma modalidade, os valores de ganho para cada perfil são incorporados nas palavras de ganho de DRC que são computadas com base em uma análise de determinadas características do sinal de áudio, tais como o pico, pico real, o nível de intensidade do som de curta duração do diálogo ou o nível de intensidade do som de curta duração total ou uma combinação (híbrida) de ambos, para computar os ganhos estáticos com base em um perfil escolhido (isto é, a característica ou a curva de transferência) assim como as constantes do tempo necessárias para executar o ataque rápido/lento e a liberação rápida/lenta dos ganhos de DRC finais para cada perfil do dispositivo e/ou nível de intensidade do som alvo possível. Tal como indicado acima, estes perfis podem ser pré-ajustados no codificador, decodificador, ou ser gerados externamente e carregados do criador de conteúdo no codificador através de metadados externos.

[00107] Em uma modalidade, os valores de ganho podem ser o ganho de banda larga que aplica o mesmo ganho em todas as frequências do conteúdo de áudio. Alternativamente, o ganho pode compreender valores de ganho de múltiplas bandas de maneira tal que valores de ganho diferentes são aplicados às frequências diferentes ou às faixas de frequência do conteúdo de áudio. No caso de múltiplos canais, cada perfil pode constituir uma matriz de valores de ganho

indicando ganhos para faixas de frequência diferentes em vez de um único valor de ganho.

[00108] Com referência à FIG. 10, em uma modalidade, as informações a respeito das propriedades ou características do ambiente de audição e/ou das capacidades e configurações dos dispositivos de reprodução são fornecidas pelo estágio do decodificador ao estágio do codificador pela conexão de retroalimentação 330. A informação de perfil 332 também é inserida no codificador 304. Em uma modalidade, o decodificador analisa os metadados no fluxo de bits para determinar se um parâmetro do nível de intensidade do som para um primeiro grupo de dispositivos de reprodução de áudio está disponível no fluxo de bits. Em caso positivo, ele transmite os parâmetros a jusante para o uso na definição do áudio. Em caso contrário, o codificador analisa determinadas características dos dispositivos para derivar os parâmetros. Estes parâmetros são então enviados a um componente de definição a jusante para a reprodução. O codificador também determina um dispositivo de saída (ou um grupo de dispositivos de saída que incluem o dispositivo de saída) que deve definir o fluxo de áudio recebido. Por exemplo, o dispositivo de saída pode ser determinado como sendo um telefone celular ou pertencendo a um grupo tais como dispositivos portáteis. Em uma modalidade, o decodificador usa a conexão de retroalimentação 330 para indicar ao codificador o dispositivo de saída determinado ou o grupo de dispositivos de saída. Para esta retroalimentação, um módulo conectado ao dispositivo de saída (por exemplo, um módulo em uma placa de som conectada aos fones de ouvido ou conectada aos alto-falantes em um laptop) pode indicar ao decodificador uma identidade do dispositivo de saída ou uma identidade de um grupo de dispositivos que inclua o dispositivo de saída. O decodificador transmite esta informação ao codificador através da conexão de retroalimentação 330. Em uma modalidade, o decodificador

executa os parâmetros do nível de intensidade do som e de DRC. Em uma modalidade, o decodificador determina os parâmetros do nível de intensidade do som e de DRC. Nesta modalidade, em vez de transmitir a informação pela conexão de retroalimentação 330, o decodificador usa a informação sobre o dispositivo determinado ou o grupo de dispositivos de saída para determinar os parâmetros do nível de intensidade do som e de DRC. Em uma outra modalidade, uma outra unidade de processamento de áudio determina os parâmetros do nível de intensidade do som e de DRC e o decodificador transmite as informações a essa unidade de processamento de áudio em vez do decodificador.

[00109] A FIG. 12 é um diagrama de blocos de um sistema de otimização da faixa dinâmica, sob uma modalidade. Tal como mostrado na FIG. 12, um codificador 402 recebe o áudio de entrada 401. O áudio codificado é combinado no multiplexador 409 com os parâmetros 404 gerados de uma curva de compressão selecionada 422 e um valor de dialnorm 424. O fluxo de bits resultante é transmitido ao desmultiplexador 411 que gera os sinais de áudio que são decodificados pelo decodificador 406. Os parâmetros e os valores de dialnorm são usados pela unidade de cálculo de ganho 408 para gerar os níveis de ganho que impelem o amplificador 410 para a amplificação da saída do decodificador. A FIG.12 ilustra como o controle da faixa dinâmica é colocado sob parâmetros e é inserido no fluxo de bits. O nível de intensidade do som também pode ser colocado sob parâmetros e inserido no fluxo de bits ao usar componentes similares. Em uma modalidade, um controle de nível de referência de saída (não mostrado) também pode ser fornecido ao decodificador. Embora a figura ilustre os parâmetros do nível de intensidade do som e da faixa dinâmica como sendo determinados e inseridos no codificador, uma determinação similar pode ser executada em outras unidades de processamento de

áudio tais como um pré-processador, um decodificador e um pós-processador.

[00110] A FIG. 15 ilustra exemplos de perfis do nível de intensidade do som para tipos diferentes de conteúdo de áudio, sob uma modalidade. Tal como mostrado na FIG. 15, as curvas exemplificadoras 600 e 602 traçam o nível de intensidade do som de entrada (em LKFS) contra o ganho centrado em torno de 0 LKFS. Os tipos diferentes de curvas exibem conteúdos diferentes, tal como mostrado na FIG. 15, em que a curva 600 pode representar a fala e a curva 602 pode representar o conteúdo de filme padrão. Tal como mostrado na FIG. 15, o conteúdo de fala é sujeito a uma quantidade maior de ganho em relação ao conteúdo do filme. A FIG.15 se presta como um exemplo de curvas representativas do perfil para determinados tipos de conteúdo de áudio, e outras curvas de perfil também podem ser usadas. Determinados aspectos das características do perfil, tal como mostrado na FIG. 15, são usados para derivar os parâmetros relevantes para o sistema de otimização. Em uma modalidade, estes parâmetros incluem: largura de faixa nula, razão de corte, razão de intensificação, intensificação máxima, ataque de FS, degradação de FS, contenção, limite de pico e nível de intensidade do som do nível alvo. Outros parâmetros podem ser usados além de ou alternativamente a pelo menos alguns destes parâmetros dependendo dos requisitos da aplicação e das restrições do sistema.

[00111] A FIG. 16 é um fluxograma que ilustra um método de otimização do nível de intensidade do som e da faixa dinâmica através dos dispositivos de reprodução e das aplicações, sob uma modalidade. Embora a figura ilustre a otimização do nível de intensidade do som e da faixa dinâmica como sendo executada no codificador, uma otimização similar pode ser executado em outras unidades de processamento de áudio tais como um pré-processador, um

decodificador e um pós-processador. Tal como mostrado no processo 620, o método começa com um estágio do codificador recebendo um sinal de entrada de uma fonte 603. O codificador ou um componente de pré-processamento determina então se o sinal da fonte foi ou não submetido a um processo que atinge um nível de intensidade do som e/ou uma faixa dinâmica 604 alvo. O nível de intensidade do som alvo corresponde ao nível de intensidade do som de longa duração e pode ser definido externa ou internamente. Se o sinal da fonte não foi submetido ao processo para atingir o nível de intensidade do som e/ou a faixa dinâmica alvo, o sistema executa a operação de controle do nível de intensidade do som e/ou da faixa dinâmica 608 apropriado; em caso contrário, se o sinal da fonte foi submetido a esta operação de controle do nível de intensidade do som e/ou da faixa dinâmica, o sistema entra em um modo de desvio para pular as operações de controle do nível de intensidade do som e/ou da faixa dinâmica para permitir que o processo original dite o nível de intensidade do som de longa duração e/ou a faixa dinâmica 606 apropriados. Os valores de ganho apropriados para o modo de desvio 606 ou então o modo executado 608 (que podem ser valores de ganho de banda larga simples ou valores de ganho de múltiplas bandas dependentes da frequência) são aplicados então no decodificador 612.

Formato de Fluxo de Bits

[00112] Tal como indicado anteriormente, o sistema para otimizar o nível de intensidade do som e a faixa dinâmica emprega um formato de metadados extensível seguro para assegurar que os metadados e o conteúdo de áudio transmitido no fluxo de bits entre o codificador e o decodificador, ou entre a fonte e os dispositivos de definição/reprodução não sejam separados uns dos outros ou então corrompidos durante a transmissão por redes ou outros equipamentos proprietários, tal como a interface do fornecedor de serviço, e assim por diante. Este fluxo de bits

fornece um mecanismo para sinalizar os componentes do codificador e/ou do decodificador para adaptar o nível de intensidade do som e a faixa dinâmica do sinal de áudio para adequar o conteúdo de áudio e as características do dispositivo de saída através das informações de perfil apropriadas. Em uma modalidade, o sistema é configurado para determinar um fluxo de bits codificado com baixa taxa de bits a ser transmitido entre o codificador e o decodificador, e a informação do nível de intensidade do som codificada através dos metadados compreende características para um ou mais perfis de saída. Segue uma descrição de um formato do fluxo de bits para o uso com um sistema de otimização do nível de intensidade do som e da faixa dinâmica sob uma modalidade.

[00113] Um fluxo de bits codificado AC-3 compreende os metadados e um a seis canais de conteúdo de áudio. O conteúdo de áudio consiste em dados de áudio que foram comprimidos ao usar uma codificação de áudio perceptual. Os metadados incluem vários parâmetros de metadados de áudio que são destinados para o uso na mudança do som de um programa entregue a um ambiente de audição. Cada quadro de um fluxo de bits de áudio codificado AC-3 contém o conteúdo de áudio e metadados para 1.536 amostras de áudio digital. Para uma taxa de amostragem de 48 kHz, isto representa 32 milissegundos de áudio digital ou uma taxa de 31,25 quadros por segundo do áudio.

[00114] Cada quadro de um fluxo de bits de áudio codificado E-AC-3 contém o conteúdo de áudio e metadados de áudio para 256, 512, 768 ou 1536 amostras de áudio digital, dependendo se o quadro contém um, dois, três ou seis blocos de dados de áudio, respectivamente. Para uma taxa da amostragem de 48 kHz, isto representa 5,333, 10,667, 16 ou 32 milissegundos de áudio digital, respectivamente, ou uma taxa de 189,9, 93,75, 62,5 ou 31,25 quadros por segundo de áudio, respectivamente.

[00115] Tal como indicado na Fig. 4, cada quadro de AC-3 é dividido

em seções (segmentos), as quais incluem: uma seção de informação de sincronização (SI) que contém (tal como mostrado na Fig. 5) uma palavra de sincronização (SW) e a primeira de duas palavras de correção de erro (CRC1); uma seção de Informação de Fluxo de Bits (BSI) que contém a maior parte dos metadados; seis blocos de áudio (AB0 a AB5) que contêm o conteúdo de áudio comprimido em dados (e também pode incluir metadados); bits restantes (W) que contêm todos os bits não utilizados restantes que o conteúdo de áudio é comprimido; uma seção de informação Auxiliar (AUX) que pode conter mais metadados; e a segunda de duas palavras de correção de erro (CRC2).

[00116] Tal como indicado na Fig. 7, cada quadro de E-AC-3 é dividido em seções (segmentos), as quais incluem: uma seção de informação de sincronização (SI) que contém (tal como mostrado na Fig. 5) uma palavra de sincronização (SW); uma seção de Informação de Fluxo de Bits (BSI) que contém a maior parte dos metadados; entre um e seis blocos de áudio (AB0 a AB5) que contêm o conteúdo de áudio comprimido em dados (e também pode incluir metadados); bits restantes (W) que contêm todos os bits não utilizados restantes depois que o conteúdo de áudio é comprimido; uma seção de informação Auxiliar (AUX) que pode conter mais metadados; e uma palavra de correção de erro (CRC).

[00117] Em um fluxo de bits AC-3 (ou E-AC-3) há vários parâmetros de metadados de áudio que são destinados especificamente para o uso na mudança do som do programa entregue a um ambiente de audição. Um dos parâmetros de metadados é o parâmetro de dialnorm, que é incluído no segmento de BSI.

[00118] Tal como mostrado na FIG. 6, o segmento de BSI de um quadro de AC-3 inclui um parâmetro de cinco bits ("dialnorm") que indica o valor de dialnorm para o programa. Um parâmetro de cinco bits ("dialnorm2") que indica o valor de dialnorm para um segundo programa

de áudio contido dentro do mesmo quadro de AC-3 é incluído se o modo de codificação de áudio ("acmod") do quadro de AC-3 for "0", indicando que uma configuração de canal dual-mono ou "1+1" está em uso.

[00119] O segmento do BSI também inclui um sinalizador ("addbsie") que indica a presença (ou a ausência) da informação do fluxo de bits adicional que segue a informação o bit "addbsie", um parâmetro ("addbsil") que indica o comprimento de qualquer informação do fluxo de bits adicional que segue o valor de "addbsil", e até 64 bits de informação do fluxo de bits adicional ("addbsi") que segue o valor de "addbsil". O segmento de BSI pode incluir outros valores de metadados não mostrados especificamente na FIG. 6.

[00120] Os aspectos de uma ou mais modalidades descritas no presente documento podem ser implementados em um sistema de áudio que processe sinais de áudio para a transmissão através de uma rede que inclua um ou mais computadores ou dispositivos de processamento que executam instruções de software. Algumas das modalidades descritas podem ser usadas sozinhas ou umas em conjunto com as outras em qualquer combinação. Embora várias modalidades possam ter sido motivadas por várias deficiências com a técnica anterior, as quais podem ser discutidas ou aludidas em um ou mais lugares no relatório descritivo, as modalidades não são focadas necessariamente em qualquer uma destas deficiências. Em outras palavras, as modalidades diferentes podem ser focadas em deficiências diferentes que podem ser discutidas no relatório descritivo. Algumas modalidades podem ser focadas somente parcialmente em algumas deficiências ou em apenas uma deficiência que podem ser discutidas no relatório descritivo, e algumas modalidades podem não ser focadas em qualquer uma destas deficiências.

[00121] Os aspectos dos sistemas descritos no presente documento podem ser implementados em um ambiente de processamento de som

por um ambiente de rede de processamento de som baseado em computador para processar arquivos de áudio digitais ou digitalizados. As partes do sistema de áudio adaptável podem incluir uma ou mais redes que compreendem qualquer número desejado de máquinas individuais, incluindo um ou mais roteadores (não mostrados) que servem para armazenar temporariamente e distribuir os dados transmitidos entre os computadores. Tal rede pode ser construída em vários protocolos de rede diferentes, e pode ser a Internet, uma rede de área larga (WAN), uma rede de área local (LAN), ou qualquer combinação destas.

[00122] Um ou mais dos componentes, blocos, processos ou outros componentes funcionais podem ser implementados através de um programa de computador que controla a execução de um dispositivo de computando baseado em processador do sistema. Também deve ser anotado que as várias funções divulgadas no presente documento podem ser descritas ao usar qualquer número de combinações de hardware, firmware, e/ou dados e/ou instruções incorporados em vários meios que podem ser lidos por máquinas ou por computadores, em termos de suas características comportamentais, de transferência de registro, de componentes lógicos, e/ou outras características. O meio que pode ser lido por computador em que tais dados e/ou instruções formatados podem ser incorporados incluem, mas sem ficar a eles limitados, meios físicos (não transitórios), meios de armazenamento não volátil em várias formas, tais como meios de armazenamento ópticos, magnéticos ou semicondutores.

[00123] A menos que o contexto requeira claramente de alguma outra maneira, por toda a descrição e as concretizações, as palavras "compreendem", "que compreendem" e outras do gênero devem ser interpretadas em um sentido inclusivo ao contrário de um sentido exclusivo ou exaustivo; o que equivale dizer, em um sentido de "incluir,

mas sem ficar limitado a". As palavras que usam o número singular ou plural também incluem o número plural ou singular, respectivamente. Além disso, as palavras "aqui", "abaixo", "acima", "em baixo" e as palavras de importância similar referem-se a esta aplicação como um todo e não a quaisquer partes particulares desta aplicação. Quando a palavra "ou" for usada em referência a uma lista de dois ou mais itens, essa palavra cobre todas as seguintes interpretações da palavra: qualquer um dos itens na lista, todos os itens na lista e qualquer combinação dos itens na lista.

[00124] Embora uma ou mais implementações tenham sido descritas a título de exemplo e em termos das modalidades específicas, deve ser compreendido que uma ou mais implementações não ficam limitadas às modalidades divulgadas. Pelo contrário, pretende-se cobrir várias modificações e arranjos similares tal como deve ser aparente aos elementos versados na técnica. Portanto, o âmbito das concretizações deve ser adaptado à definição mais ampla de modo a englobar todos tais modificações e arranjos similares.

REIVINDICAÇÕES

1. Método, **caracterizado pelo fato de que** compreende as etapas de:

receber, em um fluxo de bits que compreende dados de áudio, metadados associados com os dados de áudio;

analisar os metadados no fluxo de bits para determinar se um parâmetro do nível de intensidade do som para um primeiro grupo de dispositivos de reprodução de áudio (510, 512, 514, 516) está disponível no fluxo de bits;

em resposta à determinação de que o parâmetro está presente para o primeiro grupo, usar o parâmetro e dados de áudio para renderizar o áudio; e

em resposta à determinação de que o parâmetro do nível de intensidade do som não está presente para o primeiro grupo, analisar uma ou mais características do primeiro grupo, e determinar o parâmetro com base em uma ou mais características,

em que as uma ou mais características compreendem níveis de ganho para perfis diferentes dos dados de áudio.

2. Método, de acordo com a reivindicação 1, **caracterizado pelo fato de que** os níveis de ganho definem pelo menos um de: um perfil de faixa dinâmica artístico que controla a maneira na qual os dados de áudio serão reproduzidos para um programa definido, e um perfil da faixa dinâmica de curta duração que confere proteção contra sobrecarga para partes do programa definido.

3. Método, de acordo com a reivindicação 1, **caracterizado pelo fato de que** o uso do parâmetro e dos dados de áudio para renderizar o áudio compreende a transmissão do parâmetro e dos dados de áudio a um módulo a jusante que renderiza o áudio para a reprodução.

4. Método, de acordo com a reivindicação 1, **caracterizado**

pelo fato de que o uso do parâmetro e dos dados de áudio para renderizar o áudio compreende renderizar os dados de áudio com base no parâmetro e nos dados de áudio.

5. Método, de acordo com a reivindicação 1, **caracterizado pelo fato de que** compreende adicionalmente:

determinar um dispositivo de saída que deve renderizar o fluxo de áudio recebido; e

determinar se o dispositivo de saída pertence ou não ao primeiro grupo de dispositivos de reprodução de áudio, em que a etapa de análise dos metadados no fluxo para determinar se o parâmetro do nível de intensidade do som para o primeiro grupo de dispositivos de reprodução de áudio está disponível é executada depois da etapa de determinação do dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio.

6. Método, de acordo com a reivindicação 5, **caracterizado pelo fato de que** a etapa de determinação de que o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio compreende:

receber uma indicação de um módulo conectado ao dispositivo de saída que indica uma identidade do dispositivo de saída ou que indica uma identidade de um grupo de dispositivos que incluem o dispositivo de saída; e

determinar que o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio com base na indicação recebida.

7. Meio legível em computador **caracterizado pelo fato de que** configurado para controlar a execução de um dispositivo de computação com base em processador de modo que ele execute o método conforme definido em qualquer uma das reivindicações 1 a 6.

8. Aparelho, **caracterizado pelo fato de que** compreende:

uma interface configurada para receber um fluxo de bits que inclui dados de áudio e metadados associados com os dados de áudio;

um analisador acoplado à interface e configurado para analisar os metadados no fluxo de bits para determinar se um parâmetro do nível de intensidade do som para um primeiro grupo de dispositivos de reprodução de áudio (510, 512, 514, 516) está disponível no fluxo de bits; e

um componente de renderização configurado para usar os parâmetros e os dados de áudio para renderizar o áudio em resposta à determinação pelo analisador de que os parâmetros estão presentes para o primeiro grupo, o componente de renderização sendo configurado ainda para analisar uma ou mais características do primeiro grupo e determinar o parâmetro com base nas uma ou mais características em resposta à determinação pelo analisador de que os parâmetros do nível de intensidade do som não estão presentes para o primeiro grupo,

em que as uma ou mais características compreendem níveis de ganho para diferentes perfis dos dados de áudio.

9. Aparelho, de acordo com a reivindicação 8, **caracterizado pelo fato de que** o componente de renderização que usa os parâmetros e os dados de áudio para renderizar o áudio compreende a transmissão pelo componente de renderização do parâmetro e dos dados de áudio a um módulo a jusante que renderiza o áudio para a reprodução.

10. Aparelho, de acordo com a reivindicação 9, **caracterizado pelo fato de que** o componente de renderização que usa o parâmetro e os dados de áudio para renderizar o áudio compreende a renderização, pelo componente de renderização, dos dados de áudio com base no parâmetro e nos dados de áudio.

11. Aparelho, de acordo com a reivindicação 10, **caracterizado pelo fato de que** compreende adicionalmente um

segundo componente configurado para determinar um dispositivo de saída que renderiza o fluxo de áudio recebido e determina se ou o dispositivo de saída pertence ou não ao primeiro grupo de dispositivos de reprodução de áudio; em que o analisador analisa os metadados no fluxo para determinar se o parâmetro do nível de intensidade do som para o primeiro grupo de dispositivos de reprodução de áudio está disponível depois que o segundo componente determina se o dispositivo de saída pertence ou não ao primeiro grupo de dispositivos de reprodução de áudio.

12. Aparelho, de acordo com a reivindicação 11, **caracterizado pelo fato de que** compreende adicionalmente uma interface configurada para receber uma indicação de um módulo conectado ao dispositivo de saída que indica uma identidade do dispositivo de saída ou que indica uma identidade de um grupo de dispositivos que incluem o dispositivo de saída, e para determinar que o dispositivo de saída pertence ao primeiro grupo de dispositivos de reprodução de áudio com base na indicação recebida.

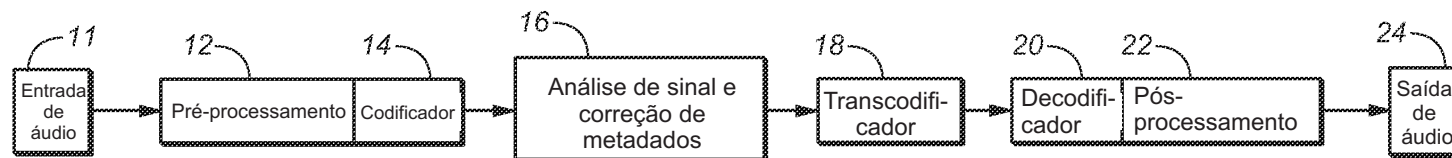


FIG. 1

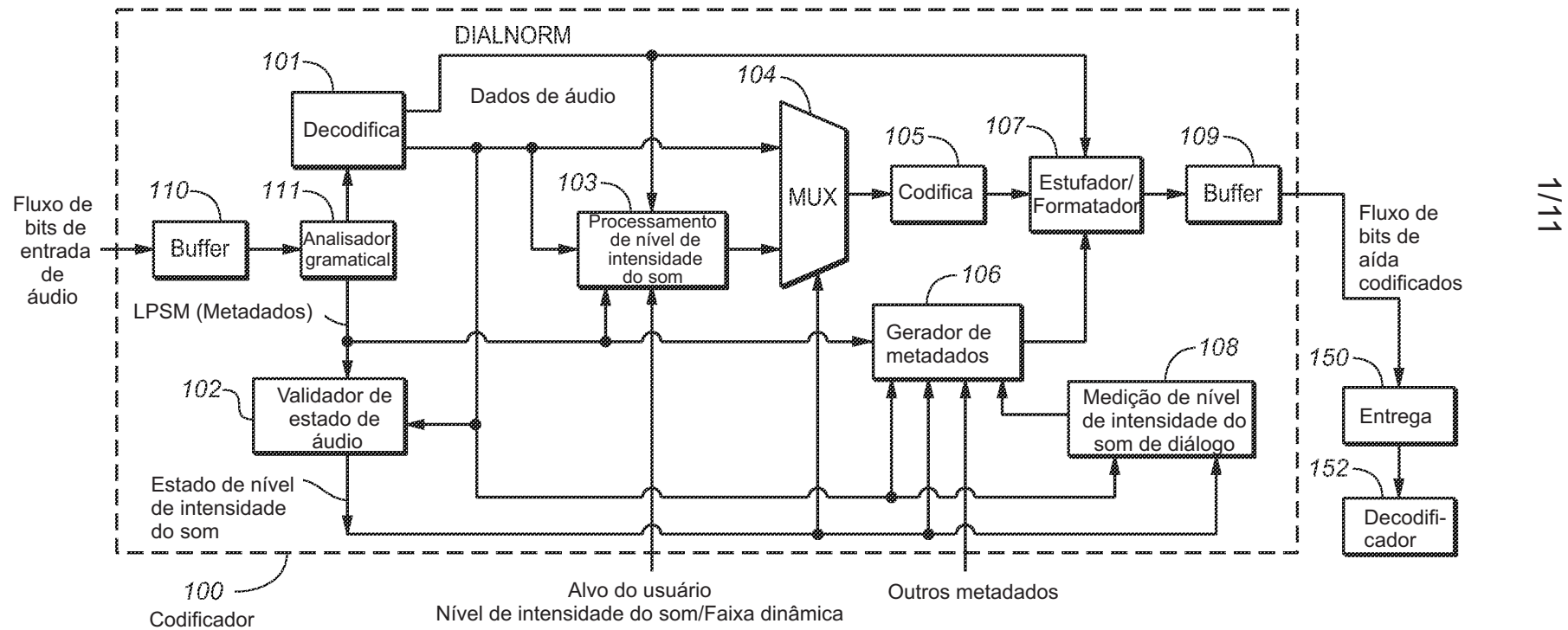


FIG. 2

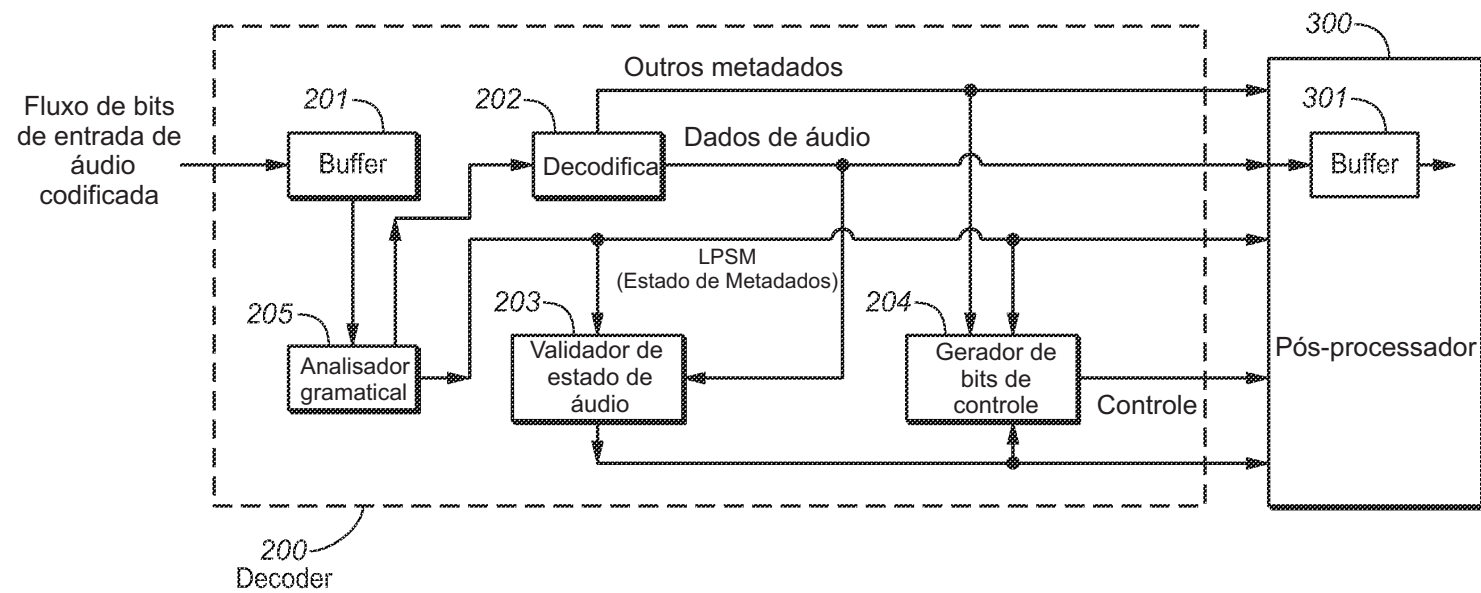


FIG. 3



FIG. 4

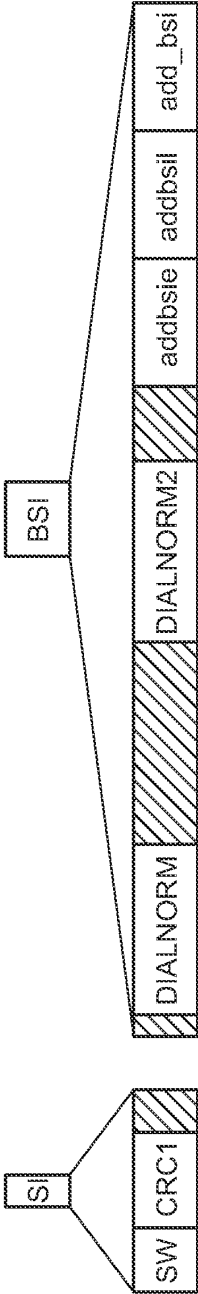


FIG. 5

FIG. 6



FIG. 7

FIGURA 8

Parâmetro	Descrição	Compulsório/Opcional
SYNC [ID]		M
Versão do elemento de núcleo		M
Extensão do elemento de núcleo		M
Período do elemento de núcleo (xxx)		M
Contagem do elemento de núcleo	Indica o número de elementos de metadados estendidos associado com o elemento de núcleo. Este valor pode incrementar/decrementar à medida que o fluxo de bits é designado da produção através da distribuição e emissão final.	M
Associação de subfluxo	Descreve com qual(is) subfluxo(s) o elemento de núcleo está associado.	M
Assinatura (compilação de HMAC)	Compilação HMAC de 256 bits (usando o algoritmo SHA-2) computada através dos dados de áudio, do elemento de núcleo e de todos os elementos expandidos, de todo o quadro.	M
Contagem de limite de PGM	O campo só aparece para algum número de quadros na cabeça ou na cauda de um arquivo/fluxo de programa de áudio. Desse modo, uma mudança de versão de elemento de núcleo pode ser usada para sinalizar a inclusão deste parâmetro.	O
Impressão digital de áudio	Impressão Digital de Áudio tomada sobre algum número de amostras de áudio PCM representadas pelo campo do período de elemento de núcleo.	O
Impressão digital de vídeo	Impressão Digital de Vídeo tomada sobre algum número de amostras de vídeo comprimidas (caso haja alguma) representadas pelo campo do período de elemento de núcleo.	O
URL/UUID	Este campo é definido para conter uma URL e/ou UUID (pode ser redundante para a impressão digital) que faz referência a um a localização estendida de conteúdo de programa adicional (essência) e/ou metadados associados com o fluxo de bits.	O

TABELA 1

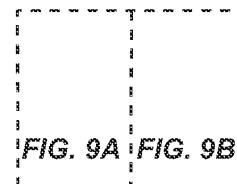
FIGURA 9A

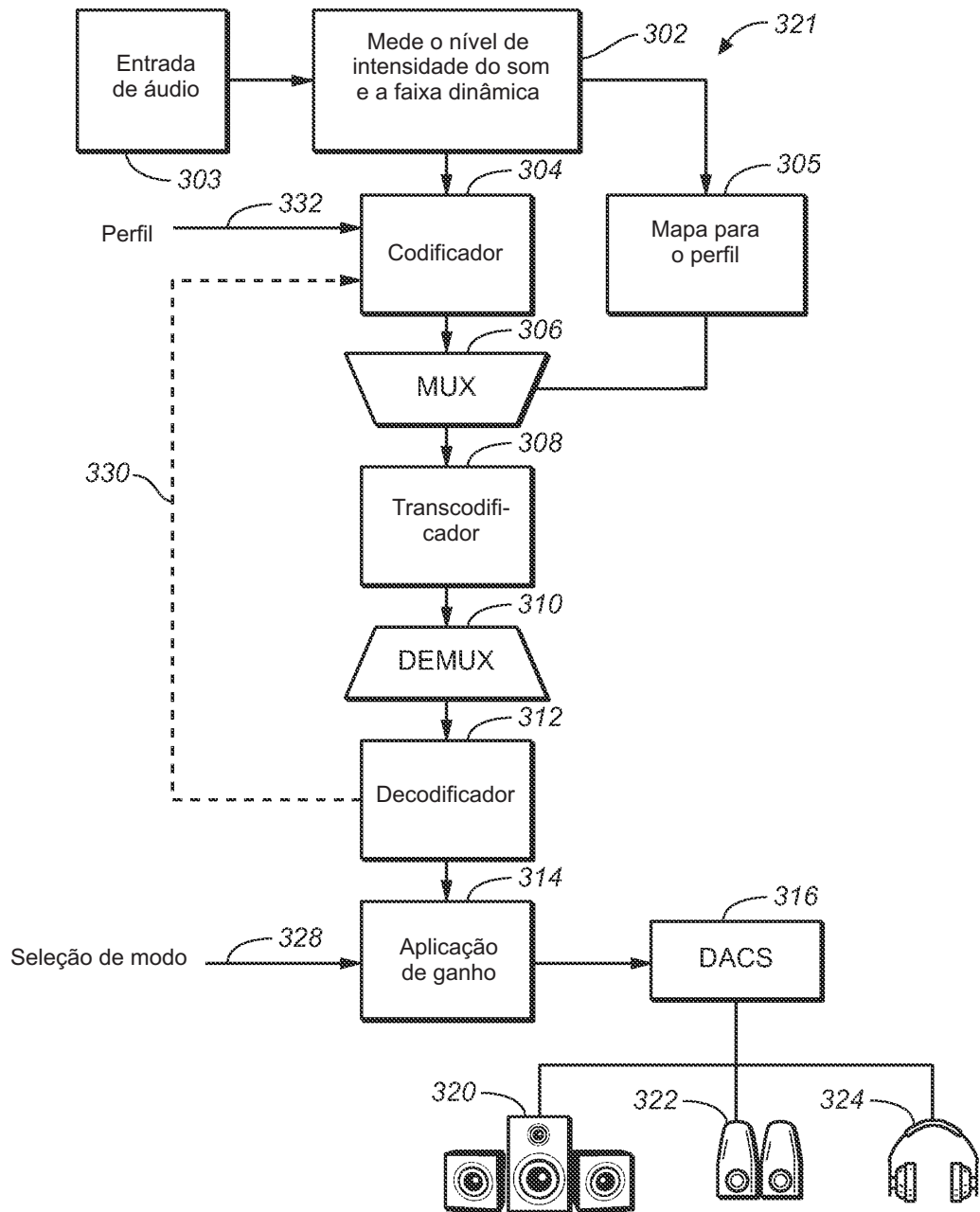
Parâmetro LPSM [Nível de intensidade do som inteligente]	Descrição
Versão de LPSM	
Período de LPSM (xxx)	Aplicável apenas a campos xxx
Contagem de LPSM	
Associação de subfluxo de LPSM	
Canal(is) de diálogo	Indica qual combinação de canais de áudio L, C e R contém fala pelo 0,5 segundo precedente. Quando a fala não está presente em nenhuma combinação de L, C ou R, então este parâmetro deve indicar "sem diálogo"
Tipo de regulação de nível de intensidade do som	Indica que o fluxo de dados de áudio associado está em conformidade com um conjunto específico de regulações (por exemplo, ATSC A/85 ou EBU R128)
Sinalizador de correção de nível de intensidade do som conectado com diálogo	Indica se o fluxo de áudio associado foi corrigido com base na conexão de diálogo
Tipo de correção de nível de intensidade do som	Indica se o fluxo de áudio associado foi corrigido com uma previsão infinita (baseada em arquivo) ou com um controlador de nível de intensidade do som e faixa dinâmica em tempo real (RT)
Nível de intensidade do som conectado relativo de ITU (INF)	Indica o nível de intensidade do som integrado ITU-R BS.1770-2 do fluxo de áudio associado com/ou os metadados aplicados (por exemplo, 7 bits: -58 → +5,5 LKFS etapas de 0,5 LKFS)
Nível de intensidade do som conectado de fala de ITU (INF)	Indica o nível de intensidade do som integrado ITU-R BS.1770-2 do fluxo de áudio associado com/ou os metadados aplicados (por exemplo, 7 bits: -58 → +5,5 LKFS etapas de 0,5 LKFS)
Nível de intensidade do som de 3 s de curta duração de ITU (EBU 3341)	Indica o nível de intensidade do som integrado ITU não conectado de 3 segundos do fluxo de áudio associado com/ou os metadados aplicados (janela deslizante) a uma taxa de
	inserção de ~ 10 Hz (por exemplo, 8 bits: 116 → +11,5 LKFS etapas de 0,5 LKFS)
Valor de Pico real	Indica o valor de pico real de Anexo 2 ITU-R BS.1770 (dB TP) do fluxo de áudio associado com/ou os metadados aplicados (ou seja, o maior valor no período do quadro assinalado no campo do período do elemento) 116 → +11,5 LKFS etapas de 0,5 LKFS)

TABELA 2

Número de estados singulares	Compulsório/Opcional	
	M	
	M	
	M	
	M	
8	M	~0,5 segundo (típico)
8	M	Quadro
2	O (presente apenas se Lodness_Type_Regulation indica que o áudio correspondente NÃO ESTÁ CORRIGIDO)	Quadro
2	O (presente apenas se Lodness_Type_Regulation indica que o áudio correspondente NÃO ESTÁ CORRIGIDO)	Quadro
128	O	1 s
128	O	1 s
256	O	0,1 s
256	O	0,5 s

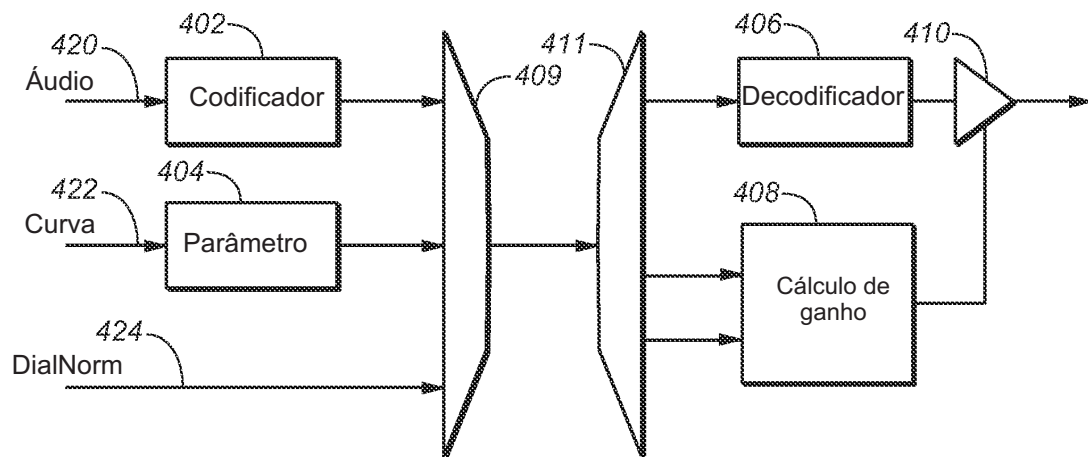
TABELA 2

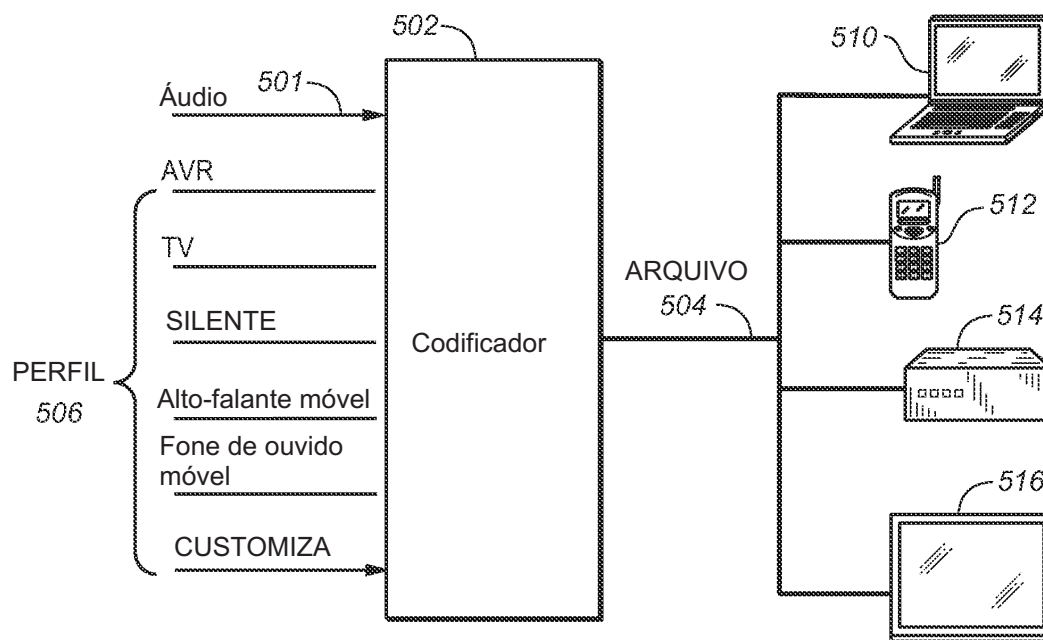
**FIG. 9****FIG. 9B**

**FIG. 10**

	Condição de Palyback	
Alto-falantes	Silentes	Barulhentos
Pequeno/Mini	Faixa dinâmica baixa	Faixa dinâmica baixa
Fones de ouvido	Faixa dinâmica média	Faixa dinâmica baixa
Grande	Faixa dinâmica alta	Faixa dinâmica baixa

TABELA 3

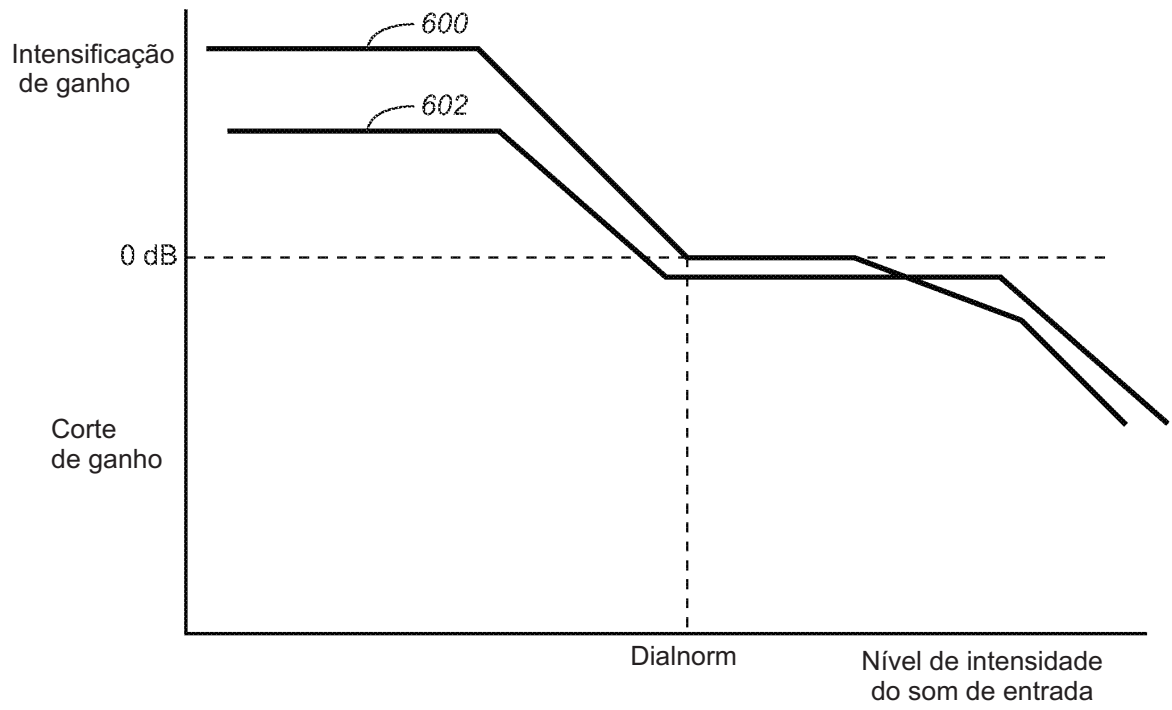
FIG. 11**FIG. 12**

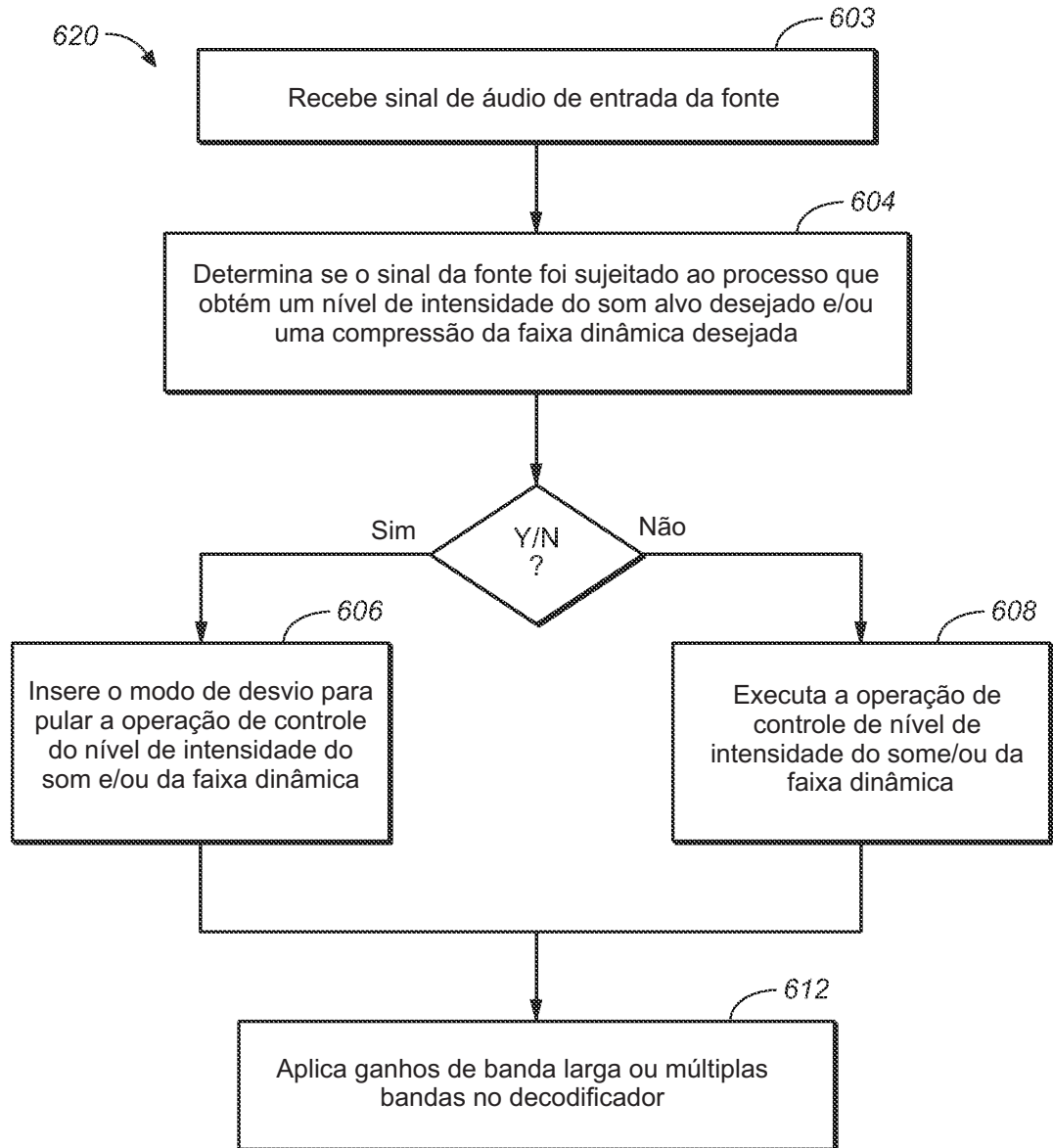
**FIG. 13**

Nível de intensidade do som de longa duração	Ganhos de DRC	Compressão de sobrecarga
1	Perfil 1	1
2	Perfil 2	2
3	Perfil 3	3
⋮	⋮	⋮
N	Perfil N	N

TABELA 4

FIG. 14

**FIG. 15**

**FIG. 16**