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

(11) BR 122020006972-4 B1

(22) Data do Depósito: 17/03/2014

(45) Data de Concessão: 17/05/2022

(54) Título: MÉTODO DE NORMALIZAÇÃO DE VOLUME COM BASE EM UM VALOR DE VOLUME ALVO, APARELHO DE PROCESSAMENTO DE ÁUDIO CONFIGURADO PARA NORMALIZAR O VOLUME COM BASE EM UM VALOR DE VOLUME ALVO E DISPOSITIVO DE ARMAZENAMENTO DE MÉTODO IMPLEMENTADO POR COMPUTADOR LEGÍVEL POR UMA MÁQUINA

(51) Int.Cl.: H03G 3/30; H03G 7/00.

(30) Prioridade Unionista: 11/04/2013 US 61/811,072; 26/03/2013 CN 201310100422.1.

(73) Titular(es): DOLBY LABORATORIES LICENSING CORPORATION.

(72) Inventor(es): JUN WANG; LIE LU; ALAN J. SEEFELDT.

(86) Pedido PCT: PCT US2014030385 de 17/03/2014

(87) Publicação PCT: WO 2014/160542 de 02/10/2014

(85) Data do Início da Fase Nacional: 07/04/2020

(62) Pedido Original do Dividido: BR112015024037-2 - 17/03/2014

(57) Resumo: São divulgados um controlador de nivelador de volume e um método de controle. Em uma modalidade, um controlador de nivelador de volume inclui um classificador de conteúdo de áudio para identificar o tipo de conteúdo de um sinal de áudio em tempo real; e uma unidade de ajuste para ajustar um nivelador de volume de maneira contínua com base no tipo de conteúdo tal como identificado. A unidade de ajuste pode ser configurada para correlacionar positivamente o ganho dinâmico do nivelador de volume com tipos informativos de conteúdo do sinal de áudio, e correlacionar negativamente o ganho dinâmico do nivelador de volume com tipos interferentes de conteúdo de sinal de áudio.

Relatório Descritivo da Patente de Invenção para **"MÉTODO DE NORMALIZAÇÃO DE VOLUME COM BASE EM UM VALOR DE VOLUME ALVO, APARELHO DE PROCESSAMENTO DE ÁUDIO CONFIGURADO PARA NORMALIZAR O VOLUME COM BASE EM UM VALOR DE VOLUME ALVO E DISPOSITIVO DE ARMAZENAMENTO DE MÉTODO IMPLEMENTADO POR COMPUTADOR LEGÍVEL POR UMA MÁQUINA"**.

Dividido do BR112015024037-2 depositado em 17 de março de 2014.

Referência cruzada a pedidos relacionados

[001] Este pedido reivindica prioridade ao Pedido de Patente chinês Nº 201310100422.1, depositado em 26 de março de 2013 e ao Pedido de Patente Provisória dos Estados Unidos Nº 61/811.072, depositado em 11 de abril de 2013, cada um deles sendo incorporado aqui por referência em sua totalidade.

Campo Técnico

[002] O presente pedido refere-se, de modo geral, a um processamento de sinal de áudio. Especificamente, as modalidades do presente pedido se relacionam a aparelhos e métodos para classificação e processamento de áudio, especialmente para controlar o otimizador de diálogo, o virtualizador de surround, o nivelador de volume e o equalizador.

Antecedentes

[003] Alguns dispositivos de melhoria de áudio tendem a modificar os sinais de áudio, em domínio temporal ou espectral, para melhorar a qualidade geral do áudio e aperfeiçoar a experiência dos usuários de maneira correspondente. Vários dispositivos de melhoria de áudio foram desenvolvidos para várias finalidades. Alguns exemplos típicos de dispositivos de melhoria de áudio incluem:

[004] Otimizador de diálogo: O diálogo é o componente mais importante em um programa de rádio ou TV para compreender a

história. Métodos foram desenvolvidos para otimizar os diálogos a fim de aumentar sua clareza e inteligibilidade, sobretudo para idosos com capacidade auditiva decrescente.

[005] Virtualizador de surround: Um virtualizador de surround possibilita que um sinal de som surround (multicanais) seja distribuído aos alto-falantes internos do PC ou nos fones de ouvido. Isto é, com o dispositivo estéreo (como alto-falantes ou fones de ouvido), ele cria um efeito surround virtualmente e proporciona uma experiência cinemática aos consumidores.

[006] Nivelador de volume: Um nivelador de volume direciona-se ao ajuste do volume do conteúdo de áudio em playback e mantê-lo praticamente consistente ao longo de um período de tempo com base no valor alvo de intensidade.

[007] Equalizador: Um equalizador fornece consistência de balanceamento espectral, mais conhecida como "tom" ou "timbre", e permite que os usuários configurem o perfil geral (curva ou formato) da resposta de frequência (ganho) em cada banda de frequência individual, a fim de enfatizar determinados sons ou remover sons indesejados. Em um equalizador tradicional, diferentes pré-ajustes do equalizador podem ser proporcionados a diferentes sons, como gêneros musicais diferentes. A partir do momento em que o pré-ajuste é selecionado, ou que um perfil de equalização é estabelecido, os mesmos ganhos de equalização serão aplicados ao sinal até que o perfil de equalização seja modificado manualmente. Em contraste, um equalizador dinâmico atinge uma consistência de balanceamento espectral pelo monitoramento constante do balanceamento espectral do áudio, comparando-o a um tom desejado e ajustando dinamicamente um filtro de equalização para transformar o tom original do áudio em um tom desejado.

[008] Em geral, dispositivos de melhoria de áudio têm seu próprio

contexto/cenário de aplicação. Isto é, um dispositivo de melhoria de áudio pode ser apropriado para apenas uma determinada série de conteúdos, mas não para todos os sinais de áudio possíveis, pois diferentes conteúdos podem precisar ser processados de maneiras diferentes. Por exemplo, um método de otimização de diálogo costuma ser aplicado ao conteúdo do filme. Se for aplicado à música em que não existem diálogos, pode impulsionar falsamente algumas sub-bandas de frequência e inserir pesadas mudanças de timbre e inconsistência perceptual. De maneira similar, se um método de supressão de ruído for aplicado aos sinais da música, fortes artefatos serão audíveis.

[009] Entretanto, para um sistema de processamento de áudio que geralmente compreende uma série de dispositivos de melhoria de áudio, sua entrada poderia ser inevitavelmente de todos os tipos possíveis de sinais de áudio. Por exemplo, um sistema de processamento de áudio, integrado em um PC, receberá o conteúdo de áudio de uma variedade de fontes, incluindo filmes, músicas, VoIP e jogos. Assim, identificar ou diferenciar o conteúdo a ser processado se torna importante para aplicar melhores algoritmos ou melhores parâmetros de cada algoritmo ao conteúdo correspondente.

[0010] A fim de diferenciar o conteúdo de áudio e aplicar melhores parâmetros ou melhores algoritmos de melhora de áudio, de maneira correspondente, os sistemas tradicionais costumam pré-projetar uma série de pré-ajustes e solicita-se que os usuários selecionem um pré-ajuste para o conteúdo a ser tocado. Um pré-ajuste geralmente codifica uma série de algoritmos de melhoria de áudio e/ou seus melhores parâmetros que serão aplicados, como pré-ajuste "Filme" e um pré-ajuste "Música" que é projetado especificamente para um filme ou playback de música.

[0011] Todavia, a seleção manual é inconveniente para os

usuários. Os usuários, em geral, não alternam com frequência entre os pré-ajustes predefinidos, mas simplesmente mantêm-se usando um pré-ajuste para todo o conteúdo. Além disso, mesmo em algumas soluções automáticas, os parâmetros ou a configuração de algoritmos nos pré-ajustes são geralmente discretas (como ligar ou desligar em um algoritmo específico em relação a um conteúdo específico), eles não podem ajustar parâmetros de uma maneira contínua com base no conteúdo.

Sumário

[0012] O primeiro aspecto do presente pedido é configurar automaticamente os dispositivos de melhoria de áudio de uma maneira contínua, com base no conteúdo de áudio no playback. Com esse modo "automático", os usuários podem simplesmente aproveitar o conteúdo sem se preocupar em selecionar diferentes pré-ajustes. Por um lado, um ajuste contínuo é mais importante para evitar artefatos audíveis em pontos de transição.

[0013] De acordo com uma modalidade do primeiro aspecto, um aparelho de processamento de áudio inclui um classificador de áudio para classificar um sinal de áudio em pelo menos um tipo de áudio em tempo real; um dispositivo de melhoria de áudio para melhorar a experiência da audiência; e uma unidade de ajuste para ajustar pelo menos um parâmetro do dispositivo de melhoria de áudio de uma maneira contínua com base no valor de confiança de um tipo de áudio.

[0014] O dispositivo de melhoria de áudio pode ser qualquer otimizador de diálogo, visualizador surround, nivelador de volume e equalizador.

[0015] De maneira correspondente, um método de processamento de áudio inclui: classificar um sinal de áudio em pelo menos um tipo de áudio em tempo real; e ajustar pelo menos um parâmetro para a melhora de áudio de maneira contínua com base no valor de confiança

de pelo menos um tipo de áudio.

[0016] De acordo com uma outra modalidade do primeiro aspecto, um controlador de nivelador de volume inclui um classificador de conteúdo de áudio para identificar o tipo de conteúdo de um sinal de áudio em tempo real; e uma unidade de ajuste para ajustar um nivelador de volume de maneira contínua com base no tipo de conteúdo, conforme identificado. A unidade de ajuste pode ser configurada para se correlacionar de maneira coletiva a um ganho dinâmico do nivelador de volume com tipos de conteúdo informativo do sinal de áudio e correlacionar negativamente o ganho dinâmico do nivelador de volume com os tipos de conteúdo de interferência do sinal de áudio.

[0017] É divulgado, ademais, um aparelho de processamento de áudio que compreende um controlador de nivelador de volume, conforme afirmado acima.

[0018] De maneira correspondente, um método de controle do nivelador de volume inclui: a identificação do tipo de conteúdo de um sinal de áudio em tempo real; e o ajuste de um nivelador de volume de maneira contínua com base no tipo de conteúdo, conforme identificado, pela correlação positiva do ganho dinâmico do nivelador de volume com os tipos de conteúdo informativo do sinal de áudio e pela correlação negativa do ganho dinâmico do nivelador de volume com os tipos de conteúdo de interferência do sinal de áudio.

[0019] De acordo com outra modalidade do primeiro aspecto, ainda, um controlador de equalizador inclui um classificador de áudio para identificar o tipo de áudio de um sinal de áudio em tempo real; e uma unidade de ajuste para ajustar um equalizador de uma maneira contínua com base no valor de confiança do tipo de áudio, conforme identificado.

[0020] É divulgado, ademais, um aparelho de processamento de

áudio que compreende um controlador de equalizador, conforme afirmado acima.

[0021] De maneira correspondente, um método de controle de equalizador inclui: identificar um tipo de áudio de um sinal de áudio em tempo real; e uma unidade de ajuste para ajustar um equalizador de uma maneira contínua com base no valor de confiança do tipo de áudio, conforme identificado.

[0022] O presente pedido fornece ainda um meio legível em computador com instruções de programa de computador gravadas nele, quando executado por um processador, as instruções permitindo que o processador execute o método de processamento de áudio supracitado, ou o método de controle do nivelador de volume, ou o método de controle do equalizador.

[0023] De acordo com as modalidades do primeiro aspecto, o dispositivo de melhoria de áudio, que pode ser: o otimizador de diálogo, o virtualizador de surround, o nivelador de volume e o equalizador, pode ser ajustado continuamente de acordo com o tipo de sinal de áudio e/ou o valor de confiança do tipo.

[0024] O segundo aspecto do presente pedido é desenvolver um componente de identificação de conteúdo para identificar múltiplos tipos de áudio, e os resultados detectados podem ser usados para controlar/orientar os comportamentos de vários dispositivos de melhoria de áudio, encontrando melhores parâmetros de maneira contínua.

[0025] De acordo com uma modalidade do segundo aspecto, um classificador de áudio inclui: um extrator de recursos de curta duração para extrair recursos de curta duração de segmentos de áudio de curta duração, cada um compreendendo uma sequência de frames de áudio; um classificador de curta duração para classificar uma sequência de segmentos de curta duração em um segmento de áudio

de longa duração em tipos de áudio de curta duração utilizando seus respectivos recursos de curta duração; um extrator de estatísticas para calcular as estatísticas dos resultados do classificador de curta duração em relação à sequência de segmentos de curta duração no segmento de áudio de longa duração, como recursos de longa duração; e um classificador de longa duração para classificar, usando recursos de longa duração, o segmento de áudio de longa duração em tipos de áudio de longa duração.

[0026] É divulgado também um aparelho de processamento de áudio que compreende um classificador de áudio, como indicado acima.

[0027] De maneira correspondente, um método de classificação de áudio inclui: a extração de recursos de curta duração de segmentos de áudio de curta duração, cada um compreendendo uma sequência de frames de áudio; a classificação de uma sequência de segmentos de curta duração em um segmento de áudio de longa duração em tipos de áudio de curta duração utilizando os respectivos recursos de curta duração; o cálculo das estatísticas dos resultados da operação de classificação em relação à sequência de segmentos de curta duração no segmento de áudio de longa duração, como recursos de longa duração; e a classificação do segmento de áudio de longa duração em tipos de áudio de longa duração utilizando os recursos de longa duração.

[0028] De acordo com outra modalidade do segundo aspecto, um classificador de áudio inclui: um classificador de conteúdo de áudio para identificar o tipo de conteúdo de um segmento de curta duração de um sinal de áudio e um classificador de contexto de áudio para identificar um tipo de contexto do segmento de curta duração pelo menos parcialmente com base no tipo de contexto identificado pelo classificador do conteúdo de áudio.

[0029] É divulgado, ademais, um aparelho de processamento de áudio que compreende um classificador de áudio, conforme indicado acima.

[0030] De maneira correspondente, um método de classificação de áudio inclui: a identificação de um tipo de conteúdo de um segmento de curta duração de um sinal de áudio; e a identificação de um tipo de contexto do segmento de curta duração, pelo menos parcialmente com base no tipo de conteúdo, conforme identificado.

[0031] O presente pedido fornece ainda um meio legível em computador com instruções de programa de computador gravadas nele, quando executado por um processador, as instruções permitindo que o processador execute os métodos de classificação de áudio supramencionados.

[0032] De acordo com as modalidades do segundo aspecto, um sinal de áudio pode ser classificado em diferentes tipos de contexto ou de longa duração, que são diferentes dos tipos de conteúdo ou de curta duração. Os tipos de sinal de áudio e/ou o valor de confiança dos tipos podem ser utilizados ainda para ajustar um dispositivo de melhoria de áudio, como um otimizador de diálogo, um virtualizador de surround, um nivelador de volume ou um equalizador.

Breve Descrição das Figuras

[0033] O presente pedido é ilustrado a título de exemplo, e não a título de limitação, nas figuras das ilustrações em anexo, em que numerais de referência semelhantes se referem a elementos semelhantes e em que:

[0034] A Figura 1 é um diagrama que ilustra um aparelho de processamento de áudio, de acordo com uma modalidade do pedido;

[0035] As Figuras 2 e 3 são diagramas que ilustram variantes da modalidade, como mostrado na Figura 1;

[0036] As Figuras 4-6 são diagramas que ilustram a possível

arquitetura dos classificadores para identificar múltiplos tipos de áudio e o cálculo do valor de confiança;

[0037] As Figuras 7-9 são diagramas que ilustram mais modalidades do aparelho de processamento de áudio do presente pedido;

[0038] A Figura 10 é um diagrama que ilustra uma transição entre diferentes tipos de áudio;

[0039] As Figuras 11-14 são fluxogramas que ilustram um método de processamento de áudio de acordo com modalidades do presente pedido;

[0040] A Figura 15 é um diagrama que ilustra um controlador de otimizador de diálogo de acordo com uma modalidade do presente pedido;

[0041] As Figuras 16 e 17 são fluxogramas que ilustram o uso do método de processamento de áudio de acordo com o presente pedido no controle de um otimizador de diálogo;

[0042] A Figura 18 é um diagrama que ilustra um controlador do virtualizador de surround, de acordo com uma modalidade do presente pedido;

[0043] A Figura 19 é um fluxograma que ilustra o uso do método de processamento de áudio, de acordo com o presente pedido, no controle de um virtualizador de surround;

[0044] A Figura 20 é um diagrama que ilustra um controlador de nivelador de volume, de acordo com uma modalidade do presente pedido;

[0045] A Figura 21 é um diagrama que ilustra o efeito do controlador de nivelador de volume, de acordo com o presente pedido;

[0046] A Figura 22 é um diagrama que ilustra um controlador de equalizador, de acordo com uma modalidade do presente pedido;

[0047] A Figura 23 ilustra diversos exemplos de pré-ajustes de

balanceamento espectral desejados;

[0048] A Figura 24 é um diagrama que ilustra um classificador de áudio, de acordo com uma modalidade do presente pedido;

[0049] As Figuras 25 e 26 são diagramas que ilustram alguns recursos a serem usados pelo classificador de áudio do presente pedido;

[0050] As Figuras 27-29 são diagramas que ilustram mais modalidades do classificador de áudio, de acordo com o presente pedido;

[0051] As Figuras 30-33 são fluxogramas que ilustram um método de classificação de áudio de acordo com modalidades do presente pedido;

[0052] A Figura 34 é um diagrama que ilustra um classificador de áudio, de acordo com outra modalidade do presente pedido;

[0053] A Figura 35 é um diagrama que ilustra um classificador de áudio, de acordo com outra modalidade do presente pedido, ainda;

[0054] A Figura 36 é um diagrama que ilustra as normas heurísticas usadas no classificador de áudio do presente pedido;

[0055] As Figuras 37 e 38 são diagramas que ilustram mais modalidades do classificador de áudio, de acordo com o presente pedido;

[0056] As Figuras 39 e 40 são fluxogramas que ilustram um método de classificação de áudio de acordo com modalidades do presente pedido;

[0057] A Figura 41 é um diagrama de blocos que ilustra um sistema exemplar para implementar modalidades do presente pedido.

Descrição Detalhada

[0058] As modalidades do presente pedido são descritas abaixo referindo-se às figuras. Deve ser notas que, para fins de clareza, as representações e descrições quanto a esses componentes e

processos conhecidos por aqueles versados na técnica, mas não necessárias para compreender o presente pedido, são omitidas nas figuras e na descrição.

[0059] Como será apreciado por uma pessoa versada na técnica, os aspectos do presente pedido podem ser incorporados como um sistema, um dispositivo (por exemplo, um telefone celular, um reprodutor de mídia portátil, um computador pessoal, um servidor, um set-top box de televisão ou um gravador de vídeo digital, ou qualquer outro reprodutor de mídia), um método ou um produto de programa de computador. Com efeito, os aspectos do presente pedido podem assumir a forma de uma modalidade de hardware, uma modalidade de software (incluindo firmware, software residente, microcódigos, etc.) ou uma modalidade combinando aspectos de software e hardware que podem todos, de maneira geral, ser referidos aqui como "circuito", "módulo" ou "sistema". Além disso, os aspectos do presente pedido podem assumir a forma de um produto de programa de computador incorporado em um ou mais meios legíveis em computador com um código de programa legível em computador a eles acoplados.

[0060] Qualquer combinação de um ou mais meios legíveis em computador pode ser utilizada. O meio legível em computador pode ser um meio de sinal legível em computador ou um meio de armazenamento legível em computador. Um meio de armazenamento legível por computador pode ser, por exemplo, mas sem se limitar a, um sistema, aparelho ou dispositivo eletrônico, magnético, óptico, eletromagnético, infravermelho, ou de semicondutores, ou qualquer combinação adequada dos supracitados. Exemplos mais específicos (uma lista não exaustiva) de meio de armazenamento legível por computador incluem o seguinte: uma conexão elétrica que tem um ou mais fios, um disquete de computador portátil, um disco rígido, uma memória de acesso aleatório (RAM), uma memória somente leitura

(ROM), uma memória programável apagável somente de leitura (EPROM ou memória Flash), uma fibra óptica, uma disco compacto portátil somente de leitura (CD-ROM), um dispositivo de armazenamento óptico, um dispositivo de armazenamento magnético, ou qualquer combinação adequada dos anteriores. No contexto deste documento, um meio de armazenamento legível em computador pode ser qualquer meio tangível que possa conter ou armazenar um programa para uso por um em conexão com um dispositivo, aparelho ou sistema de execução de instruções.

[0061] Um meio de sinal legível em computador pode incluir um sinal de dados propagado com um código de programa legível em computador a ele incorporado, por exemplo, em banda de base ou como parte de uma onda de transportadora. Esse sinal propagado pode assumir qualquer uma dentre uma variedade de formas, incluindo, sem se limitar a, um sinal ótico ou eletromagnético ou qualquer combinação apropriada destes.

[0062] Um meio de sinal legível em computador pode ser qualquer meio legível em computador que não seja um meio de armazenamento legível em computador e possa comunicar, propagar ou transportar um programa para uso por ou em conexão com um dispositivo, aparelho ou sistema de execução de instruções.

[0063] O código de programa incorporado a um meio legível em computador pode ser transmitindo utilizando qualquer meio apropriado, incluindo, sem se limitar a, cabo de fibra ótica, RF, linha com fio, sem fio, etc., ou qualquer combinação apropriada destes.

[0064] O código do programa de computador para realizar operações para os aspectos do presente pedido pode ser escrito em qualquer combinação de uma ou mais linguagens de programação, incluindo uma linguagem de programação orientada para o objeto, como Java, Smalltalk, C++ ou afins e linguagens de programação

processuais convencionais, como linguagem de programação "C" ou linguagens de programação semelhantes. O código do programa pode executar inteiramente no computador do usuário como um pacote de software autônomo ou parcialmente no computador do usuário e parcialmente em um computador remoto ou inteiramente no servidor ou computador remoto. No último cenário, o computador remoto pode estar conectado ao computador do usuário através de qualquer tipo de rede, incluindo uma rede de área local (LAN) ou uma rede de longa distância (WAN), ou a conexão pode ser feita para um computador externo (por exemplo, através da Internet usando um Provedor de Serviços de Internet).

[0065] Os aspectos do presente pedido encontram-se descritos abaixo com referência às ilustrações do fluxograma e/ou aos diagramas de bloco dos métodos, aparelhos (sistemas) e produtos de programa de computador de acordo com modalidades do pedido. Será entendido que cada bloco (também referido como bloco) do das ilustrações do fluxograma e/ou diagramas de bloco e as combinações de blocos nas ilustrações do fluxograma e/ou diagramas de bloco pode ser implementado por instruções de programa legíveis por computador. Estas instruções de programa de computador podem ser fornecidas a um processador de um computador com finalidade geral, computador com finalidade especial, ou outro aparelho de processamento de dados programáveis para produzir uma máquina, de modo que as instruções, que executam via processador do computador ou outro aparelho de processamento de dados programáveis, criam meios para a implementação de funções/atos especificados no fluxograma e/ou no bloco ou nos blocos do diagrama de bloco.

[0066] Essas instruções de programa de computador também podem ser armazenadas em um meio legível em computador que

pode direcionar um computador, outros aparelhos de processamento de dados programáveis ou outros dispositivos para funcionar de um jeito específico, de modo que as instruções armazenadas no meio legível em computador produzam um artigo de produção incluindo instruções que implementem a função/o ato especificado no fluxograma e/ou no bloco ou blocos do diagrama de bloco.

[0067] As instruções de programa de computador podem também ser carregadas em um computador, outros aparelhos de processamento de dados, ou outros dispositivos para fazer com que uma série de operações operacionais seja realizada no computador, outros aparelhos ou dispositivos para produzir um processo implementado por computador, de modo que as instruções executadas no computador ou outros aparelhos programáveis forneçam processos para implementar funções/atos especificados no fluxograma e/ou blocos ou bloco do diagrama de bloco.

[0068] Abaixo serão descritas em detalhes as modalidades do presente pedido. Para maior clareza, a descrição é organizada na seguinte arquitetura:

[0069] Parte 1: Aparelhos e métodos de processamento de áudio

[0070] Seção 1.1 Tipos de áudio

[0071] Seção 1.2 Valores de confiança de tipos de áudio e arquitetura dos classificadores Seção 1.3 Unificação dos valores de confiança dos tipos de áudio

[0072] Seção 1.4 Ajuste de parâmetro

[0073] Seção 1.5 Uniformização de parâmetro

[0074] Seção 1.6 Transição de tipos de áudio

[0075] Seção 1.7 Combinação das modalidades e cenários de aplicação

[0076] Seção 1.8 Método de processamento de áudio

[0077] Parte 2: Controlador do otimizador de diálogo e método de

controle

[0078] Seção 2.1 Nível de otimização do diálogo

[0079] Seção 2.2 Limites para determinar bandas de frequência a serem potencializadas

[0080] Seção 2.3 Ajuste ao nível de fundo

[0081] Seção 2.4 Combinação das modalidades e cenários de aplicação

[0082] Seção 2.5 Método de controle do otimizador de diálogo

[0083] Parte 3: Controlador do Virtualizador de surround e Método de Controle

[0084] Seção 3.1 Quantidade de Aumento de Surround

[0085] Seção 3.2 Frequência inicial

[0086] Seção 3.3 Combinação das modalidades e cenários de aplicação

[0087] Seção 3.4 Método de controle do virtualizador de surround

[0088] Parte 4: Controlador do nivelador de volume e método de controle

[0089] Seção 4.1 Tipos informativos de conteúdo de intervenção

[0090] Seção 4.2 Tipos de conteúdo em diferentes contextos

[0091] Seção 4.3 Tipos de contexto

[0092] Seção 4.4 Combinação das modalidades e cenários de aplicação

[0093] Seção 4.5 Método de controle do nivelador de volume

[0094] Parte 5: Controlador do equalizador e método de controle

[0095] Seção 5.1 Controle com base no tipo de conteúdo

[0096] Seção 5.2 Probabilidade de fontes dominantes na música

[0097] Seção 5.3 Pré-ajustes do equalizador

[0098] Seção 5.4 Controle com base no tipo de contexto

[0099] Seção 5.5 Combinação das modalidades e cenários de aplicação

[00100] Seção 5.6 Método de controle do equalizador

[00101] Parte 6: Classificadores de áudio e métodos de classificação

[00102] Seção 6.1 Classificador de contexto com base na classificação do tipo de conteúdo

[00103] Seção 6.2 Extração de característica de longa duração

[00104] Seção 6.3 Extração de características de curta duração

[00105] Seção 6.4 Combinação das modalidades e cenários de aplicação

[00106] Seção 6.5 Métodos de classificação de áudio

[00107] Parte 7: Classificadores de VoIP e métodos de classificação

[00108] Seção 7.1 Classificação de contexto com base no segmento de curta duração

[00109] Seção 7.2 Classificação usando fala de VoIP e ruído VoIP

[00110] Seção 7.3 Flutuação de uniformização

[00111] Seção 7.4 Combinação das modalidades e cenários de aplicação

[00112] Seção 7.5 Métodos de classificação de VoIP

[00113] Parte 1: Aparelho de processamento de áudio e métodos

[00114] A Figura 1 ilustra uma estrutura geral de um aparelho de processamento de áudio adaptativo ao conteúdo 100 que suporta uma configuração automática de pelo menos um dispositivo de melhoria de áudio 400 com parâmetros melhorados com base no conteúdo de áudio em playback. Ela compreende três componentes principais: um classificador de áudio 200, uma unidade de ajuste 300 e um dispositivo de melhoria de áudio 400.

[00115] O classificador de áudio 200 é para classificar um sinal de áudio em pelo menos um tipo de áudio em tempo real. Ele identifica automaticamente os tipos de áudio do conteúdo em playback. Qualquer uma das tecnologias de classificação de áudio, como através

de processamento de sinal, aprendizagem de máquina e reconhecimento de padrão, podem ser aplicadas para identificar o conteúdo de áudio. Os valores de confiança, que representam as probabilidades do conteúdo de áudio quanto a uma série de tipos de áudio alvo pré-definidos, são estimados realmente ao mesmo tempo.

[00116] O dispositivo de melhoria de áudio 400 é para melhorar a experiência da audiência, realizando um processamento sobre o sinal de áudio, e será discutido em detalhes posteriormente.

[00117] A unidade de ajuste 300 é para ajustar pelo menos um parâmetro do dispositivo de melhoria de áudio de uma maneira contínua com base no valor de confiança de pelo menos um tipo de áudio. Ela é projetada para controlar o comportamento do dispositivo de melhoria de áudio 400. Ela estima os parâmetros mais apropriados do dispositivo de melhoria de áudio correspondente com base nos resultados obtidos a partir do classificador de áudio 200.

[00118] Vários dispositivos de melhoria de áudio podem ser aplicados nesse aparelho. A Figura 2 mostra um sistema de exemplo que compreende quatro dispositivos de melhoria de áudio, incluindo o otimizador de Diálogo (DE) 402, o Virtualizador Surround (SV) 404, o Nivelador de Volume (VL) 406 e o Equalizador (EQ) 408. Cada dispositivo de melhoria de áudio pode ser ajustado automaticamente de maneira contínua, com base nos resultados (valores de confiança e/ou tipos de áudio) obtidos no classificador de áudio 200.

[00119] Naturalmente, o aparelho de processamento de áudio pode não incluir, necessariamente, todos os tipos de dispositivos de melhoria de áudio, mas pode incluir apenas um ou mais deles. Por outro lado, os dispositivos de melhoria de áudio não são limitados àqueles dispositivos fornecidos na presente divulgação e podem incluir mais tipos de dispositivos de melhoria de áudio que também fazem parte do escopo do presente pedido. Além disso, os nomes daqueles

dispositivos de melhoria de áudio discutidos na presente divulgação, incluindo otimizador de Diálogo (DE) 402, Virtualizador surround (SV) 404, Nivelador de Volume (VL) 406 e Equalizador (EQ) 408, não devem constituir uma limitação e cada um deles deve ser considerado como abrangendo quaisquer outros dispositivos que realizam funções semelhantes ou as mesmas funções.

1.1 Tipos de áudio

[00120] Para controlar corretamente vários tipos de dispositivo de melhoria de áudio, o presente pedido fornece ainda uma nova arquitetura de tipos de áudio, ainda que esses tipos de áudio do estado da técnica também sejam aplicáveis aqui.

[00121] Especificamente, os tipos de áudio de diferentes níveis semânticos são modelos, incluindo elementos de áudio de nível baixo que representam os componentes fundamentais nos sinais de áudio e nos gêneros de áudio de nível alto que representam os conteúdos de áudio mais populares nas aplicações de entretenimento do usuário na vida real. O anterior também pode ser denominado "tipo de conteúdo". Os tipos de conteúdo de áudio fundamentais podem incluir fala, música (incluindo sons), sons de fundo (ou efeitos sonoros) e ruídos.

[00122] O significado de fala e músicas é auto-evidente. O ruído no presente pedido significa ruído físico, não ruído semântico. O ruído físico no presente pedido pode incluir os ruídos, por exemplo, de ar condicionados e aqueles ruídos que se originam de motivos técnicos, como ruídos rosas por conta da passagem de transmissão de sinal. Em contraste, os "sons de fundo" no presente pedido são aqueles efeitos sonoros que podem ser eventos auditivos que ocorrem em torno do alvo principal da atenção do ouvinte. Por exemplo, em um sinal de áudio em uma ligação telefônica, além da voz de quem fala, também pode haver outros sons não intencionais, como vozes de outras pessoas irrelevantes para a ligação telefônica, sons de teclado,

sons de passos e assim por diante. Esses sons indesejados são referidos como "sons de fundo", não ruídos. Em outras palavras, pode-se definir "sons de fundo" como aqueles sons que não são o alvo (ou o alvo principal da atenção do ouvinte), ou mesmo não desejados, mas que ainda têm valor semântico; enquanto que "ruídos" podem ser definidos como sons indesejados, exceto pelos sons alvo e pelos sons de fundo.

[00123] Às vezes os sons do fundo são realmente não "indesejados"< mas criados intencionalmente e portam algumas informações úteis, como aqueles sons de fundo em um filme, programa de TV ou programa de rádio. Assim, às vezes eles também podem ser definidos como "efeitos sonoros". Doravante, na presente divulgação, apenas "sons de fundo" é usado por questão de concisão e pode ser abreviado ainda como "fundo".

[00124] Ademais, a música pode ser classificada ainda como música sem fontes dominantes e música com fontes dominantes. Se houver uma fonte (voz ou um instrumento) que é muito mais intensa do que as outras fontes em uma parte da música, ela é denominada "música com fonte dominante"; do contrário, ela é denominada "música sem fonte dominante". Por exemplo, em uma música polifônica acompanhada de vocal e vários instrumentos, se estiver harmonicamente equilibrado, ou a energia de várias fontes mais salientes forem comparáveis umas às outras, considera-se uma música sem fonte dominante; em contraste, se uma fonte (por exemplo, voz) for muito mais alta e as outras forem muito mais baixas, considera-se que contém uma fonte dominante. Como um outro exemplo, os tons de instrumentos singulares e distintivos são "músicas com fonte dominante".

[00125] A música pode ser classificada ainda como diferentes tipos com base em diferentes padrões. Pode ser classificado com base nos

gêneros da música, como rock, jazz, rap e folk, mas não se limita a estes. Pode também ser classificado com base nos instrumentos, como música vocal e instrumental. A música instrumental pode incluir várias músicas reproduzidas com diferentes instrumentos, como piano e violão. Outros padrões exemplificativos incluem ritmo, tempo, timbre da música e/ou qualquer outro atributo musical, de modo que a música possa ser agrupada com base na semelhança de tais atributos. Por exemplo, de acordo com o timbre, o vocal pode ser classificado como tenor, barítono, baixo, soprano, mezzo soprano e alto.

[00126] O tipo de conteúdo de um sinal de áudio pode ser classificado em relação aos segmentos de áudio de curta duração, como compreendido em uma pluralidade de frames. Geralmente, um frame de áudio é uma extensão de múltiplos milissegundos, como 20 ms, e a extensão de um segmento de áudio de curta duração a ser classificado pelo classificador de áudio pode ter uma extensão de várias centenas de milissegundos a vários segundos, como 1 segundo.

[00127] Para controlar o dispositivo de melhoria de áudio de uma maneira adaptativa ao conteúdo, o sinal de áudio pode ser classificado em tempo real. Para o tipo de conteúdo indicado acima, o tipo de conteúdo do presente segmento de áudio de curta duração representa o tipo de conteúdo do presente sinal de áudio. Visto que a extensão de um segmento de áudio de curta duração não é muito longa, o sinal de áudio pode ser dividido como segmentos de áudio de curta duração não-sobrepostos, um após o outro. Entretanto, os segmentos de áudio de curta duração também podem ser amostrados continuamente/semi-continuamente ao longo da linha de tempo do sinal de áudio. Isto é, os segmentos de áudio de curta duração podem ser amostrados com uma janela com uma extensão pré-determinada (extensão pretendida do segmento de áudio de curta duração) que se move ao longo da linha de

tempo do sinal de áudio a um tamanho de etapa de um ou mais frames.

[00128] Os gêneros de áudio de nível elevado podem ser denominados "tipo de contexto", pois indicam um tipo de longa duração do sinal de áudio e podem ser considerados um ambiente ou contexto do evento sonoro do momento, o qual pode ser classificado em tipos de conteúdo, como indicado acima. De acordo com o presente pedido, o tipo de contexto pode incluir as aplicações de áudio mais populares, como mídias do tipo filme, músicas (incluindo sons), jogos e VoIP (Voice on Internet Protocol).

[00129] O significado de músicas, jogos e VoIP é auto-evidente. As mídias do tipo filme podem incluir filmes, programas de TV, programas de rádio ou qualquer outra mídia de áudio semelhante supramencionada. A característica principal das mídias do tipo filme é uma mistura de possíveis falas, músicas e vários tipos de sons de fundo (efeitos sonoros).

[00130] Pode-se notar que o tipo de conteúdo e o tipo de contexto incluem músicas (incluindo sons). Doravante, no presente pedido, foram utilizados, os termos "músicas de curta duração" e "músicas de longa duração" para distingui-las, respectivamente.

[00131] Para algumas modalidades do presente pedido, algumas outras arquiteturas de tipo de contexto são propostas.

[00132] Por exemplo, um sinal de áudio pode ser classificado como áudio de alta qualidade (como mídias do tipo filme e CDs de música) ou áudio de baixa qualidade (como VoIP, áudio de streaming online de taxa de bits baixa e conteúdo gerado por usuário), que pode ser coletivamente denominado "tipos de qualidade de áudio".

[00133] Como outro exemplo, um sinal de áudio pode ser classificado como VoIP e não-VoIP, que pode ser considerado como uma transformação na arquitetura de tipo de 4 contextos mencionada acima (VoIP, mídias do tipo filme, músicas (de longa duração) e

jogos). Em relação ao contexto de VoIP ou de non-VoIP, um sinal de áudio pode ser classificado como tipos de conteúdo de áudio relacionado a VoIP, como fala em VoIP, fala não-VoIP, ruído VoIP e ruído não-VoIP. A arquitetura de tipos de conteúdo de áudio em VoIP é especialmente útil para diferenciar contextos VoIP e não-VoIP, visto que o contexto VoIP geralmente é o cenário de aplicação mais desafiador de um nivelador de volume (um tipo de dispositivo de melhoria de áudio).

[00134] Geralmente, o tipo de contexto de um áudio de sinal pode ser classificado em relação a segmentos de áudio de longa duração mais longos do que os segmentos de áudio de curta duração. Um segmento de áudio de longa duração é compreendido de uma pluralidade de frames em uma quantidade superior à quantidade de frames em um segmento de áudio de curta duração. Um segmento de áudio de longa duração também pode ser compreendido de uma pluralidade de segmentos de áudio de curta duração. Geralmente, um segmento de áudio de longa duração pode ter uma extensão na ordem de segundos, como de vários segundos a dezenas de segundos, por exemplo, 10 segundos.

[00135] Para controlar o dispositivo de melhoria de áudio de uma maneira adaptativa, o sinal de áudio pode ser classificado em tipos de contexto em tempo real. De modo semelhante, o tipo de contexto do presente segmento de áudio de longa duração representa o tipo de contexto do presente sinal de áudio. Visto que a extensão de um segmento de áudio de longa duração é relativamente longa, o sinal de áudio pode ser amostrado continuamente/semi-continuamente ao longo da linha de tempo do sinal de áudio para evitar uma mudança abrupta de seu tipo de contexto e, portanto, uma mudança abrupta dos parâmetros de funcionamento do(s) dispositivo(s) de melhoria de áudio. Isto é, os segmentos de áudio de longa duração podem ser amostrados com uma

janela com uma extensão pré-determinada (extensão pretendida do segmento de áudio de longa duração) que se move ao longo da linha de tempo do sinal de áudio a um tamanho de etapa de um ou mais frames ou de um ou mais segmentos de curta duração.

[00136] Acima, foram descritos o tipo de conteúdo e o tipo de contexto. Nas modalidades do presente pedido, a unidade de ajuste 300 pode ajustar pelo menos um parâmetro do(s) dispositivo(s) de melhoria de áudio com base em pelo menos um dentre vários tipos de conteúdo e/ou pelo menos um dentre vários tipos de contexto. Consequentemente, como mostrado na Figura 3, em uma variante da modalidade mostrada na Figura 1, o classificador de áudio 200 pode compreender um classificador de conteúdo de áudio 202 ou um classificador de contexto de áudio 204, ou ambos.

[00137] Foram mencionados acima diferentes tipos de áudio com base em diferentes padrões (como para os tipos de contexto), bem como diferentes tipos de áudio em diferentes níveis hierárquicos (como para os tipos de conteúdo). Entretanto, os padrões e os níveis hierárquicos são apenas para a conveniência da presente descrição e definitivamente não são limitadores. Em outras palavras, no presente pedido, qualquer um dentre dois ou mais tipos de áudio supramencionados podem ser identificados pelo classificador de áudio 200 ao mesmo tempo e ser considerados pela unidade de ajuste 300 ao mesmo tempo, conforme será descrito posteriormente. Em outras palavras, todos os tipos de áudio em diferentes níveis hierárquicos podem ser paralelos ou estar no mesmo nível.

1.2 Valores da confiança de tipos de áudio e arquitetura dos classificadores

[00138] O classificador de áudio 200 pode enviar os resultados de decisão complicada ou a unidade de ajuste 300 pode considerar os resultados do classificador de áudio 200 como resultados de decisão

complicada. Mesmo para a decisão complicada, os múltiplos tipos de áudio podem ser atribuídos a um segmento de áudio. Por exemplo, um segmento de áudio pode ser rotulado como "fala" e "música de curta duração", visto que pode ser um sinal de mistura de fala ou de música de curta duração. Os rótulos obtidos podem ser usados diretamente para controlar o(s) dispositivo(s) de melhoria de áudio 400. Um exemplo simples é ativar o otimizador de diálogo 402 quando houver fala e desligá-lo quando não houver fala. Entretanto, esse método de decisão complicada pode trazer certa artificialidade nos pontos de transição de um tipo de áudio a outro, se não houver um esquema cauteloso de uniformização (que será discutido posteriormente).

[00139] A fim ter mais flexibilidade e ajustar os parâmetros dos dispositivos de melhoria de áudio de uma maneira contínua, o valor de confiança de cada tipo de áudio alvo pode ser estimado (decisão branda). Um valor de confiança representa o nível combinado entre o conteúdo de áudio a ser identificado e o tipo de áudio alvo, com valores de 0 a 1.

[00140] Como indicado antes, muitas técnicas de classificação podem enviar diretamente valores de confiança. O valor de confiança também pode ser calculado a partir de vários métodos, que podem ser considerados como uma parte do classificador. Por exemplo, se os modelos de áudio forem treinados por algumas tecnologias probabilísticas de modelagem, como os Modelos de Mistura Gaussiana (GMM), uma probabilidade posterior pode ser usada para representar o valor de confiança, como

$$p(c_i \mid x) = \frac{p(x \mid c_i)}{\sum_{i=1}^{N} p(x \mid c_i)} \qquad (1)$$

[00141] onde x é uma parte do segmento de áudio; C é um tipo de áudio alvo, N é o número de tipos de áudios alvo, p(xlci) é a probabilidade de que o segmento de áudio X seja do tipo de áudio c; e

p (c¡lx) é a probabilidade posterior correspondente.

[00142] Por outro lado, se os modelos de áudio forem treinados a partir de alguns métodos discriminativos, como a Máquina de Vetor de Suporte (SVM) e adaBoost, são obtidas apenas contagens (valores reais) a partir da comparação modelo. Nesses casos, uma função sigmoide é geralmente utilizada para mapear a contagem obtida (teoricamente, de -∞ a ∞) para a confiança esperada (de 0 a 1):

$$conf = \frac{1}{1+e^{Ay+B}} \qquad (2)$$

[00143] onde y é a contagem de saída da SVM ou adaBoost, A e B são dois parâmetros que necessitam ser estimados a partir de uma série de dados de treinamento usando algumas tecnologias bem conhecidas.

[00144] Para algumas modalidades do presente pedido, a unidade de ajuste 300 pode usar mais de dois tipos de conteúdo e/ou mais de dois tipos de contexto. Em seguida, o classificador de conteúdo de áudio 202 precisa identificar mais de dois tipos de conteúdo e/ou o classificador de contexto de áudio 204 precisa identificar mais de dois tipos de contexto. Em tal situação, o classificador de conteúdo de áudio 202 ou o classificador de contexto de áudio 204 podem ser um grupo de classificadores organizados em uma certa arquitetura.

[00145] Por exemplo, se a unidade de ajuste 300 necessitar de todos os quatro tipos de mídia de tipo filme de tipos de contexto, música de longa duração, jogos e VoIP, então o classificador de contexto de áudio 204 pode ter as diferentes arquiteturas a seguir:

[00146] Primeiramente, o classificador de contexto de áudio 204 pode compreender 6 classificadores binários de um a um (cada classificador discrimina um tipo de áudio alvo de outro tipo de áudio alvo) organizados conforme mostrado na Figura 4, 3 classificadores binários de um para outros (cada classificador discrimina um tipo de áudio alvo dos outros) organizados conforme mostrado na Figura 5 e 4

classificadores de um para outros organizados conforme mostrado na Figura 6. Há também outras arquiteturas, como a arquitetura de Gráfico Acíclico Direcionada para a Decisão (DDAG). Note que nas Figuras 4-6 e a descrição correspondente abaixo, utiliza-se "filme" em vez de "mídia do tipo filme" por questão de concisão.

[00147] Cada classificador binário produzirá uma contagem de confiança H(x) para sua saída (x representa um segmento de áudio). Depois que as saídas de cada classificador binário forem obtidas, será necessário mapeá-las para os valores finais de confiança dos tipos de contexto identificados.

[00148] Geralmente, supondo que o sinal de áudio deva ser classificado nos tipos de contexto M (M é um número inteiro positivo). A arquitetura um a um convencional constrói M(M— 1)/2 classificadores, onde cada um é treinado nos dados de duas classes, em seguida, cada classificador um a um lança um voto para sua classe preferida e o resultado final é a classe com o maior número de votos entre as classificações dos M(M— 1)/2 classificadores. Em comparação à arquitetura um a um convencional, a arquitetura hierárquica da Figura 4 também precisa construir M(M— 1)/2 classificadores. Todavia, as iterações de teste podem ser abreviadas para M— 1, visto que o segmento x será determinado como estando/não estando na classe correspondente a cada nível hierárquico e a contagem geral de nível é M— 1. Os valores finais de confiança para vários tipos do contexto podem ser calculados a partir da confiança de classificação binária H_k (x), por exemplo, (k=1, 2,… 6, representando diferentes tipos de contexto):

$$C_{MOVIE} \quad H_1(x)) \cdot (1 - H_3(x)) \cdot (1 - H_6(x))$$

$$C_{VOIP} = H_1(x) \cdot H_2(x) \cdot H_4(x)$$

$$C_{MUSIC} \quad \cdot (1 - H_2(x)) \cdot (1 - H_5(x)) + H_3(x) \cdot (1 - H_1(x)) \cdot (1 - H_5(x))$$

$$+ H_6(x) \cdot (1 - H_1(x)) \cdot (1 - H_3(x))$$

$$C_{GAME} \quad \cdot H_2(x) \cdot (1 - H_4(x)) + H_1(x) \cdot H_5(x) \cdot (1 - H_2(x)) + H_3(x) \cdot H_5(x)$$

$$\cdot (1 - H_1(x))$$

[00149] Na arquitetura mostrada na Figura5, a função de mapeamento da classificação binária resulta em $H_k(x)$ para que os valores finais de confiança possam ser definidos como o exemplo a seguir:

$$C_{MOVIE} = H_1(x)$$
$$C_{MUSIC} = H_2(x) \cdot \big(1 - H_1(x)\big)$$
$$C_{VOIP} = H_3(x) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_1(x)\big)$$
$$C_{GAME} = \big(1 - H_3(x)\big) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_1(x)\big)$$

[00150] Na arquitetura ilustrada na Figura 6, os valores finais de confiança podem ser iguais aos resultados de classificação binária correspondentes H_k (x) ou se a soma dos valores de confiança para todas as classes precisar ser 1, então os valores finais de confiança podem ser simplesmente normalizados com base no $H_k(x)$ estimado:

$$C_{MOVIE} = H_1(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{MUSIC} = H_2(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{VOIP} = H_3(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$
$$C_{GAME} = H_4(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

[00151] Esse um ou mais de um com os valores máximos de confiança pode ser determinado como sendo a classe final identificada.

[00152] Deve-se notar que nas arquiteturas mostradas nas Figuras4-6, a sequência de classificadores binários diferentes não são necessariamente como mostrado, mas podem ser outras sequências, as quais podem ser selecionadas por atribuição manual ou aprendizagem automática, de acordo com diferentes requisitos de várias aplicações.

[00153] As descrições acima são direcionadas aos classificadores de contexto de áudio 204. Quanto ao classificador de conteúdo de áudio 202, a situação é semelhante.

[00154] De maneira alternativa, o classificador de conteúdo de

áudio 202 ou o classificador de contexto de áudio 204 podem ser implementados como um único classificador identificando todos os tipos de conteúdo/de contexto ao mesmo tempo e dando os valores de confiança correspondentes ao mesmo tempo. Existem muitas técnicas para fazer isso.

[00155] Utilizando o valor de confiança, a saída do classificador de áudio 200 pode ser representada como um vetor, com cada dimensão representando o valor de confiança de cada tipo de áudio alvo. Por exemplo, se os tipos de áudio alvo estiverem (fala, música de curta duração, ruído, fundo) em sequência, um resultado de saída exemplificativo poderia ser (0,9, 0,5, 0,0, 0,0), indicando que é 90% certo o conteúdo de áudio é de fala, e 50% certo que o áudio é música. Nota-se que a soma de todas as dimensões no vetor de saída não precisa ser um (por exemplo, os resultados da Figura 6 não são necessariamente normalizados), o que significa que o sinal de áudio pode ser um sinal de mistura de fala e de música de curta duração.

[00156] Posteriormente, nas Partes 6 e 7, uma nova implementação da classificação do contexto de áudio e da classificação do conteúdo de áudio serão discutidas em detalhes.

1.3 Uniformização dos valores de confiança dos tipos de áudio

[00157] Opcionalmente, depois que cada segmento de áudio foi classificado em tipos de áudio pré-definidos, uma etapa adicional é uniformizar os resultados de classificação ao longo da linha de tempo para evitar um salto abrupto de um tipo para outro e fazer uma estimativa de uniformização dos parâmetros nos dispositivos de melhoria de áudio. Por exemplo, um excerto longo é classificado como mídia de tipo filme, exceto por apenas um segmento classificado como VoIP, então a decisão VoIP abrupta pode ser revisada para a mídia de tipo filme por uniformização.

[00158] Assim, em uma variante da modalidade, conforme

mostrado na Figura 7, uma unidade de uniformização de tipo 712 é fornecida ainda para, para cada tipo de áudio, uniformizar o valor de confiança do sinal de áudio no momento atual.

[00159] Um método de uniformização comum é baseado na média pesada, como o cálculo de uma soma ponderada do valor de confiança atual no presente e um valor de confiança uniformizado do tempo mais recente, conforme segue:

$$smoothConf(t) = \beta \cdot smoothConf(t-1) + (1-\beta) \cdot conf(t) \quad (3)$$

[00160] onde t representa o tempo atual (o segmento de áudio atual), t-1 representa o tempo mais recente (o último segmento de áudio), β é o peso, conf e smoothConf são os valores de confiança antes e depois da uniformização, respectivamente.

[00161] Do ponto de vista dos valores de confiança, os resultados da decisão complicada dos classificadores também podem ser representados com valores de confiança, com os valores sendo de 0 ou 1. Isto é, se um tipo de áudio alvo é escolhido e atribuído a um segmento de áudio, a confiança correspondente é 1; do contrário, a confiança é 0. Assim, mesmo se o classificador de áudio 200 não der o valor de confiança mas der apenas uma decisão complicada quanto ao tipo de áudio, um ajuste contínuo da unidade de ajuste 300 ainda é possível através da operação de uniformização da unidade de uniformização de tipo 712.

[00162] O algoritmo de uniformização pode ser "assimétrico", utilizando um peso de uniformização diferente para casos diferentes. Por exemplo, os pesos para calcular a soma ponderada podem ser alterados de maneira adaptativa com base no valor de confiança do tipo de áudio do sinal de áudio. O valor de confiança do segmento atual sendo maior, seu peso é maior

[00163] De outro ponto da vista, os pesos para calcular a soma ponderada podem ser alterados de maneira adaptativa com base nos

diferentes pares de transição de um tipo de áudio para outro tipo de áudio, sobretudo quando o(s) dispositivo(s) de melhoria de áudio forem ajustados com base nos múltiplos tipos de conteúdo, como identificado pelo classificador de áudio 200, em vez de baseados na presença ou ausência de um único tipo de conteúdo. Por exemplo, para uma transição de um tipo de áudio que aparece com mais frequência em um certo contexto para outro tipo de áudio que não aparece com tanta frequência no contexto, o valor de confiança deste pode ser uniformizado, de modo que não aumentará tão depressa, porque pode ser apenas uma interrupção ocasional.

[00164] Um outro fator é a tendência a alteração (aumento ou diminuição), incluindo a taxa de alteração. Supondo que seja mais importante a latência quando um tipo de áudio se torna presente (isto é, quando seu valor de confiança aumenta), pode-se projetar o algoritmo de uniformização da seguinte maneira:

$$smoothConf(t) = \begin{cases} conf(t) & conf(t) \geq smoothConf(t-1) \\ \beta \cdot smoothConf(t-1) + (1-\beta) \cdot conf(t) & otherwi \end{cases} \quad (4)$$

[00165] A fórmula acima permite que o valor de confiança uniformizado responda rapidamente ao estado atual quando o valor de confiança aumenta e uniformiza lentamente quando o valor de confiança diminui. As variantes das funções de uniformização podem ser facilmente projetadas de maneira semelhante. Por exemplo, a fórmula (4) pode ser revisada de modo que o peso do conf(t) se torne maior quando conf(t)>=smoothConf(t-1). Na verdade, na fórmula (4), pode-se considerar que β =0 e o peso de conf(t) se torna o maior, isto é: 1.

[00166] De um ponto de vista diferente, considerar a tendência a alteração de determinado tipo de áudio é apenas um exemplo específico de considerar diferentes pares de transição dos tipos de áudio. Por exemplo, o aumento do valor de confiança do tipo A pode ser considerado como uma transição de não-A para A e a diminuição

do valor de confiança de tipo A pode ser considerado como uma transição de A para não-A.

1.4 Ajuste de parâmetro

[00167] A unidade de ajuste 300 é projetada para estimar ou ajustar parâmetros apropriados para o(s) dispositivo(s) de melhoria de áudio 400 com base nos resultados obtidos a partir do classificador de áudio 200. Diferentes algoritmos de ajuste podem ser projetados para diferentes dispositivos de melhoria de áudio utilizando o tipo de conteúdo ou o tipo de contexto, ou ambos, para uma decisão conjunta. Por exemplo, com informações do tipo de contexto, como mídia de tipo filme e música de longa duração, os pré-ajustes, como mencionado anteriormente, podem ser selecionados automaticamente ou aplicados sobre o conteúdo correspondente. Com as informações do tipo de conteúdo disponíveis, os parâmetros de cada dispositivo de melhoria de áudio podem ser sintonizados de uma maneira mais fina, como mostrado nas partes subsequentes. As informações do tipo de conteúdo e as informações de contexto podem ser usadas de maneira conjunta, ainda, na unidade de ajuste 300 para balancear as informações de curta e de longa duração. O algoritmo de ajuste específico para um dispositivo de melhoria de áudio específico pode ser considerado como uma unidade de ajuste separada ou os algoritmos de ajuste diferentes podem ser considerados coletivamente como uma unidade de ajuste unida.

[00168] Isto é, a unidade de ajuste 300 pode ser configurado para ajustar pelo menos um parâmetro do dispositivo de melhoria de áudio com base no valor de confiança de pelo menos um tipo de conteúdo e/ou no valor de confiança de pelo menos um tipo de contexto. Para um dispositivo de melhoria de áudio específico, alguns tipos de áudio são informativos e alguns tipos de áudio são de intervenção. Com efeito, os parâmetros do dispositivo de melhoria de áudio específico

pode ser positivamente ou negativamente correlacionado ao(s) valor(es) de confiança do(s) tipo(s) de áudio informativo ou do(s) tipo(s) de áudio de intervenção. Aqui, "positivamente correlacionado" quer dizer que o parâmetro aumenta ou diminui com o aumento ou a diminuição do valor de confiança do tipo de áudio, de maneira linear ou não-linear. "Negativamente correlacionado" quer dizer que o parâmetro aumenta ou diminui com, respectivamente, a diminuição ou o aumento do valor de confiança do tipo de áudio, de maneira linear ou não-linear.

[00169] Aqui, a diminuição e o aumento do valor de confiança são diretamente "transferidos" aos parâmetros a serem ajustados pela correlação positiva ou negativa. Em matemática, tal correlação ou "transferência" pode ser incorporada como uma proporção linear ou proporção inversa, operação de mais ou menos (adição ou subtração), operação de multiplicação ou divisão ou função não-linear. Todas essas formas de correlação podem ser denominadas "função de transferência". Para determinar o aumento ou a diminuição do valor de confiança, pode-se também comparar o valor de confiança atual ou sua transformação matemática com o último valor de confiança ou com uma pluralidade de valores de confiança do histórico, ou suas transformações matemáticas. No contexto do presente pedido, o termo "comparar" significa comparação através de operação de subtração ou comparação através de operação de divisão. Pode-se determinar um aumento ou uma diminuição determinando se a diferença é superior a 0 ou se a razão é superior a 1.

[00170] Em implementações específicas, pode-se relacionar diretamente os parâmetros com os valores de confiança ou suas razões ou diferenças através do algoritmo apropriado (como a função de transferência) e não é necessário que um "observador externo" saiba explicitamente se um valor de confiança específico e/ou um

parâmetro específico aumentou ou diminuiu. Alguns exemplos específicos serão dados nas partes 2-5 subsequentes quanto aos dispositivos de melhoria de áudio específicos.

[00171] Como indicado na seção anterior, em relação ao mesmo segmento de áudio, o classificador 200 pode identificar múltiplos tipos de áudio com os respectivos valores de confiança, cujos valores de confiança podem não necessariamente chegar a 1, visto que o segmento de áudio pode compreender múltiplos componentes ao mesmo tempo, como música e fala e sons de fundo. Em tal situação, os parâmetros dos dispositivos de melhoria de áudio devem ser equilibrados entre os diferentes tipos de áudio. Por exemplo, a unidade de ajuste 300 pode ser configurada para considerar pelo menos alguns dos múltiplos tipos de áudio através da pesagem dos valores de confiança de pelo menos um tipo de áudio com base na importância de pelo menos um tipo de áudio. Quanto mais importante for um tipo específico de áudio, mais importante são os parâmetros influenciados por este.

[00172] O peso também pode refletir o efeito informativo e de intervenção do tipo de áudio. Por exemplo, para um tipo de áudio de intervenção, um peso de menos pode ser dado. Alguns exemplos específicos serão dados nas partes 2-5 subsequentes quanto aos dispositivos de melhoria de áudio específicos.

[00173] Observe-se que no contexto do presente pedido, "peso" tem um significado mais amplos do que os coeficientes em uma multinomial. Além dos coeficientes em uma multinomial, ele também pode assumir o formato de uma potência ou um expoente. Quando os coeficientes estiverem em uma multinomial, os coeficientes de pesagem pode ser ou não normalizados. Em resumo, o peso representa apenas quanta influência o objeto pesado tem sobre o parâmetro a ser ajustado.

[00174] Em algumas outras modalidades, para os múltiplos tipos de áudio contidos no mesmo segmento de áudio, os valores de confiança destes podem ser convertidos para pesos através de sua normalização, e então o parâmetro final pode ser determinado através do cálculo de uma soma de valores pré-ajustados de parâmetro para cada tipo de áudio e pesado pelos pesos com base nos valores de confiança. Isto é, a unidade de ajuste 300 pode ser configurada para considerar os múltiplos tipos de áudio através da pesagem dos efeitos dos múltiplos tipos de áudio com base nos valores de confiança.

[00175] Como um exemplo específico de pesagem, a unidade de ajuste é configurada para considerar que pelo menos um tipo de áudio dominante com base nos valores de confiança. Para aqueles tipos de áudio com valores de confiança muito baixos (menos do que um limite), eles podem não ser considerados. Isso equivale aos pesos dos outros tipos de áudio cujos valores de confiança forem inferiores ao limite serem estabelecidos como zero. Alguns exemplos específicos serão dados nas partes 2-5 subsequentes quanto aos dispositivos de melhoria de áudio específicos.

[00176] O tipo de conteúdo e de contexto podem ser considerados em conjunto. Em uma modalidade, eles podem ser considerados como no mesmo nível e seus valores de confiança podem ser seus respectivos pesos. Em outra modalidade, tal como a denominação sugere, o "tipo de contexto" é o contexto ou ambiente em que o "tipo de contexto" se situa e, assim, a unidade de ajuste 200 pode ser configurada de modo que a um tipo de conteúdo em um sinal de áudio de um tipo de contexto diferente é atribuído um peso diferente, dependendo do tipo de contexto do sinal de áudio. Falando de modo geral, qualquer tipo de áudio pode constituir um contexto de outro tipo de áudio e, assim, a unidade de ajuste 200 pode ser configurada para modificar o peso de um tipo de áudio com o valor de confiança de

outro tipo de áudio. Alguns exemplos específicos serão dados nas partes 2-5 subsequentes quanto aos dispositivos de melhoria de áudio específicos.

[00177] No contexto do presente pedido, "parâmetro" tem um significado mais amplo do que seu significado literal. Além de um parâmetro com um valor único, isso também significa um pré-ajuste, como mencionado antes, que inclui uma série de parâmetros diferentes, um vetor compreendido de diferentes parâmetros ou um perfil. Especificamente, nas partes 2-5 subsequentes os parâmetros a seguir serão discutidos, mas o presente pedido não se limita a isso: o nível de otimização de diálogo, os limites para a determinação das bandas de frequência a terem os diálogos otimizados, o nível de fundo, a quantidade de boost do surround, a frequência inicial do virtualizador de surround, o ganho dinâmico ou a amplitude do ganhou dinâmico de um nivelador de volume, os parâmetros indicativos do grau de sinal de áudio sendo um novo evento de áudio perceptível, o nível de equalização, os perfis de equalização e os pré-ajustes de balanceamento espectral.

1.5 Uniformização de parâmetro

[00178] Na Seção 1.3, foi discutida a uniformização do valor de confiança de um tipo de áudio para evitar sua mudança abrupta e, portanto, a mudança abrupta dos parâmetros do(s) dispositivo(s) de melhoria de áudio. Outras medidas também são possíveis. Uma é uniformizar o parâmetro ajustado com base no tipo de áudio, que será discutido nesta seção; a outra é configurar o classificador de áudio e/ou a unidade de ajuste de modo a retardar a mudança dos resultados do classificador de áudio, e isso será discutido na Seção 1.6.

[00179] Em uma modalidade, o parâmetro pode ser uniformizado, ainda, para evitar uma mudança rápida, a qual pode inserir artefatos

audíveis nos pontos de transição, como

$$\tilde{L}(t) = \tau\tilde{L}(t-1) + (1-\tau)L(t) \qquad (3')$$

[00180] onde $\tilde{L}(t)$ é o parâmetro uniformizado, *L(t)* é o parâmetro não-uniformizado, τ é um coeficiente que representa uma constante de tempo, t é o tempo atual e t-1 é o tempo mais recente.

[00181] Isto é, como mostrado na Figura 8, o aparelho de processamento de áudio pode compreender uma unidade de uniformização de parâmetro 814 para, para um dispositivo de melhoria de áudio (como pelo menos um dentre o otimizador de diálogo 402, o virtualizador surround 404, o nivelador de volume 406 e o equalizador 408) ajustado pela unidade de ajuste 300, o valor de parâmetro de uniformização determinado pela unidade de ajuste 300 no tempo atual calculando-se uma soma ponderada do valor de parâmetro determinado pela unidade de ajuste no tempo atual e um valor de parâmetro uniformizado do tempo mais recente.

[00182] A constante de tempo τ pode ser um valor fixo com base no requisito específico de uma aplicação e/ou a implementação do dispositivo de melhoria de áudio 400. Ela também poderia ser alterada de maneira adaptativa com base no tipo de áudio, especialmente com base em diferentes tipos de transição de um tipo de áudio para outro, como de música para fala e de fala para música.

[00183] Tome o equalizador como exemplo (mais detalhes podem ser referidos na parte 5). A equalização é boa para aplicar no conteúdo musical, mas não no conteúdo de fala. Assim, para uniformizar o nível de equalização, a constante de tempo pode ser relativamente pequena quando o sinal de áudio transita de música para fala, de modo que um nível de equalização menor possa ser aplicado ao conteúdo mais rapidamente. Por um lado, a constante de tempo para a transição de fala para música pode ser relativamente maior a fim de evitar artefatos audíveis nos pontos de transição.

[00184] Para estimar o tipo da transição (por exemplo, de fala para música ou de música para fala), os resultados de classificação de conteúdo podem ser usados diretamente. Isto é, a classificação do conteúdo de áudio em música ou fala simplifica a obtenção do tipo de transição. Para estimar a transição de uma maneira mais contínua, pode-se confiar também no nível de equalização não-uniformizado estimado em vez de comparar diretamente as decisões complicadas dos tipos de áudio. A ideia geral é, se o nível de equalização não-uniformizado estiver aumentando, ele indica uma transição de fala para música (ou mais do tipo música); do contrário, é mais uma transição de música para fala (ou mais do tipo fala). Pela diferenciação de diferentes tipos de transição, a constante de tempo pode ser ajustada de maneira correspondente, um exemplo é:

$$\tau(t) = \begin{cases} \tau_1 & L(t) \geq L(t-1) \\ \tau_2 & L(t) < L(t-1) \end{cases} \qquad (4')$$

[00185] onde o τ (t) é a constante de tempo variante com o tempo que depende do conteúdo, τ 1 e τ 2 são dois valores constantes de tempo pré-ajustados, geralmente satisfazendo τ1> τ 2. Intuitivamente, a função acima indica uma transição relativamente lenta quando o nível de equalização estiver aumentando e uma transição relativamente rápida quando o nível de equalização estiver diminuindo, mas o presente pedido não se limita a isso. Ademais, o parâmetro não se limita ao nível de equalização, mas podem ser outros parâmetros. Isto é, a unidade de uniformização de parâmetro 814 pode ser configurada de modo que os pesos para calcular a soma ponderada são alterados de maneira adaptativa com base em uma tendência de aumento ou de diminuição do valor de parâmetro determinado pela unidade de ajuste 300.

1.6 Transição de tipos de áudio

[00186] Em referência às Figuras 9 e 10, será descrito outro

esquema para evitar a mudança abrupta do tipo de áudio e, assim, evitar a mudança abrupta dos parâmetros do(s) dispositivo(s) de melhoria de áudio.

[00187] Como mostrado na Figura 9, o aparelho de processamento de áudio 100 pode compreender ainda um temporizador 916 para medir o tempo de duração durante o qual o classificador de áudio 200 envia continuamente o mesmo tipo de áudio novo, em que a unidade de ajuste 300 pode ser configurada para continuar a usar o tipo de áudio atual até a extensão do tempo de duração do novo tipo de áudio alcançar um limite.

[00188] Em outras palavras, uma fase de observação (ou sustentação) é inserida, como ilustrado na Figura 10. Com a fase de observação (que corresponde ao limite da extensão do tempo de duração), a mudança no tipo de áudio é monitorada, ainda, quanto a uma quantidade consecutiva de tempo para confirmar se o tipo de áudio de fato mudou, antes da unidade de ajuste 300 de fato usar o novo tipo de áudio.

[00189] Como mostrado na Figura 10, a seta (1) ilustra a situação onde o estado atual é de tipo A e o resultado do classificador de áudio 200 não muda.

[00190] Se o estado atual for de tipo A e o resultado do classificador de áudio 200 se tornar de tipo B, então o temporizador 916 começa a contar ou, como mostrado na Figura 10, o processo entra em uma fase de observação (a seta (2)) e um valor inicial da contagem de sobra cnt é ajustada, indicando a quantidade da duração da observação (igual ao limite).

[00191] Então, se o classificador de áudio 200 enviar continuamente o tipo B, então cnt diminui continuamente (a seta (3)) até que cnt seja igual a 0 (isto é, a extensão do tempo de duração do novo tipo B atingir o limite), então a unidade de ajuste 300 pode usar o

novo tipo B do áudio (a seta (4)) ou, em outras palavras, só até agora o tipo de áudio pode ser considerado como tendo de fato mudado para o tipo B.

[00192] Caso contrário, se antes que cnt se torne zero (antes que a extensão do tempo de duração atingir o limite) a saída do classificador de áudio 200 voltar ao tipo A antigo, então a fase de observação é terminada e a unidade de ajuste 300 ainda usa o tipo A antigo (a seta (5)).

[00193] A mudança do tipo B para o tipo A pode ser similar ao processo descrito acima.

[00194] No processo acima, o limite (ou contagem de sobra) pode ser ajustada com base no requisito do pedido. Pode ser um valor fixo pré-definido. Também pode ser ajustado de modo adaptativo. Em uma variante, o limite é diferente para diferentes pares de transição de um tipo de áudio para outro tipo de áudio. Por exemplo, ao mudar do tipo A para o tipo B, o limite pode ser de um primeiro valor; e ao mudar do tipo B para o tipo A, o limite pode ser de um segundo valor.

[00195] Em outra variante, a contagem de sobra (limite) pode ser negativamente correlacionada ao valor de confiança do novo tipo de áudio. A ideia geral é, se a confiança apresentar confusão entre os dois tipos (por exemplo, quando o valor de confiança é apenas em torno de 0,5), a duração da observação precisa ser longa; caso contrário, a duração pode ser relativamente curta. Seguindo essa orientação, uma contagem de sobra de exemplo pode ser ajustada pela seguinte fórmula,

$$HangCnt = C \cdot |0.5 - Conf| + D$$

[00196] onde HangCnt é a duração da sobra ou o limite, C e D são dois parâmetros que podem ser ajustados com base no requisito do pedido, C geralmente é negativo e D é um valor positivo.

[00197] Incidentalmente, o temporizador 916 (e, portanto, o

processo de transição descrito acima) foi descrito acima como uma parte do aparelho de processamento de áudio, mas fora do classificador de áudio 200. Em algumas outras modalidades, pode ser considerado como uma parte do classificador de áudio 200, como descrito na Seção 7.3.

1.7 Combinação de modalidades e cenários de aplicação

[00198] Todas as modalidades e variantes discutidas acima podem ser implementadas em qualquer combinação e quaisquer componentes mencionados em diferentes partes/modalidades, mas com funções iguais ou semelhantes, podem ser implementadas como componentes iguais ou separados.

[00199] Especificamente, ao descrever as modalidades e suas variações anteriores, aqueles componentes com sinais de referência semelhantes àqueles já descritos em modalidades anteriores ou variantes são omitidos e componentes diferentes são descritos. Inclusive, esses componentes diferentes podem ser combinados com os componentes de outras modalidades ou variantes ou constituir somente soluções separadas. Por exemplo, duas ou mais das soluções descritas com referência às Figuras de 1 a 10 podem ser combinadas umas com as outras. Como a solução mais completa, o aparelho de processamento de áudio pode compreender o classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204, bem como a unidade de uniformização de tipo 712, a unidade de uniformização de parâmetro 814 e o temporizador 916.

[00200] Como mencionado anteriormente, os dispositivos de melhoria de áudio 400 podem incluir o otimizador de diálogo 402, o virtualizador surround 404, o nivelador de volume 406 e o equalizador 408. O aparelho de processamento de áudio 100 pode incluir qualquer um ou mais deles, com a unidade de ajuste 300 adaptada a eles. Ao envolver os múltiplos dispositivos de melhoria de áudio 400, a unidade

de ajuste 300 pode ser considerada como incluindo múltiplas subunidades 300A a 300D (Figuras 15, 18, 20 e 22) específicas para os respectivos dispositivos de melhoria de áudio 400 ou ainda ser consideradas como uma unidade de ajuste unida. Quando específica para um dispositivo de melhoria de áudio, a unidade de ajuste 300 em conjunto com o classificador de áudio 200, bem como outros componentes possíveis, pode ser considerada como o controlador do dispositivo de melhoria de áudio específico, que será discutido em detalhes nas partes 2-5 subsequentes.

[00201] Ademais, os dispositivos de melhoria de áudio 400 não são limitados aos exemplos, conforme mencionado, e podem incluir qualquer outro dispositivo de melhoria de áudio.

[00202] Ademais, quaisquer soluções já discutidas ou quaisquer combinações destes podem ser combinadas ainda com qualquer modalidade descrita ou implicada em outras partes desta divulgação. Especialmente, as modalidades dos classificadores de áudio, como serão discutidas nas partes 6 e 7, podem ser usadas no aparelho de processamento de áudio.

1.8 Método de processamento de áudio

[00203] No processo de descrição do aparelho de processamento de áudio nas modalidades anteriores, são divulgadas de modo aparente, também, alguns processos e métodos. Doravante, um resumo desses métodos é fornecido sem repetir alguns dos detalhes já discutidos anteriormente, mas deve-se notar que embora os métodos sejam divulgados no processo de descrição do aparelho de processamento de áudio, os métodos não necessariamente adotam aqueles componentes descritos ou não são necessariamente executados por aqueles componentes. Por exemplo, as modalidades do aparelho de processamento de áudio podem ser executadas parcialmente ou completamente com hardware e/ou firmware,

enquanto é possível que o método de processamento de áudio discutido abaixo pode ser executado totalmente por um programa executável em computador, embora os métodos também possam adotar o hardware e/ou o firmware do aparelho de processamento de áudio.

[00204] Os métodos serão descritos abaixo com referência às Figuras 11-14. Observe-se que em correspondência à propriedade de streaming do sinal de áudio, as várias operações são repetidas quando o método é implementado em tempo real e diferentes operações não são necessariamente em relação ao mesmo segmento de áudio.

[00205] Em uma modalidade, como mostrado na Figura 11, um método de processamento de áudio é fornecido. Primeiramente, o sinal de áudio a ser processado é classificado em pelo menos um tipo de áudio em tempo real (operação 1102). Baseado no valor de confiança de pelo menos um tipo de áudio, pelo menos um parâmetro para a melhoria de áudio pode ser ajustado continuamente (operação 1104). A melhoria de áudio pode ser: otimização de diálogo (operação 1106), virtualização surround (operação 1108), nivelamento de volume (1110) e/ou equalização (operação 1112). De modo correspondente, pelo menos um parâmetro pode compreender pelo menos um parâmetro para pelo menos um processamento de otimização de diálogo, processamento de virtualização surround, processamento de nivelamento de volume e processamento de equalização.

[00206] Aqui, "em tempo real" e "continuamente" quer dizer o tipo de áudio e, assim, o parâmetro mudará em tempo real com o conteúdo específico do sinal de áudio e "continuamente" também quer dizer que o ajuste é um ajuste contínuo com base no valor de confiança em vez de um ajuste discreto ou abrupto.

[00207] O tipo de áudio pode compreender o tipo de conteúdo e/ou

o tipo de contexto. Correspondentemente, a operação de ajuste 1104 pode ser configurada para ajustar pelo menos um parâmetro com base no valor de confiança de pelo menos um tipo de conteúdo ou o valor de confiança de pelo menos um tipo de contexto. O tipo de conteúdo pode compreender ainda pelo menos um dos tipos de conteúdo de música de curta duração, fala, som de fundo e ruído. O tipo de contexto pode compreender ainda pelo menos um dos tipos de contexto de música de curta duração, mídia de tipo filme, jogo e VoIP.

[00208] Alguns outros esquemas de tipo de contexto também são propostos, como tipos de contexto relacionados a VoIP, incluindo VoIP e não-VoIP e tipos de qualidade de áudio que incluem áudio de alta qualidade ou áudio de baixa qualidade.

[00209] A música de curta duração pode ser classificada ainda em subtipos, de acordo com diferentes padrões. Dependendo da presença da fonte dominante, ela pode compreender uma música sem fontes dominantes e música com fontes dominantes. Além disso, a música de curta duração pode compreender pelo menos um agrupamento com base em gênero ou pelo menos um agrupamento com base em instrumento ou pelo menos um agrupamento de música classificado com base no ritmo, no tempo, no timbre da música e/ou qualquer outro atributo musical.

[00210] Quando os tipos de conteúdo e os tipos de contexto são identificados, a importância de um tipo de conteúdo pode ser determinada pelo tipo de contexto em que o tipo de conteúdo se situa. Isto é, o tipo de conteúdo em um sinal de áudio de um tipo de contexto diferente é designado com um peso diferente dependendo do tipo de contexto do sinal de áudio. De modo mais geral, um tipo de áudio pode influenciar ou pode ser a premissa de outro tipo de áudio. Assim, a operação de ajuste 1104 pode ser configurada para modificar o peso de um tipo de áudio com o valor de confiança de outro tipo de áudio.

[00211] Quando um sinal de áudio for classificado em múltiplos tipos de áudio ao mesmo tempo (isto é, com relação ao mesmo segmento de áudio), a operação de ajuste 1104 pode considerar alguns ou todos os tipos de áudio identificados para ajustas o(s) parâmetros de melhoria daquele segmento de áudio. Por exemplo, a operação de ajuste 1104 pode ser configurada para considerar pelo menos alguns dos múltiplos tipos de áudio através da pesagem dos valores de confiança de pelo menos um tipo de áudio com base na importância de pelo menos um tipo de áudio. Ou a operação de ajuste 1104 pode ser configurada para considerar pelo menos alguns dos tipos de áudio pesando-os com base em seus valores de confiança. Em um caso especial, a operação de ajuste 1104 pode ser configurada para considerar que pelo menos um tipo de áudio dominante com base nos valores de confiança.

[00212] Para evitar mudanças abruptas nos resultados, podem ser introduzidos esquemas de uniformização.

[00213] A valor de parâmetro ajustado pode ser uniformizado (operação 1214 na Figura 12). Por exemplo, o valor de parâmetro determinado pela operação de ajuste 1104 no tempo atual pode ser substituído por uma soma de valores de parâmetros determinada pela operação de ajuste no tempo atual e um valor de parâmetro uniformizado do tempo mais recente. Assim, com a operação de uniformização iterada, o valor de parâmetro é uniformizado na linha de tempo.

[00214] Os pesos para calcular a soma ponderada podem ser mudados de modo adaptativo com base no tipo de áudio do sinal de áudio ou com base nos diferentes pares de transição de um tipo de áudio para outro tipo de áudio. De maneira alternativa, os pesos para calcular a soma ponderada são alterados de maneira adaptativa com base em uma tendência de aumento ou de diminuição do valor de

parâmetro determinado pela operação de ajuste.

[00215] Outro esquema de uniformização é mostrado na Figura 13. Isto é, o método pode compreender ainda, para cada tipo de áudio, a uniformização do valor de confiança do sinal de áudio no tempo atual através do cálculo de uma soma ponderada do valor de confiança real no valor de confiança uniformizado e presente do tempo mais recente (operação 1303). Similarmente à operação de uniformização de parâmetro 1214, os pesos para calcular a soma ponderada podem ser mudados de maneira adaptativa com base no valor de confiança do tipo de áudio do sinal de áudio ou com base em pares de transição diferentes de um tipo de áudio para outro tipo de áudio.

[00216] Outro esquema de uniformização é um mecanismo de buffer para retardar a transição de um tipo de áudio para outro tipo de áudio, mesmo se a saída da operação de classificação de áudio 1102 mudar. Isto é, a operação de ajuste 1104 não usa o novo tipo de áudio imediatamente, mas aguarda a estabilização da saída da operação de classificação de áudio 1102.

[00217] Especificamente, o método pode compreender a medição do tempo de duração durante o qual a operação de classificação envia continuamente o mesmo novo tipo de áudio (operação 1403 na Figura 14), em que a operação de ajuste 1104 é configurada para continuar usando o tipo de áudio atual ("N" na operação 14035 e operação 11041) até que a extensão do tempo de duração do novo tipo de áudio atinja um limite ("Y" na operação 14035 e operação 11042). Especificamente, quando a saída do tipo de áudio da operação de classificação de áudio 1102 muda em relação ao tipo de áudio atual usado na operação de ajuste de parâmetro de áudio 1104 ("Y" na operação 14031), então a contagem inicia (operação 14032). Se a operação de classificação de áudio 1102 continuar a enviar o novo tipo de áudio, isto é, se o julgamento na operação 14031 continuar sendo

"Y", a contagem prossegue (operação 14032). Por fim, quando o tempo de duração do novo tipo de áudio atingir um limite ("Y" na operação 14035), a operação de ajuste 1104 utiliza o novo tipo de áudio (operação 11042) e a contagem é reinicializada (operação 14034) para preparar-se para a próxima mudança do tipo de áudio. Antes de atingir o limite ("N" na operação 14035), a operação de ajuste 1104 continua a usar o tipo de áudio atual (operação 11041).

[00218] Aqui, a contagem pode ser implementada com o mecanismo de um temporizador (contagem progressiva ou regressiva). Se depois que a contagem começar, mas antes de atingir o limite, a saída da operação de classificação de áudio 1102 retorna ao tipo de áudio atual usado na operação de ajuste 1104, deve-se considerar que não há mudanças ("N" na operação 14031) quanto ao tipo de áudio atual usado pela operação de ajuste 1104. Mas o resultado da classificação atual (correspondente ao segmento de áudio atual a ser classificado no sinal de áudio) muda em relação à saída anterior (correspondente ao segmento de áudio anterior a ser classificado no sinal de áudio) da operação de classificação de áudio 1102 ("Y" na operação 14033), assim, a contagem é reiniciada (operação 14034), até que a próxima mudança ("Y" na operação 14031) inicie a contagem. Naturalmente, se o resultado de classificação da operação de classificação de áudio 1102 não mudar em relação ao tipo de áudio atual usado pela operação de ajuste de parâmetro de áudio 1104 ("N" na operação 14031) nem mudar em relação à classificação anterior ("N" na operação 14033), ela mostra que a classificação de áudio está em um estado estável e o tipo de áudio atual continua a ser usado.

[00219] O limite usado aqui também pode ser diferente para diferentes pares de transição de um tipo de áudio para outro tipo de áudio, visto que quando o estado não for tão estável, geralmente

pode-se preferir que o dispositivo de melhoria de áudio esteja em suas condições padrão em vez de outras. Por outro lado, se o valor de confiança do novo tipo de áudio for relativamente elevado, é mais seguro transitar ao novo tipo de áudio. Assim, o limite pode ser correlacionado negativamente com o valor de confiança do novo tipo de áudio. Quanto maior for o valor de confiança, menor é o limite, o que significa que o tipo de áudio pode passar para o novo, mais rápido.

[00220] Semelhante às modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações são práticas, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações podem ter soluções separadas. Especialmente, em todos os métodos de processamento de áudio, os métodos de classificação de áudio, como discutido nas partes 6 e 7, podem ser usados.

[00221] Parte 2: Controlador de otimizador de diálogo e método de controle

[00222] Um exemplo de dispositivo de melhoria de áudio é o otimizador de Diálogo (DE), que se direciona ao monitoramento contínuo do áudio em playback, à detecção da presença de diálogo e à otimização do diálogo para aumentar sua clareza e inteligibilidade (tornar o diálogo mais fácil de ser ouvido e compreendido), sobretudo para idosos com capacidade auditiva decrescente. Além de detectar se um diálogo está presente, as frequências mais importantes para a inteligibilidade também são detectadas se um diálogo estiver presente e, então, potencializadas de modo correspondente (com rebalanceamento espectral dinâmico). Um método de otimização de diálogo é apresentado em H. Muesch. "Speech Enhancement in Entertainment Audio" publicado como WO 2008/106036 A2, cuja

totalidade é incorporada aqui por referência.

[00223] Uma configuração manual comum do Otimizador de Diálogo é que ela é geralmente habilitada em um conteúdo de mídia de tipo filme, mas desabilitada no conteúdo da música, porque a otimização de diálogo pode ativar falsamente em excesso nos sinais da música.

[00224] Com as informações de tipo de áudio disponíveis, o nível de otimização de diálogo e outros parâmetros podem ser ajustados com base nos valores de confiança dos tipos de áudio identificados. Como um exemplo específico do aparelho de processamento de áudio e método discutidos anteriormente, o otimizador de diálogo pode fazer uso de todas as modalidades discutidas na parte 1 e todas as combinações dessas modalidades. Especificamente, no caso de controlar o otimizador de diálogo, o classificador de áudio 200 e a unidade de ajuste 300 no aparelho de processamento de áudio 100, como mostrado nas Figuras 1-10, podem constituir um controlador do otimizador de diálogo 1500, como mostrado na Figura 15. Nesta modalidade, já que a unidade de regulação é específica para o otimizador de diálogo, ele pode ser referido como 300A. E, como discutido na parte anterior, o classificador de áudio 200 pode compreender pelo menos um dentre o classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204 e o controlador do otimizador de diálogo 1500 ainda pode incluir pelo menos um dentre: a unidade de uniformização de tipo 712, a unidade de uniformização de parâmetro 814 e o temporizador 916.

[00225] Portanto, nesta parte, não serão repeditos, os conteúdos já descritos na parte na parte anterior e só dar alguns exemplos específicos respectivos.

[00226] Para um otimizador de diálogo, os parâmetros ajustáveis incluem, mas não se limitam, ao nível de otimização de diálogo, o nível

de fundo e os limites das bandas de frequência de determinação a serem otimizados. Ver H. Muesch. "Speech Enhancement in Entertainment Audio" publicado como WO 2008/106036 A2, cuja totalidade é incorporada aqui por referência.

2.1 Nível de otimização do diálogo

[00227] Ao envolver o nível de otimização de diálogo, a unidade de ajuste 300A pode ser configurada para correlacionar de maneira positivo o nível de otimização de diálogo do otimizador de diálogo com o valor de confiança da fala. Adicionalmente ou alternativamente, o nível pode ser correlacionado negativamente ao valor de confiança dos outros tipos de conteúdo. Assim, o nível de otimização de diálogo pode ser ajustado para ser proporcional (linearmente ou não-linearmente) à confiança da fala, de modo que a otimização de diálogo é menos efetiva em sinais não-de fala, como som de música ou de fundo (efeitos sonoros).

[00228] Quanto ao tipo de contexto, a unidade de ajuste 300A pode ser configurada para correlacionar de maneira positiva o nível de otimização de diálogo do otimizador de diálogo com o valor de confiança da mídia de tipo filme e/ou VoIP e/ou correlacionar negativamente o nível de otimização de diálogo do otimizador de diálogo com o valor de confiança da música de longa duração e/ou do jogo. Por exemplo, o nível de otimização de diálogo pode ser ajustado para ser proporcional (linearmente ou não-linearmente) ao valor de confiança da mídia de tipo filme. Quando o valor de confiança de mídia de tipo filme for 0 (por exemplo, no conteúdo da música, o nível de otimização de diálogo também é 0, que é equivalente a desabilitar o otimizador de diálogo.

[00229] Como descrito na parte precedente, o tipo de conteúdo e de contexto podem ser considerados em conjunto.

2.2 Limites para determinar bandas de frequência a serem

potencializadas

[00230] Durante o funcionamento do otimizador de diálogo, há um limite (geralmente limite de intensidade ou energia) para cada banda de frequência para determinar se precisa ser otimizado, isto é, aquelas bandas de frequência acima dos respectivos limites de intensidade/energia serão otimizadas. Para ajustar os limites, a unidade de ajuste 300A pode ser configurada para correlacionar positivamente os limites com um valor de confiança de música de curta duração e/ou sons de ruído e/ou de fundo, e/ou negativamente correlacionar os limites com um valor de confiança de fala. Por exemplo, os limites podem ser baixados se a confiança da fala for alta, presumindo uma detecção de fala mais confiável, para permitir que mais bandas de frequência sejam otimizadas; por outro lado, quando o valor de confiança da música for alto, os limites podem ser aumentados para fazer com que menos bandas de frequência sejam otimizadas (e, assim, menos artefatos).

2.3 Ajuste ao nível de fundo

[00231] Outro componente no otimizador de diálogo é a unidade de rastreamento mínima 4022, como mostrado na Figura 15, que é usado para estimar o nível de fundo no sinal de áudio (para estimativa de SNR e a estimativa de limite de banda de frequência, como mencionado na Seção 2.2). Ele também pode ser ajustado com base nos valores de confiança dos tipos de conteúdo de áudio. Por exemplo, se a confiança de fala for alta, a unidade de rastreamento mínimo pode ser de mais confiança para ajustar o nível de fundo ao mínimo atual. Se a confiança da música for alta, o nível de fundo pode ser estabelecido como um pouco mais elevado do que o mínimo atual ou, de outra maneira, estabelecida a uma média pesada de um mínimo atual e a energia do frame atual, com um amplo peso sobre o mínimo atual. Se o ruído e a confiança de fundo for alta, o nível de

fundo pode ser estabelecido como muito mais elevado do que o valor mínimo atual ou, de outra maneira, estabelecido a uma média pesada do mínimo atual e a energia do frame atual, com um pequeno peso sobre o mínimo atual.

[00232] Desse modo, a unidade de ajuste 300A pode ser configurada para atribuir um ajustamento ao nível de fundo estimado pela unidade de mapeamento mínimo, em que a unidade de ajuste é configurada ainda para correlacionar positivamente o ajuste com um valor de confiança de música de curta duração e/ou ruído e/ou som de fundo, e/ou negativamente correlacionar o ajuste com um valor de confiança de fala. Em uma variante, a unidade de ajuste 300A pode ser configurada para correlacionar o ajuste com o valor de confiança de ruído e/ou de fundo mais positivamente do que a música de curta duração.

2.4 Combinação das modalidades e cenários de aplicação

[00233] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser implementadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou iguais.

[00234] Por exemplo, quaisquer duas ou mais das soluções descritas nas seções 2.1 até a 2.3 podem ser combinadas uma com a outra. E essas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita na Parte 1 e outra partes que serão descritas posteriormente. Especialmente, muitas fórmulas são de fato aplicáveis a cada tipo de dispositivo de melhoria de áudio ou método, mas não são necessariamente mencionadas ou discutidas em cada parte desta divulgação. Em tal situação, pode-se fazer referência cruzada entre as partes dessa divulgação para aplicar uma

fórmula específica discutida em uma parte à outra parte, apenas com parâmetro(s), coeficiente(s), potência(s) (expoentes) e peso(s) relevante(s) sendo ajustados apropriadamente de acordo com requisitos específicos para o pedido específico.

2.5 Método de controle do otimizador de diálogo

[00235] De modo semelhante à Parte 1, no processo de descrição do controlador do otimizador de diálogo nas modalidades precedentes, são divulgados também, aparentemente, alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00236] Primeiramente, as modalidades do método de processamento de áudio, como discutido na Parte 1, podem ser usadas em um otimizador de diálogo, cujo(s) parâmetro(s) é/são um dos alvos a ser ajustado pelo método de processamento de áudio. Deste ponto da vista, o método de processamento de áudio também é um método de controle de otimizador de diálogo.

[00237] Nesta seção, apenas aqueles aspectos específicos para o controle do otimizador de diálogo serão discutidos. Para aspectos gerais do método de controle pode ser feita referência à parte 1.

[00238] De acordo com uma modalidade, o método de processamento de áudio adicionalmente pode incluir processamento de otimização de diálogo, e a operação de ajuste 1104 compreende correlacionar positivamente o nível da potencialização do diálogo com o valor de confiança da mídia similar a filme e/ou VoIP, e/ou negativamente correlacionar o nível de otimização de diálogo com o valor de confiança de música e/ou jogo de longa duração. Ou seja, otimização de diálogo é principalmente orientada para o sinal de áudio no contexto da mídia similar a filme ou VoIP.

[00239] Mais especificamente, a operação de ajuste 1104 pode incluir correlacionar positivamente o nível da otimização do diálogo do

otimizador de diálogo com o valor de confiança da fala.

[00240] O presente pedido também pode ajustar as bandas de frequência de modo a serem reforçadas no processamento de otimização de diálogo. Como mostrado na Figura 16, os limiares (geralmente a energia ou intensidade) para determinar se bandas de frequências respectivas devem ser reforçadas podem ser ajustados com base em valores de confiança de tipos de áudio identificados (operação 1602) de acordo com o presente pedido. Então, dentro do otimizador de diálogo, com base nos limiares ajustados, bandas de frequência acima de respectivos limiares são selecionadas (operação 1604) e potencializadas (operação 1606).

[00241] Especificamente, a operação de ajuste 1104 pode incluir positivamente correlacionar os limiares com um valor de confiança de música de curta duração e/ou sons de ruído e/ou de fundo, e/ou negativamente correlacionar os limiares com um valor de confiança de fala.

[00242] O método de processamento de áudio (especialmente o processamento de otimização de diálogo) geralmente ainda compreende estimar o nível de fundo no sinal de áudio, que geralmente é implementado por uma unidade de rastreamento mínima 4022 realizada no otimizador de diálogo 402 e usada em estimativa SNR ou estimativa de limiar de banda de frequência. O presente pedido também pode ser usado para ajustar o nível de fundo. Em tal situação, após o nível de fundo ser estimado (operação 1702), ele primeiro é ajustado com base em valor(es) de confiança de tipo(s) de áudio (operação 1704) e em seguida é usado na estimativa SNR e/ou a estimativa de limiar de banda de frequência (operação 1706). Especificamente, a operação de ajuste 1104 pode ser configurada para atribuir um ajustamento ao nível estimado de fundo, em que a operação de ajuste 1104 pode ser configurada ainda para

correlacionar positivamente o ajuste com um valor de confiança de música de curta duração e/ou ruído e/ou som de fundo, e/ou negativamente correlacionar o ajuste com um valor de confiança de fala.

[00243] Mais especificamente, a operação de ajuste 1104 pode ser configurada para correlacionar o ajuste com o valor de confiança de ruído e/ou de fundo mais positivamente do que a música de curta duração.

[00244] Semelhante às modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações são práticos, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações pode ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita na parte 1 e as outras partes que serão descritas mais tarde.

Parte 3: Controlador do Virtualizador Surround e Método de Controle

[00245] Um virtualizador surround possibilita que um sinal de som surround (como multicanais 5.1 e 7.1) seja distribuído aos alto-falantes internos do PC ou nos fones de ouvido. Isto é, com dispositivos estéreo como alto-falantes internos de laptop ou fones de ouvido, ele cria um efeito surround virtualmente e proporciona uma experiência cinemática aos consumidores. Funções de Transferência Relacionadas à Cabeça (HRTFs) são normalmente utilizadas no virtualizador surround para simular a chegada do som nas orelhas proveniente de vários locais de alto-falante associados com o sinal de áudio multicanal.

[00246] Embora o virtualizador surround atual funcione bem em

fones de ouvido, funciona de forma diferente em diferentes conteúdos com alto-falantes internos. Em geral, conteúdo de mídia similar a filme permite Virtualizador Surround para alto-falantes, enquanto a música não o faz já que pode soar com pouca ressonância.

[00247] Já que os mesmos parâmetros em Virtualizador Surround não podem criar boa imagem sonora para filme similar a mídia e conteúdo de música simultaneamente, parâmetros precisam ser sintonizados com base no conteúdo mais precisamente. Com informações de tipo de áudio disponíveis, especialmente o valor de confiança de música e valor de confiança de fala, bem como algumas outras informações de tipo de conteúdo e informações de contexto, o trabalho pode ser feito com o pedido atual.

[00248] Semelhante à parte 2, como um exemplo específico do aparelho de processamento de áudio e método discutido na parte 1, o virtualizador surround 404 pode fazer uso de todas as modalidades discutidas na parte 1 e todas as combinações dessas modalidades divulgadas ali. Especificamente, no caso de controlar o virtualizador surround 404, o classificador de áudio 200 e unidade de ajuste 300 no aparelho de processamento de áudio 100 como mostrado nas Figuras 1-10 pode constituir um controlador de virtualizador surround 1800 como mostrado na Figura 18. Na presente modalidade, já que a unidade de regulação é específica para o virtualizador surround 404, ela pode ser referida como 300B. Semelhante à parte 2, o classificador áudio 200 pode incluir pelo menos um dentre o classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204 e o controlador do virtualizador surround 1800 ainda pode incluir pelo menos um da unidade de uniformização de tipo 712, a unidade de uniformização de parâmetro 814 e o temporizador 916.

[00249] Portanto, nesta parte, não serão repetidos os conteúdos já descritos na parte 1 e só dar alguns exemplos específicos respectivos.

[00250] Para um virtualizador surround, os parâmetros ajustáveis incluem, mas não estão limitados à quantidade de aumento de surround e a frequência de início para o virtualizador surround 404.

3.1 Quantidade de Aumento de Surround

[00251] Quando envolvendo a quantidade de aumento de surround, a unidade de ajuste 300B pode ser configurada para correlacionar positivamente a quantidade de aumento de surround do virtualizador surround 404 com um valor de confiança de ruído e/ou fundo e/ou voz e/ou negativamente correlacionar a quantidade de aumento de surround com um valor de confiança de música de curta duração.

[00252] Especificamente modificar o virtualizador surround 404 a fim de que música (tipo de conteúdo) soe aceitável, um exemplo de implementação da unidade de ajuste 300B poderia ajustar a quantidade de aumento de surround baseado no valor de confiança de música de curta duração, tal como:

$$\mathrm{SB} \propto (1 - Conf_{music}) \tag{5}$$

[00253] onde SB indica a quantidade de aumento de surround, $Conf_{music}$ é o valor de confiança da música de curta duração.

[00254] Ajuda a diminuir o aumento de surround para música e impedi-la de soar abafada.

[00255] Da mesma forma, o valor de confiança da fala pode também ser utilizado, por exemplo:

$$\mathrm{SB} \propto (1 - Conf_{music}) * Conf_{speech}^{\alpha} \tag{6}$$

[00256] onde $Conf_{speech}$ é o valor de confiança da fala, a é um coeficiente de ponderação na forma de expoente e pode estar no intervalo de 1-2. Esta fórmula indica que a quantidade de aumento de surround será elevada para pura fala apenas (alta confiança de fala e confiança de música baixa).

[00257] Ou pode-se considerar apenas o valor de confiança da fala:

$$\mathrm{SB} \propto Conf_{speech} \qquad (7)$$

[00258] Várias variantes podem ser projetadas de forma semelhante. Especialmente, para o ruído ou som de fundo, podem ser construídas fórmulas semelhantes a fórmulas 5 a 7. Além disso, os efeitos dos quatro tipos de conteúdo podem ser considerados juntos em qualquer combinação. Em tal situação, ruído e fundo são sons de ambiência e são mais seguros para se ter uma quantidade grande de aumento; fala pode ter uma quantidade média de aumento, supondo um falador geralmente se sentar na frente da tela; e a música usa menos quantidade de aumento. Portanto, a unidade de ajuste 300B pode ser configurada para correlacionar a quantidade de aumento de surround com o valor de confiança do ruído e/ou fundo mais positivamente do que a fala de tipo de conteúdo.

[00259] Supondo que foi predefinida, uma quantidade esperada de aumento (que é equivalente a um peso) para cada tipo de conteúdo, pode ser também aplicada outra alternativa:

$$\hat{a} = \frac{a_{speech} \cdot Conf_{speech} + a_{music} \cdot Conf_{music} + a_{noise} \cdot Conf_{noise} + a_{bkg} \cdot Conf_{bkg}}{Conf_{speech} + Conf_{music} + Conf_{noise} + Conf_{bkg}} \qquad (8)$$

[00260] onde â é a quantidade estimada do aumento, a com um subscrito do tipo de conteúdo é a quantidade de aumento esperada/predefinida (peso) do tipo de conteúdo, *Conf* com um subscrito do tipo de conteúdo é o valor de confiança do tipo de conteúdo (onde bkg representa "som de fundo"). Dependendo de situações, a_{music} pode ser (mas não necessariamente) definida como 0, indicando que o virtualizador surround 404 será desativado para música pura (tipo de conteúdo).

[00261] De outro ponto de vista, *a* com um subscrito do tipo de conteúdo na fórmula (8) é a quantidade de aumento

esperado/predefinido de tipo de conteúdo, e o quociente do valor de confiança do tipo de conteúdo correspondente dividido pela soma dos valores de confiança todos os tipos de conteúdo identificados podem ser considerados como peso normalizado da quantidade predefinida/esperada de aumento do tipo de conteúdo correspondente. Ou seja, a unidade de ajuste 300B pode ser configurada para considerar pelo menos alguns dos vários tipos de conteúdo através de ponderação das quantidades de aumentos predefinidas dos vários tipos de conteúdo com base em valores de confiança.

[00262] Quanto ao tipo de contexto, a unidade de ajuste 300B pode ser configurada para correlacionar positivamente a quantidade de aumento de surround do virtualizador surround 404 com um valor de confiança da mídia similar a filme e/ou jogo, e/ou negativamente, correlacionar a quantidade de aumento de surround com um valor de confiança de música de longa duração e/ou VoIP. Em seguida, podem ser construídas fórmulas semelhantes às fórmulas 5 a 8.

[00263] Como um exemplo especial, o virtualizador surround 404 pode ser habilitado para pura mídia similar a filme e/ou o jogo, mas desabilitado para música e/ou VoIP. Enquanto isso, a quantidade de aumento do virtualizador surround 404 pode ser definida de forma diferente para mídia similar a filme e jogo, mídia similar a filme utiliza uma maior quantidade de aumento, e jogos usam menos. Portanto, a unidade de ajuste 300B pode ser configurada para correlacionar a quantidade de aumento de surround com o valor de confiança da mídia similar a filme mais positivamente do que jogo.

[00264] Semelhante ao tipo de conteúdo, a quantidade de aumento de um sinal de áudio também pode ser definida para uma média ponderada dos valores de confiança dos tipos de contexto:

$$\hat{a} = \frac{a_{MOVIE} \cdot Conf_{MOVIE} + a_{MUSIC} \cdot Conf_{MUSIC} + a_{GAME} \cdot Conf_{GAME} + a_{VOIP} \cdot Conf_{VOIP}}{Conf_{MOVIE} + Conf_{MUSIC} + Conf_{GAME} + Conf_{VOIP}} \qquad (9)$$

[00265] onde *â* é a quantidade estimada do aumento, *a* com um subscrito do tipo de contexto é a quantidade de aumento esperada/predefinida (peso) do tipo de contexto, Conf com um subscrito do tipo de contexto é o valor de confiança do tipo de contexto. Dependendo de situações, a_{MUSIC} e o a_{VOIP} podem ser (mas não necessariamente) definidos como 0, indicando que o virtualizador surround 404 será desativado para música pura (tipo de conteúdo) e/ou VoIP puro.

[00266] De novo, similar ao tipo de conteúdo, com um subscrito do tipo de conteúdo na fórmula (9) é a quantidade de aumento esperado/predefinido de tipo de conteúdo, e o quociente do valor de confiança do tipo de conteúdo correspondente dividido pela soma dos valores de confiança todos os tipos de conteúdo identificados podem ser considerados como peso normalizado da quantidade predefinida/esperada de aumento do tipo de conteúdo correspondente. Ou seja, a unidade de ajuste 300B pode ser configurada para considerar pelo menos alguns dos vários tipos de contexto através de ponderação das quantidades de aumentos predefinidas dos vários tipos de contexto com base em valores de confiança.

3.2 Frequência de Início

[00267] Outros parâmetros podem ser modificados também no virtualizador surround, tais como a frequência de início. Em geral, componentes de alta frequência em um sinal de áudio são mais adequados a serem renderizados espacialmente. Por exemplo, na música, isso vai soar estranho se o baixo espacialmente for renderizado para ter mais efeitos surround. Portanto, para um sinal de áudio específico, o virtualizador surround precisa determinar um limiar de frequência, os componentes acima do qual espacialmente são renderizados enquanto os componentes abaixo do qual são mantidos. O limiar de frequência é a frequência de início.

[00268] De acordo com uma modalidade do presente pedido, a frequência de início para o virtualizador de surround pode ser aumentada sobre o conteúdo de música para que mais baixo possa ser retido por sinais de música. Então, a unidade de ajuste 300B pode ser configurada para correlacionar positivamente a frequência de início do virtualizador surround com um valor de confiança da música de curta duração.

3.3 Combinação das modalidades e cenários de aplicação

[00269] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser executadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou iguais.

[00270] Por exemplo, quaisquer duas ou mais das soluções descritas nas seções 3.1 e 3.2 podem ser combinadas uma com a outra. E qualquer uma das combinações pode ainda ser combinada com qualquer modalidade descrita ou implícita na parte 1, parte 2 e as outras partes que serão descritas depois.

3.4 Método de controle do virtualizador surround

[00271] De modo semelhante à Parte 1, no processo de descrição do controlador do virtualizador surround nas modalidades precedentes, são divulgados também, aparentemente, alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00272] Primeiramente, as modalidades do método de processamento de áudio, como discutido na Parte 1, podem ser usadas em um virtualizador surround, cujo(s) parâmetro(s) é/são um dos alvos a ser ajustado pelo método de processamento de áudio. Deste ponto da vista, o método de processamento de áudio também é

um método de controle de virtualizador surround.

[00273] Nesta seção, apenas aqueles aspectos específicos para o controle do virtualizador surround serão discutidos. Para aspectos gerais do método de controle pode ser feita referência à parte 1.

[00274] De acordo com uma modalidade, o método de processamento de áudio ainda pode incluir processamento de virtualização surround e a operação de ajuste 1104 pode ser configurada para correlacionar positivamente a quantidade de aumento de surround do processamento de virtualização surround com um valor de confiança de ruído e/ou fundo e/ou voz e/ou negativamente correlacionar a quantidade de aumento de surround com um valor de confiança de música de curta duração.

[00275] Especificamente, a operação de ajustar 1104 pode ser configurada para correlacionar a quantidade de aumento de surround com o valor de confiança do ruído e/ou fundo mais positivamente do que a fala de tipo de conteúdo.

[00276] Em alternativa, ou adicionalmente, quantidade de aumento de surround ser ajustada também ser com base em valores de confiança do(s) tipo(s) de contexto(s). Especificamente, a operação de ajustar 1104 pode ser configurada para correlacionar positivamente a quantidade de aumento de surround do processamento de virtualização surround 404 com um valor de confiança da mídia similar a filme e/ou jogo, e/ou negativamente, correlacionar a quantidade de aumento de surround com um valor de confiança de música de longa duração e/ou VoIP.

[00277] Mais especificamente, a operação de ajustar 1104 pode ser configurada para correlacionar a quantidade de aumento de surround com o valor de confiança da mídia similar a filme mais positivamente do que jogo.

[00278] Outro parâmetro a ser ajustado é a frequência de início

para o processamento de virtualização surround. Como mostrado na Figura 19, a frequência de início é ajustada em primeiro lugar com base em valores de confiança de tipo(s) de áudio (operação 1902), então o virtualizador surround processa os componentes de áudio acima da frequência de início (operação 1904). Especificamente a operação de ajustar 1104 pode ser configurada para correlacionar positivamente a frequência de início do processamento de virtualização surround com um valor de confiança da música de curta duração.

[00279] Semelhante às modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações são práticos, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações pode ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em outras partes desta divulgação.

Parte 4: CONTROLADOR DE NIVELADOR DE VOLUME E MÉTODO DE CONTROLE

[00280] O volume de diferentes fontes de áudio ou peças diferentes na mesma fonte de áudio às vezes muda muito. É irritante, uma vez que os usuários têm que ajustar o volume com frequência. Um nivelador de volume (VL) direciona-se ao ajuste do volume do conteúdo de áudio em playback e mantê-lo praticamente consistente ao longo de um período de tempo com base no valor alvo de intensidade. Niveladores de Volume de exemplo são apresentados em J. Seefeldt et al. "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", publicado como US2009/0097676A1; B. G. Crockett et al. "Audio Gain Control

Using Specific-Loudness-Based Auditory Event Detection", publicado como WO2007/127023A1; e A. Seefeldt et al. "Audio Processing Using Auditory Scene Analysis and Spectral Skewness", publicado como WO 2009/011827 Al. Os três documentos são pelo presente são incorporados por referência em sua totalidade.

[00281] O nivelador de volume continuamente mede a intensidade de um sinal de áudio de alguma maneira e em seguida modifica o sinal por uma quantidade de ganho que é um fator de escala para modificar a intensidade de um sinal de áudio e geralmente é uma função do volume medido, o volume de destino desejado e vários outros fatores. Uma variedade de fatores necessários teve de ser considerada para se estimar um ganho adequado, com critérios subjacentes a ambas abordagens para o volume de destino e manter a gama dinâmica. Geralmente compreende vários subelementos, como controle de ganho automático (AGC), detecção de evento auditivo, controle de gama dinâmica (DRC).

[00282] Um sinal de controle é geralmente aplicado no nivelador de volume para controlar o "ganho" do sinal de áudio. Por exemplo, um sinal de controle pode ser um indicador da mudança na magnitude do sinal de áudio obtido por análise de puro sinal. Pode ser também um indicador de evento de áudio para representar se aparece um novo evento de áudio, através da análise psico-acústica, tais como análise de cena auditiva ou detecção de evento auditivo especificamente-com-base-em-sonoridade. Tal sinal de controle é aplicado no nivelador de volume para controlar o ganho, por exemplo, garantindo que o ganho seja quase constante dentro de um evento auditivo e confinando muito da mudança de ganho para a vizinhança de um limite de evento, a fim de reduzir possíveis artefatos audíveis devido a uma rápida mudança do ganho no sinal de áudio.

[00283] No entanto, os métodos convencionais de derivação sinais

de controle não podem diferenciar eventos auditivos informativos de eventos auditivos não-informativos (interferentes). Aqui, o evento informativo auditivo representa o evento de áudio que contém informações significativas e a ele pode ser prestada mais atenção pelos usuários, tais como diálogo e música, enquanto o sinal não-informativo não contém informações significativas para os usuários, tais como ruído em VoIP. Como uma consequência, os sinais não-informativos também podem ser aplicados por um grande ganho e aumentados para perto da sonoridade alvo. Vai ser desagradável em algumas aplicações. Por exemplo, em chamadas VoIP, o sinal de ruído que aparece na pausa de uma conversa é muitas vezes impulsionado para um volume alto, depois de processado por um nivelador de volume. Isto é não desejado pelos usuários.

[00284] A fim de resolver este problema, pelo menos em parte, o presente pedido propõe controlar o nivelador de volume baseado nas modalidades discutidas na parte 1.

[00285] Semelhante à parte 2 e 3, como um exemplo específico do aparelho de processamento de áudio e método discutido na parte 1, o nivelador de volume 406 pode fazer uso de todas as modalidades discutidas na parte 1 e todas as combinações dessas modalidades divulgadas ali. Especificamente, no caso de controlar o nivelador de volume 406, o classificador de áudio 200 e unidade de ajuste 300 no aparelho de processamento de áudio 100 como mostrado nas Figuras 1-10 pode constituir um controlador 2000 de nivelador de volume 406, como mostrado na Figura 20. Nesta modalidade, já que a unidade de regulação é específica para o nivelador de volume 406, ele pode ser referido como 300C.

[00286] Isto é, baseado na divulgação da parte 1, um controlador de nivelador de volume 2000 pode incluir um classificador de áudio 200 para continuamente identificar o tipo de áudio (como tipo de conteúdo

e/ou tipo de contexto) de um sinal de áudio; e uma unidade de ajuste 300 para ajustar um nivelador de volume de forma contínua baseado no valor de confiança do tipo de áudio identificado. Da mesma forma, o classificador de áudio 200 pode incluir pelo menos um do classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204, e o controlador de nivelador de volume 2000 ainda pode incluir pelo menos uma unidade de uniformização de tipo 712, a unidade de uniformização de parâmetro 814 e o temporizador 916.

[00287] Portanto, nesta parte, não serão repetidos os conteúdos já descritos na parte 1 e só dar alguns exemplos específicos respectivos.

[00288] Parâmetros diferentes no regulador do volume 406 podem ser ajustados adaptativamente baseado nos resultados de classificação. Podem-se ajustar os parâmetros diretamente relacionados com o ganho dinâmico ou a gama do ganho dinâmico, por exemplo, reduzindo o ganho para os sinais não-informativos. Pode-se também ajustar os parâmetros que indicam o grau do sinal sendo um novo evento de áudio perceptível e então indiretamente controlar o ganho dinâmico (o ganho vai mudar lentamente dentro de um evento de áudio, mas pode mudar rapidamente na fronteira de dois eventos de áudio). Nesta aplicação, apresentam-se várias modalidades do mecanismo de controle de nivelador de volume ou ajuste de parâmetro.

4.1 Tipos de Conteúdo Informativos e Interferentes

[00289] Como mencionado acima, em conexão com o controle do nivelador de volume, os tipos de conteúdo de áudio podem ser classificados como tipos de conteúdo informativos e tipos de conteúdo interferentes. A unidade de ajuste 300C pode ser configurada para se correlacionar de maneira coletiva a um ganho dinâmico do nivelador de volume com tipos de conteúdo informativo do sinal de áudio e correlacionar negativamente o ganho dinâmico do nivelador de volume

com os tipos de conteúdo de interferência do sinal de áudio.

[00290] Como um exemplo, supondo que o ruído está interferindo (não-informativo) e vai ser irritante aumentá-lo a um volume alto, o parâmetro diretamente controlando o ganho dinâmico ou o parâmetro indicando novos eventos de áudio pode ser definido para ser proporcional para uma função de diminuição do valor de confiança de ruído ($Conf_{noise}$), tal como

$$\text{GainControl} \propto 1 - Conf_{noise} \qquad (10)$$

[00291] Aqui, para simplificar, foi usado o símbolo GainControl para representar todos os parâmetros (ou seus efeitos) relacionados a controle de ganho no nivelador de volume, já que implementações diferentes de regulador do volume podem usar nomes diferentes de parâmetros com diferente significado subjacente. Usando o único termo GainControl pode ter uma expressão curta sem perder generalidade. Em essência, ajustar esses parâmetros é equivalente a aplicar um peso sobre o ganho original, linear ou não linear. Como um exemplo, o GainControl pode ser diretamente usado para o ganho de escala, para que o ganho seja pequeno se GainControl é pequeno. Como outro exemplo específico, o ganho indiretamente é controlado por escala com GainControl do sinal de controle de evento descrito em B.G.Grockett et al. "Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection", publicado como WO2007/127023A1, que é incorporado aqui na sua totalidade por referência. Neste caso, quando GainControl é pequeno, os controles de ganho do nivelador de volume são modificados para evitar o ganho de se alterar significativamente com o tempo. Quando GainControl é alto, os controles são modificados para que o ganho do nivelador seja permitido a mudar mais livremente.

[00292] Com o controle de ganho descrito na fórmula (10) (ou diretamente dimensionar em escala o ganho original ou o sinal de

controle do evento), o ganho dinâmico de um sinal de áudio está correlacionado (linear ou não linearmente) ao valor de confiança de ruído. Se o sinal é ruído com um valor de alta confiança, o ganho final será pequeno devido ao fator (1 - $Conf_{noise}$). Desta forma, evita a aumentar um sinal de ruído em um volume alto desagradável.

[00293] Como uma variante de exemplo de fórmula (10), se o som de fundo também não está interessado em um aplicativo (como em VoIP), pode ser tratado da mesma forma e aplicado por um pequeno ganho também. Uma função de controle pode considerar tanto o valor de confiança de ruído ($Conf_{noise}$) e o valor de confiança de fundo ($Conf_{bkg}$), por exemplo

$$\text{GainControl} \propto (1 - Conf_{noise}) \cdot (1 - Conf_{bkg}) \qquad (11)$$

[00294] Na fórmula acima, uma vez que tanto barulho e sons de fundo não são desejados, o GainControl é igualmente afetado pelo valor de confiança do ruído e o valor de confiança do fundo e pode ser considerado que ruído e sons de fundo têm o mesmo peso. Dependendo de situações, eles podem ter pesos diferentes. Por exemplo, pode-se dar aos valores de confiança de ruído e sons de fundo (ou sua diferença com 1) diferentes coeficientes ou expoentes diferentes (a e γ). Ou seja, a fórmula (11) pode ser reescrita como:

$$\text{GainControl} \propto (1 - Conf_{noise})^{\alpha} \cdot (1 - Conf_{bkg})^{\gamma} \qquad (12)$$

ou $$\text{GainControl} \propto (1 - Conf_{noise}^{\alpha}) \cdot (1 - Conf_{bkg}^{\gamma}) \qquad (13)$$

[00295] Como alternativa, a unidade de ajuste 300C pode ser configurada para considerar pelo menos um tipo de conteúdo dominante, com base nos valores de confiança. Por exemplo:

$$\text{GainControl} \propto 1 - \max(Conf_{noise}, Conf_{bkg}) \qquad (14)$$

[00296] Tanto a fórmula (11) (e suas variantes) e fórmula (14) indicam um pequeno ganho para sinais de ruído e sinais sonoros de fundo, e o comportamento original do nivelador de volume é mantido somente quando ambos confiança de ruído e confiança do fundo é pequena (como em sinal de fala e música), assim GainControl estando perto de um.

[00297] O exemplo acima considera o tipo de conteúdo interferente dominante. Dependendo da situação, a unidade de ajuste 300C também pode ser configurada para considerar o tipo de conteúdo informativo dominante, com base nos valores de confiança. Para ser mais geral, a unidade de 300 pode ser configurada para considerar pelo menos um tipo de conteúdo dominante, com base nos valores de confiança, não importa se os tipos de áudio identificados são/incluem tipos de áudio informativos e/ou interferências.

[00298] Como uma outra variante de exemplo de fórmula (10), supondo que o sinal de faça é o conteúdo mais informativo e precisa de menos modificação no comportamento padrão do nivelador de volume, a função de controle pode considerar tanto o valor de confiança de ruído ($Conf_{noise}$) e o valor de confiança da fala ($Conf_{speech}$), como

$$\text{GainControl} \propto 1 - Conf_{noise} \cdot (1 - Conf_{speech}) \qquad (15)$$

[00299] Com essa função, um pequeno GainControl é obtido apenas para aqueles sinais com ruído de alta confiança e confiança baixa de fala (por exemplo, ruído puro), e o GainControl vai estar perto de 1 se a confiança da fala é alta (e, assim, manter o comportamento original do nivelador de volume). Em geral, pode considerar-se que o peso de um tipo de conteúdo (como $Conf_{noise}$) pode ser modificado com o valor de confiança pelo menos um outro tipo de conteúdo (como $Conf_{speech}$). Na fórmula acima (15), pode ser considerado que a confiança da fala altera o coeficiente de peso de confiança do ruído

(outro tipo de peso se comparado com os pesos na fórmula (12 e 13)). Em outras palavras, na fórmula (10), o coeficiente de $Conf_{speech}$ pode ser considerado como 1; enquanto que na fórmula (15), alguns outros tipos de áudio (coma fala, mas não se limitando) afetarão a importância do valor confiança de ruído, pode dizer-se o peso de $Conf_{noise}$ é modificado pelo valor de confiança da fala. No contexto da presente divulgação, o termo "peso" deve ser interpretado para incluir isto. Ou seja, ele indica a importância de um valor, mas não necessariamente normalizado. Pode ser feita referência à seção 1.4.

[00300] De outro ponto de vista, semelhante à fórmula (12) e (13), pesos na forma de expoentes podem ser aplicados sobre os valores de confiança na função acima para indicar a prioridade (ou importância) de diferentes sinais de áudio, por exemplo, a fórmula (15) pode ser alterada para:

$$\text{GainControl} \propto 1 - Conf_{noise}^{\alpha} \cdot (1 - Conf_{speech})^{\gamma} \qquad (16)$$

[00301] onde a e γ são dois pesos, que podem ser definidos como menores se é esperado que sejam mais responsivos para modificar os parâmetros do nivelador.

[00302] As fórmulas (10)-(16) podem ser livremente combinadas para formar várias funções de controle que podem ser adequadas em aplicações diferentes. Os valores de confiança de outros tipos de conteúdo áudio, tais como valor de confiança de música, podem ser também facilmente incorporados nas funções de controle em uma maneira similar.

[00303] No caso onde o GainContrtol é usado para ajustar os parâmetros que indicam o grau do sinal sendo um novo evento de áudio perceptível e, em seguida, controlar indiretamente o ganho dinâmico (o ganho vai mudar lentamente dentro de um evento de áudio, mas pode mudar rapidamente, na fronteira de dois eventos de áudio), pode considerar-se que existe uma outra função de

transferência entre o valor de confiança dos tipos de conteúdo e o ganho dinâmico final.

4.2 Tipos de conteúdo em diferentes contextos

[00304] As fórmulas de controle acima (10)-(16) tomam em consideração os valores de confiança dos tipos de conteúdo de áudio, tais como ruídos, sons de fundo, música de curta duração e fala, mas não consideram seus contextos áudio de onde os sons vêm, como mídia similar a filme e VoIP. É possível que o mesmo tipo de conteúdo de áudio pode precisar ser processado de forma diferente em diferentes contextos de áudio, por exemplo, os sons de fundo. Som de fundo é composto por vários sons como motor de carro, explosão e aplausos. Isso pode não ser significativo em uma chamada de VoIP, mas pode ser importante em uma mídia similar a filme. Isso indica que os contextos de áudio interessados precisam ser identificados e diferentes funções de controle precisam ser projetadas para diferentes contextos de áudio.

[00305] Portanto, a unidade de ajuste 300C pode ser configurada para considerar o tipo de conteúdo o sinal de áudio como informativos ou interferentes com base no tipo de contexto do sinal de áudio. Por exemplo, considerando o valor de confiança de ruído e valor de confiança de fundo, e diferenciando os contextos de VoIP e não VoIP, uma função de controle de áudio contexto-dependente pode ser,

se *o contexto de áudio for VoIP*

$$\text{GainControl} \propto 1 - \max(Conf_{noise},\ Conf_{bkg})$$

senão (17)

$$\text{GainControl} \propto 1 - Conf_{noise}$$

[00306] Ou seja, no contexto de VoIP, ruído e sons de fundo são considerados como tipos de conteúdo interferentes; enquanto no contexto não-VoIP, som de fundo é considerado como o tipo de conteúdo informativo.

[00307] Como outro exemplo, uma função de controle de áudio de contexto-dependente considerando valores de confiança de fala, ruído e fundo e diferenciando contextos de VoIP e não-VoIP, poderia ser

se *o contexto de áudio for VoIP*

$$\text{GainControl} \propto 1 - \max(Conf_{noise}, Conf_{bkg})$$

senão (18)

$$\text{GainControl} \propto 1 - Conf_{noise} \cdot (1 - Conf_{speech})$$

[00308] Aqui, a fala é enfatizada como um tipo de conteúdo informativo.

[00309] E supondo que a música é também informação informativa importante no contexto de não-VoIP, pode-se estender a segunda parte da fórmula (18) para:

$$\text{GainControl} \propto 1 - Conf_{noise} \cdot (1 - \max(Conf_{speech}, Conf_{music})) \quad (19)$$

[00310] Na verdade, cada uma das funções de controle em (10) - (16) ou suas variantes pode ser aplicada em contextos de áudio diferentes/correspondentes. Assim, pode gerar um grande número de combinações para formar funções de controle dependentes de contexto de áudio.

[00311] Além de contextos de VoIP e não VoIP como diferenciados e utilizados na fórmula (17) e (18), além de outros contextos de áudio, tais como mídia similar a filme, música de longa duração e jogo, ou baixa qualidade de áudio e áudio de alta qualidade, podem ser utilizados de forma semelhante

4.3 Tipos de Contexto

[00312] Tipos de contexto também podem ser usados diretamente para controlar o nivelador de volume para evitar aqueles irritantes sons, tais como ruído, de serem aumentados demais. Por exemplo, o valor de confiança de VoIP pode ser usado para orientar o nivelador de volume, tornando-o menos sensível quando seu valor de confiança

é alto.

[00313] Especificamente, com o valor de confiança do VoIP de $Conf_{VOIP}$, o nível do nivelador de volume pode ser definido como proporcional a (1 -$Conf_{VOIP}$). Ou seja, o nivelador de volume é quase desativado no conteúdo de VoIP (quando o valor de confiança de VoIP é alto), que é consistente com a configuração tradicional manual (predefinição) que desativa o nivelador de volume para o contexto de VoIP.

[00314] Como alternativa, pode-se definir gamas de ganho dinâmico diferentes para diferentes contextos de sinais de áudio. Em geral, uma quantidade de VL (nivelador de volume) ajusta mais a quantidade de ganho aplicada em um sinal de áudio e pode ser vista como outro peso (não linear) sobre o ganho. Em uma modalidade, poderia ser uma configuração:

Tabela 1

	MÍDIA TIPO FILME	MÚSICA A LONGA DURAÇÃO	VOIP	JOGO
Quantidade de VL	alta	média	Desligado (ou a mais baixa)	baixa

[00315] Além disso, supondo uma quantidade esperada de VL predefinida para cada tipo de contexto. Por exemplo, a quantidade de VL é definida como 1 para Mídia similar a filme, 0 para VoIP, 0,6 para Música e 0,3 para Jogo, mas o presente pedido não está limitado aos mesmos. De acordo com o exemplo, se a gama de ganho de dinâmica da mídia similar a filme é 100%, então a escala do ganho dinâmico de VoIP é 60% e assim por diante. Se a classificação do classificador de áudio 200 baseia-se em decisão difícil, então a gama do ganho dinâmico pode ser diretamente definida como o exemplo acima. Se a

classificação do classificador de áudio 200 baseia-se em decisão leve, então a gama pode ser ajustada com base no valor de confiança do tipo de contexto.

[00316] Da mesma forma, o classificador de áudio 200 pode identificar vários tipos de contexto do sinal de áudio, e a unidade de ajuste 300C pode ser configurada para ajustar o intervalo do ganho dinâmico por ponderação dos valores de confiança dos vários tipos de conteúdo com base na importância dos vários tipos de conteúdo.

[00317] Geralmente, para o tipo de contexto, as funções semelhantes (10)-(16) podem também ser usadas aqui para definir a quantidade adequada de VL adaptativamente, com os tipos de conteúdo ali substituído com tipos de contexto, e na verdade a tabela 1 reflete a importância dos tipos de contexto diferentes.

[00318] De outro ponto de vista, o valor de confiança pode ser usado para derivar um peso normalizado, conforme discutido na seção 1.4. Supondo que uma quantidade específica é predefinida para cada tipo de contexto na tabela 1, então uma fórmula similar à fórmula (9) pode igualmente ser aplicada. Aliás, soluções semelhantes também podem ser aplicadas a vários tipos de conteúdo e de quaisquer outros tipos de áudio.

4.4 Combinação das modalidades e cenários de aplicação

[00319] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser implementadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou iguais. Por exemplo, quaisquer duas ou mais das soluções descritas nas seções 4.1 até a 4.3 podem ser combinadas uma com a outra. E qualquer uma das combinações pode ainda ser combinada com qualquer modalidade descrita ou implícita nas partes 1-3 e as outras

partes que serão descritas depois.

[00320] Figura 21 ilustra o efeito do controlador de nivelador de volume proposto no pedido, comparando um segmento de curta duração original (Figura 21(A)), o segmento de curta duração processado por um nivelador de volume convencional sem modificação de parâmetros (Figura 21(B) e o segmento de curta duração processado por um nivelador de volume conforme apresentado neste pedido (Figura 21(C)). Como visto, no nivelador de volume convencional como mostrado em Figura 21(B), o volume do ruído (segunda metade do sinal de áudio) também sofre aumento e é irritante. Em contraste, no novo nivelador de volume conforme mostrado na Figura 21(C), o volume da parte efetiva do sinal de áudio é aumentado sem aparentemente aumentar o volume do ruído, dando à audiência boa experiência.

4.5 Método de Controle de Nivelador de Volume

[00321] Semelhante à parte 1, no processo de descrever o controlador de nivelador de volume nas modalidades seguintes, são também aparentemente divulgados alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00322] Primeiramente, as modalidades do método de processamento de áudio, como discutido na Parte 1, podem ser usadas em um nivelador de volume, cujo(s) parâmetro(s) é/são um dos alvos a ser ajustado pelo método de processamento de áudio. Deste ponto da vista, o método de processamento de áudio também é um método de controle de nivelador de volume.

[00323] Nesta seção, apenas aqueles aspectos específicos para o controle do nivelador de volume serão discutidos. Para aspectos gerais do método de controle pode ser feita referência à parte 1.

[00324] De acordo com o presente pedido, um método de controle

do nivelador de volume é fornecido, que inclui a identificação do tipo de conteúdo de um sinal de áudio em tempo real e o ajuste de um nivelador de volume de maneira contínua com base no tipo de conteúdo, conforme identificado, pela correlação positiva do ganho dinâmico do nivelador de volume com os tipos de conteúdo informativo do sinal de áudio e pela correlação negativa do ganho dinâmico do nivelador de volume com os tipos de conteúdo de interferência do sinal de áudio.

[00325] O tipo de conteúdo pode incluir fala, música de curta duração, ruído e som de fundo. Geralmente, o ruído é considerado como um tipo de conteúdo interferente.

[00326] Ao se ajustar o ganho dinâmico do nivelador de volume, ele pode ser ajustado diretamente com base no valor de confiança do tipo de conteúdo, ou pode ser ajustado através de uma função de transferência do valor de confiança do tipo de conteúdo.

[00327] Como já descrito, o sinal de áudio pode ser classificado em vários tipos de áudio ao mesmo tempo. Quando envolvendo vários tipos de conteúdo, a operação de ajuste 1104 pode ser configurada para considerar pelo menos alguns dos vários tipos de conteúdo de áudio através de ponderação dos valores de confiança dos vários tipos de conteúdo com base na importância dos vários tipos de conteúdo, ou através de ponderação dos efeitos dos vários tipos de conteúdo com base em valores de confiança. Em especial, a operação de ajuste 1104 pode ser configurada para considerar pelo menos um tipo de conteúdo dominante com base nos valores de confiança. Quando o sinal de áudio contém tipo(s) de conteúdo interferente(s) e tipo(s) de conteúdo informativo(s), a operação de ajuste pode ser configurada para considerar pelo menos um tipo de conteúdo interferente dominante com base nos valores de confiança, e/ou considerar pelo menos um tipo de conteúdo informativo dominante com base nos

valores de confiança.

[00328] Diferentes tipos de áudio podem afetar um ao outro. Portanto, a operação de ajuste 1104 pode ser configurada para modificar o peso de um tipo de conteúdo com o valor de confiança de pelo menos um outro tipo de conteúdo.

[00329] Conforme descrito na parte 1, o valor de confiança do tipo de áudio do sinal de áudio pode ser unificado. Para o detalhe da operação de unificação, por favor consulte parte 1.

[00330] Método pode ainda compreender adicionalmente a identificação do tipo de contexto do sinal de áudio, em que a operação de ajuste 1104 é configurada para ajustar o intervalo do ganho dinâmico com base no valor de confiança do tipo de contexto.

[00331] O papel de um tipo de conteúdo é limitado pelo tipo de contexto onde está localizado. Portanto, quando informações de tipo de conteúdo e informações de tipo de contexto são obtidas por um sinal de áudio ao mesmo tempo (ou seja, para o mesmo segmento de áudio), o tipo de conteúdo do sinal de áudio pode ser determinado como informativo ou interferente com base no tipo de contexto do sinal de áudio. Ademais, o tipo de conteúdo em um sinal de áudio de um tipo de contexto diferente pode ser designado com um peso diferente dependendo do tipo de contexto do sinal de áudio. De outro ponto de vista, pode-se usar peso diferente (maior ou menor, mais valor ou menos valor) para refletir a natureza informativa ou natureza interferente de um tipo de conteúdo.

[00332] O tipo de contexto do sinal de áudio pode incluir VoIP, mídia similar a filme, música de longa duração e jogo. E no sinal de áudio de contexto tipo Voz sobre IP, o som de fundo é considerado como tipo de conteúdo interferente; ao passo que no sinal de áudio de contexto tipo não-Voz sobre IP, o som de fundo e/ou fala e/ou música é considerado como tipo informativo. Outros tipos de contexto podem

incluir alta qualidade de áudio ou áudio de baixa qualidade.

[00333] Semelhante para os vários tipos de conteúdo, quando o sinal de áudio é classificado em vários tipos de contexto com valores de confiança correspondentes ao mesmo tempo (em relação ao mesmo segmento de áudio), a operação de ajuste 1104 pode ser configurada para considerar pelo menos alguns de múltiplos tipos de contexto por meio de ponderação dos valores de confiança dos vários tipos de contexto com base na importância dos vários tipos de contexto, ou através de ponderação dos efeitos dos vários tipos de contexto com base nos valores de confiança. Em especial, a operação de ajuste pode ser configurada para considerar pelo menos um tipo de contexto dominante com base nos valores de confiança.

[00334] Finalmente, as modalidades do método como discutido nesta seção podem usar o método de classificação de áudio, como será discutido em partes 6 e 7, e descrição detalhada é omitida aqui.

[00335] Semelhante para as modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações é prática, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações pode ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em outras partes desta divulgação.

Parte 5: Controlador de Equalizador e Método de Controle

[00336] Equalização geralmente é aplicada em um sinal de música para ajustar ou modificar seu equilíbrio espectral, mais conhecido como "tom" ou "timbre". Um equalizador tradicional permite aos usuários configurar o perfil global (curva ou forma) da resposta de frequência (ganho) em cada banda de frequência individual, a fim de

enfatizar certos instrumentos ou remover sons indesejáveis. Tocadores de música populares, como o player de mídia do windows, geralmente fornecem um equalizador gráfico para ajustar o ganho em cada banda de frequência e também fornecem um conjunto de predefinições de equalizador para gêneros musicais diferentes, como Rock, Rap, Jazz e Folk, para obter a melhor experiência em ouvir diferentes gêneros de música. Depois que uma predefinição é selecionada, ou um perfil é definido, os mesmos ganhos de equalização serão aplicados sobre o sinal, até o perfil ser modificado manualmente.

[00337] Em contraste, um equalizador dinâmico fornece uma maneira para ajustar automaticamente os ganhos de equalização em cada banda de frequência para manter a coerência global do equilíbrio espectral no que se refere a um timbre desejado ou tom. Essa consistência é conseguida monitorando continuamente o equilíbrio espectral do áudio, comparando-o com um equilíbrio espectral predefinido desejado e ajustando dinamicamente os ganhos de equalização aplicados para transformar o equilíbrio espectral original do áudio para o desejado equilíbrio espectral. O equilíbrio espectral desejado é manualmente selecionado ou pré-definido antes do processamento.

[00338] Os dois tipos de equalizadores compartilham a seguinte desvantagem: o melhor perfil de equalização, o desejado equilíbrio espectral ou os parâmetros relacionados tem que ser selecionados manualmente, e eles não podem ser automaticamente modificados baseados no conteúdo de áudio na reprodução. Discriminar tipos de conteúdo de áudio será muito importante para fornecer qualidade boa geral para diferentes tipos de sinais de áudio. Por exemplo, peças de música diferente precisam de perfis diferentes de equalização, tais como aqueles de diferentes gêneros.

[00339] Em um sistema de equalização em que quaisquer tipos de sinais de áudio (não só música) são possíveis de serem entrados, os parâmetros de equalização precisam ser ajustados com base em tipos de conteúdo. Por exemplo, equalizador geralmente é habilitado em sinais de música, mas deficiente em sinais de fala, uma vez que pode mudar o timbre da fala demais e correspondentemente fazer o som de sinal não soar natural.

[00340] A fim de resolver este problema, pelo menos em parte, o presente pedido propõe controlar o equalizador baseado nas modalidades discutidas na parte 1.

[00341] Semelhante às partes 2 e 4, como um exemplo específico do aparelho de processamento de áudio e método discutido na parte 1, o equalizador 408 pode fazer uso de todas as modalidades discutidas na parte 1 e todas as combinações dessas modalidades divulgadas ali.

[00342] Especificamente, no caso de controlar o equalizador 408, o classificador de áudio 200 e unidade de ajuste 300 no aparelho de processamento de áudio 100 como mostrado nas Figuras 1-10 pode constituir um controlador 2200 de equalizador 408 como mostrado na Figura 22. Nesta modalidade, já que a unidade de regulação é específica para o nivelador de equalizador 408, ele pode ser referido como 300D.

[00343] Isto é, baseado na divulgação da parte 1, um equalizador 2200 pode incluir um classificador de áudio 200 para continuamente identificar o tipo de áudio de um sinal de áudio; e uma unidade de ajuste 300D para ajustar um equalizador de forma contínua baseado no valor de confiança do tipo de áudio identificado. Da mesma forma, o classificador de áudio 200 pode incluir pelo menos um do classificador de conteúdo de áudio 20 2e o classificador de contexto de áudio 204, e o controlador de equalizador 2200 ainda pode incluir pelo menos

uma unidade de uniformização de tipo 712, a unidade de uniformização de parâmetro 814 e o temporizador 916.

[00344] Portanto, nesta parte, não serão repetidos os conteúdos já descritos na parte 1 e só dar alguns exemplos específicos respectivos.

5.1 Controle com base no tipo de conteúdo

[00345] De um modo geral, para tipos de conteúdo de áudio gerais tais como música, fala, som de fundo e ruído, o equalizador deve ser definido de forma diferente em diferentes tipos de conteúdo. Semelhante à instalação tradicional, o equalizador pode ser automaticamente habilitado em sinais de música, mas desabilitado na fala; ou de forma mais contínua, definido num nível de alta equalização de sinais de música e equalização de baixo nível em sinais de fala. Desta forma, o nível de equalização de um equalizador pode automaticamente ser definido para conteúdo de áudio diferente.

[00346] Especificamente para a música, observa-se que o equalizador não funciona tão bem em uma peça de música que tem uma fonte dominante, já que o timbre da fonte dominante pode se alterar significativamente e não soar natural se for aplicada uma equalização inadequada. Considerando isso, seria melhor definir uma equalização de baixo nível em peças de música com fontes dominantes, enquanto o nível de equalização pode ser mantido alto em peças de música sem fontes dominantes. Com esta informação, o equalizador pode automaticamente definir o nível de equalização para conteúdo de música diferente.

[00347] Música também pode ser agrupada com base em propriedades diferentes, tais como gênero, instrumento e características de música gerais, incluindo ritmo, tempo e timbre. Da mesma maneira que predefinições de equalizador diferentes são utilizadas para gêneros musicais diferentes, estes grupos/conjuntos de música também podem ter seus próprios perfis de equalização ideal

ou curvas de equalização (no equalizador tradicional) ou equilíbrio espectral ideal desejado (em equalizador dinâmico).

[00348] Como mencionado acima, o equalizador geralmente é habilitado no conteúdo de música, mas com desabilitado na fala, já que o equalizador pode fazer a caixa de diálogo não soar tão bem devido à mudança de timbre. Um método para alcançar automaticamente é relacionar o nível de equalização ao conteúdo, em especial o valor de confiança de música e/ou valor de confiança de fala obtido a partir do módulo de classificação de conteúdo de áudio. Aqui, o nível de equalização pode ser explicado como o peso dos ganhos de equalizador aplicados. Quanto maior o nível, mais forte a equalização aplicada. Por exemplo, se o nível de equalização é 1, um perfil completo de equalização é aplicado; se o nível de equalização é zero, todos os ganhos são correspondentemente OdB e, portanto, não-equalização é aplicada. O nível de equalização pode ser representado por diferentes parâmetros em diferentes implementações de algoritmos de equalizador. Uma modalidade de exemplo deste parâmetro é o peso de equalizador conforme implementado em A. Seefeldt et.al. "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", publicada como US 2009/0097676 Al, que é incorporada aqui na sua totalidade por referência.

[00349] Diversos esquemas de controle podem ser projetados para ajustar o nível de equalização. Por exemplo, com as informações de tipo de conteúdo de áudio, o valor de confiança de fala ou valor de confiança de música pode ser usado para definir o nível de equalização, como

$$L_{eq} \propto Conf_{music} \quad (20)$$

Ou

$$L_{eq} \propto 1 - Conf_{speech} \quad (21)$$

[00350] onde L_{eq} é o nível de equalização *e* $Conf_{music}$ e $Conf_{speech}$ significam o valor de confiança da música e da fala.

[00351] Ou seja, unidade de ajuste 300 pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de música de curta duração, ou negativamente, correlacionar o nível de equalização com um valor de confiança da fala.

[00352] O valor de confiança da fala e o valor de confiança de música podem ser ainda usados em conjunto para definir o nível de equalização. A ideia geral é que o nível de equalização deva ser alto somente quando o tanto valor de confiança de música é alto e valor de confiança da fala é baixo, e caso contrário, o nível de equalização é baixo. Por exemplo,

$$L_{eq} = Conf_{music}(1 - Conf_{speech}{}^{\alpha}) \qquad (22)$$

[00353] valor de confiança é alimentado à ordem de a para lidar com confiança de fala diferente de zero em sinais de música que podem acontecer com frequência. Com a fórmula acima, equalização será plenamente aplicada (com o nível igual a 1) sobre os sinais de música puros sem quaisquer componentes da fala. Como dito na parte 1, *a* pode ser considerado como um coeficiente de ponderação baseado na importância do tipo de conteúdo e pode ser normalmente definido como 1 para 2.

[00354] Se posando maior peso no valor confiança da fala, a unidade de ajuste 300 pode ser configurada para desativar o equalizador 408 quando o valor de confiança para a fala de tipo de conteúdo é maior que um limiar.

[00355] Na descrição acima, os tipos de conteúdo de música e fala são tidos como exemplos. Em alternativa ou adicionalmente, também podem ser considerados os valores de confiança de som de fundo e/ou ruído. Especificamente, a unidade de ajuste 300D pode

ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de fundo e/ou negativamente correlacionar o nível de equalização com um valor de confiança de ruído.

[00356] Em outra modalidade o valor de confiança pode ser usado para derivar um peso normalizado, conforme discutido na seção 1.4. Supondo que um nível esperado de equalização é predefinido para cada tipo de conteúdo (por exemplo, 1 para a música, 0 para a fala, 0,5 para ruído e fundo), uma fórmula similar à fórmula (8) pode ser aplicada exatamente.

[00357] O nível de equalização pode ser ainda mais unificado para evitar a rápida mudança que pode introduzir artefatos audíveis em pontos de transição. Isso pode ser feito com a unidade de uniformização de parâmetro 814, conforme descrito na seção 1.5.

5.2 Probabilidade de fontes dominantes na música

[00358] Para evitar a música com fontes dominantes de ser aplicada um nível elevado de equalização, o nível de equalização pode ser ainda mais correlacionado com o valor de confiança $Conf_{dom}$ indicando se uma peça de música contém uma fonte dominante, por exemplo,

$$L_{eq} = 1 - Conf_{dom} \qquad (23)$$

[00359] Desta forma, o nível de equalização será baixo em peças de música com fontes dominantes e alto em peças de música sem fontes dominantes.

[00360] Aqui, embora o valor de confiança da música com uma fonte dominante seja descrito, pode-se usar também o valor de confiança de música sem uma fonte dominante. Ou seja, unidade de ajuste 300D pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de música de curta duração sem fontes dominantes e/ou negativamente correlacionar o

nível de equalização com um valor de confiança de música de curta duração sem fontes dominantes.

[00361] Como indicado na seção 1.1, embora a música e fala por um lado, e música com ou sem fontes dominantes por outro lado, sejam tipos de conteúdo em diferentes níveis hierárquicos, eles podem ser considerados em paralelo. Em conjunto, considerando o valor de confiança das fontes dominantes e os valores de confiança de voz e música como descrito acima, o nível de equalização pode ser definido pela combinação de pelo menos uma das fórmulas (20)-(21) com (23). Um exemplo é a combinação de todas as três fórmulas:

$$L_{eq} = Conf_{music}(1 - Conf_{speech})(1 - Conf_{dom}) \tag{24}$$

[00362] Pesos diferentes com base na importância do tipo de conteúdo podem ser ainda aplicados para valores diferentes de confiança para generalidade, tais como à maneira da fórmula (22).

[00363] Como outro exemplo, suponha que $Conf_{dom}$ é calculado apenas quando o sinal de áudio é música, uma função em etapas pode ser projetada, como

$$L_{eq} = \begin{cases} (1 - Conf_{dom}) & Conf_{music} \geq limite\ 1 \\ Conf_{music}(1 - conf_{speech}^{\alpha}) & de\ outra\ forma \end{cases} \tag{25}$$

[00364] Esta função define a equalização de nível com base no valor de confiança das pontuações dominantes, se o sistema de classificação verificar o suficiente que o áudio é música (o valor de confiança de música é maior que um limiar); caso contrário, é definido com base nos valores de confiança de música e fala. Ou seja, a unidade de ajuste 300D pode ser configurada para considerar a música de curta duração sem/com fontes dominantes, quando o valor de confiança para a música de curta duração é maior que um limiar. Claro, o primeiro ou a segunda metade na fórmula (25) pode ser

modificada da maneira das fórmulas 20 a 24.

[00365] O mesmo esquema de unificação conforme discutido na seção 1.5 pode ser aplicado, e a constante de tempo a pode ser ainda definida com base no tipo de transição, como a transição da música com fontes dominantes para música sem fontes dominantes, ou a transição da música sem fontes dominantes para música com fontes dominantes. Para este efeito, uma fórmula similar à fórmula (4') também pode ser aplicada.

5.3 Predefinições de equalizador

[00366] Além de adaptativamente ajustar do nível de equalização com base em valores de confiança de tipos de conteúdo de áudio, perfis de equalização adequados ou predefinições desejadas espectrais podem também ser automaticamente escolhidas para conteúdo de áudio diferente, dependendo do seu gênero, instrumento ou outras características. A música com o mesmo gênero, contendo o mesmo instrumento, ou com as mesmas características musicais, pode compartilhar os mesmos perfis de equalização ou predefinições desejadas de equilíbrio espectral.

[00367] Por generalidade, foi usado o termo "agrupamentos de música" para representar grupos de música com o mesmo gênero, o mesmo instrumento ou atributos musicais semelhantes, e podem ser considerados como outro nível hierárquico de tipos de conteúdo de áudio, como indicado na seção 1.1. Perfil de equalização apropriado, nível de equalização, e/ou predefinição desejada de equilíbrio espectral, podem ser associados a cada agrupamento de música. O perfil de equalização é a curva de ganho aplicada sobre o sinal de música e pode ser qualquer uma das predefinições de equalizador usadas para gêneros musicais diferentes (como a música clássica, Rock, Jazz e Folk) e a predefinição desejada de equilíbrio espectral representa o timbre desejado para cada agrupamento. A Figura 23

ilustra vários exemplos de predefinições de equilíbrio espectral desejadas conforme implementadas em tecnologias Dolby Home Theater. Cada uma descreve a forma espectral desejada em toda a gama de frequências audíveis. Esta forma é continuamente comparada à forma espectral do áudio de entrada e ganhos de equalização são calculados a partir desta comparação para transformar a forma espectral do áudio de entrada naquela da predefinição.

[00368] Para uma peça de música nova, o agrupamento mais próximo pode ser determinado (decisão difícil), ou o valor de confiança no que diz respeito a cada agrupamento de música pode ser computado (decisão simples). Com base nesta informação, perfil de equalização apropriado, ou predefinição desejada de equilíbrio espectral, podem ser determinados para a peça de determinada música. A maneira mais simples é atribuir o perfil correspondente do melhor agrupamento correspondente, como

$$P_{eq} = P_{c^*} \qquad (26)$$

[00369] onde P_{eq} é o perfil de equalização estimado ou predefinição desejada de equilíbrio espectral, e c* é o índice do agrupamento de música com melhor correspondência (o tipo dominante de áudio), que pode ser obtido por escolher o agrupamento com o valor mais alto de confiança.

[00370] Além disso, pode haver mais de um agrupamento de música tendo valor de confiança maior que zero, ou seja, a peça de música tem mais ou menos semelhantes atributos como esses agrupamentos. Por exemplo, uma peça de música pode ter vários instrumentos, ou pode ter atributos de vários gêneros. Isso inspira outra maneira de estimar o perfil de equalização apropriado, considerando todos os agrupamentos, em vez de usar somente o

agrupamento mais próximo. Por exemplo, uma soma ponderada pode ser usada:

$$P_{eq} = \sum_{c=1}^{N} w_c P_c \qquad (27)$$

[00371] onde N é o número de agrupamentos predefinidos, e w_c é o peso do perfil projetado P_c sobre cada agrupamento de música predefinido (com índice de c), que deve ser normalizado a 1, com base em seus valores correspondentes de confiança. Desta forma, o perfil estimado seria uma mistura dos perfis de agrupamentos de música. Por exemplo, para uma peça de música tendo ambos os atributos de Jazz e Rock, o perfil estimado seria algo intermediário.

[00372] Em algumas aplicações, pode-se não querer envolver todos os agrupamentos, como mostrado na fórmula (27). Apenas um subconjunto dos agrupamentos - os agrupamentos mais relacionados com a peça de música atual - precisam ser considerados, a fórmula (27) pode ser ligeiramente revisada para:

$$P_{eq} = \sum_{c'=1}^{N'} w_{c'} P_{c'} \qquad (28)$$

[00373] onde o N' é o número de agrupamentos que devem ser considerados e c' é o índice de agrupamento após triagem decrescente dos agrupamentos com base em seus valores de confiança. Usando um subconjunto, pode-se focar mais os agrupamentos mais relacionados e excluir aqueles menos relevantes. Em outras palavras, a unidade de ajuste 300D pode ser configurada para considerar pelo menos um tipo de áudio dominante, com base nos valores de confiança.

[00374] Na descrição acima, agrupamentos de música são tomados como exemplo. Na verdade, as soluções são aplicáveis aos tipos de áudio em qualquer nível hierárquico, conforme discutido na seção 1.1. Portanto, em geral, a unidade de ajuste 300D pode ser configurada

para atribuir um nível de equalização e/ou perfil de equalização e/ou predefinição de equilíbrio espectral para cada tipo de áudio.

5.4 Controle com base no tipo de contexto

[00375] Nas seções anteriores, a discussão está focada em vários tipos de conteúdo. Em modalidades ainda a se discutir nesta seção, tipo de contexto pode ser em alternativa ou adicionalmente considerado.

[00376] Em geral, o equalizador é habilitado para música, mas desabilitado para conteúdo de mídia similar a filme, já que o equalizador pode fazer diálogos na mídia similar a filme não soarem tão bem devido à mudança de timbre óbvia. Isto indica que o nível de equalização pode estar relacionado com o valor de confiança da música de longa duração e/ou o valor de confiança da mídia similar a filme:

$$L_{eq} \propto Conf_{MUSIC} \quad (29)$$

Ou

$$L_{eq} \propto 1 - Conf_{MOVIE} \quad (30)$$

[00377] onde L_{eq} é o nível de equalização, $Conf_{MUSIC}$ e $Conf_{MOVIE}$ representam o valor de confiança de música de longa duração e mídia similar a filme.

[00378] Ou seja, unidade de ajuste 300D pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de música de longa duração, ou negativamente, correlacionar o nível de equalização com um valor de confiança de mídia similar a filme.

[00379] Ou seja, para um sinal de mídia similar a filme, o valor de confiança da mídia similar a filme é alto (ou confiança de música é baixa), e, portanto, o nível de equalização é baixo; por outro lado, para um sinal de música, o valor de confiança da mídia similar a filme será baixo (ou confiança de música é alta) e, portanto, o nível de

equalização é alto.

[00380] As soluções mostradas na fórmula (29) e (30) podem ser modificadas da mesma forma como fórmulas de 22 a 25 e/ou podem ser combinadas com qualquer uma das soluções mostradas nas fórmulas de 22 a 25.

[00381] Além disso ou, alternativamente, a unidade de ajuste 300D pode ser configurada para correlacionar negativamente o nível de equalização com um valor de confiança do jogo.

[00382] Em outra modalidade o valor de confiança pode ser usado para derivar um peso normalizado, conforme discutido na seção 1.4. Supondo que um nível esperado de equalização/perfil é predefinido para cada tipo de contexto (perfis de equalização são mostrados na seguinte tabela 2), pode ser também aplicada uma fórmula similar à fórmula (9).

Tabela 2:

	MÍDIA TIPO FILME	MÚSICA A LONGA DURAÇÃO	VOIP	JOGO
Perfil de equalização	Perfil 1	Perfil 2	Perfil 3	Perfil 4

[00383] Aqui, em alguns perfis, todos os ganhos podem ser definidos como zero, como uma forma de desativar o equalizador para aquele determinado tipo de contexto, tais como a mídia similar a filme e jogo.

5.5 Combinação das modalidades e cenários de aplicação

[00384] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser implementadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou

iguais.

[00385] Por exemplo, quaisquer duas ou mais das soluções descritas nas seções 5.1 até a 5.4 podem ser combinadas uma com a outra. E qualquer uma das combinações pode ainda ser combinada com qualquer modalidade descrita ou implícita nas partes 1-4 e as outras partes que serão descritas depois.

5.6 Método de controle do equalizador

[00386] Semelhante à parte 1, no processo de descrever o controlador de equalizador nas modalidades seguintes, são também aparentemente divulgados alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00387] Primeiramente, as modalidades do método de processamento de áudio, como discutido na Parte 1, podem ser usadas em um equalizador, cujo(s) parâmetro(s) é/são um dos alvos a ser ajustado pelo método de processamento de áudio. Deste ponto da vista, o método de processamento de áudio também é um método de controle de equalizador.

[00388] Nesta seção, apenas aqueles aspectos específicos para o controle do equalizador serão discutidos. Para aspectos gerais do método de controle pode ser feita referência à parte 1.

[00389] De acordo com modalidades, um método de controle de equalizador pode incluir identificar um tipo de áudio de um sinal de áudio em tempo real; e uma unidade de ajuste para ajustar um equalizador de uma maneira contínua com base no valor de confiança do tipo de áudio, conforme identificado.

[00390] Semelhante a outras partes do presente pedido, quando envolvendo vários tipos de áudio com valores correspondentes de confiança, a operação de ajuste 1104 pode ser configurada para considerar pelo menos alguns dos vários tipos de áudio através de

ponderação dos valores de confiança dos vários tipos de áudio baseado na importância dos vários tipos de áudio, ou através de ponderação dos efeitos dos vários tipos de áudio baseado em valores de confiança. Em especial, a operação de ajuste 1104 pode ser configurada para considerar pelo menos um tipo de áudio dominante com base nos valores de confiança.

[00391] Conforme descrito na parte 1, o valor do parâmetro ajustado pode ser unificado. Pode ser feita referência à seção 1.5 e 1.8, e descrição detalhada é omitida aqui.

[00392] O tipo de áudio pode ser qualquer tipo de conteúdo ou tipo de contexto ou ambos. Quando envolvendo o tipo de conteúdo, a operação de ajuste 1104 pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de música de curta duração, e/ou negativamente correlacionar o nível de equalização com um valor de confiança da fala. Adicional ou alternativamente, a operação de ajuste pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de fundo e/ou negativamente correlacionar o nível de equalização com um valor de confiança de ruído.

[00393] Quando envolvendo o tipo de contexto, a operação de ajuste 1104 pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança de música de longa duração, e/ou negativamente correlacionar o nível de equalização com um valor de confiança da mídia similar a filme e/ou jogo.

[00394] Para o tipo de conteúdo de música de curta duração, a operação de ajuste 1104 pode ser configurada para correlacionar positivamente um nível de equalização com um valor de confiança da música de curta duração sem fontes dominantes e/ou negativamente correlacionar o nível de equalização com um valor de confiança da música de curta duração com fontes dominantes. Isso pode ser feito

somente quando o valor de confiança para a música de curta duração é maior que um limiar.

[00395] Além do ajuste do nível de equalização, outros aspectos de um equalizador podem ser ajustados com base em valores de confiança dos tipos de áudio de um sinal de áudio. Por exemplo, a operação de ajuste 1104 pode ser configurada para atribuir um nível de equalização e/ou perfil de equalização e/ou predefinição de equilíbrio espectral para cada tipo de áudio.

[00396] Sobre as instâncias específicas dos tipos de áudio, pode ser feita referência à parte 1.

[00397] Semelhante para as modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações é prática, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações pode ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em outras partes desta divulgação.

[00398] Parte 6: Classificadores de áudio e métodos de classificação

[00399] Como indicado nas seções 1.1 e 1.2, os tipos de áudio discutidos no presente pedido, incluindo vários níveis hierárquicos de tipos de conteúdo e contexto, podem ser classificados ou identificados com qualquer esquema de classificação existente, incluindo métodos baseados em aprendizado de máquina. Nesta parte e na parte posterior, o presente pedido propõe alguns aspectos novos de classificadores e métodos de classificação dos tipos de contexto, conforme mencionado na parte anterior.

6.1 Classificador de contexto com base na classificação de tipos de

conteúdo

[00400] Como indicado nas partes anteriores, o classificador de áudio 200 é usado para identificar o tipo de conteúdo de um sinal de áudio e/ou identificar o tipo de contexto do sinal de áudio. Portanto, o classificador de áudio 200 pode compreender um classificador de conteúdo de áudio 202 e/ou um classificador de contexto de áudio 204. Quando da adoção de técnicas existentes para implementar o classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204, os dois classificadores podem ser independentes um do outro, embora eles possam compartilhar algumas características e assim possam compartilhar alguns esquemas para extrair os recursos.

[00401] Nesta parte e na parte 7 subsequentes, de acordo com o aspecto novo proposto no presente pedido, o classificador de contexto de áudio 204 pode fazer uso dos resultados do classificador de conteúdo de áudio 202, ou seja, o classificador de áudio 200 pode incluir: um classificador de conteúdo de áudio 202 para identificar o tipo de conteúdo de um sinal de áudio; e um classificador de contexto de áudio 204 para identificar o tipo de contexto do sinal de áudio baseado nos resultados do classificador de conteúdo de áudio 202. Assim, os resultados de classificação do classificador de conteúdo de áudio 202 podem ser utilizados por ambos o classificador de contexto de áudio 204 e a unidade de ajuste 300 (ou as unidades de ajuste 300A a 300D), conforme discutido na parte anterior. No entanto, embora não mostrado nos desenhos, o classificador de áudio 200 também pode conter dois classificadores de conteúdo de áudio 202 para serem usados respectivamente pela unidade de ajuste 300 e o classificador de contexto de áudio 204.

[00402] Além disso, conforme discutido na seção 1.2, especialmente quando da classificação dos vários tipos de áudio, o classificador de conteúdo áudio 202 ou o classificador de contexto de

áudio 204 podem compreender um grupo de classificadores cooperando uns com os outros, embora também seja possível ser implementado como um único classificador.

[00403] Como discutido na seção 1.1, o tipo de conteúdo é um tipo de tipo de áudio com relação a segmentos de áudio de curta duração, em geral, tendo um comprimento da ordem de vários a várias dezenas de quadros (tal como 1s) e o tipo de contexto é um tipo de tipo de áudio com relação segmentos de longa duração de áudio em geral, tendo um comprimento da ordem de vários a várias dezenas de segundos (por exemplo, 10s). Portanto, correspondente ao "tipo de conteúdo" e "tipo de contexto", são usados "curto" e "longa duração" respectivamente quando necessário. No entanto, como será discutido na parte posterior 7, embora o tipo de contexto indique a propriedade do sinal de áudio em uma escala de tempo relativamente longa, pode também ser identificado com base em recursos extraídos de segmentos de áudio de curta duração.

[00404] Ocupa-se agora das estruturas do classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204 com referência à Figura 24.

[00405] Como mostrado na Figura 24, o classificador de conteúdo de áudio 202 pode incluir um extrator de recurso de curta duração 2022 para extrair recursos de curta duração de segmentos de curta duração de áudio cada um compreendendo uma sequência de quadros de áudio; e um classificador de curta duração 2024 para classificar uma sequência de segmentos de curta duração em um segmento de áudio a longo em tipos de áudio de curta duração usando os recursos de curta duração respectivos. Tanto o extrator de recursos de curta duração 2022 e o classificador de curta duração 2024 podem ser implementados com técnicas existentes, mas também algumas modificações são propostas para o extrator de recurso de curta

duração 2022 na subsequente seção 6.3.

[00406] O classificador de curta duração 2024 pode ser configurado para classificar cada um da sequência de segmentos de curta duração em pelo menos um dos seguintes tipos de áudio de curta duração (tipos de conteúdo): fala, música de curta duração, som de fundo e ruído, que foram explicados na seção 1.1. Cada um dos tipos de conteúdo pode ser ainda classificado em tipos de conteúdo em nível hierárquico inferior, como discutido na seção 1.1, mas não se limitando.

[00407] Como conhecido na técnica, valores de confiança dos tipos classificados de áudio também podem ser obtidos pelo classificador de curta duração 2024. No presente pedido, ao mencionar a operação de qualquer classificador, entende-se que os valores de confiança são obtidos ao mesmo tempo, se necessário, estando ou não explicitamente registrados. Um exemplo de classificação de tipo de áudio pode ser encontrado em L. Lu, H.-J. Zhang, and S. Li, "Content-based Audio Classification and Segmentation by Using Support Vector Machines", ACM MultimediaSystems Journal 8 (6), pp. 482-492, março de 2003, que é incorporado aqui na sua totalidade por referência.

[00408] Por outro lado, como mostrado na Figura 24, o classificador de contexto de áudio 204 pode incluir um extrator de estatísticas 2042 para calcular as estatísticas dos resultados do classificador de curta duração com relação à sequência de segmentos de curta duração no segmento de áudio de longa duração, como recursos de longa duração; e um classificador de longa duração 2044, usando os recursos de longa duração, classificando o segmento de áudio de longa duração em tipos de áudio de longa duração. Da mesma forma, tanto o extrator de estatísticas 2042 e o classificador de longa duração 2044 podem ser implementados com técnicas existentes, mas também algumas modificações são propostas para o extrator de estatísticas

2042 na subsequente seção 6.2.

[00409] O classificador de longa duração 2044 pode ser configurado para classificar o segmento de áudio de longa duração em pelo menos um dos seguintes tipos de áudio de longa duração (tipos de contexto): mídia similar a filme, música de longa duração, jogo e VoIP, que foram explicados na seção 1.1. Em alternativa ou adicionalmente, o classificador de longa duração 2044 pode ser configurado para classificar o segmento de áudio de longa duração em VoIP ou não-VoIP, que foi explicado na seção 1.1. Em alternativa ou adicionalmente, o classificador de longa duração 2044 pode ser configurado para classificar o segmento de longa duração de áudio em áudio de alta qualidade ou áudio de baixa qualidade, que tem foi explicado na seção 1.1. Na prática, vários tipos de áudio de destino podem ser escolhidos e treinados com base nas necessidades do aplicativo/sistema.

[00410] Sobre o significado e a seleção de segmento de curta duração e segmento de longa duração (bem como quadro a ser discutido na seção 6.3), pode ser feita referência à seção 1.1.

6.2 Extração de características de longa duração

[00411] Como mostrado na Figura 24, em uma modalidade, só o extrator de estatísticas 2042 é usado para extrair recursos de longa duração dos resultados de classificador de curta duração 2024. Como características de longa duração, pelo menos uma das seguintes pode ser calculada através do extrator de estatísticas 2042: média e variância dos valores de confiança dos tipos de áudio de curta duração dos segmentos de curta duração no segmento de longa duração a ser classificado, a média e a variância ponderada mediante os graus de importância dos segmentos de curta duração, frequência de ocorrência de cada tipo de áudio de curta duração e frequência de transições entre diferentes tipos de áudio de curta duração no segmento de longa

duração a ser classificado.

[00412] Foi usada, na Figura 25 a média dos valores de confiança de fala e música de curta duração em cada segmento de curta duração (de comprimento 1s). Para comparação, os segmentos são extraídos de três diferentes contextos de áudio: mídia similar a filme (Figura 25(A)), música de longa duração (Figura 25(B)) e VoIP (Figura 25(C)). Pode-se observar que, para o contexto da mídia similar a filme, valores de alta confiança são ganhados para um tipo de fala ou para o tipo de música e alternados entre esses dois tipos de áudio com frequência. Por outro lado, o segmento de música de longa duração dá um estável e alto valor de confiança de música de curta duração e um valor de confiança da fala relativamente estável e baixo. Enquanto que o segmento de VoIP dá um estável e baixo valor de confiança de música de curta duração, mas dá valores de confiança flutuantes de fala devido às pausas durante a conversa de VoIP.

[00413] A variância dos valores de confiança para cada tipo de áudio também é uma característica importante para a classificação de diferentes contextos de áudio. • Figura 26 dá histogramas da variância dos valores de confiança de fala, música de curta duração, fundo e ruído em mídia similar a filme, música de longa duração e contextos de áudio de VoIP (a abscissa é a variância dos valores de confiança em um dataset, e a ordenada é o número de ocorrências de cada bin de variâncias de valores no conjunto de dados, que pode ser normalizado para indicar a probabilidade de ocorrência de cada bin dos valores de variância). Para mídia similar a filme, todas as variações de valor de confiança de fala, música de curta duração e fundo são relativamente elevadas e amplamente distribuídas, indicando que os valores de confiança desses tipos de áudio estão mudando intensamente; para a música de longa duração, todas as variações de valor de confiança de fala, música de curta duração, fundo e ruído são relativamente baixas

e estreitamente distribuídas, indicando que os valores de confiança desses tipos de áudio estão estáveis: valor de confiança da fala mantém-se constantemente baixo e o valor de confiança de música se mantém constantemente elevado; para VoIP, as variações de valor de confiança de música de curta duração são baixas e estreitamente distribuídas, enquanto que as da fala são relativamente amplamente distribuídas, que é devido a frequentes pausas durante conversas VoIP.

[00414] Sobre as ponderações utilizadas no cálculo da média ponderada e variância, são determinadas com base no grau de importância em cada segmento de curta duração. O grau importante de um segmento de curta duração pode ser medido por sua energia ou intensidade, que pode ser estimada com muitas técnicas existentes.

[00415] A frequência de ocorrência de cada tipo de áudio de curta duração no segmento de longa duração a ser classificado é a contagem de cada tipo de áudio para o qual os segmentos de curta duração no segmento de longa duração foram classificados, normalizada com o comprimento do segmento de longa duração.

[00416] A frequência de transições entre diferentes tipos áudio de curta duração no segmento de longa duração a ser classificado é a contagem de alterações do tipo de áudio entre segmentos adjacentes de curta duração no segmento de longa duração a se classificar, normalizada com o comprimento do segmento de longa duração.

[00417] Quando se discute a média e a variância dos valores de confiança, com referência à Figura 25, a frequência de ocorrência de cada tipo de áudio de curta duração e a frequência de transição entre os diferentes tipos de áudio de curta duração também são na verdade mencionados. Esses recursos também são altamente relevantes para a classificação de contexto de áudio. Por exemplo, a música de longa duração principalmente contém tipo de áudio de música de curta

duração, então tem alta frequência de ocorrência de música de curta duração, enquanto que VoIP principalmente contém fala e faz uma pausa para que tenha frequência alta de ocorrência de fala ou ruído. Outro exemplo, a mídia similar a filme transita entre diferentes tipos de áudio de curta duração mais frequentemente do que música de longa duração ou VoIP, então geralmente tem uma maior frequência de transição entre a música de curta duração, fala e fundo; VoIP geralmente transita entre a fala e o ruído mais frequentemente do que os outros, por isso geralmente tem uma maior frequência de transição entre a fala e ruído.

[00418] Geralmente, presume-se que os segmentos de longa duração são do mesmo comprimento no mesmo aplicativo/sistema. Se este é o caso, então a contagem da ocorrência de cada tipo de áudio de curta duração, e a contagem de transição entre diferentes tipos de áudio de curta duração no segmento de longa duração pode ser usada diretamente sem normalização. Se o comprimento do segmento de longa duração é variável, então a frequência de ocorrência e a frequência de transições, como mencionado acima, devem ser usadas. E as concretizações no presente pedido devem ser entendidas como abrangendo as duas situações.

[00419] Além disso, ou alternativamente, o classificador de áudio 200 (ou o classificador de contexto de áudio 204) pode incluir mais um extrator de recurso de longa duração 2046 (Figura 27) para extrair mais recursos de longa duração do segmento de áudio de longa duração baseado nas características de curta duração da sequência de segmentos de curta duração no segmento de áudio de longa duração. Em outras palavras, o extrator de recurso de longa duração 2046 não usa os resultados de classificação do classificador de curta duração 2024, mas diretamente usa os recursos de curta duração extraídos do extrator de recursos de curta duração 2022 para derivar

algumas características de longa duração a ser usadas por classificador de longa duração 2044. O extrator de recurso de longa duração 2046 e o extrator de estatísticas 2042 podem ser usados independentemente ou em conjunto, em outras palavras, o classificador de áudio 200 pode incluir o extrator de recurso de longa duração 2046 ou o extrator de estatísticas 2042 ou ambos.

[00420] Todas as características podem ser extraídas pelo extrator de recurso de longa duração 2046. No pedido presente, propõe-se a calcular, como os recursos de longa duração, pelo menos uma das seguintes estatísticas dos recursos de curta duração do extrator de recurso de curta duração 2022: média, variância, média ponderada, variância ponderada, média alta, média baixa e razão (contraste) entre a média alta e média baixa.

[00421] Média e variância dos recursos de curta duração extraídos de segmentos de curta duração no segmento de longa duração a ser classificado;

[00422] Média ponderada e variação das características de curta duração extraídas de segmentos de curta duração no segmento de longa duração a ser classificado. Os recursos de curta duração são ponderados com base no grau de importância de cada segmento de curta duração que é medido com sua energia ou sonoridade como mencionado agora;

[00423] Média alta: uma média de recursos de curta duração selecionados extraídos de segmentos de curta duração no segmento de longa duração a ser classificado. Os recursos de curta duração são selecionados quando encontram pelo menos uma das seguintes condições: acima de um limiar; ou dentro de uma proporção predeterminada de recursos de curta duração não inferiores a todos os outros recursos de curta duração, por exemplo, os mais altos 10% dos recursos de curta duração;

[00424] Média baixa: uma média de recursos de curta duração selecionados extraídos de segmentos de curta duração no segmento de longa duração a ser classificado. Os recursos de curta duração são selecionados quando satisfazem pelo menos uma das seguintes condições: abaixo de um limiar; ou dentro de uma proporção predeterminada de recursos de curta duração não superior a todos os outros recursos de curta duração, por exemplo, os mais baixos 10% dos recursos de curta duração; e

[00425] Contraste: uma razão de contraste entre a média alta e a média baixa para representar a dinâmica dos recursos de curta duração em um segmento de longa duração.

[00426] O extrator de recursos de curta duração 2022 pode ser implementado com técnicas existentes, e todos os recursos podem ser extraídos desse modo. No entanto, algumas modificações são propostas para o extrator de recurso de curta duração 2022 na subsequente seção 6.3.

6.3 Extração de características de curta duração

[00427] Como mostrado na Figura 24 e Figura 27, o extrator de recursos de curta duração 2022 pode ser configurado para extrair, como recursos de curta duração pelo menos uma das seguintes características diretamente de cada segmento de áudio de curta duração: características rítmicas, características de interrupções/mutismo e características de qualidade de áudio de curta duração.

[00428] As características rítmicas podem incluir a força do ritmo, regularidade do ritmo, clareza de ritmo (ver L. Lu, D. Liu e H.-J. Zhang. "Automatic mood detection and tracking of music audio signals". IEEE Transactions on Audio, Speech, and Language Processing, 14(1):5 - 18, 2006, que é incorporado aqui na sua totalidade por referência) e modulação de sub-banda 2D (M.F McKinney e J. Breebaart. "Features

for audio and music classification", Proc. ISMIR, 2003, que é incorporado aqui na sua totalidade por referência).

[00429] As características de interrupções/mutismo podem incluir interrupções da fala, declínios fortes, comprimento de mutismo, silêncio não natural, média de silêncio não natural, energia total do silêncio não natural etc.

[00430] Os recursos de qualidade de áudio de curta duração são recursos de qualidade de áudio com relação a segmentos de curta duração, que são semelhantes aos recursos de qualidade de áudio extraídos de quadros de áudio, que são discutidos abaixo.

[00431] Em alternativa ou adicionalmente, como mostrado em Figura 28, o classificador de áudio 200 pode incluir um extrator de recurso de nível de quadro 2012 para extrair recursos de nível de quadro de cada um da sequência de quadros de áudio compreendida em um segmento de curta duração, e o extrator de recurso de curta duração 2022 pode ser configurado para calcular recursos de curta duração, baseados nas características do nível de quadro extraído da sequência de quadros de áudio.

[00432] Como pré-processamento, o sinal de áudio de entrada pode ser misturado para baixo a um sinal de áudio mono. O pré-processamento é desnecessário se o sinal de áudio já é um sinal mono. É então dividido em quadros com um comprimento predefinido (tipicamente de 10 a 25 milissegundos). Correspondentemente, recursos de nível de quadro são extraídos de cada quadro.

[00433] O extrator de recurso de nível de quadro 2012 pode ser configurado para extrair pelo menos um dos seguintes recursos: recursos caracterizando propriedades de vários tipos de áudio de curta duração, frequência de corte, características de razão sinal-ruído estática (SNR), características da razão sinal-ruído segmentar (SNR), descritores de fala básicos e características do trato vocal.

[00434] Os recursos que caracterizam as propriedades de vários tipos de áudio de curta duração (especialmente fala, música de curta duração, som de fundo e ruído) podem incluir pelo menos um dos seguintes recursos: energia de quadro, distribuição espectral de sub-banda, fluxo espectral, Mel-frequency Cepstral Coefficient (MFCC), baixo, informação residual, característica de Chroma e taxa de cruzamento zero.

[00435] Para detalhes de MFCC, referência pode ser feita a L. Lu, H.-J. Zhang, and S. Li, "Content-based Audio Classification and Segmentation by Using Support Vector Machines", ACM MultimediaSystems Journal 8 (6), pp. 482-492, março de 2003, que é incorporado aqui na sua totalidade por referência. Para o detalhe da característica de Chroma, pode ser feita referência a G. H. Wakefield, "Mathematical representation of joint time Chroma distributions" in SPIE, 1999, que é incorporado aqui na sua totalidade por referência.

[00436] A frequência de corte representa a mais alta frequência de um sinal de áudio acima da qual a energia do conteúdo é próxima de zero. É projetada para detectar conteúdo limitado por banda, que é útil neste pedido para classificação de contexto de áudio. Uma frequência de corte é geralmente causada por codificação, uma vez que a maioria dos programadores descarta altas frequências em bitrates médios ou baixos. Por exemplo, o codec de MP3 tem uma frequência de corte de 16kHz em 128kbps; outro exemplo, muitos codecs populares de VoIP têm uma frequência de corte de 8kHz ou 16kHz.

[00437] Além da frequência de corte, degradação do sinal durante o processo de codificação de áudio é considerada como outra característica para diferenciar os vários contextos de áudio tais como VoIP versus não-VoIP, contextos de alta qualidade vs. de áudio de baixa qualidade. Os recursos que representam a qualidade de áudio, tais como aqueles para avaliação objetiva da qualidade de fala (ver

Ludovic Malfait, Jens Berger, and Martin Kastner, "P.563- The ITU-T Standard for Single-Ended Speech Quality Assessment", IEEE Transaction on Audio, Speech, and Language Processing, VOL. 14, Nº 6, novembro de 2006, que é incorporado aqui na sua totalidade por referência), pode ser ainda mais extraídos em vários níveis para capturar as características mais ricas. Exemplos de recursos da qualidade de áudio:

[00438] a) Características de SNR estáticas incluindo nível estimado de ruído de fundo, clareza espectral, etc.

[00439] b) Características de SNR segmentares incluindo desvio espectral de nível, faixa espectral de nível, piso de ruído relativo, etc.

[00440] c) Descritores de fala básicos incluindo variação de nível de seção de fala, nível da fala, etc.

[00441] d) Características do trato vocal incluindo robotização, potência de cruzamento de tom, etc.

[00442] Para derivar os recursos de curta duração dos recursos de nível de quadro, o extrator de recurso de curta duração 2022 pode ser configurado para calcular estatísticas dos recursos de nível de quadro, como os recursos de curta duração.

[00443] Exemplos das estatísticas dos recursos de nível de quadro incluem a média e o desvio-padrão, que capta as propriedades rítmicas para diferenciar os vários tipos de áudio, tais como a música de curta duração, fundo, fala e ruído. Por exemplo, fala normalmente alterna entre sons sonoros e surdos a uma taxa de sílaba, enquanto que a música não o faz, indicando que a variação das características da fala de nível de quadro é geralmente maior do que a da música.

[00444] Outro exemplo das estatísticas é a média ponderada dos recursos de nível de quadro. Por exemplo, para a frequência de corte, a média ponderada entre as frequências de corte derivadas de todos os quadros de áudio em um segmento de curta duração com a energia

ou a altura de cada quadro como peso, seria a frequência de corte para esse segmento de curta duração.

[00445] Em alternativa ou adicionalmente, como mostrado na Figura 29, o classificador de áudio 200 pode incluir um extrator de recurso de nível de quadro 2012 para extrair recursos de nível de quadro de quadros de áudio, e um classificador de nível de quadro 2014 para a classificação de cada um da sequência de quadros de áudio para tipos de áudio de nível de quadro usando os respectivos recursos de nível de quadro, em que o extrator de recursos de curta duração 2022 pode ser configurado para calcular as características de curta duração com base nos resultados do classificador de nível de quadro 2014 no que diz respeito à sequência de quadros de áudio.

[00446] Em outras palavras, além do classificador de conteúdo de áudio 202 e o classificador de contexto de áudio 204, o classificador de áudio 200 pode incluir ainda um classificador de quadro 201. Em uma tal arquitetura, o classificador de conteúdo de áudio 202 classifica um segmento de curta duração baseado nos resultados de classificação de nível de quadro do classificador de quadro 201 e o classificador de contexto de áudio 204 classifica um segmento de longa duração baseado nos resultados de curta duração de classificação do classificador de conteúdo de áudio 202.

[00447] O classificador de nível de quadro 2014 pode ser configurado para classificar cada um da sequência de quadros de áudio em qualquer classe, que podem ser referidas como "tipos de áudio de nível de quadro". Em uma modalidade, os tipos de áudio de nível de quadro podem ter uma arquitetura semelhante à arquitetura dos tipos de conteúdo discutidos aqui, e têm também significado semelhante para os tipos de conteúdo e a única diferença é que os tipos de áudio de nível de quadro e os tipos de conteúdo são classificados em diferentes níveis do sinal de áudio, que é de nível de

quadro e de nível de curta duração. Por exemplo, o classificador de nível de quadro 2014 pode ser configurado para classificar cada um da sequência de quadros de áudio em pelo menos um dos seguintes tipos de áudio de nível de quadro: fala, música, som de fundo e ruído. Por outro lado, os tipos de áudio de nível de quadro também podem ter uma arquitetura completamente ou parcialmente diferente da arquitetura dos tipos de conteúdo, mais adequados para a classificação de nível de quadro e mais adequados para uso como os recursos de curta duração para a classificação de curta duração. Por exemplo, o classificador de nível de quadro 2014 pode ser configurado para classificar cada um da sequência de quadros de áudio em pelo menos um dos seguintes tipos de áudio de nível de quadro: sonoro, surdo e pausa.

[00448] Sobre como derivar recursos de curta duração dos resultados da classificação de nível de quadro, um esquema similar pode ser adotado referindo-se à descrição na seção 6.2.

[00449] Como uma alternativa, tanto recursos de curta duração baseados nos resultados do classificador de nível de quadro 2014 quanto recursos de curta duração diretamente baseados nos recursos de nível de quadro obtidas do extrator de recurso de nível de quadro 2012 podem ser utilizados pelo classificador de curta duração 2024. Portanto, o extrator de recursos de curta duração 2022 pode ser configurado para calcular as características de curta duração com base em ambos os recursos de nível de quadro extraídos da sequência de quadros de áudio e os resultados do classificador de nível de quadro com relação à sequência de quadros de áudio.

[00450] Em outras palavras, o extrator de recurso de nível de quadro 2012 pode ser configurado para calcular estatísticas semelhantes àquelas discutidas na seção 6.2 e esses recursos de curta duração descritos no âmbito da Figura 28, incluindo pelo menos

um dos seguintes recursos: recursos caracterizando propriedades de vários tipos de áudio de curta duração, frequência de corte, características de razão sinal-ruído estáticas, características segmentares da razão sinal-ruído, descritores de fala básicos e características do trato vocal.

[00451] Para trabalhar em tempo real, em todas as modalidades o extrator de recurso de curta duração 2022 pode ser configurado para trabalhar nos segmentos de áudio de curta duração formados com uma janela em movimento deslizante na dimensão temporal do segmento de áudio de longa duração em um comprimento de passo predeterminado. Sobre a janela em movimento para o segmento de áudio de curta duração, bem como o quadro de áudio e a janela em movimento para o segmento de áudio de longa duração, pode ser feita referência à seção 1.1 para detalhes.

6.4 Combinação das modalidades e cenários de aplicação

[00452] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser implementadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou iguais.

[00453] Por exemplo, quaisquer duas ou mais das soluções descritas nas seções 6.1 até a 6.3 podem ser combinadas uma com a outra. E qualquer uma das combinações pode ainda ser combinada com qualquer modalidade descrita ou implícita nas partes 1-5 e as outras partes que serão descritas depois. Especialmente, a unidade de uniformização de tipo 712 discutida na parte 1 pode ser usada nesta parte como um componente do classificador de áudio 200, para unificar os resultados do classificador de quadro 2014, o classificador de conteúdo de áudio 202 ou o classificador de contexto de áudio 204.

Além disso, o temporizador 916 também pode servir como um componente do classificador de áudio 200 para evitar a mudança abrupta da saída do classificador de áudio 200.

6.5 Método de classificação áudio

[00454] Semelhante à Parte 1, no processo de descrever o classificador de áudio nas modalidades seguintes, são também aparentemente divulgados alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00455] Em uma modalidade, conforme mostrado na Figura 30, um método de classificação de áudio é fornecido. Para identificar o tipo de áudio de longa duração (que é o tipo de contexto) de um segmento de áudio de longa duração, composto por uma sequência de segmentos de áudio de curta duração (sobrepostos ou não-sobrepostos uns com os outros), os segmentos de áudio de curta duração são em primeiro lugar classificados (operação 3004) em tipos de áudio de curta duração, ou seja, tipos de conteúdo, e recursos de longa duração são obtidos pelo cálculo (operação 3006) das estatísticas dos resultados da operação de classificação com relação à sequência de segmentos de curta duração no segmento de áudio de longa duração. Então a classificação de longa duração (operação 3008) pode ser realizada usando os recursos de longa duração. O segmento de áudio de curta duração pode incluir uma sequência de quadros de áudio. Claro, para identificar o tipo de áudio de curta duração dos segmentos de curta duração, recursos de curta duração precisam ser extraídos deles (operação 3002).

[00456] Os tipos de áudio de curta duração (tipos de conteúdo) podem incluir, mas não estão limitados à fala, música de curta duração, som de fundo e ruído.

[00457] Os recursos de longa duração podem incluir, mas não

estão limitados a: média e a variância dos valores de confiança de tipos de áudio de curta duração, a média e a variância ponderada mediante os graus de importância dos segmentos de curta duração, frequência de ocorrência de cada tipo de áudio de curta duração e frequência de transições entre diferentes tipos de áudio de curta duração.

[00458] Em uma variante, conforme mostrado na Figura 31, mais recursos de longa duração podem ser obtidos (operação 3107) diretamente baseado nas características de curta duração da sequência de segmentos de curta duração no segmento de áudio de longa duração. Tais características de longa duração ainda podem incluir podem incluir, mas não estão limitadas às seguintes estatísticas dos recursos de curta duração: média, variância, média ponderada, variância ponderada, média alta, média baixa e relação entre a média alta e média baixa.

[00459] Há diferentes maneiras para extrair os recursos de curta duração. Uma é extrair diretamente os recursos de curta duração do segmento de áudio de curta duração a ser classificado. Tais recursos incluem, mas não estão limitados às características rítmicas, características de interrupções/mutismo e características de qualidade de áudio de curta duração.

[00460] A segunda maneira é extrair recursos de nível de quadro dos quadros de áudio compreendidos em cada segmento de curta duração (operação 3201 na Figura 32), e em seguida calcular recursos de curta duração baseados nas características do nível de quadro, tais como estatísticas calculadas dos recursos de nível de quadro como os recursos de curta duração. Os recursos de nível de quadro podem compreender, mas sem limitação: recursos caracterizando propriedades de vários tipos de áudio de curta duração, frequência de corte, características de razão sinal-ruído estática (SNR),

características da razão sinal-ruído segmentar (SNR), descritores de fala básicos e características do trato vocal. As características que caracterizam as propriedades de vários tipos de áudio de curta duração podem incluir ainda energia de quadro, distribuição espectral de sub-banda, fluxo espectral, Mel-frequency Cepstral Coefficient, baixo, informação residual, característica de Chroma e taxa de cruzamento zero.

[00461] A terceira maneira é extrair os recursos de curta duração em uma maneira semelhante à extração dos recursos de longa duração: depois de extrair os recursos de nível do frame dos frames de áudio em um segmento de curta duração a ser classificado (operação 3201), classificando cada frame de áudio em tipos de áudio de nível do frame usando os recursos de nível do frame respectivo (operação 32011 na Figura 33); e os recursos de curta duração podem ser extraídos (operação 3002) calculando os recursos de curta duração, dependendo do tipo de áudio do nível do frame (opcionalmente incluindo os valores de confiança). Os tipos do nível do frame de áudio podem ter propriedades e uma arquitetura semelhante ao tipo de áudio de curta duração (tipo de conteúdo) e pode incluir também fala, música, som de fundo e ruído.

[00462] O segundo caminho e o terceiro caminho podem ser combinados juntos, como indicado pela seta tracejada na Figura 33.

[00463] Como discutido na Parte 1, ambos segmentos de áudio de curta duração e segmentos de áudio de longa duração podem ser amostrados com janelas em movimento. Ou seja, a operação de extração de recursos de curta duração (operação 3002) pode ser executada no segmento de áudio de curta duração formado com uma janela em movimento deslizante na dimensão temporal, do segmento de áudio de longa duração em um comprimento da etapa predeterminada, e a operação de extração de recursos de longa

duração (operação 3107) e a operação de calcular estatísticas de tipos de áudio de curta duração (operação 3006) também podem ser executadas em segmentos de áudio de longa duração formados com uma janela em movimento deslizante na dimensão temporal do sinal de áudio em um comprimento da etapa predeterminada.

[00464] Semelhante às modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações são práticas, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações podem ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em outras partes desta divulgação. Especialmente, como já discutido na Seção 6.4, os esquemas de suavização e o regime de transição dos tipos de áudio podem ser uma parte do áudio classificando o método discutido aqui.

Parte 7: Classificadores de VoIP e métodos de classificação

[00465] Na parte 6 um classificador de áudio novo é proposto para classificar um sinal de áudio em tipos de contexto áudio, pelo menos parcialmente baseado nos resultados da classificação do tipo de conteúdo. Nas modalidades discutidas na Parte 6, recursos de longa duração são extraídos de um segmento de longa duração de comprimento de várias dezenas de segundos, assim, a classificação do contexto do áudio pode causar latência longa. Isto é desejado que o contexto de áudio pode também ser classificado em tempo real ou quase em tempo real, tal como no nível do segmento de curta duração.

Seção 7.1 Classificação de Contexto Baseado no Segmento de Curta Duração

[00466] Portanto, como mostrado na Figura 34, é fornecido um classificador áudio 200A, compreendendo um classificador de conteúdo do áudio 202A para identificar um tipo de conteúdo de um segmento de curta duração, de um sinal de áudio e um classificador de contexto do áudio 204A para identificar um tipo de contexto do segmento de curta duração, pelo menos em parte com base no tipo de conteúdo identificado pelo classificador de conteúdo do áudio.

[00467] Aqui o classificador de conteúdo do áudio 202A pode adotar as técnicas já mencionadas na Parte 6, mas também pode adotar diferentes técnicas como serão discutidas abaixo na Seção 7.2. Além disso, o classificador de contexto do áudio 204A pode adotar as técnicas já mencionadas na Parte 6, com a diferença que o classificador de contexto 204A pode diretamente usar os resultados do classificador de conteúdo do áudio 202A, em vez de usar as estatísticas dos resultados do classificador de conteúdo do áudio 202A desde tanto o classificador do contexto de áudio 204A e o classificador de conteúdo do áudio 202A estão classificando o mesmo segmento de curta duração. Além disso, semelhante à Parte 6, além dos resultados do classificador de conteúdo do áudio 202A, o classificador de contexto de áudio 204A pode usar outros recursos extraídos diretamente do segmento de curta duração. Isto é, o classificador de contexto do áudio 204A pode ser configurado para classificar o segmento de curta duração om base em um modelo de aprendizado maquínico usando, como recursos, os valores de confiança dos tipos de conteúdo do segmento de curta duração e outros recursos extraídos do segmento de curta duração. Sobre os recursos extraídos do segmento de curta duração, pode ser feita referência à Parte 6.

[00468] O classificador de conteúdo do áudio 200A simultaneamente pode rotular o segmento de curta duração como mais tipos de áudio de VoIP fala/ruído e/ou não VoIP fala/ruído (VoIP

fala/ruído e não VoIP fala/ruído será discutido abaixo na Seção 7.2) e cada um dos vários tipos de áudio podem ter seu próprio valor de confiança, conforme discutido na Seção 1.2. Isto pode conseguir melhor precisão de classificação desde que as informações mais sofisticadas possam ser capturadas. Por exemplo, informação conjunta dos valores de confiança de fala e música de curta duração revela até que ponto o conteúdo de áudio é provável que seja uma mistura de fala e música de fundo para que ele pode ser discriminado de puro conteúdo VoIP.

Seção 7.2 Classificação Usando Fala de VoIP e Ruído VoIP

[00469] Este aspecto do presente pedido é especialmente útil em um sistema de classificação de VoIP/não-VoIP, que seria necessário para classificar o segmento atual de curta duração para latência de curta decisão.

[00470] Para essa finalidade, como mostrado em Figura34, o classificador do áudio 200A é projetado especialmente para classificação de VoIP/não VoIP. Para a classificação de VoIP/não VoIP, um classificador da fala de VoIP 2026 e/ou um classificador de ruído VoIP são desenvolvidos para gerar resultados intermediários para a classificação final de VoIP/não VoIP robusto pelo classificador de contexto do áudio 204A.

[00471] Um segmento de curta duração de VoIP conteria fala de VoIP e ruído VoIP como alternativa. Observa-se que esta alta precisão pode ser alcançada para classificar um segmento de curta duração da fala na fala de VoIP ou fala não VoIP, mas não tanto para a classificação de um segmento de curta duração do ruído no ruído VoIP ou ruído não-VoIP. Assim, pode concluir-se que ele irá desfocar a discriminabilidade classificando diretamente o segmento de curta duração em VoIP (compreendendo a fala de VoIP e ruído VoIP mas com a fala de VoIP e ruído VoIP não especificamente identificado) e

não VoIP sem considerar a diferença entre a fala e o ruído e, portanto, com as características destes dois tipos de conteúdo (fala e ruído) misturadas.

[00472] É razoável para classificadores alcançar maior precisão para a classificação da fala VoIP/fala não VoIP do que para a classificação de ruído VoIP/ruído não-VoIP como fala contém mais informações do que o ruído e características tais como a frequência de corte são mais eficazes para a classificação da fala. De acordo com o ranking de peso obtido do processo de formação de adaBoost, os recursos de curta duração ponderados acima para a classificação da fala de VoIP/não VoIP são: desvio-padrão de energia de logaritmo, frequência de corte, desvio-padrão da força rítmica e desvio-padrão de fluxo espectral. O desvio-padrão de energia do logaritmo, desvio-padrão da força rítmica e desvio-padrão de fluxo espectral são geralmente mais elevados para fala de VoIP do que para fala não VoIP.

[00473] Um provável motivo é que muitos segmentos da fala de curta duração, em um contexto não VoIP como um filme como mídia ou um jogo geralmente são misturados com outros sons como efeito da música ou som de fundo, dos quais os valores das características acima são mais baixos. Enquanto isso, o recurso de corte é geralmente inferior para fala de VoIP do que para fala não VoIP, que indica a frequência de corte baixo introduzida pelos muitos codecs de VoIP populares

[00474] Portanto, em uma modalidade, o classificador de conteúdo do áudio 202A pode compreender um classificador da fala de VoIP 2026 para classificar o segmento de curta duração para o tipo de conteúdo da fala de VoIP ou fala não VoIP do tipo de conteúdo; e o classificador de contexto do áudio 204A pode ser configurado para classificar o segmento de curta duração para o tipo do contexto de

VoIP ou o tipo do contexto não VoIP com base na confiança da fala de VoIP e fala não VoIP.

[00475] Em outra modalidade, o classificador de conteúdo do áudio 202A ainda pode compreender um classificador de ruído VoIP 2028 para classificar o segmento de curta duração para o ruído VoIP do tipo de conteúdo ou o ruído do tipo de conteúdo de não-VoIP; e o classificador de contexto do áudio 204A pode ser configurado para classificar o segmento de curta duração para o tipo de contexto de VoIP ou o tipo de contexto non-VoIP com base na confiança da fala de VoIP, a fala não VoIP, ruído VoIP e ruído não-VoIP.

[00476] Os tipos de conteúdo da fala de VoIP, a fala de não-VoIP, ruído VoIP e ruído de não-VoIP podem ser identificados com técnicas existentes, conforme discutido na Parte 6, Seção 1.2 e Seção 7.1.

[00477] Como alternativa, o classificador de conteúdo do áudio 202A pode ter uma estrutura hierárquica, como mostrado na Figura 35. Ou seja, tirou-se proveito dos resultados de um classificador da fala/ruído 2025 para classificar primeiro o segmento de curta duração na fala ou ruído/de fundo.

[00478] Com base na modalidade usando VoIP meramente classificador da fala 2026, se um segmento de curta duração é determinado como a fala do classificador da fala/ruído 2025 (em tal situação é apenas um classificador da fala) e, em seguida, o classificador da fala de VoIP 2026 continua a classificar se é a fala de VoIP ou fala não VoIP e calcula o resultado de classificação binária; Caso contrário pode considerar-se que o valor de confiança da fala de VoIP é baixo, ou a decisão sobre a fala de VoIP é incerta.

[00479] Com base na modalidade usando um meramente classificador de ruído VoIP 2028, se o segmento de curta duração é determinado como o ruído, o classificador da fala/ruído 2025 (em tal situação é apenas um classificador de ruído (fundo)) e, em seguida, o

classificador de ruído VoIP 2028 continua a classificá-lo no ruído VoIP ou ruído de não-VoIP e calcular o resultado de classificação binária. Caso contrário pode considerar-se que o valor de confiança de ruído VoIP é baixo, ou a decisão sobre o ruído VoIP é incerta.

[00480] Aqui, uma vez que geralmente a fala é um tipo de conteúdo informativo e ruído/de fundo é um tipo de conteúdo interferente, mesmo se um segmento de curta duração não é um ruído, na modalidade no parágrafo anterior não determinarmos definitivamente que o segmento de curta duração não é o tipo de contexto VoIP. Enquanto se um segmento de curta duração não é uma intervenção, na modalidade apenas usando o classificador de fala de VoIP 2026, provavelmente não é o tipo de contexto de VoIP, portanto, geralmente a modalidade usando meramente um classificador da fala de VoIP 2026 pode ser realizada independentemente, enquanto a outra modalidade usando meramente um classificador de ruído VoIP 2028 pode ser usado como uma modalidade complementar cooperando com, por exemplo, a modalidade usando o classificador de fala de VoIP 2026.

[00481] Ou seja, podem ser utilizados tanto classificador da fala VoIP 2026 e classificador de ruído VoIP 2028. Se um segmento de curta duração é determinado como a fala pelo classificador da fala/ruído 2025, em seguida, o classificador da fala de VoIP 2026 continua classificando se é fala de VoIP ou fala não VoIP e calcula o resultado de classificação binária. Se o segmento de curta duração é determinado como ruído pelo classificador da fala/ruído 2025, em seguida, o classificador de ruído VoIP 2028 continua a classificá-lo no ruído VoIP ou ruído de não-VoIP e calcular o resultado de classificação binária. Caso contrário, ele pode ser considerado esse segmento de curta duração pode ser classificado como não-VoIP.

[00482] A implementação do classificador da fala/ruído 2025, o

classificador da fala de VoIP 2026 e o classificador do ruído VoIP 2028 pode adotar quaisquer técnicas existentes e pode ser o classificador de conteúdo do áudio 202 discutido nas Partes 1-6.

[00483] Se o classificador de conteúdo do áudio 202A implementado de acordo com a descrição acima finalmente classifica um segmento de curta duração em nenhuma fala, o ruído e o fundo, ou nenhuma fala de VoIP, fala de não-VoIP, ruído VoIP e ruído não-VoIP, ou seja, todos os valores de confiança relevantes são baixos, então o classificador de conteúdo do áudio 202A (e o classificador de contexto do áudio 204A) podem classificar o segmento de curta duração como não-VoIP.

[00484] Para classificar o segmento de curta duração para os tipos de contexto de VoIP ou não VoIP baseado nos resultados do classificador da fala de VoIP 2026 e o classificador de ruído VoIP 2028, o classificador de contexto do áudio 204A pode adotar técnicas baseadas no aprendizado da máquina, conforme discutido na Seção 7.1, e como uma modificação, mais recursos podem ser utilizados, incluindo recursos de curta duração diretamente extraídos do segmento de curta duração e/ou resultados de outros classificadores de conteúdo do áudio direcionado para outros tipos de conteúdo do que VoIP relacionados com tipos de conteúdo , como já discutido na Seção 7.1.

[00485] Além das técnicas descritas acima baseada no aprendizado de máquina, uma abordagem alternativa para classificação de VoIP/não VoIP pode ser uma regra heurística, aproveitando-se do conhecimento do domínio e utilizando os resultados de classificação relacionados com a fala de VoIP e ruído VoIP. Um exemplar de tais regras heurísticas será ilustrado abaixo.

[00486] Se o segmento atual de curta duração de tempo t é determinado como fala de VoIP ou fala não VoIP, o resultado de

classificação diretamente é tomado como o resultado de classificação de VoIP/não-VoIP desde a classificação da fala de VoIP/não VoIP é robusta conforme discutido antes. Ou seja, se o segmento de curta duração é determinado como fala de VoIP, então é o tipo de contexto VoIP; se o segmento de curta duração é determinado como a fala não VoIP, então é o tipo de contexto de non VoIP.

[00487] Quando o classificador da fala de VoIP 2026 faz uma decisão binária sobre fala de VoIP/fala não VoIP em relação à fala determinada pelo classificador da fala/ruído 2025 como mencionado acima, os valores de confiança da fala de VoIP e fala não VoIP pode ser complementar, ou seja, que a soma deste é 1 (se 0 representa 100% não e 1 representa 100% sim) e os limites do valor de confiança para diferenciar a fala de VoIP e fala não VoIP pode indicar na verdade o mesmo ponto. Se o classificador da fala de VoIP 2026 não é um classificador binário, os valores de confiança da fala de VoIP e fala não VoIP podem não ser complementares, e os limites de valor de confiança para diferenciar a fala de VoIP e a fala não VoIP podem não indicam necessariamente o mesmo ponto.

[00488] No entanto, no caso onde a confiança da fala de VoIP ou fala não VoIP está perto de e oscila em torno do limite, os resultados da classificação de VoIP/não VoIP são possíveis alternar com demasiada frequência. Para evitar tal flutuação, um esquema de tampão que pode ser fornecido: ambos os limites para a fala de VoIP e fala não VoIP podem ser definidos a maiores, mas que não é tão fácil mudar de tipo de conteúdo presente para outro tipo de conteúdo. Para facilidade da descrição, pode-se converter o valor de confiança para a fala de não-VoIP para o valor de confiança da fala de VoIP. Ou seja, se o valor de confiança é alto, o segmento de curta duração é considerado como mais próximo à fala de VoIP, e se o valor de confiança é baixo o segmento de curta duração é considerado como o

mais perto da fala de não-VoIP. Apesar de classificador não binário como descrito acima um valor de alta confiança da fala não VoIP não significa necessariamente um valor de baixa confiança da fala de VoIP, essa simplificação também pode refletir a essência da solução e as concretizações relevantes descritas com a linguagem dos classificadores binários devem ser entendidas como abrangendo as soluções equivalentes para classificadores não binários.

[00489] O esquema do tampão é mostrado na Figura 36. Existe um tampão entre os dois limites Th1 e Th2 (Th1 > = Th2). Quando o valor de confiança v(t) da fala de VoIP cai na área, a classificação de contexto não mudará, como mostrado pelas setas nas laterais esquerda e direita na Figura 36. Somente quando o valor de confiança v(t) é maior que o limite maior de Th1, será o segmento de curta duração a ser classificado como VoIP (como indicado pela seta na parte inferior na Figura 36); e somente quando o valor de confiança não é maior que o limite menor de Th2, será o segmento de curta duração a ser classificada como não-VoIP (como indicado pela seta na parte superior na Figura 36).

[00490] Se o classificador de ruído VoIP 2028 é usado em vez disso, a situação é semelhante. Para tornar a solução mais robusta, o classificador da fala de VoIP 2026 e o classificador de ruído VoIP 2028 podem ser utilizados conjuntamente. Então, classificar o segmento de contexto de áudio 204A pode ser configurado para: classificar o segmento de curta duração como tipo de contexto VoIP se o valor de confiança da fala VoIP for maior que um primeiro limite ou se o valor de confiança do ruído VoIP for maior que um terceiro limite; classificar o segmento de curta duração como tipo de contexto não VoIP, ou se o valor de confiança da fala de VoIP não for maior que um segundo limite, em que o segundo limite não maior que o primeiro limite; ou se o valor de confiança do ruído VoIP não for maior que um quarto limite,

em que o quarto limite não maior que o terceiro limite; caso contrário, o segmento de classificar de curta duração como o tipo de contexto para o último segmento de curta duração.

[00491] Aqui, o primeiro limite pode ser igual ao segundo limite, e o terceiro limite pode ser igual ao quarto limite, especialmente para mas não ser limitado ao classificador binário da fala de VoIP e classificador binário de ruído VoIP. No entanto, como geralmente o resultado da classificação do ruído VoIP não é tão robusto, que seria melhor se o terceiro e o quarto limites não são iguais uns aos outros, e ambos devem ser longe de 0,5 (0 indica alta confiança para ser o ruído não-VoIP e 1 indica alta confiança para ser ruído VoIP).

Seção 7.3 Flutuação de uniformização

[00492] Para evitar flutuação rápida, outra solução é uniformizar o valor de confiança como determinado pelo classificador de conteúdo do áudio. Portanto, como mostrado na Figura 37, um tipo de uniformização única 203A pode ser composto no classificador de áudio 200A. Para o valor de confiança de cada um dos 4 VoIP relacionados com tipos de conteúdo, como discutido antes, os esquemas de uniformização, discutidos na Seção 1.3, podem ser adotados.

[00493] Alternativamente, semelhante à Seção 7.2, fala de VoIP e fala não VoIP pode considerado um par tendo valores complementares de confiança; e ruído VoIP e ruído de não-VoIP também pode ser considerado um par tendo valores complementares de confiança. Em tal situação, apenas um em cada par precisa ser alisado e esquemas de uniformização, discutidos na Seção 1.3, podem ser adotados.

[00494] Tirar o valor de confiança da fala de VoIP, por exemplo, a fórmula (3) pode ser reescrita como:

[00495] $v(t) = \beta \cdot v(t - 1) + (1 - \beta) \cdot voipSpeechConf(t)$ (3") onde v(t) é o valor de confiança da fala de VoIP suavizado no tempo t, v(t-l) é o valor de confiança da fala de VoIP suavizado da última vez, e

voipSpeechConf é a confiança da fala de VoIP no atual tempo t antes da uniformização, a é um coeficiente de ponderação.

[00496] Em uma variante, se existe um classificador da fala/ruído 2025 como descrito acima, se o valor de confiança de expressão para um segmento curto é baixo, então o segmento de curta duração não pode ser classificado como robustamente fala de VoIP, e pode-se definir diretamente voipSpeechConf (t)=v(t-l) sem fazer o classificador da fala de VoIP 2026 realmente funcionar.

[00497] Como alternativa, na situação descrita acima, pode-se definir voipSpeechConf (t) = 0,5 (ou outro valor não superior a 0,5, como 0,4-0,5), indicando um caso incerto (aqui, confiança = 1 indica uma alta confiança que é VoIP e confiança = 0 indica uma alta confiança, que não é um VoIP).

[00498] Entretanto, de acordo com a variante, como na Figura 37, o classificador de conteúdo de áudio 200A compreende adicionalmente um classificador fala/ruído 2025 para identificar tipos de conteúdo de fala do segmento de curta duração, e a unidade do tipo de uniformização 203A pode ser configurada para estabelecer o valor de confiança da fala de VoIP para o presente segmento de curta duração antes da uniformização conforme um valor de confiança pré-determinado (tal como 0,5 ou outro valor, tal como 0,4-0,5) ou o valor de confiança uniformizado do último segmento de curta duração onde o valor de confiança apara o tipo de conteúdo de fala tal como classificado pelo classificador fala/ruído é menor que um quinto limite. Em tal situação, o classificador da fala de VoIP 2026 pode ou não trabalhar. Como alternativa a configuração do valor de confiança pode ser feita pelo classificador da fala de VoIP 2026, isto é equivalente à solução onde o trabalho é feito pelo tipo de uniformização de unidade 203 A, e a reclamação deve ser entendida como abrangendo as duas situações. Além disso, aqui foi usada a linguagem "o valor de

confiança para a fala do tipo de conteúdo como classificadas pelo classificador da fala de ruído é inferior a um quinto limite", mas o escopo de proteção não é limitado aos mesmos, e é equivalente à situação onde o segmento de curta duração é classificado em outros tipos de conteúdo da fala.

[00499] Para o valor de confiança do ruído VoIP, a situação é semelhante e a descrição detalhada é omitida aqui.

[00500] Para evitar flutuação rápida, contudo uma outra solução é suavizar o valor de confiança como determinado pelo classificador de contexto do áudio 204A, e os esquemas de suavização, discutidos na Seção 1.3, podem ser adoptados.

[00501] Para evitar flutuação rápida, ainda uma outra solução é adiar a transição do tipo de contexto entre VoIP e não VoIP, e pode ser usado o mesmo esquema como descrito na Seção 1.6. Conforme descrito na Seção 1.6, o temporizador 916 pode ser fora do classificador de áudio ou dentro do classificador de áudio como uma parte do mesmo. Portanto, como mostrado na Figura 38, o classificador de áudio 200A ainda pode compreender o temporizador 916. E o classificador de áudio está configurado para continuar para a saída do tipo de contexto presente até o comprimento do tempo duradouro de um novo tipo de contexto atingindo um sexto limite (tipo de contexto é uma instância do tipo de áudio). Referindo-se a Seção 1.6, a descrição detalhada pode ser omitida aqui.

[00502] Em adição ou em alternativa, como um outro esquema para atrasar a transição entre VoIP e não VoIP, o primeiro e/ou segundo limites como descrito antes, para a classificação de VoIP/não-VoIP pode ser diferente dependendo do tipo de contexto do último segmento de curta duração. Ou seja, o primeiro e/ou segundo limite torna-se maior quando o tipo de contexto do novo segmento de curta duração é diferente do tipo de contexto do último segmento de curta

duração, o tempo torna-se menor quando o tipo de contexto do novo segmento de curta duração é o mesmo que o tipo de contexto do último segmento de curta duração. Por esta maneira, o tipo de contexto tende a ser mantido com o tipo de contexto atual e, portanto, a flutuação abrupta do tipo de contexto pode ser suprimida em certa medida.

7.4 Combinação das modalidades e cenários de aplicação

[00503] Semelhante à parte 1, todas as modalidades e variantes respectivas discutidas acima podem ser implementadas em qualquer combinação respectiva e todos os componentes mencionados em diferentes partes/modalidades, mas tendo as funções iguais ou similares podem ser implementados como componentes separados ou iguais.

[00504] Por exemplo, quaisquer duas ou mais das soluções descritas nas Seções 7.1 até a 7.3 podem ser combinadas uma com a outra. E qualquer uma das combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em Partes 1-6. Especialmente, as modalidades discutidas nesta parte e qualquer combinação destas podem ser combinadas com as modalidades do aparelhos/método de processamento do áudio ou o método de controlar/controlador do nivelador de volume discutido na Parte 4.

Seção 7.5 Métodos de classificação de VoIP

[00505] Semelhante à Parte 1, no processo de descrever o classificador de áudio nas modalidades seguintes, são também aparentemente divulgados alguns processos ou métodos. Doravante, um resumo desses métodos é oferecido sem repetição dos detalhes já discutidos anteriormente.

[00506] Em uma modalidade como mostrado na Figura 39, um método de classificação de áudio inclui a identificação de um tipo de conteúdo de um segmento de curta duração de um sinal de áudio

(operação 4004), em seguida, identificar um tipo de contexto do segmento de curta duração, pelo menos em parte, com base no tipo de conteúdo como identificado (operação 4008).

[00507] Para identificar o tipo de contexto de um sinal de áudio dinamicamente e rápido, o método de classificação do áudio nesta parte é especialmente útil para identificar o tipo de contexto VoIP e não VoIP. Em tal situação, o segmento de curta duração pode ser classificado primeiramente no tipo de conteúdo da fala de VoIP ou fala não VoIP do tipo de conteúdo; e o tipo de contexto de identificação ou operação é configurado para classificar o segmento de curta duração para o tipo do contexto de VoIP ou o tipo do contexto não VoIP com base na confiança da fala de VoIP e fala não VoIP.

[00508] Alternativamente, o segmento de curta duração pode ser classificado primeiramente no tipo de conteúdo do ruído VoIP ou do tipo de conteúdo do tipo de ruído não-VoIP, e a operação de identificação do tipo de contexto pode ser configurado para classificar o segmento de curta duração no tipo de contexto VoIP, ou no tipo de contexto não VoIP com base em valores de confiança da fala VoIP, fala não VoIP, ruído VoIP e ruído não-VoIP.

[00509] A fala e o ruído, podem ser considerados em conjunto. Em tal situação, a operação de identificação do tipo de contexto pode ser configurada para classificar o segmento de curta duração no tipo de contexto VoIP, ou no tipo de contexto não VoIP com base em valores de confiança de fala VoIP, fala não-VoIP, ruído VoIP e ruído não-VoIP.

[00510] Para identificar o tipo de contexto do segmento de curta duração, um modelo de aprendizado maquínico pode ser usado, tendo ambos os valores de confiança dos tipos de conteúdo do segmento de curta duração e outros recursos extraídos do segmento de curta duração como os recursos.

[00511] A operação de identificar o tipo de contexto também pode

ser realizada com base em regras heurísticas. Quando somente uma fala de VoIP e fala não VoIP são envolvidas, a regra heurística é como esta: classificando o segmento como tipo de contexto VoIP se o valor de confiança da fala VoIP é maior do que um primeiro limite; classificando o segmento de curta duração como um tipo de contexto não VoIP se o valor de confiança da fala de VoIP não for maior que um segundo limite, em que o segundo limite não é maior do que o primeiro limite; de qualquer outro modo, classificando o segmento de curta duração como o tipo de contexto para o último segmento de curta duração.

[00512] A regra heurística para a situação onde estão envolvidos apenas ruído VoIP e ruído não-VoIP é semelhante.

[00513] Quando tanto a fala quanto o ruído estão envolvidos, a regra heurística é como esta: classificado o segmento como tipo de contexto VoIP se o valor de confiança da fala de VoIP for maior que um primeiro limite ou se o valor de confiança do ruído VoIP for maior que um terceiro limite; classificar o segmento de curta duração como tipo de contexto não VoIP, ou se o valor de confiança da fala de VoIP não for maior que um segundo limite, em que o segundo limite não maior que o primeiro limite; ou se o valor de confiança do ruído VoIP não for maior que um quarto limite, em que o quarto limite não maior que o terceiro limite; caso contrário, classificando de curta duração como o tipo de contexto para o último segmento de curta duração.

[00514] O esquema de uniformização discutido na Seção 1.3 e Seção 1.8 podem ser aprovados aqui e a descrição detalhada é omitida. Como uma modificação para o esquema de uniformização na Seção 1.3, antes da operação de uniformização 4106, o método pode compreender adicionalmente a identificação de um tipo de conteúdo da fala do segmento de curta duração (operação 40040 na Figura 40), em que o valor de confiança da fala de VoIP para o presente

segmento de curta duração antes da uniformização é estabelecido como valor de confiança pré-determinado ou o valor de confiança uniformizado do último segmento de curta duração (operação 40044 na Figura 40) onde o valor de confiança para o tipo de conteúdo da fala for inferior a um quinto limite ("N" na operação 40041).

[00515] Se caso contrário a operação de identificar a fala do tipo de conteúdo robustamente julga o segmento de curta duração como fala ("Y" na operação 40041), então o segmento de curta duração é mais classificado na fala de VoIP ou fala não VoIP (operação 40042), antes da operação de uniformização 4106.

[00516] Na verdade, mesmo sem usar o esquema de uniformização, o método também pode identificar a fala do tipo de conteúdo e/ou primeiro ruído, quando o segmento de curta duração é classificado como fala ou ruído, ainda mais a classificação sendo implementado para classificar o segmento de curta duração em uma fala de VoIP e fala não VoIP, ou um ruído VoIP e ruído não-VoIP. Em seguida, é feita a operação de identificar o tipo de contexto.

[00517] Conforme mencionado na Seção 1.6 e Seção 1.8, o esquema de transição discutido nele pode ser tomado como uma parte do áudio classificando o método descrito aqui, e o detalhe é omitido. Brevemente, o método pode compreender adicionalmente a medição do tempo de duração durante o qual a operação de identificar o tipo de contexto de áudio dá saída contínua ao mesmo tipo de contexto, em que o método de classificar o áudio é configurado para continuar a dar saída ao tipo de contexto presente até que o comprimento do tempo de duração de um novo tipo de contexto atinja um sexto limiar.

[00518] Semelhantemente, sextos limites diferentes podem ser estabelecidos para diferentes pares de transição de um tipo de contexto a outro tipo de contexto. Além disso, o sexto limite pode ser correlacionado negativamente com o valor de confiança do novo tipo

de contexto.

[00519] Como uma modificação do esquema de transição no método de classificação de áudio especialmente direcionado à classificação de VoIP/não VoIP, um ou mais do primeiro ao quarto limite para o segmento atual e a curta duração podem ser criados diferentes dependendo do tipo de contexto do último segmento de curta duração.

[00520] Semelhante para as modalidades do aparelho de processamento de áudio, qualquer combinação das modalidades do método de processamento de áudio e suas variações é prática, por um lado; e, por outro lado, cada aspecto das modalidades do método de processamento de áudio e suas variações pode ter soluções separadas. Além disso, qualquer duas ou mais soluções descritas nesta seção podem ser combinadas com as outras, e estas combinações podem ainda ser combinadas com qualquer modalidade descrita ou implícita em outras partes desta divulgação. Especificamente, o método de classificação do áudio descrito aqui pode ser utilizado no método de processamento do áudio descrito antes, especialmente o regulador de volume, método de controle.

[00521] Como discutido no início da Descrição Detalhada do presente pedido, a modalidade do pedido pode ser incorporada no hardware ou no software ou em ambos. A Figura 41 é um diagrama de blocos que ilustra um sistema exemplar para implementar os aspectos do presente pedido.

[00522] Na Figura 41, uma unidade de processamento central (CPU) 4201 executa vários processos em conformidade com um programa armazenado em uma memória somente leitura (ROM) 4202 ou um programa carregado de uma seção de armazenamento 4208 para uma memória de acesso aleatório (RAM) 4203. Na RAM 4203, dados necessários quando o CPU 4201 executa vários processos ou

semelhante também armazenados conforme necessário.

[00523] O CPU 4201, 4202 a ROM e a RAM 4203 estão ligados um ao outro através de um barreamento 4204. Uma interface de entrada/saída 4205 também está conectada para o barreamento 4204.

[00524] Os seguintes componentes estão conectados à interface de entrada/saída 4205: uma seção de entrada 4206 incluindo um teclado, um mouse ou semelhantes; uma seção de saída 4207, incluindo uma exibição como um tubo de raios catódicos (CRT), um display de cristal líquido (LCD), ou semelhantes e um alto-falante ou similares; a seção de armazenamento 4208 incluindo um disco rígido ou semelhantes; e uma seção de comunicação 4209 incluindo uma placa de interface de rede como um cartão de LAN, um modem ou algo parecido. A seção de comunicação 4209 executa um processo de comunicação através da rede como a internet.

[00525] Uma unidade 4210 também está conectado para a interface de entrada/saída 4205 como requerida. Uma mídia removível 4211, tal como um disco magnético, um disco óptico, um disco magneto-óptico, uma memória de semicondutora ou similares, é montado na unidade 4210 conforme necessário, para que um programa de computador leia é instalado na seção de armazenamento 4208 conforme necessário.

[00526] No caso onde os componentes acima descritos são implementados pelo software, o programa que constitui o software é instalado a partir da rede como a internet ou meio de armazenamento, tal como a mídia removível 4211.

[00527] Nota-se que a terminologia usada aqui é com a finalidade de descrever as modalidades particulares somente e não é pretendido limitar o pedido. Conforme usado neste documento, as formas singulares "um", "uma" e "o(a)" são destinadas a incluir as formas plurais também, a menos que o contexto indique claramente o contrário. Será ainda compreendido que os termos “compreende” e/ou

“compreendendo,” quando usado nesta especificação, especifica a presença de recursos, inteiros, etapas, operações, elementos e/ou componentes indicados, mas não impossibilita a presença ou adição de um ou mais outros recursos, inteiros, etapas, operações, elementos, componentes e/ou grupos destes.

[00528] As estruturas, os materiais, os atos, e os equivalentes correspondentes de todos os meios ou operações mais elementos da função nas concretizações são pretendidos incluir toda a estrutura, material, ou ato para executar a função em combinação com outros elementos reivindicados como reivindicados especificamente. A descrição do pedido atual foi apresentada para finalidades da ilustração e da descrição, mas não é pretendida ser exaustiva ou limitada ao pedido na forma divulgada. Muitas modificações e variações serão evidentes para aqueles versados na técnica sem abandonar o escopo e o sentido do pedido. A modalidade foi escolhida e descrita para melhor explicar os princípios do pedido e a aplicação prática, e permitir que outros versados na técnica entendam o pedido de várias modalidades com várias modificações como são adequados ao uso específico contemplado.

REIVINDICAÇÕES

1. Método de normalização de volume com base em um valor de volume alvo, o método **caracterizado pelo fato de que** compreende as etapas de:

determinar parâmetros de ganho dinâmico a serem aplicados a quadros de áudio de um sinal de áudio com base em características de curto prazo ou longo prazo do sinal de áudio, a determinação incluindo determinar um parâmetro de ganho dinâmico para um primeiro quadro de áudio com base em uma característica de curto prazo do sinal de áudio e o valor de volume alvo, e determinar um parâmetro de ganho dinâmico para um segundo quadro de áudio com base em uma característica de longo prazo do sinal de áudio e o valor de volume alvo;

dimensionar o primeiro quadro de áudio pelo parâmetro de ganho dinâmico para o primeiro quadro de áudio para modificar um volume do primeiro quadro de áudio; e

dimensionar o segundo quadro de áudio pelo parâmetro de ganho dinâmico para o segundo quadro de áudio para modificar um volume do segundo quadro de áudio,

em que a característica de longo prazo é determinada de forma diferente da característica de curto prazo.

2. Método de normalização de volume, de acordo com a reivindicação 1, **caracterizado pelo fato de que** um aprimoramento de diálogo é aplicado tendo um efeito de tornar o diálogo mais proeminente dentro de um contexto particular.

3. Método de normalização de volume, de acordo com a reivindicação 1, **caracterizado pelo fato de que** uma equalização de volume é aplicada para ter um efeito em um ou mais níveis de reprodução em um equilíbrio tonal.

4. Método de normalização de volume, de acordo com a

reivindicação 1, **caracterizado pelo fato de que** uma suavização de parâmetro é aplicada aos parâmetros de ganho dinâmico.

5. Aparelho de processamento de áudio configurado para normalizar o volume com base em um valor de volume alvo, **caracterizado pelo fato de que** compreende:

pelo menos um processador; e

pelo menos uma memória armazenando um método implementado por computador;

em que a pelo menos uma memória com o método implementado por computador é configurada com o pelo menos um processador para fazer com que o aparelho de processamento de áudio pelo menos:

determine parâmetros de ganho dinâmico a serem aplicados a quadros de áudio de um sinal de áudio com base em características de curto prazo ou longo prazo do sinal de áudio, a determinação incluindo a determinação de um parâmetro de ganho dinâmico para um primeiro quadro de áudio com base em uma característica de curto prazo do sinal de áudio e o valor de volume alvo, e determinar um parâmetro de ganho dinâmico para um segundo quadro de áudio com base em uma característica de longo prazo do sinal de áudio e o valor de volume alvo;

dimensione o primeiro quadro de áudio pelo parâmetro de ganho dinâmico para o primeiro quadro de áudio para modificar um volume do primeiro quadro de áudio; e

dimensione o segundo quadro de áudio pelo parâmetro de ganho dinâmico para o segundo quadro de áudio para modificar um volume do segundo quadro de áudio,

em que a característica de longo prazo é determinada de forma diferente da característica de curto prazo.

6. Aparelho, de acordo com a reivindicação 5, **caracterizado pelo fato de que** o aprimoramento de diálogo é

aplicado tendo um efeito de tornar o diálogo mais proeminente dentro de um contexto particular.

7. Aparelho, de acordo com a reivindicação 5, **caracterizado pelo fato de que** uma equalização de volume é aplicada para ter um efeito em um ou mais níveis de reprodução em um equilíbrio tonal.

8. Aparelho, de acordo com a reivindicação 5, **caracterizado pelo fato de que** uma suavização de parâmetro é aplicada aos parâmetros de ganho dinâmico.

9. Dispositivo de armazenamento de método implementado por computador legível por uma máquina, **caracterizado pelo fato de que** incorpora de forma tangível um método implementado por computador executável pela máquina para fazer com que etapas sejam realizadas, as etapas compreendendo:

determinar parâmetros de ganho dinâmico a serem aplicados a quadros de áudio de um sinal de áudio com base em características de curto prazo ou longo prazo do sinal de áudio, a determinação incluindo determinar um parâmetro de ganho dinâmico para um primeiro quadro de áudio com base em uma característica de curto prazo do sinal de áudio e o valor de volume alvo, e determinar um parâmetro de ganho dinâmico para um segundo quadro de áudio com base em uma característica de longo prazo do sinal de áudio e o valor de volume alvo;

dimensionar o primeiro quadro de áudio pelo parâmetro de ganho dinâmico para o primeiro quadro de áudio para modificar um volume do primeiro quadro de áudio; e

dimensionar o segundo quadro de áudio pelo parâmetro de ganho dinâmico para o segundo quadro de áudio para modificar um volume do segundo quadro de áudio,

em que a característica de longo prazo é determinada de forma diferente da característica de curto prazo.

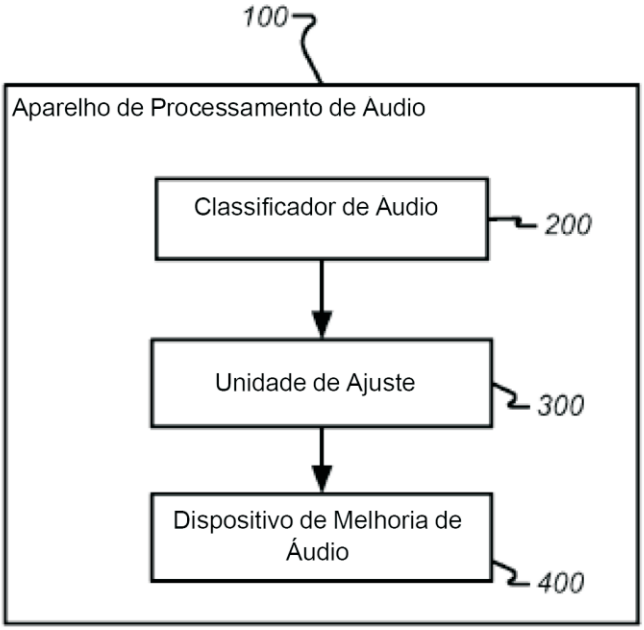
100
Aparelho de Processamento de Audio
Classificador de Audio
200
Unidade de Ajuste
300
Dispositivo de Melhoria de Áudio
400

FIG. 1

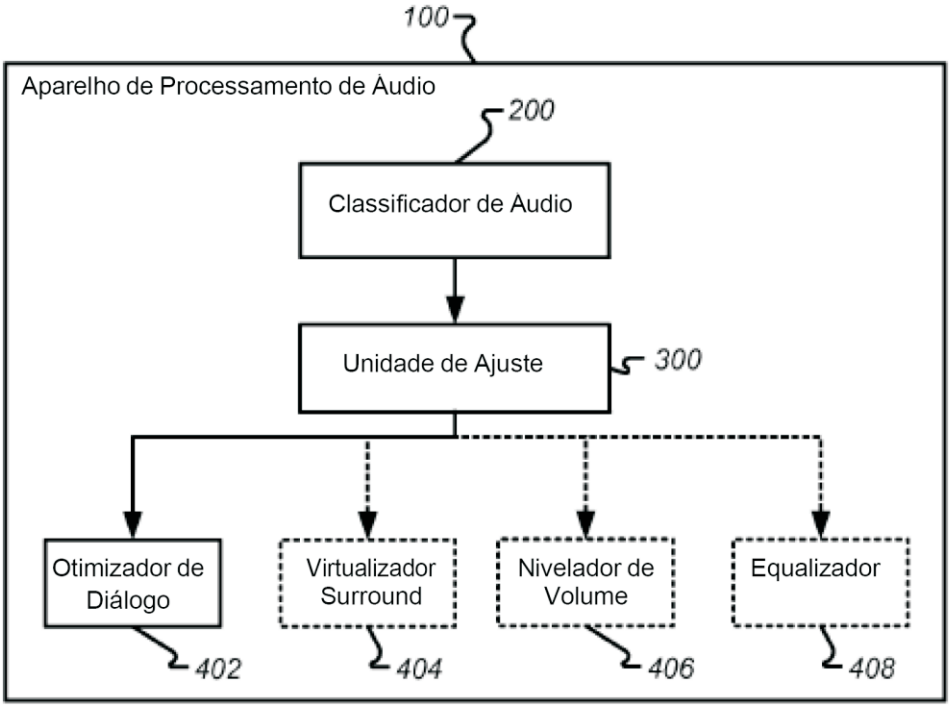
100
Aparelho de Processamento de Audio
200
Classificador de Audio
Unidade de Ajuste
300
Otimizador de Diálogo
402
Virtualizador Surround
404
Nivelador de Volume
406
Equalizador
408

FIG. 2

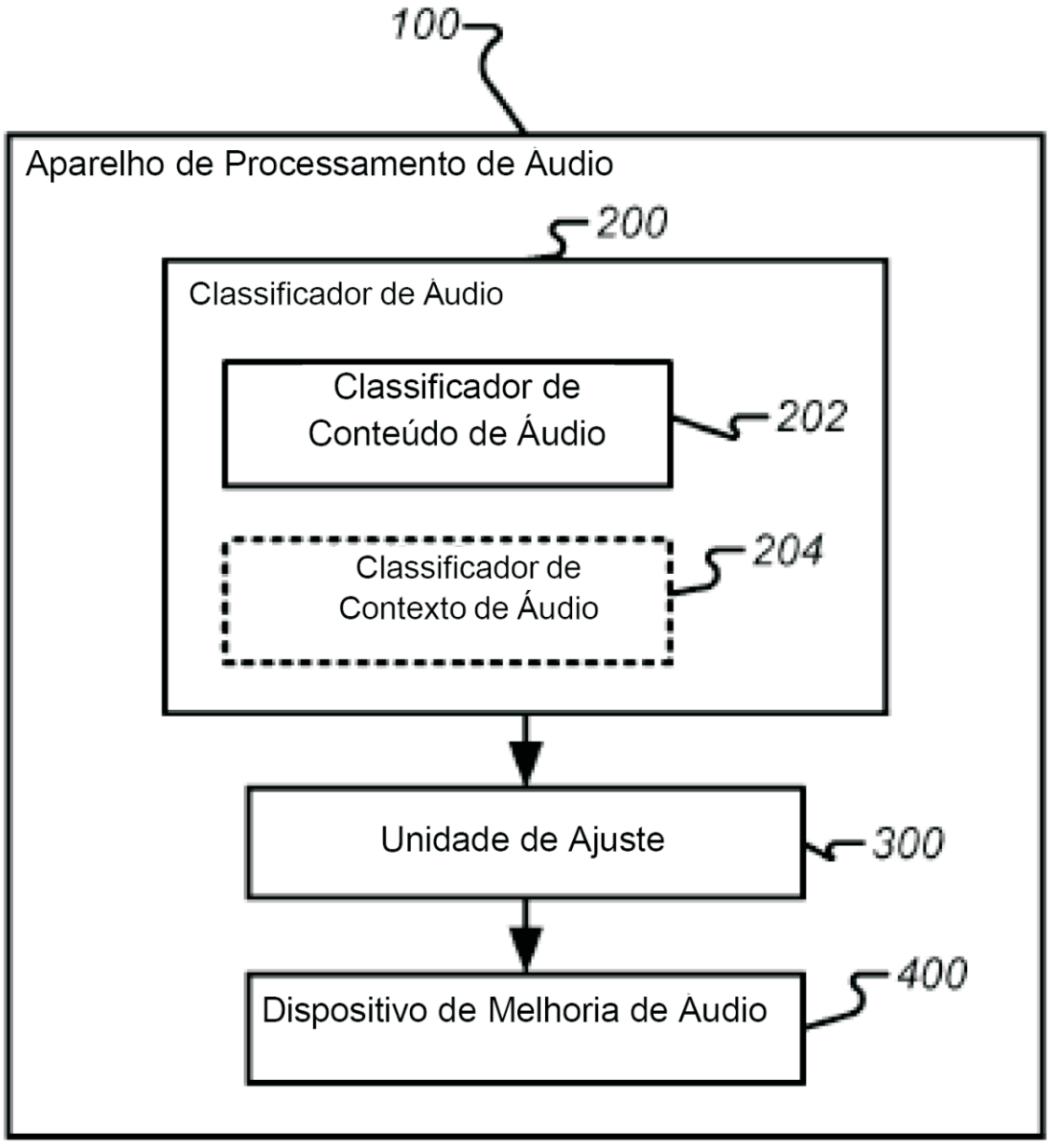
100
Aparelho de Processamento de Áudio
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
Unidade de Ajuste
300
Dispositivo de Melhoria de Áudio
400

FIG. 3

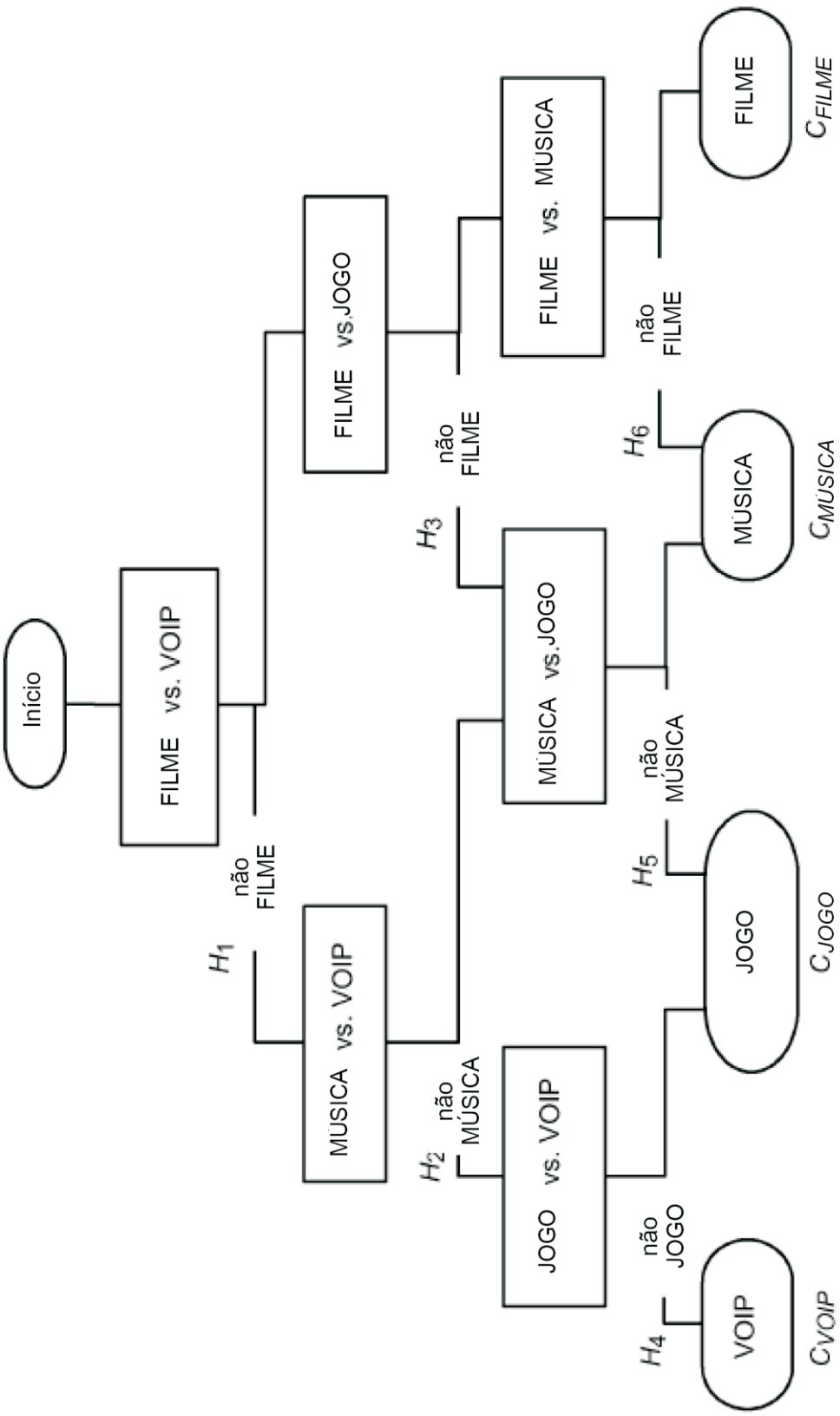
Início
FILME vs. VOIP
não FILME
H_1
MÚSICA vs. VOIP
não MÚSICA
H_2
JOGO vs. VOIP
não JOGO
H_4
VOIP
C_{VOIP}
H_5
JOGO
C_{JOGO}
não MÚSICA
MÚSICA vs.JOGO
H_3
não FILME
FILME vs.JOGO
H_6
MÚSICA
$C_{MÚSICA}$
não FILME
FILME vs. MÚSICA
FILME
C_{FILME}

FIG. 4

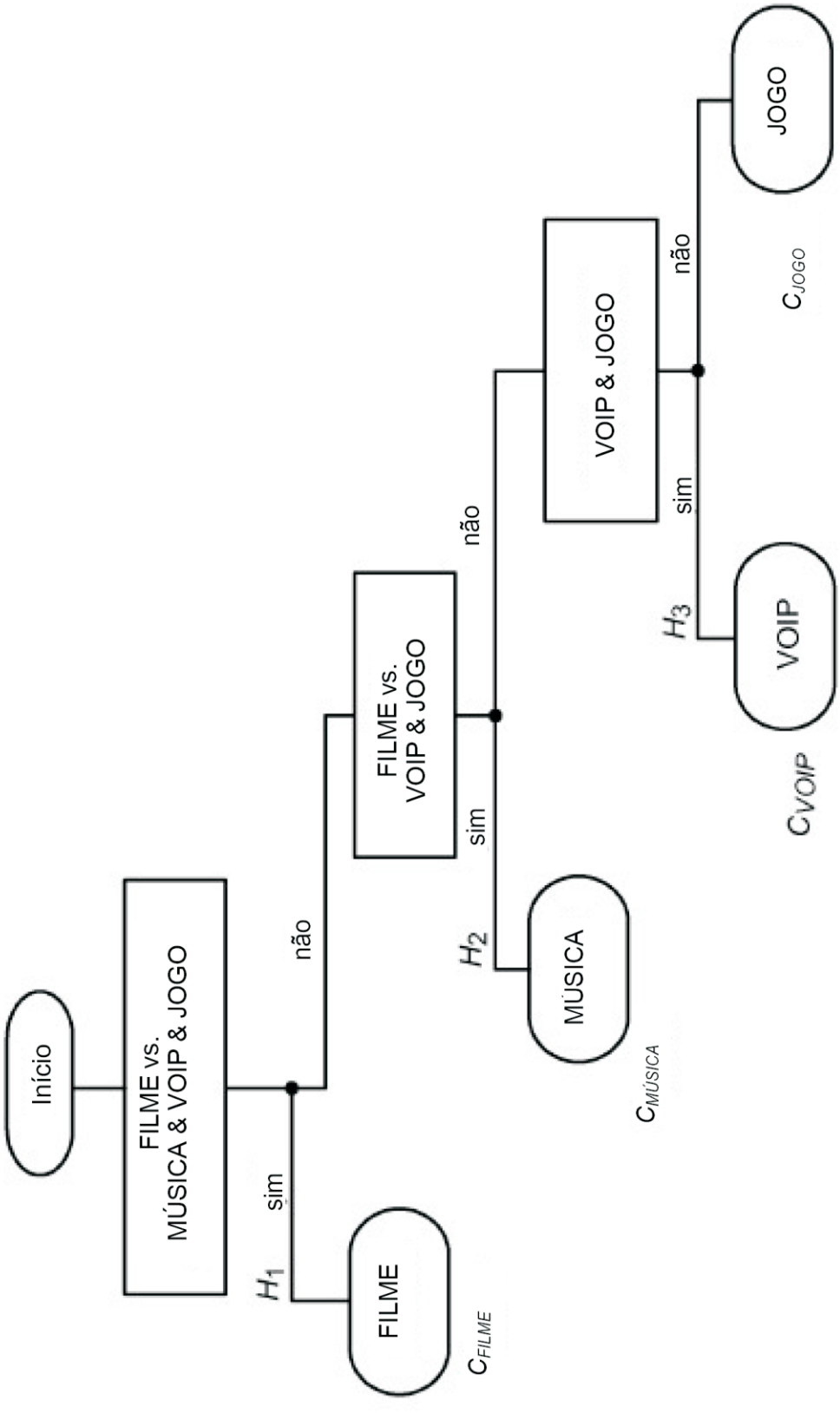
Início
FILME vs.
MÚSICA & VOIP & JOGO
sim
não
H_1
FILME
C_{FILME}
FILME vs.
VOIP & JOGO
sim
não
H_2
MÚSICA
$C_{MÚSICA}$
VOIP & JOGO
sim
não
H_3
VOIP
C_{VOIP}
JOGO
C_{JOGO}

FIG. 5

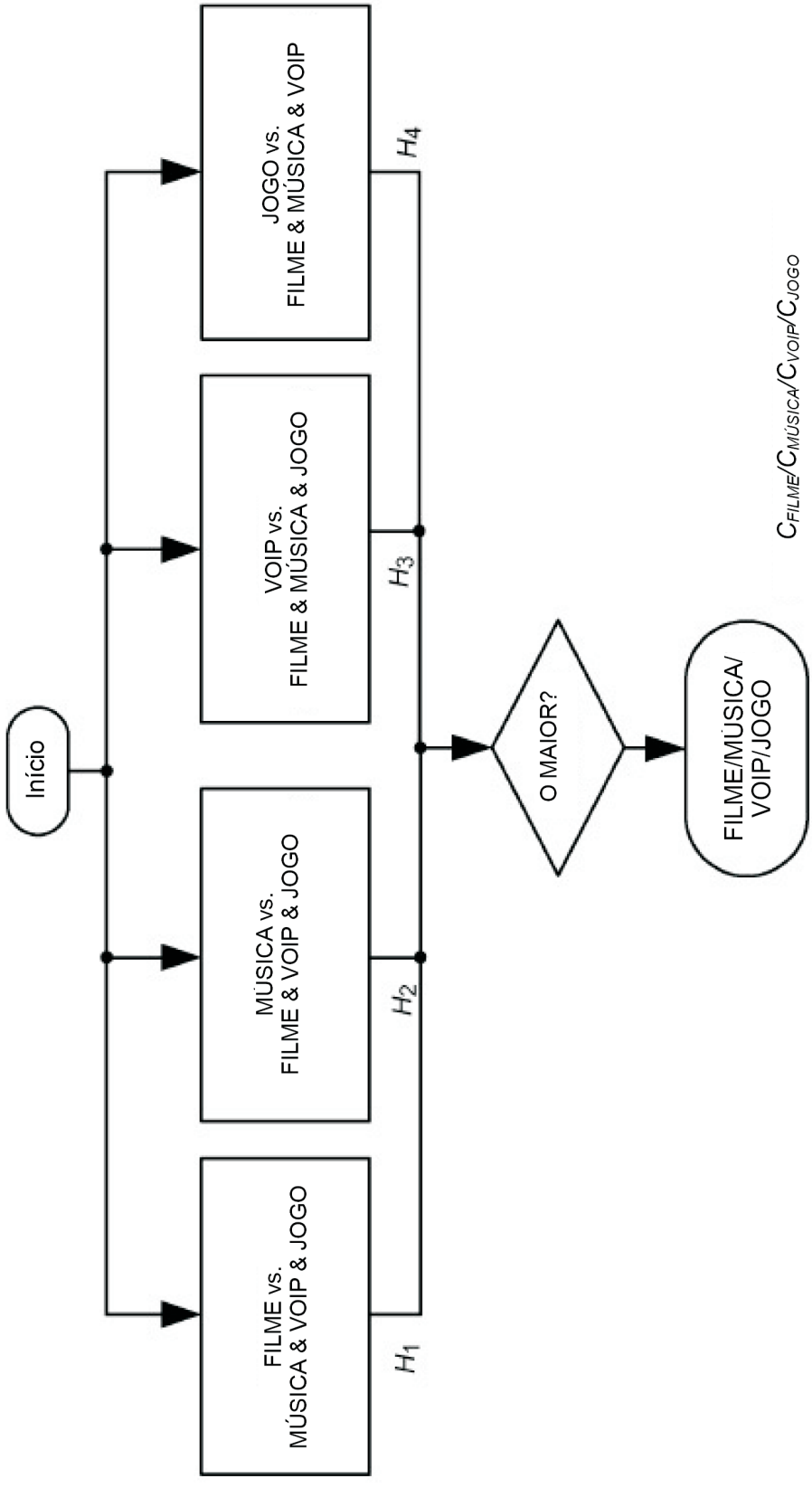
Início
FILME vs.
MÚSICA & VOIP & JOGO
MÚSICA vs.
FILME & VOIP & JOGO
VOIP vs.
FILME & MÚSICA & JOGO
JOGO vs.
FILME & MÚSICA & VOIP
H1
H2
H3
H4
O MAIOR?
FILME/MUSICA/
VOIP/JOGO
CFILME/CMÚSICA/CVOIP/CJOGO

FIG. 6

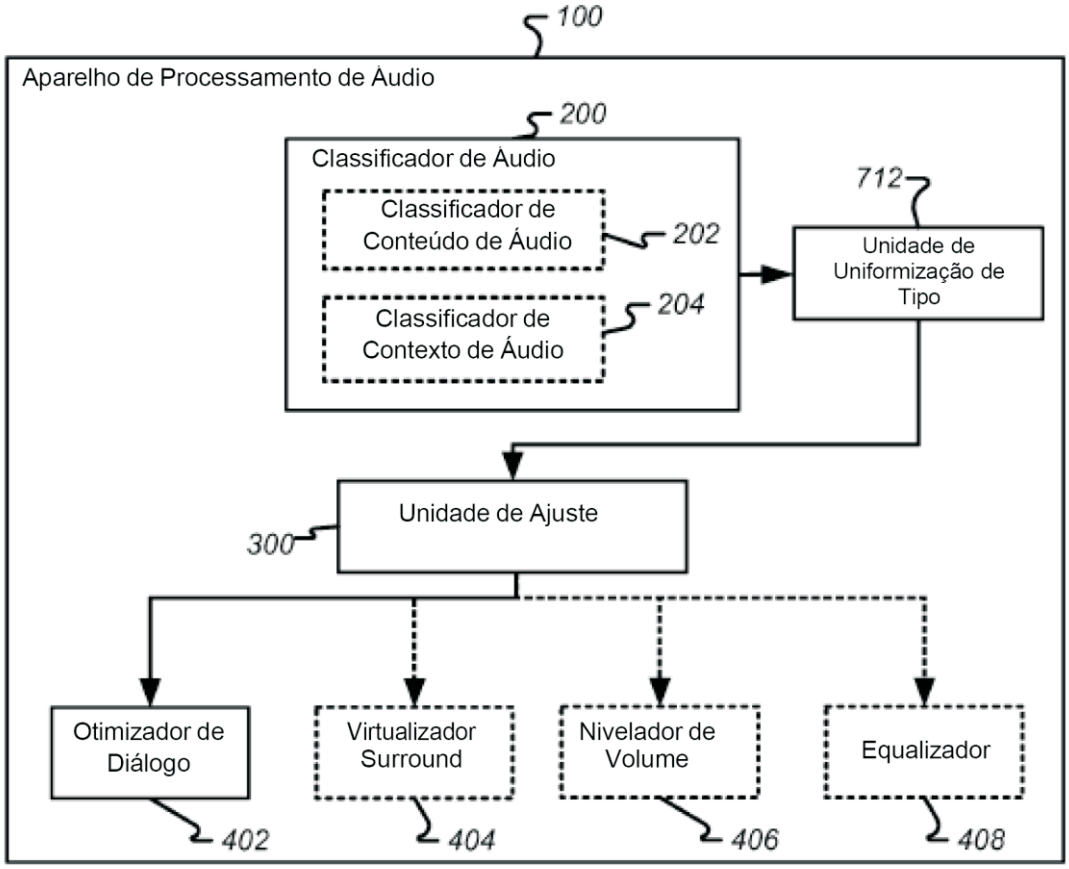
100
Aparelho de Processamento de Áudio
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
712
Unidade de Uniformização de Tipo
Unidade de Ajuste
300
Otimizador de Diálogo
Virtualizador Surround
Nivelador de Volume
Equalizador
402
404
406
408

FIG. 7

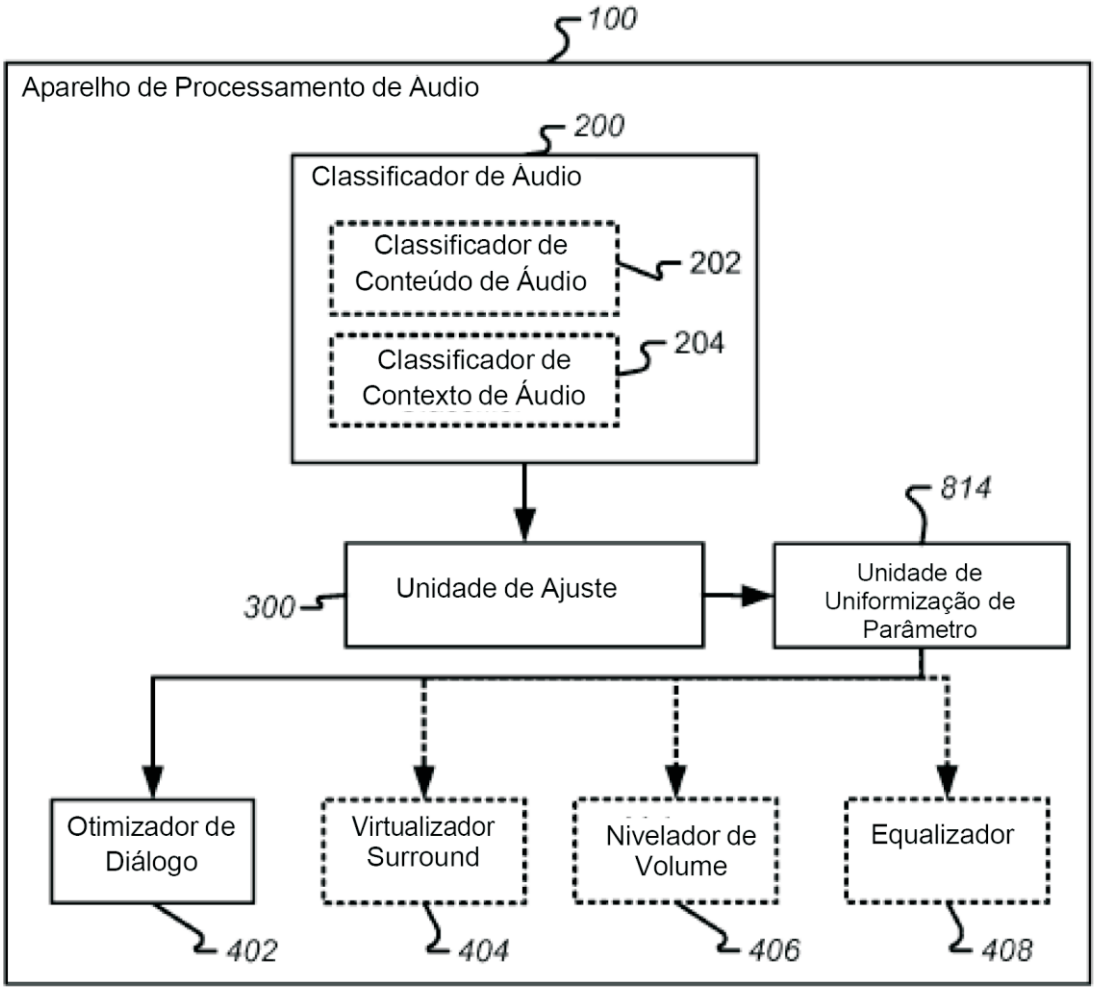
100
Aparelho de Processamento de Audio
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
814
300
Unidade de Ajuste
Unidade de Uniformização de Parâmetro
Otimizador de Diálogo
Virtualizador Surround
Nivelador de Volume
Equalizador
402
404
406
408

FIG. 8

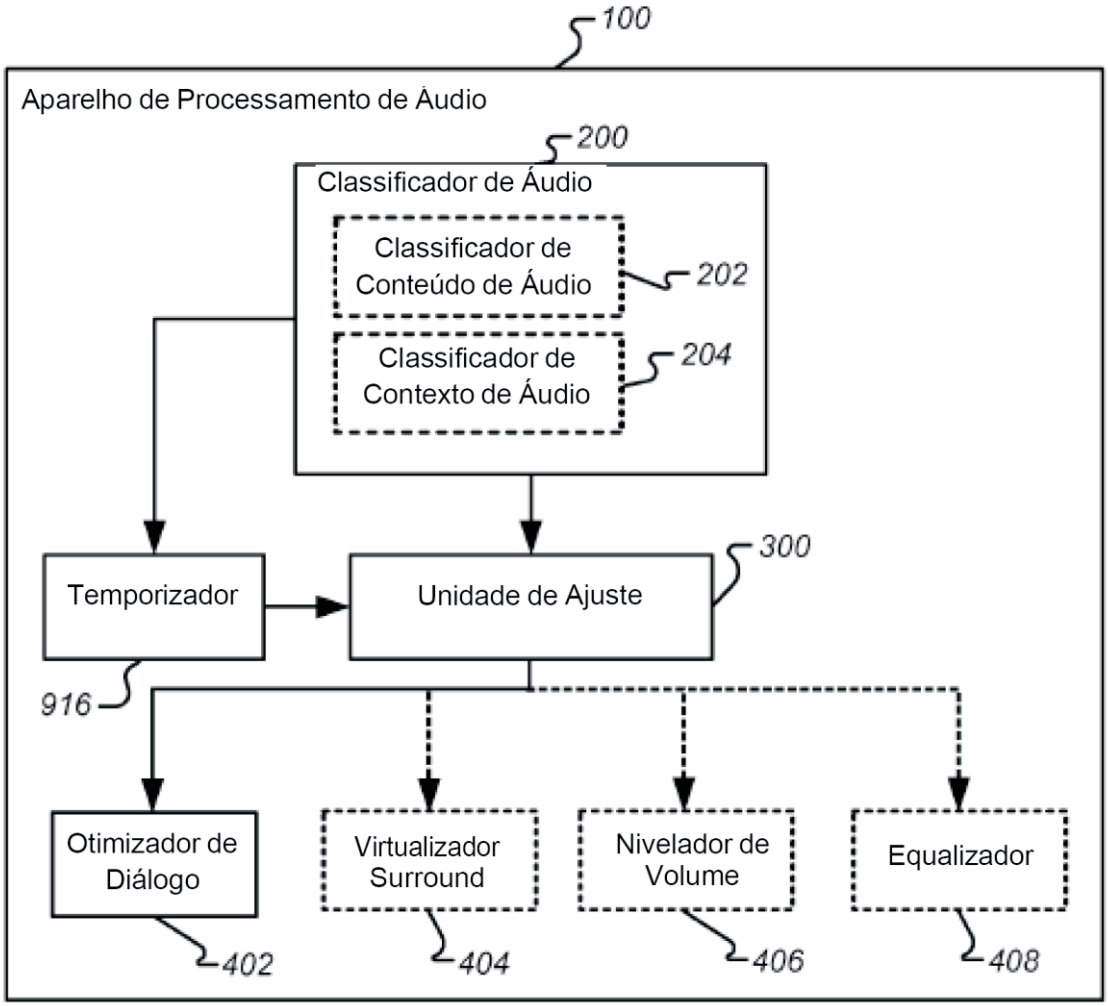
100
Aparelho de Processamento de Áudio
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
Temporizador
Unidade de Ajuste
300
916
Otimizador de Diálogo
Virtualizador Surround
Nivelador de Volume
Equalizador
402
404
406
408

FIG. 9

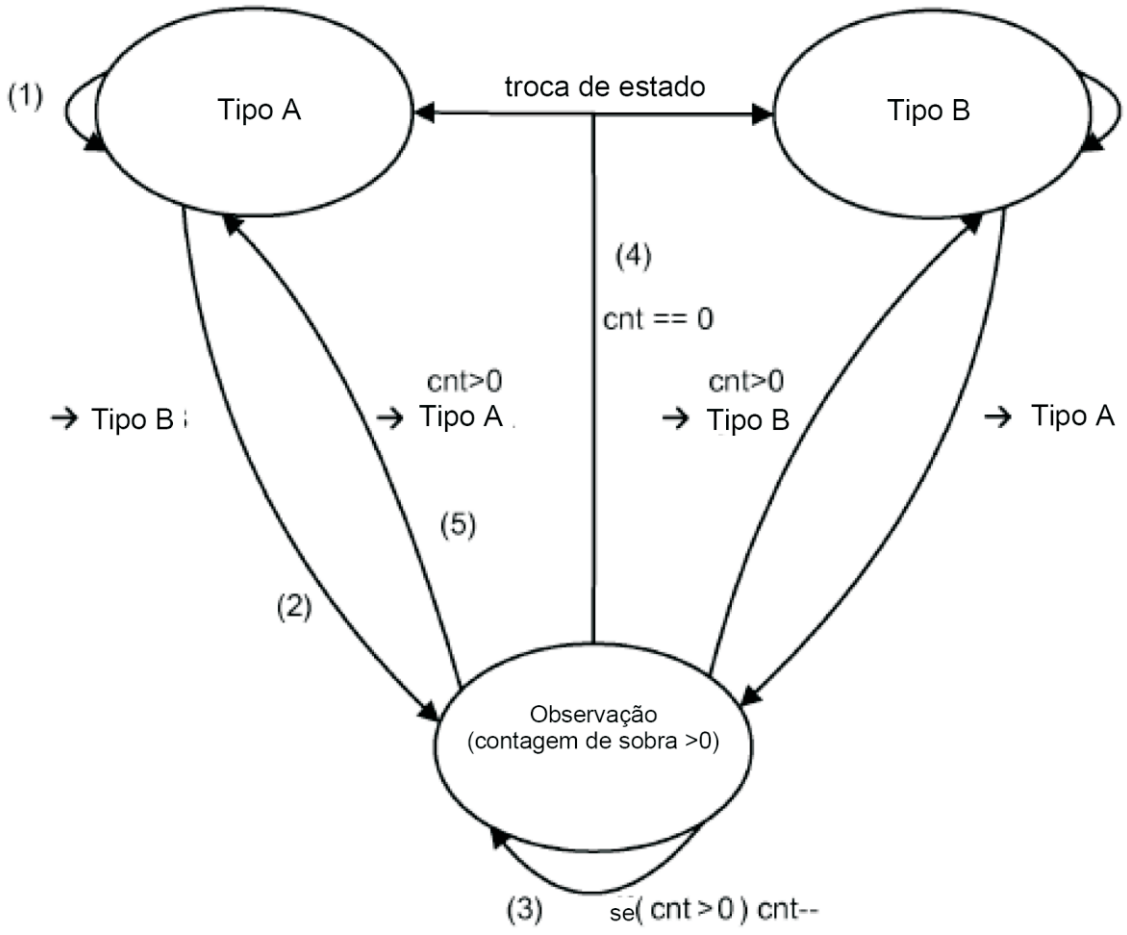
(1)
Tipo A
troca de estado
Tipo B
(4)
cnt == 0
cnt>0
→ Tipo A
cnt>0
→ Tipo B
→ Tipo B
→ Tipo A
(5)
(2)
Observação
(contagem de sobra >0)
(3)
se(cnt > 0) cnt--

FIG. 10

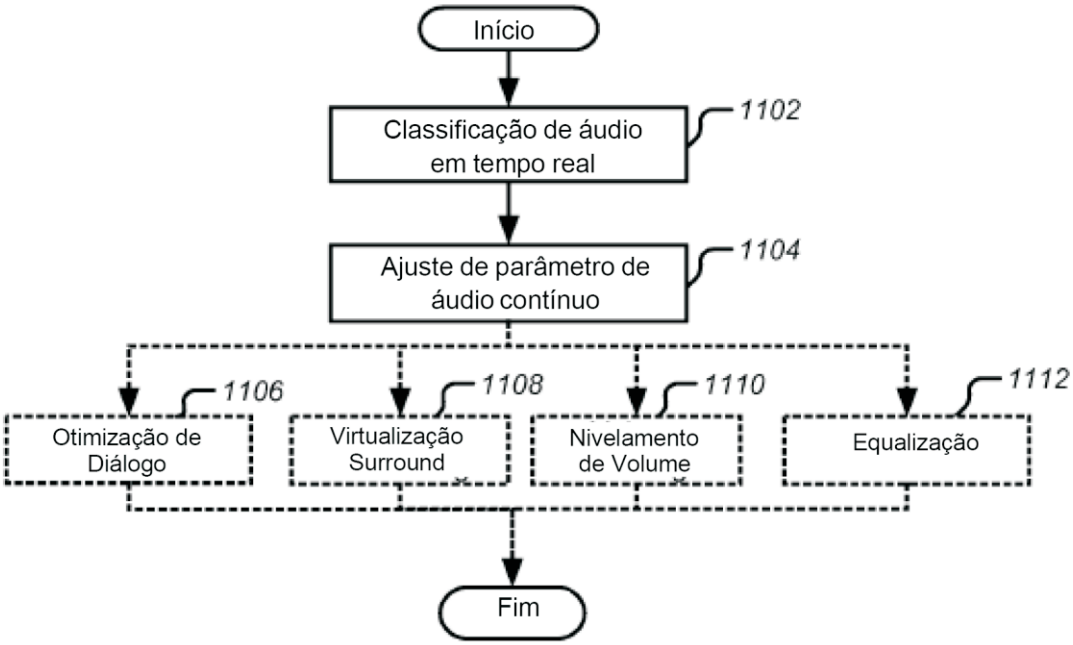

FIG. 11

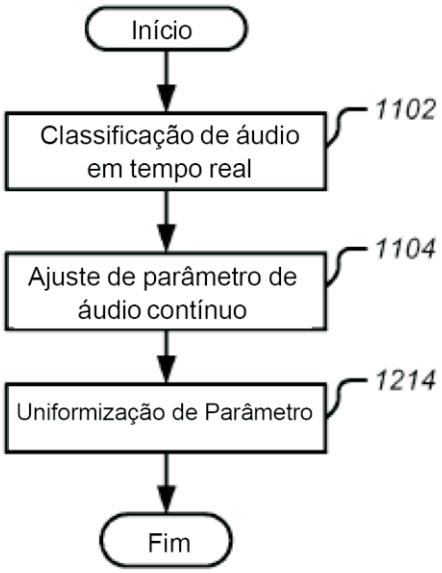

FIG. 12

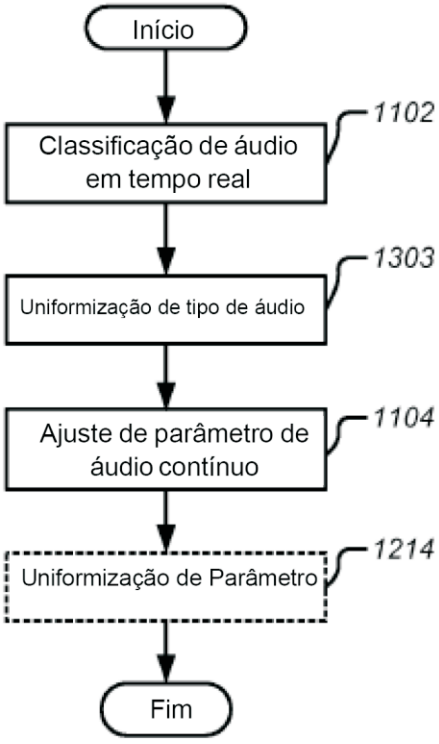

FIG. 13

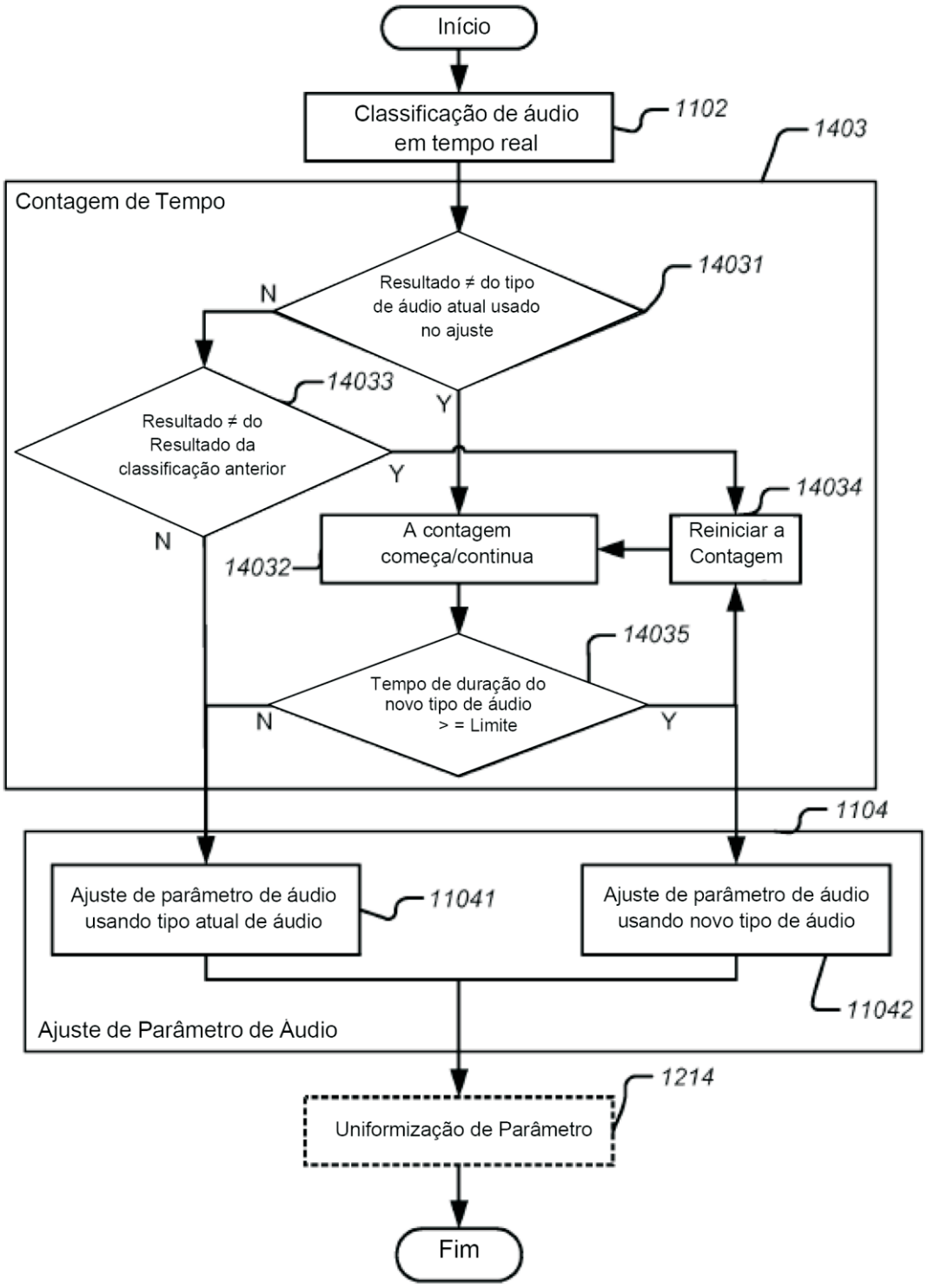
Início
Classificação de áudio em tempo real
1102
1403
Contagem de Tempo
14031
Resultado ≠ do tipo de áudio atual usado no ajuste
N
Y
14033
Resultado ≠ do Resultado da classificação anterior
Y
N
14034
A contagem começa/continua
14032
Reiniciar a Contagem
14035
Tempo de duração do novo tipo de áudio > = Limite
N
Y
1104
Ajuste de parâmetro de áudio usando tipo atual de áudio
11041
Ajuste de parâmetro de áudio usando novo tipo de áudio
11042
Ajuste de Parâmetro de Áudio
1214
Uniformização de Parâmetro
Fim

FIG. 14

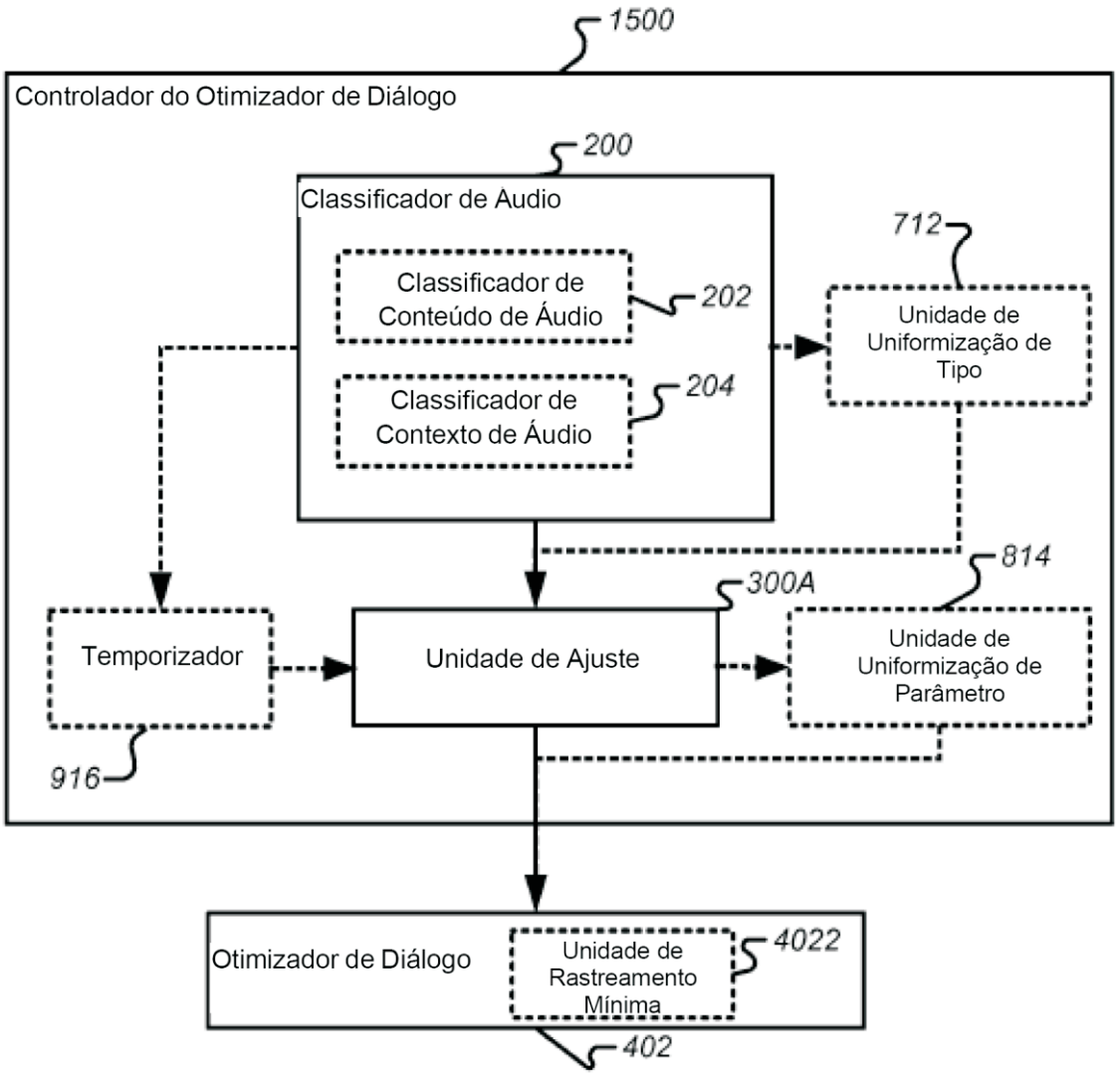
1500
Controlador do Otimizador de Diálogo
200
Classificador de Audio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
712
Unidade de Uniformização de Tipo
814
300A
Temporizador
Unidade de Ajuste
Unidade de Uniformização de Parâmetro
916
Otimizador de Diálogo
Unidade de Rastreamento Mínima
4022
402

FIG. 15

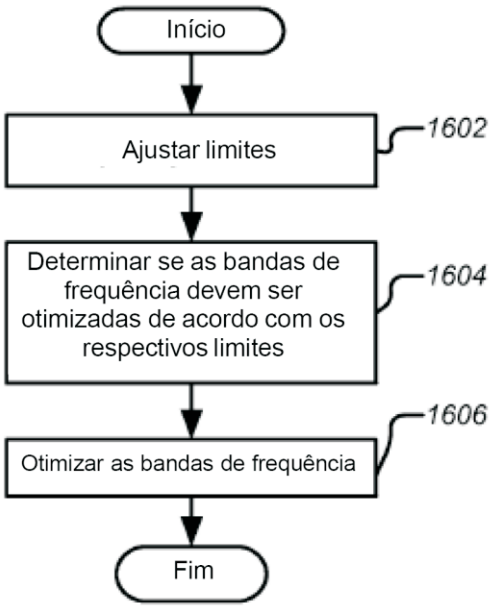

FIG. 16

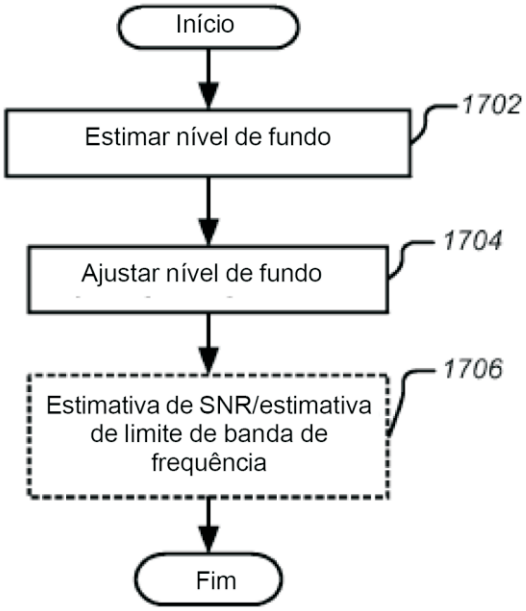

FIG. 17

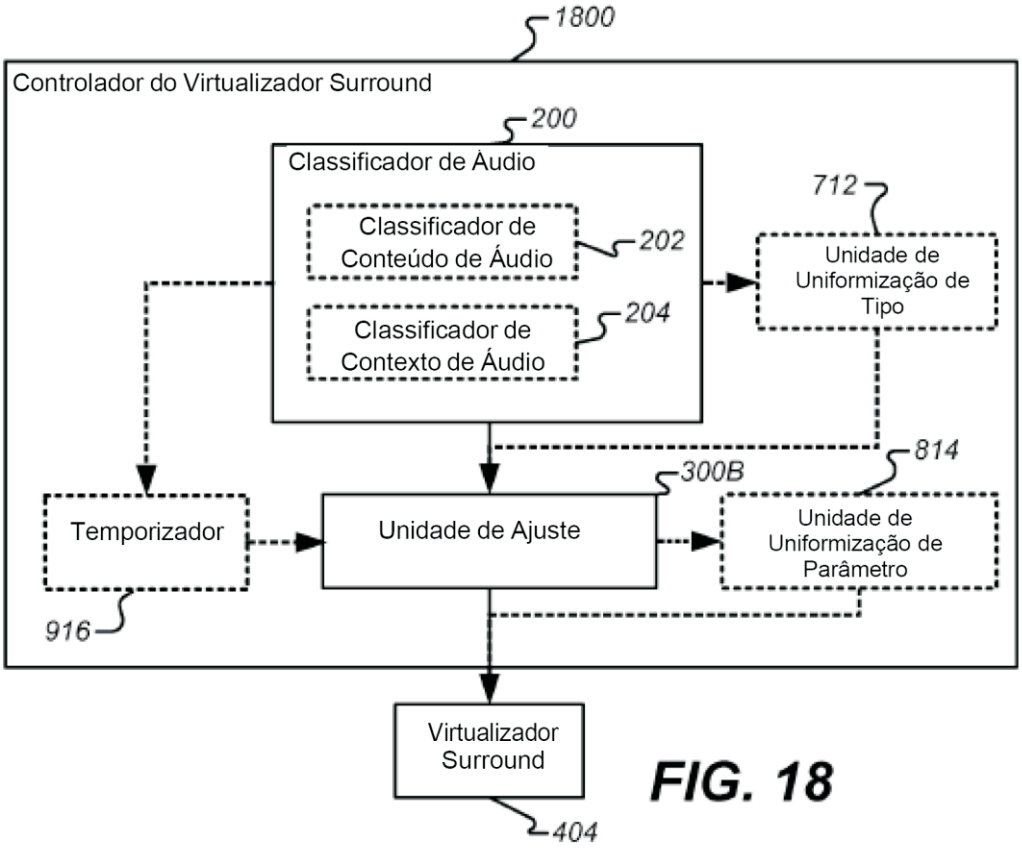

FIG. 18

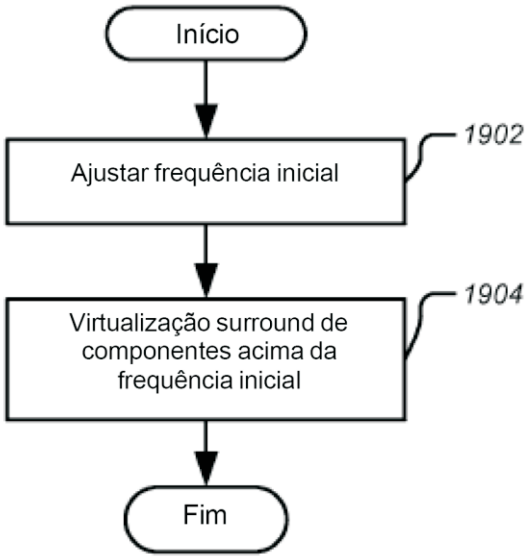
Início
Ajustar frequência inicial
1902
Virtualização surround de componentes acima da frequência inicial
1904
Fim

FIG. 19

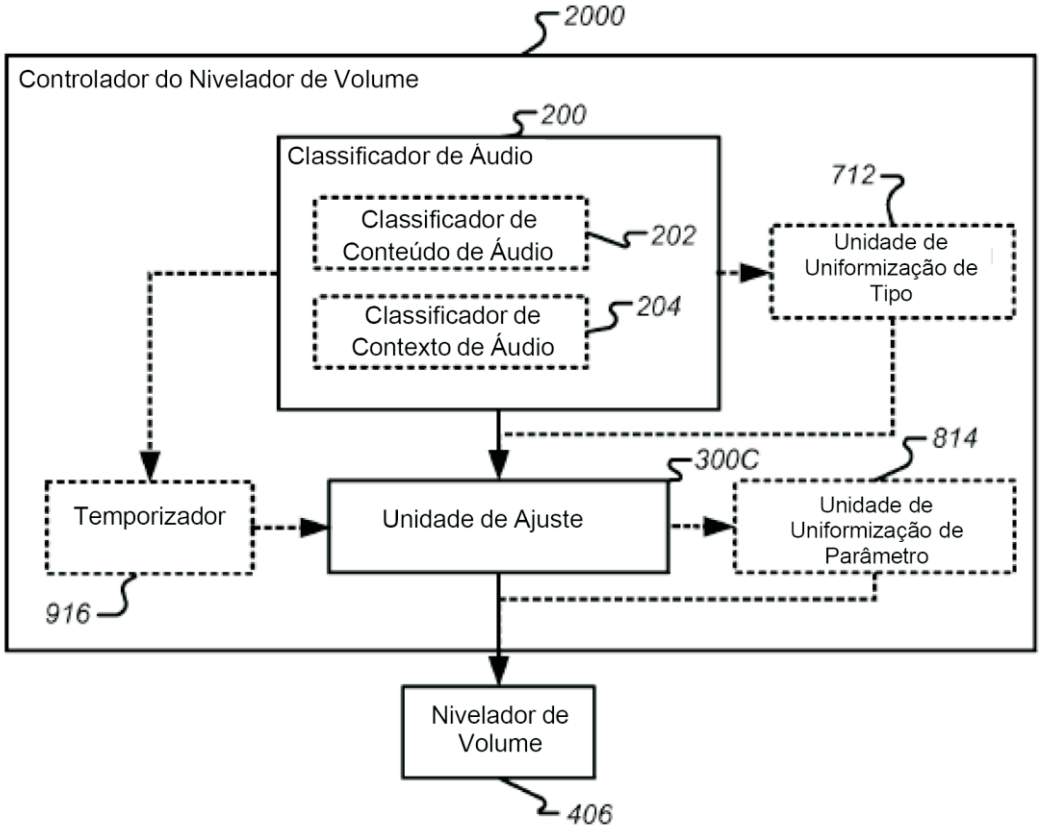
2000
Controlador do Nivelador de Volume
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
712
Unidade de Uniformização de Tipo
814
300C
Temporizador
Unidade de Ajuste
Unidade de Uniformização de Parâmetro
916
Nivelador de Volume
406

FIG. 20

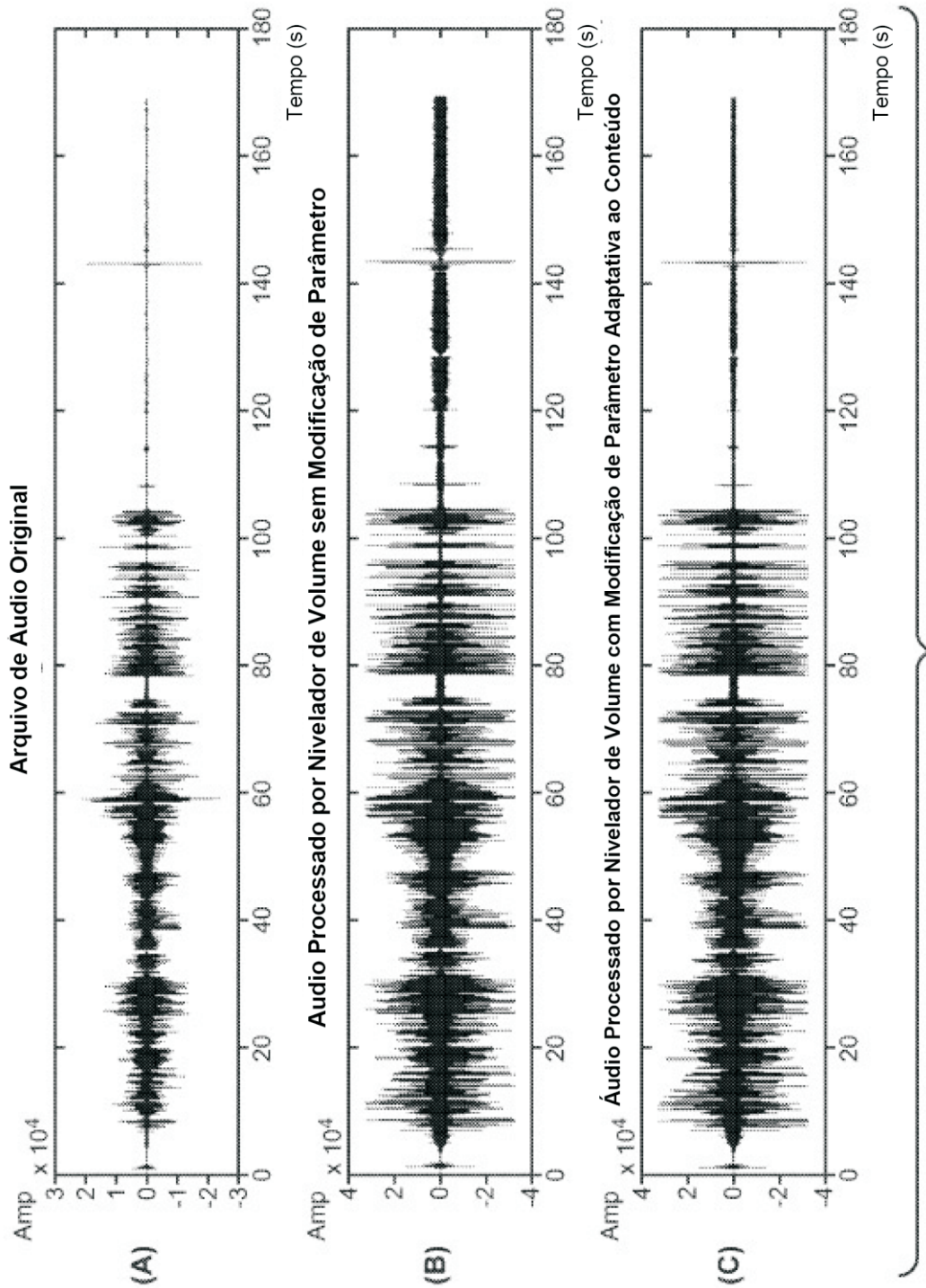
Arquivo de Audio Original
Amp
x 10^4
Tempo (s)
(A)
Audio Processado por Nivelador de Volume sem Modificação de Parâmetro
Amp
x 10^4
Tempo (s)
(B)
Áudio Processado por Nivelador de Volume com Modificação de Parâmetro Adaptativa ao Conteúdo
Amp
x 10^4
Tempo (s)
(C)

FIG. 21

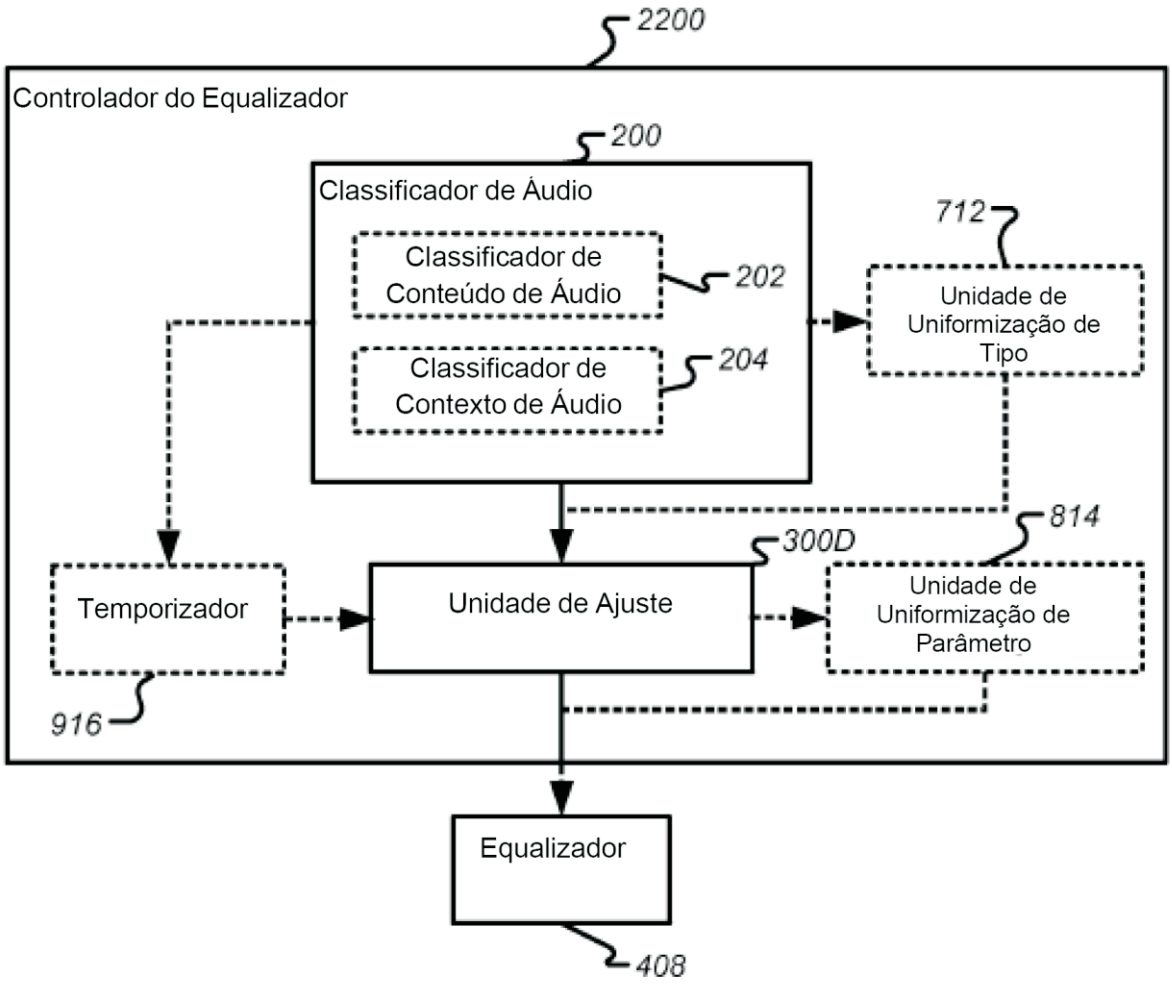
2200
Controlador do Equalizador
200
Classificador de Áudio
Classificador de Conteúdo de Áudio
202
Classificador de Contexto de Áudio
204
712
Unidade de Uniformização de Tipo
814
300D
Temporizador
Unidade de Ajuste
Unidade de Uniformização de Parâmetro
916
Equalizador
408

FIG. 22

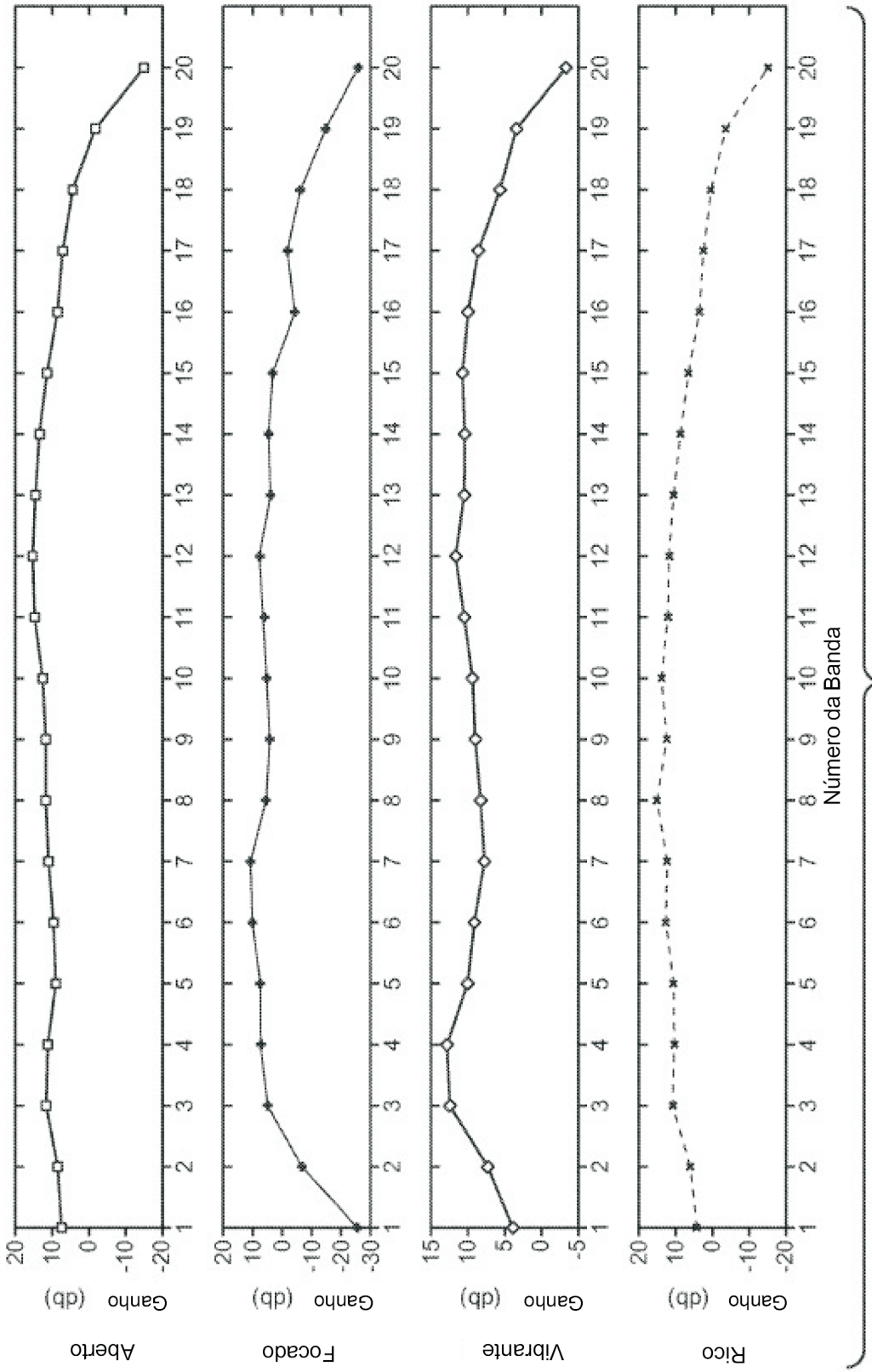

FIG. 23

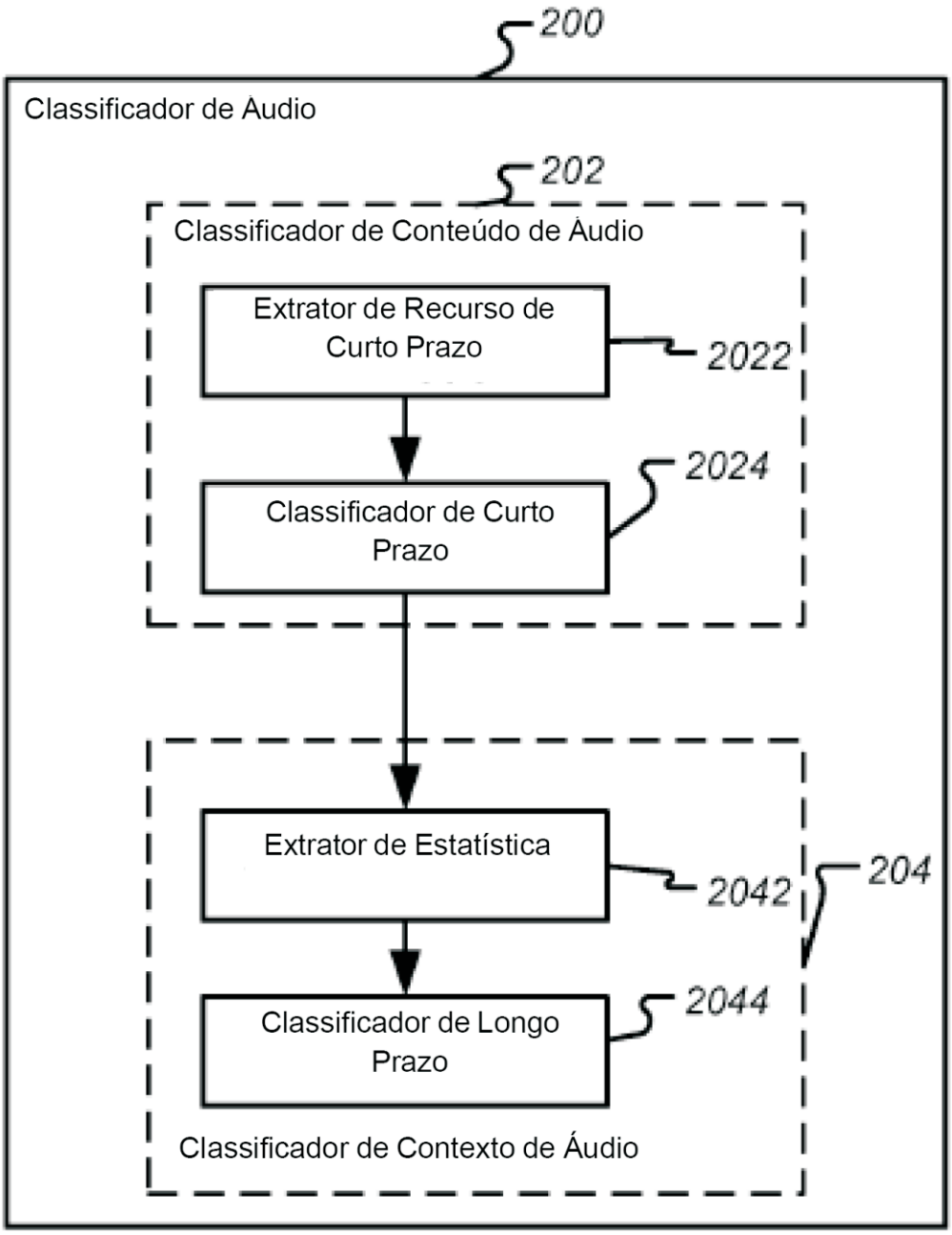
200
Classificador de Áudio
202
Classificador de Conteúdo de Áudio
Extrator de Recurso de Curto Prazo
2022
2024
Classificador de Curto Prazo
Extrator de Estatística
2042
204
2044
Classificador de Longo Prazo
Classificador de Contexto de Áudio

FIG. 24

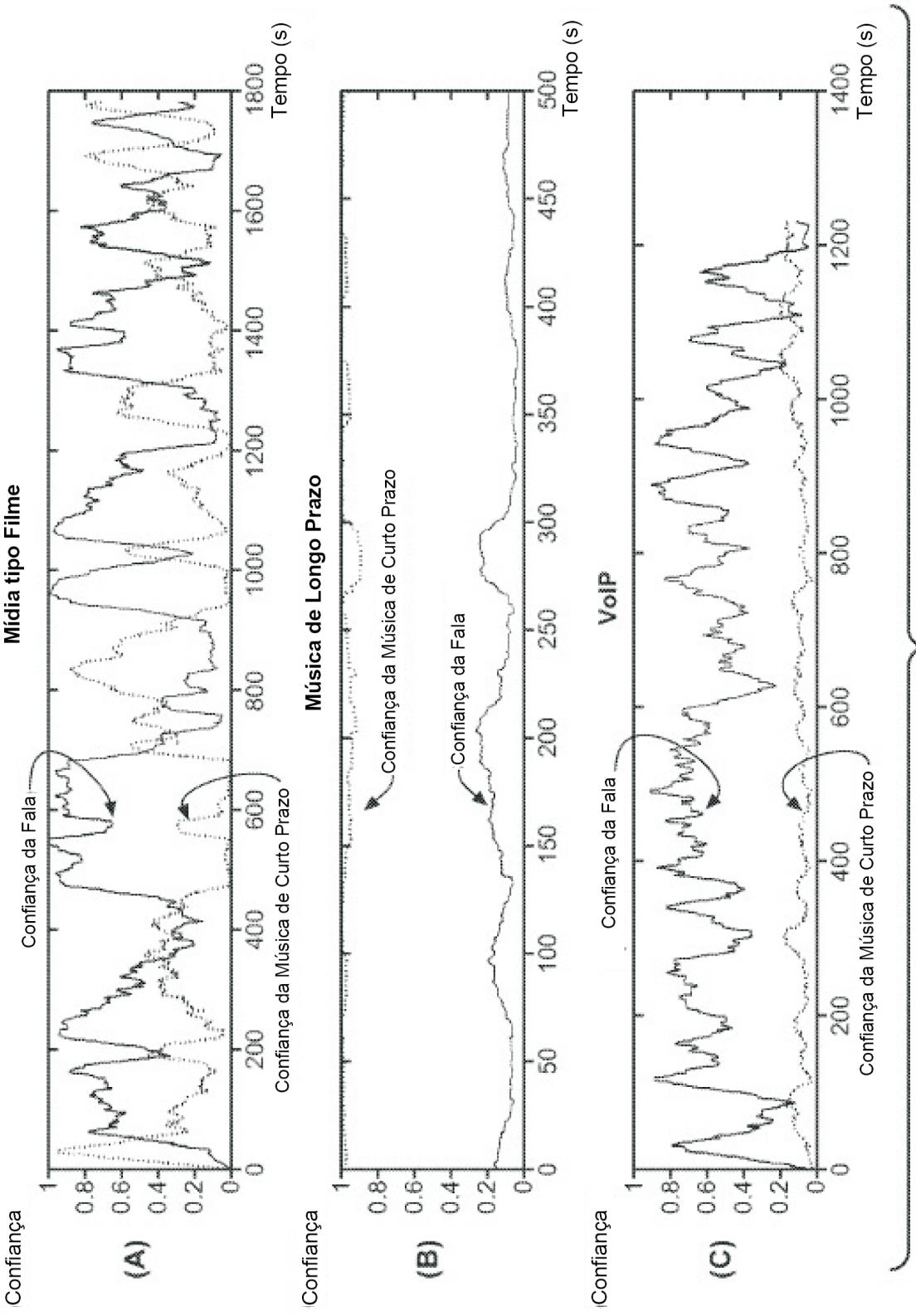
Mídia tipo Filme
Confiança
(A)
Confiança da Fala
Confiança da Música de Curto Prazo
Tempo (s)
Música de Longo Prazo
Confiança
(B)
Confiança da Música de Curto Prazo
Confiança da Fala
Tempo (s)
VoIP
Confiança
(C)
Confiança da Fala
Confiança da Música de Curto Prazo
Tempo (s)

FIG. 25

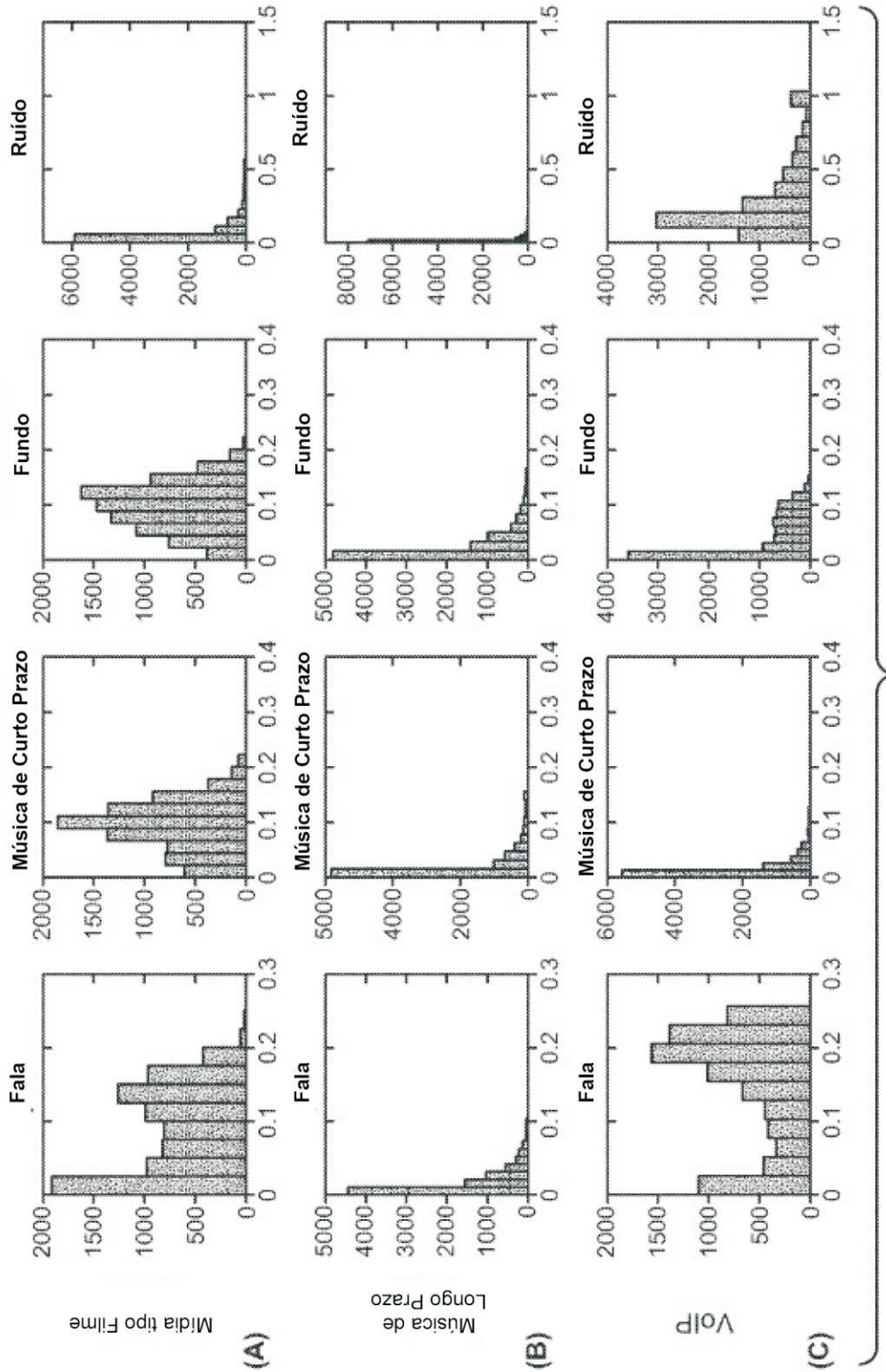
Fala
Música de Curto Prazo
Fundo
Ruído
(A)
Mídia tipo Filme
(B)
Música de Longo Prazo
(C)
VoIP

FIG. 26

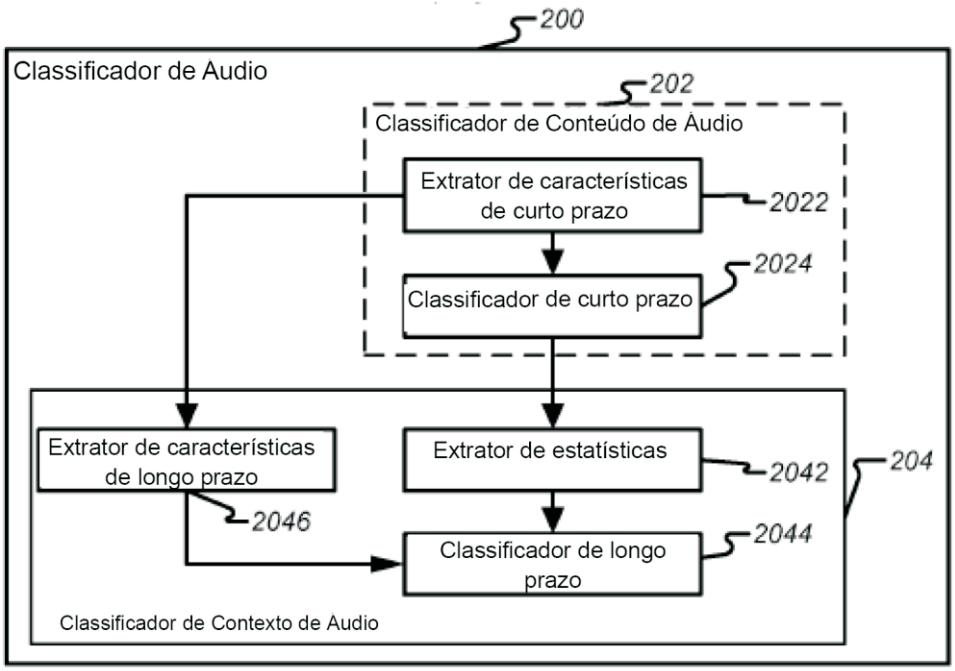
200
Classificador de Áudio
202
Classificador de Conteúdo de Áudio
Extrator de características de curto prazo
2022
2024
Classificador de curto prazo
Extrator de características de longo prazo
Extrator de estatísticas
2042
204
2046
2044
Classificador de longo prazo
Classificador de Contexto de Áudio

FIG. 27

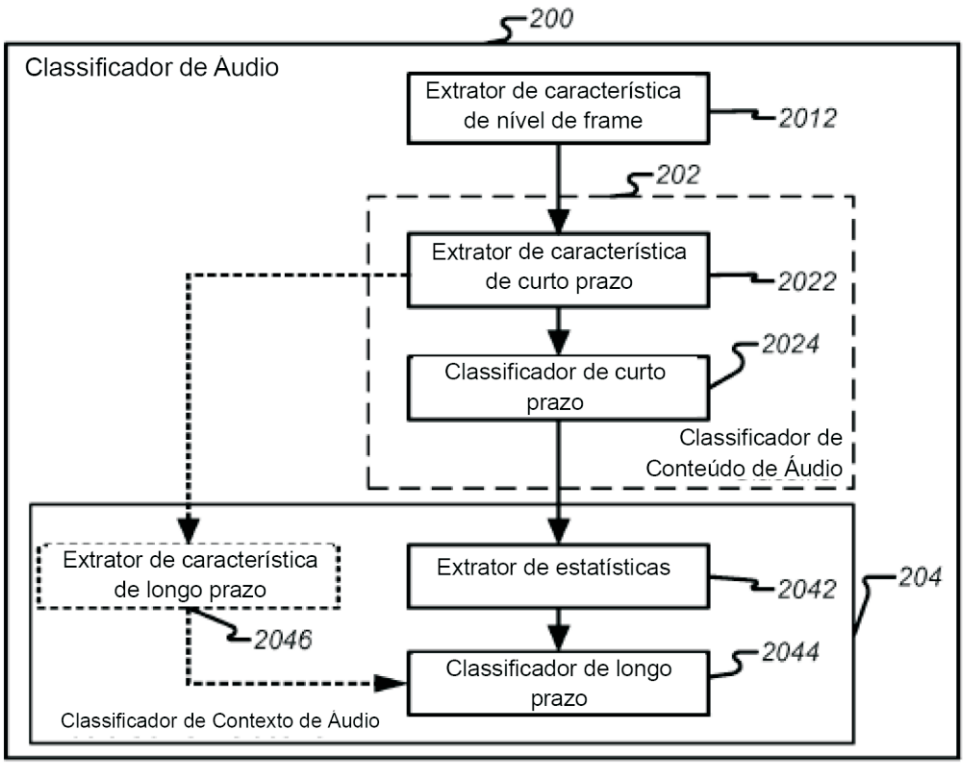
200
Classificador de Áudio
Extrator de característica de nível de frame
2012
202
Extrator de característica de curto prazo
2022
2024
Classificador de curto prazo
Classificador de Conteúdo de Áudio
Extrator de característica de longo prazo
Extrator de estatísticas
2042
204
2046
2044
Classificador de longo prazo
Classificador de Contexto de Áudio

FIG. 28

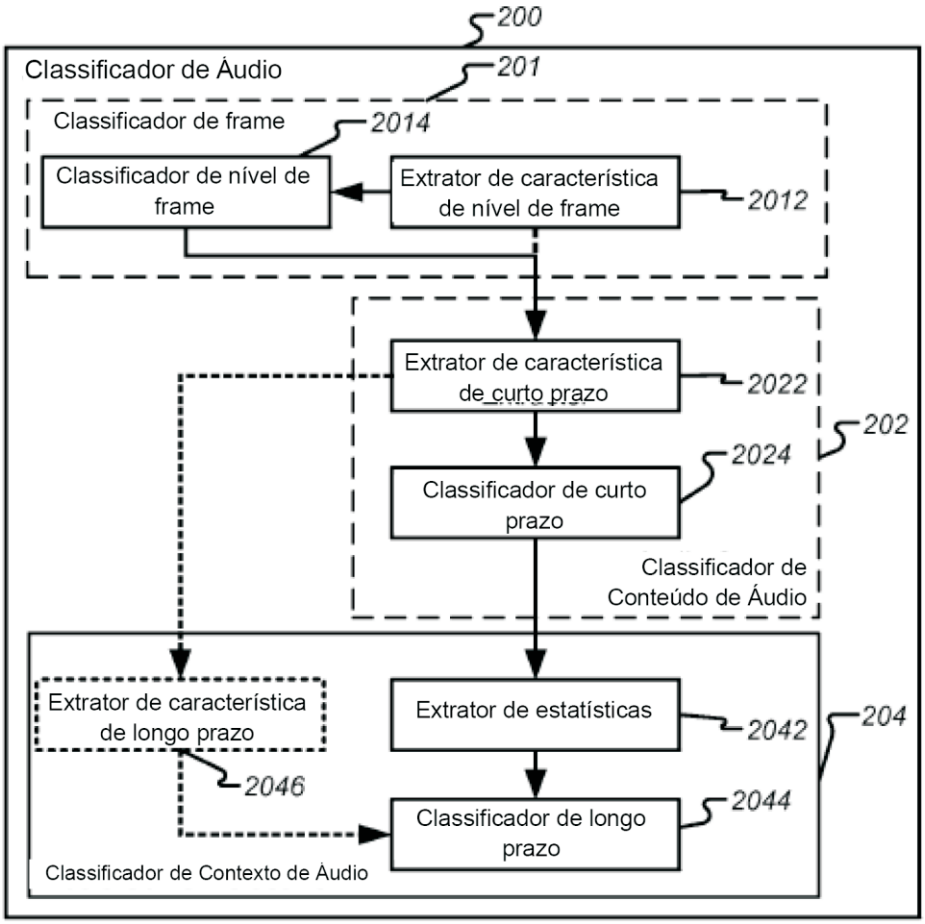
200
Classificador de Áudio
201
Classificador de frame
2014
Classificador de nível de frame
Extrator de característica de nível de frame
2012
Extrator de característica de curto prazo
2022
202
2024
Classificador de curto prazo
Classificador de Conteúdo de Áudio
Extrator de característica de longo prazo
2046
Extrator de estatísticas
2042
204
2044
Classificador de longo prazo
Classificador de Contexto de Áudio

FIG. 29

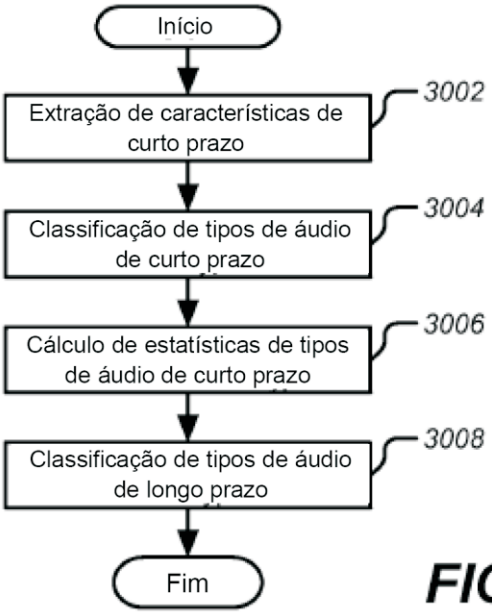
Início
Extração de características de curto prazo
3002
Classificação de tipos de áudio de curto prazo
3004
Cálculo de estatísticas de tipos de áudio de curto prazo
3006
Classificação de tipos de áudio de longo prazo
3008
Fim

FIG. 30

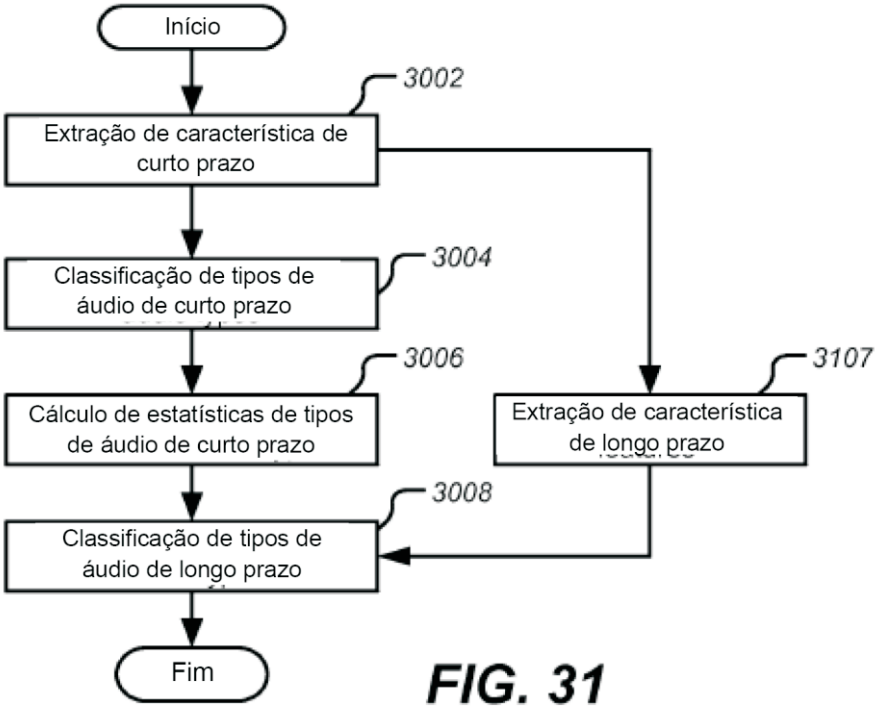
Início
3002
Extração de característica de curto prazo
3004
Classificação de tipos de áudio de curto prazo
3006
Cálculo de estatísticas de tipos de áudio de curto prazo
3107
Extração de característica de longo prazo
3008
Classificação de tipos de áudio de longo prazo
Fim

FIG. 31

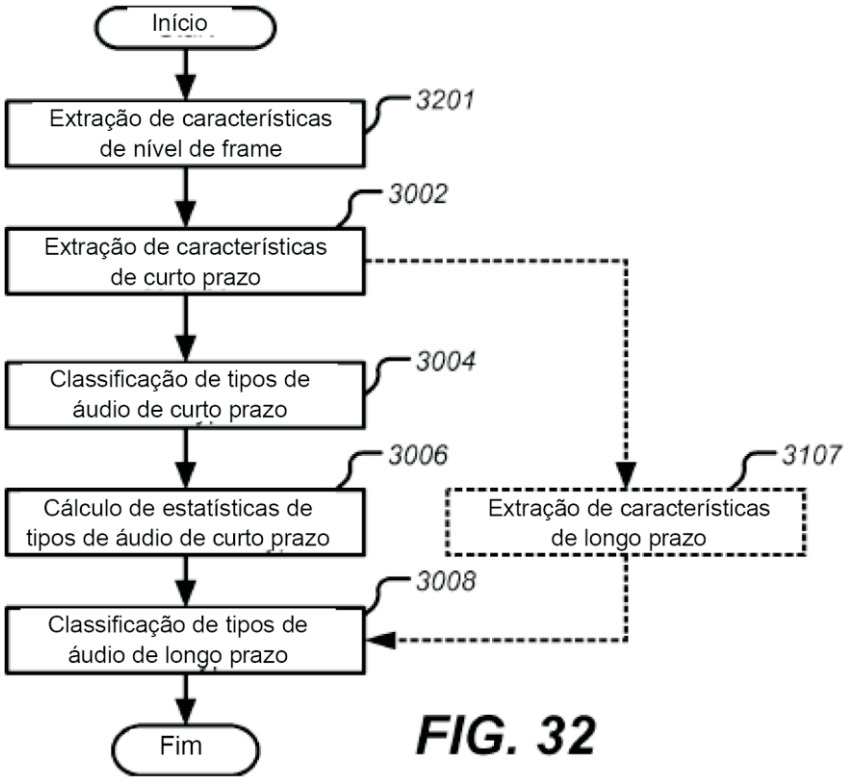
Início
3201
Extração de características de nível de frame
3002
Extração de características de curto prazo
3004
Classificação de tipos de áudio de curto prazo
3006
Cálculo de estatísticas de tipos de áudio de curto prazo
3107
Extração de características de longo prazo
3008
Classificação de tipos de áudio de longo prazo
Fim

FIG. 32

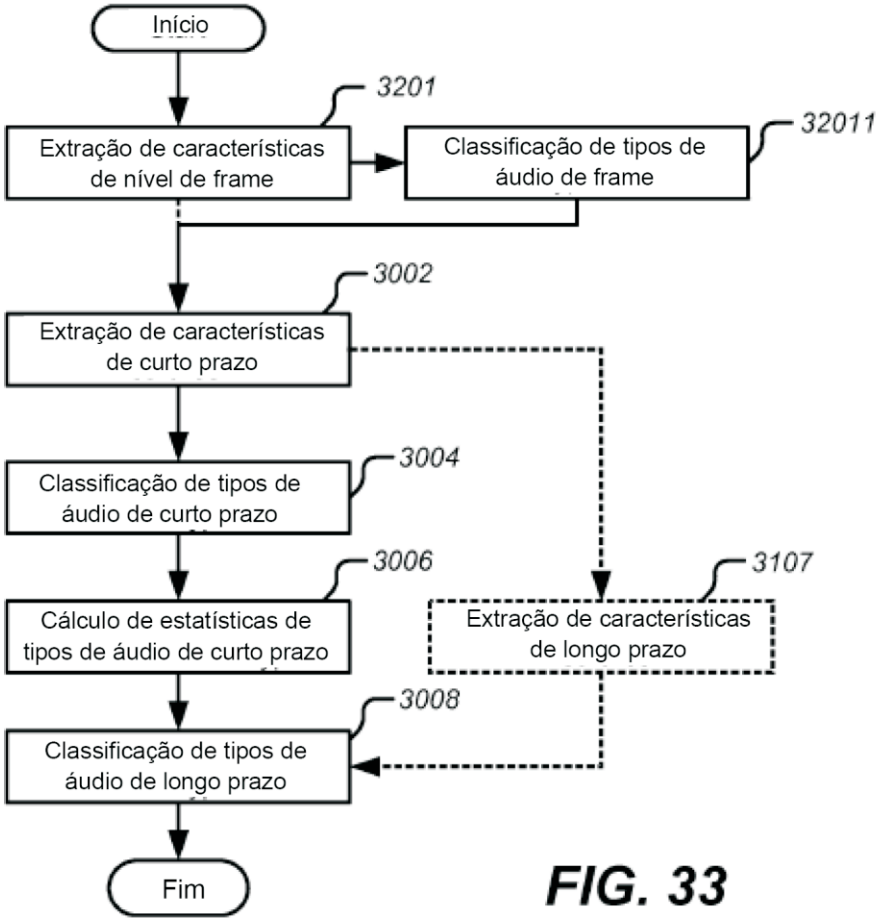
Início
3201
Extração de características de nível de frame
32011
Classificação de tipos de áudio de frame
3002
Extração de características de curto prazo
3004
Classificação de tipos de áudio de curto prazo
3006
Cálculo de estatísticas de tipos de áudio de curto prazo
3107
Extração de características de longo prazo
3008
Classificação de tipos de áudio de longo prazo
Fim

FIG. 33

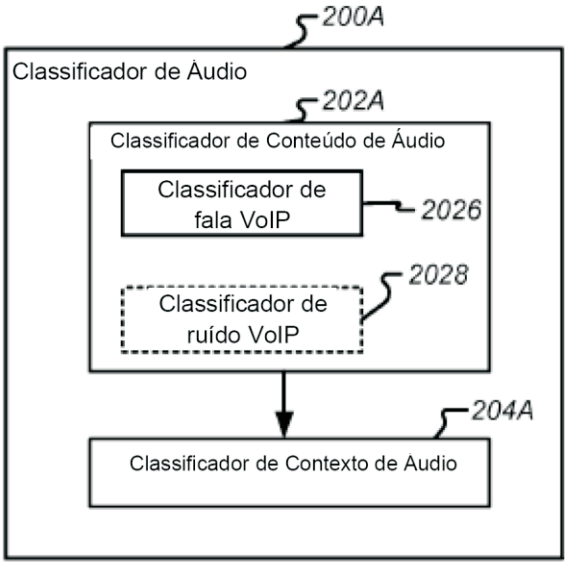
200A
Classificador de Áudio
202A
Classificador de Conteúdo de Áudio
Classificador de fala VoIP
2026
2028
Classificador de ruído VoIP
204A
Classificador de Contexto de Áudio

FIG. 34

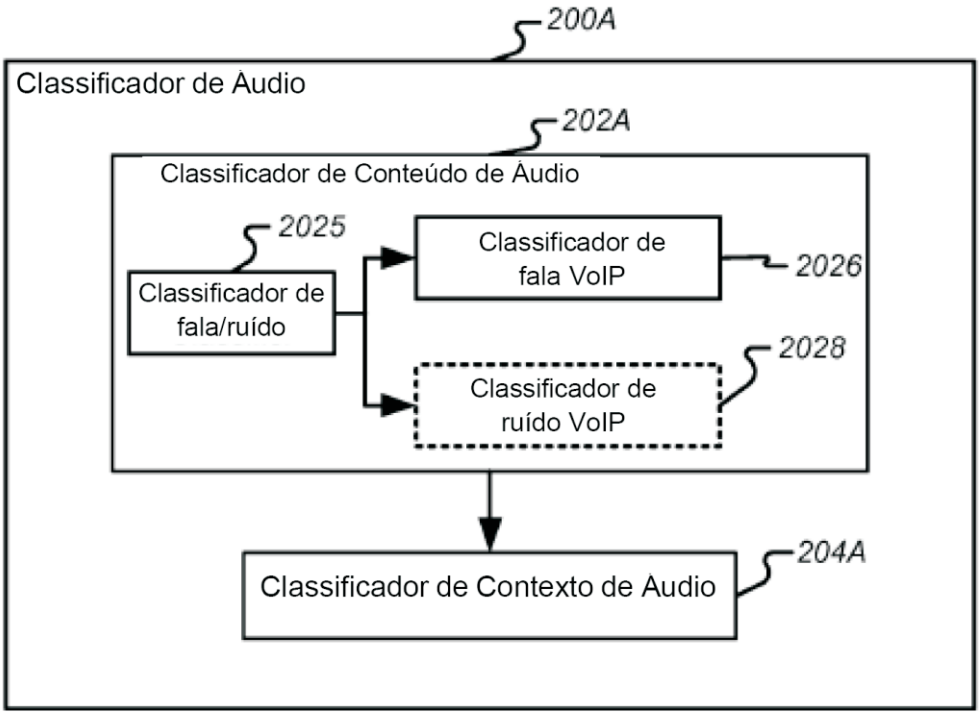
200A
Classificador de Áudio
202A
Classificador de Conteúdo de Áudio
2025
Classificador de fala/ruído
Classificador de fala VoIP
2026
2028
Classificador de ruído VoIP
204A
Classificador de Contexto de Áudio

FIG. 35

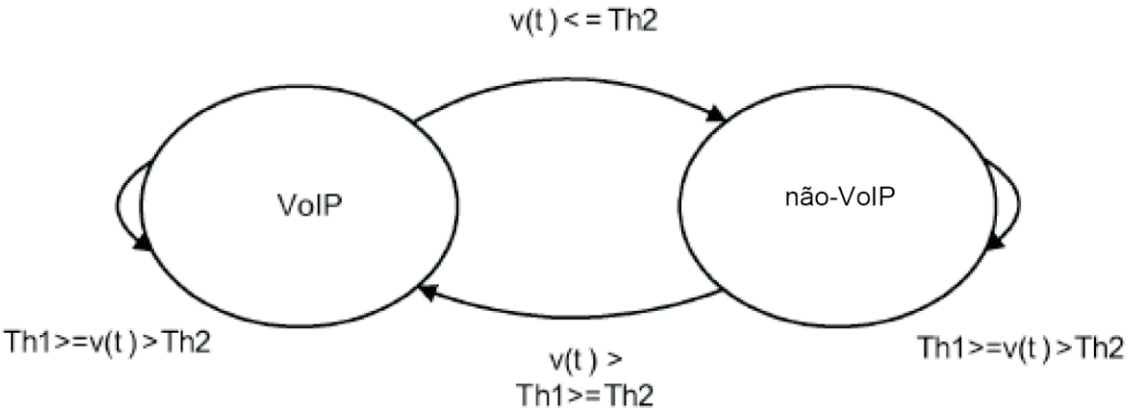
v(t) < = Th2
VoIP
não-VoIP
Th1>=v(t)>Th2
v(t) >
Th1>=Th2
Th1>=v(t)>Th2

FIG. 36

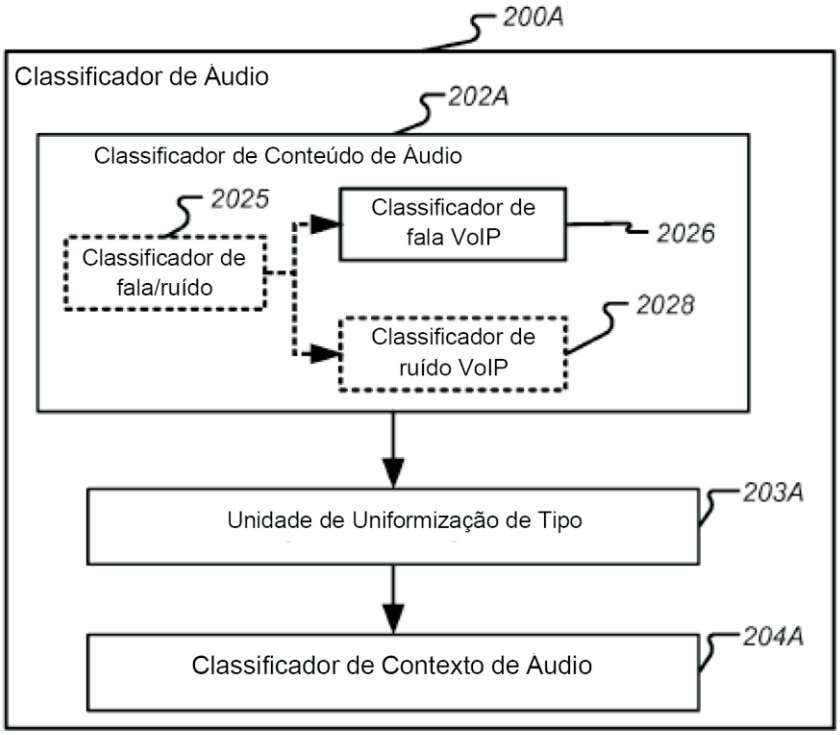
200A
Classificador de Áudio
202A
Classificador de Conteúdo de Áudio
2025
Classificador de fala/ruído
Classificador de fala VoIP
2026
2028
Classificador de ruído VoIP
Unidade de Uniformização de Tipo
203A
Classificador de Contexto de Áudio
204A

FIG. 37

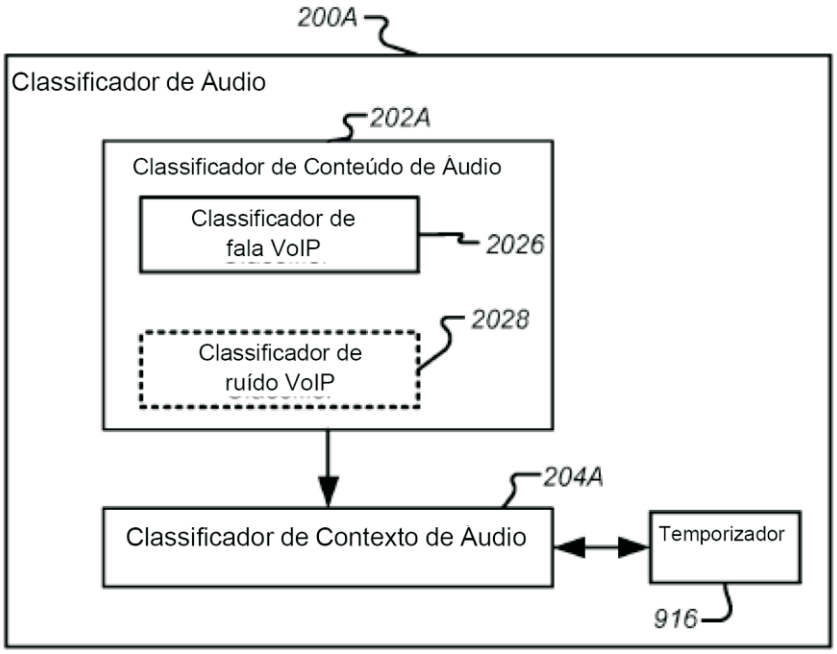
200A
Classificador de Audio
202A
Classificador de Conteúdo de Áudio
Classificador de fala VoIP
2026
2028
Classificador de ruído VoIP
204A
Classificador de Contexto de Áudio
Temporizador
916

FIG. 38

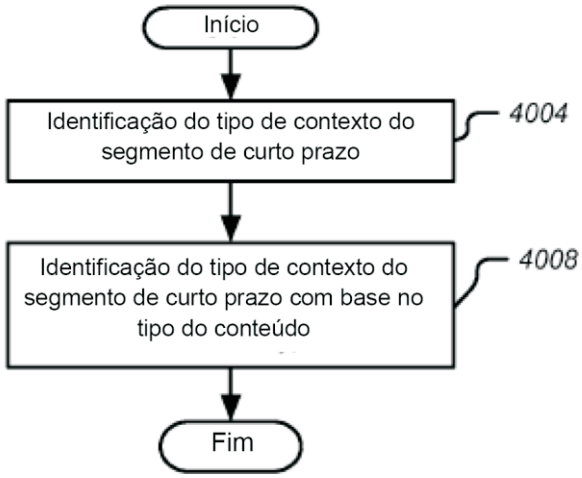
Início
Identificação do tipo de contexto do segmento de curto prazo
4004
Identificação do tipo de contexto do segmento de curto prazo com base no tipo do conteúdo
4008
Fim

FIG. 39

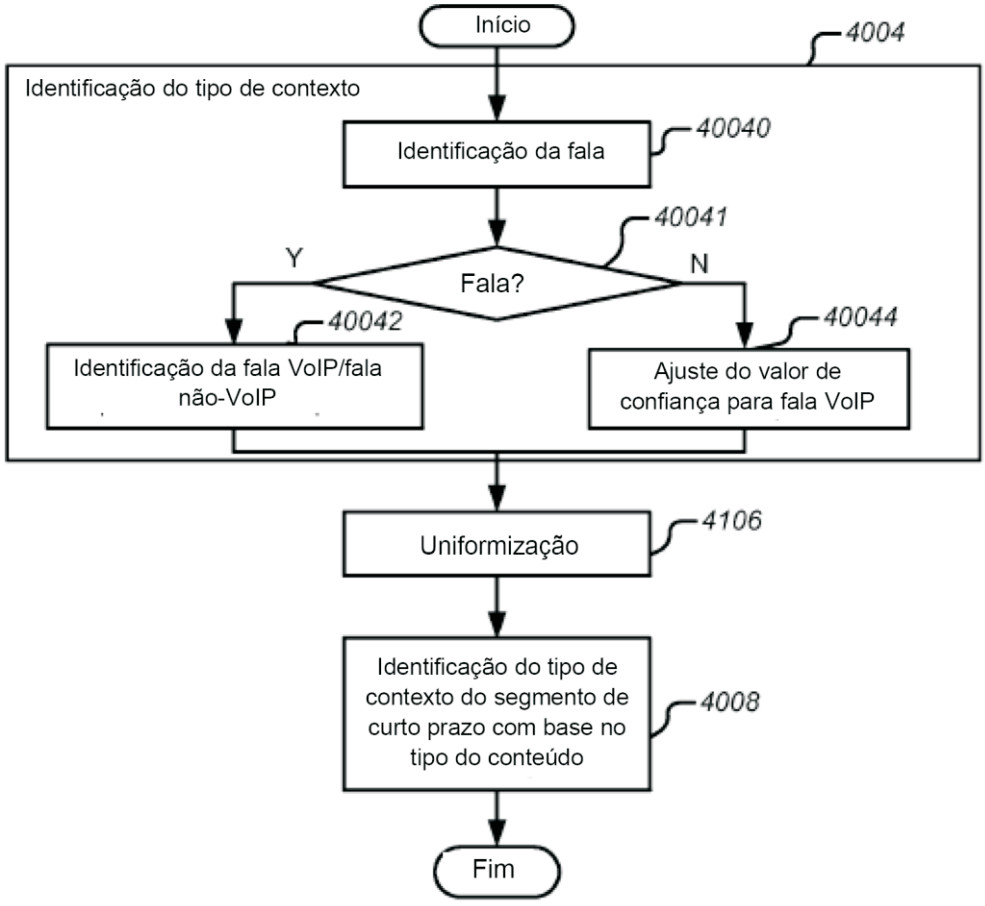
Início
4004
Identificação do tipo de contexto
Identificação da fala
40040
40041
Y
Fala?
N
40042
Identificação da fala VoIP/fala não-VoIP
40044
Ajuste do valor de confiança para fala VoIP
Uniformização
4106
Identificação do tipo de contexto do segmento de curto prazo com base no tipo do conteúdo
4008
Fim

FIG. 40

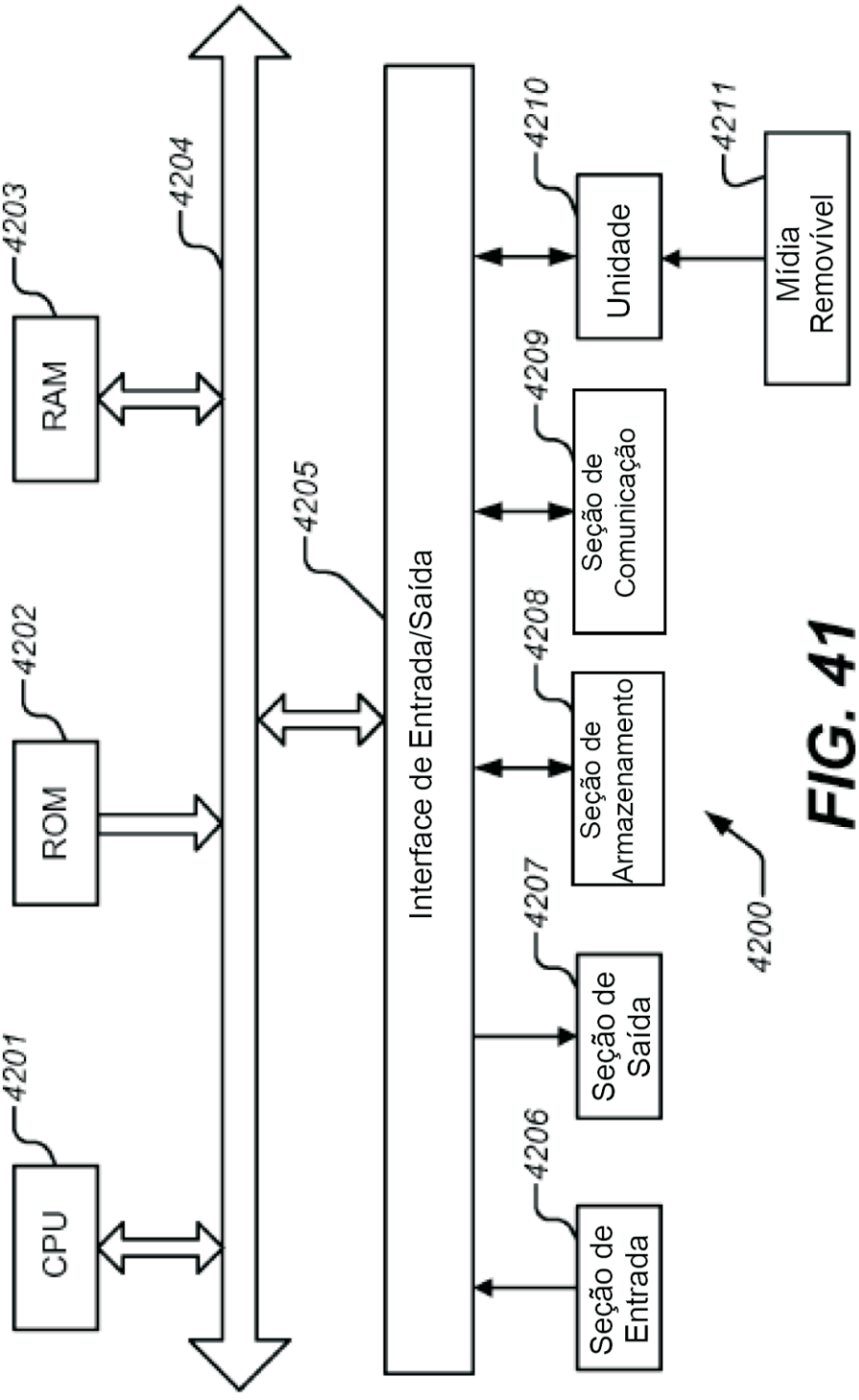
CPU
4201
ROM
4202
RAM
4203
4204
4205
Interface de Entrada/Saída
4206
Seção de Entrada
4207
Seção de Saída
4208
Seção de Armazenamento
4209
Seção de Comunicação
4210
Unidade
4211
Mídia Removível
4200

FIG. 41