РОССИЙСКАЯ ФЕДЕРАЦИЯ

(19) **RU** (11) **2 715 029**(13) **C2**

(51) МПК
H03G 3/30 (2006.01)

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК
H03G 3/3089 *(2019.08);* ***H03G 3/32*** *(2019.08);* ***H03G 5/165*** *(2019.08);* ***H03G 7/002*** *(2019.08);* ***H03G 7/007*** *(2019.08)*

(21)(22) Заявка: **2016119382, 17.03.2014**

(24) Дата начала отсчета срока действия патента:
17.03.2014

Дата регистрации:
21.02.2020

Приоритет(ы):
(30) Конвенционный приоритет:
26.03.2013 CN 201310100422.1;
11.04.2013 US 61/811,072
Номер и дата приоритета первоначальной заявки, из которой данная заявка выделена:
2015140903 26.03.2013

(43) Дата публикации заявки: **07.11.2018** Бюл. № **31**

(45) Опубликовано: **21.02.2020** Бюл. № **6**

Адрес для переписки:
123242, Москва, Кудринская площадь, 1, а/я 35, "Михайлюк, Сороколат и партнеры - патентные поверенные"

(72) Автор(ы):
ВАН Цзюнь (CN),
ЛУ Ле (CN),
СИФЕЛДТ Алан (US)

(73) Патентообладатель(и):
ДОЛБИ ЛАБОРАТОРИС ЛАЙСЭНЗИН КОРПОРЕЙШН (US)

(56) Список документов, цитированных в отчете о поиске: **US 6782361 B1, 24.08.2004. EP 517233 B1, 30.10.1996. RU 2417414 C2, 27.04.2011. US 2012/0294461 A1, 22.11.2012. US 4785418 A1, 15.11.1998. WO 2008/ 078232 A1, 03.07.2008.**

(54) КОНТРОЛЛЕР ВЫРАВНИВАТЕЛЯ ГРОМКОСТИ И СПОСОБ УПРАВЛЕНИЯ

(57) Реферат:

Изобретение относится к контроллеру выравнивателя громкости и способу звуковой классификации. Технический результат заключается в обеспечении регулировки в непрерывном режиме. В одном варианте осуществления контроллер выравнивателя громкости содержит классификатор звукового содержимого для идентификации типа содержимого звукового сигнала в реальном времени и регулирующий блок для регулировки выравнивателя громкости в непрерывном режиме в зависимости от идентифицированного типа содержимого. Регулирующий блок может выполняться с возможностью положительной корреляции коэффициента динамического усиления выравнивателя громкости с типами информативного содержимого звукового сигнала и отрицательной корреляции коэффициента динамического усиления выравнивателя громкости с типами мешающего содержимого звукового сигнала. 6 н. и 12 з.п. ф-лы, 2 табл., 41 ил.

RU 2715029 C2

RU 2715029 C2

RUSSIAN FEDERATION

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(19) **RU** (11) **2 715 029**(13) **C2**

(51) Int. Cl.
H03G 3/30 (2006.01)

(12) ABSTRACT OF INVENTION

(52) CPC
H03G 3/3089 *(2019.08);* ***H03G 3/32*** *(2019.08);* ***H03G 5/165*** *(2019.08);* ***H03G 7/002*** *(2019.08);* ***H03G 7/007*** *(2019.08)*

(21)(22) Application: **2016119382, 17.03.2014**

(24) Effective date for property rights:
17.03.2014

Registration date:
21.02.2020

Priority:
(30) Convention priority:
26.03.2013 CN 201310100422.1;
11.04.2013 US 61/811,072
Number and date of priority of the initial application, from which the given application is allocated:
2015140903 26.03.2013

(43) Application published: **07.11.2018** Bull. № **31**

(45) Date of publication: **21.02.2020** Bull. № **6**

Mail address:
123242, Moskva, Kudrinskaya ploshchad, 1, a/ya 35, "Mikhajlyuk, Sorokolat i partnery - patentnye poverennye"

(72) Inventor(s):
WANG Jun (CN),
LU Lie (CN),
SEEFELDT Alan (US)

(73) Proprietor(s):
DOLBY LABORATORIES LICENSING CORPORATION (US)

RU 2715029 C2

(54) VOLUME EQUALIZER CONTROLLER AND CONTROL METHOD

(57) Abstract:

FIELD: monitoring devices.

SUBSTANCE: invention relates to a volume equalizer controller and a sound classification method. In one embodiment, the volume equalizer controller comprises an audio content classifier for identifying the type of content of the audio signal in real time and an adjustment unit for adjusting the volume equalizer in continuous mode depending on the identified type of content. Control unit may be configured to positively correlate the dynamic gain of the volume equalizer with the types of informative content of the audio signal and negatively correlate the dynamic gain of the volume equalizer with the types of interfering content of the audio signal.

EFFECT: technical result consists in providing adjustment in continuous mode.

18 cl, 2 tbl, 41 dwg

RU 2715029 C2

Перекрестная ссылка на родственные заявки

Данная заявка заявляет приоритет согласно заявке на патент Китая №201310100422.1, поданной 26 марта 2013 года, и предварительной заявке на патент США №61/811072, поданной 11 апреля 2013 года, каждая из которых в полном объеме включена в данную заявку посредством ссылки.

Область техники

Настоящее изобретение в целом относится к обработке звуковых сигналов. В частности, варианты осуществления настоящего изобретения относятся к устройствам и способам классификации и обработки звуковых сигналов, в особенности к управлению усилителем диалога, виртуализатором окружающего звука, выравнивателем громкости и эквалайзером.

Предпосылки создания изобретения

Некоторые устройства улучшения качества звука имеют обыкновение изменять звуковые сигналы либо во временной области, либо в спектральной области с целью улучшения общего качества звука и, соответственно, улучшения восприятия пользователем. Различные устройства улучшения качества звука были разработаны для различных целей. Некоторые типичные примеры устройств улучшения звука включают:

Усилитель диалога: Диалог является наиболее важным компонентом в кинофильме и радио- или телепрограмме для понимания сюжета. Были разработаны способы для усиления диалогов с целью повышения их ясности и разборчивости, в особенности для старых людей со сниженными слуховыми способностями.

Виртуализатор окружающего звука: Виртуализатор окружающего звука позволяет представить сигнал окружающего (многоканального) звука с помощью внутренних громкоговорителей ПК или с помощью наушников. То есть посредством стерео устройства (например, громкоговорителей и наушников) он создает эффект виртуального окружения и обеспечивает кинематографический опыт для потребителей.

Выравниватель громкости: Выравниватель громкости предназначен для настройки громкости звукового содержимого при воспроизведении и поддержании ее практически постоянной по оси времени в зависимости от целевой величины громкости.

Эквалайзер: Эквалайзер обеспечивает постоянство спектрального баланса, известного как "тон" или "тембр", и позволяет пользователям настраивать общий профиль (кривую или форму) частотной характеристики (усиления) в каждом отдельном диапазоне частот с целью подчеркивания определенных звуков или удаления нежелательных звуков. В традиционном эквалайзере для разных звуков, например, разных музыкальных жанров могут предусматриваться различные предустановки эквалайзера. После того, как предустановка выбрана, или набор параметров, определяющих частотную коррекцию, установлен, к сигналу будут применяться одни и те же коэффициенты усиления частотной коррекции до тех пор, пока параметры, определяющие частотную коррекцию, не изменить вручную. В отличие от этого, динамический эквалайзер обеспечивает постоянство спектральный баланса посредством постоянного контроля спектрального баланса звукового сигнала, сравнивая его с желаемым тоном, и динамической регулировки выравнивающего фильтра для преобразования исходного тона звукового сигнала в желаемый тон.

В целом, устройства улучшения качества звука имеют свой собственный сценарий/контекст применения. То есть, устройства улучшения качества звука могут быть предназначены только для определенного набора содержимого, но не для всех возможных звуковых сигналов, так как различное содержимое может нуждаться в

обработке разными способами. Например, способ усиления диалога обычно применяется к содержимому кинофильма. Если он применяется к музыке, в которой нет диалогов, он может ложно повысить некоторые частотные поддиапазоны и ввести сильные изменения тембра и несоответственность восприятия. Точно так же, если способ подавления шума применяется к музыкальным сигналам, будут слышны сильные искажения.

Тем не менее, для системы обработки звукового сигнала, которая содержит серию устройств улучшения звука, ее входным сигналом могут быть неизбежно все возможные типы звуковых сигналов. Например, система обработки звукового сигнала, встроенная в ПК, получит звуковое содержимое из различных источников, включая кино, музыку, VoIP и игру. Таким образом, становится важной идентификация или дифференциация обрабатываемого содержимого для применения более лучших алгоритмов или более лучших параметров каждого алгоритма к соответствующему содержимому.

С целью разграничения звукового содержимого и применения лучших параметров или лучших алгоритмов улучшения качества звука соответственно в традиционных системах обычно предварительно проектируется набор предустановок, а пользователей просят выбрать предустановку воспроизводимого содержимого. Предустановка обычно кодирует набор алгоритмов улучшения качества звука и/или их наилучшие параметры, которые будут применяться, например, предустановка "Кинофильм" и предустановка "Музыка", которые специально предназначены для воспроизведения кинофильмов или музыки.

Тем не менее, ручной выбор неудобен для пользователей. Люди обычно не часто переключают между предварительно определенными перестановками, а продолжают использовать одну предустановку для всего содержимого. Кроме того, даже в некоторых автоматических решениях параметры или алгоритмы настройки в предустановках обычно являются дискретными (например, включение или выключение конкретного алгоритма в отношении конкретного содержимого), она не может регулировать параметры в непрерывном режиме.

Сущность изобретения

Первый аспект настоящего изобретения состоит в том, чтобы автоматически настраивать устройства улучшения качества звука в непрерывном режиме в зависимости от воспроизводимого звукового содержимого. С помощью данного "автоматического" режима пользователи могут просто наслаждаться содержимым, не утруждая себя выбором разных предустановок. С другой стороны, непрерывная настройка является более важной для предотвращения слышимых искажений в точках переключения.

В соответствии с вариантом осуществления первого аспекта устройство обработки звукового сигнала содержит звуковой классификатор сигнала для классификации звукового сигнала по меньшей мере по одному звуковому типу в реальном времени; устройство улучшения качества звука для улучшения восприятия аудиторией; и регулирующий блок для регулировки по меньшей мере одного параметра устройства улучшения качества звука в непрерывном режиме в зависимости от величины достоверности по меньшей мере одного звукового типа.

Устройство улучшения качества звука может быть или усилителем диалога, или виртуализатором окружающего звука, или выравнивателем громкости, или эквалайзером.

Соответственно, способ обработки звукового сигнала включает: классификацию звукового сигнала по меньшей мере по одному звуковому типу сигнала в реальном времени; и регулировку по меньшей мере одного параметра для улучшения качества

звука в непрерывном режиме в зависимости от величины достоверности по меньшей мере одного звукового типа.

Согласно другому варианту осуществления первого аспекта контроллер выравнивателя громкости содержит классификатор звукового содержимого для идентификации типа содержимого звукового сигнала в реальном времени; и регулирующий блок для регулировки выравнивателя громкости в непрерывном режиме в зависимости от идентифицированного типа содержимого. Регулирующий блок может выполняться с возможностью положительной корреляции коэффициента динамического усиления выравнивателя громкости с типами информативного содержимого звукового сигнала и отрицательной корреляции коэффициента динамического усиления выравнивателя громкости с типами мешающего содержимого звукового сигнала.

Также описано устройство обработки звукового сигнала, содержащее контроллер выравнивателя громкости, указанный выше.

Соответственно, способ управления выравнивателем громкости включает: идентификацию типа содержимого звукового сигнала в реальном времени; и регулировку выравнивателя громкости в непрерывном режиме в зависимости от идентифицированного типа содержимого посредством положительной корреляции коэффициента динамического усиления выравнивателя громкости с типами информативного содержимого звукового сигнала и отрицательной корреляции коэффициента динамического усиления выравнивателя громкости с типами мешающего содержимого звукового сигнала.

Согласно еще одному варианту осуществления первого аспекта контроллер эквалайзера содержит звуковой классификатор для идентификации звукового типа звукового сигнала в реальном времени; и регулирующий блок для регулировки эквалайзера в непрерывном режиме в зависимости от величины достоверности идентифицированного звукового типа.

Также описано устройство обработки звукового сигнала, содержащее контроллер эквалайзера, указанный выше.

Соответственно, способ управления эквалайзером включает: идентификацию звукового типа звукового сигнала в реальном времени; и регулировку эквалайзера в непрерывном режиме в зависимости от величины достоверности идентифицированного звукового типа.

В настоящем изобретении также описан машиночитаемый носитель, содержащий записанные на нем команды компьютерной программы, которые при выполнении их процессором обеспечивают процессору возможность осуществлять вышеупомянутый способ обработки звукового сигнала, или способ управления выравнивателем громкости, или способ управления эквалайзером.

В соответствии с вариантами осуществления первого аспекта устройство улучшения качества звука, которое может быть или усилителем диалога, или виртуализатором окружающего звука, или выравнивателем громкости, или эквалайзером, может непрерывно регулироваться в зависимости от типа звукового сигнала и/или величины достоверности типа.

Второй аспект настоящего изобретения состоит в том, чтобы разработать компонент идентификации содержимого, чтобы идентифицировать несколько звуковых типов, и выявленные результаты могут применяться для управления/ руководства характеристиками различных устройств улучшения качества звука посредством нахождения в непрерывном режиме лучших параметров.

В соответствии с вариантом осуществления второго аспекта, звуковой классификатор

содержит: извлекатель кратковременных признаков объекта для извлечения кратковременных признаков объекта из кратковременных звуковых сегментов, каждый из которых содержит последовательность звуковых кадров; кратковременный классификатор для классификации последовательности кратковременных сегментов в долговременном звуковом сегменте по кратковременным звуковым типам, используя соответствующие кратковременные признаки объекта; извлекатель статистических данных для расчета статистических данных результатов кратковременного классификатора в отношении последовательности кратковременных сегментов в долговременном звуковом сегменте в качестве долговременных признаков объекта; и долговременный классификатор, использующий долговременные признаки объекта, для классификации долговременного звукового сегмента по долговременным звуковым типам.

Также описано устройство обработки звукового сигнала, содержащее звуковой классификатор, указанный выше.

Соответственно, способ звуковой классификации включает: извлечение кратковременных признаков объекта из кратковременных звуковых сегментов, каждый из которых содержит последовательность звуковых кадров; классификацию последовательности кратковременных сегментов в долговременном звуковом сегменте по кратковременным звуковым типам, используя соответствующие кратковременные признаки объекта; расчет статистических данных результатов операции классификации в отношении последовательности кратковременных сегментов в долговременном звуковом сегменте долговременных признаков объекта; и классификацию долговременного звукового сегмента по долговременным звуковым типам с использованием долговременных признаков объекта.

Согласно другому варианту осуществления второго аспекта звуковой классификатор содержит: классификатор звукового содержимого для идентификации типа содержимого кратковременного сегмента звукового сигнала; и классификатор звукового контекста для определения типа контекста кратковременного сегмента в зависимости, по меньшей мере частично, от идентифицированного типа содержимого посредством классификатора звукового содержимого.

Также описано устройство обработки звукового сигнала, содержащее звуковой классификатор, указанный выше.

Соответственно, способ звуковой классификации включает: идентификацию типа содержимого кратковременного сегмента звукового сигнала; и идентификацию типа контекста кратковременного сегмента в зависимости, по меньшей мере частично, от идентифицированного типа содержимого.

Настоящее изобретение также предлагает машиночитаемый носитель, содержащий команды компьютерной программы, записанные на нем, которые при выполнении их процессором позволяют процессору осуществлять вышеупомянутые способы звуковой классификации.

В соответствии с вариантами осуществления второго аспекта звуковой сигнал может классифицироваться по разным долговременным типам или типам контекста, которые отличаются от кратковременных типов или типов содержимого. Типы звукового сигнала и/или величина достоверности типов могут дополнительно использоваться для регулировки устройства улучшения качества звука, такого как усилитель диалога, виртуализатор окружающего звука, выравниватель громкости или эквалайзер.

Краткое описание графического материала

Настоящее изобретение иллюстрируется в качестве примера, а не с целью

ограничения, фигурами прилагаемого графического материала, на которых подобные номера позиций относятся к подобным элементам, и на которых:

на фиг. 1 приведена схема, иллюстрирующая устройство обработки звукового сигнала в соответствии с вариантом осуществления изобретения;

на фиг. 2 и 3 приведены схемы, иллюстрирующие разновидности варианта осуществления, показанного на фиг. 1;

на фиг. 4-6 приведены схемы, иллюстрирующие возможную конфигурацию классификаторов для идентификации нескольких звуковых типов и расчета величины достоверности;

на фиг. 7-9 приведены схемы, иллюстрирующие несколько вариантов осуществления устройства обработки звукового сигнала согласно настоящему изобретению;

на фиг. 10 приведена схема, иллюстрирующая задержку переключения между разными звуковыми типами;

на фиг. 11-14 приведены блок-схемы, иллюстрирующие способ обработки звукового сигнала в соответствии с вариантами осуществления настоящего изобретения;

на фиг. 15 приведена схема, иллюстрирующая контроллер усилителя диалога в соответствии с вариантом осуществления настоящего изобретения;

на фиг. 16 и 17 приведены блок-схемы, иллюстрирующие применение способа обработки звукового сигнала в соответствии с настоящим изобретением при управлении усилителем диалога;

на фиг. 18 приведена схема, иллюстрирующая контроллер виртуализатора окружающего звука в соответствии с вариантом осуществления настоящего изобретения;

на фиг. 19 приведена блок-схема, иллюстрирующая применение способа обработки звукового сигнала в соответствии с настоящим изобретением при управлении виртуализатором окружающего звука;

на фиг. 20 приведена схема, иллюстрирующая контроллер выравнивателя громкости в соответствии с вариантом осуществления настоящего изобретения;

на фиг. 21 приведена схема, иллюстрирующая результат применения контроллера выравнивателя громкости в соответствии с настоящим изобретением;

на фиг. 22 приведена схема, иллюстрирующая контроллер эквалайзера в соответствии с вариантом осуществления настоящего изобретения;

на фиг. 23 представлено несколько примеров предустановок желаемых спектральных балансов;

на фиг. 24 приведена схема, иллюстрирующая звуковой классификатор в соответствии с вариантом осуществления настоящего изобретения;

на фиг. 25 и 26 приведены схемы, иллюстрирующие некоторые признаки объекта для использования звуковым классификатором в соответствии с настоящим изобретением;

на фиг. 27-29 приведены схемы, иллюстрирующие дополнительное количество вариантов осуществления звукового классификатора в соответствии с настоящим изобретением;

на фиг. 30-33 приведены блок-схемы, иллюстрирующие способ звуковой классификации в соответствии с вариантами осуществления настоящего изобретения;

на фиг. 34 приведена схема, иллюстрирующая звуковой классификатор в соответствии с другим вариантом осуществления настоящего изобретения;

на фиг. 35 приведена схема, иллюстрирующая звуковой классификатор в соответствии с еще одним вариантом осуществления настоящего изобретения;

на фиг. 36 приведена схема, иллюстрирующая эвристические правила, применяемые

в звуковом классификаторе в соответствии с настоящим изобретением;

на фиг. 37 и 38 приведены схемы, иллюстрирующие дополнительное количество вариантов осуществления звукового классификатора в соответствии с настоящим изобретением;

на фиг. 39 и 40 приведены блок-схемы, иллюстрирующие способ звуковой классификации в соответствии с вариантами осуществления настоящего изобретения;

на фиг. 41 приведена структурная схема, иллюстрирующая примерную систему для реализации вариантов осуществления настоящего изобретения.

Подробное описание

Варианты осуществления настоящего изобретения описываются ниже со ссылкой на графический материал. Следует отметить, что для ясности, объяснения и описания, те компоненты и процессы, которые известны специалистам в данной области техники, но не обязательны для понимания настоящего изобретения, опущены в графическом материале и описании.

Как будет понятно специалисту в данной области техники, аспекты настоящего изобретения могут воплощаться в виде системы, устройства (например, сотового телефона, портативного мультимедийного проигрывателя, персонального компьютера, сервера, телевизионной приставки или цифрового видеомагнитофона, или любого другого мультимедийного проигрывателя), метода или компьютерного программного продукта. Соответственно, аспекты настоящего изобретения могут принимать форму аппаратного варианта осуществления, программного варианта осуществления (в том числе аппаратно-программного обеспечения, резидентного программного обеспечения, набора микрокоманд и т.д.) или варианта осуществления, сочетающего как программные, так и аппаратные аспекты, которые все могут, как правило, упоминаться в данной заявке как "схема", "модуль" или "система". Кроме того, аспекты настоящего изобретения могут принимать форму компьютерного программного продукта, воплощенного в одном или нескольких машиночитаемых носителях, содержащих машиночитаемый программный код, воплощенный на них.

Может быть использовано любое сочетание одного или нескольких машиночитаемых носителей. Машиночитаемый носитель может представлять собой машиночитаемый носитель сигнала или машиночитаемый носитель данных. Машиночитаемый носитель данных может представлять собой, например, электронную, магнитную, оптическую, электромагнитную, инфракрасную или полупроводниковую систему, приспособление, или устройство, или любое подходящее сочетание вышеперечисленного, но не ограничивается этим. Более конкретные примеры (не исчерпывающий список) машиночитаемых носителей данных включают следующее: электрическое соединение, содержащее один или несколько проводов, портативный компьютерный гибкий диск, жесткий диск, оперативное запоминающее устройство (RAM), постоянное запоминающее устройство (ROM), стираемое программируемое постоянное запоминающее устройство (EPROM или Flash-память), оптическое волокно, портативный компакт-диск для однократной записи данных (CD-ROM), оптическое устройство хранения данных, магнитное устройство хранения данных или любое подходящее сочетание вышеизложенного. В контексте данного документа машиночитаемый носитель данных может представлять собой любой материальный носитель, который может содержать или хранить программу для использования посредством или в соединении с системой выполнения команд, устройством или приспособлением.

Машиночитаемый носитель сигнала может включать распространяемый сигнал данных с машиночитаемым программным кодом, воплощенным в нем, например, в

основной полосе частот, либо как часть несущей волны. Такой распространяемый сигнал может принимать любую из множества форм, в том числе форму электромагнитного или оптического сигнала или любого подходящего сочетания, но не ограничивается этим.

Машиночитаемый носитель сигнала может представлять сбой любой машиночитаемый носитель, который не является машиночитаемым носителем данных, который может обмениваться информацией, распространять или передавать программу для использования посредством или в соединении с системой выполнения команд, устройством или приспособлением.

Программный код, воплощенный на машиночитаемом носителе, может быть передан с использованием любого подходящего носителя, включая беспроводную, проводную линию, оптоволоконный кабель, RF и т.д. или любое подходящее сочетание вышеперечисленного, но не ограничиваясь этим.

Компьютерный программный код для выполнения операции по аспектам настоящего изобретения может быть написан на любом сочетании одного или нескольких языков программирования, в том числе объектно-ориентированном языке программирования, таком как Java, Smalltalk, C++ и т.п., и обычных процедурных языках программирования, таких как язык программирования "C" или подобные языки программирования. Программный код может выполняться полностью на компьютере пользователя в виде отдельного пакета автономного программного обеспечения, или частично на компьютере пользователя и частично на удаленном компьютере, или полностью на удаленном компьютере или сервере. В последнем сценарии удаленный компьютер может быть подключен к компьютеру пользователя посредством сети любого типа, в том числе локальной вычислительной сети (LAN) или глобальной вычислительной сети (WAN), или подключение может быть сделано к внешнему компьютеру (например, через Интернет с использованием поставщика услуг сети Интернет).

Аспекты настоящего изобретения описаны ниже со ссылкой на блок-схемы и/или структурные схемы способов, устройств (систем) и компьютерных программных продуктов в соответствии с вариантами осуществления настоящего изобретения. Следует понимать, что каждый блок изображений блок-схемы и/или структурных схем и сочетание блоков в изображениях блок-схем и/или структурных схем может реализовываться командами компьютерной программы. Эти команды компьютерной программы могут предусматриваться в процессоре компьютера общего назначения, специализированном компьютере или другом программируемом устройстве обработки данных для изготовления машины, так что команды, которые выполняются посредством процессора компьютера или другого программируемого устройства обработки данных, создают средства для реализации функций/действий, указанных в блоке или блоках блок-схемы и/или структурной схемы.

Данные команды компьютерной программы могут также храниться на машиночитаемом носителе, которые могут управлять компьютером, другим программируемым устройством обработки данных или другими устройствами для функционирования определенным образом, чтобы команды, сохраненные на машиночитаемом носителе производили готовое изделие, в том числе команды, реализующие функцию/действие, указанное в блоке или блоках блок-схемы и/или структурной схемы.

Команды компьютерной программы также могут загружаться в компьютер, другое программируемое устройство обработки данных или другие устройства, чтобы вызвать серию рабочих операций, подлежащих выполнению на компьютере, другом

программируемом устройстве или других устройствах для выполнения компьютерно-реализуемого процесса таким образом, чтобы команды, которые выполняются на компьютере или другом программируемом устройстве предусматривали процессы для реализации функций/действий, указанных в блоке или блоках блок-схемы и/или структурной схемы.

Ниже подробно будут описаны варианты осуществления настоящего изобретения. Для ясности описание организовано по следующей структуре:

Часть 1: Устройство и способы обработки звукового сигнала
Раздел 1.1 Звуковые типы
Раздел 1.2 Величины достоверности звуковых типов и конфигурация классификаторов
Раздел 1.3 Сглаживание величин достоверности звуковых типов
Раздел 1.4 Регулировка параметров
Раздел 1.5 Сглаживание параметров
Раздел 1.6 Переключение звуковых типов
Раздел 1.7 Сочетание вариантов осуществления и сценариев применения
Разделе 1.8 Способ обработки звукового сигнала
Часть 2: Контроллер усилителя диалога и способ управления
Раздел 2.1 Уровень усиления диалога
Раздел 2.2 Пороговые значения для определения диапазонов частот для усиления
Раздел 2.3 Регулировка уровня фона
Раздел 2.4 Сочетание вариантов осуществления и сценариев применения
Раздел 2.5 Способ управления усилителем диалога
Часть 3: Контроллер виртуализатора окружающего звука и способ управления
Раздел 3.1 Коэффициент повышения окружающего звука
Раздел 3.2 Начальная частота
Раздел 3.3 Сочетание вариантов осуществления и сценариев применения
Раздел 3.4 Способ управления виртуализатором окружающего звука
Часть 4: Контроллер выравнивателя громкости и способ управления
Раздел 4.1 Типы информативного и мешающего содержимого
Раздел 4.2 Типы содержимого в различных контекстах
Раздел 4.3 Типы контекста
Раздел 4.4 Сочетание вариантов осуществления и сценариев применения
Раздел 4.5 Способ управления выравнивателем громкости
Часть 5: Контроллер эквалайзера и способ управления
Раздел 5.1 Управление в зависимости от типа содержимого
Раздел 5.2 Вероятность преобладающих источников в музыке
Раздел 5.3 Предустановки эквалайзера
Раздел 5.4 Управление в зависимости от типа контекста
Раздел 5.5 Сочетание вариантов осуществления и сценариев применения
Раздел 5.6 Способ управления эквалайзером
Часть 6: Звуковой классификатор и способы классификации
Раздел 6.1 Классификатор контекста на основе классификации типа содержимого
Раздел 6.2 Извлечение долговременных признаков объекта
Раздел 6.3 Извлечение кратковременных признаков объекта
Раздел 6.4 Сочетание вариантов осуществления и сценариев применения
Раздел 6.5 Способы звуковой классификации
Часть 7: VoIP классификаторы и способы классификации
Раздел 7.1 Классификация контекста на основе кратковременного сегмента

Раздел 7.2 Классификация с применением VoIP-речи и VoIP-шума

Раздел 7.3 Сглаживание флуктуаций

Раздел 7.4 Сочетание вариантов осуществления и сценариев применения

Раздел 7.5 Способы VoIP классификации

Часть 1: Устройство и способы обработки звукового сигнала

На фиг. 1 показана общая структура адаптивного к содержимому устройства 100 обработки звукового сигнала, которое поддерживает автоматическую настройку по меньшей мере одного устройства 400 улучшения качества звука с улучшенными параметрами в зависимости от звукового содержимого при воспроизведении. Оно включает три основных компонента: звуковой классификатор 200, регулирующий блок 300 и устройство 400 улучшения качества звука.

Звуковой классификатор 200 предназначен для классификации звукового сигнала по меньшей мере по одному звуковому типу в реальном времени. Он автоматически идентифицирует звуковой тип содержимого при воспроизведении. Любые технологии звуковой классификации, такие как обработка транзитных сигналов, машинное обучение и распознавания образов, могут применяться для идентификации звукового содержимого. Величины достоверности, которые представляют вероятности звукового содержимого относительно набора предопределенных целевых звуковых типов, оцениваются в большинстве случаев одновременно.

Устройство 400 улучшения качества звука предназначено для улучшения восприятия аудитории посредством выполнения обработки звукового сигнала и подробно будет рассмотрено ниже.

Регулирующий блок 300 для регулирования по меньшей мере одного параметра устройства улучшения качества звука в непрерывном режиме в зависимости от величины достоверности по меньшей мере одного звукового типа. Он предназначен для управления характеристиками устройства 400 улучшения качества звука. Он оценивает наиболее подходящие параметры соответствующего устройства улучшения качества звука в зависимости от результатов, полученных от звукового классификатора 200.

В данном устройстве могут применяться различные устройства улучшения качества звука. На фиг. 2 показан пример системы, содержащей четыре устройства улучшения качества звука, в том числе усилитель 402 диалога (DE), виртуализатор 404 окружающего звука (SV), выравниватель 406 громкости (VL) и эквалайзер (EQ) 408. Каждое устройство улучшения качества звука может автоматически регулироваться в непрерывном режиме в зависимости от результатов (звуковых типов и/или величин достоверности), полученных в звуковом классификаторе 200.

Конечно, устройство обработки звукового сигнала не обязательно может содержать все виды устройств улучшения качества звука, а может содержать только одно или несколько из них. С другой стороны, устройства улучшения качества звука не ограничены этими устройствами, приведенными в настоящем описании, и могут включать больше видов устройств улучшения качества звука, которые также входят в объем настоящего изобретения. Кроме того, названия этих устройств улучшения качества звука, рассмотренных в настоящем описании, включая усилитель 402 диалога (DE), виртуализатор 404 окружающего звука (SV), выравниватель 406 громкости (VL) и эквалайзер (EQ) 408, не являются ограничением, и каждое из них может быть истолковано как охватывающее все другие устройства, реализующие те же или подобные функции.

1.1 Звуковые типы

Для надлежащего управления различными видами устройства улучшения качества

звука, настоящее изобретение также предусматривает новую структуру звуковых типов, хотя звуковые типы предшествующего уровня техники также применимы в данном изобретении.

В частности, звуковые типы из разных семантических уровней моделируются, включая звуковые элементы низкого уровня, представляющие основные компоненты в звуковых сигналах, и звуковые жанры высокого уровня, представляющие наиболее популярное звуковое содержимое в развлекательных приложениях реальной жизни пользователя. Предшествующим также может быть термин указанный как "тип содержимого". Основные типы звукового содержимого могут включать речь, музыку (в том числе песню), фоновые звуки (или звуковые эффекты) и шум.

Понятие речи и музыки не требует разъяснений. Шум в настоящей заявке означает физический шум, а не смысловой шум. Физический шум в настоящей заявке может включать шумы, например, от кондиционеров, и шумы, возникающие по техническим причинам, такие как розовые шумы, обусловленные трактом передачи сигнала. В противоположность этому, "фоновые звуки" в настоящей заявке представляют собой те звуковые эффекты, которые могут быть акустическими событиями, происходящими вокруг основной цели внимания слушателя. Например, в звуковом сигнале в телефонном разговоре, кроме голоса говорящего, могут быть некоторые другие нежелательные звуки, такие как голоса некоторых других лиц, не связанных с телефонным разговором, звуки клавиатуры, звуки шагов и так далее. Эти нежелательные звуки называют "фоновыми звуками", а не шумом. Другими словами, мы можем определить "фоновые звуки", как те звуки, которые не являются целью (или основной целью внимания слушателя), или даже являясь нежелательными, но все еще имеют некоторое смысловое значение; в то время как "шум" может быть определен как нежелательные звуки, кроме целевых звуков и фоновых звуков.

Иногда фоновые звуки в самом деле не являются "нежелательными", а создаются намеренно и несут некоторую полезную информацию, например, фоновые звуки в кинофильмах, телепрограмме или программе радиовещания. Таким образом, иногда они также могут называться "звуковыми эффектами". Далее в настоящем описании для краткости используется только термин "фоновые звуки", и он может в дальнейшем сокращаться как "фон".

Кроме того, музыка может дополнительно классифицироваться как музыка без преобладающих источников и музыка с преобладающими источниками. Если присутствует гораздо более сильный источник (голос или инструмент), чем другие источники в музыкальном произведении, его называют "музыкой с преобладающим источником"; в противном случае она называется "музыкой без преобладающего источника". Например, в полифонической музыке, сопровождаемой певческим голосом и различными инструментами, если она гармонически уравновешена, или энергии нескольких наиболее характерных источников сопоставимы друг с другом, она считается музыкой без преобладающего источника; в противоположность этому, если источник (например, голос) гораздо более сильный в то время, как другие гораздо более тихие, считается, что она содержит преобладающий источник. В качестве другого примера, особые или своеобразные инструментальные тона представляют собой "музыку с преобладающим источником".

Музыка может дополнительно классифицироваться на разные типы в зависимости от разных стандартов. Она может классифицироваться в зависимости от жанров музыки, таких как рок, джаз, рэп и фолк, но не ограничивается ими. Она также может классифицироваться в зависимости от инструментов, например, на вокальную музыку

и инструментальную музыку. Инструментальная музыка может включать различную музыку, исполняемую на различных инструментах, такую как фортепианная музыка и гитарная музыка. Другие примерные стандарты включают ритм, темп, тембр музыки и/или любые другие музыкальные атрибуты, таким образом, музыка может группироваться на основе подобия этих атрибутов. Например, в соответствии с тембром, вокальная музыка может классифицироваться как тенор, баритон, бас, сопрано, меццо-сопрано и альт.

Тип содержимого звукового сигнала может классифицироваться в отношении кратковременных звуковых сегментов, которые содержат множество кадров. Обычно звуковой кадр имеет длину несколько миллисекунд, например, 20 мс, а длина кратковременного сегмента звукового сигнала для классификации посредством звукового классификатора может иметь длину от нескольких сотен миллисекунд до нескольких секунд, например, 1 секунду.

Для управления устройством улучшения качества звука способом, адаптивным к содержимому, звуковой сигнал может классифицироваться в реальном времени. Для типа содержимого, указного выше, тип содержимого текущего кратковременного сегмента звукового сигнала представляет собой тип содержимого текущего звукового сигнала. Так как длина кратковременного звукового сегмента не такая большая, звуковой сигнал может делиться на не перекрывающиеся кратковременные звуковые сегменты, следующие один за другим. Тем не менее, кратковременные звуковые сегменты также могут выбираться непрерывно/полунепрерывно вдоль оси времени звукового сигнала. То есть, кратковременные звуковые сегменты могут выбираться с окном предопределенной длины (предполагаемой длины кратковременного звукового сегмента), движущимся вдоль оси времени звукового сигнала с размером шага в один или несколько кадров.

Звуковые жанры высокого уровня также могут указываться как "тип контекста", поскольку он указывает долговременный тип звукового сигнала, и может рассматриваться в качестве среды или контекста мгновенного звукового события, которое может классифицироваться по типам содержимого, как указано выше. В соответствии с настоящей заявкой тип контекста может включать большинство популярных звуковых приложений, таких как программный материал, подобный кинофильму, музыку (включая песню), игру и VoIP (голосовую связь по IP-протоколу).

Понятие музыки, игр и VoIP не требует разъяснений. Программный материал, подобный кинофильму, может включать кинофильм, телепрограмму, программу радиовещания или любой другой звуковой программный материал, подобный вышеуказанному. Основной характеристикой программного материала, подобного кинофильму, является смесь возможных речевых сигналов, музыки и различных видов фоновых звуков (звуковых эффектов).

Следует отметить, что как тип содержимого, так и тип контекста включает музыку (в том числе песню). В дальнейшем в настоящей заявке мы используем формулировки "кратковременная музыка" и "долговременная музыка", чтобы соответственно отличать их.

Для некоторых вариантов осуществления настоящего изобретения также предложены некоторые другие структуры типа контекста.

Например, звуковой сигнал может классифицироваться как звуковой сигнал высокого качества (такой как программный материал, подобный кинофильму, и музыкальный CD) или звуковой сигнал низкого качества (например, VoIP, низкоскоростное потоковое онлайн воспроизведение звука и материалы пользователей), которые могут в

совокупности именоваться "типами качества звука".

В качестве другого примера, звуковой сигнал может классифицироваться как VoIP или не VoIP, что может рассматриваться как преобразование упомянутой выше структуры из 4 типов контекста (VoIP, программный материал, подобный кинофильму, (долговременная) музыка и игра). С использованием VoIP или не VoIP-контекста звуковой сигнал может классифицироваться по типам, относящимся к VoIP-контенту, таким как VoIP-речь, не VoIP-речь, VoIP-шум и не VoIP-шум. Структура типов звукового содержимого VoIP особенно полезна для разграничения VoIP и не VoIP контекстов, так как контекст VoIP обычно является наиболее сложным сценарием применения выравнивателя громкости (одного вида устройства улучшения качества звука).

Обычно тип контекста звукового сигнала может классифицироваться в отношении долговременных звуковых сегментов дольше, чем кратковременных звуковых сегментов. Долговременный звуковой сегмент состоит из множества кадров в количестве большем, чем количество кадров в кратковременном звуковом сегменте. Долговременный звуковой сегмент может также состоять из множества кратковременный звуковых сегментов. В большинстве случаев долговременный звуковой сегмент может иметь длину порядка секунд, например, от нескольких секунд до нескольких десятков секунд, скажем 10 секунд.

Аналогичным образом для управления устройством улучшения качества звука адаптивным способом, звуковой сигнал может классифицироваться по типам контекста в реальном времени. Кроме того, тип контекста текущего долговременного сегмента звукового сигнала представляет собой тип контекста текущего звукового сигнала. Так как длина долгосрочного сегмента звукового сигнала относительно велика, звуковой сигнал может выбираться непрерывно/полу непрерывно вдоль оси времени звукового сигнала, чтобы избежать резкого изменения его типа контекста и, таким образом, резкого изменения рабочих параметров устройства(в) улучшения качества звука. То есть, долгосрочные звуковые сегменты могут выбираться посредством окна предопределенной длины (предполагаемой длины кратковременного звукового сегмента), движущимся вдоль оси времени звукового сигнала с размером шага в один или несколько кратковременных сегментов.

Выше описаны как тип содержимого, так и тип контекста. В вариантах осуществления настоящего изобретения регулирующий блок 300 может регулировать по меньшей мере один параметр устройства(в) улучшения качества звука в зависимости по меньшей мере от одного из различных типов содержимого и/или по меньшей мере от одного из различных типов контекста. Таким образом, как показано на фиг. 3, в варианте осуществления, показанном на фиг. 1, звуковой классификатор 200 может содержать либо классификатор 202 звукового содержимого, либо классификатор 204 звукового контекста, либо оба.

Выше уже упоминались различные звуковые типы, основанные на разных стандартах (например, для типов контекста), а также различные звуковые типы на разных иерархических уровнях (например, для типов содержимого). Тем не менее, в данной заявке стандарты и иерархические уровни предназначены только для удобства описания, и, безусловно, не для ограничения. Другими словами, в настоящей заявке любые два или несколько указанных выше звуковых типов могут идентифицироваться посредством звукового классификатора 200 одновременно и одновременно учитываться регулирующим блоком 300, как будет описано позже. Другими словами, все звуковые типы в разных иерархических уровнях могут быть параллельными, или находится на том же самом уровне.

1.2 Величина достоверности звуковых типов и конфигурация классификаторов

Звуковой классификатор 200 может выводить результаты жесткого решения, или регулирующий блок 300 может принимать во внимание результаты звукового классификатора 200 в качестве результатов жесткого решения. Для жесткого решения за звуковым сегментом даже могут закрепляться несколько звуковых типов. Например, звуковой сегмент может помечаться и как "речь", и как "кратковременная музыка", так как он может быть смесью сигнала речи и кратковременной музыки. Полученные метки могут использоваться непосредственно для управления устройством(ами) 400 улучшения качества звука. Простым примером является задействование усилителя 402 диалога при присутствии речи и его выключение при отсутствии речи. Тем не менее, этот способ принятия жесткого решения может внести некоторую неестественность в точках переключения от одного звукового типа к другому, если не применяется схема аккуратного сглаживания (которая будет рассмотрена позже).

Для того чтобы иметь большую гибкость и настраивать параметры устройств улучшающих качество звука в непрерывном режиме, может оцениваться величина достоверности каждого целевого звукового типа (мягкое решение). Величина достоверности представляет собой подобранный уровень между подлежащим идентификации звуковым содержимым и целевым звуковым типом со значениями от 0 до 1.

Как отмечалось ранее, многие методы классификации могут непосредственно выдавать величину достоверности. Величина достоверности также может быть рассчитана различными способами, которые могут рассматриваться как часть классификатора. Например, если звуковые модели обучены посредством некоторых вероятностных технологий моделирования, таких как модели смеси нормальных распределений (GMM), для представления величины достоверности может применяться апостериорная вероятность:

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

где x - часть звукового сегмента, c_i - целевой звуковой тип, N - число целевых звуковых типов, $p(x|c_i)$ - вероятность того, что звуковой сегмент x представляет собой звуковой тип c_i и $p(c_i|x)$ - соответствующая апостериорная вероятность.

С другой стороны, если звуковые модели обучены некоторым различающим методам, таким как метод опорных векторов (SVM) и adaBoost, то из сравнения моделей получаются только оценки (реальные значения). В этих случаях для отображения полученной оценки (теоретически от $-\infty$ до ∞) в виде расчетной достоверности $conf$ (от 0 до 1) обычно используется сигмоидальная функция:

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

где y - выходная оценка от SVM или AdaBoost, A и B - два параметра, которые должны оцениваться из набора данных обучения с применением некоторых хорошо известных технологий.

Для некоторых вариантов осуществления настоящего изобретения регулирующий блок 300 может использовать более двух типов содержимого и/или более двух типов контекста. Затем звуковой классификатор 202 должен идентифицировать более двух

типов содержимого и/или классификатор 204 звукового контекста должен идентифицировать более двух типов контекста. В такой ситуации либо классификатор 202 звукового содержимого, либо классификатор 204 звукового контекста может представлять собой группу классификаторов, организованных в виде определенной конфигурации.

Например, если регулирующему блоку 300 необходимы все четыре вида типов контекста: программный материал, подобный кинофильму, долговременная музыка, игра и VoIP, то классификатор 204 звукового контекста может иметь следующие различные конфигурации:

Во-первых, классификатор 204 звукового содержимого может содержать 6 взаимно-однозначных двоичных классификаторов (каждый классификатор отличает один целевой звуковой тип от другого целевого звукового типа), организованных, как показано на фиг. 4, 3 взаимно-однозначных двоичных классификатора (каждый классификатор отличает целевой звуковой тип от других), организованных, как показано на фиг. 5, и 4 взаимно-однозначных классификатора, организованных, как показано на фиг. 6. Также имеются другие конфигурации, такие как конфигурация разрешающего направленного ациклического графа (DDAG). Следует отметить, что на фиг. 4-6 и в соответствующем описании ниже, для краткости используется термин "кинофильм", а не "программный материал, подобный кинофильму".

Каждый двоичный классификатор даст оценку достоверности H(x) для своего выходного сигнала (x представляет собой звуковой сегмент). После того, как выходные сигналы каждого бинарного классификатора получены, мы должны отобразить их в виде конечных величин достоверности идентифицированных типов контекста.

В большинстве случаев, полагают, что звуковой сигнал должен классифицироваться по M типам контекста (M является положительным целым числом). Традиционная взаимно-однозначная конфигурация строит M(M-1)/2 классификаторов, где каждый обучается данными из двух классов, затем каждый взаимно-однозначный классификатор отдает один голос за предпочтительный класс, и окончательным результатом является класс с большинством голосов среди M(M-1)/2 классификаций классификаторов. В сравнении с традиционной взаимно-однозначной конфигурацией, иерархическая конфигурация на фиг. 4 также нуждается в построении M(M-1)/2 классификаторов. Однако, итерации тестирования могут быть сокращены до M-1, поскольку сегмент x будет определен как относящийся/не относящийся к соответствующему классу на каждом иерархическом уровне, а общее число уровней составляет M-1. Конечные величины достоверности для различных типов контекста могут рассчитываться по достоверности двоичной классификации $H_k(x)$, например (k=1, 2, …6, представляющие разные типы контекста):

$$C_{\text{Кино}} = \big(1 - H_1(x)\big) \cdot (1 - H_3(x)) \cdot (1 - H_6(x))$$

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

$$C_{\text{Музыка}} = H_1(x) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_5(x)\big) + H_3(x) \cdot \big(1 - H_1(x)\big) \cdot \big(1 - H_5(x)\big) + H_6(x) \cdot (1 - H_1(x)) \cdot (1 - H_3(x))$$

$$C_{\text{Игра}} = H_1(x) \cdot H_2(x) \cdot \big(1 - H_4(x)\big) + H_1(x) \cdot H_5(x) \cdot \big(1 - H_2(x)\big) + H_3(x) \cdot H_5(x) \cdot \big(1 - H_1(x)\big)$$

В конфигурации, показанной на фиг. 5, функция отображения результатов бинарной

классификации $H_k(x)$ в виде конечных величин достоверности может определяться как в следующем примере:

$$C_{\text{Кино}} = H_1(x)$$

$$C_{\text{Музыка}} = H_2(x) \cdot \big(1 - H_1(x)\big)$$

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

$$C_{\text{Игра}} = (1 - H_3(x)) \cdot \big(1 - H_2(x)\big) \cdot \big(1 - H_1(x)\big)$$

В конфигурации, показанной на фиг. 6, конечные величины достоверности могут быть равны соответствующим результатам двоичной классификации $H_k(x)$, или, если требуется, чтобы сумма величин достоверности для всех классов была равна 1, то конечные величины достоверности могут просто нормироваться в зависимости от расчетной $H_k(x)$:

$$C_{\text{Кино}} = H_1(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

$$C_{\text{Музыка}} = H_2(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

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

$$C_{\text{Игра}} = H_4(x)/(H_1(x) + H_2(x) + H_3(x) + H_4(x))$$

Один или несколько с максимальными величинами достоверности могут быть определены как окончательно идентифицированный класс.

Следует отметить, что в конфигурациях, показанных на фиг. 4-6, последовательность разных двоичных классификаторов является не обязательно такой, как показана, но могут быть и другие последовательности, которые могут выбираться с помощью ручного назначения или автоматического обучения согласно различным требованиям различных приложений.

Описания выше направлены на классификаторы 204 звукового контекста. Для классификаторов 202 звукового содержимого ситуация аналогична.

В альтернативном варианте либо классификатор 202 звукового содержимого, либо классификатор 204 звукового контекста может реализовываться в виде всего одного классификатора, идентифицирующего все типы содержимого/типы контекста одновременно, и выдавать соответствующие величины достоверности одновременно. Для этого существует много используемых методов.

С применением величины достоверности выходной сигнал звукового классификатора 200 может представляться в виде вектора, с каждой размерностью, представляющей величину достоверности каждого целевого звукового типа. Например, если целевые звуковые типы (речь, кратковременная музыка, шум, фон) последовательны, то примером выходного результата может быть (0,9, 0,5, 0,0, 0,0), указывая, что с достоверностью 90% звуковое содержимое является речью, а с 50% достоверностью звуковой сигнал является музыкой. Следует отметить, что сумма всех измерений в выходном векторе не обязательно должна быть равна единице (например, результаты на фиг. 6 не нужно нормализовать), это означает, что звуковой сигнал может представлять собой смесь сигналов речи и кратковременной музыки.

Позже в части 6 и части 7, будет подробно рассмотрена новая реализация классификации звукового контекста и классификации звукового содержимого.

1.3 Сглаживание величин достоверности звуковых типов

Факультативно, после того, как каждый звуковой сегмент классифицирован по

предопределенным звуковым типам, применяется дополнительный шаг для сглаживания результатов классификации вдоль оси времени, чтобы избежать резкого скачка от одного типа к другому и сделать более гладкой оценку параметров в устройствах улучшения качества звука. Например, длинный фрагмент классифицируется как фрагмент программного продукта, подобного кинофильму, за исключением только одного сегмента, классифицированного как VoIP, затем посредством сглаживания скачкообразное определение VoIP может быть пересмотрено на программный продукт, подобный кинофильму.

Таким образом, в разновидности варианта осуществления, как показано на фиг. 7, для каждого типа звукового сигнала дополнительно предусмотрен блок 712 сглаживания типа для сглаживания величины достоверности звукового сигнала в текущее время.

Общий метод сглаживания на основе средневзвешенного значения, например, вычисляет взвешенную сумму фактической величины достоверности в текущее время и сглаженную величину достоверности за предыдущее время, а именно:

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

где t представляет собой текущее время (текущий звуковой сегмент), t-1 представляет собой предыдущее время (предыдущий звуковой сегмент), β - весовой коэффициент, conf и smoothConf - величины достоверности до и после сглаживания соответственно.

С точки зрения величин достоверности результаты жесткого решения классификаторов могут также представляться величиной достоверности со значениями или 0, или 1. То есть, если целевой звуковой тип выбирается и назначается звуковому сегменту, то соответствующая достоверность равна 1; в противном случае, достоверность равна 0. Таким образом, даже если звуковой классификатор 200 не выдает величину достоверности, а только выдает жесткое решение в отношении звукового типа, все еще возможна плавная регулировка регулирующего блока 300 посредством операции сглаживания блока 712 сглаживания типа.

Алгоритм сглаживания может быть "асимметричным" с применением разных весовых коэффициентов сглаживания для разных случаев. Например, весовые коэффициенты для вычисления взвешенной суммы могут адаптивно изменяться в зависимости от величины достоверности звукового типа звукового сигнала. Величина достоверности текущего сегмента тем больше, чем больше его весовой коэффициент.

С другой точки зрения, весовые коэффициенты для расчета взвешенной суммы могут адаптивно изменяться в зависимости разных пар переключения от одного звукового типа к другому звуковому типу, особенно когда устройство(а) улучшения качества звука регулируется в зависимости от нескольких типов содержимого, идентифицированных звуковым классификатором 200, вместо того, чтобы регулироваться в зависимости от наличия или отсутствия одного типа содержимого. Например, для переключения от звукового типа, чаще встречаемого в определенном контексте, в другой звуковой тип, не так часто встречаемого в контексте, величина достоверности последнего может сглаживаться так, чтобы она не увеличивалась так быстро, потому что это может быть просто случайным прерыванием.

Еще одним фактором является изменение (увеличение или уменьшение) тенденции, включая изменение скорости. Предположим, что мы заботимся больше о времени ожидания, когда звуковой тип становится текущим (то есть, когда его величина достоверности увеличивается), мы можем разработать алгоритм сглаживания следующим образом:

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

Приведенная выше формула позволяет сглаженной величине достоверности быстро реагировать на текущее состояние, когда величина достоверности увеличивается, и медленно сглаживаться, когда величина достоверности уменьшается. Аналогичным образом могут легко создаваться разновидности функций сглаживания. Например, формула (4) может быть изменена таким образом, чтобы весовой коэффициент conf(t) становился больше, при conf(t)>=smoothConf(t-1). Фактически, в формуле (4) может считаться, что β=0, и весовой коэффициент conf(t) становится наибольшим, то есть равным 1.

С другой точки зрения, учитывая изменяющуюся тенденцию определенного звукового типа, это представляет собой просто конкретный пример, учитывающий разные пары переключения звуковых типов. Например, увеличение величины достоверности типа А может рассматриваться как переключение от не А к А, а уменьшение величины достоверности типа А может рассматриваться как переключение от А к не А.

1.4 Регулировка параметров

Регулирующий блок 300 предназначен для оценки или регулировки соответствующих параметров для устройств(а) 400 улучшения качества звука в зависимости от полученных результатов от звукового классификатора 200. Разные регулировочные алгоритмы могут предназначаться для разных устройств улучшения качества звука посредством применения либо типа содержимого, либо типа контекста, либо обоих для совместного решения. Например, с информацией о типе контекста, таком как программный материал, подобный кинофильму, и долгосрочная музыка, предустановки, как сказано выше, могут автоматически выбираться и применяться к соответствующему содержимому. С имеющейся информацией о типе содержимого параметры каждого устройства улучшения качества звука могут настраиваться более точно, как показано в последующих частях. Информация о типе содержимого и информация о контексте могут дополнительно совместно использоваться в регулирующем блоке 300, чтобы сбалансировать долговременную и кратковременную информацию. Конкретный регулирующий алгоритм для конкретного устройства улучшения качества звука может рассматриваться как отдельный регулирующий блок, или разные регулировочные алгоритмы могут рассматривать в совокупности как единый регулирующий блок.

То есть, регулирующий блок 300 может выполняться с возможностью регулировки по меньшей мере одного параметра устройства улучшения качества звука в зависимости от величины достоверности по меньшей мере одного типа содержимого и/или величины достоверности по меньшей мере одного типа контекста. Для конкретного устройства улучшения качества звука некоторые из звуковых типов являются информативными, и некоторые из звуковых типов являются мешающими. Соответственно, параметры конкретного устройства улучшения качества звука могут либо положительно, либо отрицательно коррелировать с величиной(ами) достоверности информативного звукового типа(ов) или мешающего звукового типа(ов). В данной заявке термин "положительно коррелируют" означает, что параметр увеличивается или уменьшается с увеличением или уменьшением величины достоверности звукового типа линейным образом или нелинейным образом. "Отрицательно коррелируют" означает, что параметр увеличивается или уменьшается, соответственно, с уменьшением или увеличением величины достоверности звукового типа линейным образом или нелинейным образом.

В данной заявке уменьшение и увеличение величины достоверности непосредственно

"предается" с параметрами для регулировки посредством положительной или отрицательной корреляции. В математике, например, корреляция или "передача" может воплощаться в виде линейной пропорции или обратной пропорции, операции плюс или минус (сложения или вычитания), операции умножения или деления или нелинейной функции. Все эти формы корреляции могут упоминаться как "передаточная функция". Чтобы определить увеличение или уменьшение величины достоверности, мы также можем сравнить настоящую величину достоверности или ее математическое преобразование с предыдущей величиной достоверности, или множеством изменений во времени величин достоверности, или их математическими преобразованиями. В контексте настоящего изобретения термин "сравнивать" означает либо сравнение с помощью операции вычитания, либо сравнение с помощью операции деления. Мы можем определить увеличение или уменьшение путем определения, является ли разность больше, чем 0, или является ли отношение больше, чем 1.

В конкретных реализациях мы можем непосредственно связать параметры с величинами достоверности, или их отношениями, или разностями посредством подходящего алгоритма (например, передаточной функцией), а для "внешнего наблюдателя" не является необходимым знать, увеличились ли или уменьшились ли конкретная величина достоверности и/или конкретный параметр. Некоторые конкретные примеры будут приведены в последующих частях 2-5 о конкретных устройствах улучшения качества звука.

Как указано в предыдущем разделе, в отношении того же звукового сегмента классификатор 200 может идентифицировать несколько звуковых типов с соответствующими величинами достоверности, величины достоверности которых не обязательно достигают 1, так как звуковой сегмент может содержать несколько компонентов одновременно, такие как музыка, и речь, и фоновые звуки. В такой ситуации параметры устройств улучшения качества звука должны быть сбалансированы между различными звуковыми типами. Например, регулирующий блок 300 может выполняться с возможностью учета по меньшей мере некоторых из множества звуковых типов посредством взвешивания величин достоверности по меньшей мере одного звукового типа в зависимости от важности по меньшей мере одного звукового типа. Таким образом, чем больше параметров влияют, тем более важным является конкретный звуковой тип.

Весовой коэффициент также может отражать информативное и мешающее воздействие звукового типа. Например, для мешающего звукового типа может быть задан отрицательный весовой коэффициент. Некоторые конкретные примеры будут приведены в последующих частях 2-5 о конкретных устройствах улучшения качества звука.

Обратите внимание, что в контексте настоящего изобретения термин "весовой коэффициент" имеет более широкий смысл, чем коэффициенты в многочлене. Кроме коэффициентов в многочлене он также может принимать форму экспоненты или показателя степени. Весовые коэффициенты могут или не могут быть нормализованы, если являются коэффициентами в многочлене. Вкратце, весовой коэффициент просто показывает, какое влияние имеет взвешенный объект на параметр, который следует регулировать.

В некоторых других вариантах осуществления для нескольких звуковых типов, содержащихся в том же звуковом сегменте, их величины достоверности могут преобразовываться в весовые коэффициенты посредством нормализации, затем окончательный параметр может определяться посредством вычисления суммы предустановленных значений параметров, предопределенных для каждого звукового

типа и взвешенных посредством весовых коэффициентов в зависимости от величин достоверности. То есть, регулирующий блок 300 может выполняться с возможностью учета нескольких звуковых типов посредством взвешивания воздействия нескольких звуковых типов в зависимости от величин достоверности.

В качестве конкретного примера взвешивания регулирующий блок выполнен с возможностью учета по меньшей мере одного преобладающего звукового типа в зависимости от величин достоверности. Звуковые типы, имеющие слишком низкие величины достоверности (меньше, чем пороговое значение), могут не учитываться. Это эквивалентно тому, что весовые коэффициенты других звуковых типов, величины достоверности которых меньше порогового значения, устанавливаются равными нулю. Некоторые конкретные примеры будут приведены в последующих частях 2-5 о конкретных устройствах улучшения качества звука.

Тип содержимого и тип контекста могут учитываться вместе. В одном варианте осуществления, они могут учитываться на том же уровне, а их величины достоверности могут иметь соответствующие весовые коэффициенты. В другом варианте осуществления, только как показывает указание, "тип контекста" представляет собой контекст или среду, где находится "тип содержимого", и, таким образом, регулирующий блок 200 может выполняться таким образом, что типу содержимого в звуковом сигнале другого типа контекста назначается разный вес в зависимости от типа контекста звукового сигнала. Вообще говоря, любой звуковой тип может представлять собой контекст другого звукового типа и, следовательно, регулирующий блок 200 может выполняться с возможностью изменения весового коэффициента одного звукового типа с величиной достоверности другого звукового типа. Некоторые конкретные примеры будут приведены в последующих частях 2-5 о конкретных устройствах улучшения качества звука.

В контексте настоящего изобретения термин "параметр" имеет более широкий смысл, чем его буквальное значение. Кроме параметра, имеющего одну величину, он может также означать предустановку, как упоминалось ранее, включая набор разных параметров, вектор, состоящий из разных параметров, или конфигурацию параметров. В частности, в последующих частях 2-5 будут рассмотрены следующие параметры, но настоящая заявка не ограничивается ими: уровень усиления диалога, пороговые значения для определения диапазонов частот для усиления диалога, уровень фона, коэффициент повышения окружающего звука, начальную частоту для виртуализатора окружающего звука, коэффициент динамического усиления или диапазон коэффициента динамического усиления выравнивателя громкости, параметры, указывающие степень звукового сигнала нового воспринимаемого звукового события, уровень частотной коррекции, конфигурации частотной коррекции и предустановки спектрального баланса.

1.5 Сглаживание параметров

В разделе 1.3 мы обсуждали сглаживание величин достоверности звукового типа, чтобы избежать его резкого изменения и, таким образом, избежать резкого изменения параметров устройства(в) улучшения качества звука. Другие меры также возможны. Одна из них состоит в сглаживании параметра, регулируемого в зависимости от звукового типа, и будет обсуждаться в этом разделе; другая состоит в выполнении звукового классификатора и/или регулирующего блока с возможностью задержки изменения результатов звукового классификатора и будет обсуждаться в разделе 1.6.

В одном варианте осуществления параметр может дополнительно сглаживаться, чтобы избежать быстрого изменения, которые могут вносить слышимые искажения в точках переключения, в виде

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

где $\tilde{L}(t)$ - сглаженный параметр, L(t) - несглаженный параметр, τ - коэффициент, представляющий собой постоянную времени, t - текущее время и t-1 - предыдущее время.

То есть, как показано на фиг. 8, устройство обработки звукового сигнала для параметра устройства улучшения качества звука (например, по меньшей мере одного усилителя 402 диалога, виртуализатора 404 окружающего звука, выравнивателя 406 громкости и эквалайзера 408), регулируемого регулирующим блоком 300, может содержать блок 814 сглаживания параметра для сглаживания значения параметра, определенного в текущее время регулирующим блоком 300 посредством расчета взвешенной суммы значения параметра, определенного регулирующим блоком в текущее время, и сглаженного значения параметра предыдущего времени.

Постоянная времени τ может иметь фиксированное значение в зависимости от конкретных требований применения и/или реализации устройства 400 улучшения качества звука. Она также может адаптивно изменяться в зависимости от звукового типа, в частности, в зависимости от различных типов переключения от одного звукового типа к другому, например, от музыки к речи и от речи к музыке.

Рассмотрим в качестве примера эквалайзер (дополнительные пояснения могут ссылаться на часть 5). Частотная коррекция хорошо применяется к музыкальному содержимому, но не к речевому содержимому. Таким образом, для сглаживания уровня частотной коррекции постоянная времени может быть относительно небольшой, когда звуковой сигнал переходит от музыки к речи, чтобы меньший уровень частотной коррекции мог применяться к речевому содержимому быстрее. С другой стороны, постоянная времени для перехода от речи к музыке может быть относительно большой для того, чтобы избежать слышимых искажений в точках переключения.

Чтобы оценить тип переключения (например, от речи к музыке или от музыки к речи), результаты классификации содержимого могут применяться непосредственно. То есть, классификация звукового содержимого на либо музыку, либо на речь делает его эффективным для получения типа переключения. Чтобы предварительно рассчитать переключение более непрерывным способом, мы также можем исходить из расчетного уровня несглаженной частотной коррекции, вместо непосредственного сравнения жестких решений звуковых типов. Общей идей является то, что если уровень несглаженной частотной коррекции увеличивается, это указывает на переключение от речи к музыке (или большему подобию музыки); в противном случае, это больше похоже на переключение от музыки к речи (или большему подобию речи). Посредством различия разных типов переключения соответственно может устанавливаться постоянная времени, одним из примеров является:

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

где τ(t) - постоянная времени, изменяющаяся во времени в зависимости от содержимого, τ1 и τ2 - два предустановленных значения постоянных времени, обычно удовлетворяющие условию τ1>τ2. Интуитивно, приведенная выше функция задает относительно медленный переход, когда уровень частотной коррекции увеличивается, и относительно быстрый переход, когда уровень частотной коррекции уменьшается,

но настоящее изобретение не ограничивается этим. Кроме того, параметр не ограничивается уровнем частотной коррекции, а могут быть и другие параметры. То есть, блок 814 сглаживания параметра может быть выполнен таким образом, чтобы весовые коэффициенты для расчета взвешенной суммы адаптивно изменялись в зависимости от тенденции к повышению или снижению значения параметра, определенного регулирующим блоком 300.

1.6 Переключение звуковых типов

Со ссылкой на фиг. 9 и 10 будет описана другая схема для предотвращения резкого изменения звукового типа, и, таким образом, предотвращения резкого изменения параметров устройства(в) улучшения качества звука.

Как показано на фиг. 9, устройство 100 обработки звукового сигнала может дополнительно содержать таймер 916 для измерения времени неизменности, в течение которого звуковой классификатор 200 непрерывно выдает тот же самый новый звуковой тип, в котором регулирующий блок 300 может быть выполнен с возможностью продолжения использования текущего звукового типа до тех пор, пока продолжительность времени неизменности нового звукового типа не достигает порогового значения.

Другими словами, вводится фаза наблюдения (или выдержки), как показано на фиг. 10. Благодаря фазе наблюдения (в соответствии с пороговым значением продолжительности времени неизменности) изменение звукового типа дополнительно контролируется на последовательном отрезке времени, чтобы проверить, действительно ли звуковой тип изменился, действительно ли используется перед регулирующим блоком 300 новый звуковой тип.

Как показано на фиг. 10, стрелка (1) показывает ситуацию, когда текущее состояние является типом A, и результат вычислений звукового классификатора 200 не изменился.

Если текущее состояние является типом A, а результат вычислений звукового классификатора 200 становится типом B, то таймер 916 начинает отсчет времени, или, как показано на фиг. 10, процесс входит в фазу наблюдения (стрелка (2)), и устанавливается начальное значение счетчика задержки cnt, указывающее продолжительность наблюдения (равное пороговому значению).

Затем, если звуковой классификатор 200 непрерывно выдает тип B, то cnt непрерывно уменьшается (стрелка (3)) до тех пор, пока cnt не станет равным 0 (то есть, продолжительность времени неизменности нового типа B достигает порогового значения), то регулирующий блок 300 может применять новый звуковой тип B (стрелка (4)), или, другими словами, только в настоящий момент звуковой тип может считаться в действительности изменившимся на тип B.

В противном случае, если до того, как cnt становится равным нулю (до того, как продолжительность времени неизменности достигает порогового значения) выходной сигнал звукового классификатора 200 возвращается к старому типу A, то фаза наблюдения прекращается, и регулирующий блок 300 по-прежнему применяет старый типа A (стрелка (5)).

Переход от типа B к типу A может быть аналогичен процессу, описанному выше.

В описанном выше процессе, пороговое значение (или счетчик задержки) может устанавливаться в зависимости от требований применения. Оно может быть предопределенным фиксированным значением. Также оно может быть установлено адаптивно. В одном варианте осуществления пороговое значение разное для разных пар переключения от одного звукового типа к другому звуковому типу. Например, при переключении от типа A к типу B пороговое значение может иметь первое значение;

и при переключении от типа B к типу A, пороговое значение может иметь второе значение.

В другом варианте осуществления счетчик задержки (пороговое значение) может отрицательно коррелировать с величиной достоверности нового звукового типа. Существует общее представление, что, если достоверность показывает нечеткость между двумя типами (например, когда величина достоверности составляет только приблизительно 0,5), длительность наблюдения должна быть продолжительной; в противном случае, длительность может быть относительно короткой. Следуя этому принципу, счетчик примерной блокировки может устанавливаться по следующей формуле,

$$HangCnt = C \cdot |0{,}5 - Conf| + D$$

где HangCnt - длительность блокировки или пороговое значение, C и D - два параметра, которые могут устанавливаться в зависимости требований к эксплуатации, обычно C является отрицательным, а D имеет положительное значение.

К тому же, таймер 916 (и, таким образом, процесс переключения, описанный выше) был описан выше как часть устройства обработки звукового сигнала, но как вешний для звукового классификатора 200.В некоторых других вариантах осуществления он может рассматриваться как часть звукового классификатора 200, так же, как описано в разделе 7.3.

1.7 Сочетание вариантов осуществления и сценариев применения

Все варианты осуществления и разновидности, которые обсуждались выше, могут реализовываться в любом их сочетании, а любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции, могут реализовываться как такие же или отдельные компоненты.

В частности, при описании вариантов осуществления и их вариаций, приведенных выше в данной заявке, опущены компоненты, имеющие ссылочные позиции аналогичные тем, которые уже описаны в предыдущих вариантах осуществления или разновидностях, а описаны только отличающиеся компоненты. В действительности, эти отличающиеся компоненты могут либо сочетаться с компонентами других вариантов осуществления или разновидностей, либо представлять собой отдельные решения. Например, любые два или более решений, описанные со ссылкой на фиг. 1-10, могут сочетаться друг с другом. В качестве наиболее полного решения устройство обработки звукового сигнала может содержать как классификатор 202 звукового содержимого, так и классификатор 204 звукового контента, а также блок 712 сглаживания типа, блок 814 сглаживания параметра и таймер 916.

Как упоминалось ранее, устройства 400 улучшения качества звука могут содержать усилитель 402 диалога, виртуализатор 404 окружающего звука, выравниватель 406 громкости и эквалайзер 408. Устройство 100 обработки звукового сигнала может содержать любой один или несколько из них с регулирующим блоком 300, приспособленным к ним. При задействовании нескольких устройств 400 улучшения качества звука регулирующее устройство 300 может рассматриваться, как содержащее нескольких подблоков 300A-300D (фиг. 15, 18, 20 и 22), характерных для соответствующего устройства 400 улучшения качества звука, или по-прежнему рассматриваться как один объединенный регулирующий блок. Конкретные для устройства улучшения качества звука регулирующий блок 300 вместе с звуковым классификатором 200, а также другими возможными компонентами могут рассматриваться в качестве контроллера специального устройства улучшения качества

звука, которое будет рассмотрено подробно в последующих частях 2-5.

Кроме того, устройства 400 улучшения качества звука не ограничиваются примерами, как уже упоминалось, и могут содержать любое другое устройство улучшение качества звука.

Кроме того, любые уже рассмотренные решения или любые их сочетания могут дополнительно объединяться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания. В частности, варианты осуществления звуковых классификаторов, как будет описано в части 6 и 7, могут применяться в устройстве обработки звукового сигнала.

1.8 Способ обработки звукового сигнала

В процессе описания устройства обработки звукового сигнала в вариантах осуществления, приведенных выше, также очевидным образом описываются некоторые процессы или способы. В дальнейшем в данной заявке краткий обзор этих методов дается без повторения некоторых подробностей, которые уже обсуждались выше, но следует отметить, что, хотя в процессе описания устройства обработки звукового сигнала описаны способы, способы не обязательно осваивают эти описанные компоненты или не обязательно осуществляются этими компонентами. Например, варианты осуществления устройства обработки звукового сигнала могут реализовываться частично или полностью посредством аппаратных средств и/или аппаратно-программных средств, хотя и возможно, что способ обработки, рассмотренный ниже, может реализовываться полностью с помощью компьютерной исполняемой программы, хотя способы могут также осваивать аппаратные средства и/или аппаратно-программные средства устройств обработки звукового сигнала.

Ниже со ссылкой на фиг. 11-14 будут описаны способы. Пожалуйста, обратите внимание, что в соответствии с потоковым свойством звукового сигнала повторяются различные операции при реализации способа в реальном времени, а разные операции являются необязательными в отношении того же звукового сегмента.

В одном варианте осуществления, как показано на фиг. 11, предусмотрен способ обработки звукового сигнала. Во-первых, звуковой сигнал для обработки классифицируется по меньшей мере по одному звуковому типу в реальном времени (операция 1102). В зависимости от величины достоверности по меньшей мере одного звукового типа по меньшей мере один параметр для улучшения качества звука может непрерывно регулироваться (операция 1104). Улучшение качества звука может представлять собой усиление диалога (операция 1106), виртуализацию окружающего звука (операция 1108), выравнивание громкости (1110) и/или частотную коррекцию (операция 1112). Соответственно, по меньшей мере один параметр может содержать по меньшей мере один параметр для по меньшей мере одного из: обработки усиления диалога, обработки виртуализации окружающего звука, обработки выравнивания громкости и обработки частотной коррекции.

В данном документе термины "в реальном времени" и "непрерывно" означают звуковой тип, и, таким образом, параметр будет изменяться в реальном времени с конкретным содержимым звукового сигнала, а термин "непрерывно" также означает, что регулировка является непрерывной регулировкой в зависимости от величины достоверности, а не скачкообразной или дискретной регулировкой.

Звуковой тип может включать тип содержимого и/или тип контекста. Соответственно, операция 1104 регулировки может быть выполнена с возможностью регулировки по меньшей мере одного параметра в зависимости от величины достоверности по меньшей мере одного типа содержимого и величины достоверности по меньшей мере одного

типа контекста. Тип содержимого может дополнительно включать по меньшей мере один из типов содержимого: кратковременную музыку, речь, фоновый звук и шум. Тип контекста может дополнительно включать по меньшей мере один из типов контекста: долгосрочную музыку, программный материал, подобный кинофильму, игру и VoIP.

Предложены также некоторые другие схемы типа контекста, такие как типы контекста близкие к VoIP, включающие VoIP и не VoIP, и типы качества звука, включающие звуковой сигнал высокого качества или звуковой сигнал низкого качества.

Кратковременная музыка может дополнительно классифицироваться по подтипам в соответствии с разными стандартами. В зависимости от наличия преобладающего источника она может содержать музыку без преобладающих источников и музыку с преобладающими источниками. Кроме того, кратковременная музыка может содержать по меньшей мере один кластер в зависимости от жанра, или по меньшей мере один кластер в зависимости от инструмента, или по меньшей мере один музыкальный кластер, классифицированный в зависимости от ритма, темпа, тембра музыки и/или любых других музыкальных атрибутов.

Когда как типы содержимого, так и типы контекста идентифицированы, значение типа содержимого может быть определено с помощью типа контекста, где находится тип содержимого. То есть, типу содержимого в звуковом сигнале разного типа контекста назначается разный весовой коэффициент в зависимости от типа контекста звукового сигнала. В общем, один звуковой тип может влиять или может быть предпосылкой другого звукового типа. Таким образом, операция 1104 регулировки может выполняться с возможностью изменения весового коэффициента одного звукового типа с величиной достоверности другого звукового типа.

Когда звуковой сигнал классифицируется по нескольким звуковым типам одновременно (то есть по отношению к тому же звуковому сегменту), операция регулировки 1104 может учитывать некоторые или все из идентифицированных звуковых типов для регулировки параметра(ов) для улучшения того звукового сегмента. Например, операция регулировки 1104 может быть выполнена с возможностью взвешивания величин достоверности по меньшей мере одного звукового типа в зависимости от важности по меньшей мере одного звукового типа. Или операция регулировки 1104 может быть выполнена с возможностью учета по меньшей мере некоторых звуковых типов посредством их взвешивания в зависимости от их величин достоверности. В частном случае операция регулировки 1104 может быть выполнена с возможностью учета по меньшей мере одного преобладающего звукового типа в зависимости от величин достоверности.

Для предотвращения скачкообразных изменений результатов, могут вводиться схемы сглаживания.

Значение регулируемого параметра может сглаживаться (операция 1214 на фиг. 12). Например, значение параметра, определяемое операцией 1104 регулировки в текущее время, может заменяться взвешенной суммой значений параметра, определенных посредством операции регулировки в текущее время и сглаженного значения параметра в предыдущее время. Таким образом, посредством итерационной операции сглаживания значение параметра сглаживается на линии времени.

Весовые коэффициенты для расчета взвешенной суммы могут адаптивно изменяться в зависимости от звукового типа звукового сигнала или в зависимости от различных пар переключения от одного звукового типа другому звуковому типу. Кроме того, весовые коэффициенты для вычисления взвешенной суммы адаптивно изменяются в зависимости от тенденции к увеличению или уменьшению значения параметра,

определенного с помощью операции регулировки.

Другая схема сглаживания показана на фиг. 13. То есть, способ для каждого звукового типа может дополнительно включать сглаживание величины достоверности звукового сигнала в текущее время путем расчета взвешенной суммы фактической величины достоверности в текущий момент и сглаженной величины достоверности прошедшего времени (операция 1303). По аналогии с операцией 1214 сглаживания параметра, весовые коэффициенты для расчета взвешенной суммы могут адаптивно изменяться в зависимости от величины достоверности звукового типа звукового сигнала или в зависимости от разных пар переключения от одного звукового типа к другому звуковому типу.

Другая схема сглаживания является буферным механизмом для задержки переключения от одного звукового типа к другому звуковому типу, даже если выходной сигнал операции 1102 звуковой классификации изменяется. То есть, операция 1104 регулировки не сразу использует новый звуковой тип, а ждет стабилизации на выходе операции 1102 звуковой классификации.

В частности, способ может включать измерение времени неизменности, в течение которого операция классификации непрерывно выводит тот же самый новый звуковой тип (операция 1403 на фиг. 14), причем операция 1104 регулировки выполнена с возможностью продолжения использования текущего звукового типа ("N" в операции 14035 и операции 11041) до тех пор, пока продолжительность времени неизменности нового звукового типа не достигает порогового значения ("Y" в операции 14035 и операции 11042). В частности, когда выходной сигнал звукового типа операции 1102 звуковой классификации изменяется в отношении текущего звукового типа, используемого в операции 1104 регулировки звукового параметра ("Y" в операции 14031), то начинается отсчет времени (операция 14032). Если операция 1102 звуковой классификации продолжает выводить новый звуковой тип, то есть, если решение в операции 14031 продолжает оставаться "Y", то отсчет времени продолжается (операция 14032). Наконец, когда время неизменности нового звукового типа достигает порогового значения ("Y" в операции 14035), операция 1104 регулировки применяет новый звуковой тип (операция 11042), а отчет времени сбрасывается (операция 14034) для подготовки к следующему переключению звукового типа. До достижения порогового значения ("N" в операции 14035), операция 1104 регулировки продолжает применять текущий звуковой тип (операция 11041).

Здесь отсчет времени может реализовываться посредством механизма таймера (прямой отсчет или обратный отсчет). Если после начала отчета времени, но до достижения порогового значения, выходная величина операции 1102 звуковой классификации возвращается к текущему звуковому типу, используемому в операции 1104 регулировки, следует считать, что нет никакого изменения ("N" в операции 14031) в отношении текущего звукового типа, используемого операцией 1104 регулировки. Но, если результат текущей классификации (соответствующий текущему звуковому сегменту для классификации в звуковом сигнале) изменяется в отношении предыдущей выходной величины (соответствующей предыдущему звуковому сегменту для классификации в звуковом сигнале) операции 1102 звуковой классификации ("Y" в операции 14033), то отсчет времени сбрасывается (операция 14034) до тех пор, пока следующее изменение ("Y" в операции 14031) не начнет отчет времени. Конечно, если результат классификации операции 1102 звуковой классификации не изменяется в отношении текущего звукового типа, используемого операцией 1104 регулировки звукового параметра ("N" в операции 14031), нет изменений в отношении предыдущей

классификации ("N" в операции 14033), это показывает, что звуковая классификация находится в устойчивом состоянии, и следует продолжать использовать текущий звуковой тип.

Пороговое значение, используемое здесь, также может быть разным для разных пар переключения от одного звукового типа другому звуковому типу, потому что, когда состояние не так стабильно, в большинстве случаев мы можем предпочитать, чтобы устройство улучшения качества звука находилось в состоянии по умолчанию, а не в другом. С другой стороны, если величина достоверности нового звукового типа является относительно высокой, надежнее переключиться к новому звуковому типу. Таким образом, пороговое значение может отрицательно коррелировать с величиной достоверности нового звукового типа. Чем выше величина достоверности, чем ниже пороговое значение, то есть звуковой тип может быстрее переключаться в новый звуковой тип.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, являются практически осуществимыми; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. В частности, во всех способах обработки звукового сигнала могут применяться способы звуковой классификации, как описано в частях 6 и 7.

Часть 2: Контроллер усилителя диалога и способ, управления

Одним из примеров устройства улучшения качества звука является усилитель диалога (DE), который предназначен для непрерывного контроля звуковоспроизведения, обнаружения наличия диалога и усиления диалога для увеличения его четкости и разборчивости (делания диалога легче слышимым и понятым), особенно для старших людей с уменьшенными возможностями слуха. Кроме обнаружения присутствия диалога также обнаруживаются частоты наиболее важные для разборчивости, если присутствует диалог, а затем, соответственно, усиливаются (посредством динамического спектрального повторного выравнивания). Пример способа усиления диалога представлен в документе H. Muesch. "Speech Enhancement in Entertainment Audio", опубликованном как WO 2008/106036 A2, который в полном объеме включен в данную заявку посредством ссылки.

Распространенной ручной настройкой в усилителе диалога является то, что он обычно включается для содержимого программного материала, подобного кинофильму, но отключается для музыкального содержимого, потому что усиление диалога может слишком часто ошибочно срабатывать на музыкальные сигналы.

С доступностью информации звукового типа уровень усиления диалога и других параметров может быть настроен в зависимости от величин достоверности идентифицированных звуковых типов. В качестве конкретного примера устройства и способа обработки звукового сигнала, рассмотренных ранее, усилитель диалога может использовать все варианты осуществления, рассмотренные в части 1, и любые сочетания этих вариантов осуществления. В частности, в случае управления усилителем диалога звуковой классификатор 200 и регулирующий блок 300 в устройстве 100 обработки звукового сигнала, как показано на фиг. 1-10, могут представлять собой контроллер 1500 усилителя диалога, как показано на фиг. 15. В данном варианте осуществления, поскольку регулировочный блок является специфическим для усилителя диалога, он может упоминаться как 300A. И, как описано в предыдущей части, звуковой классификатор 200 может содержать по меньшей мере одно из следующего:

классификатор 202 звукового содержимого и классификатор 204 звукового контекста, а контроллер 1500 усилителя диалога может дополнительно содержать по меньшей мере одно из следующего: блок 712 сглаживания типа, блок 814 сглаживания параметра и таймер 916.

Таким образом, в этой части мы не будем повторять содержимое, уже описанное в предыдущей части, а просто дадим некоторые его конкретные примеры.

Для усилителя диалога регулируемые параметры включают уровень усиления диалога, уровень фона и пороговые значения для определения диапазонов частот, чтобы обеспечить усиление, но не ограничиваются ими. См. документ H. Muesch. "Speech Enhancement in Entertainment Audio", опубликованный как WO 2008/106036 A2, который в полном объеме включен в данную заявку посредством ссылки.

2.1 Уровень усиления диалога

При использовании уровня усиления диалога регулирующий блок 300A может выполняться с возможностью положительной корреляции уровня усиления диалога усилителя диалога с величиной достоверности речи. В дополнение к этому или в альтернативном варианте уровень может отрицательно коррелировать с величиной достоверности других типов содержимого. Таким образом, уровень усиления диалога может устанавливаться пропорционально (линейно или нелинейно) достоверности речи, следовательно, усиление диалога является менее эффективным в неречевых сигналах, таких как музыка и фоновый звук (звуковые эффекты).

Что касается типа контекста, то регулирующий блок 300A может выполняться с возможностью положительной корреляции уровня усиления диалога усилителя диалога с величиной достоверности программного материала, подобного кинофильму, и/или VoIP и/или отрицательной корреляции уровня усиления диалога усилителя диалога с величиной достоверности долговременной музыки и/или игры. Например, уровень усиления диалога может устанавливаться пропорционально (линейно или нелинейно) величине достоверности программного материала, подобного кинофильму. Когда величина достоверности программного материала, подобного кинофильму, равна 0 (например, в музыкальном содержимом), уровень усиления диалога также равен 0, что эквивалентно отключению усилителя диалога.

Как описано в предыдущей части, тип содержимого и тип контекста могут учитываться совместно.

2.2 Пороговые значения для определения диапазонов частот для усиления

Во время работы усилителя диалога существует пороговое значение (обычно пороговое значение энергии или громкости) для каждого диапазона частот, определяющие необходимость усиления, то есть, те диапазоны частот, которые выше соответствующих пороговых значений энергии/громкости, будут усилены. Для регулировки пороговых значений регулирующий блок 300A может выполняться с возможностью положительной корреляции пороговых значений с величиной достоверности кратковременной музыки, и/или шума, и/или фоновых звуков и/или отрицательной корреляции пороговых значений с величиной достоверности речи. Например, пороговые значения могут уменьшаться, если достоверность речи высока, предполагая более надежное обнаружение речи, чтобы обеспечить возможность усиления большего числа диапазонов частот; с другой стороны, когда величина достоверности музыки высока, пороговые значения могут увеличиваться для обеспечения возможности усиления меньшего числа диапазонов частот (и, следовательно, меньшего количества искажений).

2.3 Регулировка уровня фона

Другим компонентом в усилителе диалога является блок 4022 отслеживания минимума, как показано на фиг. 15, который применяется для оценки уровня фона в звуковом сигнале (для оценки SNR и оценки порогового значения диапазона частот, как указано в разделе 2.2). Он может также настраиваться в зависимости от величин достоверности типов звукового содержимого. Например, если достоверность речи высока, то блок отслеживания минимума может более достоверно установить уровень фона на текущий минимум. Если достоверность музыки высока, то уровень фона может устанавливаться немного выше, чем тот текущий минимум, или по-другому, устанавливаться в виде средневзвешенного значения текущего минимума и энергии текущего кадра с большим весом текущего минимума. Если достоверность шума и фона высока, то уровень фона может устанавливаться значительно выше, чем текущее минимальное значение, или по-другому, устанавливаться на средневзвешенное значение текущего минимума и энергии текущего кадра с небольшим весовым коэффициентом текущего минимума.

Таким образом, регулирующий блок 300A может выполняться с возможностью назначения регулирования уровня фона, определенного блоком отслеживания минимума, при этом регулирующий блок дополнительно выполнен с возможностью положительной корреляции регулировки с величиной достоверности кратковременной музыки и/или шума, и/или фонового звука и/или отрицательной корреляции регулировки с величиной достоверности речи. Как вариант, регулирующий блок 300A может выполняться с возможностью более положительной корреляции регулировки с величиной достоверности шума и/или фона, чем кратковременной музыки.

2.4 Сочетание вариантов осуществления и сценариев применения

Аналогично части 1 все варианты осуществления и разновидности, рассмотренные выше, могут реализовываться в любом их сочетании, и любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющих одинаковые или подобные функции могут реализовываться как такие же или отдельные компоненты.

Например, любые два или более решений, описанных в разделах 2.1-2.3, могут сочетаться друг с другом. И эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в части 1 и в других частях, которые будут описаны позже. В частности, многие формулы фактически применимы к каждому виду устройства или способу улучшения качества звука, но они не обязательно изложены или описаны в каждой части данного описания. В такой ситуации перекрестные ссылки могут быть выполнены из частей данного описания для применения конкретной формулы, описанной в одной части, в другой части только со значимым параметром(ами), коэффициентом(ами), показателем(ями) степени (экспонентами) и весовым коэффициентом(ами), которые регулируются надлежащим образом в соответствии с требованиями конкретного применения.

2.5 Способ управления усилителем диалога

Аналогично части 1, в процессе описания контроллера усилителя диалога в вариантах осуществления, приведенных выше в данной заявке явно описываются также некоторые процессы или способы. Далее в данной заявке дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

Во-первых, варианты осуществления способа обработки звукового сигнала, как описывалось в части 1, могут применяться для усилителя диалога, параметр(ы) которого является одной из целей для регулировки посредством способа обработки звукового сигнала. С этой точки зрения, способ обработки звукового сигнала представляет собой также способ управления усилителем диалога.

В этом разделе, будут обсуждаться только аспекты, характерные для управления усилителем диалога. Для общих аспектов способа управления может быть сделана ссылка на часть 1.

Согласно одному варианту осуществления способ обработки звукового сигнала может дополнительно включать обработку усиления диалога, а операция 1104 регулировки включает положительную корреляцию уровня усиления диалога с величиной достоверности программного материала, подобного кинофильму, и/или VoIP, и или отрицательную корреляцию уровня усиления диалога с величиной достоверности долговременной музыки и/или игры. То есть, усиление диалога в основном направлено на звуковой сигнал в контексте программного материала, подобного кинофильму, или VoIP.

Более конкретно, операция 1104 регулировки может включать положительную корреляцию уровня усиления диалога усилителя диалога с величиной достоверности речи.

Настоящая заявка также может регулировать диапазоны частот, чтобы обеспечить усиление при обработке усиления диалога. Как показано на фиг. 16, пороговые значения (обычно энергии или громкости) для определения, должны ли усиливаться соответствующие диапазоны частот, могут регулироваться в зависимости от величины (н) достоверности идентифицированных звуковых типов (операция 1602) в соответствии с настоящей заявкой. Затем в усилителе диалога в зависимости от регулируемых пороговых значений выбираются диапазоны частот выше соответствующих пороговых значений (операция 1604) и усиливаются (операции 1606).

В частности, операция 1104 регулировки может включать положительную корреляцию пороговых значений с величиной достоверности кратковременной музыки, и/или шума, и/или фоновых звуков и/или отрицательную корреляцию пороговых значений с величиной достоверности речи.

Способ обработки звукового сигнала (особенно обработка усиления диалога) обычно дополнительно включает оценку уровня фона в звуковом сигнале, который обычно реализуется посредством блока 4022 отслеживания минимума, реализованным в усилителе 402 диалога и применяемом при оценке SNR или оценке порогового значения диапазона частот. Настоящая заявка также может применяться для регулировки уровня фона. В такой ситуации, после того, как оценивается уровень фона (операция 1702), он сначала регулируется в зависимости от величины(н) достоверности звукового типа(ов) (операция 1704), а затем применяется для оценки SNR и/или порогового значения диапазона частот (операция 1706). В частности, операция 1104 регулировки может выполняться с возможностью назначения регулировки расчетного уровня фона, при этом операция 1104 регулировки может дополнительно выполняться с возможностью положительной корреляции регулировки с величиной достоверности кратковременной музыки, и/или шума, и/или фонового звука и/или отрицательной корреляции регулировки с величиной достоверности речи.

Более конкретно, операция 1104 регулировки может выполняться с возможностью более положительной корреляции регулировки с величиной достоверности шума и/или фона, чем кратковременной музыки.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. Кроме того, любые два

или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в части 1 и других частях, которые будут описаны позже.

Часть 3: Контроллер виртуализатора окружающего звука и способ управления

Виртуализатор окружающего звука позволяет представлять окружающий звуковой сигнал (например, многоканальный 5.1 и 7.1) через встроенные громкоговорители ПК или через наушники. То есть, посредством стерео устройств, таких как внутренние громкоговорители или наушники ноутбука, он создает виртуальный эффект окружающего звучания и обеспечивает кинематографическое впечатление для потребителей. Передаточная функция головы (HRTF) обычно используется в виртуализаторе окружающего звука для имитации поступления звука в уши, идущего из различных местоположений громкоговорителей, связанных с многоканальным звуковым сигналом.

В то время как текущий виртуализатор окружающего звука хорошо работает для наушников, он работает по-разному с разным содержимым для встроенных громкоговорителей. В целом, программный материал, подобный кинофильму, предполагает возможность использования виртуализатора окружающего звука для громкоговорителей, в то время как музыка - нет, так как она может звучать на пределе.

Поскольку те же параметры в виртуализаторе окружающего звука не могут создавать хороший звуковой образ одновременно как для программного материала, подобного кинофильму, так как и для музыкального содержимого, параметры должны настраиваться более точно в зависимости от содержимого. С помощью предоставляемой информации о звуковом типе, особенно величины достоверности музыки и величины достоверности речи, а также какой-либо другой информации о типе содержимого и информации о контексте работа может быть выполнена с помощью настоящей заявки.

Аналогично части 2, в качестве конкретного примера устройства обработки звукового сигнала и способа, рассмотренных в части 1, виртуализатор 404 окружающего звука может использовать все варианты осуществления, описанные в части 1, и любые сочетания этих вариантов осуществления, описанные в данной заявке. В частности, в случае управления виртуализатором 404 окружающего звука звуковой классификатор 200 и регулирующий блок 300 в устройстве 100 обработки звукового сигнала, как показано на фиг. 1-10, могут представлять собой контроллер 1800 виртуализатора окружающего звука, как показано на фиг. 18. В этом варианте осуществления, поскольку регулирующий блок является специфическим для виртуализатора 404 окружающего звука, он может указываться как 300 В. И, аналогично части 2, звуковой классификатор 200 может содержать по меньшей мере одно из следующего: классификатор 202 звукового содержимого и классификатор 204 звукового контекста, а контроллер 1800 виртуализатора окружающего звука может дополнительно содержать по меньшей мере одно из следующего: блок 712 сглаживания типа, блок 814 сглаживания параметра и таймер 916.

Таким образом, в этой части мы не будем повторять содержимое, уже описанное в части 1, а просто приведем его конкретные примеры.

Для виртуализатора окружающего звука регулируемые параметры включают, но не ограничиваются, коэффициент повышения окружающего звука и начальную частоту для виртуализатора 404 окружающего звука.

3.1 Коэффициент повышения окружающего звука

При использовании коэффициента повышения окружающего звука, регулирующий

блок 300 B может выполняться с возможностью положительной корреляции коэффициента повышения окружающего звука виртуализатора 404 окружающего звука с величиной достоверности шума, и/или фона, и/или речи и/или отрицательной корреляции коэффициента повышения окружающего звука с величиной достоверности кратковременной музыки.

В частности, чтобы изменить виртуализатор 404 окружающего звука с целью приемлемого звучания музыки (тип содержимого), пример реализации регулирующего блока 300B мог бы настроить коэффициент повышения окружающего звука в зависимости от величины достоверности кратковременной музыки следующим образом:

$$SB \propto (1 - Conf_{музыка}) \qquad (5)$$

где SB указывает коэффициент повышения окружающего звука, $Conf_{музыка}$ - величина достоверности кратковременной музыки.

Это помогает уменьшить повышение окружающего звука для музыки и предотвратить размытие ее звучания.

Подобным образом величина достоверности речи тоже может использоваться, например:

$$SB \propto (1 - Conf_{музыка}) * Conf_{речь}^{\alpha} \qquad (6)$$

где $Conf_{речь}$ - величина достоверности речи, α - весовой коэффициент в виде показателя степени, который может находиться в диапазоне 1-2. Эта формула показывает, что коэффициент повышения окружающего звука будет высоким только для чистой речи (с высокой достоверностью речи и низкой достоверностью музыки).

Или мы можем учитывать только величину достоверности речи:

$$SB \propto Conf_{речь} \qquad (7)$$

Таким же образом могут быть разработаны различные варианты. В частности, для шума или фонового звука могут быть построены формулы, подобные формулам (5)-(7). Кроме того, воздействия четырех типов содержимого могут учитываться вместе в любом сочетании. В такой ситуации шум и фон представляют собой окружающие звуки, и они являются более безопасными, чтобы иметь большой коэффициент повышения; речь может иметь средний коэффициент повышения, предполагая, что говорящий обычно сидит перед экраном; и музыка применяет наименьший коэффициент повышения. Таким образом, регулирующий блок 300B может выполняться с возможностью более положительной корреляции коэффициента повышения окружающего звука с величиной достоверности шума и/или фона, чем тип содержимого речи.

Предположим, мы предопределили предполагаемый коэффициент повышения (который эквивалентен весовому коэффициенту) для каждого типа содержимого, тогда может применяться другой альтернативный вариант:

$$\hat{a} = \frac{a_{речь} \cdot Conf_{речь} + a_{музыка} \cdot Conf_{музыка} + a_{шум} \cdot Conf_{шум} + a_{фон} \cdot Conf_{фон}}{Conf_{речь} + Conf_{музыка} + Conf_{шум} + Conf_{фон}} \qquad (8)$$

где $\hat{a}$ - расчетный коэффициент повышения, □ с индексом типа содержимого - предполагаемый/предопределенный коэффициент (весовой коэффициент) типа содержимого, $Conf$ с индексом типа содержимого - величина достоверности типа

содержимого (где фон представляет собой "фоновый звук"). В зависимости от ситуации, $a_{\text{музыка}}$ может (но не обязательно) устанавливаться в 0, указывая, что виртуализатор 404 окружающего звука будет отключен для чистой музыки (чипа содержимого).

С другой точки зрения, □ с индексом типа содержимого в формуле (8) является предполагаемым/предопределенным коэффициентом повышения типа содержимого, а частное от деления величины достоверности соответствующего типа содержимого, деленной на сумму величин достоверности всех идентифицированных типов содержимого, можно рассматривать как нормализованный весовой коэффициент предопределенного/предполагаемого коэффициента повышения соответствующего типа содержимого. То есть, регулирующий блок 300B может выполняться с возможностью учета по меньшей мере некоторых из множества типов содержимого посредством взвешивания предопределенных коэффициентов повышения нескольких типов содержимого в зависимости от величин достоверности.

Касательно типа контекста регулирующий блок 300B может быть выполнен с возможностью положительной корреляции коэффициента повышения окружающего звука виртуализатора 404 окружающего звука с величиной достоверности программного материала, подобного кинофильму, и/или игры и/или отрицательной корреляции коэффициента повышения окружающего звука с величиной достоверности долговременной музыки и/или VoIP. Затем могут быть построены формулы, аналогичные формулам (5)-(8).

В качестве специального примера виртуализатор 404 окружающего звука может включаться для чистого программного материала, подобного кинофильму, и/или игры, а отключаться для музыки и/или VoIP. Между тем, коэффициент повышения виртуализатора 404 окружающего звука может устанавливаться по-разному для программного материала, подобного кинофильму, и игры, программный материал, подобный кинофильму, использует более высокий коэффициент повышения, а игра использует более низкий. Таким образом, регулирующий блок 300B может выполняться с возможностью более положительной корреляции коэффициента повышения окружающего звука с величиной достоверности программного материала, подобного кинофильму, чем игра.

Подобно типу содержимого коэффициент повышения звукового сигнала может также устанавливаться как средневзвешенное значение величин достоверности типов контекста:

$$\hat{a} = \frac{a_{КИНО} \cdot Conf_{КИНО} + a_{МУЗЫКА} \cdot Conf_{МУЗЫКА} + a_{ИГРА} \cdot Conf_{ИГРА} + a_{VOIP} \cdot Conf_{VOIP}}{Conf_{КИНО} + Conf_{МУЗЫКА} + Conf_{ИГРА} + Conf_{VOIP}} \qquad (9)$$

где $\hat{a}$ - расчетный коэффициент повышения, □ с индексом типа контекста - предполагаемый/предопределенный коэффициент повышения (весовой коэффициент) типа контекста, $Conf$ с индексом типа контекста - величина достоверности типа контекста. В зависимости от ситуации $a_{\text{МУЗЫКА}}$ и a_{VOIP} могут устанавливаться (но не обязательно) равными 0, указывая, что виртуализатор 404 окружающего звука будет отключен для чистой музыки (тип контекста) и/или чистого VoIP.

Опять же, аналогично типу содержимого, а с индексом типа контекста в формуле (9) представляет собой предполагаемый/предопределенный коэффициент повышения типа контекста, а частное от деления величины достоверности соответствующего типа контекста на сумму величин достоверности всех идентифицированных типов контекста может рассматриваться как нормированный весовой коэффициент предопределенного/

предполагаемого коэффициента повышения соответствующего типа контекста. То есть, регулирующий блок 300B может выполняться с возможностью учета по меньшей мере некоторых из множества типов контекста посредством взвешивания предопределенных коэффициентов повышения нескольких типов контекста в зависимости от величин достоверности.

3.2 Начальная частота

В виртуализаторе окружающего звука также могут изменяться другие параметры, такие как начальная частота. Как правило, высокочастотные составляющие в звуковом сигнале больше подходят для пространственного представления. Например, в музыке будет звучать странно, если бас пространственно представлен для создания большего количества эффектов окружающего звука. Таким образом, для конкретного звукового сигнала виртуализатор окружающего звука должен определить пороговое значение частоты, составляющие выше которой пространственно представляются, в то время как составляющие ниже которой сохраняются без изменения. Пороговое значение частоты является начальной частотой.

В соответствии с вариантом осуществления настоящей заявки начальная частота для виртуализатора окружающего звука может увеличиваться для музыкального содержимого таким образом, чтобы более низкие частоты могли сохраняться без изменения для музыкальных сигналов. Затем регулирующий блок 300B может выполняться с возможностью положительной корреляции начальной частоты виртуализатора окружающего звука с величиной достоверности кратковременной музыки.

3.3 Сочетание вариантов осуществления и сценариев применения

Аналогично части 1, все варианты осуществления и их видоизменения, рассмотренные выше, могут реализовываться в любом их сочетании, а любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции, могут реализовываться как такие же или отдельные компоненты.

Например, любые два или более решений, описанных в разделах 3.1 и 3.2, могут сочетаться друг с другом. И любое из сочетаний может дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в части 1, части 2 и других частях, которые будут описаны позже.

3.4 Способ управления виртуализатором окружающего звука

Аналогично части 1 в процессе описания контроллера виртуализатора окружающего звука в вариантах осуществления, приведенных выше в данной заявке, явно описаны также некоторые процессы или способы. Далее дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

Во-первых, варианты осуществления способа обработки звукового сигнала, как описывалось в части 1, могут применяться для виртуализатора окружающего звука, параметр(ы) которого является одной из целей для регулировки посредством способа обработки звукового сигнала. С этой точки зрения способ обработки звукового сигнала представляет собой также способ управления виртуализатором окружающего звука.

В этом разделе будут рассматриваться только аспекты, характерные для управления виртуализатором окружающего звука. Для общих аспектов способа управления может выполняться ссылка на часть 1.

Согласно одному варианту осуществления способ обработки звукового сигнала может дополнительно включать обработку виртуализации окружающего звука, и операция 1104 регулировки может выполняться с возможностью положительной корреляции коэффициента повышения окружающего звука обработки виртуализации

окружающего звука с величиной достоверности шума, и/или фона, и/или речи и/или отрицательной корреляции коэффициента повышения окружающего звука с величиной достоверности кратковременной музыки.

В частности, операция 1104 регулировки может выполняться с возможностью более положительной корреляции коэффициента повышения окружающего звука с величиной достоверности шума и/или фона, чем тип содержимого речи.

В альтернативном варианте или в дополнение к этому коэффициент повышения окружающего звука может также регулироваться в зависимости от величины(н) достоверности типа(ов) контекста. В частности, операция 1104 регулировки может выполняться с возможностью положительной корреляции коэффициента повышения окружающего звука обработки виртуализации окружающего звука с величиной достоверности программного материала, подобного кинофильму, и/или игры и/или отрицательной корреляции коэффициента повышения окружающего звука с величиной достоверности долговременной музыки и/или VoIP.

Более конкретно, операция 1104 регулировки может выполняться с возможностью более положительной корреляции коэффициента повышения окружающего звука с величиной достоверности программного материала, подобного кинофильму, чем игра.

Еще один параметр, который необходимо регулировать, представляет собой начальную частоту обработки виртуализации окружающего звука. Как показано на фиг. 19, начальная частота регулируется прежде всего в зависимости от величины(н) достоверности звукового типа(ов) (операция 1902), затем виртуализатор окружающего звука обрабатывает те составляющие звукового сигнала, которые выше начальной частоты (операция 1904). В частности, операция 1104 регулировки может выполняться с возможностью положительной корреляции начальной частоты обработки виртуализации окружающего звука с величиной достоверности кратковременной музыки.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. Кроме того, любые два или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания.

Часть 4: Контроллер выравнивателя громкости и способ управления

Громкость разных источников звука или разных фрагментов в том же источнике звука иногда сильно меняется. Это раздражает, поскольку пользователи должны часто регулировать громкость. Выравниватель громкости (VL) предназначен для настройки громкости звукового содержимого при воспроизведении и поддержании ее практически постоянной вдоль оси времени в зависимости от целевой величины громкости. Пример выравнивателя представлены в документе A.J. Seefeldt и др. "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", опубликованном как US 2009/0097676 A1; в документе B.G. Grockett и др. "Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection", опубликованном как WO 2007/127023 A1; в документе A. Seefeldt и др. "Audio Processing Using Auditory Scene Analysis and Spectral Skewness", опубликованном как WO 2009/011827 A1. Три документа в полном объеме включены в данную заявку посредством ссылки.

Выравниватель громкости непрерывно измеряет громкость звукового сигнала

определенным образом, а затем изменяет сигнал посредством коэффициента усиления, который представляет собой коэффициент масштабирования для изменения громкости звукового сигнала и обычно является функцией измеряемой громкости, желаемой целевой громкости и ряда других факторов. Необходимо учесть ряд факторов, чтобы оценить правильное усиление с основными критериями для обоих подходов к целевой громкости и поддерживать динамический диапазон. Он обычно содержит несколько подэлементов, таких как автоматическая регулировка усиления (AGC), обнаружение слуховых событий, управление динамическим диапазоном (DRC).

Управляющий сигнал, как правило, применяется в выравнивателе громкости для управления "усилением" звукового сигнала. Например, управляющий сигнал может быть индикатором изменений в амплитуде звукового сигнала, полученного посредством анализа чистого сигнала. Он также может быть индикатором звуковых событий, чтобы представлять, если появляется новое звуковое событие, с помощью психо-акустического анализа, такого как анализ слуховой сцены или выявления слухового события на основе конкретной громкости. Такой управляющий сигнал применяется в выравнивателе громкости для управления усилением, например, посредством обеспечения того, что коэффициент усиления является почти постоянной величиной в течение слухового события и посредством ограничения большей части изменения коэффициента усиления для окрестности границы события с целью уменьшения возможности слышимых искажений из-за быстрого изменения коэффициента усиления в звуковом сигнале.

Тем не менее, общепринятые способы получения управляющих сигналов не могут различать информативные слуховые события от неинформативных (мешающих) слуховых событий. В данной заявке информативное слуховое событие означает звуковое событие, которое содержит значимую информацию и может сильнее обратить внимание пользователей, например, диалог и музыка, в то время как неинформативный сигнал не содержит значимой информации для пользователей, например, шум в VoIP. Как следствие, к неинформативным сигналам может также применяться большой коэффициент усиления, и они могут повышаться почти до целевой громкости. Это будет неприятным в некоторых приложениях. Например, в VoIP-звонках, шум сигнала, который появляется в паузе разговора, часто увеличивается до большой громкости после обработки выравнивателем громкости. Это является нежелательным для пользователей.

С целью решения этой проблемы, по меньшей мере частично, настоящая заявка предлагает управлять выравнивателем громкости на основе вариантов осуществления, описанных в части 1.

Аналогично части 2 и части 3 в качестве конкретного примера устройства и способа обработки звукового сигнала, рассмотренных в части 1, выравниватель 406 громкости может использовать все варианты осуществления, описанные в части 1, и любые сочетания этих вариантов осуществления, описанных в данной заявке. В частности, в случае управления выравнивателем 406 громкости звуковой классификатор 200 и регулирующий блок 300 в устройстве 100 обработки звукового сигнала, как показано на фиг. 1-10 может представлять собой контроллер 2000 выравнивателя 406 громкости, как показано на фиг. 20. В этом варианте осуществления, поскольку регулирующий блок специфичен для выравнивателя 406 громкости, он может упоминаться как 300C.

То есть, на основе описания части 1, контроллер 2000 выравнивателя громкости может содержать звуковой классификатор 200 для непрерывного определения звукового типа (например, типа содержимого и/или типа контекста) звукового сигнала; и регулирующий блок 300C для регулировки выравнивателя громкости в непрерывном

режиме в зависимости от величины достоверности идентифицированного звукового типа. Аналогичным образом, звуковой классификатор 200 может содержать по меньшей мере одно из следующего: классификатор 202 звукового содержимого и классификатор 204 звукового контекста, а контроллер 2000 выравнивателя громкости может дополнительно содержать по меньшей мере одно из следующего: блок 712 сглаживания типа, блок 814 сглаживания параметра и таймер 916.

Таким образом, в этой части мы не будем повторять содержимое, уже описанное в части 1, а просто приведем его конкретные примеры.

Различные параметры в выравнивателе 406 громкости могут адаптивно настраиваться в зависимости от результатов классификации. Мы можем настроить параметры, непосредственно связанные с коэффициентом динамического усиления или диапазоном динамического усиления, например, путем уменьшения коэффициента усиления для неинформативных сигналов. Мы также можем настроить параметры, которые указывают на степень сигнала, нового воспринимаемого звукового события, а затем косвенно контролировать коэффициент динамического усиления (коэффициент усиления будет медленно изменяться в течение звукового события, но может быстро измениться на границе двух звуковых событий). В данной заявке представлены несколько вариантов осуществления настройки параметров или механизм управления выравнивателем громкости.

4.1 Типы информативного и мешающего содержимого

Как упоминалось выше, в связи с управлением выравнивателем громкости, типы звукового содержимого могут классифицироваться как типы информативного содержимого и типы мешающего содержимого. А регулирующий блок 300C может выполняться с возможностью положительной корреляции коэффициента динамического усиления выравнивателя громкости с типами информативного содержимого звукового сигнала и отрицательной корреляции коэффициента динамического усиления выравнивателя громкости с типами мешающего содержимого звукового сигнала.

В качестве примера предположим, что шум является мешающим (неинформативным), и он будет раздражать, будучи повышенным до большой громкости, параметр непосредственного управления динамическим усилением или параметр, указывающий новые звуковые события, может устанавливаться пропорционально убывающей функции величины достоверности шума $(Conf_{шум})$, например,

$$\text{GainControl} \propto D\,Conf_{шум} \qquad (10)$$

В данной заявке для простоты мы используем символ GainControl, чтобы представить все параметры (или их воздействия), связанные управлением усилением в уравнителе громкости, так как разные реализации выравнивателя громкости могут использовать разные названия параметров с разным скрытым смыслом. Использование единого термина GainControl может иметь короткое выражение без потери универсальности. В сущности, настройка данных параметров эквивалентна применению либо линейных, либо нелинейных весовых коэффициентов к исходному усилению. В одном примере GainControl может непосредственно применяться для масштабирования усиления таким образом, чтобы коэффициент усиления был небольшим, если GainControl мал. В качестве другого конкретного примера коэффициент усиления косвенно управляется масштабированием посредством управляющего событием сигнала GainControl, описанного в документе B.G. Grockett и др. "Audio Gain Control Using Specific-Loudness-Based Auditory Event Detection", опубликованном как в WO 2007/127023 A1, который включен в данную заявку в полном объеме посредством ссылки. В этом случае, когда

GainControl мал, элементы управления коэффициентом усиления выравнивателя громкости изменяются для предотвращения значительного изменения коэффициента усиления со временем. Когда GainControl высок, то элементы управления изменяются таким образом, что коэффициент усиления выравнивателя мог изменяться более свободно.

Посредством управления усилением, описанным в формуле (10) (либо непосредственного масштабированием исходного коэффициента усиления, либо управляющего событием сигнала), коэффициент динамического усиления звукового сигнала коррелирует (линейно или нелинейно) с величиной достоверности его шума. Если сигнал представляет собой шум с высокой величиной достоверности, окончательный коэффициент усиления будет мал из-за множителя $(1 - Conf_{шум})$. Таким образом, это позволяет избежать повышения сигнала шума до неприятно сильной громкости.

В качестве примерного варианта из формулы (10), если фоновый звук также не интересен в применении (например, в VoIP), он может трактоваться аналогичным образом и к нему также применяется малый коэффициент усиления. Функция управления может учитывать как величину достоверности шума $(Conf_{шум})$, так и величину достоверности фона $(Conf_{фон})$, например

$$\text{GainControl} \propto \text{D\_}Conf_{шум}) \cdot (1 - Conf_{фон}) \qquad (11)$$

В приведенной выше формуле, так как шум и фоновые звуки являются нежелательными, GainControl одинаково зависит от величины достоверности шума и величины достоверности фона, и можно считать, что шум и фоновые звуки имеют одинаковый весовой коэффициент. В зависимости от ситуации, они могут иметь разные весовые коэффициенты. Например, мы можем задать величины достоверности шума и фоновых звуков (или их разность с 1) разными коэффициентами или разными показателями степени (α и γ). То есть, формула (11) может быть переписана в виде:

$$\text{GainControl} \propto \_\_Conf_{шум})^{\alpha} \cdot (1 - Conf_{фон})^{\gamma} \qquad (12)$$

или

$$\text{GainControl} \propto \text{D\_}Conf_{шум}{}^{\alpha}) \cdot (1 - Conf_{фон}{}^{\gamma}) \qquad (13)$$

В альтернативном варианте регулирующий блок 300C может выполняться с возможностью учета по меньшей мере одного преобладающего типа содержимого в зависимости от величин достоверности. Например:

$$\text{GainControl} \propto \text{D\_Dd}\_\_Conf_{шум}, Conf_{фон}) \qquad (14)$$

И формула (11) (и ее варианты), и формула (14) указывают на небольшое усиление для сигналов шума и сигналов фонового звука, а исходный режим выравнивателя громкости сохраняется только тогда, когда и достоверность шума, и достоверность фона малы (например, в речи и музыкальном сигнале) настолько, что GainControl близок к единице.

Приведенный выше пример используется, чтобы учитывать преобладающий тип мешающего содержимого. В зависимости от ситуации регулирующий блок 300C может выполняться с возможностью учета преобладающего типа информативного содержимого в зависимости от величин достоверности. Чтобы быть более

универсальным, регулирующий блок 300C может выполняться с возможностью учета по меньшей мере одного преобладающего типа содержимого в зависимости от величин достоверности, независимо от того, представляют ли собой/включают ли идентифицированные звуковые типы информативные и/или мешающие звуковые типы.

В качестве другого примера варианта формулы (10), предполагая, что речевой сигнал является наиболее информативным содержимым и требует меньше изменений в стандартном режиме выравнивателя громкости, функция управления может учитывать как величину достоверности шума $(Conf_{шум})$, так и величину достоверности речи $(Conf_{речь})$, как

$$\text{GainControl} \propto Conf_{шум} \cdot (1 - Conf_{речь}) \qquad (15)$$

С помощью этой функции небольшой GainControl получается только для сигналов с высокой достоверностью шума и низкой достоверностью речи (например, для белого шума), и GainControl будет близок к 1, если достоверность речи высока (и, соответственно, сохраняется исходный режим выравнивателя громкости). В более общем смысле можно считать, что весовой коэффициент одного типа содержимого (например, $Conf_{шум}$) может изменяться с величиной достоверности по меньшей мере одного другого типа содержимого (например, $Conf_{речь}$). В приведенной выше формуле (15) можно считать, что достоверность речи изменяет весовой коэффициент достоверности шума (другой вид весового коэффициента по сравнению с весовыми коэффициентами в формулах (12 и 13)). Другими словами, в формуле (10) коэффициент $Conf_{шум}$ может рассматриваться как 1; в то время как в формуле (15) некоторые другие звуковые типы (например, речь, но не ограничиваясь ими) будут влиять на величину достоверности шума, таким образом, можно сказать, что весовой коэффициент $Conf_{шум}$ изменяется посредством величины достоверности речи. В контексте настоящего изобретения термин "весовой коэффициент" может истолковываться, чтобы включить это. То есть, он указывает на важность значения, но не обязательно нормализуется. Можно сделать ссылку на раздел 1.4.

С другой точки зрения, аналогично формулам (12) и (13), весовые коэффициенты в виде показателей степени могут применяться к величинам достоверности в приведенной выше функции, чтобы указать приоритет (или важность) различных звуковых сигналов, например, формула (15) может быть изменена на:

$$\text{GainControl} \propto D\,Conf_{шум}^{\alpha} \cdot (1 - Conf_{речь})^{\gamma} \qquad (16)$$

где α и γ - два весовых коэффициента, которые могут устанавливаться в меньшее значение, если, ожидается, что они будут более участвующими в изменении параметров выравнивателя.

Формулы (10)-(16) могут свободно комбинироваться в различные функции управления, которые могут подходить для различных применений. Аналогичным образом величины достоверности других типов звукового содержимого, например, величина достоверности музыки, могут также легко включаться в функции управления.

В случае, когда GainContrtol используется для настройки параметров, которые указывают на степень сигнала, являющегося воспринимаемым звуковым событием, а затем косвенно управляет коэффициентом динамического усиления (коэффициент усиления будет медленно изменяться в течение звукового события, но может быстро

измениться на границе двух звуковых событий), можно считать, что существует другая передаточная функция между величиной достоверности типов содержимого и окончательным коэффициентом динамического усиления.

4.2 Типы содержимого в разных контекстах

Приведенные выше функции управления в формулах (10)-(16) используют учет величин достоверности звуковых типов содержимого, таких как шум, фоновые звуки, кратковременная музыка и речь, но не учитывают их звуковые контексты, где звуки поступают от, например, программного материала, подобного кинофильму, и VoIP. Возможно, что данный одинаковый тип звукового содержимого нуждается в разной обработке в разных звуковых контекстах, например, фоновые звуки. Фоновый звук включает различные звуки, такие как двигатель автомобиля, взрыв и аплодисменты. Это может не иметь смысла в VoIP-звонке, но это может быть важно в программном материале, подобном кинофильму. Это означает, что должны идентифицироваться интересующие звуковые контексты, и разные функции управления должны предназначаться для разных звуковых контекстов.

Таким образом, регулирующий блок 300C может выполняться с возможностью отнесения типа содержимого звукового сигнала к информативному или мешающему в зависимости от типа контекста звукового сигнала. Например, посредством учета величины достоверности шума и величины достоверности фона и разграничения VoIP и не VoIP-контекстов звуковая контекстно-зависимая функция управления может быть следующей,

если *звуковой контекст является VoIP*

$$\text{GainControl} \propto \text{D\_Dd}\_\_Conf_{шум}, Conf_{фон})$$

иначе

$$\text{GainControl} \propto \text{D\_}Conf_{шум} \quad (17)$$

То есть, в контексте VoIP, шум и фоновые звуки рассматриваются как типы мешающего содержимого; в то время как в контексте не VoIP, фоновые звуки рассматриваются как тип информативного содержимого.

В качестве другого примера, звуковая контекстно-зависимая функция управления, учитывающая величины достоверности речи, шума и фона, и различающая VoIP и не VoIP-контексты, может быть следующей

если *звуковой контекст является VoIP*

$$\text{GainControl} \propto \_\_\_\_\_Conf_{шум}, Conf_{фон})$$

иначе

$$\text{GainControl} \propto \text{D}Conf_{шум} \cdot (1- Conf_{речь}) \quad (18)$$

В данной ситуации речь выделяется в качестве типа информативного содержимого.

Предположим, музыка также является важными информативными сведениями в не VoIP-контексте, мы можем распространить вторую часть формулы (18) к:

$$\text{GainControl} \propto 71717171\, Conf_{шум} \cdot (1- \max(Conf_{речь}, Conf_{музыки})) \quad (19)$$

В действительности, каждая из функций управления в (10)-(16) или их вариантах может быть применена в разных/соответствующие звуковых контекстах. Таким образом, может образовываться большое количество комбинаций для формирования звуковых контекстно-зависимых функций управления.

Кроме VoIP и не VoIP-контекстов, как разграничиваемых и применяемых в формулах (17) и (18), другие звуковые контексты, такие как программный материал, подобный кинофильму, долговременная музыка и игра или звуковой сигнал низкого качества и звуковой сигнал высокого качества, могут использоваться аналогичным образом.

4.3 Типы контекста

Типы контекста могут непосредственно использоваться для управления выравнивателем громкости, чтобы избежать неприятных звуков, таких как шум, повышенных слишком сильно. Например, величина достоверности VoIP может использоваться для управления выравнивателем громкости, делая его менее чувствительным при высокой величине достоверности.

В частности, посредством величины достоверности VoIP $Conf_{VOIP}$ уровень выравнивателя громкости может устанавливаться пропорционально $(1 - Conf_{VOIP})$. То есть, выравниватель громкости почти выключается в VoIP содержимом (когда величина достоверности VoIP высока), что согласуется с традиционной ручной настройкой (предустановкой), которая отключает выравниватель громкости для VoIP-контекста.

Кроме того, мы можем установить различные диапазоны динамического усиления для разных контекстов звуковых сигналов. В общем, коэффициент VL (выравнивателя громкости) дополнительно регулирует коэффициент усиления, применяемый к звуковому сигналу, и может рассматриваться как другой (нелинейный) весовой коэффициент к коэффициенту усиления. В одном варианте осуществления установка может быть следующей:

Таблица 1

	ПРОГРАММНЫЙ МАТЕРИАЛ, ПОДОБНЫЙ КИНОФИЛЬМУ	ДОЛГОВРЕМЕННАЯ МУЗЫКА	VOIP	ИГРА
Коэффициент VL	высокий	средний	Выкл. (или самый низкий)	низкий

Кроме того, если допустить, что предполагаемый коэффициент VL предопределен для каждого типа контекста. Например, коэффициент VL устанавливается как 1 для программного материала, подобного кинофильму, 0 для VoIP, 0,6 для музыки, и 0,3 для игры, но настоящая заявка не ограничивается этим. В соответствии с примером, если диапазон динамического усиления программного материала, подобного кинофильму, составляет 100%, то диапазон динамического усиления VoIP составляет 60%, и так далее. Если классификация звукового классификатора 200 основана на жестком решении, то диапазон динамической усиления может непосредственно устанавливаться как в примере выше. Если классификация звукового классификатора 200 основана на мягком решении, то диапазон может регулироваться в зависимости от величины достоверности типа контекста.

Аналогичным образом, звуковой классификатор 200 может идентифицировать различные типы контекстов из звукового сигнала, а регулирующий блок 300C может выполняться с возможностью регулировки диапазона динамического усиления путем взвешивания величин достоверности нескольких типов содержимого в зависимости от важности нескольких типов содержимого.

В целом, для типа контекста функции, аналогичные (10)-(16), могут также применяться здесь для установки соответствующего коэффициента VL адаптивно к типам содержимого, замененных типами контекста, и фактически таблица 1 отражает важность разных типов контекста.

С другой точки зрения, значение достоверности может применяться для получения нормализованного весового коэффициента, как описано в разделе 1.4. Предположим, что конкретный коэффициент предопределяется для каждого типа контекста в таблице 1, а затем также может применяться по формуле, аналогичной формуле (9). В связи с этим, подобные решения также могут применяться к нескольким типам содержимого и любым другим звуковым типам.

4.4 Сочетание вариантов осуществления и сценариев применения

Аналогично части 1 все варианты осуществления и разновидности, рассмотренные выше, могут реализовываться в любом их сочетании, и любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции могут реализовываться как такие же или отдельные компоненты. Например, любые два или более решений, описанных в разделах 4.1-4.3, могут сочетаться друг с другом. И любое из сочетаний может дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в частях 1-3 и других частях, которые будут описаны позже.

Фиг. 21 иллюстрирует результат применения контроллера выравнивателя громкости, предлагаемого в заявке, путем сравнения кратковременного сегмента (фиг. 21(A)), кратковременный сегмент обрабатывался общепринятым выравнивателем громкости без изменения параметров (фиг. 21(B)), и кратковременный сегмент обрабатывался выравнивателем громкости, представленным в данной заявке (фиг. 21(C)). Как видно, в общепринятых выравнивателях громкости, как показано на фиг. 21(B), громкость шума (вторая половина звукового сигнала) также повышена и раздражает. В противоположность этому, в новом выравнивателе громкости, как показано на фиг. 21(C), громкость полезной части звукового сигнала повышается без явного повышения громкости шума, давая хорошее восприятие слушателем.

4.5 Способ управления выравнивателем громкости

Аналогично части 1 в процессе описания контроллера выравнивателя громкости в вариантах осуществления, приведенных выше в данной заявке, явно описаны также некоторые процессы или способы. Далее дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

Во-первых, варианты осуществления способа обработки звукового сигнала, как описывалось в части 1, могут применяться для выравнивателя громкости, параметр(ы) которого является одной из целей для регулировки посредством способа обработки звукового сигнала. С этой точки зрения способ обработки звукового сигнала представляет собой также способ управления выравнивателем громкости.

В этом разделе будут рассматриваться только аспекты, характерные для управления выравнивателем громкости. Для общих аспектов способа управления может выполняться ссылка на часть 1.

В соответствии с настоящей заявкой предлагается способ управления выравнивателем

громкости, включающий идентификацию типа содержимого звукового сигнала в реальном времени, и регулировку выравнивателя громкости в непрерывном режиме в зависимости от идентифицированного типа содержимого посредством положительной корреляции коэффициента динамического усиления выравнивателя громкости с типами информативного содержимого звукового сигнала и отрицательной корреляции коэффициента динамического усиления выравнивателя громкости с типами мешающего содержимого звукового сигнала.

Тип содержимого может включать речь, кратковременную музыку, шум и фоновый звук. Как правило, шум рассматривается как тип мешающего содержимого.

При регулировке коэффициента динамического усиления выравнивателя громкости, он может регулироваться непосредственно в зависимости величины достоверности типа содержимого или может регулироваться с помощью передаточной функции величины достоверности типа содержимого.

Как уже было описано, звуковой сигнал может классифицироваться по нескольким звуковым типам одновременно. При использовании нескольких типов содержимого, операция 1104 регулировки может выполняться с возможностью учета по меньшей мере некоторых из множества типов звукового содержимого посредством взвешивания величин достоверности нескольких типов содержимого в зависимости от важности нескольких типов содержимого либо посредством взвешивания воздействий нескольких типов содержимого в зависимости от величин достоверности. В частности, и операция 1104 регулировки может выполняться с возможностью учета по меньшей мере одного преобладающего типа содержимого в зависимости от величин достоверности.

Когда звуковой сигнал содержит как тип(ы) мешающего содержимого, так и тип(ы) информативного содержимого, операция регулировки может выполняться с возможностью учета по меньшей мере одного преобладающего типа мешающего содержимого в зависимости от величин достоверности и/или учета по меньшей мере одного преобладающего типа информативного содержимого в зависимости от величин достоверности.

Разные звуковые типы могут влиять друг на друга. Таким образом, операция 1104 регулировки может выполняться с возможностью изменения весового коэффициента одного типа содержимого с величиной достоверности по меньшей мере одного другого типа содержимого.

Как описано в части 1, величина достоверности звукового типа звукового сигнала может сглаживаться. За подробностями операции сглаживания, пожалуйста, обратитесь к части 1.

Способ может дополнительно включать идентификацию типа контекста звукового сигнала, при этом операция 1104 регулировки может выполняться с возможностью регулировки диапазона динамического усиления в зависимости от величины достоверности типа контекста.

Роль типа содержимого ограничена типом контекста, где он расположен. Таким образом, когда и информация о типе содержимого, и информация о типе контекста получаются для звукового сигнала одновременно (то есть для того же звукового сегмента), тип содержимого звукового сигнала может определяться как информативный или мешающий в зависимости от типа контекста звукового сигнала. Кроме того, типу содержимого в звуковом сигнале отличающегося типа контекста может назначаться разный весовой коэффициент в зависимости от типа контекста звукового сигнала. С другой точки зрения, мы можем применять разный весовой коэффициент (больший или меньший, положительное значение или отрицательное значение), чтобы отразить

информативный характер или мешающий характер типа содержимого.

Тип контекста звукового сигнала может включать VoIP, программный материал, подобный кинофильму, долговременную музыку и игру. И в звуковом сигнале типа контекста VoIP фоновый звук рассматривается как тип мешающего содержимого; в то время как в звуковом сигнале типа контекста не VoIP фоновый звук, и/или речь, и/или музыка рассматриваются в качестве типа информативного содержимого. Другие типы контекста могут включать звуковой сигнал высокого качества или звуковой сигнал низкого качества.

Подобно нескольким типам содержимого, когда звуковой сигнал классифицируется по нескольким типам контекста с соответствующими величинами достоверности одновременно (в отношении того же звукового сегмента), операция 1104 регулировки может выполняться с возможностью учета по меньшей мере некоторых из множества типов контекста посредством взвешивания величин достоверности нескольких типов контекста в зависимости от важности нескольких типов контекста или посредством взвешивания воздействий нескольких типов в зависимости от величин достоверности. В частности, операция регулировки может выполняться с возможностью учета по меньшей мере одного преобладающего типа контекста в зависимости от величин достоверности.

Наконец, варианты осуществления способа, как описано в этом разделе, могут использовать способ звуковой классификации, который будет рассмотрен в частях 6 и 7, а в данной части подробное описание опущено.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. Кроме того, любые два или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания.

Часть 5: Контроллер эквалайзера и способ управления

Частотная коррекция обычно применяется к музыкальному сигналу для регулировки или изменения его спектрального баланса, известного как "тон" или "тембр". Традиционный эквалайзер позволяет пользователям настраивать общий профиль (кривую или форму) частотной характеристики (усиления) на каждого отдельного диапазона частот с целью выделения определенных инструментов или удаления нежелательных звуков. Популярные музыкальные проигрыватели, такие как Windows Media Player, обычно предусматривают графический эквалайзер для регулировки усиления в каждом диапазоне частот, а также предусматривают набор предустановок эквалайзера для разных музыкальных жанров, таких как рок, рэп, джаз и фолк, чтобы получить наилучшие впечатления при прослушивании разных жанров музыки. После выбора предустановки или установки набора параметров к сигналу будут применяться те же коэффициенты усиления частотной коррекции до тех пор, пока набор параметров не будет изменен вручную.

В отличие от этого, динамический эквалайзер предусматривает автоматическую регулировку коэффициентов усиления частотной коррекции в каждом диапазоне частот с целью сохранения общей согласованности спектрального баланса с учетом желаемого тембра или тона. Данная согласованность достигается постоянным контролем спектрального баланса звукового сигнала, сравнением его с желаемым

предустановленным спектральным балансом и динамической регулировкой применяемых коэффициентов усиления частотной коррекции для преобразования исходного спектрального баланса звукового сигнала к желаемому спектральному балансу. Желаемый спектральный баланс выбирается вручную или предварительно устанавливается перед обработкой.

Оба вида эквалайзеров совместно обладают следующим недостатком: лучший набор параметров частотной коррекции, желаемый спектральный баланс или связанные параметры должны выбираться вручную, и они не могут автоматически изменяться в зависимости от звукового содержимого при воспроизведении. Способность различать типы звукового содержимого будет очень важной для обеспечения общего хорошего качества для разных видов звуковых сигналов. Например, различным музыкальным произведениям необходимы разные наборы параметров эквалайзера, например, различные жанры.

В системе эквалайзера, в которой любые виды звуковых сигналов (а не только музыкальные) возможны для ввода, параметры эквалайзера должны регулироваться в зависимости от типов содержимого. Например, эквалайзер, обычно включается при музыкальных сигналах, но отключается при речевых сигналах, так как он может изменить тембр речи слишком сильно, и, соответственно, сделать звучание сигнала неестественным.

С целью решения этой проблемы по меньшей мере частично настоящая заявка предлагает управлять эквалайзером на основе вариантов осуществления, описанных в части 1.

Аналогично части 2-4, в качестве конкретного примера устройства и способа обработки звукового сигнала, рассмотренных в части 1, эквалайзер 408 может использовать все варианты осуществления, описанные в части 1, и любые сочетания этих вариантов осуществления, описанные в данной заявке. В частности, в случае управления эквалайзером 408 звуковой классификатор 200 и регулирующий блок 300 в устройстве 100 обработки звукового сигнала, как показано на фиг. 1-10, могут представлять собой контроллер 2200 эквалайзера, как показано на фиг. 22. В этом варианте осуществления, поскольку регулирующий блок специфичен для эквалайзера 408, он может упоминаться как 300D.

То есть, на основе описания части 1, контроллер 2200 эквалайзера может содержать звуковой классификатор 200 для непрерывного определения звукового типа звукового сигнала; и регулирующий блок 300D для регулировки эквалайзера в непрерывном режиме в зависимости от величины достоверности идентифицированного звукового типа. Аналогичным образом, звуковой классификатор 200 может содержать по меньшей мере одно из следующего: классификатор 202 звукового содержимого и классификатор 204 звукового контекста, а контроллер 2200 громкости эквалайзера может дополнительно содержать по меньшей мере одно из следующего: блок 712 сглаживания типа, блок 814 сглаживания параметра и таймер 916.

Таким образом, в этой части мы не будем повторять содержимое, уже описанное в части 1, а просто приведем его конкретные примеры.

5.1 Управление в зависимости от типа содержимого

В общем случае для общих типов звукового содержимого, таких как музыка, речь, фоновый звук и шум, эквалайзер должен настраиваться по-разному для разных типов содержимого. Аналогично традиционной настройке эквалайзер может автоматически включаться для музыкальных сигналов, но отключаться для речи; или более непрерывным способом устанавливается высокий уровень частотной коррекции для

музыкальных сигналов и низкий уровень частотной коррекции для речевых сигналов. Таким образом, уровень частотной коррекции эквалайзера может автоматически устанавливаться для разного звукового содержимого.

Специально для музыки, было отмечено, что эквалайзер не работает так хорошо для музыкального произведения, которое имеет преобладающий источник, так как тембр преобладающего источника может значительно изменяться и звучать неестественно, если применяется неподходящая частотная коррекция. Учитывая это, было бы лучше устанавливать низкий уровень частотной коррекции для музыкальных произведений с преобладающими источниками, при том, что уровень частотной коррекции может быть высоким для музыкальных произведений без преобладающих источников. С помощью этой информации эквалайзер может автоматически установить уровень частотной коррекции для разного музыкального содержимого.

Музыка также может группироваться в зависимости от различных свойств, таких как жанр, инструмент, а также общих музыкальных характеристик, включая ритм, темп и тембр. Подобно тому, как различные предустановки эквалайзера применяются для различных музыкальных жанров, данные музыкальные группы/кластеры могут также иметь свои собственные наборы параметров оптимальной частотной коррекции, или кривые эквалайзера (в традиционном эквалайзере), или оптимальный желаемый спектральный баланс (в динамическом эквалайзере).

Как уже упоминалось выше, эквалайзер, как правило, включается для музыкального содержимого, но отключается для речи, так как эквалайзер может сделать диалог звучащим не так хорошо из-за изменения тембра. Одним из способов достижения этого автоматически является сопоставление уровня частотной коррекции с содержимым, в частности, величины достоверности музыки и/или величины достоверности речи, полученных из блока классификации звукового содержимого. Здесь уровень частотной коррекции может истолковываться как весовой коэффициент применяемых коэффициентов усиления эквалайзера. Чем выше уровень, тем сильнее применяемая частотная коррекция. Для примера, если уровень частотной коррекции равен 1, то будет применяться полный набор параметров частотной коррекции; если уровень частотной коррекции равен нулю, то все коэффициенты усиления составляют соответственно 0 дБ и, таким образом, частотная коррекция не применяется. Уровень частотной коррекции может представляться разными параметрами в разных реализациях алгоритмов эквалайзера. Примерный вариант осуществления данного параметра представляет собой весовой коэффициент эквалайзера, как это реализовано в документе A. Seefeldt и др. "Calculating and Adjusting the Perceived Loudness and/or the Perceived Spectral Balance of an Audio Signal", опубликованном как US 2009/0097676 A1, который в полном объеме включен в данную заявку посредством ссылки.

Различные управляющие схемы могут предназначаться для настройки уровня частотной коррекции. Например, с информацией о типе звукового содержимого для установки уровня частотной коррекции может применяться или величина достоверности речи, или величина достоверности музыки, как

$$L_{ЧК} \propto Conf_{музыка} \qquad (20)$$

Или

$$L_{ЧК} \propto 1 - Conf_{речь} \qquad (21)$$

где $L_{ЧК}$ - уровень частотной коррекции и $Conf_{музыка}$ и $Conf_{речь}$ означают величину

достоверности музыки и речи.

То есть, регулирующий блок 300D может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки или отрицательной корреляции уровня частотной коррекции с величиной достоверности речи.

Величина достоверности речи и величина достоверности музыки далее могут использоваться совместно для установки уровня частотной коррекции. Существует общее мнение, что уровень частотной коррекции должен быть высоким, только когда величина достоверности музыки высокая, а величина достоверности речи низкая, а в противном случае уровень частотной коррекции низкий. Например,

$$L_{ЧК} = Conf_{музыка}(1 - Conf_{речь}{}^{\alpha}) \qquad (22)$$

где величина достоверности речи возводится в степень a с целью решения проблем, связанных с ненулевой величиной достоверности речи в музыкальном сигнале, что может часто случаться. С учетом приведенной выше формулы, частотная коррекция будет полностью применяться (с уровнем, равным 1) для чистых музыкальных сигналов без каких-либо компонентов речи. Как указано в части 1, a может рассматриваться как весовой коэффициент, зависящий от важности типа содержимого, и может, как правило, устанавливаться от 1 до 2.

Если задается больший весовой коэффициент для величины достоверности речи, то регулирующий блок 300D может выполняться с возможностью отключения эквалайзера 408, когда величина достоверности для типа содержимого речи больше, чем пороговое значение.

В приведенном выше описании типы содержимого музыки и речи взяты в качестве примеров. В альтернативном или дополнительном варианте величины достоверности фонового звука и/или шума могут также учитываться. В частности, регулирующий блок 300D может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности фона и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности шума.

В другом варианте осуществления величина достоверности может применяться для получения нормализованного весового коэффициента, как описано в разделе 1.4. Предположим, предполагаемый уровень частотной коррекции предопределяется для каждого типа содержимого (например, 1 для музыки, 0 для речи, 0,5 для шума и фона), формула, аналогичная формуле (8), может верно применяться.

Уровень частотной коррекции может дополнительно сглаживаться, чтобы избежать быстрого изменения, которые могут вносить слышимые искажения в точках переключения. Это может выполняться с помощью блока 814 сглаживания параметра, как описано в разделе 1.5.

5.2 Вероятность преобладающих источников в музыке

С целью избегания применения высокого уровня частотной коррекции к музыке с преобладающими источниками, уровень частотной коррекции может дополнительно коррелировать с величиной достоверности $Conf_{преоблад.}$, указывающей, если музыкальное произведение содержит преобладающий источник, например,

$$L_{ЧК} = 1 - Conf_{преоблад.} \qquad (23)$$

Таким образом, уровень частотной коррекции будет низким для музыкальных произведений с преобладающими источниками и высоким для музыкальных

произведений без преобладающих источников.

В данной формуле хотя описывается величина достоверности музыки с преобладающим источником, мы также можем использовать величину достоверности музыки без преобладающего источника. То есть, регулирующий блок 300D может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки без преобладающих источников и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки с преобладающими источниками.

Как указано в разделе 1.1, хотя музыка и речь, с одной стороны, и музыка с или без преобладающих источников, с другой стороны, являются типами содержимого на разных иерархических уровнях, они могут учитываться параллельно. Посредством совместного учета величины достоверности преобладающих источников и величин достоверности речи и музыки, как описано выше, уровень частотной коррекции может устанавливаться посредством объединения по меньшей мере одной из формул (20)-(21) с (23). Примером является объединение всех трех формул:

$$L_{ЧК} = Conf_{музыка}(1 - Conf_{речь})(1 - Conf_{преоблад.}) \qquad (24)$$

Разные весовые коэффициенты, зависящие от важности типа содержимого, могут дополнительно применяться к различным величинам достоверности для универсальности, наподобие формулы (22).

В качестве другого примера, предположим, что $Conf_{преоблад.}$ вычисляется только тогда, когда звуковой сигнал представляет собой музыку, ступенчатая функция может быть синтезирована, как

$$L_{ЧК} = \begin{cases} (1 - Conf_{преоблад.}) & Conf_{музыка} > порог \\ Conf_{музыка}(1 - conf_{речь}^{a}) & иначе \end{cases} \qquad (25)$$

Эта функция устанавливает уровень частотной коррекции в зависимости от величины достоверности преобладающих оценок, если система классификации четко устанавливает, что звуковой сигнал представляет собой музыку (величина достоверности музыки больше, чем пороговое значение); в противном случае, она устанавливается в зависимости от величин достоверности музыки и речи. То есть, регулирующий блок 300D может выполняться с возможностью учета кратковременной музыки без/с преобладающими источниками, когда величина достоверности для кратковременной музыки больше, чем пороговое значение. Конечно, первая или вторая половины в формуле (25) могут быть изменены наподобие формул (20)-(24).

Такая же схема сглаживания, описанная в разделе 1.5, также может применяться, и постоянная времени a может дополнительно устанавливаться в зависимости от типа переключения, такого как переключение от музыки с преобладающими источниками к музыке без преобладающих источников или переключение от музыки без преобладающих источников к музыке с преобладающими источниками. Для этой цели также может примеряться формула, аналогичная формуле (4').

5.3 Предустановки эквалайзера

Кроме адаптивной настройки уровня частотной коррекции в зависимости от величин достоверности типов звукового содержимого, надлежащие наборы параметров частотной коррекции или предустановки желаемого спектрального баланса также могут выбираться автоматически для разного звукового содержимого в зависимости от его жанра, инструмента или других характеристик. Музыка с таким же жанром,

содержащая тот же самый инструмент или имеющая те же самые музыкальные характеристики, может использовать совместно одни и те же наборы параметров эквалайзера или предустановки желаемого спектрального баланса.

Для универсальности мы используем термин "музыкальные кластеры", чтобы представлять музыкальные группы с тем же жанром, с тем же инструментом или подобными музыкальными атрибутами, и они могут рассматриваться как другой иерархический уровень типов звукового содержимого, как указано в разделе 1.1. Надлежащий набор параметров частотной коррекции, уровень частотной коррекции и/или предустановки желаемого спектрального баланса могут ассоциироваться с каждым музыкальным кластером. Набор параметров частотной коррекции представляет собой кривую усиления, применяемую к музыкальному сигналу, и может быть одной из переустановок эквалайзера, применяемых для различных музыкальных жанров (таких как классика, рок, джаз и фолк), и предустановки желаемого спектрального баланса представляет собой желаемый тембр для каждого кластера. Фиг. 23 иллюстрирует несколько примеров предустановок желаемого спектрального баланса, реализованных по технологии Dolby Home Theater. Каждый пример описывает желаемую спектральную форму во всем слышимом диапазоне частот. Эта форма непрерывна по сравнению со спектральной формой входящего звукового сигнала, и коэффициенты усиления частотной коррекции вычисляются из этого сравнения для преобразования спектральной формы звукового сигнала в ту, которая определяется предустановкой.

Для нового музыкального произведения может определяться ближайший кластер (жесткое решение), или величина достоверности может вычисляться в отношении каждого музыкального кластера (мягкое решение). Основываясь на этой информации, надлежащий набор параметров частотной коррекции или предустановки желаемого спектрального баланса могут определяться для данного музыкального произведения. Самый простой способ состоит в том, чтобы присвоить ему соответствующий набор параметров кластеров с наилучшим соответствием, как

$$P_{ЧК} = P_{c*} \qquad (26)$$

где $P_{ЧК}$ - расчетный набор параметров частотной коррекции или предустановки желаемого спектрального баланса, c* - индекс музыкального кластера с наилучшим соответствием (преобладающий звуковой тип), который может получаться посредством подбора кластера с самой высокой величиной достоверности.

Кроме того, может быть более одного музыкального кластера, имеющего величину достоверности больше нуля, это означает, что музыкальное произведение имеет более или менее схожие тем кластерам атрибуты. Например, музыкальное произведение может иметь несколько инструментов, или оно может иметь атрибуты нескольких жанров. Это является причиной еще одного способа для оценивания надлежащего набора параметров частотной коррекции посредством учета всех кластеров, вместо использования только ближайшего кластера. Например, может применяться взвешенная сумма:

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

где N - количество предопределенных кластеров и w_c - весовой коэффициент проектного набора параметров P_c в отношении каждого заранее определенного

музыкального кластера (с индексом c), который должен быть нормализован относительно 1 на основе соответствующих величин достоверности. Таким образом, расчетный набор параметров будет представлять собой смесь из наборов параметров музыкальных кластеров. Например, для музыкального произведения, имеющего как атрибуты джаза, так и рока, расчетный набор параметров будет приблизительно промежуточным.

В некоторых применениях мы можем не хотеть затрагивать все кластеры, как показано в формуле (27). Только подмножество кластеров - кластеры наиболее связанные с текущим музыкальным произведением - должно учитываться, формула (27) может быть немного пересмотрена:

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

где N' - количество кластеров, которые должны учитываться и c' - индекс кластера после сортировки кластеров в порядке убывания в зависимости от их величин достоверности. При использовании подмножества мы можем сильнее сосредоточиться на наиболее связанных кластерах и исключить те, кто менее значимы. Другими словами, регулирующий блок 300D может выполняться с возможностью учета по меньшей мере одного преобладающего звукового типа в зависимости от величин достоверности.

В приведенном выше описании музыкальные кластеры взяты в качестве примера. В действительности, решения применимы к звуковым типам на любом иерархическом уровне, как описано в разделе 1.1. Таким образом, в общем случае регулирующий блок 300D может выполняться с возможностью назначения уровня частотной коррекции, и/или набора параметров частотной коррекции, и/или предустановки спектрального баланса для каждого звукового типа.

5.4 Управление в зависимости от типа контекста

В предыдущих разделах рассмотрение сосредоточено на различных типах содержимого. В дополнительном количестве вариантов осуществления, которые будут рассматриваться в этом разделе, тип контекста может учитываться альтернативно или дополнительно.

В большинстве случаев эквалайзер включается для музыки, но отключается для программного материла, подобного кинофильму, так как эквалайзер может сделать диалоги в программном материале, подобном кинофильму, звучание не таким хорошим из-за явного изменения тембра. Это означает, что уровень частотной коррекции может связываться с величиной достоверности долговременной музыки и/или величиной достоверности программного материла, подобного кинофильму:

$$L_{ЧК} \propto Conf_{МУЗЫКА} \qquad (29)$$

Или

$$L_{ЧК} \propto 1 - Conf_{МУЗЫКА} \qquad (30)$$

где $L_{ЧК}$ - уровень частотной коррекции, $Conf_{МУЗЫКА}$ и $Conf_{КИНО}$ означает величину достоверности долговременной музыки и программного материла, подобного кинофильму.

То есть, регулирующий блок 300D может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности

долговременной музыки или отрицательной корреляции уровня частотной коррекции с величиной достоверности программного материла, подобного кинофильму.

То есть, для программного материла, подобного кинофильму, величина достоверности программного материла, подобного кинофильму, высока (или достоверность музыки низка), и, таким образом уровень частотной коррекции низкий; с другой стороны, для музыкального сигнала величина достоверности программного материла, подобного кинофильму, будет низкой (или достоверность музыки высокой) и, таким образом, уровень частотной коррекции высокий.

Решения, показанные в формулах (29) и (30) могут изменяться таким же образом, как формулы (22)-(25), и/или могут комбинироваться с любым одним из решений, показанных в формулах (22)-(25).

В дополнение к этому или в альтернативном варианте регулирующий блок 300D может выполняться с возможностью отрицательной корреляции уровня частотной коррекции с величиной достоверности игры.

В другом варианте осуществления величина достоверности может применяться для получения нормализованного весового коэффициента, как описано в разделе 1.4. Предположим, предполагаемый уровень/набор параметров частотной коррекции предопределен для каждого типа контекста (наборы параметров частотной коррекции приведены в следующей таблице 2), также может применяться формула, аналогичная формуле (9).

Таблица 2.

	ПРОГРАММНЫЙ МАТЕРИАЛ, ПОДОБНЫЙ КИНОФИЛЬМУ	ДОЛГОВРЕМЕННАЯ МУЗЫКА	VoIP	ИГРА
набор параметров частотной коррекции	Набор параметров 1	Набор параметров 2	Набор параметров 3	Набор параметров 4

Здесь в некоторых наборах параметров все коэффициенты усиления могут устанавливаться равными нулю в качестве способа отключения эквалайзера для определенного типа контекста, такого как программный материл, подобный кинофильму, и игра.

5.5 Сочетание вариантов осуществления и сценариев применения

Аналогично части 1 все варианты осуществления и разновидности, рассмотренные выше, могут реализовываться в любом их сочетании, и любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции, могут реализовываться как такие же или отдельные компоненты.

Например, любые два или более решений, описанных в разделах 5.1-5.4, могут сочетаться друг с другом. И любое из сочетаний может дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в частях 1-4 и других частях, которые будут описаны позже.

5.6 Способ управления эквалайзером

Аналогично части 1 в процессе описания контроллера эквалайзера в вариантах осуществления, приведенных выше в данной заявке, явно описаны также некоторые процессы или способы. Далее дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

Во-первых, варианты осуществления способа обработки звукового сигнала, как описывалось в части 1, могут применяться для эквалайзера, параметр(ы) которого является одной из целей для регулировки посредством способа обработки звукового сигнала. С этой точки зрения способ обработки звукового сигнала представляет собой также способ управления эквалайзером.

В этом разделе будут рассматриваться только аспекты, характерные для управления эквалайзером. Для общих аспектов способа управления может выполняться ссылка на часть 1.

Соответственно, способ управления эквалайзером может включать идентификацию звукового типа звукового сигнала в реальном времени и регулировку эквалайзера в непрерывном режиме в зависимости от величины достоверности идентифицированного звукового типа.

Аналогично другим частям настоящей заявки, при использовании нескольких типов содержимого с соответствующими величинами достоверности, операция 1104 регулировки может выполняться с возможностью учета по меньшей мере некоторых из множества звуковых типов посредством взвешивания величин достоверности нескольких типов содержимого в зависимости от важности нескольких типов содержимого либо посредством взвешивания воздействий нескольких типов содержимого в зависимости от величин достоверности. В частности, операция 1104 регулировки может выполняться с возможностью учета по меньшей мере одного преобладающего звукового типа в зависимости от величин достоверности.

Как описано в части 1, значение регулируемого параметра может сглаживаться. Можно сослаться на раздел 1.5 и раздел 1.8, и подробное описание здесь опущено.

Звуковой тип может являться либо типом содержимого, либо типом контекста, либо обоими. При использовании типа содержимого операция 1104 регулировки может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности речи. В дополнение к этому или в альтернативном варианте операция регулировки может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности фона и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности шума.

При использовании типа контекста операция 1104 регулировки может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности долговременной музыки и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности программного материла, подобного кинофильму, и/или игры.

Для типа содержимого краткосрочной музыки операция 1104 регулировки может выполняться с возможностью положительной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки без преобладающих источников и/или отрицательной корреляции уровня частотной коррекции с величиной достоверности кратковременной музыки с преобладающими источниками. Это может быть выполнено только тогда, когда величина достоверности для кратковременной музыки больше, чем пороговое значение.

Кроме регулировки уровня частотной коррекции другие аспекты эквалайзера могут регулироваться в зависимости от величины(н) достоверности звукового типа(ов) звукового сигнала. Например, операция 1104 регулировки может выполняться с возможностью назначения уровня частотной коррекции, и/или наборы параметров

частотной коррекции, и/или предустановки спектрального баланса для каждого звукового типа.

На конкретные примеры звуковых типов может быть сделана ссылка на часть 1.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. Кроме того, любые два или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания.

Часть 6: Звуковые классификаторы и способы классификации

Как указано в разделах 1.1 и 1.2, звуковые типы, рассмотренные в настоящей заявке, включающие различные иерархические уровни типов содержимого и типов контекста, могут классифицироваться или идентифицироваться с помощью любой существующей схемы классификации, включающей способы на основе машинного обучения. В этой части и последующей части настоящая заявка предлагает некоторые новые аспекты классификаторов и способы классификации типов контекста, как указано в предыдущих частях.

6.1 Классификатор контекста на основе классификации типа содержимого

Как указано в предыдущих частях, звуковой классификатор 200 применяется для идентификации типа содержимого звукового сигнала и/или идентифицирует тип контекста звукового сигнала. Таким образом, звуковой классификатор 200 может содержать классификатор 202 звукового содержимого и/или классификатор 204 звукового контекста. При принятии существующих методов для реализации классификатора 202 звукового содержимого и классификатора 204 звукового контекста два классификатора могут быть независимыми друг от друга, хотя они могут совместно использовать некоторые признаки объекта и, следовательно, могут совместно использовать некоторые схемы для извлечения признаков объекта.

В этой части и последующей части 7 в соответствии с новым аспектом, предлагаемым в настоящей заявке, классификатор 204 звукового контекста может использовать результаты классификатора 202 звукового содержимого, то есть, звуковой классификатор 200 может содержать: классификатор 202 звукового содержимого для идентификации типа содержимого звукового сигнала; и классификатор 204 звукового контекста для идентификации типа контекста звукового сигнала на основе результатов классификатора 202 звукового содержимого. Таким образом, результаты классификации классификатора 202 звукового содержимого могут использоваться как классификатором 204 звукового контекста, так и регулирующим блоком 300 (или регулирующими блоками 300A-300D), рассмотренным в предыдущих частях. Тем не менее, несмотря на то, что это не приведено в графическом материале, звуковой классификатор 200 может также содержать два классификатора 202 звукового содержимого, которые будут использоваться соответственно регулирующим блоком 300 и классификатором 204 звукового контекста.

Кроме того, как отмечалось в разделе 1.2, особенно при классификации нескольких звуковых типов, либо классификатор 202 звукового содержимого, либо классификатор 204 звукового контекста может входить в группу классификаторов взаимодействующих друг с другом, хотя также возможно это реализовать в виде одного классификатора.

Как отмечалось в разделе 1.1, тип содержимого представляет собой вид звукового

типа в отношении кратковременных звуковых сегментов, в большинстве случаев имеющих длительность порядка от нескольких кадров до нескольких десятков кадров (например, 1 с), а тип контекста представляет собой вид звукового типа в отношении долговременных звуковых сегментов, в большинстве случаев имеющих длительность порядка от нескольких секунд до нескольких десятков секунд (например, 10 с). Таким образом, в соответствии с терминами "тип содержимого" и "тип контекста" при необходимости мы используем, соответственно, термины «кратковременный» и «долговременный». Однако, как это будет описано в последующей части 7, несмотря на то, что тип контекста используется для указания свойства звукового сигнала в течение относительно длительного срока, он также может определяться на основе признаков объекта, извлеченных из кратковременных звуковых сегментов.

Теперь обратимся к структурам классификатора 202 звукового содержимого и классификатора 204 звукового контекста со ссылкой на фиг. 24.

Как показано на фиг. 24, классификатор 202 звукового содержимого может содержать выделитель 2022 кратковременных признаков объекта для извлечения кратковременных признаков объекта из кратковременных звуковых сегментов, каждый из которых содержит последовательность звуковых кадров; и кратковременный классификатор 2024 для классификации последовательности кратковременных сегментов в долговременном звуковом сегменте по кратковременным звуковым типам с использованием соответствующих кратковременных признаков объекта. И выделитель 2022 кратковременных признаков объекта, и кратковременный классификатор 2024 могут быть реализованы с помощью существующих методов, но также предложены некоторые модификации для выделителя 2022 кратковременных признаков объекта в последующем разделе 6.3.

Кратковременный классификатор 2024 может выполняться с возможностью классификации каждого из последовательности кратковременных сегментов по меньшей мере одного из следующих кратковременных звуковых типов (типов содержимого): речи, кратковременной музыки, фонового звука и шума, что пояснялось в разделе 1.1. Каждый из типов содержимого может дополнительно классифицироваться по типам содержимого на низком иерархическом уровне таким образом, как описано в разделе 1.1, но не ограничивается ими.

Как известно в данной области техники, величины достоверности классифицированных звуковых типов могут получаться с помощью кратковременного классификатора 2024. В настоящей заявке при упоминании работы любого классификатора должно быть понятно, что величины достоверности получаются одновременно, если это необходимо, независимо от того записываются ли они явно. Пример классификации звуковых типов может быть найден в документе L. Lu, H.-J. Zhang, and S. Li, "Content-based Audio Classification and Segmentation by Using Support Vector Machines", ACM Multimedia Systems Journal 8 (6), pp. 482-492, March, 2003, который включен в данную заявку в полном объеме посредством ссылки.

С другой стороны, как показано на фиг. 24, классификатор 204 звукового контекста может содержать выделитель 2042 статистических данных для расчета статистических данных результатов кратковременного классификатора в отношении последовательности кратковременных сегментов в долговременном звуковом сегменте в качестве долговременных признаков объекта; и использующий долговременные признаки объекта долговременный классификатор 2044 для классификации долговременного звукового сегмента по долговременным звуковым типам. Аналогично с помощью существующих методов могут реализовываться как выделитель 2042

статистических данных, так и долговременный классификатор 2044, но для выделителя 2042 статистических данных в последующем разделе 6.2 также предложены некоторые модификации.

Долговременный классификатор 2044 может выполняться с возможностью классификации долговременных звуковых сегментов по меньшей мере одного из следующих долговременных звуковых типов (типов контекста): программного материала, подобного кинофильму, долговременной музыки, игры и VoIP, которые пояснялись в разделе 1.1. В альтернативном варианте или в дополнение к этому долговременный классификатор 2044 может выполняться с возможностью классификации долговременного звукового сегмента на VoIP или не VoIP, что было объяснено в разделе 1.1. В альтернативном варианте или в дополнение к этому долговременный классификатор 2044 может выполняться с возможностью классификации долговременного звукового сегмента на звуковой сигнал высокого качества или звуковой сигнал низкого качества, что было объяснено в разделе 1.1. На практике различные целевые звуковые типы могут выбираться и обучаться на основе обязательных требований приложения/системы.

Относительно значения и выбора кратковременного сегмента и долговременного сегмента (а также кадра, который будет рассматриваться в разделе 6.3) может быть сделана ссылка на раздел 1.1.

6.2 Извлечение долговременных признаков объекта

Как показано на фиг. 24, в одном варианте осуществления применяется только выделитель 2042 статистических данных для извлечения долговременных признаков объекта из результатов кратковременного классификатора 2024. В качестве долговременных признаков объекта может рассчитываться посредством выделителя 2042 статистических данных по меньшей мере одно из следующего: среднее значение и дисперсия величин достоверности кратковременных звуковых типов кратковременных сегментов в классифицируемом долговременном, среднее значение и дисперсия взвешенных по степени важности кратковременных сегментов, частота появления каждого кратковременного звукового типа и частота переключения между разными кратковременными звуковыми типами в классифицируемом долговременном сегменте.

Проиллюстрируем на фиг. 25 среднее значение величин достоверности речи и кратковременной музыки в каждом кратковременном сегменте (длительностью 1 с). Для сравнения сегменты извлекаются из трех разных звуковых контекстов: программного материала, подобного кинофильму (фиг. 25(A)), долговременной музыки (фиг. 25(B)) и VoIP (фиг. 25(C)). Можно заметить, что для контекста программного материала, подобного кинофильму, высокие величины достоверности получаются или для типа речи, или для типа музыки, и часто чередуется между этими двумя звуковыми типами. В противоположность этому, сегмент долговременной музыки дает стабильную и высокую величину достоверности кратковременной музыки и относительно стабильную и низкую величину достоверности речи. В то время как сегмент VoIP дает стабильную и низкую величину достоверности кратковременной музыки, но дает колеблющуюся из-за пауз во время VoIP-разговора величину достоверности речи.

Дисперсия величин достоверности для каждого звукового типа также является важным признаком объекта для классификации разных звуковых контекстов. На фиг. 26 изображены гистограммы дисперсии величин достоверности речи, кратковременной музыки, фона и шума в звуковых контекстах программного материала, подобного кинофильму, долговременной музыки и VoIP (по оси абсцисс отложена дисперсия величин достоверности в наборе данных, а по оси ординат - число вхождений в каждый

интервал значений дисперсии s в наборе данных, которые могут быть нормализованы для указания вероятности появления в каждом интервале значений дисперсии). Для программного материала, подобного кинофильму, все дисперсии величины достоверности речи, кратковременной музыки и фона являются относительно высокими и имеют широкое распределение, указывающее, что величины достоверности данных звуковых типов меняются интенсивно; для долговременной музыки все дисперсии величины достоверности речи, кратковременной музыки, фона и шума являются относительно низкими и имеют узкое распределение, указывающее, что величины достоверности данных звуковых типов сохраняют стабильность: величина достоверности речи держится постоянно низкой, а величина достоверности музыки держится постоянно высокой; для VoIP дисперсия величины достоверности кратковременной музыки является низкой и имеет узкое распределение, в то время как, для речи имеет сравнительно широкое распределение, что связано с частыми паузами во время VoIP-разговоров.

Относительно весовых коэффициентов, используемых при расчете средневзвешенного значения и дисперсии, они определяются в зависимости степени важности каждого кратковременного сегмента. Степень важности кратковременного сегмента может измеряться посредством его энергии или громкости, которая может оцениваться с помощью многих существующих методов.

Частота появления каждого кратковременного звукового типа в долговременном сегменте, который должен классифицироваться, представляет собой нормализованную по длительности долговременного сегмента встречаемость каждого звукового типа, к которому были отнесены кратковременные сегменты в долговременном сегменте.

Частота переключения между различными кратковременными звуковыми типами в долговременном сегменте, который должен классифицироваться, представляет собой нормализованную по длительности долговременного сегмента встречаемость изменений звуковых типов между смежными кратковременными сегментами в долговременном сегменте, который должен классифицироваться.

При описании среднего значения и дисперсии величин достоверности со ссылкой на фиг. 25, частота появления каждого кратковременного звукового типа и частота переключения между этими разными кратковременными звуковыми типами также фактически сравниваются. Данные признаки объекта также весьма значимы для классификации звукового контекста. Например, долговременная музыка в основном содержит звуковой тип кратковременной музыки и, таким образом, имеет высокую частоту появления кратковременной музыки, в то время как VoIP в основном содержит речь и паузы и, таким образом, имеет высокую частоту появления речи или шума. В качестве другого примера программный материал, подобный кинофильму, переключается между разными кратковременными звуковыми типами чаще, чем долговременная музыка или VoIP, таким образом, он в целом имеет более высокую частоту переключения между кратковременной музыкой, речью и фоном; VoIP обычно переключается между речью и шумом чаще, чем это делают другие, и, таким образом, в целом имеет более высокую частоту переключения между речью и шумом.

Как правило, мы предполагаем, что долговременные сегменты имеют одинаковую длительность в том же приложении/системе. Если это так, то число появлений каждого кратковременного звукового типа и число переключений между различными кратковременными звуковыми типами в долговременном сегменте может применяться непосредственно без нормализации. Если длительность долговременного сегмента является переменной величиной, то должны использоваться частота появления и частота

переключений, как указано выше. А формулу изобретения в настоящей заявке следует толковать, как охватывающую обе ситуации.

В дополнение к этому или в альтернативном варианте звуковой классификатор 200 (или классификатор 204 звукового контекста) может дополнительно содержать выделитель 2046 долговременных признаков объекта (фиг. 27) для дополнительного извлечения долговременных признаков объекта из долговременного звукового сегмента в зависимости от кратковременных признаков объекта последовательности кратковременных сегментов в долговременном звуковом сегменте. Другими словами, выделитель 2046 долговременных признаков объекта не использует результаты классификации кратковременного классификатора 2024, а непосредственно использует кратковременные признаки объекта, извлеченные выделителем 2022 кратковременных признаков объекта, для получения некоторых долговременных признаков объекта для использования долговременным классификатором 2044. Выделитель 2046 долговременных признаков объекта и выделитель 2042 статистических данных могут применяться раздельно или совместно. Другими словами, звуковой классификатор 200 может содержать либо выделитель 2046 долговременных признаков объекта, либо выделитель 2042 статистических данных, либо оба.

Любые признаки объекта могут извлекаться с помощью выделителя 2046 долговременных признаков объекта. В настоящей заявке, предлагается вычислять в качестве долговременных признаков объекта по меньшей мере одно из следующих статистических данных кратковременных признаков объекта из выделителя 2022 кратковременных признаков объекта: среднее значение, дисперсию, средневзвешенное значение, взвешенную дисперсию, наибольшее среднее, наименьшее среднее и отношение (контраст) между наибольшим средним и наименьшим средним.

Среднее значение и дисперсия кратковременных признаков объекта извлечены из кратковременных сегментов в долговременном сегменте, который должен классифицироваться;

Средневзвешенное значение и дисперсия кратковременных признаков объекта извлечены из кратковременных сегментов в долговременном сегмента, который должен классифицироваться. Кратковременные признаки объекта взвешиваются по степени важности каждого кратковременного сегмента, которая измеряется посредством его энергии или громкости, как только что упоминалось;

Наибольшее среднее: среднее значение выбранных кратковременных признаков объекта, извлеченных из кратковременных сегментов в долговременном сегменте, который должен классифицироваться. Кратковременные признаки объекта выбираются при выполнении по меньшей мере одного из следующих условий: больше, чем пороговое значение; или в пределах предопределенной доли кратковременных признаков объекта не ниже, чем все другие кратковременные признаки объекта, например, самые высокие 10% кратковременных признаков объекта;

Наименьшее среднее: среднее значение выбранных кратковременных признаков объекта, извлеченных из кратковременных сегментов в долговременном сегменте, который должен классифицироваться. Кратковременные признаки объекта выбираются при удовлетворении по меньшей мере одного из следующих условий: меньше, чем пороговое значение; или в пределах предопределенной доли кратковременных признаков объекта не выше, чем все другие кратковременные признаки объекта, например, самый низкие 10% кратковременных признаков объекта; и

Контраст: отношение между наибольшим средним и наименьшим средним для представления динамики кратковременных признаков объекта в долговременном

сегменте.

Выделитель 2022 кратковременных признаков объекта может реализовываться с помощью существующих методов, и им могут извлекаться любые признаки объекта. Тем не менее, в последующем разделе 6.3 предлагаются некоторые модификации для выделителя 2022 кратковременных признаков объекта.

6.3 Извлечение кратковременных признаков объекта

Как показано на фиг. 24 и фиг. 27, выделитель 2022 кратковременных признаков объекта может выполняться с возможностью извлечения в качестве кратковременных признаков объекта по меньшей мере одного из следующих признаков объекта непосредственно из каждого кратковременного звукового сегмента: ритмических характеристик, характеристик прерываний/приглушений и кратковременных признаков качества звукового сигнала.

Ритмические характеристики могут включать интенсивность ритма, равномерность ритма, четкость ритма (см. документ L. Lu, D. Liu, and H.-J. Zhang. "Automatic mood detection and tracking of music audio signals". EEE Transactions on Audio, Speech, and Language Processing, 14(1):5 - 18, 2006, который в полном объеме включен в данную заявку посредством ссылки) и двухмерную модуляцию поддиапазонов (см. документ M.F. McKinney and J. Breebaart. "Features for audio and music classification", Proc. ISMIR, 2003, который в полном объеме включен в данную заявку посредством ссылки).

Характеристики прерываний/приглушений могут включать перерывы речи, резкие спады, беззвучный интервал, неестественную тишину, среднее значение неестественной тишины, полную энергию неестественной тишины и т.д.

Кратковременные признаки качества звукового сигнала представляют собой признаки качества в отношении к кратковременных сегментов, которые подобны признакам качества звукового сигнала, извлеченным из звуковых кадров, которые будут рассматриваться ниже.

В альтернативном варианте или в дополнение к этому, как показано на фиг. 28, звуковой классификатор 200 может содержать выделитель 2012 признаков на уровне кадра для извлечения признаков на уровне кадра из каждого кадра последовательности звуковых кадров, содержащихся в кратковременном сегменте, а выделитель 2022 кратковременных признаков объекта может выполняться с возможностью вычисления кратковременных признаков объекта в зависимости от признаков на уровне кадра, извлеченных из последовательности звуковых кадров.

В качестве предварительной обработки входной звуковой сигнал может быть микширован с понижением до монофонического звукового сигнала. Предварительная обработка не требуется, если звуковой сигнал уже является монофоническим сигналом. Затем он делится на кадры с предопределенной длительностью (как правило, от 10 до 25 миллисекунд). Соответственно, признаки на уровне кадра извлекаются из каждого кадра.

Выделитель 2012 признаков на уровне кадра может выполняться с возможностью извлечения по меньшей мере одного из следующих признаков объекта: признаков объекта, характеризующих свойства различных кратковременных звуковых типов, частоты среза, статических характеристик отношения сигнал/шум (SNR), сегментных характеристик отношения сигнал/шум (SNR), основных речевых дескрипторов и характеристик речевого тракта.

Признаки объекта, характеризующие свойства различных кратковременных звуковых типов (особенно речи, кратковременной музыки, фонового звуку и шума) могут содержать по меньшей мере один из следующих признаков объекта: энергию кадра,

спектральное распределение поддиапазонов, спектральный поток,

коэффициенты косинусного преобразования Фурье для частот чистых тонов (MFCC), низкую звуковую частоту, остаточную информацию, цветовой признак и частоту переходов через нуль.

Для подробностей о MFCC может быть сделана ссылка на документ L. Lu, H.-J. Zhang, and S. Li, "Content-based Audio Classification and Segmentation by Using Support Vector Machines", ACM Multimedia Systems Journal 8 (6), pp. 482-492, March, 2003, который в полном объеме включен в данную заявку посредством ссылки. Для подробностей о цветовом признаке может быть сделана ссылка на документ G.H. Wakefield, "Mathematical representation of joint time Chroma distributions" in SPIE, 1999, который в полном объеме включен в данную заявку посредством ссылки.

Частота среза представляет наивысшую частоту звукового сигнала, выше которой энергия содержимого близка к нулю. Она предназначена для обнаружения содержимого с ограниченной полосой частот, что является результативным в данной заявке для классификации звукового контекста. Частота среза обычно обусловлена кодированием, так как большинство кодеров отбрасывают высокие частоты при низких или средних скоростях передачи данных. Например, MP3-кодек имеет частоту среза 16 кГц при 128 кбит/с; в качестве другого примера многие популярные кодеки VoIP имеют частоту среза 8 кГц или 16 кГц.

Кроме частоты среза деградация сигнала в процессе кодирования звукового сигнала считается другой характеристикой для разграничения различных звуковых контекстов, таких как контексты VoIP в противовес не VoIP, контексты звукового сигнала высокого качества в противовес контекстам звукового сигнала низкого качества. Признаки объекта, представляющие качество звукового сигнала, например, для объективной оценки качества речи (см. документ 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, November 2006, который в полном объеме включен в данную заявку посредством ссылки) могут дополнительно извлекаться на нескольких уровнях для получения более полных характеристик. Примеры признаков качества звукового сигнала включают:

Статические характеристики SNR, включающие расчетный уровень фонового шума, спектральную четкость и т.д.

Сегментные SNR характеристики, включающие отклонения спектрального уровня, диапазон спектрального уровня, относительный минимальный уровень шума и т.д.

Основные речевые дескрипторы, включающие усреднение шага, изменение средней чувствительности речевого отрезка, среднюю чувствительность микрофона и т.д.

Характеристики речевого тракта, включающие роботизацию, шаг взаимного энергетического спектра и т.д.

Для получения кратковременных признаков объекта из признаков на уровне кадра выделитель 2022 кратковременных признаков объекта может выполняться с возможностью вычисления статистических данных признаков на уровне кадра в качестве кратковременных признаков объекта.

Примеры статистических данных признаков на уровне кадра включают среднее значение и стандартное отклонение, которые охватывают ритмические свойства для разграничения различных звуковых типов, таких как кратковременная музыка, речь, фон и шум. Например, речь, обычно чередуется между вокализированными и невокализированными звуками в размере слога, тогда как музыка нет, указывая, что изменение признака речи на уровне кадра обычно больше, чем музыки.

Другой пример статистических данных представляет собой средневзвешенное значение признаков на уровне кадра. Например, для частоты среза, средневзвешенное значение, полученное из частот среза из каждого звукового кадра в кратковременном сегменте с энергией или громкостью каждого кадра в качестве весового коэффициента, может быть частотой среза для этого кратковременного сегмента.

В альтернативном варианте или в дополнение к этому, как показано на фиг. 29, звуковой классификатор 200 может содержать выделитель 2012 признаков на уровне кадра для извлечения признаков на уровне кадра из звуковых кадров и классификатор 2014 на уровне кадра для классификации каждого из последовательности звуковых кадров по звуковым типам на уровне кадра с применением соответствующих признаков на уровне кадра, при этом выделитель 2022 кратковременных признаков объекта может выполняться с возможностью вычисления кратковременных признаков объекта в зависимости от результатов классификатора 2014 на уровне кадра в отношении последовательности звуковых кадров.

Другими словами, в дополнение к классификатору 202 звукового содержимого и классификатору 204 звукового контекста звуковой классификатор 200 может дополнительно содержать классификатор 201 кадров. В такой конфигурации классификатор 202 звукового содержимого классифицирует кратковременные сегменты в зависимости от результатов классификации на уровне кадра классификатора 201 кадров, и классификатор 204 звукового контекста классифицирует долговременный сегмент в зависимости от результатов кратковременной классификации классификатора 202 звукового содержимого.

Классификатор 2014 на уровне кадра может выполняться с возможностью классификации каждого из последовательности звуковых кадров по любым классам, которые могут называться "звуковые типы на уровне кадра". В одном варианте осуществления звуковые типы на уровне кадра могут иметь структуру, подобную структуре типов содержимого, рассмотренную выше, и также иметь значение подобное типам содержимого, и разница лишь в том, что звуковые типы на уровне кадра и типы содержимого классифицируются на разных уровнях звукового сигнала, то есть на уровне кадра и уровне кратковременного сегмента. Например, классификатор 2014 на уровне кадра может выполняться с возможностью классификации каждого из последовательности звуковых кадров по меньшей мере одного из следующих звуковых типов на уровне кадра: речи, музыки, фонового звука и шума. С другой стороны, звуковые типы на уровне кадра могут также иметь структуру, частично или полностью отличающуюся от структуры типов содержимого, более подходящую для классификации на уровне кадра и более подходящую для использования в качестве кратковременных признаков объекта для кратковременной классификации. Например, классификатор 2014 на уровне кадра может выполняться с возможностью классификации каждого из последовательности звуковых кадров по меньшей мере одного из следующих звуковых типов на уровне кадра: вокализированный, невокализованный и пауза.

Относительно получения кратковременных признаков объекта из результатов классификации на уровне кадра может быть принята аналогичная схема со ссылкой на описание в разделе 6.2.

В качестве альтернативы как кратковременные признаки объекта, зависящие от результатов классификатора 2014 на уровне кадра, так и кратковременные признаки объекта, непосредственно зависящие от признаков на уровне кадра, получаемые с помощью выделителя 2012 признаков на уровне кадра, могут применяться в кратковременном классификаторе 2024. Таким образом, выделитель 2022

кратковременных признаков объекта может выполняться с возможностью вычисления кратковременных признаков объекта в зависимости от как признаков на уровне кадра, извлеченных из последовательности звуковых кадров, так и результатов классификатора на уровне кадра в отношении последовательности звуковых кадров.

Другими словами, выделитель 2012 признаков на уровне кадра может выполняться с возможностью вычисления как статистических данных, аналогичных тем, которые рассматривались в разделе 6.2, так и тем кратковременным признакам объекта, описанным с использованием фиг. 28, включающим по меньшей мере один из следующих признаков: признаки, характеризующие свойства различных кратковременных звуковых типов, частоту среза, статические характеристики отношения сигнал/шум, сегментные характеристики отношения сигнал/шум, основные речевые дескрипторы и характеристики речевого тракта.

Для работы в реальном времени во всех вариантах осуществления выделитель 2022 кратковременных признаков объекта может выполняться с возможностью работы на кратковременных звуковых сегментах, образованных посредством скольжения скользящего окна во временном измерении долговременного звукового сегмента при предопределенной длине шага. В отношении скользящего окна для кратковременного звукового сегмента, а также звукового кадра и скользящего окна для долговременного звукового сегмента для подробностей может быть сделана ссылка на раздел 1.1.

6.4 Сочетание вариантов и сценариев применения

Аналогично части 1 все варианты осуществления и разновидности, рассмотренные выше, могут реализовываться в любом их сочетании, и любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции могут реализовываться как такие же или отдельные компоненты.

Например, любые два или более решений, описанных в разделах 6.1-6.3, могут сочетаться друг с другом. И любое из сочетаний может дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в частях 1-5 и других частях, которые будут описаны позже. В частности, блок 712 сглаживания типа, рассмотренный в части 1, может применяться в этой части в качестве компонента звукового классификатора 200 для сглаживания результатов классификатора 2014 кадра, или классификатора 202 звукового содержимого, или классификатора 204 звукового контекста.

Кроме того, таймер 916 может также служить в качестве компонента звукового классификатора 200 для избегания резкого изменения выходного сигнала звукового классификатора 200.

6.5 Способ звуковой классификации

Аналогично части 1 в процессе описания звукового классификатора в вариантах осуществления, приведенных выше в данной заявке, явно описаны также некоторые процессы или способы. Далее дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

В одном варианте осуществления, как показано на фиг. 30, предлагается способ звуковой классификации. Для идентификации долговременного звукового типа (то есть, типа контекста) долговременного звукового сегмента, состоящего из последовательности кратковременных звуковых сегментов (либо перекрывающихся, либо не перекрывающихся друг с другом), кратковременные звуковые сегменты прежде всего классифицируются (операция 3004) по кратковременным звуковым типам, то есть, типам содержимого, и долговременные признаки объекта получаются путем расчета (операция 3006) статистических данные результатов операции классификации

в отношении последовательности кратковременных сегментов в долговременном звуковом сегменте. Затем может осуществляться долговременная классификация (операция 3008) с использованием долговременных признаков объекта. Кратковременный звуковой сегмент может содержать последовательность звуковых кадров. Конечно, для идентификации кратковременного звукового типа кратковременных сегментов из них необходимо извлечь кратковременные признаки объекта (операция 3002).

Кратковременные звуковые типы (типы содержимого) могут включать речь, кратковременную музыку, фоновый звук и шум, но не ограничиваются ими.

Долговременные признаки объекта могут включать: среднее значение и дисперсию величин достоверности кратковременных звуковых типов, среднее значение и дисперсию, взвешенную по степени важности кратковременных сегментов, частоту появления каждого кратковременного звукового типа и частоту переключения между различными кратковременными звуковыми типами, но не ограничиваются ими.

В варианте, показанном на фиг. 31, могут получаться дополнительные долговременные признаки объекта (операция 3107), непосредственно зависящие от кратковременных признаков объекта последовательности кратковременных сегментов в долговременном звуковом сегменте. Такие дополнительные долговременные признаки объекта могут включать, но не ограничивается следующими статистическими данными кратковременных признаков объекта: среднее значение, дисперсия, средневзвешенное значение, взвешенная дисперсия, наибольшее среднее, наименьшее среднее и отношение между наибольшим средним и наименьшим средним.

Существую различные способы для извлечения кратковременных признаков объекта. Одним из них является непосредственное извлечение кратковременных признаков объекта из кратковременного звукового сегмента, который должен классифицироваться. Такие особенности включают ритмические характеристики, характеристики прерываний/приглушений и кратковременные признаки качества звукового сигнала, но не ограничиваются ими.

Второй способ заключается в извлечении признаков на уровне кадра из звуковых кадров, входящих в каждый кратковременный сегмент (операция 3201 на фиг. 32), и последующем расчете кратковременных признаков объекта в зависимости от признаков на уровне кадра, например расчете статистических данных признаков на уровне кадра в качестве кратковременных признаков объекта. Признаки на уровне кадра могут включать: признаки объекта, характеризующие свойства различных кратковременных звуковых типов, частоту среза, статические характеристики отношения сигнал/шум, сегментные характеристики отношения сигнал/шум, основные речевые дескрипторы и характеристики речевого тракта, но не ограничиваются ими. Признаки объекта, характеризующие свойства различных кратковременных звуковых типов могут дополнительно включать энергию кадра, спектральное распределение поддиапазонов, спектральный поток, коэффициенты косинусного преобразования Фурье для частот чистых тонов, низкую звуковую частоту, остаточную информацию, цветовой признак и частоту переходов через нуль.

Третий способ заключается в извлечении кратковременных признаков объекта способом, аналогичным извлечению долговременных признаков объекта: после извлечения признаков на уровне кадра из звуковых кадров в классифицируемом кратковременном сегменте (операция 3201), классификация каждого звукового кадра по звуковым типам на уровне кадра с использованием соответствующих признаков на уровне кадра (операция 32011 на фиг. 33); и кратковременные признаки могут

извлекаться (операция 3002) посредством вычисления кратковременных признаков объекта на основе звуковых типов на уровне кадра (факультативно включающих величины достоверности). Звуковые типы на уровне кадра могут иметь свойства и структуру, аналогичные кратковременным звуковым типам (типам содержимого), а также может включать речь, музыку, фоновый звук и шум.

Второй способ и третий способ могут комбинироваться вместе, как показано пунктирной стрелкой на фиг. 33.

Как упоминалось в части 1, и кратковременные звуковые сегменты, и долговременные звуковые сегменты могут выбираться посредством скользящего окна. То есть, операция извлечения кратковременных признаков объекта (операция 3002) может выполняться на кратковременном звуковом сегменте, образованном посредством скользящего окна, скользящего во временном измерении долговременного звукового сегмента при предопределенной длине шага, и операция извлечения долговременных признаков объекта (операция 3107), и операция вычисления статистических данных кратковременных звуковых типов (операция 3006) также могут выполняться на долговременном звуковом сегменте, образованном посредством скользящего окна, скользящего во временном измерении долговременного звукового сегмента при предопределенной длине шага.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их видоизменения могут представлять собой отдельные решения. Кроме того, любые два или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания. В частности, как уже упоминалось в разделе 6.4, сглаживающие схемы и схема переключения звуковых типов могут быть частью способа звуковой классификации, описанного в данной заявке.

Часть 7: Классификаторы VoIP и способы классификации

В части 6 предлагается новый звуковой классификатор для классификации звукового сигнала по типам звукового контекста, по меньшей мере частично, в зависимости от результатов классификации типа содержимого. В вариантах осуществления, рассматриваемых в части 6, долговременные признаки объекта извлекаются из долговременного сегмента длительностью от нескольких секунд до нескольких десятков секунд, таким образом, классификация звукового контекста может привести к долгому времени ожидания. Желательно, чтобы звуковой контекст мог также классифицироваться в реальном времени или почти в реальном времени, например, на уровне кратковременного сегмента..

7.1 Классификация контекста на основе кратковременного сегмента

Поэтому, предлагается, показанный на фиг. 34, звуковой классификатор 200A, содержащий классификатор 202A звукового содержимого для идентификации типа содержимого кратковременного сегмента звукового сигнала и классификатор 204A звукового контекста для идентификации типа контекста кратковременного сегмента, по меньшей мере частично, в зависимости от типа содержимого, идентифицированного классификатором звукового содержимого.

В данном случае классификатор 202A звукового содержимого может принимать методы, уже рассмотренные в части 6, но может также принимать различные методы, которые будут описаны ниже в разделе 7.2. Кроме того, классификатор 204A звукового

контекста может принимать методы, уже описанные в части 6, с той разницей, что классификатор 204A контекста может непосредственно использовать результаты классификатора 202A звукового содержимого, а не использовать статистические данные результатов из классификатора 202A звукового содержимого, поскольку классификатор 204A звукового контекста и классификатор 202A звукового содержимого классифицируют тот же кратковременный сегмент. Кроме того, аналогично части 6, в дополнение к результатам из классификатора 202A звукового содержимого классификатор 204A звукового контекста может использовать другие признаки объекта, непосредственно извлеченные из кратковременного сегмента. То есть, классификатор 204A звукового контекста может выполняться с возможностью классификации кратковременного сегмента на основе модели машинного обучения, используя в качестве признаков объекта величины достоверности типов содержимого кратковременного сегмента и другие признаки объекта, извлеченные из кратковременного сегмента. В отношении признаков объекта, извлеченных из кратковременного сегмента, может быть сделана ссылка на часть 6.

Классификатор 200A звукового содержимого может одновременно помечать кратковременный сегмент в качестве большего количества звуковых типов, чем VoIP-речь/шум и/или не VoIP-речь/шум (VoIP-речь/шум и не VoIP-речь/шум будут рассмотрены ниже в разделе 7.2), и каждый из множества звуковых типов может иметь свою собственную величину достоверности, как описано в разделе 1.2. Это позволяет достичь большей точности классификации, так как может быть получена более полная информация. Например, объединенная информация о величинах достоверности речи и кратковременной музыки показывает, в какой степени звуковое содержимое, вероятно, будет смесью речи и фоновой музыки, вследствие того, оно может отличаться от содержимого чистого VoIP.

7.2 Классификация с применением VoIP-речи и VoIP-шума

Данный аспект настоящей заявки особенно полезен в системе классификации VoIP/не VoIP, которая потребовалась бы для классификации текущего кратковременного сегмента для принятия решения с коротким временем ожидания.

Для этой цели, как показано на фиг. 34, звуковой классификатор 200A специально предназначен для классификации VoIP/не VoIP. Для классификации VoIP/не VoIP классификатор 2026 VoIP-речи и/или классификатор VoIP-шума разработаны для генерации промежуточных результатов для окончательной надежной классификации VoIP/не VoIP посредством классификатора 204A звукового контекста.

Кратковременный сегмент VoIP будет поочередно содержать VoIP-речь и VoIP-шум. Следует отметить, что высокая точность может достигаться для классификации кратковременного сегмента речи на VoIP-речь или не VoIP-речь, но не достигаться для классификации кратковременного сегмента шума на VoIP-шум или не VoIP-шум. Таким образом, можно сделать вывод, что будет нечеткость различимости с помощью непосредственной классификации кратковременного сегмента на VoIP (включая VoIP-речь и VoIP-шум, но с VoIP-речью и VoIP-шумом, неточно идентифицированными) и не VoIP без учета разницы между речью и шумом, и, следовательно, с признаками этих двух типов содержимого (речи и шума), смешанными вместе.

Это приемлемо для классификаторов для достижения более высоких точностей для классификации VoIP-речи/не VoIP-речи, чем для классификации VoIP-шума/не VoIP-шума, так как речь содержит больше информации, чем шум, и такие признаки объекта, как частота среза, более эффективны для классификации речи. В соответствии с весовым ранжированием, полученным из процесса обучения AdaBoost, кратковременными

признаками объекта с наибольшим весом для классификации VoIP/не VoIP-речи являются: стандартное отклонение логарифма энергии, частота среза, стандартное отклонение ритмической интенсивности и стандартное отклонение спектрального потока. Стандартное отклонение логарифма энергии, стандартное отклонение ритмической интенсивности и стандартное отклонение спектрального потока в большинстве случаев выше для VoIP-речи, чем для не VoIP-речи.

Одна из возможных причин заключается в том, что многие кратковременные сегменты речи в контексте не VoIP, такие как программный материал, подобный кинофильму, или игра, обычно смешиваются с другими звуками, такими как фоновая музыка или звуковой эффект, значения вышеуказанных признаков объекта которых являются ниже. Между тем, признак среза, в большинстве случаев ниже для VoIP-речи, чем для не VoIP-речи, что указывает на низкую частоту среза, применяемую во многих популярных кодеках VoIP.

Таким образом, в одном варианте осуществления классификатор 202A звукового содержимого может содержать классификатор 2026 VoIP-речи для классификации кратковременного сегмента как типа содержимого VoIP-речи или типа содержимого не VoIP-речи; и классификатор 204A звукового контекста может выполняться с возможностью классификации кратковременного сегмента как типа контекста VoIP или типа контекста не VoIP в зависимости от величин достоверности VoIP-речи и не VoIP-речи.

В другом варианте осуществления классификатор 202A звукового содержимого может дополнительно содержать классификатор 2028 VoIP-шума для классификации кратковременного сегмента как типа содержимого VoIP-шума или типа содержимого не VoIP-шума; и классификатор 204A звукового контекста может выполняться с возможностью классификации кратковременного сегмента как типа контекста VoIP или типа контекста не VoIP в зависимости от величин достоверности VoIP-речи, не VoIP-речи, VoIP-шума и не VoIP-шума.

Типы содержимого VoIP-речи, не VoIP-речи, VoIP-шума и не VoIP-шума могут идентифицироваться с помощью существующих методов, как описано в части 6, разделе 1.2 и разделе 7.1.

В альтернативном варианте классификатор 202A звукового содержимого может иметь иерархическую структуру, как показано на фиг. 35. То есть, мы пользуемся результатами классификатора 2025 речи/шума для классификации сначала кратковременных сегментов как речи или шума/фона.

На основе варианта осуществления с применением исключительно классификатора 2026 VoIP-речи, если кратковременный сегмент определяется как речь классификатором 2025 речи/шума (в такой ситуации это просто классификатор речи), то классификатор 2026 VoIP-речи продолжает классифицировать, является ли он VoIP-речью или не VoIP-речью, и вычисляет двоичный результат классификации; в противном случае можно считать, что величина достоверности VoIP-речи низкая, или решение о VoIP-речи является неопределенным.

На основе варианта осуществления с применением исключительно классификатора 2028 VoIP-шума, если кратковременный сегмент определяется как шум классификатором 2025 речи/шума (в такой ситуации это просто классификатор шума (фона)), то классификатор 2028 VoIP-шума продолжает классифицировать его на VoIP-шум или не VoIP-шум и вычисляет двоичный результат классификации. В противном случае можно считать, что величина достоверности VoIP-шума низкая, или решение о VoIP-шуме является неопределенным.

В данном случае, поскольку, как правило, речь является типом информативного содержимого, а шум/фон - типом мешающего содержимого, даже если кратковременный сегмент не является шумом, в варианте осуществления в предыдущем абзаце, мы не можем точно определить, что кратковременный сегмент не является типом контекста VoIP. В то время как, если кратковременный сегмент не является речью, в варианте осуществления с применением исключительно классификатора 2026 VoIP-речи, он, вероятно, не является типом контекста VoIP. Таким образом, в большинстве случаев вариант осуществления с применением исключительно классификатора 2026 VoIP-речи может реализовываться независимо, в то время как другой вариант осуществления с применением исключительно классификатора 2028 VoIP-шума может использоваться в качестве дополнительного варианта осуществления, взаимодействующего, например, с вариантом осуществления с применением классификатора 2026 VoIP-речи.

То есть, могут применяться как классификатор 2026 VoIP-речи, так и классификатор 2028 VoIP-шума. Если кратковременный сегмент определяется как речь классификатором 2025 речи/шума, то классификатор 2026 VoIP-речи продолжает классифицировать, является ли он VoIP-речью или не VoIP-речью, и вычисляет двоичный результат классификации. Если кратковременный сегмент определяется как шум классификатором 2025 речи/шума, то классификатор 2028 VoIP-шума продолжает классифицировать его как VoIP-шум или не VoIP-шум и вычисляет двоичный результат классификации. В противном случае, можно считать, что кратковременный сегмент может классифицироваться как не VoIP.

Реализация классификатора 2025 речи/шума, классификатор 2026 VoIP-речи и классификатор 2028 VoIP-шума могут принимать любые существующие методы, и могут представлять собой классификатор 202 звукового содержимого, рассмотренный в частях 1-6.

Если классификатор 202A звукового содержимого, реализованный в соответствии с описанием выше, окончательно классифицирует кратковременный сегмент ни как речь, шум и фон, или ни как VoIP-речь, не VoIP-речь, VoIP-шум и не VoIP-шум, что означает, что все значимые величины достоверности низкие, тогда классификатор 202A звукового содержимого (и классификатор 204A звукового контекста) может классифицировать кратковременный сегмент как не VoIP.

Для классификации кратковременного сегмента в качестве типов контекста VoIP или не VoIP в зависимости от результатов классификатора 2026 VoIP-речи и классификатора 2028 VoIP-шума, классификатор 204A звукового контекста может принять методы, основанные на машинном обучении, как описано в разделе 7.1, а в качестве модификации, может использоваться больше признаков объекта, в том числе кратковременные признаки объекта, непосредственно извлеченные из кратковременного сегмента и/или результаты другого классификатора(ов) звукового содержимого, направленного на другие типы содержимого, чем связанные с VoIP типы содержимого, что уже рассматривалось в разделе 7.1.

Кроме описанных выше методов на основе машинного обучения, альтернативный подход к классификации VoIP/не VoIP может представлять собой эвристическое правило, пользующееся знаниями в конкретной области и применяющее результаты классификации в отношении VoIP-речи и VoIP-шума. Пример такого эвристического правила будет показан ниже.

Если текущий кратковременный сегмент времени t определяется как VoIP-речь или не VoIP-речь, то результат классификации непосредственно принимается в качестве результата классификации VoIP/не VoIP, поскольку классификация VoIP/не VoIP-речи

является надежной, как обсуждалось ранее. То есть, если кратковременный сегмент определяется как VoIP-речь, то он имеет тип контекста VoIP; если кратковременный сегмент определяется как не VoIP-речь, то он имеет тип контекста не VoIP.

Когда классификатор 2026 VoIP-речи принимает бинарное решение относительно VoIP-речи/не VoIP-речи в отношении к речи, определяемой классификатором 2025 речи/шума, как упоминалось выше, величины достоверности VoIP-речи и не VoIP-речи могут быть взаимодополняющими, то есть их сумма равна 1 (если 0 означает 100% нет, а 1 означает 100% да), и пороговое значение величины достоверности для разграничения VoIP-речи и не VoIP-речи может указывать в действительности ту же точку. Если классификатор 2026 VoIP-речи не является двоичным классификатором, величины достоверности VoIP-речи и не VoIP-речи могут не быть взаимодополняющими, и пороговое значение величины достоверности для разграничения VoIP-речи и не VoIP-речи не обязательно указывает на ту же точку.

Тем не менее, в случае, когда достоверность VoIP-речи или не VoIP-речи близка к и колеблется вокруг порогового значения, переключение результатов VoIP/не VoIP классификации возможно слишком часто. Чтобы избежать такой неустойчивости, может предусматриваться буферная схема: оба пороговых значения для VoIP-речи и для не VoIP-речи могут устанавливаться большими, настолько, чтобы было не так легко переключаться с текущего типа содержимого к другому типу содержимого. Для простоты описания, можно преобразовать величину достоверности для не VoIP-речи к величине достоверности VoIP-речи. То есть, если величина достоверности является высокой, то кратковременный сегмент рассматривается в качестве более близкого к VoIP-речи, а если величина достоверности является низкой, то кратковременный сегмент рассматривается как более близкий к не VoIP-речи. Хотя для недвоичного классификатора, как описано выше, высокая величина достоверности не VoIP-речи не обязательно означает низкую величину достоверности VoIP-речи, такое упрощение вполне может отражать суть решения и значимые требования, описанные языком бинарных классификаторов, должны толковаться как охватывающие эквивалентные решения для недвоичных классификаторов.

Буферная схема показана на фиг. 36. Существует буферная зона между двумя пороговыми значениями Th1 и Th2 (Th1>=Th2). Когда величина достоверности v(t) VoIP-речи падает в зоне, классификация контекста не изменится, как показано стрелками на левой и правой сторонах фиг. 36. Только тогда, когда величина достоверности v(t) больше, чем большее пороговое значение Th1, кратковременный сегмент будет классифицирован как VoIP (как показано стрелкой в нижней части фиг. 36); и только тогда, когда величина достоверности не больше, чем меньшее пороговое значение Th2, кратковременный сегмент будет классифицирован как не VoIP (как показано стрелкой в верхней части фиг. 36).

Ситуация аналогична, если вместо этого применяется классификатор 2028 VoIP-шума. Для принятия более надежного решения классификатор 2026 VoIP-речи и классификатор 2028 VoIP-шума могут применяться совместно. Затем классификатор 204A звукового контекста может выполняться с возможностью: классификации кратковременного сегмента в качестве типа контекста VoIP, если величина достоверности VoIP-речи больше, чем первое пороговое значение, или если величина достоверности VoIP-шума больше, чем третье пороговое значение; классификации кратковременного сегмента в качестве типа контекста не VoIP, если величина достоверности VoIP-речи не больше, чем второе пороговое значение, при этом второе пороговое значение не больше, чем первое пороговое значение, или если величина

достоверности VoIP-шума не больше, чем четвертое пороговое значение, при этом четвертое пороговое значение не больше, чем третье пороговое значение; в противном случае, классификации кратковременного сегмента в качестве типа контекста предыдущего кратковременного сегмента.

В данном случае первое пороговое значение может быть равно второму пороговому значению, а третье пороговое значение может быть равно четвертому пороговому значению, особенно для двоичного классификатора VoIP-речи и двоичного классификатора VoIP-шума, но не ограничиваясь ими. Однако, поскольку, в большинстве случаев результат классификации VoIP-шума не такой надежный, было бы лучше, если третье и четвертое пороговое значения были бы не равны друг другу, и оба должны быть далекими от 0,5 (0 указывает на высокую достоверность не VoIP-шума, а 1 указывает на высокую достоверность VoIP-шума).

7.3 Сглаживание колебаний

Для избежания быстрых колебаний другое решение состоит в том, чтобы сгладить величины достоверности, определенные классификатором звукового содержимого. Поэтому, как показано на фиг. 37, блок 203A сглаживания типа может входить в звуковой классификатор 200A. Для величины достоверности каждого из 4 типов содержимого, связанного с VoIP, как описывалось ранее, могут приниматься сглаживающие схемы, описанные в разделе 1.3.

Кроме того, аналогично разделу 7.2, VoIP-речь и не VoIP-речь могут рассматриваться как пара, имеющая взаимодополняющие величины достоверности; и VoIP-шум и не VoIP-шум могут также рассматриваться как пара, имеющая взаимодополняющие величины достоверности. В такой ситуации, только один выходной сигнал из каждой пары должен сглаживаться, и могут приниматься сглаживающие схемы, рассмотренные в разделе 1.3.

Возьмем в качестве примера величину достоверности VoIP-речи, формула (3) может быть переписана в виде:

$$v(t) = \beta \cdot v(t-1) + (1-\beta) \cdot voipSpeechConf(t) \qquad (3'')$$

где ν(t) - сглаженная величина достоверности VoIP-речи в момент времени t, ν(t-1) - сглаженная величина достоверности VoIP-речи в предыдущее время и $voipSpeechConf$ - достоверность VoIP-речи в текущий момент времени t до сглаживания, α - весовой коэффициент.

В варианте, если присутствует классификатор 2025 речи/шума, как описано выше, если величина достоверности речи для кратковременного сегмента низкая, то кратковременный сегмент не может надежно классифицироваться как VoIP-речь, и мы можем непосредственно установить voipSpeechConf (t)=ν(t-1), не заставляя классификатор 2026 VoIP-речи работать в действительности.

Кроме того, в ситуации, описанной выше, мы могли бы установить voipSpeechConf (t) = 0,5 (или другое значение, не превышающее 0,5, например, 0,4-0,5), указывающее на неопределенный случай (в данном случае, достоверность = 1 указывает на высокую достоверность, что это VoIP, а достоверность = 0 указывает на высокую достоверность, что это не VoIP).

Таким образом, в соответствии с вариантом, показанным на фиг. 37, классификатор 200A звукового содержимого может дополнительно содержать классификатор 2025 речи/шума для идентификации типа содержимого речи кратковременного сегмента, и блок 203A сглаживания типа может выполняться с возможностью установки значения достоверности VoIP-речи для текущего кратковременного сегмента до сглаживания в

качестве предопределенной величины достоверности (такой как 0,5 или другое значение, например, 0,4-0,5) или сглаженной величины достоверности предыдущего кратковременного сегмента, где величина достоверности для типа содержимого речи классифицированная классификатором речь/шум ниже, чем пятое пороговое значение. В такой ситуации классификатор 2026 VoIP-речи может работать или может не работать. Альтернативная установка величины достоверности может делаться с помощью классификатора 2026 VoIP-речи, это эквивалентно решению, где работа выполняется блоком 203A сглаживания типа, и формула изобретения должна толковаться как охватывающая обе ситуации. Кроме того, в данном случае мы используем выражение "величина достоверности для типов содержимого речи как классифицированная классификатором речи/шума ниже, чем пятое пороговое значение", но объем охраны не ограничивается этим, и это эквивалентно тому, где кратковременный сегмент классифицируется по другим типам содержимого, чем речь.

Для величины достоверности VoIP-шума ситуация аналогична и подробное описание в данном случае опущено.

Для избежания быстрых колебаний, еще одно решение состоит в том, чтобы сгладить величину достоверности как определенную классификатором 204A звукового контекста, и могут приниматься сглаживание схемы, описанные в разделе 1.3.

Для избежания быстрых колебаний, еще одно решение состоит в том, чтобы задержать переключение типа контекста между VoIP и не VoIP, и могут применяться те же схемы, что описаны в разделе 1.6. Как описано в разделе 1.6, таймер 916 может находиться вне звукового классификатора или в пределах звукового классификатора в качестве его части. Поэтому, как показано на фиг. 38, звуковой классификатор 200A может дополнительно содержать таймер 916. И звуковой классификатор выполнен с возможностью продолжения вывода текущего типа контекста до тех пор, пока длительность времени неизменности нового типа контекста не достигнет шестого пороговое значения (тип контекста является частным случаем звукового типа). Ссылаясь на раздел 1.6, подробное описание в данном случае может быть опущено.

В дополнение или в качестве альтернативы, в качестве еще одной схемы задержки переключения между VoIP и не VoIP, первое и/или второе пороговое значение, как описано выше, для классификации VoIP/не VoIP могут быть разными в зависимости от типа контекста предыдущего кратковременного сегмента. То есть, первое и/или второе пороговое значение становится больше, когда тип контекста нового кратковременного сегмента отличается от типа контекста предыдущего кратковременного сегмента, и становится меньше, когда тип контекста нового кратковременного сегмента такой же, как типа контекста предыдущего кратковременного сегмента. Таким образом, тип контекста имеет склонность сохранять текущий тип контекста и, таким образом резкое колебание типа контекста может быть подавлено до некоторой степени.

7.4 Сочетание вариантов осуществления и сценариев применения

Аналогично части 1 все варианты осуществления и разновидности, рассмотренные выше, могут реализовываться в любом их сочетании, и любые компоненты, упоминаемые в разных частях/вариантах осуществления, но имеющие одинаковые или подобные функции могут реализовываться как такие же или отдельные компоненты.

Например, любые два или более решений, описанных в разделах 7.1-7.3, могут сочетаться друг с другом. И любое из сочетаний может дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в частях 1-6. В частности, варианты осуществления, рассмотренные в этой части, и любое их сочетание могут сочетаться с вариантами осуществления устройства/способа обработки звукового

сигнала или контроллера выравнивателя громкости /способа управления, описанных в части 4.

7.5 Способ классификации VoIP

Аналогично части 1 в процессе описания звукового классификатора в вариантах осуществления, приведенных выше в данной заявке, явно описаны также некоторые процессы или способы. Далее дается краткий обзор этих способов без повторения некоторых подробностей, которые уже рассматривались выше.

В одном варианте осуществления, как показано на фиг. 39, способ звуковой классификации включает идентификацию типа содержимого в кратковременном сегменте звукового сигнала (операция 4004), а затем идентификацию типа контекста кратковременного сегмента, по меньшей мере частично, в зависимости от идентифицированного типа содержимого (операции 4008).

Для динамичной и быстрой идентификации типа контекста звукового сигнала способ звуковой классификации в этой части особенно результативен в идентификации типа контекста VoIP и не VoIP. В такой ситуации кратковременный сегмент может прежде всего классифицироваться на тип содержимого VoIP-речи или тип содержимого не VoIP-речи, и операция идентификации типа контекста выполнена с возможностью классификации кратковременного сегмента на тип контекста VoIP или тип контекста не VoIP в зависимости от величин достоверности VoIP-речи и не VoIP-речи.

Кроме того, кратковременный сегмент может прежде всего классифицироваться на тип содержимого VoIP-шума или тип содержимого не VoIP-шума, и операция идентификации типа контекста может выполняться с возможностью классификации кратковременного сегмента на тип контекста VoIP или тип контекста не VoIP в зависимости от величин достоверности VoIP-шума и не VoIP-шума.

Речь и шум могут учитываться совместно. В такой ситуации, операция идентификации типа контекста может выполняться с возможностью классификации кратковременного сегмента на тип контекста VoIP или тип контекста не VoIP в зависимости от величин достоверности VoIP-речи, не VoIP-речи, VoIP-шума и не VoIP-шума.

Для идентификации типа контекста кратковременного сегмента может использоваться модель машинного обучения, принимая как величины достоверности типов содержимого кратковременного сегмента, так и другие признаки объекта, извлеченные из кратковременного сегмента в качестве признаков объекта.

Операция идентификации типа контекста может реализовываться также на основе эвристических правил. Когда содержится только VoIP-речь и не VoIP-речь, эвристическое правило представляет собой следующее: классифицировать кратковременный сегмент в качестве типа контекста VoIP, если величина

достоверности VoIP-речи больше, чем первое пороговое значение; классифицировать кратковременный сегмент в качестве типа контекста не VoIP, если величина достоверности VoIP-речи не больше, чем второе пороговое значение, причем второе пороговое значение не больше, чем первое пороговое значение; в противном случае, классифицировать кратковременный сегмент в качестве типа контекста предыдущего кратковременного сегмента.

Эвристическое правило для ситуации, когда содержатся только VoIP-шум и не VoIP-шум, аналогично.

Когда содержатся и речь, и шум, эвристическое правило представляет собой следующее: классифицировать кратковременный сегмент в качестве типа контекста VoIP, если величина достоверности VoIP-речи больше, чем первое пороговое значение, или если величина достоверности VoIP-шума больше, чем третье пороговое значение;

классифицировать кратковременный сегмент в качестве типа контекста не VoIP, если величина достоверности VoIP-речи не больше, чем второе пороговое значение, причем второе пороговое значение не больше, чем первое пороговое значение, или если величина достоверности VoIP-шума не больше, чем четвертое пороговое значение, причем четвертое пороговое значение не больше, чем третье пороговое значение; в противном случае, классифицировать кратковременный сегмент в качестве типа контекста предыдущего кратковременного сегмента.

Сглаживающая схема, рассмотренная в разделе 1.3 и разделе 1.8, может приниматься в данном случае, а подробное описание опущено. В качестве модификации к сглаживающей схеме, описанной в разделе 1.3, перед операцией сглаживания 4106 способ может дополнительно включать идентификацию типа содержимого речи из кратковременного сегмента (операция 40040 на фиг. 40), при этом величина достоверности VoIP-речи для текущего кратковременного сегмента до сглаживания установлена в качестве предопределенной величины достоверности или сглаженной величины достоверности предыдущего кратковременного сегмента (операция 40044 на фиг. 40), при этом величина достоверности для типа содержимого речи ниже, чем пятое пороговое значение ("N" в операции 40041).

Если в противном случае операция идентификации типа содержимого речи надежно оценивает кратковременный сегмент как речь ("Y" в операции 40041), то кратковременный сегмент дополнительно классифицируется на VoIP-речь или не VoIP-речь (операция 40042) перед операцией сглаживания 4106.

В действительности, даже без использования сглаживающей схемы способ может также идентифицировать тип содержимого речи и/или шума, сначала, когда кратковременный сегмент классифицируется как речь или шум, дальнейшая классификация реализована для классификации кратковременного сегмента как VoIP-речь или не VoIP-речь, или как VoIP-шум или не VoIP-шум. Затем выполняется операция идентификации типа контекста.

Как отмечалось в разделе 1.6 и разделе 1.8, схема переключения, рассмотренная в них, может приниматься в качестве части способа звуковой классификации, описанном здесь, а подробности опущены. Кратко, способ может дополнительно включать измерение времени неизменности, в течение которого операция идентификации типа контекста непрерывно выводит тот же тип контекста, при этом способ звуковой классификации выполнен с возможностью продолжения выводить текущий тип контекста до тех пор, пока продолжительность времени непрерывности новый типа контекста не достигает шестого порогового значения.

Подобным образом разные шестые пороговые значения могут устанавливаться разных пар переключения от одного типа контекста к другому типу контекста. Помимо этого, шестое пороговое значение может отрицательно коррелировать с величиной достоверности нового типа контекста.

В качестве модификации схемы переключения в способе звуковой классификации специально направленном на VoIP/не VoIP классификацию любое одно или несколько из первого по четвертое порогового значения для текущего кратковременного сегмента может быть установлено разным в зависимости от типа контекста предыдущего кратковременного сегмента.

Подобно вариантам осуществления устройства обработки звукового сигнала, любое сочетание вариантов осуществления способа обработки звукового сигнала и их видоизменения, с одной стороны, применяются на практике; а, с другой стороны, каждый аспект вариантов осуществления способа обработки звукового сигнала и их

видоизменения могут представлять собой отдельные решения. Кроме того, любые два или более решений, описанных в этом разделе, могут сочетаться друг с другом, и эти сочетания могут дополнительно сочетаться с любым вариантом осуществления, описанным или подразумеваемым в других частях настоящего описания. В частности, метод звуковой классификации, описанный в данной заявке, может применяться в способе обработки звукового сигнала, описанном выше, в частности, в способе управления выравнивателем громкости.

Как упоминалось в начале подробного описания настоящей заявки, вариант осуществления заявки может реализовываться либо аппаратными средствами, либо программными средствами, или обоими. На фиг. 41 показана структурная схема, иллюстрирующая примерную систему для реализации аспектов настоящей заявки.

На фиг. 41, центральный процессор (CPU) 4201 выполняет различные процессы в соответствии с программой, сохраненной в постоянном запоминающем устройстве (ROM) 4202 или программой, загруженной из секции 4208 хранения данных в запоминающее устройство 4203 с произвольной выборкой (RAM). В RAM 4203, данные, необходимые при выполнении CPU 4201 различных процессов и тому подобного, также хранятся в соответствии с требованиями.

CPU 4201, ROM 4202 и RAM 4203 соединены друг с другом через шину 4204. Интерфейс 4205 ввода/вывода также подключен к шине 4204.

Следующие компоненты соединены с интерфейсом 4205 ввода/вывода: секция 4206 ввода данных, включающая клавиатуру, мышь или тому подобное; секция 4207 вывода данных, включающая дисплей, например, электронно-лучевую трубку (CRT), жидкокристаллический дисплей (LCD) или тому подобное и громкоговоритель или тому подобное; секция 4208 хранения данных в том числе жесткий диск или тому подобное; и секцию связи 4209 в том числе сетевой карты, таких как карты LAN, модем, и тому подобное. Секция 4209 обмена данными выполняет процесс обмена данными через сеть, такую как Интернет.

Накопитель 4210 также соединен с интерфейсом 4205 ввода/вывода в соответствии с требованиями. Съемный носитель 4211, такой как магнитный диск, оптический диск, магнитооптический диск, полупроводниковое запоминающее устройство или тому подобное, устанавливается на накопитель 4210 по мере необходимости таким образом, чтобы компьютерная программа, считанная с него, устанавливалась в секцию 4208 хранения данных в соответствии с требованиями.

В случае, когда описанные выше компоненты реализованы с помощью программного обеспечения, программа, которая представляет собой программное средство, устанавливается из сети, такой как Интернет, или носителя данных, такого как съемный носитель 4211.

Обратите внимание, что используемая в данной заявке терминология предназначена для описания конкретных вариантов осуществления и не предназначена для ограничения заявки. В данном контексте формы единственного числа предназначены также для включения форм множественного числа, если из контекста явно не следует иное. Следует также понимать, что термины "содержит" и/или "содержащий" при использовании в данном описании указывают на наличие указанных признаков, целых чисел, операций, этапов, элементов и/или компонентов, но не исключают присутствия или добавления одного или нескольких других признаков, целых чисел, операций, этапов, элементов, компонентов и/или их групп.

Соответствующие структуры, материалы, действия и эквиваленты всех средств или элементов операций плюс функций в поле ниже формулы изобретения включают любую

структуру, материал или действие для выполнения функции в сочетании с другими заявленными элементами в качестве особым образом заявленных. Описание настоящей заявки было представлено в целях иллюстрации и описания, но не предназначено быть исчерпывающим или ограничивающим заявку в описанной форме. Многие модификации и видоизменения будут понятны специалистам в данной области техники без отступления от объема и сущности данной заявки. Вариант осуществления был выбран и описан с целью наилучшего объяснения принципов заявки и практического применения и обеспечения возможности другим специалистам в данной области техники понимать заявку в различных вариантах осуществления с различными модификациями, подходящими для конкретно предусмотренного применения.

(57) Формула изобретения

1. Контроллер выравнивателя громкости, содержащий:

классификатор звукового содержимого для идентификации типа содержимого звукового сигнала в реальном времени и

регулирующий блок для регулировки выравнивателя громкости в непрерывном режиме в зависимости от идентифицированного типа содержимого.

2. Устройство обработки звукового сигнала, содержащее контроллер выравнивателя громкости по п. 1.

3. Способ звуковой классификации, включающий:

идентификацию типа содержимого в кратковременном сегменте звукового сигнала и

идентификацию типа контекста кратковременного сегмента в зависимости, по меньшей мере частично, от идентифицированного типа содержимого.

4. Способ звуковой классификации по п. 3, отличающийся тем, что операция классификации типа содержимого включает классификацию кратковременного сегмента на тип содержимого VoIP-речи или тип содержимого не VoIP-речи и

операция идентификации типа контекста организована с возможностью классификации кратковременного сегмента на тип контекста VoIP или тип контекста не VoIP в зависимости от величин достоверности VoIP-речи и не VoIP-речи.

5. Способ звуковой классификации по п. 4, отличающийся тем, что операция классификации типа содержимого дополнительно включает

классификацию кратковременного сегмента на тип содержимого VoIP-шума или тип содержимого не VoIP-шума и

операция идентификации типа контекста организована с возможностью классификации кратковременного сегмента на тип контекста VoIP или тип контекста не VoIP в зависимости от величин достоверности VoIP-речи, не VoIP-речи, VoIP-шума и не VoIP-шума.

6. Способ звуковой классификации по п. 4, отличающийся тем, что операция идентификации типа контекста организована с возможностью:

классификации кратковременного сегмента в качестве типа контекста VoIP, если величина достоверности VoIP-речи больше, чем первое пороговое значение;

классификации кратковременного сегмента в качестве типа контекста не VoIP, если величина достоверности VoIP-речи не больше, чем второе пороговое значение, причем второе пороговое значение не больше, чем первое пороговое значение; в противном случае,

классификации кратковременного сегмента как типа контекста для предыдущего кратковременного сегмента.

7. Способ звуковой классификации по п. 3, отличающийся тем, что операция идентификации типа контекста организована с возможностью:

классификации кратковременного сегмента в качестве типа контекста VoIP, если величина достоверности VoIP-речи больше, чем первое пороговое значение, или если величина достоверности VoIP-шума больше, чем третье пороговое значение;

классификации кратковременного сегмента в качестве типа контекста не VoIP, если величина достоверности VoIP-речи не больше, чем второе пороговое значение, причем второе пороговое значение не больше, чем первое пороговое значение; или если величина достоверности VoIP-шума не больше, чем четвертое пороговое значение, причем четвертое пороговое значение не больше, чем третье пороговое значение; в противном случае,

классификации кратковременного сегмента как типа контекста для предыдущего кратковременного сегмента.

8. Способ звуковой классификации по п. 4, дополнительно включающий сглаживание величины достоверности типа содержимого в текущее время в зависимости от предыдущих величин достоверности типа содержимого.

9. Способ звуковой классификации по п. 8, отличающийся тем, что операция сглаживания организована с возможностью определения сглаженной величины достоверности текущего кратковременного сегмента посредством вычисления взвешенной суммы величины достоверности текущего кратковременного сегмента и сглаженной величины достоверности предыдущего кратковременного сегмента.

10. Способ звуковой классификации по п. 9, дополнительно включающий идентификацию типа содержимого речи из кратковременного сегмента, причем величину достоверности VoIP-речи для текущего кратковременного сегмента до сглаживания устанавливают в качестве предопределенной величины достоверности или сглаженной величины достоверности предыдущего кратковременного сегмента, где величина достоверности для типа содержимого речи ниже, чем пятое пороговое значение.

11. Способ звуковой классификации по п. 3, отличающийся тем, что операция идентификации типа контекста организована с возможностью классификации кратковременного сегмента на основе модели машинного обучения с применением в качестве признаков объекта величин достоверности типов содержимого кратковременного сегмента и других признаков объекта, извлеченных из кратковременного сегмента.

12. Способ звуковой классификации по п. 6, дополнительно включающий измерение времени неизменности, в течение которого операция идентификации типа контекста непрерывно выводит тот же тип контекста, при этом способ звуковой классификации организован с возможностью продолжения выводить текущий тип контекста до тех пор, пока продолжительность времени неизменности нового типа контекста не достигает шестого порогового значения.

13. Способ звуковой классификации по п. 12, отличающийся тем, что разные шесть пороговых значений устанавливают для разных пар переключения от одного типа контекста к другому типу контекста.

14. Способ звуковой классификации по п. 12, отличающийся тем, что шестое пороговое значение отрицательно коррелирует с величиной достоверности нового типа контекста.

15. Способ звуковой классификации по п. 6, отличающийся тем, что первое и/или второе пороговое значение отличаются в зависимости от типа контекста предыдущего кратковременного сегмента.

16. Звуковой классификатор, содержащий:

классификатор звукового содержимого для идентификации типа содержимого кратковременного сегмента звукового сигнала и

классификатор звукового контекста для идентификации типа контекста кратковременного сегмента по меньшей мере частично в зависимости от типа содержимого, идентифицированного классификатором звукового содержимого;

при этом звуковой классификатор настроен на осуществление способа по п. 3.

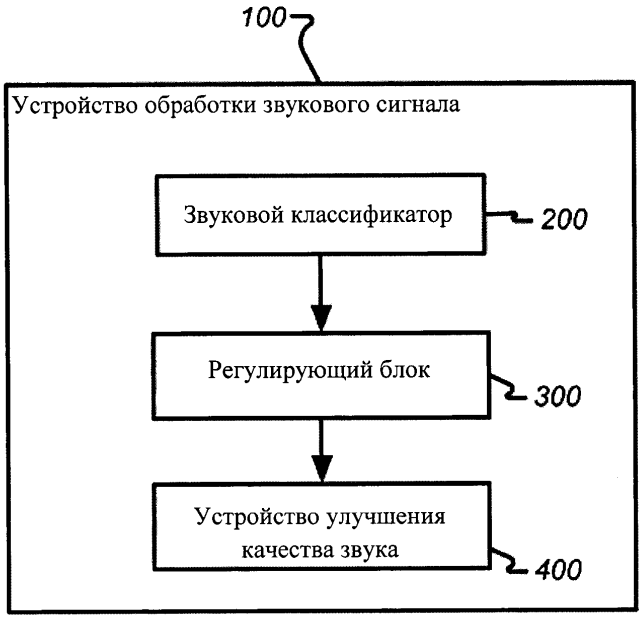
17. Устройство обработки звукового сигнала, содержащее звуковой классификатор, содержащий:

классификатор звукового содержимого для идентификации типа содержимого кратковременного сегмента звукового сигнала и

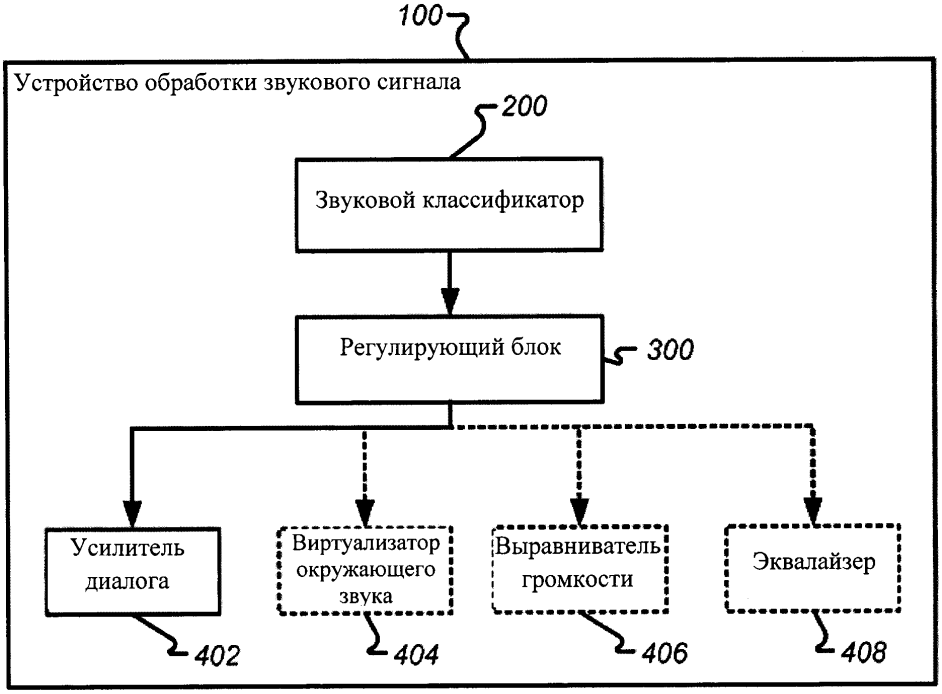
классификатор звукового контекста для идентификации типа контекста кратковременного сегмента по меньшей мере частично в зависимости от типа содержимого, идентифицированного классификатором звукового содержимого по п. 3.

18. Энергонезависимый машиночитаемый носитель с командами, хранящимися на нем, который при выполнении одним или более процессорами заранее определяет способ звуковой классификации по п. 3.

1

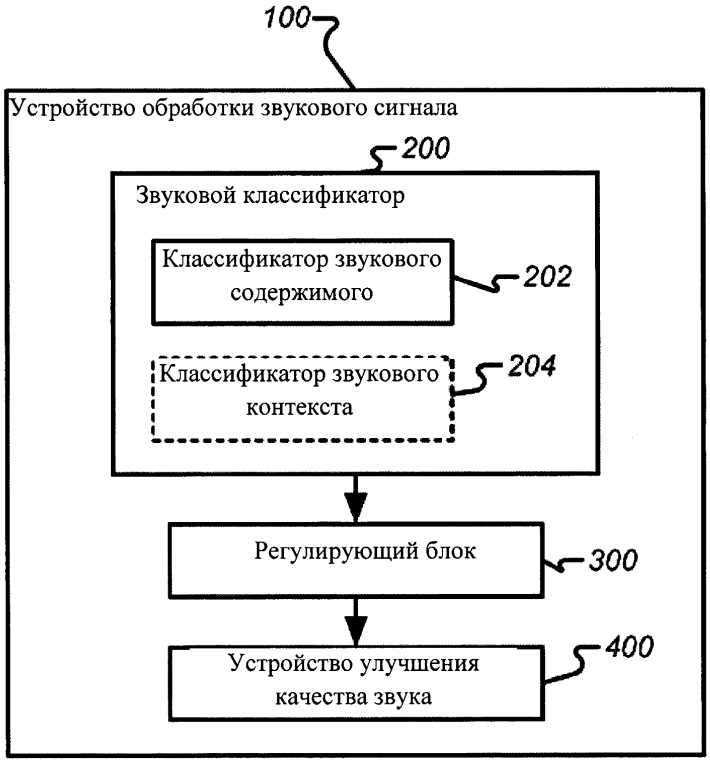

100
Устройство обработки звукового сигнала
Звуковой классификатор
200
Регулирующий блок
300
Устройство улучшения качества звука
400

Фиг. 1


100
Устройство обработки звукового сигнала
200
Звуковой классификатор
Регулирующий блок
300
Усилитель диалога
402
Виртуализатор окружающего звука
404
Выравниватель громкости
406
Эквалайзер
408

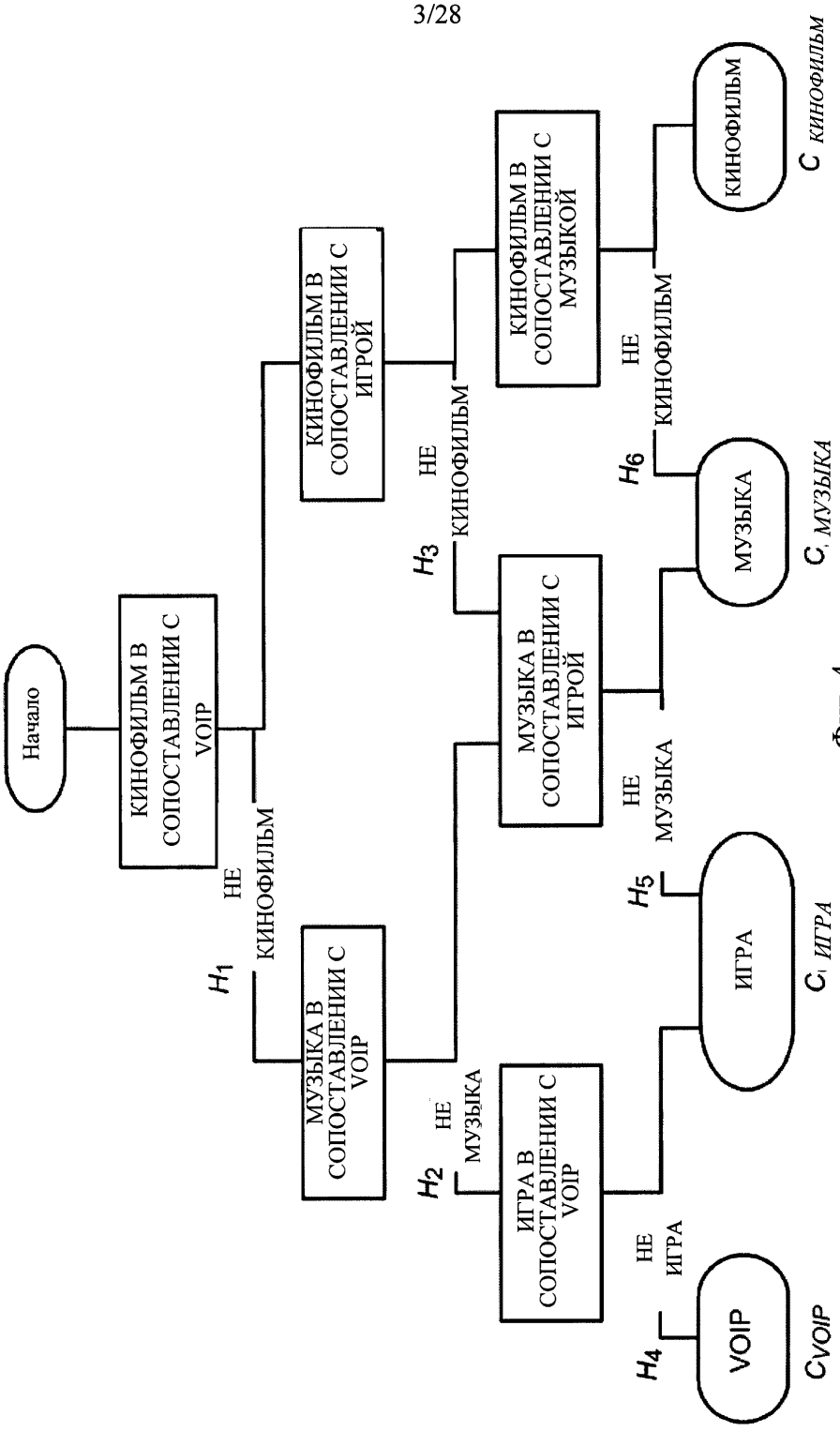
Фиг. 2

2


100
Устройство обработки звукового сигнала
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
Регулирующий блок
300
Устройство улучшения качества звука
400

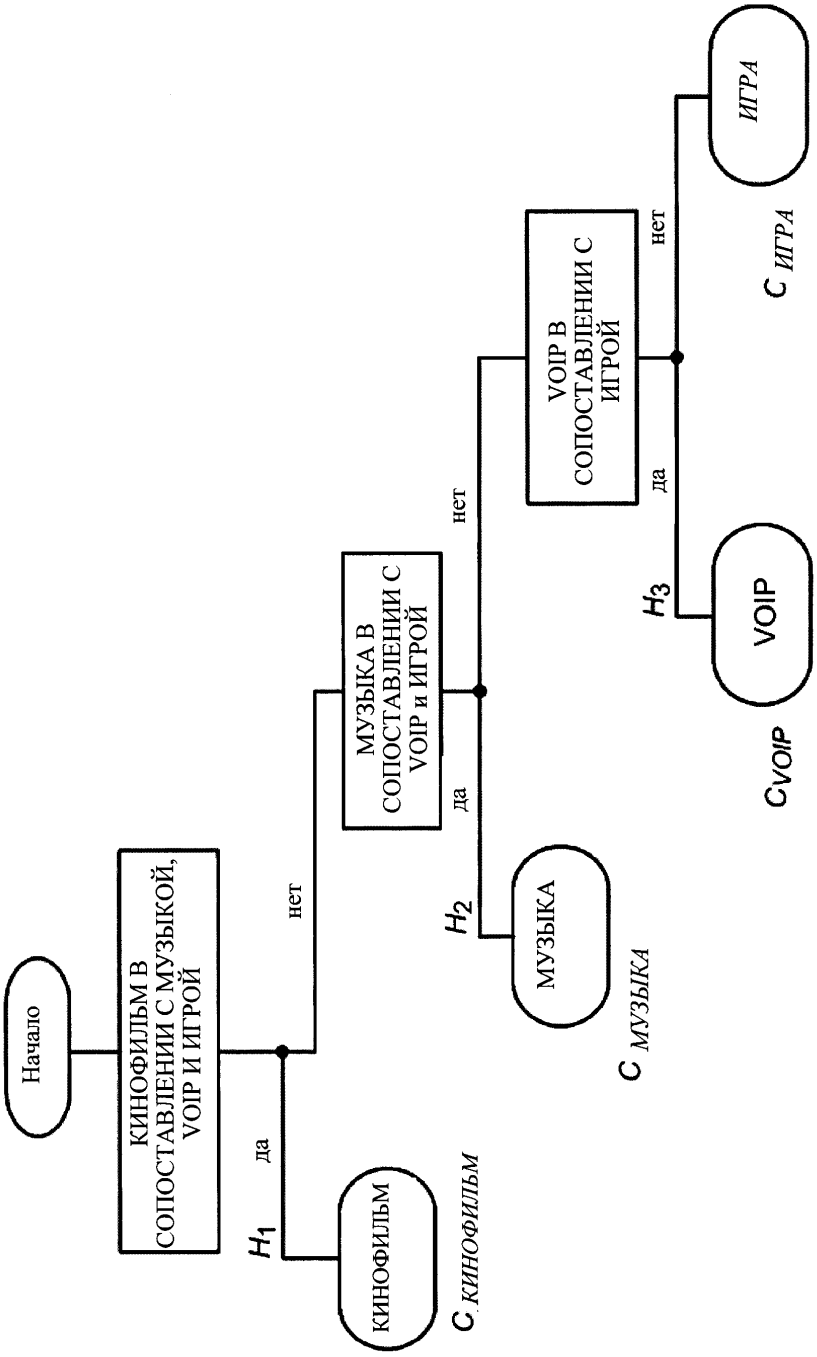
Фиг. 3

3

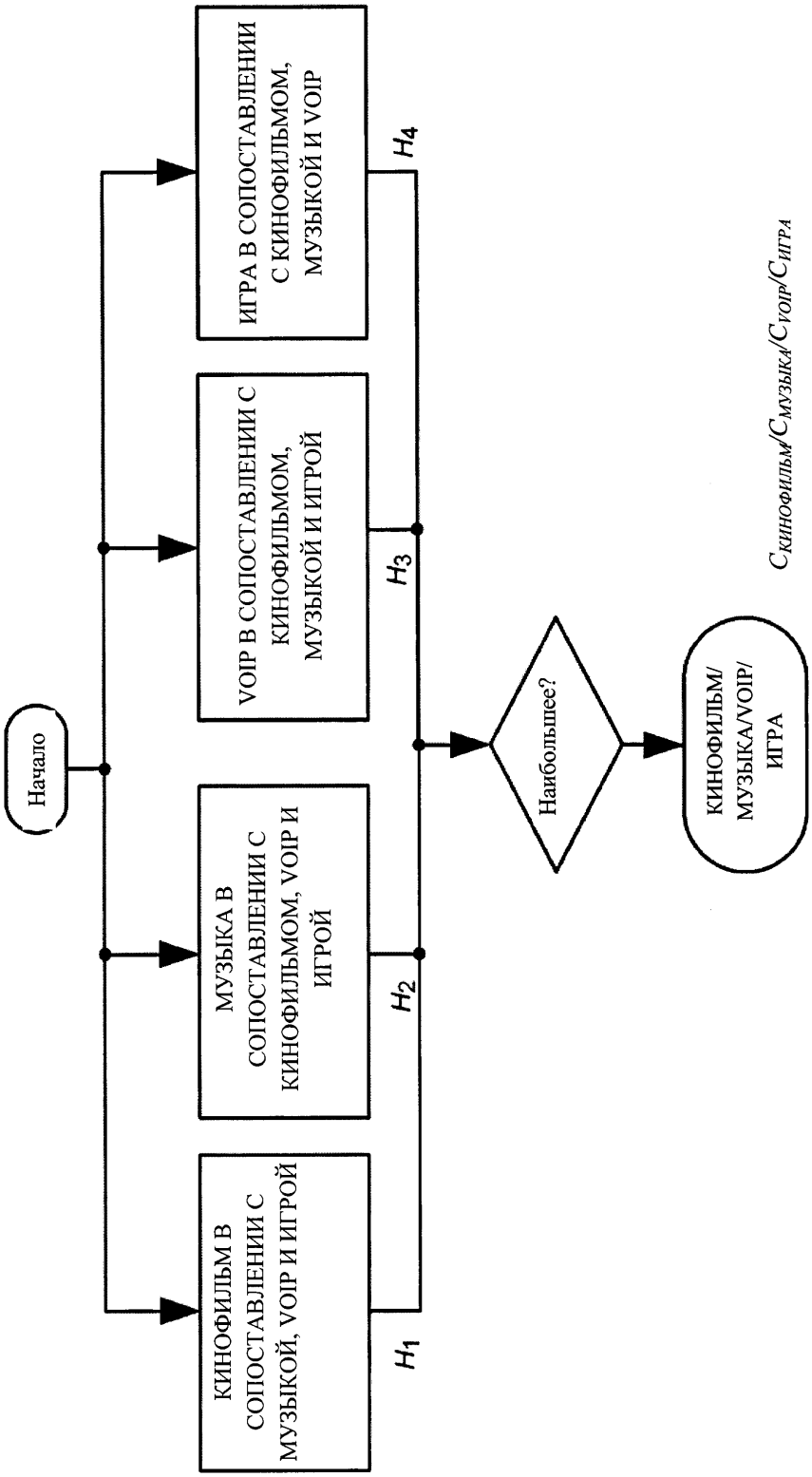

Начало
КИНОФИЛЬМ В СОПОСТАВЛЕНИИ С VOIP
НЕ КИНОФИЛЬМ
H1
МУЗЫКА В СОПОСТАВЛЕНИИ С VOIP
КИНОФИЛЬМ В СОПОСТАВЛЕНИИ С ИГРОЙ
НЕ КИНОФИЛЬМ
H3
МУЗЫКА В СОПОСТАВЛЕНИИ С ИГРОЙ
КИНОФИЛЬМ В СОПОСТАВЛЕНИИ С МУЗЫКОЙ
НЕ КИНОФИЛЬМ
H6
КИНОФИЛЬМ
C КИНОФИЛЬМ
НЕ МУЗЫКА
H5
МУЗЫКА
C, МУЗЫКА
H2
НЕ МУЗЫКА
ИГРА В СОПОСТАВЛЕНИИ С VOIP
НЕ ИГРА
H4
ИГРА
C, ИГРА
VOIP
CVOIP

Фиг. 4

4


Начало
КИНОФИЛЬМ В СОПОСТАВЛЕНИИ С МУЗЫКОЙ, VOIP И ИГРОЙ
да
нет
H1
КИНОФИЛЬМ
C КИНОФИЛЬМ
МУЗЫКА В СОПОСТАВЛЕНИИ С VOIP и ИГРОЙ
да
нет
H2
МУЗЫКА
C МУЗЫКА
VOIP В СОПОСТАВЛЕНИИ С ИГРОЙ
да
нет
H3
VOIP
CVOIP
ИГРА
C ИГРА

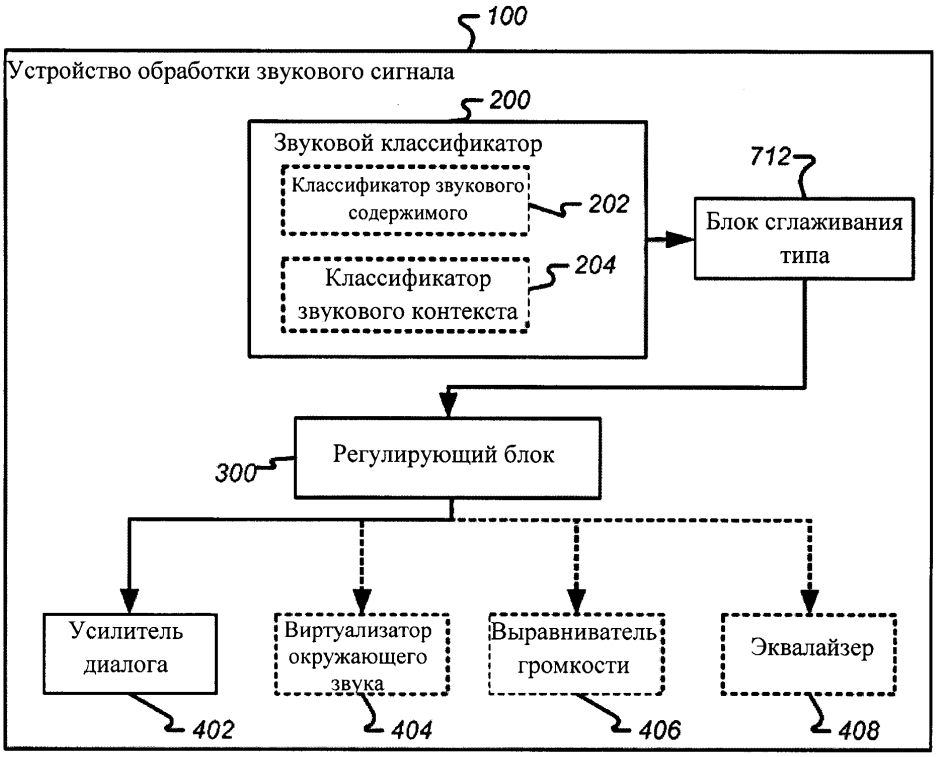
Фиг. 5


Начало
КИНОФИЛЬМ В СОПОСТАВЛЕНИИ С МУЗЫКОЙ, VOIP И ИГРОЙ
МУЗЫКА В СОПОСТАВЛЕНИИ С КИНОФИЛЬМОМ, VOIP И ИГРОЙ
VOIP В СОПОСТАВЛЕНИИ С КИНОФИЛЬМОМ, МУЗЫКОЙ И ИГРОЙ
ИГРА В СОПОСТАВЛЕНИИ С КИНОФИЛЬМОМ, МУЗЫКОЙ И VOIP
H1
H2
H3
H4
Наибольшее?
КИНОФИЛЬМ/ МУЗЫКА/VOIP/ ИГРА
$C_{КИНОФИЛЬМ}/C_{МУЗЫКА}/C_{VOIP}/C_{ИГРА}$

Фиг. 6

6

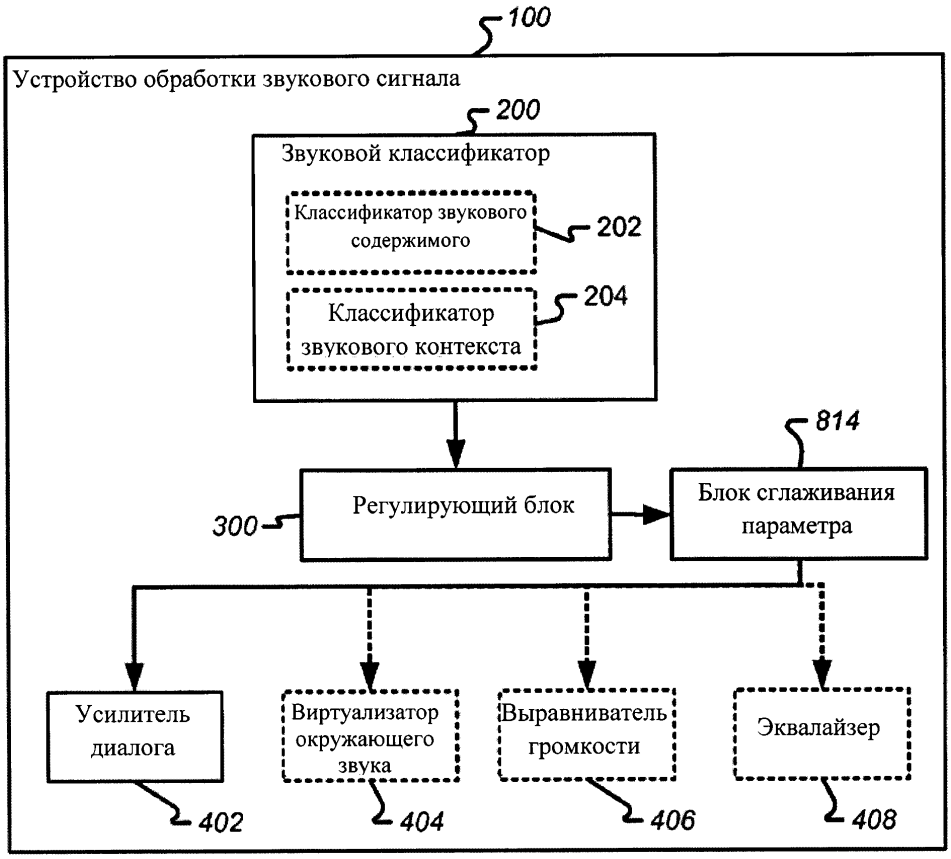
6/28


100
Устройство обработки звукового сигнала
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
712
Блок сглаживания типа
300
Регулирующий блок
Усилитель диалога
402
Виртуализатор окружающего звука
404
Выравниватель громкости
406
Эквалайзер
408

Фиг. 7

7

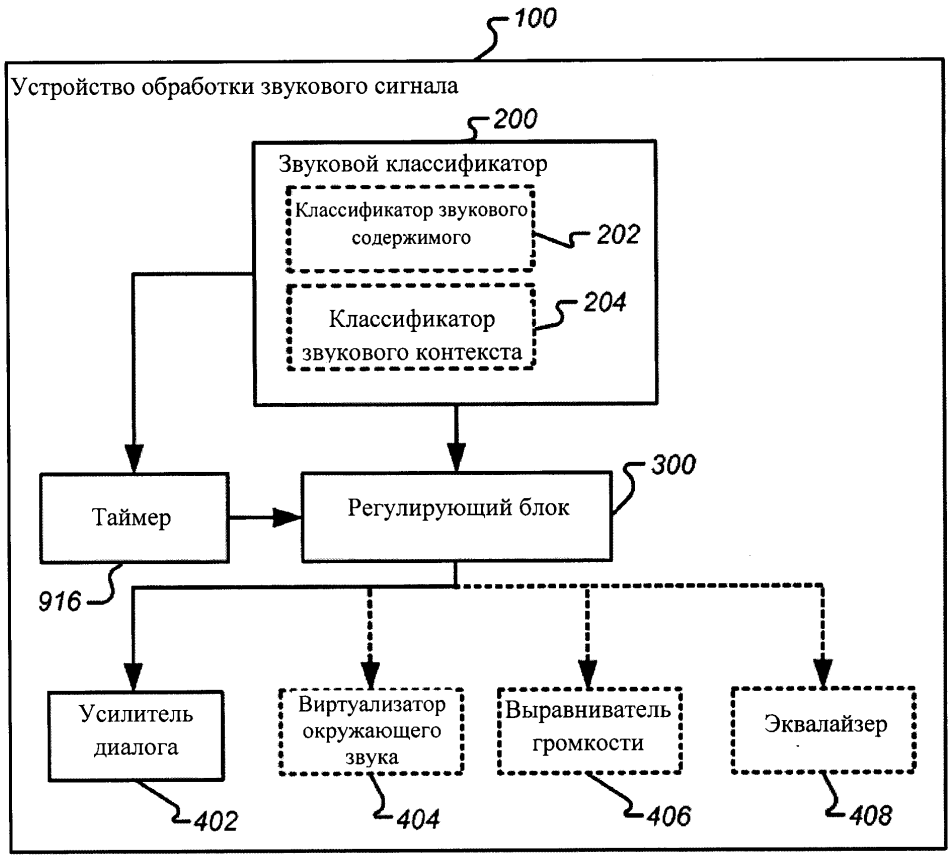
7/28


100
Устройство обработки звукового сигнала
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
814
300
Регулирующий блок
Блок сглаживания параметра
Усилитель диалога
Виртуализатор окружающего звука
Выравниватель громкости
Эквалайзер
402
404
406
408

Фиг. 8

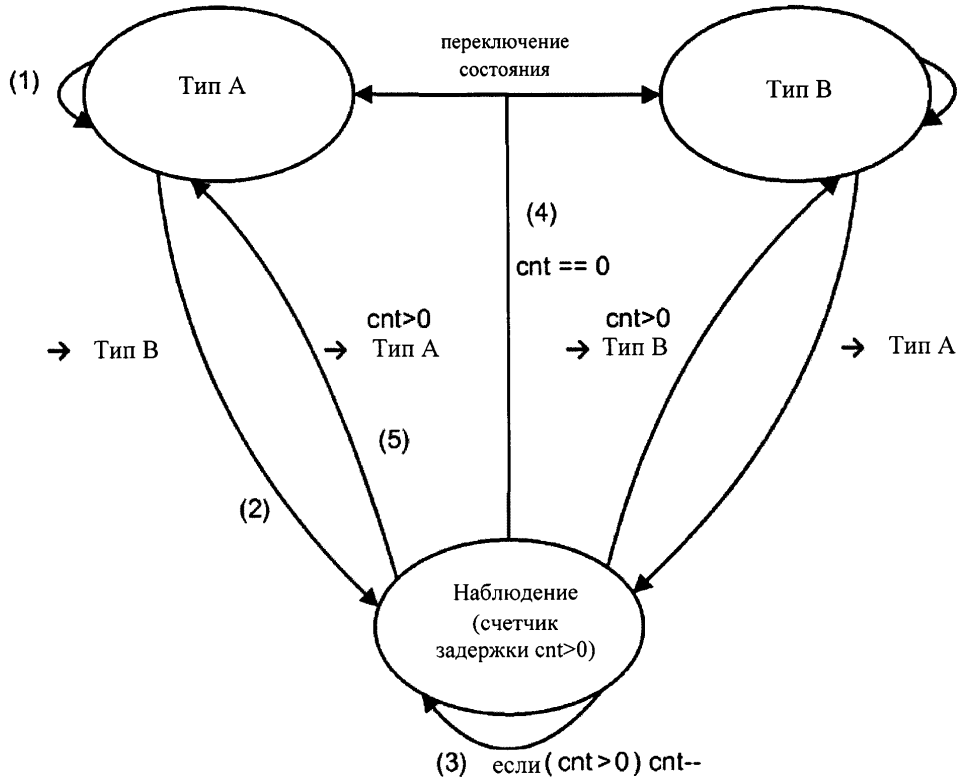
8

8/28


100
Устройство обработки звукового сигнала
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
Таймер
Регулирующий блок
300
916
Усилитель диалога
Виртуализатор окружающего звука
Выравниватель громкости
Эквалайзер
402
404
406
408

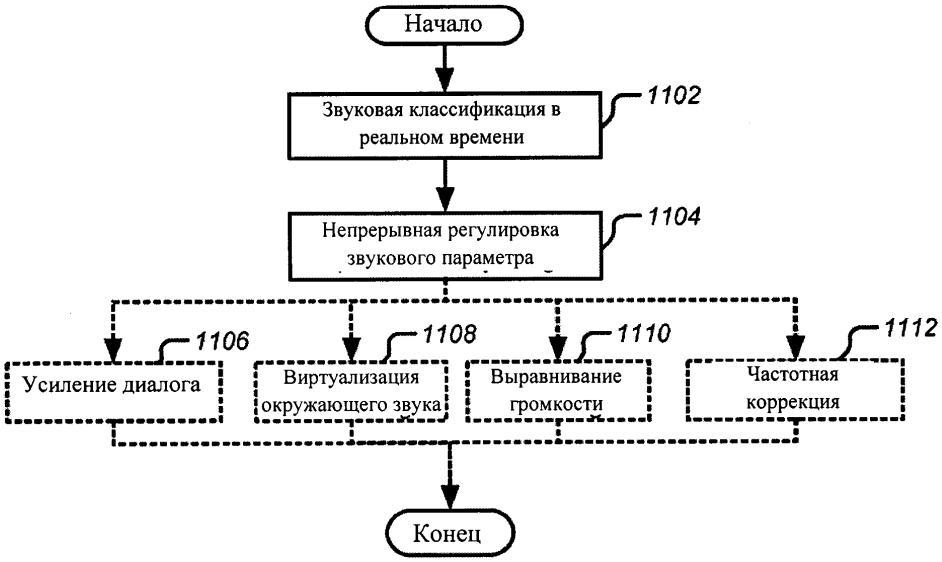
Фиг. 9

9

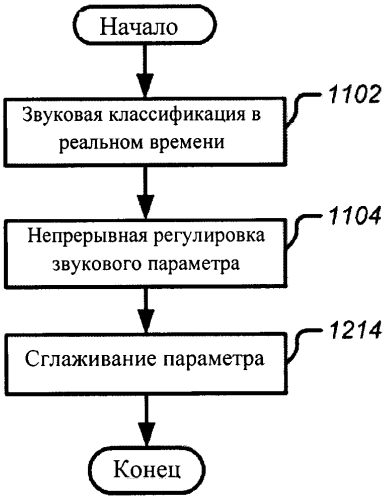

(1)
Тип A
переключение
состояния
Тип B
(4)
cnt == 0
cnt>0
Тип B
Тип A
cnt>0
Тип B
Тип A
(5)
(2)
Наблюдение
(счетчик
задержки cnt>0)
(3) если (cnt > 0) cnt--

Фиг. 10

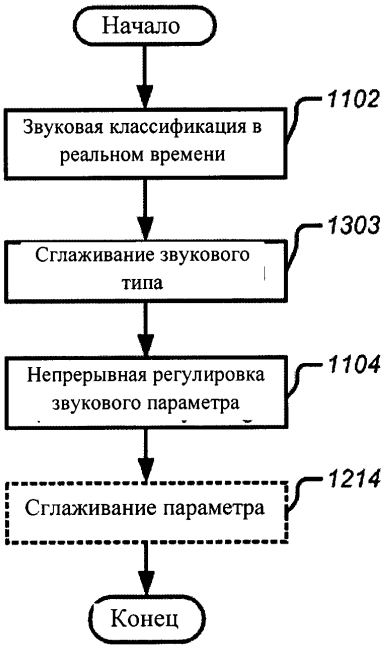
10



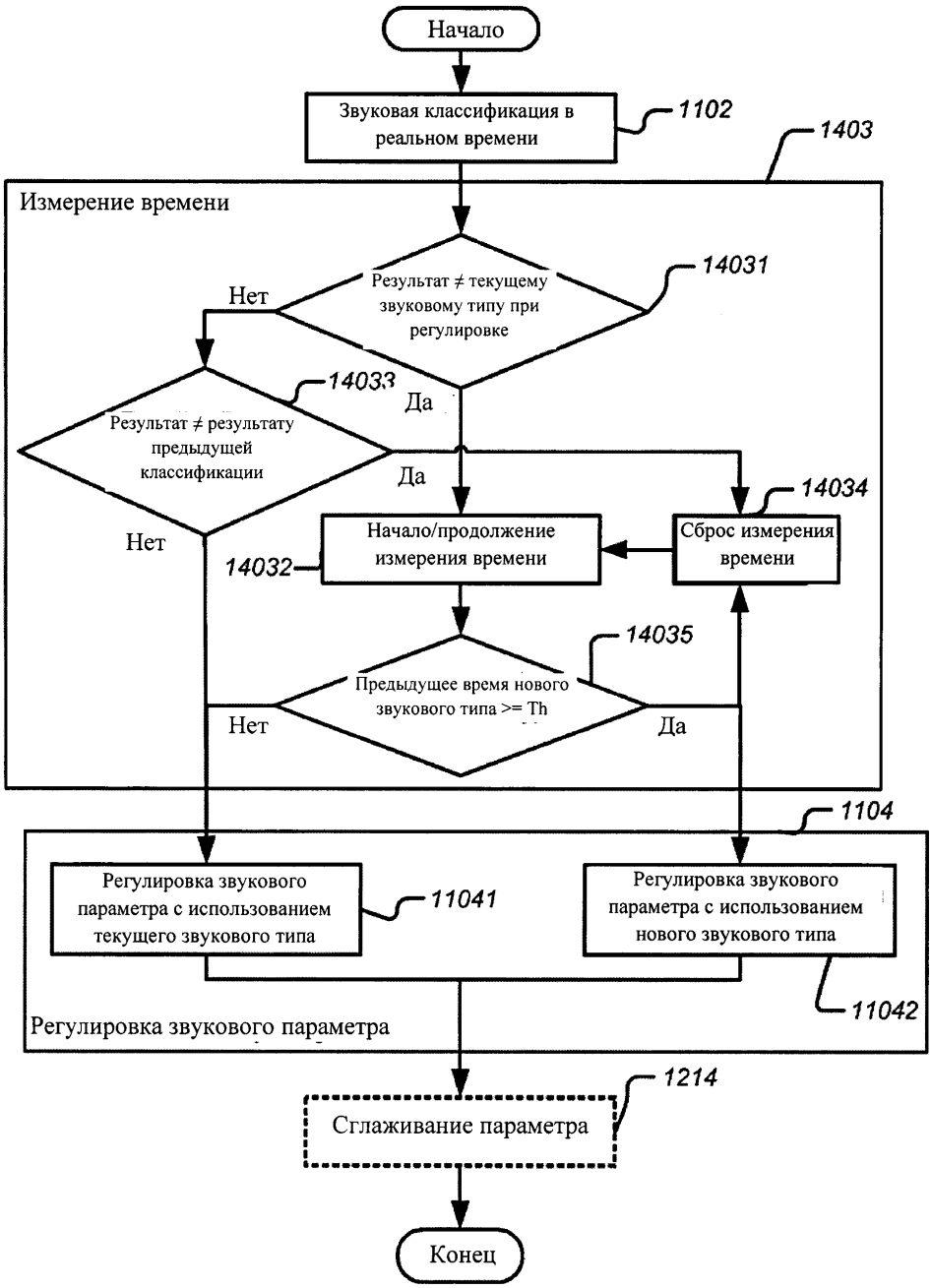
Фиг. 11



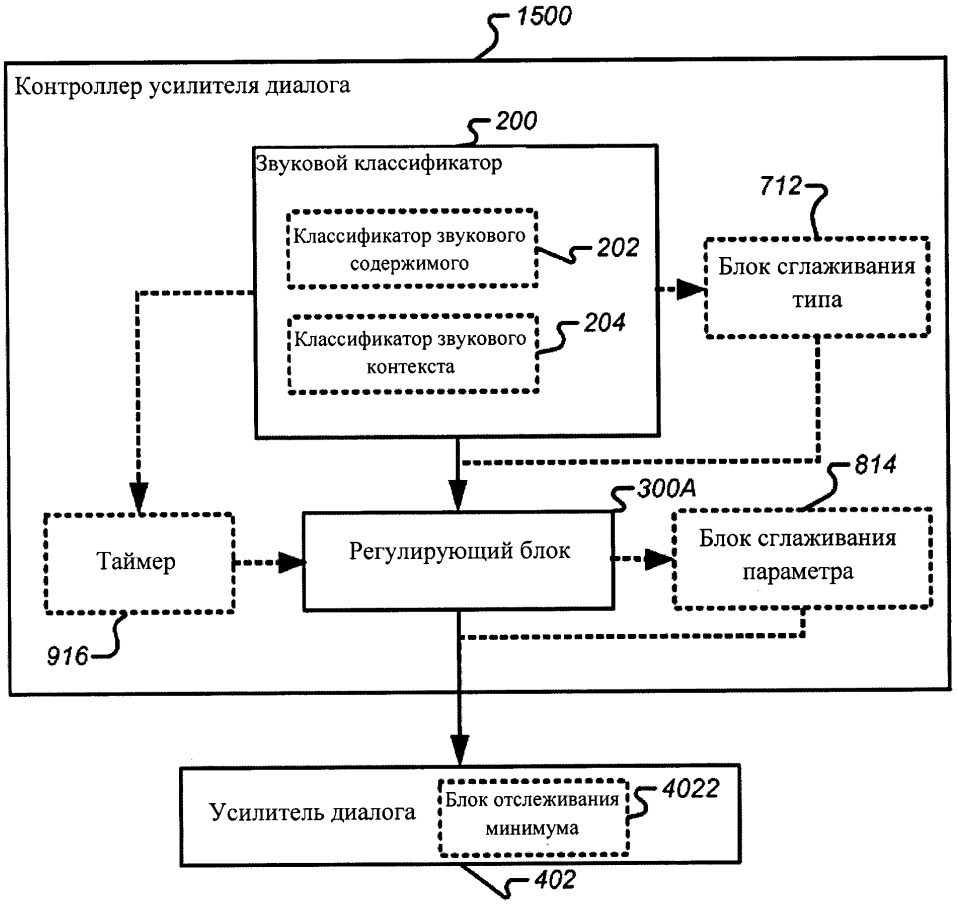
Фиг. 12



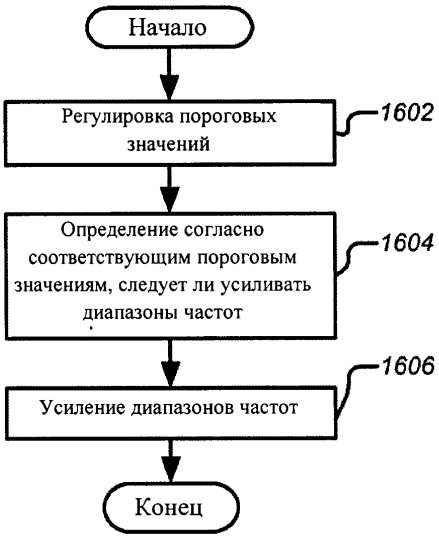
Фиг. 13


Начало
Звуковая классификация в реальном времени
1102
1403
Измерение времени
Результат ≠ текущему звуковому типу при регулировке
14031
Нет
Да
14033
Результат ≠ результату предыдущей классификации
Да
Нет
14034
Начало/продолжение измерения времени
Сброс измерения времени
14032
14035
Предыдущее время нового звукового типа >= Th
Нет
Да
1104
Регулировка звукового параметра с использованием текущего звукового типа
11041
Регулировка звукового параметра с использованием нового звукового типа
11042
Регулировка звукового параметра
1214
Сглаживание параметра
Конец

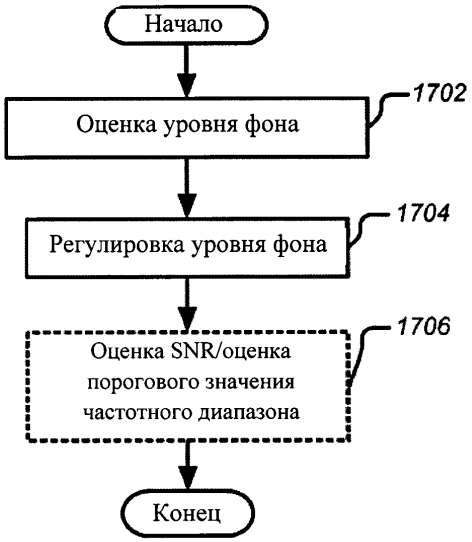
Фиг. 14


1500
Контроллер усилителя диалога
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
712
Блок сглаживания типа
814
300A
Таймер
Регулирующий блок
Блок сглаживания параметра
916
Усилитель диалога
Блок отслеживания минимума
4022
402

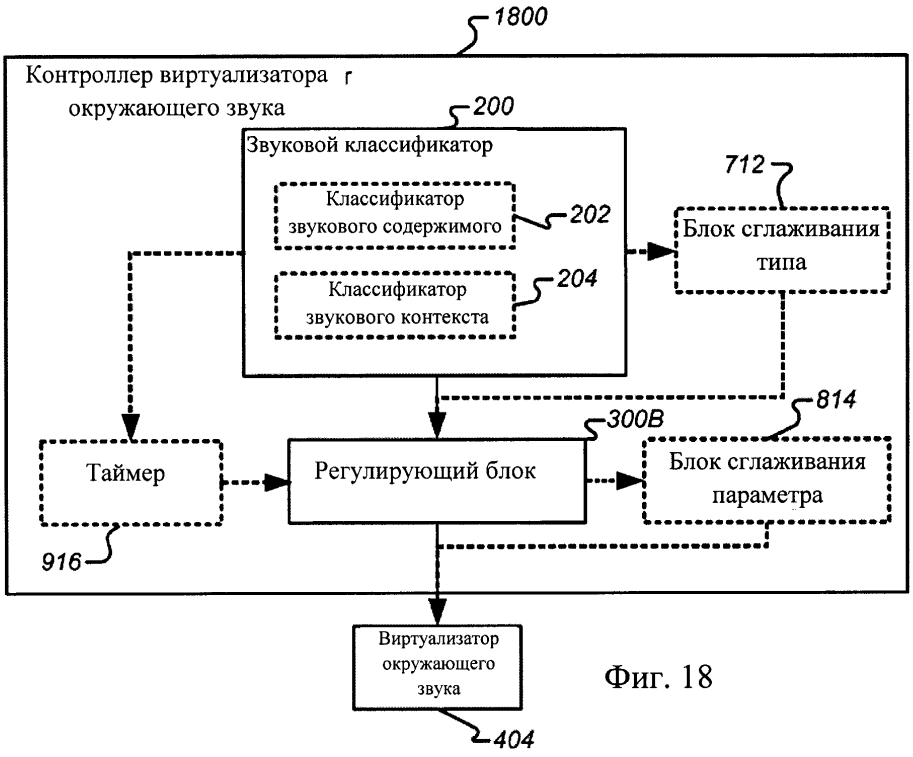
Фиг. 15



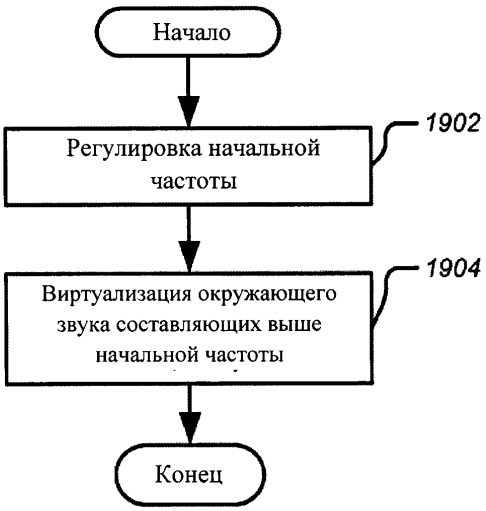
Фиг. 16



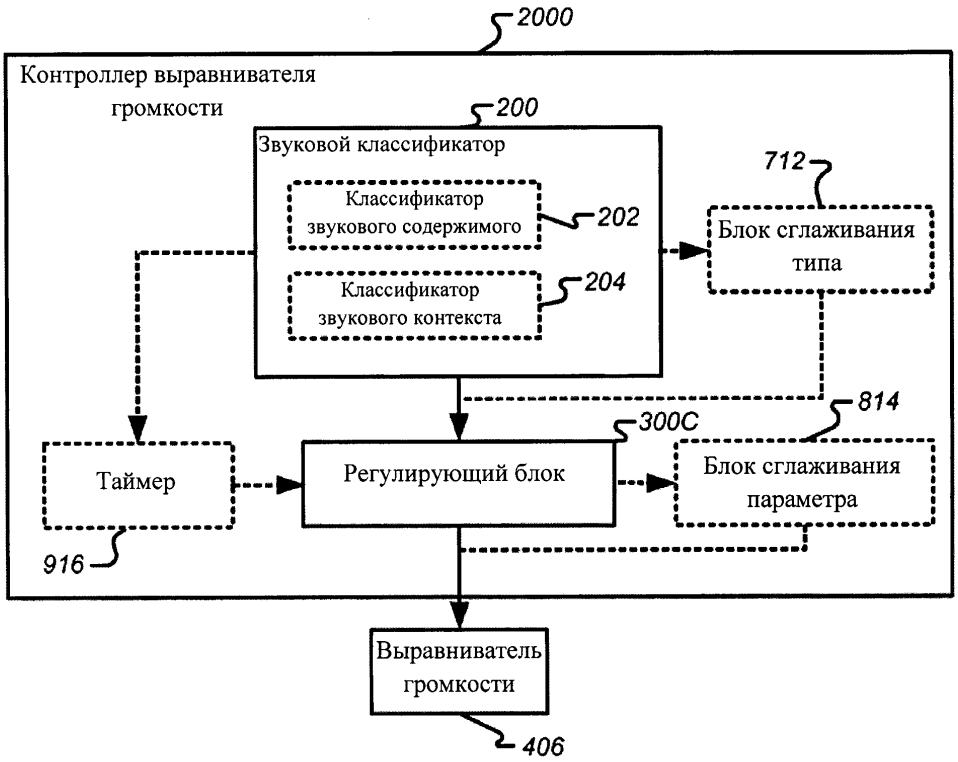
Фиг. 17



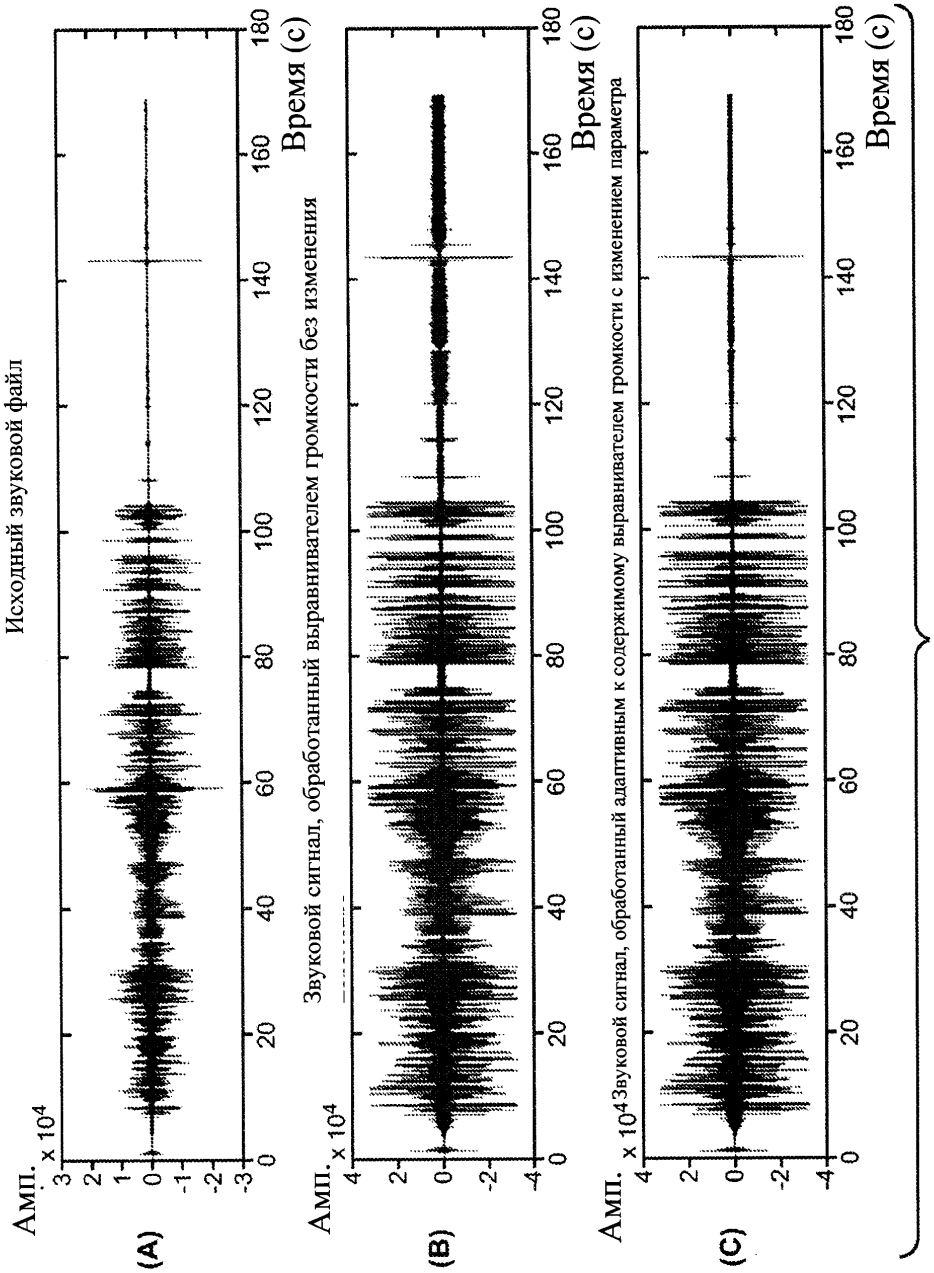
Фиг. 18


Начало
Регулировка начальной частоты
1902
Виртуализация окружающего звука составляющих выше начальной частоты
1904
Конец

Фиг. 19


2000
Контроллер выравнивателя громкости
200
Звуковой классификатор
Классификатор звукового содержимого
202
Классификатор звукового контекста
204
712
Блок сглаживания типа
814
300C
Таймер
Регулирующий блок
Блок сглаживания параметра
916
Выравниватель громкости
406

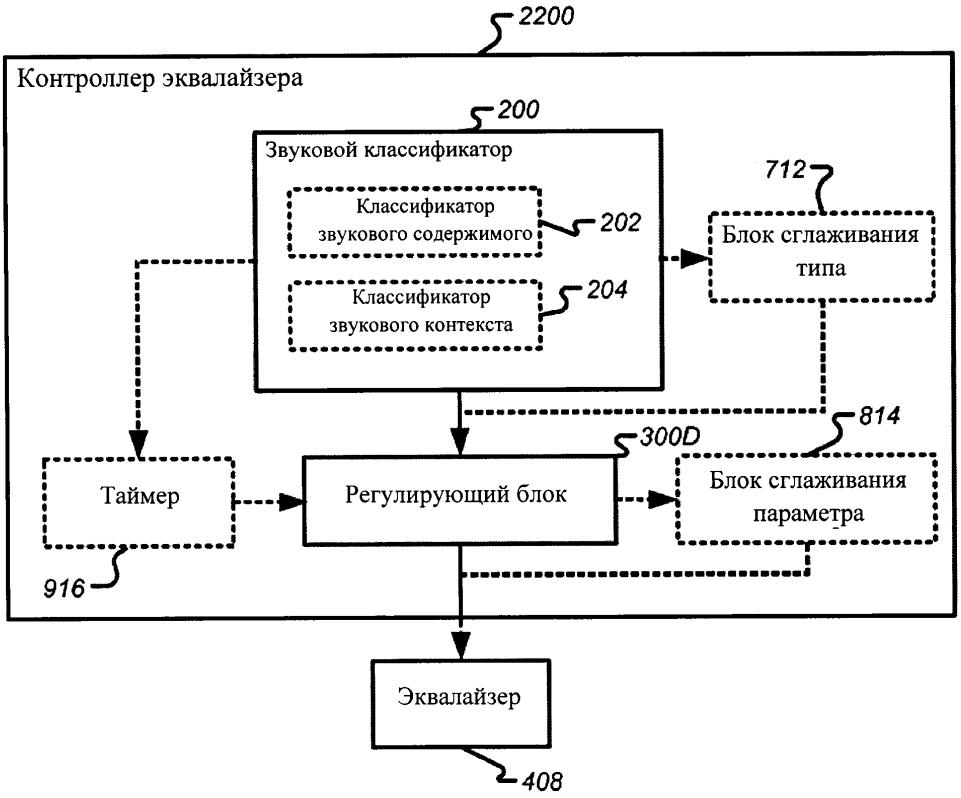
Фиг. 20


Исходный звуковой файл
Амп.
x 10^4
(A)
Время (с)
Звуковой сигнал, обработанный выравнивателем громкости без изменения
Амп.
x 10^4
(B)
Время (с)
Звуковой сигнал, обработанный адаптивным к содержимому выравнивателем громкости с изменением параметра
Амп.
x 10^4
(C)
Время (с)

Фиг. 21

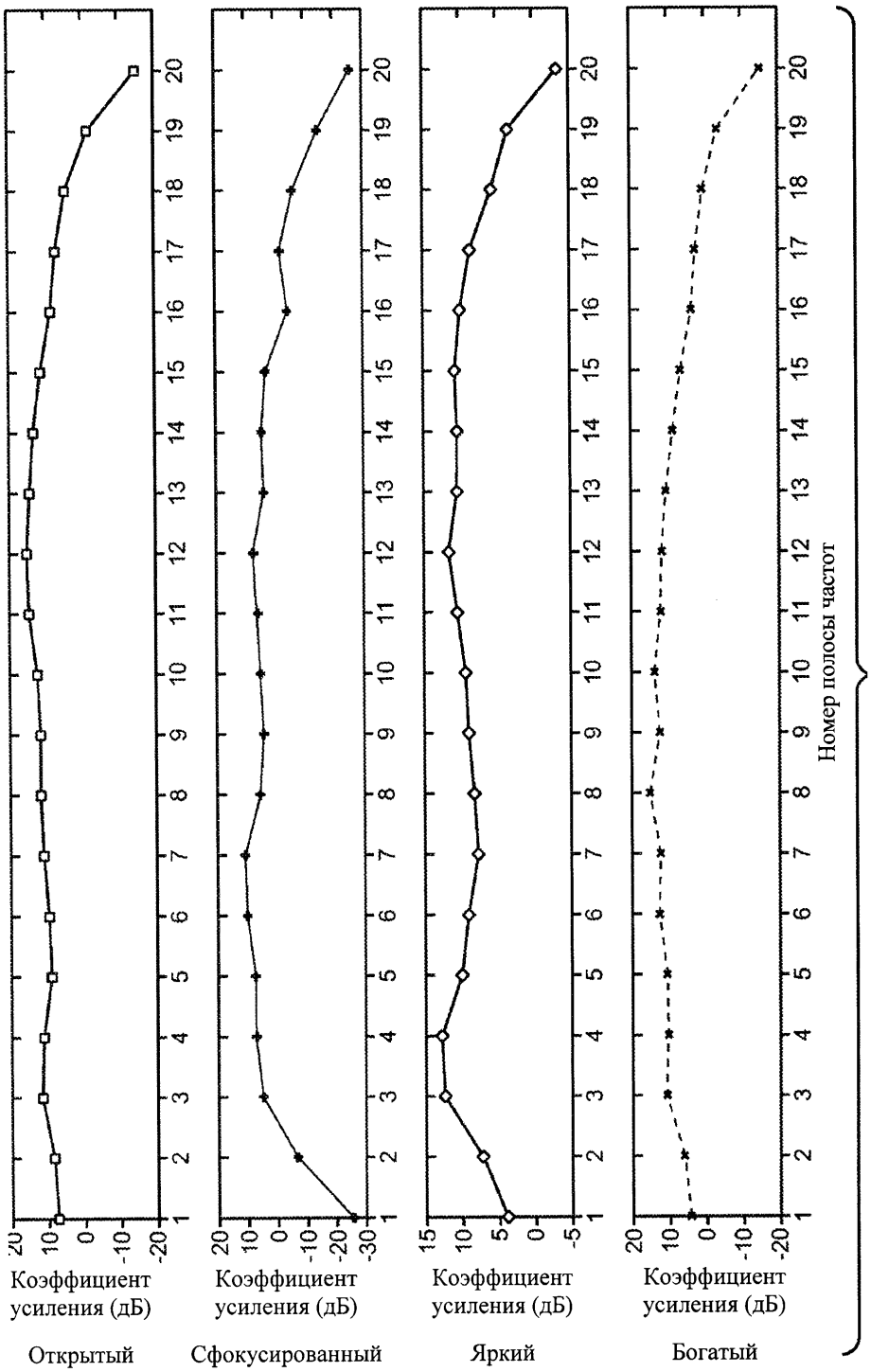
16

16/28

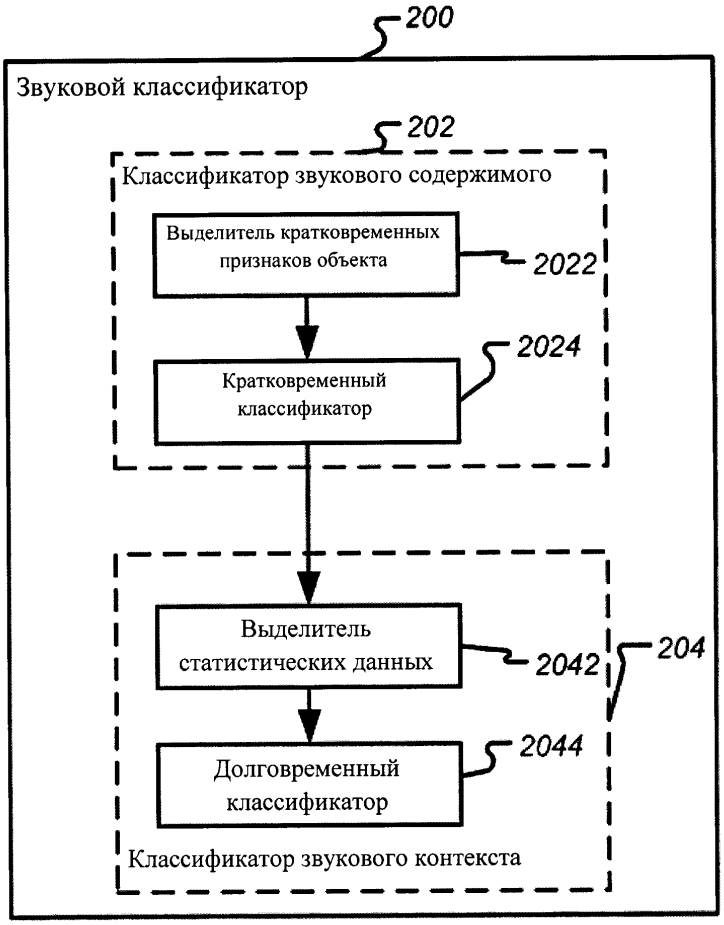


Фиг. 22

17

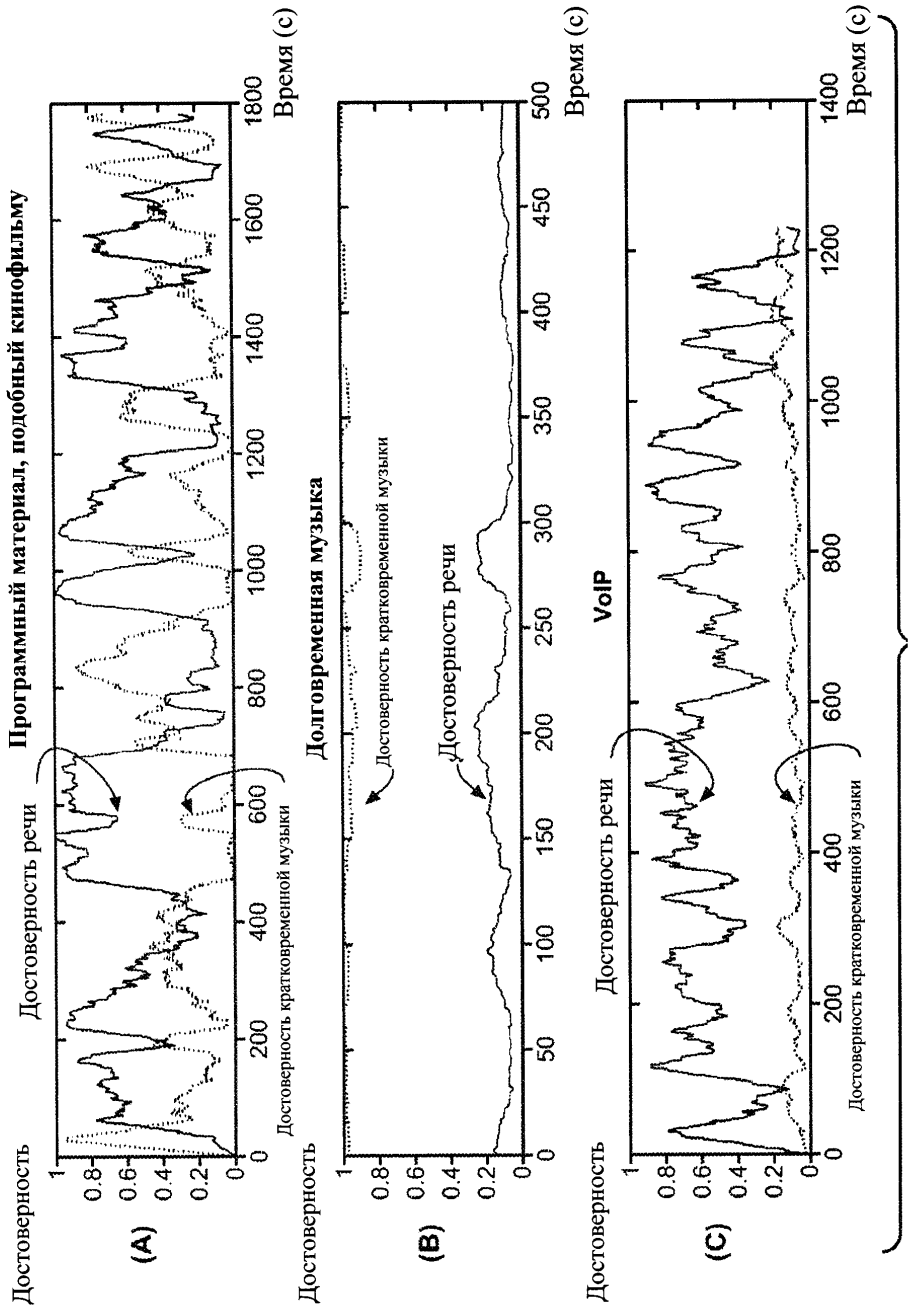

Коэффициент усиления (дБ)
Открытый
Коэффициент усиления (дБ)
Сфокусированный
Коэффициент усиления (дБ)
Яркий
Коэффициент усиления (дБ)
Богатый
Номер полосы частот

Фиг. 23

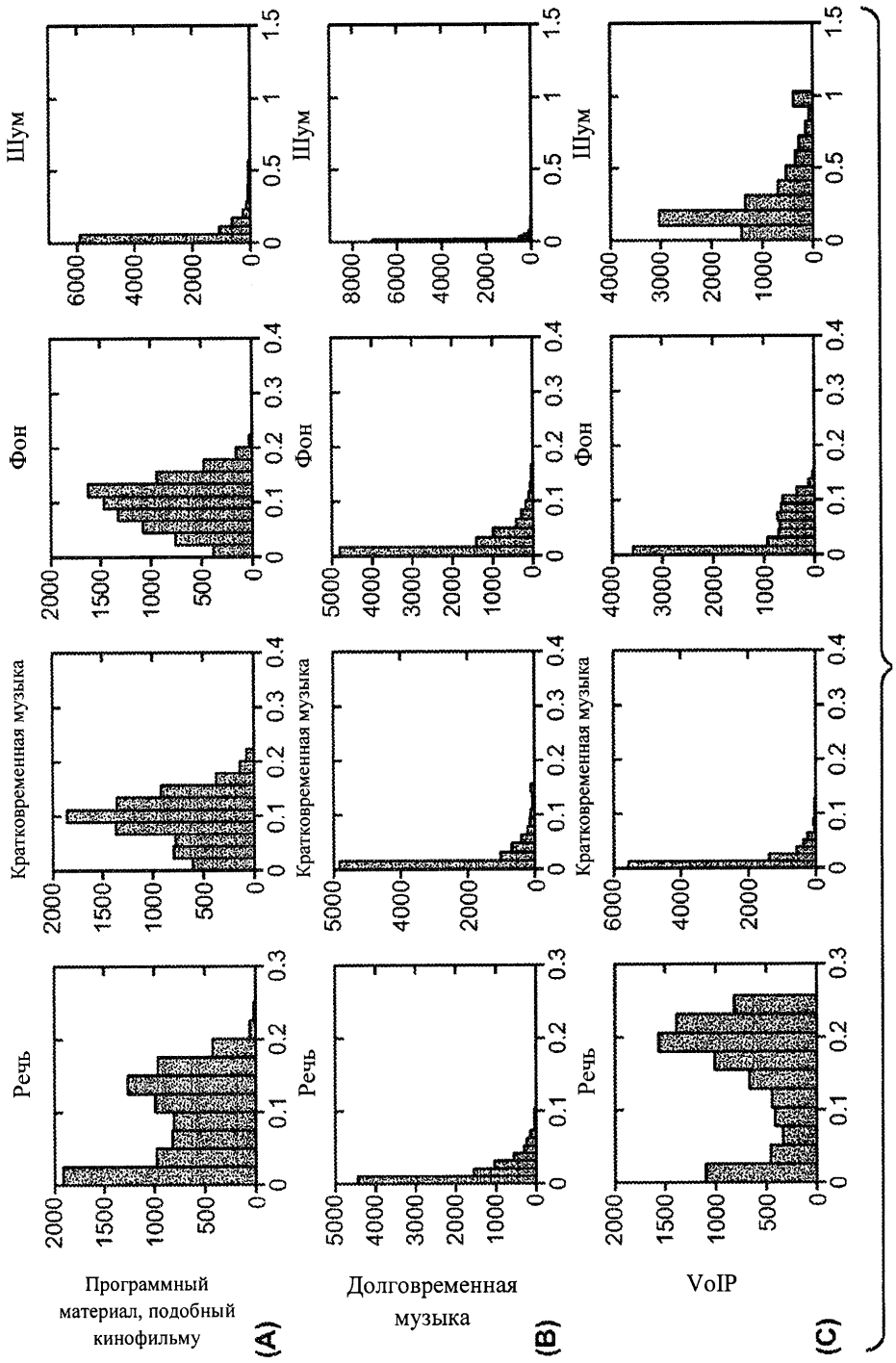

200
Звуковой классификатор
202
Классификатор звукового содержимого
Выделитель кратковременных признаков объекта
2022
2024
Кратковременный классификатор
Выделитель статистических данных
2042
204
2044
Долговременный классификатор
Классификатор звукового контекста

Фиг. 24

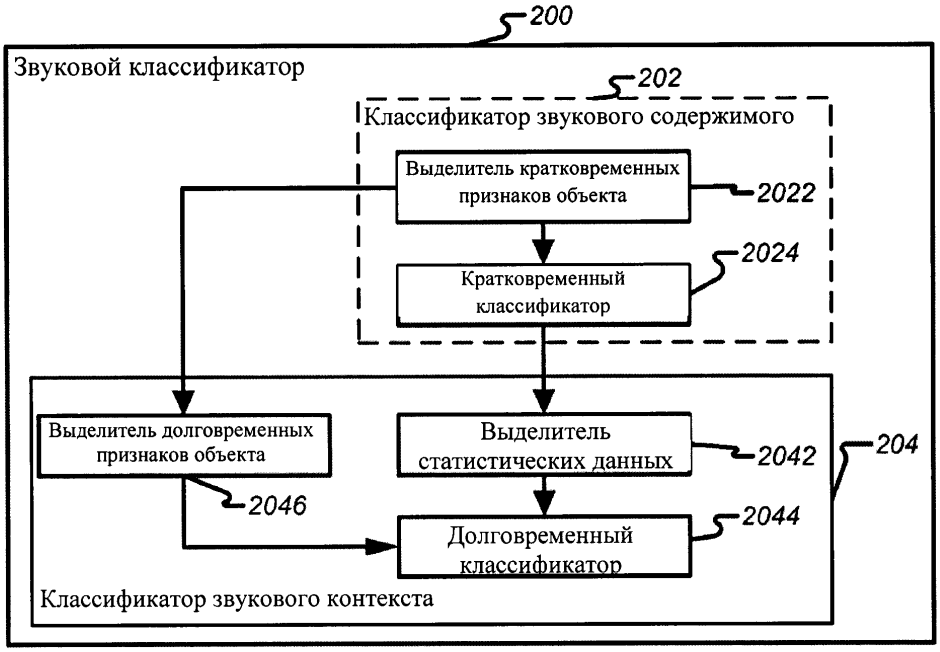
19


Программный материал, подобный кинофильму
Достоверность
Достоверность речи
Достоверность кратковременной музыки
Время (с)
(A)
Долговременная музыка
Достоверность
Достоверность кратковременной музыки
Достоверность речи
Время (с)
(B)
VoIP
Достоверность
Достоверность речи
Достоверность кратковременной музыки
Время (с)
(C)

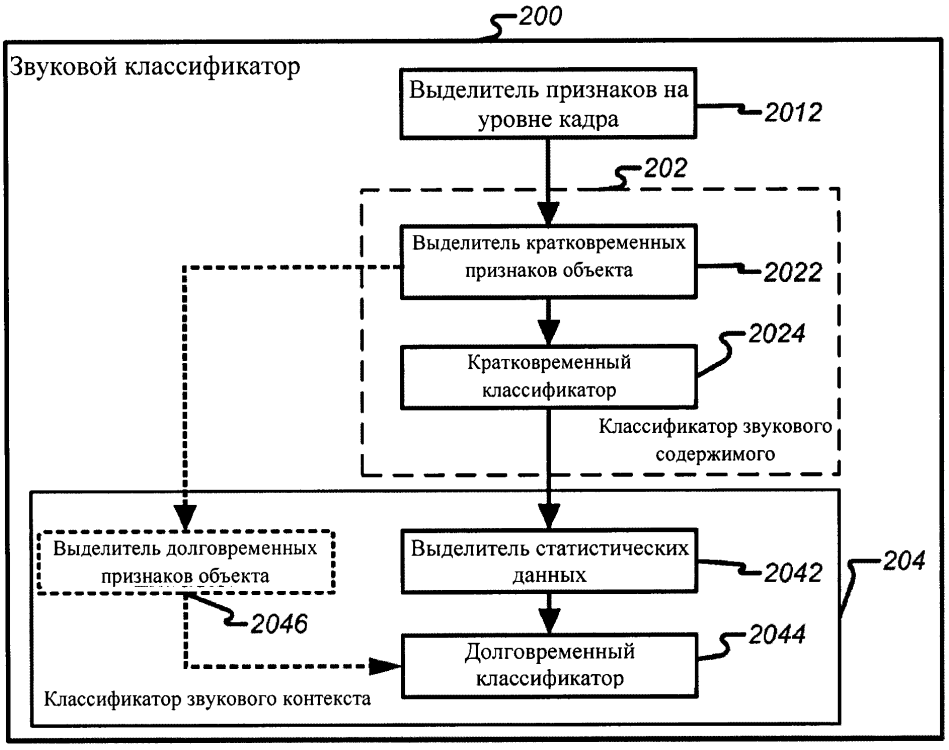
Фиг. 25


Речь
Кратковременная музыка
Фон
Шум
Программный материал, подобный кинофильму
(A)
Долговременная музыка
(B)
VoIP
(C)

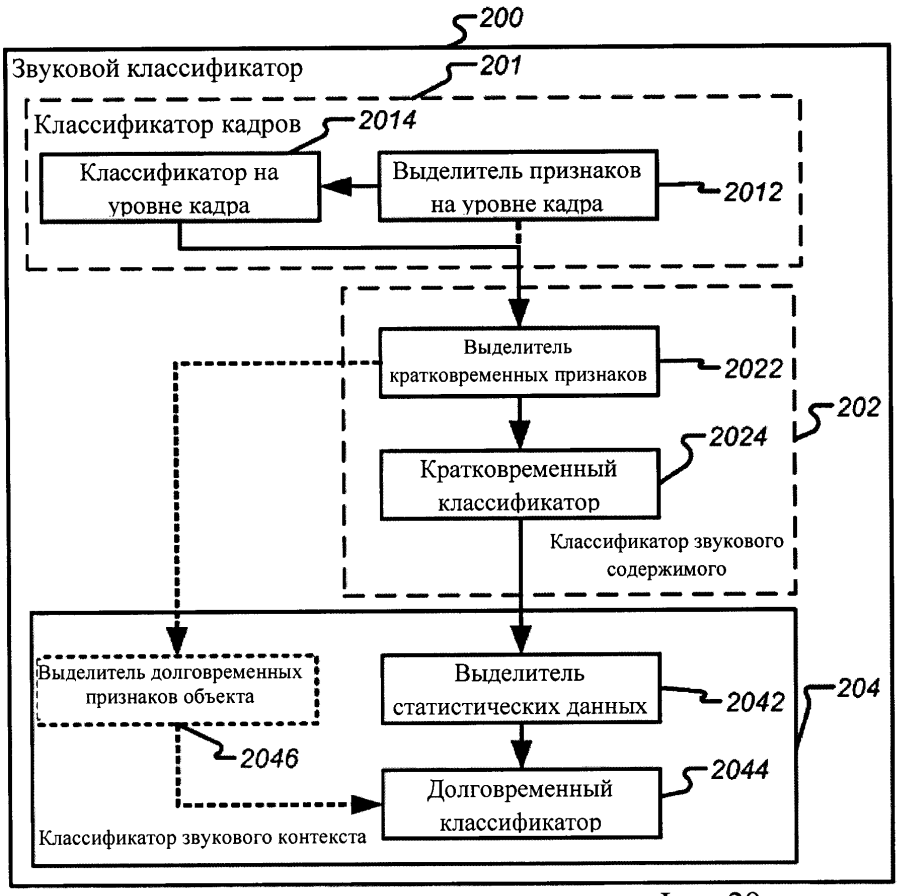
Фиг. 26


200
Звуковой классификатор
202
Классификатор звукового содержимого
Выделитель кратковременных признаков объекта
2022
2024
Кратковременный классификатор
Выделитель долговременных признаков объекта
Выделитель статистических данных
2042
204
2046
2044
Долговременный классификатор
Классификатор звукового контекста

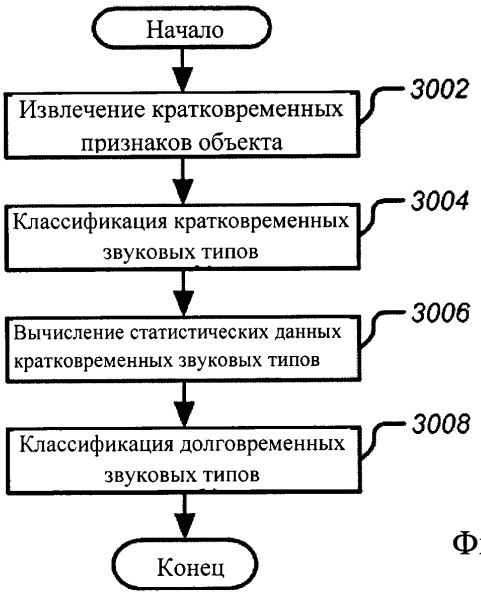
Фиг. 27


200
Звуковой классификатор
Выделитель признаков на уровне кадра
2012
202
Выделитель кратковременных признаков объекта
2022
2024
Кратковременный классификатор
Классификатор звукового содержимого
Выделитель долговременных признаков объекта
Выделитель статистических данных
2042
204
2046
2044
Долговременный классификатор
Классификатор звукового контекста

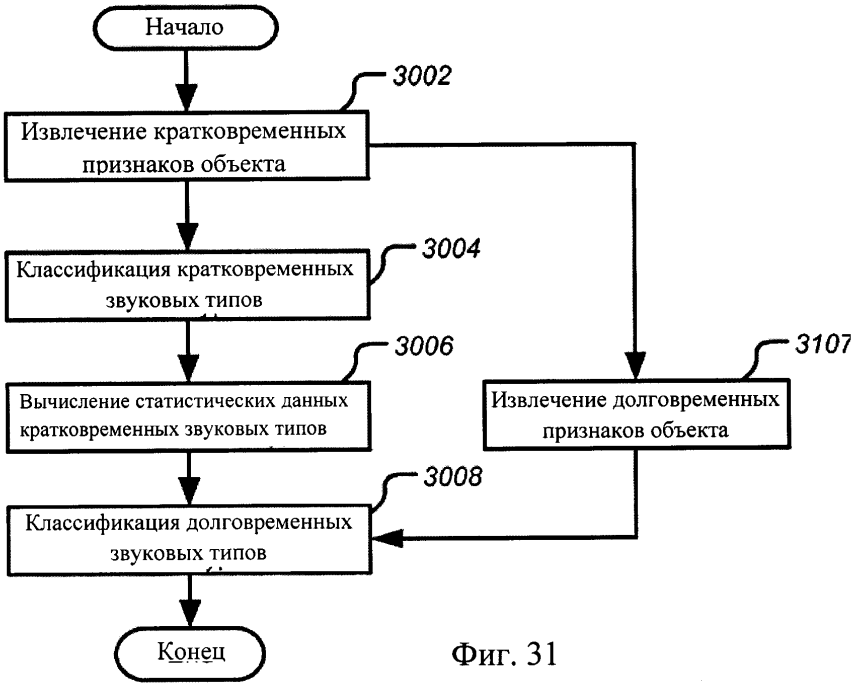
Фиг. 28


200
Звуковой классификатор
201
Классификатор кадров
2014
Классификатор на уровне кадра
Выделитель признаков на уровне кадра
2012
Выделитель кратковременных признаков
2022
202
2024
Кратковременный классификатор
Классификатор звукового содержимого
Выделитель долговременных признаков объекта
Выделитель статистических данных
2042
204
2046
2044
Долговременный классификатор
Классификатор звукового контекста

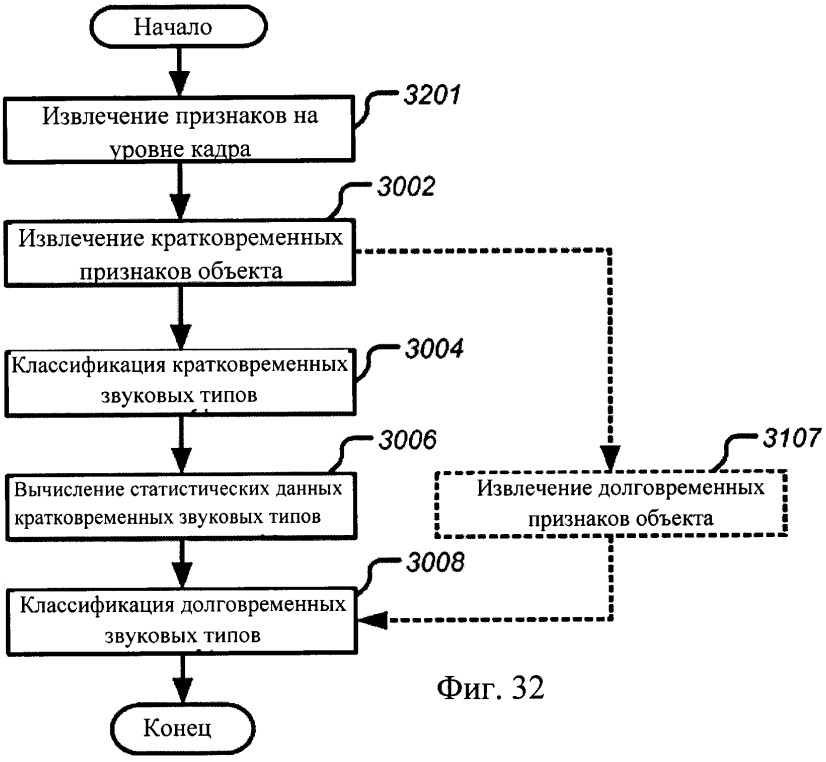
Фиг. 29


Начало
Извлечение кратковременных признаков объекта
3002
Классификация кратковременных звуковых типов
3004
Вычисление статистических данных кратковременных звуковых типов
3006
Классификация долговременных звуковых типов
3008
Конец

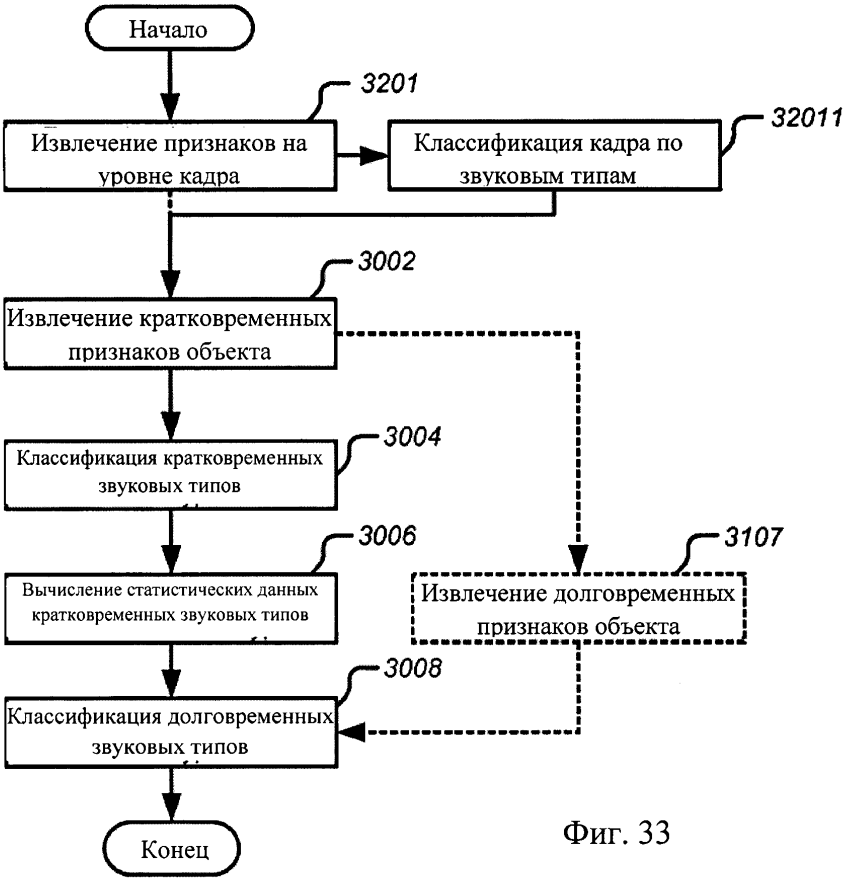
Фиг. 30


Начало
3002
Извлечение кратковременных признаков объекта
3004
Классификация кратковременных звуковых типов
3006
Вычисление статистических данных кратковременных звуковых типов
3107
Извлечение долговременных признаков объекта
3008
Классификация долговременных звуковых типов
Конец

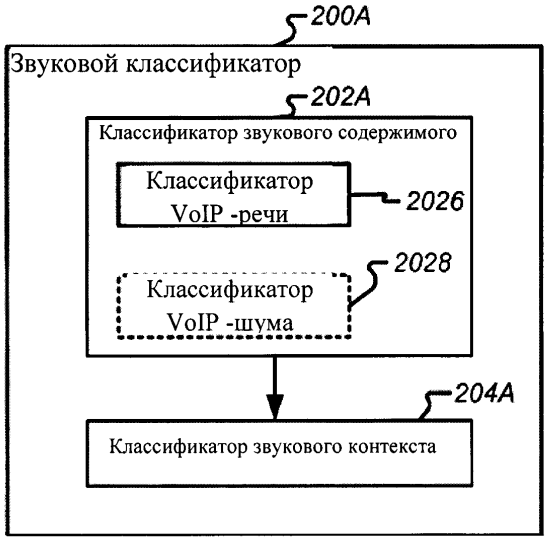
Фиг. 31


Начало
3201
Извлечение признаков на уровне кадра
3002
Извлечение кратковременных признаков объекта
3004
Классификация кратковременных звуковых типов
3006
Вычисление статистических данных кратковременных звуковых типов
3107
Извлечение долговременных признаков объекта
3008
Классификация долговременных звуковых типов
Конец

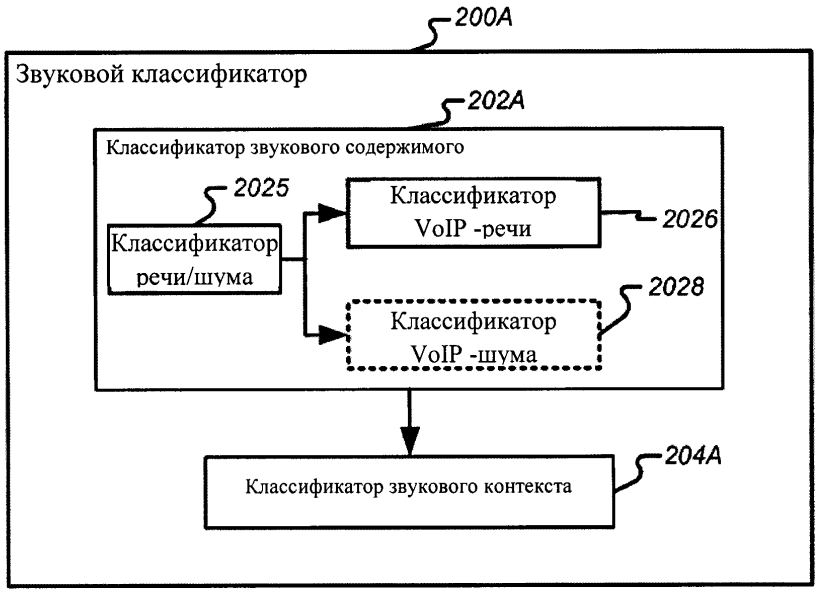
Фиг. 32


Начало
3201
Извлечение признаков на уровне кадра
32011
Классификация кадра по звуковым типам
3002
Извлечение кратковременных признаков объекта
3004
Классификация кратковременных звуковых типов
3006
Вычисление статистических данных кратковременных звуковых типов
3107
Извлечение долговременных признаков объекта
3008
Классификация долговременных звуковых типов
Конец

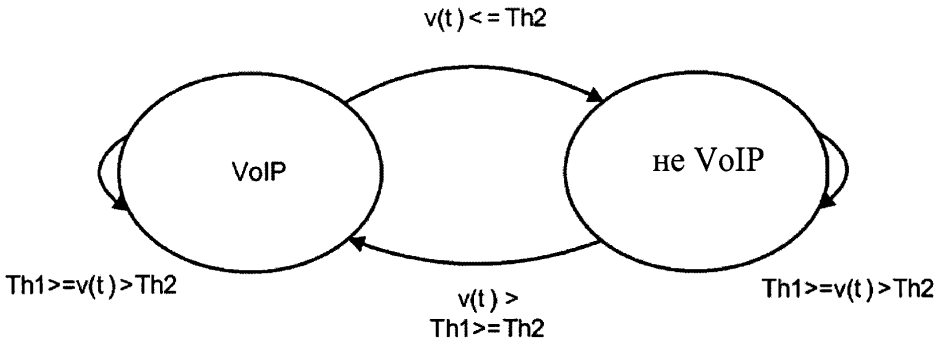
Фиг. 33


200A
Звуковой классификатор
202A
Классификатор звукового содержимого
Классификатор VoIP -речи
2026
2028
Классификатор VoIP -шума
204A
Классификатор звукового контекста

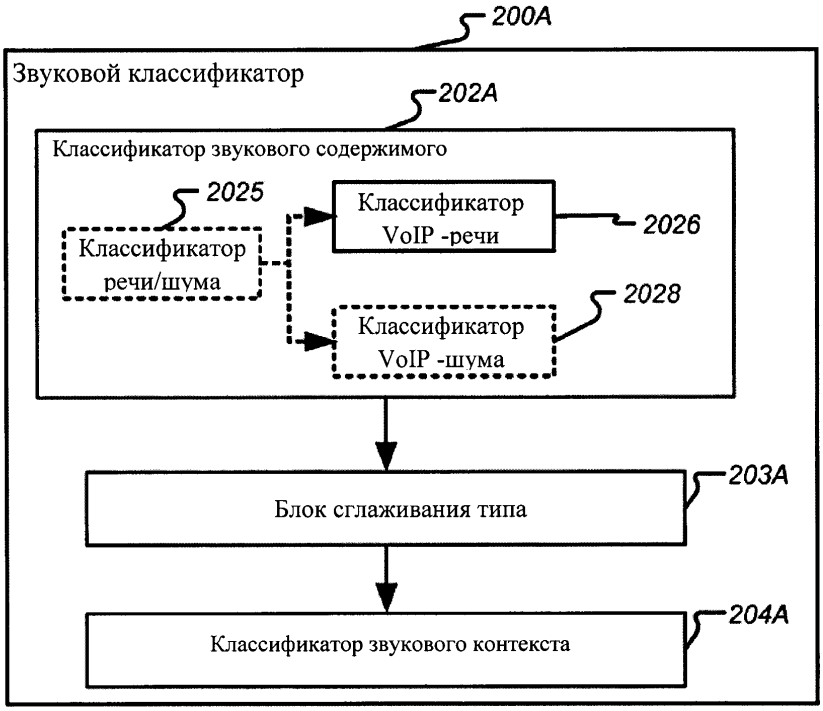
Фиг. 34


200A
Звуковой классификатор
202A
Классификатор звукового содержимого
2025
Классификатор речи/шума
Классификатор VoIP -речи
2026
2028
Классификатор VoIP -шума
204A
Классификатор звукового контекста

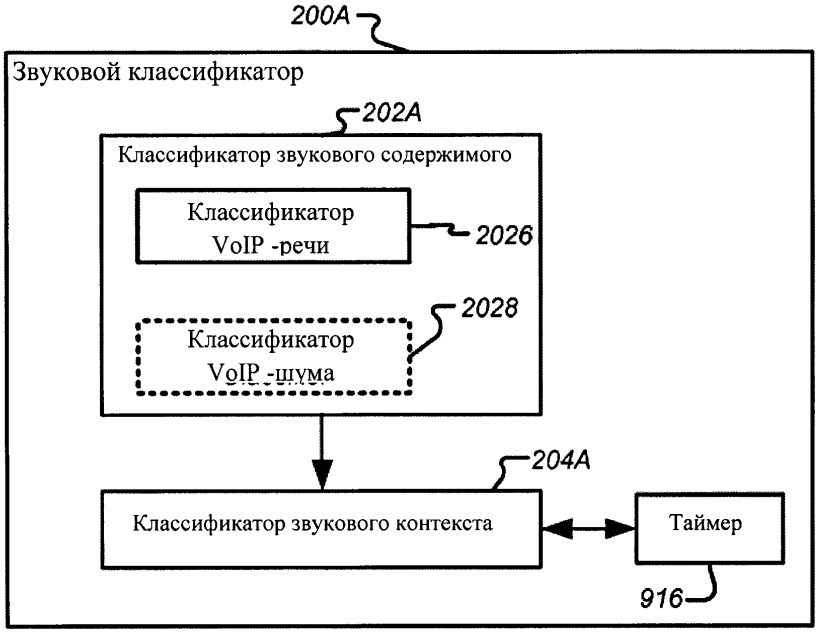
Фиг. 35


v(t) <= Th2
VoIP
не VoIP
Th1>=v(t)>Th2
v(t) > Th1>=Th2
Th1>=v(t)>Th2

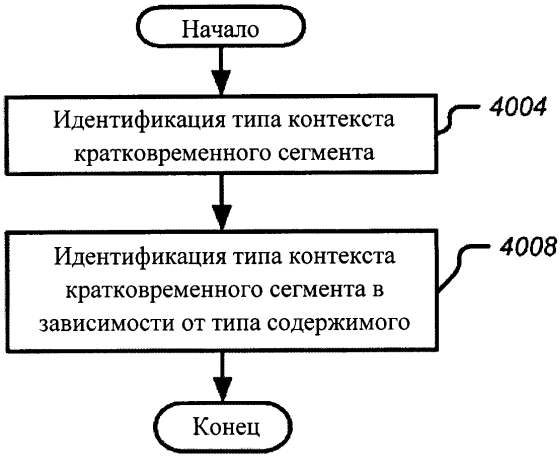
Фиг. 36


200A
Звуковой классификатор
202A
Классификатор звукового содержимого
2025
Классификатор речи/шума
Классификатор VoIP -речи
2026
2028
Классификатор VoIP -шума
203A
Блок сглаживания типа
204A
Классификатор звукового контекста

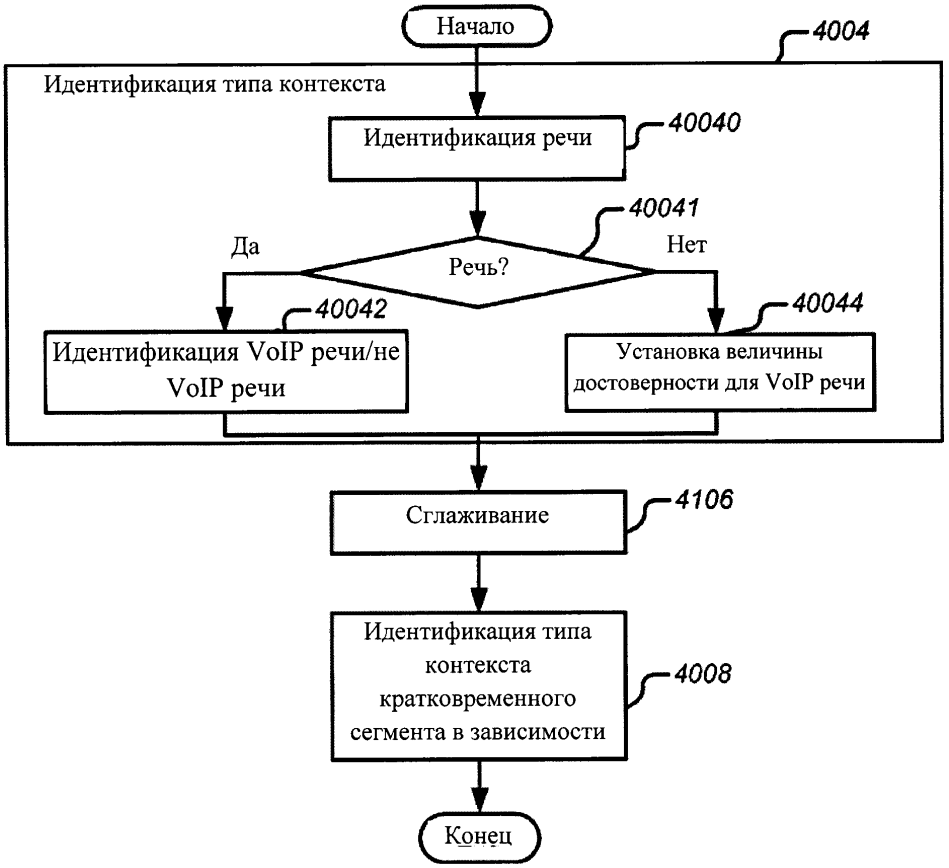
Фиг. 37


200A
Звуковой классификатор
202A
Классификатор звукового содержимого
Классификатор VoIP -речи
2026
2028
Классификатор VoIP -шума
204A
Классификатор звукового контекста
Таймер
916

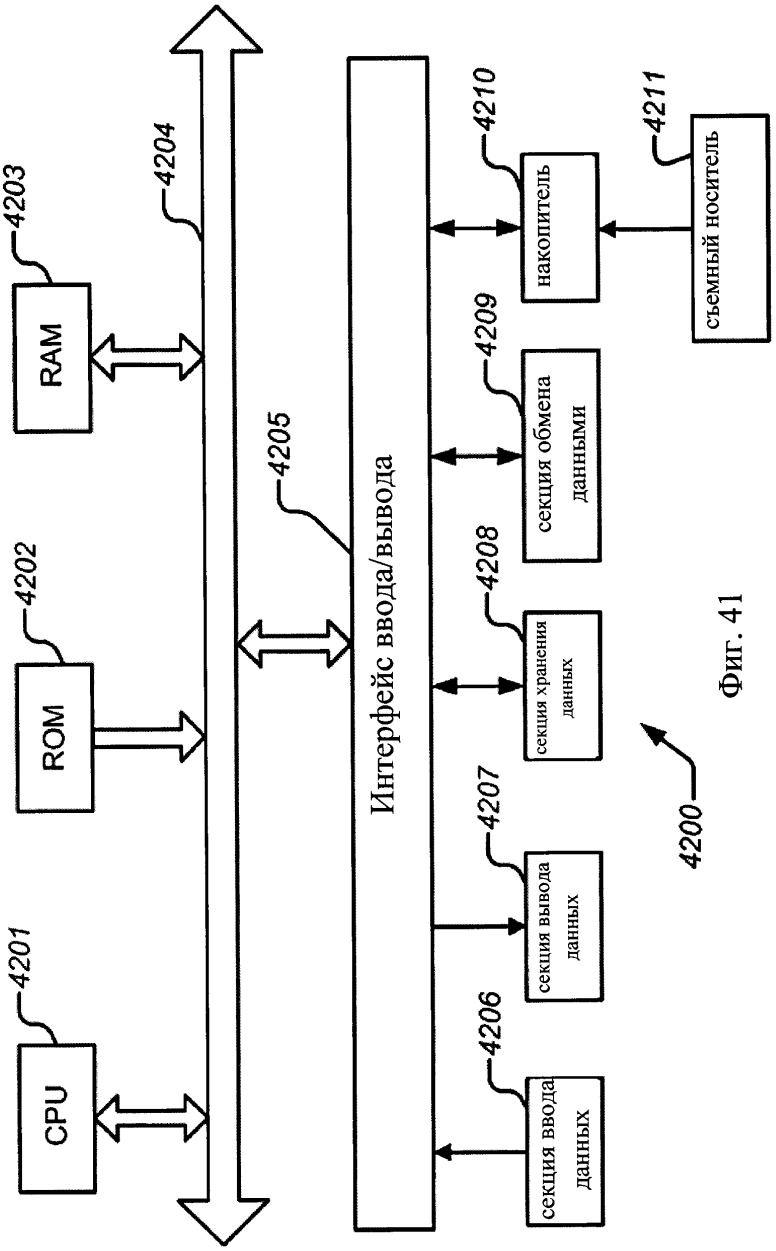
Фиг. 38


Начало
Идентификация типа контекста кратковременного сегмента
4004
Идентификация типа контекста кратковременного сегмента в зависимости от типа содержимого
4008
Конец

Фиг. 39


Начало
4004
Идентификация типа контекста
Идентификация речи
40040
40041
Да
Речь?
Нет
40042
Идентификация VoIP речи/не VoIP речи
40044
Установка величины достоверности для VoIP речи
Сглаживание
4106
Идентификация типа контекста кратковременного сегмента в зависимости
4008
Конец

Фиг. 40


4201
CPU
4202
ROM
4203
RAM
4204
4205
Интерфейс ввода/вывода
4206
секция ввода данных
4207
секция вывода данных
4208
секция хранения данных
4209
секция обмена данными
4210
накопитель
4211
съемный носитель
4200

Фиг. 41