(19) Евразийское патентное ведомство

(11) 041269 (13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2022.10.03

(21) Номер заявки
202092875

(22) Дата подачи заявки
2020.12.23

(51) Int. Cl. ***G10L 17/02*** (2006.01)
G10L 25/03 (2006.01)

(54) СПОСОБ И УСТРОЙСТВО ДИАРИЗАЦИИ АУДИОСИГНАЛА

(31) **2020134876**
(32) **2020.10.23**
(33) **RU**
(43) **2022.04.29**

(71)(73) Заявитель и патентовладелец:
ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО "СБЕРБАНК РОССИИ" (ПАО СБЕРБАНК) (RU)

(72) Изобретатель:
Литвак Юрий Николаевич,
Василенко Алексей Алексеевич,
Песня Станислав Михайлович,
Малых Сергей Владимирович (RU)

(74) Представитель:
Герасин Б.В. (RU)

(56) US-A1-20180299527
US-A1-20150310863
US-A1-20110119060
US-A1-20170359666
RU-C2-2716846

(57) Представленное техническое решение относится в общем к измерительной технике, в частности к способу и устройству диаризации аудиосигнала, и предназначено для разделения поступающего аудиопотока на однородные сегменты в соответствии с принадлежностью аудиопотока тому или иному говорящему (диктору). Техническим результатом, достигаемым при решении вышеуказанной технической проблемы или технической задачи, является обеспечение возможности разметки (сегментации) аудиосигнала с малой погрешностью и с малым энергопотреблением на основе данных, полученных с 2 микрофонов, в том числе, в режиме реального времени. Указанный технический результат достигается благодаря осуществлению способа диаризации речевого аудиосигнала, содержащего этапы, на которых получают цифровые аудиосигналы, содержащие данные голоса, синхронно регистрируемые по меньше мере двумя микрофонами; определяют разностный сигнал для сигналов двух микрофонов на основе данных цифровых аудиосигналов, полученных от упомянутых микрофонов; определяют значения огибающей функции разностного сигнала; определяют значения огибающей функции исходного аудиосигнала на основе данных цифрового аудиосигнала, полученного от одного из микрофонов; на основе значения огибающей функции разностного сигнала и значения огибающей функции исходного аудиосигнала определяют характеристическое значение аудиосигнала; на основе характеристического значения аудиосигнала осуществляют разметку данных цифрового аудиосигнала, указывающую на то, к какому источнику звукового сигнала относится соответствующий блок данных цифрового аудиосигнала.

041269 B1

041269 B1

Область техники

Представленное техническое решение относится в общем к измерительной технике, в частности к способу и устройству диаризации аудиосигнала, и предназначено для разделения поступающего аудиопотока на однородные сегменты в соответствии с принадлежностью аудиопотока тому или иному говорящему (диктору).

Уровень техники

С научной точки зрения представленное техническое решение относится к категории решений, использующих для диаризации пространственные признаки источника сигнала. Устройства, использующие для выделения требуемого сигнала его пространственные признаки, могут быть как стационарными (микрофонами, зафиксированными в определенных точках помещения), так и портативными (средствами акустической разведки). Общей особенностью устройств такого типа является большое количество микрофонов (до 40 и более) и, соответственно, необходимость обеспечения широкой пропускной способности канала передачи данных для передачи уловленного сигнала. С точки зрения вычислительной сложности реализация алгоритма формирования "луча" (направления, коэффициент усилия звука, исходящего из которого, будет максимальным) является сложной задачей, требующей поиск решения в режиме реального времени. Однако с ростом вычислительной мощности портативных устройств и увеличением пропускной способности каналов передачи данных реализация данной функциональности стала возможной и доступной широкому кругу лиц, в связи с чем упрощенные системы подобного типа находят применение в устройствах с голосовым управлением с заранее не предопределенным расположением оператора.

Известными примерами сенсоров такого типа для бытового использования являются ФМР (фазированная микрофонная решетка) Microsoft Kinnect, а также ФМР, используемые в составе т.н. "умных колонок", например Apple Home Pod, Amazon Echo, Яндекс Станция и др. Особенностью ФМР, используемых в изделиях такого типа, является расположение микрофонов в горизонтальной плоскости (в случае Kinnect - линейный массив микрофонов, т.к. поиск источника сигнала ведется только в одной полуплоскости), что обусловлено необходимостью решения задачи по "определению" источника звука и выделению его из общего звукового потока (усилению сигнала в заданном пространственном направлении). Как правило, в процессе голосового управления устройствами реализация алгоритма выделения слов целевого диктора из аудиопотока выполняется в 3 шага:

1) в режиме ожидания устройство постоянно анализирует поступающий аудиопоток с одного микрофона на предмет наличия слова-триггера включения, сохраняя при этом в буфер сигналы всех микрофонов решетки;

2) в момент детектирования слова-триггера система производит вычисления для формирования "луча": осуществляется поиск таких величин фазовой задержки сигналов каждого из микрофонов, при добавлении которой к функции полученного сигнала каждым из микрофонов и сложении сигналов микрофонов с определенными коэффициентами будет обеспечиваться максимальное усиление звука, исходящего от направления целевого диктора (детектированное слово-триггер), по сравнению с фоновыми звуками, предположительно поступающими с разных направлений;

3) обработка получаемого в дальнейшем аудиопотока от всех микрофонов решетки осуществляется с учетом полученных на шаге 2 величин фазовой задержки, за счет чего результирующий сигнал ("луча") обладает большим соотношением сигнал/шум по сравнению с сигналом отдельного микрофона (это позволяет осуществить последующее преобразование речевого сигнала в текст с меньшей вероятностью ошибки).

Достоинством такой реализации алгоритма формирования луча является возможность значительного усиления речи целевого диктора (по сравнению с акустическим фоном), однако отсутствие предопределенности делает необходимым расположение массива микрофонов в плоскости сканирования. Сегментация речи по критерию принадлежности ее разным дикторам посредством данной технологии предполагает решение только 2 части этой задачи, т.е. поиска такого набора фазовых (временных) задержек для сигналов отдельных микрофонов, когда будет достигнуто максимальное усиление звуков речи диктора. Данный набор фазовых задержек сигналов отдельных микрофонов будет определять направление источника речи относительно приемника, которое, в свою очередь, будет являться характерным признаком для сегментации речевого аудиосигнала. Недостатком данного способа является низкая разрешающая способность при условии сочетания малых габаритов приемного устройства (и, соответственно, расстоянии между микрофонами, например, порядка единиц см) и низкой частоты дискретизации (например, 16 кГц, которая является стандартом для систем речевой аналитики). Это обусловлено тем, что при работе с цифровыми сигналами фазовые (временные) задержки сигнала от источника, принимаемого разными микрофонами, могут составлять только целое число дискрет. Решение задачи диаризации является актуальным и для другого класса устройств предполагающих персональное использование: телефоны, гарнитуры и т.п. На сегодняшний день для обеспечения малых габаритов и веса устройств типичным решением является применение в таких устройствах микрофонов, выполненных по технологии МЭМС. Достоинством микрофонов данного типа является широкий динамический диапазон, высокая чувствительность и широкая диаграмма направленности (ДН) (вплоть до сферической), что важно для обеспечения

возможности работы устройства, как в режиме "трубки", так и в режиме "громкой связи". По причине широкой ДН съем звука таким устройством будет ненаправленным. Для того чтобы обеспечить комфорт собеседника (исключить из передаваемого сигнала посторонние шумы в период пауз), а также снизить объем передаваемых данных, запись и передачу звука гарнитурой или телефоном целесообразно осуществлять только в момент произнесения пользователем реплики, что также требует решения задачи по определению принадлежности речевого сигнала пользователю. Выделение речи пользователя из поступающих звуковых сигналов может осуществляться либо по уровню звукового давления, либо посредством фиксации колебаний поверхности тела пользователя в момент произнесения речи. Последнее может достигаться, например, за счет встраивания в устройство МЭМС-акселерометра, который формирует физический триггер-сигнал о том, что в заданный момент времени говорит именно пользователь (см., например, решение, раскрытое в заявке US 2014093093A1, опубл. 03.04.2014). Достоинством такого принципа реализации функции выделения речи пользователя из общего звукового потока, поступающего к микрофону, является органически присущая данному способу высокая точность определения принадлежности речевого сигнала пользователю, а также отсутствие необходимости выполнения сложных вычислений. Это, в свою очередь, обеспечивает высокое быстродействие алгоритмов сегментации и потенциально малое энергопотребление устройствами, их реализующими. Существенным ограничением данного способа является необходимость обеспечения достаточного для передачи колебаний механического контакта устройства с поверхностями тела оператора, которые испытывают вибрацию в процессе произнесения речи (боковая область черепа, слуховой проход, гортань, грудь).

Сущность технического решения

Технической проблемой или технической задачей, поставленной в данном техническом решении, является создание нового эффективного, простого и надежного способа диаризации аудиосигнала, обеспечивающего возможность принятия решения о принадлежности аудиозаписи конкретному диктору.

Техническим результатом, достигаемым при решении вышеуказанной технической проблемы или технической задачи, является обеспечение возможности разметки (сегментации) аудиосигнала с малой погрешностью и с малым энергопотреблением, на основе данных, полученных с 2 микрофонов, в том числе, в режиме реального времени.

Указанный технический результат достигается благодаря осуществлению способа диаризации речевого аудиосигнала, содержащего этапы, на которых

получают цифровые аудиосигналы, содержащие данные голоса, синхронно регистрируемые по меньшей мере двумя микрофонами;

определяют разностный сигнал для сигналов двух микрофонов на основе данных цифровых аудиосигналов, полученных от упомянутых микрофонов;

определяют значения огибающей функции разностного сигнала;

определяют значения огибающей функции исходного аудиосигнала на основе данных цифрового аудиосигнала, полученного от одного из микрофонов;

на основе значения огибающей функции разностного сигнала и значения огибающей функции исходного аудиосигнала определяют характеристическое значение аудиосигнала;

на основе характеристического значения аудиосигнала осуществляют разметку данных цифрового аудиосигнала, указывающую на то, к какому источнику звукового сигнала относится соответствующий блок данных цифрового аудиосигнала.

В одном из частных примеров осуществления способа два микрофона разнесены относительно друг друга по вертикали.

В другом частном примере осуществления способа разностный сигнал определяют посредством посэмплового вычета (для синхронно полученных сигналов) значения величины сигнала, определенного для аудиосигнала, поступившего с одного из микрофонов, из значения величины сигнала, определенного для аудиосигнала, поступившего с другого микрофона.

В другом частном примере осуществления способа характеристическое значение аудиосигнала (sp) определяется по формуле

$$sp = Env(11\text{-}12)/Env(11),$$

где Env(11-12) - значение огибающей функции разностного сигнала, а

Env(11) - значение огибающей функции исходного аудиосигнала, полученного от одного из микрофонов.

В другом частном примере осуществления способа разметка данных цифрового аудиосигнала осуществляется посредством сравнения характеристического значения аудиосигнала с заранее заданным пороговым значением, причем если характеристическое значение аудиосигнала больше порогового значения, то соответствующий блок данных цифрового аудиосигнала размечается как относящийся к первому источнику звукового сигнала, а если характеристическое значение аудиосигнала меньше порогового значения, то соответствующий блок данных цифрового аудиосигнала размечается как относящийся ко второму источнику звукового сигнала.

В другом частном примере осуществления способа разметка данных цифрового аудиосигнала осуществляется посредством разделения записанного аудиопотока речи дикторов по каналам стерео.

В другом частном примере осуществления способа разметка данных цифрового аудиосигнала осуществляется посредством создания дополнительного блока данных с указанием временных меток, характеризующих время записи реплик по меньшей мере одного диктора.

В другом предпочтительном варианте осуществления заявленного решения представлено устройство диаризации речевого аудиосигнала, содержащее по меньшей мере одно вычислительное устройство и по меньшей мере одно устройство памяти, содержащее машиночитаемые инструкции, которые при их исполнении по меньшей мере одним вычислительным устройством выполняют вышеуказанный способ.

Краткое описание чертежей

Признаки и преимущества настоящего технического решения станут очевидными из приводимого ниже подробного описания изобретения и прилагаемых чертежей, на которых

на фиг. 1 представлена общая схема расположения микрофонов;

на фиг. 2 представлен пример параметров речевого сигнала;

на фиг. 3 представлена схема устройства диаризации аудиосигнала;

на фиг. 4 подставлен пример общего вида вычислительного устройства.

Осуществление изобретения

Ниже будут описаны понятия и термины, необходимые для понимания данного технического решения.

В данном техническом решении под системой подразумевается, в том числе компьютерная система, ЭВМ (электронно-вычислительная машина), ЧПУ (числовое программное управление), ПЛК (программируемый логический контроллер), компьютеризированные системы управления и любые другие устройства, способные выполнять заданную, четко определенную последовательность операций (действий, инструкций).

Под устройством обработки команд подразумевается электронный блок, вычислительное устройство, либо интегральная схема (микропроцессор), исполняющая машинные инструкции (программы).

Устройство обработки команд считывает и выполняет машинные инструкции (программы) с одного или более устройств хранения данных. В роли устройства хранения данных могут выступать, но не ограничиваясь, жесткие диски (HDD), флеш-память, ПЗУ (постоянное запоминающее устройство), твердотельные накопители (SSD), оптические приводы.

Программа - последовательность инструкций, предназначенных для исполнения устройством управления вычислительной машины или устройством обработки команд.

Блок данных - последовательность битов, имеющая фиксированную длину и используемая для представления данных в памяти или для их пересылки. На фиг. 1 представлена схема расположения микрофонов относительно источников звукового сигнала. В частности, на схеме изображены первый источник звукового сигнала - оператор 1, второй источник звукового сигнала - клиент 2, устройство 10 регистрации звукового сигнала, содержащее микрофоны 11 и 12. В качестве устройства 10 регистрации звукового сигнала может быть использовано любое известное вычислительное устройство, модифицированное в программно-аппаратной части такими образом, чтобы обеспечить сбор, обработку и хранение данных звукового сигнала. Например, упомянутое устройство 10 может быть выполнено в виде цифрового бейджа/"знака отличия", заранее размещенного определенным образом относительно первого источника звукового сигнала - оператора 1. Например, устройство 10 регистрации звукового сигнала может быть размещено на груди оператора с левой стороны. Поскольку расположение устройства 10 относительно первого источника звукового сигнала заранее известно, соответственно, решение задачи определения направления звукового сигнала голоса оператора является избыточным и для диаризации достаточно выполнить проверку на соответствие направления источника звука ожидаемому направлению. Такое упрощение алгоритма позволяет для данного класса устройств (с предопределенным сценарием использования) осуществлять диаризацию записей, используя при этом минимальные вычислительные ресурсы. Эта особенность наряду с применением радиоэлектронных компонентов с низким энергопотреблением обеспечивает возможность снизить общее энергопотребление устройством, что, соответственно, позволяет обеспечить длительное время работы при малом весе и объеме аккумуляторной батареи. Логика работы алгоритма диаризации основывается на том факте, что источники звукового сигнала (рот оператора 1 и рот клиента 2) располагаются в разных точках пространства и рот оператора 1 смещен незначительно относительно прямой, на которой расположены микрофоны 11 и 12 (предположительно лежащей в вертикальной плоскости). В связи с этим временная задержка между сигналами двух расположенных друг под другом на заданном расстоянии микрофонов 11 и 12 от звуков речи оператора 1 будет максимальна в то время, как временная задержка между сигналами тех же микрофонов от звуков речи клиента 2 будет значительно меньше.

Например, если расстояние между микрофонами будет составлять примерно 50 мм (см. фиг 1), бейдж будет находиться на груди оператора на расстоянии примерно 20 см от его рта и расстояние между оператором и клиентом примерно = 1 м, то звуковой волне от клиента 2 до нижнего микрофона 12 нужно будет дополнительно пройти около 10 мм, после поступления звуковой волны на микрофон 11, в то время как для звуковой волны, обусловленной речью оператора, разность акустического пути между микрофонами 11 и 12 составит примерно 50 мм. Таким образом, при условии равенства характеристик

обоих микрофонов при вычитании сигнала одного из микрофонов из сигнала другого микрофона амплитуда результирующего сигнала для низких частот (до 300 Гц) будет пропорциональна фазовой разности, обусловленной временной задержкой при распространении сигнала.

На фиг. 2 представлен пример зависимости величины звукового давления от времени для регистрируемых звуковых волн, на которой изображено

sin(pi*x) - исходный сигнал (сигнал, зарегистрированный микрофоном 11),

sin(pi*(x-1*k)) - сигнал, полученный вторым микрофоном (12), для случая когда сигнал пришел бы из точки, где расположен клиент (малая задержка между моментами достижения фронтом звуковой волны микрофонов 11 и 12),

sin(pi*(x-5*k)) - сигнал, полученный вторым микрофоном (12), для случая когда сигнал пришел бы из точки, где расположен рот оператора (большая задержка между моментами достижения фронтом звуковой волны микрофонов 11 и 12),

delta1 - разностный сигнал, в случае если источником сигнала является голос клиента {sin(pi*(x)-sin(pi*(x-1*k))},

delta2 - разностный сигнал, в случае если источником сигнала является голос оператора {sin(pi*(x)-sin(pi*(x-5*k))}.

pi - число "Пи",

x - переменная времени для иллюстрации графика зависимости величины звукового давления, воспринимаемого микрофонами, от времени;

k - величина задержки, обусловленная наличием для звуковой волны разности акустического пути от источника до микрофонов (11) и (12) при условии, что ее источником является клиент (2) (в случае когда задержка равна 1*k) или оператор (1) (в случае когда задержка равна 5*k).

Далее будет описан способ диаризации аудиосигнала со ссылкой на фиг. 3, на котором представлен пример схемы устройства 100 диаризации аудиосигнала. Устройство 100 диаризации аудиосигнала может быть реализовано на базе устройства 10 регистрации звукового сигнала и содержать модуль 101 обработки сигналов, модуль 102 определения характеристик сигнала, модуль 103 разметки аудиосигнала и модуль 104 хранения данных. Перечисленные модули могут быть реализованы на базе программно-аппаратных средств устройства 100 диаризации аудиосигнала, выполненных в программной части таким образом, чтобы выполнять приписанные им ниже функции.

Соответственно цифровые аудиосигналы, содержащие данные голоса оператора 1 или клиента 2 и синхронно регистрируемые микрофонами 11 и 12, поступают в буфер модуля 101 обработки сигналов в виде потока данных. Из полученных цифровых аудиосигналов модуль 101 формирует массив данных.

Сформированный массив данных цифрового аудиосигнала направляется модулем 101 в модуль 102 определения характеристик сигнала.

Далее модуль 102 определения характеристик сигнала определяет разностный сигнал для сигналов двух микрофонов посредством посэмплового вычета (для синхронно полученных сэмплов) значения величины сигнала, определенного для аудиосигнала, поступившего с микрофона 12, из значения величины сигнала, определенного для аудиосигнала, поступившего с микрофона 11. После этого известными из уровня техники методами модуль 102 обработки сигналов определяет значения огибающей функции разностного сигнала и огибающей функции исходного аудиосигнала, полученного от микрофона 11 или 12, и на основе полученных значений определяет характеристические значения аудиосигнала. Например, характеристические значения аудиосигнала (sp) могут быть определены по формуле

$$sp=Env(11\text{-}12)/Env(11),$$

где Env(11-12) - значение огибающей функции разностного сигнала, а

Env(11) - значение огибающей функции исходного аудиосигнала, полученного от микрофона 11.

Для того чтобы характеристическое значение аудиосигнала обладало физическим смыслом необходимо, чтобы микрофоны 11 и 12 рассматриваемой пары были разнесены относительно друг друга по вертикали, а устройство 10 регистрации звукового сигнала располагалось на груди оператора таким образом, чтобы рот оператора находился на прямой, проходящей через центр апертур микрофонов 11 и 12 или с незначительным отклонением от нее. Выполнение шага по определению характеристического значения аудиосигнала является необходимым для нормирования сигнала на собственную амплитуду, чтобы характеристическое значение аудиосигнала было зависимым только от расположения источников звука, а не от амплитуды сигнала.

Далее данные характеристических значений аудиосигнала и массив данных цифрового аудиосигнала упомянутый модуль 102 направляет в модуль 103 разметки аудиосигнала, который на основе характеристического значения аудиосигнала осуществляет разметку данных цифрового аудиосигнала, указывающую на то, к какому источнику звукового сигнала блоки данных цифрового аудиосигнала относится. Например, характеристические значения аудиосигнала упомянутым модулем 103 могут быть сравнены с заранее заданным пороговым значением, и если характеристические значения аудиосигнала больше порогового значения, то блок данных цифрового аудиосигнала размечается как относящийся к первому источнику звукового сигнала - оператору (1). Если характеристические значения аудиосигнала меньше порогового значения, то блок данных цифрового аудиосигнала размечается как относящийся ко второму

источнику звукового сигнала - клиенту (2). Разметка аудиоданных может осуществляться как посредством разделения записанного аудиопотока речи дикторов по каналам стерео (речь одного из дикторов - в правый канал, другого - в левый) с последующим сохранением их в виде аудиофайла, либо в виде дополнительного блока данных (отдельного файла либо дополнительной дорожки вышеуказанного аудиофайла) с указанием временных меток, характеризующих время записи реплик одного либо обоих дикторов.

Полученный аудиофайл с разметкой и/или дополнительный файл разметки аудиоданных может быть сохранен в памяти модуля 104 хранения данных для его передачи в дальнейшем на внешние устройства и системы обработки данных через соответствующие интерфейсы вывода данных, которые будут раскрыты далее в тексте описания.

Таким образом, за счет того что источник звукового сигнала определяется на основе характеристического значения аудиосигнала, полученного на основе значения огибающей функции разностного сигнала и значения огибающей функции исходного аудиосигнала, повышается точность его определения и снижается погрешность при разметки данных цифрового аудиосигнала на основе характеристического значения аудиосигнала. Также снижается энергопотребление при разметке данных цифрового аудиосигнала за счет того, что для разметки данных цифрового аудиосигнала не требуется осуществлять дополнительную обработку полученных с микрофонов цифровых аудиосигналов. В общем виде (см. фиг. 3) вычислительное устройство (200) содержит объединенные общей шиной информационного обмена один или несколько процессоров (201), средства памяти, такие как ОЗУ (202) и ПЗУ (203), интерфейсы ввода/вывода (204), устройства ввода/вывода (205) и устройство для сетевого взаимодействия (206).

Процессор (201) (или несколько процессоров, многоядерный процессор и т.п.) может выбираться из ассортимента устройств, широко применяемых в настоящее время, например, таких производителей, как Intel™, AMD™, Apple™, Samsung Exynos™, MediaTEK™, Qualcomm Snapdragon™ и т.п. Под процессором или одним из используемых процессоров в системе (200) также необходимо учитывать графический процессор, например GPU NVIDIA с программной моделью, совместимой с CUDA, или Graphcore, тип которых также является пригодным для полного или частичного выполнения способа, а также может применяться для обучения и применения моделей машинного обучения в различных информационных системах.

ОЗУ (202) представляет собой оперативную память и предназначено для хранения исполняемых процессором (201) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (202), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.). При этом в качестве ОЗУ (202) может выступать доступный объем памяти графической карты или графического процессора.

ПЗУ (203) представляет собой одно или более устройств постоянного хранения данных, например жесткий диск (HDD), твердотельный накопитель данных (SSD), флэш-память (EEPROM, NAND и т.п.), оптические носители информации (CD-R/RW, DVD-R/RW, BlueRay Disc, MD) и др.

Для организации работы компонентов устройства (200) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (204). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь, PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SATA, IDE, Lightning, USB (2.0, 3.0, 3.1, micro, mini, type C), TRS/Audio jack (2.5, 3.5, 6.35), HDMI, DVI, VGA, Display Port, RJ45, RS232 и т.п.

Для обеспечения взаимодействия пользователя с устройством (200) применяются различные средства (205) В/В информации, например клавиатура, дисплей (монитор), сенсорный дисплей, тач-пад, джойстик, манипулятор мышь, световое перо, стилус, сенсорная панель, трекбол, динамики, микрофон, средства дополненной реальности, оптические сенсоры, планшет, световые индикаторы, проектор, камера, средства биометрической идентификации (сканер сетчатки глаза, сканер отпечатков пальцев, модуль распознавания голоса) и т.п. Средство сетевого взаимодействия (206) обеспечивает передачу данных посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (206) может использоваться, но не ограничиваться, Ethernet карта, GSM модем, GPRS модем, LTE модем, 5G модем, модуль спутниковой связи, NFC модуль, Bluetooth и/или BLE модуль, Wi-Fi модуль и др.

Дополнительно могут применяться также средства спутниковой навигации в составе устройства (200), например GPS, ГЛОНАСС, BeiDou, Galileo.

Конкретный выбор элементов устройства (200) для реализации различных программно-аппаратных архитектурных решений может варьироваться с сохранением обеспечиваемого требуемого функционала.

Модификации и улучшения вышеописанных вариантов осуществления настоящего технического решения будут ясны специалистам в данной области техники. Предшествующее описание представлено только в качестве примера и не несет никаких ограничений. Таким образом, объем настоящего технического решения ограничен только объемом прилагаемой формулы изобретения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ диаризации аудиосигнала, выполняемый по меньшей мере одним вычислительным устройством, содержащий этапы, на которых

получают цифровые аудиосигналы, содержащие данные голоса, синхронно регистрируемые по меньшей мере двумя микрофонами;

определяют разностный сигнал для сигналов двух микрофонов на основе данных цифровых аудиосигналов, полученных от упомянутых микрофонов;

определяют значения огибающей функции разностного сигнала;

определяют значения огибающей функции исходного аудиосигнала на основе данных цифрового аудиосигнала, полученного от одного из микрофонов;

на основе значения огибающей функции разностного сигнала и значения огибающей функции исходного аудиосигнала определяют характеристическое значение аудиосигнала;

на основе характеристического значения аудиосигнала осуществляют разметку данных цифрового аудиосигнала, указывающую на то, к какому источнику звукового сигнала относится соответствующий блок данных цифрового аудиосигнала.

2. Способ по п.1, характеризующийся тем, что по меньшей мере два микрофона разнесены относительно друг друга по вертикали.

3. Способ по п.1, характеризующийся тем, что разностный сигнал определяют посредством посэмплового вычета (для синхронно полученных сигналов) значения величины сигнала, определенного для аудиосигнала, поступившего с одного из микрофонов, из значения величины сигнала, определенного для аудиосигнала, поступившего с другого микрофона.

4. Способ по п.1, характеризующийся тем, что характеристическое значение аудиосигнала (sp) определяется по формуле

$$sp=Env(11\text{-}12)/Env(11),$$

где Env(11-12) - значение огибающей функции разностного сигнала, а

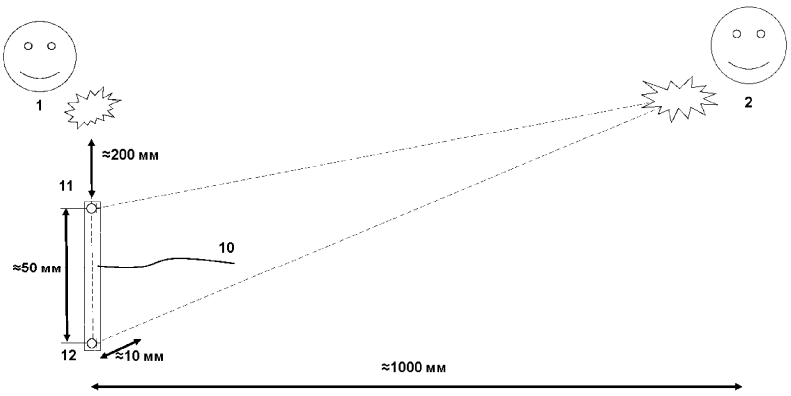
Env(11) - значение огибающей функции исходного аудиосигнала, полученного от одного из микрофонов.

5. Способ по п.1, характеризующийся тем, что разметка данных цифрового аудиосигнала осуществляется посредством сравнения характеристического значения аудиосигнала с заранее заданным пороговым значением, причем если характеристическое значение аудиосигнала больше порогового значения, то соответствующий блок данных цифрового аудиосигнала размечается как относящийся к первому источнику звукового сигнала, а если характеристическое значение аудиосигнала меньше порогового значения, то соответствующий блок данных цифрового аудиосигнала размечается как относящийся ко второму источнику звукового сигнала.

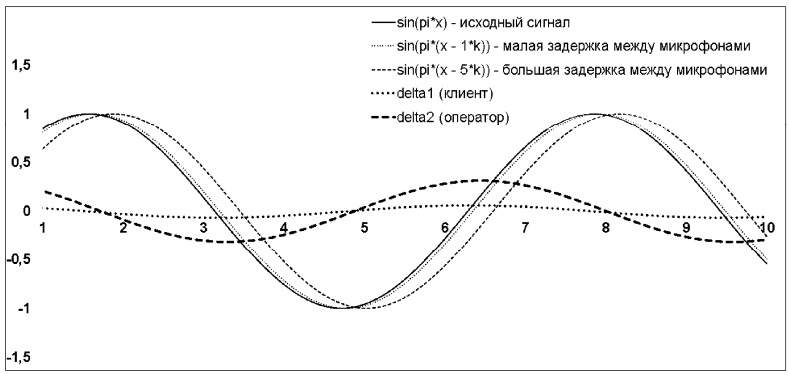
6. Способ по п.1, характеризующийся тем, что разметка данных цифрового аудиосигнала осуществляется посредством разделения записанного аудиопотока речи дикторов по каналам стерео.

7. Способ по п.1, характеризующийся тем, что разметка данных цифрового аудиосигнала осуществляется посредством создания дополнительного блока данных с указанием временных меток, характеризующих время записи реплик по меньшей мере одного диктора.

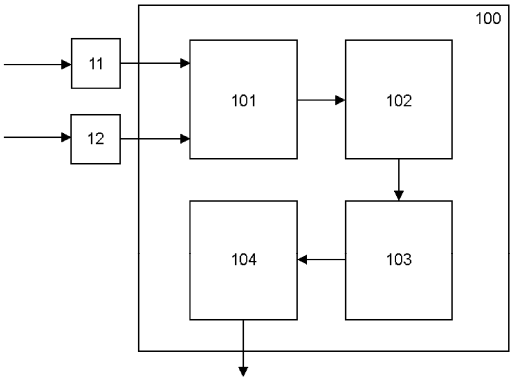
8. Устройство диаризации речевого аудиосигнала, содержащее по меньшей мере одно вычислительное устройство и по меньшей мере одно устройство памяти, содержащее машиночитаемые инструкции, которые при их исполнении по меньшей мере одним вычислительным устройством выполняют способ по любому из пп.1-7.



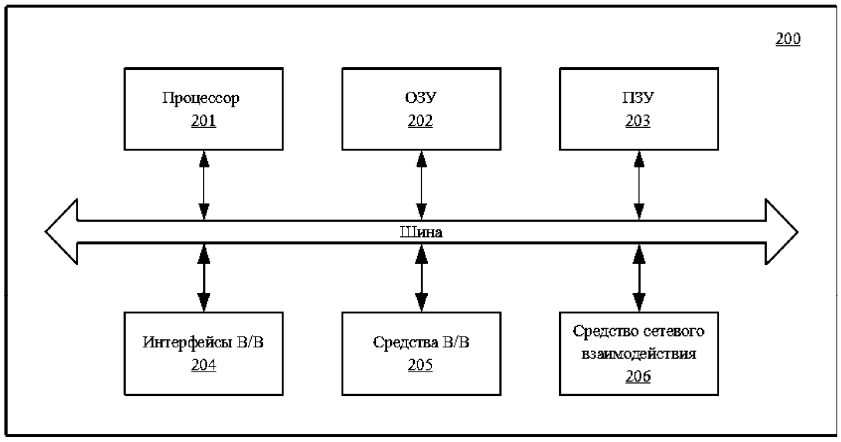
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4