



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G10L 15/02 (2023.01)

(21)(22) Заявка: 2022103928, 16.02.2022

(24) Дата начала отсчета срока действия патента:
27.05.2021

Дата регистрации:
28.02.2023

Приоритет(ы):

(62) Номер и дата подачи первоначальной заявки,
из которой данная заявка выделена:
2021115228 27.05.2021

(45) Опубликовано: 28.02.2023 Бюл. № 7

Адрес для переписки:
117997, Москва, ул. Вавилова, 19, ПАО
Сбербанк, Правовой департамент

(72) Автор(ы):

Оболенский Иван Александрович (RU),
Вышегородцев Кирилл Евгеньевич (RU),
Губанов Дмитрий Николаевич (RU),
Богданов Илья Владимирович (RU)

(73) Патентообладатель(и):

Публичное акционерное общество "Сбербанк
России" (ПАО Сбербанк) (RU)

(56) Список документов, цитированных в отчете
о поиске: US 20150288791 A1, 08.10.2015. WO
2017218243 A2, 21.12.2017. US 10841424 B1,
17.11.2020. EP 2418643 A1, 15.02.2012. US
2004013253 A1, 22.01.2004.

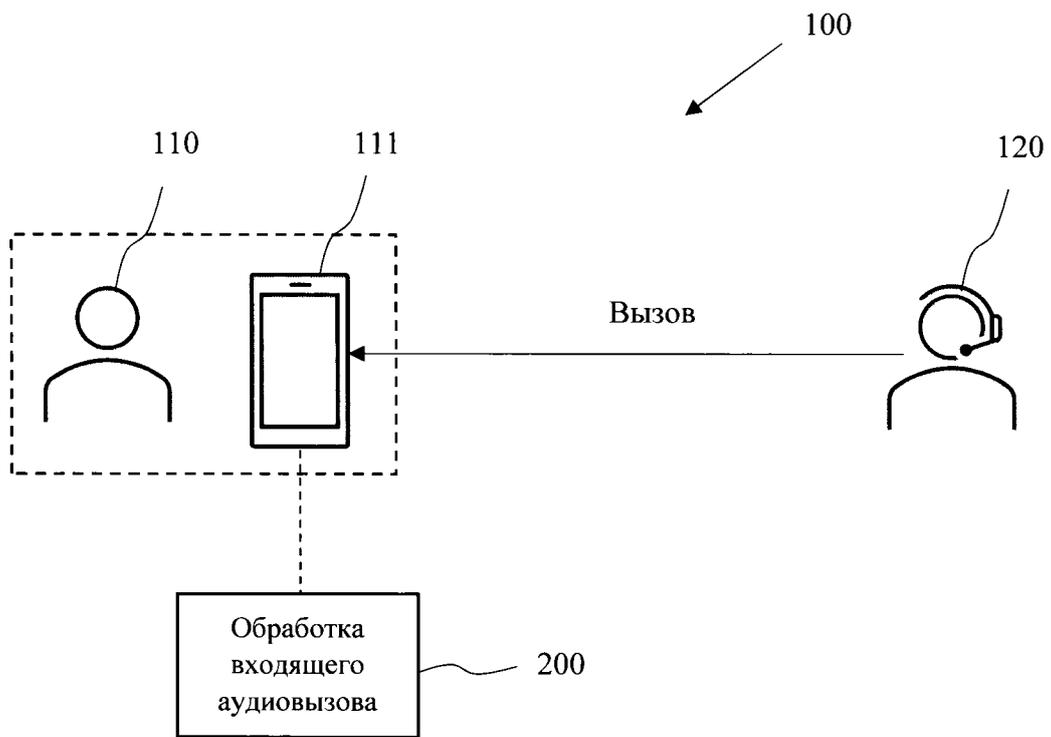
(54) СПОСОБ И СИСТЕМА АНАЛИЗА ГОЛОСОВЫХ ВЫЗОВОВ НА ПРЕДМЕТ ВЫЯВЛЕНИЯ И
ПРЕДОТВРАЩЕНИЯ СОЦИАЛЬНОЙ ИНЖЕНЕРИИ

(57) Реферат:

Настоящее техническое решение относится к области вычислительной техники, в частности к обработке данных входящих аудиовызовов для классификации наличия состава мошеннических действий. Техническим результатом является повышение эффективности и точности распознавания мошеннической активности входящих аудиовызовов за счет комбинированного анализа аудиопотока и семантики паттерна диалога. Заявленный технический результат достигается за счет выполнения компьютерно-реализуемого способа анализа диалога во время аудиовызовов на предмет выявления мошеннической активности, выполняемого с помощью процессора и содержащего этапы, на которых: получают

входящий аудиопоток, поступающий от вызывающей стороны; осуществляют обработку входящего аудиопотока с помощью по меньшей мере одной модели машинного обучения, в ходе которой:

преобразовывают входящий аудиопоток в векторную форму; выполняют сравнение векторной формы аудиопотока с ранее сохраненными векторами, характеризующими мошенническую активность; осуществляют транскрибирование аудиопотока для анализа диалога вызывающей стороны на по меньшей мере семантический состав информации и паттерн ведения диалога; осуществляют классификацию входящего аудиопотока на основании выполненной обработки. 2 н. и 13 з.п. ф-лы, 5 ил.



Фиг. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**

(52) CPC

G10L 15/02 (2023.01)(21)(22) Application: **2022103928, 16.02.2022**(24) Effective date for property rights:
27.05.2021Registration date:
28.02.2023

Priority:

(62) Number and date of filing of the initial application,
from which the given application is allocated:
2021115228 27.05.2021(45) Date of publication: **28.02.2023** Bull. № 7

Mail address:

**117997, Moskva, ul. Vavilova, 19, PAO Sberbank,
Pravovoj departament**

(72) Inventor(s):

**Obolenskij Ivan Aleksandrovich (RU),
Vyshegorodtsev Kirill Evgenevich (RU),
Gubanov Dmitrij Nikolaevich (RU),
Bogdanov Ilya Vladimirovich (RU)**

(73) Proprietor(s):

**Publichnoe aktsionernoe obshchestvo "Sberbank
Rossii" (PAO Sberbank) (RU)**(54) **METHOD AND SYSTEM FOR ANALYZING VOICE CALLS FOR SOCIAL ENGINEERING DETECTION AND PREVENTION**

(57) Abstract:

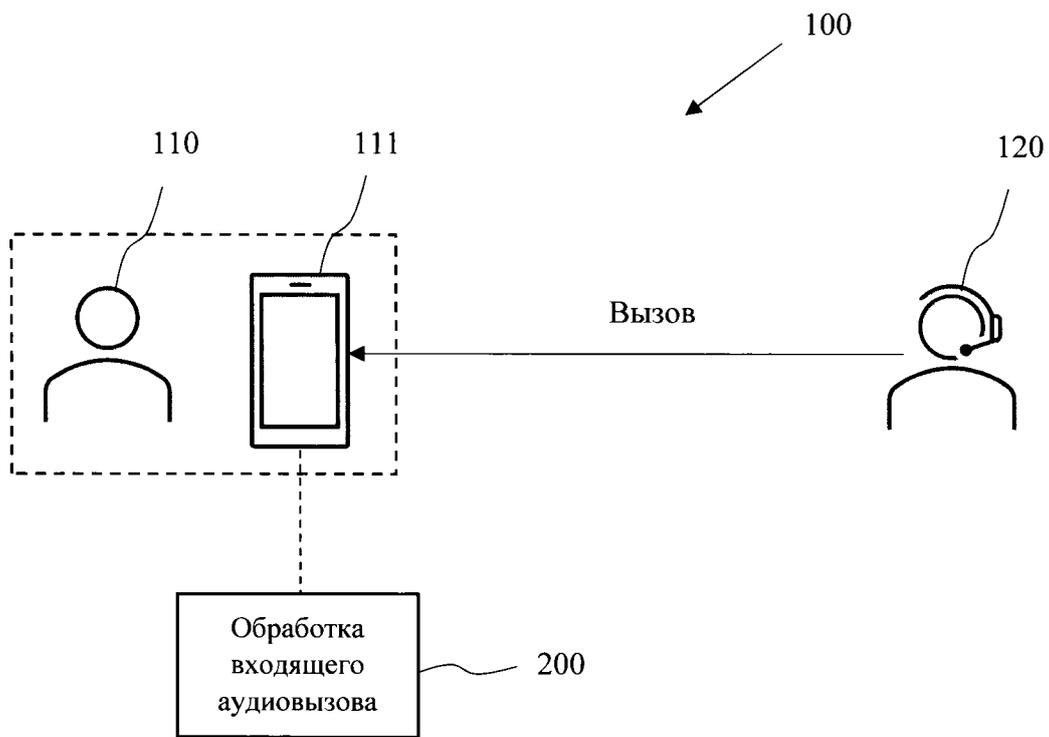
FIELD: computer technology.

SUBSTANCE: processing of data from incoming audio calls to classify the presence of fraudulent activities. The claimed effect of the invention is achieved by a computer-implemented method of analysing the dialogue during audio calls to detect fraudulent activity, powered by a processor and containing the following steps: receive an incoming audio stream coming from the calling party; process the incoming audio stream using at least one machine-learning model, which includes: converting the incoming audio stream into a vector form; comparing

the vector form of the audio stream with previously stored vectors characterizing fraudulent activity; transcribing the audio stream to analyse the dialogue of the calling party for at least the semantic composition of the information and the pattern of the dialogue; classifying the incoming audio stream based on the processing.

EFFECT: increased efficiency and accuracy of recognizing fraudulent activity in incoming audio calls due to the combined analysis of the audio stream and the semantics of the dialogue pattern.

15 cl, 5 dwg



Фиг. 1

ОБЛАСТЬ ТЕХНИКИ

[0001] Настоящее техническое решение относится к области вычислительной техники, в частности к обработке данных входящих аудиовызовов для классификации наличия состава мошеннических действий.

УРОВЕНЬ ТЕХНИКИ

[0002] Применение различных методов в части анализа аудиопотоков для их последующей классификации является достаточно распространенным подходом, применяемым в различных областях техники и бизнеса. Возросшая активность преступлений в области кибербезопасности особенно часто находит свое отражение в сфере финансов, что негативно сказывается как на благосостоянии клиентов, так и на репутации финансовых институтов. Наиболее частым приемом, применяемым мошенниками при телефонных звонках, является социальная инженерия, при котором клиента вводят в заблуждение и вынуждают самостоятельно совершить определенные действия, приводящие, как правило, к хищению денежных средств.

[0003] Одним из примеров решений, направленных на борьбу с мошеннической активностью, является способ определения риск-балла звонка, который заключается в анализе речевой информации звонящего и ее классификации на наличие заданных триггеров, свидетельствующих о намерениях звонящего (US 20170142252 A1, 18.05.2017).

[0004] Другим примером подходов является обнаружение изменения голоса звонящего или формирование синтетической речи, воспроизводимой роботом или ботом, на основе выделения из звуковой дорожки характерных признаков, свидетельствующих о синтетической природе звука (US 10944864 B2, 09.03.2021).

[0005] Основным недостатком известных решений является отсутствие комплексного подхода, позволяющего проводить многосторонний анализ аудиопотока на предмет выявления ряда характеристик, в частности помимо анализа звуковой составляющей диалога осуществлять транскрибирование звуковой информации для обработки паттерна диалога звонящего. Также, недостатком является отсутствие автоматизированных способов защиты абонента от мошеннических действий при входящих вызовах, а также автоматическое получение мошеннических аудиопотоков.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

[G006] Решаемой технической проблемой с помощью заявленного изобретения является повышение эффективности распознавания мошеннической активности.

[0007] Техническим результатом является повышение эффективности и точности распознавания мошеннической активности входящих аудиовызовов, за счет комбинированного анализа аудиопотока и семантики паттерна диалога.

[0008] Заявленный технический результат достигается за счет выполнения компьютерно-реализуемого способа анализа диалога во время аудиовызовов на предмет выявления мошеннической активности, выполняемого с помощью процессора и содержащего этапы, на которых:

- получают входящий аудиопоток, поступающий от вызывающей стороны;
- осуществляют преобразование входящего аудиопотока в векторную форму;
- осуществляют обработку преобразованного аудиопотока с помощью первой модели машинного обучения, в ходе которой выполняют сравнение векторной формы аудиопотока с ранее сохраненными векторами, характеризующими мошенническую активность;
- осуществляют транскрибирование аудиопотока и его последующую обработку с помощью второй модели машинного обучения, которая выполняет анализ диалога вызывающей стороны, при этом в ходе упомянутого анализа осуществляется:

семантический состав информации и паттерн ведения диалога, при этом паттерн ведения диалога включает в себя анализ слов, используемых в разговоре, анализ построения фраз, анализ следование фраз друг за другом;

наличие и длительность пауз в диалоге входящего аудиопотока;

5 - осуществляют классификацию входящего аудиопотока на основании выполненной обработки первой и второй моделями машинного обучения.

[0009] В одном из частных примеров реализации способа при семантическом анализе транскрибированного диалога выполняется выявление слов, присущих мошеннической активности.

10 [0010] В другом частном примере реализации способа дополнительно входящий аудиопоток анализируется на по меньшей мере одно из: тональность, эмотивность, просодия или их сочетания.

[0011] В другом частном примере реализации способа векторная форма входящего аудиопотока анализируется на предмет наличия признаков, выбираемых из группы: 15 изменение голоса, синтетическое формирование голоса, наложение фонового аудиопотока или их сочетания.

[0012] В другом частном примере реализации способа дополнительно анализируют исходящий аудиопоток.

20 [0013] В другом частном примере реализации способа выполняют разделение исходящего и входящего аудиопотоков.

[0014] В другом частном примере реализации способа дополнительно анализируется по меньшей мере один параметр входящего аудиопотока, выбираемый из группы: 25 высота тембра, сила звука, интенсивность речи, длительность произнесения слов, придыхание, глоттализация, палатализация, тип примыкания согласного к гласному или их сочетания.

[0015] В другом частном примере реализации способа дополнительно анализируется наличие посторонних шумов во входящем аудиопотоке.

[0016] В другом частном примере реализации способа выполняется на устройстве пользователя, представляющим собой смартфон, планшет или компьютер.

30 [0017] В другом частном примере реализации способа при получении входящей аудиодорожки выполняется генерирование синтетического исходящего голосового аудиопотока.

[0018] В другом частном примере реализации способа генерирование исходящего аудиопотока выполняется до момента классификации входной аудиодорожки.

35 [0019] В другом частном примере реализации способа генерирование синтетического аудиопотока осуществляется на основании голосового образца пользователя устройства.

[0020] В другом частном примере реализации способа при классификации входящего аудиопотока как мошеннического выполняется сохранение его векторного 40 представления.

[0021] В другом частном примере реализации способа при классификации входящего аудиопотока как мошеннического выполняется генерирование сообщения о статусе, отображаемое на дисплее устройства.

[0022] Заявленный технический результат также достигается с помощью системы анализа диалога во время аудиовызовов на предмет выявления мошеннической 45 активности, которая содержит по меньшей мере один процессор и по меньшей мере одну память, хранящую машиночитаемые инструкции, которые при их выполнении процессором реализуют вышеуказанный способ.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0023] Фиг. 1 иллюстрирует общую схему заявленного решения.

[0024] Фиг. 2А иллюстрирует блок-схему общего процесса анализа аудиопотока вызова.

5 [0025] Фиг. 2Б иллюстрирует блок-схему процесса анализа аудиопотока на предмет синтетических изменений.

[0026] Фиг. 3 иллюстрирует блок-схему процесса формирования синтетического исходящего аудиопотока для ведения диалога.

[0027] Фиг. 4 иллюстрирует общую схему вычислительного устройства.

ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

10 [0028] На Фиг. 1 представлена общая схема (100) заявленного решения. Решение основано на базе программно-аппаратного комплекса, реализуемого на одном или нескольких вычислительных устройствах, например, на смартфоне (111) пользователя (110), или связанном с ним устройством, которое может обеспечивать обработку входящих аудиовызовов, поступающих от стороннего абонента (120). Под
15 аудиовызовами следует понимать, например, звонки посредством телефонной связи, звонки, осуществляемые посредством мессенджеров (WhatsApp, Viber, Telegram, Facebook Messenger и др.) через сеть Интернет, в том числе видеовызовы.

[0029] Поступающие аудиовызовы от абонентов (120) поступают на дальнейшую обработку (200), выполняемую с помощью программной логики, реализуемой
20 вычислительным устройством, например, смартфоном (111). Обработка (200) выполняется посредством одной или нескольких моделей машинного обучения, которые обучены выполнять обработку входящего аудиопотока (аудиодорожки) для анализа на предмет риска мошеннической активности со стороны абонента (120).

[0030] На Фиг. 2А представлена схема выполнения способа (200) обработки
25 аудиопотока, выполняемые при получении входящего вызова. На первом этапе (201) осуществляется получение аудиовызова и захват входящего аудиопотока. Захват может осуществляться широко известными из уровня техники средствами записи диалога, например, с помощью специализированного ПО (Voice Recorder, Cube ACR и т.п.). Полученный аудиопоток обрабатывается параллельно для одновременного анализа
30 как аудио составляющей, так и семантики диалога.

[0031] Полученный на этапе (201) аудиопоток преобразовывается в векторный формат (эмбединг, от англ.) на этапе (202) для последующей передачи в модель машинного обучения на этапе (203) для анализа на наличие совпадения с ранее
35 зафиксированными голосовыми эмбедингами мошенников. Преобразование входного аудиопотока может выполняться с помощью технологии IBM Audio Embedding Generator (<https://developer.ibm.com/technologies/artificial-intelligence/models/max-generator/>).

[0032] Ранее известные векторные представления аудиопотоков, для которых была зафиксирована мошенническая активность, могут храниться в базе данных (БД). БД
40 указанных эмбедингов может размещаться на удаленном сервере, связь с которым во время аудиовызова устанавливается через смартфон (111). При этом БД может также дублироваться непосредственно на само устройство (111).

[0033] На этапе (204) по итогам обработки эмбединга с помощью модели машинного обучения, обеспечивающей классификацию входящего аудиопотока, принимается
45 решение о характере аудиовызова звонящего. Если сравнение эмбедингов показывает, что найдено совпадение, выше, чем установленный порог для классификации моделью машинного обучения, то аудиовызов классифицируется как имеющий мошеннический характер (этап 210). В противном случае аудиовызов классифицируется как безопасный (этап 220).

[0034] Примером такой модели может быть модель, построенная на основе «метода опорных векторов», модель на основе линейной или нелинейной регрессии, модель на основе метода «k-соседей». В одном из вариантов реализации использует поиск одной ближайшей записи на основе Евклидова расстояния между векторами. В другом варианте реализации может использоваться расстояние Махаланобиса. Также, в одном из частных примеров реализации может использоваться косинусное расстояние, коэффициент корреляции Пирсона, расстояние Минковского r -степени и прочее.

[0035] Параллельно с выполнением этапа (202) осуществляется транскрибирование аудиопотока на этапе (205), для чего входящий аудиопоток преобразуется в текстовый формат. Данная процедура может выполняться различными известными алгоритмами, обеспечивающими преобразование аудиодорожки в текст, например, технологии Speech-To-Text. Также может применяться модель машинного обучения для осуществления процедуры транскрибирования.

[0036] Для выполнения анализа аудиопотоков применяется также алгоритм по разделению голосов собеседников в многоголосовом диалоге, который обеспечивает очистку звуковых дорожек от шумов и другого вида артефактов, что обеспечивает более четкий аудиосигнал. Как пример, для этого можно применить подходы, основанные на NMF-разложении (Non-negative matrix factorization) исходного или преобразованного сигнала, использование сверточных искусственных нейронных сетей (Convolutional Neural Network), моделей «Cone of Silence» и иные подходы.

[0037] Переведенный в текстовую форму аудиопоток анализируется на этапе (206) на предмет классификации паттерна ведения диалога звонящим абонентом (120). Классификация может осуществляться с применением технологий анализа естественного языка (NLP - Natural Language Processing), в том числе могут применяться технологии на базе машинного обучения. С помощью обученной модели на этапе (206) выполняется анализ текстовых данных для их последующего отнесения к классам, характеризующим мошенническое поведение, например, свидетельствующих о факте социальной инженерии. Примером социальной инженерии могут служить фразы, в которых от клиента (110) требуют срочно перевести его деньги на чужой счет, просят сообщить полный номер карты, требуют взять кредит, спрашивают CVV-код, код подтверждения или код из смс и т.п.

[0038] Под «классом» или «классами» понимается по меньшей мере класс с содержанием данных по мошенникам или класс с данными не мошенников. Также, классификация может являться нечеткой, когда нельзя однозначно осуществить классификацию - мошенник и не мошенник (2 класса); 3 класса - мошенник, не мошенник, неизвестно; несколько классов - мошенник типа А, мошенник типа Б и так далее.

[0039] Выходом работы модели на этапе (206) является классификация паттерна ведения диалога на этапе (207). Под паттерном следует понимать, в частности, слова, используемые в разговоре, построение фраз, следование фраз друг за другом и т.п. Модель классификации обучена на примерах диалогов, подтвержденного факта мошеннической активности, в частности на паттернах, позволяющих осуществить последующую классификацию данных при обработке входных аудиопотоков.

[0040] Модель анализа паттерна диалога на этапе (206) обучена характеризовать степень достоверности утверждения, что прямой источник текстовых данных является мошенником или не мошенником. Такую оценку модель может проводить на основе выявления, совокупного анализа, сопоставления по близости к устойчивым семантическим конструкциям речи, типичным репликам, паттернам общего смысла диалога. По итогу классификации модели на этапе (207) принимается решение об

отнесении входящего аудиовызова к мошеннической активности (210) или к безопасной (220).

[0041] Дополнительно при выполнении способа (200) анализ аудиопотока осуществляется помощью эмотивно-просодической модели (модель с анализом эмотивности и просодии), которая позволяет по меньшей мере характеризовать степень достоверности утверждения, что прямой источник аудиозаписи является мошенником или не мошенником на основе, как минимум одной из следующей характеристики: выделения общих имманентных свойств языка по выражению психологического (эмоционального) состояния и переживания человека при совершении им мошеннического звонка, выделению общих особенностей мошенников в произношении, например таких как высота, сила/интенсивность, длительность, придыхание, глоттализация, палатализация, тип примыкания согласного к гласному и других признаков, являющиеся дополнительными к основной артикуляции звука, акценте, интонации в общем и других особенностей речи, а также особенностей фонового сопровождения речи, элементов постороннего шума и подобного. Ключевой особенностью модели является то, что она позволяет выявлять и анализировать общие особенности аудиодорожек, в которых присутствуют элементы мошеннических действий, диалогов и прочей информации, свидетельствующей в той или иной степени о мошеннической активности.

[0042] Данная модель обучается на основе примеров аудиопотоков ранее отмеченных как мошеннические, по обратной информации от потерпевших в мошеннических схемах. Также возможно расширение базы данных через аугментацию данных или на основе самостоятельной генерации мошеннических диалогов. Такую генерацию можно провести через запись диалогов, в которых будут активно использоваться приемы и методы мошенников, выявленные по имеющимся данным или сформированные самостоятельно.

[0043] При классификации входящего аудиовызова может формироваться уведомление о статусе, отображаемое на экране смартфона (111). Также может применяться вибросигнал, передача информации на внешнее устройство, связанное со смартфоном, например, смарт-часы, и другие типы оповещения, позволяющие информировать пользователя (110) о статусе входящего звонка.

[0044] На Фиг. 2Б представлена блок-схема этапов дополнительной обработки аудиовызовов, при их преобразовании в векторную форму на этапе (202). Дополнительная обработка выполняется с помощью нескольких моделей машинного обучения на этапе (230), которые позволяют выявить те или иные изменения аудиопотока. На этапе (230) выполняется анализ аудиопотока на предмет изменения голоса (231), синтетического формирования голоса (232), наличия наложения фона (233), наличие посторонних шумов (234).

[0045] На этапах (231, 232) модель анализирует факт программного изменения голоса звонящего абонента (120), например, с помощью применения алгоритмов Deep Fake Voice, алгоритмы клонирования голоса и т.п. Модель осуществляет оценку соответствия входной аудиодорожки естественной записи голоса человека и его окружающего пространства или наличие в ней дополнительной электронной обработки, элементов искусственной генерации звуков, полного или частичного синтеза записи. Реализация данного выявления может основываться на выявлении синтетических особенностей и машинных артефактов при искусственной генерации речи человека. Примерами таких особенностей и артефактов могут быть неестественная монотонность в речи, скрипы в произношении, множество помех и прочее. Данная модель позволяет по меньшей мере характеризовать вероятность наличия намеренных искажений в естественной

записи или ее искусственной генерации. Одним из примеров реализации функционала модели может выступать анализ графического представления спектрограмм аудиозаписи или использование архитектур «трансформеров», например, на основании нейронных сетей. Данный пример реализации при этом не ограничивает другие частные формы

5 воплощения реализации функционала вышеуказанной модели машинного обучения.

[0046] На этапе (233) выполняется анализ факта наложения фона на входящий аудиопоток, например, для формирования звуковой активности офиса, колл-центра и т.п. Данный подход может применяться мошенниками для маскирования звуковой дорожки и сокрытия места реального осуществления вызова, что может быть также

10 установлено посредством посторонних шумов при звонке. Обученная модель на этапе (233) анализирует артефакты, присущие синтетическим звуковым сигналам, нехарактерным для реальной обстановки.

[0047] На этапе (234) выполняется анализ наличия посторонних шумов в аудиодорожке при входящем вызове, например, при синтезе речи, как правило,

15 наблюдается треск в записи, помехи и т.п. Модель, обеспечивая заданный функционал, также может осуществлять анализ с помощью сравнения спектрограмм или по иному принципу, позволяющему установить «нехарактерные» для обычного звонка аудиоданные.

[0048] Применяемая модель на этапе (230) позволяет сверхаддитивно (синергетически)

20 объединять и анализировать по меньше мере двух любых выходов с применяемых моделей. Отличительной особенностью является то, что подобная модель позволяет анализировать в совокупности выходные данные от предыдущих моделей и получать более достоверные оценки о наличии мошеннических элементов в аудиозаписи, чем при каком-либо использовании выходов с моделей самостоятельно или простым

25 обобщении, таком как расчет среднего, извлечение максимального и подобного. Данный эффект может быть достигнут за счет объединения нескольких выходов в общий числовой вектор (упорядоченную последовательность) и использовании в качестве классификатора нейронных сетей, получении характерных объектов каждого класса через метод опорных векторов или к-соседей, построение ансамблей или бустингов

30 деревьев решений.

[0049] Итогом отработки одной или нескольких моделей на этапе (230) является дополнительная классификация входящего аудиозвонка на предмет мошеннической активности (210) или отсутствия таковой (220).

[0050] На Фиг. 3 представлен частный случай выполнения способа (300) защиты

35 абонента (110) от мошеннических действий при входящих вызовах. При получении входящего вызова на этапе (301) с помощью устройства пользователя (110), например, смартфона (111), выполняется активация синтетического исходящего аудиопотока на этапе (302), который выполняет роль роботизированного собеседника (бота) со стороны пользователя (110). Специальное программное обеспечение активирует

40 заданный алгоритм ведения диалога при входящем аудиовызове. Это необходимо для того, что собирать данные и анализировать входящий звонок от абонента (120) на предмет мошеннической активности. Генерирование синтетической исходящей со стороны пользователя (110) аудиодорожки (аудиопотока) может выполняться на основании клонирования или синтеза по голосовому образцу пользователя

45 (110). Для этого также могут применяться различные известные решения по формированию аудиоданных из заданных образцов, например, AI Voice Generator или похожие решения.

[0051] На этапе (303) захватываемая с помощью бота аудиодорожка входящего

аудиовызова проходит этапы обработки вышеописанного способа (200). Программный бот может выполняться на базе технологий голосовых помощников с применением моделей машинного обучения для того, чтобы фиксировать входящие фразы и генерировать соответствующие ответные голосовые команды. На этапе (304) происходит

5 итоговая классификация входящего звонка и пользователю (110) формируется уведомление о статусе звонка, например, с помощью отображения на экране смартфона (111). Диалог ботом может вестись заданное количество времени, необходимое для классификации входящего звонка. Временной диапазон может варьироваться исходя из диалога абонента (120), а также при срабатывании одной или нескольких моделей

10 машинного обучения при выполнении способа классификации, приведенного на Фиг. 2А - 2Б, и вынесении точного суждения, в зависимости от установленного порогового значения классификации типа звонка.

[0052] Заявленный способ может также применяться для сбора векторного представления мошеннических голосовых дорожек, паттернов диалогов и иной

15 информации, которая накапливается и применяется для последующих тренировок моделей машинного обучения, а также формирования стоп-листов, идентифицирующих мошенников.

[0053] На Фиг. 4 представлен общий вид вычислительного устройства (400), пригодного для выполнения способов (200, 300). Устройство (400) может представлять

20 собой, например, сервер или иной тип вычислительного устройства, который может применяться для реализации заявленного технического решения, в том числе: смартфон, планшет, ноутбук, компьютер и т.п. Устройство (400) может также входить в состав облачной вычислительной платформы.

[0054] В общем случае вычислительное устройство (400) содержит объединенные

25 общей шиной информационного обмена один или несколько процессоров (401), средства памяти, такие как ОЗУ (402) и ПЗУ (403), интерфейсы ввода/вывода (404), устройства ввода/вывода (405), и устройство для сетевого взаимодействия (406).

[0055] Процессор (401) (или несколько процессоров, многоядерный процессор) могут

30 выбираться из ассортимента устройств, широко применяемых в текущее время, например, компаний Intel™, AMD™, Apple™, Samsung Exynos™, MediaTek™, Qualcomm Snapdragon™ и т.п. В качестве процессора (401) может также применяться графический процессор, например, Nvidia, AMD, Graphcore и пр.

[0056] ОЗУ (402) представляет собой оперативную память и предназначено для

35 хранения исполняемых процессором (401) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (402), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.).

[0057] ПЗУ (403) представляет собой одно или более устройств постоянного хранения

40 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.

[0058] Для организации работы компонентов устройства (400) и организации работы

45 внешних подключаемых устройств применяются различные виды интерфейсов В/В (404). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь: PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SAT A, 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.

[0059] Для обеспечения взаимодействия пользователя с вычислительным устройством (400) применяются различные средства (405) 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.

[0060] Средство сетевого взаимодействия (406) обеспечивает передачу данных устройством (400) посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (406) 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.

[0061] Дополнительно могут применяться также средства спутниковой навигации в составе устройства (400), например, GPS, ГЛОНАСС, BeiDou, Galileo.

[0062] Представленные материалы заявки раскрывают предпочтительные примеры реализации технического решения и не должны трактоваться как ограничивающие иные, частные примеры его воплощения, не выходящие за пределы испрашиваемой правовой охраны, которые являются очевидными для специалистов соответствующей области техники.

(57) Формула изобретения

1. Компьютерно-реализуемый способ анализа диалога во время аудиовызовов на предмет выявления мошеннической активности, выполняемый с помощью процессора и содержащий этапы, на которых:

- получают входящий аудиопоток, поступающий от вызывающей стороны;
- осуществляют преобразование входящего аудиопотока в векторную форму;
- осуществляют обработку преобразованного аудиопотока с помощью первой модели машинного обучения, в ходе которой выполняют сравнение векторной формы аудиопотока с ранее сохраненными векторами, характеризующими мошенническую активность;
- осуществляют транскрибирование аудиопотока и его последующую обработку с помощью второй модели машинного обучения, которая выполняет анализ диалога вызывающей стороны, при этом в ходе упомянутого анализа осуществляется семантический состав информации и паттерн ведения диалога, при этом паттерн ведения диалога включает в себя анализ слов, используемых в разговоре, анализ построения фраз, анализ следования фраз друг за другом;
- осуществляют классификацию входящего аудиопотока на основании выполненной обработки первой и второй моделями машинного обучения.

2. Способ по п. 1, характеризующийся тем, что при семантическом анализе транскрибированного диалога выполняется выявление слов, присущих мошеннической активности.

3. Способ по п. 1, характеризующийся тем, что дополнительно входящий аудиопоток анализируется на по меньшей мере одно из: тональность, эмотивность, просодия или их сочетания.

4. Способ по п. 1, характеризующийся тем, что векторная форма входящего аудиопотока анализируется на предмет наличия признаков, выбираемых из группы:

изменение голоса, синтетическое формирование голоса, наложение фонового аудиопотока или их сочетания.

5. Способ по п. 1, характеризующийся тем, что дополнительно анализируют исходящий аудиопоток.

5 6. Способ по п. 5, характеризующийся тем, что выполняют разделение исходящего и входящего аудиопотоков.

7. Способ по п. 1, характеризующийся тем, что дополнительно анализируется по меньшей мере один параметр входящего аудиопотока, выбираемый из группы: высота тембра, сила звука, интенсивность речи, длительность произнесения слов, придыхание, 10 глоттализация, палатализация, тип примыкания согласного к гласному или их сочетания.

8. Способ по п. 1, характеризующийся тем, что дополнительно анализируется наличие посторонних шумов во входящем аудиопотоке.

9. Способ по п. 1, характеризующийся тем, что выполняется на устройстве пользователя, представляющем собой смартфон, планшет или компьютер.

15 10. Способ по п. 9, характеризующийся тем, что при получении входящей аудиодорожки выполняется генерирование синтетического исходящего голосового аудиопотока.

11. Способ по п. 10, характеризующийся тем, что генерирование исходящего аудиопотока выполняется до момента классификации входной аудиодорожки.

20 12. Способ по п. 10, характеризующийся тем, что генерирование синтетического аудиопотока осуществляется на основании голосового образца пользователя устройства.

13. Способ по п. 1, характеризующийся тем, что при классификации входящего аудиопотока как мошеннического выполняется сохранение его векторного представления.

25 14. Способ по п. 12, характеризующийся тем, что при классификации входящего аудиопотока как мошеннического выполняется генерирование сообщения о статусе, отображаемое на дисплее устройства.

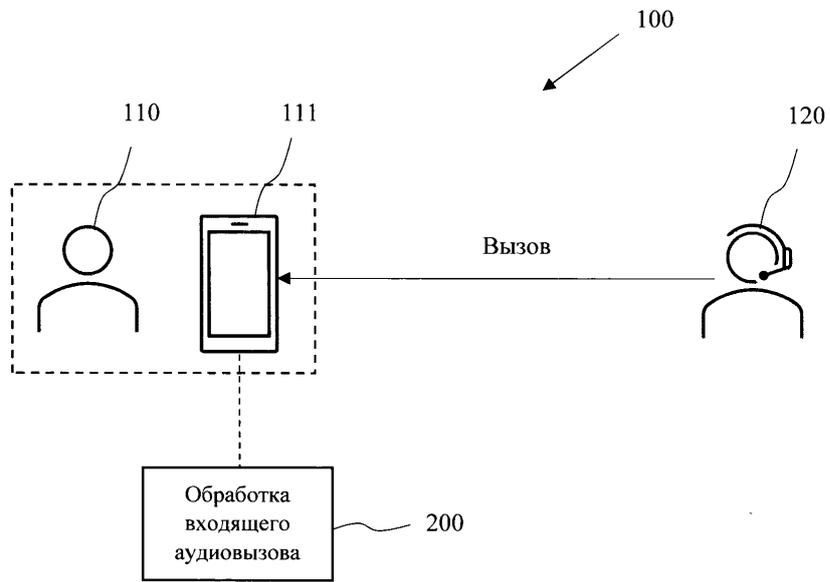
15. Система анализа диалога во время аудиовызовов на предмет выявления мошеннической активности, содержащая по меньшей мере один процессор и по меньшей 30 мере одну память, хранящую машиночитаемые инструкции, которые при их выполнении процессором реализуют способ по любому из пп. 1-14.

35

40

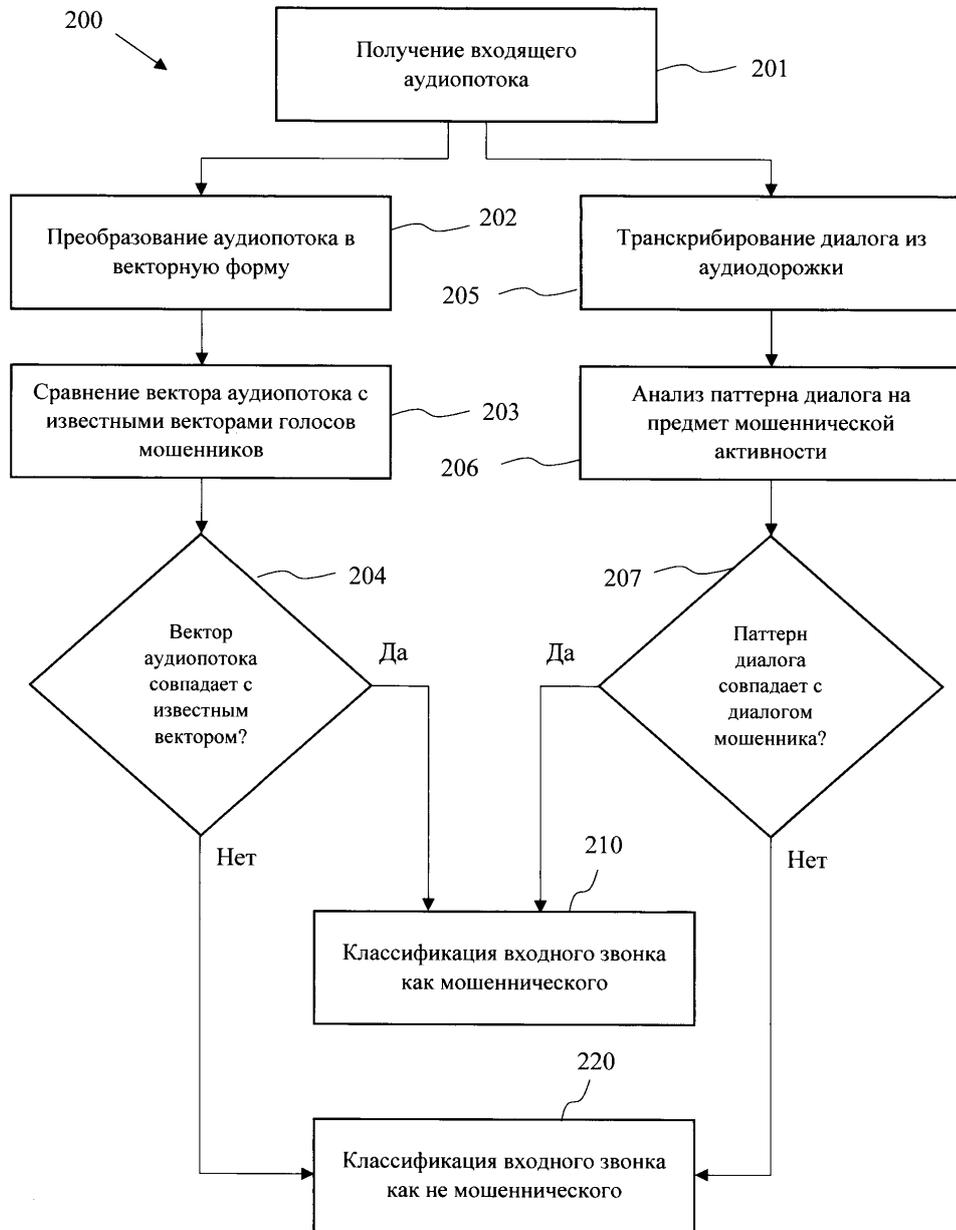
45

1

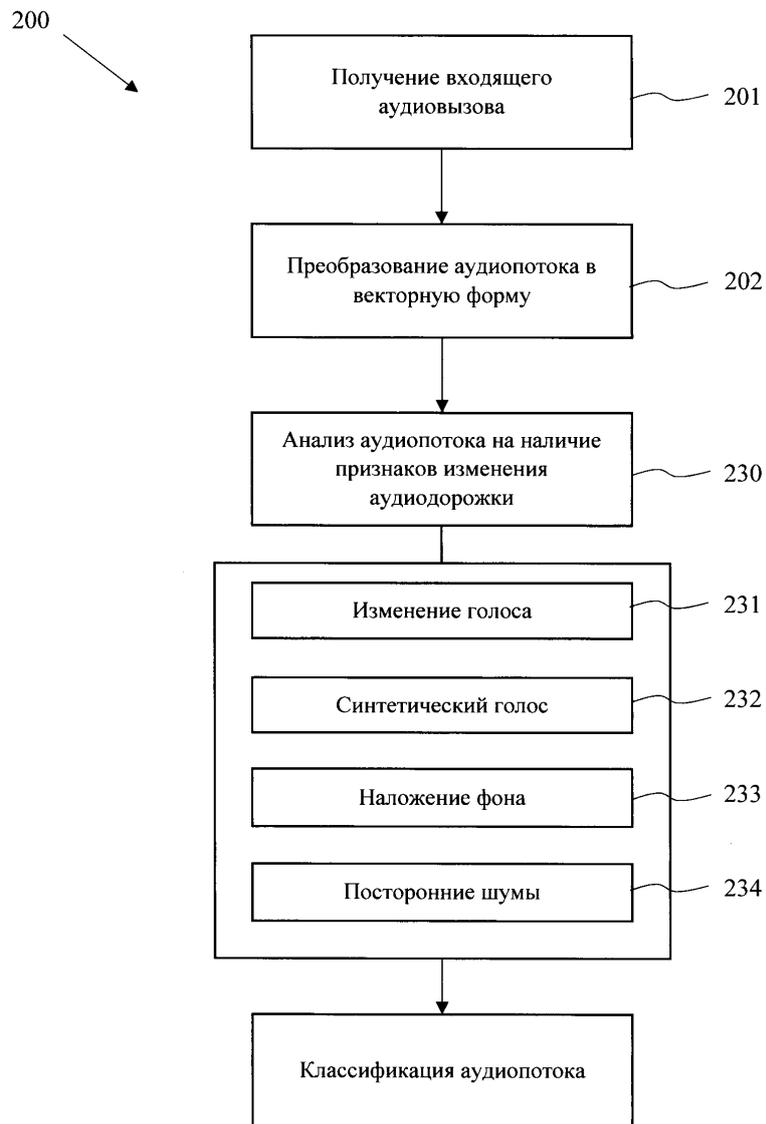


Фиг. 1

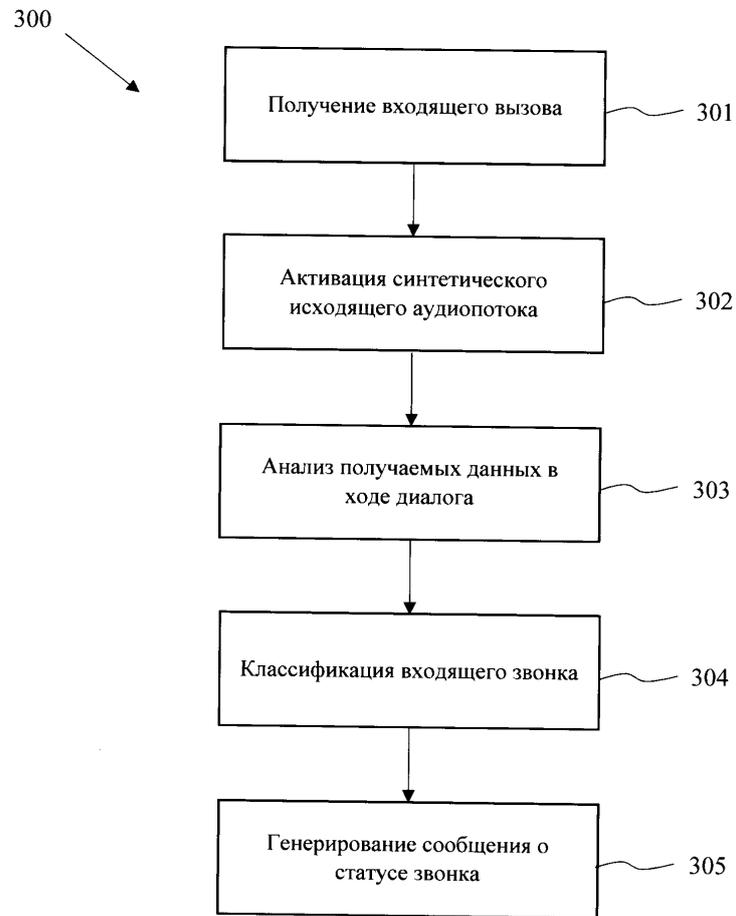
2



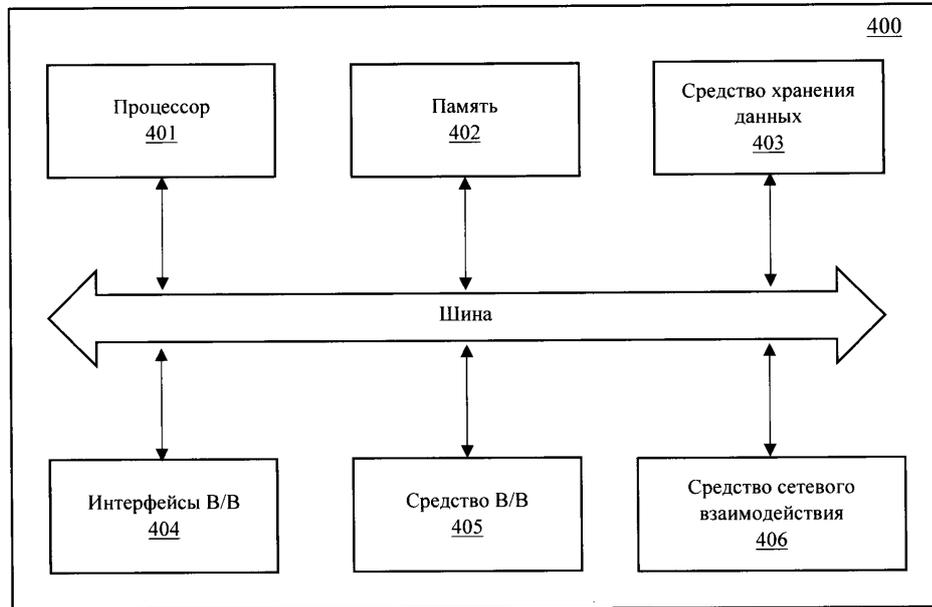
Фиг. 2А



Фиг. 2Б



Фиг. 3



Фиг. 4