(19) Евразийское патентное ведомство

(11) 046585 (13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2024.03.27

(21) Номер заявки
202091595

(22) Дата подачи заявки
2017.12.27

(51) Int. Cl. ***G10L 25/30*** (2013.01)
G10L 25/75 (2013.01)

(54) СПОСОБ И УСТРОЙСТВО ДЛЯ ПОСТРОЕНИЯ ГОЛОСОВОЙ МОДЕЛИ ЦЕЛЕВОГО ДИКТОРА

(43) **2020.09.18**
(86) **PCT/RU2017/000990**
(87) **WO 2019/132690 2019.07.04**

(71)(73) Заявитель и патентовладелец:
ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "ЦЕНТР РЕЧЕВЫХ ТЕХНОЛОГИЙ" (RU)

(72) Изобретатель:
Новоселов Сергей Александрович, Козлов Александр Викторович, Румянцев Дмитрий Александрович, Кудашев Олег Юрьевич (RU)

(74) Представитель:
Нилова М.И. (RU)

(56) US-B2-9043207
US-A-5659662
RU-C1-2530314

(57) Изобретение относится к области голосовой биометрии, в частности к задаче автоматической оценки голосовых моделей дикторов по записям их телефонных переговоров с автоматической привязкой голосовой модели диктора к номеру телефона. Способ получения голосовой модели целевого диктора, согласно которому осуществляют сегментацию по голосам дикторов по меньшей мере двух фонограмм телефонных переговоров с получением сегментов речи; строят голосовые модели дикторов по полученным сегментам речи; осуществляют кластеризацию построенных голосовых моделей дикторов с использованием метаданных телефонных переговоров с получением кластеров; определяют связи между кластерами на основании фонограмм телефонных переговоров и выделяют кластер с наибольшим количеством связей как кластер целевого диктора. Также предложено устройство для получения голосовой модели целевого диктора.

046585 B1

046585 B1

Область техники

Изобретение относится к области голосовой биометрии, в частности к задаче автоматической оценки голосовых моделей дикторов по записям их телефонных переговоров с автоматической привязкой голосовой модели диктора к номеру телефона.

Уровень техники

Известен способ кластеризация дикторов "без учителя" для автоматической идентификации диктора с записанных аудиоданных (US 5659662), осуществляющий следующие этапы: предоставление фрагмента аудиоданных, содержащих речь всех дикторов; формирование начальных кластеров путем деления фрагмента аудиоданных на сегменты, каждый из которых включает в себя упорядоченный набор данных; вычисление попарного расстояния между каждой парой кластеров с использованием коэффициента правдоподобия, независимого от набора данных в сегментах; и объединение в новый кластер двух кластеров с минимальным попарным расстоянием. Эти шаги повторяются до тех пор, пока не будет получено количество кластеров, равное количеству дикторов.

Недостатком известного способа является то, что при кластеризации используется только один фрагмент аудиоданных с голосами дикторов. Такое решение позволяет получить сегменты речи диктора только в одних условиях, что исключает разнообразие получаемых сегментов и уменьшает информативность получаемого кластера.

Известен способ анализа набора вызовов одного или нескольких центров приёма звонков (US 20030154072), осуществляющий прием лексического содержимого телефонного звонка, обрабатываемого оператором центра приёма звонков, и идентификацию одного или нескольких признаков (характеристик) телефонного звонка на основе полученного лексического содержимого. Этот способ также осуществляет совместный анализ сохраненных характеристик наряду с сохраненными характеристиками других телефонных звонков и предоставляет отчет по результатам анализа. Также возможен прием такой информации, как: продолжительность вызова, время вызова, идентификатор вызывающего абонента и идентификатор оператора. Кроме этого, анализ может включать в себя представление по меньшей мере некоторых вызовов в модели векторного пространства. Анализ может дополнительно включать определение кластеров вызовов в модели векторного пространства, например с использованием кластеризации k-средних. Анализ может также включать отслеживание кластеров вызовов с течением времени (например, определение новых кластеров и/или определение изменений в кластере). Анализ может дополнительно включать использование модели векторного пространства для идентификации вызовов, похожих на вызов с указанными свойствами, например для идентификации вызовов, аналогичных указанному вызову. В других конфигурациях, в частности обеспечивающих на выходе только объединенные голоса оператора и клиента, аппаратное и/или программное обеспечение может отделять голоса операторов и клиента.

Недостатком данного способа является то, что кластеризацию и идентификацию вызовов осуществляют на основе проанализированного лексического содержимого телефонных звонков. Другими словами, идентификацию оператора и клиентов осуществляют по записанным ранее и проанализированным на основе частоты употребления определенным словам или репликам оператора и клиентов, что предполагает высокую вероятность ошибок. Таким образом, известный способ не обеспечивает высокого качества кластеризации и идентификации.

Наиболее близким аналогом предложенного изобретения является способ распознавания дикторов по телефону (US 20120232900), содержащий этапы получения и хранения голосовой модели по меньшей мере одного целевого диктора, а также этапы получения голосовой модели по меньшей мере одного неизвестного диктора и сравнения ее с сохраненной голосовой моделью указанного по меньшей мере одного целевого диктора для определения того, идентичен ли неизвестный диктор сохраненному целевому диктору. В данном способе голосовая модель целевого диктора может быть получена путем обработки множества речевых выборок целевого диктора. Речевые выборки могут быть получены по меньшей мере из одного телефонного звонка целевого диктора. В частности, голосовая модель целевого диктора может быть получена путем захвата микрофоном или телефоном изолированных слов или непрерывной речи, в последующем преобразуемой в аналоговые сигналы, которые затем оцифровываются. Также в данном способе предусмотрена возможность использования информации о телефонных номерах.

Недостатком данного способа является то, что прежде чем получить голосовую модель целевого диктора, используя множество телефонных звонков, необходимо обеспечить получение речевых выборок целевого диктора из каждого телефонного звонка в отдельности путем захвата микрофоном или телефоном изолированных слов или непрерывной речи, что усложняет и замедляет процесс получения голосовой модели целевого диктора с использованием множества телефонных звонков.

Таким образом, в известных способах кластеризации дикторов недостаточно проработана возможность использования множества фрагментов аудиоданных или фонограмм, поученных в разных условиях записи, для кластеризации и последующего выделения модели голоса целевого диктора, оператора, клиента. Общим недостатком известных способов является то, что они не обеспечивают создание голосовой модели, отвечающей высокому уровню качества в полностью автоматическом режиме.

Ввиду имеющихся недостатков известных способов распознания и идентификации дикторов по голосу технической проблемой настоящего изобретения является создание способа получения голосовой

модели целевого диктора в массивах неразмеченных данных в полностью автоматическом режиме для дальнейшей эффективной идентификации личности.

Раскрытие сущности изобретения

Поставленная задача решается благодаря тому, что, согласно предлагаемому способу получения голосовой модели целевого диктора, осуществляют сегментацию по голосам дикторов по меньшей мере двух фонограмм телефонных переговоров с получением сегментов речи, строят голосовые модели дикторов по полученным сегментам речи. Далее осуществляют кластеризацию построенных голосовых моделей дикторов с использованием метаданных телефонных переговоров с получением кластеров. Затем определяют связи между кластерами на основании фонограмм телефонных переговоров и выделяют кластер с наибольшим количеством связей как кластер целевого диктора.

Предлагаемый способ позволяет достичь технического результата в виде автоматизации построения голосовой модели целевого диктора, увеличения скорости построения голосовой модели диктора и точности кластеризации, а также в виде увеличения информативности получаемой голосовой модели целевого диктора.

В предлагаемом способе получения голосовой модели диктора осуществляют сегментацию по голосам дикторов по меньшей мере двух фонограмм телефонных переговоров, причем фонограммы телефонных переговоров предпочтительно набираются в разных условиях записи целевого диктора с несколькими неизвестными дикторами. Голосовая модель целевого диктора, построенная с использованием нескольких фонограмм (по меньшей мере двух), набранных таким образом, отличается высокой информативностью, т.к. для ее построения использовано большое количество обучающих данных. Полученные голосовые модели дикторов с использованием одной фонограммы обладают относительно низкой информативностью. Также осуществление сегментации фонограмм по голосу дикторов позволяет выделять сегменты речи дикторов в неразмеченных массивах фонограмм телефонных переговоров в полностью автоматическом режиме.

Кроме этого, использование в способе по меньшей мере двух фонограмм телефонных переговоров предоставляет возможность автоматического построения голосовой модели целевого диктора путем определения связей между кластерами на основании фонограмм телефонных переговоров и последующего выделения кластера с наибольшим количеством связей как кластера целевого диктора. Причем получение кластеров осуществляют кластеризацией голосовых моделей дикторов, построенных по полученным сегментам речи в результате сегментации.

Согласно частному случаю реализации, голосовые модели представлены в виде векторов пространства полной изменчивости. Голосовые модели, построенные в виде векторов пространства полной изменчивости, обладают признаками, позволяющими в дальнейшем производить их качественное сравнение при проведении кластеризации, т.е. позволяющими с высокой точностью оценивать вероятность совпадения голосов на сравниваемых звуковых фрагментах голосовой модели Также такие голосовые модели являются компактным математическим представлением голоса диктора в виде набора чисел.

Согласно частному случаю реализации, голосовые модели могут быть построены с использованием высокоуровневых признаков нейронных сетей. Наиболее подходящими для кластеризации являются конволюционные или рекурентные нейронные сети. Построение голосовых моделей с помощью высокоуровневых признаков нейронных сетей позволяет улучшить результаты кластеризации благодаря возможности моделирования нелинейных зависимостей скрытых факторов сложных пространств голосовых моделей.

Согласно частному случаю реализации, построение голосовых моделей осуществляют с учетом длительности речевых сегментов. Длительность речевых сегментов влияет на качество голосовых моделей, получаемых при использовании данных сегментов. Более длинный речевой сегмент позволяет статистически более точно оценить особенности голоса диктора в нем и представить этот голос в виде качественной голосовой модели. При использовании короткого речевого сегмента может быть получена недостоверная оценка голосовой модели. Следует отметить, что различная длительность речевых сегментов при построении голосовых моделей приводит к смещению оценок сравнения голосов дикторов. Такое смещение возможно компенсировать при учете длительностей речи сегментов.

Согласно частному случаю реализации, две голосовые модели, полученные из одной фонограммы, распределяют в разные кластеры. Данное решение обеспечивает сокращение ошибок при построении кластеров по голосовым моделям дикторов, что повышает точность кластеризации.

Согласно частному случаю реализации, кластеризацию осуществляют методом "Сдвиг среднего". Данный метод не требует априорного задания количества кластеров для кластеризации и использует преимущества голосовых моделей, состоящих из сегментов, полученных из разных фонограмм, что обеспечивает высокую точность кластеризации с получением кластеров, содержащих голосовые модели.

Согласно частному случаю реализации, кластеризацию осуществляют методом агломеративной иерархической кластеризации. Данный метод обеспечивает высокую точность кластеризации.

Согласно частному случаю реализации, кластеризацию осуществляют методом обучения модели смесей гауссовских распределений. Использование данного метода обеспечивает повышение скорости кластеризации.

Согласно частному случаю реализации, при кластеризации используют вероятностный линейный дискриминантный анализ. Использование вероятностного линейного дискриминантного анализа (PLDA) для голосовых моделей, представленных в виде векторов пространства полной изменчивости, обеспечивает высокую точность оценки вероятности сходства голосовых моделей дикторов и, как следствие, способствует повышению точности кластеризации.

Согласно частному случаю реализации, при кластеризации используют косинусную метрику схожести. Косинусная метрика схожести способна увеличить скорость распределения голосовых моделей дикторов по кластерам.

Согласно частному случаю реализации способа, при кластеризации используют метрику L-1 норма. Согласно еще одному из частных вариантов реализации, при кластеризации используют метрику L-2 норма. Использование предложенных метрик обеспечивает повышение скорости кластеризации.

Согласно частному случаю реализации, при выделении кластера целевого диктора учитывают размеры кластеров. Размер кластера определяется количеством содержащихся в нем голосовых моделей диктора. Кластер, содержащий наибольшее количество голосовых моделей, с высокой вероятностью является кластером целевого диктора. Учитывая это обстоятельство возможно увеличить точность выделения кластера целевого диктора.

Устройство, реализующее заявленный способ получения голосовой модели диктора, содержит средства сегментации фонограмм, выполненные с возможностью сегментации фонограмм телефонных переговоров по голосам дикторов с получением сегментов речи, средства построения голосовых моделей, выполненные с возможностью построения голосовых моделей дикторов по полученным сегментам речи, средства кластеризации, выполненные с возможностью кластеризации построенных голосовых моделей дикторов с использованием метаданных телефонных переговоров с получением кластеров, и средства анализа, выполненные с возможностью определения связей между кластерами на основании фонограмм телефонных переговоров и с возможностью выделения кластера с наибольшим количеством связей как кластера целевого диктора. Разработанное устройство отвечает всем требованиям надежности, является технологичным и простым в реализации.

В одном из вариантов осуществления средства кластеризации выполнены с возможностью выполнения указанной кластеризации таким образом, что обеспечено распределение голосовых моделей, полученных из одной фонограммы, в разные кластеры.

В различных вариантах осуществления устройство содержит средства, обеспечивающие реализацию любой комбинации частных случаев реализации заявленного способа.

Краткое описание чертежей

Сущность изобретения более подробно поясняется на неограничительных примерах его осуществления со ссылкой на прилагаемые чертежи, среди которых:

фиг. 1 - функциональная схема устройства для получения голосовой модели целевого диктора;

фиг. 2 - схема построения голосовых моделей дикторов по полученным сегментам речи из фонограмм телефонных переговоров, согласно одному из вариантов осуществления изобретения;

фиг. 3 - схема построения одного кластера, согласно одному из вариантов осуществления изобретения;

фиг. 4 - схема построения связей между кластерами, согласно одному из вариантов осуществления изобретения.

Подробное описание

В настоящем описании под термином "целевой диктор" понимается диктор, для которого необходимо построить голосовую модель, содержащую в себе сегменты речи, полученные из разных телефонных разговоров с другими (неизвестными) дикторами.

Способ получения голосовой модели целевого диктора в соответствии с настоящим изобретением может быть осуществлен посредством устройства для получения голосовой модели целевого диктора, реализованного, например, с использованием известных компьютерных или мультипроцессорных систем. В других вариантах реализации заявленный способ может быть реализован посредством специализированных программно-аппаратных средств.

На фиг. 1 представлена функциональная схема устройства в соответствии с одним из вариантов осуществления настоящего изобретения. Устройство содержит блок 1 сегментации фонограмм, который предназначен для получения сегментов голосов дикторов из массива неразмеченных фонограмм телефонных переговоров. Блок 1 связан с блоком 2 построения голосовых моделей, а блок 2 связан с блоком 3 кластеризации, который в свою очередь связан с блоком 4 анализа, который предназначен для получения голосовой модели целевого диктора.

При получении голосовой модели целевого диктора предложенное устройство выполняет следующую последовательность операций.

Формирование базы фонограмм осуществляют с использованием информационной системы, позволяющей записывать телефонные переговоры между дикторами (собеседниками) и отправлять полученную информацию на сервер. В качестве такой информационной системы может быть использована любая подходящая информационная система, принципы функционирования которой известны специалисту,

а потому не рассматриваются подробно в настоящем описании. Сервер хранит эту информацию в базе фонограмм для последующего использования. Также сервер может хранить дополнительную информацию в отношении каждой фонограммы, такую как длительность переговоров и номера телефонов дикторов, голоса которых записаны на фонограмме. Фонограммы телефонных переговоров, на основании которых должна быть получена голосовая модель целевого диктора, извлекают, например с использованием сетевого протокола, из базы фонограмм (не показана), и направляют в блок 1 в виде массива неразмеченных данных. Выбор фонограмм телефонных переговоров для извлечения осуществляется, например с учетом номера телефона целевого диктора. При этом используемые фонограммы могут быть записаны как в стерео-, так и монорежиме, в разных условиях записи и с разными дикторами. Необходимо отметить, что для осуществления предложенного способа использование пяти фонограмм является предпочтительным, так как данное количество фонограмм обеспечивает информативность получаемой голосовой модели целевого диктора, достаточно высокую для большинства приложений. При этом использование пяти фонограмм, полученных в разных условиях записи, например в разное время суток, при разной активности целевого диктора, в разных погодных условиях, в разных психологических состояниях целевого диктора, увеличивает информативность голосовой модели за счет того, что она будет содержать сегменты голоса разных эмоциональных и физических состояний диктора. В других варрантах реализации возможно использование фонограмм, полученных в других условиях записи, известных специалисту в данной области техники.

При этом в других вариантах реализации, в которых требуется сравнительно меньшая информативность голосовой модели диктора, возможно использование меньшего количества фонограмм, например двух, трёх или четырёх фонограмм, с одновременным уменьшением времени получения голосовой модели и/или количества потребных ресурсов.

В некоторых вариантах реализации для ещё большего увеличения информативности возможно использование и более пяти фонограмм. Однако, так как в этом случае возможно увеличение времени получения голосовой модели и/или количества потребных ресурсов, увеличение количества используемых фонограмм предпочтительно в приложениях с повышенными требованиями к информативности голосовых моделей.

Блок 1 осуществляет сегментацию пяти фонограмм телефонных переговоров по голосам дикторов. Особенностью фонограмм телефонных переговоров является частая смена дикторов на фонограмме, что необходимо учитывать при сегментации.

Сегментацию в данном случае производят с использованием вероятностного линейного дискриминантного анализа, который учитывает указанные особенности. В других вариантах реализации возможно использование других методов сегментации, учитывающие особенности сегментации фонограмм телефонных переговоров, например сегментацию на основе вариационного байесовского анализа. На выходе блок 1 выдаёт сегменты речи дикторов, то есть массив данных, соответствующих фонограмме и имеющих разметку, соответствующую участкам фонограммы с записью речи конкретного диктора. Сегменты речи содержат метаданные, в частности содержат информацию о номерах телефонов дикторов.

Блок 2 принимает полученные сегменты речи, по которым осуществляет построение голосовых моделей дикторов. Голосовая модель диктора, построенная по многим сегментам его речи, набранных из разных условий записи голоса диктора, называется "мультисессионной". На фиг. 2 прямоугольниками обозначены голосовые модели дикторов №1, №2, №3, №4 и №5, построенные по сегментам речи каждой фонограммы телефонных переговоров. Построение голосовых моделей осуществляется с привязкой к ним номеров телефонов в зависимости от содержащихся в них сегментов речи. Важно отметить, что построение голосовых моделей осуществляется с учетом длительности речевых сегментов. Длительность речевых сегментов влияет на качество голосовых моделей, получаемых при использовании данных сегментов. Более длинный речевой сегмент позволяет статистически более точно оценить особенности голоса диктора в нем и представить этот голос в виде качественной голосовой модели. При использовании короткого речевого сегмента полученная голосовая модель может оказаться недостоверной. Следует отметить, что различная длительность речевых сегментов при построении голосовых моделей приводит к смещению оценок сравнения голосов дикторов. Таким образом, для повышения качества мультисессионной модели необходимо учитывать различия длительностей сегментов речи, по которым строят голосовые модели дикторов и использовать весовые коэффициенты, зависящие от длительностей сегментов речи (1).

$$w_{avg} = \frac{1}{\sum_i C(t_i^{dur})} \sum_{i=1}^{N} C(t_i^{dur}) w_i \qquad (1)$$

где w_{avg} - мультиссесионная модель, W_i - сегменты речи, $C(t_i^{dur})$ - весовой коэффициент, зависящий от длительности речи.

Предпочтительным вариантом построения голосовых моделей дикторов является представление их в виде векторов пространства полной изменчивости. Голосовые модели, построенные таким образом, являются компактным математическим представлением голоса диктора в виде набора чисел (2).

$$\vec{M}(s,h) = \vec{m} + \hat{T}\vec{w}(s,h) \qquad (2)$$

где $\vec{M}$ - супервектор гауссовской смеси распределений дикторской модели акустического пространства мелкепстральных признаков, $\vec{m}$ - дикторо- и каналонезависимый супервектор средних, $\hat{T}$ - прямоугольная матрица низкого ранга и $\vec{w}$ - случайный вектор, имеющий стандартное гауссово распределение N(0,Î). Компоненты вектора $\vec{w}$ - это факторы полной изменчивости, этот вектор часто называют i-вектором (вектор пространства полной изменчивости).

В блоке 3 осуществляется кластеризация построенных голосовых моделей. В результате кластеризации должно получиться определенное количество кластеров, равное количеству дикторов. Однако кластеризация проводится в условиях отсутствия априорной информации о количестве дикторов. Метод "Сдвиг среднего" не требует априорного задания количества кластеров для кластеризации. С помощью метода "Сдвиг среднего" и метрики вероятностного линейного дискриминантного анализа в блоке 3 осуществляется кластеризация по всем построенным в блоке 2 голосовым моделям дикторов. В качестве примера на фиг. 3 показан один кластер, содержащий голосовую модель диктора № 1.

Осуществление кластеризации производят с учетом метаданных. Каждая полученная голосовая модель содержит информацию о телефонном номере диктора, чьи сегменты речи находятся в данной голосовой модели. Таким образом, в одном из вариантов реализации обеспечивается возможность кластеризации голосовых моделей по кластерам с учетом телефонных номеров дикторов, т.е. голосовые модели с разными телефонными номерами распределяются в разные кластеры, что позволяет сформировать кластеры, состоящие из голосовых моделей, содержащих номер одного диктора. По такому же принципу распределяются в разные кластеры голосовые модели, полученные из одной фонограммы телефонных переговоров, т.е. они не могут попасть в один кластер. Таким образом, в результате кластеризации образуются кластеры, содержащие голосовые модели дикторов и номера их телефонов.

После проведения кластеризации в блоке 4 определяет связи между полученными кластерами. Определение связей происходит с использованием информации о телефонных номерах, содержащихся в каждом кластере, и на основании фонограмм телефонных переговоров, в частности метаданных. С помощью фонограмм телефонных переговоров можно определить, между какими телефонными номерами происходили разговоры дикторов. На фиг. 4 проиллюстрирована схема построения связей между кластерами, содержащими голосовые модели дикторов № 1, № 2, № 3, № 4 и № 5. На схеме прямыми линиями отображены переговорные межкластерные связи, а пунктирной линией показана ложная связь, получившаяся в результате ошибок кластеризации.

Кроме этого, блок 4 выделяет кластер с наибольшим количеством связей. Все фонограммы изначально содержали голосовую модель целевого диктора и, как следствие, целевой диктор общался с наибольшим количеством других дикторов. Таким образом, кластер с наибольшим количеством связей с другими кластерами содержит голосовую модель целевого диктора. На фиг. 4 кластер, содержащий голосовую модель диктора №1, имеет связи с кластерами, содержащими голосовые модели дикторов № 2, № 3, № 4 и № 5, то есть имеет наибольшее количество связей, а значит, именно он является кластером, содержащим голосовую модель целевого диктора.

Стоит отметить, что при выделении кластера целевого диктора учитывают размеры кластеров. Размер кластера определяется количеством содержащихся в нем голосовых моделей диктора. Кластер, содержащий наибольшее количество голосовых моделей, с высокой вероятностью является кластером целевого диктора. Учитывая это обстоятельство, возможно увеличить точность выделения кластера целевого диктора. В других вариантах реализации, в которых при определении связей между кластерами присутствует высокий процент ошибок, возможно выделение кластера целевого диктора учитывая только размеры кластеров.

При этом в других вариантах реализации, в которых требуется сравнительно меньшая точность выделения кластера целевого диктора, возможно выделение кластера целевого диктора без учета размеров кластеров, с одновременным уменьшением времени получения кластера, содержащего голосовую модель целевого диктора и/или количества потребных ресурсов.

Таким образом, предложенный способ позволяет добиться автоматизации и увеличения скорости построения голосовой модели целевого диктора за счет обработки по меньшей мере двух фонограмм телефонных переговоров, с использованием которых возможно определить связи между кластерами, полученными в результате кластеризации голосовых моделей, на основании которых возможно выделить кластер, содержащий голосовую модель целевого диктора. При использовании менее чем двух фонограмм телефонных переговоров целевого диктора определить в автоматическом режиме основной кластер будет невозможно. Кроме того, сегментация по меньшей мере двух фонограмм телефонных переговоров обеспечивает получение голосовой модели целевого диктора, обладающей высокой информативностью. Также предложенный способ позволяет увеличить точность кластеризации за счет использования при ее проведении метаданных телефонных переговоров.

В способе построения голосовой модели целевого диктора дополнительно возможно реализовать построение голосовых моделей с использованием высокоуровневых признаков нейронных сетей, напри-

мер с использованием конволюционных или рекупентных нейронных сетей. Также возможно осуществлять кластеризацию методом агломеративной иерархической кластеризации или методом обучения модели смесей гауссовских распределений. Кроме того, при кластеризации возможно использование косинусной метрики схожести или метрики L-1 норма или метрики L-2 норма. Также предложенный способ возможно реализовать, используя фонограммы телефонных переговоров для построения голосовой модели целевого диктора, не определенного заранее. В данном случае под термином "целевой диктор" будет пониматься диктор, разговаривающий с большинством других (неизвестных) дикторов.

Необходимо отметить, что возможными областями применения настоящего изобретения являются использование его для автоматического построении образцов голоса оператора центра приема звонков с целью дальнейшего автоматического выделения речи клиента на монозаписях диалогов, для автоматического построения образцов голоса клиента центра приёма звонков для идентификации в дальнейшем, для поиска повторяющихся голосов разных клиентов (в частности мошенников) в массивах записей разговоров государственных учреждений, андеррайтинговых агентств и отделов банков, бюро кредитных историй и т.д., для автоматического построения речевого образца объекта идентификации в системах безопасности, для кластеризации дикторов в массивах фонограмм для оценки их количества, тендерного, возрастного распределения и иных статистических данных и др.

Настоящее изобретение не ограничено конкретными вариантами реализации, раскрытыми в описании в иллюстративных целях, и охватывает все возможные модификации и альтернативы, входящие в объем настоящего изобретения, определенный формулой изобретения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Компьютерно-реализуемый способ получения голосовой модели целевого диктора, согласно которому

осуществляют сегментацию по голосам дикторов по меньшей мере двух фонограмм телефонных переговоров с получением сегментов речи;

строят голосовые модели дикторов по полученным сегментам речи;

осуществляют кластеризацию построенных голосовых моделей дикторов с использованием метаданных телефонных переговоров с получением кластеров;

определяют связи между кластерами на основании фонограмм телефонных переговоров и

выделяют кластер с наибольшим количеством связей как кластер целевого диктора.

2. Способ по п.1, согласно которому голосовые модели строят в виде векторов пространства полной изменчивости.

3. Способ по п.1, согласно которому голосовые модели строят с использованием высокоуровневых признаков нейронных сетей.

4. Способ по любому из пп.1-3, согласно которому строят голосовые модели с учетом длительности сегментов речи.

5. Способ по любому из пп.1-4, согласно которому при кластеризации голосовые модели, полученные из одной фонограммы, распределяют в разные кластеры.

6. Способ по любому из пп.1-5, согласно которому кластеризацию осуществляют методом "Сдвиг среднего".

7. Способ по любому из пп.1-5, согласно которому кластеризацию осуществляют методом агломеративной иерархической кластеризации.

8. Способ по любому из пп.1-5, согласно которому кластеризацию осуществляют методом обучения модели смесей гауссовских распределений.

9. Способ по любому из пп.1-6, согласно которому при кластеризации используют метрику вероятностного линейного дискриминантного анализа.

10. Способ по любому из пп.1-6, согласно которому при кластеризации используют косинусную метрику схожести.

11. Способ по любому из пп.1-6, согласно которому при кластеризации используют метрику L-1 норма.

12. Способ по любому из пп.1-6, согласно которому при кластеризации используют метрику L-2 норма.

13. Способ по любому из пп.1-12, согласно которому при выделении кластера целевого диктора учитывают размеры кластеров.

14. Устройство для получения голосовой модели целевого диктора, содержащее

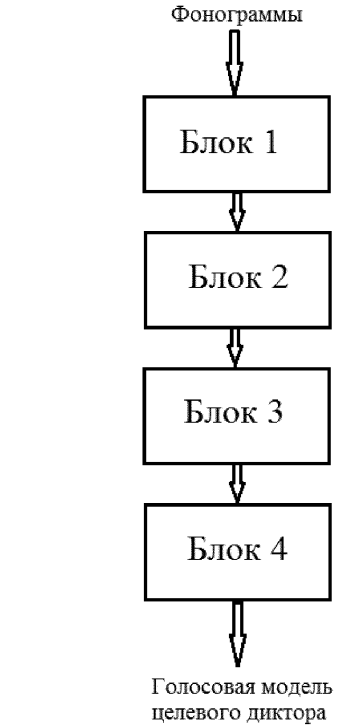
средства сегментации фонограмм, выполненные с возможностью сегментации фонограмм телефонных переговоров по голосам дикторов с получением сегментов речи;

средства построения голосовых моделей, выполненные с возможностью построения голосовых моделей дикторов по полученным сегментам речи;

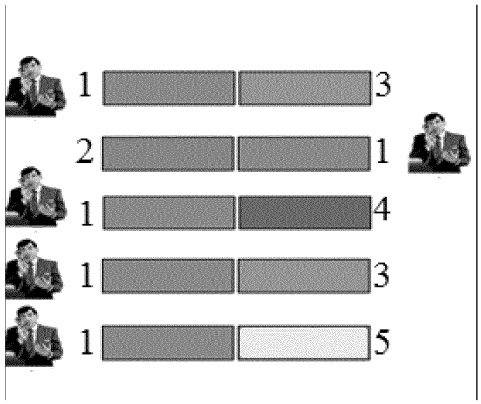
средства кластеризации, выполненные с возможностью кластеризации построенных голосовых моделей дикторов с использованием метаданных телефонных переговоров с получением кластеров;

средства анализа, выполненные с возможностью определения связей между кластерами на основании фонограмм телефонных переговоров и с возможностью выделения кластера с наибольшим количеством связей как кластера целевого диктора.

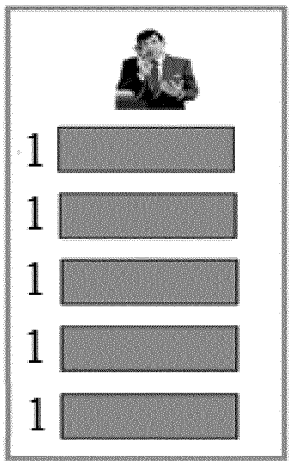
15. Устройство по п.14, в котором средства кластеризации, выполненные с возможностью выполнения указанной кластеризации таким образом, что обеспечено распределение голосовых моделей, полученных из одной фонограммы, в разные кластеры.



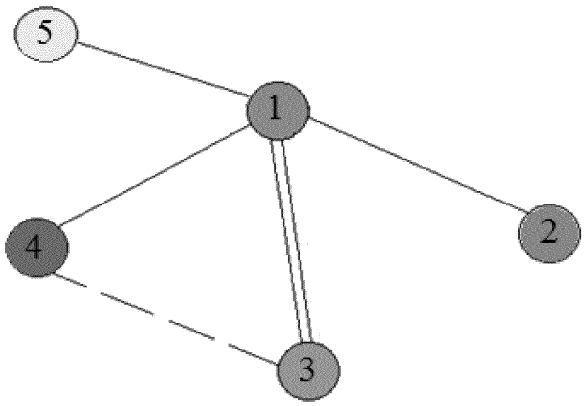
Фиг. 1



Фиг. 2



Фиг. 3



Фиг. 4