

ОПИСАНИЕ ПОЛЕЗНОЙ МОДЕЛИ К ПАТЕНТУ

(12)

РЕСПУБЛИКА БЕЛАРУСЬ



НАЦИОНАЛЬНЫЙ ЦЕНТР
ИНТЕЛЛЕКТУАЛЬНОЙ
СОБСТВЕННОСТИ

(19) ВУ (11) 6229

(13) U

(46) 2010.06.30

(51) МПК (2009)

G 10L 15/00

(54) УСТРОЙСТВО АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ДИКТОРА

(21) Номер заявки: u 20090860

(22) 2009.10.20

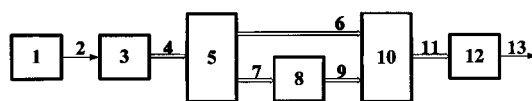
(71) Заявитель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(72) Авторы: Зельманский Олег Борисович; Давыдов Андрей Геннадьевич; Давыдов Геннадий Владимирович; Лыньков Леонид Михайлович (ВУ)

(73) Патентообладатель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(57)

Устройство для автоматического распознавания диктора, включающее источник речевого сигнала, блок определения параметрического описания речевого сигнала в виде выделителя начала/конца речевого сигнала, сегментатора речевого сигнала на последовательность сегментов, блока умножения на взвешивающее окно, блока добавления к сигналу в сегменте нулей, вычислителя преобразования Фурье и формирователя параметрических описаний входного речевого сигнала, соединенных последовательно, коммутатор, блок сравнения параметрических описаний эталона и входного речевого сигнала, блок принятия решения о распознаваемом дикторе и запоминающее устройство, при этом источник речевого сигнала подключен к блоку определения параметрического описания речевого сигнала, выход которого соединен с входом коммутатора, первый выход которого подключен к первому входу блока сравнения параметрических описаний эталона и входного речевого сигнала, а второй выход соединен с входом запоминающего устройства, выход которого подключен к второму входу блока сравнения параметрических описаний эталона и входного речевого сигнала, выход которого соединен с входом блока принятия решения о распознаваемом дикторе, выход которого является выходом устройства в целом, отличающееся тем, что блок определения параметрического описания речевого сигнала дополнительно содержит блок вычисления спектральной оценки сигнала, вход которого соединен с выходом вычислителя преобразования Фурье, а выход - с первым входом формирователя параметрических описаний входного речевого сигнала, блок вычисления среднеквадратического значения сигнала, блок вычисления среднего числа переходов сигнала через нулевой уровень, блок выделения частоты основного тона сигнала, при этом их входы соединены с выходом блока умножения на взвешивающее окно, а выходы подключены к остальным входам формирователя параметрических описаний



Фиг. 1

ВУ 6229 U 2010.06.30

входного речевого сигнала, блок сравнения параметрических описаний эталона и входного речевого сигнала выполнен в виде блоков сравнения значений спектральной оценки сигнала, среднеквадратического значения сигнала, среднего числа переходов сигнала через нулевой уровень, частоты основного тона сигнала с эталонными значениями, входы которых соединены с соответствующими входами блока сравнения параметрических описаний эталона и входного речевого сигнала, а выходы соединены с входом блока принятия решения о распознаваемом дикторе.

(56)

1. Патент РФ 2 161 826 С2, МПК G 10L 17/00, G 10L 11/04, G 10L 11/06, G 10L 15/06, G 10L 15/16, 2001.

2. Патент США 7.065.487 В2, МПК G 10L 19/10, 2006.

3. Патент РФ 2 230 375 С2, МПК G 10L 15/00, G 10L 17/00, 2004.

Полезная модель относится к области техники анализа речи, в частности к устройствам автоматического распознавания диктора, и может быть использована для контроля и ограничения несанкционированного доступа к материальным и информационным ресурсам, а также для обеспечения удаленного распознавания пользователей на основе биометрических параметров в телекоммуникационных системах.

Известен способ автоматической идентификации личности по особенностям произношения парольной фразы этой личностью [1], заключающийся в том, что речевой сигнал разбивают на вокализованные зоны, выделяют временные интервалы в вокализованных зонах, для этого определяют параметры речевого сигнала, сравнивают их с эталонами, которые формируют с учетом математических ожиданий и допустимых разбросов этих параметров, для чего в конце первой, начале последней, в начале и конце остальных вокализованных зон выделяют временные интервалы, длительность временных интервалов устанавливают кратной периоду основного тона речевого сигнала, определяют оценки коэффициентов корреляции параметров речевого сигнала, которые включают в число сравниваемых с эталонами, при формировании эталонов дополнительно учитывают коэффициенты корреляции параметров речевого сигнала. Кроме того, дополнительно вводят и обрабатывают речевой сигнал с ларингофона, контактирующего с телом говорящего. На основании отклика нейронной сети, на входы которой одновременно предъявляют параметры речевого сигнала и оценки коэффициентов корреляции этих параметров, принимают решение по идентификации личности.

Недостатками известного способа идентификации личности [1] являются низкая помехоустойчивость метода, так как для его работы требуется выделение во входном речевом сигнале точного положения границ основного тона голоса, что в условиях наличия акустических помех (шум большого офисного помещения, улицы и т.п.) практически невозможно, использование дополнительного оборудования, а именно ларингофона, и обработка дополнительной информации, получаемой от него, использование нейронной сети для принятия решения, что усложняет алгоритм и требует значительных затрат при его реализации.

Известны метод и устройство распознавания речи [2], использующие множественные акустические модели, и аппарат, реализующий этот метод и содержащий источник речевых сигналов (микрофон и процессор предварительной обработки), блок первичного выделения параметров речи, блок выделения сегментов шума и сегментов речи, блок определения типа шума и выбора акустической модели, устройство хранения данных о параметрах, блок устранения шума, блок вторичного анализа параметров речи, блок уст-

ранения шума по методу нормализации усреднением кепстра, блок хранения акустических моделей, блок хранения моделей языка и блок распознавания речи.

Недостатками предложенных в [2] метода, программы и аппарата распознавания речи являются многоступенчатая процедура анализа входных воздействий, сходимость которой к конкретному результату в патенте не проанализирована, и отсутствие конкретных сведений о надежности распознавания речи. В патенте нет сведений о возможности применения предложенных в нем средств для распознавания (или идентификации) дикторов.

Наиболее близкими к предлагаемому устройству являются способ и устройство распознавания диктора [3], совпадающее с заявляемым устройством по наибольшему числу существенных признаков и принятое за прототип. В основу заявленного в [3] способа распознавания диктора положено сравнение входного речевого сигнала неизвестного диктора с заранее сохраненными эталонами, представляющими собой речевой сигнал голосовых паролей, произносимых заранее известными дикторами. Известное устройство-прототип включает в себя источник речевого сигнала, блок определения параметрического описания речевого сигнала в виде выделителя начала/конца речевого сигнала, сегментатора речевого сигнала на последовательность сегментов, блока умножения на взвешивающее окно, блока добавления к сигналу в сегменте нулей, вычислителя преобразования Фурье, вычислителя спектра мощности сигнала в сегменте, блока определения формантного вектора текущего сегмента и формирователя параметрических описаний входного речевого сигнала, соединенных последовательно, коммутатор, блок сравнения параметрических описаний эталона и входного речевого сигнала, блок принятия решения о распознаваемом дикторе и запоминающее устройство.

Недостатком такого устройства является то, что оно все еще не обеспечивает запросы практики по надежности распознавания звуков речи и дикторов.

Задачей данной полезной модели является при сохранении в ее составе основных функциональных узлов и достоинств устройства-прототипа увеличение помехоустойчивости, уменьшение вероятности ошибок распознавания звуков речи и повышение на этой основе надежности распознавания речевых сигналов и дикторов, включая возможность их идентификации.

Поставленная задача решается тем, что в устройстве для автоматического распознавания диктора, включающем источник речевого сигнала, блок определения параметрического описания речевого сигнала в виде выделителя начала/конца речевого сигнала, сегментатора речевого сигнала на последовательность сегментов, блока умножения на взвешивающее окно, блока добавления к сигналу в сегменте нулей, вычислителя преобразования Фурье и формирователя параметрических описаний входного речевого сигнала, соединенных последовательно, коммутатор, блок сравнения параметрических описаний эталона и входного речевого сигнала, блок принятия решения о распознаваемом дикторе и запоминающее устройство, при этом источник речевого сигнала подключен к блоку определения параметрического описания речевого сигнала, выход которого соединен с входом коммутатора, первый выход которого подключен к первому входу блока сравнения параметрических описаний эталона и входного речевого сигнала, а второй выход соединен с входом запоминающего устройства, выход которого подключен ко второму входу блока сравнения параметрических описаний эталона и входного речевого сигнала, выход которого соединен с входом блока принятия решения о распознаваемом дикторе, выход которого является выходом устройства в целом, блок определения параметрического описания речевого сигнала дополнительно содержит блок вычисления спектральной оценки сигнала, включенный между вычислителем преобразования Фурье и формирователем параметрических описаний входного речевого сигнала, блоки вычисления среднеквадратического значения сигнала, среднего числа переходов сигнала через нулевой уровень, частоты основного тона сигнала, включенные параллельно друг другу между блоком умножения на взвешивающее окно и формирователем параметрических описаний входного речевого

ВУ 6229 U 2010.06.30

сигнала, блок сравнения параметрических описаний эталона и входного речевого сигнала выполнен в виде блоков сравнения значений спектральной оценки сигнала, среднеквадратического значения сигнала, среднего числа переходов сигнала через нулевой уровень, частоты основного тона сигнала с эталонными значениями, входы которых подключены к соответствующим входам блока сравнения параметрических описаний эталона и входного речевого сигнала, а выходы соединены с входом блока принятия решения о распознаваемом дикторе.

Сущностью устройства является то, что в нем рассчитываются классификационные параметры сигнала как во временной области, а именно кратковременное среднеквадратичное значение сигнала, значение функции среднего числа переходов через нуль и частота основного тона сигнала, так и параметр в частотной области, а именно изменение спектра сигнала.

На фиг. 1 представлена функциональная схема устройства автоматического распознавания диктора.

На фиг. 2 приведена функциональная схема входящего в состав устройства автоматического распознавания диктора блока определения параметрического описания речевого сигнала.

В соответствии с фиг. 1 устройство автоматического распознавания диктора содержит источник речевого сигнала в цифровой форме 1, например микрофон и аналого-цифровой преобразователь, с выходом 2 на блок определения параметрического описания речевого сигнала 3 с выходом 4 на коммутатор 5, блок запоминающего устройства для хранения эталонов параметрического описания речевого сигнала заранее известных дикторов 8 с входом 7, блок сравнения параметрических описаний эталона и входного речевого сигнала 10 с входами 6 и 9 и выходом 11, блок принятия решения о распознаваемом дикторе 12, выход 13 которого является выходом системы в целом.

Для увеличения достоверности распознавания диктора в блок определения параметрического описания речевого сигнала 3 устройства автоматического распознавания диктора (фиг. 2), содержащий выделитель начала/конца речевого сигнала 14, сегментатор речевого сигнала на последовательность сегментов 15, блок умножения на взвешивающее окно 16, блок добавления к сигналу в сегменте нулей 17, вычислитель преобразования Фурье 19 и формирователь параметрических описаний входного речевого сигнала 23, введены блок вычисления спектральной оценки сигнала 21, блок вычисления среднеквадратического значения сигнала 18, блок вычисления среднего числа переходов сигнала через нулевой уровень 20, блок выделения частоты основного тона сигнала 22.

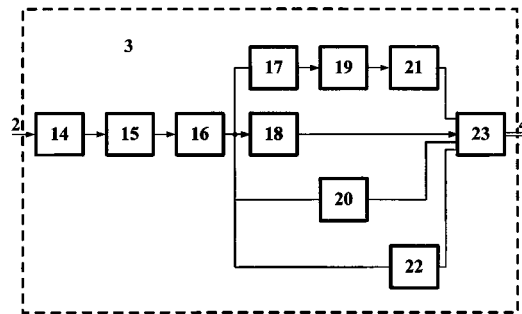
Работа блока определения параметрического описания речевого сигнала 3 происходит следующим образом.

Из аудиосигнала, поступающего на вход 2 блока определения параметрического описания речевого сигнала (фиг. 1), при помощи выделителя начала/конца речевого сигнала 14 происходит формирование речевых участков, которые передаются в сегментатор речевого сигнала на последовательность сегментов 15. Сегментатор осуществляет формирование временных сегментов, содержащих отсчеты входного сигнала и фильтрацию сигнала в этих сегментах с помощью полосового фильтра с полосой пропускания 0,3-3,4 кГц, которая соответствует частотному диапазону речевого сигнала. Оптимальным размером сегмента является 0,07 секунды при частоте дискретизации сигнала 22050 Гц. После умножения на взвешивающее окно в блоке 16 сегменты речевого сигнала поступают в блок добавления к сигналу в сегменте нулей 17, а также в блоки расчета классификационных параметров, а именно блок вычисления среднеквадратического значения сигнала 18, блок вычисления среднего числа переходов сигнала через нулевой уровень 20, блок выделения частоты основного тона сигнала 22. Добавление к сигналу в сегменте нулей обусловлено требованием четности для осуществления быстрого дискретного преобразования Фурье, выполняемого в вычислителе преобразования Фурье 19, результаты которого передаются

ВУ 6229 U 2010.06.30

в блок вычисления спектральной оценки сигнала 21, результатом работы которого является оценка динамики изменения спектра сигнала. Рассчитанные значения классификационных параметров поступают на входы формирователя параметрических описаний входного речевого сигнала 23, выход которого соединен с входом коммутатора 5.

Совместное применение методов спектрального и временного анализа позволило разработать быстродействующее и надежное устройство автоматического распознавания диктора, совмещающее эффективность и компактность вычислений и обеспечивающее результаты, обладающие высокой степенью достоверности и точности. Данное устройство функционирует в режиме реального времени благодаря высокой скорости обработки аудиосигнала, при этом отсутствует необходимость в применении дополнительных датчиков.



Фиг. 2