

**ОПИСАНИЕ
ПОЛЕЗНОЙ
МОДЕЛИ К
ПАТЕНТУ**
(12)

РЕСПУБЛИКА БЕЛАРУСЬ



НАЦИОНАЛЬНЫЙ ЦЕНТР
ИНТЕЛЛЕКТУАЛЬНОЙ
СОБСТВЕННОСТИ

(19) **ВУ** (11) **4214**
(13) **U**
(46) **2008.02.28**
(51) МПК (2006)
G 10L 15/00

(54)

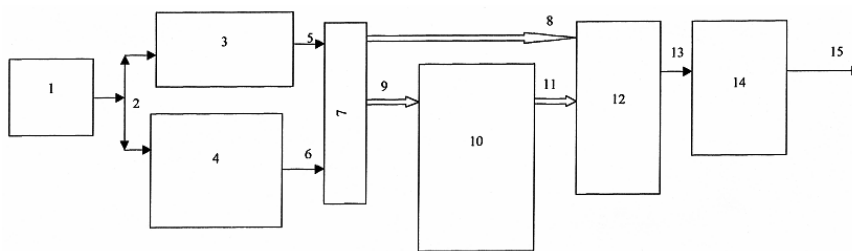
УСТРОЙСТВО РАСПОЗНАВАНИЯ ДИКТОРА

(21) Номер заявки: u 20070153
(22) 2007.02.28
(71) Заявитель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(72) Авторы: Воробьев Василий Иванович; Давыдов Андрей Геннадьевич; Давыдов Геннадий Владимирович; Лыньков Леонид Михайлович; Шамгин Юрий Васильевич (ВУ)
(73) Патентообладатель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(57)

Устройство распознавания диктора, содержащее источник речевого сигнала, блок определения параметрического описания речевого сигнала, соединенные последовательно, коммутатор, блок сравнения параметрических описаний эталона и входного речевого сигнала, блок принятия решения о распознаваемом дикторе и запоминающее устройство эталонов параметрических описаний речевых сигналов дикторов и фазовых квазиинвариантов, при этом источник речевого сигнала подключен ко входу блока определения параметрического описания речевого сигнала, выход которого соединен с первым 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 эталонов параметрических описаний речевых сигналов дикторов и фазовых квазиинвариантов, выход которого подключен ко второму входу блока сравнения параметрических описаний эталона и входного сигнала, выход которого соединен со входом блока принятия решения о распознаваемом дикторе, выход которого является выходом устройства в целом, **отличающееся** тем, что дополнительно содержит блок межкомпонентной фазовой обработки речевого сигнала, выполняющий вычисление фазового квазиинварианта, под которым понимается совокупность разностей фаз между квазигармоническим колебанием на частоте основного тона и обертонами, вход которого подключен к выходу источника речевых сигналов, а выход - ко второму входу коммутатора.



Фиг. 1

ВУ 4214 U 2008.02.28

(56)

1. Мясников Л.Л. Объективное распознавание звуков речи // Журнал технической физики. - Т. XIII. - Вып. № 3. - 1943. - С. 709-715.

2. Мясников Л.Л., Мясникова Е.Н. Автоматическое распознавание звуковых образов. - Л.: Энергия, 1970. - С. 183; - С. 135-138.

3. Патент US 7,065,487 B2, МПК G 10L 19/10.

4. Патент RU 2 230 375 C2, МПК G 10L 15/00, 17/00.

5. Воробьев В.И., Давыдов Г.В., Шамгин Ю.В. Фазовые соотношения между основным тоном и обертонами гласных звуков // Доклады БГУИР. - № 2. - 2006. - С. 64-68.

6. Воробьев В.И. Межкомпонентная фазовая обработка речевых сигналов для их распознавания и идентификации дикторов. Акустика речи. Медицинская и биологическая акустика. Архитектурная и строительная акустика: Шумы и вибрации. Сб. тр. XVIII сессии Российского акустического общества. Т3. - М.: ГЕОС, 2006. - С. 48-51.

Полезная модель относится к устройствам распознавания диктора и может быть использована для распознавания дикторов с целью ограничения несанкционированного доступа к материальным и информационным ресурсам, а также, например, для снижения вероятности ошибок при преобразовании речи в текст.

Известно устройство автоматического распознавания звуков речи [1, 2], содержащее источник речевых сигналов, подключенный к входу усилителя, выход которого подключен к попарному набору полосовых фильтров, предназначенных для выделения первой, второй и третьей формант звуков речи. Выходы фильтров через выпрямители подключены к кодирующему устройству, выполненному на электронных балансных реле. Выходы кодирующего устройства подключены к блоку принятия решения о распознаваемом звуке.

Недостатком такого устройства является сравнительно низкая (75-80) % вероятность распознавания гласных звуков [1] и лишь предположительная возможность преобразования речи в печатный текст. Кроме того, это устройство целевым образом ориентировано на независящее от диктора качество работы и, следовательно, не может обеспечивать распознавание и идентификацию дикторов.

Известны метод и устройство распознавания речи [3], 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 устранения шума по методу нормализации усреднением кепстра, блок хранения акустических моделей, блок хранения моделей языка и блок распознавания речи.

Недостатками предложенных в [3] метода, программы и аппарата распознавания речи являются многоступенчатая процедура анализа входных воздействий, сходимость которой к конкретному результату (в особенности, при распознавании русскоязычной речи) 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 близким к предлагаемому устройству являются способ и устройство распознавания диктора [4], совпадающее с заявляемым устройством по наибольшему числу существенных признаков и принятое за прототип. В основу заявленного в [4] 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 параметрическо-

го 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 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 снижение чувствительности к флуктуациям уровня речевого сигнала, увеличение помехоустойчивости, уменьшение вероятности ошибок распознавания звуков речи и повышение на этой основе надежности распознавания речевых сигналов и дикторов (включая возможность их идентификации).

Указанная задача решается тем, что в устройство распознавания диктора, содержащее источник речевого сигнала, блок определения параметрического описания речевого сигнала, соединенные 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 сигнала, выход которого соединен со входом блока принятия решения о распознаваемом дикторе, выход которого является выходом устройства в целом, дополнительно включен блок межкомпонентной фазовой обработки речевого сигнала, выполняющий вычисление фазового квазиинварианта, под которым понимается совокупность разностей фаз между квазигармоническим колебанием на частоте основного тона и обертонами, вход которого подключен к выходу источника речевых сигналов, а выход - ко второму входу коммутатора. При этом в анализ колебаний частоты основного тона и обертонов речевых сигналов включен ранее не использовавшийся частотный диапазон от 60 до 300 Гц. Для увеличения достоверности распознавания тональных звуков речи (гласных звуков речи) и звонких согласных в систему распознавания введен вычислитель разностей фаз между квазигармоническим колебанием на частоте основного тона и обертонами, вход которого подключен к выходу источника речевых сигналов, а выход - к входу блока принятия решения о распознаваемом звуке.

Важно, что межкомпонентный анализ разностей фаз между квазигармоническим колебанием на частоте основного тона и обертонами отличается высокой помехоустойчивостью и малой чувствительностью к изменениям уровня анализируемых колебаний.

Функциональная схема устройства распознавания дикторов представлена на фиг. 1. На фиг. 2 помещена функциональная схема входящего в состав устройства распознавания дикторов блока межкомпонентной фазовой обработки речевого сигнала.

Устройство распознавания диктора (фиг. 1) 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 и блок межкомпонентной фазовой обработки речевого сигнала 4, коммутатор 7 с двумя входами 5 и 6, блок запоминающего устройства для хранения эталонов параметрического описания речевого сигнала заранее известных дикторов и фазовых инвариантов 12 со входом 9, блок сравнения параметрических описаний эталона и входного речевого сигнала 12 со входами 8 и 11 и выходом 13, блок принятия решения о распознаваемом дикторе 14, выход которого 15 является выходом системы в целом.

Для увеличения достоверности распознавания тональных звуков речи (гласных звуков речи) и звонких согласных и распознавания диктора в устройство распознавания дикторов введен блок межкомпонентной фазовой обработки 4, осуществляющий вычисление фазового квазиинварианта (ФКИ), представляющего собой совокупность разностей фаз между квазигармоническим колебанием на частоте основного тона и обертонами.

Функциональная схема блока 4 представлена на фиг. 2.

Работа блока происходит следующим образом.

Из пароля, поступающего с выхода 2 источника 1 речевых сигналов в цифровой форме (см. фиг. 1), в блоке сегментации тональных звуков парольной фразы 16 выделяются вокализованные участки речевого сигнала. Сегментирование таких участков достаточно распространено в речевых технологиях и дополнительных пояснений не требует. В рассматриваемом случае работа 16 существенно облегчается тем, что пароль известен заранее.

В блоке оценки частоты основного тона 17 реализация каждого из выделенных тональных звуков пропускается через временное окно Хэннинга с длительностью, равной длительности этой реализации. Далее вычисляется кепстр мощности реализации, по которому определяется усредненная на длительности анализируемого звука оценка значения частоты его основного тона F_{k0} . По величине вычисляются примерные значения средних частот ближайших обертонов pF_0 ; $p = 2, 3, 4$.

В блоке синтеза полосовых фильтров и фильтрации 18 осуществляется расчет полосовых фильтров с центральными частотами $F_0, 2F_0, 3F_0, 4F_0$, полосами прозрачности 0,1 от центральных частот. Специфическим требованием к этим фильтрам является недопустимость внесения сдвига фаз между входными и выходными квазигармоническими колебаниями на центральных частотах $F_0, 2F_0, 3F_0, 4F_0$.

В блоке 18 помимо синтеза полосовых фильтров осуществляется и вычисление их откликов на поступающие с выхода 2 источника 1 речевых сигналов воздействия. Полосовые фильтры синтезируются для каждого тонального звука пароля и каждого произносящего его диктора.

Колебания с выходов фильтров поступают на блок оценки фазового квазиинварианта 19.

Понятие фазового квазиинварианта (ФКИ) и способ его оценки требуют следующих пояснений.

Колебание с ЧОТ и обертоны реализации обрабатываемого тонального звука $x(t)$, рассматриваемые как квазигармонические процессы, можно представить в виде

$$x(t) = \sum_{p=1}^{p=N} A_p(t) \cos(2\pi F_0 p t + \Phi_p(t)), \quad (1)$$

где $A_p(t)$ и $\Phi_p(t)$ - медленно меняющиеся амплитуда и фаза p -ой квазигармонической составляющей для реализации звука $x(t)$; F_0 - ЧОТ в реализации; N - число выбранных для анализа квазигармонических составляющих.

В формуле (1) аргумент косинуса представляет собой текущее значение полной фазы p -го квазигармонического колебания в реализации звука $x(t)$, равное

$$\psi_p(t) = 2\pi F_0 p t + \Phi_p(t); \quad p = \overline{1, N}. \quad (2)$$

Для определения разности фаз $\Delta\psi_1^p(t)$ между колебанием с ЧОТ и p -ой квазигармонической составляющей ($p = 2, 3, 4$) вычисленные по формуле (2) значения $\psi_p(t)$ делятся на p и результат вычитается из полной фазы $\psi_1(t)$ колебания с ЧОТ ($p = 1$)

$$\Delta\psi_1^p(t) = \psi_1(t) - \psi_p(t)/p, \quad p = \overline{1, N}. \quad (3)$$

Функция $\Delta\psi_1^p(t)$ не содержит линейно нарастающих слагаемых. Для достижения взаимного уничтожения линейно нарастающих слагаемых требуется непрерывность функций

$\Phi_1(t)$ и $\Phi_p(t)$, которая обеспечивается применением известной процедуры их "сшивания" в точках квазипериодически возникающих скачков фаз на величину 2π .

Диапазоном однозначного определения величины $\Delta\psi_1^p(t)$ является отрезок $[0; 2\pi/p]$. Поэтому вычисляемые по формуле (3) значения $\Delta\psi_1^p(t)$ необходимо нормировать по модулю $|2\pi/p|$:

$$\Delta\psi_1^p(t) = \left[\Phi_1(t) - \Phi_p(t)/p \right] \Big|_{\frac{2\pi}{p}}, \quad p = \overline{1, N}. \quad (4)$$

При вычислении разности фаз между любыми q -ой и p -ой квазигармонической составляющими формула (4) переписывается в виде

$$\Delta\psi_p^q(t) = \left[\Phi_p(t) - \Phi_q(t)p/q \right] \Big|_{\frac{2\pi p}{q}}, \quad p = \overline{1, N}; \quad q = \overline{2, N}. \quad (5)$$

Полные фазы $\psi_p(t)$ отфильтрованных и доступных для преобразований по отдельности квазигармонических компонентов $A_p(t) \cos(\psi_p(t))$, ($p = \overline{1, N}$) целесообразно определять с помощью перехода к аналитическим сигналам с использованием преобразования Гильберта.

Для придания межкомпонентным фазовым характеристикам разных тональных звуков речи у различных дикторов компактной формы используется вектор

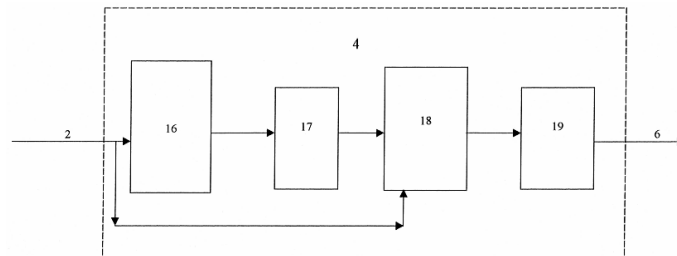
$$\left\{ \Delta\psi_1^2(t); \Delta\psi_1^3(t); \Delta\psi_2^4(t) \right\},$$

который имеет смысл называть фазовым квазиинвариантом (ФКИ).

Вычисление ФКИ в блоке 19 для каждого тонального звука пароля и каждого диктора производится с использованием формул (1-5).

Фазовый квазиинвариант по линии связи 6 поступает на блок принятия решения о распознаваемом дикторе 14 и является дополнительным информационным признаком при принятии решения о распознаваемом дикторе [5, 6].

Введение в устройство блока межкомпонентной фазовой обработки речевого сигнала позволило повысить достоверность распознавания диктора до 98 %.



Фиг. 2