

(19)



Евразийское
патентное
ведомство

(11) 041537

(13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2022.11.03

(21) Номер заявки
201900297

(22) Дата подачи заявки
2019.06.19

(51) Int. Cl. G10L 15/06 (2006.01)
G10L 15/187 (2006.01)

(54) РАСПОЗНАВАТЕЛЬ ФОНЕМ

(43) 2020.12.30

(96) 201900062 (RU) 2019.06.19

(71)(72)(73) Заявитель, изобретатель и
патентовладелец:

КРУПКО ВИКТОР ПАВЛОВИЧ (RU)

(56) RU-C1-2466468
RU-C2-2403628
US-A1-20170154620

(57) Распознаватель фонем, предназначенный для записи в память ЭВМ в виде двоичных чисел образов фонем, поступающих на микрофон. Этим подготавливается база для работы устройства Фонетик по распознаванию фонем путем сравнения образов фонем, заранее записанных в память ЭВМ, с текущими образами фонем, поступающих на микрофон Фонетика. Изделие Фонетик может быть реализовано как автономно, когда в роли компьютера может выступать специализированный микропроцессор, так и в форме приставки, сопрягаемой с компьютером через специальный вход, или в форме специализированной платы, соединяемой с компьютером через слот.

041537 B1

041537 B1

041537

B1

Область техники

Изобретение относится к области распознавания фоновыми методами технической кибернетики.

Предшествующий уровень техники

За последние десятилетия была предпринята не одна попытка реализации алгоритмов распознавания, при этом применялись различные способы от марковских цепей до нейронных сетей, а также алгоритмы на основе результата выбеливания решетчатых фильтров, коэффициентов отражения и логарифмов отношения площадей сечений голосового тракта. Реализуются также алгоритмы, в которых спектрограммы обрабатываются как изображения с помощью последовательности сверточных слоев, соединяющихся с последовательностью рекуррентных блоков.

Некоторые из этих предложений реализованы в ряде изобретений, защищенных патентами.

Патент РФ 2268504 "Способ распознавания фоновыми речью и устройство для реализации способа" предполагает повышение точности распознавания фоновыми речью, достигающейся тем, что формирование признаков фоновыми для распознавания фоновыми речью основано на применении биспектрального анализа, базирующегося на преобразовании цифровых последовательностей кодов, соответствующих речевым сигналам, в биспектральную область, характеризующую взаимодействия между значениями компонентов Фурье на разных частотах в диапазоне речью. К недостаткам можно отнести громоздкий алгоритм, что вызывает большой расход вычислительного ресурса и, как следствие, снижение быстродействия.

Патент РФ 2434245 решает задачу, которая заключается в обнаружении, достоверном распознавании (классификации) и точном определении пространственных координат "акустически малозаметных" - с малым уровнем подводных акустических шумов и слабой акустической отражательной способностью. К недостаткам можно отнести большие требования к вычислительным ресурсам.

Патент РФ 2230375 "Метод распознавания диктора и устройство для его осуществления" относится к области техники анализа речью, в частности к системам ограничения несанкционированного доступа к материальным или информационным ресурсам на основе биометрической информации о говорящем. Используют нестандартную неевклидову меру близости при сравнении форматных векторов двух отдельных сегментов с последующим ее применением к определению статистического сходства сегментных формантных структур речевого сигнала сравниваемых произнесений на основе поиска наилучшего приближения. Вследствие больших требований к вычислительным ресурсам проблематична реализация автономно без сопряжения с мощной вычислительной системой, что сужает область применения.

Патент RU 2268504 "Способ распознавания фоновыми речью и устройство для реализации способа" относится к области анализа и распознавания речевых сигналов и касается способа распознавания фоновыми речью. Техническим результатом является повышение точности распознавания фоновыми речью, достигающейся тем, что формирование признаков фоновыми для распознавания фоновыми речью основано на применении биспектрального анализа, базирующегося на преобразовании цифровых последовательностей кодов, соответствующих речевым сигналам, в биспектральную область, характеризующую взаимодействия между значениями компонентов Фурье на разных частотах в диапазоне речью. Характеризуется большим расходом памяти, что снижает быстродействие и функциональность устройства.

Документ WO 2013002674 A1 "Система и способ распознавания речью" относится к распознаванию речью и обеспечивает дикторонезависимое распознавание слитной речью в режиме реального времени. Система и способ распознавания речью осуществляют прием речевого сигнала, обработку речевого сигнала аналого-цифровым преобразователем с предустановленной частотой дискретизации и разделением на сегменты, спектральный анализ сегментов речевого сигнала и нормализацию спектра на высоких частотах. Отличаются большим расходом вычислительного ресурса и сравнительно невысокой достоверностью распознавания речевых образов.

Сущность изобретения

Распознаватель фоновыми (далее Фонетик) представляет собой аппаратно-программный комплекс, предназначенный для распознавания фоновыми путем сравнения фрагмента определенной длины числового кода образа фонемы, поступающей на микрофон, с фрагментом той же длины заранее записанного в память компьютера числового кода образа этой фонемы. При совпадении этих фрагментов фонема считается распознанной.

Подробное описание изобретения

Устройство для записи образов фоновыми.

Устройство записи образов фоновыми (далее "устройство") представляет собой аппаратно-программный комплекс, предназначенный для записи в память ЭВМ в виде двоичных чисел образов фоновыми, поступающих на микрофон. При этом каждое число соответствует определенной фонеме и является образом этой фонемы.

Устройство используется для создания массивов образов фоновыми для конкретного оператора-носителя конкретного языка. Для различных операторов или для одного оператора, но использующего разные языки, создаются индивидуальные массивы образов фоновыми.

Технически устройство может быть реализовано как отдельная приставка, соединенная своим выходом со входом ЭВМ или микропроцессора, обладающего необходимым объемом памяти, либо как специализированная плата в составе ЭВМ. А массивы образов фоновыми можно рассматривать как само-

стоятельный информационный продукт.

Если под фонемой понимать звуковой сигнал определённой частоты, длительности и уровня, то можно создавать массивы образов фонем, издаваемых как техническими устройствами, так и живыми организмами, например птицами или животными, включая язык дельфинов.

Функциональная схема устройства изображена на фиг. 1. Работа устройства осуществляется следующим образом.

В исходном состоянии на выходе микрофона сигнал отсутствует, генератор тактовых импульсов (ГТИ) выключен; в регистр, а также в элементы массивов образов гласных и согласных фонем записано число нуль.

При поступлении сигнала с микрофона на клиппер происходит его усиление и ограничение на уровне, достаточном для преобразования в сигнал, напоминающий сигнал типа меандр, но только не периодический и с различной длительностью импульсов. При этом теряется индивидуальная окраска речи, но полностью сохраняются точки перехода сигнала через нулевой уровень.

С появлением первого импульса на выходе клиппера запускается ГТИ, чем осуществляется фазовая привязка ГТИ к началу фонемы.

Частота следования импульсов ГТИ выбирается в соответствии с требованием теоремы Котельникова и уточняется опытным путем; длительность импульсов этого генератора подбирается опытным путем.

Одновременно сигнал с выхода клиппера подается на вход схемы совпадения "И", а на второй вход этой схемы подаются тактовые импульсы с выхода ГТИ.

С выхода схемы совпадения "И" сигнал в виде двоичного числа в последовательном коде подается в компьютер на вход регистра и записывается в него с помощью импульсов ГТИ.

С регистра двоичное число записывается как элемент массива образов гласных или согласных фонем соответственно.

По окончании записи устройство приводится в исходное состояние с помощью стандартной кнопки "Установка нуля", а номер фонемы вводится вручную с клавиатуры.

Распознаватель фонем.

Функциональная схема Фонетика изображена на фиг. 2. Его работа осуществляется следующим образом.

В исходном состоянии на выходе микрофона сигнал отсутствует, генератор тактовых импульсов (ГТИ) выключен, в регистр компьютера записано число нуль, а в массивы образов гласных и согласных фонем записаны с помощью устройства записи образов фонем, как элементы этих массивов, двоичные числа, представляющие собой образы фонем. Количество элементов массивов равно количеству фонем, подлежащих распознаванию.

При поступлении сигнала с выхода микрофона на вход клиппера происходит его усиление и ограничение на уровне, достаточном для преобразования в сигнал, напоминающий сигнал типа меандр, но только не периодический и с различной длительностью импульсов. При этом теряется индивидуальная окраска речи, но полностью сохраняются точки перехода сигнала через нулевой уровень.

С появлением первого импульса на выходе клиппера запускается ГТИ, чем осуществляется фазовая привязка ГТИ к началу фонемы. Частота следования импульсов ГТИ выбирается в соответствии с требованием теоремы Котельникова и уточняется опытным путем; длительность импульсов этого генератора подбирается опытным путем.

Одновременно сигнал с выхода клиппера подается на вход схемы совпадения "И", а на второй вход этой схемы подаются тактовые импульсы с выхода ГТИ.

С выхода схемы совпадения "И" сигнал в виде двоичного числа в последовательном коде подается в компьютер на вход регистра и записывается в него с помощью импульсов ГТИ.

Одновременно сигнал с выхода микрофона подается на входы дискриминаторов по частоте и по уровню сигнала. С выходов этих дискриминаторов сигнал подается на измеритель длительности фонемы (τ_f), с выхода которого сигнал поступает для запуска программной части комплекса Фонетик.

Если $\tau_f \geq 5 \cdot 10^{-2}$ с, то выбирается первый элемент массива образов гласных фонем, в противном случае выбирается первый элемент массива образов согласных фонем.

Далее выделяется фрагмент определенной длины двоичного кода числа, выбранного из массива образов фонем. Подобным же образом выделяется фрагмент той же длины двоичного кода числа, записанного в регистр компьютера. Длина фрагментов кода подбирается опытным путем.

После этого выделенные фрагменты двоичных кодов чисел сравниваются. Если эти фрагменты совпадают, то фонема, поступившая на микрофон, считается распознанной, выдается ее номер и поступает сигнал на один из входов схемы "ИЛИ", с выхода которой поступает сигнал на установку аппаратной части Фонетика в исходное состояние.

При несовпадении фрагментов проверяется количество сдвигов вправо на 1 разряд числа, поступившего из массива образов фонем. Максимальное (max) количество сдвигов числа вправо на 1 разряд определяется исходя из разрядности числа и длины фрагмента кода числа.

Если количество сдвигов числа вправо не достигло $\max$, то число сдвигается вправо на 1 разряд, далее из полученного числа выделяется фрагмент кода той же длины и передается для сравнения с фрагментом кода числа, полученного с регистра. В противном случае проверяется, достигнут ли максимальный ($\max$) номер элемента массива.

Если он не достигнут, то осуществляется переход к следующему элементу массива образов фонем и процесс распознавания продолжается. В противном случае фонема считается нераспознанной и сигнал подается на второй вход схемы "ИЛИ" для приведения в исходное состояние аппаратной части Фонетика.

Предлагаемый аппаратно-программный комплекс-распознаватель фонем позволяет подойти к распознаванию речи через распознавание фонем, так сказать, "первокирпичиков" речи, что значительно снижает размер задачи и сужает область поиска.

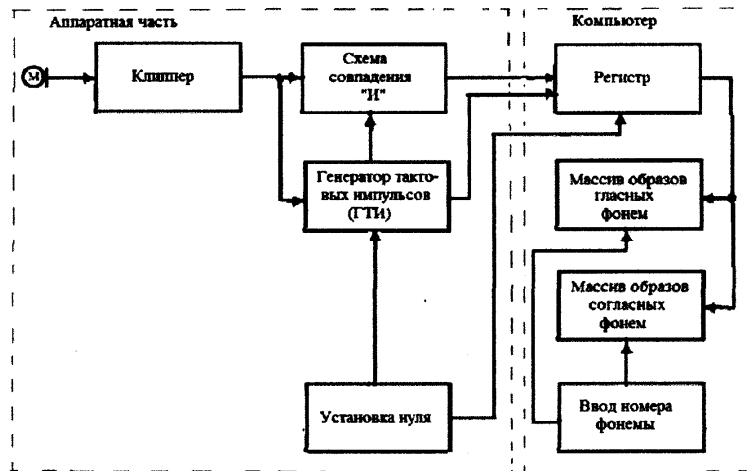
После распознавания фонем и передачи Фонетиком соответствующих букв в компьютер может быть реализована дальнейшая обработка полученной информации по алгоритмам, сходным с используемыми современными поисковиками. Тем самым будет открыт путь к возможности полноценного вербального общения с искусственными высокоорганизованными материальными системами, которые, в некоторой степени некорректно, принято называть искусственным разумом.

В автономном исполнении изделие Фонетик может быть легко реализовано с массогабаритными и энергетическими показателями, не превышающими показатели телевизионного пульта управления, и использовано в широких областях.

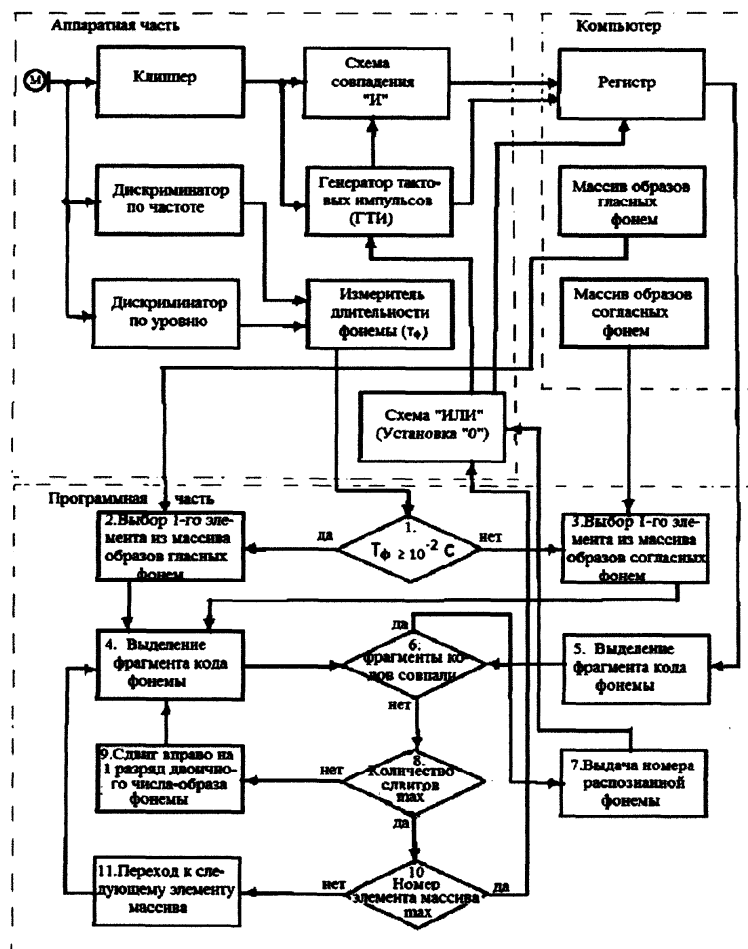
ФОРМУЛА ИЗОБРЕТЕНИЯ

Распознаватель фонем, отличающийся тем, что используется соединенный с микрофоном клиппер, выполненный с возможностью усиливать поступающий с микрофона сигнал и ограничивать его по амплитуде, с выхода которого клиппированный речевой сигнал подается на запуск генератора тактовых импульсов (ГТИ) и на первый вход схемы совпадения, на второй вход которой подаются тактовые импульсы с выхода ГТИ, а с выхода схемы совпадения импульсы в форме двоичного числа в последовательном коде подаются на вход регистра компьютера, соединенного с выходом схемы совпадения и ГТИ, после чего это число с регистра компьютера поступает в программную часть распознавателя; ГТИ и регистр соединены с выходами схемы "ИЛИ" для выключения ГТИ и записи в регистр компьютера числа нуль, а на входы схемы "ИЛИ" подаются сигналы из программной части распознавателя; с выходом микрофона соединены также входы дискриминаторов по частоте и по уровню сигнала, выполненных с возможностью выдачи сигнала на выходе при изменениях частоты или уровня входного сигнала соответственно, а выходы этих дискриминаторов соединены со входами измерителя длительности фонемы (τ_f), выполненного по любой из известных схем (интегратор по нарастанию напряжения на конденсаторе, интегратор по нарастанию тока в катушке, счетчик и т.п.), с выхода которого сигнал подается для запуска программной части распознавателя, и программную часть распознавателя, отличающуюся тем, что с целью распознавания фонем применяется сравнение заранее записанных образов фонем с образом текущей фонемы, поступающей на вход системы, осуществляемое по следующему алгоритму:

- 1) проверка длительности фонемы; при $\tau_f \geq 5 \cdot 10^{-2}$ с перейти к этапу 2), в противном случае перейти к этапу 3);
- 2) выбор 1-го элемента из массива образов гласных фонем;
- 3) выбор 1-го элемента из массива образов согласных фонем;
- 4) выделение фрагмента кода, выбранного из массива образов фонем;
- 5) выделение фрагмента кода, записанного в регистр компьютера;
- 6) сравнение выделенных на этапах 4) и 5) фрагментов кодов; если фрагменты кодов совпали, то перейти к этапу 7), в противном случае перейти к этапу 8);
- 7) выдача номера распознанной фонемы, а также сигнала на вход схемы "ИЛИ" в аппаратной части распознавателя; с выхода схемы "ИЛИ" сигналы подаются на выключение ГТИ и запись в регистр компьютера числа нуль;
- 8) проверка количества сдвигов вправо на 1 разряд числа, выбранного из массивов образов фонем; если количество сдвигов не достигло $\max$, то перейти к 9), в противном случае перейти к 10);
- 9) сдвиг вправо на 1 разряд двоичного числа-образа фонемы и переход к этапу 4);
- 10) проверка номера элемента массива; если номер элемента массива не достиг $\max$, то перейти к этапу 11), в противном случае выдать сигнал на второй вход схемы "ИЛИ" в аппаратной части распознавателя;
- 11) перейти к следующему элементу массива, находящегося в работе, а затем перейти к этапу 4).



Фиг. 1



Фиг. 2



Евразийская патентная организация, ЕАПО

Россия, 109012, Москва, Малый Черкасский пер., 2