(19) **RU** (11) **2 180 974** (13) **C2**

(51) МПК[7] **G 10 L 15/02, 19/02**

РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка: 2000107717/09, 29.03.2000

(24) Дата начала действия патента: 29.03.2000

(46) Дата публикации: 27.03.2002

(56) Ссылки: RU 2136059 C1, 27.08.1999. JP 04-003200 A, 08.01.1992. US 5268991 A, 07.12.1993. US 5933803 A, 03.08.1999. DE 3808038 AI, 28.09.1989. JP 02-050199 A, 20.02.1990. JP 07-160294 A, 23.06.1995.

(98) Адрес для переписки:
443010, г. Самара, ул. Л.Толстого, 23, ПГАТИ

(71) Заявитель:
Поволжская государственная академия телекоммуникаций и информатики

(72) Изобретатель: Брайнина И.С., Кузнецов М.В.

(73) Патентообладатель:
Поволжская государственная академия телекоммуникаций и информатики

(54) СПОСОБ СЖАТИЯ ИЗОЛИРОВАННЫХ СЛОВ

(57)

Изобретение относится к цифровой обработке речи. Его использование в системах речевого ответа абонентам телефонной сети, в справочных службах, для озвучивания объявлений на транспорте, в иных общественных местах обеспечивает достижение технического результата в виде сокращения объема памяти для хранения и воспроизведения требуемой речевой информации. Способ заключается в том, что разделяют предварительно записанный в оперативное запоминающее устройство сигнал очередного слова на отрезки равной длины, вычисляют в каждом из них средний модуль этого сигнала и число смен знака в нем, определяют по этим данным два образа обрабатываемого слова, описывающие характер изменения его сигнала во времени по уровню и мгновенной частоте. Технический результат достигается благодаря тому, что определяют внутри слова участки локальной стационарности, на которых одновременно уровень сигнала и его мгновенная частота почти не изменяются, выделяют внутри каждого такого участка отрезок сигнала, служащий эталонным периодом основного тона речи, переписывают данные отрезки сигнала один за другим в постоянное запоминающее устройство, снабжая каждый из них паролем, содержащим информацию о продолжительности данного отрезка сигнала, числе его повторений при воспроизведении слова и величине адаптивного шага квантования, пропорционального среднему модулю сигнала на данном участке локальной стационарности. 4 ил.

RU 2180974 C2

RU 2180974 C2

(19) **RU** (11) **2 180 974** (13) **C2**

(51) Int. Cl.[7] **G 10 L 15/02, 19/02**

RUSSIAN AGENCY
FOR PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION

(21), (22) Application: **2000107717/09, 29.03.2000**

(24) Effective date for property rights: **29.03.2000**

(46) Date of publication: **27.03.2002**

(98) Mail address:
443010, g. Samara, ul. L.Tolstogo, 23, PGATI

(71) Applicant:
Povolzhskaja gosudarstvennaja akademija telekommunikatsij i informatiki

(72) Inventor: **Brajnina I.S.,**
Kuznetsov M.V.

(73) Proprietor:
Povolzhskaja gosudarstvennaja akademija telekommunikatsij i informatiki

(54) PROCESS OF COMPRESSION OF INSULATED LAYERS

(57) Abstract:

FIELD: digital processing of speech. SUBSTANCE: process is intended for use in systems of speech answer to subscribers of telephone network, in information services, for sounding announcements on transport or in other public places. Process includes division of signal of due word preliminary recorded in on-line storage into sections of equal length, computation of mean modulus of this signal and number of changes of sign in each of them. These data are used to define two images of processed word that describe character of change of its signal by level and instantaneous frequency in time. Decreased capacity of speech information needed for storage and playback is obtained by determination of sections of local stationarity in which level of signal and its instantaneous frequency almost do not change simultaneously, by isolation of section of signal which serves as standard period of basic tone of speech inside each such section, by re-recording of given sections of signal one after another into permanent storage supplying each of them with password containing information on duration of given section of signal, on number of its repetitions during of playback of word and on value of adaptive period of quantization proportional to mean modulus of signal across given section of local stationarity. EFFECT: decreased capacity of information storage used for memorization and playback of required speech information. 4 dwg

RU 2180974 C2

RU 2180974 C2

Изобретение относится к технике цифровой обработки речи и может быть использовано в различных приложениях, например в системах речевого ответа абонентам телефонной сети (автоответчики), в справочных службах, для озвучивания объявлений на транспорте, в общественных местах и т.д.

Известен алгоритм сжатия данных звука ISO/MPEG (MUSICAM), использующий информационное сжатие для передачи с высоким качеством сигналов звукового сопровождения телевизионных программ, а также программ цифрового спутникового радиовещания [1] . Этот алгоритм основан на особенностях восприятия звуков ухом человека - так называемом психоакустическом эффекте. Доказано, что человек воспринимает примерно 10% информации, содержащейся в звуковом сигнале, остальные 90% являются избыточными и их можно не передавать по каналу связи. Сигнал определенной частоты (тон), воздействуя на ухо человека, не позволяет различать (маскирует) другие тоны, близкие к данному по частоте и меньшие по уровню. В реальном звуковом сигнале одновременно присутствуют несколько маскирующих тонов на различных частотах. Компоненты сигнала, уровни которых ниже порога маскирования, ухом не воспринимаются и являются избыточными.

Недостатком описанного алгоритма сжатия звукового сигнала является сложность в его реализации. Как и в частотных вокодерах, для сжатия осуществляется анализ мгновенного энергетического спектра сигнала, в данном случае с помощью гребенки фильтров, разделяющих спектр на 32 частотных полосы. В каждой из них по отдельности выполняется аналого-цифровое преобразование, обработке подвергаются поочередно "кадры" сигнала длительностью 24 мсек, с частотой выборки отсчетов, равной 48 кГц. Устранение из передаваемого "кадра" частотных полос с уровнем сигнала ниже порога маскирования, в сочетании с динамическим распределением битов информации между оставшимися полосами, позволяет достичь почти 6-кратного сжатия спектра стереофонического сигнала, с сохранением практически неизменного, очень высокого качества звучания.

Сложность в реализации алгоритма [1] оправдана необходимостью получения высокого качества звучания радиовещательного стереофонического сигнала, которое вовсе не требуется в устройствах типа автоответчиков, для справочных служб и т.д., где достаточно только обеспечить высокую разборчивость и натуральность речи.

Наиболее близким техническим решением (прототипом) является алгоритм цифрового преобразования звукового сигнала на примере изолированных слов, произносимых произвольным диктором, описанный в [2]. Этот адаптивный алгоритм, использующий избыточность речевого сигнала во временной области, позволяет обеспечить распознавание произвольного голоса, независимо от его громкости, темпа речи и частоты основного тона. В отличие от [1], где так же, как и во всех известных вокодерных системах передачи речи, сжатие сигнала достигается путем устранения его частотной избыточности, в [2] показана возможность использования временной избыточности речевого сигнала. Эта избыточность проявляется в сильных корреляционных связях, охватывающих до (10-12) соседних отсчетов речевого сигнала, взятых с частотой дискретизации, равной 8 кГц. В свою очередь, связи между соседними отсчетами вызваны резкой неравномерностью спектра мощности речевого сигнала, имеющего максимум в области (400-500) Гц и быстро спадающего на высоких частотах со скоростью от 6 до 12 дБ на октаву.

Особенно заметно проявляется избыточность речи при произнесении так называемых вокализованных (гласных) звуков, которым соответствуют участки локальной стационарности протяженностью до (150-200) мсек. На каждом таком участке размещаются десятки почти однотипных отрезков сигнала с периодом основного тона речи, индивидуального для каждого голоса. Для мужских голосов этот период колебания голосовых связок составляет (5-20) мсек, а для высоких женских и детских голосов он изменяется в диапазоне (2-5) мсек.

Произнесение звонких согласных (типа "б", "в", "г", "д" и т.д.), а также "м", "н" тоже сопровождается периодическим повторением отрезков основного тона речи, изменяется только их форма, амплитуда и количество на протяжении звучания данной согласной. И только глухие согласные (например, "п", "ф", "к", т" и другие) не содержат в своем составе периодических отрезков, звуковой сигнал напоминает шум и отличается низким уровнем звучания, малой протяженностью во времени и высокой частотой смены знака сигнала (переходов через ноль).

Недостатком прототипа [2] является то, что алгоритм его функционирования направлен, главным образом, на распознавание изолированных (разделенных паузами) слов, произносимых любым голосом, а задача сжатия речи путем устранения избыточности рассматривается как вспомогательная и решается лишь частично. При этом устраняется та небольшая часть избыточности, которая связана с индивидуальными особенностями голоса диктора. Основная же часть избыточности, вызванная многократной повторяемостью отрезков сигнала на периодах основного тона речи, сохраняется неизменной.

Техническим результатом предлагаемого изобретения является сокращение объема памяти, удешевление и уменьшение габаритов постоянных запоминающих устройств, необходимых для хранения и воспроизведения требуемой речевой информации, содержащей изолированные (разделенные паузами) слова.

Предлагаемый способ сжатия изолированных слов, заключающийся в том, что по превышению заданного порогового уровня определяются начало и конец очередного слова, оно предварительно записывается в оперативное запоминающее устройство (ОЗУ) и подразделяется на отрезки равной длины, в каждом из которых вычисляется средний модуль сигнала и число смен знака, по этим данным определяются два "образа" слова, описывающие характер изменения сигнала во времени по уровню и мгновенной частоте, отличается тем, что определяются участки локальной

RU 2180974 C2

RU 2180974 C2

стационарности внутри слова, на которых одновременно уровень сигнала и его мгновенная частота почти не изменяются, внутри каждого такого участка выделяется отрезок сигнала, служащий эталонным периодом основного тона речи, данные отрезки сигнала один за другим переписываются в постоянное запоминающее устройство (ПЗУ), при этом каждый из них снабжается "паролем", содержащим информацию о продолжительности данного отрезка сигнала, числе его повторений при воспроизведении слова и величине адаптивного шага квантования, пропорционального среднему модулю сигнала на данном участке локальной стационарности.

В соответствии с предлагаемым способом опишем подход к сжатию речевых сигналов, основанный на упрощенном описании одного эталонного, из числа периодически повторяющихся на каждом из участков локальной стационарности, отрезка с периодом основного тона речи, и дальнейшем синтезе речи по этим отрезкам.

Сжатие происходит на временной основе, при этом используется избыточность квазистационарных участков вокализованной речи и устраняются малые уровни, т.е. сигнал в паузах приравнивается к нулю. Речь разбивается на отрезки, равные 16 мсек, не превышающие половины интервала локальной стационарности (порядка 40 мсек). На каждом отрезке определяется средний модуль, число переходов через ноль и устанавливается адаптивный шаг квантования по уровню, равный половине среднего модуля. Использование адаптивного шага квантования позволяет снизить разрядность кода отсчета речевого сигнала без заметных потерь почти в 3 раза.

На вокализованном участке в процессе синтеза слова воспроизводится один период основного тона речи столько раз, сколько звучит этот участок слова.

Невокализованные участки речи сжимаются в меньшей степени, но они короче гласных звуков, поэтому, взяв за основу средний период основного тона речи и его повторяемость, можно сжать сигнал приблизительно в 10 раз, что в сочетании с использованием адаптивного шага квантования и снижением разрядности кода отсчета дает выигрыш порядка 30 раз.

На фиг. 1 и 2 изображены гистограммы распределения во времени среднего модуля сигнала и количества переходов через ноль на интервалах анализа длиной 16 мсек, при произнесении женским голосом слова "НОЛЬ". Эти гистограммы представляют собой два "образа" слова, описывающие характер изменения сигнала во времени по уровню и мгновенной частоте. На фиг. 2 отчетливо заметны участки локальной стационарности, соответствующие звукам "Н" (с 1 по 9), "О" (с 10 по 20) и "ЛЬ" (с 21 по 26). Этим участкам на фиг. 1 соответствуют разные средние модули сигнала.

На фиг. 3 представлены временные диаграммы реального речевого сигнала при произнесении слова "НОЛЬ". Разделение всего слова на интервалы локальной стационарности, соответствующие звукам "Н" длиной 144 мсек, "О" длиной 176 мсек, "ЛЬ" длиной 96 мсек основано на данных гистограммы на фиг. 2.

Общая протяженность слова составляет 416 мсек или 3328 отсчетов частоты дискретизации 8 кГц. На фиг. 3 явно прослеживаются периоды основного тона речи, для данного голоса порядка 3 мсек, или 24 отсчета.

На периодах основного тона речи имеется два перехода через ноль для невокализованных звуков "Н" и "ЛЬ". Для вокализованного звука "О" частота переходов через ноль вдвое выше. Средние уровни сигнала на невокализованных участках в 1,5-2 раза ниже, чем на вокализованном. Каждый из участков локальной стационарности характеризуется своей формой сигнала, многократно повторяющейся с периодом основного тона речи.

Используя гистограмму на фиг. 1, можно подразделить каждый звук на несколько участков, близких по уровню. Внутри каждого из них можно выделить один период основного тона и повторить его в соответствии с длиной данного участка.

Дополнительное сжатие достигается путем снижения разрядности кода отсчета сигнала. Для этого шаг квантования по уровню устанавливается адаптивным, пропорционально среднему модулю сигнала на интервале анализа. Моделирование на ПК алгоритма сжатия показало, что удовлетворительное качество звучания достигается при использовании трехразрядного кода. Это соответствует передаче знакового разряда отсчета и двух разрядов модуля, то есть чисел ± 3, поэтому выбран адаптивный шаг квантования, равный половине среднего модуля сигнала.

На фиг. 4 приведены временные диаграммы синтезированного по описанному алгоритму слова. В соответствии с гистограммой на фиг. 1, для описания звука "Н" выбраны четыре эталонных периода основного тона речи с разной величиной адаптивного шага квантования. Для вокализованного звука "О" выбрано всего три эталонных периода основного тона, один на границе звуков "Н-О", второй в середине звука "О" и третий на переходе звука "О" в "ЛЬ". Поскольку звук "О" наиболее протяженный и составляет почти половину слова, на нем достигается наиболее эффективное сжатие. Звук "ЛЬ" наименее протяженный и для его описания достаточно трех эталонных периодов основного тона речи, а именно на переходе "О" в "ЛЬ", в середине звука и в конце, на участке затухания сигнала.

Всего на протяжении слова были выбраны и запомнены десять эталонных периодов основного тона речи общей протяженностью 240 трехразрядных отсчетов, то есть 90 байт. С учетом необходимых описаний ("пароля") каждого эталонного периода, по два байта на период, на все синтезированное слово потребуется 110 байт.

До сжатия слово содержало 3328 восьмиразрядных отсчетов, то есть для его описания требовалось 3328 байт. Таким образом, предложенный алгоритм обеспечил 30-кратное сжатие необходимого объема памяти. При этом сохранилась узнаваемость по голосу, качество звучания соответствовало экспертной оценке в 3 балла по 5-балльной шкале.

Отметим также, что предложенный

RU 2180974 C2

RU 2180974 C2

алгоритм сжатия позволяет с легкостью осуществлять обмен степени сжатия на качество звучания путем изменения разрядности кода отсчетов сигнала и количества эталонных периодов основного тона речи, входящих в состав синтезированного слова.

Литература

1. Алгоритм сжатия данных звука ISO/MPEG (MUSICAM). Глеб Высоцкий, [Image] GS Урал, июль 1998 г., статья в сети Интернет.

2. Брайнина И. С., Кузнецов М.В. Устройство для распознавания изолированных слов. Патент 2136659, 6 G 10 L 7/04, БИ 24, 1999.

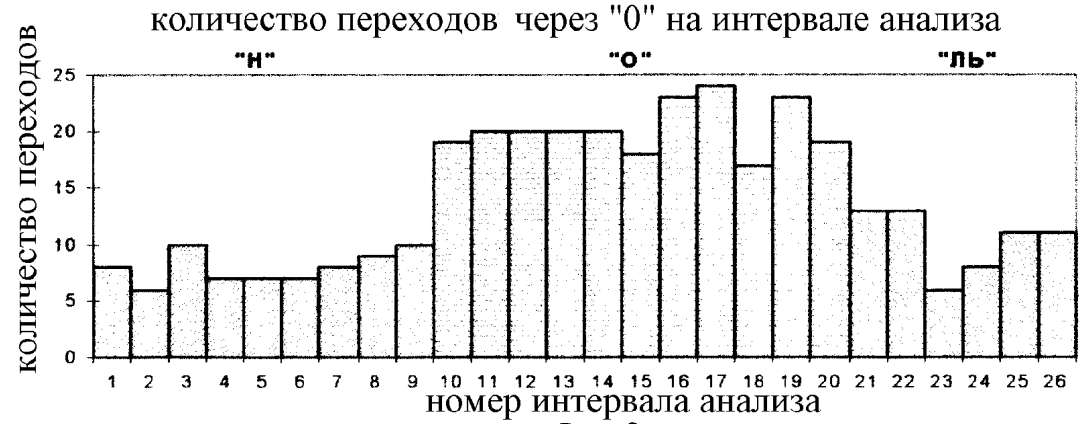
Формула изобретения:

Способ сжатия изолированных слов при цифровой обработке речи, заключающийся в том, что разделяют предварительно записанный в оперативное запоминающее устройство сигнал очередного слова на отрезки равной длины, вычисляют в каждом из них средний модуль этого сигнала и число смен знака в нем, определяют по этим данным два образа обрабатываемого слова, описывающих характер изменения его сигнала во времени по уровню и мгновенной частоте, отличающийся тем, что определяют внутри слова участки локальной стационарности, на которых одновременно уровень сигнала и его мгновенная частота почти не изменяются, выделяют внутри каждого такого участка отрезок сигнала, служащий эталонным периодом основного тона речи, переписывают данные отрезки сигнала один за другим в постоянное запоминающее устройство, снабжая каждый из них паролем, содержащим информацию о продолжительности данного отрезка сигнала, числе его повторений при воспроизведении слова и величине адаптивного шага квантования, пропорционального среднему модулю сигнала на данном участке локальной стационарности.

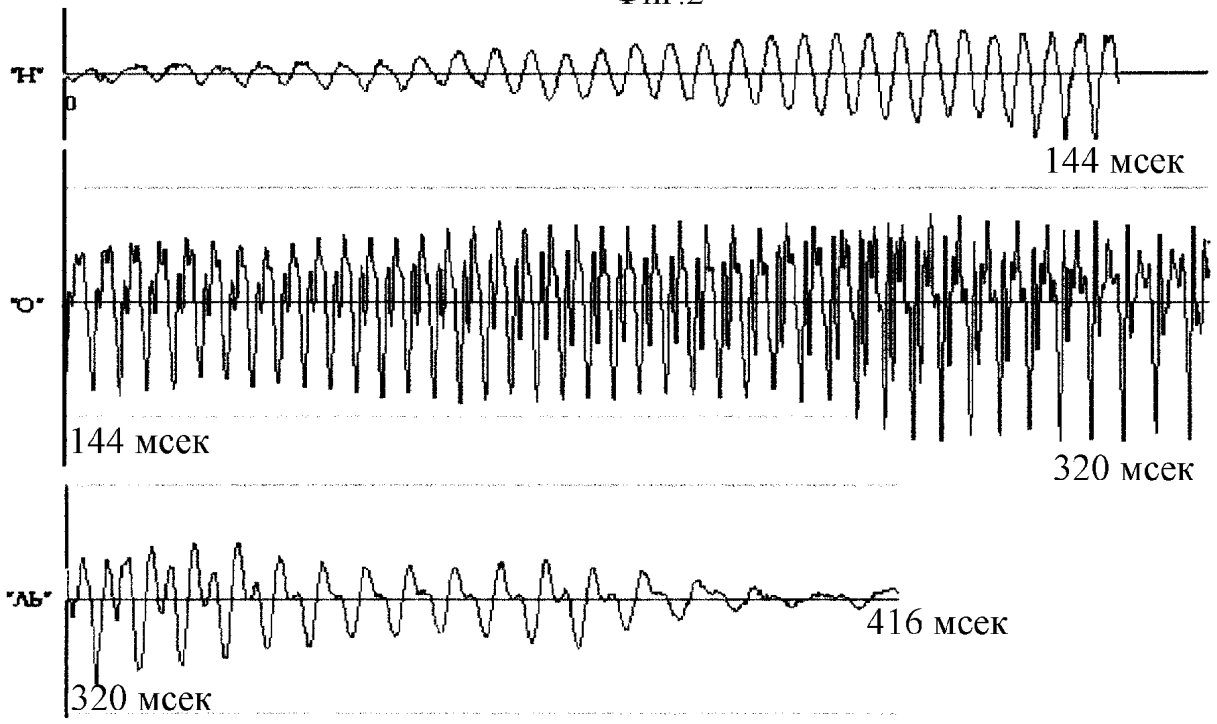
RU 2180974 C2

RU 2180974 C2

Фиг.1

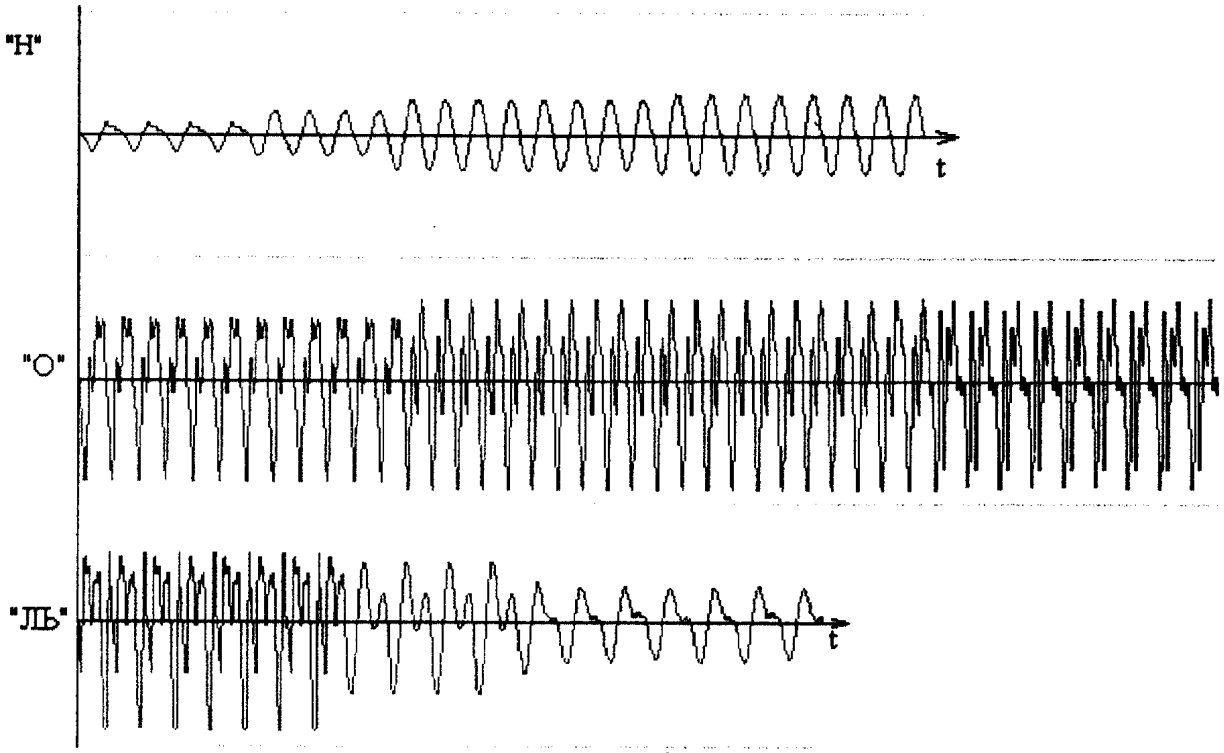


Фиг.2



Фиг.3

RU 2180974 C2



Фиг.4

RU 2180974 C2

RU 2180974 C2