



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ,
ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21), (22) Заявка: 2005113877/09, 09.10.2003

(24) Дата начала отсчета срока действия патента:
09.10.2003(30) Конвенционный приоритет:
11.10.2002 US 60/417,667

(43) Дата публикации заявки: 10.10.2005

(45) Опубликовано: 20.08.2008 Бюл. № 23

(56) Список документов, цитированных в отчете о
поиске: WO 01/22402 A1, 29.03.2001. RU 2107951
C1, 27.03.1998. RU 2146394 C1, 10.03.2000. US
5911128 A, 08.06.1999. RU 97119637 A,
20.09.1999.(85) Дата перевода заявки РСТ на национальную фазу:
11.05.2005(86) Заявка РСТ:
CA 03/01571 (09.10.2003)(87) Публикация РСТ:
WO 2004/034379 (22.04.2004)

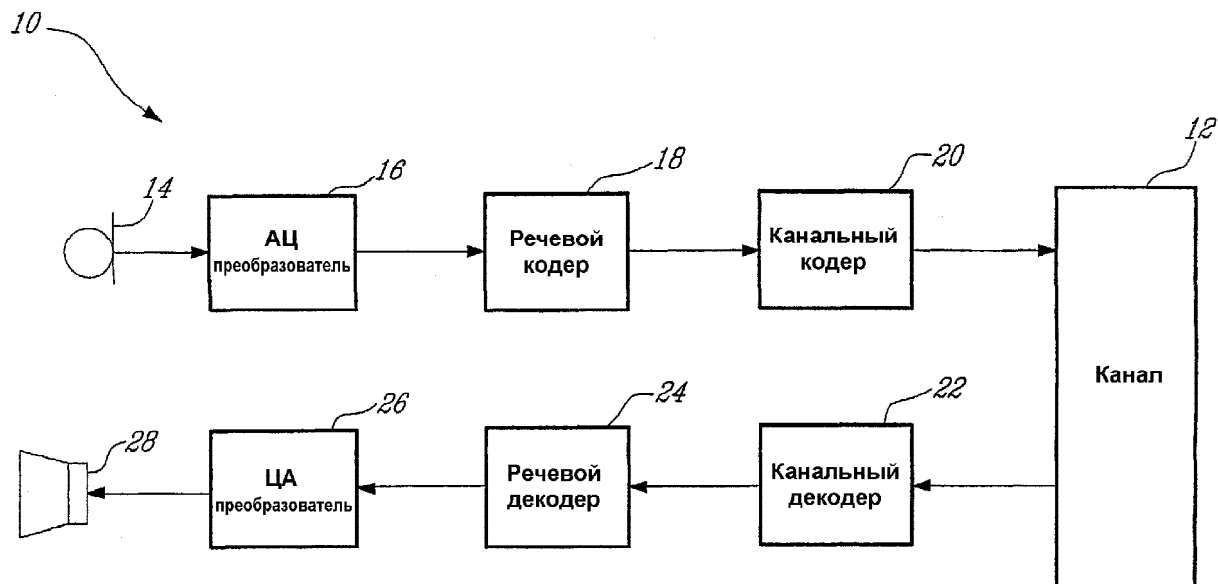
Адрес для переписки:
129010, Москва, ул. Б. Спасская, 25, стр.3,
ООО "Юридическая фирма Городисский и
Партнеры", пат.пов. Ю.Д.Кузнецову, рег.№ 595

(72) Автор(ы):
ЖЕЛИНЕК Милан (CA)(73) Патентообладатель(и):
НОКИА КОРПОРЕЙШН (FI)(54) СПОСОБЫ И УСТРОЙСТВА УПРАВЛЯЕМОГО ИСТОЧНИКОМ ШИРОКОПОЛОСНОГО
КОДИРОВАНИЯ РЕЧИ С ПЕРЕМЕННОЙ СКОРОСТЬЮ В БИТАХ

(57) Реферат:

Изобретение относится к цифровому кодированию звуковых сигналов. Предложены системы и способы классификации и кодирования речевых сигналов. Классификацию сигналов проводят в три этапа, на каждом из которых обеспечивают различение конкретного класса сигналов. Сначала детектор речевой активности осуществляет различение между активными и неактивными речевыми кадрами. Если обнаруживают неактивный речевой кадр, то классификацию заканчивают, а кадр кодируют посредством генерирования комфортного шума. Если обнаруживают активный речевой кадр, то этот кадр подвергают воздействию второго классификатора, предназначенного для

различения невокализованных кадров. Если классификатор классифицирует кадр как невокализованной речевой сигнал, то классификацию заканчивают, а кадр кодируют с помощью способа кодирования, оптимизированного для невокализованных сигналов. В противном случае речевой кадр пропускают через модуль классификации "устойчивых вокализованных" сигналов. Если кадр классифицируется как устойчивый вокализованный кадр, то этот кадр кодируют с помощью способа кодирования, оптимизированного для устойчивых вокализованных сигналов. В противном случае, если кадр содержит неустойчивый речевой сегмент, например вокализованный начальный или быстро эволюционирующий сигнал, то используют



Фиг. 1

RU 2331933 C2

RU 2331933 C2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY,
PATENTS AND TRADEMARKS

(51) Int. Cl.
G10L 19/14 (2006.01)

(12) ABSTRACT OF INVENTION

(21), (22) Application: **2005113877/09, 09.10.2003**

(24) Effective date for property rights: **09.10.2003**

(30) Priority:
11.10.2002 US 60/417,667

(43) Application published: **10.10.2005**

(45) Date of publication: **20.08.2008 Bull. 23**

(85) Commencement of national phase: **11.05.2005**

(86) PCT application:
CA 03/01571 (09.10.2003)

(87) PCT publication:
WO 2004/034379 (22.04.2004)

Mail address:
**129010, Moskva, ul. B. Spasskaja, 25, str.3,
OOO "Juridicheskaja firma Gorodisskij i
Partnery", pat.pov. Ju.D.Kuznetsovu, reg.№ 595**

(72) Inventor(s):
ZhELINEK Milan (CA)

(73) Proprietor(s):
NOKIA KORPOREJShN (FI)

(54) METHODS AND DEVICES OF SOURCE-GUIDED BROADBAND SPEECH CODING AT VARIABLE BIT RATE

(57) Abstract:

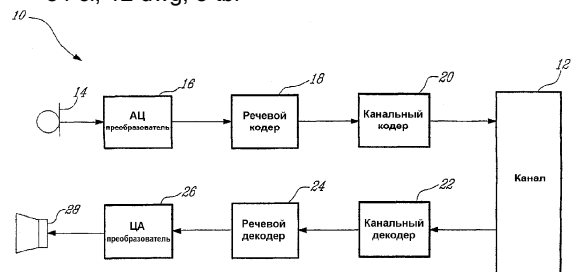
FIELD: physics.

SUBSTANCE: invention claims systems and methods of speech signal classification and coding. Signal classification passes over three stages, with recognition of a definite signal class at each stage. First, active speech detector recognises active and inactive speech frames. If an inactive speech frame is found, the classification is finished, and the frame is encoded by comfortable noise generation. If an active speech frame is found, it undergoes second classification recognising non-vocalised frames. If the frame is recognised as non-vocalised speech signal, the classification is finished, and the frame is encoded by a method optimised for non-vocalised signals. In the opposite case, the speech frame is directed to 'stable vocalised' signal classification module. If the

frame is classified as a stable vocalised frame, it is encoded by a method optimised for stable vocalised signals. In the opposite case, if the frame contains instable speech segment, e.g. vocalised initial or rapidly evolving signal, then a speech coder is applied.

EFFECT: improved speech quality at a given average data transfer speed.

84 cl, 12 dwg, 5 tbl



Фиг. 1

Область техники

Настоящее изобретение относится к цифровому кодированию звуковых сигналов, в частности, но не исключительно, речевого сигнала, принимая во внимание передачу и синтез этого звукового сигнала. Более конкретно, настоящее изобретение относится к способам классификации сигналов и выбора скорости, предназначенным для кодирования речи с переменной скоростью в битах (VBR-кодирования речи).

Предшествующий уровень техники

Спрос на эффективные цифровые методы узкополосного и широкополосного кодирования речи с приемлемым компромиссом между субъективным качеством и скоростью в битах растет в различных областях применения, таких как организация телеконференций, мультимедиа и беспроводная связь. Вплоть до недавнего времени полосу пропускания телефонной связи, ограниченную диапазоном 200-3400 Гц, использовали в основном для приложений кодирования речи. Однако широкополосные речевые приложения обеспечивают повышенную разборчивость и естественность связи по сравнению с обычной полосой пропускания телефонной связи. Обнаружено, что полоса пропускания в диапазоне 50-7000 Гц является достаточной для предоставления надлежащего качества обслуживания, создающего впечатление общения лицом к лицу. В случае обычных аудиосигналов эта полоса пропускания дает приемлемое субъективное качество, но все же более низкое, чем качество средств радиосвязи с частотной модуляцией или звучания компакт-дисков, которые работают в диапазонах 20-16000 Гц и 20-20000 Гц соответственно.

Речевой кодер преобразует речевой сигнал в цифровой поток битов, который передается по каналу связи или хранится на носителе информации. Речевой сигнал преобразуется в цифровую форму, то есть дискретизируется и квантуется, обычно с 16 битами на выборку. Речевой кодер выполняет функцию представления этих цифровых выборок меньшим количеством битов, поддерживая при этом надлежащее субъективное качество речи. Речевой декодер или синтезатор обрабатывает переданный или сохраненный поток битов и преобразует его обратно в звуковой сигнал.

Хорошо известным путем достижения надлежащего компромисса между субъективным качеством и скоростью в битах является кодирование методом линейного предсказания с кодовым возбуждением по коду (CELP). Этот метод кодирования является основой нескольких стандартов кодирования речи как в беспроводных, так и в проводных приложениях. В случае CELP-кодирования дискретизированный речевой сигнал обрабатывается в последовательных блоках выборок длительностью L , обычно называемых кадрами, где L - предварительно определенное число, в типичном случае соответствующее 10-30 мс. Фильтр линейного предсказания (LP-фильтр) рассчитывается и передается в каждом кадре. Расчет LP-фильтра, как правило, требует упреждающего сегмента речи длительностью 5-15 мс из последующего кадра. Кадр выборки длительностью L делится на меньшие блоки, называемый субкадрами. Количество субкадров обычно составляет три или четыре, что приводит к получению субкадров длительностью 4-10 мс. В каждом субкадре обычно получается сигнал возбуждения из двух компонентов - прошлого возбуждения и нового возбуждения - по фиксированной кодовой книге. Компонент, образуемый из прошлого возбуждения, часто называют возбуждением по адаптивной кодовой книге или возбуждением основного тона. Параметры, характеризующие сигнал возбуждения, кодируются и передаются в декодер, где восстановленный сигнал возбуждения используется в качестве входного для LP-фильтра.

В беспроводных системах, использующих технологию множественного доступа с кодовым разделением каналов (CDMA), применение управляемого источником кодирования речи с переменной скоростью в битах (VBR-кодирования речи) значительно увеличивает пропускную способность системы. При управляемом источником VBR-кодировании речи кодер работает на нескольких скоростях в битах, а для определения скорости в битах, используемой при кодировании каждого речевого кадра на основании характера этого речевого кадра (например, вокализованного, невокализованного,

переходного, фонового шума), используется модуль выбора скорости. Целью является достижение наивысшего качества речи при некоторой заданной средней скорости в битах, называемой также средней скоростью передачи данных (ADR). Кодек может работать в разных режимах за счет настройки модуля выбора скорости на достижение разных ADR в разных режимах, при этом рабочая характеристика кодека улучшается при повышенных ADR. Режим работы задается системой в зависимости от канальных условий. Это обеспечивает кодек механизмом достижения компромисса между качеством речи и пропускной способностью системы.

В типичном случае при VBR-кодировании для CDMA-систем используется одна восьмая скорости для кодирования кадров без речевой активности (т.е. кадров пауз или только шума). Когда кадр является стационарно вокализованным или стационарно невокализованным, используется половинная скорость или четвертная скорость, в зависимости от рабочего режима. Если можно использовать половинную скорость, то применяется CELP-модель без кодовой книги основных тонов в случае невокализованного кадра, а в случае вокализованного кадра применяется модификация сигнала для повышения периодичности и уменьшения количества битов для индексов основного тона. Если рабочий режим обуславливает четвертную скорость, согласование сигналов обычно невозможно, поскольку количество битов оказывается недостаточным, обычно применяется параметрическое кодирование. Полная скорость используется для начальных кадров, переходных кадров и смешанных вокализованных кадров (как правило, используется типичная CELP-модель). В дополнение к управляемой источником работе кодека в CDMA-системах система может ограничивать максимальную скорость в битах в некоторых речевых кадрах для посылки информации внутриволновой сигнализации (называемой размерно-пакетной сигнализацией) или во время плохих канальных условий (например, вблизи границ ячейки), чтобы повысить робастность кодека. Эта скорость упоминается как максимум, равный половинной скорости. Когда модуль выбора скорости выбирает кодируемый кадр как кадр полной скорости, а система накладывает ограничение, например, предписывая кадр половинной скорости (HR-кадра), рабочая характеристика речи ухудшается, потому что специализированные режимы половинной скорости (HR-режимы) не могут обеспечить эффективное кодирование начальных сигналов и переходных сигналов. Для обработки в этих специальных случаях можно предусмотреть другую модель кодирования с половинной скоростью (HR) (или с четвертной скоростью (QR)).

Как можно заметить из вышеизложенного описания, классификация сигналов и определение скорости очень важны для эффективного VBR-кодирования. Выбор скорости является ключевой частью достижения наименьшей средней скорости передачи данных с наилучшим возможным качеством.

Задачи изобретения

Задача настоящего изобретения в целом состоит в том, чтобы разработать усовершенствованные способы классификации сигналов и выбора скорости для широкополосного кодирования речи с переменной скоростью, в частности в том, чтобы разработать усовершенствованные способы классификации сигналов и выбора скорости для многорежимного широкополосного кодирования речи с переменной скоростью, пригодного для CDMA-систем.

Сущность изобретения

Применение управляемого источником VBR-кодирования речи значительно увеличивает пропускную способность многих систем связи, особенно беспроводных систем, использующих технологию CDMA. При управляемом источником VBR-кодировании речи кодер может работать на нескольких скоростях в битах, а для определения скорости в битах, используемой при кодировании каждого речевого кадра на основе характера этого речевого кадра (например, вокализованного, невокализованного, переходного, фонового шума), используется модуль выбора скорости. Целью является достижение наивысшего качества речи при некоторой заданной средней скорости передачи данных. Кодек может

работать в разных режимах за счет настройки модуля выбора скорости на достижение разных ADR в разных режимах, при этом рабочая характеристика кодека улучшается при повышенных ADR. В некоторых системах режим работы задается системой в зависимости от канальных условий. Это обеспечивает кодек механизмом достижения компромиссов между качеством речи и пропускной способностью системы.

Алгоритм классификации сигналов анализирует входной речевой сигнал и классифицирует каждый речевой кадр с отнесением последнего к одному из набора предварительно определенных классов (например, к фоновому шуму, вокализованным, невокализованным, смешанным вокализованным, переходным кадрам и т.п.). Алгоритм выбора скорости принимает решение, какую скорость в битах и какую модель кодирования следует использовать, на основе класса речевого кадра и желаемой средней скорости передачи данных.

При многорежимном VBR-кодировании различные рабочие режимы, соответствующие различным средним скоростям в битах, получают путем определения процентной доли использования отдельных скоростей в битах. Таким образом, алгоритм выбора скорости принимает решение, какая скорость в битах должна использоваться для некоторого речевого кадра, на основе характера речевого кадра (классификационной информации) и требуемой средней скорости передачи данных.

В некоторых конкретных вариантах осуществления принимаются во внимание три рабочих режима: высококачественный, стандартный и экономичный режимы, обсуждаемые в первоисточнике [7]. Высококачественный режим гарантирует самое высокое достижимое качество при использовании самой высокой ADR. Экономичный режим максимизирует пропускную способность системы путем использования самой низкой ADR, по-прежнему гарантируя высококачественную широкополосную речь. Стандартный режим представляет собой компромисс между пропускной способностью системы и качеством речи, и в этом режиме используется ADR, значение которой заключено между значениями ADR для высококачественного и экономичного режимов.

Многорежимный широкополосный кодек с переменной скоростью в битах, предложенный для работы в системах стандартов CDMA-один и CDMA2000, именуется в нижеследующем тексте VMR-WB-кодеком.

Более конкретно, в соответствии с первым аспектом настоящего изобретения предложен способ цифрового кодирования звука, заключающийся в том, что

- i) обеспечивают кадр сигнала из дискретизированной версии звука,
- ii) определяют, является ли кадр сигнала активным речевым кадром или неактивным речевым кадром,

iii) если кадр сигнала является неактивным речевым кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования фонового шума с низкой скоростью в битах,

iv) если кадр сигнала является активным речевым кадром, то определяют, является ли активный речевой кадр невокализованным кадром или нет,

v) если кадр сигнала является невокализованным кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования невокализованных сигналов, а

vi) если кадр сигнала не является невокализованным кадром, то определяют, является кадр сигнала устойчивым вокализованным кадром или нет,

vii) если кадр сигнала является устойчивым вокализованным кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования устойчивых вокализованных сигналов,

viii) если кадр сигнала не является невокализованным кадром и кадр сигнала не является устойчивым вокализованным кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования родовых сигналов.

В соответствии со вторым аспектом настоящего изобретения также предложен способ цифрового кодирования звука, заключающийся в том, что

- i) обеспечивают кадр сигнала из дискретизированной версии звука,

ii) определяют, является ли кадр сигнала активным речевым кадром или неактивным речевым кадром,

iii) если кадр сигнала является неактивным речевым кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования фонового шума с низкой скоростью в битах,

iv) если кадр сигнала является активным речевым кадром, то определяют, является ли активный речевой кадр невокализованным кадром или нет,

v) если кадр сигнала является невокализованным кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования невокализованных сигналов, а

vi) если кадр сигнала не является невокализованным кадром, то кодируют этот кадр сигнала с использованием алгоритма кодирования родовой речи.

В соответствии с третьим аспектом настоящего изобретения предложен способ классификации невокализованных сигналов, при осуществлении которого используют, по меньшей мере, три из следующих параметров для классификации невокализованного кадра:

а) меру (\bar{x}) звучания,

б) меру (e_t) спектрального наклона,

в) вариацию (dE) энергии в пределах кадра сигнала и

г) относительную энергию (E_{rel}) кадра сигнала.

Способы, соответствующие настоящему изобретению, обеспечивают создание VBR-кодеков, способных эффективно работать в беспроводных системах, основанных на технологии множественного доступа с кодовым разделением каналов (CDMA), а также систем на основе Internet-протоколов (IP).

И, наконец, в соответствии с четвертым аспектом настоящего изобретения предложено устройство для кодирования звукового сигнала, содержащее

речевой кодер для приема преобразованного в цифровую форму звукового сигнала, отображающего упомянутый звуковой сигнал, причем преобразованный в цифровую форму звуковой сигнал включает в себя, по меньшей мере, один кадр сигнала, а речевой кодер включает в себя

классификатор первого уровня для различения между активными и неактивными речевыми кадрами,

генератор комфортного шума для кодирования неактивных речевых кадров,

классификатор второго уровня для различения между вокализованными и невокализованными кадрами,

кодер невокализованной речи,

классификатор третьего уровня для различения между устойчивыми и неустойчивыми вокализованными кадрами,

оптимизированный кодер вокализованной речи и

кодер родовой речи,

при этом речевой кодер конфигурирован с обеспечением возможности выдачи двоичного представления параметров кодирования.

Вышеуказанные и другие задачи, преимущества и признаки настоящего изобретения поясняются в нижеследующем неограничительном описании иллюстративных вариантов осуществления изобретения, приводимых лишь в качестве примера, со ссылками на прилагаемые чертежи.

Краткое описание чертежей

Фиг.1 - блок-схема системы речевой связи, иллюстрирующая применение устройств кодирования и декодирования речи в соответствии с первым аспектом настоящего изобретения;

фиг.2 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в соответствии с первым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

фиг.3 - схема последовательности операций, иллюстрирующая способ различения

невокализованного кадра в соответствии с иллюстративным вариантом осуществления третьего аспекта настоящего изобретения,

фиг.4 - схема последовательности операций, иллюстрирующая способ различения устойчивого вокализованного кадра в соответствии с иллюстративным вариантом

5 осуществления четвертого аспекта настоящего изобретения,

фиг.5 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в высококачественном режиме в соответствии со вторым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

10 фиг.6 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в стандартном режиме в соответствии с третьим иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

фиг.7 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в экономичном режиме в соответствии с четвертым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

15 фиг.8 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в режиме, обеспечивающем возможность взаимодействия, в соответствии с пятым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

20 фиг.9 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в высококачественном или стандартном режиме во время работы на максимуме, равном половинной скорости, в соответствии с шестым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

25 фиг.10 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в экономичном режиме во время работы на максимуме, равном половинной скорости, в соответствии с седьмым иллюстративным вариантом осуществления второго аспекта настоящего изобретения,

30 фиг.11 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала в режиме, обеспечивающем возможность взаимодействия, во время работы на максимуме, равном половинной скорости, в соответствии с восьмым иллюстративным вариантом осуществления второго аспекта настоящего изобретения, и

35 фиг.12 - схема последовательности операций, иллюстрирующая способ цифрового кодирования звукового сигнала с обеспечением взаимодействия между многорежимным широкополосным кодеком с переменной скоростью в битах (VMR-WB-кодеком) и адаптивным многорежимным широкополосным кодеком (AMR-WB-кодеком) в соответствии с иллюстративным вариантом осуществления пятого аспекта настоящего изобретения.

Подробное описание изобретения

На фиг.1 изображена система 10 речевой связи, демонстрирующая применение кодирования и декодирования речи в соответствии с иллюстративным вариантом осуществления первого аспекта настоящего изобретения. Система 10 речевой связи 40 поддерживает передачу и воспроизведение речевого сигнала по каналу 12 связи. Канал 12 связи может представлять собой провод, оптическую либо волоконную линию или радиочастотную линию. Канал 12 связи также может представлять собой комбинацию различных сред передачи, например быть частично волоконной линией, а частично - радиочастотной линией. Радиочастотная линия может обеспечивать поддержку множества 45 одновременных речевых передач, требующих наличия совместно используемых ресурсов полосы пропускания, например, в сотовой телефонии. В качестве альтернативы, канал связи может быть заменен запоминающим устройством (не показано) в варианте осуществления системы связи с одним таким устройством, которое обеспечивает запись и хранение кодированного речевого сигнала для последующего воспроизведения.

50 Система 10 связи включает в себя кодирующее устройство, состоящее из микрофона 14, аналого-цифрового преобразователя 16, речевого кодера 18 и канального кодера 20 на излучающей (передающей) стороне канала 12 связи, а также канальный декодер 22, речевой декодер 24, цифроаналоговый преобразователь 26 и динамик 28 на принимающей

стороне.

Микрофон 14 вырабатывает аналоговый речевой сигнал, который подается в аналогово-цифровой (АЦ) преобразователь 16 для преобразования этого сигнала в цифровую форму. Речевой кодер 18 кодирует преобразованный в цифровую форму речевой сигнал, вырабатывая набор параметров, которые кодируются с получением их двоичного представления и подаются в канальный кодер 20. Устанавливаемый по выбору канальный кодер 20 вносит избыточность в двоичное представление параметров кодирования перед передачей их по каналу 12 связи. Кроме того, в некоторых приложениях, таких как приложения в пакетных сетях, осуществляется пакетирование кодированных кадров перед передачей.

На принимающей стороне канальный декодер 22 использует избыточную информацию в принимаемом потоке битов для обнаружения и исправления канальных ошибок, возникающих при передаче. Речевой декодер 24 преобразует поток битов, принимаемый из канального декодера 20, обратно в набор параметров кодирования, чтобы создать синтезированный речевой сигнал. Синтезированный речевой сигнал, восстановленный в речевом декодере 24, преобразуется в аналоговую форму в цифроаналоговом (ЦА) преобразователе 26 и воспроизводится в блоке 28 динамика.

Микрофон 14 и/или АЦ преобразователь 16 могут быть заменены в некоторых конкретных вариантах осуществления другими источниками речи для речевого кодера 18.

Кодер 20 и декодер 22 конфигурированы с обеспечением возможности воплощения способа кодирования речевого сигнала в соответствии с настоящим изобретением, как описано ниже.

Классификация сигналов

На фиг.2 проиллюстрирован способ 100 цифрового кодирования речевого сигнала в соответствии с первым иллюстративным вариантом осуществления первого аспекта настоящего изобретения. Способ 100 включает в себя классификацию речевых сигналов в соответствии с иллюстративным вариантом осуществления второго аспекта настоящего изобретения. Следует отметить, что выражение "речевой сигнал" относится к голосовым сигналам, а также к любому мультимедийному сигналу, который может включать в себя звучащую часть, например аудиочасть с речевым содержанием (речь между фрагментами музыки, речь с фоновой музыкой, речь со специальными звуковыми эффектами и т.д.).

Как показано на фиг.2, классификацию сигналов проводят в три этапа - 102, 106 и 110, на каждом из которых обеспечивают различение конкретного класса сигналов. Сначала, на этапе 102, классификатор первого уровня в форме детектора речевой активности (VAD) (не показан) осуществляет различение между активными и неактивными речевыми кадрами. Если обнаруживается неактивный речевой кадр, то способ 100 кодирования оканчивается кодированием текущего кадра, например, посредством генерирования комфортного шума (CNG) (этап 104). Если на этапе 102 обнаруживается активный речевой кадр, то этот кадр подвергается воздействию классификатора второго уровня (не показан), конфигурированного с обеспечением возможности различения невокализованных кадров. Если классификатор на этапе 106 классифицирует кадр как невокализованной речевой сигнал, то способ 100 кодирования оканчивается на этапе 108, где кадр кодируется способом кодирования, оптимизированным для невокализованных сигналов. В противном случае речевой кадр на этапе 110 пропускается через классификатор третьего уровня (не показан) в форме модуля классификации "устойчивых вокализованных" сигналов (не показан). Если текущий кадр классифицируется как устойчивый вокализованный кадр, то этот кадр кодируется способом кодирования, оптимизированным для устойчивых вокализованных сигналов (этап 112). В противном случае кадр, вероятно, содержит неустойчивый речевой сегмент, например вокализованный начальный сигнал или часть быстро эволюционирующего сигнала, и этот кадр кодируется с помощью речевого кодера общего назначения с высокой скоростью в битах, позволяющей поддерживать надлежащее субъективное качество (этап 114). Отметим, что если относительная энергия кадра ниже, чем некоторый порог, то эти кадры

можно кодировать общим способом кодирования с более низкой скоростью, чтобы дополнительно уменьшить среднюю скорость передачи данных.

Классификаторы и кодеры могут принимать многие формы - от электронных схем до однокристалльного процессора.

- 5 В нижеследующем описании приведено более подробное пояснение классификации различных типов речевого сигнала, а также описаны способы классификации невокализованной и вокализованной речи.

Различение неактивных речевых кадров (с помощью детектора речевой активности)

- 10 Различение неактивных речевых кадров осуществляется на этапе 102 с помощью детектора речевой активности (VAD). Схема VAD хорошо известна специалисту в данной области техники, так что более подробное описание ее здесь будет опущено. Пример VAD описан в работе [5].

Различение невокализованных активных речевых кадров

- 15 Невокализованные части речевого сигнала характеризуются отсутствием периодичности и могут быть дополнительно подразделены на неустойчивые кадры, в которых энергия и спектр претерпевают быстрое изменение, и устойчивые кадры, в которых упомянутые характеристики остаются относительно неизменными.

На этапе 106 различение невокализованных кадров осуществляется с использованием, по меньшей мере, трех из следующих параметров:

- 20 меры звучания, которую можно вычислить как усредненную нормализованную корреляцию (\bar{r}_x),

меры (e_t) спектрального наклона,

отношения (dE) энергии сигнала, используемого для оценки вариации энергии кадра в пределах кадра, а значит - и устойчивости кадра, и относительной энергии кадра.

- 25 Мера звучания

На фиг.3 иллюстрируется способ 200 различения невокализованного кадра в соответствии с иллюстративным вариантом третьего аспекта настоящего изобретения.

Нормализованная корреляция, используемая для определения меры звучания, вычисляется как часть модуля 214 исследования основного тона при разомкнутом контуре.

- 30 В иллюстративном варианте осуществления согласно фиг.3 используются кадры длительностью 20 мс. Модуль исследования основного тона при разомкнутом контуре выдает оценку r основного тона при разомкнутом контуре каждые 10 мс (дважды за кадр).

При осуществлении способа 200 этот модуль также используется для выдачи мер r_x

- 35 нормализованных корреляций. Эти нормализованные корреляции вычисляются по взвешенной речи и по прошлой взвешенной речи с задержкой основного тона при разомкнутом контуре.

Взвешенный речевой сигнал $s_w(n)$ вычисляется в перцепционном

взвешивающем фильтре 212. В этом иллюстративном варианте осуществления используют

перцепционный взвешивающий фильтр 212 с фиксированным знаменателем, пригодный

для широкополосных сигналов. Нижеследующее соотношение представляет собой пример

- 40 передаточной функции для перцепционного взвешивающего фильтра 212:

$$W(z) = A(z/\gamma_1)/(1-\gamma_2 z^{-1}), \text{ где } 0 < \gamma_2 < \gamma_1 \leq 1$$

где $A(z)$ - передаточная функция фильтра с линейным предсказанием (LP-фильтра), вычисляемая в модуле 218, которая задается следующим соотношением:

45
$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

Мера звучания задается средней корреляцией \bar{r}_x , которая определяется как

$$\bar{r}_x = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2)) \quad (1)$$

- 50

где $r_x(0)$, $r_x(1)$ и $r_x(2)$ соответственно представляют собой нормализованную корреляцию первой половины текущего кадра, нормализованную корреляцию второй половины текущего кадра и нормализованную корреляцию упреждающей выборки (начала

следующего кадра).

К нормализованной корреляции в уравнении (1) можно прибавить поправку r_e на шум, чтобы учесть присутствие фонового шума. В присутствии фонового шума средняя нормализованная корреляция уменьшается. Вместе с тем, применительно к классификации 5 сигналов, это уменьшение не оказывает влияния на распознавание вокализованных и невокализованных кадров, так что упомянутое уменьшение компенсируется введением r_e . Следует отметить, что в случае использования надлежащего алгоритма снижения шума величина r_e является практически нулевой.

При осуществлении способа 200 используется упреждающая выборка длительностью 13 10 мс. Нормализованная корреляция $r_x(k)$ вычисляется следующим образом:

$$r_x(k) = \frac{r_{xy}}{\sqrt{r_{xx} \cdot r_{yy}}}, \quad (2)$$

15 где

$$\begin{aligned} r_{xy} &= \sum_{i=0}^{L_k-1} x(t_k + i)x(t_k + i - p_k) \\ r_{xx} &= \sum_{i=0}^{L_k-1} x^2(t_k + i) \\ r_{yy} &= \sum_{i=0}^{L_k-1} x^2(t_k + i - p_k) \end{aligned}$$

При осуществлении способа 200 вычисление корреляций происходит следующим 25 образом. Корреляции $r_x(k)$ вычисляются по взвешенному речевому сигналу $s_w(n)$. Моменты t_k относятся к началу текущего полукадра и составляют 0, 128 и 256 выборок соответственно для $k = 0, 1$ и 2 при частоте дискретизации 12800 Гц. Значения $p_k = T_{OL}$ представляют собой выбранные оценки основного тона при разомкнутом контуре для полукадров. Протяженность L_k вычисления автокорреляции зависит от периода основного 30 тона. Сводка значений L_k в первом варианте осуществления приведена ниже (для частоты дискретизации, составляющей 12,8 кГц):

$L_k = 80$ выборок для $p_k \leq 62$ выборки;

$L_k = 124$ выборки для $62 \text{ выборки} < p_k \leq 122$ выборки;

$L_k = 230$ выборок для $p_k > 122$ выборки.

35 Эти протяженности предполагают, что длина коррелированного вектора включает в себя, по меньшей мере, один период основного тона, что способствует робастному обнаружению основного тона в разомкнутом контуре. При длинных периодах основного тона ($p_1 > 122$ выборки) $r_x(1)$ и $r_x(2)$ идентичны, т.е. вычисляется только одна корреляция, поскольку коррелированные векторы достаточно длинны для того, чтобы анализ по 40 упреждающей выборке оказался больше ненужным.

В альтернативном варианте взвешенный речевой сигнал можно подвергнуть децимации вдвое, чтобы упростить поиск основного тона в разомкнутом контуре. Взвешенный речевой сигнал можно подвергнуть фильтрации нижних частот перед децимацией. В этом случае значения L_k задаются следующим образом:

45 $L_k = 40$ выборок для $p_k \leq 31$ выборке;

$L_k = 62$ выборки для $31 \text{ выборки} < p_k \leq 61$ выборке;

$L_k = 115$ выборок для $p_k > 61$ выборки.

Для вычисления корреляций можно использовать и другие способы. Например, можно 50 вычислять всего одно значение нормализованной корреляции для всего кадра вместо усреднения нескольких нормализованных корреляций. Кроме того, корреляции можно вычислять по сигналам, не являющимся взвешенными речевыми, например, по остаточному сигналу, речевому сигналу, или остаточному речевому сигналу, или взвешенному речевому сигналу, подвергнутому фильтрации нижних частот.

Спектральный наклон

Параметр спектрального наклона содержит информацию о частотном распределении энергии. При осуществлении способа 200 спектральный наклон оценивают в частотной области как отношение между энергией, сконцентрированной на низких частотах, и энергией, сконцентрированной на высоких частотах. Вместе с тем спектральный наклон можно оценивать и по-другому, например, как отношение между двумя первыми коэффициентами автокорреляции речевого сигнала.

При осуществлении способа 200 используют дискретное преобразование Фурье для осуществления спектрального анализа в модуле 210, показанном на фиг.10. Частотный анализ и вычисление наклона проводятся дважды за кадр. Используют 256-точечное быстрое преобразование Фурье (БПФ) с перекрытием 50 процентов. Окна анализа располагают так, что используется вся упреждающая выборка. В начале первого окна находятся 24 выборки, следующие после начала текущего кадра. Во втором окне находятся еще 128 выборок. Можно использовать разные окна для взвешивания входного сигнала в целях частотного анализа. Используется корень квадратный из (ширины) окна Хэмминга (что эквивалентно синусоидальному окну). В частности, это окно весьма подходит для способов с перекрытием и суммированием, вследствие чего этот конкретный спектральный анализ можно использовать в реализуемом по выбору алгоритме подавления шумов на основании спектрального вычитания и анализа/синтеза с перекрытием и суммированием. Поскольку такие алгоритмы подавления шумов считаются хорошо известными в данной области техники, более подробное описание их здесь опущено.

Энергия на высоких частотах и на низких частотах вычисляется в следующих перцепционных критических полосах [6]:

Критические полосы = {100,0, 200,0, 300,0, 400,0, 510,0, 630,0, 770,0, 920,0, 1080,0, 1270,0, 1480,0, 1720,0, 2000,0, 2320,0, 2700,0, 3150,0, 3700,0, 4400,0, 5300,0, 6350,0} Гц.

Энергия на высоких частотах вычисляется как среднее значение энергий последних двух критических полос:

$$\bar{E}_h = 0,5(E_{CB}(18) + E_{CB}(19)),$$

где $E_{CB}(i)$ - средние энергии, приходящиеся на критическую полосу, вычисленные как

$$E_{CB}(i) = \frac{1}{N_{CB}(i)} \sum_{k=0}^{N_{CB}(i)-1} (X_R^2(k + j_i) + X_I^2(k + j_i)), \quad i=0, \dots, 19$$

где $N_{CB}(i)$ - количество интервалов дискретизации по частоте в i -той полосе, а $X_R(k)$ и $X_I(k)$ - соответственно действительная и мнимая части k -того интервала дискретизации по частоте и j_i - индекс первого интервала дискретизации в i -той критической полосе.

Энергия на низких частотах вычисляется как средняя из энергий в первых 10-ти критических полосах. Критические полосы средних частот исключены из вычисления для улучшения различения между кадрами с высокой концентрацией энергии на низких частотах (обычно - вокализованными) и высокой концентрацией энергии на высоких частотах (обычно - невокализованными). Между этими частотами запас энергии не характеризует никакие классы и увеличивает путаницу при принятии решения.

Энергия на низких частотах вычисляется по-разному для длинных периодов основного тона и коротких периодов основного тона. Для вокализованных охватывающих сегментов речи используется гармоническая структура спектра для улучшения различения вокализованных и невокализованных кадров. Так, для коротких периодов основного тона (на каждом интервале дискретизации) поинтервально вычисляется E_l , а при суммировании учитываются только интервалы дискретизации по частоте, достаточно близкие к гармоникам речи. То есть

$$\bar{E}_l = \frac{1}{c \pi t} \sum_{k=0}^{24} E_{BIN}(k) W_h(k)$$

где $E_{BIN}(k)$ - энергии интервалов дискретизации в первых 25 интервалах дискретизации по частоте (постоянная составляющая не учитывается). Отметим, что эти 25 интервалов

дискретизации соответствуют первым 10 критическим полосам. При вышеуказанном суммировании учитываются только члены, связанные с интервалами дискретизации, близкими к гармоникам основного тона, так что значение $w_h(k)$ задается равным 1, если расстояние между интервалом дискретизации и ближайшей гармоникой не превышает

5 некоторый порог частоты (50 Гц), и задается равным 0 в противном случае. Подсчет cnt - это количество ненулевых членов при суммировании. Учитываются только интервалы дискретизации, находящиеся ближе, чем в 50 герцах, к ближайшим гармоникам. В данном случае, если структура является гармонической на низких частотах, в сумму будут

10 включаться только члены высоких энергий. С другой стороны, если структура не является гармонической, выбор членов будет случайным, а сумма будет меньше. Таким образом, можно обнаружить даже невокализованные сигналы с высоким запасом энергии на низких частотах. Эту обработку нельзя провести для более длительных периодов основного тона, поскольку разрешение по частоте оказывается недостаточным. Для значений длительности основного тона, превышающих 128, или для априори невокализованных сигналов энергия

15 на низких частотах, приходящаяся на критическую полосу, вычисляется как

$$\bar{E}_l = \frac{1}{10} \sum_{k=0}^9 E_{CB}(k)$$

Априори невокализованные сигналы определяются, когда $r_x(0)+r_x(1)+r_e < 0,6$, где значение r_e - это поправка, прибавляемая к нормализованной корреляции, как описано

20 выше.

Результирующие энергии на низких и высоких частотах получают путем вычитания оцененной энергии шумов из значений E_l и E_h , вычисленных выше. То есть

$$E_h = \bar{E}_h - N_h;$$

$$E_l = \bar{E}_l - N_l;$$

25 где N_h и N_l - усредненные энергии шумов в последних 2-х критических полосах и первых 10-ти критических полосах соответственно. Оцененные энергии шумов прибавлены к результату вычисления наклона, чтобы учесть присутствие фонового шума.

И, наконец, спектральный наклон задается следующим образом:

$$30 \quad e_{\text{tilt}}(i) = \frac{E_l}{E_h}$$

Отметим, что вычисление спектрального наклона проводят дважды за кадр, чтобы получить значения $e_{\text{tilt}}(0)$ и $e_{\text{tilt}}(1)$, соответствующие обоим спектральным анализам за кадр. Средний спектральный наклон, используемый в классификации невокализованных кадров,

35 задают следующим образом:

$$e_t = \frac{1}{3} (e_{\text{old}} + e_{\text{tilt}}(0) + e_{\text{tilt}}(1))$$

где e_{old} - наклон, полученный в результате второго спектрального анализа предыдущего кадра.

40 Вариация энергии, dE

Вариация dE энергии оценивается по речевому сигналу $s(n)$, из которого устранен шум, где $n=0$ соответствует началу текущего кадра. Энергия сигнала оценивается дважды за субкадр, т.е. 8 раз за кадр, на основании кратковременных сегментов длиной по 32 выборки. Кроме того, вычисляется также кратковременные энергии последних 32-х выборок

45 из следующего кадра. Максимальные кратковременные энергии вычисляются как

$$E_{\text{st}}^{(1)}(j) = \max_{i=0}^{31} (s^2(i + 32j)), \quad j = -1, \dots, 8$$

где $j=-1$ и $j=8$ соответствуют концу предыдущего кадра и началу следующего кадра. Еще один набор из 9-ти максимальных энергий вычисляют путем сдвига индексов речи на 16

50 выборок. То есть

$$E_{\text{st}}^{(1)}(j) = \max_{i=0}^{31} (s^2(i + 32j - 16)), \quad j = 0, \dots, 8$$

Максимальная вариация dE энергии между последовательными кратковременными

сегментами вычисляется как максимум следующих выражений:

$$E_{st}^{(1)}(0)/E_{st}^{(1)}(-1), \text{ если } E_{st}^{(1)}(0) > E_{st}^{(1)}(-1)$$

$$E_{st}^{(1)}(7)/E_{st}^{(1)}(8), \text{ если } E_{st}^{(1)}(7) > E_{st}^{(1)}(8)$$

$$5 \quad \frac{\max(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))}{\min(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))} \text{ для } j=1...7$$

$$\frac{\max(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))}{\min(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))} \text{ для } j=1...8$$

10 В альтернативном варианте, можно использовать другие способы, чтобы оценить вариацию энергии в кадре.

Относительная энергия E_{rel}

Относительная энергия кадра задается разностью между энергией кадра, выраженной в децибелах, и долговременной средней энергией кадра. Энергия кадра вычисляется как

$$15 \quad E_t = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \text{ ДБ,}$$

где $E_{CB}(i)$ - средние энергии, приходящиеся на критическую полосу, как описано выше.

Долговременная средняя энергия кадра задается следующим образом:

$$\bar{E}_f = 0,99 \bar{E}_f + 0,01 E_t,$$

20 с начальным значением $\bar{E}_f = 45$ ДБ.

Таким образом, относительная энергия кадра задается следующим образом:

$$E_{rel} = E_t - \bar{E}_f.$$

Относительная энергия кадра используется для того, чтобы идентифицировать кадры с

25 низкой энергией, которые не классифицированы как кадры фонового шума или невокализованные кадры. Эти кадры могут кодироваться общим кодером половинной скорости (общим HR-кодером), чтобы уменьшить среднюю скорость (ADR).

Классификация невокализованной речи

Классификация невокализованных речевых кадров основана на вышеописанных

30 параметрах, а именно мере \bar{x} звучания, мере e_t спектрального наклона, вариации dE энергии в пределах кадра сигнала и относительной энергии E_{rel} кадра сигнала. Решение принимается на основе, по меньшей мере, трех из этих параметров. Пороги решения

устанавливаются на основе рабочего режима (требуемой средней скорости передачи данных). В основном, для рабочих режимов с более низкими желаемыми скоростями

35 передачи данных пороги задаются как более благоприятствующие классификации невокализованной речи (поскольку для кодирования кадра будет использоваться кодирование с половинной скоростью или четвертной скоростью). Невокализованные кадры обычно кодируются кодером невокализованной речи половинной скорости (HR-

кодером). Однако в случае экономичного режима можно использовать кодер

40 невокализованной речи четвертной скорости (QR-кодер), чтобы дополнительно уменьшить ADR, если удовлетворяются некоторые дополнительные условия.

В высококачественном режиме кадр кодируется как невокализованный HR-кадр, если удовлетворяется следующее условие:

$$(\bar{x} < th_1) \text{ и } (e_t < th_2) \text{ и } (dE < th_3),$$

45 где $th_1 = 0,5$, $th_2 = 1$ и

$$th_3 = \begin{cases} -4 & \text{для } \bar{E}_f > 34 \\ 0 & \text{для } 21 < \bar{E}_f \leq 34 \\ 4 & \text{в противном случае} \end{cases}$$

50 Во время принятия решения по речевой активности используется задержка решения.

Так, после периодов активной речи, когда в соответствии с алгоритмом принимается решение, что кадр является неактивным речевым кадром, локальное решение по речевой активности (локальное PPA) устанавливается на нуль, а флаг фактического PPA

устанавливается на нуль только спустя некоторое количество кадров (т.е. период задержки). Это позволяет избежать "обрезания" смещений речи. Как в стандартном, так и в экономичном режиме, если локальное PPA является нулевым, кадр классифицируется как невокализованной кадр.

5 В стандартном режиме кадр кодируется как невокализованный HR-кадр, если локальное PPA равно 0 или если удовлетворяется следующее условие:

$$(\bar{x} < th_4) \text{ и } (e_t < th_5) \text{ и } ((dE < th_6) \text{ или } (E_{rel} < th_7)),$$

где $th_4 = 0,695$, $th_5 = 4$, $th_6 = 40$ и $th_7 = -14$.

10 В экономичном режиме кадр квалифицируется как невокализованный кадр, если локальное PPA равно 0 или если удовлетворяется следующее условие:

$$(\bar{x} < th_8) \text{ и } (e_t < th_9) \text{ и } ((dE < th_{10}) \text{ или } (E_{rel} < th_{11})),$$

где $th_8 = 0,695$, $th_9 = 4$, $th_{10} = 60$ и $th_{11} = -14$.

15 В экономичном режиме невокализованные кадры обычно кодируют как невокализованные HR-кадры. Однако они также могут кодироваться как невокализованные QR-кадры, если удовлетворяются следующие дополнительные условия: если последний кадр является вокализованным кадром или кадром фонового шума и если в конце кадра энергия сконцентрирована на высоких частотах, а на упреждающей выборке не обнаруживается потенциальная вокализованная начальная часть, то этот кадр кодируется как невокализованный QR-кадр. Последние два условия определяются следующим

20 образом:

$$(r_x(2) < th_{12}) \text{ и } (e_{tit}(1) < th_{13}), \text{ где } th_{12} = 0,73, th_{13} = 3.$$

Отметим, что $r_x(2)$ - это нормализованная корреляция на интервале упреждающей выборки, а $e_{tit}(1)$ - это наклон во втором спектральном анализе, охватывающем конец кадра и упреждающую выборку.

25 Конечно, для различения невокализованного кадра можно использовать способы, отличающиеся от способа 200.

Различение устойчивых вокализованных речевых кадров

В случае стандартного и экономичного режимов устойчивые вокализованные кадры могут кодироваться способом кодирования вокализованных HR-сигналов.

30 Кодирование вокализованных HR-сигналов предусматривает использование модификации сигналов для эффективного кодирования устойчивых вокализованных кадров.

35 Методы модификации сигналов обеспечивают настройку основного тона сигнала с достижением предварительно определенного профиля задержки. Затем посредством долговременного предсказания прошлый сигнал возбуждения отображается на текущий субкадр с использованием этого профиля задержки и масштабируется с помощью параметра усиления. Профиль задержки получают непосредственным интерполированием между двумя оценками основного тона в разомкнутом контуре, первая из которых получена в предыдущем кадре, а вторая - в текущем кадре. Интерполяция дает значение задержки кадра для каждого момента времени. После получения профиля задержки основной тон в субкадре, подлежащем кодированию в текущий момент, настраивается для отслеживания

40 профиля задержки посредством деформации, изменения временного масштаба сигнала. В случае прерывистой деформации [1, 4, 5] сегмент сигнала сдвигается влево или вправо без изменения длины сегмента. Прерывистая деформация требует процедуры обработки

45 перекрывающихся или пропущенных частей результирующих сигналов. Для уменьшения артефактов в этих операциях допустимое изменение временного масштаба поддерживается малым. Кроме того, деформация обычно осуществляется с использованием остаточного сигнала LP-фильтра или взвешенного речевого сигнала, чтобы уменьшить получаемые искажения. Использование этих сигналов вместо речевого

50 сигнала облегчает обнаружение импульсов основного тона и областей малой мощности между ними, а значит - и определение сегментов сигнала для деформации. Фактический модифицированный речевой сигнал формируется путем инверсной фильтрации. После осуществления модификации сигнала для данного субкадра кодирование может

производиться обычным образом, за исключением того, что возбуждение по адаптивной кодовой книге формируется с использованием предварительно определенного профиля задержки.

5 В рассматриваемом иллюстративном варианте осуществления модификация сигналов по основному тону и кадру осуществляется одновременно, то есть адаптируя один сегмент
цикла основного тона в некоторый момент времени в текущем кадре таким образом, что
10 последующий речевой кадр начинается при полном временном согласовании с исходным сигналом. Сегменты цикла основного тона ограничены границами кадра. Это предотвращает перенос временного сдвига через границы кадра, упрощая реализацию
кодера и уменьшая риск артефактов в модифицированном речевом сигнале. Это также
упрощает работу с переменной скоростью в битах при переходах между типами
кодирования с разрешенной и заблокированной модификацией сигналов, поскольку каждый
новый кадр начинается во временном согласовании с исходным сигналом.

15 Как показано на фиг.2, если кадр не классифицирован ни как неактивный речевой кадр, ни как невокализованный кадр, то проверяют, является ли он устойчивым вокализованным кадром (этап 110). Классификация устойчивых вокализованных кадров выполняется с использованием метода разомкнутого контура в связи с процедурой модификации сигналов, используемой для кодирования устойчивых вокализованных кадров.

20 На фиг.4 проиллюстрирован способ 300 для различения устойчивого вокализованного кадра в соответствии с иллюстративным вариантом осуществления четвертого аспекта настоящего изобретения.

Субпроцедуры, выполняемые при модификации сигналов, обеспечивают получение показателей, количественно характеризующих достижимую рабочую характеристику
25 долговременного предсказания в текущем кадре. Если любой из этих показателей оказывается вне допустимых для него пределов, то процедура модификации сигналов завершается одним из логических блоков. В этом случае, исходный сигнал сохраняется нетронутым, а кадр не классифицируется как устойчивый вокализованный кадр. Эта комплексная логика обеспечивает максимизацию качества модифицированного речевого сигнала после модификации сигнала и кодирования с низкой скоростью в битах.

30 Процедура поиска импульсов основного тона на этапе 302 обеспечивает получение нескольких показателей исходя из периодичности основного кадра. Поэтому логический блок, следующий за этим этапом, является важным компонентом логики классификации. Наблюдается эволюция длительности цикла основного тона. Логический блок сравнивает
35 расстояние до обнаруженных позиций импульсов основного тона с интерполированной оценкой основного тона в разомкнутом контуре, а также с расстоянием до ранее обнаруженных импульсов основного тона. Процедура модификации сигналов завершается, если различие с оценкой основного тона в разомкнутом контуре или с длительностями предыдущего цикла основного тона оказывается слишком большой.

40 Выбор профиля задержки на этапе 304 дает дополнительную информацию об эволюции циклов основного тона и периодичности текущего речевого кадра. Процедура модификации сигналов продолжается с этого блока, если удовлетворяется условие $|d_n - d_{n-1}| < 0,2d_n$, где d_n и d_{n-1} - задержки основного тона в текущем и прошлом кадрах. По существу, это означает, что при классификации текущего кадра как устойчиво вокализованного допускается лишь малое изменение задержки.

45 Когда кадры, подвергнутые модификации сигналов, кодируются с низкой скоростью передачи в битах, форма сегментов цикла основного тона поддерживается одной и той же на протяжении кадра, чтобы обеспечить точное моделирование сигнала посредством долговременного предсказания, а значит - и кодирование с низкой скоростью передачи в битах без ухудшения субъективного качества. При модификации сигналов на этапе 306
50 сходство последовательных сегментов можно количественно охарактеризовать посредством нормализованной корреляции между текущим сегментом и целевым сигналом при оптимальном сдвиге. Сдвиг сегментов цикла основного тона, максимизирующий их корреляцию с целевым сигналом, повышает периодичность и дает значительный выигрыш

от долговременного предсказания, если используется модификация сигналов. Успех этой процедуры гарантируется требованием, в соответствии с которым все значения корреляции должны быть больше, чем предварительно определенный порог. Если это условие не выполняется для всех сегментов, процедура модификации сигналов завершается, а

5 исходный сигнал сохраняется нетронутым. В общем случае несколько меньший диапазон порогов допустим для мужских голосов при одинаковой эффективности кодирования. Пороги усиления можно изменять в разных рабочих режимах VBR-кодека, чтобы регулировать использование режимов кодирования, в которых применяется модификация сигналов, следовательно, изменять целевую среднюю скорость в битах.

10 Как описано выше, вся логика выбора скорости передачи в соответствии со способом 100 предусматривает три этапа, каждый из которых обеспечивает различие конкретного класса сигналов. Один из этапов включает в себя алгоритм модификации сигналов в качестве неотъемлемой части. Во-первых, VAD осуществляет различие между активными и неактивными речевыми кадрами. Если обнаруживается неактивный речевой

15 кадр, то способ классификации заканчивается, так как кадр рассматривается как кадр фонового шума и кодируется, например, с помощью генератора комфортного шума. Если обнаруживается активный речевой кадр, то этот кадр подвергается обработке на втором этапе, предназначенном для различения невокализованных кадров. Если кадр классифицируется как невокализованной речевой сигнал, то цепочка классификации

20 заканчивается, а кадр кодируется в режиме кодирования, предназначенном для невокализованных кадров. На последнем этапе речевой кадр обрабатывают посредством предложенной процедуры модификации сигналов, которая разрешает модификацию, если соблюдаются условия, описанные ранее в этом подразделе. В таком случае кадр классифицируется как устойчивый речевой кадр, основной тон исходного сигнала

25 настраивается в соответствии с определенным контуром искусственно создаваемой задержки, а кадр кодируется с использованием конкретного режима, оптимизированного для этих типов кадров. В противном случае кадр, вероятно, содержит неустойчивый речевой сегмент, например вокализованный начальный или быстро эволюционирующий речевой сигнал. Эти кадры, как правило, требуют более обобщенной модели кодирования.

30 Эти кадры обычно кодируются с использованием FR-кодирования общего типа. Вместе с тем, если относительная энергия кадра меньше, чем некоторый порог, то эти кадры можно кодировать посредством HR-кодирования общего типа, чтобы дополнительно уменьшить ARD.

Кодирование речи и выбор скорости передачи для многорежимных VBR-систем, использующих технологию CDMA (CDMA-VBR-систем)

35 Ниже описаны способы выбора скорости передачи и цифрового кодирования звука для звуковых многорежимных CDMA-VBR-систем, которые могут работать в оболочке Rate Set II, в соответствии с иллюстрируемыми вариантами осуществления настоящего изобретения.

40 Описанный кодек основан на адаптивном многоскоростном широкополосном (AMR-WB) речевом кодеке, который недавно выбран Сектором стандартизации электросвязи Международного союза электросвязи (ITU-T) для нескольких вариантов широкополосных речевых услуг, а группой "Проект партнерства в создании систем третьего поколения" (3GPP) - для беспроводных систем третьего поколения, работающих в соответствии со

45 стандартами Глобальной системы мобильной электросвязи (GSM) и широкополосного множественного доступа с кодовым разделением каналов (W-CDMA). AMR-WB-кодек предусматривает девять скоростей в битах, а именно 6,6, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 и 23,85 кбит/с. Управляемый источником VBR-кодек на основе AMR-WB-кодека, предназначенный для CDMA-систем, обеспечивает возможность взаимодействия

50 между CDMA-системами и другими системами, использующими AMR-WB-кодек. В качестве общей скорости между широкополосным VBR-кодеком CDMA-системы и AMR-WB-кодеком можно использовать скорость в битах AMR-WB-кодека, составляющую 12,65 кбит/с и являющуюся той скоростью, которая лучше всего согласуется с полной скоростью,

составляющей 13,3 кбит/с, в оболочке Rate Set II, что обеспечит возможность взаимодействия не требуя преобразования кода (которое ухудшает качество речи). Специально для решения проблемы широкополосного VBR-кодирования в CDMA-системах разработаны типы кодирования с более низкими скоростями, чтобы обеспечить эффективную работу в оболочке Rate Set II. В таком случае кодек может работать в нескольких CDMA-специфичных режимах, используя все скорости, но при этом будет иметь режим, который гарантирует взаимодействие с системами, в которых используется AMR-WB-кодек.

Способы кодирования в соответствии с вариантами осуществления настоящего изобретения сведены в таблицу 1, в которой они все будут именоваться типами кодирования.

Типы кодирования, используемого в иллюстративных вариантах осуществления, с соответствующими скоростями в битах:			
Тип кодирования	Скорость в битах [кбит/с]	Количество битов на кадр длительностью 20 мс	
Общий FR	13,3	266	
FR с обеспечением возможности взаимодействия	13,3	266	
HR вокализованных сигналов	6,2	124	
HR невокализованных сигналов	6,2	124	
HR с обеспечением возможности взаимодействия	6,2	124	
Общий HR	6,2	124	
QR невокализованных сигналов	2,7	54	
CNG QR	2,7	54	
CNG ER	1,0	20	

Типы кодирования с полной скоростью (FR) основаны на применении стандартного AMR-WB-кодека, работающего на скорости 12,65 кбит/с. Использование скорости 12,65 кбит/с AMR-WB-кодека позволяет осуществить кодек, работающий с переменной скоростью в битах, для CDMA-системы, выполненной с возможностью взаимодействия с другими системами путем использования стандарта AMR-WB-кодека. Для согласования с полной скоростью, предусматриваемой оболочкой Rate Set II при использовании технологии CDMA, составляющей 13,3 кбит/с, добавляют избыточные 13 битов на кадр. Эти биты используют для повышения робастности кодека в случае стертых кадров и получения существенного различия между типом общего FR и FR с обеспечением возможности взаимодействия (упомянутые биты не используются для типа FR с возможностью взаимодействия). Типы FR-кодирования основаны на модели линейного предсказания с возбуждением по алгебраическому коду (ACELP-модели), оптимизированной для обычных широкополосных речевых сигналов. Эта модель работает на речевых кадрах длительностью 20 мс с частотой дискретизации 16 кГц. Перед дальнейшей обработкой входной сигнал подвергают субдискретизации с частотой дискретизации 12,8 кГц и предварительной обработке. Параметры LP-фильтра кодируются один раз за кадр с использованием 46 бит. Затем этот кадр делится на четыре субкадра, в которых индексы и усиления адаптивной и фиксированной кодовой книг кодируются один раз за субкадр. Фиксированная кодовая книга формируется с использованием структуры алгебраической кодовой книги, в которой 64 позиции в субкадре разделены на 4 дорожки чередующихся позиций и в которой 2 импульса со знаками помещены в каждую дорожку. Эти два импульса, приходящиеся на дорожку, кодируются с использованием 9 битов, что дает в сумме 36 битов на субкадр. Более подробное описание AMR-WB-кодека содержится в работе [1]. Распределения битов для типов FR-кодирования приведены в таблице 2.

Распределение битов для типов кодирования с полной скоростью (FR) общего типа и с обеспечением возможности взаимодействия в оболочке Rate Set II стандарта CDMA2000 на основании стандарта на AMR-WB-кодеки при скорости 12,65 кбит/с			
Параметр	Количество битов на кадр		
	Общий FR	FR с обеспечением возможности взаимодействия	
Класс информации	-	-	
Бит PPA	-	1	
Параметры	46	46	
Задержка основного тона	30	30	

Фильтрация основного тона	4	4
Усиления	28	28
Алгебраическая кодовая книга	144	144
Биты FER-защиты	14	-
Неиспользованные биты	-	13
Итого	266	266

5

В случае устойчивых вокализованных кадров используется кодирование вокализованных сигналов с половинной скоростью. Распределение битов вокализованных кадров для половинной скорости приведено в таблице 3. Поскольку кадры, подлежащие кодированию в этом режиме связи, являются - что характерно - весьма периодичными, для поддержания хорошего субъективного качества, например, по сравнению со случаем переходных кадров, оказывается достаточной значительно более низкая скорость в битах. Используют модификацию сигналов, которая обеспечивает удовлетворительное кодирование задерживаемой информации с использованием лишь девяти битов на кадр длительностью 20 мс, что позволяет экономить значительную долю ресурса битов для задания других параметров кодирования сигналов. При модификации сигналов сигнал вынужденно следует некоторому профилю основного тона, который может передаваться с использованием по 9 битов на кадр. Хорошие показатели долговременного предсказания позволяют использовать лишь 12 битов на субкадр длительностью 5 мс для возбуждения по фиксированной кодовой книге без ухудшения субъективного качества речи. Фиксированная кодовая книга представляет собой алгебраическую кодовую книгу и содержит две дорожки, на каждую из которых приходится по одному импульсу, при этом каждая дорожка имеет 32 возможных позиции.

10

15

20

25

30

35

Параметр	Количество битов на кадр			
	Общий HR	HR вокализованных сигналов	HR невокализованных сигналов	HR с обеспечением возможности взаимодействия
Класс информации	1	3	2	3
Бит PPA	-	-	-	1
Параметры LP	36	36	46	46
Задержка основного тона	13	9	-	30
Фильтрация основного тона	-	2	-	4
Усиления	26	26	24	28
Алгебраическая кодовая книга	48	48	52	-
Биты FER-защиты	-	-	-	-
Неиспользованные биты	-	-	-	12
Итого	124	124	124	124

40

45

50

В случае невокализованных кадров адаптивная кодовая книга (или кодовая книга основного тона) не используется. В каждом субкадре используется 13-битовая гауссова кодовая книга, причем усиление этой кодовой книги кодируется с использованием 6 битов на субкадр. Следует отметить, что в случаях, когда нужно дополнительно уменьшить среднюю скорость в битах, можно использовать кодирование невокализованных сигналов с четвертной скоростью в случае устойчивых невокализованных кадров.

Для сегментов с низкой энергией используется режим кодирования с половинной скоростью общего типа. Этот режим общего HR также можно использовать при работе на максимуме, равном половинной скорости, как пояснено ниже. Распределение битов для режима общего HR приведено в таблице 3.

Например, в случае классификационной информации для разных HR-кодеров в случае общего HR 1 бит используется, чтобы указать, является ли кадр кадром общего HR или другого HR. В случае HR невокализованных сигналов для классификации используется 2 бита: первый бит - чтобы указать, что кадр не является кадром общего HR, а второй бит - чтобы указать, что он является невокализованным HR-кадром, а не вокализованным HR-кадром или HR-кадром взаимодействия (это поясняется ниже). В случае HR-кодирования

вокализованных сигналов используются 3 бита: первые 2 бита указывают, что кадр не является кадром общего HR или невокализованным HR-кадром, а третий бит указывает, является ли кадр невокализованным кадром или HR-кадром взаимодействия.

В экономичном режиме большинство невокализованных кадров можно кодировать с использованием QR-кодера невокализованных сигналов. В этом случае индексы гауссовой кодовой книги генерируют случайным образом, а усиление кодируется с использованием лишь 5 битов на субкадр. Кроме того, коэффициенты LP-фильтра квантуются с меньшей скоростью в битах. 1 бит используется для различения между двумя типами кодирования с четвертной скоростью: QR-кодированием невокализованных кадров и CNG QR-кодированием. Распределение битов для типов кодирования невокализованных сигналов приведено в [6].

Тип HR-кодирования с обеспечением возможности взаимодействия позволяет осуществлять обработку в ситуациях, когда CDMA-система задает половинную скорость (HR) как максимальную скорость для конкретного кадра, а этот кадр классифицирован как кадр с полной скоростью. HR-кадр взаимодействия получают непосредственно из кодера полной скорости за счет отбрасывания индексов фиксированной кодовой книги после кодирования кадра как кадра с полной скоростью (таблица 4). На декодирующей стороне индексы фиксированной кодовой книги могут генерироваться случайным образом, и декодер будет работать как в режиме полной скорости. Эта схема имеет преимущество, заключающееся в том, что она минимизирует воздействие принудительно устанавливаемого режима половинной скорости во время бестандемной работы между CDMA-системой и другими системами, использующими стандарт на AMR-WB-кодеки (такими как GSM-система или беспроводная W-CDMA-система третьего поколения). Как упоминалось ранее, тип FR-кодирования с обеспечением возможности взаимодействия или CNG QR-кодирования используются для бестандемной работы (TFO) с AMR-WB-кодеком. В линии связи с направлением от CDMA2000-системы к системе, использующей AMR-WB-кодек, когда подуровень мультиплексирования указывает запрос режима половинной скорости, VMR-WB-кодек будет использовать тип HR-кодирования с обеспечением возможности взаимодействия. В системном интерфейсе, когда принимается HR-кадр взаимодействия, к потоку битов добавляются индексы алгебраической кодовой книги, что позволяет получить скорость 12,65 кбит/с. AMR-WB-декодер на принимающей стороне будет интерпретировать это как обычный кадр со скоростью 12,65 кбит/с. В другом направлении, т.е. в линии связи от системы, использующей AMR-WB-кодек, к CDMA2000-системе, если в системном интерфейсе принимается запрос режима половинной скорости, то индексы алгебраической кодовой книги отбрасываются, а биты режима, указывающие тип HR-кадра взаимодействия, добавляются. Декодер на стороне, соответствующей стандарту CDMA2000, работает как устройство, предусматривающее тип HR-кодирования с обеспечением возможности взаимодействия, что является частью решения, предусматривающего VMR-WB-кодирование. Без признаков HR-кодирования с обеспечением возможности взаимодействия, принудительно устанавливаемый режим половинной скорости интерпретировался бы как стирание кадров.

Для обработки неактивных речевых кадров используют метод генерирования комфортного шума (CNG). Во время работы в рамках CDMA-системы для кодирования неактивных речевых кадров используют тип кодирования с одной восьмой скорости (ER-кодирования) посредством CNG. При вызове, требующем взаимодействия с аппаратурой, соответствующей стандарту AMR-WB-кодирования речи, использовать CNG ER-кодирование нельзя, потому что обуславливаемая этим стандартом скорость в битах ниже, чем скорость в битах, необходимая для передачи информации обновления для CNG-кодера, соответствующего стандарту AMR-WB-кодирования речи [3]. В этом случае используют CNG QR-кодирование. Вместе с тем AMR-WB-кодек часто работает в режиме прерывистой передачи (DTX). Во время прерывистой передачи информация о фоновом шуме не обновляется в каждом кадре. Как правило, передается только один кадр из 8-ми последовательных неактивных речевых кадров. Этот кадр обновления называют

дескриптором паузы (SID) [4]. В CDMA-системе, где кодируется каждый кадр, работа в режиме DTX не используется. Следовательно, методом CNG QR-кодирования CDMA-стороне нужно кодировать только SID-кадры, а остальные кадры можно по-прежнему кодировать методом CNG ER-кодирования, чтобы снизить среднюю скорость передачи данных (ADR), поскольку эти кадры не используются VMR-WB-аппаратурой на противоположной стороне. При CNG-кодировании только параметры LP-фильтра и усиление кодируются один раз за кадр. Распределение битов для CNG QR-кодирования приведено в таблице 4, а распределение битов для CNG ER-кодирования приведено в таблице 5.

10

Параметр	QR-кодирование невокализованных сигналов	CNG QR-кодирование
Биты выбора	1	1
Параметры LP	32	28
Усиления	20	6
Неиспользованные биты	1	19
Итого	54	54

15

Параметр	CNG ER-кодирование
	Количество битов на кадр
Параметры LP	14
Усиления	6
Неиспользованные биты	-
Итого	20

20

Классификация сигналов и выбор скорости в высококачественном режиме

На фиг.5 проиллюстрирован способ 400 цифрового кодирования звукового сигнала в соответствии со вторым иллюстративным вариантом осуществления второго аспекта настоящего изобретения. Следует отметить, что способ 400 представляет собой конкретное приложение способа 100 в высококачественном режиме, предусмотренное для максимизации качества синтезированной речи, достигаемого на доступных скоростях в битах (следует отметить, что случай, когда система ограничивает максимальную доступную скорость для конкретного кадра, будет описан в отдельном подразделе). Следовательно, большинство активных речевых кадров кодируются с полной скоростью, т.е. 13,3 кбит/с.

35 Аналогично способу 100, проиллюстрированному на фиг.2, детектор речевой активности (VAD) осуществляет различие между активными и неактивными речевыми кадрами (этап 102). Алгоритм принятия решения о речевой активности (PPA) аналогичен для всех режимов работы. Если обнаруживается неактивный речевой кадр (сигнал фонового шума), то способ классификации заканчивается, а кадр кодируется с использованием типа CNG ER-кодирования со скоростью 1,0 кбит/с в соответствии с оболочкой Rate Set II при использовании технологии CDMA (этап 402). Если обнаруживается активный речевой кадр, то этот кадр подвергают воздействию второго классификатора, предназначенного специально для различения невокализованных кадров (этап 404). Поскольку целью высококачественного режима является достижение наилучшего возможного качества, то различие невокализованных кадров является очень строгим, и выбираются только очень стабильные невокализованные кадры. Правила классификации невокализованных кадров и пороги принятия решений являются такими же, как приведенные выше. Если второй классификатор классифицирует кадр как невокализованной речевой сигнал, то способ классификации заканчивается, а кадр кодируется с использованием типа HR-кодирования невокализованных сигналов (этап 408), оптимизированного для невокализованных сигналов (6,2 кбит/с в соответствии с оболочкой Rate Set II при использовании технологии CDMA). Все остальные кадры обрабатываются с использованием типа общего FR-кодирования на основе стандарта на AMR-WB-кодеки со скоростью 12,65 кбит/с (этап

406).

Классификация сигналов и выбор скорости в стандартном режиме

На фиг.6 проиллюстрирован способ 500 цифрового кодирования звукового сигнала в соответствии с третьим иллюстративным вариантом осуществления второго аспекта
 5 настоящего изобретения. Способ 500 обеспечивает классификацию речевого сигнала и его кодирование в стандартном режиме.

На этапе 102 VAD осуществляет различение между активными и неактивными речевыми кадрами. Если обнаруживается неактивный речевой кадр, то осуществление способа классификации заканчивается, а кадр кодируется с одной восьмой скорости посредством
 10 CNG (т.е. как CNG-ER-кадр) (этап 510). Если обнаруживается активный речевой кадр, то этот кадр подвергается воздействию классификатора второго уровня для различения невокализованных кадров (этап 404). Правила классификации невокализованных кадров и пороги принятия решений являются такими же, как описанные выше. Если классификатор второго уровня классифицирует кадр как невокализованной речевой сигнал, то способ
 15 классификации заканчивается, а кадр кодируется с использованием типа HR-кодирования невокализованных сигналов (этап 508). В противном случае речевой кадр пропускается через модуль классификации "устойчивых вокализованных" кадров (этап 502). Различение вокализованных кадров является неотъемлемым признаком алгоритма модификации сигналов, как описано выше. Если кадр пригоден для модификации сигналов, он
 20 классифицируется как устойчивый вокализованный кадр и кодируется с использованием типа HR-кодирования вокализованных сигналов (этап 506) в модуле, оптимизированном для устойчивых вокализованных сигналов (6,2 кбит/с в соответствии с оболочкой Rate Set II при использовании технологии CDMA). В противном случае кадр, вероятно, содержит неустойчивый речевой сегмент, например вокализованный начальный или
 25 быстро эволюционирующий сигнал. Такие кадры обычно требуют высокой скорости в битах для поддержания надлежащего субъективного качества. Вместе с тем, если энергия кадра ниже, чем некоторый порог, то эти кадры можно кодировать с использованием типа общего HR-кодирования. Если на этапе 512 классификатор четвертого уровня обнаруживает сигнал с низкой энергией, то кадр кодируется с использованием типа общего HR-
 30 кодирования (этап 514). В противном случае речевой кадр кодируется как FR-кадр общего типа (13,3 кбит/с в соответствии с оболочкой Rate Set II при использовании технологии CDMA) (этап 504).

Классификация сигналов и выбор скорости в экономичном режиме

На фиг.7 проиллюстрирован способ 600 цифрового кодирования звукового сигнала в соответствии с четвертым иллюстративным вариантом осуществления первого аспекта
 35 настоящего изобретения. Способ 600 обеспечивает классификацию речевого сигнала и его кодирование в экономичном режиме.

Экономичный режим обеспечивает максимальную пропускную способность системы и при этом по-прежнему дает высококачественную широкополосную речь. Логика
 40 определения скорости аналогична стандартному режиму, за исключением того, что используется также тип QR-кодирования невокализованных кадров, а использование FR-кодирования уменьшается.

Во-первых, на этапе 102, VAD осуществляет различение между активными и неактивными речевыми кадрами. Если обнаруживается неактивный речевой кадр, то
 45 способ классификации заканчивается, а кадр кодируется как CNG-ER-кадр (этап 402). Если обнаруживается активный речевой кадр, то этот кадр подвергается воздействию второго классификатора для различения невокализованных кадров (этап 106). Правила классификации невокализованных кадров и пороги принятия решений являются такими же, как описанные выше. Если второй классификатор классифицирует кадр как
 50 невокализованной речевой сигнал, то речевой кадр пропускается в первый классификатор третьего уровня (этап 602). Этот классификатор третьего уровня проверяет, является ли кадр переходным от вокализованного сигнала к невокализованному сигналу, с помощью правил, описанных выше. В частности, этот классификатор третьего уровня проверяет,

является ли последний кадр либо невокализованным кадром, либо кадром фонового шума, и имеет ли место ситуация, в которой энергия в конце кадра сконцентрирована на высоких частотах, а в упреждающей выборке не обнаруживается потенциальная вокализованная начальная часть. Как пояснялось выше, последние два условия

5 определяются следующим образом:

$$(r_x(2) < th_{12}) \text{ и } (e_{\text{tilt}}(1) < th_{13}) \text{ при } th_{12} = 0,73, th_{13} = 3,$$

где $r_x(2)$ - корреляция в упреждающей выборке, а $e_{\text{tilt}}(1)$ - наклон во втором спектральном анализе, который распространяется на конец кадра и упреждающую выборку.

10 Если кадр содержит переход от вокализованного сигнала к невокализованному сигналу, то этот кадр на этапе 508 кодируется с использованием типа HR-кодирования невокализованных сигналов. В противном случае кадр кодируется с использованием типа QR-кодирования невокализованных сигналов (этап 604). Кадры, не классифицированные как невокализованные, пропускаются через модуль классификации "устойчивых
15 вокализованных" кадров, который является вторым классификатором третьего уровня (этап 110). Различение вокализованных кадров является неотъемлемым признаком алгоритма модификации сигналов, как описано выше. Если кадр пригоден для модификации сигналов, он классифицируется как устойчивый вокализованный кадр и кодируется с использованием типа HR-кодирования вокализованных сигналов на этапе 506. Аналогично стандартному
20 режиму, остальные кадры (не классифицированные как невокализованные или стабильные вокализованные) проверяются на малое содержание энергии. Если на этапе 512 обнаруживается сигнал с низкой энергией, этот кадр кодируется на этапе 514 с использованием HR-кодирования общего типа. В противном случае речевой кадр кодируется как FR-кадр общего типа (13,3 кбит/с в соответствии с оболочкой Rate Set
25 II при использовании технологии CDMA) (этап 504).

Классификация сигналов и выбор скорости в режиме, обеспечивающем возможность взаимодействия

На фиг.8 проиллюстрирован способ 700 цифрового кодирования звукового сигнала в соответствии с пятым иллюстративным вариантом осуществления второго аспекта
30 настоящего изобретения. Способ 700 обеспечивает классификацию речевого сигнала и его кодирование в режиме, обеспечивающем возможность взаимодействия.

Режим, обеспечивающий возможность взаимодействия, обеспечивает бестандемную работу между CDMA-системой и другими системами, использующими стандарт на AMR-WB-кодеки, на скорости 12,65 кбит/с (или более низких скоростях). В отсутствие
35 ограничения скорости, накладываемого CDMA-системой, используется только FR-кодирование с обеспечением возможности взаимодействия и генераторы комфортного шума.

Во-первых, на этапе 102, VAD осуществляет различение между активными и неактивными речевыми кадрами. Если обнаруживается неактивный речевой кадр, на этапе
40 702 принимается решение, следует ли кодировать этот кадр как SID-кадр. Как упоминалось ранее, SID-кадры служат для обновления параметров CNG на AMR-WB-стороне во время работы в режиме DTX [4]. Во время периодов пауз, как правило, кодируется лишь один из 8 неактивных речевых кадров. Однако после активного речевого сегмента информация об обновлении SID должна посылатся уже в 4-м кадре (см. [4]).
45 Поскольку для кодирования SID-кадра одной восьмой скорости (ER) оказывается недостаточно, SID-кадры кодируют методом CNG QR-кодирования на этапе 704. Кадры, иные, чем неактивные SID-кадры, кодируются методом CNG EQ-кодирования на этапе 402. Если линия связи реализуется в направлении от VMR-WB-аппаратуры, использующей технологию CDMA, к AMR-WB-аппаратуре, конфигурированной для бестандемной работы
50 (TFO), то CNG-ER-кадры отбрасываются в системном интерфейсе, поскольку AMR-WB-аппаратура не использует их. В противоположном направлении эти кадры не передаются (AMR-WB-аппаратура генерирует только SID-кадры) и квалифицируются как стирания кадров. Все активные речевые кадры обрабатываются с использованием типа FR-

кодирования с обеспечением возможности взаимодействия (этап 706), который, по существу является стандартом AMR-WB-кодирования, со скоростью 12,65 кбит/с.

Классификация сигналов и выбор скорости при работе на максимуме, равном половинной скорости

- 5 На фиг.9 проиллюстрирован способ 800 цифрового кодирования звукового сигнала в соответствии с шестым иллюстративным вариантом осуществления второго аспекта настоящего изобретения. Способ 800 обеспечивает классификацию речевого сигнала и его кодирование при работе на максимуме, равном половинной скорости, для высококачественного и стандартного режимов.
- 10 Как описано выше, CDMA-система задает максимальную скорость в битах для конкретного кадра. Чаще всего максимальная скорость в битах, задаваемая системой, ограничивается половинной скоростью (HR). Вместе с тем система может задавать и меньшие скорости.
- 15 Все активные речевые кадры, которые обычно должны классифицироваться как FR-кадры во время нормальной работы, теперь кодируются с использованием типа HR-кодирования. Тогда механизм классификации и выбора скорости классифицирует все такие локализованные кадры с использованием типа HR-кодирования локализованных сигналов (это кодирование происходит на этапе 506), а все такие не локализованные кадры - с использованием типа HR-кодирования не локализованных сигналов (это кодирование
- 20 происходит на этапе 408). Все остальные кадры, которые должны классифицироваться как FR-кадры во время нормальной работы, кодируются с использованием HR-кодирования общего типа на этапе 514, за исключением того, что в режиме, обеспечивающем возможность взаимодействия, используется тип HR-кодирования, обеспечивающего возможность взаимодействия (этап 908 на фиг.11).
- 25 Как можно заметить на фиг.9, механизм классификации сигналов и выбора скорости аналогичен тому, который действует при обычной работе в стандартном режиме. Однако используется HR-кодирование общего типа (этап 514) вместо FR-кодирования общего типа (этап 406 на фиг.5), а пороги, используемые для различения локализованных и не локализованных кадров, являются более низкими, чтобы обеспечить кодирование как
- 30 можно большего количества кадров с использованием типов HR-кодирования не локализованных сигналов и HR-кодирования локализованных сигналов. Как правило, в случае работы на максимуме, равном половинной скорости, в высококачественном и стандартном режимах используются пороги, предназначенные для экономичного режима.
- 35 На фиг.10 проиллюстрирован способ 900 цифрового кодирования речевого сигнала в соответствии с седьмым иллюстративным вариантом осуществления первого аспекта настоящего изобретения. Способ 900 обеспечивает классификацию речевого сигнала и кодирование при работе на максимуме, равном половинной скорости, для экономичного режима. Способ 900, проиллюстрированный на фиг.10, аналогичен способу 600, проиллюстрированному на фиг.7, за исключением того, что все кадры, кодировавшиеся
- 40 методом FR-кодирования общего типа, теперь кодируются методом HR-кодирования общего типа (при работе на максимуме, равном половинной скорости, нет необходимости в классификации кадров с низкой энергией). На фиг.11 проиллюстрирован способ 920 цифрового кодирования речевого сигнала в соответствии с восьмым иллюстративным вариантом осуществления первого аспекта настоящего изобретения. Способ 920
- 45 обеспечивает классификацию речевого сигнала и определение скорости в режиме, обеспечивающем возможность взаимодействия, при работе на максимуме, равном половинной скорости. Поскольку способ 920 очень похож на способ 700, проиллюстрированный на фиг.8, ниже будет приведено описание лишь различий между этими двумя способами.
- 50 В случае способа 920 не могут использоваться типы кодирования, специфичные для сигналов (HR-кодирование не локализованных сигналов и HR-кодирование локализованных сигналов), потому что эти типы будут не понятны для AMR-WB-аппаратуры на противоположном конце системы; также нельзя использовать и HR-кодирование общего

типа. Поэтому активные речевые кадры при работе на максимуме, равном половинной скорости, кодируются с использованием типа HR-кодирования, обеспечивающего возможность взаимодействия.

Если система ограничивает максимальную скорость в битах величиной, меньшей, чем половинная скорость (HR), то кодирование общего типа не предусматривается для обработки в этих случаях; в сущности, потому, что эти случаи исключительно редки, а такие кадры могут квалифицироваться как стирания кадров. Вместе с тем, если система ограничивает максимальную скорость в битах величиной, равной четвертной скорости (QR), можно использовать QR-кодирование невокализированных сигналов. Однако это возможно только в режимах, специфичных для CDMA (высококачественном, стандартном, экономичном), потому что AMR-WB-аппаратура на противоположном конце системы не способна интерпретировать QR-кадры.

Эффективное взаимодействие между AMR-WB-кодеком и VMR-WB-кодеком, работающим в оболочке Rate Set II

Ниже, со ссылками на фиг.12 описан способ 1000 кодирования речевого сигнала для взаимодействия между AMR-WB- и VMR-WB-кодеками в соответствии с иллюстративным вариантом осуществления четвертого аспекта настоящего изобретения.

Более конкретно, способ 1000 обеспечивает бестандемную работу между стандартным AMR-WB-кодеком и управляемым источником VBR-кодеком, предназначенным, например, для CDMA2000-систем (и именуемым далее VMR-WB-кодеком). В гарантируемом способом 1000 режиме, обеспечивающем возможность взаимодействия, VMR-WB-кодек использует скорости в битах, которые могут интерпретироваться AMR-WB-кодеком и при этом остаются в пределах скоростей в битах, характерных для оболочки Rate Set II и используемых, например, в CDMA-кодеке.

Поскольку характерные для оболочки Rate Set II скорости в битах составляют 13,3 кбит/с (полная скорость, FR), 6,2 кбит/с (половинная скорость, HR), 2,7 кбит/с (четвертная скорость, QR) и 1,0 кбит/с (одна восьмая скорости, ER), то для AMR-WB-кодека можно использовать скорости в битах, величины которых в режиме полной скорости составляют 12,65, 8,85 или 6,6 кбит/с, SID-кадры можно кодировать со скоростью 1,75 кбит/с в режиме четвертной скорости. AMR-WB-кодек, работающий на скорости 12,65 кбит/с, является ближайшим к кодеку стандарта CDMA2000, работающему на полной скорости 13,3 кбит/с (CDMA2000-FR-кодеку), и используется в качестве FR-кодека в этом иллюстративном варианте осуществления. Вместе с тем, когда AMR-WB-кодек используется в GSM-системах, алгоритм адаптации линии связи может способствовать снижению скорости в битах до 8,85 или 6,6 кбит/с, в зависимости от канальных условий (чтобы предоставить больше битов для канального кодирования). Таким образом, работа AMR-WB-кодека на скоростях в битах 8,85 и 6,6 кбит/с может представлять собой часть режима, обеспечивающего возможность взаимодействия, а эти скорости можно использовать в CDMA2000-приемнике в случае, если GSM-система решила использовать одну из этих скоростей в битах. В иллюстративном варианте осуществления, показанном на фиг.12, используются три типа кодирования с полной скоростью, обеспечивающей возможность взаимодействия (I-FR), соответствующие скоростям AMR-WB-кодека, составляющим 12,65, 8,85 и 6,6 кбит/с, которые далее обозначаются символами I-FR-12, I-FR-8 и I-FR-6 соответственно. В случае I-FR-12 имеются 13 неиспользованных битов. Первые 8 битов используются для различения I-FR-кадров и FR-кадров общего типа (которые используют дополнительные биты для улучшения маскировки стирания кадров). Другие 5 битов используются для сигнализации о трех типах I-FR-кадров. При обычной работе используется I-FR-12, а более низкие скорости используются в случае, если это требуется для адаптации линии связи в GSM-системе.

В CDMA2000-системе средняя скорость передачи данных речевого кодека непосредственно связана с пропускной способностью системы. Следовательно, достижение той наименьшей средней скорости передачи данных (ADR), которая возможна при минимальной потере качества речи, приобретает первостепенное значение. AMR-WB-

кодек предназначен для основных GSM-систем и радиосвязи третьего поколения на основе эволюции GSM. Таким образом, режим, обеспечивающий возможность взаимодействия, для CDMA2000-системы может привести к повышенной ADR по сравнению с VBR-кодеком, специально предназначенным для CDMA2000-систем.

5 Основными причинами этого являются следующие:

недостаточность режима половинной скорости на уровне 6,2 кбит/с в AMR-WB-кодеке; скорость в битах для CID-кадров в AMR-WB-кодеке составляет 1,75 кбит/с, что не согласуется с одной восьмой скорости (ER) в оболочке Rate Set II;

10 при работе в режиме прерывистой передачи (DTX) детектора речевой активности (VAD) AMR-WB-кодека используются несколько кадров (кодированных как речевые кадры), чтобы вычислить первый кадр дескриптора паузы (SID_FIRST-кадр).

Способ кодирования речевого сигнала для взаимодействия между AMR-WB- и VMR-WB-кодеками обеспечивает устранение вышеупомянутых ограничений и приводит к сниженной ADR режима, обеспечивающей возможность взаимодействия, так что он оказывается эквивалентным CDMA-специфичным режимам при сравнимом качестве речи. Ниже приводится описание способов для обоих направлений работы: "VMR-WB-кодирование - AMR-WB-декодирование" и "AMR-WB-кодирование - VMR-WB-декодирование".

"VMR-WB-кодирование - AMR-WB-декодирование"

20 При кодировании на стороне VMR-WB-кодека в CDMA-системе операция генерирования комфортного шума в режиме прерывистой передачи для детектора речевой активности (VAD/DTX/CNG) согласно стандарту AMR-WB не требуется. VAD надлежащим образом установлен в VMR-WB-кодек и работает точно так же, как в других CDMA2000-специфичных режимах, т.е. задержка решений по речевой активности (PPA) используется по мере необходимости, чтобы не пропустить невокализованные паузы, так что всякий раз, когда 25 флаг PPA устанавливается в нуль (PPA_флаг=0) (что классифицируется как фоновый шум), применяется кодирование посредством генерирования комфортного шума (CNG).

Операция генерирования комфортного шума для VAD (VAD/CNG) максимально приближена к операции в режиме прерывистой передачи AMR-кодека (AMR/DTX). Операция генерирования комфортного шума в режиме прерывистой передачи для VAD 30 (VAD/DTX/CNG) в AMR-WB-кодеке проводится следующим образом. Семь кадров фонового шума после периода активной речи кодируются как речевые кадры, а бит PPA задается равным нулю (переход DTX). Затем посылают SID_FIRST-кадр. В SID_FIRST-кадре сигнал не кодируется, а параметры CNG получают в результате DTX-перехода (7 речевых кадров) в декодере. Следует отметить, что AMR-WB-кодек не использует DTX-переход 35 после периодов активной речи, которые короче, чем 24 кадра, чтобы уменьшить перегрузку при DTX-переходе. После SID_FIRST-кадра посылают два кадра как кадры без данных (NO_DATA-кадры), за которыми следует кадр обновления SID (SID_UPDATE-кадр) (1,75 кбит/с). После этого посылают 7 NO_DATA-кадров, за которыми следует SID_UPDATE-кадр, и т.д. Это продолжается до тех пор, пока не будет обнаружен активный 40 речевой кадр (PPA_флаг=1) [4].

В иллюстративном варианте, показанном на фиг.12, VAD в VMR-WB-кодеке не использует DTX-переход. Первый кадр фонового шума после периода активной речи кодируется со скоростью 1,75 кбит/с и передается как QR-кадр, затем кодируют 2 кадра со скоростью 1 кбит/с (одной восьмой скорости), а затем следующий кадр передается со 45 скоростью 1,75 кбит/с как QR-кадр. После этого 7 кадров передаются как ER-кадры, за которыми следует один QR-кадр, и т.д. Это приблизительно соответствует операции прерывистой передачи (DTX) AMR-WB-кодека, за тем исключением, что DTX-переход используется для уменьшения средней скорости передачи данных (ADR).

Хотя операция VAD/CNG в VMR-WB-кодеке, описываемая в этом иллюстративном 50 варианте осуществления, близка к операции DTX в AMR-WB-кодеке, можно использовать и другие способы, которые способствуют дальнейшему уменьшению ADR. Например, QR-кадры генерирования комфортного шума (CNG-QR-кадры) могут передаваться не так часто, например, по одному на каждые 12 кадров. Кроме того, могут оцениваться

вариации шума в кодере, а CNG-QR-кадры могут передаваться лишь тогда, когда характеристики шума изменяются (а не по одному на каждые 8-12 кадров).

Чтобы снять ограничение отсутствия половинной скорости на уровне 6,2 кбит/с в AMR-WB-кодере, в режиме, обеспечивающем возможность взаимодействия, предусматривается работа с I-HR - половинной скоростью, при этом предусматривается кодирование кадра как кадра с полной скоростью и последующее отбрасывание битов, соответствующих индексам алгебраической кодовой книги (144 бита на кадр в AMR-WB-кодере при 12,65 кбит/с). Это уменьшает скорость в битах до 5,45 кбит/с, что соответствует половинной скорости в оболочке Rate Set II согласно стандарту CDMA2000. Перед декодированием отбрасываемые биты могут генерироваться либо случайным образом (т.е. с использованием генератора случайных чисел), либо псевдослучайным образом (т.е. путем повторения части существующего потока битов), либо некоторым предварительно определенным образом. I-HR можно использовать, когда CDMA2000-система передает запрос размерно-пакетной сигнализации или работы на максимуме, равном половинной скорости. Это препятствует оценке речевого кадра как потерянного кадра. I-HR также может использоваться VMR-WB-кодеком в режиме, обеспечивающем возможность взаимодействия, для кодирования невокализованных кадров или кадров, в которых вклад алгебраической кодовой книги в качество синтезируемой речи является минимальным. Это приводит к уменьшенной ADR. Следует заметить, что в этом случае кодер может выбрать кодирование кадров в I-HR-режиме и тем самым минимизировать ухудшение качества речи за счет использования таких кадров.

Как показано на фиг.12, в направлении "VMR-WB-кодирование - AMR-WB-декодирование" речевые кадры кодируют в режиме, обеспечивающем возможность взаимодействия, VMR-WB-кодера 1002, что позволяет получить одну из следующих скоростей в битах: полную скорость в режиме, обеспечивающем возможность взаимодействия, обозначаемую символом I-FR, для активных речевых кадров (т.е. скорость I-FR-12, I-FR-8 или I-FR-6), половинную скорость в режиме, обеспечивающем возможность взаимодействия, обозначаемую символом I-HR, в случае размерно-пакетной сигнализации, или - по выбору - для кодирования некоторых невокализованных кадров или кадров, в которых вклад алгебраической кодовой книги в качество синтезируемой речи является минимальным, четвертную скорость (QR) при генерировании комфортного шума (CNG) для кодирования кадров фонового шума (одного из восьми кадров фонового шума, как описано выше, или в случае, когда обнаруживается вариация в характеристике фонового шума), и одну восьмую скорости (ER) при CNG с получением CNG-ER-кадров для большинства кадров фонового шума (кадры фонового шума не кодируются как CNR-QR-кадры). В системном интерфейсе, который выполнен в форме шлюза, выполняются следующие операции.

Во-первых, проверяется достоверность кадра, принимаемого шлюзом из VMR-WB-кодера. Если он недостоверен, то после этого VMR-WB-кадр посылается как стертый (тип потери речи для AMR-WB-декодера). Кадр считается недостоверным, например, если возникает одно из следующих состояний:

- если принимается кадр со всеми нулями (используемый сетью в случае пробела и пакета), то этот кадр стирается;
- в случае FR-кадров, если 13 битов преамбулы не соответствуют I-FR-12, I-FR-8 или I-FR-6, или если неиспользованные биты не являются нулевыми, то кадр стирается; кроме того, I-FR устанавливает бит PPA равным 1, так что если бит PPA принимаемого кадра не равен 1, то этот кадр стирается;
- в случае HR-кадров, аналогично FR-кадрам, если биты преамбулы не соответствуют I-HR-12, I-HR-8 или I-HR-6, или если неиспользованные биты не являются нулевыми, то кадр стирается; то же самое имеет место для бита PPA;
- в случае QR-кадров, если биты преамбулы не соответствуют QR для CNG, то кадр стирается; кроме того, VMR-WB-кодер устанавливает бит SID_UPDATE равным 1, а биты запроса режима - равными 0010; в противном случае кадр стирается;

- в случае ER-кадров, если принимается кадр со всеми единицами, то этот кадр стирается; кроме того, VMR-WB-кодер использует битовую комбинацию фильтра Internet-сервера (ISF) со всеми нулями (первые 14 битов) для сигнализации о кадрах пробелов; если принимается эта комбинация, то кадр стирается.

5 Если принимаемый кадр является достоверным кадром в режиме, обеспечивающем возможность взаимодействия, то выполняются следующие операции:

- I-FR-кадры посылаются в AMR-WB-декодер как кадры, закодированные на скорости 12,65, 8,8 или 6,6 кбит/с, в зависимости от типа I-FR-кадров;

- CNG-QR-кадры посылаются в AMR-WB-декодер как SID_UPDATE-кадры;

10 - CNG-ER-кадры посылаются в AMR-WB-декодер как NO_DATA-кадры;

- I-HR-кадры преобразуются в кадры, кодируемые со скоростью 12,65, 8,85 или 6,6 кбит/с (в зависимости от типа кадров) путем генерирования опускаемых индексов алгебраической кодовой книги на этапе 1010; эти индексы можно генерировать случайным образом, либо посредством повторения части существующих кодовых битов, либо некоторым предварительно определенным образом; при этом также отбрасываются биты, характеризующие тип I-HR-кадров (биты, используемые для различения разных типов кадров, кодируемых с половинной скоростью, в VMR-WB-кодеке).

"AMR-WB-кодирование - VMR-WB-декодирование"

В этом направлении способ 1000 ограничивается операцией в режиме прерывистой передачи (DTX) на стороне AMR-WB-кодека. Вместе с тем во время кодирования активной речи имеется один бит (1-ый бит данных) в потоке битов, указывающий PPA_флаг (0 - для периода DTX-перехода, 1 - для активной речи). Следовательно, работу в шлюзе можно вкратце охарактеризовать следующим образом:

- SID_UPDATE-кадры пересылаются как CNG-QR-кадры;

25 - SID_FIRST-кадры и NO_DATA-кадры пересылаются как ER-кадры пробелов;

- стертые кадры (потери речи) пересылаются как ER-кадры стирания;

- первый кадр после активной речи, у которого PPA_флаг=0 (что удостоверяется на этапе 1012), сохраняется как FR-кадр, а следующие кадры, у которых PPA_флаг=0, пересылаются как ER-кадры пробелов;

30 - если шлюз на этапе 1014 принимает запрос работы на максимуме, равном половинной скорости (сигнализация уровня кадров), и при этом принимает FR-кадры, то кадр преобразуется в I-HR-кадр; это заключается в отбрасывании битов, соответствующих индексам алгебраической кодовой книги, и добавлении битов режима, указывающих тип I-HR-кадров.

35 В этом иллюстративном варианте первые два бита в ER-кадрах пробелов устанавливаются на 0x00, а в ER-кадрах стирания первые два бита устанавливаются на 0x04. В основном, первые 14 битов соответствуют индексам ISF, а для указания кадров пробелов (все нули) или кадров стирания (все нули, за исключением 14-го бита, установленного на 1, что составляет 0x04 в шестнадцатеричной системе счисления) резервируются две комбинации. Когда в VMR-WB-декодере 1004 обнаруживаются ER-кадры, они обрабатываются посредством декодера генерирования комфортного шума (CNG-декодера) с использованием последних принятых удовлетворительными параметрами CNG. Исключением является случай первого принимаемого ER-кадра пробела (инициализация CNG-декодера, когда "старые" параметры CNG еще не известны).

45 Поскольку первый кадр, у которого PPA_флаг=0, передается как FR-кадр, параметры этого кадра, а также параметры прошлого CNG используются для инициализации операции CNG. В случае ER-кадров стирания декодер использует ту процедуру маскировки, которая используется для стертых кадров.

50 Отметим, что в иллюстративном варианте осуществления, показанном на фиг.12, для FR-кадров используется скорость 12,65 кбит/с. Однако с тем же успехом можно использовать скорости 8,85 и 6,6 кбит/с в соответствии с алгоритмом адаптации линии связи, который требует использования меньших скоростей в случае плохих канальных условий. Например, при взаимодействии между CDMA2000- и GSM-системами модуль

адаптации линии связи в GSM-системе может принять решение уменьшить скорость в битах до 8,85 и 6,6 кбит/с в случае плохих канальных условий. В таком случае эти уменьшенные скорости в битах должны быть включены в решение о работе VMR-WB-аппаратуры в соответствии с технологией CDMA.

5 Работа CDMA-VMR-WB-кодека в оболочке Rate Set I

Скорости, используемые в оболочке Rate Set I, составляют: 8,55 кбит/с - полная скорость (FR); 4,0 кбит/с - половинная скорость (HR); 2,0 кбит/с - четвертная скорость (QR) и 800 бит/с - одна восьмая скорости (ER). В этом случае AMR-WB-кодек можно использовать только на скорости 6,6 кбит/с в качестве FR, а CNG-кадры можно
 10 посылать либо как QR-кадры (SID_UPDATE), либо как ER-кадры в качестве других кадров фонового шума (аналогично вышеописанной работе в оболочке Rate Set II). Чтобы снять ограничение по низкому качеству для скорости 6,6 кбит/с, предусмотрена скорость 8,55 кбит/с, обеспечивающая взаимодействие с AMR-WB-кодеком, работающим со скоростью 8,85 кбит/с. Эта скорость будет далее именоваться полной скоростью, обеспечивающей
 15 возможность взаимодействия в оболочке Rate Set I (I-FR-I), AMR-WB-кодека. Распределение битов для скорости 8,85 кбит/с и две возможные конфигурации для I-FR-I-кодирования показаны в таблице 6.

Таблица 6			
Распределение битов для типов I-FR-I-кодирования в конфигурации оболочки Rate Set I			
Параметр	AMR-WB с 8,85 кбит/с	I-FR-I с 8,55 кбит/с (конфигурация 1)	I-FR-I с 8,55 кбит/с (конфигурация 2)
	Количество битов на кадр	Количество битов на кадр	Количество битов на кадр
Биты режима половинной скорости	-	-	
Флаг PPA	1	0	0
Параметры LP	46	41	46
Задержка основного тона	26=8+5+8+5	26	26
Усиления	24=6+6+6+6	24	24
Алгебраическая кодовая книга	80=20+20+20+20	80	75
Итого	177	171	171

При I-FR-I-кодировании бит "PPA_флаг" и дополнительные 5 битов отбрасываются, чтобы получить скорость 8,55 кбит/с. Отбрасываемые биты можно легко ввести в декодере
 30 или системном интерфейсе, так что можно использовать декодер, работающий со скоростью 8,85 кбит/с. Для отбрасывания 5 битов можно использовать несколько способов, которые оказывают малое влияние на качество речи. В конфигурации 1, показанной в таблице 6, 5 битов отбрасываются, исходя из квантования параметров линейного предсказания (LP). В AMR-WB-кодеке используются 46 битов для квантования
 35 параметров LP в домене пар спектра иммитансов (ISP) (с использованием удаления среднего значения и предсказания скользящего среднего). Остаточный (после предсказания) вектор ISP размером 16 квантуется с использованием многостадийного квантования вектора с разбиением. Вектор разбивается на 2 субвектора с размерами 9 и 7 соответственно. Эти 2 субвектора квантуется в две стадии. На первой стадии каждый
 40 субвектор квантуется 8 битами. На втором этапе векторы с погрешностью квантования разбиваются на 3 и 2 субвектора соответственно. Субвекторы, получаемые на второй стадии, имеют размеры 3, 3, 3, 3 и 4 соответственно и квантуются 6, 7, 7, 5 и 5 битами соответственно. В предлагаемом I-FR-I-режиме 5 битов последнего субвектора, получаемого на второй стадии, отбрасываются. Они оказывают наименьшее влияние,
 45 поскольку соответствуют высокочастотной части спектра. Отбрасывание этих 5-ти битов реализуется путем фиксации индекса последнего субвектора, получаемого на второй стадии, на конкретном значении, которое не нужно передавать. Факт фиксации этого 5-битового индекса легко учесть во время квантования в VMR-WB-кодеке. Этот фиксированный индекс вводится либо в системном интерфейсе (т.е. во время работы VMR-
 50 WB-кодера и AMR-WB-декодера), или в декодере (т.е. во время работы AMR-WB-кодера и VMR-WB-декодера). Таким образом, для декодирования I-FR-I-кадра в оболочке Rate Set I AMR-WB-декодер используется на скорости 8,85 кбит/с.

Во второй конфигурации согласно иллюстрируемому варианту осуществления 5 битов

отбрасываются, исходя из индексов алгебраической кодовой книги. В AMR-WB-кодеке, работающем со скоростью 8,85 кбит/с, кадр делится на субкадры длиной по 64 выборки. Алгебраическая кодовая книга основана на делении субкадра на 4 дорожки по 16 позиций и размещении импульса со знаком на каждой дорожке. Каждый импульс кодируется 5 битами: 4 бита - для положения, а 1 бит - для знака. Таким образом, для каждого субкадра используется 20-битовая алгебраическая кодовая книга. Один путь отбрасывания пяти битов заключается в том, чтобы отбрасывать один импульс из некоторого субкадра. Например, 4-й импульс в 4-й дорожке с позициями в 4-м субкадре. В VMR-WB-кодеке этот импульс можно фиксировать на предварительно определенном значении (т.е. в определенной позиции и с определенным знаком) во время поиска в кодовой книге. Этот известный индекс импульса затем может добавляться в системном интерфейсе и посылаться в AMR-WB-декодер. В другом направлении индекс этого импульса отбрасывается в системном интерфейсе, а в CDMA-VMR-WB-декодеке этот индекс импульса может генерироваться случайным образом. Для отбрасывания этих битов можно использовать и другие способы.

Для обработки запроса размерно-пакетной сигнализации или в случае максимума, равного половинной скорости, посылаемого CDMA2000-системой, HR-режим, обеспечивающий возможность взаимодействия, предусмотрен также для кодека, работающего в оболочке Rate Set I (на скорости, обозначаемой символами I-HR-I). Аналогично случаю оболочки Rate Set II несколько битов должны отбрасываться в системном интерфейсе во время операции "AMR-WB-кодирования - VMR-WB-декодирования" или должны генерироваться в системном интерфейсе во время операции "VMR-WB-кодирования - AMR-WB-декодирования". Распределение битов для скорости 8,85 кбит/с и примерной конфигурации для I-HR-I-кодирования показано в таблице 7.

Параметр	AMR-WB с 8,85 кбит/с	I-HR-I с 4,0 кбит/с
	Количество битов на кадр	Количество битов на кадр
Биты режима половинной скорости	-	-
Флаг PPA	1	0
Параметры LP	46	36
Задержка основного тона	26=8+5+8+5	20
Усиления	24=6+6+6+6	24
Алгебраическая кодовая книга	80=20+20+20+20	0
Итого	177	80

При предлагаемом I-HR-I-режиме 10 битов последних 2 субвекторов, получаемых на второй ступени, при квантовании параметров LP-фильтра отбрасываются или генерируются в системном интерфейсе аналогично вышеописанной работе в оболочке Rate Set II. Задержка основного тона кодируется только с целочисленным разрешением и с распределением битов, обеспечивающим наличие 7, 3, 7 и 3 битов в четырех субкадрах. Вследствие этого работа AMR-WB-кодера и VMR-WB-декодера обеспечивает отбрасывание дробной части основного тона в системном интерфейсе и добавление дифференциальной задержки в 3 бита для 2-го и 4-го субкадров. Индексы алгебраической кодовой книги отбрасываются аналогично тому, как это делается при реализации I-HR-I-решения в оболочке Rate Set II. Информация об энергии сигнала остается нетронутой. Остальная работа в режиме, обеспечивающем возможность взаимодействия, в оболочке Rate Set I аналогична работе в таком же режиме в оболочке Rate Set II, описанной со ссылкой на фиг.12 (применительно к операции VAD/DTX/CNG), и поэтому подробно не описывается.

Хотя настоящее изобретение описано выше путем рассмотрения иллюстративных вариантов его осуществления, в него можно вносить изменения в рамках сущности и объема настоящего изобретения, охарактеризованных в прилагаемой формуле изобретения. Например, хотя иллюстративные варианты осуществления настоящего изобретения описаны применительно к кодированию речевого сигнала, следует иметь в

виду, что эти варианты также применимы к звуковым сигналам, не являющимся речевыми.

ЛИТЕРАТУРА

[1] ITU-T Recommendation G.722.2 "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)", Geneva, 2002.

5 [2] 3GPP TS 26.190, "AMR Wideband Speech Codec; Transcoding Functions," 3GPP Technical Specification.

[3] 3GPP TS 26.192, "AMR Wideband Speech Codec; Comfort Noise Aspects," 3GPP Technical Specification.

10 [4] 3GPP TS 26.193: "AMR Wideband Speech Codec; Source Controlled Rate operation," 3GPP Technical Specification.

[5] M. Jelinek and F. Labonte, "Robust Signal/Noise Discrimination for Wideband Speech and Audio Coding," Proc. IEEE Workshop on Speech Coding, pp. 151-153, Delavan, Wisconsin, USA, September 2000.

15 [6] J. D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE Jour, on Selected Areas in Communications, vol. 6, no. 2, pp. 314-323.

[7] 3GPP2 C.S0030-0, "Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems", 3GPP2 Technical Specification.

20 [8] 3GPP2 C.S0014-0, "Enhanced Variable Rate Codec (EVRC)", 3GPP2 Technical Specification

[9] TIA/EIA/IS-733, "High Rate Speech Service option 17 for Wideband Spread Spectrum Communication Systems". Also 3GPP2 Technical Specification C.S0020-0.

Формула изобретения

1. Способ кодирования дискретизированного речевого сигнала, содержащего речевые кадры, заключающийся в том, что

определяют, является ли текущий кадр дискретизированного речевого сигнала активным речевым кадром или неактивным речевым кадром,

если упомянутый текущий кадр сигнала является активным речевым кадром, то проводят процедуру классификации, чтобы определить, является ли текущий кадр

30 невокализованным кадром, причем упомянутая процедура классификации предусматривает проверку, по меньшей мере, трех из следующих параметров, чтобы определить, является ли текущий кадр невокализованным кадром:

а) меры (r_x , \bar{r}_x) звучания,

б) меры (e_{tilt} , e_t) спектрального наклона,

35 в) вариации (dE) энергии в пределах текущего кадра и

г) относительной энергии (E_{rel}) текущего кадра,

а если текущий кадр классифицирован как невокализованный кадр посредством упомянутой процедуры классификации, кодируют этот текущий кадр с помощью алгоритма кодирования невокализованных сигналов.

40 2. Способ по п.1, в котором меру (\bar{r}_x) звучания определяют как

$$\bar{r}_x = \frac{1}{3}(r_x(0) + r_x(1) + r_x(2)),$$

45 где $r_x(0)$, $r_x(1)$ и $r_x(2)$ соответственно представляют собой нормализованную корреляцию первой половины упомянутого текущего кадра, нормализованную корреляцию первой половины кадра, следующего за упомянутым текущим кадром.

3. Способ по п.2, дополнительно предусматривающий прибавление поправки (r_e) на шум к упомянутой мере (\bar{r}_x) звучания.

4. Способ по п.1, предусматривающий определение количества перцепционных критических полос, характеризующих диапазоны частот в пределах спектра энергии текущего кадра, упорядоченных в соответствии с увеличением частоты от первой перцепционной критической полосы, соответствующей диапазону самых низких частот, до последней перцепционной критической полосы, соответствующей диапазону самых высоких частот, и проведение спектрального анализа текущего кадра для определения

распределения энергии по перцепционным критическим полосам.

5. Способ по п.1, в котором спектральный наклон пропорционален отношению между энергией текущего кадра на низких частотах и энергией текущего кадра на высоких частотах.

5 6. Способ по п.4, предусматривающий вычисление меры (\bar{E}_h), характеризующей энергию текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных критических полос.

7. Способ по п.4, предусматривающий вычисление меры (\bar{E}_1), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос.

8. Способ по п.4, предусматривающий вычисление меры (\bar{E}_1), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос, за исключением первой перцепционной критической полосы.

15 9. Способ по п.7, дополнительно предусматривающий определение периода основного тона речи, а для периодов основного тона речи, которые меньше, чем предварительно определенное значение, - вычисление меры (\bar{E}_1) энергии на низких частотах путем суммирования в пределах интервалов дискретизации по частоте, получаемых в результате спектрального анализа текущего кадра, и учета только интервалов дискретизации по частоте, достаточно близких к гармоникам речи, при суммировании по формуле

$$\bar{E}_1 = \frac{1}{\text{cnt}} \sum_{k=K_{\min}}^{24} E_{\text{BIN}}(k) w_h(k),$$

25 где $E_{\text{BIN}}(k)$ - энергии в пределах интервалов дискретизации, K_{\min} - индекс первого интервала дискретизации по частоте, учитываемый при суммировании, cnt - количество ненулевых членов при суммировании, а $w_h(k)$ задают равным 1, если расстояние между интервалом дискретизации по частоте и ближайшей гармоникой не превышает предварительно определенный порог частоты, и задают $w_h(k)$ равным нулю в противном случае.

30 10. Способ по п.7, дополнительно предусматривающий определение периода основного тона речи, а для периодов основного тона речи, которые больше, чем предварительно определенное значение, - вычисление меры (\bar{E}_1) энергии на низких частотах по формуле

$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{\text{CB}}(k),$$

35 где $E_{\text{CB}}(k)$ - энергия перцепционной критической полосы k .

11. Способ по п.7, дополнительно предусматривающий определение априори невокализованного звука, когда

$$r_x(0)+r_x(1)+r_e < 0,6,$$

40 и вычисление меры (\bar{E}_1) энергии на низких частотах по формуле

$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{\text{CB}}(k),$$

где $E_{\text{CB}}(k)$ - энергия перцепционной критической полосы k .

45 12. Способ по любому из пп.6-11, дополнительно предусматривающий вычисление меры (N_h), характеризующей энергию шума текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных критических полос, вычисление меры (N_l), характеризующей энергию шума текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос,

50 вычитание меры (N_h) шума на высоких частотах из меры (\bar{E}_h) энергии на высоких частотах для получения энергии (E_h) на высоких частотах,

вычитание меры (N_l) шума на низких частотах из меры (\bar{E}_1) энергии на низких частотах для получения энергии (E_l) на низких частотах и

вычисление меры (e_{tilt}) спектрального наклона как отношения, представляющего собой энергию (E_l) на низких частотах, деленную на энергию (E_h) на высоких частотах.

13. Способ по п.12, предусматривающий проведение спектрального анализа по п.4

5 дважды за текущий кадр, один раз для первой половины текущего кадра и один раз для второй половины текущего кадра, и дополнительно предусматривающий вычисление меры (e_{tilt}) спектрального наклона дважды за текущий кадр, по одному разу для каждого спектрального анализа, чтобы получить первую меру ($e_{\text{tilt}}(0)$) спектрального наклона для первой половины текущего кадра и вторую меру ($e_{\text{tilt}}(1)$) спектрального наклона для второй половины текущего кадра.

10 14. Способ по п.13, дополнительно предусматривающий вычисление среднего спектрального наклона (e_t) по формуле

$$e_t = \frac{1}{3}(e_{\text{old}} + e_{\text{tilt}}(0) + e_{\text{tilt}}(1)),$$

15 где e_{old} - мера спектрального наклона, полученная в результате спектрального анализа второй половины предыдущего кадра.

15. Способ по п.1, предусматривающий вычисление относительной энергии (E_{rel}) текущего кадра как разности между энергией (E_t) кадра в дБ и значения (\bar{E}_f) долгосрочной средней энергии кадра.

16. Способ по п.15, предусматривающий вычисление энергии (E_t) кадра по формуле

$$20 E_t = 10 \log \left(\sum_{i=0}^{19} E_{\text{CB}}(i) \right), \text{ дБ,}$$

где $E_{\text{CB}}(i)$ - средние энергии, приходящиеся на критическую полосу.

25 17. Способ по п.15, предусматривающий вычисление значения долгосрочной средней энергии по формуле

$$\bar{E}_f = 0,99\bar{E}_f + 0,01E_t,$$

где \bar{E}_f имеет начальное значение 45 дБ.

30 18. Способ по п.1, дополнительно предусматривающий выбор скорости кодирования в битах из набора доступных скоростей кодирования в битах и кодирование текущего кадра в соответствии с выбранной скоростью кодирования в битах.

19. Способ по п.18, в котором набор доступных скоростей кодирования в битах включает в себя полную скорость кодирования в битах, половинную скорость кодирования в битах, четвертную скорость кодирования в битах и одну восьмую скорости кодирования в битах.

35 20. Способ по п.19, в котором, когда текущий кадр классифицирован как невокализованный кадр, кодируют этот текущий кадр с упомянутой половинной скоростью кодирования в битах с помощью алгоритма кодирования невокализованных сигналов с половинной скоростью.

40 21. Способ по п.19, в котором упомянутая процедура классификации, проводимая, чтобы определить, является ли текущий кадр невокализованным кадром, дополнительно включает в себя определение того, является ли текущий кадр переходным между вокализованной речью и невокализованной речью, и, если текущий кадр является переходным между вокализованной речью и невокализованной речью, кодируют этот текущий кадр с упомянутой половинной скоростью кодирования в битах с помощью

45 алгоритма кодирования невокализованных сигналов с половинной скоростью, а если текущий кадр классифицирован как невокализованная речь и не является переходным между вокализованной речью и невокализованной речью, кодируют этот текущий кадр с упомянутой четвертной скоростью кодирования в битах с помощью алгоритма кодирования невокализованных сигналов с четвертной скоростью.

50 22. Способ по п.1, предусматривающий использование алгоритма генерирования комфортного шума, если определено, что текущий кадр является неактивным речевым кадром.

23. Способ по п.1, предусматривающий использование режима прерывистой передачи,

если определено, что текущий кадр является неактивным речевым кадром.

24. Способ по п.19, предусматривающий определение набора рабочих режимов, причем каждый рабочий режим обеспечивает предварительно определенную среднюю скорость в битах, выбор рабочего режима и кодирование дискретизированного речевого сигнала в соответствии с выбранным рабочим режимом.

25. Способ по п.24, в котором набор рабочих режимов содержит высококачественный режим, имеющий самую высокую среднюю скорость в битах, стандартный режим, имеющий промежуточную среднюю скорость в битах, и экономичный режим, имеющий самую низкую среднюю скорость в битах.

26. Способ по п.25, в котором, когда дискретизированный речевой сигнал кодируют в высококачественном режиме, а текущий кадр классифицирован как невокализованный кадр, этот текущий кадр кодируют с упомянутой половинной скоростью кодирования в битах, когда удовлетворяются следующие условия:

упомянутая мера звучания меньше, чем предварительно определенное первое пороговое значение, и

упомянутая мера спектрального наклона меньше, чем предварительно определенное второе пороговое значение, и

упомянутая вариация энергии меньше, чем предварительно определенное третье пороговое значение.

27. Способ по п.25, в котором, когда дискретизированный речевой сигнал кодируют в стандартном режиме, а текущий кадр классифицирован как невокализованный кадр, этот текущий кадр кодируют с упомянутой половинной скоростью кодирования в битах, когда удовлетворяются следующие условия:

упомянутая мера звучания меньше, чем предварительно определенное четвертое пороговое значение, и

упомянутая мера спектрального наклона меньше, чем предварительно определенное пятое пороговое значение, и

упомянутая вариация энергии меньше, чем предварительно определенное шестое пороговое значение, или упомянутая относительная энергия меньше, чем предварительно

определенное седьмое пороговое значение.

28. Способ по п.27, в котором упомянутое четвертое пороговое значение составляет 0,695, упомянутое пятое пороговое значение составляет 4, упомянутое шестое пороговое значение составляет 40, а упомянутое седьмое пороговое значение составляет -14.

29. Способ по п.25, в котором, когда дискретизированный речевой сигнал кодируют в экономичном режиме, а текущий кадр классифицирован как невокализованный кадр, этот текущий кадр кодируют с упомянутой половинной скоростью кодирования в битах, когда удовлетворяются следующие условия:

упомянутая мера звучания меньше, чем предварительно определенное восьмое пороговое значение, и

упомянутая мера спектрального наклона меньше, чем предварительно определенное девятое пороговое значение, и

упомянутая вариация энергии меньше, чем предварительно определенное десятое пороговое значение, или упомянутая относительная энергия меньше, чем предварительно определенное одиннадцатое пороговое значение.

30. Способ по п.29, в котором упомянутое восьмое пороговое значение составляет 0,695, упомянутое девятое пороговое значение составляет 4, упомянутое десятое пороговое значение составляет 60, а упомянутое одиннадцатое пороговое значение составляет -14.

31. Способ по п.25, в котором, когда дискретизированный речевой сигнал кодируют в экономичном режиме, а текущий кадр классифицирован как невокализованный кадр, этот текущий кадр кодируют с упомянутой четвертной скоростью кодирования в битах, когда удовлетворяются следующие дополнительные условия:

нормализованная корреляция ($r_x(2)$) в кадре упреждающей выборки меньше, чем

предварительно определенное двенадцатое пороговое значение, и вторая мера ($e_{\text{tilt}}(1)$) спектрального наклона для второй половины текущего кадра меньше, чем предварительно определенное тринадцатое пороговое значение.

32. Способ по п.31, в котором упомянутое двенадцатое пороговое значение составляет 0,73, а упомянутое тринадцатое пороговое значение составляет 3.

33. Устройство для кодирования дискретизированного речевого сигнала, содержащее речевые кадры, содержащее

детектор речевой активности для определения того, являются ли кадры дискретизированного речевого сигнала активными речевыми кадрами или неактивными речевыми кадрами,

блок классификации, выполненный с возможностью проведения процедуры классификации на активных речевых кадрах для определения того, являются ли упомянутые активные речевые кадры невокализованными кадрами, причем упомянутая процедура классификации предусматривает проверку, по меньшей мере, трех из следующих параметров, чтобы определить, является ли текущий кадр невокализованным кадром:

а) меры (r_x , \bar{r}_x) звучания,

б) меры (e_{tilt} , e_t) спектрального наклона,

в) вариации (dE) энергии в пределах текущего кадра и

г) относительной энергии (E_{rel}) текущего кадра,

при этом упомянутое устройство выполнено с возможностью кодирования текущего кадра с помощью алгоритма кодирования невокализованных сигналов, если блок классификации классифицирует текущий кадр как невокализованный кадр.

34. Устройство по п.33, в котором мера (\bar{r}_x) звучания определяется как

$$\bar{r}_x = \frac{1}{3}(r_x(0) + r_x(1) + r_x(2)),$$

где $r_x(0)$, $r_x(1)$ и $r_x(2)$ соответственно представляют собой нормализованную корреляцию первой половины упомянутого текущего кадра, нормализованную корреляцию второй половины упомянутого текущего кадра и нормализованную корреляцию первой половины кадра, следующего за упомянутым текущим кадром.

35. Устройство по п.34, выполненное с дополнительной возможностью прибавления поправки (r_e) на шум к упомянутой мере (\bar{r}_x) звучания.

36. Устройство по п.33, выполненное с возможностью определения количества перцепционных критических полос, характеризующих диапазоны частот в пределах спектра энергии текущего кадра, упорядоченных в соответствии с увеличением частоты от первой перцепционной критической полосы, соответствующей диапазону самых низких частот, до последней перцепционной критической полосы, соответствующей диапазону самых высоких частот, и проведения спектрального анализа текущего кадра для определения распределения энергии по перцепционным критическим полосам.

37. Устройство по п.33, в котором спектральный наклон пропорционален отношению между энергией текущего кадра на низких частотах и энергией текущего кадра на высоких частотах.

38. Устройство по п.36, выполненное с возможностью вычисления меры (\bar{E}_h), характеризующей энергию текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных критических полос.

39. Устройство по п.36, выполненное с возможностью вычисления меры (\bar{E}_l), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос.

40. Устройство по п.36, выполненное с возможностью вычисления меры (\bar{E}_1), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос, за исключением первой перцепционной критической полосы.

41. Устройство по п.39, выполненное с возможностью определения периода основного тона речи, а для периодов основного тона речи, которые меньше, чем предварительно определенное значение, - с возможностью вычисления меры (\bar{E}_1) энергии на низких частотах путем суммирования в пределах интервалов дискретизации по частоте, получаемых в результате спектрального анализа текущего кадра, и учета только интервалов дискретизации по частоте, достаточно близких к гармоникам речи, при суммировании по формуле

$$\bar{E}_1 = \frac{1}{\text{cnt}} \sum_{k=K_{\min}}^{24} E_{\text{BIN}}(k) w_h(k),$$

где $E_{\text{BIN}}(k)$ - энергии в пределах интервалов дискретизации, K_{\min} - индекс первого интервала дискретизации по частоте, учитываемый при суммировании, cnt - количество ненулевых членов при суммировании, а $w_h(k)$ задано равным 1, если расстояние между интервалом дискретизации по частоте и ближайшей гармоникой не превышает предварительно определенный порог частоты, и задано $w_h(k)$ равным нулю в противном случае.

42. Устройство по п.39, выполненное с дополнительной возможностью определения периода основного тона речи, а для периодов основного тона речи, которые больше, чем предварительно определенное значение, - с возможностью вычисления меры (\bar{E}_1) энергии на низких частотах по формуле

$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{\text{CB}}(k),$$

где $E_{\text{CB}}(k)$ - энергия перцепционной критической полосы k.

43. Устройство по п.39, выполненное с дополнительной возможностью определения априори невокализованного звука, когда

$$r_x(0)+r_x(1)+r_e < 0,6,$$

и вычисления меры (\bar{E}_1) энергии на низких частотах по формуле

$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{\text{CB}}(k),$$

где $E_{\text{CB}}(k)$ - энергия перцепционной критической полосы k.

44. Устройство по любому из пп.38-43, выполненное с дополнительными возможностями вычисления меры (N_h), характеризующей энергию шума текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных критических полос,

вычисления меры (N_l), характеризующей энергию шума текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос,

вычитания меры (N_h) шума на высоких частотах из меры (\bar{E}_h) энергии на высоких частотах для получения энергии (E_h) на высоких частотах,

вычитания меры (N_l) шума на низких частотах из меры (\bar{E}_l) энергии на низких частотах для получения энергии (E_l) на низких частотах и

вычисления меры (e_{tilt}) спектрального наклона как отношения, представляющего собой энергию (E_l) на низких частотах, деленную на энергию (E_h) на высоких частотах.

45. Устройство по п.44, выполненное с возможностью проведения спектрального анализа по п.36 дважды за текущий кадр, один раз для первой половины текущего кадра и один раз для второй половины текущего кадра, а также с возможностью вычисления меры (e_{tilt}) спектрального наклона дважды за текущий кадр, по одному разу для каждого спектрального анализа, чтобы получить первую меру ($e_{\text{tilt}}(0)$) спектрального наклона для первой половины текущего кадра и вторую меру ($e_{\text{tilt}}(1)$) спектрального наклона для второй половины текущего кадра.

46. Устройство по п.45, выполненное с дополнительной возможностью вычисления среднего спектрального наклона (e_t) по формуле

$$e_t = \frac{1}{3}(e_{old} + e_{tit}(0) + e_{tit}(1)),$$

где e_{old} - мера спектрального наклона, полученная в результате спектрального анализа второй половины предыдущего кадра.

5 47. Устройство по п.33, выполненное с возможностью вычисления относительной энергии (E_{rel}) текущего кадра как разности между энергией (E_t) кадра в дБ и значения (\bar{E}_r) долгосрочной средней энергии кадра.

48. Устройство по п.47, выполненное с возможностью вычисления энергии (E_t) кадра по формуле

$$10 E_t = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \text{ дБ,}$$

где $E_{CB}(i)$ - средние энергии, приходящиеся на критическую полосу.

49. Устройство по п.47, выполненное с возможностью вычисления значения долгосрочной средней энергии по формуле

$$15 \bar{E}_r = 0,99\bar{E}_r + 0,01E_t,$$

где \bar{E}_r имеет начальное значение 45 дБ.

50. Устройство по п.33, выполненное с возможностью выбора скорости кодирования в битах из набора доступных скоростей кодирования в битах и кодирования текущего кадра в соответствии с выбранной скоростью кодирования в битах.

20 51. Устройство по п.50, в котором набор доступных скоростей кодирования в битах включает в себя полную скорость кодирования в битах, половинную скорость кодирования в битах, четвертную скорость кодирования в битах и одну восьмую скорости кодирования в битах.

25 52. Устройство по п.51, выполненное с возможностью кодирования текущего кадра с упомянутой половинной скоростью кодирования в битах с помощью алгоритма кодирования невокализованных сигналов с половинной скоростью, когда этот текущий кадр классифицирован как невокализованный кадр.

30 53. Устройство по п.51, выполненное с дополнительной возможностью определения того, является ли текущий кадр переходным между вокализованной речью и невокализованной речью, и если текущий кадр является переходным между вокализованной речью и невокализованной речью, - с возможностью кодирования этого текущего кадра с упомянутой половинной скоростью кодирования в битах с помощью алгоритма кодирования невокализованных сигналов с половинной скоростью, а если

35 54. Устройство по п.33, выполненное с возможностью использования алгоритма генерирования комфортного шума, когда определено, что текущий кадр является неактивным речевым кадром.

40 55. Устройство по п.33, выполненное с возможностью использования режима прерывистой передачи, когда определено, что текущий кадр является неактивным речевым кадром.

45 56. Устройство по п.51, выполненное с возможностью определения набора рабочих режимов, причем каждый рабочий режим обеспечивает предварительно определенную среднюю скорость в битах, а также с возможностью выбора рабочего режима и кодирования дискретизированного речевого сигнала в соответствии с выбранным рабочим режимом.

50 57. Устройство по п.56, в котором набор рабочих режимов содержит высококачественный режим, имеющий самую высокую среднюю скорость в битах, стандартный режим, имеющий промежуточную среднюю скорость в битах, и экономичный режим, имеющий самую низкую среднюю скорость в битах.

58. Устройство по п.57, выполненное с возможностью кодирования текущего кадра с

упомянутой половинной скоростью кодирования в битах, когда дискретизированный речевой сигнал кодируется в высококачественном режиме, а текущий кадр классифицирован как невокализованный кадр, и удовлетворяются следующие условия:

- 5 упомянутая мера звучания меньше, чем предварительно определенное первое пороговое значение, и
 упомянутая мера спектрального наклона меньше, чем предварительно определенное второе пороговое значение, и
 упомянутая вариация энергии меньше, чем предварительно определенное третье пороговое значение.

10 59. Устройство по п.57, выполненное с возможностью кодирования текущего кадра с упомянутой половинной скоростью кодирования в битах, когда дискретизированный речевой сигнал кодируется в стандартном режиме, а текущий кадр классифицирован как невокализованный кадр, и удовлетворяются следующие условия:

- 15 упомянутая мера звучания меньше, чем предварительно определенное четвертое пороговое значение, и
 упомянутая мера спектрального наклона меньше, чем предварительно определенное пятое пороговое значение, и
 упомянутая вариация энергии меньше, чем предварительно определенное шестое пороговое значение, или упомянутая относительная энергия меньше, чем предварительно
 20 определенное седьмое пороговое значение.

60. Устройство по п.59, в котором упомянутое четвертое пороговое значение составляет 0,695, упомянутое пятое пороговое значение составляет 4, упомянутое шестое пороговое значение составляет 40, а упомянутое седьмое пороговое значение составляет -14.

25 61. Устройство по п.57, выполненное с возможностью кодирования текущего кадра с упомянутой половинной скоростью кодирования в битах, когда дискретизированный речевой сигнал кодируется в экономичном режиме, а текущий кадр классифицирован как невокализованный кадр, и удовлетворяются следующие условия:

- 30 упомянутая мера звучания меньше, чем предварительно определенное восьмое пороговое значение, и
 упомянутая мера спектрального наклона меньше, чем предварительно определенное девятое пороговое значение, и
 упомянутая вариация энергии меньше, чем предварительно определенное десятое пороговое значение, или упомянутая относительная энергия меньше, чем предварительно
 35 определенное одиннадцатое пороговое значение.

62. Устройство по п.61, в котором упомянутое восьмое пороговое значение составляет 0,695, упомянутое девятое пороговое значение составляет 4, упомянутое десятое пороговое значение составляет 60, а упомянутое одиннадцатое пороговое значение составляет - 14.

40 63. Устройство по п.57, выполненное с возможностью кодирования текущего кадра с упомянутой четвертной скоростью кодирования в битах, когда дискретизированный речевой сигнал кодируется в экономичном режиме, а текущий кадр классифицирован как невокализованный кадр, и удовлетворяются следующие дополнительные условия:

- 45 нормализованная корреляция ($r_x(2)$) в кадре упреждающей выборки меньше, чем предварительно определенное двенадцатое пороговое значение, и
 вторая мера ($e_{\text{tilt}}(1)$) спектрального наклона для второй половины текущего кадра меньше, чем предварительно определенное тринадцатое пороговое значение.

64. Устройство по п.63, в котором упомянутое двенадцатое пороговое значение составляет 0,73, а упомянутое тринадцатое пороговое значение составляет 3.

50 65. Устройство для кодирования дискретизированного речевого сигнала, содержащего речевые кадры, содержащее средство для определения того, является ли текущий кадр дискретизированного речевого сигнала активным речевым кадром или неактивным речевым кадром,

средство, реагирующее на упомянутый текущий кадр, являющийся активным речевым кадром, для проведения процедуры классификации для определения того, является ли текущий кадр невокализованным кадром, причем упомянутая процедура классификации предусматривает проверку, по меньшей мере, трех из следующих параметров, чтобы

5 определить, является ли текущий кадр невокализованным кадром:

- а) меры (r_x , \bar{r}_x) звучания,
- б) меры (e_{tilt} , e_t) спектрального наклона,
- в) вариации (dE) энергии в пределах текущего кадра и
- г) относительной энергии (E_{rel}) текущего кадра,

10 и средство для кодирования текущего кадра с помощью алгоритма кодирования невокализованных сигналов, когда текущий кадр классифицирован как невокализованный кадр посредством упомянутой процедуры классификации.

66. Речевой кодер, реагирующий на текущий кадр, классифицируемый как активный речевой кадр, для кодирования упомянутого текущего кадра, причем упомянутый кодер

15 содержит средство для выполнения процедуры классификации для определения, является ли текущий кадр активным невокализованным речевым кадром путем проверки, по меньшей мере, трех параметров, выбранных из набора: мера (r_x , \bar{r}_x) звучания, мера (e_{tilt} , e_t) спектрального наклона, вариация (dE) энергии в пределах текущего кадра и

20 относительная энергия (E_{rel}) текущего кадра, и

средство для кодирования текущего кадра с использованием алгоритма кодирования невокализованных сигналов, если текущий кадр классифицирован как активный невокализованный речевой кадр посредством процедуры классификации.

67. Носитель информации, содержащий машиночитаемые команды, исполняемые

25 цифровым процессором данных, для осуществления действий, направленных на кодирование дискретизированного речевого сигнала, содержащего речевые кадры, причем эти действия включают

определение, является ли текущий кадр дискретизированного речевого сигнала

30 активным речевым кадром или неактивным речевым кадром, выполнение процедуры классификации на активном речевом кадре, чтобы определить, является ли текущий кадр невокализованным кадром, причем упомянутая процедура классификации предусматривает проверку, по меньшей мере, трех из следующих параметров, чтобы определить, является ли текущий кадр невокализованным кадром:

- а) меры (r_x , \bar{r}_x) звучания,
- б) меры (e_{tilt} , e_t) спектрального наклона,
- в) вариации (dE) энергии в пределах текущего кадра и
- г) относительной энергии (E_{rel}) текущего кадра,

и кодирование текущего кадра с помощью алгоритма кодирования невокализованных сигналов, если текущий кадр классифицирован как невокализованный кадр посредством

40 упомянутой процедуры классификации.

68. Носитель информации по п.67, в котором мера (\bar{r}_x) звучания определяется как

$$\bar{r}_x = \frac{1}{3}(r_x(0) + r_x(1) + r_x(2)),$$

45 где $r_x(0)$, $r_x(1)$ и $r_x(2)$ соответственно представляют собой нормализованную корреляцию первой половины упомянутого текущего кадра, нормализованную корреляцию второй половины упомянутого текущего кадра и нормализованную корреляцию первой половины кадра, следующего за упомянутым текущим кадром.

69. Носитель информации по п.68, в котором упомянутые действия дополнительно включают прибавление поправки (r_e) на шум к упомянутой мере (\bar{r}_x) звучания.

50 70. Носитель информации по п.67, в котором упомянутые действия дополнительно включают определение количества перцепционных критических полос, характеризующих диапазоны частот в пределах спектра энергии текущего кадра, упорядоченных в соответствии с увеличением частоты от первой перцепционной критической полосы,

соответствующей диапозону самых низких частот, до последней перцепционной критической полосы, соответствующей диапозону самых высоких частот, и выполнение спектрального анализа текущего кадра для определения распределения энергии по перцепционным критическим полосам.

5 71. Носитель информации по п.67, в котором спектральный наклон пропорционален отношению между энергией текущего кадра на низких частотах и энергией текущего кадра на высоких частотах.

72. Носитель информации по п.70, в котором упомянутые действия дополнительно включают вычисление меры (\bar{E}_n), характеризующей энергию текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных критических полос.

10 73. Носитель информации по п.70, в котором упомянутые действия дополнительно включают вычисление меры (\bar{E}_1), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос.

15 74. Носитель информации по п.70, в котором упомянутые действия дополнительно включают вычисление меры (\bar{E}_1), характеризующей энергию текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос, за исключением первой перцепционной критической полосы.

20 75. Носитель информации по п.73, в котором упомянутые действия дополнительно включают определение периода основного тона речи, а для периодов основного тона речи, которые меньше, чем предварительно определенное значение, - вычисление меры (\bar{E}_1) энергии на низких частотах путем суммирования в пределах интервалов дискретизации по частоте, получаемых в результате спектрального анализа текущего кадра, и учета только

25 интервалов дискретизации по частоте, достаточно близких к гармоникам речи, при суммировании по формуле

$$\bar{E}_1 = \frac{1}{cnt} \sum_{k=K_{min}}^{24} E_{BIN}(k) w_h(k),$$

30 где $E_{BIN}(k)$ - энергии в пределах интервалов дискретизации, K_{min} - индекс первого интервала дискретизации по частоте, учитываемый при суммировании, cnt - количество ненулевых членов при суммировании, а $w_h(k)$ задается равным 1, если расстояние между интервалом дискретизации по частоте и ближайшей гармоникой не превышает

35 предварительно определенный порог частоты, и задается $w_h(k)$ равным нулю в противном случае.

76. Носитель информации по п.73, в котором упомянутые действия дополнительно включают определение периода основного тона речи, а для периодов основного тона речи, которые больше, чем предварительно определенное значение, - вычисление меры (\bar{E}) энергии на низких частотах по формуле

40
$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k),$$

где $E_{CB}(k)$ - энергия перцепционной критической полосы k .

77. Носитель информации по п.73, в которой упомянутые действия дополнительно включают определение априори невокализованного звука, если

45
$$r_x(0)+r_x(1)+r_e < 0,6,$$

и вычисление меры (\bar{E}_1) энергии на низких частотах по формуле

50
$$\bar{E}_1 = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k),$$

где $E_{CB}(k)$ - энергия перцепционной критической полосы k .

78. Носитель информации по любому из пп.72-77, в котором упомянутые действия дополнительно включают

вычисление меры (N_h), характеризующей энергию шума текущего кадра на высоких частотах, путем вычисления средней из энергий последних двух перцепционных

критических полос,

вычисление меры (N_i), характеризующей энергию шума текущего кадра на низких частотах, путем вычисления средней из энергий первых i перцепционных критических полос,

5 вычитание меры (N_h) шума на высоких частотах из меры (\bar{E}_h) энергии на высоких частотах для получения энергии (E_h) на высоких частотах,

вычитание меры (N_l) шума на низких частотах из меры (\bar{E}_l) энергии на низких частотах для получения энергии (E_l) на низких частотах и

10 вычисление меры (e_{tilt}) спектрального наклона как отношения, представляющего собой энергию (E_l) на низких частотах, деленную на энергию (E_h) на высоких частотах.

79. Носитель информации по п.78, в котором упомянутые действия дополнительно включают проведение спектрального анализа по п.36 дважды за текущий кадр, один раз - для первой половины текущего кадра и один раз - для второй половины текущего кадра, и дополнительно включают вычисление меры (e_{tilt}) спектрального наклона дважды за

15 текущий кадр, по одному разу для каждого спектрального анализа, чтобы получить первую меру ($e_{\text{tilt}}(0)$) спектрального наклона для первой половины текущего кадра и вторую меру ($e_{\text{tilt}}(1)$) спектрального наклона для второй половины текущего кадра.

80. Носитель информации по п.79, в котором упомянутые действия дополнительно включают вычисление среднего спектрального наклона (e_t) по формуле

$$20 e_t = \frac{1}{3}(e_{\text{old}} + e_{\text{tilt}}(0) + e_{\text{tilt}}(1)),$$

где e_{old} - мера спектрального наклона, полученная в результате спектрального анализа второй половины предыдущего кадра.

81. Носитель информации по п.67, в котором упомянутые действия дополнительно

25 включают вычисление относительной энергии (E_{rel}) текущего кадра как разности между энергией (E_t) кадра в дБ и значения (\bar{E}_r) долгосрочной средней энергии кадра.

82. Носитель информации по п.81, в котором упомянутые действия дополнительно включают вычисление энергии (E_t) кадра по формуле

$$30 E_t = 10 \log \left(\sum_{i=0}^{19} E_{\text{CB}}(i) \right), \text{ дБ},$$

где $E_{\text{CB}}(i)$ - средние энергии, приходящиеся на критическую полосу.

83. Носитель информации по п.81, в котором упомянутые действия включают

35 вычисление значения долгосрочной средней энергии по формуле

$$\bar{E}_r = 0,99\bar{E}_r + 0,01E_t,$$

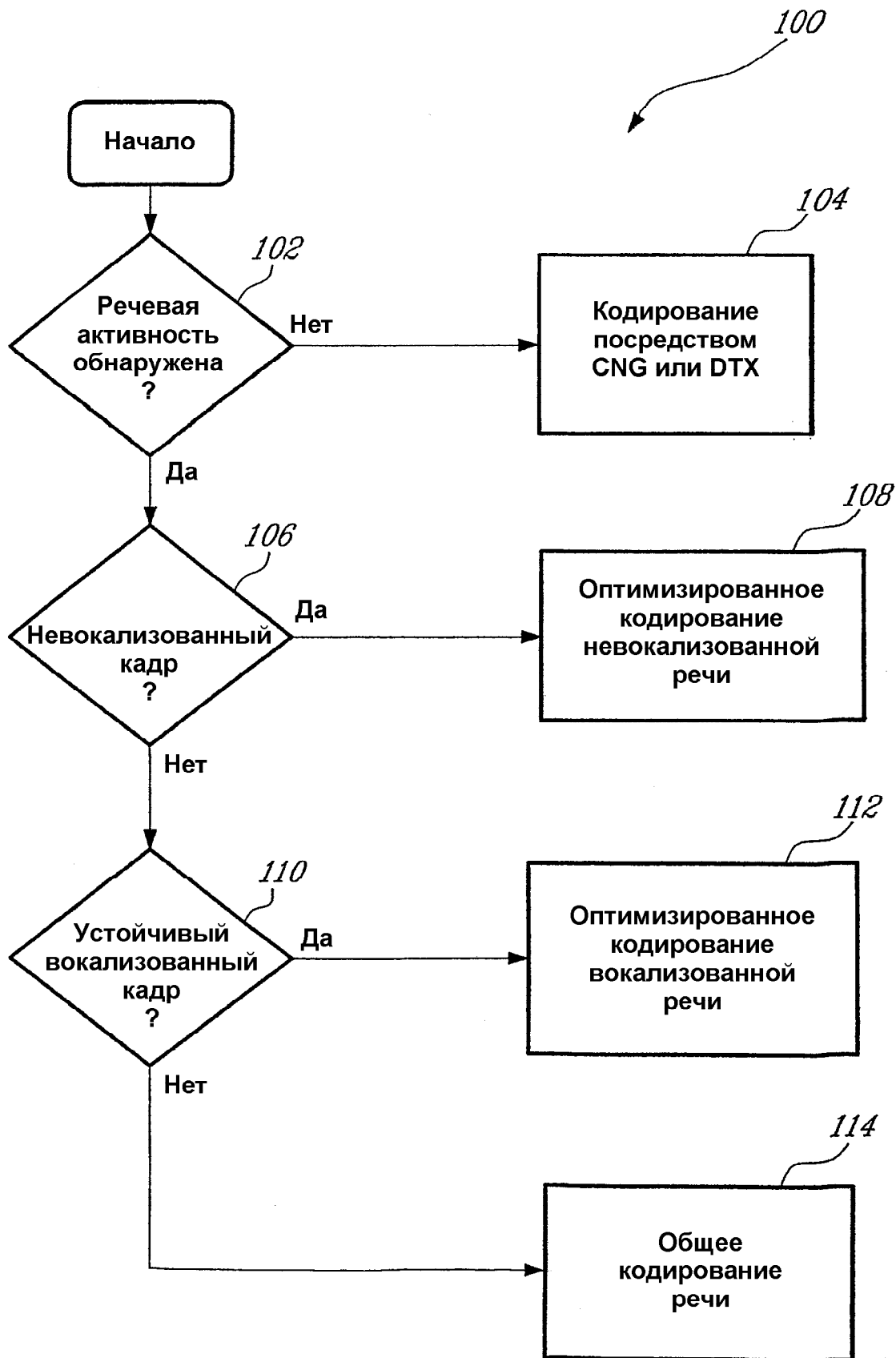
где \bar{E}_r имеет начальное значение 45 дБ.

84. Носитель информации по п.67, причем упомянутый носитель информации и цифровой процессор данных размещены внутри мобильной станции.

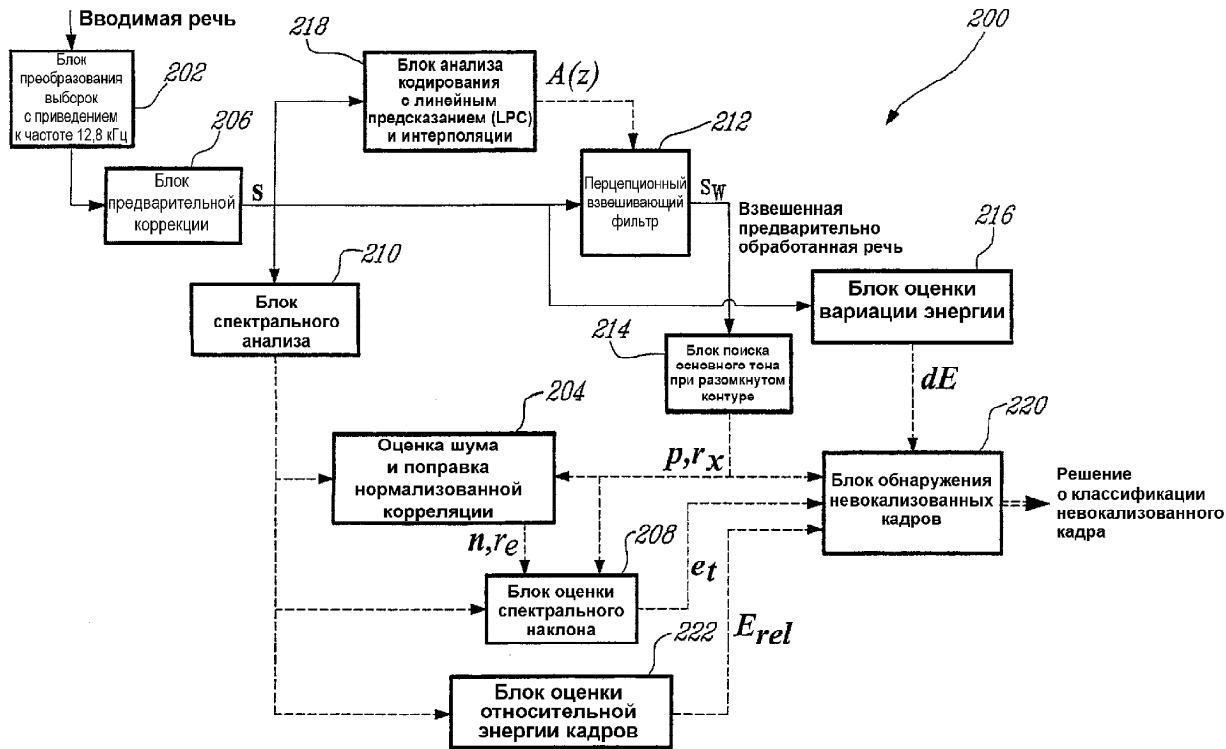
40

45

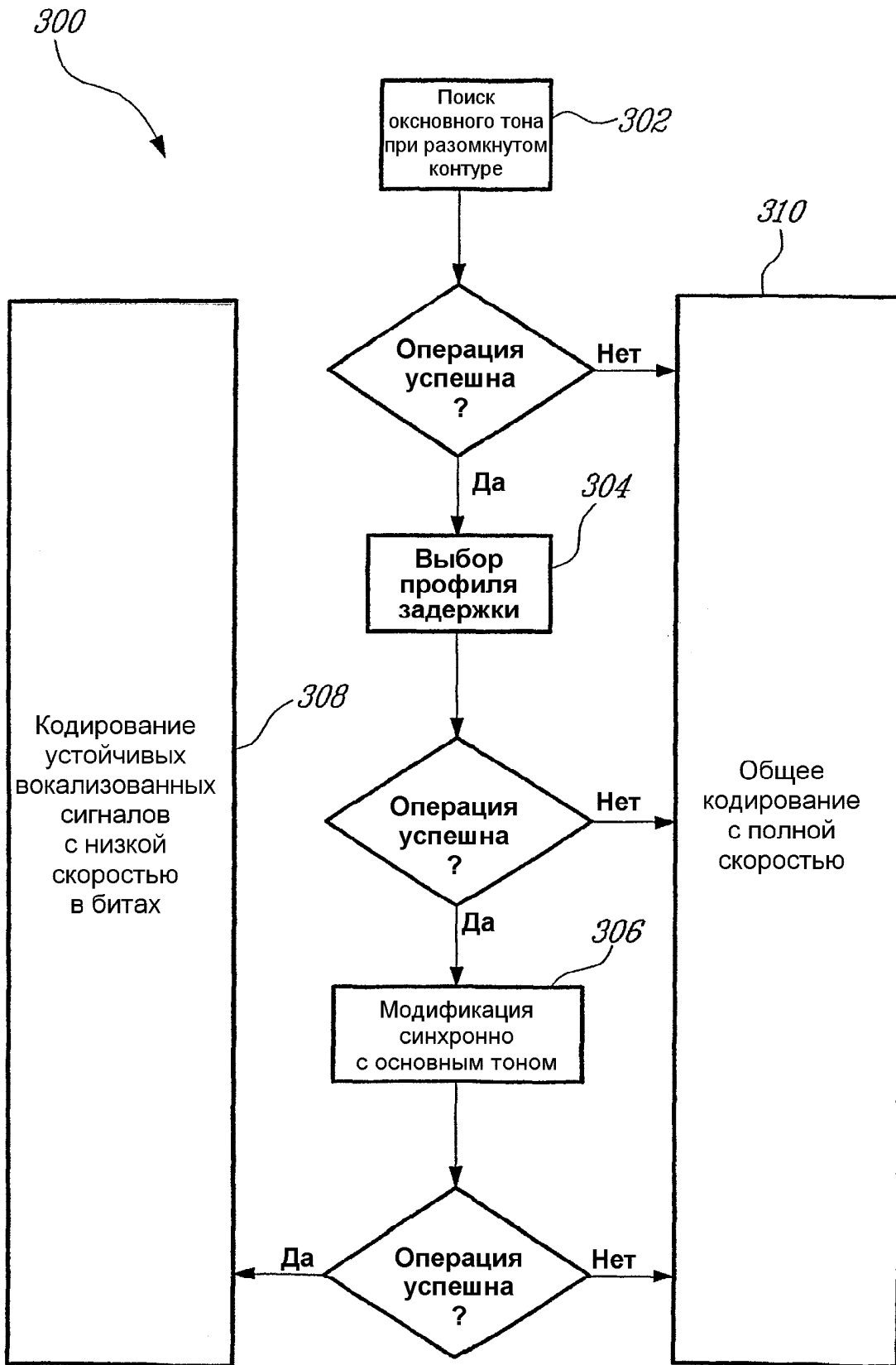
50



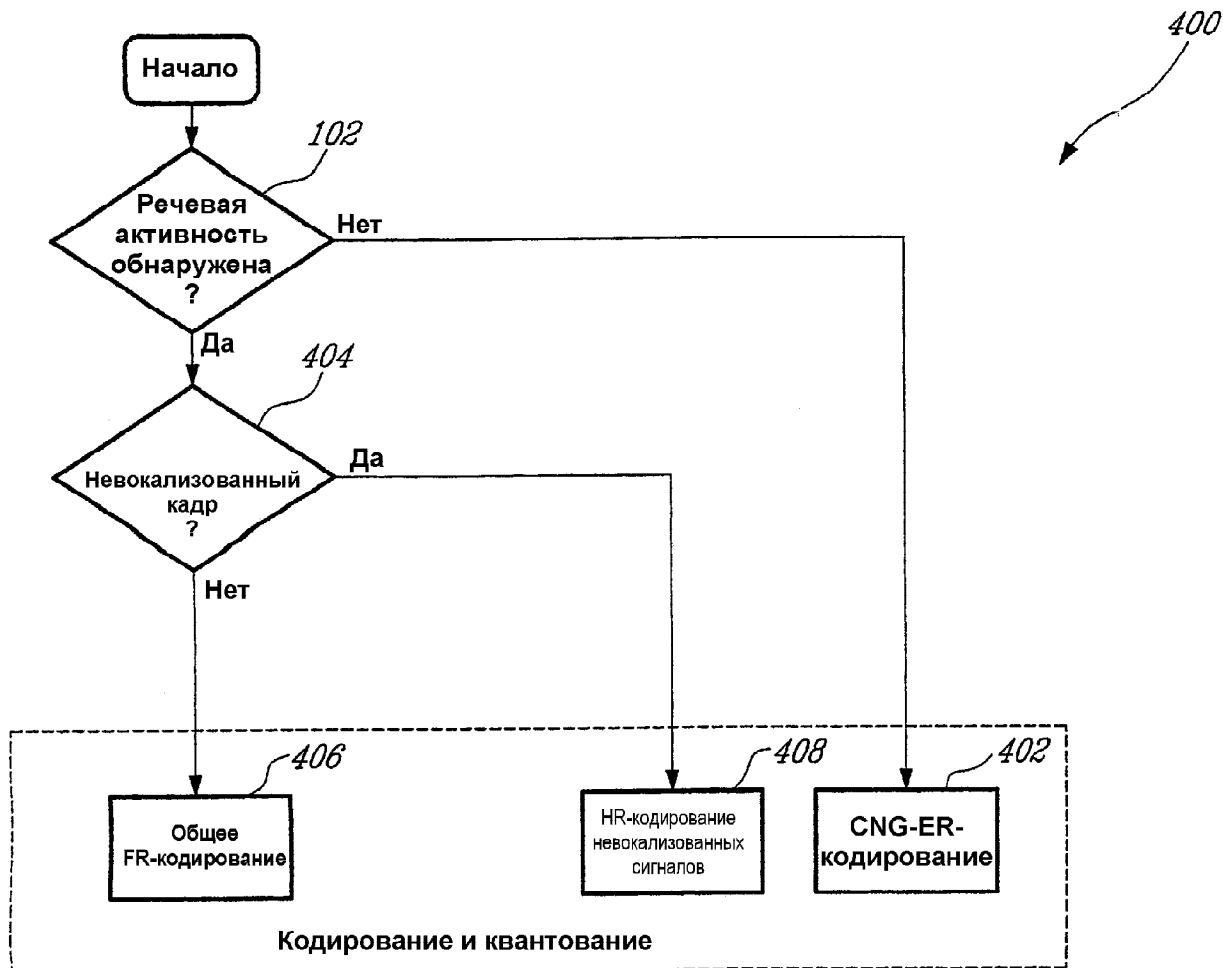
Фиг. 2



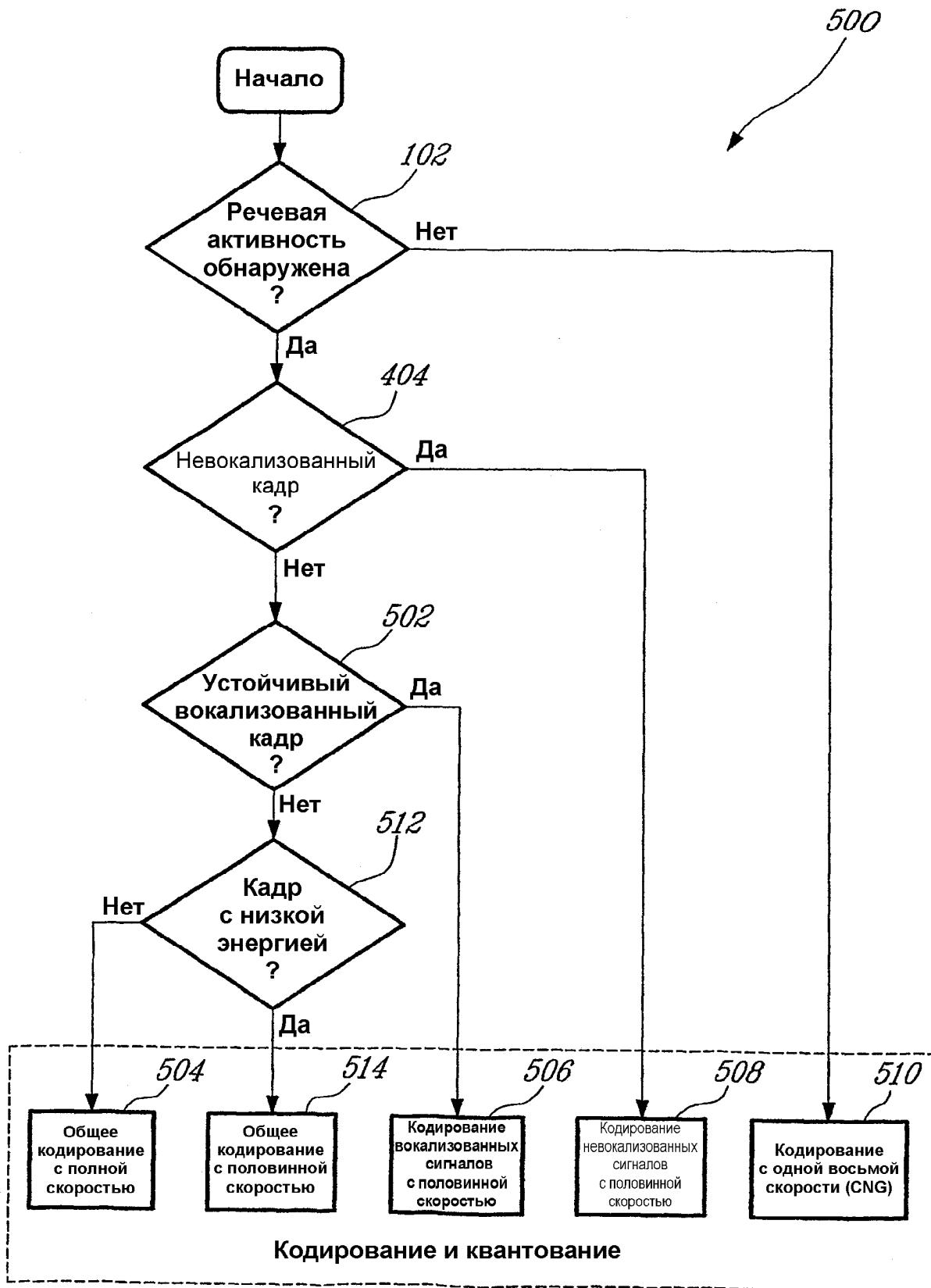
Фиг. 3



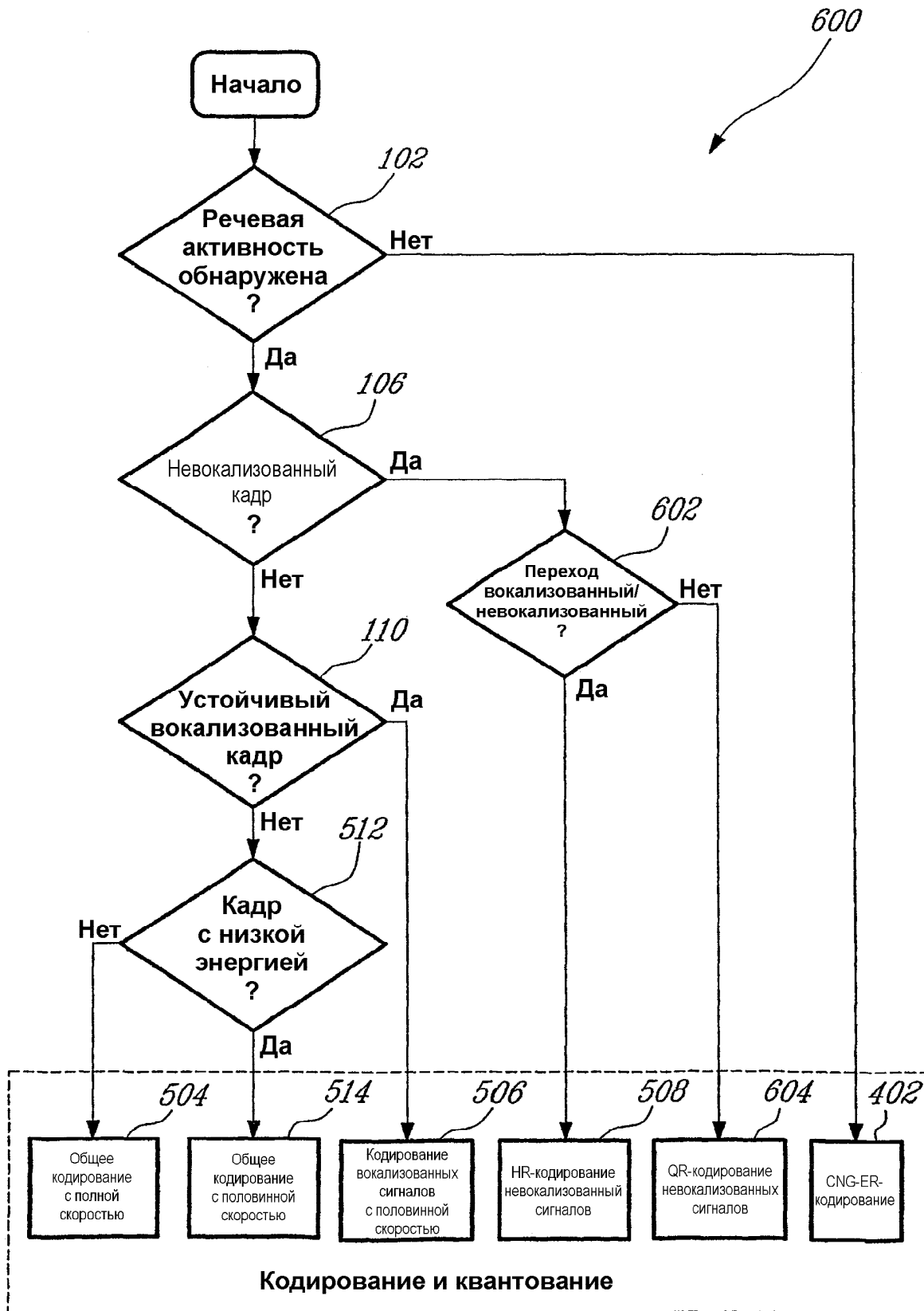
Фиг. 4



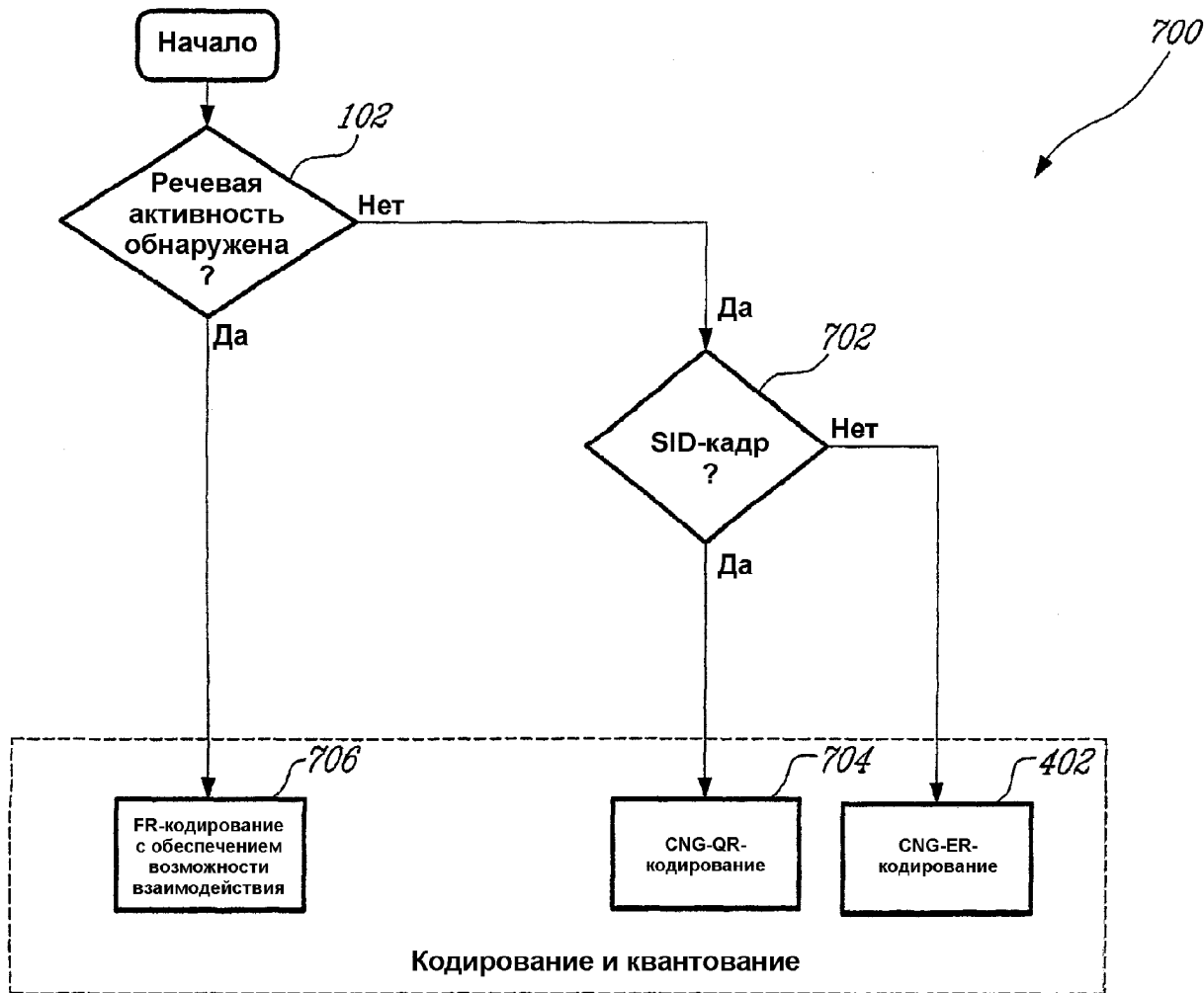
Фиг. 5



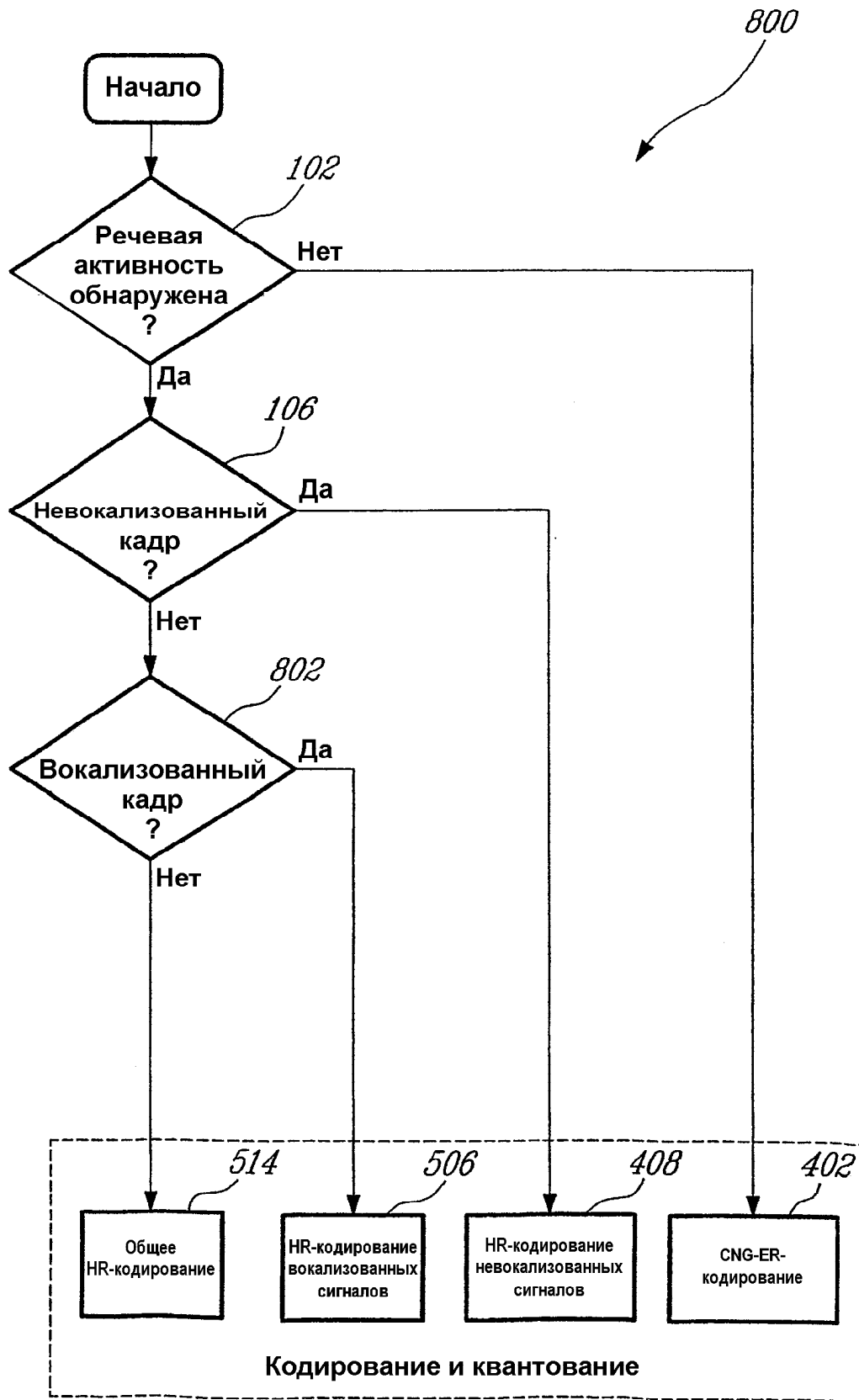
Фиг. 6



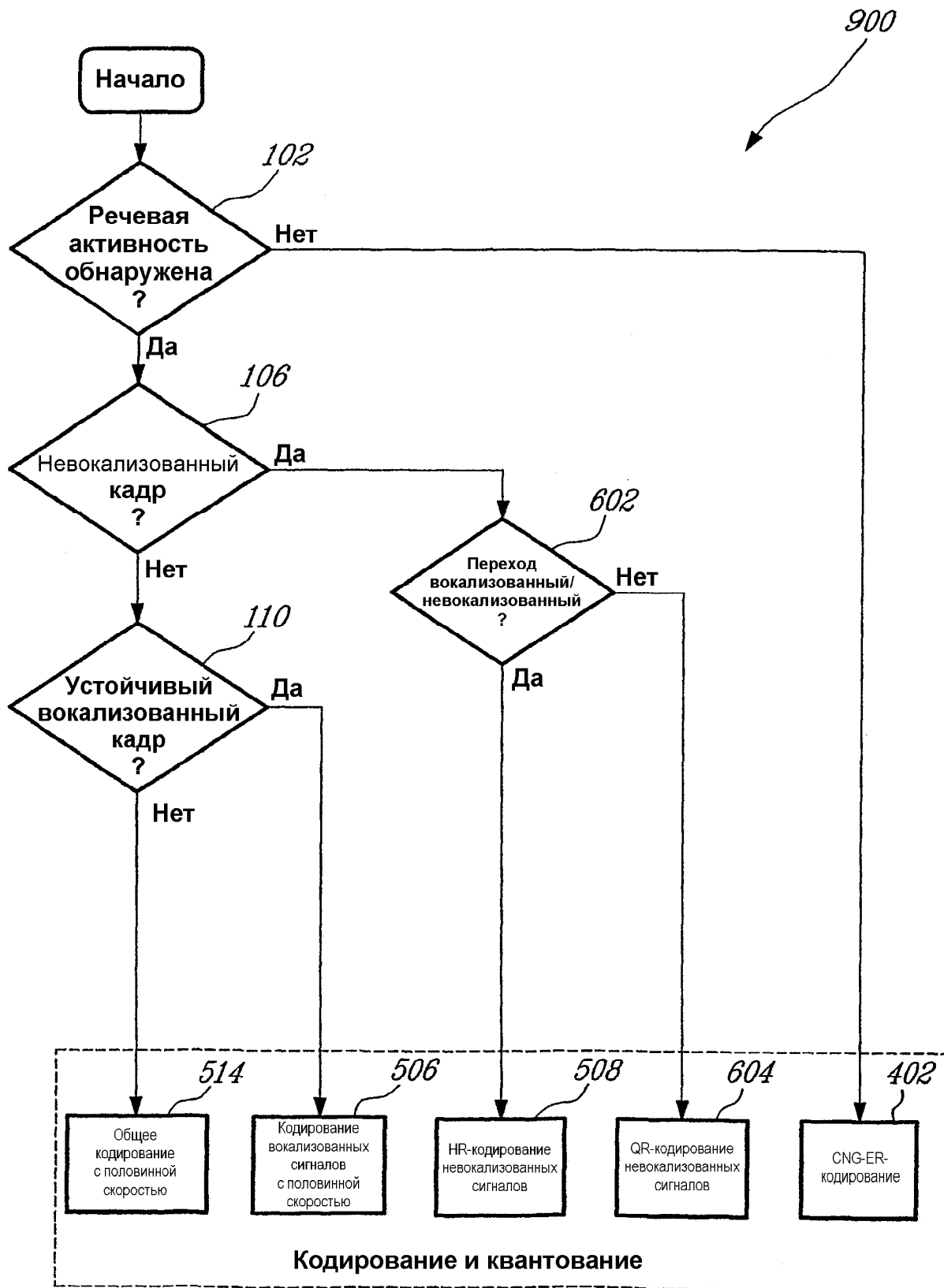
Фиг. 7



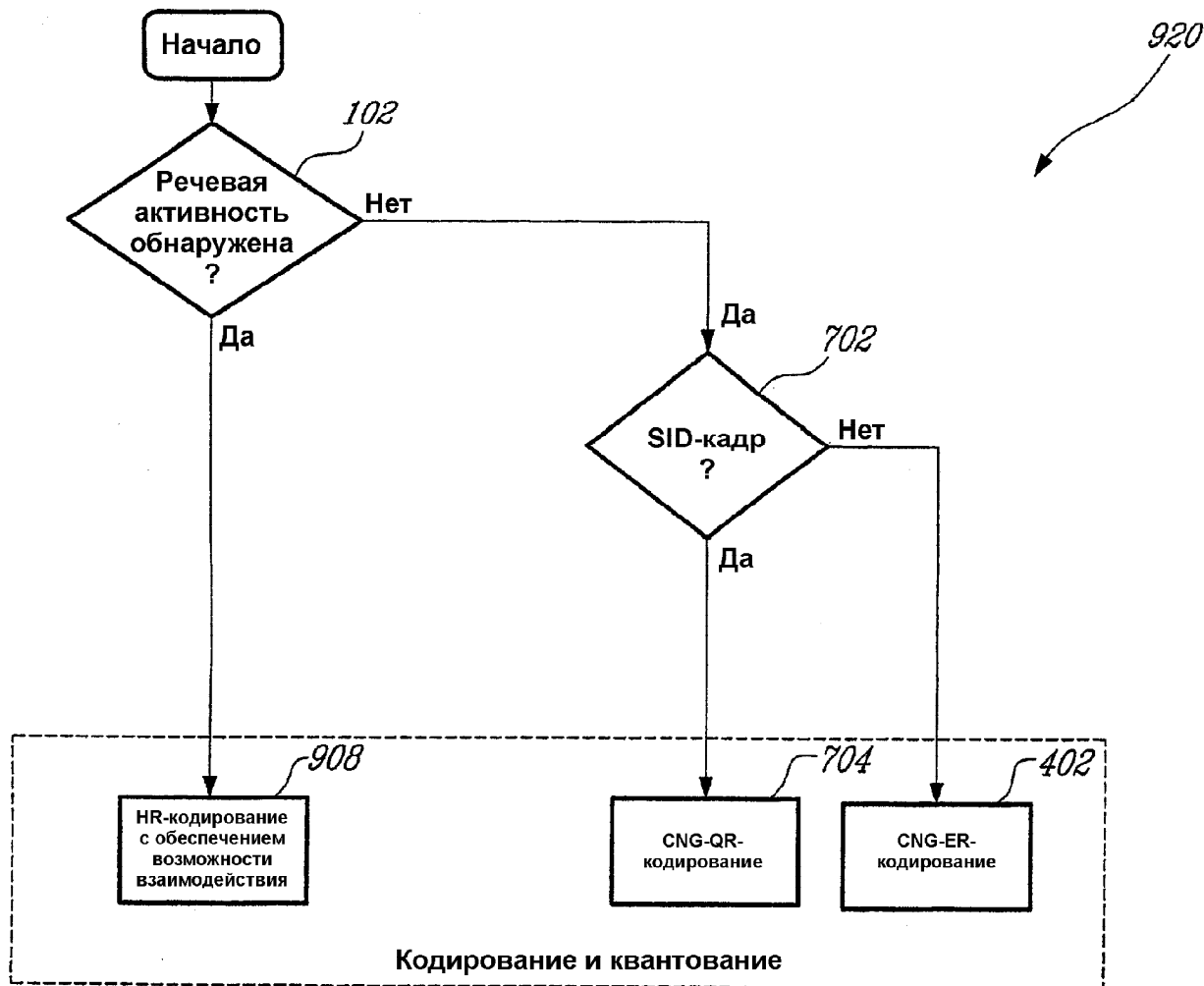
Фиг. 8



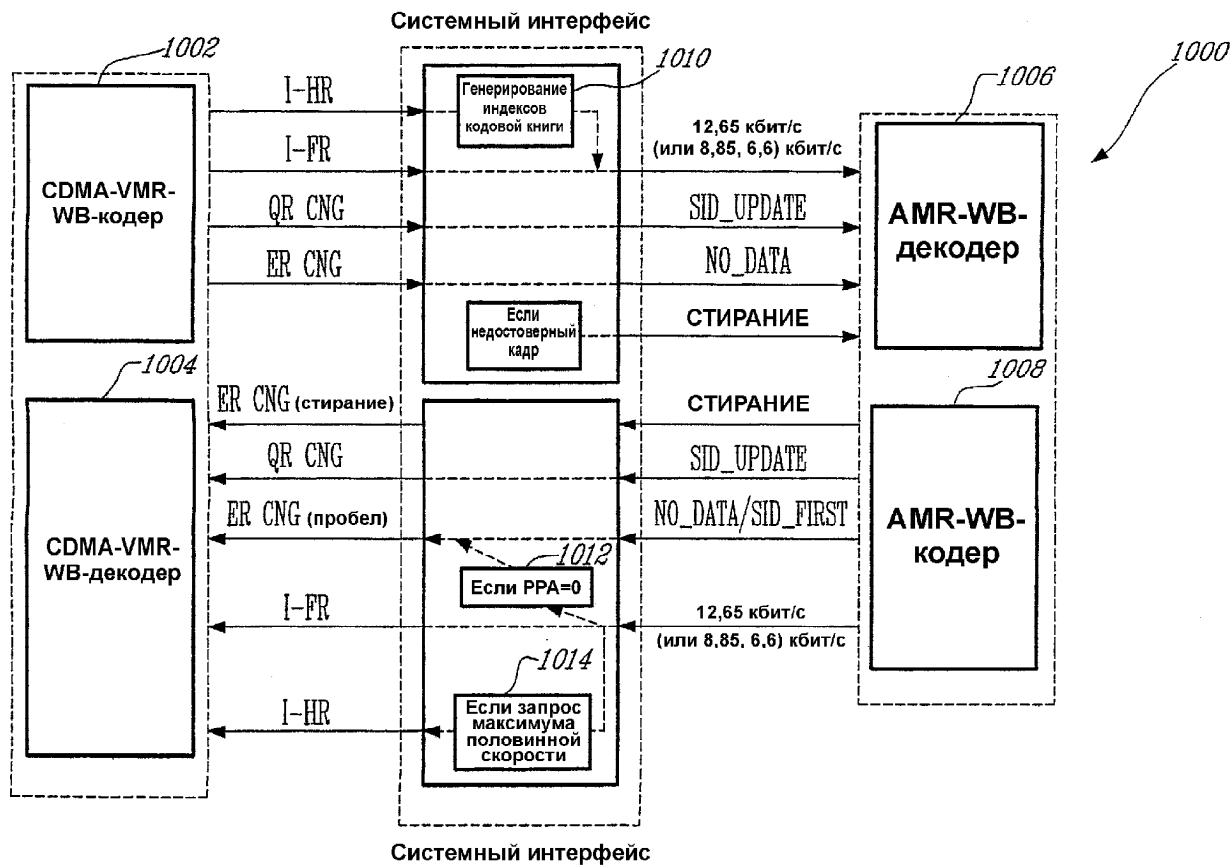
Фиг. 9



Фиг. 10



Фиг. 11



Фиг. 12