



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2015142108, 09.01.2014

(24) Дата начала отсчета срока действия патента:
09.01.2014Дата регистрации:
15.12.2017

Приоритет(ы):

(30) Конвенционный приоритет:
04.03.2013 US 61/772,037

(43) Дата публикации заявки: 11.04.2017 Бюл. № 11

(45) Опубликовано: 15.12.2017 Бюл. № 35

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 05.10.2015(86) Заявка РСТ:
СА 2014/000014 (09.01.2014)(87) Публикация заявки РСТ:
WO 2014/134702 (12.09.2014)Адрес для переписки:
129090, Москва, ул. Б. Спасская, 25, стр. 3, ООО
"Юридическая фирма Городисский и Партнеры

(72) Автор(ы):

**ВАЙАНКУР Томми (СА),
ЕЛИНЕК Милан (СА)**

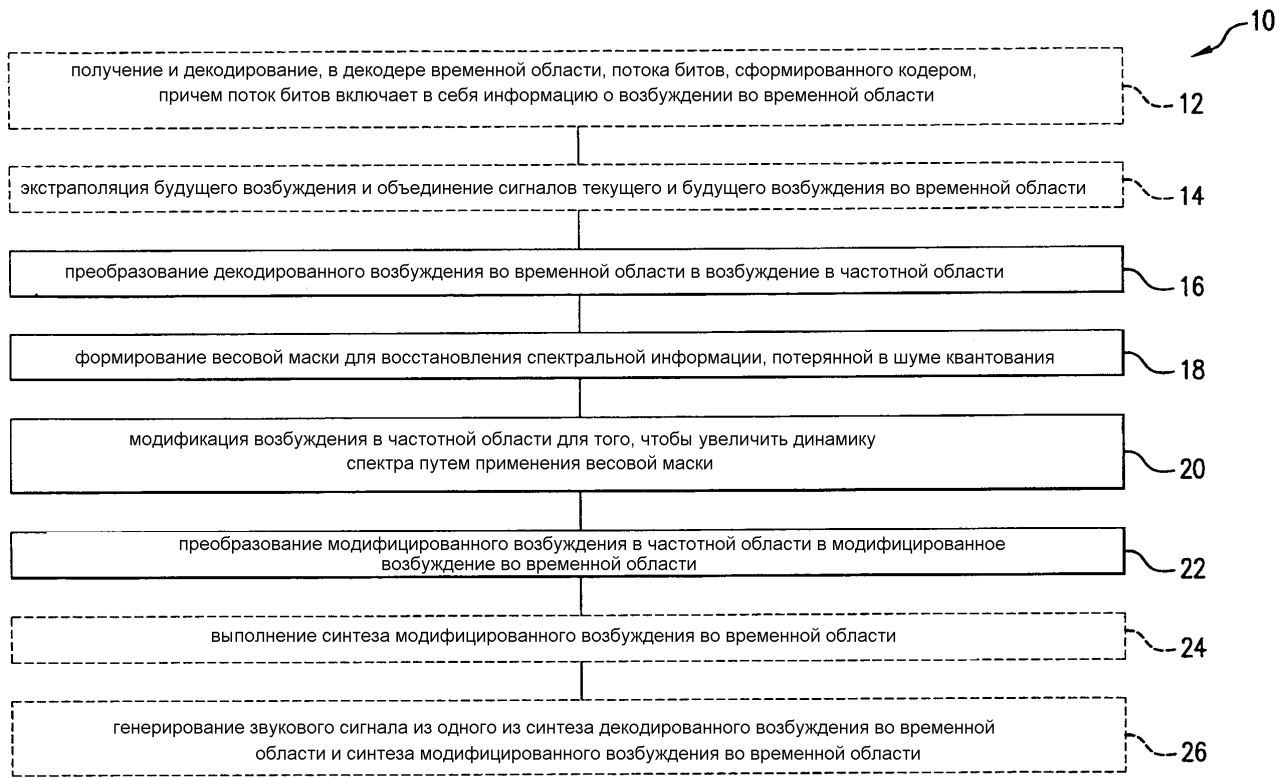
(73) Патентообладатель(и):

ВОЙСЭЙДЖ КОРПОРЕЙШН (СА)(56) Список документов, цитированных в отчете
о поиске: US 2011/0046947 A1, 24.02.2011. WO
2007/073604 A1, 05.07.2007. US 2011/0002266
A1, 06.01.2011. US 2006/0271354 A1, 30.11.2006.
US 5659661 A, 19.08.1997. WO 2009/109050 A1,
11.09.2009. WO 03/102921 A1, 11.12.2003. US
2007/0094016 A1, 26.04.2007. RU 2224302 C2,
20.02.2004.**(54) УСТРОЙСТВО И СПОСОБ ДЛЯ УМЕНЬШЕНИЯ ШУМА КВАНТОВАНИЯ В ДЕКОДЕРЕ
ВРЕМЕННОЙ ОБЛАСТИ**

(57) Реферат:

Изобретение относится к средствам для уменьшения шума квантования в сигнале, содержащемся в возбуждении во временной области, декодируемом декодером временной области. Технический результат заключается в повышении качества кодируемого речевого сигнала. Декодированное возбуждение во временной области преобразовывается в возбуждение в частотной области. Весовая маска формируется для восстановления спектральной информации, потерянной в шуме квантования. Возбуждение в частотной области модифицируется для того, чтобы увеличить динамику спектра путем применения весовой

маски. Модифицированное возбуждение в частотной области преобразовывается в модифицированное возбуждение во временной области. Способ и устройство могут использоваться для того, чтобы улучшить воспроизведение музыкального контента кодеками на основе линейного предсказания (LP). Опционально синтез декодированного возбуждения во временной области может быть классифицирован на одно из первого набора категорий возбуждения и второго набора категорий возбуждения. 2 н. и 29 з.п. ф-лы, 4 табл., 4 ил.



ФИГ. 1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
G10L 19/03 (2013.01)
G10L 21/0232 (2013.01)

(12) **ABSTRACT OF INVENTION**

(21)(22) Application: **2015142108, 09.01.2014**

(24) Effective date for property rights:
09.01.2014

Registration date:
15.12.2017

Priority:

(30) Convention priority:
04.03.2013 US 61/772,037

(43) Application published: **11.04.2017** Bull. № 11

(45) Date of publication: **15.12.2017** Bull. № 35

(85) Commencement of national phase: **05.10.2015**

(86) PCT application:
CA 2014/000014 (09.01.2014)

(87) PCT publication:
WO 2014/134702 (12.09.2014)

Mail address:

**129090, Moskva, ul. B. Spasskaya, 25, str. 3, OOO
"Yuridicheskaya firma Gorodisskij i Partnery"**

(72) Inventor(s):

**VAJANKUR Tommi (CA),
ELINEK Milan (CA)**

(73) Proprietor(s):

VOJSEJDZH KORPOREJSHN (CA)

(54) **DEVICE AND METHOD FOR REDUCING QUANTIZATION NOISE IN DECODER OF TEMPORAL AREA**

(57) Abstract:

FIELD: physics.

SUBSTANCE: decoded excitation in the time domain is converted into an excitation in the frequency domain. A weight mask is formed to reconstruct the spectral information lost in the quantization noise. The excitation in the frequency domain is modified in order to increase the dynamics of the spectrum by applying a weight mask. The modified excitation in the frequency domain is converted into a modified excitation in the

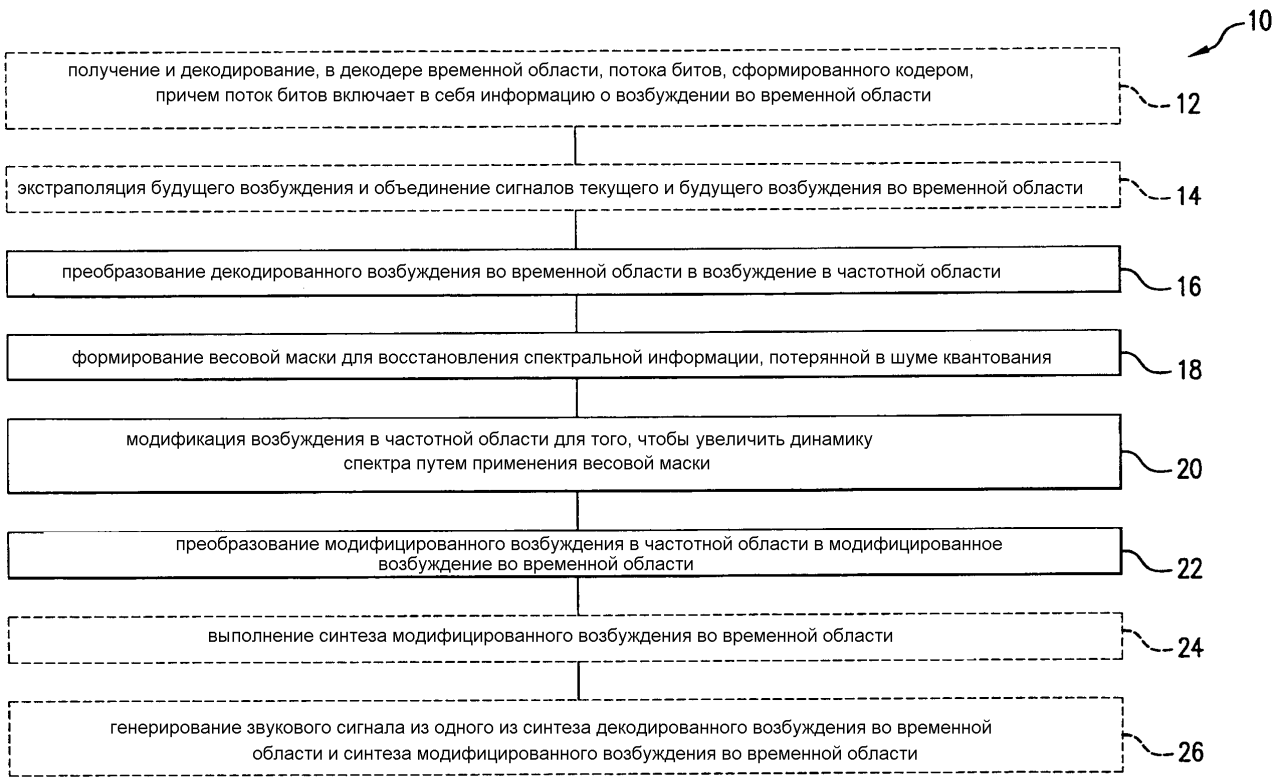
time domain. The method and device can be used to improve reproduction of music content by codecs based on linear prediction (LP). Optionally, the synthesis of the decoded excitation in the time domain can be classified into one of the first set of excitation categories and the second set of excitation categories.

EFFECT: improving the quality of the encoded speech signal.

31 cl, 4 tbl, 4 dwg

C 2
4
7
4
4
2
6
3
8
7
4
4
R U

R U
2
6
3
8
7
4
4
C 2



ФИГ. 1

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

[0001] Настоящее изобретение относится к области обработки звука. Более конкретно, настоящее изобретение относится к уменьшению шума квантования в звуковом сигнале.

УРОВЕНЬ ТЕХНИКИ

5 [0002] Современные разговорные кодеки представляют с очень хорошим качеством чисто речевые сигналы при скоростях передачи приблизительно 8 Кбит/с и приближаются к незаметности для пользователя при скорости передачи 16 Кбит/с. Для того чтобы поддерживать это высокое качество речи при низкой скорости передачи, обычно используется мультимодальная схема кодирования. Обычно входной сигнал
10 расщепляется на различные категории, отражающие его характеристику. Эти различные категории включают в себя, например, вокализованную речь, невокализованную речь, вокализованные вступления и т.д. Кодек затем использует различные режимы кодирования, оптимизированные для этих категорий.

[0003] Основанные на модели речи кодеки обычно не очень хорошо воспроизводят
15 общие сигналы звуковой частоты, такие как музыку. Следовательно, некоторые развернутые кодеки для разговорных сигналов не представляют музыку с хорошим качеством, особенно при низких скоростях передачи. Когда кодек развернут, трудно модифицировать кодер из-за того, что поток битов стандартизован, и любые изменения в потоке битов нарушили бы функциональную совместимость кодека.

20 [0004] Следовательно, имеется потребность в улучшении воспроизведения музыкального контента основанными на модели речи кодеками, например кодеками на основе линейного предсказания (LP).

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

[0005] В соответствии с настоящим изобретением предлагается устройство для
25 уменьшения шума квантования в сигнале, содержащем во временной области возбуждение, декодируемое декодером временной области. Это устройство включает в себя преобразователь декодированного возбуждения во временной области в возбуждение в частотной области. Устройство также включает в себя блок формирования маски для формирования весовой маски для восстановления
30 спектральной информации, потерянной в шуме квантования. Устройство также включает в себя модификатор возбуждения в частотной области для того, чтобы увеличить динамику спектра путем применения весовой маски. Устройство дополнительно включает в себя преобразователь модифицированного возбуждения в частотной области в модифицированное возбуждение во временной области.

35 [0006] Настоящее изобретение также относится к способу для уменьшения шума квантования в сигнале, содержащем во временной области возбуждение, декодируемое декодером временной области. Декодированное возбуждение во временной области преобразовывается в возбуждение в частотной области декодером временной области. Весовая маска формируется для восстановления спектральной информации, потерянной
40 в шуме квантования. Возбуждение в частотной области модифицируется для того, чтобы увеличить динамику спектра путем применения весовой маски. Модифицированное возбуждение в частотной области преобразовывается в модифицированное возбуждение во временной области.

[0007] Вышеперечисленные и другие признаки станут более ясными после прочтения
45 последующего не ограничивающего описания иллюстративных вариантов их осуществления, представленных только в качестве примеров со ссылками на сопроводительные чертежи.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0008] Далее варианты осуществления настоящего изобретения будут описаны только в качестве примеров со ссылками на сопроводительные чертежи, на которых:

[0009] Фиг. 1 представляет собой блок-схему, показывающую операции способа для уменьшения шума квантования в сигнале, содержащемся в возбуждении во временной области, декодированном декодером временной области, в соответствии с одним вариантом осуществления;

[0010] Фиг. 2a и 2b, совместно упоминаемые как Фиг. 2, представляют собой упрощенную принципиальную схему декодера, имеющего возможности постобработки в частотной области для уменьшения шума квантования в музыкальных сигналах и других звуковых сигналах; и

[0011] Фиг. 3 представляет собой упрощенную блок-схему примерной конфигурации аппаратных компонентов, формирующих декодер, изображенный на Фиг. 2.

ПОДРОБНОЕ ОПИСАНИЕ

[0012] Различные аспекты настоящего изобретения в целом решают одну или более проблем улучшения воспроизведения музыкального контента кодеками на основе модели речи, например кодеками на основе линейного предсказания (LP), путем уменьшения шума квантования в музыкальном сигнале. Следует учесть, что настоящее изобретение может также применяться к другим звуковым сигналам, например к общим сигналам звуковой частоты, отличающимся от музыки.

[0013] Модификации декодера могут улучшить воспринимаемое качество на стороне приемника. Настоящее изобретение раскрывает подход к реализации на стороне декодера постобработки для музыкальных сигналов и других звуковых сигналов в частотной области, который уменьшает шум квантования в спектре синтезируемого декодированного сигнала. Постобработка может быть осуществлена без какой-либо дополнительной задержки кодирования.

[0014] Принцип удаления в частотной области шума квантования между гармониками спектра и частотной постобработки, используемый в настоящем документе, основан на патентной публикации PCT WO 2009/109050 A1 автора Vaillancourt и др., датированной 11 сентября 2009 г. (в дальнейшем упоминаемой как «Vaillancourt '050»), раскрытие которой включено в настоящий документ посредством ссылки. В большинстве случаев такая частотная постобработка применяется к синтезируемому декодированному сигналу и требует увеличения задержки обработки для того, чтобы включить перекрытие и добавить процесс для получения значительного выигрыша в качестве. Более того, при традиционной постобработке в частотной области чем короче добавляемая задержка (то есть чем короче окно преобразования), тем менее эффективной является постобработка благодаря ограниченному частотному разрешению. В соответствии с настоящим изобретением частотная постобработка достигает более высокого частотного разрешения (используется более длинное частотное преобразование) без добавления задержки к синтезу. Кроме того, информация, присутствующая в энергии спектра прошлых кадров, используется для создания весовой маски, которая применяется к спектру текущего кадра для того, чтобы восстановить, то есть улучшить, спектральную информацию, потерянную в шуме кодирования. Для того, чтобы достичь этой постобработки без добавления задержки к синтезу, в этом примере используется симметричное трапецидальное окно. Это окно центрируется на текущем кадре, причем окно является плоским (оно имеет постоянное значение, равное 1), и экстраполяция используется для того, чтобы создать будущий сигнал. В то время как постобработка обычно может быть применена непосредственно к сигналу синтеза любого кодека, настоящее изобретение представляет иллюстративный вариант осуществления, в

котором постобработка применяется к сигналу возбуждения в рамках кодека линейного предсказания с кодовым возбуждением (CELP), описанного в технической спецификации (TS) 26.190 Программы Партнерства 3-го поколения (3GPP), озаглавленной как «Адаптивный многоскоростной широкополосный (AMR-WB) речевой кодек; Функции транскодирования», доступной на веб-сайте 3GPP, полное содержание которой включено в настоящий документ посредством ссылки. Преимущество работы над сигналом возбуждения, а не над сигналом синтеза, состоит в том, что любые потенциальные разрывы, вводимые постобработкой, сглаживаются последующим применением фильтра синтеза CELP.

[0015] В настоящем изобретении для целей иллюстрации используется AMR-WB с внутренней частотой оцифровки 12,8 кГц. Однако настоящее изобретение может быть применено к другим речевым декодерам с низкой скоростью передачи, где синтез получается с помощью сигнала возбуждения, отфильтрованного через фильтр синтеза, например фильтр синтеза LP. Это может быть также применено на мультимодальных кодеках, где музыка кодируется с помощью комбинации возбуждения во временной области и в частотной области. Следующие строки суммируют работу постфильтра. Затем следует подробное описание иллюстративного варианта осуществления, использующего AMR-WB.

[0016] Сначала полный битовый поток декодируется, и текущий синтезированный кадр обрабатывается классификатором первого этапа, подобным тому, который раскрывается в патентной публикации PCT WO 2003/102921 A1 автора Jelinek и др., датированной 11 декабря 2003 г., в патентной публикации PCT WO 2007/073604 A1 автора Vaillancourt и др., датированной 5 июля 2007 г., и в международной заявке PCT/CA2012/001011, зарегистрированной 1 ноября 2012 автора Vaillancourt и др. (в дальнейшем упоминаемой как «Vaillancourt '011»), раскрытия которых включены в настоящий документ посредством ссылки. Для целей данного раскрытия этот классификатор первого этапа анализирует кадр и обособленно устанавливает НЕАКТИВНЫЕ кадры и НЕВОКАЛИЗИРОВАННЫЕ кадры, например кадры, соответствующие активной НЕВОКАЛИЗИРОВАННОЙ речи. Все кадры, которые не категоризируются как НЕАКТИВНЫЕ кадры или как НЕВОКАЛИЗИРОВАННЫЕ кадры на первом этапе, анализируются с помощью классификатора второго этапа. Классификатор второго этапа решает, применять ли постобработку, и в какой степени. Когда постобработка не применяется, обновляется только память, относящаяся к постобработке.

[0017] Для всех кадров, которые не категоризированы классификатором первого этапа как НЕАКТИВНЫЕ кадры или как кадры с активной НЕВОКАЛИЗИРОВАННОЙ речью, формируется вектор с использованием прошлого декодированного возбуждения, декодированного возбуждения текущего кадра и экстраполяции будущего возбуждения. Длина прошлого декодированного возбуждения и экстраполируемого возбуждения является одинаковой и зависит от желаемого разрешения частотного преобразования. В этом примере длина используемого частотного преобразования составляет 640 отсчетов. Создание вектора с использованием прошлого и экстраполируемого возбуждения позволяет увеличить частотное разрешение. В представленном примере длина прошлого и экстраполируемого возбуждения является одинаковой, но для эффективной работы постфильтра не обязательно требуется симметрия окна.

[0018] Энергетическая устойчивость частотного представления объединенного возбуждения (включающего прошлое декодированное возбуждение, декодированное

возбуждение текущего кадра и экстраполяцию будущего возбуждения) затем анализируется с помощью классификатора второго этапа для того, чтобы определить вероятность присутствия музыки. В этом примере определение присутствия музыки выполняется в ходе двухэтапного процесса. Однако обнаружение музыки может быть выполнено различными путями, например, оно может быть выполнено в единственной операции, предшествующей частотному преобразованию, или даже определено в кодере и передано в потоке битов.

[0019] Межгармонический шум квантования уменьшается так же, как и в публикации Vaillancourt'050, путем оценки соотношения сигнал/шум (SNR) для каждого элемента разрешения по частоте и применения усиления к каждому элементу разрешения по частоте в зависимости от значения его SNR. В настоящем изобретении, однако, оценка энергии шумов выполняется не так, как описано в публикации Vaillancourt'050.

[0020] Затем используется дополнительная обработка, которая восстанавливает информацию, потерянную в шуме кодирования, и дополнительно увеличивает динамику спектра. Этот процесс начинается с нормализации энергетического спектра диапазоном от 0 до 1. Затем постоянное смещение прибавляется к нормализованному энергетическому спектру. Наконец, степень 8 применяется к каждому элементу разрешения по частоте модифицированного энергетического спектра. Получаемый масштабированный энергетический спектр обрабатывается усредняющей функцией вдоль частотной оси, от низких частот до высоких частот. Наконец, долговременное сглаживание спектра во времени выполняется элемент за элементом разрешения.

[0021] Эта вторая часть обработки приводит к маске, в которой пики соответствуют важной информации о спектре, а впадины соответствуют кодирующему шуму. Эта маска затем используется для того, чтобы отфильтровать шум и увеличить динамику спектра путем небольшого увеличения амплитуды элементов разрешения спектра в пиковых областях, ослабляя амплитуду элементов разрешения во впадинах, и, следовательно, увеличивая отношение пиков ко впадинам. Эти две операции выполняются с использованием высокого частотного разрешения, но без добавления задержки к синтезу выхода.

[0022] После того как частотное представление объединенного вектора возбуждения улучшено (его шум уменьшен, а его динамика спектра увеличена), выполняется обратное частотное преобразование, для того, чтобы создать улучшенную версию объединенного возбуждения. В настоящем изобретении часть окна преобразования, соответствующая текущему кадру, является по существу плоской, и только те части окна, которые применяются к прошлому и экстраполируемому сигналу возбуждения, нуждаются в сужении. Это делает возможным уничтожение повышенного возбуждения в текущем кадре после обратного преобразования. Эта последняя манипуляция аналогична умножению повышенного возбуждения во временной области на прямоугольное окно в положении текущего кадра. В то время как эта операция не может быть выполнена в области синтеза без добавления важных блочных артефактов, это может быть альтернативно сделано в области возбуждения, потому что фильтр синтеза LP помогает сглаживать переходы от одного блока к другому, как показано в публикации Vaillancourt'011.

Описание иллюстративного варианта осуществления AMR-WB

[0023] Описанная здесь постобработка применяется к декодированному возбуждению фильтра синтеза LP для таких сигналов, как музыка или реверберирующая речь. Решение о природе сигнала (речь, музыка, реверберирующая речь и т.п.) и решение о применении постобработки могут быть сообщены кодером, который посылает декодеру информацию

о классификации как часть потока битов AMR-WB. Если это не так, то классификация сигнала альтернативно может быть сделана на стороне декодера. В зависимости от компромисса между сложностью и надежностью классификации фильтр синтеза может опционально быть применен к текущему возбуждению для того, чтобы получить временный синтез и более хороший анализ классификации. В этой конфигурации синтез перезаписывается, если классификация приводит к категории, в которой применяется постфильтрация. Для того чтобы минимизировать добавленную сложность, классификация может также быть выполнена на синтезе прошлого кадра, и фильтр синтеза тогда применяется однократно после постобработки.

10 [0024] Обращаясь теперь к чертежам, Фиг. 1 представляет собой блок-схему, показывающую операции способа для уменьшения шума квантования в сигнале, содержащемся в возбуждении во временной области, декодированном декодером временной области, в соответствии с одним вариантом осуществления. На Фиг. 1 последовательность 10 включает в себя множество операций, которые могут выполняться в переменном порядке, некоторые из этих операций могут выполняться параллельно, и некоторые из этих операций могут быть опциональными. В операции 12 декодер временной области, получает и декодирует поток битов, сформированный кодером, включающий в себя информацию о возбуждении во временной области в форме параметров, которые можно использовать для того, чтобы реконструировать возбуждение во временной области. Для этого декодер временной области, может получать поток битов через интерфейс входа или считывать поток битов из памяти. Декодер временной области, преобразовывает декодированное возбуждение во временной области в возбуждение в частотной области в операции 16. Прежде, чем преобразовать сигнал возбуждения из временной области в частотную область в операции 16, будущее возбуждение во временной области может быть экстраполировано в операции 14 так, чтобы преобразование возбуждения во временной области в возбуждение в частотной области можно было сделать без задержки. Таким образом, выполняется лучший частотный анализ без потребности в дополнительной задержке. С этой целью прошлый, текущий и предсказанный будущий сигнал возбуждения во временной области могут быть объединены перед преобразованием в частотную область. Декодер временной области формирует затем весовую маску для того, чтобы восстановить спектральную информацию, потерянную в шуме квантования, в операции 18. В операции 20 декодер временной области, модифицирует возбуждение в частотной области для того, чтобы увеличить динамику спектра путем применения весовой маски. В операции 22 декодер временной области, преобразовывает модифицированное возбуждение в частотной области в модифицированное возбуждение во временной области. Декодер временной области, может затем выполнить синтез модифицированного возбуждения во временной области в операции 24 и сгенерировать звуковой сигнал из одного из синтеза декодированного возбуждения во временной области и синтеза модифицированного возбуждения во временной области в операции 26.

45 [0025] Способ, проиллюстрированный на Фиг. 1, может быть адаптирован с использованием нескольких дополнительных особенностей. Например, синтез декодированного возбуждения во временной области может быть классифицирован на одно из первого набора категорий возбуждения и второго набора категорий возбуждения, в которых второй набор категорий возбуждения включает в себя НЕАКТИВНУЮ или НЕВОКАЛИЗИРОВАННУЮ категории, в то время как первый набор категорий возбуждения включает в себя ДРУГУЮ категорию. Преобразование

декодированного возбуждения во временной области в возбуждение в частотной области может быть применено к декодированному возбуждению во временной области, классифицированному как первый набор категорий возбуждения. Восстановленный поток битов может включать в себя информацию о классификации, которая может использоваться для того, чтобы классифицировать синтез декодированного возбуждения во временной области как первый набор или как второй набор категорий возбуждения. Для генерирования звукового сигнала выходной синтез может быть выбран как синтез декодированного возбуждения во временной области, когда возбуждение во временной области классифицируется как второй набор категорий возбуждения, или как синтез модифицированного возбуждения во временной области, когда возбуждение во временной области классифицируется как первый набор категорий возбуждения. Возбуждение в частотной области может быть проанализировано для того, чтобы определить, содержит ли возбуждение в частотной области музыку. В частности, определение того, что возбуждение в частотной области содержит музыку, может основываться на сравнении с некоторым порогом статистической девиации разностей спектральных энергий возбуждения в частотной области. Весовая маска может быть сформирована с использованием усреднения во времени, или частотного усреднения, или их комбинации. Величина отношения сигнал/шум может быть оценена для выбранного диапазона декодированного возбуждения во временной области, и шумоподавление в частотной области может быть выполнено на основе оценки отношения сигнал/шум.

[0026] Фиг. 2а и 2б, совместно упоминаемые как Фиг. 2, представляют собой упрощенную принципиальную схему декодера, имеющего возможности постобработки в частотной области для уменьшения шума квантования в музыкальных сигналах и других звуковых сигналах. Декодер 100 включает в себя несколько элементов, проиллюстрированных на Фиг. 2а и 2б, эти элементы соединены, как показано стрелками, некоторые из взаимосвязей проиллюстрированы с использованием соединителей А, В, С, D и E, которые показывают, как некоторые элементы, изображенные на Фиг. 2а, соединяются с другими элементами, изображенными на Фиг. 2б. Декодер 100 включает в себя приемник 102, который получает поток битов AMR-WB от кодера, например через интерфейс радиосвязи. Альтернативно декодер 100 может быть оперативно соединен с памятью (не показана), хранящей поток битов. Демультимплексор 103 извлекает из потока битов параметры возбуждения во временной области для того, чтобы реконструировать возбуждение во временной области, информацию о задержке высоты тона и информацию об определении присутствия голосового сигнала (VAD). Декодер 100 включает в себя декодер 104 возбуждения во временной области, получающий параметры возбуждения во временной области для того, чтобы декодировать возбуждение во временной области существующего кадра, буферную память 106 прошлого возбуждения, два (2) фильтра 108 и 110 синтеза LP, классификатор 112 сигнала первого этапа, включающий в себя блок 114 оценки классификации сигнала, который получает сигнал VAD и контрольную точку 116 выбора класса, блок 118 экстраполяции возбуждения, который получает информацию о задержке высоты тона, блок 120 объединения возбуждения, модуль 122 кадрирования и частотного преобразования, анализатор энергетической устойчивости как классификатор 124 сигнала второго этапа, блок 126 оценки уровня шума в диапазоне, блок 128 уменьшения шума, блок 130 формирования маски, включающий в себя блок 131 нормализации спектральной энергии, блок 132 усреднения энергии и блок 134 сглаживания энергии, блок 136 модификации динамики спектра, блок 138

преобразования из частотной области во временную область, блок 140 извлечения возбуждения кадра, блок 142 перезаписи, включающий в себя контрольную точку 144 принятия решения, управляющую переключателем 146, и фильтр устранения предсказаний и передискретизатор 148. Решение о перезаписи, принимаемое контрольной точкой 144 принятия решения, основывается на НЕАКТИВНОЙ или НЕВОКАЛИЗИРОВАННОЙ классификации, получаемой из классификатора 112 сигнала первого этапа, и на категории звукового сигнала e_{SAT} , получаемой из классификатора 124 сигнала второго этапа, независимо от того, подается ли к фильтру устранения предсказаний и передискретизатору 148 сигнал 150 основного синтеза от фильтра 108 синтеза LP, или модифицированный, то есть улучшенный сигнал 152 синтеза от фильтра 110 синтеза LP. Выход фильтра устранения предсказаний и передискретизатора 148 подается к цифро-аналоговому (D/A) преобразователю 154, который обеспечивает аналоговый сигнал, усиленный усилителем 156 и подаваемый далее к громкоговорителю 158, который генерирует слышимый звуковой сигнал.

Альтернативно выход фильтра устранения предсказаний и передискретизатора 148 может быть передан в цифровом формате по коммуникационному интерфейсу (не показан) или сохранен в цифровом формате в памяти (не показана), на компакт-диске или на любом другом носителе цифрового накопителя. В качестве другой альтернативы, выход цифроаналогового преобразователя 154 может быть подан в наушники (не показаны), непосредственно или через усилитель. В качестве еще одной альтернативы, выход цифроаналогового преобразователя 154 может быть записан на аналоговом носителе (не показан) или передан через коммуникационный интерфейс (не показан) как аналоговый сигнал.

[0027] Следующие параграфы описывают подробности операций, выполняемых различными компонентами декодера 100, изображенного на Фиг. 2.

1) Классификация первого этапа

[0028] В иллюстративном варианте осуществления классификация первого этапа выполняется в декодере в классификаторе 112 первого этапа в ответ на параметры определения присутствия голосового сигнала VAD от демультимплексора 103.

Классификация первого этапа декодера аналогична тому, что описано в публикации Vaillancourt'011. Следующие параметры используются для классификации в блоке 114 оценки классификации сигнала декодера: нормализованная корреляция r_x , мера спектрального наклона e_t счетчика устойчивости высоты тона r_s , относительная энергия кадра сигнала в конце текущего кадра E_s , а также счетчик нулевых пересечений z_s .

Вычисление этих параметров, которые используются для классификации сигнала, объясняется ниже.

[0029] Нормализованная корреляция r_x вычисляется в конце кадра на основе сигнала синтеза. Используется задержка высоты тона последнего подкадра.

[0030] Нормализованная корреляция r_x вычисляется одновременно с высотой тона как

$$r_x = \frac{\sum_{i=0}^{T-1} x(t+i)x(t+i-T)}{\sqrt{\sum_{i=0}^{T-1} x^2(t+i) \sum_{i=0}^{T-1} x^2(t+i-T)}} \quad (1)$$

[0031] где T является задержкой высоты тона последнего подкадра, $t=L-T$, и L является размером кадра. Если задержка высоты тона последнего подкадра больше, чем $3N/2$ (где N - размер подкадра), T устанавливается равным средней задержке высоты тона последних двух подкадров.

5 [0032] Корреляция r_x вычисляется с использованием сигнала синтеза $x(i)$. Для задержки высоты тона ниже, чем размер подкадра (64 отсчета) нормализованная корреляция вычисляется дважды в моменты времени $t=L-T$ и $t=L-2T$, а r_x задается как среднее значение этих двух вычислений.

10 [0033] Параметр спектрального наклона e_t содержит информацию о частотном распределении энергии. В существующем иллюстративном варианте осуществления спектральный наклон в декодере оценивается как первый нормализованный коэффициент автокорреляции сигнала синтеза. Он вычисляется на основе последних 3 подкадров как

$$15 \quad e_t = \frac{\sum_{i=N}^{L-1} x(i)x(i-1)}{\sum_{i=N}^{L-1} x^2(i)} \quad (2)$$

20 [0034] где $x(i)$ является сигналом синтеза, N является размером подкадра, а L является размером кадра ($N=64$ и $L=256$ в этом иллюстративном варианте осуществления).

[0035] Счетчик устойчивости высоты тона оценивает вариацию периода высоты тона. Она вычисляется в декодере следующим образом:

$$pc = |p_3 + p_2 - p_1 - p_0| \quad (3)$$

25 [0036] Значения p_0 , p_1 , p_2 и p_3 соответствуют задержке высоты тона в замкнутом цикле от этих 4 подкадров.

[0037] Относительная энергия E_s кадра вычисляется как разность между текущей энергией кадра в дБ и его долгосрочным средним значением

$$30 \quad E_s = E_f - E_{lt} \quad (4)$$

[0038] где энергия E_f кадра является энергией сигнала синтеза s_{out} в дБ, вычисляемой синхронно с высотой тона в конце кадра как

$$35 \quad E_f = 10 \log_{10} \left(\frac{1}{T} \sum_{i=0}^{T-1} s_{out}^2(i+L-T) \right) \quad (5)$$

40 [0039] где $L=256$ является длиной кадра, а T является средней задержкой высоты тона последних двух подкадров. Если значение T меньше, чем размер подкадра, тогда значение T устанавливается равным $2T$ (энергии, вычисленной с использованием двух периодов высоты тона для коротких задержек высоты тона).

[0040] Долгосрочная усредненная энергия обновляется на активных кадрах с использованием следующего соотношения:

$$45 \quad E_{lt} = 0.99E_{lt} + 0.01E_f \quad (6)$$

[0041] Последний параметр является параметром z_c нулевых пересечений, вычисленным на одном кадре сигнала синтеза. В этом иллюстративном варианте осуществления счетчик нулевых пересечений z_c подсчитывает количество раз, которое знак сигнала меняется с положительного на отрицательный во время этого интервала.

[0042] Для того, чтобы сделать классификацию первого этапа более надежной,

параметры классификации рассматриваются вместе, формируя функцию выгоды f_m . С этой целью параметры классификации сначала масштабируются с использованием линейной функции. Рассмотрим параметр p_x , масштабированная версия которого

$$p^s = k_p \cdot p_x + c_p \quad (7)$$

[0043] Масштабированный параметр устойчивости высоты тона обрезается между 0 и 1. Коэффициенты функции k_p и c_p были найдены экспериментально для каждого из параметров. Значения, используемые в этом иллюстративном варианте осуществления, приведены в Таблице 1.

Параметр	Значение	k_p	c_p
r_x	Нормализованная корреляция	0,8547	0,2479
e_t	Спектральный наклон	0,8333	0,2917
p_c	Счетчик устойчивости высот тона	-0,0357	1,6074
E_s	Относительная энергия кадра	0,04	0,56
z_c	Счетчик нулевых пересечений	-0,04	2,52

[0044] Функция выгоды была определена как

$$f_m = \frac{1}{6} (2 \cdot r_x^s + e_t^s + p_c^s + E_s^s + z_c^s) \quad (8)$$

[0045] где верхний индекс s указывает масштабированную версию параметров.

[0046] Классификация затем выполняется (контрольная точка 116 выбора класса) с использованием функции f_m выгоды, следуя правилам, приведенным в Таблице 2.

Класс предыдущего кадра	Правило	Класс текущего кадра
ДРУГОЙ	$f_m \geq 0,39$	ДРУГОЙ
	$f_m < 0,39$	НЕВОКАЛИЗИРОВАННЫЙ
НЕВОКАЛИЗИРОВАННЫЙ	$f_m > 0,45$	ДРУГОЙ
	$f_m \leq 0,45$	НЕВОКАЛИЗИРОВАННЫЙ
	VAD = 0	НЕАКТИВНЫЙ

[0047] В дополнение к этой классификации первого этапа, информация об определении присутствия голосового сигнала (VAD) кодером может быть передана в потоке битов, как это имеет место в случае иллюстративного примера на основе AMR-WB. Таким образом, один бит посылается в потоке битов для того, чтобы определить, рассматривает ли кодер текущий кадр как активный контент (VAD = 1) или НЕАКТИВНЫЙ контент (фоновый шум, VAD = 0). Когда контент рассматривается как НЕАКТИВНЫЙ, тогда классификация перезаписывается как НЕВОКАЛИЗИРОВАННЫЙ. Схема классификации первого этапа также включает в себя обнаружение ОБЩЕГО ЗВУКА. Категория ОБЩИЙ ЗВУК включает в себя музыку, ревербирующую речь и может также включать фоновую музыку. Для того, чтобы идентифицировать эту категорию, используются два параметра. Одним из этих параметров является общая энергия E_f кадра, выражаемая уравнением (5).

[0048] Сначала модуль определяет разность энергий ΔE^t двух смежных кадров, в

частности разность между энергией текущего кадра E_f^t и энергией предыдущего кадра. Затем вычисляется средняя разность энергий E_{df} по прошлым 40 кадрам, используя следующее соотношение:

$$\overline{E_{df}} = \frac{\sum_{t=40}^{t-1} \Delta_E^t}{40}; \quad \text{где: } \Delta_E^t = E_f^t - E_f^{(t-1)} \quad (9)$$

[0049] Затем модуль определяет статистическую девиацию вариации энергии a_E для последних пятнадцати (15) кадров, используя следующее соотношение:

$$\sigma_E = p \sqrt{\sum_{t=-15}^{t-1} \frac{(\Delta_E^t - \overline{E_{df}})^2}{15}} \quad (10)$$

[0050] При практической реализации иллюстративного варианта осуществления масштабный коэффициент p был найден экспериментально и установлен равным приблизительно 0,77. Получаемая девиация a_E указывает на энергетическую устойчивость декодированного синтеза. Как правило, музыка имеет более высокую энергетическую устойчивость, чем речь.

[0051] Результат классификации первого этапа далее используется для того, чтобы подсчитать количество кадров N_{uv} между двумя кадрами, классифицированными как НЕВОКАЛИЗИРОВАННЫЕ. При практической реализации подсчитываются только кадры с энергией E_f выше чем -12 дБ. Обычно счетчик N_{uv} инициализируется нулем, когда кадр классифицируется как НЕВОКАЛИЗИРОВАННЫЙ. Однако, когда кадр классифицируется как НЕВОКАЛИЗИРОВАННЫЙ, и его энергия E_f больше, чем -9 дБ, и долгосрочная средняя энергия E_{lt} ниже 40 дБ, тогда счетчик инициализируется значением 16 для того, чтобы придать небольшое смещение в сторону музыкального решения. В противном случае, если кадр классифицируется как НЕВОКАЛИЗИРОВАННЫЙ, но долгосрочная средняя энергия E_{lt} выше 40 дБ, счетчик уменьшается на 8 для того, чтобы обеспечить схождение к речевому решению. При практической реализации счетчик ограничивается диапазоном от 0 до 300 для активного сигнала; счетчик также ограничивается диапазоном от 0 до 125 для НЕАКТИВНОГО сигнала для того, чтобы получить быструю сходимость к речевому решению, когда следующий активный сигнал является речевым. Эти диапазоны не являются ограничивающими, и другие амплитуды также могут быть рассмотрены в конкретной реализации. Для этого иллюстративного примера решение между активным и НЕАКТИВНЫМ сигналом выводится из решения о речевой активности (VAD), включенного в поток битов.

[0052] Долгосрочное среднее число N_{uv} выводится из этого счетчика НЕВОКАЛИЗИРОВАННЫХ кадров для активного сигнала следующим образом: $N_{uvlt} = 0,9 \cdot N_{uvlt} + 0,1 \cdot N_{uv}$

$$\overline{N}_{uv}^t = 0,9 \cdot \overline{N}_{uv}^{(t-1)} + 0,1 \cdot N_{uv}, \quad (11)$$

[0053] и для НЕАКТИВНОГО сигнала следующим образом:

$$\overline{N}_{uv}^t = 0,95 \cdot \overline{N}_{uv}^{(t-1)}. \quad (12)$$

[0054] где t является индексом кадра. Следующий псевдокод иллюстрирует функциональность счетчика НЕВОКАЛИЗИРОВАННЫХ кадров и его долгосрочное среднее значение:

```

5      if (UNVOICED &  $E_f > 9\text{ dB}$ )
          if ( $E_{it} \leq 40$ )
               $N_{uv} = 16$ 
          else
10          $N_{uv} = N_{uv} - 8$ 
        elseif ( $E_f > 12$ )
             $N_{uv} = N_{uv} + 1$ 

15      $N_{uv} = \max(\min(300, N_{uv}), 0)$ 
        if ( $VAD = 0$ )
             $\overline{N}_{uv} = 0.95 \cdot \overline{N}_{uv}$ 
             $N_{uv} = \min(125, N_{uv})$ 
20     else
             $\overline{N}_{uv} = 0.9 \cdot \overline{N}_{uv} + 0.1 \cdot N_{uv}$ 

```

[0055] Кроме того, когда долгосрочное среднее значение N_{uv} является очень высоким и девиация σ_E также является высокой в некотором кадре ($N_{uv} > 140$ и $\sigma_E > 5$ в текущем примере), что означает, что текущий сигнал навряд ли будет музыкой, долгосрочное среднее значение обновляется в этом кадре по-другому. Обновление осуществляется так, чтобы оно сходилось к значению 100 и смещало решение в сторону речи. Это делается, как показано ниже:

$$30 \quad \overline{N}_{uv}^t = 0.2 \cdot \overline{N}_{uv}^{(t-1)} + 80 \quad (13)$$

[0056] Этот параметр на долгосрочном среднем значении ряда кадров, находящихся между кадрами, классифицированными как НЕВОКАЛИЗИРОВАННЫЕ, используется для определения того, должен ли этот кадр рассматриваться как ОБЩИЙ ЗВУК или нет. Чем ближе друг к другу по времени НЕВОКАЛИЗИРОВАННЫЕ кадры, тем более вероятно, что сигнал имеет речевую характеристику (менее вероятно, что он является ОБЩИМ ЗВУКОВЫМ сигналом). В иллюстративном примере порог для принятия решения о том, что кадр следует рассматривать как ОБЩИЙ ЗВУК G_A , определяется следующим образом:

$$40 \quad \text{Кадр является ОБЩИМ ЗВУКОМ } G_A, \text{ если: } N_{uv} > 100 \text{ и } \Delta_E^t < 12 \quad (14)$$

[0057] Параметр Δ_E^t , определенный в уравнении (9), используется в условии (14) для того, чтобы избежать классификации большой энергетической вариации в качестве ОБЩЕГО ЗВУКА.

45 [0058] Постобработка, выполняемая на возбуждении, зависит от классификации сигнала. Для некоторых типов сигналов модуль постобработки вообще не используется. Следующая таблица показывает все случаи, в которых выполняется постобработка.

Таблица 3 Категории сигнала для модификации возбуждения	
Классификация кадра	Использовать ли модуль постобработки, да/нет
ВОКАЛИЗИРОВАННЫЙ	да
ОБЩИЙ ЗВУК	да
НЕВОКАЛИЗИРОВАННЫЙ	нет
НЕАКТИВНЫЙ	нет

[0059] Когда используется модуль постобработки, другой анализ энергетической устойчивости, описываемый ниже, выполняется на спектральной энергии объединенного возбуждения. Аналогично описанному в публикации Vaillancourt'050, этот второй анализ энергетической устойчивости дает указание, где именно в спектре должна начаться постобработка и в какой степени она должна быть применена.

2) Создание вектора возбуждения

[0060] Для того, чтобы увеличить частотное разрешение, используется частотное преобразование более длинное, чем длина кадра. Чтобы сделать это, в иллюстративном варианте осуществления в блоке 120 объединения возбуждения создается объединенный вектор возбуждения $e_c(n)$ путем объединения последних 192 отсчетов предыдущего кадра возбуждения, сохраненного в буферной памяти 106 прошлого возбуждения, декодированного возбуждения текущего кадра $e(n)$ из декодера 104 возбуждения во временной области, и экстраполяции 192 отсчетов возбуждения будущего кадра $e_x(n)$ из блока 118 экстраполяции возбуждения. Это описывается ниже, где L_w является длиной прошлого возбуждения, а также длиной экстраполируемого возбуждения, а L является длиной кадра. Это соответствует 192 и 256 отсчетам соответственно, давая полную длину $L_c = 640$ отсчетов в иллюстративном варианте осуществления:

$$e_c(n) = \begin{cases} e(n) & n = -L_w, \dots, -1 \\ e(n) & n = 0, \dots, L-1 \\ e_x(n) & n = L, \dots, L+L_w-1 \end{cases} \quad (15)$$

[0061] В декодере CELP сигнал $e(n)$ возбуждения во временной области задается формулой

$$e(n) = bv(n) + gc(n)$$

[0062] где $v(n)$ является вкладом адаптивной кодировочной книги, b является усилением адаптивной кодировочной книги, $c(n)$ является вкладом фиксированной кодировочной книги, и g является усилением фиксированной кодировочной книги. Экстраполяция будущих отсчетов возбуждения $e_x(n)$ вычисляется в блоке 118

экстраполяции возбуждения путем периодического расширения сигнала возбуждения $e(n)$ текущего кадра из декодера 104 возбуждения во временной области с использованием декодированной фракционной высоты тона последнего подкадра текущего кадра. Учитывая фракционное разрешение задержки высоты тона, повышающая дискретизация возбуждения текущего кадра выполняется с использованием кадрирующей синусоидальной функции Хэмминга длиной 35 отсчетов.

3) Кадрирование

[0063] В модуле 122 кадрирования и частотного преобразования перед преобразованием из временной в частотную область выполняется кадрирование объединенного возбуждения. Выбранное окно $w(n)$ имеет плоскую вершину, соответствующую текущему кадру, и уменьшается по функции Хэмминга до 0 на каждом конце. Следующее уравнение представляет используемое окно:

$$w(n) = \begin{cases} 0.5 \left(1 - \cos \left(\frac{2\pi(n+L_w)}{2L_w-1} \right) \right) & n = -L_w, \dots, -1 \\ 1.0 & n = 0, \dots, L-1 \\ 0.5 \left(1 - \cos \left(\frac{2\pi((n-L)+L_w)}{2L_w-1} \right) \right) & n = L, \dots, L+L_w-1 \end{cases} \quad (16)$$

[0064] При применении к объединенному возбуждению при практической реализации получается вход для частотного преобразования, имеющий полную длину $L_c=640$ отсчетов ($L_c=2L_w+L$). Кадрованное объединенное возбуждение $e_{wc}(n)$ центруется на текущем кадре и представляется следующим уравнением:

$$e_{wc}(n) = \begin{cases} e(n)w(n) & n = -L_w, \dots, -1 \\ e(n)w(n) & n = 0, \dots, L-1 \\ e_x(n)w(n) & n = L, \dots, L+L_w-1 \end{cases} \quad (17)$$

4) Частотное преобразование

[0065] Во время фазы постобработки в частотной области объединенное возбуждение представляется в домене преобразования. В этом иллюстративном варианте осуществления преобразование из временной в частотную область достигается в модуле 122 кадрования и частотного преобразования, использующем дискретное косинусное преобразование типа II, дающее разрешение 10 Гц, однако может использоваться любое другое преобразование. В случае, если используется другое преобразование (или другая длина преобразования), частотное разрешение (определенное выше), количество полос и количество элементов разрешения на полосу (определенное ниже), может быть соответственно пересмотрено. Частотное представление объединенного и кадрованного возбуждения CELP во временной области f_c определяется следующим образом:

$$f_e(k) = \begin{cases} \sqrt{\frac{1}{L_c}} \cdot \sum_{n=0}^{L_c-1} e_{wc}(n), & k = 0 \\ \sqrt{\frac{2}{L_c}} \cdot \sum_{n=0}^{L_c-1} e_{wc}(n) \cdot \cos \left(\frac{\pi}{L_c} \left(n + \frac{1}{2} \right) k \right), & 1 \leq k \leq L_c - 1 \end{cases} \quad (18)$$

[0066] Где $e_{wc}(n)$ представляет собой объединенное и кадрованное возбуждение во временной области, а L_c является длиной частотного преобразования. В этом иллюстративном варианте осуществления длина кадра L составляет 256 отсчетов, но длина частотного преобразования L_c составляет 640 отсчетов для соответствующей внутренней частоты оцифровки, равной 12,8 кГц.

5) Анализ энергии на полосу и на элемент разрешения

[0067] После дискретного косинусного преобразования получаемый спектр делится на полосы критических частот (практическая реализация использует 17 критических полос в частотном диапазоне 0-4000 Гц и 20 полос критических частот в частотном диапазоне 0-6400 Гц). Используемые зоны критических частот являются максимально возможно близкими к тому, что определяется в публикации J. D. Johnston, «Transform coding of audio signal using perceptual noise criteria», IEEE J. Select. Areas Commun., vol. 6,

pp. 314-323, Feb. 1988, содержание которой включено в настоящий документ посредством ссылки, и их верхние границы определяются следующим образом:

$C_B = \{100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400\}$ Гц.

5 [0068] Дискретное косинусное преобразование с 640 точками дает частотное разрешение 10 Гц (6400 Гц / 640 точек). Количество частотных элементов разрешения на полосу критической частоты составляет

$M_{cb} = \{10, 10, 10, 10, 11, 12, 14, 15, 16, 19, 21, 24, 28, 32, 38, 45, 55, 70, 90, 110\}$.

10 [0069] Средняя спектральная энергия на полосу критической частоты $E_b(i)$ вычисляется следующим образом:

$$E_b(i) = \frac{1}{L_c M_{CB}(i)} \sum_{h=0}^{M_B(i)-1} (f_e(h + j_i))^2, \quad i = 0, \dots, 20 \quad (19)$$

15 [0070] где $f_e(h)$ представляет h-й элемент разрешения по частоте критической полосы, а j_i является индексом первого элемента разрешения в i-й критической полосе, определяемым как

$j_i = \{0, 10, 20, 30, 40, 51, 63, 77, 92, 108, 127, 148, 172, 200, 232, 270, 315, 370, 440, 530\}$.

20 [0071] Спектральный анализ также вычисляет энергию спектра на элемент разрешения по частоте $E_{BIN}(k)$ с использованием следующего соотношения:

$$E_{BIN}(k) = \frac{1}{L_c} f_e(k)^2, \quad k = 0, \dots, 639 \quad (20)$$

25 [0072] Наконец, спектральный анализ вычисляет полную спектральную энергию E_c объединенного возбуждения как сумму спектральных энергий первых 17 полос критических частот с использованием следующего соотношения:

$$E_c = 10 \log_{10} \left(\sum_{i=0}^{16} E_b(i) \right) - 3.0103, \text{ dB} \quad (21)$$

30 б) Классификация сигнала возбуждения второго этапа

[0073] Как описано в публикации Vaillancourt'050, способ для улучшения декодированного общего звукового сигнала включает в себя дополнительный анализ сигнала возбуждения, спроектированный для того, чтобы дополнительно максимизировать эффективность межгармонического шумоподавления путем
35 идентификации того, какой кадр хорошо подходит для межтонального шумоподавления.

[0074] Классификатор 124 сигнала второго этапа не только дополнительно разделяет декодированное объединенное возбуждение на категории звукового сигнала, но также дает инструкции блоку 128 межгармонического уменьшения шума относительно
40 максимального уровня затухания и минимальной частоты, где может начинаться это уменьшение.

[0075] В представленном иллюстративном примере классификатор 124 сигнала второго этапа был сохранен настолько простым, насколько это возможно, и очень похож на классификатор типа сигнала, описанный в публикации Vaillancourt'050. Первая операция заключается в выполнении анализа энергетической устойчивости аналогично
45 тому, как это делается в уравнениях (9) и (10), но используя в качестве ввода полную спектральную энергию объединенного возбуждения E_c , как сформулировано в уравнении (21):

$$\bar{E}_d = \frac{\left(\sum_{t=-40}^{t=-1} \Delta_{E_c}^t \right)}{40}, \quad \text{where } \Delta_{E_c}^t = E_c^t - E_c^{(t-1)} \quad (22)$$

5 [0076] где E_d представляет среднюю разницу энергий объединенных векторов возбуждения двух смежных кадров, E_c^t представляет энергию объединенного возбуждения текущего кадра t , и $E_c^{(t-1)}$ представляет энергию объединенного возбуждения предыдущего кадра $t-1$. Среднее значение вычисляется по последним 40 кадрам.

10 [0077] Затем статистическая девиация σ_c вариации энергии по последним пятнадцати (15) кадрам вычисляется с использованием следующего соотношения:

$$15 \quad \sigma_c = p \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{(\Delta_{E_c}^t - \bar{E}_d)^2}{15}} \quad (23)$$

[0078] где при практической реализации масштабный коэффициент p находится экспериментально и устанавливается равным приблизительно 0,77. Получаемая девиация σ_c сравнивается с четырьмя (4) плавающими порогами для определения того, в какой степени шум между гармониками может быть уменьшен. Выход этого классификатора 124 сигнала второго этапа расщепляется на пять (5) категорий звукового сигнала e_{SAT} , называемых категориями звукового сигнала 0-4. Каждая категория звукового сигнала имеет свою собственную настройку межтонального шумоподавления.

25 [0079] Пять (5) категорий звукового сигнала 0-4 могут быть определены, как указано в следующей Таблице.

Таблица 4 Выходные характеристики классификатора возбуждения		
Категория	Улучшенная полоса (широкая полоса)	Допустимое уменьшение
e_{SAT}	Гц	дБ
0	неприменимо	0
1	[920, 6400]	6
2	[920, 6400]	9
3	[770, 6400]	12
4	[630, 6400]	12

30 [0080] Категория 0 звукового сигнала является категорией нетонального, неустойчивого звукового сигнала, которая не модифицируется методом межтонального шумоподавления. Эта категория декодированного звукового сигнала имеет самую большую статистическую девиацию вариации спектральной энергии, и в большинстве случаев включает в себя речевой сигнал.

40 [0081] Категория 1 звукового сигнала (самая большая статистическая девиация вариации спектральной энергии после категории 0) обнаруживается, когда статистическая девиация σ_c вариации спектральной энергии ниже Порога 1, и последняя обнаруженная категория звукового сигнала ≥ 0 . Тогда максимальное уменьшение шума квантования декодированного тонального возбуждения в пределах полосы частот от 920 Гц до $F_s/2$ Гц (6400 Гц в этом примере, где F_s является частотой оцифровки) ограничивается максимальным шумоподавлением R_{max} с величиной 6 дБ.

[0082] Категория 2 звукового сигнала обнаруживается, когда статистическая девиация

σ_c вариации спектральной энергии ниже Порога 2, и последняя обнаруженная категория звукового сигнала ≥ 1 . Тогда максимальное уменьшение шума квантования декодированного тонального возбуждения в пределах полосы частот от 920 Гц до $F_s/2$ Гц ограничивается максимумом в 9 дБ.

[0083] Категория 3 звукового сигнала обнаруживается, когда статистическая девиация σ_c вариации спектральной энергии ниже Порога 3, и последняя обнаруженная категория звукового сигнала ≥ 2 . Тогда максимальное уменьшение шума квантования декодированного тонального возбуждения в пределах полосы частот от 770 Гц до $F_s/2$ Гц ограничивается максимумом в 12 дБ.

[0084] Категория 4 звукового сигнала обнаруживается, когда статистическая девиация σ_c вариации спектральной энергии ниже Порога 4, и последняя обнаруженная категория звукового сигнала ≥ 3 . Тогда максимальное уменьшение шума квантования декодированного тонального возбуждения в пределах полосы частот от 630 Гц до $F_s/2$ Гц ограничивается максимумом в 12 дБ.

[0085] Плавающие пороги 1-4 помогают предотвратить неправильную классификацию типа сигнала. Как правило, декодированный тональный звуковой сигнал, представляющий музыку, получает намного более низкую статистическую девиацию вариации своей спектральной энергии, чем речь. Однако даже музыкальный сигнал может содержать сегмент более высокой статистической девиации, и аналогичным образом речевой сигнал может содержать сегменты с более низкой статистической девиацией. Тем не менее маловероятно, чтобы речь и музыкальный контент регулярно чередовались от одного кадра к другому. Плавающие пороги добавляют гистерезис решения и действуют как усиление предыдущего состояния для того, чтобы по существу предотвратить ошибочную классификацию, которая может привести к неоптимальной эффективности блока 128 межгармонического уменьшения шума.

[0086] Счетчики последовательных кадров категории 0 звукового сигнала и счетчики последовательных кадров категории 3 или 4 звукового сигнала используются для того, чтобы соответственно уменьшить или увеличить эти пороги.

[0087] Например, если счетчик подсчитывает серию из более чем 30 кадров звукового сигнала категории 3 или 4, все плавающие пороги (1-4) увеличиваются на predetermined значение с целью разрешения рассматривать большее количество кадров как категорию 4 звукового сигнала.

[0088] Обратное также справедливо для категории 0 звукового сигнала. Например, если насчитывается серия из более чем 30 кадров звукового сигнала категории 0, все плавающие пороги (1-4) уменьшаются с целью разрешения рассматривать большее количество кадров как категорию 0 звукового сигнала. Все плавающие пороги 1-4 ограничиваются абсолютными максимальными и минимальными значениями для того, чтобы гарантировать, что классификатор сигнала не блокируется на фиксированной категории.

[0089] В случае стирания кадра все пороги 1-4 вновь устанавливаются равными их минимальным величинам, и выход классификатора второго этапа рассматривается как нетональный (категория 0 звукового сигнала) для трех (3) последовательных кадров (включая потерянный кадр).

[0090] Если доступна информация от детектора речевой активности (VAD), и она указывает на отсутствие речевой активности (наличие тишины), решение классификатора второго этапа насильно устанавливается в категорию 0 звукового сигнала ($\epsilon_{CAT} = 0$).

7) Межгармоническое шумоподавление в домене возбуждения

[0091] Межтональное или межгармоническое шумоподавление выполняется на частотном представлении объединенного возбуждения как первая операция улучшения. Уменьшение шума межтонального квантования выполняется в блоке 128 уменьшения шума путем масштабирования спектра в каждой критической полосе масштабирующим усилением g_s , ограниченным минимальным и максимальным усилением g_{\min} и g_{\max} . Масштабирующее усиление выводится из оценки отношения сигнал/шум (SNR) в этой критической полосе. Эта обработка выполняется на основе частотных элементов разрешения, а не на основе критических полос. Таким образом, масштабирующее усиление применяется ко всем частотным элементам разрешения, и оно выводится из SNR, вычисленного с использованием энергии элемента разрешения, деленной на оценку энергии шумов критической полосы, включающей в себя этот элемент разрешения. Эта особенность позволяет сохранить энергию на частотах около гармоник или тонов, таким образом по существу предотвращая искажение, одновременно с этим сильно уменьшая шум между гармониками.

[0092] Межтональное шумоподавление выполняется на поэлементной основе по всем 640 элементам разрешения. После применения межтонального шумоподавления к спектру выполняется другая операция улучшения спектра. Затем обратное дискретное косинусное преобразование используется для того, чтобы реконструировать сигнал улучшенного объединенного возбуждения e_{td} , как будет описано позже.

[0093] Минимальное масштабирующее усиление g_{\min} выводится из максимально допустимого межтонального шумоподавления R_{\max} в дБ. Как описано выше, второй этап классификации дает максимально допустимое понижение в диапазоне от 6 до 12 дБ. Таким образом минимальное масштабирующее усиление определяется как

$$g_{\min} = 10^{-R_{\max}/20} \quad (24)$$

[0094] Масштабирующее усиление вычисляется относительно значения SNR на элемент разрешения. Затем поэлементное шумоподавление выполняется, как упомянуто выше. В текущем примере поэлементная обработка применяется ко всему спектру до максимальной частоты 6400 Гц. В этом иллюстративном варианте осуществления шумоподавление начинается в 6-й критической полосе (то есть никакого шумоподавления не выполняется ниже 630 Гц). Для того чтобы уменьшить негативное воздействие метода, классификатор второго этапа может сместить начальную критическую полосу вплоть до 8-й полосы (920 Гц). Это означает, что первая критическая полоса, на которой выполняется шумоподавление, находится между 630 Гц и 920 Гц, и это может изменяться от кадра к кадру. В более консервативной реализации минимальная полоса, где начинается шумоподавление, может быть установлена выше.

[0095] Масштабирование для определенного элемента разрешения по частоте k вычисляется как функция SNR, определяемая выражением

$$g_s(k) = \sqrt{k_s \text{SNR}(k) + c_s}, \text{ ограниченным как } g_{\min} \leq g_s \leq g_{\max} \quad (25)$$

[0096] Обычно значение g_{\max} равно 1 (то есть усиление не выполняется), затем определяются значения k_s и c_s таким образом, что $g_s = g_{\min}$ для SNR = 1 дБ, и $g_s = 1$ для SNR = 45 дБ. Таким образом, для значения SNR 1 дБ и ниже масштабирование ограничивается величиной g_{\min} , а для значения SNR 45 дБ и выше никакое шумоподавление не выполняется ($g_s = 1$). Таким образом, учитывая эти две конечных

точки, значения k_s и c_s в Уравнении (25) определяются как

$$k_s = (1 - g_{\min}^2) / 44 \quad \text{и} \quad c_s = (45g_{\min}^2 - 1) / 44 \quad (26)$$

5 [0097] Если значение g_{\max} имеет величину выше 1, то оно позволяет процессу слегка усиливать тоны, имеющие самую высокую энергию. Это может использоваться для того, чтобы компенсировать тот факт, что кодек CELP, используемый в практической реализации, не полностью выравнивает энергию в частотной области. Это обычно имеет место для сигналов, отличающихся от вокализированной речи.

10 [0098] Значение SNR на элемент разрешения в определенной критической полосе i вычисляется как

$$\text{NRF}_{\text{BIN}}(h) = \frac{0.3E_{\text{BIN}}^{(1)}(h) + 0.7E_{\text{BIN}}^{(2)}(h)}{N_B(i)}, \quad h = j_i, \dots, j_i + M_B(i) - 1 \quad (27)$$

15 [0099] где $E_{\text{BIN}}^{(1)}(h)$ и $E_{\text{BIN}}^{(2)}(h)$ обозначают энергию на элемент разрешения по частоте для спектрального анализа прошлого и текущего кадра соответственно, вычисленную по уравнению (20), $N_B(i)$ обозначает оценку энергии шумов критической полосы i , j_i является индексом первого элемента разрешения в i -й критической полосе, и $M_B(i)$ является количеством элементов разрешения в критической полосе i , как определено выше.

[00100] Коэффициент сглаживания является адаптивным, и он сделан обратно относящимся к самому усилению. В этом иллюстративном варианте осуществления коэффициент сглаживания задается выражением $\alpha_{gs} = 1 - g_s$. Иначе говоря, сглаживание является более сильным для меньших усилений g_s . Этот подход по существу предотвращает искажение в сегментах с высоким значением SNR, которым предшествуют кадры с низким значением SNR, как это имеет место для вокализированных вступлений. В иллюстративном варианте осуществления процедура сглаживания способна быстро адаптироваться и использовать более низкие масштабирующие усиления на вступлениях.

30 [00101] В случае поэлементной обработки в критической полосе с индексом i , после определения масштабирующего усиления, как в уравнении (25), и использования значения SNR, как определено в уравнениях (27), фактическое масштабирование выполняется с использованием сглаженного масштабирующего усиления $g_{\text{BIN},LP}$, обновляемого при каждом частотном анализе следующим образом

$$g_{\text{BIN},LP}(k) = \alpha_{gs} g_{\text{BIN},LP}(k) + (1 - \alpha_{gs}) g_s \quad (28)$$

40 [00102] Временное сглаживание усилений по существу предотвращает слышимые колебания энергии, в то время как управление сглаживанием с использованием α_{gs} по существу предотвращает искажение в сегментах с высоким значением SNR, которым предшествуют кадры с низким значением SNR, как это имеет место для вокализированных вступлений или атак.

45 [00103] Масштабирование в критической полосе i выполняется как

$$f_e'(h + j_i) = g_{\text{BIN},LP}(h + j_i) f_e(h + j_i), \quad h = 0, \dots, M_B(i) - 1 \quad (29)$$

[00104] где j_i является индексом первого элемента разрешения в критической полосе

i , а $M_B(i)$ является количеством элементов разрешения в этой критической полосе.

[00105] Сглаженные масштабирующие усиления $g_{BIN,LP}(k)$ первоначально устанавливаются равными 1. Каждый раз, когда обрабатывается нетональный звуковой кадр, $e_{CAT}=0$, сглаженные масштабирующие усиления вновь устанавливаются равными 1,0 для того, чтобы уменьшить любое возможное понижение в следующем кадре.

[00106] Следует отметить, что при каждом спектральном анализе сглаженные масштабирующие усиления $g_{BIN,LP}(k)$ обновляется для всех частотных элементов разрешения во всем спектре. Также следует отметить, что в случае низкоэнергетического сигнала межтональное шумоподавление ограничено величиной -1,25 дБ. Это происходит, когда максимальная энергия шумов во всех критических полосах $\max(N_b(i))$, $i = 0 \dots, 20$, меньше или равна 10.

8) Оценка шума межтонального квантования

[00107] В этом иллюстративном варианте осуществления энергия шумов межтонального квантования на полосу критической частоты оценивается в блоке 126 оценки уровня шума в полосе как средняя энергия этой полосы критической частоты за исключением максимальной энергии элемента разрешения этой же самой полосы. Следующая формула суммирует оценку энергии шумов квантования для конкретной полосы i :

$$N_B(i) = \frac{1}{q(i)} \left(\frac{\left(E_B(i) M_B(i) - \max_h (E_{BIN}(h + j_i)) \right)}{(M_B(i) - 1)} \right), \quad h = 0, \dots, M_B(i) - 1 \quad (30)$$

[00108] где j_i является индексом первого элемента разрешения в критической полосе i , $M_B(i)$ является количеством элементов разрешения в этой критической полосе, $E_B(i)$ является средней энергией полосы i , $E_{BIN}(h+j_i)$ является энергией конкретного элемента разрешения, и $N_B(i)$ является получаемой оценкой энергии шумов конкретной полосы i . В уравнении (30) оценки шума величина $q(i)$ представляет шумовой масштабирующий коэффициент на полосу, который находится экспериментально и может модифицироваться в зависимости от реализации, в которой используется постобработка. При практической реализации шумовой масштабирующий коэффициент устанавливается так, чтобы больше шума могло быть удалено на низких частотах и меньше шума могло быть удалено на высоких частотах, как показано ниже:

$$q = \{10, 10, 10, 10, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 15, 15, 15, 15, 15\}.$$

9) Увеличение динамики спектра возбуждения

[00109] Вторая операция частотной постобработки обеспечивает возможность восстановления частотной информации, которая была потеряна в шумах кодирования. Кодеки CELP, особенно когда они используются при низких скоростях передачи, не очень эффективны для кодирования частотного контента выше 3,5-4 кГц. Главная идея здесь состоит в том, чтобы использовать преимущество того факта, что музыкальный спектр зачастую не изменяется существенно от кадра к кадру. Следовательно, может быть сделано долгосрочное усреднение, и часть шума кодирования может быть устранена. Следующие операции выполняются для того, чтобы определить частотно-зависимую функцию усиления. Эта функция затем используется для того, чтобы дополнительно улучшить возбуждение перед его обратным преобразованием во временную область.

а. Поэлементная нормализация энергии спектра

[00110] Первая операция заключается в создании в блоке 130 формирования маски весовой маски на основе нормализованной энергии спектра объединенного возбуждения. Нормализация выполняется в блоке 131 нормализации энергии спектра так, чтобы тона (или гармоники) имели значение выше 1,0, а впадины имели значение ниже 1,0. Для того чтобы сделать это, энергетический спектр $E_{BIN}(k)$ элемента разрешения нормализуется в диапазоне от 0,925 до 1,925, с тем чтобы получить нормализованный энергетический спектр $E_n(k)$, с использованием следующего уравнения:

$$E_n(k) = \frac{E_{BIN}(k)}{\max(E_{BIN})} + 0.925, \quad k = 0, \dots, 639 \quad (31)$$

[00111] где $E_{BIN}(k)$ представляет энергию элемента разрешения, вычисленную в соответствии с уравнением (20). Так как нормализация выполняется в энергетической области, многие элементы разрешения имеют очень низкие значения. При практической реализации смещение 0,925 было выбрано так, чтобы только небольшая часть нормализованных энергетических элементов разрешения имела значение меньше 1,0. Как только нормализация выполнена, полученный нормализованный энергетический спектр обрабатывается с помощью степенной функции для того, чтобы получить масштабированный энергетический спектр. В этом иллюстративном примере используется степень 8 для того, чтобы ограничить минимальные величины масштабированного энергетического спектра величиной приблизительно 0,5, как показано в следующей формуле:

$$E_p(k) = E_n(k)^8 \quad k = 0, \dots, 639 \quad (32)$$

[00112] где $E_n(k)$ является нормализованным энергетическим спектром, а $E_p(k)$ является масштабированным энергетическим спектром. Более агрессивная степенная функция может использоваться для того, чтобы еще больше уменьшить шум квантования, например, может быть выбрана степень 10 или 16, возможно со смещением, более близким к единице. Однако попытка удалить слишком много шума может также привести к потере важной информации.

[00113] Использование степенной функции без ограничения ее выхода быстро приводит к насыщению для значений энергетического спектра, больших единицы. Максимальный предел масштабированного энергетического спектра, таким образом, устанавливается равным 5 при практической реализации, создавая отношение между максимальным и минимальным значениями нормализованной энергии, равное приблизительно 10. Это является полезным, учитывая, что доминирующий элемент разрешения может иметь слегка отличающееся положение от одного кадра к другому, так что предпочтительно, чтобы весовая маска была относительно устойчивой от одного кадра к следующему кадру. Следующее уравнение показывает, как применяется эта функция:

$$E_{pl}(k) = \min(5, E_p(k)) \quad k = 0, \dots, 639 \quad (33)$$

[00114] где $E_{pl}(k)$ представляет ограниченный масштабированный энергетический спектр, а $E_p(k)$ является масштабированным энергетическим спектром, как определено в уравнении (32).

б. Сглаживание масштабированного энергетического спектра вдоль частотной оси и временной оси

[00115] В ходе последних двух операций начинает формироваться положение большинства энергетических импульсов. Применение степени 8 к элементам разрешения нормализованного энергетического спектра является первой операцией, которая создаст эффективную маску для увеличения динамики спектра. Следующие две (2) операции
5 дополнительно улучшают эту маску спектра. Сначала масштабированный энергетический спектр сглаживается в блоке 132 усреднения энергии вдоль частотной оси от низких частот до высоких частот с использованием усредняющего фильтра. Затем полученный спектр обрабатывается в блоке 134 сглаживания энергии вдоль оси временной области для того, чтобы сгладить значения элементов разрешения от кадра
10 к кадру.

[00116] Сглаживание масштабированного энергетического спектра вдоль частотной оси может быть описано следующей функцией:

$$\bar{E}_{pl}(k) = \begin{cases} \frac{E_{pl}(k) + E_{pl}(k+1)}{2}, & k = 0 \\ \frac{E_{pl}(k-1) + E_{pl}(k) + E_{pl}(k+1)}{3}, & k = 1, \dots, 638 \\ \frac{E_{pl}(k-1) + E_{pl}(k)}{2}, & k = 639 \end{cases} \quad (34)$$

[00117] Наконец, сглаживание вдоль оси времени приводит к усредненной во времени весовой маске G_m усиления/ослабления, которая должна быть применена к спектру f'_e . Весовая маска, также называемая маской усиления, описывается следующим уравнением:

$$G'_m(k) = \begin{cases} 0.95 \cdot G_m^{(t-1)}(k) + 0.05 \bar{E}_{pl}(k), & k = 0, \dots, 319 \\ 0.85 \cdot G_m^{(t-1)}(k) + 0.15 \bar{E}_{pl}(k), & k = 320, \dots, 639 \end{cases} \quad (35)$$

[00118] где E_{pl} является масштабированным энергетическим спектром, сглаженным
30 вдоль частотной оси, t является индексом кадра, а G_m является усредненной во времени весовой маской.

[00119] Более медленная скорость адаптации была выбрана для более низких частот для того, чтобы по существу предотвратить колебания усиления. Более быстрая скорость адаптации обеспечивается для более высоких частот, так как положения тонов с большей
35 вероятностью быстро меняются в более высокой части спектра. При усреднении, выполняемом на частотной оси, и долгосрочном сглаживании, выполняемом вдоль оси времени, конечный вектор, полученный в выражении (35), используется в качестве весовой маски, применяемой непосредственно к улучшенному спектру объединенного возбуждения f'_e уравнения (29).

10) Применение весовой маски к улучшенному спектру объединенного возбуждения

[00120] Весовая маска, определенная выше, применяется по-разному блоком 136 модификации динамики спектра в зависимости от выхода классификатора возбуждения второго этапа (от значения e_{SAT} , показанного в таблице 4). Весовая маска не
45 применяется, если возбуждение классифицируется как категория 0 ($e_{SAT} = 0$; т.е. высокая вероятность наличия речи). Когда скорость передачи кодека является высокой, уровень шума квантования является в большинстве случаев низким, и он изменяется в зависимости от частоты. Это означает, что усиление тонов может быть ограничено в зависимости от положений импульсов в спектре и закодированной скорости передачи.

При использовании способа кодирования, отличающегося от CELP, например, если сигнал возбуждения включает в себя комбинацию компонентов, закодированных во временной области и в частотной области, использование весовой маски может быть скорректировано для каждого конкретного случая. Например, усиление импульса
5 может быть ограничено, но способ может все еще использоваться для уменьшения шума квантования.

[00121] Для первого 1 кГц (первые 100 элементов разрешения в практической реализации) маска применяется, если возбуждение не классифицируется как возбуждение категории 0 ($e_{CAT} \neq 0$). Ослабление возможно, однако никакого усиления не выполняется
10 в этом частотном диапазоне (максимальное значение маски ограничено величиной 1,0).

[00122] Если больше чем 25 последовательных кадров классифицируются как кадры категории 4 ($e_{CAT} = 4$; то есть высокая вероятность музыкального контента), но не более 40 кадров, тогда весовая маска применяется без усиления для всех остающихся элементов разрешения (элементы разрешения 100-639) (максимальное усиление G_{max0}
15 ограничивается величиной 1,0, и нет никакого ограничения на минимальное усиление).

[00123] Когда более 40 кадров классифицируются как кадры категории 4, для частот между 1 и 2 кГц (элементы разрешения 100-199 в практической реализации) максимальное усиление G_{max1} устанавливается равным 1,5 для скоростей передачи ниже 12650 бит в
20 секунду (бит/с). В противном случае максимальное усиление G_{max1} устанавливается равным 1,0. В этой полосе частот минимальное усиление G_{min1} устанавливается равным 0,75, только если скорость передачи является более высокой, чем 15850 бит/с, в противном случае нет никакого ограничения на минимальное усиление.

[00124] Для полосы от 2 до 4 кГц (элементы разрешения 200-399 в практической реализации), максимальное усиление G_{max2} ограничивается величиной 2,0 для скоростей
25 передачи ниже 12650 бит/с, и ограничивается величиной 1,25 для скоростей передачи, равных или выше чем 12650 бит/с и меньше 15850 бит/с. В противном случае максимальное усиление G_{max2} ограничивается величиной 1,0. В этой же полосе частот
30 минимальное усиление G_{min2} устанавливается равным 0,5, только если скорость передачи является более высокой, чем 15850 бит/с, в противном случае нет никакого ограничения на минимальное усиление.

[00125] Для полосы от 4 до 6,4 кГц (элементы разрешения 400-639 в практической реализации), максимальное усиление G_{max3} ограничивается величиной 2,0 для скоростей
35 передачи ниже 15850 бит/с, и величиной 1,25 в противном случае. В этой полосе частот минимальное усиление G_{min3} устанавливается равным 0,5, только если скорость передачи является более высокой, чем 15850 бит/с, в противном случае нет никакого ограничения на минимальное усиление. Следует отметить, что другие настройки максимального и
40 минимального усиления могут быть подходящими в зависимости от характеристик кодека.

[00126] Следующий псевдокод показывает, как воздействует на окончательный спектр объединенного возбуждения f'_e применение весовой маски G_m к улучшенному спектру f' . Следует отметить, что первая операция улучшения спектра (как описано в
45 секции 7) не является абсолютно необходимой для того, чтобы выполнить эту вторую операцию улучшения путем поэлементной модификации усиления.

if ($e_{CAT} \neq 0$)

if ($e_{CAT} == 4 \forall t = -1, \dots -40$)

$$f''_e(k) = \begin{cases} f'_e(k) \min(G_m(k), G_{max0}), & k = 0, \dots, 99 \\ f'_e(k) \max(\min(G_m(k), G_{max1}), G_{min1}), & k = 100, \dots, 199 \\ f'_e(k) \max(\min(G_m(k), G_{max2}), G_{min2}), & k = 200, \dots, 399 \\ f'_e(k) \max(\min(G_m(k), G_{max3}), G_{min3}), & k = 400, \dots, 639 \end{cases}$$

else if ($e_{CAT} == 4 \forall t = -1, \dots -25$)

$$f''_e(k) = f'_e(k) \min(G_m(k), 1.0), \quad k = 0, \dots, 639$$

else

$$f''_e(k) = f'_e(k), \quad k = 0, \dots, 639 \quad (36)$$

[00127] Здесь f'_e представляет спектр объединенного возбуждения, предварительно улучшенный относящейся к SNR функцией $g_{BINLP}(k)$ уравнения (28), G_m является весовой маской, вычисленной в уравнении (35), G_{max} и G_{min} являются максимальным и минимальным усилениями частотного диапазона, определенными выше, t является индексом кадра, где $t=0$ соответствует текущему кадру, и, наконец, f''_e представляет собой окончательный улучшенный спектр объединенного возбуждения.

11) Обратное частотное преобразование

[00128] После того как улучшение в частотной области завершено, обратное преобразование из частотной области во временную область выполняется в блоке 138 преобразования из частотной области во временную область для того, чтобы вернуть улучшенное возбуждение во временную область. В этом иллюстративном варианте осуществления преобразование из частотной области во временную область достигается с помощью того же самого дискретного косинусного преобразования типа II, которое используется для преобразования и временной области в частотную область. Модифицированное возбуждение e'_{id} во временной области получается как

$$e'_{id}(n) = \begin{cases} \sqrt{\frac{1}{L_c}} \cdot \sum_{k=0}^{L_c-1} f''_e(k), & n = 0 \\ \sqrt{\frac{2}{L_c}} \cdot \sum_{k=0}^{L_c-1} f''_e(k) \cdot \cos\left(\frac{\pi}{L_c} \left(k + \frac{1}{2}\right) n\right), & 1 \leq n \leq L_c - 1 \end{cases} \quad (37)$$

[00129] где f''_e является частотным представлением модифицированного возбуждения, e'_{id} является улучшенным объединенным возбуждением, а L_c является длиной объединенного вектора возбуждения.

12) Фильтрация синтеза и перезапись текущего синтеза CELP

[00130] Так как нежелательно добавлять задержку к синтезу, было решено избегать алгоритма перекрытия и добавления при практической реализации. Практическая реализация берет точную длину конечного возбуждения e_f , используемого для генерирования синтеза, непосредственно из улучшенного объединенного возбуждения, без перекрытия, как показано в уравнении ниже:

$$e_f(n) = e'_{id}(n + L_w), \quad n = 0, \dots, 255 \quad (38)$$

[00131] Здесь L_w представляет длину кадрирования, применяемую к прошлому возбуждению перед частотным преобразованием, как объяснено в уравнении (15). Как только модификация возбуждения выполнена и правильная длина улучшенного модифицированного возбуждения во временной области из блока 138 преобразования из частотной области во временную область извлечена из объединенного вектора с использованием блока 140 извлечения возбуждения кадра, модифицированное возбуждение во временной области обрабатывается с помощью фильтра 110 синтеза для того, чтобы получить улучшенный сигнал синтеза для текущего кадра. Этот улучшенный синтез используется для того, чтобы перезаписать первоначально декодированный синтез из фильтра 108 с тем, чтобы улучшить качество восприятия. Решение о перезаписи принимается блоком 142 перезаписи, включающим в себя контрольную точку 144 принятия решения, управляющую переключателем 146, как описано выше, в ответ на информацию от контрольной точки 116 выбора класса и от классификатора 124 сигнала второго этапа.

[00132] Фиг. 3 представляет собой упрощенную блок-схему примерной конфигурации аппаратных компонентов, формирующих декодер, изображенный на Фиг. 2. Декодер 200 может быть осуществлен как часть мобильного терминала, как часть портативного медиапроигрывателя, или в любом другом подобном устройстве. Декодер 200 включает в себя вход 202, выход 204, процессор 206 и память 208.

[00133] Вход 202 выполнен с возможностью получения потока 102 битов AMR-WB. Вход 202 является обобщением приемника 102, изображенного на Фиг. 2.

Неограничивающие примеры реализации входа 202 включают в себя радиоинтерфейс мобильного терминала, физический интерфейс, такой как, например, порт универсальной последовательной шины (USB) портативного медиапроигрывателя и т.п. Выход 204 является обобщением цифроаналогового преобразователя 154, усилителя 156 и громкоговорителя 158, изображенных на Фиг. 2, и может включать в себя аудиоплеер, громкоговоритель, записывающее устройство и т.п. Альтернативно выход 204 может включать в себя интерфейс, способный соединяться с аудиоплеером, с громкоговорителем, с записывающим устройством и т.п. Вход 202 и выход 204 могут быть осуществлены в общем модуле, например в устройстве последовательного ввода-вывода.

[00134] Процессор 206 оперативно соединяется со входом 202, с выходом 204 и с памятью 208. Процессор 206 реализуется как один или более процессоров для выполнения кодовых инструкций для поддержания функций декодера 104 возбуждения во временной области, фильтров 108 и 110 синтеза LP, классификатора 112 сигнала первого этапа и его компонентов, блока 118 экстраполяции возбуждения, блока 120 объединения возбуждения, модуля 122 кадрирования и частотного преобразования, классификатора 124 сигнала второго этапа, блока 126 оценки уровня шума в полосе, блока 128 уменьшения шума, блока 130 формирования маски и его компонентов, блока 136 модификации динамики спектра, блока 138 преобразования из частотной области во временную область, блока 140 извлечения возбуждения кадра, блока 142 перезаписи и его компонентов, а также фильтра устранения предисказаний и передискретизатора 148.

[00135] Память 208 хранит результаты различных операций постобработки. Более конкретно, память 208 включает в себя буферную память 106 прошлого возбуждения. В некоторых вариантах результаты промежуточной обработки различных функций

процессора 206 могут быть сохранены в памяти 208. Память 208 может дополнительно включать в себя постоянную память для хранения кодовых инструкций, исполняемых процессором 206. Память 208 может также сохранять сигнал звуковой частоты от фильтра устранения предыскажений и передискретизатора 148, подавая хранящийся

5 сигнал звуковой частоты на выход 204 по запросу от процессора 206.

[00136] Специалист в данной области техники поймет, что описание устройства и способа для уменьшения шума квантования в музыкальном сигнале или другом сигнале, содержащемся в возбуждении во временной области, декодируемом декодером временной области, является всего лишь иллюстративным и никоим образом не является

10 ограничивающим. Другие варианты осуществления могут быть легко сформированы специалистами в данной области техники на основе представленного раскрытия. Кроме того, раскрытые устройство и способ могут быть специализированы для того, чтобы предложить ценные решения для существующих потребностей и проблем улучшения воспроизведения музыкального контента кодеками на основе линейного предсказания

15 (LP).

[00137] В интересах ясности показаны и описаны не все обычные признаки реализаций устройства и способа. Следует, конечно, иметь в виду, что при разработке любой такой фактической реализации устройства и способа для уменьшения шума квантования в музыкальном сигнале, содержащемся в возбуждении во временной области,

20 декодируемом декодером временной области, возможно, должны быть приняты многочисленные специфичные для реализации решения, чтобы достигнуть конкретных целей разработчика, таких как соответствие ограничениям, относящимся к применению, системе, сети и организации, и что эти конкретные цели будут изменяться от одной реализации к другой и от одного разработчика к другому. Более того, следует иметь в

25 виду, что опытно-конструкторские работы могут быть сложными и отнимающими много времени, но тем не менее будут представлять собой повседневную деятельность специалистов в области обработки звука, пользующихся выгодами представленного раскрытия.

[00138] В соответствии с настоящим изобретением описанные в настоящем документе компоненты, операции процесса, и/или структуры данных могут быть осуществлены

30 с использованием различных типов операционных систем, вычислительных платформ, сетевых устройств, компьютерных программ и/или машин общего назначения. В дополнение к этому, специалист в данной области техники поймет, что также могут использоваться устройства менее общего назначения, такие как аппаратные устройства,

35 программируемые пользователем вентильные матрицы (FPGA), специализированные интегральные схемы (ASIC) и т.п. Там, где способ, включающий в себя ряд операций процесса, осуществляется компьютером или машиной, и эти операции процесса могут быть сохранены как последовательность машиночитаемых инструкций, они могут быть сохранены на материальном носителе.

[00139] Хотя настоящее изобретение было описано выше посредством не ограничивающих иллюстративных вариантов его осуществления, эти варианты осуществления могут модифицироваться по желанию в рамках прилагаемой формулы изобретения без отступлений от сущности и природы настоящего изобретения.

45 (57) Формула изобретения

1. Устройство, реализованное в декодере линейного предсказания с кодовым возбуждением (CELP) для уменьшения шума квантования в звуковом сигнале, содержащемся в декодированном CELP возбуждении во временной области, подлежащем

обработке фильтром синтеза линейного предсказания (LP) для выполнения его синтеза, при этом упомянутое устройство включает в себя:

преобразователь декодированного CELP возбуждения во временной области до синтеза в возбуждение в частотной области;

5 блок формирования маски для, в ответ на возбуждение в частотной области, формирования весовой маски для восстановления спектральной информации, потерянной в шуме квантования;

модификатор возбуждения в частотной области для увеличения динамики спектра путем применения весовой маски к возбуждению в частотной области; и

10 преобразователь модифицированного возбуждения в частотной области в модифицированное CELP возбуждение во временной области, содержащее версию с уменьшенным шумом квантования звукового сигнала.

2. Устройство по п. 1, включающее в себя:

15 фильтр синтеза LP для выполнения синтеза декодированного CELP возбуждения во временной области; и

классификатор синтеза декодированного CELP возбуждения во временной области на одно из первого набора категорий возбуждения и второго набора категорий возбуждения;

при этом второй набор категорий возбуждения включает в себя категории
20 НЕАКТИВНЫЙ или НЕВОКАЛИЗИРОВАННЫЙ; и

первый набор категорий возбуждения включает в себя категорию ДРУГОЙ.

3. Устройство по п. 2, в котором преобразователь декодированного CELP возбуждения во временной области в возбуждение в частотной области применяется к декодированному CELP возбуждению во временной области, когда синтез
25 декодированного CELP возбуждения во временной области классифицирован в первый набор категорий возбуждения.

4. Устройство по любому из пп. 2 или 3, в котором классификатор синтеза декодированного CELP возбуждения во временной области на одно из первого набора категорий возбуждения и второго набора категорий возбуждения использует
30 информацию о классификации, передаваемую от кодера декодеру CELP и извлекаемую в декодере CELP из декодируемого потока битов.

5. Устройство по любому из пп. 2 или 3, включающее в себя первый фильтр синтеза LP для выполнения синтеза модифицированного CELP возбуждения во временной области.

35 6. Устройство по п. 1, включающее в себя второй фильтр синтеза LP для выполнения синтеза декодированного CELP возбуждения во временной области.

7. Устройство по п. 5, включающее в себя фильтр устранения предыскажений и передискретизатор для формирования звукового сигнала из одного из синтеза декодированного CELP возбуждения во временной области и синтеза
40 модифицированного CELP возбуждения во временной области.

8. Устройство по п. 5, включающее в себя двухэтапный классификатор для выбора выходного синтеза как:

10 синтеза декодированного CELP возбуждения во временной области, когда синтез декодированного CELP возбуждения во временной области классифицируется во второй набор категорий возбуждения; и

синтеза модифицированного CELP возбуждения во временной области, когда синтез декодированного CELP возбуждения во временной области классифицируется в первый набор категорий возбуждения.

9. Устройство по любому из пп. 1-3, включающее в себя анализатор возбуждения в частотной области для определения того, содержит ли возбуждение в частотной области музыку.

10. Устройство по п. 9, в котором анализатор возбуждения в частотной области определяет, что возбуждение в частотной области содержит музыку, путем сравнения статистической девиации разностей спектральных энергий возбуждения в частотной области с порогом.

11. Устройство по любому из пп. 1-3, включающее в себя экстраполятор возбуждения для оценки возбуждения будущих кадров, посредством чего преобразование модифицированного возбуждения в частотной области в модифицированное CELP возбуждение во временной области выполняется без задержки.

12. Устройство по п. 11, содержащее блок объединения возбуждений во временной области прошлых кадров, текущих кадров и экстраполируемых будущих кадров, подаваемых в преобразователь декодированного CELP возбуждения во временной области, в возбуждение в частотной области.

13. Устройство по любому из пп. 1-3, в котором блок формирования маски формирует весовую маску, используя временное усреднение, или частотное усреднение, или комбинацию временного и частотного усреднения.

14. Устройство по любому из пп. 1-3, включающее в себя уменьшитель шума для оценки соотношения сигнал/шум в выбранной полосе декодированного CELP возбуждения во временной области и для выполнения уменьшения шума в частотной области на основе соотношения сигнал/шум.

15. Устройство по любому из пп. 1-3, в котором блок формирования маски включает в себя:

блок нормализации энергии спектра возбуждения в частотной области для получения масштабированного энергетического спектра;

блок усреднения масштабированного энергетического спектра вдоль частотной оси;

и

блок сглаживания усредненного энергетического спектра вдоль оси временной области для того, чтобы сгладить значения частотного спектра от кадра к кадру.

16. Устройство по п. 15, в котором упомянутый блок нормализации производит нормализованный энергетический спектр, применяет значение степени к нормализованному энергетическому спектру для получения масштабированного энергетического спектра и ограничивает значение масштабированного энергетического спектра до максимального предела.

17. Способ, реализованный в декодере линейного предсказания с кодовым возбуждением (CELP), для уменьшения шума квантования в звуковом сигнале, содержащемся в декодированном CELP возбуждении во временной области, подлежащем обработке фильтром синтеза линейного предсказания (LP) для выполнения его синтеза, при этом упомянутый способ включает в себя:

преобразование с использованием преобразователя из временной области в частотную область, декодированного CELP возбуждения во временной области до синтеза в возбуждение в частотной области;

формирование с использованием блока формирования маски, в ответ на возбуждение в частотной области, весовой маски для восстановления спектральной информации, потерянной в шуме квантования;

модификацию возбуждения в частотной области для увеличения динамики спектра путем применения весовой маски к возбуждению в частотной области; и

преобразование с использованием преобразователя из частотной области во временную область модифицированного возбуждения в частотной области в модифицированное CELP возбуждение во временной области, содержащее версию с уменьшенным шумом квантования звукового сигнала.

5 18. Способ по п. 17, включающий в себя:

обработку декодированного CELP возбуждения во временной области фильтром синтеза LP для выполнения синтеза декодированного CELP возбуждения во временной области; и

10 классификацию синтеза декодированного CELP возбуждения во временной области на одно из первого набора категорий возбуждения и второго набора категорий возбуждения;

при этом второй набор категорий возбуждения включает в себя категории НЕАКТИВНЫЙ или НЕВОКАЛИЗИРОВАННЫЙ; и

первый набор категорий возбуждения включает в себя категорию ДРУГОЙ.

15 19. Способ по п. 18, включающий в себя применение преобразования декодированного CELP возбуждения во временной области в возбуждение в частотной области к декодированному CELP возбуждению во временной области, когда синтез декодированного CELP возбуждения во временной области классифицирован в первый набор категорий возбуждения.

20 20. Способ по любому из пп. 18 или 19, включающий в себя использование информации о классификации, передаваемой от кодера декодеру CELP и извлекаемой в декодере CELP из декодируемого потока битов, для классификации синтеза декодированного CELP возбуждения во временной области на одно из первого набора категорий возбуждения и второго набора категорий возбуждения.

25 21. Способ по любому из пп. 18 или 19, включающий в себя выполнение синтеза модифицированного CELP возбуждения во временной области.

22. Способ по п. 21, включающий в себя генерирование звукового сигнала из одного из синтеза декодированного CELP возбуждения во временной области и синтеза модифицированного CELP возбуждения во временной области.

30 23. Способ по п. 21, включающий в себя выбор выходного синтеза как:

синтеза декодированного CELP возбуждения во временной области, когда синтез декодированного CELP возбуждения во временной области классифицируется во второй набор категорий возбуждения; и

35 синтеза модифицированного CELP возбуждения во временной области, когда синтез декодированного CELP возбуждения во временной области классифицируется в первый набор категорий возбуждения.

24. Способ по любому из пп. 17-19, включающий в себя анализ возбуждения в частотной области для определения того, содержит ли возбуждение в частотной области музыку.

40 25. Способ по п. 24, включающий в себя определение того, что возбуждение в частотной области содержит музыку, путем сравнения статистической девиации разностей спектральных энергий возбуждения в частотной области с порогом.

26. Способ по любому из пп. 17-19, включающий в себя оценку экстраполированного возбуждения будущих кадров, посредством чего преобразование модифицированного возбуждения в частотной области в модифицированное CELP возбуждение во временной области выполняется без задержки.

27. Способ по п. 26, включающий в себя объединение возбуждений во временной области прошлых кадров, текущих кадров и экстраполируемых будущих кадров для

преобразования в возбуждение в частотной области.

28. Способ по любому из пп. 17-19, в котором весовая маска формируется с использованием временного усреднения, или частотного усреднения, или комбинации временного и частотного усреднения.

5 29. Способ по любому из пп. 17-19, включающий в себя:

оценку соотношения сигнал/шум в выбранной полосе декодированного CELP возбуждения во временной области; и

выполнение уменьшения шума в частотной области на основе оцененного соотношения сигнал/шум.

10 30. Способ по любому из пп. 17-19, в котором формирование весовой маски включает в себя:

нормализацию энергии спектра возбуждения в частотной области для получения масштабированного энергетического спектра;

усреднение масштабированного энергетического спектра вдоль частотной оси; и

15 сглаживание усредненного энергетического спектра вдоль оси временной области для того, чтобы сгладить значения частотного спектра от кадра к кадру.

31. Способ по п. 30, в котором нормализация энергии спектра возбуждения в частотной области содержит производство нормализованного энергетического спектра, применение значения степени к нормализованному энергетическому спектру для
20 получения масштабированного энергетического спектра и ограничение значения масштабированного энергетического спектра до максимального предела.

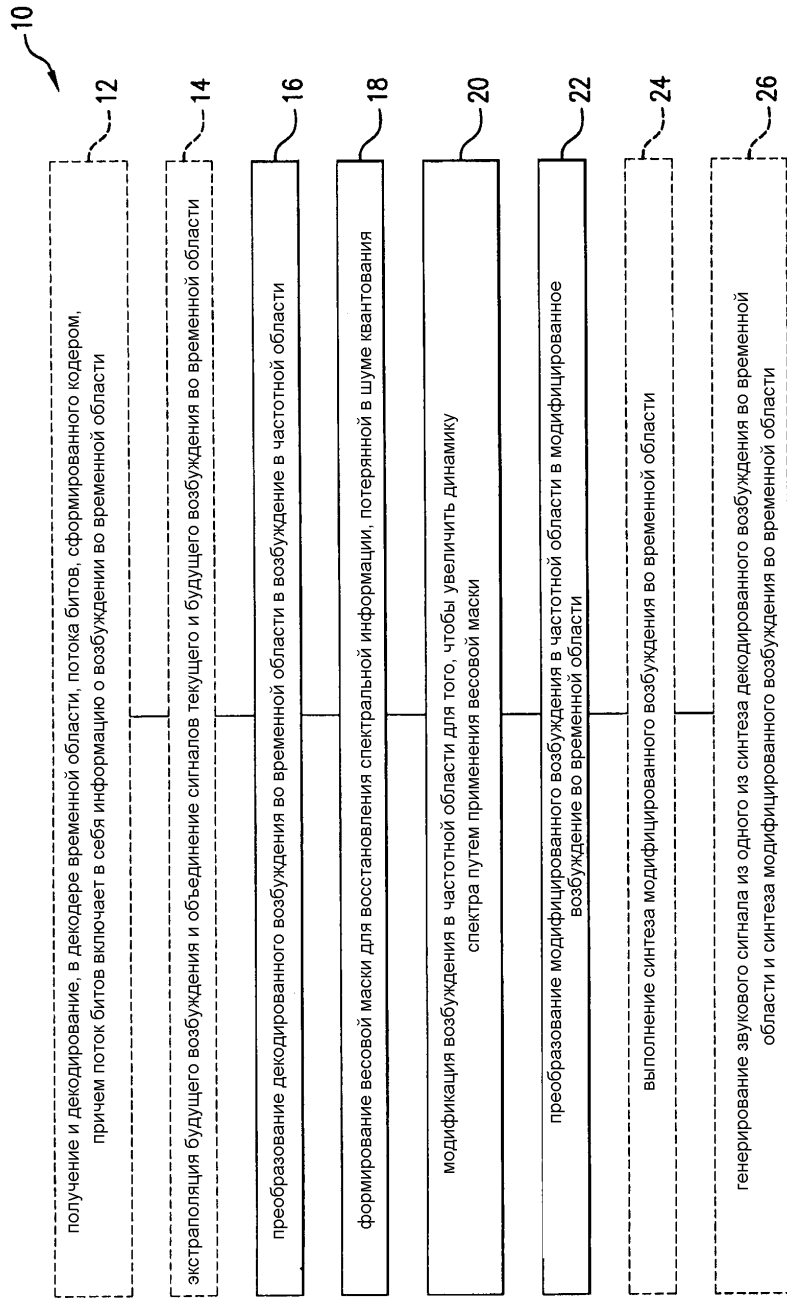
25

30

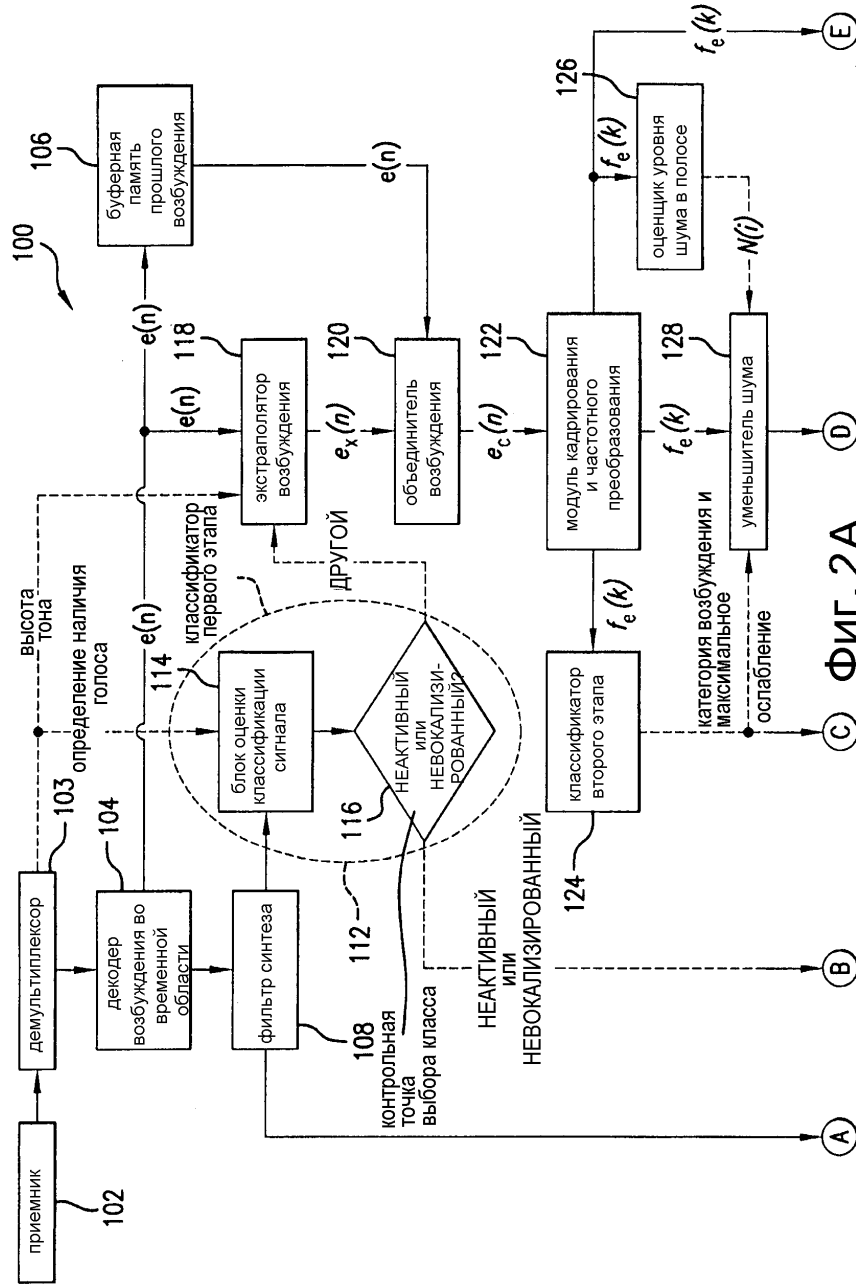
35

40

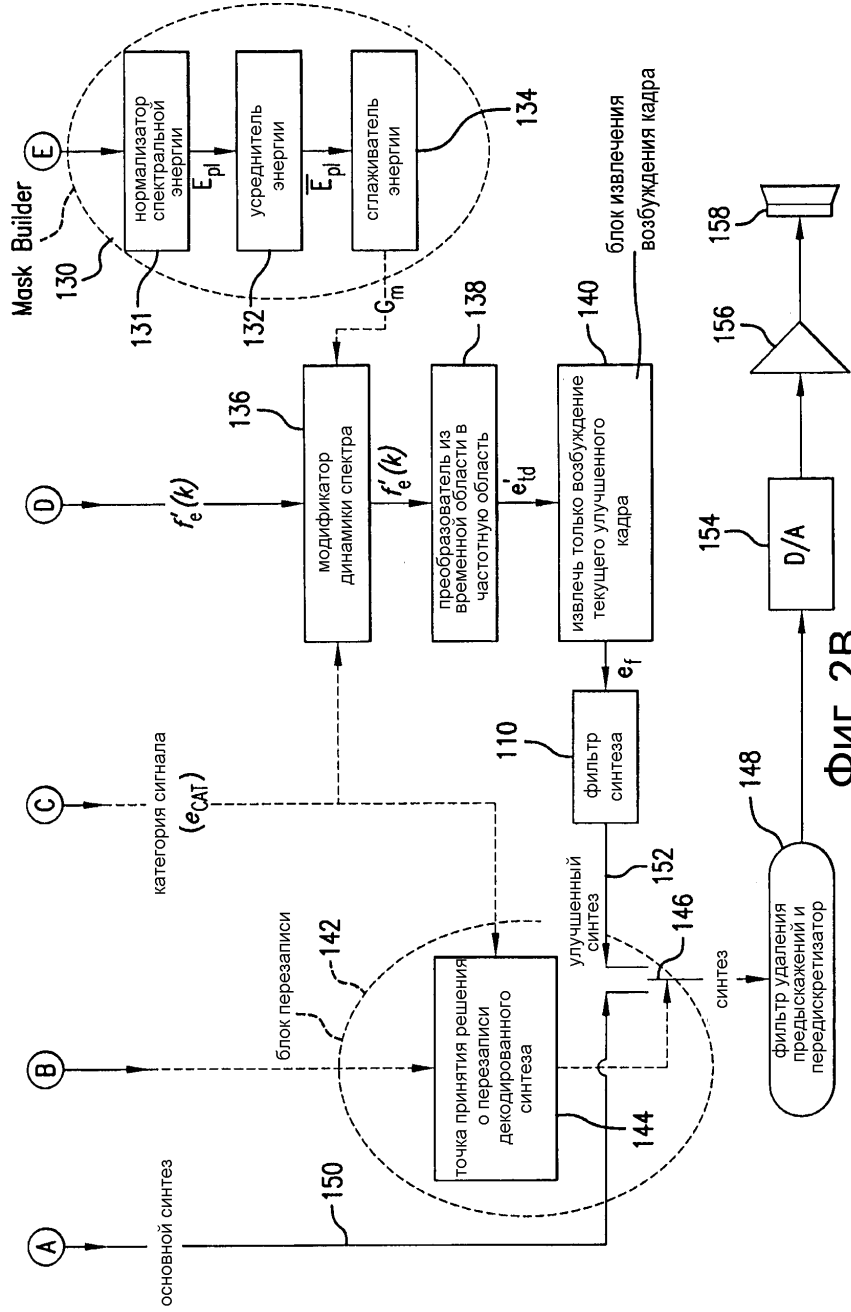
45



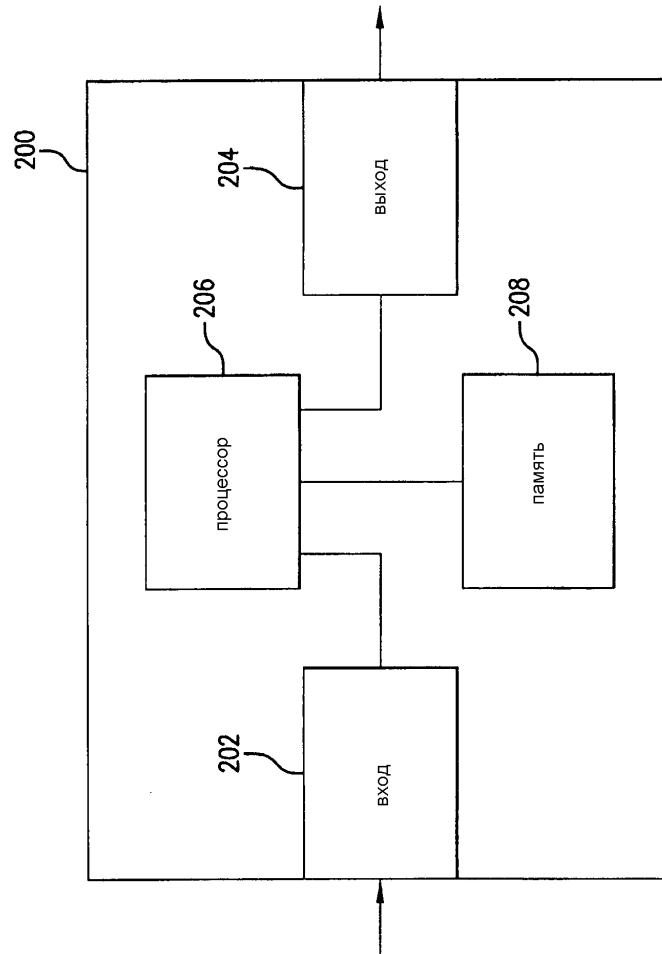
ФИГ. 1



ФИГ. 2А



ФИГ. 2В



ФИГ. 3