



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК

G10L 21/0388 (2006.01); G10L 19/265 (2006.01); G10L 19/002 (2006.01); G10L 25/69 (2006.01)

(21)(22) Заявка: 2017109527, 28.01.2014

(24) Дата начала отсчета срока действия патента:
28.01.2014

Дата регистрации:
26.12.2018

Приоритет(ы):

(30) Конвенционный приоритет:
29.01.2013 US 61/758,092

Номер и дата приоритета первоначальной заявки,
из которой данная заявка выделена:
2015136789 29.01.2013

(45) Опубликовано: 26.12.2018 Бюл. № 36

Адрес для переписки:
129090, Москва, ул. Большая Спасская, д. 25,
строение 3, ООО "Юридическая фирма
Городисский и Партнеры"

(72) Автор(ы):

НАГЕЛЬ Фредерик (DE),
ДИШ Сапа (DE),
НИДЕРМАЙЕР Андреас (DE)

(73) Патентообладатель(и):

ФРАУНХОФЕР-ГЕЗЕЛЛЬШАФТ ЦУР
ФЕРДЕРУНГ ДЕР АНГЕВАНДТЕН
ФОРШУНГ Е.Ф. (DE)

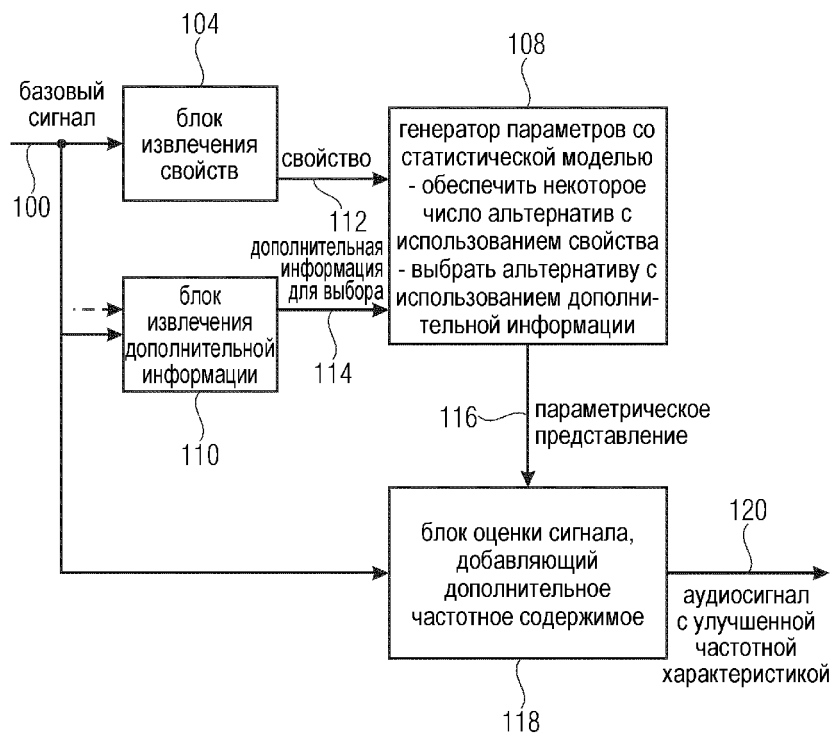
(56) Список документов, цитированных в отчете
о поиске: EP 2239732 A1, 13.10.2010. P
BAUER et al. "A STATISTICAL
FRAMEWORK FOR ARTIFICIAL
BANDWIDTH EXTENSION EXPLOITING
SPEECH WAVEFORM AND PHONETIC
TRANSCRIPTION", 01.04.2010. US 2007/
0255572 A1, 01.11.2007. WO 2011/047886 A1,
28.04.2011. US 2007/0094027 A1, 26.04.2007.
RU 2455710 C2, 10.07.2012. RU 2011101616 A,
27.07.2012.

(54) ДЕКОДЕР ДЛЯ ФОРМИРОВАНИЯ АУДИОСИГНАЛА С УЛУЧШЕННОЙ ЧАСТОТНОЙ ХАРАКТЕРИСТИКОЙ, СПОСОБ ДЕКОДИРОВАНИЯ, КОДЕР ДЛЯ ФОРМИРОВАНИЯ КОДИРОВАННОГО СИГНАЛА И СПОСОБ КОДИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ КОМПАКТНОЙ ДОПОЛНИТЕЛЬНОЙ ИНФОРМАЦИИ ДЛЯ ВЫБОРА

(57) Реферат:

Изобретение относится к средствам для кодирования аудиосигнала. Технический результат заключается в повышении эффективности кодирования аудиоданных. Извлекают свойство из базового сигнала. Извлекают дополнительную информацию для выбора, ассоциированную с базовым сигналом. Формируют параметрическое представление для оценки спектрального диапазона аудиосигнала с улучшенной частотной характеристикой, не определяемого базовым сигналом. Причем обеспечивают некоторое количество альтернативных параметрических представлений

в ответ на упомянутое свойство, и при этом выбирают одно из альтернативных параметрических представлений в качестве параметрического представления в ответ на дополнительную информацию для выбора. Оценивают аудиосигнал с улучшенной частотной характеристикой с использованием выбранного параметрического представления. При формировании параметрического представления принимают параметрическую информацию улучшения частотной характеристики, ассоциированной с базовым сигналом. 6 н. и 11 з.п. ф-лы, 16 ил.



ФИГ.1



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(51) Int. Cl.
G10L 21/0388 (2013.01)
G10L 25/69 (2013.01)

(12) **ABSTRACT OF INVENTION**

(52) CPC

G10L 21/0388 (2006.01); G10L 19/265 (2006.01); G10L 19/002 (2006.01); G10L 25/69 (2006.01)(21)(22) Application: **2017109527, 28.01.2014**(24) Effective date for property rights:
28.01.2014Registration date:
26.12.2018

Priority:

(30) Convention priority:
29.01.2013 US 61/758,092Number and date of priority of the initial application,
from which the given application is allocated:
2015136789 29.01.2013(45) Date of publication: **26.12.2018 Bull. № 36**

Mail address:

**129090, Moskva, ul. Bolshaya Spasskaya, d. 25,
stroenie 3, OOO "Yuridicheskaya firma
Gorodisskij i Partnery"**

(72) Inventor(s):

**NAGEL, Frederik (DE),
DISCH, Sascha (DE),
NIEDERMEIER, Andreas (DE)**

(73) Proprietor(s):

**FRAUNHOFER-GESELLSCHAFT ZUR
FOERDERUNG DER ANGEWANDTEN
FORSCHUNG E.V. (DE)**

(54) **DECODER FOR FORMATION OF AUDIO SIGNAL WITH IMPROVED FREQUENCY CHARACTERISTIC, DECODING METHOD, ENCODER FOR FORMATION OF ENCODED SIGNAL AND ENCODING METHOD USING COMPACT ADDITIONAL INFORMATION FOR SELECTION**

(57) Abstract:

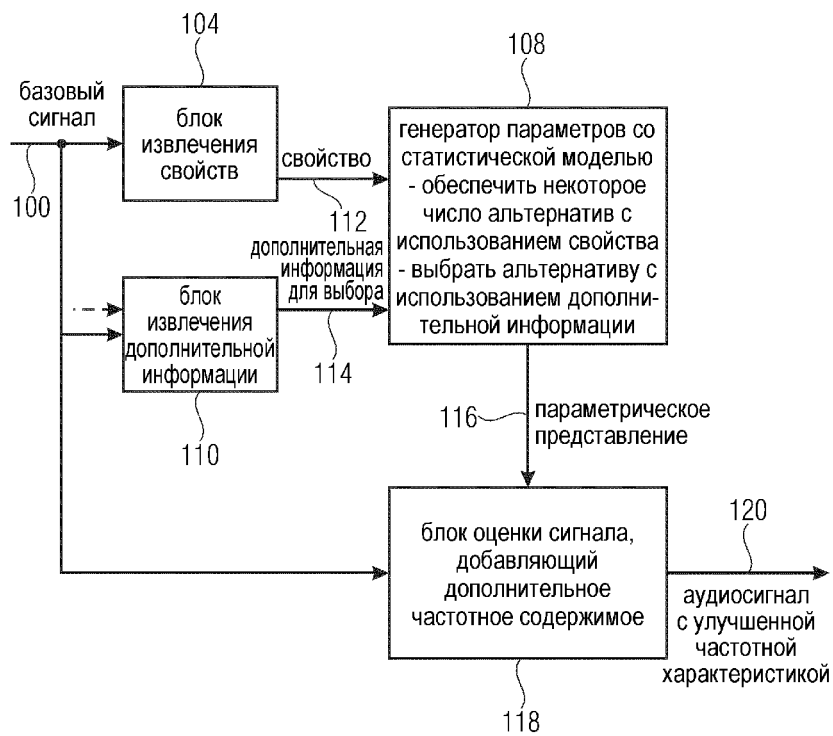
FIELD: data processing.

SUBSTANCE: invention relates to means for encoding an audio signal. Extract the property from the base signal. Extract additional information for the selection associated with the base signal. Form a parametric representation for estimating the spectral range of an audio signal with an improved frequency response that is not defined by the base signal. Provide a number of alternative parametric representations in response to the mentioned property, and choose one of the alternative parametric representations as a

parametric representation in response to additional information for selection. Audio signal with improved frequency response is evaluated using the selected parametric representation. When forming the parametric representation, the parametric information of improving the frequency response associated with the base signal is received.

EFFECT: technical result is to improve the efficiency of audio data encoding.

17 cl, 16 dwg



ФИГ.1

Настоящее изобретение относится к аудиокодированию и, в частности, к аудиокодированию в контексте улучшения частотной характеристики, т.е. того, что выходной сигнал декодера имеет большее число полос частот по сравнению с кодированным сигналом. Такие процедуры содержат расширение полосы частот, спектральную репликацию или интеллектуальное заполнение пробелов.

Современные системы кодирования голосовых данных способны улучшать широкополосное (WB) цифровое аудиосодержимое, то есть сигналы с частотами до 7-8 кГц, при скоростях передачи данных до минимум 6 кбит/с. Наиболее широко обсуждаемыми примерами являются рекомендации G.722.2 [1] ITU-T, а также более недавно разработанные G.718 [4, 10] и документ Unified Speech and Audio Coding (USAC) [8] MPEG-D. Оба из них, то есть G.722.2, также известный как AMR-WB, и G.718 используют технологии расширения полосы частот (BWE) между 6,4 и 7 кГц, чтобы позволить лежащему в основе базовому кодеру ACELP «сосредоточиться» на более значимых с точки зрения восприятия нижних частотах (в частности тех частотах, при которых система слуха человека является фазочувствительной), и таким образом достигают достаточного качества в особенности при очень низких скоростях передачи данных. В профиле расширенного высокоэффективного усовершенствованного аудиокодирования USAC (xHE-AAC) используется улучшенная репликация спектральной полосы (eSBR) для увеличения ширины полосы частот аудиоданных за пределы ширины полосы частот базового кодера, которая обычно составляет менее 6 кГц при 16 кбит/с. Существующие в настоящее время процессы BWE могут быть в общем разделены на два принципиальных подхода:

- «Слепое» или искусственное BWE, в котором высокочастотные (ВЧ) компоненты восстанавливают только из декодированного низкочастотного (НЧ) сигнала базового кодера, т.е. без необходимости передачи дополнительной информации из кодера. Эта схема используется в AMR-WB и G.718 при 16 кбит/с и ниже, а также в некоторых обратно совместимых средствах последующей обработки BWE, работающих с традиционными телефонными голосовыми данными с узкой полосой частот [5, 9, 12] (пример: Фиг. 15).

- Направленное BWE, которое отличается от «слепого» BWE тем, что некоторые из параметров, используемых для восстановления ВЧ содержимого передаются декодеру в качестве дополнительной информации, а не оцениваются из декодированного базового сигнала. AMR-WB, G.718, xHE-AAC, а также некоторые другие кодеки [2, 7, 11] используют данный подход, но не при очень низких скоростях передачи данных (Фиг. 16).

На Фиг. 15 проиллюстрировано такое «слепое» или искусственное расширение полосы частот, описанное в публикации Bernd Geiser, Peter Jax и Peter Vary: “ROBUST WIDEBAND ENHANCEMENT OF SPEECH BY COMBINED CODING AND ARTIFICIAL BANDWIDTH EXTENSION”, Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), 2005. Самостоятельный алгоритм расширения полосы частот, проиллюстрированный на Фиг. 15, содержит процедуру 1500 интерполяции, анализирующий фильтр 1600, расширение 1700 сигнала возбуждения, синтезирующий фильтр 1800, процедуру 1510 извлечения свойств, процедуру 1520 оценки огибающей и статистическую модель 1530. После интерполяции узкополосного сигнала в широкополосную частоту выборки вычисляют вектор свойств. Затем посредством предварительно обученной статистической скрытой марковской модели (СММ) определяют оценку для широкополосной спектральной огибающей с точки зрения коэффициентов линейного прогнозирования (LP). Эти широкополосные коэффициенты

используются для анализирующей фильтрации интерполированного узкополосного сигнала. После расширения итогового сигнала возбуждения применяют инверсный синтезирующий фильтр. Выбор расширения сигнала возбуждения, который не изменяет узкополосный сигнал, является прозрачным по отношению к компонентам

5 узкополосного сигнала.

На Фиг. 16 проиллюстрировано расширение полосы частот с дополнительной информацией, описанное в вышеупомянутой публикации, причем расширение полосы частот содержит телефонный фильтр 1620 полосы пропускания, блок 1610 извлечения дополнительной информации, (комбинированный) кодер 1630, декодер 1640 и блок

10 1650 расширения полосы частот. Эта система для широкополосного улучшения голосового сигнала полосы ошибок посредством комбинированного кодирования и расширения полосы частот проиллюстрирована на Фиг. 16. В передающем терминале анализируется спектральная огибающая полосы высоких частот широкополосного входного сигнала и определяется дополнительная информация. Итоговое сообщение

15 m кодируется либо отдельно, либо совместно с узкополосным голосовым сигналом. В приемнике дополнительная информация для декодера используется для поддержки оценки огибающей широкополосного сигнала в алгоритме расширения полосы частот. Сообщение m получают посредством нескольких процедур. Пространственное представление частот от 3,4 кГц до 7 кГц извлекают из широкополосного сигнала,

20 доступного только на передающей стороне.

Эту огибающую поддиапазона вычисляют путем избирательного линейного прогнозирования, т.е. вычисления спектра мощности широкополосного сигнала, за которым следует обратное дискретное преобразование Фурье (IDFT) компонентов его

верхней полосы частот и последующий рекурсивный алгоритм Левинсона-Дарбина 8

25 порядка. Итоговые коэффициенты LP для поддиапазона преобразуют в кепстральную область и наконец квантуют посредством векторного квантователя с помощью кодовой таблицы размера $M = 2^N$. Для длины кадра в 20 мс это приводит к скорости передачи данных дополнительной информации в 300 бит/с. Комбинированный оценочный подход

30 расширяет вычисление апостериорных вероятностей и повторно вводит зависимости от свойства узкополосного сигнала. Таким образом получается улучшенная форма маскировки ошибок, в которой для оценки ее параметров используется более одного источника информации.

При низких скоростях передачи данных, как правило ниже 10 кбит/с, в WB кодеках может наблюдаться определенная дилемма в отношении качества. С одной стороны,

35 такие скорости уже являются слишком низкими, чтобы оправдать передачу даже умеренных объемов данных BWE, исключая обычные системы направленного BWE с 1 кбит/с или более дополнительной информации. С другой стороны, оказывается, что допустимое «слепое» BWE звучит значительно хуже в случае по меньшей мере некоторых

40 видов голосового или музыкального материала вследствие невозможности надлежащего прогнозирования параметров из базового сигнала. Это в особенности верно для некоторых звуков речи, таких как фрикативные согласные с низкой корреляцией между ВЧ и НЧ. Поэтому желательно уменьшить скорость передачи дополнительной информации для схемы направленного BWE до уровня значительно менее 1 кбит/с, что

45 позволило бы использовать данную схему даже при кодировании с очень низкой скоростью передачи данных.

В последние годы документированы многоступенчатые подходы к BWE [1-10]. Все они в общем случае являются либо полностью «слепыми», либо полностью направленными в определенной рабочей точке, безотносительно к моментальным

характеристикам входного сигнала. Кроме того, многие системы «слепого» BWE [1, 3, 4, 5, 9, 10] оптимизированы в особенности для голосовых сигналов, а не для музыки, и поэтому могут обеспечивать неудовлетворительные результаты в случае с музыкой.

Наконец, большинство реализаций BWE являются относительно сложными в плане вычислений, поскольку используют преобразования Фурье, вычисления фильтров коэффициентов LP (LPC) или векторное квантование дополнительной информации (векторное кодирование с прогнозированием в USAC MPEG-D [8]). Это может быть недостатком при внедрении новой технологии кодирования на рынках мобильных телекоммуникаций при том, что большинство мобильных устройств обеспечивает очень ограниченную вычислительную мощность и емкость аккумуляторных батарей.

Подход, в котором «слепое» BWE расширено за счет малого объема дополнительной информации, представлен в [12] и проиллюстрирован на Фиг. 16. Однако дополнительная информация “m” ограничивается передачей спектральной огибающей диапазона частот с расширенной полосой частот.

Другая проблема процедуры, проиллюстрированной на Фиг. 16, заключается в очень сложном способе оценки огибающей с использованием, с одной стороны, низкочастотного свойства и, с другой стороны, дополнительной информации по огибающей. Оба вида входных данных, т.е. низкочастотное свойство и дополнительная высокочастотная огибающая, влияют на статистическую модель. Это приводит к сложной реализации на стороне декодера, что особенно проблематично для мобильных устройств ввиду повышенного потребления мощности. Кроме того, статистическую модель даже еще сложнее обновить ввиду того, что на нее влияют не только дополнительные данные высокочастотной огибающей.

Задача настоящего изобретения состоит в создании усовершенствованной концепции кодирования/декодирования аудиоданных.

Данная задача решается декодером по пункту 1 формулы изобретения, кодером по пункту 15 формулы изобретения, способом декодирования по пункту 20 формулы изобретения, способом кодирования по пункту 21 формулы изобретения, компьютерной программой по пункту 22 формулы изобретения или кодированным сигналом по пункту 23 формулы изобретения.

Настоящее изобретение основано на наблюдении о том, что для еще большего уменьшения объема дополнительной информации и, кроме того, для того, чтобы сделать весь кодер/декодер не чрезмерно сложным, параметрическое кодирование высокочастотной части согласно уровню техники должно быть заменено или по меньшей мере улучшено дополнительной информацией для выбора, фактически относящейся к статистической модели, используемой вместе с блоком извлечения свойств в декодере с улучшением частотной характеристики. Ввиду того, что извлечение свойств в сочетании со статистической моделью обеспечивает альтернативные параметрические представления, которые имеют неопределенности конкретно для определенных частей голосовых данных, было обнаружено, что фактическое управление статистической моделью в генераторе параметров на стороне декодера в отношении того, какая из имеющихся альтернатив будет наилучшей, превосходит фактическое параметрическое кодирование определенной характеристики сигнала конкретно в применениях с очень низкой скоростью передачи данных, при которых дополнительная информация для расширения полосы частот является ограниченной.

Таким образом улучшается «слепое» BWE, которое использует модель источника для кодированного сигнала, путем расширения с небольшим объемом добавленной дополнительной информации, в частности если сам сигнал не допускает реконструкцию

высокочастотного (ВЧ) содержимого на приемлемом уровне воспринимаемого качества. Таким образом, данная процедура объединяет параметры модели источника, которые формируются из кодированного содержимого от базового кодера, посредством дополнительной информации. Это полезно, в частности, для повышения

5 воспринимаемого качества звуков, которые трудно кодировать в такой модели источника. Такие звуки обычно демонстрируют низкую корреляцию между ВЧ и НЧ содержимым.

Настоящее изобретение направлено на решение проблем традиционного BWE при кодировании аудиосигнала с очень низкой скоростью передачи данных и на устранение

10 недостатков существующих, известных из уровня техники технологий BWE. Решение вышеописанной дилеммы в отношении качества обеспечено путем предложения в минимальной степени направленного BWE в качестве адаптируемого по отношению к сигналу сочетания «слепого» и направленного BWE. BWE согласно изобретению добавляет к сигналу некоторый небольшой объем дополнительной информации, который

15 позволяет дополнительно различать кодированные сигналы, которые в ином случае являются проблематичными. При кодировании голосовых данных это применимо, в частности, к сибилантам или фрикативным звукам.

Было обнаружено, что в WB кодеках спектральная огибающая ВЧ области выше области базового кодера представляет наиболее важные данные, необходимые для

20 выполнения BWE с приемлемым воспринимаемым качеством. Все прочие параметры, такие как спектральная огибающая тонкой структуры и временная огибающая, зачастую могут довольно точно выводиться из декодированного базового сигнала или обладают невысокой важностью в плане восприятия. Однако для фрикативных звуков часто отсутствует надлежащее воспроизведение в сигнале BWE. Таким образом,

25 дополнительная информация может включать в себя добавочную информацию, различающую различные сибиланты или фрикативные звуки, такие как «ф», «с», «ч» и «ш».

Другая проблематичная акустическая информация для расширения полосы частот возникает, когда встречаются взрывные звуки или аффрикаты, такие как «т» или «ч».

Настоящее изобретение позволяет использовать лишь эту дополнительную

30 информацию и фактически передавать эту дополнительную информацию, когда это необходимо, и не передавать эту дополнительную информацию, когда в статистической модели не ожидается неопределенность.

Кроме того, в предпочтительных вариантах выполнения настоящего изобретения

35 используется лишь малый объем дополнительной информации, такой как три или менее бита на кадр, комбинированное обнаружение голосовой активности/обнаружение голосовых/неголосовых данных для управления блоком оценки сигнала, различные статистические модели, определяемые классификатором сигнала или альтернативными параметрическими представлениями, относящимися не только к оценке огибающей,

40 но также относящимися к другим инструментам расширения полосы частот или улучшения параметров расширения полосы частот или добавления новых параметров к уже имеющимся и фактически передаваемым параметрам расширения полосы частот.

Предпочтительные варианты выполнения настоящего изобретения описаны ниже в контексте сопровождающих чертежей и также представлены в зависимых пунктах

45 формулы изобретения.

Фиг. 1 иллюстрирует декодер для формирования аудиосигнала с улучшенной частотной характеристикой;

Фиг. 2 иллюстрирует предпочтительную реализацию в контексте блока извлечения

дополнительной информации по Фиг. 1;

Фиг. 3 иллюстрирует таблицу, соотносящую число битов дополнительной информации для выбора с числом альтернативных параметрических представлений;

Фиг. 4 иллюстрирует предпочтительную процедуру, выполняемую в генераторе параметров;

Фиг. 5 иллюстрирует предпочтительную реализацию блока оценки сигнала, управляемого детектором голосовой активности или детектором голосовых/неголосовых данных;

Фиг. 6 иллюстрирует предпочтительную реализацию генератора параметров, управляемого классификатором сигнала;

Фиг. 7 иллюстрирует пример результата для статистической модели и соответствующую дополнительную информацию для выбора;

Фиг. 8 иллюстрирует примерный кодированный сигнал, содержащий кодированный базовый сигнал и соответствующую дополнительную информацию;

Фиг. 9 иллюстрирует схему обработки сигнала расширения полосы частот для улучшения оценки огибающей;

Фиг. 10 иллюстрирует другую реализацию декодера в контексте процедур репликации спектральной полосы;

Фиг. 11 иллюстрирует другой вариант выполнения декодера в контексте дополнительно передаваемой дополнительной информации;

Фиг. 12 иллюстрирует вариант выполнения кодера для формирования кодированного сигнала;

Фиг. 13 иллюстрирует реализацию генератора дополнительной информации для выбора по Фиг. 12;

Фиг. 14 иллюстрирует другую реализацию генератора дополнительной информации для выбора по Фиг. 12;

Фиг. 15 иллюстрирует самостоятельный алгоритм расширения полосы частот из уровня техники; и

Фиг. 16 иллюстрирует общий вид передающей системы с добавочным сообщением.

Фиг. 1 иллюстрирует декодер для формирования аудиосигнала 120 с улучшенной частотной характеристикой. Декодер содержит блок 104 извлечения свойств для извлечения (по меньшей мере) свойства из базового сигнала 100. В общем случае блок извлечения свойств может извлекать одно свойство или множество свойств, т.е. два или более свойств, и даже предпочтительно, чтобы блок извлечения свойств извлекал множество свойств. Это применимо не только к блоку извлечения свойств в декодере, но и к блоку извлечения свойств в кодере.

Кроме того, предусмотрен блок 110 извлечения дополнительной информации для извлечения дополнительной информации 114 для выбора, ассоциированной с базовым сигналом 100. Кроме того, генератор 108 параметров соединен с блоком 104 извлечения свойств посредством линии 112 передачи свойств и с блоком 110 извлечения дополнительной информации посредством дополнительной информации 114 для выбора. Генератор 108 параметров выполнен с возможностью формирования параметрического представления для оценки спектрального диапазона аудиосигнала с улучшенной частотной характеристикой, не определяемого базовым сигналом. Генератор 108 параметров выполнен с возможностью обеспечения некоторого числа альтернативных параметрических представлений в ответ на свойства 112 и выбора одного из альтернативных параметрических представлений в качестве упомянутого параметрического представления в ответ на дополнительную информацию 114 для

выбора. Кроме того, декодер содержит блок 118 оценки сигнала для оценки аудиосигнала с улучшенной частотной характеристикой с использованием параметрического представления, выбранного блоком выбора, т.е. параметрического представления 116.

5 В частности, блок 104 извлечения свойств может быть реализован с возможностью извлечения свойств из декодированного базового сигнала, как показано на Фиг. 2. Тогда интерфейс 210 ввода выполнен с возможностью приема кодированного входного сигнала 200. Этот кодированный входной сигнал 200 вводится в интерфейс 210, и затем интерфейс 210 отделяет дополнительную информацию для выбора от кодированного базового сигнала. Таким образом, интерфейс 210 ввода действует как блок 110
10 извлечения дополнительной информации по Фиг. 1. Кодированный базовый сигнал 201, выдаваемый интерфейсом 210 ввода, затем вводится в базовый декодер 124 для обеспечения декодированного базового сигнала, который может быть базовым сигналом 100.

15 Однако, в качестве альтернативы, блок извлечения свойств также может действовать или извлекать свойство из кодированного базового сигнала. Обычно кодированный базовый сигнал содержит представление коэффициентов масштабирования для полос частот или любое другое представление аудиоинформации. В зависимости от вида извлечения свойств кодированное представление аудиосигнала представляет
20 декодированный базовый сигнал и поэтому свойства могут быть извлечены. В качестве альтернативы или дополнения, свойство может быть извлечено не только из полностью декодированного базового сигнала, но также из частично декодированного базового сигнала. При кодировании в частотной области кодированный сигнал представляет представление в частотной области, содержащее последовательность спектральных
25 кадров. Таким образом, кодированный базовый сигнал может быть лишь частично декодирован для получения декодированного представления последовательности спектральных кадров перед выполнением собственно спектрально-временного преобразования. Таким образом, блок 104 извлечения свойств может извлекать свойства либо из кодированного базового сигнала, либо из частично декодированного базового
30 сигнала или полностью декодированного базового сигнала. Блок 104 извлечения свойств может быть реализован по отношению к извлекаемым им свойствам так, как это известно в данной области техники и, например, блок извлечения свойств может быть реализован так, как это делается в технологиях создания «цифровых отпечатков» аудиосигналов или идентификации (ID) аудиосигналов.

35 Предпочтительно дополнительная информация 114 для выбора содержит число N битов на кадр базового сигнала. Фиг. 3. иллюстрирует таблицу для различных альтернатив. Число битов для дополнительной информации для выбора либо является фиксированным, либо выбирается в зависимости от числа альтернативных параметрических представлений, обеспечиваемых статистической моделью в ответ на
40 извлеченное свойство. Один бит дополнительной информации для выбора достаточен, когда только два альтернативных параметрических представления обеспечены статистической моделью в ответ на упомянутое свойство. Когда статистическая модель обеспечивает максимально четыре альтернативы, для дополнительной информации для выбора необходимы два бита. Три бита дополнительной информации для выбора
45 допускают максимально восемь одновременных альтернативных параметрических представлений. Четыре бита дополнительной информации для выбора фактически допускают 16 альтернативных параметрических представлений, и пять битов дополнительной информации для выбора допускают 32 одновременных альтернативных

параметрических представления. Предпочтительно использовать три или менее трех битов дополнительной информации для выбора на кадр, что приводит к скорости передачи дополнительной информации в 150 битов в секунду, когда секунда разделена на 50 кадров. Эта скорость передачи дополнительной информации может даже быть снижена ввиду того, что дополнительная информация для выбора необходима только тогда, когда статистическая модель фактически обеспечивает альтернативные параметрические представления. Таким образом, когда статистическая модель обеспечивает только одну альтернативу для свойства, бит дополнительной информации для выбора вовсе не нужен. С другой стороны, когда статистическая модель обеспечивает только четыре альтернативных параметрических представления, необходимы только два бита, а не три бита дополнительной информации для выбора. Таким образом, в типичных случаях скорость передачи добавочной дополнительной информации может быть снижена даже менее 150 битов в секунду.

Кроме того генератор параметров выполнен с возможностью обеспечения не более чем количества альтернативных параметрических представлений, равного 2^N . С другой стороны, когда генератор 108 параметров обеспечивает, например, только пять альтернативных параметрических представлений, тем не менее требуется три бита дополнительной информации для выбора.

Фиг. 4 иллюстрирует предпочтительную реализацию генератора 108 параметров. В частности, генератор 108 параметров выполнен таким образом, что свойство 112 по Фиг. 1 вводится в статистическую модель, как обозначено на этапе 400. Затем, как обозначено на этапе 402 модель обеспечивает множество альтернативных параметрических представлений.

Кроме того, генератор 108 параметров выполнен с возможностью получения дополнительной информации 114 для выбора из блока извлечения дополнительной информации, как обозначено на этапе 404. Затем на этапе 406 выбирают конкретное альтернативное параметрическое представление с использованием дополнительной информации 114 для выбора. Наконец на этапе 408 выбранное альтернативное параметрическое представление выдают в блок 118 оценки сигнала.

Предпочтительно генератор 108 параметров выполнен с возможностью использования при выборе одного из альтернативных параметрических представлений предварительно заданного порядка альтернативных параметрических представлений или, в качестве альтернативы, порядка альтернатив по сигналу кодера. Для этой цели обратимся к Фиг. 7. Фиг. 7 иллюстрирует результат обеспечения статистической моделью четырех альтернативных параметрических представлений 702, 704, 706, 708. Также проиллюстрирован соответствующий код дополнительной информации для выбора. Альтернатива 702 соответствует битовой структуре 712. Альтернатива 704 соответствует битовой структуре 714. Альтернатива 706 соответствует битовой структуре 716, и альтернатива 708 соответствует битовой структуре 718. Таким образом, когда генератор 108 параметров или, например, этап 402 получает четыре альтернативы 702-708 в порядке, проиллюстрированном на Фиг. 7, дополнительная информация для выбора, имеющая битовую структуру 716, будет уникальным образом идентифицировать альтернативное параметрическое представление 3 (ссылочная позиция 706), и тогда генератор 108 параметров выберет эту третью альтернативу. Однако когда битовая структура дополнительной информации для выбора является битовой структурой 712, будет выбрана первая альтернатива 702.

Таким образом, предварительно заданный порядок альтернативных параметрических представлений может быть порядком, в котором статистическая модель фактически

выдает альтернативы в ответ на извлеченное свойство. В качестве альтернативы, если отдельная альтернатива имеет различные ассоциированные вероятности, которые, однако, весьма близки друг к другу, предварительно заданный порядок может состоять в том, что параметрическое представление с наибольшей вероятностью следует первым и так далее. В качестве альтернативы, порядок может сигнализироваться, например, одним битом, но для того, чтобы сэкономить даже этот бит, предпочтительным является предварительно заданный порядок.

Далее обратимся к Фиг. 9-11.

В варианте выполнения по Фиг. 9 изобретение в особенности приспособлено для голосовых сигналов, поскольку для извлечения параметров используется специализированная голосовая модель источника. Однако изобретение не ограничено кодированием голосовых данных. В различных вариантах выполнения могут использоваться также и другие модели источника.

В частности, дополнительная информация 114 для выбора также называется «информацией о фрикативных звуках», поскольку такая дополнительная информация для выбора различает проблематичные сибиланты и фрикативные звуки, такие как «ф», «с» или «ш». Таким образом, дополнительная информация для выбора обеспечивает ясное определение одной из трех проблематичных альтернатив, которые, например, обеспечены статистической моделью 904 в процессе оценки 902 огибающей, причем оба действия выполняются в генераторе 108 параметров. Итогом оценки огибающей является параметрическое представление спектральной огибающей для спектральных участков, не включенных в базовый сигнал.

Таким образом, блок 104 может соответствовать блоку 1510 по Фиг. 15. Кроме того, блок 1530 по Фиг. 15 может соответствовать статистической модели 904 по Фиг. 9.

Кроме того, предпочтительно, чтобы блок 118 оценки сигнала содержал анализирующий фильтр 910, блок 912 расширения сигнала возбуждения и синтезирующий фильтр 914. Таким образом, блоки 910, 912, 914 могут соответствовать блокам 1600, 1700 и 1800 по Фиг. 15. В частности, анализирующий фильтр 910 представляет собой анализирующий фильтр LPC. Блок 902 оценки огибающей управляет коэффициентами фильтра для анализирующего фильтра 910 таким образом, что результат блока 910 представляет собой сигнал возбуждения фильтра. Этот сигнал возбуждения фильтра расширен по отношению к частоте для получения сигнала возбуждения на выходе блока 912, который не только имеет частотный диапазон декодера 124 для выходного сигнала, но также имеет частотный или спектральный диапазон, не определяемый базовым кодером и/или превышающий спектральный диапазон базового сигнала. Таким образом, аудиосигнал 909 на выходе декодера подвергается повышающей дискретизации и интерполируется интерполятором 900 и затем интерполированный сигнал подвергается обработке в блоке 118 оценки сигнала. Таким образом, интерполятор 900 по Фиг. 9 может соответствовать интерполятору 1500 по Фиг. 15. Однако предпочтительно в отличие от Фиг. 15 извлечение 104 свойств выполняется с использованием не интерполированного сигнала, а неинтерполированного сигнала, как показано на Фиг. 15. Это полезно по той причине, что блок 104 извлечения свойств работает более эффективно ввиду того, что неинтерполированный аудиосигнал 909 имеет меньшее число выборок по сравнению с определенным временным участком аудиосигнала, сравниваемого с подвергнутым повышающей дискретизации и интерполированным сигналом на выходе блока 900.

Фиг. 10 иллюстрирует другой вариант выполнения настоящего изобретения. В отличие от Фиг. 9, Фиг. 10 содержит статистическую модель 904, которая не только

обеспечивает оценку огибающей, как на Фиг. 9, но также обеспечивает дополнительные параметрические представления, содержащие информацию для формирования отсутствующих тонов 1080 или информацию для инверсной фильтрации 1040 или информацию для маскирующего шума (шумовой завесы) 1020, который необходимо
 5 добавить. Блоки 1020, 1040, процедуры формирования 1060 спектральной огибающей и отсутствующих тонов 1080 описаны в стандарте MPEG-4 в контексте HE-AAC (высокоэффективного усовершенствованного аудиокодирования).

Таким образом, другие сигналы, отличные от голосовых данных, также могут кодироваться, как проиллюстрировано на Фиг. 10. В таком случае может быть не
 10 достаточно кодировать только спектральную огибающую 1060, но также и другую дополнительную информацию, такую как тональность (1040), уровень шума (1020) или отсутствующие синусоиды (1080), как это делается в технологии репликации спектральной полосы (SBR), проиллюстрированной в [6].

Другой вариант выполнения проиллюстрирован на Фиг. 11, на которой
 15 дополнительная информация 114, т.е. дополнительная информация для выбора, используется в дополнение к дополнительной информации SBR, проиллюстрированной в блоке 1100. Таким образом, дополнительная информация для выбора, содержащая, например, информацию относительно обнаруженных звуков речи, добавляется к уже имеющейся дополнительной информации 1100 SBR. Это помогает более точно
 20 регенерировать высокочастотное содержимое для звуков голоса, таких как сibilанты, а также фрикативные, взрывные, или таких как гласные звуки. Таким образом, процедура, проиллюстрированная на Фиг. 11, имеет преимущество, состоящее в том, что дополнительно передаваемая дополнительная информация 114 для выбора поддерживает классификацию (фонем) на стороне декодера для обеспечения адаптации
 25 параметров SBR или BWE (расширения полосы частот) на стороне декодера. Таким образом, в отличие от Фиг. 10 вариант выполнения по Фиг. 11 обеспечивает уже имеющуюся дополнительную информацию SBR в качестве дополнения к дополнительной информации для выбора.

Фиг. 8 иллюстрирует примерное представление кодированного входного сигнала.
 30 Кодированный входной сигнал состоит из последовательных кадров 800, 806, 812. Каждый кадр имеет кодированный базовый сигнал. В качестве примера, кадр 800 имеет голосовые данные в качестве кодированного базового сигнала. Кадр 806 имеет музыку в качестве кодированного базового сигнала, а кадр 812 опять же имеет голосовые данные в качестве кодированного базового сигнала. В качестве примера, кадр 800
 35 имеет в качестве дополнительной информации только дополнительную информацию для выбора, но не имеет дополнительной информации SBR. Таким образом, кадр 800 соответствует Фиг. 9 или Фиг. 10. В качестве примера, кадр 806 содержит информацию SBR, но не содержит какой-либо дополнительной информации для выбора. Кроме того, кадр 812 содержит кодированный голосовой сигнал и, в отличие от кадра 800, кадр 812
 40 не содержит какую-либо дополнительную информацию для выбора. Это вызвано тем, что дополнительная информация для выбора не нужна, поскольку на стороне кодера не обнаружены какие-либо неопределенности в процессе извлечения свойств/статистической модели.

Далее описана Фиг. 5. Применяется детектор голосовой активности или детектор
 45 500 голосовых/неголосовых данных, работающий с базовым сигналом для определения того, следует ли применять технологию улучшения полосы частот или частотной характеристики согласно изобретению или другую технологию расширения полосы частот. Таким образом, когда детектор голосовой активности или детектор голосовых/

неголосовых данных обнаруживает голос или речь, используется первая технология расширения полосы частот BWEXT.1, проиллюстрированная позицией 511, которая работает, например, как описано в отношении Фиг. 1, 9, 10, 11. Таким образом, переключатели 502, 504 устанавливаются так, что принимаются параметры от генератора параметров со входа 512 и переключатель 504 соединяет эти параметры с блоком 511. Однако когда детектор 500 обнаруживает ситуацию, которая не указывает на какие-либо голосовые сигналы, но указывает, например, на музыкальные сигналы, параметры 514 расширения полосы частот из битового потока вводятся предпочтительно в процедуру 513 другой технологии расширения полосы частот. Таким образом, детектор 500 обнаруживает то, следует ли применять технологию 511 расширения полосы частот согласно изобретению. Для неголосовых сигналов кодер может переключаться на другие технологии расширения полосы частот, проиллюстрированные блоком 513, такие как те, что упомянуты [6, 8]. Таким образом, блок 118 оценки сигнала по Фиг. 5 выполнен с возможностью переключения на другую процедуру расширения полосы частот и/или использования других параметров, извлекаемых из кодированного сигнала, когда детектор 500 обнаруживает неголосовую активность или неголосовой сигнал. Для этой другой технологии 513 расширения полосы частот дополнительная информация для выбора предпочтительно отсутствует в битовом потоке и также не используется, что обозначено на Фиг. 5 путем переключения переключателя 502 на вход 514.

Фиг. 6 иллюстрирует другую реализацию генератора 108 параметров. Генератор 108 параметров предпочтительно имеет множество статистических моделей, таких как первая статистическая модель 600 и вторая статистическая модель 602. Кроме того, предусмотрен блок 604 выбора, управляемый дополнительной информацией для выбора для обеспечения правильного альтернативного параметрического представления. То, какая статистическая модель является активной, регулируется дополнительным классификатором 606 сигнала, принимающим на входе базовый сигнал, т.е. тот же сигнал, что вводится в блок 104 извлечения свойств. Таким образом, статистическая модель по Фиг. 10 или по любым другим чертежам может быть различной в зависимости от кодированного содержимого. Для голосовых данных применяется статистическая модель, которая представляет модель источника для формирования голосовых данных, в то время как для других сигналов, таких как музыкальные сигналы, согласно, например, классификации посредством классификатора 606 сигнала, используется другая модель, которая обучена на основании большого набора музыкальных данных. Кроме того, различные статистические модели полезны для различных языков и т.д.

Как описано выше, Фиг. 7 иллюстрирует множество альтернатив, получаемых статистической моделью, такой как статистическая модель 600. Таким образом, выходные данные блока 600 существуют, например, для различных альтернатив, как показано параллельной линией 605. Таким же образом вторая статистическая модель 602 может также выдавать множество альтернатив, таких как альтернативы, показанные линией 606. В зависимости от конкретной статистической модели предпочтительно, чтобы выводились только те альтернативы, которые обладают довольно высокой вероятностью по отношению к блоку 104 извлечения свойств. Таким образом, в ответ на упомянутое свойство статистическая модель обеспечивает множество альтернативных параметрических представлений, причем каждое альтернативное параметрическое представление обладает вероятностью, идентичной вероятностям других различных альтернативных параметрических представлений или отличной от вероятностей других параметрических представлений менее чем на 10 %. Таким образом, в варианте

выполнения выдается только параметрическое представление, обладающее наибольшей вероятностью, и некоторое число других альтернативных параметрических представлений, которые обладают вероятностью, лишь на 10% меньшей, чем вероятность наиболее подходящей альтернативы.

5 Фиг. 12 иллюстрирует кодер для формирования кодированного сигнала 1212. Кодер содержит базовый кодер 1200 для кодирования исходного сигнала 1206 для получения кодированного базового аудиосигнала 1208, имеющего информацию о меньшем числе полос частот по сравнению с исходным сигналом 1206. Кроме того, предусмотрен генератор 1202 дополнительной информации для выбора для формирования
10 дополнительной информации 1210 для выбора (SSI - дополнительная информация для выбора). Дополнительная информация 1210 для выбора указывает на определенное альтернативное параметрическое представление, обеспеченное статистической моделью в ответ на свойство, извлеченное из исходного сигнала 1206 или из кодированного аудиосигнала 1208 или из декодированной версии кодированного аудиосигнала. Кроме
15 того, кодер содержит интерфейс 1204 вывода для вывода кодированного сигнала 1212. Кодированный сигнал 1212 содержит кодированный аудиосигнал 1208 и дополнительную информацию 1210 для выбора. Предпочтительно генератор 1202 дополнительной информации для выбора реализован как показано на Фиг. 13. Для этой цели генератор 1202 дополнительной информации для выбора содержит базовый
20 декодер 1300. Предусмотрен блок 1302 извлечения свойств, который работает с декодированным базовым сигналом, выдаваемым блоком 1300. Свойство вводится в процессор 1304 статистических моделей для формирования некоторого числа альтернативных параметрических представлений для оценки спектрального диапазона сигнала с улучшенной частотной характеристикой, не определяемого декодированным
25 базовым сигналом, выдаваемым блоком 1300. Все эти альтернативные параметрические представления 1305 вводятся в блок 1306 оценки сигнала для оценки аудиосигнала 1307 с улучшенной частотной характеристикой. Затем эти оцениваемые аудиосигналы 1307 с улучшенной частотной характеристикой вводятся в блок 1308 сравнения для сравнения аудиосигналов 1307 с улучшенной частотной характеристикой с исходным сигналом
30 по Фиг. 12. Генератор 1202 дополнительной информации для выбора дополнительно выполнен с возможностью установления дополнительной информации 1210 для выбора таким образом, что дополнительная информация для выбора уникальным образом идентифицирует альтернативное параметрическое представление, обеспечивающее аудиосигнал с улучшенной частотной характеристикой, который наилучшим образом
35 соответствует исходному сигналу согласно критерию оптимизации. Критерий оптимизации может представлять собой критерий, основанный на MMSE (минимальной среднеквадратической ошибке), критерий, минимизирующий разность между выборками, или предпочтительно психоакустический критерий, минимизирующий воспринимаемое искажение или любой другой критерий оптимизации, известный специалистам в данной
40 области техники.

В то время как Фиг. 13 иллюстрирует процедуру с замкнутым циклом или процедуру «анализа через синтез», Фиг. 14 иллюстрирует альтернативную реализацию генератора 1202 дополнительной информации для выбора, в большей степени подобную процедуре с незамкнутым циклом. В варианте выполнения по Фиг. 14 исходный сигнал 1206
45 содержит ассоциированную метаинформацию для генератора 1202 дополнительной информации для выбора, описывающую последовательность акустической информации (например, аннотаций) для последовательности выборок исходного аудиосигнала. В этом варианте выполнения генератор 1202 дополнительной информации для выбора

содержит блок 1400 извлечения метаданных для извлечения последовательности метаданных и, кроме того, блок интерпретации метаданных, обычно обладающий информацией о статистической модели, используемой на стороне декодера для интерпретации последовательности метаданных в последовательность
 5 дополнительной информации 1210 для выбора, ассоциированной с исходным аудиосигналом. Метаданные, извлеченные блоком 1400 извлечения метаданных, отбрасываются в кодере и не передаются в кодированном сигнале 1212. Вместо этого в кодированном сигнале передается дополнительная информация 1210 для выбора
 10 вместе с кодированным аудиосигналом 1208, сформированным базовым кодером, которая имеет другое частотное содержимое и обычно меньшее частотное содержимое по сравнению с формируемым в итоге декодированным сигналом или по сравнению с исходным сигналом 1206.

Дополнительная информация 1210 для выбора, сформированная генератором 1202 дополнительной информации для выбора, может иметь любую из характеристик,
 15 описанных в контексте предыдущих чертежей.

Хотя настоящее изобретение было описано в контексте блок-схем, в которых блоки представляют фактические или логические компоненты аппаратного обеспечения, настоящее изобретение может также быть реализовано посредством способа, реализуемого компьютером. В последнем случае блоки представляют соответствующие
 20 этапы способа, причем эти этапы обозначают функции, выполняемые соответствующими логическими или физическими блоками аппаратного обеспечения.

Хотя некоторые аспекты описаны в контексте устройства, ясно, что эти аспекты также представляют собой описание соответствующего способа, причем блок или устройство соответствуют этапу способа или признаку этапа способа. Аналогичным
 25 образом, аспекты, описанные в контексте этапа способа, также представляют собой описание соответствующего блока или элемента или признака соответствующего устройства. Некоторые или все этапы способа могут быть выполнены посредством (или с использованием) устройства аппаратного обеспечения, такого как, например, микропроцессор, программируемый компьютер или электронная схема. В некоторых
 30 вариантах выполнения один или более из некоторых наиболее важных этапов способа могут быть выполнены посредством такого устройства.

Передаваемый или кодированный сигнал согласно изобретению может быть сохранен на цифровом носителе данных или может быть передан в среде передачи, такой как беспроводная среда передачи или проводная среда передачи, такая как Интернет.

В зависимости от различных требований к реализации варианты выполнения изобретения могут быть реализованы в аппаратном или программном обеспечении. Реализация может быть осуществлена с использованием цифрового носителя данных, например гибкого магнитного диска, DVD, диска Blu-Ray, CD, ROM, PROM и EPROM, EEPROM или FLASH-памяти, на которых сохранены считываемые электронными
 40 средствами управляющие сигналы, которые взаимодействуют (или способны взаимодействовать) с программируемой компьютерной системой таким образом, что выполняется соответствующий способ. Таким образом, цифровой носитель данных может быть машиночитаемым.

Некоторые варианты выполнения согласно изобретению содержат носитель данных, имеющий считываемые электронными средствами управляющие сигналы, которые способны взаимодействовать с программируемой компьютерной системой таким образом, что выполняется один из способов, описанных в настоящем документе.

В общем случае, варианты выполнения настоящего изобретения могут быть

реализованы в виде компьютерного программного продукта с программным кодом, причем программный код выполнен с возможностью выполнения одного из способов, когда компьютерная программа выполняется на компьютере. Программный код может, например, быть сохранен на машиночитаемом носителе.

5 Другие варианты выполнения содержат компьютерную программу для выполнения одного из способов, описанных в настоящем документе, сохраненную на машиночитаемом носителе.

Другими словами, вариант выполнения способа согласно изобретению, таким образом, представляет собой компьютерную программу, имеющую программный код
10 для выполнения одного из способов, описанных в настоящем документе, когда компьютерная программа выполняется на компьютере.

Другой вариант выполнения способа согласно изобретению, таким образом, представляет собой носитель данных (или постоянный носитель данных, такой как цифровой носитель данных или машиночитаемый носитель), содержащий записанную
15 на нем компьютерную программу для выполнения одного из способов, описанных в настоящем документе. Носитель данных, цифровой носитель данных или носитель записи обычно являются материальными и/или постоянными.

Другой вариант выполнения способа согласно изобретению, таким образом, представляет собой поток данных или последовательность сигналов, представляющие
20 компьютерную программу для выполнения одного из способов, описанных в настоящем документе. Поток данных или последовательность сигналов могут, например, быть выполнены с возможностью их передачи посредством соединения для передачи данных, например через Интернет.

Другой вариант выполнения содержит средство обработки, например компьютер
25 или программируемое логическое устройство, конфигурированное или выполненное с возможностью выполнения одного из способов, описанных в настоящем документе.

Другой вариант выполнения содержит компьютер, на котором установлена компьютерная программа для выполнения одного из способов, описанных в настоящем документе.

30 Другой вариант выполнения согласно изобретению содержит устройство или систему, выполненные с возможностью передачи (например, электронными или оптическими средствами) компьютерной программы для выполнения одного из способов, описанных в настоящем документе, в приемник. Приемник может быть, например, компьютером, мобильным устройством, запоминающим устройством или тому подобным. Устройство
35 или система могут содержать, например, файловый сервер для передачи компьютерной программы в приемник.

В некоторых вариантах выполнения может использоваться программируемое логическое устройство (например, программируемая вентильная матрица) для выполнения некоторых или всех функций способов, описанных в настоящем документе.

40 В некоторых вариантах выполнения программируемая вентильная матрица может взаимодействовать с микропроцессором для выполнения одного из способов, описанных в настоящем документе. В общем случае способы предпочтительно выполняются любым аппаратным устройством.

Вышеописанные варианты выполнения являются лишь иллюстрацией принципов настоящего изобретения. Следует понимать, что другим специалистам в данной области техники будут очевидны модификации и изменения в конфигурациях и подробностях, описанных в настоящем документе. Таким образом, подразумевается ограничение
45 только объемом нижеследующей формулы изобретения, но не конкретными

подробностями, представленными в настоящем документе в качестве описания и пояснения вариантов выполнения изобретения.

Список литературы

- В. Bessette и др., "The Adaptive Multi-rate Wideband Speech Codec (AMR-WB)," IEEE Trans. on Speech and Audio Processing, том 10, №. 8, ноябрь 2002 г.
- В. Geiser и др., "Bandwidth Extension for Hierarchical Speech and Audio Coding in ITU-T Rec. G.729.1," IEEE Trans. on Audio, Speech, and Language Processing, том 15, № 8, ноябрь 2007 г.
- В. Iser, W. Minker и G. Schmidt, Bandwidth Extension of Speech Signals, Springer Lecture Notes in Electrical Engineering, том 13, Нью-Йорк, 2008 г.
- М. Jelínek и R. Salami, "Wideband Speech Coding Advances in VMR-WB Standard," IEEE Trans. on Audio, Speech, and Language Processing, том 15, №. 4, май 2007 г.
- И. Katsir, I. Cohen и D. Malah, "Speech Bandwidth Extension Based on Speech Phonetic Content and Speaker Vocal Tract Shape Estimation," в Proc. EUSIPCO 2011, Барселона, Испания, сентябрь 2011 г.
- Е. Larsen и R. M. Aarts, Audio Bandwidth Extension: Application of Psychoacoustics, Signal Processing and Loudspeaker Design, Уайли, Нью-Йорк, 2004 г.
- Ж. Mäkinen и др., "AMR-WB+: A New Audio Coding Standard for 3rd Generation Mobile Audio Services," в Proc. ICASSP 2005, Филадельфия, США, март 2005 г.
- М. Neuendorf и др., "MPEG Unified Speech and Audio Coding - The ISO/MPEG Standard for High-Efficiency Audio Coding of All Content Types," в Proc. 132nd Convention of the AES, Будапешт, Венгрия, апрель 2012 г. Также опубликовано в Журнале AES в 2013 г.
- Н. Pulakka и Р. Alku, "Bandwidth Extension of Telephone Speech Using a Neural Network and a Filter Bank Implementation for Highband Mel Spectrum," IEEE Trans. on Audio, Speech, and Language Processing, том 19, № 7, сентябрь 2011 г.
- Т. Vaillancourt и др., "ITU-T EV-VBR: A Robust 8-32 kbit/s Scalable Coder for Error Prone Telecommunications Channels," в Proc. EUSIPCO 2008, Лозанна, Швейцария, август 2008 г.
- Л. Miao и др., "G.711.1 Annex D and G.722 Annex B: New ITU-T Superwideband codecs," в Proc. ICASSP 2011, Прага, Чехия, май 2011 г.
- Bernd Geiser, Peter Jax и Peter Vary: "ROBUST WIDEBAND ENHANCEMENT OF SPEECH BY COMBINED CODING AND ARTIFICIAL BANDWIDTH EXTENSION", Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC), 2005 г.

(57) Формула изобретения

1. Декодер для формирования аудиосигнала (120) с улучшенной частотной характеристикой, содержащий:

блок (104) извлечения свойств для извлечения свойства из базового сигнала (100);

блок (110) извлечения дополнительной информации для извлечения дополнительной информации для выбора, ассоциированной с базовым сигналом;

генератор (108) параметров для формирования параметрического представления для оценки спектрального диапазона аудиосигнала (120) с улучшенной частотной характеристикой, не определяемого базовым сигналом (100), причем генератор (108) параметров выполнен с возможностью обеспечения некоторого количества альтернативных параметрических представлений (702, 704, 706, 708) в ответ на упомянутое свойство (112), и при этом генератор (108) параметров выполнен с возможностью выбора одного из альтернативных параметрических представлений в качестве параметрического представления в ответ на дополнительную информацию

(712-718) для выбора; и

блок (118) оценки сигнала для оценки аудиосигнала (120) с улучшенной частотной характеристикой с использованием выбранного параметрического представления;

причем генератор (108) параметров выполнен с возможностью приема

5 параметрической информации (1100) улучшения частотной характеристики, ассоциированной с базовым сигналом (100), причем параметрическая информация улучшения частотной характеристики содержит группу отдельных параметров,

причем генератор (108) параметров выполнен с возможностью обеспечения выбранного параметрического представления в дополнение к параметрической

10 информации улучшения частотной характеристики,

причем выбранное параметрическое представление содержит параметр, не включенный в группу отдельных параметров, или значение изменения параметра для изменения параметра в группе отдельных параметров, и

15 при этом блок (118) оценки сигнала выполнен с возможностью оценки аудиосигнала с улучшенной частотной характеристикой с использованием выбранного параметрического представления и параметрической информации (1100) улучшения частотной характеристики.

2. Декодер по п. 1, дополнительно содержащий:

интерфейс (210) ввода для приема кодированного входного сигнала (200),

20 содержащего кодированный базовый сигнал (201) и дополнительную информацию (114) для выбора; и

базовый декодер (124) для декодирования кодированного базового сигнала для получения базового сигнала (100).

3. Декодер по п. 1, в котором генератор (108) параметров выполнен с возможностью 25 использования при выборе одного из альтернативных параметрических представлений предварительно заданного порядка альтернативных параметрических представлений или сигнализируемого кодером порядка альтернативных параметрических представлений.

4. Декодер по п. 1, в котором генератор (108) параметров выполнен с возможностью 30 обеспечения представления огибающей в качестве параметрического представления,

причем дополнительная информация (114) для выбора указывает на один из множества различных сибилантов или фрикативных звуков, и

35 при этом генератор (108) параметров выполнен с возможностью обеспечения представления огибающей, идентифицируемого дополнительной информацией для выбора.

5. Декодер по п. 1,

в котором блок (118) оценки сигнала содержит интерполятор (900) для интерполяции базового сигнала (100), и

40 при этом блок (104) извлечения свойств выполнен с возможностью извлечения свойства из неинтерполированного базового сигнала (100).

6. Декодер по п. 1,

в котором блок (118) оценки сигнала содержит:

анализирующий фильтр (910) для анализа базового сигнала или интерполированного базового сигнала для получения сигнала возбуждения;

45 блок (912) расширения сигнала возбуждения для формирования улучшенного сигнала возбуждения, имеющего спектральный диапазон, не включенный в базовый сигнал (100); и

синтезирующий фильтр (914) для фильтрации расширенного сигнала возбуждения;

причем анализирующий фильтр (910) или синтезирующий фильтр (914) определяется выбранным параметрическим представлением.

7. Декодер по п. 1,

в котором блок (118) оценки сигнала содержит процессор расширения спектральной полосы частот для формирования расширенной спектральной полосы, соответствующей спектральному диапазону, не включенному в базовый сигнал, с использованием по меньшей мере спектральной полосы базового сигнала и параметрического представления,

причем параметрическое представление содержит параметры для по меньшей мере одного из регулирования (1060) спектральной огибающей, добавления (1020) маскирующего шума, инверсной фильтрации (1040) и добавления (1080) отсутствующих тонов,

причем генератор параметров выполнен с возможностью обеспечения, для упомянутого свойства, множества альтернативных параметрических представлений, причем каждое альтернативное параметрическое представление имеет параметры для по меньшей мере одного из регулирования (1060) спектральной огибающей, добавления (1020) маскирующего шума, инверсной фильтрации (1040) и добавления (1080) отсутствующих тонов.

8. Декодер по п. 1, дополнительно содержащий:

детектор голосовой активности или детектор (500) голосовых/неголосовых данных, причем блок (118) оценки сигнала выполнен с возможностью оценки сигнала с улучшенной частотной характеристикой с использованием параметрического представления, только когда детектор голосовой активности или детектор (500) голосовых/неголосовых данных указывает на голосовую активность или голосовой сигнал.

9. Декодер по п. 8,

в котором блок (118) оценки сигналов выполнен с возможностью переключения (502, 504) с процедуры (511) улучшения частотной характеристики на другую процедуру (513) улучшения частотной характеристики или использования других параметров (514), извлеченных из кодированного сигнала, когда детектор голосовой активности или детектор (500) голосовых/неголосовых данных указывает на неголосовой сигнал или на сигнал, не содержащий голосовой активности.

10. Декодер по п. 1,

в котором статистическая модель выполнена с возможностью обеспечения, в ответ на упомянутое свойство, множества альтернативных параметрических представлений (702-708),

причем каждое альтернативное параметрическое представление обладает вероятностью, идентичной вероятности другого альтернативного параметрического представления или отличной от вероятности упомянутого альтернативного параметрического представления менее чем на 10% от максимальной вероятности.

11. Декодер по п. 1,

в котором дополнительная информация для выбора включена только в кадр (800) кодированного сигнала, когда генератор (108) параметров обеспечивает множество альтернативных параметрических представлений, и

причем дополнительная информация для выбора не включена в другой кадр (812) кодированного аудиосигнала, в котором генератор (108) параметров обеспечивает только одно альтернативное параметрическое представление в ответ на упомянутое свойство (112).

12. Кодер для формирования кодированного сигнала (1212), содержащий:

базовый кодер (1200) для кодирования исходного сигнала (1206), чтобы получить кодированный аудиосигнал (1208), содержащий информацию о меньшем количестве полос частот по сравнению с исходным сигналом (1206);

5 генератор (1202) дополнительной информации для выбора для формирования дополнительной информации (1210) для выбора, указывающей на определенное альтернативное параметрическое представление (702-708), обеспеченное статистической моделью в ответ на свойство (112), извлеченное из исходного сигнала (1206), или из кодированного аудиосигнала (1208), или из декодированной версии кодированного аудиосигнала (1208); и

интерфейс (1204) вывода для вывода кодированного сигнала (1212), причем кодированный сигнал содержит кодированный аудиосигнал (1208) и дополнительную информацию (1210) для выбора;

15 причем исходный сигнал содержит ассоциированную метаинформацию, описывающую последовательность акустической информации для последовательности выборок исходного аудиосигнала,

причем генератор (1202) дополнительной информации для выбора содержит:

блок (1400) извлечения метаданных для извлечения последовательности метаинформации и

20 блок (1402) интерпретации метаданных для интерпретации последовательности метаинформации в последовательность дополнительной информации (1210) для выбора.

13. Кодер по п. 12,

в котором интерфейс (1204) вывода выполнен с возможностью включать дополнительную информацию (1210) для выбора в кодированный сигнал (1212), только
25 когда статистическая модель обеспечивает множество альтернативных параметрических представлений, и не включать какую-либо дополнительную информацию для выбора в кадр кодированного аудиосигнала (1208), в котором статистическая модель выполнена с возможностью обеспечения только одного параметрического представления в ответ на упомянутое свойство.

30 14. Способ формирования аудиосигнала (120) с улучшенной частотной характеристикой, содержащий этапы, на которых:

извлекают (104) свойство из базового сигнала (100);

извлекают (110) дополнительную информацию для выбора, ассоциированную с базовым сигналом;

35 формируют (108) параметрическое представление для оценки спектрального диапазона аудиосигнала (120) с улучшенной частотной характеристикой, не определяемого базовым сигналом (100), причем обеспечивают некоторое количество альтернативных параметрических представлений (702, 704, 706, 708) в ответ на упомянутое свойство (112), и при этом выбирают одно из альтернативных
40 параметрических представлений в качестве параметрического представления в ответ на дополнительную информацию (712-718) для выбора; и

оценивают (118) аудиосигнал (120) с улучшенной частотной характеристикой с использованием выбранного параметрического представления;

45 причем при формировании параметрического представления принимают параметрическую информацию (1100) улучшения частотной характеристики, ассоциированной с базовым сигналом (100), причем параметрическая информация улучшения частотной характеристики содержит группу отдельных параметров, причем при формировании параметрического представления генератор (108)

параметров обеспечивает выбранное параметрическое представление в дополнение к параметрической информации улучшения частотной характеристики,

причем выбранное параметрическое представление содержит параметр, не включенный в группу отдельных параметров, или значение изменения параметра для изменения параметра в группе отдельных параметров,

при этом при оценке оценивают аудиосигнал с улучшенной частотной характеристикой с использованием выбранного параметрического представления и параметрической информации (1100) улучшения частотной характеристики.

15. Способ формирования кодированного сигнала (1212), содержащий этапы, на которых:

кодируют (1200) исходный сигнал (1206) для получения кодированного аудиосигнала (1208), содержащего информацию о меньшем количестве полос частот по сравнению с исходным сигналом (1206);

формируют (1202) дополнительную информацию (1210) для выбора, указывающую на альтернативное параметрическое представление (702-708), обеспеченное статистической моделью в ответ на свойство (112), извлеченное из исходного сигнала (1206), или из кодированного аудиосигнала (1208), или из декодированной версии кодированного аудиосигнала (1208); и

выводят (1204) кодированный сигнал (1212), причем кодированный сигнал содержит кодированный аудиосигнал (1208) и дополнительную информацию (1210) для выбора; причем исходный сигнал содержит ассоциированную метаинформацию, описывающую последовательность акустической информации для последовательности выборок исходного аудиосигнала,

причем формирование дополнительной информации для выбора содержит этапы, на которых:

извлекают (1400) метаданные последовательность метаинформации; и

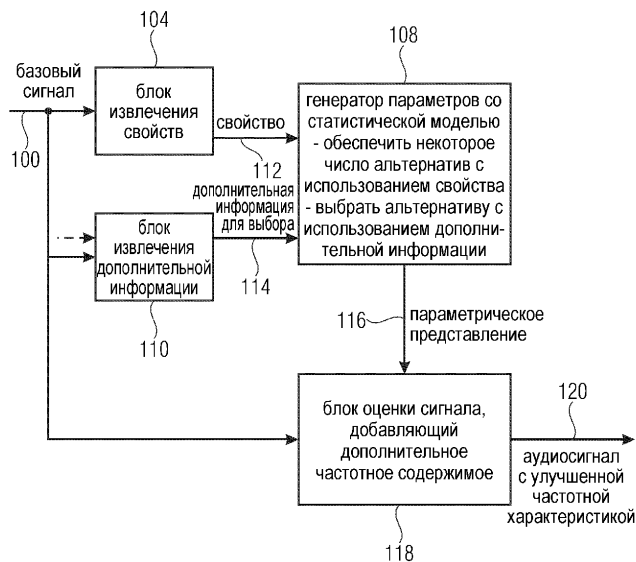
интерпретируют (1402) последовательность метаинформации в последовательность дополнительной информации (1210) для выбора.

16. Машиночитаемый носитель, имеющий сохраненную на нем компьютерную программу для выполнения, при исполнении на компьютере или в процессоре, способа по п. 14.

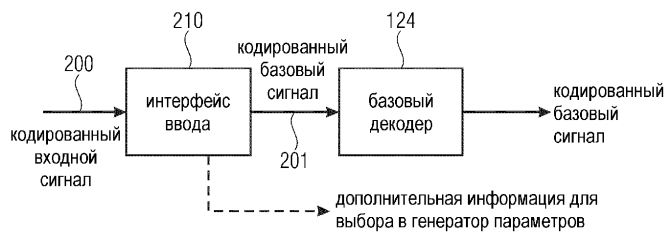
17. Машиночитаемый носитель, имеющий сохраненную на нем компьютерную программу для выполнения, при исполнении на компьютере или в процессоре, способа по п. 15.

527288

1/13



ФИГ.1

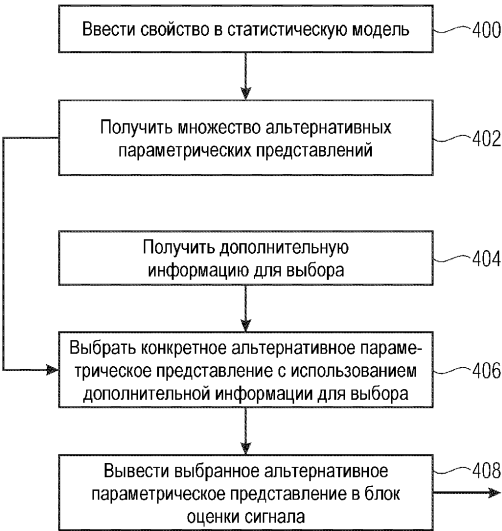


ФИГ.2

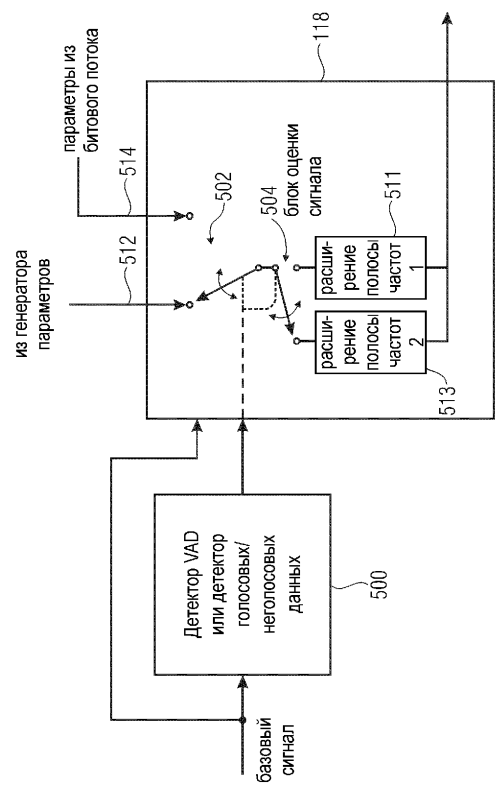
2/13

количество битов дополнительной информации для выбора	количество альтернативных параметрических представлений (самое большое)
1	2
2	4
3	8
4	16
5	32

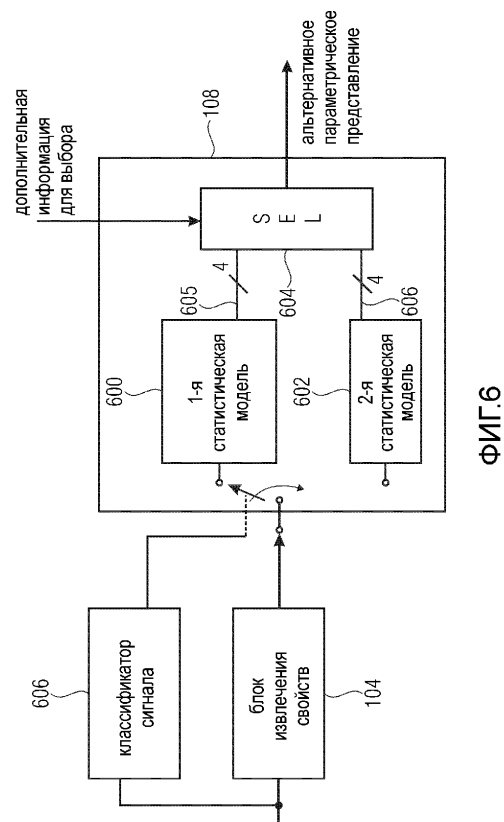
ФИГ.3



ФИГ.4

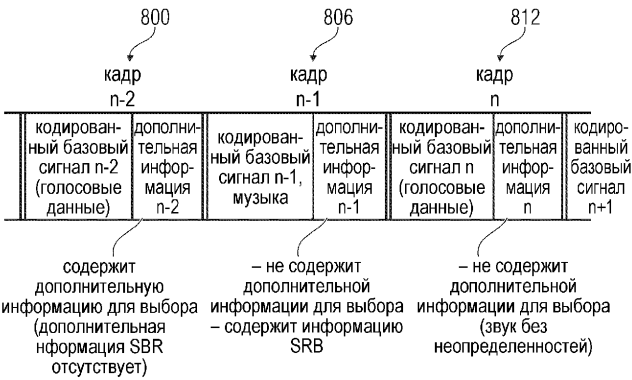


ФИГ.5

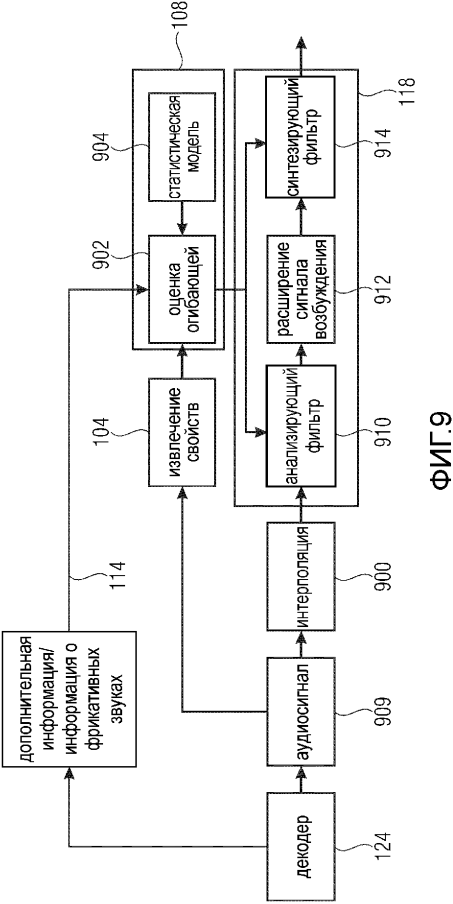


результат статистической модели		дополнительная информация для выбора	
АЛЬТЕРНАТИВА 1	702	00	712
АЛЬТЕРНАТИВА 2	704	01	714
АЛЬТЕРНАТИВА 3	706	10	716
АЛЬТЕРНАТИВА 4	708	11	718

ФИГ.7

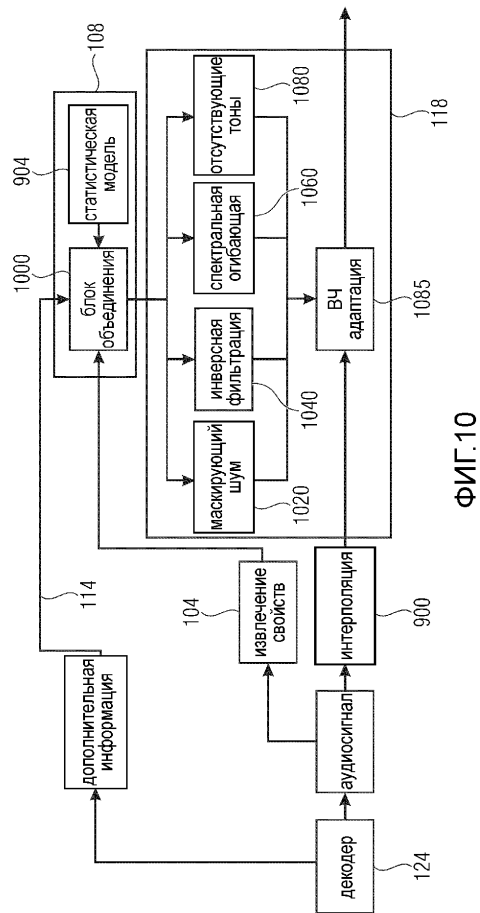


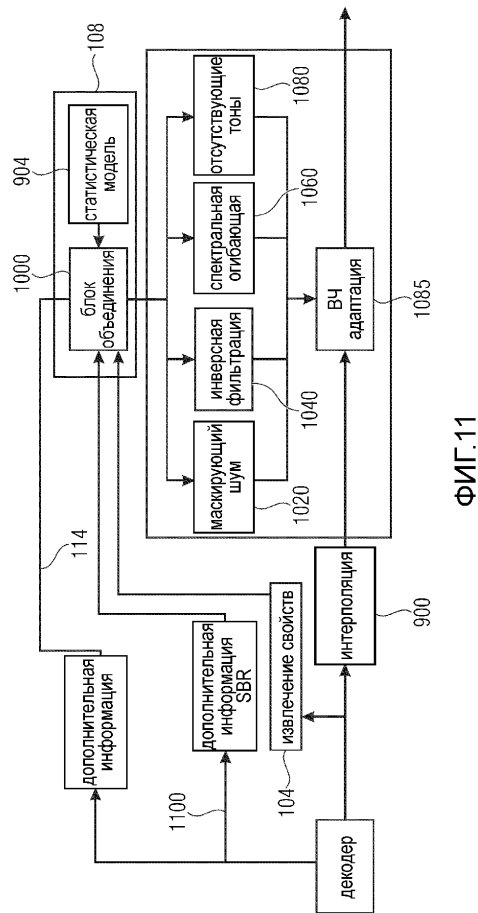
ФИГ.8



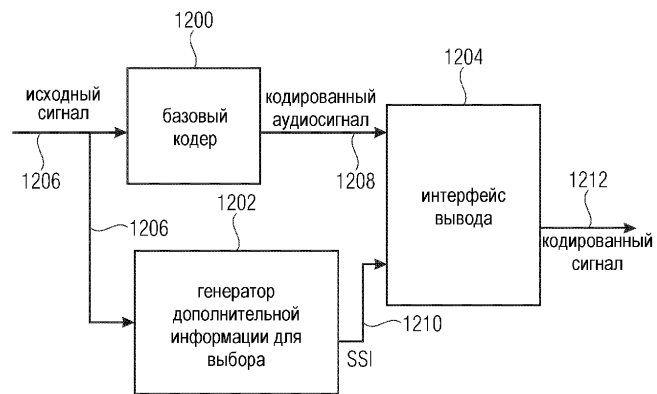
ФИГ.9

7/13



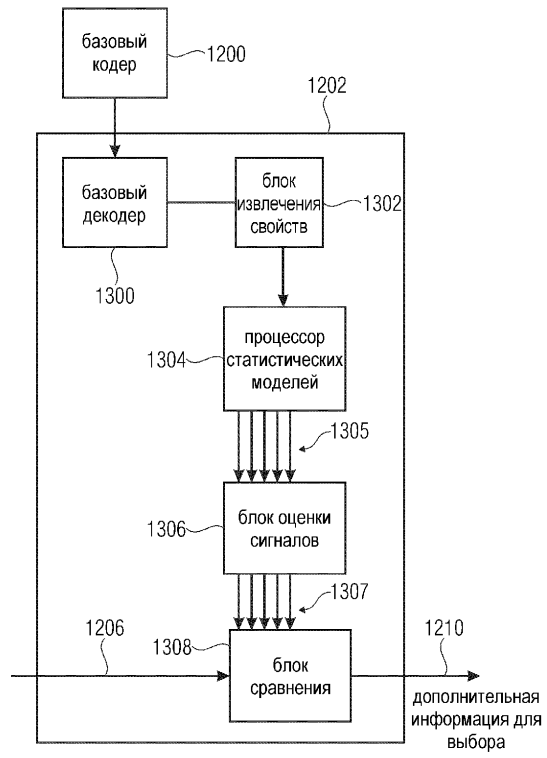


9/13

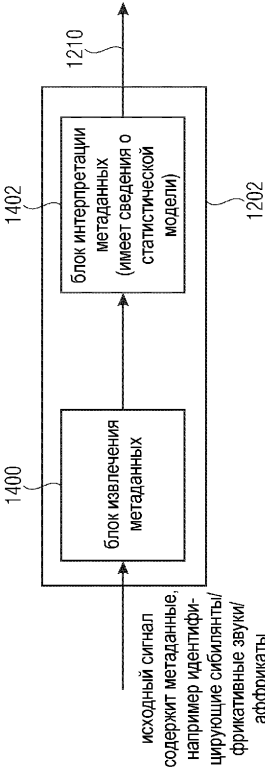


ФИГ.12

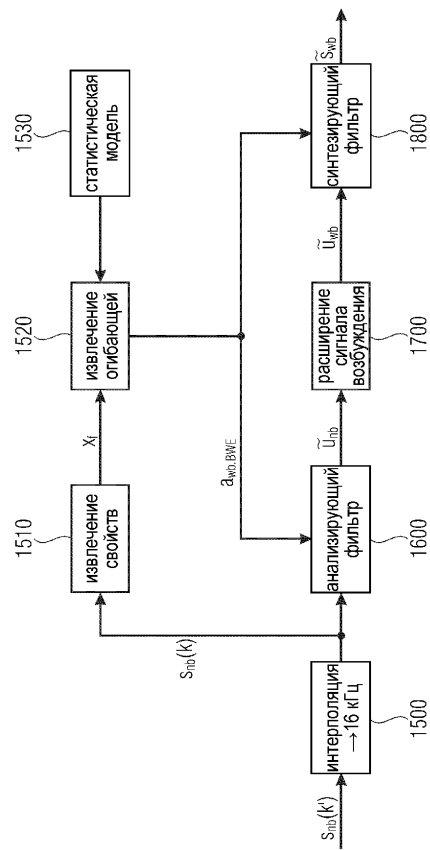
10/13



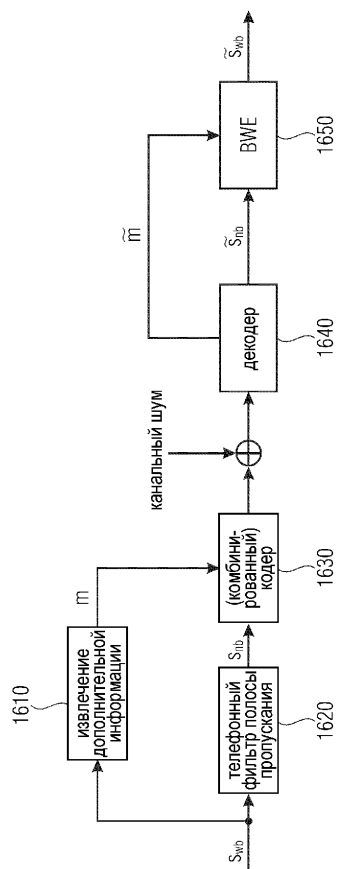
ФИГ.13



ФИГ.14



ФИГ.15
УРОВЕНЬ ТЕХНИКИ



ФИГ.16
УРОВЕНЬ ТЕХНИКИ