



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ,
ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21), (22) Заявка: **96111955/09, 19.06.1996**

(24) Дата начала действия патента: **19.06.1996**

(30) Приоритет: **20.06.1995 JP P07-153723**

(43) Дата публикации заявки: **27.09.1998**

(45) Опубликовано: **27.06.2005 Бюл. № 18**

(56) Список документов, цитированных в отчете о поиске: **SU 1314373, 30.05.1987. EP 0506394 A2, 30.09.1992. US 5299281 A, 29.03.1994. WO 94/18669 A1, 18.08.1994.**

Адрес для переписки:

**103735, Москва, ул. Ильинка, 5/2, ООО
"Союзпатент", пат.пов. С.Б.Фелицыной**

(72) Автор(ы):

НИСИГУТИ Масаюки (JP)

(73) Патентообладатель(ли):

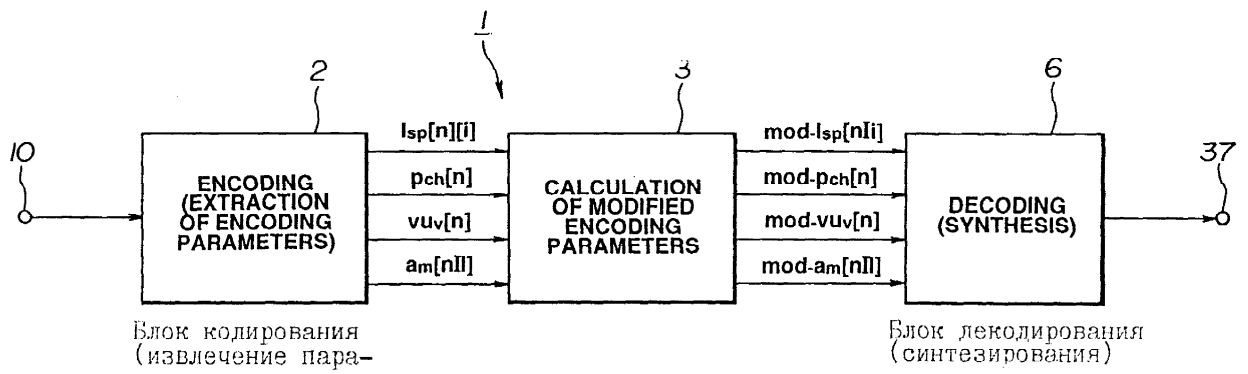
СОНИ КОРПОРЕЙШН (JP)

(54) СПОСОБ И УСТРОЙСТВО ВОСПРОИЗВЕДЕНИЯ РЕЧЕВЫХ СИГНАЛОВ И СПОСОБ ИХ ПЕРЕДАЧИ

(57) Реферат:

Изобретение касается способа и устройства воспроизведения речевых сигналов, в которых речевой сигнал делят на множество кадров и кодируют для обнаружения параметров кодирования, основанных на том, какие, по меньшей мере, гармонические волны синтезируются для воспроизведения речевого сигнала. Сущность изобретения состоит в том, что при кодировании разделяют речевые сигналы на кадры и кодируют разделенные сигналы на кадровой основе для вывода параметров кодирования, таких, как параметры линейной спектральной пары, высоты тона, вокализованный - невокализованный сигналы или спектральная

амплитуда. При вычислении видеоизмененных параметров кодирования интерполируют параметры кодирования для вычисления видеоизмененных параметров кодирования, связанных с основанными на кадрах временными периодами. При декодировании синтезируют гармонические волны и шум на основании видеоизмененных параметров кодирования и выделяют синтезированные речевые сигналы. Технический результат, достигаемый при осуществлении изобретения, состоит в том, что при воспроизведении речевых сигналов осуществляют управление скоростью в широком диапазоне с высоким качеством звука при неизменных фонах и высоты тона. 3 н. и 9 з.п. ф-лы, 24 ил.



Блок кодирования
(извлечение параметров кодирования)

Фиг.1

Блок декодирования
(синтезирования)

RU 2255380 C2

RU 2255380 C2



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY,
PATENTS AND TRADEMARKS

(12) **ABSTRACT OF INVENTION**

(21), (22) Application: **96111955/09, 19.06.1996**
 (24) Effective date for property rights: **19.06.1996**
 (30) Priority: **20.06.1995 JP P07-153723**
 (43) Application published: **27.09.1998**
 (45) Date of publication: **27.06.2005 Bull. 18**
 Mail address:
103735, Moskva, ul. Il'inka, 5/2, OOO
"Sojuzpatent", pat.pov. S.B.Felitsynoj

(72) Inventor(s):
NISIGUTI Masajuki (JP)
 (73) Proprietor(s):
SONI KORPOREJShN (JP)

(54) **METHOD AND DEVICE FOR REPRODUCING SPEECH SIGNALS AND METHOD FOR TRANSFERRING SAID SIGNALS**

(57) Abstract:

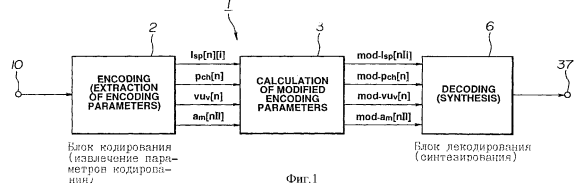
FIELD: speech recording/reproducing devices.

SUBSTANCE: during encoding speech signals are separated on frames and separated signals are encoded on frame basis for output of encoding parameters like parameters of linear spectral couple, tone height, vocalized/non-vocalized signals or spectral amplitude. During calculation of altered parameters of encoding, encoding parameters are interpolated for calculation of altered encoding parameters, connected to temporal periods based on frames. During decoding

harmonic waves and noise are synthesized on basis of altered encoding parameters and synthesized speech signals are selected.

EFFECT: broader functional capabilities, higher efficiency.

3 cl, 24 dwg



RU 2 2 5 5 3 8 0 C 2

RU 2 2 5 5 3 8 0 C 2

Предпосылки создания изобретения

Область техники, к которой относится изобретение

Настоящее изобретение касается способа и устройства воспроизведения речевых сигналов, в которых входной речевой сигнал делят на множество кадров в качестве
5 элементов и кодируют для обнаружения параметров кодирования, основанных на том, какие, по меньшей мере, гармонические волны синтезируются для воспроизведения речевого сигнала. Изобретение касается также способа передачи видоизмененных параметров кодирования, полученных при интерполировании параметров кодирования,

Описание родственной техники

10 В настоящее время известно множество способов кодирования, предназначенных для сжатия сигналов посредством использования статических свойств сигналов звуковых частот, включающих речевые сигналы и сигналы звукового сопровождения, во временной области и в частотной области и относящиеся к психологии слухового восприятия характеристики системы человеческого слухового аппарата. Эти способы кодирования
15 грубо классифицируются на кодирование по временной области, кодирование по частотной области и кодирование посредством анализа-синтеза.

Между тем, при высокоэффективном способе кодирования речи посредством обработки сигналов на временной оси, иллюстрируемого линейным предсказанием с возбуждением
20 кода (ЛПВК), встречаются трудности при преобразовании (изменении) скорости временной оси скорее из-за длительных операций по обработке сигналов, выводимых с декодирующего устройства.

Кроме того, вышеупомянутый способ нельзя использовать, например, для преобразования частоты основного тона, потому что управление скоростью выполняется в декодируемом линейном диапазоне.

25 Ввиду вышеизложенного, целью настоящего изобретения является обеспечить способ и устройство, предназначенные для воспроизведения речевых сигналов, в которых управление скоростью произвольной частоты в широком диапазоне можно легко выполнять с высоким качеством при оставлении неизменными фонемы и высоты тона.

В одном аспекте, настоящее изобретение обеспечивает способ воспроизведения
30 входного речевого сигнала на основании параметров кодирования, получаемых посредством разделения входного речевого сигнала исходя из заранее установленных кадров на временной оси и кодирования такого разделенного входного речевого сигнала на кадровой основе, включающий в себе этапы интерполирования параметров кодирования, предназначенных для определения видоизмененных параметров
35 кодирования, связанных с требуемыми временными точками, и выработки отличия модифицированного речевого сигнала по частоте от упомянутого входного речевого сигнала на основании видоизмененных параметров кодирования. Таким образом, управление скоростью на произвольной частоте в широком диапазоне можно легко выполнять с высоким качеством сигнала при оставлении неизменными фонемы и высоты
40 тона.

В другом аспекте, настоящее изобретение обеспечивает устройство, предназначенное для воспроизведения речевого сигнала, в котором входной речевой сигнал
45 восстанавливается на основании параметров кодирования, полученных посредством разделения входного речевого сигнала исходя из заранее установленных кадров на временной оси, и кодирования такого разделенного входного речевого сигнала на кадровой основе, включающее в себя средство интерполирования, предназначенное для интерполирования параметров кодирования для обнаружения видоизмененных параметров кодирования, связанных с требуемыми временными точками, и средство выработки речевого сигнала, предназначенное для выработки отличия видоизмененного
50 речевого сигнала по частоте от упомянутого входного речевого сигнала на основании видоизмененных параметров кодирования. Таким образом, становится возможным регулировать скорость передачи битов. Следовательно, управление скоростью на произвольной частоте в широком диапазоне можно легко выполнять с высоким качеством

сигнала, при оставлении неизменными фонемы и высоты тона.

Еще в одном аспекте, настоящее изобретение обеспечивает способ передачи речевых сигналов, в котором параметры кодирования обнаруживаются посредством разделения входного речевого сигнала исходя из заранее установленных кадров на временной оси в качестве элементов и посредством кодирования этого разделенного входного речевого сигнала на кадровой основе с целью обнаружения параметров кодирования, причем обнаруженные таким образом параметры кодирования интерполируются с целью определения видеоизмененных параметров кодирования, связанных с требуемой временной точкой, и видеоизмененные параметры кодирования передаются, обеспечивая, таким образом, возможность регулирования скорости передачи битов.

Благодаря делению входного речевого сигнала исходя из предварительно установленных кадров на временной оси и кодирования основанного на кадрах сигнала с целью обнаружения параметров кодирования, благодаря интерполированию параметров кодирования с целью определения видеоизмененных параметров кодирования и благодаря синтезированию, по меньшей мере, гармонических волн на основании видеоизмененных параметров кодирования для восстановления речевых сигналов, становится возможной регулировка скорости на произвольной частоте.

Краткое описание чертежей

Фиг.1 представляет структурную блок-схему, иллюстрирующую компоновку устройства воспроизведения речевого сигнала, соответствующего первому варианту осуществления настоящего изобретения.

Фиг.2 представляет структурную блок-схему, иллюстрирующую компоновку показанного на фиг.1 устройства воспроизведения речевого сигнала.

Фиг.3 представляет блок-схему, иллюстрирующую кодирующее устройство показанного на фиг.1 устройства воспроизведения речевого сигнала.

Фиг.4 представляет блок-схему, иллюстрирующую компоновку схемы анализа многополосного возбуждения (МПВ) в качестве иллюстративного примера схемы кодирования гармоник и шума кодирующего устройства.

Фиг.5 иллюстрирует компоновку векторного квантователя.

Фиг.6 представляет график, иллюстрирующий средние значения входного сигнала \bar{x} вокализованного звука, невокализованного звука и собранных вместе вокализованного и невокализованного звуков.

Фиг.7 представляет график, иллюстрирующий средние значения весового множителя $w' \|\bar{x}\|$ для вокализованного звука, невокализованного звука и для собранных вместе вокализованного и невокализованного звуков.

Фиг.8 представляет график, иллюстрирующий способ формирования кодового словаря векторного квантования для вокализованного звука, невокализованного звука и для собранных вместе вокализованного и невокализованного звуков.

Фиг.9 представляет алгоритм, иллюстрирующий схематическую работу схемы вычисления видеоизмененных параметров кодирования, используемой в показанном на фиг.1 устройстве воспроизведения речевого сигнала.

Фиг.10 представляет схематический вид, иллюстрирующий видеоизмененные параметры кодирования, получаемые посредством схемы вычисления видеоизмененных параметров на временной оси.

Фиг.11 представляет алгоритм, иллюстрирующий подробную работу схемы вычисления видеоизмененных параметров кодирования, используемой в показанном на фиг.1 устройстве воспроизведения речевых сигналов.

Фиг.12А, 12В и 12С представляют схематические виды, показывающие иллюстративную работу схемы вычисления видеоизмененных параметров кодирования.

Фиг.13А, 13В и 13С представляют схематические виды, показывающие другую иллюстративную работу схемы вычисления видеоизмененных параметров кодирования.

Фиг.14 представляет блок-схему, иллюстрирующую декодирующее устройство, используемое в устройстве воспроизведения речевых сигналов.

Фиг.15 представляет электрическую блок-схему, иллюстрирующую компоновку схемы синтеза многополосного возбуждения (МПВ) в виде иллюстративного примера схемы синтеза гармоник и шума, используемой в декодирующем устройстве.

Фиг.16 представляет блок-схему, иллюстрирующую устройство передачи речевого сигнала в виде второго варианта осуществления настоящего изобретения.

Фиг.17 представляет алгоритм, иллюстрирующий работу передающей стороны устройства передачи речевых сигналов.

Фиг.18А, 18В и 18С иллюстрируют работу устройства передачи речевых сигналов.

Описание предпочтительных вариантов осуществления изобретения

Ниже будет приведено подробное описание со ссылкой на чертежи предпочтительных вариантов осуществления соответствующих настоящему изобретению способа и устройства, предназначенных для воспроизведения речевых сигналов, и способа передачи речевых сигналов.

Вначале приведем описание устройства, предназначенного для воспроизведения речевых сигналов, в котором применены соответствующие настоящему изобретению способ и аппаратура для воспроизведения речевых сигналов. На фиг.1 показана блок-схема устройства воспроизведения речевых сигналов 1, в котором входные речевые сигналы разделяются исходя из заранее установленных кадров в качестве элементов на временной оси и кодируются на кадровой основе с целью обнаружения параметров кодирования. На основании этих параметров кодирования синтезируются гармонические волны и шум с целью воспроизведения речевых сигналов.

В частности, в случае настоящего устройства воспроизведения речевых сигналов 1, параметры кодирования интерполируются с целью определения видоизмененных параметров кодирования, связанных с требуемыми временными точками, и на основании этих видоизмененных параметров кодирования синтезируются гармонические волны и шум. Хотя на основании модифицированных параметров кодирования синтезируются гармонические волны и шум, можно также синтезировать, по меньшей мере, гармонические волны.

В данном случае устройство воспроизведения сигналов звуковой частоты включает в себя блок кодирования 2, предназначенный для разделения речевых сигналов, поступающих на входную клемму 10, на кадры в качестве элементов и для кодирования речевых сигналов на кадровой основе с целью вывода параметров кодирования, таких как параметры линейной спектральной пары (ЛСП), тон, вокализованные (V)-невокализованные (UV) сигналы или спектральные амплитуды A_m . Устройство воспроизведения сигналов звуковой частоты 1 включает в себя также блок вычисления 3, предназначенный для интерполирования параметров кодирования с целью определения видоизмененных параметров кодирования, связанных с требуемыми временными точками, и блок декодирования 6, предназначенный для синтеза гармонических волн и шума на основании видоизмененных параметров кодирования с целью вывода синтезированных речевых параметров на выходную клемму 37. Блоком кодирования 2, блоком вычисления 3, предназначенным для вычисления видоизмененных параметров кодирования, и блоком декодирования 6 управляет контроллер (не показанный).

Блок вычисления 3, предназначенный для вычисления видоизмененных параметров кодирования устройства воспроизведения речевых сигналов 1, включает в себя схему изменения периода 4, предназначенную для сжатия-расширения временной оси параметров кодирования, получаемых в каждом заранее установленном кадре, с целью изменения периода вывода параметров кодирования, и схему интерполирования 5, предназначенную для интерполирования параметров с измененным периодом с целью создания видоизмененных параметров кодирования, связанных с основанными на кадрах временными периодами, как показано, например, на фиг.2. Далее будет приведено описание блока вычисления 3, предназначенного для вычисления видоизмененных параметров кодирования.

Вначале приведем описание блока кодирования 2. Блок кодирования 3 и блок

декодирования б представляют остаточные значения краткосрочных предсказаний, например, остаточные значения кодирования с линейным предсказанием (КЛП), исходя из кодирования гармоник и шума. В качестве альтернативы, блок кодирования 3 и блок декодирования б выполняют кодирование многополосного возбуждения (МПВ) или анализ

5 многополосного возбуждения (МПВ).

В случае обычного кодирования с линейным предсказанием возбуждаемого кода (ЛПВК), остаточные значения КЛП подвергаются направленному векторному квантованию в виде формы сигнала во времени. Поскольку блок кодирования 2 кодирует остаточные значения посредством кодирования гармоник или анализа МПВ, более ровную синтезированную

10 форму волны можно получить при векторном квантовании амплитуд спектральной огибающей гармоник при меньшем количестве двоичных разрядов, тогда как выходной фильтр синтезированной формы волны КЛП также весьма соответствует качеству звука. Между тем, амплитуды спектральной огибающей квантуются с использованием метода пространственного преобразования или преобразования количества данных,

15 предложенного настоящим заявителем в японской публикации патента Kokaі JP-A-51800. То есть, амплитуды спектральной огибающей подвергаются векторному квантованию заранее установленным количеством векторных размеров.

На фиг.3 показана иллюстративная схема блока кодирования 2. Речевые сигналы, поступающие на входную клемму 10, освобождаются от сигналов ненужного диапазона

20 частот посредством фильтра 11 и затем подаются на схему анализа кодирования с линейным предсказанием (КЛП) 12 и схему обратного фильтрования 21.

В схеме анализа КЛП 12 применена взвешивающая функция Хэмминга к входной форме волны при ее длине порядка 256 выборок в качестве блока, чтобы посредством метода автокорреляции обнаруживать коэффициенты линейного предсказания, то есть так

25 называемые α -параметры. Интервал кодирования в виде блока вывода данных составляет порядка 160 выборок. Если частота выборки составляет, например, 8 кГц, интервал кодирования 160 выборок соответствует 20 миллисекунд.

α - параметр со схемы анализа КЛП 12 подается на схему преобразования α - параметра в ЛСП 13, с целью преобразования в параметры линейной спектральной пары (ЛСП). То есть, α - параметры, обнаруженные в качестве коэффициентов фильтра

30 направленного типа, преобразуются, например, в десять, то есть в пять пар, параметров ЛСП. Это преобразование выполняется с использованием, например, метода Ньютона-Рафсона. Причина преобразования α - параметров в параметры ЛСП состоит в том, что параметры ЛСП превосходят α - параметры по характеристикам интерполирования.

Параметры ЛСП со схемы преобразования параметров в ЛСП 13 подвергаются векторному квантованию посредством векторного квантователя ЛСП 14. В это время можно обнаруживать междукадровую разницу до перехода к векторному квантованию. В качестве

35 альтернативы можно собирать и квантовать множество кадров посредством матричного квантования. Для квантования, рассчитываемые каждые 20 мс параметры ЛСП

40 подвергаются векторному квантованию, при длительности одного кадра, равной 20 мс.

Квантованный выходной сигнал векторного квантователя ЛСП 14, который является показателем векторного квантователя ЛСП, выводится на клемму 15. Квантованные векторы ЛСП подаются на схему интерполирования ЛСП 16.

Схема интерполирования ЛСП 16 интерполирует векторы ЛСП, обеспечиваемые векторным квантованием каждые 20 мс, для обеспечения восьмикратной скорости. То есть, векторы ЛСП располагают таким образом, чтобы их можно было корректировать каждые

45 2,5 мс. Причина состоит в том, что если остаточная форма волны обрабатывается посредством анализа-синтеза методом кодирования-декодирования многополосного возбуждения (МПВ), огибающая синтезированной формы волны представляет чрезвычайно

50 ровную форму волны, так что если коэффициенты кодирования с линейным предсказанием (КЛП) действительно изменяются каждые 20 мс, то появляется тенденция создания своеобразных звуков. Образованию таких своеобразных звуков может создаваться препятствие, если коэффициенты КЛП постоянно изменяются каждые 2,5 миллисекунды.

Для обратного фильтрования входного речевого сигнала, используя интерполированные таким образом векторы ЛСП с интервалом 2,5 мс, параметры ЛСП преобразуются посредством схемы преобразования ЛСП в α - параметры, которые представляют собой коэффициенты фильтра направленного типа, например, десяти последовательностей.

5 Выходные сигналы схемы 17 преобразования ЛСП в α подаются на схему обратного фильтрования 21, чтобы обеспечить выполнение обратного фильтрования при скорректированном α - параметре на интервале 2,5 мс с целью создания ровного

10 выходного сигнала. Выходной сигнал схемы обратного фильтрования 21 поступает на схему кодирования гармоник и шума 22, а именно на схему анализа многополосного возбуждения (МПВ).

Схема кодирования гармоник и шума (схема анализа МПВ) 22 анализирует выходной сигнал схемы обратной фильтрации 21 методом, аналогичным методу анализа МПВ. То есть, схема кодирования гармоник-шума 22 детектирует тон и рассчитывает амплитуду A_m каждой гармоники. Схема кодирования гармоник-шума 22 позволяет также выполнять

15 установление различия вокализованного (V) - невокализованного (UV) речевого сигнала и преобразует количество амплитуд A_m гармоник, которое изменяется с изменением тона до постоянного количества посредством пространственного преобразования. Для определения тона используется автокоррелирование входных остаточных величин КЛП, как объясняется ниже.

20 На фиг.4 показан пример схемы анализа кодирования многополосного возбуждения (МПВ) в виде схемы кодирования гармоник и шума 22.

В случае показанной на фиг.4 схемы анализа МПВ разработано моделирование при предположении, что здесь присутствует вокализованная часть и невокализованная часть в частотном диапазоне той же временной точки, которая представляет тот же блок

25 или кадр.

Остаточные значения КЛП или остаточные значения кодирования с линейным предсказанием (КЛП) со схемы обратного фильтрования 21 подаются на показанную на фиг. 4 входную клемму III. Таким образом, схема анализа МПВ выполняет анализ МПВ и кодирование входных остаточных значений КЛП.

30 Остаточные значения кодирования с линейным предсказанием (КЛП), поступающие на входную клемму III, подаются на блок извлечения тона 113, блок взвешивания 114 и блок вычисления энергии субблока 126, как описано ниже.

Поскольку входной сигнал блока извлечения тона 113 представляет собой остаточное значение КЛП, определение тона можно выполнять посредством детектирования

35 максимального значения автокорреляции остаточных значений. Блок извлечения тона 113 выполняет поиск тона посредством поиска с разомкнутым циклом. Извлеченные данные тона поступают на блок точного поиска тона 116, где точный поиск тона выполняется посредством поиска тона замкнутым циклом.

В блоке взвешивания 114 применяется заранее установленная функция взвешивания, например, взвешивающая функция Хэмминга, к каждому блоку из N выборок, для

40 последовательного перемещения взвешенного блока по временной оси с интервалом между кадрами из α - выборок. Последовательность данных временной области с блока взвешивания 114 обрабатывается посредством блока ортогонального преобразования, например, посредством быстрого преобразования Фурье (БПФ).

45 Если обнаруживается, что все полосы в блоке невокализованы (UV), блок вычисления энергии субблока 126 извлекает характеристическую величину, представляющую огибающую формы сигнала во времени невокализованного звукового сигнала блока.

На блок точного поиска тона 116 подаются грубые данные тона в виде целых чисел, извлекаемые блоком выделения тона 113, и данные частотной области, вырабатываемые

50 БПФ посредством блока ортогонального преобразования 115. Блок точного поиска тона 116 выполняет качание на \pm несколько выборок с интервалом от 0,2 до 0,5 относительно значения грубых данных тона в качестве центра для приведения к точным данным тона с оптимальной десятичной запятой (плавающей). При способе точного поиска используется

анализ методом синтезирования и выбирается шаг, который дает энергетический спектр при синтезировании, который находится ближе всего к первоначальному энергетическому спектру.

То есть, количество значений тона выше и ниже грубого тона, определяемого блоком выделения тона 113 в качестве центра, обеспечиваются с интервалом, например, равным 0,25. Для тех значений тона, которые ежеминутно отличаются друг от друга, определяется сумма погрешностей $\sum \epsilon_m$. В этом случае, если устанавливается тон, то устанавливается ширина полосы, так что используя энергетический спектр по данным частотной области и спектру сигнала возбуждения, определяется погрешность ϵ_m . Таким образом, можно определить сумму погрешностей $\sum \epsilon_m$ для общего количества полос. Эта сумма погрешностей $\sum \epsilon_m$ определяется для каждого значения тона, и в качестве оптимального тона выбирается тон, соответствующий минимальной сумме погрешностей. Таким образом определяется оптимальный точный тон с интервалом, примерно равным 0,25, посредством блока поиска точного тона, и определяется амплитуда $|A_{118}|$ для оптимальной высоты тона. Значение амплитуды вычисляется посредством блока оценки амплитуды 118 V для вокализованного звукового сигнала.

В вышеприведенном описании поиска точной высоты тона предполагается, что вся совокупность полос вокализована. Однако, поскольку используемая в системе анализа-синтеза МПВ модель такова, что в ней в той же временной точке на частотной оси имеется невокализованная область, становится необходимым в каждой последовательной полосе осуществлять установление различия вокализованных - невокализованных сигналов.

Оптимальный тон с блока поиска точного тона 116 и данные об амплитуде $|A_{118}|$ с блока оценки амплитуды для вокализованного звука 118 V поступают на блок установления различия вокализованного и невокализованного сигналов 117, в котором выполняется установление различия между вокализованным звуковым сигналом и невокализованным звуковым сигналом в каждой последовательной полосе. Для этого установления различия используется отношение сигнала к шуму (ОСШ).

Между тем, поскольку количество полос, которые разделяются на основании фундаментальной частоты тона, то есть количество гармоник, колеблется в диапазоне примерно от 8 до 63, в зависимости от тона звукового сигнала, аналогичным образом колеблется количество признаков V/UV в каждой последовательной полосе. Таким образом, в настоящем варианте осуществления группируются или разлагаются результаты определения различия V и UV для каждого из заранее установленного количества полос постоянной ширины. В частности, предварительно установленный частотный диапазон, например, равный 0-4000 Гц, включающий звуковой диапазон, разделяется на N_B полос, например 12 полос, и устанавливается различие взвешенного среднего значения величины ОСШ каждой полосы с заранее установленным пороговым значением Th_2 для оценки различия V и UV в каждой последовательной полосе.

На блок оценки амплитуды 118 U для невокализованного звукового сигнала подаются данные частотной области с блока ортогонального преобразования 115, данные точного тона с блока поиска тона 116, данные амплитуды $|A_{118}|$ с блока оценки амплитуды для вокализованного звукового сигнала 118 V и данные установления различия между вокализованными и невокализованными (V/UV) звуками с блока установления различия вокализованного - невокализованного звуковых сигналов 117. И здесь блок оценки амплитуды 118 U для невокализованного звука обнаруживает амплитуду для полосы, определяемой блоком установления различия вокализованного - невокализованного сигналов 117 в виде невокализованного (UV) сигнала посредством осуществления переоценки амплитуды. Блок оценки амплитуды 118 U для невокализованного звука непосредственно выдает входное значение с блока оценки амплитуды вокализованного звука 118 V для полосы, обнаруживаемой в виде вокализованной.

Данные с блока оценки амплитуды 118 U невокализированного звука поступают на блок оценки количества данных 119, который представляет собой преобразователь частоты выборки. Блок преобразования количества данных 119 используется для образования постоянного количества данных, вследствие того, что количество разделенных полос из частотного спектра и количество данных, прежде всего количество амплитудных данных, в различных звуковых тонах различается. То есть, если эффективный частотный диапазон составляет, например, до 3400 кГц, то этот эффективный частотный диапазон разделяется на 8-63 полосы, в зависимости от тона, так что количество данных $m_{MX}+1$ амплитуд $|A_m|$, включающих амплитуду $|A_m|_{UV}$ полосы UV, изменяется в диапазоне от 8 до 63. Таким образом, блок преобразования количества данных 119 преобразует амплитудные данные с переменным количеством данных $m_{MX}+1$ в постоянное количество данных M, например 44.

Блок преобразования количества данных 119 добавляет к амплитудным данным, соответствующим одному эффективному блоку на частотной оси, такие фиктивные данные, которые интерполируют значения от последних данных в блоке к первым данным в блоке для увеличения количества данных до N_F . Блок преобразования количества данных 119 в этом случае выполняет избыточную дискретизацию типа ограничения ширины полосы с коэффициентом избыточной дискретизации O_s , например, равным 8, для обнаружения O_s - кратного количества амплитудных данных. Это O_s - кратное количество $((m_{MX}+1) \times O_s)$ амплитудных данных линейно интерполируется с целью создания еще большего количества N_M данных, например, 2048 данных. Количество N_M данных прореживается с целью преобразования в заранее установленное постоянное количество M, например, 44 данных.

Данные (амплитудные данные с заранее установленным постоянным количеством M) с блока преобразования количества данных 119 поступают на векторный квантователь 23 с целью обеспечения вектора, имеющего количество данных M, или собираются в вектор, имеющий заранее установленное количество данных, для векторного квантования.

Данные о тоне с блока точного поиска тона 116 поступают через неподвижный контакт a переключателя 27 на входную клемму 28. Этот способ, раскрытый в нашей японской заявке на патент № 5-185325 (1993 г.), состоит из переключения с информации, представляющей характеристическое значение, соответствующее форме сигнала во времени невокализированного сигнала, на информацию о тоне, если вся совокупность полос в блоке невокализирована (UV) и, следовательно, тоновая информация становится ненужной.

Эти данные получают посредством обработки данных N-го количества, например, 256 выборок. Поскольку блок продвигается по временной оси на основе вышеупомянутого кадра из α выборок в качестве элемента, передаваемые данные получают на кадровой основе. То есть, данные о тоне, данные установления различия V-UV и амплитудные данные корректируются в течение периода следования кадра. В качестве данных установления различия V-UV с блока установления различия между V и UV 117 можно использовать данные, количество полос которых уменьшено или сокращено до 12, или использовать данные, определяющие одно или более положения границ между вокализированными (V) и невокализированными (UV) областями во всем частотном диапазоне. В качестве альтернативы, всю совокупность полос можно представить одной из V и UV либо установление различия между V и UV можно выполнять на кадровой основе.

Если обнаруживается, что блок полностью невокализирован (UV), то один блок, например из 256 выборок, можно дополнительно разделить на множество субблоков, каждый из которых состоит из 32 выборок, которые поступают на блок вычисления энергии субблока 126.

Блок вычисления энергии субблока 126 вычисляет пропорцию или отношение средней мощности или среднеквадратичного значения совокупности выборок в блоке, например, 256 выборок, к средней мощности или среднеквадратичному значению каждой выборки в каждом субблоке.

То есть, определяется средняя мощность, например, k -го субблока и средняя мощность одного полного блока и рассчитывается квадратный корень отношения средней мощности всего блока к средней мощности $p(k)$ k -го субблока.

5 Полагают, что определяемое таким образом значение квадратного корня представляет вектор заранее установленного размера для того, чтобы выполнять векторное квантование в векторном квантователе 127, расположенном рядом с блоком вычисления мощности субблока.

10 Векторный квантователь 127 осуществляет 8-мерное 8-разрядное прямое векторное квантование (объем кодового словаря равен 256 выборов). Выходной показатель UV- E этого векторного квантователя, то есть код, представляющий вектор, подается на неподвижную клемму \underline{v} переключателя 27. На неподвижную клемму \underline{a} переключателя 27 поступают данные о тоне с блока точного определения тона 116, тогда как выходной сигнал переключателя 27 поступает на выходную клемму 28.

15 Управление работой переключателя 27 осуществляется выходным сигналом установления различия с блока установления различия вокализованного-невокализованного сигналов 117, так что неподвижный контакт переключателя 27 устанавливается на неподвижные контакты \underline{a} и \underline{v} , когда обнаруживается, что, по меньшей мере, одна из полос в блоке вокализована (V), и когда обнаруживается, что вся совокупность полос вокализована, соответственно.

20 Таким образом, выходные сигналы векторного квантования нормализованных на субблочной основе среднеквадратичных значений передаются посредством введения в интервалы, по существу используемые для передачи тоновой информации. То есть, если обнаруживается, что вся совокупность полос в блоке невокализована (UV), то информация о тоне оказывается ненужной, поэтому, если, и только если, обнаруживается, что признаки установления различия V-UV являются полностью невокализованными, вместо информации о тоне передается указатель выходного сигнала векторного квантования VU-E.

30 Далее приводится описание со ссылкой на фиг.3 взвешиваемого векторного квантования спектральной огибающей (A_m) в векторном квантователе 23.

Векторный квантователь 23 представляет собой 2-каскадную α -мерную, например, 44-мерную, конфигурацию.

35 То есть, сумма выходных векторов кодового словаря векторного квантования, который является 44-мерным и имеет объем кодового словаря, равный 32, умножается на коэффициент усиления g_i и полученное произведение используется в качестве квантованного значения 44-мерного вектора \underline{x} спектральной огибающей. На фиг.5 позициями СВ0 и GB1 показаны кодовые словари двух форм, выходными векторами которых являются \underline{s}_{0i} и \underline{s}_{1j} соответственно, где $0 \leq i$ и $j \leq 31$. Выходным сигналом кодового словаря СВ g коэффициента усиления является g_e , который представляет скалярную величину, где $0 \leq e \leq 31$. Конечное выходное значение становится равным $g_i (\underline{s}_{0i} + \underline{s}_{1j})$.

40 Спектральная огибающая A_m , получаемая при анализе многополосного возбуждения (МПВ) остаточных величин кодирования с линейным предсказанием (КЛП) и преобразуемая в заранее установленный размер, устанавливается на значение \underline{x} . Критическим является способ эффективного квантования \underline{x} .

45 Энергия ошибок квантования определяется следующим выражением:

$$E = \left\| W \left\{ Hx - Hg_e (\underline{s}_{0i} + \underline{s}_{1j}) \right\} \right\|^2 =$$

$$50 = \left\| WH \left\{ x - g_e (\underline{s}_{0i} + \underline{s}_{1j}) \right\} \right\|^2,$$

где H и W представляют, соответственно, место для характеристик на частотной оси синтезирующего фильтра КЛП и матрицу для взвешивания, представляющую характеристики взвешивания слухового восприятия на частотной оси.

Энергию ошибки квантования обнаруживают посредством выборки соответствующих α -мерных, например, 44-мерных, точек из частотных характеристик по формуле:

$$H(z) = \frac{1}{1 + \sum_{i=1}^P \alpha_i z^{-i}},$$

где α_i при $1 \leq i \leq P$ представляет α -параметры, получаемые посредством анализа КЛП текущего кадра.

Для расчета O_s заполняются после 1, $\alpha_1, \alpha_2, \dots, \alpha_P, 0, 0, \dots, 0$ с целью обеспечения, например, 256-точечных данных. После этого выполняется 256-точечное быстрое преобразование Фурье и вычисляются значения $(r_e^2 + I_m^2)^{1/2}$ для точек, соответствующих периоду $0-\pi$. Далее определяются обратные величины полученных в результате расчета значений $(r_e^2 + I_m^2)^{1/2}$ и прореживаются, например, до 44 точек.

Матрица, диагональные элементы которой соответствуют этим обратным величинам, определяется следующим образом:

$$H = \begin{bmatrix} h(1) & & & 0 \\ & h(2) & & \\ & & \dots & \\ 0 & & & h(L) \end{bmatrix}$$

Матрица взвешивания слухового восприятия W определяется следующим образом:

$$W(z) = \frac{1 + \sum_{i=1}^P \alpha_i \lambda_b^i z^{-i}}{i + \sum_{i=1}^P \alpha_i \lambda_a^i z^{-i}},$$

где α_i представляет результат анализа КЛП выходных данных, а λ_a, λ_b являются постоянными величинами, например, такими, как $\lambda_a=0,4, \lambda_b=0,9$.

Матрицу W можно определить из частотных характеристик уравнения (3). В качестве примера обеспечиваются 1, $\alpha_1, \lambda_b, \alpha_2 \lambda_b^2, \dots, \alpha_P \lambda_b^P, 0, 0, \dots, 0$ целью получения 256-точечных данных, для которых используется быстрое преобразование Фурье с целью определения $(r_e^2[i] + I_m^2[i])^{1/2}$, где $0 \leq i \leq 128$. Затем обеспечиваются $1, \alpha_1, \lambda_a, \alpha_2 \lambda_a^2, \dots, \alpha_P \lambda_a^P, 0, 0, \dots, 0$ и вычисляются частотные характеристики знаменателя 256-точечным быстрым преобразованием Фурье по 128 точкам для области $0-\pi$. В результате этого получаются значения $(r_e'^2[i] + I_m'^2[i])^{1/2}$, где $0 \leq i \leq 128$.

Частотные характеристики вышеприведенного уравнения (3) можно определять посредством уравнения:

$$WQ[i] = \frac{\sqrt{r_e^2[i] + I_m^2[i]}}{\sqrt{r_e'^2[i] + I_m'^2[i]}}$$

где $0 \leq i \leq 128$.

Частотные характеристики определяются следующим способом для соответствующих точек, например, 44-мерного вектора. Хотя для получения более точных результатов

необходимо использовать линейное интерполирование, при подстановке в следующем примере используются значения ближайших точек.

То есть, $\omega[i]=\omega_0[\text{nint}(128i/L)],$

где $0 \leq i \leq L$, а $\text{nint}(x)$ является функцией, которая отражает целое число, ближайшее к x .

5 Что касается величин $H, h(1), h(2), \dots, h(L)$, то они определяются аналогичным способом. То есть

$$10 \quad H = \begin{bmatrix} h(1) & & & 0 \\ & h(2) & & \\ & & \dots & \\ 0 & & & h(L) \end{bmatrix} \quad W = \begin{bmatrix} w(1) & & & 0 \\ & w(2) & & \\ & & \dots & \\ 0 & & & w(L) \end{bmatrix}$$

15 так что

$$20 \quad WH = \begin{bmatrix} h(1)w(1) & & & 0 \\ & h(2)w(2) & & \\ & & \dots & \\ 0 & & & h(L)w(L) \end{bmatrix} \dots \quad (4)$$

25 В качестве видоизмененного варианта осуществления, частотные характеристики можно определять, с целью уменьшения количества операций быстрого преобразования Фурье, после первого определения $H(z) W(z)$.

То есть

$$30 \quad H(z) W(z) = \frac{1}{1 + \sum_{i=1}^P \alpha_i z^{-i}} \cdot \frac{1 + \sum_{i=1}^P \alpha_i \lambda_i^i z^{-i}}{1 + \sum_{i=1}^P \alpha_i \lambda_i^i z^{-i}} \quad (5)$$

Знаменатель уравнения (5) раскладывается следующим образом:

$$35 \quad \left(1 + \sum_{i=1}^P \alpha_i z^{-i}\right) \left(1 + \sum_{i=1}^P \alpha_i \lambda_i^i z^{-i}\right) = 1 + \sum_{i=1}^{2P} \beta_i z^{-i}$$

40 Посредством установки $1, \beta_1, \beta_2, \dots, \beta_{2P}, 0, 0, \dots, 0$, образуют, например, 256-точечные данные. Затем выполняют 256-точечное быстрое преобразование Фурье с целью обеспечения частотных характеристик амплитуд, так что

$$\text{rms}[i] = \sqrt{\text{re}^{i^2}[i] + \text{Im}^{i^2}[i]},$$

где $0 \leq i \leq 128$. Отсюда выполняется следующее уравнение:

$$45 \quad \mathbb{W}h_0[i] = \frac{\sqrt{\text{re}^2[i] + \text{Im}^2[i]}}{\sqrt{\text{re}^{i^2}[i] + \text{Im}^{i^2}[i]}}$$

где $0 \leq i \leq 128$.

50 Это значение определяется для каждой из соответствующих точек α -мерного вектора. Если количество точек БПФ незначительное, то необходимо использовать линейное интерполирование. Однако, в этом случае используются близкие значения. То есть

$$wh [i] = wh_0 \left[n \operatorname{int} \left(\frac{128}{L} \right) \cdot i \right]$$

где $1 \leq i \leq L$.

5 Матрица W , имеющая эти самые близкие значения в качестве диагональных элементов, определяется следующим выражением:

$$10 \quad W^{-1} = \begin{bmatrix} wh(1) & & & 0 \\ & wh(2) & & \\ & & \dots & \\ 0 & & & wh(L) \end{bmatrix}$$

15 Вышеприведенное уравнение (6) представляет такую же матрицу, как и уравнение (4). Используя эту матрицу, то есть частотные характеристики взвешивающего синтезирующего фильтра, уравнение (1) можно переписать следующим образом:

$$E = \left\| W^{-1} (\underline{x} - g e (\underline{s}_{oc} + \underline{s}_{lk})) \right\|^2 \quad (7)$$

20 Далее приводится описание способа обучения кодового словаря форм и кодового словаря усиления.

Во-первых, для всех кадров, которые выбирают кодовый вектор \underline{s}_{oc} соответствующий СВО, минимизируется ожидаемое значение искажения. Если имеются M таких кадров, для минимизирования достаточно:

$$25 \quad J = \frac{1}{M} \sum_{k=1}^M \left\| W'_k (\underline{x}_k - g_k (\underline{s}_{oc} + \underline{s}_{lk})) \right\|^2 \quad (8)$$

В этом уравнении (8) W'_k , \underline{x}_k , g_k и \underline{s}_{lk} обозначают вес k -го кадра,

30 входной сигнал k -го кадра, коэффициент усиления k -го кадра и выходной сигнал кодового словаря СВИ для k -го кадра, соответственно.

Для минимизирования, уравнение (8) можно записать следующим образом

$$35 \quad J = \frac{1}{M} \sum_{k=1}^M \left\{ (\underline{x}_k^T - g_k (\underline{s}_{oc}^T + \underline{s}_{lk}^T)) W_k'^T W'_k (\underline{x}_k - g_k (\underline{s}_{oc} + \underline{s}_{lk})) \right\} =$$

$$= \frac{1}{M} \sum_{k=1}^M \left\{ \underline{x}_k^T W_k'^T W'_k \underline{x}_k - 2g_k (\underline{s}_{oc}^T + \underline{s}_{lk}^T) W_k'^T W'_k \underline{x}_k + g_k^2 (\underline{s}_{oc}^T + \underline{s}_{lk}^T) W_k'^T W'_k (\underline{s}_{oc} + \underline{s}_{lk}) \right\} =$$

$$40 \quad = \frac{1}{M} \sum_{k=1}^M \left\{ \underline{x}_k^T W_k'^T W'_k \underline{x}_k - 2g_k (\underline{s}_{oc}^T + \underline{s}_{lk}^T) W_k'^T W'_k \underline{x}_k + g_k^2 \underline{s}_{oc}^T W_k'^T W'_k \underline{s}_{oc} + 2g_k^2 \underline{s}_{oc}^T W_k'^T W'_k \underline{s}_{lk} + \right.$$

$$\left. + g_k^2 \underline{s}_{lk}^T W_k'^T W'_k \underline{s}_{lk} \right\} \quad (9)$$

поэтому

$$45 \quad \frac{\partial J}{\partial \underline{s}_{oc}} = \frac{1}{M} \sum_{k=1}^M \left\{ -2g_k W_k'^T W'_k \underline{x}_k + 2g_k^2 W_k'^T W'_k \underline{s}_{oc} + 2g_k^2 W_k'^T W'_k \underline{s}_{lk} \right\} = 0 \quad (10)$$

$$50 \quad \sum_{k=1}^M \left\{ g_k W_k'^T W'_k \underline{x}_k - g_k^2 W_k'^T W'_k \underline{s}_{lk} \right\} = \sum_{k=1}^M g_k^2 W_k'^T W'_k \underline{s}_{oc}$$

отсюда

$$\underline{S}_{0c} = \left\{ \sum_{k=1}^M g_k^2 \underline{W}_k' \underline{W}_k \right\}^{-1} \cdot \left\{ \sum_{k=1}^M g_k \underline{W}_k' \underline{W}_k (\underline{x}_k - g_k \underline{S}_{1k}) \right\} \quad (11)$$

где $\{\}^{-1}$ означает обратную матрицу, а \underline{W}_k' означает транспонированную матрицу \underline{W}_k .

Далее рассмотрим оптимизирование в отношении коэффициента усиления.

Ожидаемое значение J_g искажения для k -го кадра, выбирающего кодовое слово g_c коэффициента усиления, определяется следующим образом.

Решая уравнение

$$J_g = \frac{1}{M} \sum_{k=1}^M \left\| \underline{W}_k' (\underline{x}_k - g_c (\underline{S}_{0k} + \underline{S}_{1k})) \right\|^2,$$

получим

$$= \frac{1}{M} \sum_{k=1}^M \left\{ \underline{x}_k^T \underline{W}_k' \underline{W}_k \underline{x}_k - 2g_c \underline{x}_k^T \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k}) + g_c^2 (\underline{S}_{0k}^T + \underline{S}_{1k}^T) \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k}) \right\}$$

$$\frac{\partial J_g}{\partial g_c} = \frac{1}{M} \sum_{k=1}^M \left\{ -2 \underline{x}_k^T \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k}) + 2g_c (\underline{S}_{0k}^T + \underline{S}_{1k}^T) \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k}) \right\} = 0$$

$$\sum_{k=1}^M \underline{x}_k^T \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k}) = \sum_{k=1}^M g_c (\underline{S}_{0k}^T + \underline{S}_{1k}^T) \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k})$$

$$g_c = \frac{\sum_{k=1}^M \underline{x}_k^T \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k})}{\sum_{k=1}^M (\underline{S}_{0k}^T + \underline{S}_{1k}^T) \underline{W}_k' \underline{W}_k (\underline{S}_{0k} + \underline{S}_{1k})} \quad (12)$$

Вышеприведенные уравнения дают оптимальное центроидное условие для формы \underline{S}_{0i} , \underline{S}_{1i} и коэффициента усиления g_i , где $0 \leq i \leq 31$, то есть оптимальный выходной сигнал декодирования. Оптимальный выходной сигнал декодирования для \underline{S}_{1i} можно определить таким же образом, как и в случае для \underline{S}_{0i} .

Далее рассмотрим оптимальное условие кодирования (ближайшее соседнее условие).

Формы \underline{S}_{0i} , \underline{S}_{1i} которые минимизируют уравнение (7) для измерения искажения, то есть

$$E = \left\| \underline{W}' (\underline{x} - g_e (\underline{S}_{0i} + \underline{S}_{1j})) \right\|^2,$$

определяются каждый раз, когда получаются входной сигнал \underline{x} и весовая матрица \underline{W} ,

то есть для каждого кадра.

По своему существу, E необходимо определять для всех

сочетаний g_e ($0 \leq e \leq 31$), \underline{S}_{0i} ($0 \leq i \leq 31$) и \underline{S}_{1j} ($0 \leq j \leq 31$), то есть $32 \times 32 \times 32$ сочетаний, методом

круговой системы, с целью определения набора g_e , \underline{S}_{0i} , \underline{S}_{1j} , который дает последнее

значение E . Однако, поскольку это ведет к объемистому количеству арифметических

операций, блок кодирования 2 выполняет последовательный поиск формы и коэффициента усиления. Поиск по круговой системе необходимо выполнять для $32 \times 32 = 1024$ сочетаний

\underline{S}_{0i} , \underline{S}_{1j} . В последующем объяснении $\underline{S}_{0i} + \underline{S}_{1j}$ для простоты записываются в виде \underline{S}_m .

Вышеприведенное уравнение можно записать в виде $E = \left\| \underline{W}' (\underline{x} - g_e \underline{S}_m) \right\|^2$. Для

дальнейшего упрощения, обозначив $\underline{x}_w = \underline{W}' \underline{x}$ и $\underline{S}_w = \underline{W}' \underline{S}_m$, получим

$$E = \left\| \underline{x}_w - g_e \underline{S}_w \right\|^2 \quad (13)$$

$$E = \|\underline{x}_w\|^2 + \|\underline{s}_w\|^2 \left(g_1 - \frac{\underline{x}_w^T \cdot \underline{s}_w}{\|\underline{s}_w\|^2} \right) - \frac{(\underline{x}_w^T \cdot \underline{s}_w)^2}{\|\underline{s}_w\|^2} \quad (14)$$

5 Таким образом, полагая, что для g_e обеспечивается достаточная точность, поиск можно выполнять в два этапа:

1) поиск \underline{s}_w , который доводится до максимума

$$10 \frac{(\underline{x}_w^T \cdot \underline{s}_w)^2}{\|\underline{s}_w\|^2}$$

и 2) поиск g_e , который оказывается ближе всего к

$$15 \frac{\underline{x}_w^T \cdot \underline{s}_w}{\|\underline{s}_w\|^2}$$

Если вышеприведенные уравнения переписать, используя первоначальное представление, поиск можно выполнять в два этапа:

1)' поиск группы \underline{s}_{0i} , \underline{s}_{1j} , которая доводит до максимума

$$20 \frac{\underline{x}_w^T \cdot T_w \cdot (\underline{s}_{0i} + \underline{s}_{1j})}{\|T_w \cdot (\underline{s}_{0i} + \underline{s}_{1j})\|^2} \quad (15)$$

и 2)' поиск g_i , ближайшего к

$$25 \frac{\underline{x}_w^T \cdot T_w \cdot (\underline{s}_{0i} + \underline{s}_{1j})}{\|T_w \cdot (\underline{s}_{0i} + \underline{s}_{1j})\|^2}$$

Уравнение (15) дает оптимальное условие кодирования (условие ближайшей соседней записи).

30 Используя центроидное условие уравнений (11) и (12) и условие уравнения (15), можно одновременно тренировать кодовые словари СВ0, СВ1 и СВг посредством обобщенного алгоритма Ллойда (ОАЛ).

Рассматривая фиг.3, отметим, что векторный квантователь 23 подсоединен через переключатель 24 к кодовому словарю для вокализованного звукового сигнала 25 V и к словарю для невокализованного звука 25 U. Управляя переключением переключателя 24 в зависимости от выходного сигнала установления различия V-UV со схемы кодирования гармоник и шума 22, выполняется векторное квантование вокализованного звука и невокализованного звука, используя кодовый словарь для вокализованного звука 25 V и кодовый словарь для невокализованного звука 25 U, соответственно.

Причина переключения кодовых словарей в зависимости от оценки а отношении вокализованного звука (V) и невокализованного звука (UV) состоит в том, что поскольку взвешиваемое усреднение параметров w_k^i и g_e выполняется при вычислении новых центроидов, соответствующих уравнениям (11), (12), нежелательно определять w_k^i и g_e значения которых существенно отличается.

45 Между тем, в блоке кодирования 2 используется w' , деленный на норму входного сигнала \underline{x} . То есть, при продвижении вперед во время обработки уравнений (11), (12) и (15), вместо w' подставляют $w' / \|\underline{x}\|$.

При переключении между кодовыми словарями в зависимости от установления различия V и UV аналогичным образом распределяются обучающие данные с целью подготовки из соответствующих обучающих данных кодового словаря для вокализованного звука и кодового словаря для невокализованного звука.

Для уменьшения количества двоичных разрядов в V-UV в блоке кодирования 2 используется однопольное возбуждение (ОПВ), и данный кадр считается

вокализированным (V) кадром и невокализированным кадром (UV), если отношение V превышает 50 % и противоположное соотношение, соответственно.

На фиг.6 и 7 показаны средние значения $w' / \|\underline{x}\|$ входного сигнала \underline{x} и среднего значения w' весового коэффициента для вокализированного звука, для невокализированного звука и для сочетания вокализированного и невокализированного звуков, то есть без учета различия между вокализированным и невокализированным звуками.

На фиг.6 видно, что распределение энергии самого \underline{x} на частотной оси не сильно отличается при V и UV, хотя среднее значение коэффициента усиления ($\|\underline{x}\|$) сильно

различается между V и UV. Однако, из фиг.7 видно, что форма весового коэффициента различается между V и UV и весовой коэффициент такой, что он увеличивает назначение двоичных разрядов для низкого диапазона в случае V по сравнению с UV. Это объясняет возможность разработки кодового словаря с более высокими характеристиками посредством разделения обучения для V и UV.

На фиг.8 показан способ обучения для трех примеров, то есть для вокализированного звука (V), невокализированного звука (UV) и для объединенных вместе вокализированного и невокализированного звуков. То есть, кривые \underline{a} , \underline{b} и \underline{c} на фиг.8 устанавливают способ обучения только V, только UV и объединенных величин V и UV, причем конечные значения \underline{a} , \underline{b} и \underline{c} равны 3, 72, 7,011 и 6,25, соответственно.

На фиг.8 видно, что разделение обучения кодового словаря для V и кодового словаря для UV ведет к уменьшенному ожидаемому значению искажения выходного сигнала. Хотя состояние ожидаемого значения несколько ухудшается в случае кривой \underline{b} только для UV, ожидаемое значение в целом улучшается, поскольку область для V длиннее области для UV. В качестве примера частоты появления V и UV, полученные в результате измерений значения длин областей только для V и только для UV составляют 0,538 и 0,462 для длины обучающих данных 1. Таким образом, из конечных значений кривых \underline{a} , \underline{b} фиг.8 ожидаемое значение суммарного искажения определяется величиной:

$$3,72 \times 0,538 + 7,011 \times 0,462 = 5,24,$$

которая представляет улучшение примерно на 0,76 дБ по сравнению с ожидаемым значением, равным 6,25 для обучения объединенных вместе V и UV.

Исходя из способа тренировки, улучшение ожидаемого значения составляет порядка 0,76 дБ. Однако обнаружили, что если обрабатываются выборки речи четырех участников дискуссии из числа мужчин и четырех участников дискуссии из числа женщин вне обучающей группы с целью обнаружения ОСШ для случая, при котором квантование не выполняется, разделение на V и UV ведет к улучшению сегментного ОСШ порядка 1,3 дБ. Причина этого, вероятно, состоит в том, что отношение для V значительно выше отношения для UV.

Следует отметить, что хотя весовой коэффициент w' , используемый для взвешивания слухового восприятия при векторном квантовании векторным квантователем 23, как описано вышеприведенным уравнением (6), посредством определения текущего весового коэффициента w' , учитывающего прошлый w' , можно определить весовой коэффициент w' , учитывающий временное маскирующее действие.

Что касается элементов $wh(1), wh(2), \dots, wh(L)$ в вышеприведенном уравнении (6), то рассчитанные в момент времени n эти элементы, то есть, рассчитанные для n -го кадра, они обозначаются элементами $wh_n(1), wh_n(2), \dots, wh_n(L)$.

Учитывающий предыдущее значение в момент времени n весовой коэффициент определяется величиной $A_n(i)$, где $1 \leq i \leq L$. В этом случае

$$A_n(i) = \lambda A_{n-1}(i) + (1-\lambda)wh(i) \text{ для } wh_n(i) \leq A_{n-1}(i) = wh_n(i)$$

$$\text{для } wh_n(i) > A_{n-1}(i),$$

где λ можно устанавливать таким образом, что, например, $\lambda=0,2$. $A_n(i)$, где $1 \leq i \leq L$, можно использовать в качестве диагональных элементов матрицы, которая используется в качестве вышеуказанных весовых коэффициентов.

Возвращаясь к фиг.1, отметим, что здесь представлен блок вычисления

видоизмененных параметров кодирования 3. Устройство воспроизведения речевых сигналов 1 видоизменяет параметры кодирования, выводимые с блока кодирования 2 с определенной скоростью, посредством блока вычисления видоизмененных параметров кодирования 3, предназначенного для вычисления видоизмененных параметров кодирования, и декодирует видоизмененные параметры кодирования посредством блока декодирования с целью воспроизведения содержаний со сплошной записью со скоростью, в два раза выше скорости в реальном масштабе времени. Поскольку высота тона и фонема остаются неизменными, несмотря на более высокую скорость воспроизведения, записанное содержание можно слышать, даже если его воспроизводить на повышенной скорости.

Поскольку параметры кодирования видоизменены по скорости, блок вычисления видоизмененных параметров кодирования 3 не нуждается в обработке после декодирования и вывода сигналов и способен легко согласовываться с различными фиксированными скоростями при аналогичном алгоритме.

Рассматривая алгоритмы на фиг.9 и 11, отметим, что здесь приводится подробное описание работы блока вычисления видоизмененных параметров кодирования 3 устройства воспроизведения речевых сигналов 1. Как описывалось со ссылкой на фиг.2, блок вычисления видоизмененных параметров кодирования 3 состоит из схемы изменения периода 4 и схемы интерполирования 5.

Вначале, на этапе S1 фиг.9 на схему изменения периода 4 поступают через входные клеммы 15, 28, 29 и 26 параметры кодирования, такие как ЛСП, высота тона, V-UV и A_m . Высота тона устанавливается на величину $P_{ch}[n]$, V-UV устанавливается на $vu_v[n]$, A_m устанавливается на $a_m[n][e]$ и ЛСП устанавливается на $l_{sp}[n][i]$. Вычисленные в конечном итоге блоком вычисления видоизмененных параметров кодирования видоизмененные параметры кодирования устанавливаются на значения $mod_P_{ch}[m]$, $mod_vu_v[m]$, $mod_a_m[m][l]$ и $mod_l_{sp}[m][i]$, где l обозначает количество гармоник, i обеспечивает число порядка ЛСП, а n и m соответствует номерам кадров, соответствующим, в свою очередь, индексу временной оси до и после преобразования временной оси, соответственно. Между тем, $0 \leq n < N_1$ и $0 \leq m < N_2$, где каждый из элементов n и m представляет указатель кадра при кадровом интервале, например, равном 20 мс.

Как описывалось выше, l обозначает количество гармоник. Вышеуказанные условия можно выполнять после восстановления количества гармоник до величины $a_m[n][e]$, соответствующей реальному количеству гармоник, или можно также выполнять в состоянии $a_m[n][e]$ ($e=0-43$). То есть преобразование данных количества можно выполнять до или после декодирования декодирующим устройством.

На этапе S2 схема изменения периода 4 устанавливает количество кадров, соответствующих первоначальной длительности времени, на N_1 , устанавливая в то же время количество кадров, соответствующих длительности времени после изменения, на N_2 . Затем на этапе S3 схема изменения периода 4 сжимает временную ось скорости N_1 до скорости N_2 . То есть коэффициент сжатия временной оси spd схемой изменения периода 4 определяется отношением N_2/N_1 .

Далее на этапе S 4 схема интерполирования 5 устанавливает m , соответствующую количеству кадров, в свою очередь, соответствующему показателю временной оси после преобразования временной оси, равную 2.

Далее на этапе S 5 схема интерполирования 4 определяет два кадра fr_0 и fr_1 и различия "левого" и "правого" между двумя кадрами fr_0 и fr_1 um/spd . Если параметры кодирования P_{ch} , vu_v , a_m и l_{sp} обозначить звездочкой (*), то $mod_*[m]$ можно выразить в виде общей формулы

$$mod_*[m] = *[m/spd],$$

где $0 \leq m < N_2$. Однако, поскольку отношение m/spd является не целым числом, видоизмененный параметр кодирования для m/spd создается посредством интерполирования из двух кадров $fr_0=Lm/spd$ и $fr_1=f_0+1$. Следует отметить, что между

кадром fr_0 , m/spd и кадром fr_1 существует связь, как показано на фиг.10, то есть сохраняется связь, определяемая выражениями

левый = $m/spdL$,

правый = $fr_1 - m/spd$.

5 Параметр кодирования для m/spd на фиг.10, то есть видеоизмененный параметр кодирования, создается посредством интерполирования, как показано на этапе 6. Видеоизмененный параметр кодирования можно определить просто линейным интерполированием в виде:

$$mod_ * [m] = * [fr_0] \times \text{правый} + [fr_1] \times \text{левый} .$$

10 Однако, если в случае интерполирования между fr_0 и fr_1 эти два кадра различаются в отношении V-UV; то есть если один из двух кадров является V, а другой UV, то нельзя применять вышеприведенную общую формулу. Следовательно, схема интерполирования 5 изменяет способ определения параметров кодирования в связи с вокализованными и невокализованными характеристиками этих двух кадров fr_0 и fr_1 , как показано на этапе 15 S11 следующей фиг.11.

Вначале следует принять решение в отношении того, являются ли два кадра fr_0 и fr_1 вокализованными (V) или невокализованными (UV). Если обнаруживается, что оба кадра fr_0 и fr_1 вокализованы (V), программа переходит к этапу S 12, где все параметры линейно интерполируются, а видеоизмененные параметры кодирования представляются 20 следующим образом:

$$mod_P_{ch}[m] = P_{ch}[fr_0] \times \text{правый} + P_{ch}[fr_1] \times \text{левый}$$

$$mod_a_m[m][l] = a_m[fr_0][l] \times \text{правый} + a_m[fr_1][l] \times \text{левый}$$

где $0 \leq l \leq L$. Следует отметить, что L обозначает максимально возможное количество, 25 которое можно принять в качестве гармоник, и что "0" заполняется в $a_m[n][l]$, где отсутствуют гармоники. Если количество гармоник различается в кадрах fr_0 и fr_1 , то считается, что при вышеописанном интерполировании значение эквивалентных гармоник равно нулю. Перед прохождением через блок преобразования количества данных, количество L может быть постоянным, например, $L=43$ при $0 \leq l < L$.

30 Кроме того, видеоизмененные параметры кодирования также воспроизводятся следующим образом:

$$mod_l_{sp}[m][i] = l_{sp}[fr_0][i] \times \text{правый} + l_{sp}[fr_1][i] \times \text{левый}$$

где $0 \leq i \leq 1$, и l показывает количество порядков ЛСП и обычно равно 10, и

$$mod_vu[m] = 1$$

35 Следует понимать, что при установлении различия V-UV, 1 и 0 показывают вокализованный (V) и невокализованный (UV) кадры, соответственно.

Если на этапе S11 принимается решение, что ни один из двух кадров fr_0 и fr_1 не вокализуется (V), дается оценка, аналогичная оценке, полученной на этапе S 13, то 40 есть оценка в отношении того, невокализованы (UV) ли оба кадра fr_0 и fr_1 . Если результат оценки положительный (ДА), то есть если оба кадра невокализируются (UV), схема интерполирования 5 устанавливает P_{ch} на постоянную величину и определяет a_m и l_{sp} посредством линейного интерполирования следующим образом

$$mod_P_{ch}[m] = \text{MaxPitch} \quad (\text{максимальный тон})$$

45 для фиксирования значения тона на постоянной величине, например, на максимальной величине для невокализованного звука, равной $\text{MaxPitch}=148$;

$$mod_a_m[m][l] = a_m[fr_0][l] \times \text{правый} + a_m[fr_1][l] \times \text{левый}$$

где : $0 \leq l < \text{MaxPitch}$

$$mod_l_{sp}[m][i] = l_{sp}[fr_0][i] \times \text{правый} + l_{sp}[fr_1][i] \times \text{левый}$$

50 где : $0 \leq i < I$, и

$$mod_vu[m] = 0$$

Если оба кадра fr_0 и fr_1 невокализованы, программа переходит к этапу S15, где

принимается решение в отношении того, вокализируется (V) ли кадр fr_0 и не вокализируется (UV) кадр fr_1 . Если результат оценки положительный (ДА), то есть если кадр fr_0 вокализируется (V), а кадр fr_1 не вокализируется (UV), программа переходит к этапу S16. Если результат оценки отрицательный (НЕТ), то есть если кадр fr_0 не вокализируется (UV), а кадр fr_1 вокализируется (V), программа переходит к этапу S17.

Обработка на следующем этапе S 16 относится к случаям, когда fr_0 и fr_1 отличаются в отношении V-UV, то есть когда один из кадров вокализирован, а другой невокализирован. Это учитывает то, что интерполирование параметров между двумя кадрами fr_0 и fr_1 , отличающимися в отношении V-UV, не имеет значения. В таком случае, используется значение параметра кадра, более близкого к времени m/spd без выполнения интерполирования.

Если кадр fr_0 вокализирован (V), а кадр fr_1 не вокализирован (UV), программа переходит к этапу S 16, где сравниваются друг с другом размеры "левого" ($=m/spd-fr_0$) и "правого" ($=fr_1-m/spd$) кадров, как показано на фиг.10. Это позволяет дать оценку в отношении того, который из кадров fr_0 и fr_1 находится ближе к m/spd . Вычисление видеоизмененных параметров кодирования осуществляется с использованием значений параметров кадра, находящегося ближе к m/spd .

Если результат оценки на этапе S16 положительный (ДА), то это означает, что "правый" размер больше, а отсюда кадр fr_1 находится дальше от m/spd . Таким образом, на этапе S18 определяются видеоизмененные параметры кодирования, используя параметры кадра fr_0 , находящиеся ближе к m/spd следующим образом:

$$\begin{aligned} \text{mod_P}_{ch}[m] &= P_{ch}[fr_0] \\ \text{mod_a}_m[m][l] &= a_m[fr_0][l] \quad (\text{где } 0 \leq l < L) \\ \text{mod_l}_{sp}[m][i] &= l_{sp}[fr_0][i] \quad \left(\text{где } 0 \leq i < L \right) \\ \text{mod_vu}_v[m] &= 1 \end{aligned}$$

Если результат оценки на этапе S16 отрицательный (НЕТ), то размеры "левый" \geq "правого", а отсюда кадр fr_1 ближе к m/spd , поэтому программа переходит к этапу S19, где величина тона доводится до максимума и, используя параметры для кадра fr_1 , устанавливаются видеоизмененные параметры следующим образом:

$$\begin{aligned} \text{mod_P}_{ch}[m] &= \text{MaxPitch} \quad (\text{максимальный тон}) \\ \text{mod_a}_m[m][l] &= a_m[fr_1][l] \quad (\text{где } 0 \leq l \leq \text{MaxPitch} / 2) \\ \text{mod_l}_{sp}[m][i] &= l_{sp}[fr_1][i] \quad (\text{где } 0 \leq i < L) \\ \text{mod_vu}_v[m] &= 0 \end{aligned}$$

Далее, на этапе S17, под действием оценки на этапе S 15, состоящей в том, что два кадра fr_0 и fr_1 являются невокализированным (UV) и вокализированным (V), соответственно, дается оценка, аналогичная оценке на этапе S16. То есть, в этом случае, интерполирование не выполняется, а используются значения параметров кадра, находящегося ближе к времени m/spd .

Если результат оценки на этапе S 17 положительный (ДА), то высота тона доводится до максимума по величине на этапе S20 и, используя параметры более близкого кадра fr_0 для остальных параметров, устанавливаются видеоизмененные параметры кодирования следующим образом:

$$\begin{aligned} \text{mod_P}_{ch}[m] &= \text{MaxPitch} \quad (\text{максимальный тон}) \\ \text{mod_a}_m[m][l] &= a_m[fr_0][l] \quad (\text{где } 0 \leq l \leq \text{MaxPitch}) \\ \text{mod_l}_{sp}[m][i] &= l_{sp}[fr_0][i] \quad (\text{где } 0 \leq i < I) \\ \text{mod_vu}_v[m] &= 0 \end{aligned}$$

Если результат оценки на этапе S17 отрицательный (НЕТ), то, поскольку размеры "левого" \geq "правого", а отсюда кадр fr_1 ближе к m/spd , программа переходит к этапу S 21, где с помощью параметров для кадра fr_1 устанавливаются видеоизмененные параметры кодирования следующим образом:

5 $mod_P_{ch}[m] = MaxPitch$
 $mod_a_m[m][l] = a_m[fr_1][l]$ (где $0 \leq l < MaxPitch / 2$)
 $mod_l_{sp}[m][i] = l_{sp}[fr_1][i]$ (где $0 \leq i < L$)
 10 $mod_vu_v[m] = 0$

Таким образом, схема интерполирования 5 выполняет различные интерполирующие операции на этапе S 6 фиг.9 в зависимости от соотношения вокализованных (V) и невокализованных (UV) характеристик между двумя кадрами fr_0 и fr_1 . После окончания операции интерполирования на этапе S 6, программа переходит к этапу S 7, где
 15 производится приращение параметра m . Действия в соответствии с этапами S 5 и S 6 повторяются, пока значение m не станет равным N_2 .

Кроме того, последовательность краткосрочных среднеквадратичных значений невокализованных (UV) частей обычно используется для управления усилением шума. Однако, этот параметр здесь устанавливается на 1.

20 Работа блока вычисления видеоизмененных параметров кодирования схематически показана на фиг.12. Модель параметров кодирования, извлекаемых каждые 20 мс блоком кодирования 2, показана на фиг.12А. Схема изменения периода 4 блока вычисления видеоизмененных параметров кодирования 3 устанавливает период на 15 мс и выполняет
 сжатие по временной оси, как показано на фиг.12В. Показанные на фиг.12С

25 видеоизмененные параметры кодирования вычисляются посредством интерполирующего устройства, соответствующего установкам V-UV двух кадров fr_0 , fr_1 , как описано выше.

Схема вычисления видеоизмененных параметров кодирования 3 может также изменять на обратную последовательность, в которой выполняются операции схемой изменения
 30 периода 4 и схемой интерполирования, то есть выполнять интерполирование параметров кодирования, показанное на фиг.13А, как показано на фиг.13В, и выполнять сжатие для вычисления видеоизмененных параметров кодирования, как показано на фиг.13С.

Видеоизмененные параметры кодирования со схемы вычисления видеоизмененных параметров кодирования 3 поступают на схему декодирования 6, показанную на фиг.1. Схема декодирования 6 синтезирует гармонические волны и шум на основании
 35 видеоизмененных параметров кодирования и выводит синтезированный звуковой сигнал на выходную клемму 37.

Описание работы схемы декодирования производится со ссылкой на фиг.14 и 15. В целях объяснения предполагается, что поступающие на схему декодирования 6 параметры являются обычными параметрами кодирования.

40 На фиг.14 на клемму 31 поступает выходной сигнал с векторным квантованием линейной спектральной пары (ЛСП), соответствующий выходному сигналу на клемме 15 фиг.3, то есть так называемый указатель.

Этот входной сигнал поступает на инверсный векторный квантователь ЛСП 32 для инверсного векторного квантования с целью выработки данных линейной
 45 спектральной пары (ЛСП), которые затем поступают на схему интерполирования ЛСП 33 для интерполирования ЛСП. Получающиеся интерполированные данные преобразуются схемой преобразования ЛСП в α 32 в α - параметры кодов с линейным предсказанием (КЛП), Эти α - параметры поступают на синтезирующий фильтр 35.

На клемму 41 фиг.14 поступают данные показателя взвешенного кодового слова с векторным квантованием спектральной огибающей (A_m), соответствующей выходному
 50 сигналу на клемме 26 кодирующего устройства, показанного на фиг.3. На клемму 43 поступает информация о тоне с клеммы 28 фиг.3 и данные, показывающие характеристическое качество формы сигнала во времени в блоке UV, тогда как на клемму

46 поступают данные установления различия V-UV с клеммы 29 фиг.3.

Данные с векторным квантованием амплитуды A_m с клеммы 41 поступают на инверсный векторный квантователь 42 для обратного векторного квантования. Получающиеся данные спектральной огибающей поступают на схему синтезирования гармоник и шума или схему синтезирования многополосного возбуждения (МПВ) 45. На схему синтезирования 45 подаются данные с клеммы 43, которая переключается переключателем 44 между данными о тоне и данными, показывающими характеристическое значение формы сигнала для кадра UV в зависимости от данных установления различия V-UV. На схему синтезирования 45 поступают также данные установления различия V-UV с клеммы 46.

Ниже приводится описание со ссылкой на фиг.15 компоновки схемы синтезирования МПВ в качестве иллюстративной компоновки схемы синтезирования 45.

Со схемы синтезирования 45 берутся остаточные данные КЛП, соответствующие выходному сигналу инверсной фильтрующей схемы 21 фиг.3. Полученные таким образом остаточные данные поступают на схему синтезирования 35, где выполняется синтезирование КЛП с целью создания данных формы сигнала по времени, которые фильтруются последующим фильтром 36, так что на выходную клемму 37 выводятся воспроизведенные сигналы формы волны временной области.

Иллюстрируемый пример схемы синтезирования МПВ в качестве примера схемы синтезирования 45 описывается со ссылкой на фиг.15.

На фиг.15 показано, что данные спектральной огибающей с инверсного векторного квантователя 42 фиг.14 фактически данные спектральной огибающей остаточных значений КЛП подаются на входную клемму 131. Данные, поступающие на клеммы 43, 46, такие же, как и данные, показанные на фиг.14. Данные, поступающие на клемму 43, выбираются переключателем 44 таким образом, что данные о тоне и данные, показывающие характеристическое качество формы сигнала UV, поступают на блок синтезирования вокализованного звука 137 и на инверсный векторный квантователь 152, соответственно.

Данные спектральной амплитуды остаточных значений КЛП с клеммы 131 поступают на схему обратного преобразования количества данных 136 для обратного преобразования. Схема обратного преобразования количества данных 136 выполняет обратное преобразование, которое представляет обратную функцию преобразования, выполняемого блоком преобразования количества данных 119. Получающиеся амплитудные данные подаются на блок синтезирования вокализованного звука 137 и на блок синтезирования невокализованного звука 138. Данные о тоне, полученные с клеммы 43 через неподвижную клемму \bar{a} переключателя 44, подаются на блоки синтезирования 137, 138.

На блоки синтезирования 137, 138 поступают также данные установления различия V-UV с клеммы 46.

Блок синтезирования вокализованного звука 137 синтезирует форму волны вокализованного звука временной области, например, посредством синтезирования косинусоидальной или синусоидальной волны, тогда как блоку синтезирования невокализованного звука 138 фильтрует, например, белый шум посредством полосового фильтра с целью синтезирования невокализованной формы сигнала временной области. Вокализованная форма сигнала и невокализованная форма сигнала суммируются вместе сумматором 141 таким образом, чтобы их можно было вывести на выходную клемму 142.

Если в качестве данных установления различия V и UV передается код V и UV, все полосы можно разделить в единственной точке разграничения на вокализованную (V) область и невокализованную (UV) область, и на основании этой точки разграничения можно получить основанные на полосе данные установления различия V-UV. Если количество полос уменьшается на стороне анализа (кодирующего устройства) до постоянного количества, равного 12 полосам, это уменьшение можно аннулировать для обеспечения изменяющегося количества полос при ширине полосы, соответствующей первоначальному тону.

Ниже приводится описание действия синтезирования невокализованного звука блоком

синтезирования невокализированного звука 138.

5 Форма волны сигнала белого шума временной области с генератора белого шума 143 подается на блок взвешивания 144 для взвешивания с использованием соответственной финитной функции, например, взвешивающей функции Хэмминга, с заранее установленной
 10 длительностью, например, равной 256 выборок. Затем взвешенная форма волны сигнала подается на схему краткосрочного преобразования Фурье (КСПФ) 145 для краткосрочного преобразования Фурье с целью создания энергетического спектра частотной области белого шума. Энергетический спектр блока краткосрочного преобразования Фурье 145 подается на блок обработки амплитуды полосы 146, где полагают, что полосы
 15 невокализированы (UV) и умножаются на амплитуду $|A_m|_{UV}$, тогда как полагают, что ширина других полос, представляющих V , устанавливается на 0. На блок обработки амплитуды полосы 146 поступают амплитудные данные, данные о тоне и данные установления различия V и UV .

15 Выходной сигнал блока обработки амплитуды полосы частот 146 подается на блок инверсного краткосрочного преобразования Фурье (КСПФ), где производится инверсное КСПФ, используя в качестве фазы первоначальный белый шум, с целью преобразования в сигналы временной области. Выходной сигнал блока инверсного КСПФ 147 подается через блок формирования распределения мощности 156 и умножитель 157, описываемый ниже,
 20 на блок совмещения и сложения 148, где совмещение и добавление повторяются с соответствующим взвешиванием на временной оси для обеспечения возможности восстановления первичной непрерывной формы сигнала. Таким образом, непрерывная форма волны во временной области создается посредством синтеза. Выходной сигнал блока совмещения и добавления 148 поступает на сумматор 141.

25 Если, по меньшей мере, одна из полос в блоке вокализирована (V), то вышеупомянутая обработка выполняется в соответственном блоке синтезирования 137, 138. Если обнаруживается, что все полосы в блоке невокализированы, подвижный контакт 44 переключателя 44 установлен на неподвижной клемме \underline{v} , так что информация о форме сигнала во времени невокализированного сигнала подается вместо тоновой информации
 30 на блок инверсного векторного квантования 152.

То есть, на блок векторного деквантования 152 поступают данные, соответствующие данным, поступающим с блока векторного квантования 127 на фиг.4. Эти данные подвергаются инверсному векторному квантованию с целью вывода данных для извлечения характеристического качества формы волны невокализированного сигнала.

35 Выходной сигнал блока инверсного КСПФ 147 перед подачей на умножитель 157 подвергается распределению энергии временной области, регулируемому блоком формирования распределения мощности 156. Умножитель 157 перемножает выходной сигнал блока инверсного КСПФ 147 с сигналом, выводимым с блока векторного деквантования 152 через блок сглаживания 153. Быстрые изменения коэффициента усиления, которые кажутся резко выраженными, может подавить блок сглаживания 153.

40 Синтезированный таким образом невокализированный звуковой сигнал снимается с блока синтезирования невокализированного звукового сигнала 138 и подается на сумматор 141, где он складывается с сигналом, поступающим с блока синтезирования вокализированного звукового сигнала 137, так что с выходной клеммы 142 снимаются остаточные сигналы КЛП в качестве синтезированных выходных сигналов МПВ.

45 Эти остаточные сигналы КЛП поступают на синтезирующий фильтр 35 фиг.14 для создания конечного речевого звукового сигнала воспроизведения.

Устройство воспроизведения речевого сигнала 1 заставляет блок вычисления видеоизмененных параметров кодирования 3 производить вычисление видеоизмененных параметров кодирования под управлением контроллера (не показанного) и синтезирует
 50 речевой звуковой сигнал, который представляет компандированный по временной оси первоначальный речевой сигнал с добавлением видеоизмененных параметров кодирования.

В этом случае сигнал $mod_l_{sp}[m][i]$ с блока вычисления видеоизмененных параметров кодирования 3 используется вместо выходного сигнала схемы инверсного векторного

квантователя ЛСП. Видоизмененный параметр

кодирования $\text{mod_l}_{sp}[m][i]$ используется вместо значения, присущего векторному квантованию. Видоизмененный параметр кодирования $\text{mod_l}_{sp}[m][i]$ подается на

5 схему интерполирования ЛСП 33 для интерполирования линейной спектральной пары (ЛСП), а отсюда подается на схему преобразования ЛСП в α 34, где он преобразуется в α - параметр кодов с линейным предсказанием (КЛП), с которой поступает на синтезирующий фильтр 35.

10 С другой стороны, видоизмененный параметр кодирования $\text{mod_a}_m[m][1]$ подается вместо выходного сигнала или входного сигнала схемы преобразования количества данных 136. На клеммы 43, 46 поступают сигналы $\text{mod_F}_h[m]$ и $\text{mod_v}_v[m]$ соответственно.

15 Видоизмененный параметр кодирования $\text{mod_a}_m[m][1]$ поступает на схему синтезирования гармоник и шума 45 в качестве данных спектральной огибающей. На схему синтезирования 45 поступает сигнал $\text{mod_F}_h[m]$ с клеммы 43 через переключатель 44 в зависимости от данных установления различия, тогда как на нее также поступает сигнал $\text{mod_v}_v[m]$ с клеммы 46.

20 Посредством вышеописанной показанной на фиг.15 схемы синтезируются компандированные по временной оси первоначальные речевые сигналы, используя вышеупомянутые видоизмененные параметры кодирования таким образом, чтобы их можно было выводить на выходную клемму 37.

25 Таким образом, устройство воспроизведения речевых сигналов 1 декодирует матрицу видоизмененных параметров кодирования $\text{mod_}^*[m]$ (где $0 \leq m < N_2$) вместо неотъемлемой матрицы $^*[n]$ ($0 \leq n < N_1$). Интервал между кадрами во время декодирования может быть постоянным, обычно равным, например 20 мс. Таким образом, если $N_2 < N_1$ или $N_2 > N_1$, производится сжатие временной оси с увеличением скорости или расширение временной оси, соответственно.

30 Если выполняется изменение временной оси, как описано выше, мгновенный спектр и тон остаются неизменными, так что несмотря на значительное изменение диапазона в пределах $0,5 \leq \text{spd} \leq 2$, едва создается ухудшение.

35 В случае такой системы, поскольку в конечном итоге полученная последовательность параметров декодируется после расположения их в определенном порядке с неотъемлемым интервалом 20 мс, можно легко реализовать произвольное управление скоростью в направлении увеличения или уменьшения. С другой стороны, увеличение и уменьшение скорости можно выполнять посредством такой же обработки без точек перехода.

40 Таким образом, плотно записанное содержимое можно воспроизводить со скоростью, в два раза выше скорости в реальном масштабе времени. Поскольку тон и фонема остаются неизменными, несмотря на увеличенную скорость воспроизведения, плотно записанное содержание можно слышать, если воспроизведение выполняется на более высокой скорости. С другой стороны, что касается речевого кодека, то можно исключить дополнительные, например, арифметические операции после декодирования и выведения сигналов, которые требуются при использовании кодирования с линейным предсказанием возбуждаемого кода (ЛПВК).

45 Хотя блок вычисления видоизмененных параметров кодирования 3 изолируется при вышеописанном первом способе осуществления от блока декодирования 6, блок вычисления 3 можно также обеспечивать в блоке декодирования 6.

50 При вычислении параметров блоком вычисления видоизмененных параметров кодирования 3 в устройстве воспроизведения речевых сигналов 1, интерполирующие операции по параметру a_m выполняют на значении с векторным квантованием или на значении с обратным векторным квантованием.

Далее приводится описание устройства передачи речевых сигналов 50, предназначенного для выполнения соответствующего настоящему изобретению способа

передачи звуковых сигналов. На фиг.16 видно, что устройство передачи речевых сигналов 50 включает в себя передающее устройство 51, предназначенное для разделения входного речевого сигнала исходя из заранее установленных кадров временной области в качестве элементов, и кодирования входного речевого сигнала на кадровой основе для обнаружения параметров кодирования, интерполирования параметров кодирования с целью обнаружения видеоизмененных параметров кодирования и для передачи видеоизмененных параметров кодирования. Устройство передачи речевых сигналов 50 включает в себя также приемное устройство 56, предназначенное для приема видеоизмененных параметров кодирования и для синтезирования гармонических колебаний и шума.

То есть, передающее устройство 51 включает в себя кодирующее устройство 53, предназначенное для разделения входного речевого сигнала исходя из заранее установленных кадров временной области в качестве элементов и кодирования речевого сигнала на кадровой основе для извлечения параметров кодирования, интерполятор 54, предназначенный для интерполирования параметров кодирования с целью определения видеоизмененных параметров кодирования, и блок передачи, предназначенный для передачи видеоизмененных параметров кодирования. Приемное устройство 56 включает в себя блок приема 57, интерполятор 58, предназначенный для интерполирования видеоизмененных параметров кодирования, и блок декодирования 59, предназначенный для синтезирования гармонических колебаний и шума на основании интерполированных параметров для вывода синтезированных речевых сигналов на выходную клемму 60.

Основная работа блока кодирования 53 и блока декодирования 59 аналогична работе таких же блоков в устройстве воспроизведения речевого сигнала 1, и поэтому здесь, для простоты, подробное их описание опускается.

Описание работы передающего устройства 51 производится со ссылкой на представленный на фиг.17 алгоритм, в котором совместно показаны действия кодирования блоком кодирования 53 и интерполирования интерполятором 54.

Блок кодирования 53 извлекает параметры кодирования, состоящие из ЛСП, тона P_{ch} , $V-UV$ и a_m на этапах S31 и S33. В частности ЛСП интерполируется и переупорядочивается интерполятором 54 на этапе S 31 и квантуется на этапе S 32, тогда как тон P_{ch} , $V-UV$ и a_m интерполируются и переупорядочиваются на этапе S 34 и квантуются на этапе S 35. Эти квантованные данные передаются посредством передающего устройства 55 на приемное устройство 56.

Квантованные данные, принимаемые посредством блока приема 57 в приемном устройстве 56, подаются на блок интерполирования 58, где параметры интерполируются и переупорядочиваются на этапе S 36. На этапе S 37 данные синтезируются посредством блока декодирования 59.

Таким образом, для увеличения скорости посредством сжатия временной оси, устройство передачи речевых сигналов 50 интерполирует параметры и изменяет интервал между кадрами параметров во время передачи. Между тем, поскольку воспроизведение выполняется во время приема посредством обнаружения параметров с постоянным интервалом между кадрами, равным 20 мс, алгоритм управления скоростью можно использовать непосредственно для преобразования скорости передачи битов.

То есть, предполагается, что если интерполирование параметров используется для управления скоростью, то это интерполирование выполняется в декодирующем устройстве. Однако, если эта обработка выполняется в кодирующем устройстве, так что данные со сжатой (прореженной) временной осью кодируются и расширяется (интерполируется) временная ось декодирующим устройством, скорость передачи битов можно регулировать по соотношению spr .

Если скорость передачи составляет, например, 1,975 килобитов в секунду, а кодирование выполняется на удвоенной скорости посредством такой установки, что $spr=0,5$, то поскольку кодирование выполняется со скоростью 5 секунд вместо присущей скорости 10 секунд, скорость передачи становится равной $1,975 \times 0,5$ килобитов в секунду.

Кроме того, параметры кодирования, полученные в блоке кодирования 53, показанные

на фиг.18А, интерполируются и переупорядочиваются интерполятором 54 с произвольным интервалом, например, равным 30 мс, как показано на фиг.18В. Затем параметры кодирования интерполируются и переупорядочиваются интерполятором 58 приемного устройства 56 до 20 мс, как показано на фиг.18С, и синтезируются блоком декодирования 59.

Если аналогичную схему обеспечить в декодирующем устройстве, то можно восстанавливать скорость до первоначального значения, хотя речевой звуковой сигнал можно также слышать на высокой или низкой скорости. То есть, устройство регулирования скорости можно использовать в качестве кодера-декодера переменной скорости передачи битов.

Формула изобретения

1. Способ воспроизведения входного речевого сигнала на основании параметров кодирования, получаемых посредством разделения входного речевого сигнала, исходя из заранее установленных кадров на временной оси и посредством кодирования входного речевого сигнала на кадровой основе, содержащий этапы интерполирования параметров кодирования для определения видоизмененных параметров кодирования, связанных с основанными на кадрах временными периодами, и вырабатывания видоизмененного речевого сигнала по временной оси от входного речевого сигнала на основе видоизмененных параметров кодирования.

2. Способ по п. 1, отличающийся тем, что видоизмененный речевой сигнал создают посредством, по меньшей мере, синтеза гармонических волн на основании видоизмененных параметров кодирования для воспроизведения речевых сигналов.

3. Способ по п.2, отличающийся тем, что период вывода параметров кодирования изменяют посредством сжатия-расширения временной оси параметров кодирования, получаемых в каждом заранее установленном кадре до или после этапа интерполирования.

4. Способ по п. 1, отличающийся тем, что интерполирование параметров кодирования выполняют линейным интерполированием параметров линейной спектральной пары, высоты тона и остаточной спектральной огибающей, содержащихся в параметрах кодирования.

5. Способ по п.1, отличающийся тем, что параметры кодирования определяют посредством представления остаточных значений краткосрочных предсказаний входного речевого сигнала в качестве синтезированной гармонической волны и шума и посредством кодирования частотной спектральной информации каждого из синтезированных гармонических волн и шума.

6. Устройство воспроизведения речевого сигнала, в котором входной речевой сигнал восстанавливают на основании параметров кодирования, получаемых посредством разделения входного речевого сигнала, исходя из заранее установленных кадров на временной оси и кодирования основанного на кадрах входного речевого сигнала для обнаружения параметров кодирования, содержащее средство интерполирования параметров кодирования, предназначенное для определения видоизмененных параметров кодирования, связанных с основанными на кадрах временными периодами, и средство вырабатывания видоизмененного речевого сигнала, предназначенное для выработки видоизмененного речевого сигнала, отличающегося по временной оси от входного речевого сигнала, на основе видоизмененных параметров кодирования.

7. Устройство по п.6, отличающееся тем, что средство вырабатывания видоизмененного речевого сигнала выполнено с возможностью, по меньшей мере, синтеза гармонической волны в соответствии с видоизмененными параметрами кодирования.

8. Устройство по п.7, отличающееся тем, что дополнительно содержит средство изменения периода, предназначенное для сжатия-расширения временной оси параметров кодирования, получаемых в каждом заранее установленном кадре, для изменения периода вывода параметров кодирования, и установленное до или после средства

интерполирования.

9. Устройство по п.6, отличающееся тем, что средство интерполирования параметров кодирования выполнено с возможностью линейного интерполирования параметров линейной спектральной пары, высоты тона и остаточной спектральной огибающей, содержащихся в параметрах кодирования.

10. Устройство по п.6, отличающееся тем, что параметры кодирования определяют посредством представления остаточных величин краткосрочного предсказания входного речевого сигнала в качестве синтезированной гармонической волны и шума и посредством кодирования частотно-спектральной информации каждого из синтезированных гармонических волн и шума.

11. Способ передачи входных речевых сигналов, заключающийся в том, что параметры кодирования получают посредством разделения входного речевого сигнала, исходя из заранее установленных кадров на временной оси в качестве элементов и посредством кодирования разделенного входного речевого сигнала на кадровой основе, причем определенные таким образом параметры кодирования интерполируют для определения видоизмененных параметров кодирования, связанных с основанными на кадрах временными периодами, и видоизмененные параметры кодирования передают.

12. Способ по п. 11, отличающийся тем, что параметры кодирования определяют посредством представления остаточных величин краткосрочных предсказаний входного речевого сигнала в качестве синтезированной гармонической волны и шума и посредством кодирования частотно-спектральной огибающей каждого из синтезированных гармонической волны и шума.

25

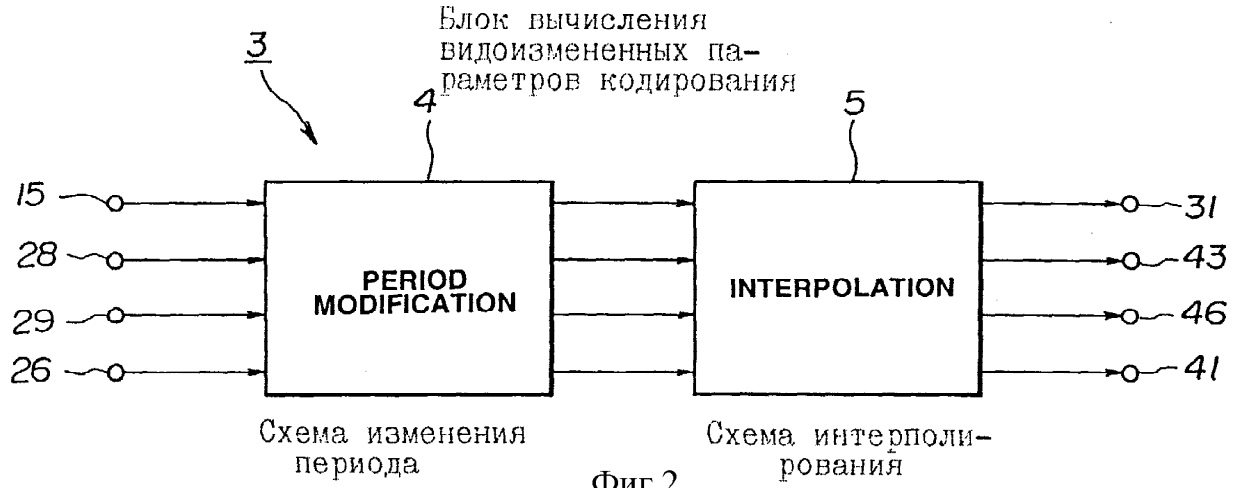
30

35

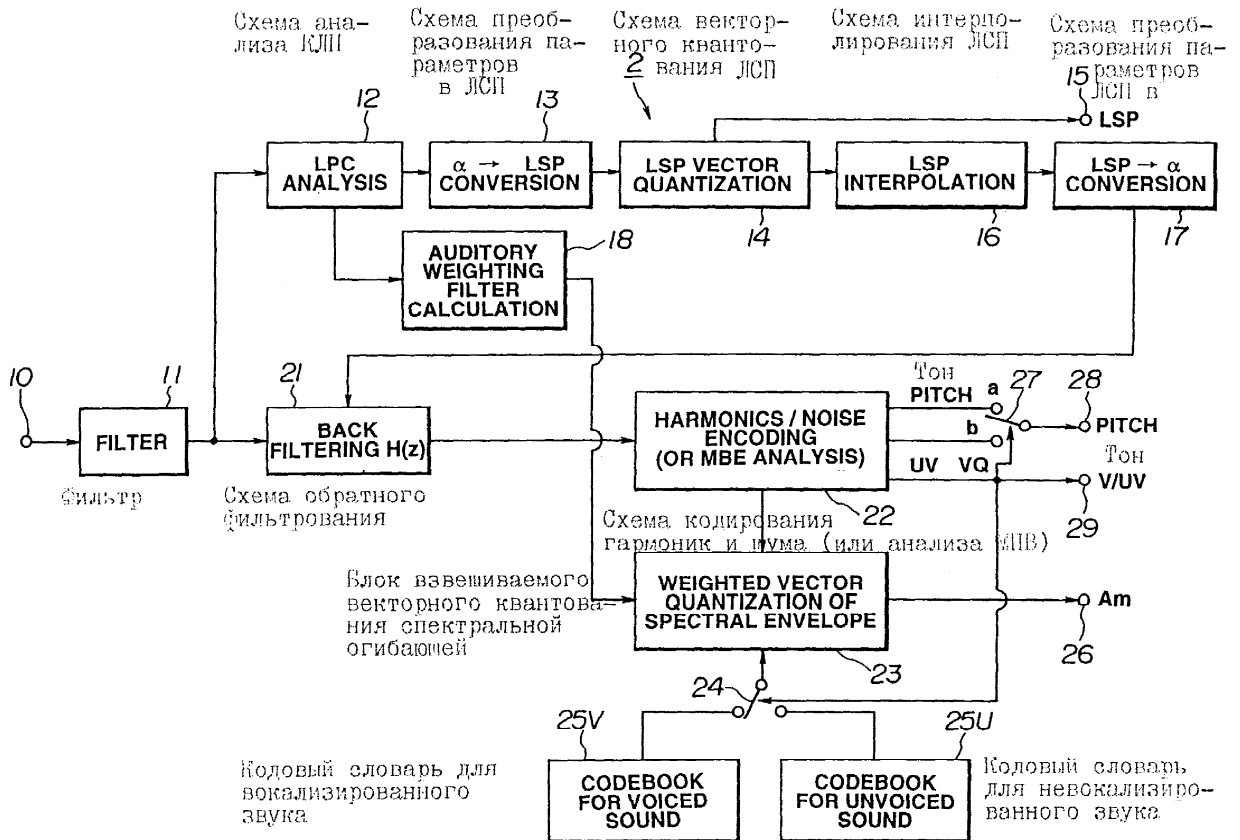
40

45

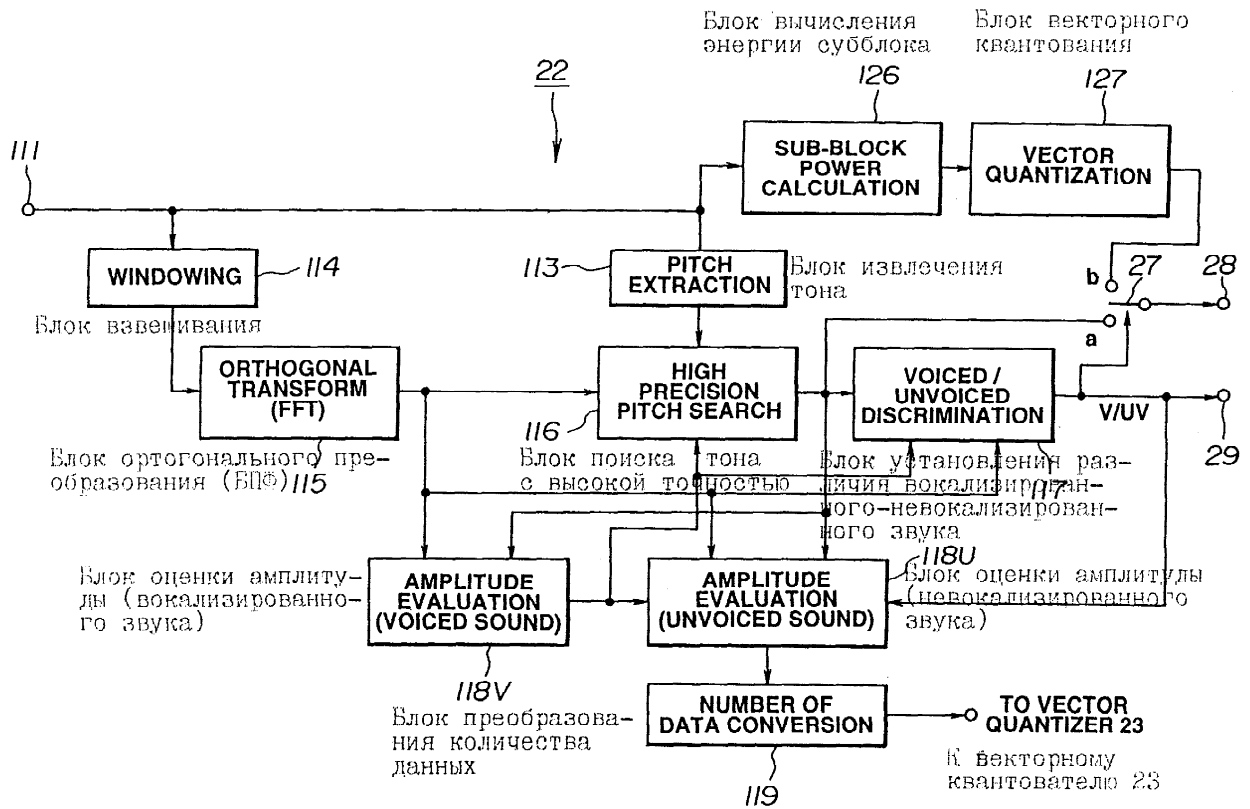
50



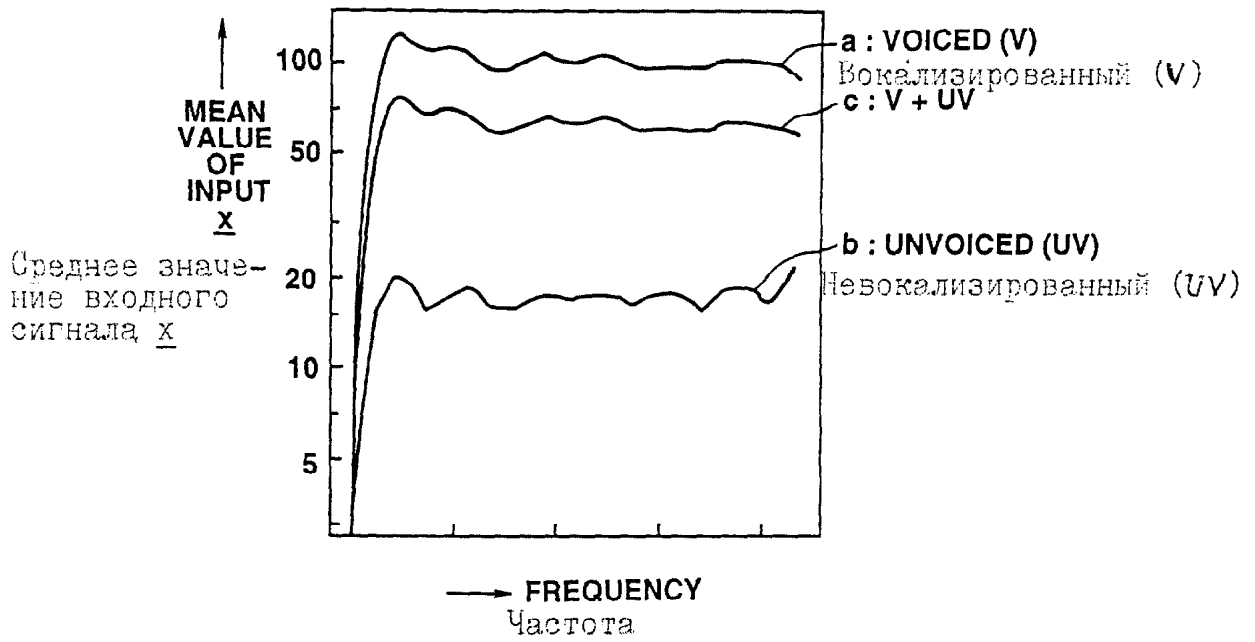
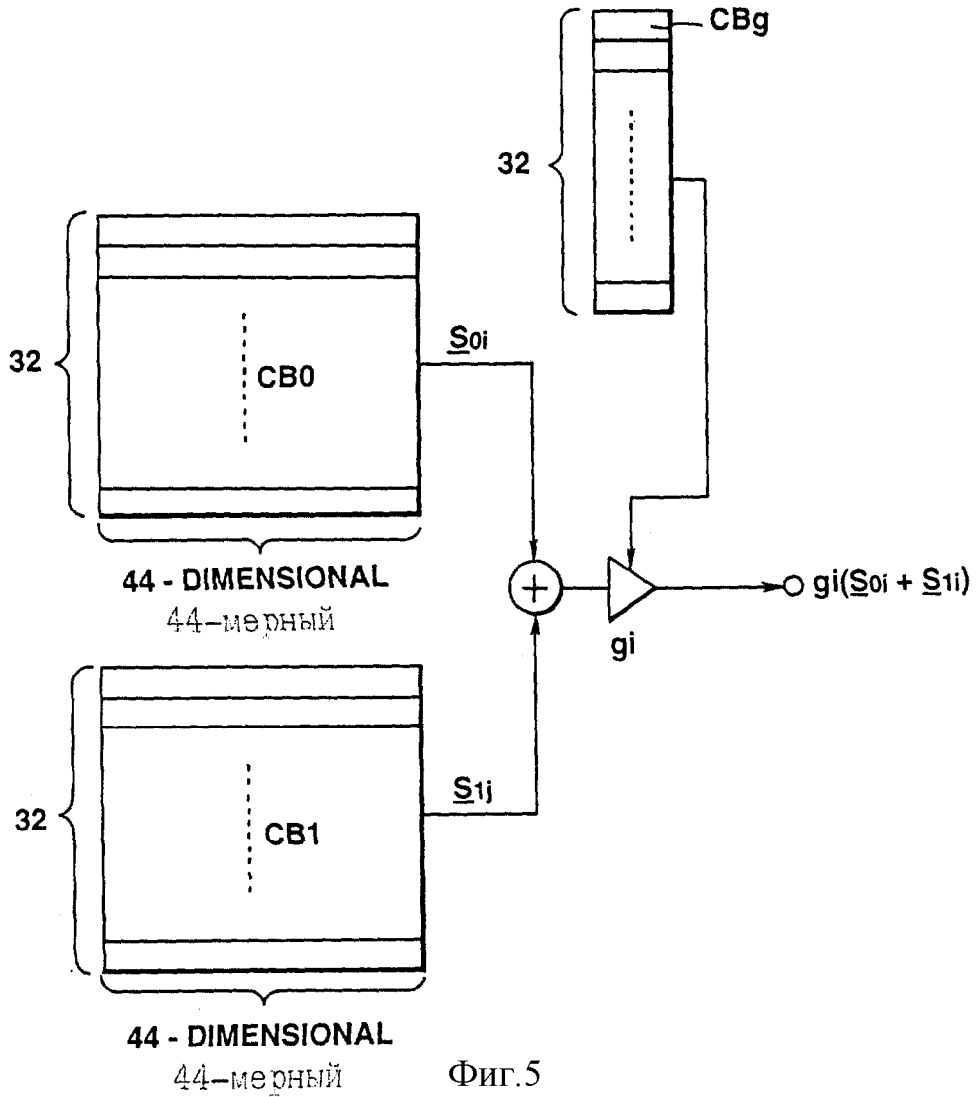
Фиг.2

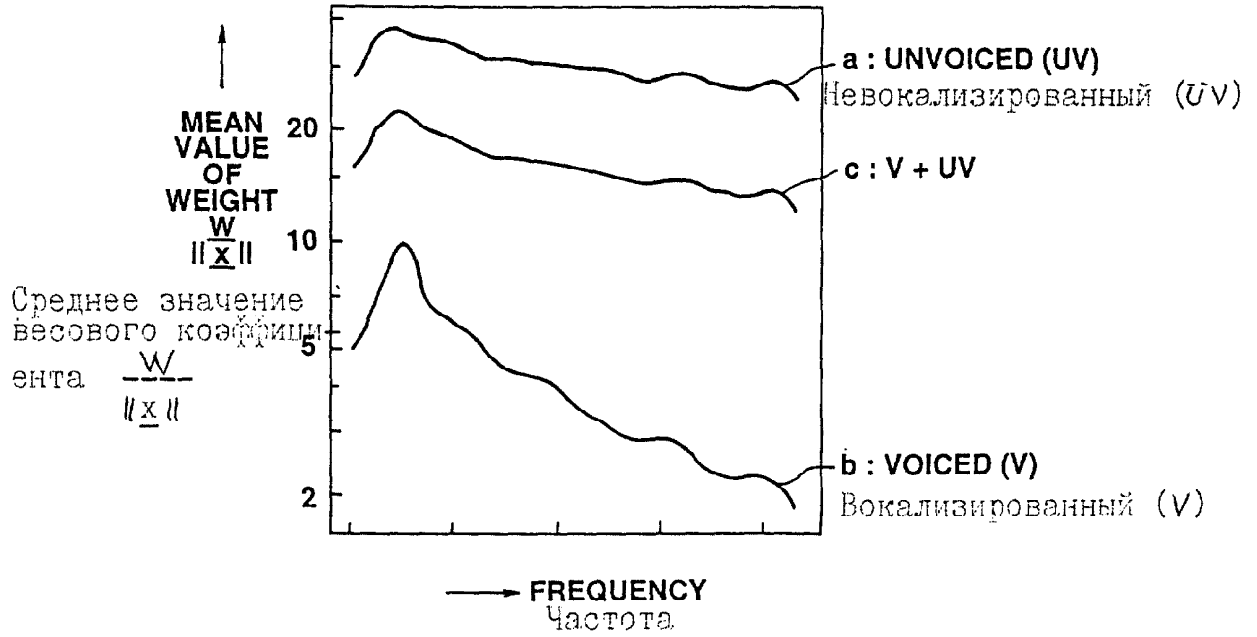


Фиг.3

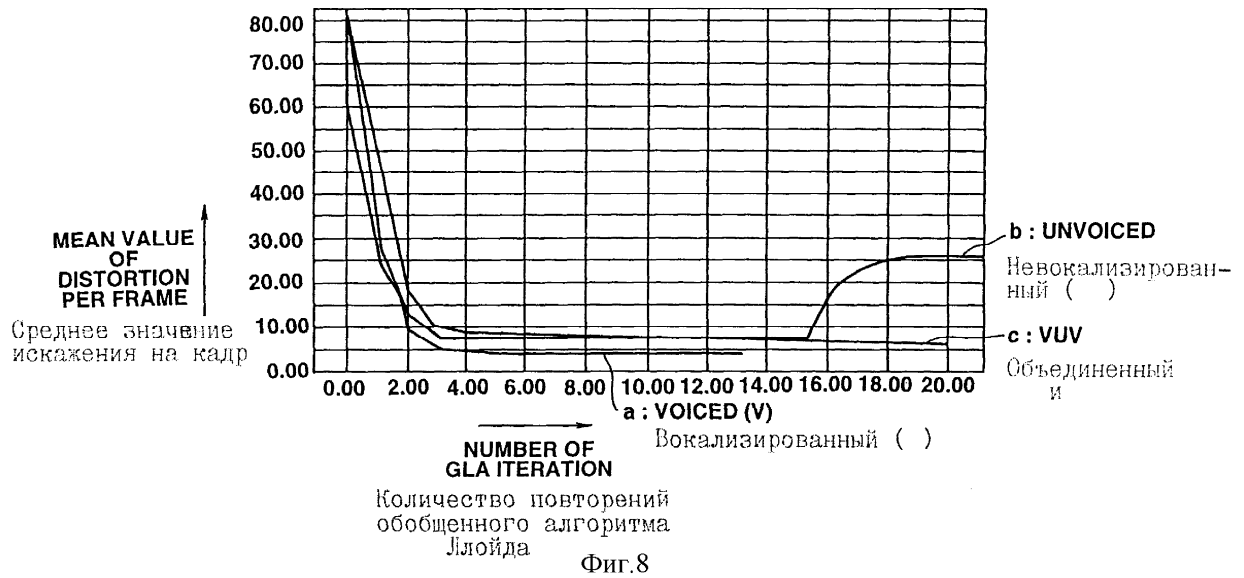


Фиг.4

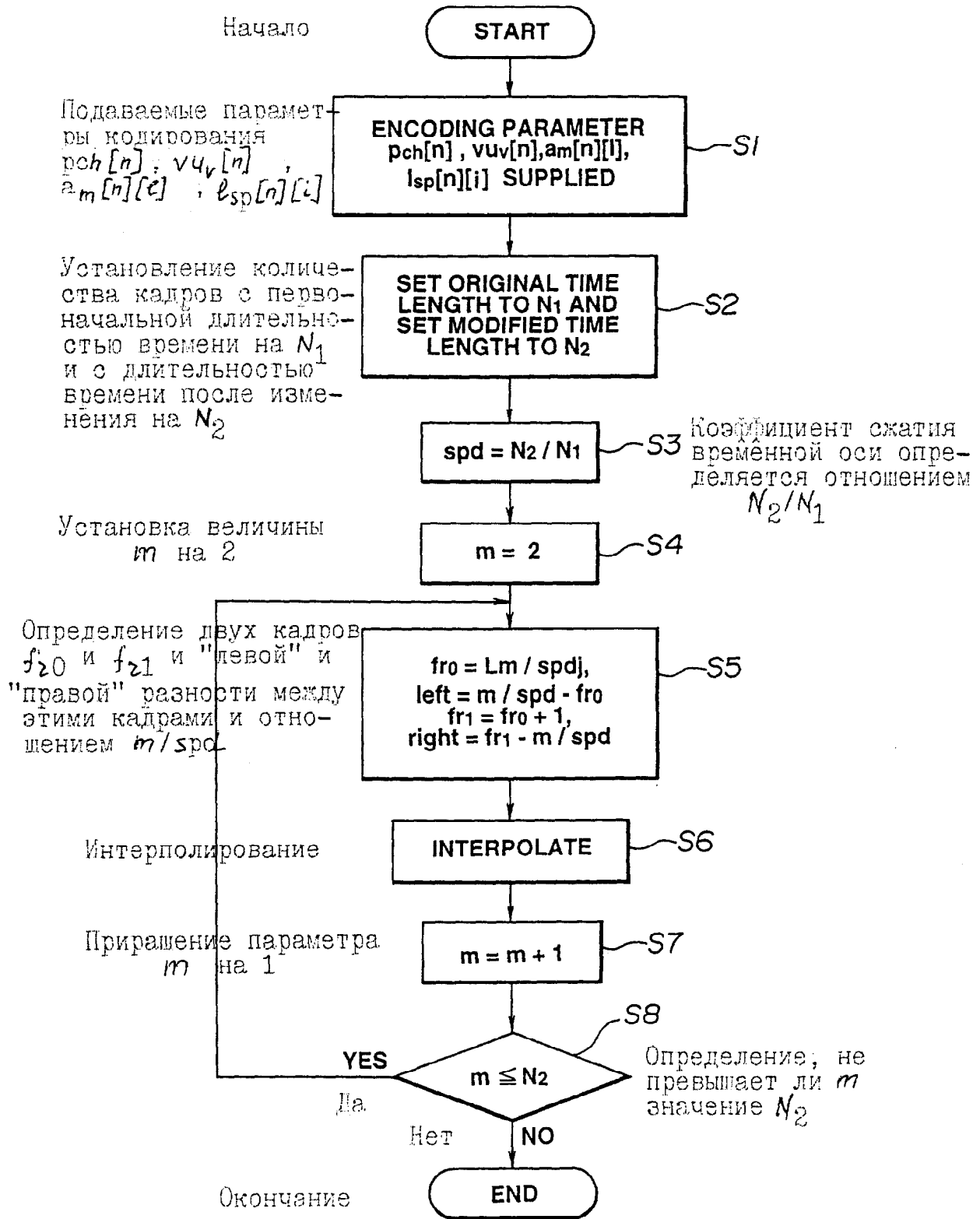




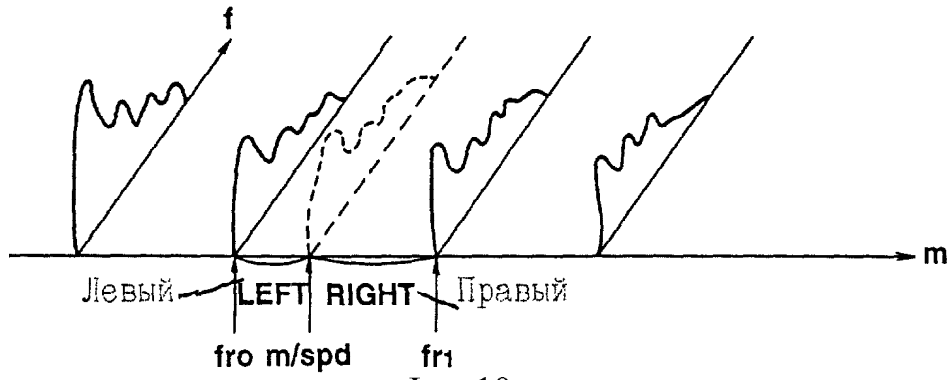
Фиг.7



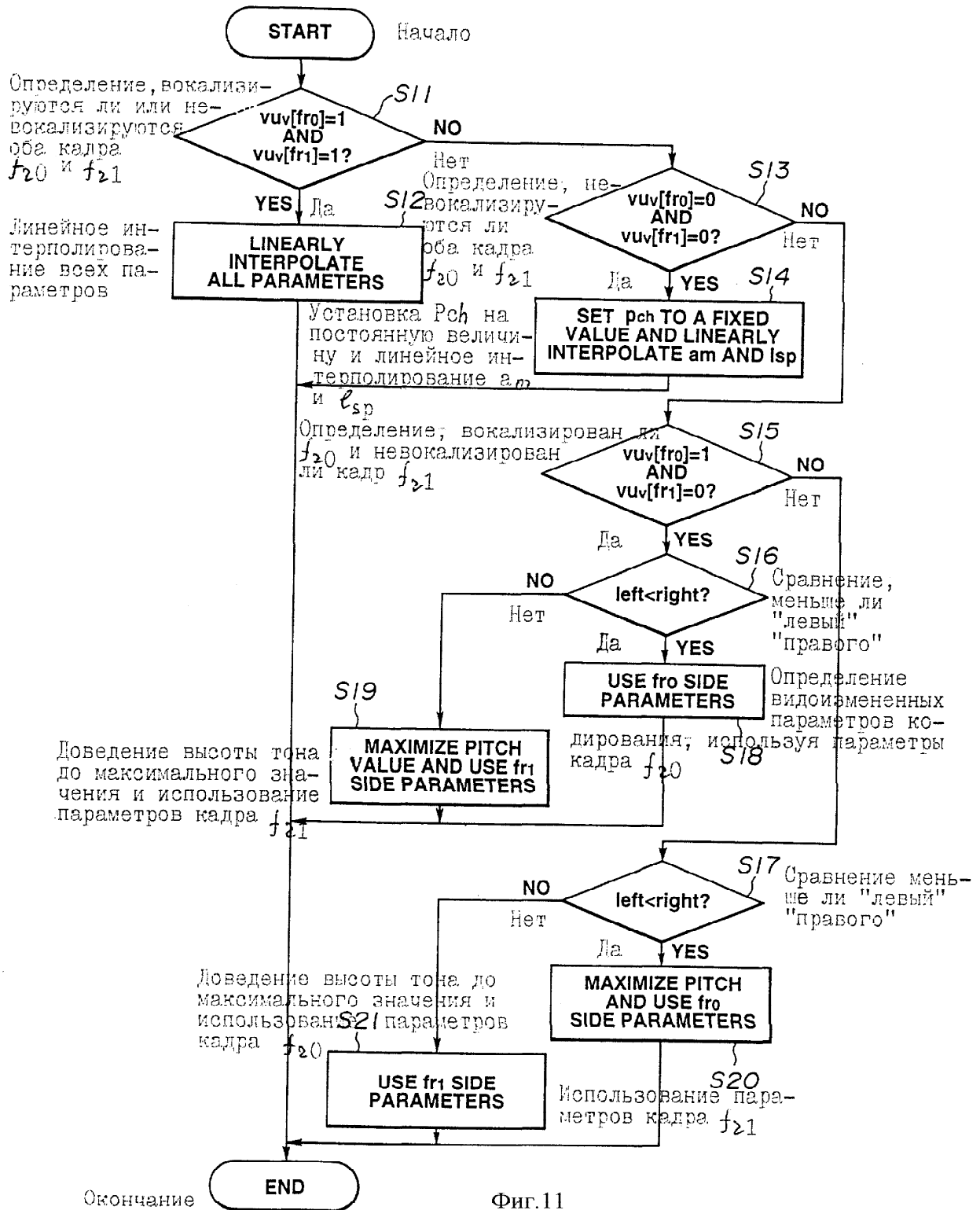
Фиг.8



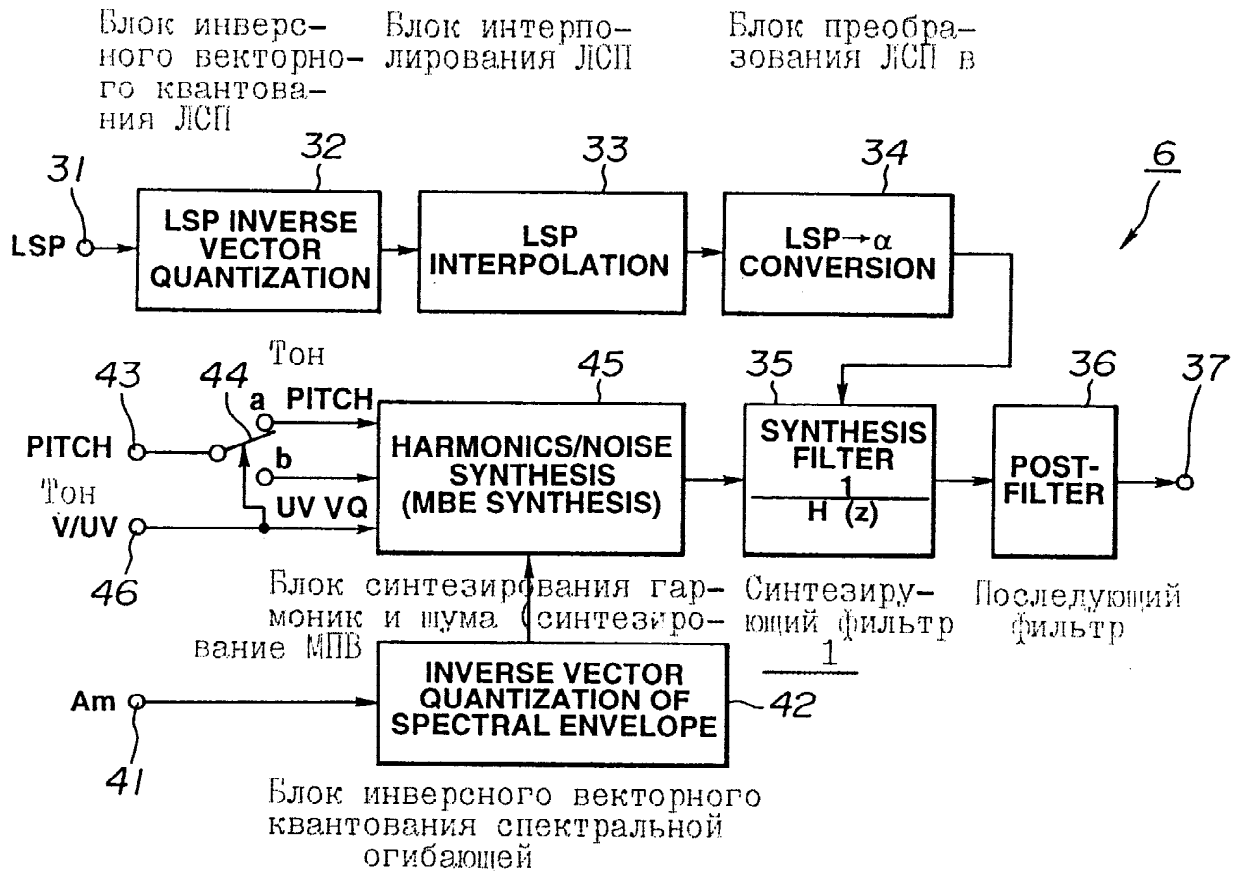
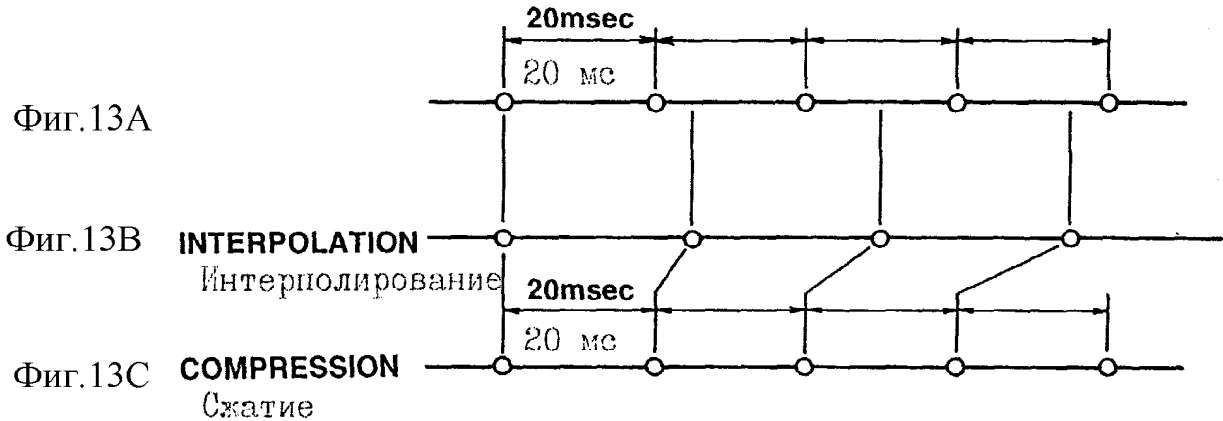
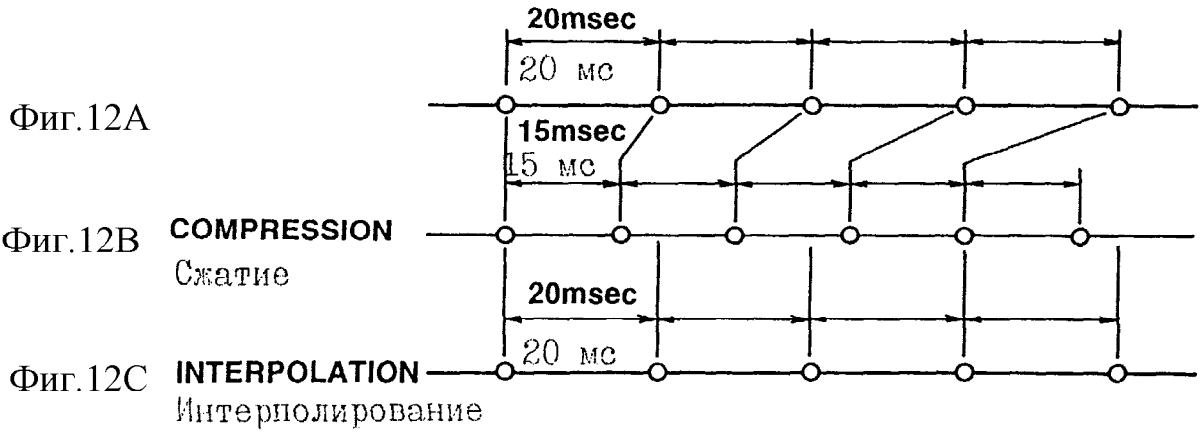
Фиг.9



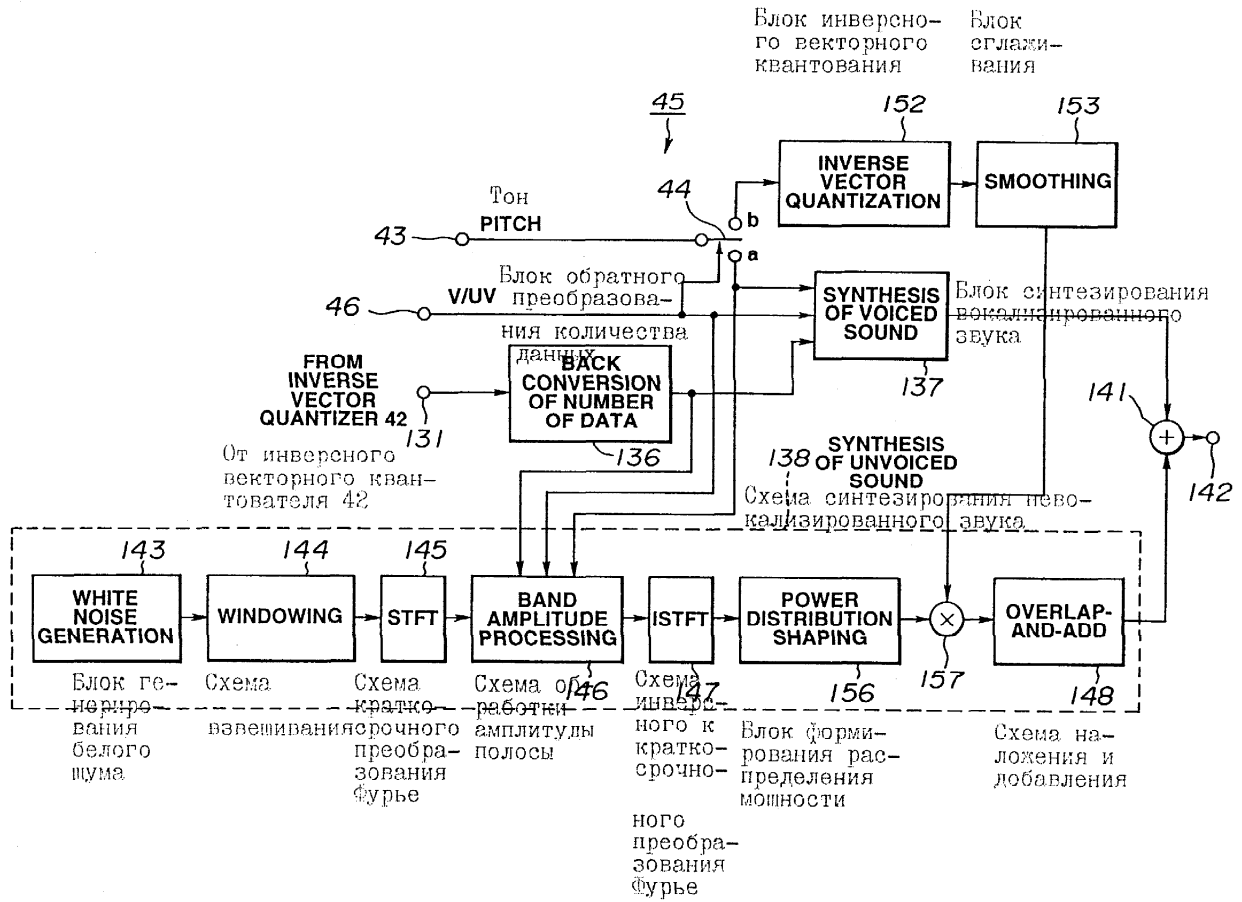
Фиг.10



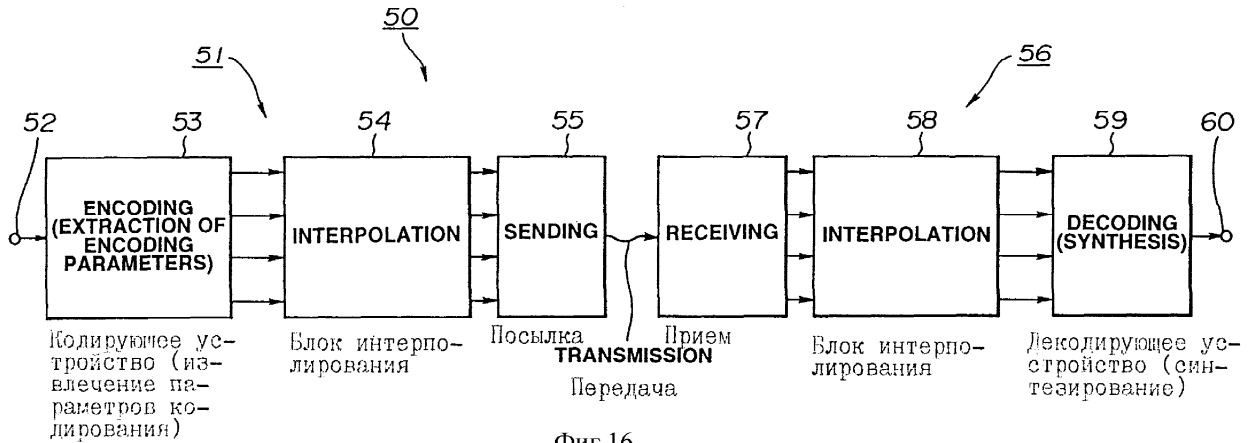
Фиг.11



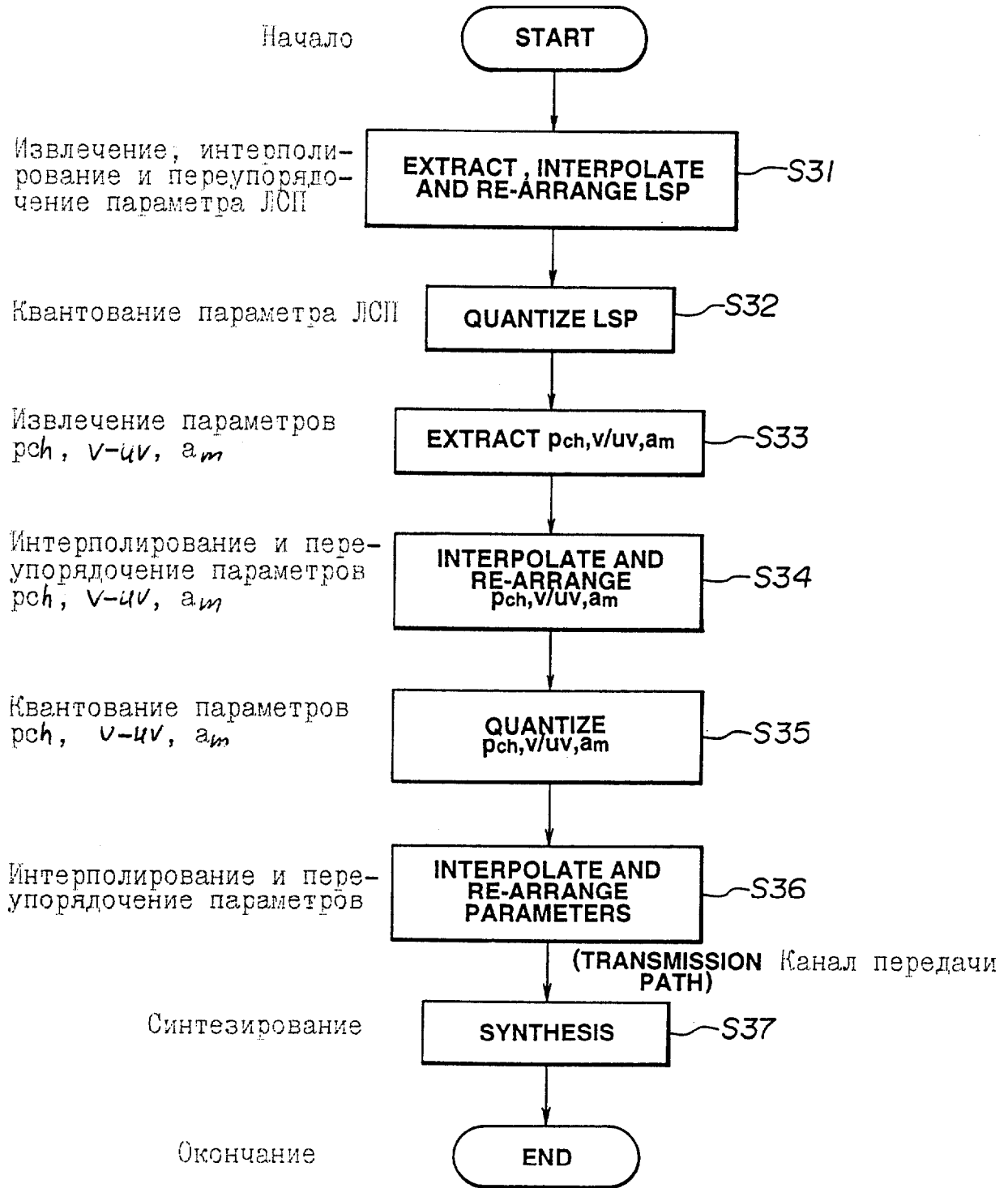
Фиг.14



Фиг. 15

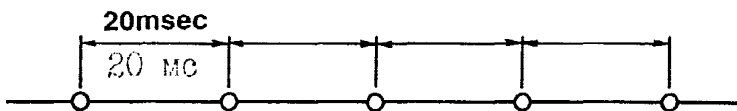


Фиг. 16



Фиг.17

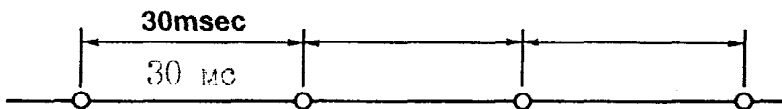
Фиг.18А



Фиг.18В

INTERPOLATION

Интерполирование



Фиг.18С

INTERPOLATION

Интерполирование

