(19) **RU** (11) **2 121 173** (13) **C1**

(51) МПК[6] **G 10 L 9/14**

РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка: 96122985/09, 27.04.1995

(30) Приоритет: 29.04.1994 US 08/236,765

(46) Дата публикации: 27.10.1998

(56) Ссылки: US 4969192 A (CHE N et al.) 06.11.90. US 5307441 A (TZENG) 26.04.94. US 5293449 A (TZENG) 08.03.94. SU 1316030 A1 (Акустический ин-т им. акад. Н.П. Андреева) 07.06.82.

(71) Заявитель:
Аудиокоудс, Лтд. (IL)

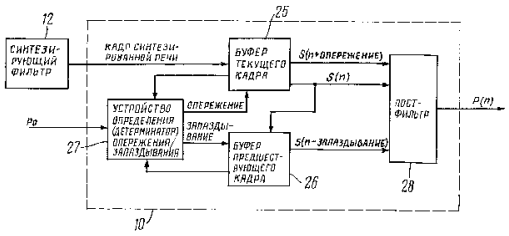
(72) Изобретатель: **Биалик Леон (IL), Фломен Феликс (IL)**

(73) Патентообладатель:
Аудиокоудс, Лтд. (IL)

(54) СПОСОБ ПОСТФИЛЬТРАЦИИ ОСНОВНОГО ТОНА СИНТЕЗИРОВАННОЙ РЕЧИ И ПОСТФИЛЬТР ОСНОВНОГО ТОНА

(57) Реферат:

В соответствии с настоящим изобретением синтезированную речь пропускают через постфильтр, который производит вычисления на основании будущих и предшествующих данных. Кадры данных разделены на субкадры для назначения точек вычисления. Технический результат заключается в улучшении соответствия синтезированной и исходной речи. 2 с. и 8 з.п. ф-лы. 3 ил.



Фиг.1

RU 2121173 C1

RU 2121173 C1

(19) **RU** (11) **2 121 173** (13) **C1**

(51) Int. Cl.[6] **G 10 L 9/14**

RUSSIAN AGENCY
FOR PATENTS AND TRADEMARKS

(12) ABSTRACT OF INVENTION

(21), (22) Application: 96122985/09, 27.04.1995

(30) Priority: 29.04.1994 US 08/236,765

(46) Date of publication: 27.10.1998

(71) Applicant:
Audiokouds, Ltd. (IL)

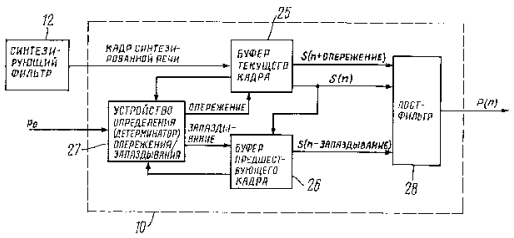
(72) Inventor: **Bialik Leon (IL),**
Flomen Feliks (IL)

(73) Proprietor:
Audiokouds, Ltd. (IL)

(54) METHOD FOR POST-FILTRATION OF FUNDAMENTAL TONE OF SYNTHESIZED SPEECH AND FUNDAMENTAL TONE POST-FILTER

(57) Abstract:

FIELD: computer engineering. SUBSTANCE: method involves passing synthesized voice through post-filter which operates using next and previous data. Data frames are split into sub-frames for setting calculation points. EFFECT: increased correlation of synthesized and source voices. 10 cl, 5 dwg



Фиг.1

RU 2121173 C1

RU 2121173 C1

Изобретение главным образом имеет отношение к созданию систем обработки речевого сигнала и, в частности, касается создания систем с использованием постфильтрации.

Обработка речевого сигнала широко известна и часто применяется для сжатия поступающего речевого сигнала как для хранения, так и для последующей передачи. Обработка речевого сигнала обычно связана с делением входного речевого сигнала на блоки данных (кадры), с последующим анализом каждого блока данных для нахождения его компонентов. Затем полученные компоненты используются как для хранения, так и для последующей передачи.

Если есть желание восстановить исходный речевой сигнал, то каждый кадр декодируют и осуществляют операции синтеза, которые обычно в основном являются инверсией анализа. Полученная при этом синтезированная речь обычно не в полной мере аналогична исходной речи. Поэтому для "улучшения" звукового сигнала обычно осуществляют операции постфильтрации.

Одним из видов постфильтрации является постфильтрация основного тона, при которой информация основного тона, полученная на выходе кодирующего устройства, используется для фильтрации синтезированного сигнала. В известных ранее постфильтрах основного тона производится просмотр ранее полученных выборок участка синтезированного речевого сигнала p_0, где p_0 представляет собой значение основного тона. Субкадр ранее полученной речи, который наилучшим образом совпадает с текущим субкадром, комбинируется с текущим субкадром, обычно в соотношении 1 : 0,25 (то есть предшествующий сигнал ослабляется на три четверти).

К сожалению, речевые сигналы не всегда содержат основной тон. Это относится к случаю промежутка между словами; кроме того, в конце или в начале слова основной тон может изменяться. Так как в известных ранее постфильтрах основного тона производится комбинирование ранее полученной речи с текущим субкадром и так как основной тон ранее полученный речи не совпадает с основным тоном текущего субкадра, то на выходе такого постфильтра основного тона можно получить слабый выходной сигнал в начале слов. Аналогичное справедливо и для субкадра с окончаниями произнесенных слов. Если же большинство субкадров относится к молчанию или к шумам (например, если речь уже закончилась), то основной тон ранее полученного сигнала вообще не может быть использован.

Авторы данного изобретения обратили внимание на то, что декодеры речи обычно создают речевые кадры между их оперативными элементами, в то время как постфильтры основного тона оперируют только с субкадрами речевым сигналов.

Так, например для некоторых из субкадров имеется информация, касающаяся будущих речевых образов.

В связи с изложенным, задачей настоящего изобретения является создание постфильтра основного тона и способа, в которых используется будущая и предыдущая информация по меньшей мере для некоторых из субкадров.

В соответствии с предпочтительным вариантом осуществления настоящего изобретения, постфильтр основного тона получает кадр синтезированной речи, и для каждого субкадра, входящего в кадр синтезированной речи, вырабатывает сигнал, который является функцией субкадра и окон более ранней и более поздней синтезированной речи. Каждое окно используется только при его приемлемом совпадении с субкадром.

В частности, в соответствии с предпочтительным вариантом осуществления настоящего изобретения, постфильтр основного тона осуществляет проверку совпадения окна более ранней синтезированной речи с субкадром и затем принимает (допускает) это окно более ранней синтезированной речи только в том случае, если ошибка между субкадром и взвешенной версией окна мала. Если имеется достаточный промежуток более поздней синтезированной речи, то постфильтр основного тона осуществляет также проверку совпадения окна более поздней синтезированной речи с субкадром и затем принимает (допускает) это окно, если ошибка мала. В этом случае выходной сигнал является функцией субкадра и окон более ранней и более поздней синтезированной речи, если они были приняты.

Далее, в соответствии с предпочтительным вариантом осуществления настоящего изобретения, оценка совпадения предусматривает определение более раннего и более позднего усиления соответственно для окон более ранней и более поздней синтезированной речи.

Кроме того, в соответствии с предпочтительным вариантом осуществления настоящего изобретения, функция для выходного сигнала является суммой субкадра, более раннего окна синтезированной речи, взвешенного более ранним усилением и первым разрешающим весом, и более позднего окна синтезированной речи, взвешенного более поздним усилением и вторым разрешающим весом.

Наконец, также в соответствии с предпочтительным вариантом осуществления настоящего изобретения, первый и второй разрешающие веса зависят от результатов шагов приятия решения (о совпадении).

Настоящее изобретение может быть более полно понятно и оценено из последующего подробного его описания, приведенного со ссылкой на чертежи.

На фиг. 1 показана структурная схема системы, которая включает в себя постфильтр в соответствии с настоящим изобретением; на фиг. 2 - схема, позволяющая лучше понять работу постфильтра в соответствии с фиг. 1; на фиг. 3 - блок-схема операций постфильтра в соответствии с фиг.1.

Обратимся теперь к рассмотрению фиг. 1, 2 и 3, которые позволяют понять работу постфильтра в соответствии с настоящим изобретением.

Как показано на фиг. 1, постфильтр основного тона, обозначенный позицией 10, в соответствии с настоящим изобретением

RU 2121173 C1

RU 2121173 C1

получает кадры синтезированной речи от синтезирующего фильтра 12, такого как синтезирующий фильтр с коэффициентом линейного прогнозирования (LPC). Постфильтр основного тона 10 получает также значение основного тона, которое было получено ранее в декодере речи. Постфильтр основного тона 10 не должен быть обязательно первым постфильтром; он может также получать постфильтрованные кадры синтезированной речи. Постфильтр основного тона 10 включает в себя буферное устройство (буфер) текущего кадра 25, буфер предшествующего кадра 26, детерминатор (устройство определения) опережения/запаздывания 27 и постфильтр 28. Буфер текущего кадра 25 запоминает текущий кадр синтезированной речи и его разбивание на субкадры. Буфер предшествующего кадра 26 запоминает предшествующие кадры синтезированной речи. Детерминатор опережения/запаздывания 27 определяет указанные ранее показатели опережения и запаздывания относительно значения основного тона p_0. Постфильтр 28 получает субкадр s [n] и будущее окно s [n + LEAD] от буфера текущего кадра 25 и вырабатывает из них постфильтрованный сигнал.

Следует иметь в виду, что синтезирующий фильтр 12 синтезирует кадры синтезированной речи и подает их на постфильтр основного тона 10. Аналогично известным ранее постфильтрам основного тона, постфильтр в соответствии с настоящим изобретением оперирует с субкадрами синтезированной речи. Однако, так как заявители настоящего изобретения поняли, что при обработке субкадров на буфере текущего кадра 25 имеется полный кадр синтезированной речи, то постфильтр основного тона 10 в соответствии с настоящим изобретением также использует будущую информацию по меньшей мере для некоторых из субкадров.

Это показано на фиг.2, где изображены восемь субкадров 20a - 20h двух кадров 22a и 22b, которые запоминаются соответственно в буфере текущего кадра 25 и в буфере предыдущего кадра 26. Показаны также местоположения, из которых могут быть взяты аналогичные субкадры данных для более поздних субкадров 20e - 20h. Как показывает стрелка 24e, для первого субкадра 20e данные могут быть взяты из предыдущих субкадров 20d, 20e и 20b, а также из будущих субкадров 20e, 20f и 20g. Как показывает стрелка 24f, для второго субкадра 20f данные могут быть взяты из предыдущих субкадров 20e, 20d и 20c, а также из будущих субкадров 20f, 20g и 20h. Следует иметь в виду, что для более поздних субкадров 20g и 20h имеется меньше будущих данных, которые могли бы быть использованы (так как при этом субкадр 20h отсутствует), однако при этом имеется некоторое количество прошлых данных, которое может быть использовано.

Детерминатор опережения/запаздывания 27 в соответствии с настоящим изобретением производит поиск в прошлом и будущем сигналов синтезированной речи, раздельно определяет для них положение выборок запаздывания и опережения или соответственно показатель, в соответствии с которым окна длины субкадра прошедшего и будущего сигналов, начинающиеся соответственно при выборах запаздывания и опережения, наиболее полно совпадают с текущим субкадром. Если совпадение плохое, то окно не используется. Обычно поиск проводят в пределах 20 - 146 выборок раньше и позже текущего субкадра, как это показано стрелками 24. Диапазон поиска сокращен для будущих данных (то есть для субкадров 20g и 20h).

После этого постфильтр 28 производит постфильтрацию сигнала синтезированной речи с использованием любого или обоих совпавших окон.

Один из вариантов построения постфильтра в соответствии с настоящим изобретением иллюстрируется фиг. 3, на которой приведена блок-схема его операций для одного субкадра. Операции 30 - 74 осуществляются при помощи детерминатора опережения/запаздывания 27, а операции 76 и 78 осуществляются при помощи постфильтра 28.

Операции способа в соответствии с настоящим изобретением начинаются с инициализации (шаг 30), при которой устанавливают в качестве величины минимального критерия минимальные и максимальные значения опережения/запаздывания. В соответствии с этим вариантом, минимальным опережением/запаздыванием является минимальный (основной тон - дельта, 20), а максимальным опережением/запаздыванием является (максимальный (основной тон + дельта, 146). В соответствии с этим вариантом дельта равняется трем.

При проведении операций 34 - 44 определяют значение опережения, а при проведении операций 60 - 70 определяют значение запаздывания, если они имеются. Обе секции осуществляют одинаковые операции, при этом предыдущие данные запоминают в буфере предшествующего кадра 26, а будущие данные запоминают в буфере текущего кадра 25. Поэтому операции будут описаны ниже только для одной из секций. Однако полученные уравнения являются разными, как это указано далее.

При проведении операции 32 устанавливают минимальную величину показателя запаздывания M_g, а при проведении операций 34 и 36 с показателем запаздывания M_g объединяют усиление g_g и определяют критерий E_g для этого показателя запаздывания. Усиление g_g является отношением кросс-корреляция субкадра s [n] и предыдущего окна s[n-M_g] с автокорреляцией предыдущего окна s[n-M_ g] в соответствии с выражением

$$g_g = \Sigma s[n]^{*}s[n-M_g]/\Sigma s^{2}[n-M_g], \quad 0 \le n \le 59. \quad (1)$$

Критерий E_ g представляет собой энергию сигнала ошибки s[n] - g_g * s[n-M _g] в соответствии с выражением

$$E_g = \Sigma(s[n]-g_g^{*}s[n-M_g])^{2}, \quad 0 \le n \le 59. \quad (2)$$

Если результирующий критерий меньше ранее определенного минимального значения (операция 38), то производят запоминание текущего показателя запаздывания M_g и условия g_g, а также устанавливают минимальное значение текущего усиления (операция 40). Показатель запаздывания увеличивают на единицу (операция 42) и процесс повторят до достижения

RU 2121173 C1

RU 2121173 C1

максимального значения показателя запаздывания.

При проведении операций 46 - 50 результат определения запаздывания принимают только в том случае, если усиление запаздывания, которое найдено при проведении операций 34 - 44, больше заданной величины порога или равно ей, причем порог, например, может быть равен 0,625. При проведении операции 46 признак разрешения запаздывания устанавливают на 0, а при проведении операции 48 проверяют порог усиления запаздывания g _ g. При проведении операции 50 принимают результат установкой признака разрешения запаздывания на 1. Таким образом, для предыдущего речевого сигнала, который не аналогичен текущему субкадру, например, если в текущем субкадре имеется речь, а в предыдущем ее нет, данные для предыдущего субкадра использоваться не будут.

При проведении операций 52 - 56 признак разрешения опережения устанавливают только в том случае, когда сумма текущего положения N, длительность субкадра (длина которого обычно составляет 60 выборок) и максимального значения запаздывания/опережения, составляет меньше длительности кадра (длина которого обычно составляет 240 выборок). Таким образом, будущие данные используют только тогда, когда они имеются в достаточном количестве. При проведении операции 52 признак разрешения опережения устанавливают на 0, а при проведении операции 54 производят проверку приемлемости суммы, и, если она приемлема, то при проведении операции 56 признак разрешения опережения устанавливают на 1.

При проведении операции 58 повторно инициализируют минимальное значение и показатель опережения устанавливают на минимальное значение опережения. Как уже упоминалось ранее, операции 60 - 70 аналогичны операциям 34 - 44; при их проведении определяют показатель опережения, который наилучшим образом совпадает с представляющим интерес субкадром. Опережение обозначают M _ d, усилие обозначают g _ d, а критерий обозначают E _ d; они определены уравнениями 3 и 4 в следующем виде:

$$g_d = \Sigma s[n]^* s[n-M_d]/\Sigma s^2[n-M_d], \quad 0 \le n \le 59; \quad (3)$$

$$E_d = \Sigma(s[n]-g_d^* s[n-M_d])^2, \quad 0 \le n \le 59. \quad (4)$$

При проведении операции 60 находят усилие g_d, при проведении операции 62 находят критерий E_d, при проведении операции 64 проверяют, что критерий E_ d меньше, чем минимальное значение, а при проведении операции 66 производят запоминание опережения M_d и усиление опережения g_d, а также обновляют минимальное значение E_d. При проведении операции 68 увеличивают показатель опережения на единицу, а при проведении операции 70 определяют, превышает или нет показатель опережения значение максимального показателя опережения.

При проведении операций 72 и 74 признак разрешения опережения блокируют (операция 74), если усиление опережения, которое найдено при проведении операций 60 - 70, является слишком низким (например, ниже заданного порога), причем эту проверку осуществляют при проведении операции 72.

При проведении операции 76 определяют веса запаздывания w_g и опережения w_ d соответственно из признаков разрешения запаздывания и опережения. Веса w_g и w_d определяют вклады, если они есть, будущих и предыдущих данных.

В соответствии с этим вариантом вес запаздывания w_g представляет собой максимум выражения (разрешение запаздывания - (0,5 * разрешение опережения)) и 0, умноженного на 0,25. Вес опережения w_d представляет собой максимум выражения (разрешение опережения - (0,5 * разрешение запаздывания)) и 0, умноженного на 0,25. Другими словами, веса запаздывания w_g и опережения w_d оба равны 0,125 в том случае, когда имеются как будущие, так и предшествующие данные, которые совпадают с текущим субкадром; равны 0,25 в том случае, когда только один из этих весов совпадает с текущим субкадром; равны 0, когда нет совпадения.

При проведении операции 78 получают выходной сигнал p[n], который является функцией сигнала s[n], более раннего окна s[n-M_g] и более позднего окна s[n -M_d]. Величина M_g и M_d представляет собой соответственно показатели запаздывания и опережения, которые хранятся в памяти. Функция сигнала p[n] для данного варианта может быть определена в соответствии со следующими выражениями 5 и 6:

$$p[n]=g_p^*\{s[n]+w_g^* g_g^* s[n-M_g]+w_d^* g_d^* s[n-M_d]\}= g_p^* p'[n];$$

$$g_p = \mathrm{sqrt}(\Sigma s^2[n]/\Sigma p'^2[n]), \quad 0 \le n \le 59. \quad (6)$$

Операции 30 - 78 повторяют для каждого субкадра.

Следует иметь в виду, что настоящее изобретение включает в себя все постфильтры основного тона, в которых используется как будущая, так и предшествующая информация.

Несмотря на то, что был описан предпочтительный вариант осуществления изобретения, совершено ясно, что в него специалистами в данной области могут быть внесены изменения и дополнения, которые не выходят однако из рамки приведенной далее формулы изобретения.

Формула изобретения:

1. Способ постфильтрации основного тона синтезированной речи, включающий в себя следующие операции: прием кадра синтезированной речи, который разделен на множество субкадров, и значения основного тона, объединенного с указанным кадром, и для каждого субкадра указанного кадра синтезированной речи выработку выходного сигнала, который является постфильтрованной версией основного тона текущего субкадра, отфильтрованной при помощи одного из выбранных элементов группы, состоящей из предшествующих и будущих данных указанной синтезированной речи, а также из будущих данных указанной синтезированной речи, причем указанные предшествующие данные имеют запаздывание относительно текущего субкадра на показатель запаздывания, а указанные будущие данные имеют опережение относительно текущего субкадра

RU 2121173 C1

RU 2121173 C1

на показатель опережения, при этом указанные показатели запаздывания и опережения основаны на указанном значении основного тона.

2. Способ по п. 1, отличающийся тем, что указанная операция выработки включает в себя следующие операции: проверка совпадения субкадра с предшествующим окном указанной предшествующей синтезированной речи, начинающимся при указанном показателе запаздывания, принятие указанного совпавшего предшествующего окна только в том случае, когда ошибка между указанным субкадром и взвешенной версией указанного предшествующего окна лежит ниже порога, и, если имеется достаточно будущей синтезированной речи, проверка совпадения субкадра с будущим окном указанной будущей синтезированной речи, начинающимся при указанном показателе опережения, принятие указанного совпавшего будущего окна только в том случае, когда ошибка между указанным субкадром и взвешенной версией указанного будущего окна лежит ниже порога, и создание указанного выходного сигнала путем постфильтрации указанного субкадра при помощи одного из выбранных элементов группы, состоящей из указанных предшествующих и будущих окон и указанного будущего окна.

3. Способ по п. 2, отличающийся тем, что указанные операции определения совпадения предусматривают операции определения более раннего и более позднего усиления соответственно для окон более ранней и более поздней синтезированной речи.

4. Способ по п. 3, отличающийся тем, что указанная операция выработки включает в себя определение сигнала, который является суммой указанного субкадра, более раннего окна синтезированной речи, взвешенного более ранним усилением и первым разрешающим весом, и более позднего окна синтезированной речи, взвешенного более поздним усилением и вторым разрешающим весом.

5. Способ по п. 4, отличающийся тем, что указанные первый и второй разрешающие веса зависят от результата указанных операций принятия решения.

6. Постфильтр основного тона для постфильтрации синтезированной речи, отличающийся тем, что он включает в себя средство для приема кадра синтезированной речи, который разделен на множество субкадров, и значения основного тона, объединенного с указанным кадром, и средство выработки, для каждого субкадра указанного кадра синтезированной речи, выходного сигнала, который является постфильтрованной версией основного тона текущего субкадра, отфильтрованной при помощи одного из выбранных элементов группы, состоящей из предшествующих и будущих данных указанной синтезированной речи, а также из будущих данных указанной синтезированной речи, причем указанные предшествующие данные имеют запаздывание относительно текущего субкадра на показатель запаздывания, а указанные будущие данные имеют опережение относительно текущего субкадра на показатель опережения, при этом указанные показатели запаздывания и опережения основаны на указанном значении основного тона.

7. Фильтр по п. 6, отличающийся тем, что указанное средство выработки включает в себя первое средство сравнения проверки совпадения субкадра с предшествующим окном указанной предшествующей синтезированной речи, начинающимся при указанном показателе запаздывания, первое средство сравнения для принятия указанного совпавшего предшествующего окна только в том случае, когда ошибка между указанным субкадром и взвешенной версией указанного предшествующего окна лежит ниже порога, второе средство проверки совпадения, которое работает только в том случае, если имеется достаточно будущей синтезированной речи, для проверки совпадения субкадра с будущим окном указанной будущей синтезированной речи, начинающимся при указанном показателе опережения, второе средство сравнения для принятия указанного совпавшего будущего окна только в том случае, когда ошибка между указанным субкадром и взвешенной версией указанного будущего окна лежит ниже порога, и средство фильтрации для создания указанного выходного сигнала путем постфильтрации указанного субкадра при помощи одного из выбранных элементов группы, состоящей из указанных предшествующих и будущих окон и указанного будущего окна.

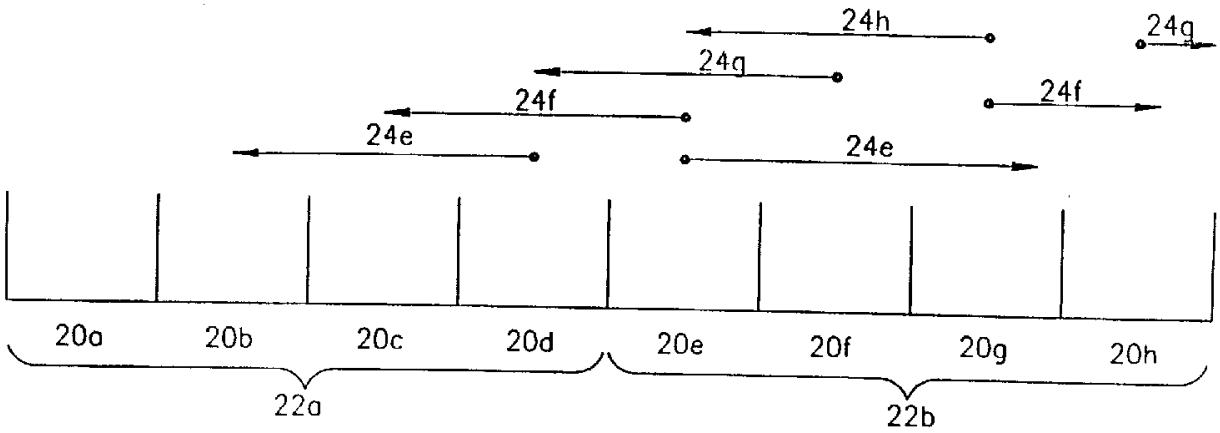
8. Фильтр по п. 7, отличающийся тем, что указанные первое и второе средства определения совпадения включают в себя детерминаторы усиления для определения более раннего и более позднего усиления соответственно для окон более ранней и более поздней синтезированной речи.

9. Фильтр по п. 8, отличающийся тем, что указанное средство фильтрации включает в себя средство для определения сигнала, который является суммой указанного субкадра, более раннего окна синтезированной речи, взвешенного более ранним усилением и первым разрешающим весом, и более позднего окна синтезированной речи, взвешенного более поздним усилением и вторым разрешающим весом.

10. Фильтр по п. 9, отличающийся тем, что указанные первый и второй разрешающие веса зависят от выходных сигналов первого и второго средств сравнения.

RU 2121173 C1

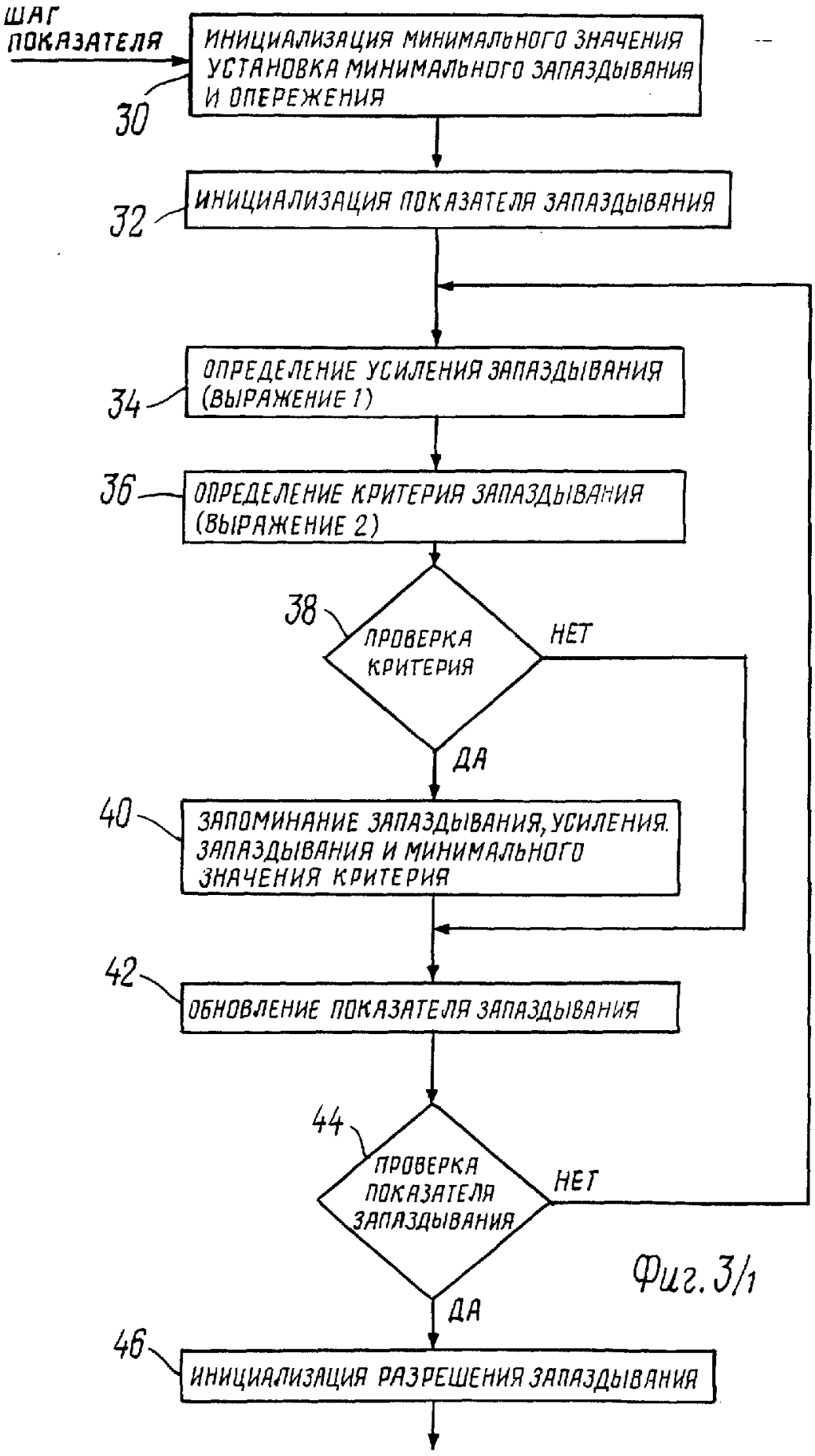
RU 2121173 C1



Фиг.2

RU 2121173 C1

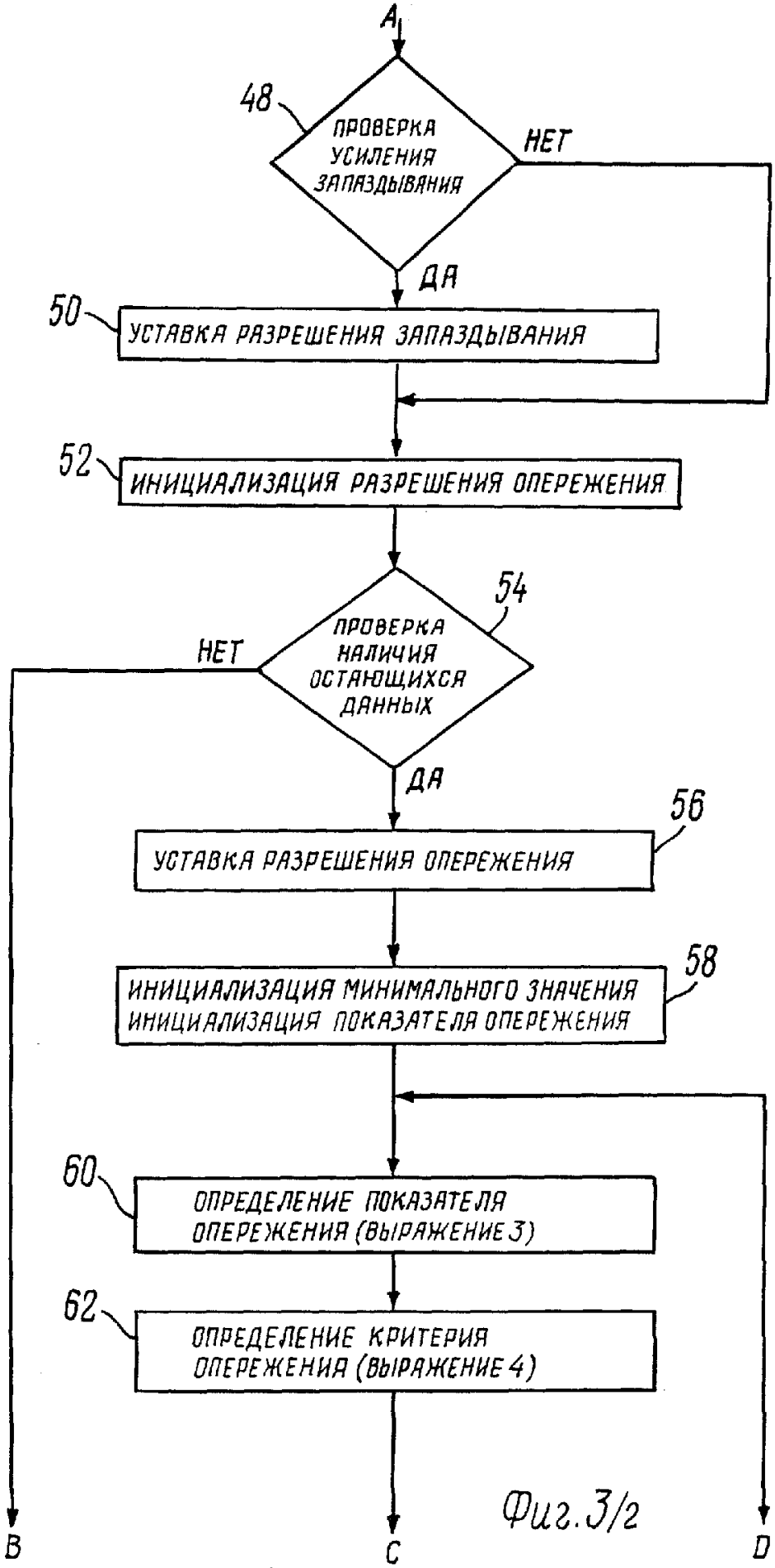
RU 2121173 C1



Фиг. 3/1

RU 2121173 C1

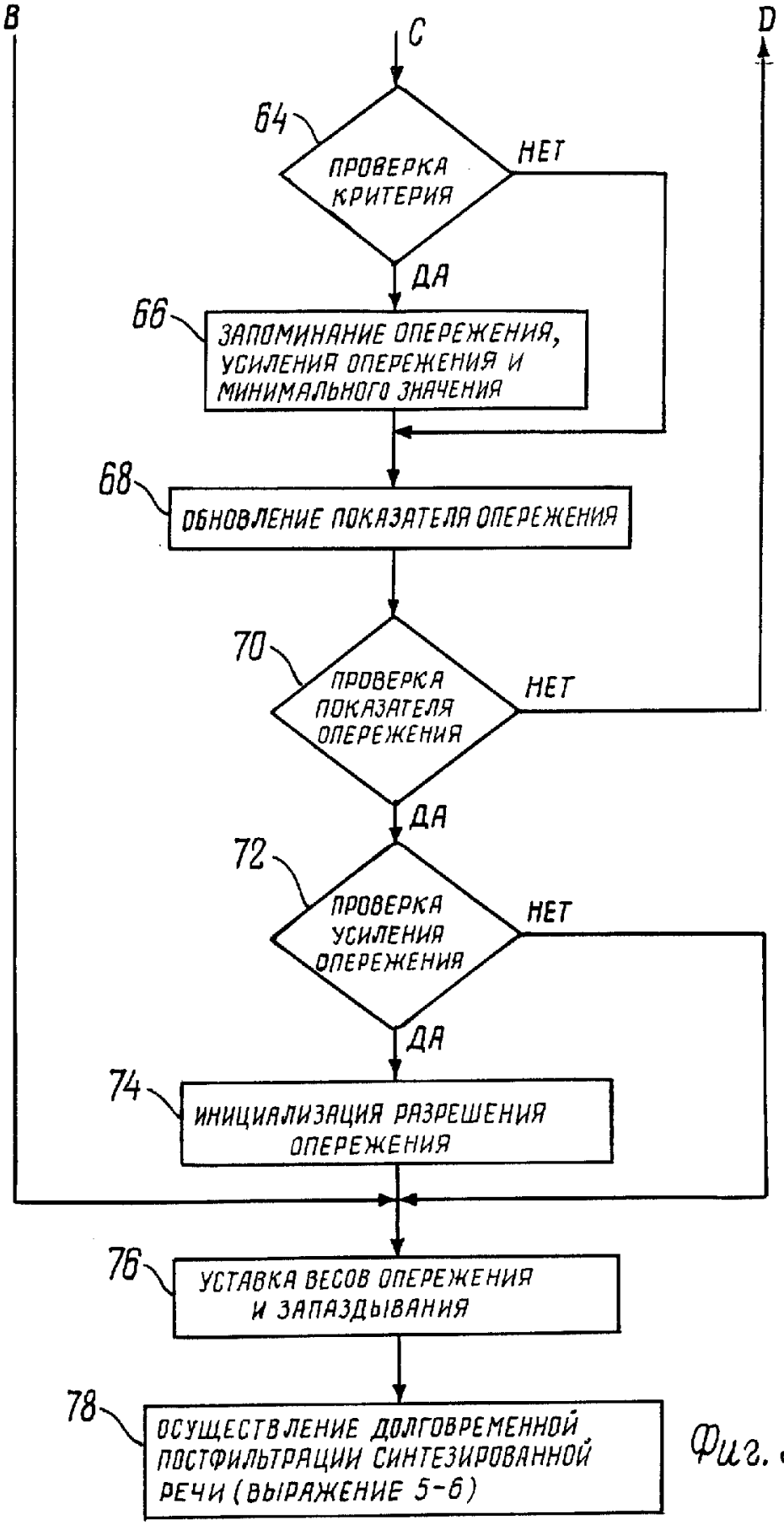
RU 2121173 C1



Фиг.3/2

RU 2121173 C1

RU 2121173 C1



Фиг. 3/3

RU 2121173 C1