

# ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(12)

РЕСПУБЛИКА БЕЛАРУСЬ



НАЦИОНАЛЬНЫЙ ЦЕНТР  
ИНТЕЛЛЕКТУАЛЬНОЙ  
СОБСТВЕННОСТИ

(19) ВУ (11) 15755

(13) С1

(46) 2012.04.30

(51) МПК

G 10L 11/02 (2006.01)

(54)

## СПОСОБ ДЕТЕКТИРОВАНИЯ РЕЧЕВОГО СИГНАЛА

(21) Номер заявки: а 20081472

(22) 2008.11.20

(43) 2010.06.30

(71) Заявитель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(72) Авторы: Зельманский Олег Борисович; Давыдов Андрей Геннадьевич; Давыдов Геннадий Владимирович; Лыньков Леонид Михайлович (ВУ)

(73) Патентообладатель: Учреждение образования "Белорусский государственный университет информатики и радиоэлектроники" (ВУ)

(56) US 5459814 А, 1995.

ВУ 4214 U, 2008.

RU 2251750 С2, 2005.

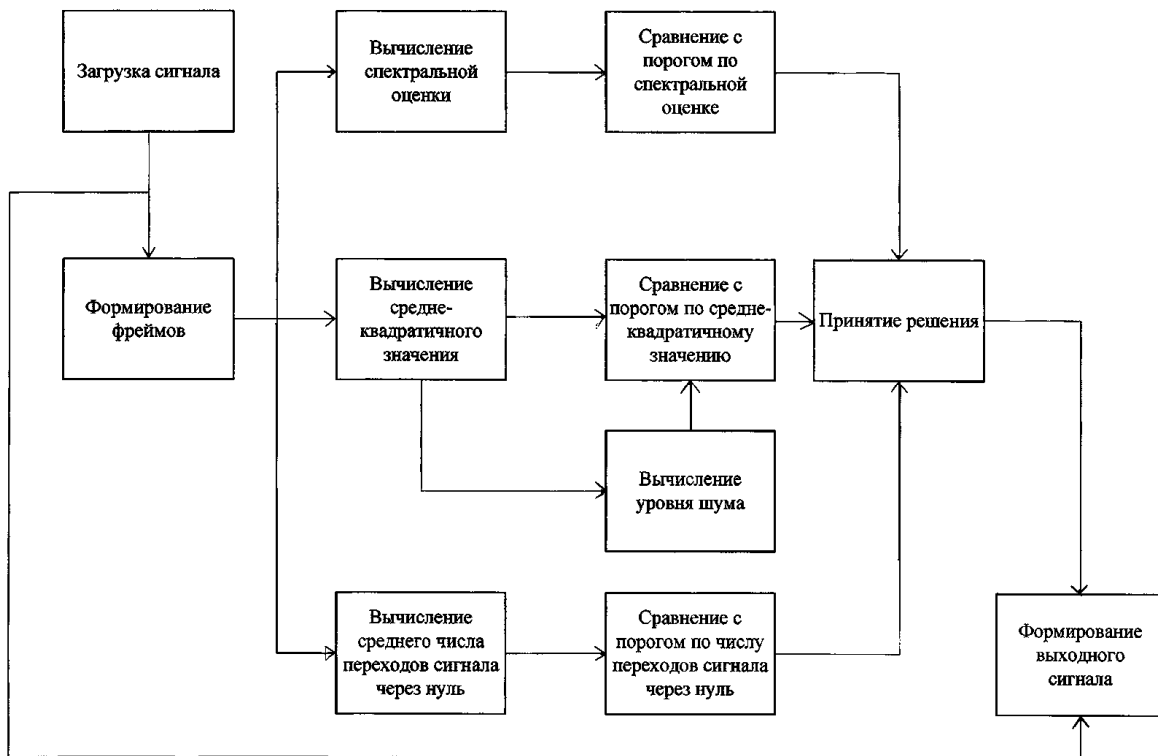
RU 2005106251 А, 2005.

JP 10274991 А, 1998.

US 6289309 В1, 2001.

(57)

Способ детектирования речевого сигнала, при котором разделяют дискретизированный аудиосигнал на временные окна, для каждого из которых вычисляют значения классификационных параметров сигнала и сравнивают их с соответствующими установленными



Фиг. 1

ВУ 15755 С1 2012.04.30

порогами, вычисляют и обновляют уровень шума и значения порогов при изменении условий внешней акустической обстановки, **отличающийся** тем, что после разделения дискретизированного аудиосигнала на временные окна осуществляют его буферизацию, для каждого окна вычисляют среднее число переходов сигнала через нулевой уровень, среднеквадратичное значение сигнала и значение изменения его спектра, сравнивают полученные величины с соответствующими установленными порогами и принимают решение о наличии речевых участков в сигнале при превышении всеми упомянутыми величинами установленных порогов, а уровень шума вычисляют путем отслеживания минимума среднеквадратического значения сигнала на заданном интервале и обновляют независимо от того, является текущее анализируемое временное окно речевым или неречевым.

---

Изобретение относится к способам обработки и анализа аудиосигналов, в частности к способам обнаружения наличия или отсутствия речи в аудиосигналах, и может использоваться в системах с голосовым управлением, системах с биометрическим контролем доступа, системах кодирования и распознавания речи, системах повышения качества речи, системах передачи речевых сигналов, системах прослушивания для того, чтобы избежать постоянного анализа эфира вычислительной техникой, приводящего к износу техники, ложным срабатываниям, повышенному энергопотреблению и прочим нежелательным последствиям.

Известен способ детектирования речи [1], при реализации которого наряду с микрофоном предусмотрено использование инфракрасного датчика и ларингофона, что позволяет уловить движение лица, вибрации челюсти, движение шеи и рта диктора. Недостатком этого способа является применение дополнительных датчиков.

Известен способ детектирования речи на основе классификационных параметров сигнала [2], использующий метод линейного предсказания. Недостатками данного способа являются сложность в реализации и прекращение оценки уровня шума в речевых периодах. Быстрое изменение уровня шума на речевых участках сигнала может привести к неверным результатам. Данный способ работает только на ограниченном интервале стационарности шума.

Известен также способ детектирования речи [3], при котором анализ классификационных параметров сигнала осуществляется с помощью функций распределения плотности вероятности этих параметров. Недостатком такого способа является то, что функции распределения для действительных голосовых сигналов не известны и обычно предполагается, что они просты, например имеют распределение Гаусса. В результате такого предположения могут возникать неточности в ходе детектирования.

Близким к предлагаемому является известный способ детектирования речи [4], в котором наряду со сравнением значений классификационных параметров сигнала с их пороговыми значениями используется конечный автомат для сглаживания решения о наличии речи. Недостатками данного способа являются сложность в реализации и большой объем вычислений.

Наиболее близким к предлагаемому является способ детектирования речевого сигнала [5], совпадающий с заявляемым способом по наибольшему числу существенных признаков и принятый за прототип. В основу заявленного в [5] способа детектирования речевого сигнала положено сравнение значений временных классификационных параметров сигнала, таких как средняя энергия сигнала, среднее число переходов сигнала через нулевой

уровень, крутизна сигнала, с их пороговыми значениями, которые периодически обновляются. При этом решение о наличии речи в сигнале принимается в случае, если значение средней энергии сигнала превышает значение верхней границы порога или если оно не превышает значение верхней, но превышает значение нижней границы порога, и, в то же время, значение среднего числа переходов сигнала через нулевой уровень превышает значение порога, и значение крутизны сигнала превышает значение порога. Вычисление и обновление уровня шума на речевых участках осуществляет реже, чем на неречевых участках.

К недостаткам такого способа можно отнести отсутствие спектрального анализа сигнала, в результате чего в процессе детектирования речи не используются частотные характеристики сигнала, отсутствие механизма сглаживания переходов между речевыми и неречевыми участками, снижение частоты обновления уровня шума на речевых участках по сравнению с неречевыми участками, что может приводить к неверным результатам. Принятие решения о наличии речи в сигнале на основании превышения значением средней энергии сигнала верхней границы порога также является недостатком известного способа, поскольку может быть причиной ложных срабатываний при быстром увеличении уровня окружающего шума или возникновении импульсных помех. Кроме того, к недостаткам также относится использование функции кратковременной энергии сигнала, которая является чувствительной к большим уровням сигнала, поскольку каждый отсчет возводится в квадрат, вследствие чего значительно искажается соотношение между отсчетами сигнала.

Задачей предлагаемого способа является достоверное определение наличия речи в аудиосигнале на фоне изменяющейся акустической обстановки и выделение из этого сигнала участков, содержащих только речь, в режиме реального времени.

Указанная задача решается тем, что способ детектирования речевого сигнала, основанный на сравнении значений классификационных параметров сигнала с их пороговыми значениями, включающий операции: разбиение дискретизированного сигнала на временные окна, вычисление среднего числа переходов сигнала через нулевой уровень для каждого временного окна, сравнение вычисленного значения среднего числа переходов сигнала через нулевой уровень с пороговым значением среднего числа переходов сигнала через нулевой уровень, принятие решения о наличии речи в анализируемом сигнале, вычисление и обновление уровня шума, обновление значений порогов классификационных параметров сигнала, отличается тем, что после операции разбиения дискретизированного сигнала на временные окна вводят операции: буферизацию анализируемого сигнала, вычисление среднеквадратичного значения сигнала, сравнение вычисленного среднеквадратичного значения сигнала с пороговым среднеквадратичным значением, вычисление значения изменения спектра сигнала, сравнение вычисленного значения изменения спектра сигнала с пороговым значением изменения спектра сигнала, а в операции принятия решения решение о наличии речи в анализируемом сигнале принимают в случае, если вычисленные значения всех классификационных параметров сигнала больше соответствующих пороговых значений, операцию вычисления уровня шума производят путем отслеживания минимума среднеквадратического значения сигнала на заданном интервале сигнала, операцию обновления уровня шума осуществляют независимо от того, является ли текущее анализируемое временное окно речевым или неречевым.

Сущностью изобретения является то, что в данном способе детектирования речевого сигнала используют параметры как во временной области, а именно кратковременное среднеквадратичное значение сигнала и кратковременную функцию среднего числа пере-

## ВУ 15755 С1 2012.04.30

ходов через нуль, так и параметр в частотной области, а именно изменение спектра сигнала, при этом для определения порога кратковременного среднеквадратичного значения сигнала отслеживают минимум среднеквадратичного значения сигнала на заданном интервале, а для сглаживания переходов между речевыми и неречевыми участками применяют механизм предварительного и пост-буферов.

На фиг. 1 представлена структурная схема способа детектирования речевого сигнала.

На фиг. 2 приведена последовательность действий при обработке сигнала в частотной области.

На фиг. 3 приведена последовательность действий при обработке сигнала во временной области.

В соответствии с фиг. 1 способ детектирования речевого сигнала включает в себя последовательность этапов. Первый этап представляет собой загрузку сигнала, т.е. формирование последовательности отсчетов входного дискретизированного сигнала. Далее осуществляется формирование временных фреймов, содержащих отсчеты входного сигнала, и фильтрация сигнала в этих фреймах с помощью полосового фильтра с полосой пропускания 0,3-3,4 кГц, которая соответствует частотному диапазону речевого сигнала. Оптимальным размером фрейма является 0,07 секунды при частоте дискретизации сигнала 22050 Гц. Для повышения качества детектирования речи в сигнале в предлагаемом способе используется три классификационных параметра сигнала, а именно: среднеквадратичное значение сигнала, среднее число переходов сигнала через нуль и изменение спектра сигнала. Эти три параметра рассчитываются для каждого временного фрейма. Вычисление значения изменения спектра более подробно представлено на фиг. 2 и представляет собой фильтрацию сигнала с помощью окна Хэмминга, быстрое дискретное преобразование Фурье, оценку динамики изменения спектра сигнала, которая позволяет классифицировать текущий фрейм как содержащий стационарный сигнал или содержащий нестационарный сигнал. Речь относится к нестационарным сигналам, таким образом, на этапе сравнения значения изменения спектра сигнала с пороговым значением принимается решение о возможном присутствии речи в анализируемом временном фрейме речевого сигнала. Наряду с вычислением изменения спектра осуществляется вычисление среднеквадратичного значения сигнала и среднего числа переходов сигнала через нулевой уровень для каждого временного фрейма, которое более подробно представлено на фиг. 3. Вычисленное значение среднего числа переходов сигнала через нулевой уровень проверяется на принадлежность интервалу 0,3-3,4 кГц, и принимается решение, является ли текущий временной фрейм речевым. После вычисления среднеквадратичного значения сигнала производится расчет среднеквадратичного значения шума путем отслеживания минимального значения среднеквадратичного значения сигнала на протяжении заданного интервала сигнала, включающего как речевые, так и неречевые участки. При этом минимальное значение целесообразно рассчитывать как среднее из 10 минимальных среднеквадратичных значений сигнала, вычисленных на интервале 600 секунд при частоте дискретизации сигнала 22050 Гц. На этапе сравнения среднеквадратичного значения сигнала с его пороговым значением осуществляется получение коэффициента соответствия текущего временного фрейма входного сигнала речи, который рассчитывается как разность среднеквадратичного значения сигнала, вычисленного для текущего фрейма, и среднеквадратичного значения шума, деленная на коэффициент отношения речь/шум. В свою очередь, полученный таким образом коэффициент соответствия для анализируемого фрейма сравнивается с заданным порогом, принадлежащим интервалу (0,1). При этом 1 соответствует 100 % вероятности нахождения в исследуемом сигнале речи. На этапе при-

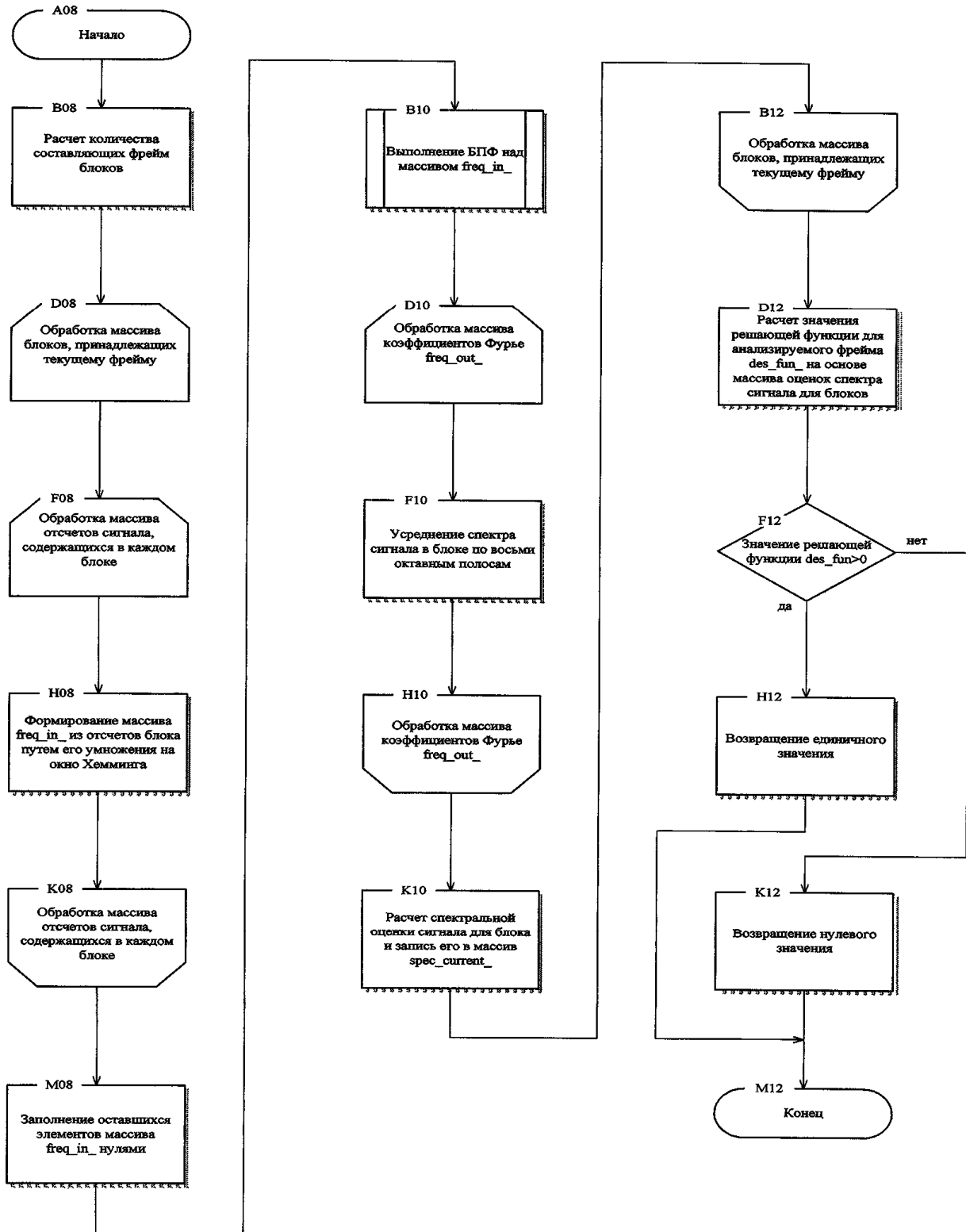
## ВУ 15755 С1 2012.04.30

нятия решения на основании результатов сравнения трех классификационных параметров сигнала с их пороговыми значениями принимается окончательное решение о наличии речи в текущем временном фрейме. Это решение представляет собой бинарное значение нуль или единица и рассчитывается как конъюнкция трех результатов сравнения с пороговыми значениями классификационных параметров. Исходя из полученного решения осуществляется формирование выходного сигнала, содержащего только речевые участки. При этом для сглаживания переходов между речевыми и неречевыми участками применяется механизм предварительного и пост-буферов. Механизм предварительного буфера заключается в том, что перед первым речевым участком сохраняется идущий перед ним в исследуемом аудиосигнале неречевой участок. Механизм пост-буфера состоит в том, что после последнего речевого участка сохраняется следующий за ним в исследуемом аудиосигнале неречевой участок. Оптимальными размерами предварительного и пост-буферов будут 3 секунды.

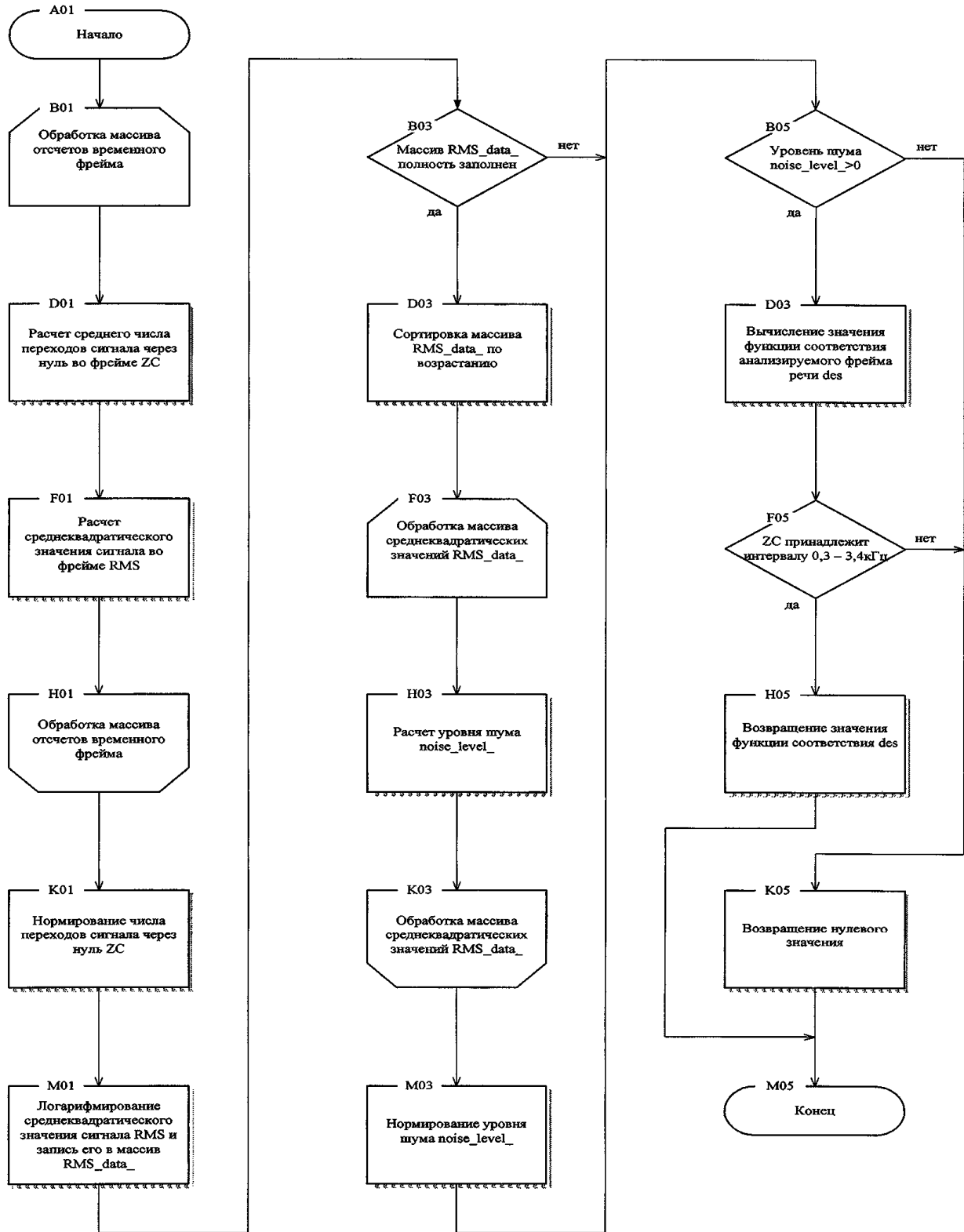
Совместное применение методов спектрального и временного анализа при сравнении значений классификационных параметров сигнала с их пороговыми значениями позволило разработать рациональный, быстродействующий и надежный способ детектирования речевого сигнала, совмещающий эффективность и компактность вычислений и обеспечивающий результаты, обладающие высокой степенью достоверности и точности. Данный способ обеспечивает детектирование речевого сигнала в режиме реального времени благодаря высокой скорости обработки аудиосигнала. Важное значение имеет продолжение оценки уровня шума в речевых периодах путем отслеживания минимального значения среднеквадратичного значения сигнала на протяжении заданного интервала сигнала, что позволяет своевременно обновлять и адаптировать порог среднеквадратического значения сигнала к изменяющимся условиям внешней акустической обстановки. Кроме того, механизм предварительного и пост-буферов позволяет решить проблему сглаживания переходов между речевыми и неречевыми участками. При этом отсутствует необходимость в применении дополнительных датчиков.

### Источники информации:

1. Пат. заявка RU 2004.123.352 А, МПК G 10L 11/00, 27.01.2006. Система детектирования речи.
2. Пат. US 4.672.669, МПК G 10L 11/02, H 04J 3/16, H 04J 3/17, G 10L 11/00, G 10L 5/00, 9.6.1987. Voice activity detection process and means for implementing said process.
3. Пат. US 6.993.481 В2, МПК G 10L 15/20, G 10L 15/08, G 10L 15/06, G 10L 15/04, 7.11.2002. Detection of speech activity using feature model adaptation.
4. Пат. US 6.453.285, МПК G 10L 11/02, G 10L 21/02, 10.8.1999. Speech activity detection for use in noise reduction system, and methods therefore.
5. Пат. US 5.459.814, МПК G 10L 009/00, 26.3.1993. Voice activity detector for speech signals in variable background noise.



Фиг. 2



Фиг. 3