УДК 004.934

РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ ОТДЕЛЬНЫХ СЛОВ

## А. Н. Савин, Н. Е. Тимофеева, А. С. Гераськин, Ю. А. Мавлютова

Савин Александр Николаевич, кандидат физико-математических наук, доцент кафедры дискретной математики и информационных технологий, Саратовский национальный исследовательский госу- дарственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, [savinan@info.sgu.ru](mailto:savinan@info.sgu.ru)

Тимофеева Надежда Евгеньевна, заведующий лабораторией теоретических проблем информатики и ее приложений кафедры дискретной математики и информационных технологий, Саратовский на- циональный исследовательский государственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, [timofeevane@yandex.ru](mailto:timofeevane@yandex.ru)

Гераськин Алексей Сергеевич, кандидат педагогических наук, доцент кафедры теоретических основ компьютерной безопасности и криптографии, Саратовский национальный исследовательский госу- дарственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, [gerascinas@mail.ru](mailto:gerascinas@mail.ru)

Мавлютова Юлия Альбертовна, старший лаборант лаборатории теоретических проблем информатики и ее приложений кафедры дискретной математики и информационных технологий, Саратовский на- циональный исследовательский государственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, [yuliyamav@yandex.ru](mailto:yuliyamav@yandex.ru)

Приведены результаты разработки программных модулей, реализующих систему распознавания речи на основе скрытых Марковских моделей отдельных слов и использования линейного предсказания при кодировании признаков звукового сигнала. Обосновывается структура системы распознавания речи, использующая скрытые марковские модели отдельных слов, состоящая из четырех модулей: модуль выделения слов из звукового потока, модуль анализа признаков слова, модуль обучения скрытых марковских моделей и модуль распознавания слов. Приводятся алгоритмы формирования скрытых марковских моделей с лево-правой топологией для отдельных слов требуемого словаря команд системы управления объекта, основанные на кодировании признаков звукового сигнала, использующего линейные предсказания. Приведены результаты оценки достоверности последова- тельности наблюдений, соответствующих отдельным словам, получаемым с помощью предложенного алгоритма обработки. Разработанные программные модули позволяют эффективно подготавливать необходимые исходные данные и формировать таким образом требуемый словарь команд системы управления объекта, строить скрытые марковские модели отдельных слов, проводить их обучение с помощью алгоритма Баума – Велша. Построенные словари команд предполагается использовать в интеллектуальных системах управления различными объектами.

*Ключевые слова:* скрытые марковские модели, кепстральный анализ, распознавание речи, метод Баума – Велша.

DOI: 10.18500/1816-9791-2017-17-4-452-464

# ВВЕДЕНИЕ

Распознавание голоса в компьютерных системах весьма распространено. Распо- знавание речи и, как следствие, голосовая идентификация нашли свое применение

◯*c Савин А. Н., Тимофеева Н. Е., Гераськин А. С., Мавлютова Ю. А., 2017*

во всех сферах человеческой деятельности. Благодаря системам распознавания ре- чи обеспечивается безопасность от несанкционированного проникновения в защи- щенную зону. Такие системы содержат базу данных голосов сотрудников, имеющих доступ к защищаемой зоне, и предотвращают допуск людей, чьих голосов в ней нет [1, 2].

В настоящее время широко разрабатываются и внедряются интеллектуальные системы управления различными объектами, которые позволяют осуществлять кон- троль за объектами в реальном времени. Управление такими системами можно осу- ществлять различными способами, одним из них является метод голосовых команд. При этом защиту объекта от несанкционированного доступа можно решить, исполь- зуя индивидуальные особенности голоса каждого человека.

Уровень развития современной микропроцессорной техники (например, мобиль- ные устройства связи) позволяет использовать сложные вычислительные алгоритмы, основанные на цифровой потоковой обработке статистических данных в реальном времени. Поэтому разработка таких алгоритмов является весьма актуальной.

Одним из путей решения вышеуказанных задач является использование для рас- познавания фрагментов речи математического аппарата скрытых марковских моде- лей (СММ) [3]. Данная работа посвящена разработке алгоритма и соответствующего программного модуля, осуществляющего формирование СММ для отдельных слов требуемого словаря команд системы управления объекта, на основе кодирования признаков звукового сигнала, использующего линейные предсказания.

# СТРУКТУРА СИСТЕМЫ РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ СММ

Рассмотрим дискретную систему, имеющую конечное множество из *N* состоя- ний — *S* = {*s*1*, . . . , sN* }, в каждом из которых она может принимать одно из *M* значений из набора наблюдаемых параметров *V* = {*v*1*, . . . , vM* } — алфавита. Состоя- ние системы *qt* в момент времени *t*, принимающее одно из *N* значений множества *S*, зависит только от её состояния *qt*−1 в момент времени *t* −1, а значение наблюдаемого параметра *ot* в момент времени *t* зависит только от состояния *qt*, т. е. не зависит от времени.

Вероятности переходов между состояниями системы задаются матрицей *A*. Ве- роятности выпадения каждого из *M* значений наблюдаемого параметра системы в каждом из *N* состояний системы задаются набором векторов *B*. Вероятность по- явления некоторого начального состояния системы задаётся вектором *π*. При этом последовательность состояний, в которых пребывает система *Q* = *q*1*, . . . , qT* , внеш- нему наблюдателю не видна, а видит он только последовательность наблюдений *O* = *o*1*, . . . , oT* (здесь *T* — длина последовательности), т. е. система ведёт себя как

«чёрный ящик». Модель такой системы получила название СММ и в компактной записи обозначается *λ* = (*A, B, π*) [2].

Для моделирования отдельного слова может быть выбрана лево- правая СММ (рис. 1) на основе предположения о том, что в каждый момент времени система переходит в



новое состояние [4]. Соответственно

неизвестное число скрытых состоя- ний *N* в этом случае определяется

Рис. 1. Структура лево-правой СММ

Fig. 1. The structure of the left-right hidden Markov models (HMM)

длиной и количеством сегментов, на которые слово разбивается при анализе его признаков. Процесс распознавания с использованием СММ предполагает два этапа (рис. 2).



Построение и обучение СММ по речевой базе, содержащей записи слов диктора

Распознавание команд в звуковом потоке

Распознавание слова

Анализ признаков

Выделение слова

Звуковой поток/файл

База СММ слов

Обучение СММ

Анализ признаков

Выделение последовательности одинаковых слов

Речевая база

Код слова

Рис. 2. Структура системы распознавания речи на основе использования СММ Fig. 2. Structure of the speech recognition system based on the use of HMM

В режиме обучения элементы системы имеют следующее функциональное назна- чение:

•

речевая база содержит записи слов, повторяющихся несколько раз для обеспе-

чения адекватности получаемых СММ, которые будут доступны для распозна- вания;

выделение последовательности одинаковых слов из файла речевой базы с по- мощью предварительной обработки (подавление шума, фильтрация и т. д.); анализ признаков и определение алфавита слова *V* , по которому формируется последовательность наблюдений *O*;

•

•

обучение СММ — подбор параметров СММ, чтобы она как можно лучше опи- сывала реальную наблюдаемую последовательность *O* символов алфавита *V* анализируемого слова;

•

сохранение СММ в базе — словаре.

•

В режиме распознавания:

выделение слова из входного звукового потока с помощью предварительной

•

обработки;

анализ признаков распознаваемого слова и формирование соответствующей по- следовательности наблюдений *O*;

•

распознавание слова с использованием базы СММ и генерация кода распозна- ваемого слова.

•

Таким образом, для реализации данной структуры необходимо всего 4 модуля:

модуль выделения слов из звукового потока, модуль анализа признаков слова, модуль обучения СММ с базой моделей, модуль распознавания слов.

# АЛГОРИТМ ВЫДЕЛЕНИЯ ОТДЕЛЬНЫХ СЛОВ ИЗ ЕДИНОГО ЗВУКОВОГО ФАЙЛА

В режиме обучения файл должен содержать несколько раз произнесенное одним диктором требуемое слово. Это необходимо для получения достоверной последо- вательности наблюдений *O*, соответствующей данному слову. На рис. 3 приведена блок-схема алгоритма предварительной обработки звукового файла, основанного на

Начало процедуры выделения команд из звукового файла

Вычисление среднего значения длительности команд и округление его до ближайшего числа, кратного «Шагу сегментов (*MА* )»(см. далее «Анализ признаков»)

Завершение процедуры выделения команд из звукового файла

Вырезка из входного файла отрезков одинаковой длины,

соответствующих командам, и сборка их в массив для дальнейшей обработки

Загрузка файла с повторяющимся словом

записанного ранее одним диктором

Удаление коротких «пауз» (меньше минимальной длины команды), возникающих внутри команды, например, между слогами слова

Удаление коротких «команд» (меньше минимальной длины команды), возникающих внутри паузы, например, при наличии импульсных шумов в паузах между командами

Вычисление индексов, в которых огибающая начинает превышать заданный уровень шума

(начало «команды»), а затем становится ниже уровня шума (конец «команды» - начало «паузы»)

Нормировка входного звукового файла дана таким образом, чтобы максимальная амплитуда была равна единице.

Выделение огибающей входного сигнала:

1. «выпрямление» сигнала с помощью операции «Модуль»;
2. фильтрация «выпрямленного» сигнала с помощью НЧ FIR фильтра с частотой среза примерно 10 Гц

Удаление первых 0.1 с из записи для исключения переходных процессов записи и фильтрации

Изменение частоты дискретизации звукового файла, так как для потокового распознавания необходимо, чтобы она была примерно равна 7 Ks/с, т.е. не очень большая, но достаточная, чтобы оцифровать сигналы с частотой до 3.5 КГц

Ограничение спектра сигнала звуковым диапазоном голоса с помощью полосно-

пропускающего фильтра Батерворта. При этом убирается постоянная составляющая из сигнала

Рис. 3. Блок-схема алгоритма выделения отдельных повторяющихся слов из звукового файла

Fig. 3. A flowchart of an algorithm for selecting separate repetitive words from a sound file

вычислении огибающей и выделении на ее основе участков файла соответствующих повторяющимся словам.

При этом на выходе формируется массив отрезков звукового файла одинаковой длины, соответствующих повторяемому слову, что позволяет использовать усреднен- ные входные данные при обучении СММ слова, делая её тем самым более адекват- ной.

Алгоритм выделения отдельных слов из звукового потока встроен в модуль построения СММ слов, реализованный в среде графического программирования LabVIEW компании National Instruments [5]. На рис. 4 показан процесс выделе- ния команды из звукового файла, содержащего десять раз повторяющееся слово

«Вперёд».

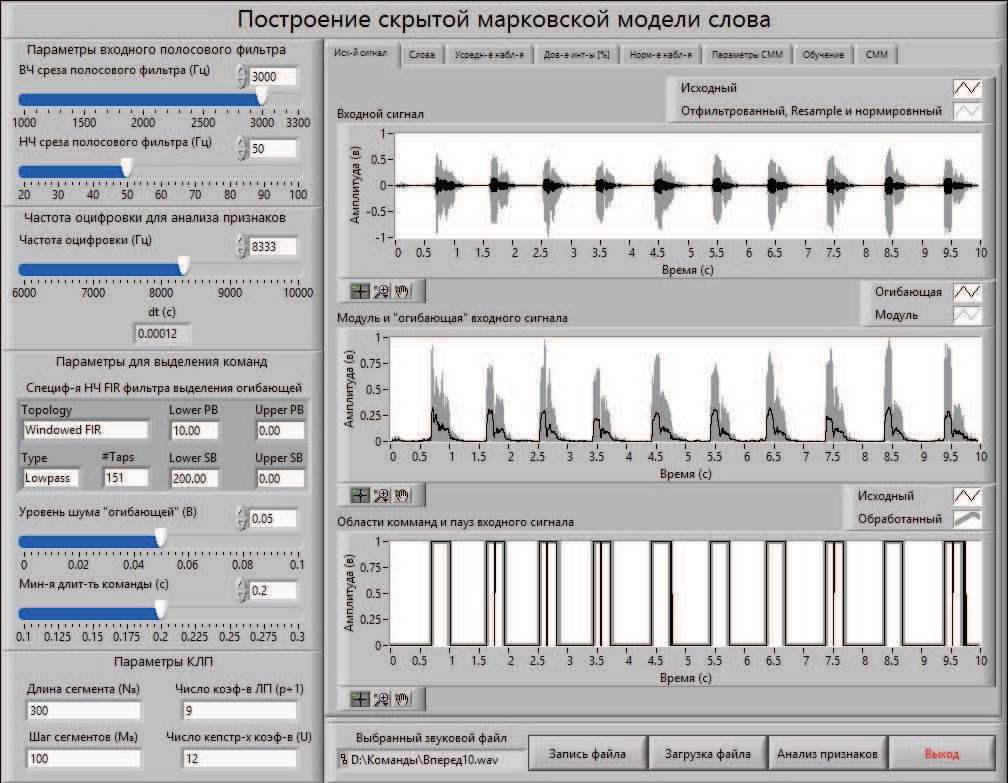


Рис. 4. Выделение повторяющихся слов из звукового файла Fig. 4. Selecting duplicate words from a sound file

На первом графике рис. 4 чёрным изображен исходный сигнал, серым — от- фильтрованный, с измененной частотой дискретизации и нормированный. Из запи- си убираются первые 0.1 с, соответствующие переходным процессам при включении микрофона и предварительной фильтрации.

Выделение слова осуществляется путем анализа огибающей сигнала. Индексы, в которых огибающая начинает превышать заданный в начале уровень шума, соот- ветствуют началу команды. Индексы, в которых огибающая становится ниже уровня шума, — концу команды. Паузы внутри команды отсеиваются с помощью задан- ной заранее минимальной длительности команд. Таким же образом отсеиваются и лишние шумы.

Элементы управления модуля (граничные частоты среза входного полосового фильтра, частота дискретизации сигнала для анализа признаков, параметры НЧ

фильтра огибающей, уровень шума огибающей, минимальная длительность команды) позволяют подбирать требуемые параметры на этапе выделения команд для обеспе- чения построения адекватных СММ слов.

# АЛГОРИТМ АНАЛИЗА ПРИЗНАКОВ СЛОВ

Для системы распознавания речи каждому слову необходимо сопоставить набор признаков. Этот процесс в [4] предложено осуществлять на основе анализа периодич- ности спектра фрагментов звукового сигнала (кепстральный анализ), предварительно обработанного с помощью алгоритмов линейного предсказания. Такой процесс назы- вается кодированием на основе линейного предсказания (КЛП). Алгоритм анализа признаков на основе КЛП, используемый при распознавании отдельных слов, приве- ден на рис. 5, 6.

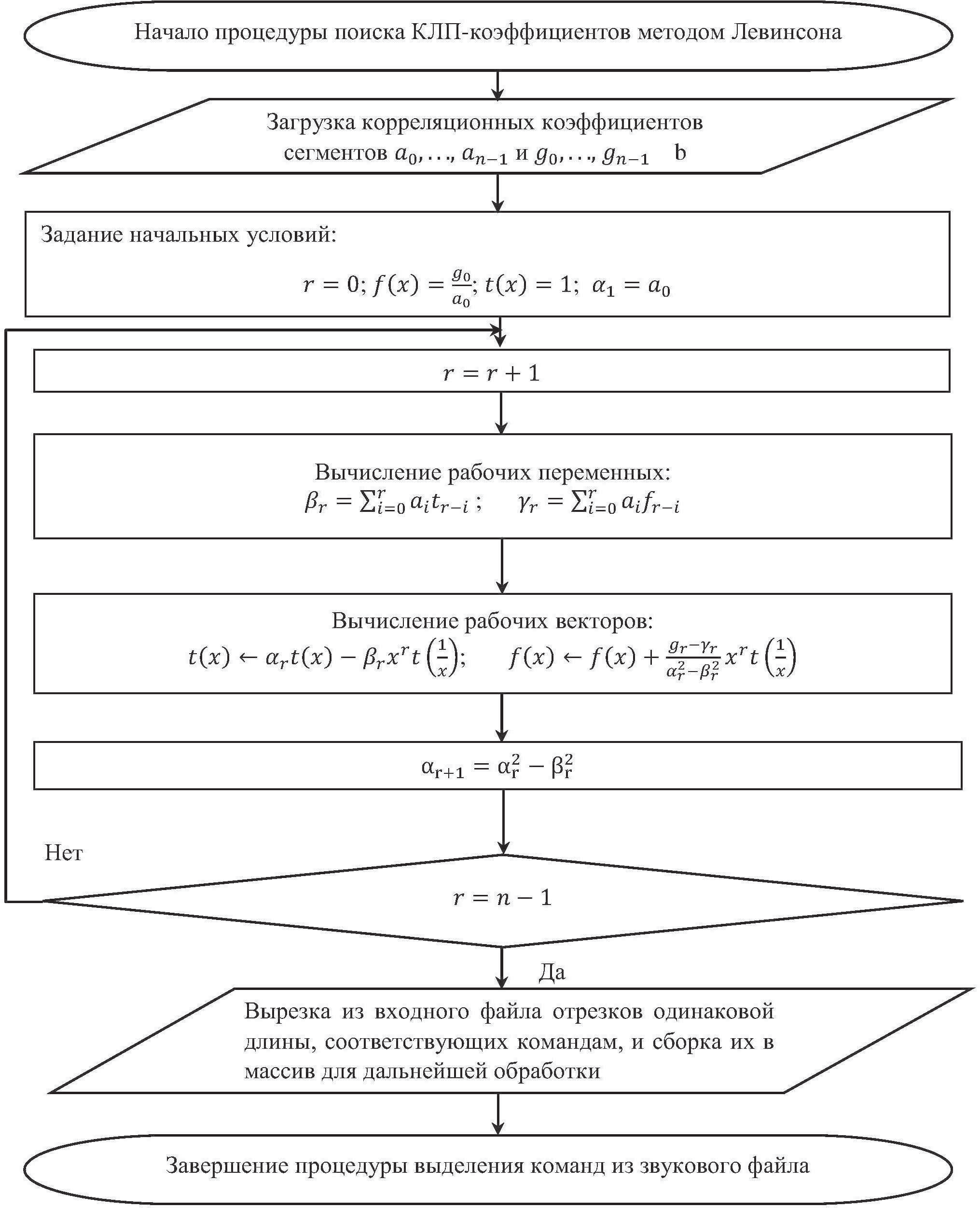


Рис. 5. Блок-схема процедуры алгоритма Левинсона вычисления КЛП-коэффициентов Fig. 5. Block diagram of the procedure of Levinson’s algorithm for calculating linear prediction

coding (LPC) coefficients

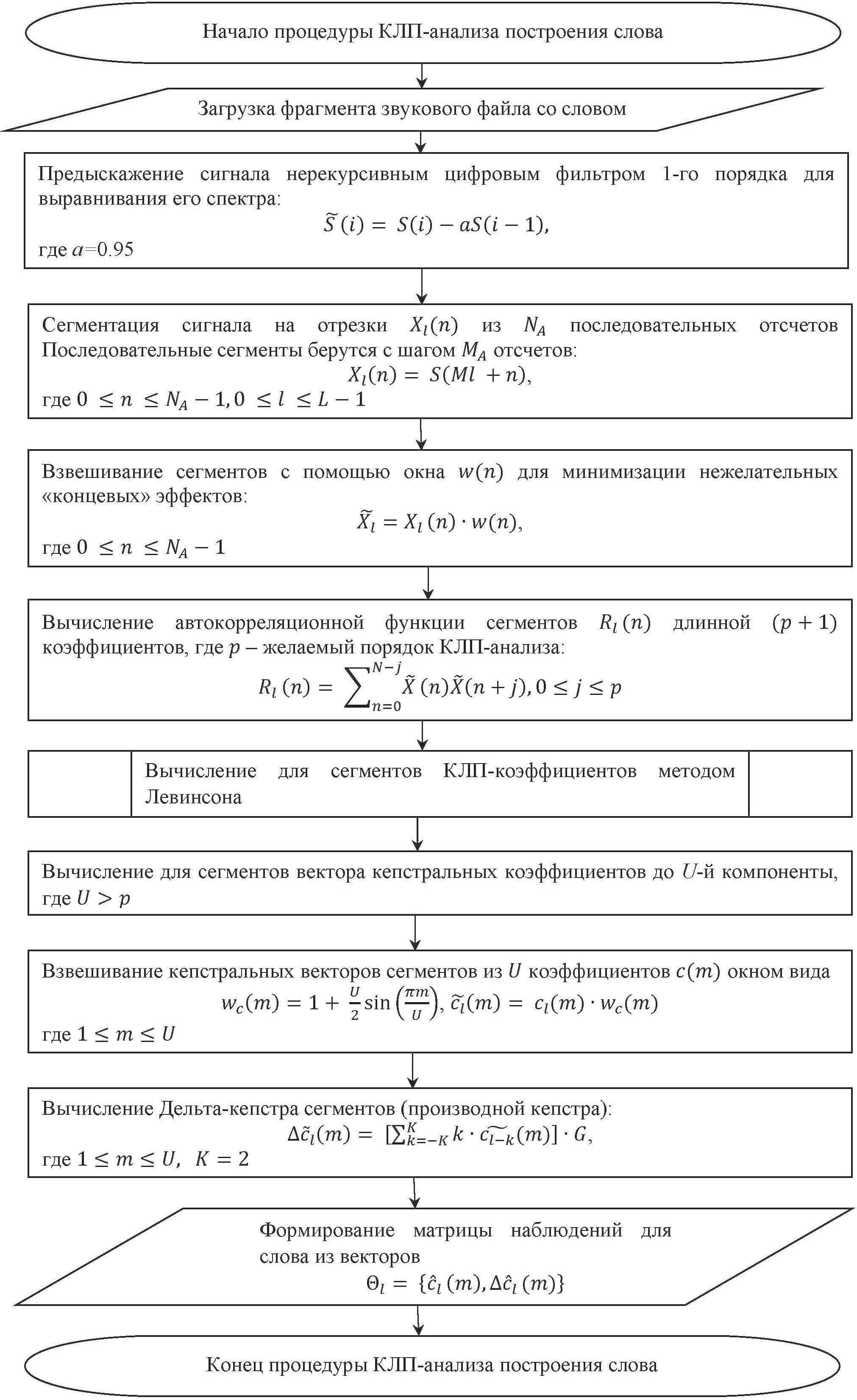


Рис. 6. Блок-схема алгоритма анализа признаков слова на основе КЛП Fig. 6. Flowchart of word analysis algorithm based on LPC

Достоинством спектральной обработки звуковых сигналов является то, что при переходе из временной области в частотную представление информации становится

более наглядным, компактным. Причем, чем более «простым» является сигнал во временной области, тем в большей степени происходит сжатие информации.

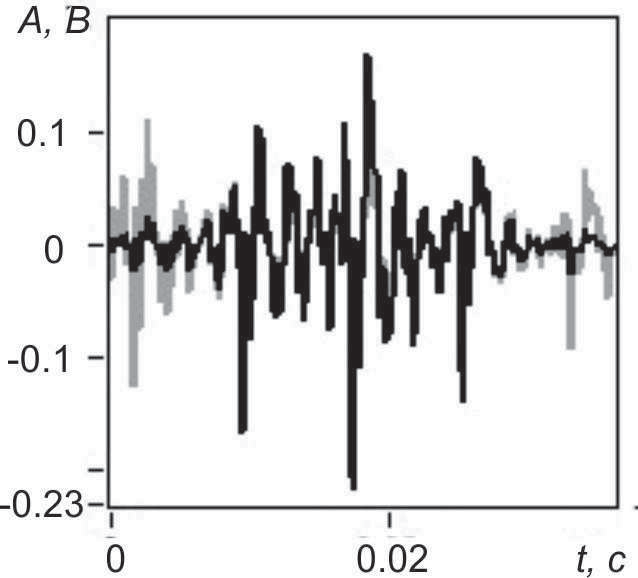
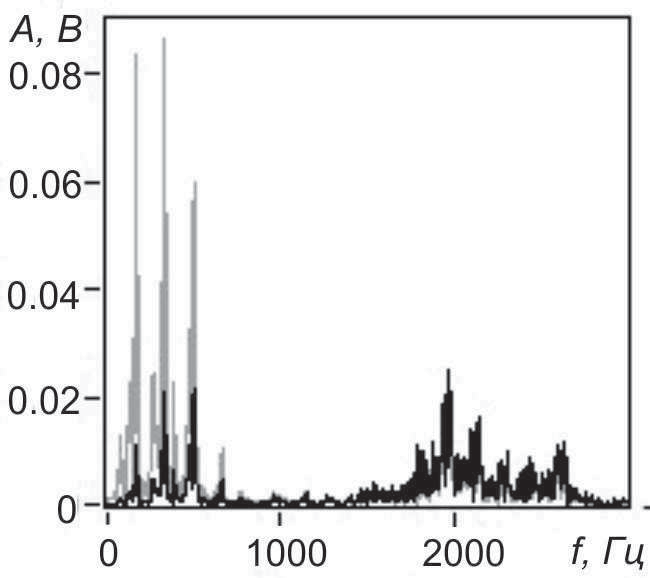
Выявление периодичности в спектре (кепстральный анализ) позволяет более до- стоверно и точно охарактеризовать особенности произношения дикторов. При этом спектральная информация представляется еще более компактно. Каждый гармони- ческий ряд исходного спектра представляется в идеале всего одной составляющей в кепстре [4].

Использование линейного предсказания, основанного на автокорреляционной фильтрации, должно улучшать отношение сигнал – шум исходного сигнала и убирать из него случайные артефакты. Вычисление коэффициентов линейного предсказания осуществляется с помощью алгоритма Левинсона (см. рис. 5) [6].

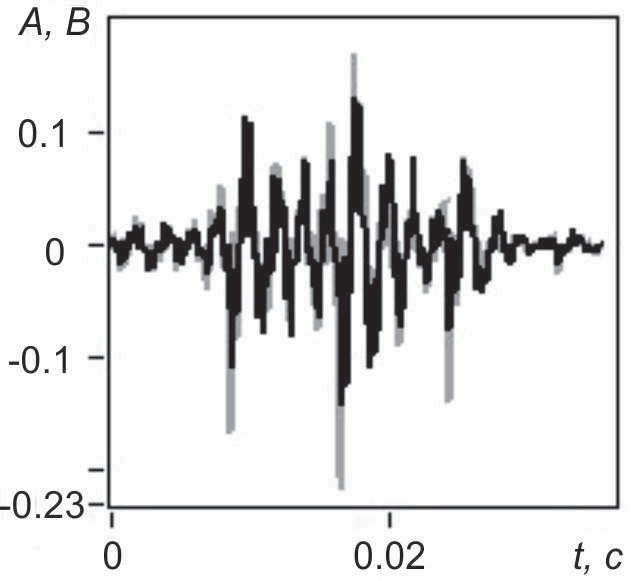
В процессе анализа признаков слова каждый участок, выделенный ранее из файла и соответствующий повторяющемуся слову, разбивается на небольшие перекрываю- щиеся отрезки – сегменты и затем обрабатывается согласно алгоритму, приведенному на рис. 5, 6. Как видно (рис. 7, а), в результате предискажения сигнала происходит выравнивание спектра, что обеспечивает равноценность спектральных компонент при анализе признаков.

Оконное взвешивание (рис. 7, *б*) уменьшает сигнал на концах сегментов и уве- личивает в центре, минимизируя нежелательные концевые эффекты.

Линейное предсказание (рис. 7, *в*) на основе алгоритма Левинсона (см. рис. 5) убирает сглаживает выбросы и случайные артефакты в анализируемом сигнале.



а / *a б / b*



*в / с*

Рис. 7. Спектр исходного (**–**) и предискаженного (–) сигналов (а), исходный (–) и взвешенный (**–**) сегменты сигнала (*б*), исходный взвешенный (–) и предсказанный (**–**) сегменты сигнала (*в*)

Fig. 7. The spectrum of the initial (**–**) and pre-faded (–) signals (*a*), the initial (–) and weighted (**–**) signal segments (*b*), the initial weighted (–) and predicted (**–**) signal segments (*c*)

На выходе алгоритма формируется необходимая для распознавания слова матри- ца, строки которой образуются конкатенацией взвешенного кепстрального и соот- ветствующего взвешенного дельта-кепстрального векторов сегментов. Каждая такая строка является набором признаков сегмента — вектором наблюдений и соответству- ет одному символу из алфавита *V* СММ слова в последовательности наблюдений *O*. Количество строк определяет число состояний *N* , в которых находилась лево-правая СММ слова.

Настройка параметров КЛП (число отсчётов в сегменте *NA*, число отсчётов в смещении сегментов *MA*, порядок КЛП-анализа *p*, число кепстральных коэффици- ентов *U* ) осуществляется посредством соответсвующих элементов (см. рис. 5).

# ПОСТРОЕНИЕ СММ СЛОВА

В процессе построения СММ слова для повышения её адекватности необходимо использовать матрицу векторов наблюдений, полученную статистическим усреднени- ем матриц векторов наблюдений повторяющихся слов. При этом усредняются векто- ра наблюдений, соответствующие одним и тем же моментам времени повторяющихся слов.

Оценка достоверности выборочных средних значений признаков сегментов (эле- ментов векторов наблюдений) повторяющихся слов проводится с помощью довери- тельных интервалов, вычисляемых при уровне статистической значимости *α* = 0*,* 05. Сравнение степени разброса, т. е. оценка однородности выборочных дисперсий значений элементов векторов наблюдений, вычисленных по результатам анализа при- знаков повторяющихся слов, осуществляется с использованием критерия Кохрена [7] Обеспечение статистически значимых выборочных средних значений элементов векторов наблюдений и однородности их дисперсий, т. е. получение достоверной по- следовательности наблюдений для анализируемого слова, достигается подбором па-

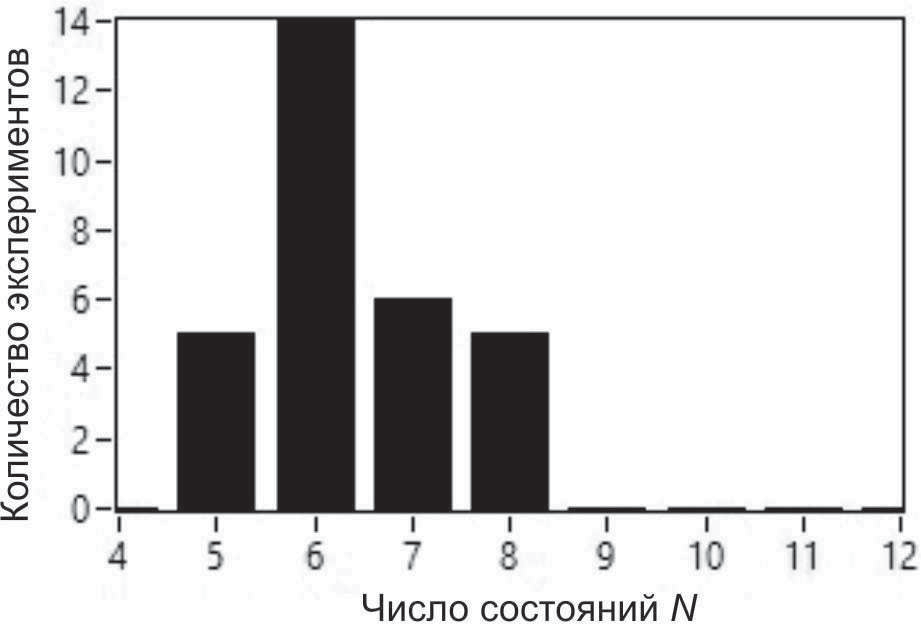
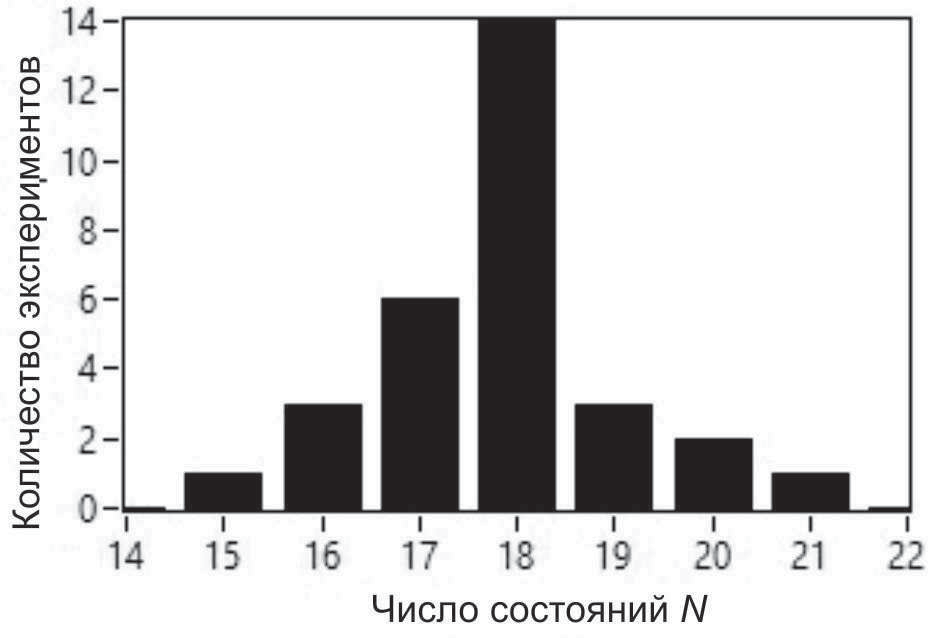
раметров обработки входного сигнала и параметров КЛП (см. п. 2, 3).

Оценку расстояния между символами алфавита *V* слова — усреднёнными век- торами наблюдений — было предложено делать с помощью евклидовой нормы. При этом для обеспечения равнозначности признаков при вычислении расстояния про- водилась их нормировка. В качестве нормирующего для каждого элемента вектора наблюдений использовался диапазон его изменения в матрице, расширенный с учё- том доверительного интервала.

Формирование алфавита СММ слова осуществляется удалением повторяющихся строк из нормированной матрицы средних значений наблюдений, если такие имеют- ся. При этом сравнивается расстояние между текущей строкой и остальными. Если оно меньше некоторой заданной величины, то строка с большим индексом удаля- ется, так как считаем, что эти строки соответствуют одному и тому же символу. Соответственно число строк получившейся прореженной матрицы определяет коли- чество символов *M* алфавита *V* , индексы строк являются значениями алфавита, а сами строки — признаками символов.

Последовательность наблюдений слова определяется сравнением строк матрицы алфавита *V* (прореженной) с исходной нормированной матрицей средних значений наблюдений. Если расстояние между строками меньше некоторой заданной величи- ны, использованной при построении алфавита, то индекс строки матрицы алфави- та *V* записывается в последовательность наблюдений *O*. Длина последовательности наблюдений *T* равна числу строк исходной матрицы, а число состояний *N* равно *T* в случае лево-правой СММ слова.

На рис. 8 приведены результаты экспериментов по определению числа состоя- ний *N* для слов «Вперёд» и «Стоп». На вход модуля построения СММ для каждого из этих слов подавалось тридцать звуковых файлов, содержащих по пятнадцать повторений, произнесённых одним диктором. Во всех случаях число состояний *N* совпадало с количеством символов *M* алфавита *V* , что соответствует предположе- нию о лево-правой структуре СММ (см. рис. 1) для этих слов из-за отсутствия в них повторяющихся звуков.



а / *a б / b*

Рис. 8. Количество полученных в экспериментах состояний для слов «Вперёд» (а) и «Стоп» (*б*)

Fig. 8. The number of states obtained in the experiments for the words „Forward“ (*a*) and „Stop“ (*b*)

Как видно из рис. 8, закон распределения отклонений числа состояний *N* от средних значений близок к нормальному. Соответственно для слова «Вперёд» сред- нее значение *N* составило 18 2*.*6% при 95% -й доверительной вероятности, а для слова «Стоп» — 6 5*.*7%.

±

±

Следовательно, данные, полученные с помощью предложенного алгоритма пред- варительной обработки звукового файла в совокупности с КЛП-анализом, являются достаточно надёжными и их можно использовать для построения СММ слова.

Процесс построения начинается генерированием по известному числу состояний *N* и количеству символов *M* алфавита *V* исходной СММ *λ* = (*A, B, π*), имеющей случайные параметры. При этом матрица вероятностей переходов между состояни- ями A, матрица вероятностей каждого наблюдения в каждом состоянии B, а также вектор вероятностей начального состояния *π* должны удовлетворять стохастическим ограничениям [3].

Далее необходимо так подобрать параметры исходной СММ, чтобы вероят- ность соответствия последовательности наблюдений, сгенерированной этой СММ *λ*∗ = (*A*∗*, B*∗*, π*∗) и полученной ранее последовательности наблюдений *O* = *o*1*, . . . , oT* слова, была максимально возможной. То есть исходную СММ *λ* = (*A, B, π*), име- ющую вероятность *p*(*O*|*λ*) генерирования заданной последовательности наблюде- ния *O* = *o*1*, . . . , oT* слова, надо обучить по этой последовательности наблюде- ния *O* = *o*1*, . . . , oT* , чтобы вероятность *p*(*O λ*∗) генерирования последовательности

|

*O* = *o*1*, . . . , oT* , обученной СММ *λ*∗ = (*A*∗*, B*∗*, π*∗), была максимально возможной.

Одним из вариантов обучения СММ *λ* = (*A, B, π*) по заданной последовательно-

сти наблюдений *O* = *o*1*, . . . , oT* является применение алгоритма Баума – Велша [3].

Алгоритм позволяет уточнять параметры исходной СММ *λ* = (*A, B, π*) таким образом, чтобы у уточнённой СММ *λ*∗ = (*A*∗*, B*∗*, π*∗) вероятность *p*(*O λ*∗) увеличи- валась. Итеративное применении алгоритма до схождения в одной точке позволяет максимизировать *p*(*O*|*λ*∗), т. е. настроить СММ *λ*∗ = (*A*∗*, B*∗*, π*∗) на заданную после- довательность наблюдений *O* = *o*1*, . . . , oT* слова. На рис. 9 приведены зависимости изменений вероятности *p*(*O λ*∗) и её приращения ∆*p*(*O λ*∗) на каждом итерационном шаге при настройке методом Баума – Велша СММ на слово «Стоп». Эти зависимости имеют характерный для метода Баума – Велша вид.

|

| |

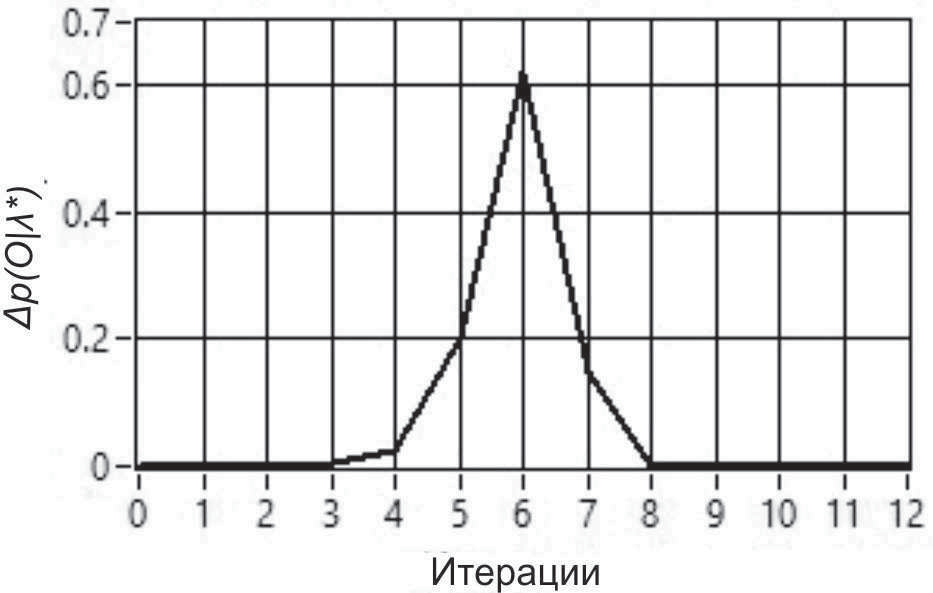
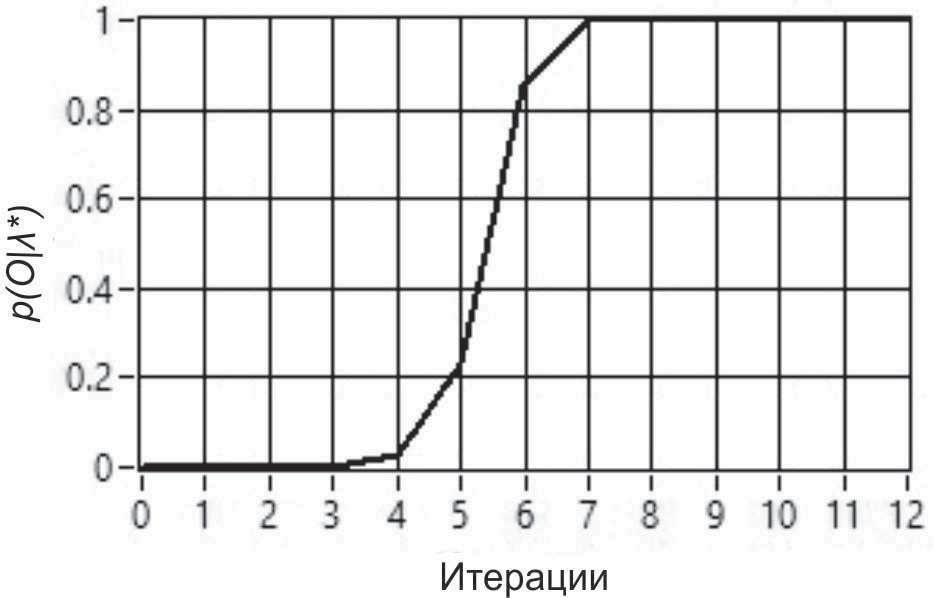
В начале обучения значения вероятности *p*(*O λ*∗) имеют, как правило, величи- ны меньшие или сравнимые с используемой для оценки сходимости положительной величиной *ε* (см. рис. 9, а), но разность значений *p*(*O*|*λ*∗) на каждом шаге увеличи- вается, т. е. приращение ∆*p*(*O*|*λ*∗) *>* 0 растет (см. рис. 9, *б*).

|

При завершении обучения значения вероятности *p*(*O λ*∗), как правило, сходятся к некоторой величине, при этом приращение ∆*p*(*O λ*∗) *>* 0, но оно начинает умень- шаться и стремиться к 0.

|

|



а / *a б / b*

Рис. 9. Изменения вероятности СММ *p*(*O*|*λ*∗): а — и её приращения ∆*p*(*O*|*λ*∗); *б* — при обучении методом Баума – Велша

Fig. 9. Changes in the probability of the HMM *p*(*O*|*λ*∗): *a* — and its increment ∆*p*(*O*|*λ*∗);

*b* — when learning by the Baum – Welsh method

Следовательно, для корректной оценки сходимости итерационного процесса обу- чения необходимо контролировать не только величину приращения ∆*p*(*O λ*∗), но и знак его изменения, т. е. для завершения обучения должно выполняться условие

|

∆*p*(*O*|*λ*∗) ≤ *ε* при уменьшении ∆*p*(*O*|*λ*∗).

В разработанном модуле построения СММ для оценки сходимости процесса обу- чения использовано значение *ε* = 1*.*1 ∗ 10−19, соответствующее машинной точности. Недостатком алгоритма Баума – Велша при обучении СММ является поиск ло- кального максимума *p*(*O λ*∗), а не глобального. Поэтому для достижения хорошего результата требуется, как правило, несколько запусков при различных начальных

|

условиях.

Таким образом, используя последовательности наблюдений и алфавит моделиру- емых слов, получаемые на первых этапах обработки, с помощью алгоритма Баума – Велша можно строить соответствующие адекватные СММ для систем распознавания речи.

# ЗАКЛЮЧЕНИЕ

Разработанный программный модуль позволяет эффективно подготавливать необ- ходимые исходные данные на основе кодирования признаков звукового сигнала, ис- пользующего линейные предсказания, строить СММ отдельных слов и проводить их обучение с помощью алгоритма Баума – Велша. Построенные СММ слов предполага- ется использовать в интеллектуальных системах управления различными объектами.

## Библиографический список

1. *Жиляков Е. Г., Бабаринов С. Л., Чадюк П. В.* Исследование сервиса компании Google Inc. по распознаванию русской речи // Научные ведомости БелГУ. Сер. История. По- литология. Экономика. Информатика. 2013. № 15(158), вып. 27/1. С. 247–255.
2. *Титов Ю. Н.* Современные технологии распознавания речи // Вестн. Тамб. ун-та. Сер. Естественные и технические науки. 2006. Т. 11, вып. 4. С. 571–574.
3. *Савин А. Н., Тимофеева Н. Е., Гераськин А. С., Мавлютова Ю. А.* Разработка ком- понентов программного комплекса для потоковой фильтрации аудиоконтента на основе использования скрытых марковских моделей // Изв. Сарат. ун-та. Нов. сер. Сер. Мате- матика. Механика. Информатика. 2015. Т. 15, вып. 3. С. 340–350. DOI: 10.18500/1816-

9791-2015-15-3-340-350.

1. *Рабинер Л. Р.* Скрытые Марковские модели и их применение в избранных приложениях при распознавании речи : Обзор // ТИИЭР. 1989. Т. 77, № 2. С. 86–120.
2. Портал компании National Instruments Russia. URL: [http://www.labview.ru](http://www.labview.ru/) (дата обра- щения: 15.12.2017).
3. *Блейхут Р.* Быстрые алгоритмы цифровой обработки сигналов. М. : Мир, 1989. 448 с.
4. *Адлер Ю. П., Маркова Е. В., Грановский Ю. В.* Планирование эксперимента при поиске оптимальных условий. М. : Наука, 1976. 279 с.

**Образец для цитирования:**

*Савин А. Н., Тимофеева Н. Е., Гераськин А. С., Мавлютова Ю. А.* Разработка системы распознавания речи на основе скрытых марковских моделей отдельных слов // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2017. Т. 17, вып. 4. С. 452–464. DOI: 10.18500/1816-9791-2017-17-4-452-464.

**Development of Speech Recognition Systems Based on Hidden Markov Models of Individual Words**

## A. N. Savin, N. E. Timofeeva, A. S. Geraskin, Yu. A. Mavlutova

Alexander N. Savin, orcid.org/0000-0001-5148-9166, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, [savinan@info.sgu.ru](mailto:savinan@info.sgu.ru)

Nadezhda E. Timofeeva, orcid.org/0000-0002-3976-3115, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, [timofeevane@yandex.ru](mailto:timofeevane@yandex.ru)

Aleksej S. Geraskin, orcid.org/0000-0002-3118-1022, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, [gerascinas@mail.ru](mailto:gerascinas@mail.ru)

Yuliya A. Mavlutova, orcid.org/0000-0002-1190-90064, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, [yuliyamav@yandex.ru](mailto:yuliyamav@yandex.ru)

The results of the development of software modules implementing the speech recognition system based on the hidden Markov models of individual words and the use of linear prediction in the coding of signs of an audio signal are presented. The structure of the speech recognition system is based on the hidden Markov models of individual words, consisting of four modules: a module for extracting words from the sound stream, a module for analyzing the features of a word, a module for learning the hidden Markov models, and a word recognition module. Algorithms for the formation of hidden Markov models with left-right topology for individual words of the required dictionary of commands of the object control system are based on the coding of signs of a sound signal using linear predictions. Results of an estimation of reliability of a sequence of observations corresponding to separate words obtained with the help of the proposed processing algorithm are given. The developed software modules allow to prepare efficiently the necessary initial data and thus form the required dictionary of commands of the object management system, build hidden Markov models of individual words, and conduct their training using the Baum – Welsh algorithm. The designed command dictionaries are supposed to be used in intelligent control systems for various objects.

*Key words:* Hidden Markov models, cepstral analysis, speech recognition, method of Baum – Welch.

## References

1. Zhilyakov E. G., Babarinov S. L., Chadyuk P. V. Google Inc. Russian Speech Recognition Service Research. *Belgorod State University Scientific Bulletin. Ser. History. Political science. Economics. Information technologies*, 2013, no. 15(158), iss. 27/1, pp. 247–255 (in Russian).
2. Titov Y. N. Modern technologies of speech recognition. *Tambov University Reports. Ser. Natural and Technical Sciences*, 2006, vol. 11, iss. 4, pp. 571–574 (in Russian).
3. Savin A. N., Timofeeva N. E., Geraskin A. S., Mavlutova Yu. A. The development of software components for streaming audio content filtering through the use of hidden Markov models. *Izv. Saratov Univ. (N.S.), Ser. Math. Mech. Inform.*, 2015, vol. 15, iss. 3,

pp. 340–350. DOI: 10.18500/1816-9791-2015-15-3-340-350.

1. Rabiner L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, vol. 77, no. 2. pp. 257–286. DOI: 10.1109/5.18626.
2. *Portal of company National Instruments Russia*. Available at: [http://www.labview.ru](http://www.labview.ru/) (accessed 25 December, 2012).
3. Blahut R. E. *Fast Algorithms for Digital Signal Processing*. Addison-Wesley Publ. Co; Repr. with corrections edition, 1987. 441 p. (Russ. ed. : Moscow, Мir, 1989. 448 p.).
4. Adler Ju. P., Markova E. V., Granovskij Ju. V. *Planirovanie eksperimenta pri poiske optimal’nykh uslovii* [Planning an experiment searching for optimal conditions]. Moscow, Nauka, 1976. 279 p. (in Russian).

**Cite this article as:**

Savin A. N., Timofeeva N. E., Geraskin A. S., Mavlutova Yu. A. Development of Speech Recog- nition Systems Based on Hidden Markov Models of Individual Words. *Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform.*, 2017, vol. 17, iss. 4, pp. 452–464 (in Russian). DOI: 10.18500/1816- 9791-2017-17-4-452-464.