

УДК 004.934

DOI: 10.15827/2311-6749.18.2.1

СИСТЕМА РАСПОЗНАВАНИЯ РЕЧИ НА ОСНОВЕ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ ОТДЕЛЬНЫХ СЛОВ

Н.Е. Тимофеева, зав. лабораторией, timofeevane@yandex.ru; А.С. Гераськин, к.п.н., доцент, gerascinas@mail.ru; Ю.А. Мавлютова, старший лаборант, yuliyatav@yandex.ru (Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, ул. Астраханская, 83, г. Саратов, 410012, Россия)

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. Разработанные программные модули позволяют эффективно подготавливать необходимые исходные данные и, таким образом, формировать 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, National Instruments LabVIEW, звуковой поток.*

Распознавание речи и, как следствие, голосовая идентификация нашли свое применение во всех сферах человеческой деятельности. Благодаря системам распознавания речи обеспечивается безопасность от несанкционированного проникновения в защищенную зону. Такие системы содержат БД голосов сотрудников, имеющих доступ к защищаемой зоне, и предотвращают допуск людей, чьи голоса в ней отсутствуют [1, 2].

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

Уровень развития современной микропроцессорной техники (например, мобильные устройства связи) позволяет использовать сложные вычислительные алгоритмы, основанные на цифровой потоковой обработке статистических данных в реальном времени.

Одним из путей решения вышеуказанных задач является использование для распознавания фрагментов речи математического аппарата *скрытых марковских моделей* (СММ) [3]. Данная работа посвящена разработке программного модуля, осуществляющего формирование СММ для отдельных слов требуемого словаря команд системы управления объектом на основе кодирования признаков звукового сигнала, использующего линейные предсказания.

Описание решения

Создание ПО для решения поставленной задачи осуществлялось в среде графического программирования LabVIEW компании National Instruments [4].

Процесс распознавания с использованием СММ предполагает два этапа (рис. 1):

- построение и обучение СММ по речевой базе, содержащей записи слов диктора;
- распознавание команд в звуковом потоке.

В рамках этих этапов можно выделить основные структурные элементы системы:

- речевая база содержит записи слов, которые будут доступны для распознавания;
- модуль анализа признаков формирует алфавит слова, по которому формируется последовательность наблюдений;
 - по полученной последовательности наблюдений обучается СММ, наиболее подходящая из которых сохраняется в базе;
 - из входного звукового потока или файла (в данной работе для простоты рассматривается второй вариант) с помощью предварительной обработки выделяются слова (подавляется шум, отбрасываются лишние звуки и т.д.);
 - модуль анализа признаков формирует алфавит полученного слова, по которому формируется последовательность наблюдений;

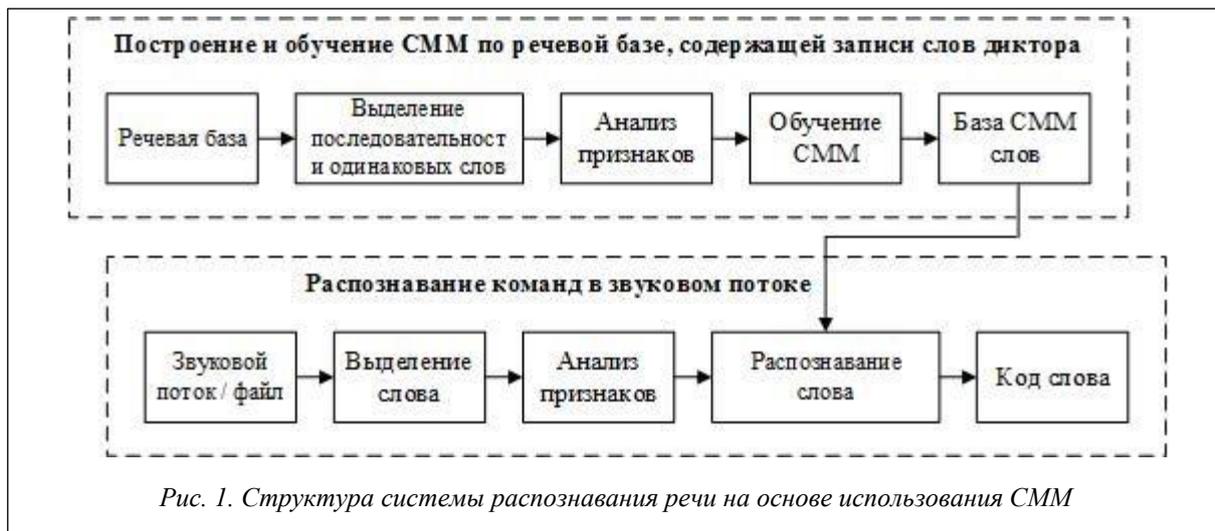


Рис. 1. Структура системы распознавания речи на основе использования СММ

– полученная последовательность поочередно подается моделям СММ, хранящимся в базе; вычисляется вероятность того, что данная последовательность подходит для очередной СММ; слово, которому принадлежит СММ с наибольшей вероятностью, считается искомым словом.

Таким образом, для реализации данной структуры необходимы всего четыре модуля: модуль выделения слов из звукового потока, модуль анализа признаков слова, модуль обучения СММ с базой моделей, модуль распознавания слов.

Основными источниками ошибок при голосовой идентификации являются среда записи (уровень и тип шума), среда представления и среда канала. Поэтому так важно произвести качественную предварительную обработку сигнала – от нее зависит, сможет ли система правильно распознать слово [5].

Для системы распознавания речи каждое слово должно переводиться в набор признаков для дальнейшего анализа. Таким образом, полученные в результате применения алгоритма выделения команд из звукового wav-файла отрезки одинаковой длины, соответствующие повторяемому слову (рис. 2), подвергаются на основе *линейного предсказания* (КЛП) анализу для построения матрицы наблюдений.

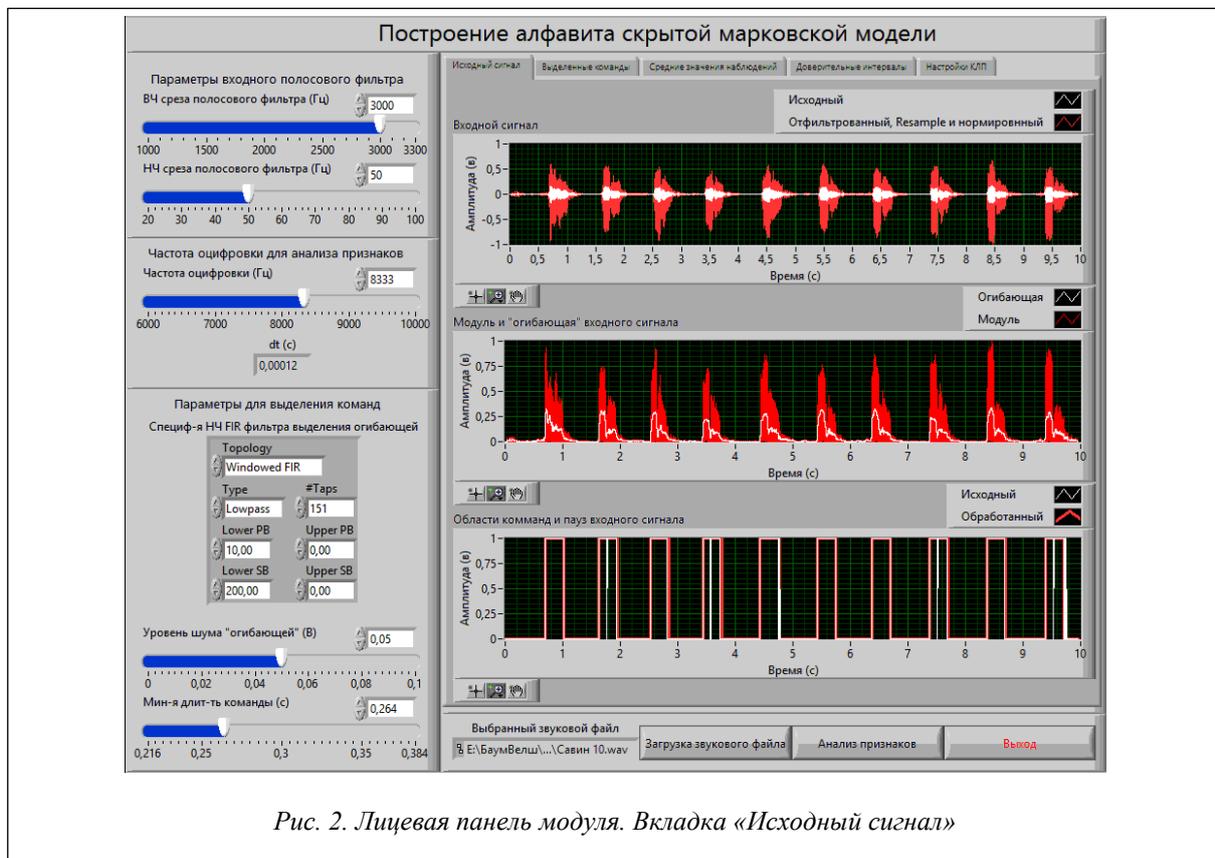


Рис. 2. Лицевая панель модуля. Вкладка «Исходный сигнал»

Полученные после КЛП анализа для нескольких реализаций одним диктором одного слова матрицы наблюдений S усредняются, и при этом строки усредненной матрицы являются символами алфавита СММ. По нему находится последовательность наблюдений (рис. 3).

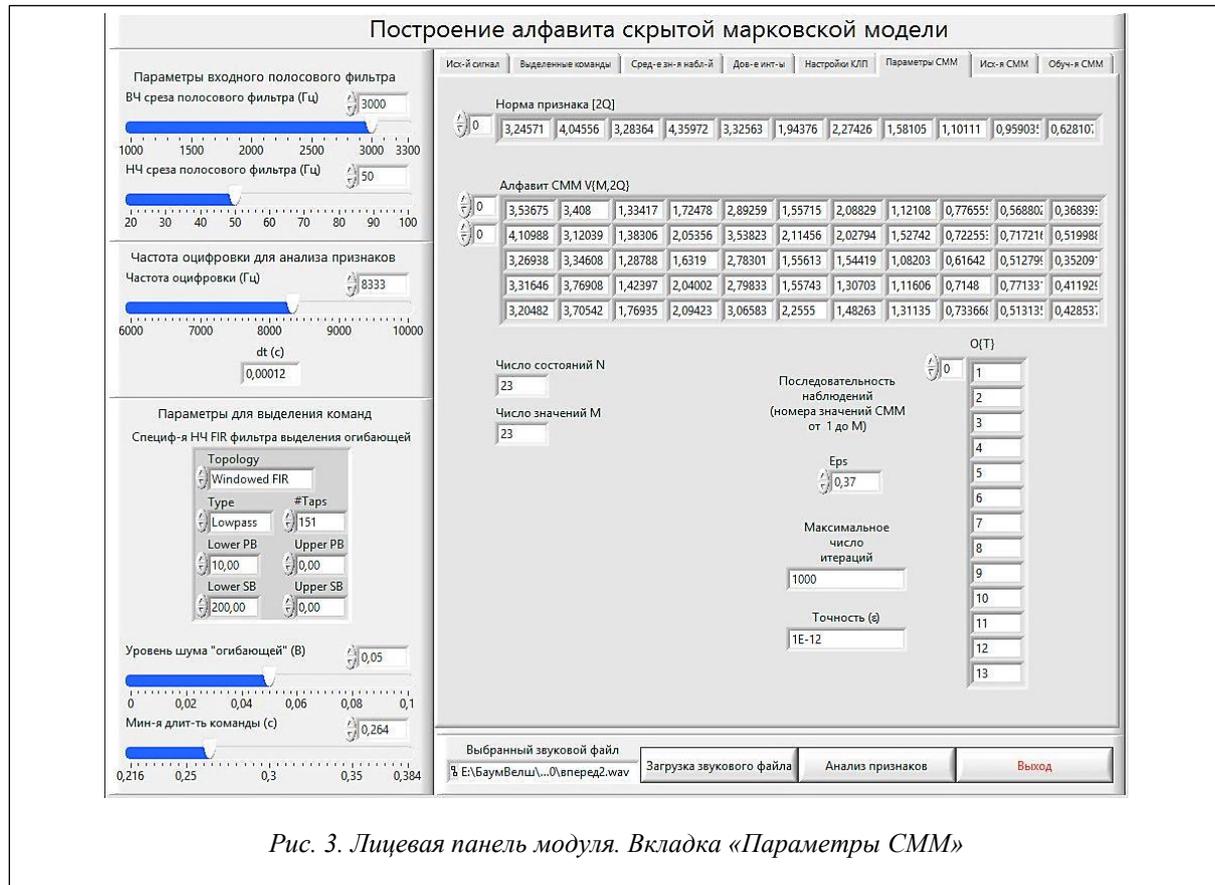


Рис. 3. Лицевая панель модуля. Вкладка «Параметры СММ»

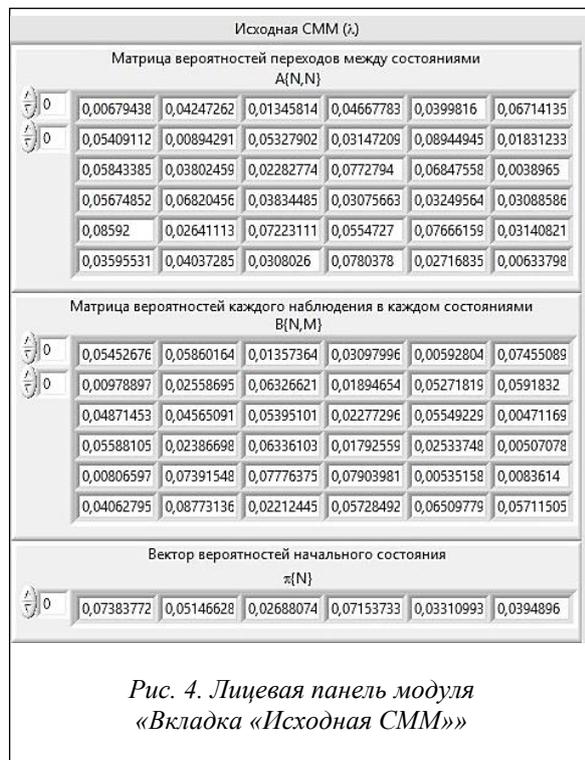


Рис. 4. Лицевая панель модуля «Вкладка «Исходная СММ»»

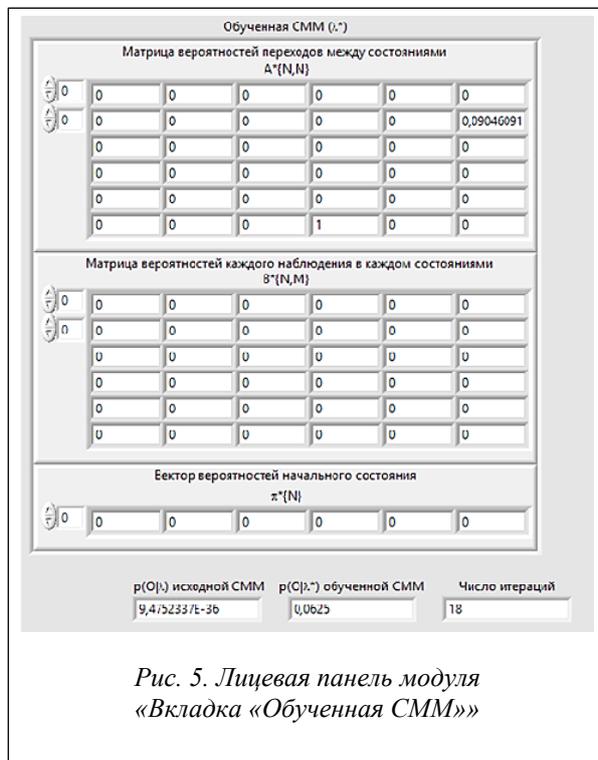


Рис. 5. Лицевая панель модуля «Вкладка «Обученная СММ»»

Для реализации алгоритма обучения СММ Баума–Велша [4] сначала генерируются матрица вероятностей переходов между состояниями, матрица вероятностей каждого наблюдения в каждом состоянии, а также вектор вероятностей начального состояния, отвечающие стохастическим ограничениям (рис. 4).

По полученной ранее последовательности наблюдений $O(T)$ обучается СММ, и полученные результаты отображаются в соответствующей вкладке (рис. 5). Сравнивая вероятности соответствия полученной последовательности наблюдений сгенерированной и обученной СММ, можно подобрать такую СММ, которая давала бы наилучшие результаты.

Заключение

Разработанный программный модуль в среде графического программирования LabVIEW компании National Instruments позволяет эффективно подготавливать необходимые исходные данные на основе кодирования признаков звукового сигнала, использующего линейные предсказания, строить СММ отдельных слов и проводить их обучение с помощью алгоритма Баума–Велша. Построенные СММ слов предполагается использовать в интеллектуальных системах управления различными объектами.

Литература

1. Жилияков Е.Г., Баринов С.Л., Чадюк П.В. Исследование сервиса компании Google Inc. по распознаванию русской речи // Науч. ведомости Белгородского гос. ун-та: Сер. Экономика. Информатика. 2013. Т. 27. Вып. 15-1. С. 247–255.
2. Титов Ю.Н. Современные технологии распознавания речи // Вестн. Тамбов. 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 компонентов программного комплекса для потоковой фильтрации аудиоконтента на основе использования скрытых марковских моделей // Изв. Саратов. ун-та: Сер. Математика. Механика. Информатика. 2015. Т. 15. Вып. 3. С. 340–350. DOI: 10.18500/1816-9791-2015-15-3-340-350.
4. Портал компании National Instruments Russia. URL: <http://www.labview.ru> (дата обращения: 14.04.2018).
5. Матвеев Ю.Н. Технологии биометрической идентификации личности по голосу и другим модальностям // Вестн. МГТУ им. Н.Э. Баумана: Сер. Приборостроение. 2012. С. 46–61.