

*На правах рукописи*

ГРЕБНОВ Сергей Викторович

**РАЗРАБОТКА ЭФФЕКТИВНЫХ МЕТОДОВ И КОМПЛЕКСОВ  
ПРОГРАММ РАСПОЗНАВАНИЯ РЕЧИ В СИСТЕМАХ  
ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ**

Специальность 05.13.18 – Математическое моделирование, численные  
методы и комплексы программ

**АВТОРЕФЕРАТ**  
диссертация на соискание ученой степени  
кандидата технических наук

Иваново 2011

Работа выполнена в Государственном образовательном учреждении высшего профессионального образования «Ивановский государственный энергетический университет имени В. И. Ленина» (ИГЭУ).

Научный руководитель: доктор технических наук, профессор  
Пантелеев Евгений Рафаилович

Официальные оппоненты: доктор технических наук, профессор  
Жуков Владимир Павлович

доктор технических наук, профессор  
Коробов Николай Анатольевич

Ведущая организация: ООО «Нейрософт», г. Иваново

Защита состоится 25 февраля 2011 года в 11-00 часов на заседании диссертационного совета Д 212.064.03 при Ивановском государственном энергетическом университете по адресу: 153003, г. Иваново, ул. Рабфаковская, д. 34, корпус «Б», аудитория 237.

С диссертацией можно ознакомиться в библиотеке Ивановского государственного энергетического университета, с авторефератом – на сайте ИГЭУ [www.ispu.ru](http://www.ispu.ru)

Автореферат разослан « » января 2011 года.

Ученый секретарь  
диссертационного совета

Шульпин А.А.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** Одним из перспективных способов организации человеко-машинного взаимодействия является передача компьютерной системе инструкций пользователя в формате речевых команд. Голосовой интерфейс является необходимой компонентой, когда речь идет о создании комфортных условий жизни для людей с нарушениями опорно-двигательного аппарата. Такие системы со временем войдут в повседневный быт в процессе реализации концепции так называемых «умных домов». Кроме того, возможно их применение и на производстве в составе комплексов управления исполнительными механизмами.

В развитие этого научного направления внесли вклад такие ученые, как Рабинер, заложивший научные основы распознавания речи статистическими методами, Wilpon, Lee, Higgins, внесшие существенный вклад в развитие методов распознавания речевых команд, Винцюк, Карпов, Ронжин, занимающиеся распознаванием слитной русской речи. Анализ их работ позволил установить, что для организации человеко-машинного взаимодействия при помощи речевых команд система распознавания речи (СРР) должна отвечать следующим требованиям:

- возможность работы в режиме реального времени;
- достаточное качество распознавания (не менее 95% правильно распознанных команд в условиях отсутствия шумовой составляющей – соотношение сигнал/шум 25дБ);
- расширяемость словаря СРР без перепрограммирования.

Последнее требование связано с тем, что для повышения надежности распознавания речи часто создаются системы с тщательно подобранным закрытым словарем команд, который включает точную настройку грамматических конструкций и подбор специальных слов в составе команд. Однако расширение или изменение словаря команд подобных систем может быть выполнено лишь силами разработчиков СРР и связано с дополнительными временными и финансовыми затратами.

Существующие методы распознавания голосовых команд не отвечают всем заявленным требованиям. Это обстоятельство определяет актуальность исследований в этом направлении.

**Объект исследования** – речевой сигнал.

**Предмет исследования** – модели, методы и алгоритмы распознавания речи в системах человеко-машинного взаимодействия.

**Цель диссертационной работы** – повышение эффективности и качества распознавания речи в СРР с динамически расширяемым словарем команд.

**Задачи исследования.**

1. Анализ существующих моделей, методов и алгоритмов распознавания речи с целью выявления степени их соответствия современным требованиям и выбора прототипов для собственных исследований.

2. Разработка моделей, методов, и алгоритмов распознавания речи, обеспечивающих достижение следующих показателей распознавания голосовых команд:

- скорость работы, достаточная для использования в режиме реального времени (в два раза быстрее режима реального времени для словаря в 10 команд);
- высокое качество распознавания (95% правильно распознанных речевых команд в условиях отсутствия шумовой составляющей – соотношение сигнал/шум 25дБ);
- легкость модификации словаря команд: возможность добавления новых слов и команд без перепрограммирования системы.

3. Программная реализация предлагаемых алгоритмов и проведение экспериментальных исследований, подтверждающих их эффективность.

**Методы исследований.** В работе использовались методы теории вероятности, теории случайных процессов, математического анализа, цифровой обработки сигналов, спектрального анализа Фурье, теории оптимизации (динамическое программирование) и теории формальных языков.

#### **Научная новизна.**

1. Предложен гибридный метод распознавания речевых команд. Новизна метода заключается в поэтапном использовании алгоритмов распознавания слитной речи и ключевых слов, применяемых к разным частям команды.

2. Для распознавания ключевых слов разработан алгоритм ограниченного перебора множества путей в скрытой марковской модели (СММ), новизной которого является отсечение путей не на заключительном этапе, а в процессе их распознавания, что позволяет значительно сократить пространство поиска.

3. Предложены новые функции правдоподобия, используемые алгоритмом ограниченного перебора для отсечения перспективных вариантов: функция правдоподобия на основе оптимальных порогов и комплексная функция правдоподобия. Функция правдоподобия на основе оптимальных порогов отличается тем, что учитывает не только длину пройденного пути в СММ, но и конкретные фонемы, пройденные этим путем. Комплексная функция правдоподобия, в дополнение к этому, оценивает соответствие всех промежуточных состояний пути в СММ локальному критерию правдоподобия с общим для всех фонем порогом.

**Обоснованность** положений диссертации обеспечивается корректным использованием математических методов. **Достоверность** подтверждается результатами экспериментов на реальном речевом материале.

**Практическая ценность результатов.** Применение предложенного метода распознавания по сравнению с подходом, использующим единый метод распознавания, позволяет:

- ослабить зависимость скорости распознавания от количества слов в словаре команд;
- использовать команды, которые плохо распознаются алгоритмом распознавания ключевого слова.

Применение разработанного алгоритма распознавания ключевых слов с использованием предложенных функций правдоподобия позволяет повысить вычислительную эффективность распознавания за счет раннего отсека неперспективных вариантов.

**Реализация результатов исследований.** Разработанные модели, методы и алгоритмы были использованы при построении программного комплекса голосового управления роботом Lego Mindstorms NXT на кафедре Электроники и микропроцессорных систем ИГЭУ (лабораторная работа «Изучения методов человеко-машинного взаимодействия на основе голосового управления», предмет «Электромеханотроника», 5 курс, специальность 210106 «Промышленная электроника»), а также внедрены в составе проекта «Системы безопасности помещений и личности на базе компьютерного интеллекта» ООО «НИИ Спецлаб». Проведен ряд экспериментов, показавших возможность практического использования предложенных методов. Потенциальная область применения, обусловленная характеристиками разработанных методов, включает сферу управления техническими устройствами различного назначения при производственной и бытовой деятельности.

**Апробация работы.** Полученные в работе научные и практические результаты докладывались и обсуждались на XV Международной научно-технической конференции «Бенардосовские чтения» и IX Международной научной конференции «Теоретические основы энерго-ресурсосберегающих процессов, оборудования и экологически безопасных производств».

**Публикации** по материалам диссертации – 9 печатных работ, в том числе три в журналах, рекомендованных ВАК РФ. Получено 1 свидетельство об официальной регистрации программы для ЭВМ в Федеральной службе по интеллектуальной собственности, патентам и товарным знакам (Роспатенте); свидетельство №2010615606 (зарегистрировано 30.08.2010).

**Структура и объём работы.** Диссертация состоит из введения, четырех глав, заключения, пяти приложений и библиографического списка из 122 наименований. Общий объём работы составляет 120 страниц, в том числе 19 рисунков и двух таблиц.

## СОДЕРЖАНИЕ РАБОТЫ

**Во введении** обоснована актуальность диссертационной работы, сформулированы цель и задачи исследования, сформулирована научная новизна исследований и показана практическая значимость полученных результатов.

**В первой главе** выполнен анализ архитектуры систем голосового управления, а также современных методов распознавания речи и оценивания достоверности результатов распознавания.

Большинство современных систем автоматизированного распознавания используют модульную архитектуру (рис. 1) с использованием блоков шумочистки, детектора голоса, спектрального анализа и главного модуля, включающего алгоритм распознавания ключевого слова. Цифровой сигнал сначала поступает в модуль шумочистки; здесь повышается качество сигнала вследствие удаления шумов и внесенного каналом искажения. Затем детектор голоса

выделяет участки сигнала, содержащие речь. Эти участки с помощью модуля спектрального анализа преобразуются в структуры данных, пригодные для распознавания. Наиболее распространенным способом представления данных являются векторы Мел-кепстральных коэффициентов. В соответствии с методом, предложенным Европейским Институтом Телекоммуникационных Стандартов (ETSI), они рассчитываются для речевого сигнала с шагом дискретизации 25 мс, а для их получения используются свойства человеческого восприятия. Далее эти коэффициенты поступают в модуль, в котором происходит непосредственное определение наличия и распознавания команды.



Рис. 1. Система распознавания речевых команд

Методы распознавания речи, реализованные в составе модуля распознавания (рис. 1), и являются предметом исследований в данной диссертационной работе.

Анализ методов моделирования и автоматического распознавания речи в контексте задачи разработки системы голосового управления показал, что распознавание слитной речи наиболее успешно решается с помощью вероятностного подхода на основе СММ, предложенного Рабинером (рис. 2).

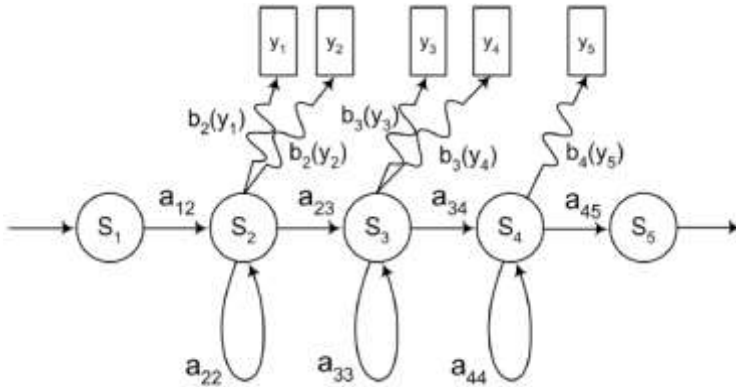


Рис. 2. Пример скрытой марковской модели

Скрытая марковская модель состоит из марковской цепи с конечным числом скрытых состояний и матрицей переходных (транзитивных) вероятностей  $a_{ij}$ , а

также конечного множества наблюдаемых состояний  $y_k$  и вероятностей  $b_n(y_k)$ , эмиссии наблюдения  $y_k$  в состоянии  $S_i$ . Эмиссионная вероятность  $b_n(y_k)$  моделируется несколькими Гауссовыми компонентами  $G_m$ , каждая из которых задана вектором математического ожидания  $\mu$ , вектором дисперсии  $\sigma^2$  и весовым коэффициентом  $p_m$  (вклад данной Гауссовой компоненты в общую функцию плотности вероятности). Таким образом,

$$b_n(y_k) = \sum_{m=1}^M p_m G_m,$$

$$b_n(y_k) = \sum_{m=1}^M p_m G(y_k, \mu_m, \sigma_m^2),$$

где  $G(x, \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ ,  $b_n(y_k)$  - вероятность появления наблюдения  $y_k$  для состояния  $S_n$ ;  $p_m$  и  $G_m$  - весовой коэффициент и величина  $m$ -ой Гауссовой компоненты;  $\mu$  и  $\sigma^2$  - векторы математического ожидания и дисперсии.

В задаче распознавания речи СММ используются для распознавания звуков речи. Каждое состояние  $S_n$  - фаза звука речи - характеризуется собственной функцией распределения вероятности появления того или иного символа наблюдения  $y_k$ . Последовательность, в которой звуки речи объединяются в распознаваемые слова (модель языка), при этом задается в виде словаря цепочек СММ.

Задача автоматического распознавания речи на основе вероятностного подхода сформулирована так: найти самое вероятное предложение (цепочку слов)  $S^*$  языка  $L$ , которое соответствует последовательности векторов признаков  $Y$  речевого высказывания, т.е.

$$S^* = \arg \max_{S \in L} P(S | Y) = \arg \max_{S \in L} P(Y | S)P(S) \quad .$$

Проведено исследование существующих подходов распознавания речевых команд на основе алгоритмов распознавания речи с помощью СММ:

- 1) поиск команд с использованием локальных мер сходства (например, оценок правдоподобия), наиболее распространенным методом является метод скользящего окна (sliding window);
- 2) полное моделирование инородной речи, метод моделей-заполнителей (filler models).

В заключение главы проводится сравнительный анализ рассмотренных методов в соответствии со сформулированным списком критериев (табл. 1). Сравнение методов позволило определить их основные недостатки: первый метод имеет большую вычислительную сложность, качество распознавания существенно зависит от фонетического состава слова, второй - требует подробного дополнительного моделирования посторонней речи, что не дает возможности динамически изменять словарь команд.

Сравнение методов распознавания речевых команд

	Метод скользящего окна	Метод моделей-заполнителей
Скорость	Низкая (зависит от количества слов в словаре)	Средняя
Качество распознавания	< 90% (зависит от фоновомного состава слов)	Высокое (90-95%)
Возможность модификации словаря	Да	Нет
Работа в зашумленных условиях	Да (снижается качество распознавания)	Да (снижается качество распознавания)

В результате анализа установлено следующее.

1. Ни один из методов не удовлетворяет всем заявленным требованиям.

2. Более перспективно развитие метода скользящего окна, так как он:

– обеспечивает модификацию словаря без перепрограммирования;

– не требует моделирования обобщенных слов, что упрощает задачу создания системы.

**Во второй главе** проведено исследование возможности повышения вычислительной эффективности и качества распознавания путем выделения в структуре команды специальной фразы, распознавание которой, подобно триггеру, активирует процесс интерпретации оставшейся части команды.

Проанализированы возможные варианты применения ключевой фразы.

1. Для управления состояниями системы (разработки Центра Речевых Технологий, продукты «ГОРЫНЫЧ» и Dragon Naturally Speaking):

– система активна, готова распознавать и выполнять голосовые команды (активации данного состояния осуществляется одной из ключевых фраз, например «Навигатор»);

– система неактивна, распознавание команд не производится. В данное состояние система переходит при отсутствии команд в течение предопределенного времени или произнесении специальной ключевой фразы, например «Отбой».

Использование состояний активности позволяет уменьшить количество ложных срабатываний и снизить нагрузку на процессор во время бездействия системы.

2. Для явного определения двухуровневой структуры речевой команды (IBM ViaVoice и Home Automation Living), состоящей из ключевого слова и собственно команды (определяющей действие, которое должна выполнить машина). Примером такой команды может служить фраза «Робот, открыть схват, где «Робот» – ключевое слово, «открыть схват» – непосредственно команда для выполнения. Отличительной особенностью данного варианта является то, что ключевое слово (фраза) присутствует в каждой речевой команде.



Показано, что второй способ подачи голосовых команд является предпочтительным, так как позволяет добиться лучшего качества распознавания, кроме этого является более удобным для использования.

Рассмотрен стандартный способ распознавания такой структуры команды:

– этап 1: определение наличия ключевого слова с помощью алгоритма распознавания ключевых слов;

– этап 2: если ключевое слово присутствует, то распознавание оставшейся части тем же алгоритмом, но с использованием расширенного словаря команд, включающего полный список команд для распознавания.

Предложен **гибридный метод распознавания** двухуровневой структуры речевой команды, обеспечивающий лучшую расширяемость словаря команд и качество распознавания за счет использования для интерпретации собственно команды (этап 2) алгоритма распознавания слитной речи, алгоритма Витерби.

Новизна метода заключается в полном разделении процесса определения присутствия речевой команды и определения непосредственного смысла команды и использования разных алгоритмов распознавания для разных частей команды:

1) детектирование ключевого слова с помощью рассматриваемого далее нового алгоритма распознавания ключевых слов (процесс определения присутствия речевой команды).

2) распознавание оставшейся части речевой команды с помощью стандартного алгоритма распознавания слитной речи, алгоритма Витерби (процесс определения непосредственного смысла команды).

Аналитически показаны преимущества данного метода распознавания:

– ослабление зависимости скорости распознавания от количества слов в слове команд;

– возможность использования команд, которые плохо распознаются алгоритмом распознавания ключевого слова.

**Третья глава** посвящена разработке алгоритма распознавания ключевого слова – **алгоритма ограниченного перебора**, основанного на комплексной функции правдоподобия. Алгоритм относится к классу методов поиска вхождения ключевых слов с использованием величин локальных мер сходства. Новизной алгоритма является использование функции правдоподобия для сравнения и отсекания путей в процессе их распознавания, что позволяет значительно сократить пространство поиска.

Описан разработанный алгоритм ограниченного перебора, который выполняет поиск самой вероятной последовательности состояний СММ, соответствующих ключевому слову  $S = \{s_1, s_2, \dots, s_K\}$ , во входном речевом сигнале, представленном последовательностью векторов наблюдений  $Y = \{y_1, y_2, \dots, y_N\}$ . Такая последовательность называется путем. На рис. 3 представлены четыре возможных пути:  $P_0, P_1, P_2, P_3$ . Путь может начинаться с любого участка речевого сигнала и только с самого первого состояния ключевого слова –  $s_1$ . Каждый путь характеризуется численным значением функции правдоподобия, которая пересчитывается на каждой итерации алгоритма.

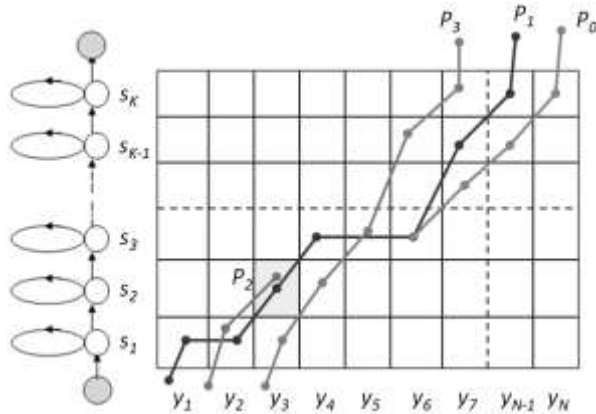


Рис. 3. Сокращение путей на основе функции правдоподобия

Для поиска возможных вариантов вхождения ключевого слова алгоритм начинает работу с начала речевого сигнала (дискретный момент времени  $t=1$ , вектор наблюдения  $y = y_1$ ), последовательно переходя к следующему дискретному моменту времени и вектору наблюдения пока не обработает речевой сигнал целиком ( $t=N$ ,  $y=y_N$ ). В каждый дискретный момент ( $t$ ) этого алгоритма выполняется этап ветвления и этап отсеечения путей.

Задачей этапа ветвления является построение всех возможных вариантов расположения ключевого слова во входном речевом сигнале. Для этого:

- стартует новый путь из текущей позиции сигнала ( $y_t$ ) в начальное состояние ключевого слова (в начальное состояние СММ первой фонемы ключевого слова,  $s_1$ );
- каждый существующий путь дублируется: один остается в текущем состоянии (переход  $s_i \rightarrow s_i$ ), второй переходит в следующее состояние ключевого слова ( $s_i \rightarrow s_{i+1}$ );
- для каждого из построенных путей пересчитывается функция правдоподобия.

На этапе отсеечения происходит сокращение путей с использованием функции правдоподобия. Для каждого состояния остается путь, которому соответствует наибольшее значение функции правдоподобия, остальные пути отбрасываются.

Алгоритм заканчивает работу в момент времени  $t=N$ . В этот момент состояние ключевого слова  $s_K$  содержит информацию о пути с наибольшим значением функции правдоподобия. Ключевое слово считается распознанным, если значение функции правдоподобия превышает пороговое.

Качество работы алгоритма распознавания ключевого слова определяется выбором эффективной функции правдоподобия. Предложены следующие модификации алгоритма и функции правдоподобия:

- использование оптимальных порогов для каждой из фонем (функция правдоподобия на основе оптимальных порогов).

– введение локальных порогов срабатывания (комплексная функция правдоподобия).

Исследование применяемых мер правдоподобия показало, что наиболее распространенной является нормированное по длине значение суммы локальных мер сходства:

$$Confidence(P_w) = \frac{1}{N_w} \sum_{j=1}^{N_w} pst_j,$$

либо отношение эмиссионной вероятности текущего ( $pst_j$ ) и лучшего состояния ( $pst_{best}$ ):

$$Confidence(P_w) = \frac{1}{N_w} \sum_{j=1}^{N_w} \frac{pst_j}{pst_{best}},$$

где  $N_w$  – длительность (число наблюдений) слова  $w$ ,  $pst_j$  – эмиссионная вероятность  $b_i(y_j)$  – вероятность появления наблюдения  $y_j$  в условиях данной акустической модели  $s_i$  слова  $w$  в момент времени  $j$ ;  $pst_{best}$  – эмиссионная вероятность  $b_n(y_j)$  лучшего состояния для рассматриваемого наблюдения  $y_j$ ,  $pst_{best} = \text{MAX}_i pst_i$ .

Коэффициент  $1/N_w$  выполняет нормирующую (усредняющую) роль, делая значение функции правдоподобия независимым от длины пути.

В ходе анализа такого подхода установлен следующий недостаток: функции правдоподобия принимают во внимание только длину пройденного пути, но не учитывают конкретные фонемы, пройденные этим путем. **Функция правдоподобия на основе оптимальных порогов** использует индивидуальный порог срабатывания для каждой фонемы вместо единого порога на все слово.

Оптимальный порог  $\Theta_{ph}$  для каждой из фонемы определяется как минимум линейной свертки функций количества ложных срабатываний (FP) и пропущенных фонем (FN) в зависимости от величины порога (рис. 4).

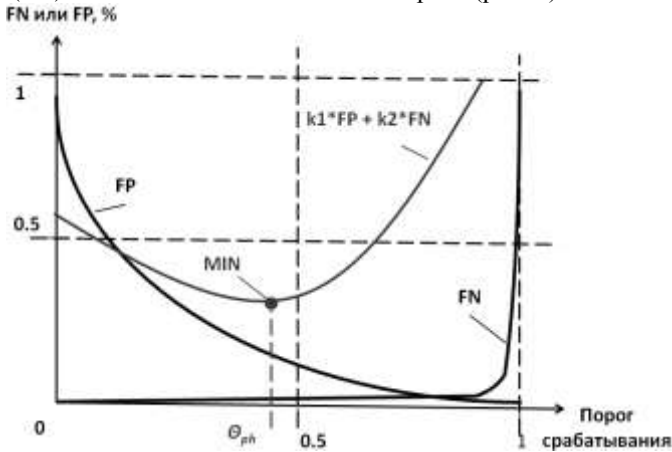


Рис. 4. Пример зависимости количества ложных срабатываний (FP) и пропущенных фонем (FN) в зависимости от порога

Так как на этапе отсеечения рассматриваемого алгоритма ограниченного перебора возможна потеря потенциально перспективных путей, предложена **комплексная функция правдоподобия (КФП)**.

КФП представляет собой функцию правдоподобия на основе оптимальных порогов с дополнительным условием фильтрации: все промежуточные состояния пути должны удовлетворять критерию отсеечения на основе локальной функции правдоподобия с общим для всех фонем порогом.

КФП использует дополнительную локальную функцию правдоподобия  $LS_{s,j}$  и дополнительное условие перехода пути в состояние  $s$  на этапе  $j$ :  $LS_{s,j} > \Theta_{local}$ . Здесь  $\Theta_{local}$  – единый локальный порог правдоподобия,  $0 < \Theta_{local} < 1$ ,  $LS_{s,j}$  – локальная функция правдоподобия пути для состояния  $s$  на этапе  $j$ ,  $LS_{s,j} = pst_{s,j}/pst_{best,j}$ .

Таким образом, комплексная функция правдоподобия для пути  $P_w$  ключевого слова  $w$  имеет вид:

$$ConfidenceEx(P_w) = \text{Min}_{ph \in w} Confidence(ph)$$

при одновременном выполнении следующих условий:

$$\begin{cases} \forall_{s,j \in P_w} : LS_{s,j} > \Theta_{local}, \\ \forall_{ph \in w} : Confidence(ph) > \Theta_{ph} \end{cases},$$

где  $Confidence(ph)$  – функция правдоподобия на основе оптимальных порогов,  $\Theta_{ph}$  – порог для фонемы  $ph$ .

Ниже приведен псевдокод итогового метода ограниченного перебора на основе данной функции правдоподобия:

**// поиск лучшего пути**  $P_{final}$

для каждого участка входного сигнала  $x_i \in X$

**// этап распространения путей**

создать новый путь для состояния  $s_1$  ключевого слова

для всех текущих путей  $p \in P$

для всех возможных переходов  $s_{new}$  из текущего состояния пути  $s_{current}$

**// переход внутри одной фонемы**

если (Фонема [ $s_{new}$ ] == Фонема [ $s_{current}$ ])

если выполняется локальный критерий ( $LS_{s,j} > \Theta_{local}$ )

осуществить переход  $p$  в  $s_{new}$

увеличить длину пути  $p$

пересчитать правдоподобие  $p$

иначе удалить  $p$  **//сокращение путей**

иначе **// переход от одной фонемы к другой**

$ph = \text{Фонема} [s_{current}]$

если  $Confidence(ph) > \Theta_{ph}$  **// сокращение путей**

если ( $ph ==$  последняя фонема ключевого слова)

$p_{final} = \text{best}(p_{final}, p)$

иначе

создать новый путь  $p_{new}$  для состояния  $s_{new}$  ключевого слова

добавить  $p_{new}$  в  $P$

если  $p_{final}$  найден, то вернуть  $p_{final}$

иначе вернуть "ключевое слово не обнаружено"

**В четвертой главе** выполнено исследование эффективности предложенных в работе методов и алгоритмов.

Приведено описание корпуса речевых данных `slSpeechCorp`, на котором выполнялись практические применения. Эта речевая БД содержит 10 часов речевого материала, записанного 40 людьми в возрасте от 18 до 50 лет и транскрибированного вручную. Для транскрибирования использовался алфавит Russian SAMPA (Speech Assessment Methods Phonetic Alphabet) и разделение гласных фонем на ударные (*stressed*) и безударные (*unstressed*). Всего 50 фонем.

Описана конфигурация базовой системы голосового управления, на которой проводились эксперименты. В качестве алгоритма спектрального анализа за основу взят алгоритм, предложенный европейским институтом стандартов телекоммуникации (ETSI). Включены предусмотренные стандартом модификации, направленные на улучшение качества работы: лифтирование (*Liftering*), вычет среднего кепстрального значения (*Cepstral Mean Subtraction, CMS*), нормализация энергии (*Energy Normalization*). На выходе данного блока для каждого участка сигнала в 25 мс формируется вектор из 39 параметров. Первыми 13-ю из них являются кепстральные коэффициенты (12 *mel-frequency cepstral coefficient*) и логарифм энергии (*logE*), а остальные – производные (1-го и 2-го порядка) этих коэффициентов (они показывают динамику изменения).

Проведено обучение СММ моделей и проведено сравнение эффективности разработанных методов с методом скользящего окна (Wilpon, 1999). В качестве мер эффективности использовались оценки Вычислительной сложности (*xRT*) и Качества распознавания (*Acc%/FP%*). Тестирование проводилось для распознавания слитной речи для словаря в 10 слов (достаточный размер словаря для управления большинством технических устройств различного назначения при производственной и бытовой деятельности). Результаты приведены в табл. 2.

Таблица 2

Сравнение разработанного метода распознавания речевых команд с методом скользящего окна

Показатель	Значение			
	V1	V2a	V2b	V2c
xRT, средний показатель для одного ключевого слова	1.33 xRT	0.15 xRT	0.16 xRT	0.24 xRT
xRT, средний показатель для всего словаря	12.4 xRT	0.2 xRT	0.21 xRT	0.29 xRT
Время работы (участок 4,1 с)	43 с	0.79 с	0.84 с	1.1 с
Качество распознавания, Acc%/FP%	86.1% /5%	92.4/0.5%	96.3/0.01%	98% /0.01%

Здесь V1 – «Метод скользящего окна» (Wilpon, 1999), V2a – «Метод ограниченного перебора, базовый», V2b – «Метод ограниченного перебора, основан-

ный на оптимальных порогах», V2c – «Метод ограниченного перебора, основанный на комплексной функции правдоподобия».

Результаты показывают превосходство разработанного метода, как в плане скорости, так и в плане точности распознавания.

1. Разработанный метод ограниченного перебора в среднем работает быстрее в 5(V2c)-8 (V2a) раз в зависимости от модификации.

2. Двухуровневая структура речевой команды в сочетании с гибридным методом распознавания позволили добиться 40-кратного преимущества в скорости на словаре в 10 команд.

3. Благодаря использованию единственного ключевого слова повысилось качество распознавания: 86.1%/5% (V1) против 98%/0.01% (V2c).

4. Использование индивидуальных порогов для каждой фонемы позволило сократить ошибку распознавания на 51% и снизить количество ложных срабатываний в 10 раз (V2b).

5. Применение локальных порогов позволило добиться дальнейшего сокращения ошибки распознавания на 46% за счет увеличения времени распознавания на 30%.

В заключение главы приведены результаты практического использования разработанных методов и моделей, включающие построение и экспериментальное исследование программного комплекса голосового управления роботом Lego Mindstorms NXT на кафедре Электроники и микропроцессорных систем ИГЭУ (лабораторная работа «Изучения методов человеко-машинного взаимодействия на основе голосового управления», предмет «Электромеханотроника», 5 курс, специальность 210106 «Промышленная электроника»).

Применение предложенного метода распознавания речевых команд обеспечило:

- возможность легкой модификации словаря команд с помощью специального интерфейса пользователя (для того чтобы добавить новую команду достаточно записать ее в виде звуковых единиц, фонем, во встроенном редакторе команд, предоставляемом программным комплексом);

- требуемое качество распознавания, позволяющее осуществлять голосовое управление роботом без предварительного обучения или настройки под конкретные характеристики голоса.

**В приложениях** приведены акты о внедрении результатов диссертационной работы.

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ**

1. Предложен **метод распознавания речевых команд**, обеспечивающий расширяемость словаря команд и улучшение качества распознавания за счет поэтапного использования алгоритмов распознавания слитной речи и ключевых слов, применяемых к разным частям команды.

2. Разработан **алгоритм распознавания ключевого слова**, реализующий ограниченный перебор на основе эффективной **комплексной функции правдоподобия**. Алгоритм обеспечивает повышение вычислительной эффективности и качества распознавания речи.

3. Разработанные модели, методы и алгоритмы реализованы в виде **вычислительного комплекса**, их преимущества подтверждены экспериментально.

## **ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **Публикации в изданиях, рекомендованных ВАК РФ**

1. Гребнов, С. В. Разработка и реализация двухуровневого метода голосового управления на основе скрытых марковских моделей / С. В. Гребнов // Информационные Технологии. – 2009. – №9. – С. 40-46.

2. Гребнов, С. В. Аналитический обзор методов распознавания речи в системах голосового управления // Вестник ИГЭУ. – 2009. – Вып. 3. – С. 83-85.

3. Гребнов, С. В. Двухуровневый метод распознавания голосовой команды // Вестник ИГЭУ. – 2009. – Вып. 3. – С. 90-93.

### **Публикации в прочих изданиях**

4. Гребнов, С. В. О новом подходе голосового управления / С. В. Гребнов // Тезисы докладов Международной научно-технической конференции "Состояние и перспективы развития электротехнологии" (XV Бенардосовские чтения), 27-29 мая / Федеральное агентство по образованию, ИГЭУ. – Иваново. – 2009 – Т.1 – С. 61-62.

5. Гребнов, С. В. Методы шумоочистки в задачах распознавания речи. / С. В. Гребнов // Тезисы докладов Международной научно-технической конференции "Состояние и перспективы развития электротехнологии" (XV Бенардосовские чтения), 27-29 мая / Федеральное агентство по образованию, ИГЭУ. – Иваново. – 2009. – Т.1 – С. 60-61.

6. Гребнов, С. В. Разработка речевого интерфейса / С. В. Гребнов // Тезисы докладов Международной научно-технической конференции "Состояние и перспективы развития электротехнологии" (XV Бенардосовские чтения), 27-29 мая / Федеральное агентство по образованию, ИГЭУ. – Иваново. – 2009. – Т.1 – С. 62-63.

7. Варков, А. А. Исследование возможности применения голосового управления манипуляционным роботом / А.А. Варков, С. В. Гребнов // Тезисы докладов Международной научно-технической конференции "Состояние и перспективы развития электротехнологии" (XV Бенардосовские чтения), 27-29 мая / Федеральное агентство по образованию, ИГЭУ. – Иваново. – 2009. – Т.1 – С. 271-272.

8. Гребнов, С. В. Перспективы использования речевого интерфейса в экологически безопасных производствах / С. В. Гребнов // Тезисы докладов IX Международной научно-технической конференции "Теоретические основы энерго-ресурсосберегающих процессов, оборудования и экологически безопасных производств", 28-30 сентября / Федеральное агентство по образованию, ИГХТУ. – Иваново. – 2010. – С. 268.

9. Свидетельство об официальной регистрации программы для ЭВМ №2010615606. Программный комплекс голосового управления роботом Lego Mindstorms NXT / Гребнов С. В.; патентообладатель ГОУ ВПО ИГЭУ. – №2010615606; заявл. 30.06.2010; зарегистрировано в реестре программ для ЭВМ 30.08.2010.

ГРЕБНОВ Сергей Викторович

РАЗРАБОТКА ЭФФЕКТИВНЫХ МЕТОДОВ И КОМПЛЕКСОВ  
ПРОГРАММ РАСПОЗНАВАНИЯ РЕЧИ В СИСТЕМАХ  
ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени

кандидата технических наук

Подписано в печать 14.01.11 Формат 60x84 1/16.  
Печать плоская. Усл.печ.л. 0,93. Тираж 100 экз. Заказ № 114.

Государственное образовательное учреждение  
высшего профессионального образования  
«Ивановский государственный энергетический  
университет имени В.И. Ленина»  
153003, г. Иваново, ул. Рабфаковская, 34.  
Отпечатано в УИУНЛ ИГЭУ.