

УДК 004.896+004.934

Ю.Н. Тесля, А.Ю. Чёрный

Киевский национальный университет строительства и архитектуры, Киев

РЕФЛЕКТОРНАЯ СИСТЕМА ГОЛОСОВОГО УПРАВЛЕНИЯ ТЕХНИЧЕСКИМИ УСТРОЙСТВАМИ (РСГУ)

Изложены результаты применения интроформационного подхода теории несилового взаимодействия к построению системы голосового управления. Разработана и представлена структура рефлекторной системы голосового управления техническими устройствами и алгоритм ее функционирования.

Ключевые слова: *интроформационный метод, рефлекторная система, голосовое управление, фонетический стенограф, понимание речи, несиловое взаимодействие*

Вкладено результати застосування інтроформаційного підходу теорії несилової взаємодії до побудови системи голосового управління. Розроблено та представлено структуру рефлекторної системи голосового керування технічними пристроями та алгоритм її функціонування.

Ключові слова: *інтроформаційний метод, рефлекторна система, голосове керування, фонетичний стенограф, розуміння мови, несилова взаємодія*

Considers the results of using introformacionno approach to the theory of nesilovogo interaction to build the system voice control. Designed and presented the structure of the reflexive voice control devices and an algorithm of its operation.

Keywords: *introformacionnoy method, reflex system, voice control, phonetic stenograf, speech understanding, nesilovoe interaction*

Постановка проблемы

Историческое развитие технических систем показало, что наиболее простым оказалось решение задачи передачи человеком управляющего воздействия с использованием рук. Однако более удобной и естественной для человека является речевая коммуникация. Именно поэтому в последнее десятилетие получили развитие области науки, связанные с созданием голосовых информационно-управляющих систем и систем распознавания речи.

Но, несмотря на колоссальные вложения трудовых и финансовых ресурсов, до настоящего времени не существует эффективных общезначимых инструментов обработки устной речи. С этой задачей хорошо справляется мозг. Но, пока неизвестны механизмы его работы как сложной структуры, состоящей из огромного количества взаимодействующих объектов, трудно говорить о прорыве в этом направлении.

Значительный вклад в решение этой проблемы может оказать теория несилового взаимодействия

(ТНВ) и разработанный в её рамках интроформационный метод.

Во-первых, потому что в ТНВ обоснованы законы, которые могли привести к возникновению такой сложной структуры как мозг.

Во-вторых, потому что интроформационный метод очень эффективен при построении искусственных интеллектуальных систем – позволяет произвести оценку совместной условной вероятности по частным. Использование этой отличительной особенности метода позволило создать, так называемые, рефлекторные системы ИИ, которые хорошо зарекомендовали себя на практике.

Учитывая практическое отсутствие эффективных и дешёвых систем голосового управления техникой, их разработка относится к актуальным задачам. Именно применение рефлекторного подхода к построению таких систем позволит получить стойкий результат при незначительных затратах на создание.

Анализ основных исследований и публикаций

Теория несилового взаимодействия и ее приложения к разработке искусственных интеллектуальных систем рассмотрены в работах [1-3]. Основным выводом из этих работ является простота интроформационных методов и алгоритмов, а также эффективность разработанных на их основе средств. Что касается проблематики, связанной с созданием систем голосового управления, то в последнее время эта область получила большую популярность благодаря коммерчески успешным системам от гигантов мировой индустрии программного обеспечения – Nuance Communications, Google, Apple.

Самое известное достижение в этой области – совместный продукт компании Apple и Nuance – персональный помощник и вопросно-ответная система *Siri*. Данное приложение использует обработку естественной речи, чтобы отвечать на вопросы и давать рекомендации. Для *Siri* Apple использовала результат 40-летних исследований «Центра Искусственного Интеллекта» и работы исследовательских групп из известнейших университетов мира. Это исследование описывается, возможно, как самый большой проект искусственного интеллекта на сегодняшний момент. Компания Google также имеет очень качественное решение для распознавания речи и голосового управления под управлением операционной системы Android – система Voice Actions. Все эти решения созданы для мобильных платформ.

Эти же компании создали системы голосового управления для телевизоров – компания Nuance с платформой Dragon TV и Google с Google TV. Обе системы представляют собой приложение к телевизору, реализованное в виде отдельного аппаратного модуля с соответствующей программной платформой.

Все перечисленные системы объединяет одно качество – они построены с привлечением подхода, ориентированного на огромные объемы обучающих данных и использование «облачных» вычислительных центров. Это подчеркивает их основную слабость – необходимость постоянного подключения к сети Интернет с целью передачи голосового поискового образа для обработки. Также перечисленные системы плохо поддерживают украинский и русский языки.

Но задача распознавания устной речи в условиях стационарного функционирования частично решена. Именно поэтому необходима разработка инструментов дикторонезависимого распознавания речи, которые бы просто

реализовывались и не требовали больших вычислительных ресурсов. Возникновение теории несилового взаимодействия может помочь в решении задачи распознавания устной речи при незначительных затратах.

В настоящее время отсутствуют работы, посвященные применению теории несилового взаимодействия к построению систем голосового управления и распознавания устной речи.

Формулировка целей статьи

Целью статьи является популяризация теории несилового взаимодействия и подтверждение ее истинности через демонстрацию применимости интроформационного метода к построению системы голосового управления.

Основной материал исследований

Созданная авторами рефлекторная система голосового управления (РСГУ) предназначена для решения задач свободно-языкового ввода управляющих команд и контента в техническое устройство. Кроме того РСГУ повышает эффективность взаимодействия с техническим устройством, предоставляя голосовой интерфейс, который разгружает руки оператора.

Программный прототип РСГУ имеет модульную архитектуру и функционирует как локальное приложение, не требующее сетевого подключения. Система состоит из двух основных модулей: автоматического фонетического стенографа и ядра РСГУ, текущая реализация которых определяет условия их использования (рис. 1).

Фонетический стенограф [4] выполнен в виде бинарного приложения, набора библиотек и конфигурационных файлов для платформы Windows. Ядро РСГУ реализует интроформационный метод выработки рефлексов. Оно разработано и исполняется в среде MS Access на всех операционных системах семейства Windows.

Входной информацией для системы есть голосовая команда, представленная в виде звуковой волны (рис. 1).

Выходная информация – управляющее воздействие на объект управления, т. е. выполнение распознанной команды в соответствии с заданными голосом параметрами.

Также система в процессе работы генерирует необходимые визуальные и голосовые информационные сообщения, что предоставляет возможность отслеживать процесс распознавания команд, реакции на них и, кроме того, позволяет в реальном масштабе времени изменять поведение системы в случае необходимости.

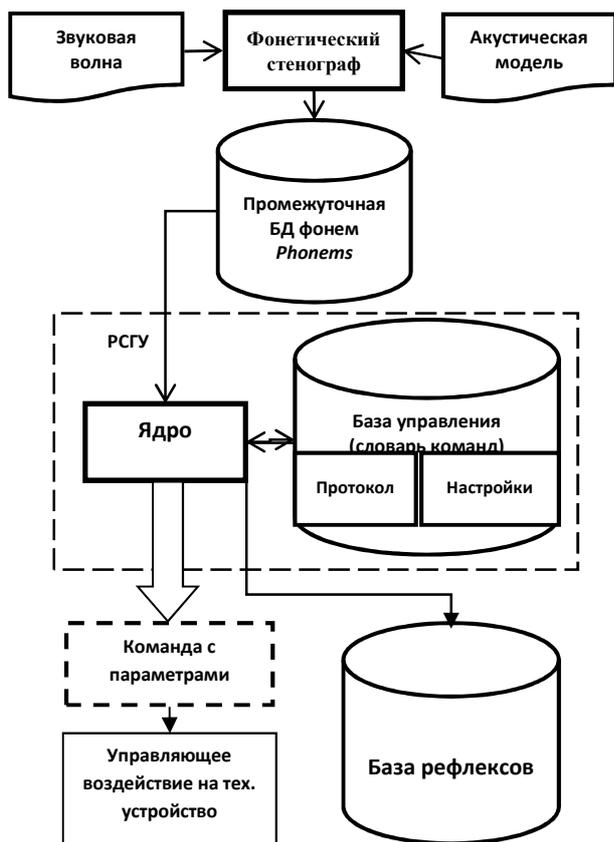


Рис. 1. Структура РСГУ

Рассмотрим схему работы системы голосового управления техническими устройствами на примере управления телевизором. Диктор в свободной форме озвучивает необходимые для него действия системы. Например, по отношению к голосовому управлению телевизором: «Найди-ка мне новости». Или «Отключись через 20 минут». Или «Сделай громче звук, пожалуйста». Программная платформа управления голосом передает необходимую команду на техническое устройство, или озвучивает диктору информацию, затребованную в его команде. При обучении диктор сам выполняет соответствующее действие, и у системы вырабатывается рефлекс на подобное обращение. Если диктор говорит «по-разному», то вырабатывается устойчивый рефлекс именно на информативную часть голосовой команды.

1. Звуковая команда. Диктор голосом обращается с просьбой к техническому устройству.

Выходная информация: звуковая волна.

Пример: включи мне, пожалуйста, восемнадцатый канал быстренько.

2. Акустическая модель. Включает статистическое описание распознаваемого языка и особенностей речи диктора. Статистическое описание формируется в процессе обучения настройкой на дикторов. В качестве акустических

моделей используются скрытые Марковские модели. 56 украинских контекстно-независимых фонем моделируются тремя состояниями Марковской цепи без пропуска.

Словарь транскрипций создается автоматически из орфографического словаря с использованием контекстно-независимых правил.

3. Фонетический стенограф. Служит для преобразования входящего оцифрованного звукового сигнала, содержащего устную речь (акустической модели), в набор фонем.

Алгоритм фонетического стенографа позволяет строить последовательность фонем для речевого сигнала без использования какого-либо словаря. Для этой цели строится некоторая генеративная грамматика, которая может синтезировать все возможные модельные сигналы непрерывной речи для любой последовательности фонем. В рамках построенной модели строится алгоритм фонемного распознавания для неизвестного сигнала. Используются те же контекстно-независимые модели фонем, как и в базовом распознавателе.

Надежность обнаружения фонемы на правильном месте для известной реализации равна приблизительно 70%.

Выходная информация: промежуточная база данных фонем – результат распознавания входных звуковых волн.

Пример: ф п л ! у ч ! ь н н ! ь п ж ! л с т ь в а с ! ь н ь ц ь т ь й ! к а н а + л у п ь с т ! т р ! н ! к ь э

4. Ядро РСГУ. Предназначено для моделирования системы голосового управления техническим устройством. Содержит программную реализацию интроформационного метода, а также алгоритмов выделения комбинаций фонем и обучения (накопления статистики). Информационная база управления содержит словарь команд, протокол работы, настройки системы.

5. Команда с параметрами. Результатом ее работы является команда с параметрами, которую необходимо реализовать техническому устройству.

Выходная информация: выполняемая команда.

Пример: Диктор (Тесля); Команда (Канал); Название канала (); Уровень (Десятки); Десятки (10); Единицы (8).

6. Управляющее воздействие. Содержит программную реализацию алгоритма управления техническим устройством.

Входная информация: преобразованная в вид формулы: команда с параметрами.

Результат: изменение параметров технического устройства.

Пример: переключение телевизора на 18-ый канал.

Система очень проста. Реализует рефлекторную модель поведения, показанную на рис. 2.

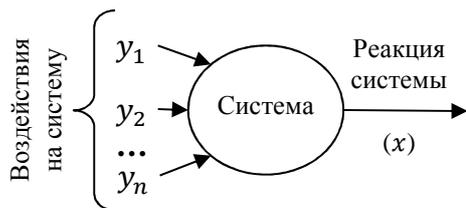


Рис. 2. Схема реакции системы на несиловые воздействия

Функционирует в двух режимах: **обучения** и **управления**. В режиме обучения с учителем формируется база рефлексов. В режиме управления РСГУ вырабатывает реакцию на обращение диктора. Также в этом режиме реализуется режим самообучения (если реакция не удовлетворила диктора).

Основная часть системы – база рефлексов. База рефлексов содержит статистику входных воздействий (комбинаций фонем) и реакций системы, разделенных на классы: диктор, команда, название канала, уровень числа, десятки, единицы (рис. 3).

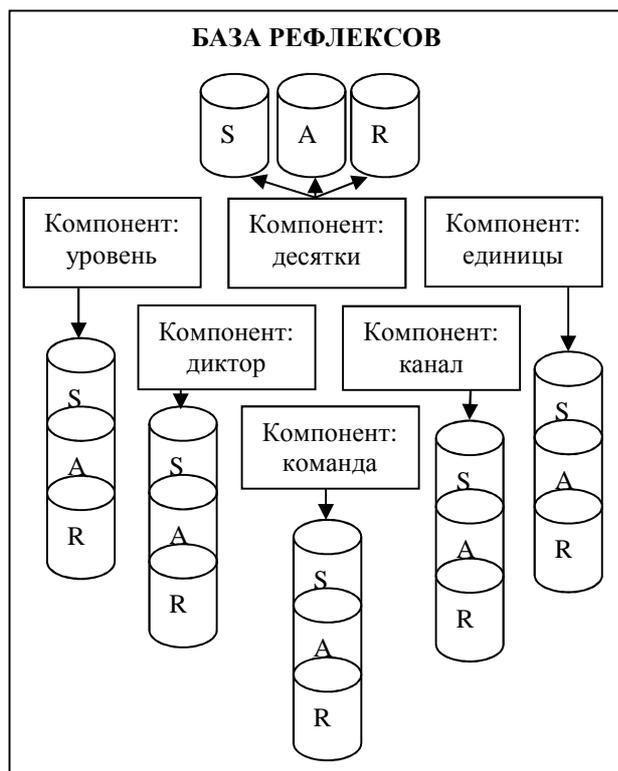


Рис. 3. Структура базы рефлексов

Каждый класс рефлексов реализуется в отдельном компоненте системы. Каждый компонент системы можно представить как отдельный «интроформационный нейрон». На входе – полный входной набор фонем, и/или реакция других

«интроформационных нейронов». На выходе – выработанная им реакция, поступающая на другие «интроформационные нейроны», или техническое устройство.

В каждом компоненте информация хранится в таблицах:

1. Таблица S – комбинации фонем. Хранятся комбинации подряд идущих фонем длиной от 2-х до 10 символов и информация о том, сколько раз они встречались.

2. Таблица R – реакции РСГУ. Содержит перечень действий, которые необходимо выполнить РСГУ или техническому устройству, частоту пользования и определенность данной реакции. Реакция типа «Не знаю» обеспечивает открытость системы.

3. Таблица A – предназначена для связи таблиц S и R. Содержит информацию о том, сколько раз и какая реакция была востребована в случае, когда на входе был некоторый набор фонем. Кроме того, таблица содержит определенность реакции, связанную с этим набором фонем.

В режиме обучения в приведенных таблицах накапливается информация о том, какие входные фразы к каким реакциям приводили. Потом эта информация используется в режиме управления для выработки соответствующих обращению диктора реакций интроформационным методом [2]. Алгоритм реализации режима управления следующий:

1. Старт алгоритма.

1.1. В поступившем на вход потоке фонем выделяются фрагменты (наборы), множество M , содержащие от 2 до 10 рядом стоящих символов.

2. Рассматривается класс команд.

2.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

2.2. Используя таблицы A и R, в соответствии с интроформационным методом, вычисляется определенность реакций (команд). Команда, имеющая наибольшую определенность, выбирается для реализации.

3. Если в команде есть обращение к названию канала (знак #), рассматривается класс названий каналов.

3.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

3.2. Используя таблицы A и R, в соответствии с интроформационным методом вычисляется определенность реакций (телевизионных каналов). Канал, имеющий наибольшую определенность, выбирается и вставляется в текст команды вместо значка #. Если канала не отобран, вставляется знак «?».

4. Если в команде есть обращение к числовому значению (знак №), рассматриваются классы уровень числа, десятки, единицы.

4.1. Класс уровня числа.

4.1.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

4.1.2. Используя таблицы A и R, в соответствии с интроформационным методом вычисляется определенность уровня числа. Варианты: нет десятков (числа от 0 до 9), есть десятки (числа больше 9).

4.2. Если уровень числа «Есть десятки», активизируются таблицы, входящие в класс десятков.

4.2.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

4.2.2. Используя таблицы A и R, в соответствии с интроформационным методом вычисляется определенность номера десятка. Если десятка не определена, в команду вставляется знак «?».

4.3. Класс единиц.

4.3.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

4.3.2. Используя таблицы A и R, в соответствии с интроформационным методом вычисляется определенность второй цифры в числе. Если цифра не определена, в команду вставляется знак «?».

5. Компонент определения диктора.

5.1. В таблице S осуществляется отбор записей, соответствующих сформированным наборам фонем (входящим в множество M).

5.2. Используя таблицы A и R, в соответствии с интроформационным методом вычисляется определенность диктора (соответствия обучающей выборке).

6. Если в команде есть знак «?» задается вопрос к диктору с просьбой уточнить обращение. Если такого знака нет – выполнение команды.

7. Завершение алгоритма.

Таким образом, в каждый компонент системы поступает весь входной набор фонем. Без выделения слов, команд, предложений и т.д. Также как у человека. Слушая устную речь, или читая письмо, он, не распознавая буквы и слова, распознает смысл. Так и в рефлекторной системе голосового управления. Не нужно создавать никаких словарей, выполнять морфологический, синтаксический, семантический анализ текста, а также выделять слова и команды. Реакции вырабатываются на звуковой поток, из которого система, как и человек сама «умеет выделять» информативную часть по максимальной определенности [2].

Для проверки этих идей РСГУ была воплощена в систему голосового управления телевизором (GUT) [5]. Для обучения системы GUT были

использованы две выборки. Для построения акустической модели обучающая выборка включала 500 команд, содержащих только информативные слова. Для выработки рефлексов обучающая выборка включала около 2000 по-разному построенных предложений, с разными словами и разным порядком их произнесения. Эта обучающая выборка имела некоторые особенности:

- использовалась спонтанная с разным уровнем звука и помех речь;

- предложения произносились в разном темпе.

Надежность разработанной системы оценивалась по контрольной выборке, включающей 600 команд: 200 простых команд, 200 предложений от 3 до 5 слов, содержащих команды и 200 длинных предложений (не менее 8 слов), также содержащих команды. Результаты экспериментов:

1. Для простых команд – 98% правильно распознанных действий.

2. Для коротких предложений – 90%.

3. Для длинных предложений – 86%.

Но даже не эти, достаточно хорошие показатели являются главным достоинством предложенной системы. Она была построена в экстремально короткие сроки (3 месяца), коллективом всего из 3-х человек и позволила выявить ряд преимуществ рефлекторных систем. Поскольку не нужны словари, не надо анализировать слова (грамматику), не надо создавать модели интеллектуального анализа текста, то разработанная авторами система относительно традиционных обладает следующим преимуществом. Она:

- проще;

- дешевле;

- не ограничивает естественный язык;

- использует универсальный алгоритм, который можно использовать для создания многих устройств распознавания, и не только речи;

- легко обучается;

- хорошо умеет выделять из произвольного речевого потока полезную информацию;

- не реагирует на шум, обращение, не содержащее никакой команды, на разговор людей между собой;

- выуживает полезную информацию в потоке свободной речи;

- позволяет естественно добавлять новые команды;

- адекватно реагирует на окружающий речевой и не речевой фон (т.е. на постороннюю информацию, не имеющую отношение к управлению техническим устройством);

- терпима к ошибкам в речи диктора;

- позволяет перепрограммировать поведение системы голосом;

- учится на собственных ошибках;

- выполняет фиксацию, анализ и исправление ошибок распознавания;
- выполняет голосовую идентификацию диктора;
- подстраивается под манеру и привычки диктора.

Такого набора преимуществ нет ни у одной системы голосового управления в мире!

Система позволяет управлять и использовать весь функционал современных телевизоров с помощью голосовых команд. В первую очередь она используется для повышения удобства управления телевизором как замена стандартному пульту дистанционного управления. Кроме выполнения обычных команд (переключения по номеру или названию канала, включения на определенное время таймера или его отключения, включение, отключение, увеличение, уменьшение звука) она может поддерживать беседу с пользователем на произвольную тему, игнорирует речь не обращенную к ней, учиться новым командам и реакциям на обращение пользователя непосредственно с речи (без нажатия каких либо кнопок).

Возможными областями применения РСГУ является:

- **Голосовое управления автоматизированным и call-центрами и IVR-системами.** *Потенциальные покупатели технологии:* операторы связи; интеграторы коммуникационных сетей; производители IVR-систем и платформ построения call-центров;
- **Заказы билетов, товаров, услуг через телефон.** Заказ железнодорожных, театральные билетов, бронирование гостиниц, авиабилетов, ресторанов, заказ такси и т. д. *Потенциальные покупатели технологии:* представители железной дороги, авиакомпаний, разработчики онлайн касс и т.д.
- **Голосовое управление банковскими счетами (по телефону).** *Потенциальные покупатели технологии:* банки; разработчики систем дистанционного банковского обслуживания.
- **Голосовое управление системой «умный дом».** *Потенциальные покупатели технологии:* зарубежные производители систем «умный дом».
- **Управление торговыми автоматами и вендинговыми машинами.** *Потенциальные покупатели технологии:* производители данных устройств.
- **Приложения голосового управления компьютером для общества слепых.** *Потенциальные покупатели технологии:* представители общества слепых, компании, разрабатывающие программы и устройства для слепых.
- **Интеграция с Microsoft Speech API.** *Потенциальные покупатели технологии:* производители ПО для платформ компании Microsoft, нуждающиеся в голосовом управлении своими приложениями.

• **Голосовой доступ к популярным веб-сервисам.** *Потенциальные покупатели технологии:* владельцы популярных социальных сетей, блогов, веб-ориентированных систем управления задачами и т. д.

• **Голосовое управление мультимедийными устройствами.** *Потенциальные покупатели технологии:* производители телевизоров, ТВ-тюнеров, игровых консолей, разработчики программного обеспечения для данного класса устройств.

Именно широкая адаптивность рефлекторных систем, построенных на основе интроформационного метода теории несилового взаимодействия, позволила постулировать столь широкие области применения построенной технологии.

Выводы и перспективы дальнейших исследований

Несмотря на то, что используемая теория несилового взаимодействия является гипотетической, для нас важно, что, основываясь на несиловой модели взаимодействия в природе, можно создавать принципиально новые системы искусственного интеллекта для многих областей деятельности человека. В том числе и действующие системы голосового управления.

Представленная в этой статье система голосового управления техническими устройствами является ярким примером возможностей практического применения теории несилового взаимодействия. И мы надеемся, что эта статья поможет многим исследователям, инженерам применить знания авторов для дальнейшего прогресса в решении основной задачи кибернетики – раскрытия законов построения и механизмов работы мозга и создания на этой основе искусственных систем, не уступающих своим «интеллектом» человеку.

Список литературы

1. Тесля Ю.Н. Несиловое взаимодействие / Юрий Тесля // Монография. – К.: Кондор, 2005. – 196 с.
2. Тесля Ю.Н. Введение в информатику природы / Юрий Тесля // Монография. – К.: Маклаут, 2010. – 256 с.
3. Тесля Ю.Н. Динамические свойства интроформации / Ю.Н. Тесля // Управление развитием сложных систем. – К.: КНУСА. – 2012. – № 9. – С. 44 – 49.
4. Пилипенко В.В. Автоматизированный стенограф украинской речи / В.В. Пилипенко, В.В. Робейко // Штучний інтелект. – 2008. – № 4. – С. 768-775.
5. Тесля Ю.Н. Демонстрация системы GUT [Электронный ресурс] / Ю.Н. Тесля. – Режим доступа: http://www.youtube.com/watch?v=m_pYXVndpbc

Статья поступила в редколлегию 12.09.2013

Рецензент: д-р техн. наук, проф. С.Д. Бушуев, Киевский национальный университет строительства и архитектуры, Киев.