ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ РЕЧЕВОГО ИНТЕРФЕЙСА С ПРИМЕНЕНИЕМ КОГНИТИВНЫХ И ЛИНГВИСТИЧЕСКИХ ЗНАНИЙ

Фархадов М. П. [1], **Петухова Н. В.** [2], **Васьковский С. В.** [3]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)
Фархадова М. Э. [4]
(*Российский университет дружбы народов, Москва*)

В современном информационном обществе ежедневно накапливается огромный объем речевых данных, и автоматизация их обработки может обеспечить принципиально новые возможности для анализа систем автоматизации принятия управленческих решений, поэтому разработка методов и средств автоматизации систем распознавания и анализа речевой информации имеет важное научное и государственное значение. Статья посвящена исследованию проблематики речевых интерфейсов: предложены методы решения возникающих проблем при проектировании человеко-машинного интерфейса с применением речевых технологий, сформулированы правила создания эффективных интерфейсов, основанные на проведенных исследованиях и опыте реализации речевых приложений. Анализируются когнитивные особенности и роль лингвистических знаний в создании речевых диалоговых систем. С точки зрения взаимодействия человека и информационной среды наиболее естественным является речевой интерфейс. Следует отметить, что сегодня при проектировании речевого интерфейса недостаточно применяются результаты как прикладных, так и теоретических исследований в данной области. Подобные исследования являются весьма актуальными, поскольку приложения с речевыми интерфейсами получают все большее прикладное распространение в различных сферах информационных технологий, и разработчики прикладных приложений нуждаются в практических рекомендациях для повышения, эффективности надежности и устойчивости систем с распознаванием речи.

Ключевые слова: речевой интерфейс, дизайн интерфейса, когнитивная нагрузка, управление диалогом, лингвистический процессор.

[1] *Маис Паша оглы Фархадов, д.т.н., г.н.с.* (*mais@ipu.ru*).
[2] *Нина Васильевна Петухова, с.н.с.* (*nvpet@ipu.ru*).
[3] *Сергей Владимирович Васьковский, к.т.н., с.н.с.* (*v63v@yandex.ru*).
[4] *Мухаббат Эргашевна Фархадова, к.ф.н., доцент* (*farkhadova_me@pfur.ru*).

1. Введение

В современном обществе огромную роль играют новейшие информационно-телекоммуникационные технологии. Поэтому при построении и реализации информационных систем используются последние достижения в этой области [1, 14, 17]. Важную роль во взаимодействии человека с различными информационно-сервисными ресурсами может играть речевой интерфейс [6–10, 15].

Разработаны модели автоматического распознавания речи на основе анализа звуковой и визуальной информации и компьютерного синтеза аудиовизуальной русской речи по произвольному тексту [7]. Сформулированы основные типы ограничений, влияющих на организацию человеко-машинного взаимодействия и конфигурирование программно-аппаратных решений, которые необходимо учитывать при проектировании многомодальных интерфейсов интерактивных приложений [15]. Важными являются когнитивные исследования ассистивного многомодального интерфейса для бесконтактного человеко-машинного взаимодействия. Изложены методики и результаты количественной оценки производительности бесконтактного человеко-машинного взаимодействия с применением элементов когнитивных экспериментов и сравнение с результатами для стандартных контактных способов указательного ввода информации [8]. В связи с модернизацией и интеллектуализацией нефтегазовой промышленности, а также со сложностью решения промышленных задач традиционными методами в данной отрасли существует большое число примеров применения нейросетей, в частности, для прогнозирования работы скважин [10, 11], при обработке экспериментальных данных процесса коксования [15], прогнозирования эффективности технологических процессов [6, 8, 17], оптимизации процесса добычи нефти [25] и также в большом количестве других направлений [9].

Другим направлением является проектирование речевых интерфейсов для информационно-управляющих систем [10].

Несмотря на большое количество работ [5, 13, 18, 22, 24, 25] по проектированию речевого интерфейса, в них недостаточно внимания уделяется практической реализации результатов

выполненных исследований. Остановимся на общих принципах проектирования человеко-машинных интерфейсов любого типа и рассмотрим актуальные задачи разработки речевого интерфейса интеллектуальной системы [11, 13, 26].

2. *Общие принципы проектирования человеко-машинных интерфейсов*

С точки зрения потребителя, интерфейс является конечным продуктом, олицетворяющим собой всю систему, и его необходимо проектировать уже на ранних стадиях разработки информационных комплексов. При этом интерфейс должен быть ориентирован на человека, т.е. отвечать его нуждам.

Речевой интерфейс, подобно графическому или любому другому человеко-машинному интерфейсу, должен выполнять следующие основные функции: предоставление клиенту исходной информации о сервисах, ввода данных со стороны клиента, вывода информации для клиента, а также исправление ошибок.

Последняя функция принципиально важна для речевого интерфейса, поскольку отличает его от других типов интерфейсов, и, в частности, от графического, так как источником ошибок при вводе исходных данных являются не только возможные ошибки человека, но и ошибки распознавателя. Основная задача проектирования – создать полнофункциональный речевой интерфейс на русском языке с применением системы массового обслуживания, обеспечить реально-временное взаимодействие с базами данных и определить методы повышения качества и устойчивости работы.

Узкие места систем распознавания речи предлагается устранять путем создания методов адаптивного управления диалогом, коррекции ошибок, разработкой удобного речевого человеко-машинного интерфейса. С этой целью было проведено исследование свойств распознавателей, а также существующих методов, моделей и подходов к реализации в этой области.

Исследование свойств и особенностей распознавателей, а также человеческого фактора в речевом взаимодействии является важным и чрезвычайно актуальным. Для исследования характеристик распознавателей и выявления зависимостей разра-

ботан аппаратно-программный технологический комплекс, обеспечивающий многоканальный доступ к реальным распознавателям через телефон. Схема этого комплекса показана на рис. 1.

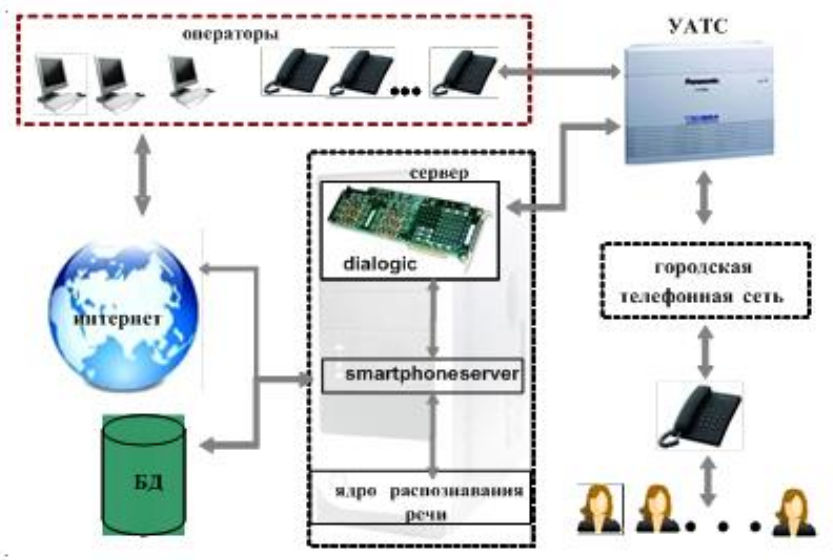

Рис. 1. Программно-технический комплекс для автоматизации исследований речевого интерфейса в многоканальном режиме

Показателем критерия качества распознавания является относительное число правильно распознанных слов WRR (WordRecognitionRate):

$$WRR = R/N,$$

где R – число правильно распознанных слов; N – общее число произнесенных слов.

Получены зависимости относительного числа правильно распознанных слов от различных факторов: от сложности блока распознавания, от продолжительности высказывания, от типа моделей, от влияния шума, стиля произношения, темпа речи, громкости голоса и др. Также получена зависимость качества распознавания речи от нескольких параметров (рис. 2).

Исследовано влияние на результаты распознавания параметров распознавателей и величины порога уверенности распознавателя. Установлено, что наибольшее влияние на результат имеют настройка параметра, определяющего чувствительность распознавателя к «лишним» словам, т.е. словам, не входящим в состав словаря, а также величина порога уверенности в гипотезе, определяющая ее состоятельность.

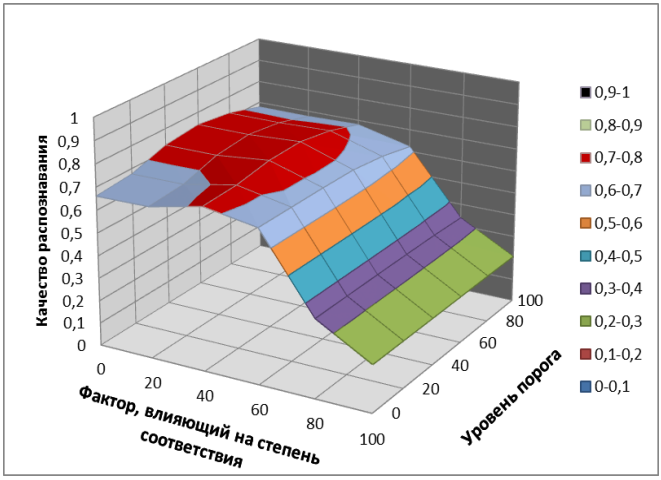

Рис. 2. Зависимость качества распознавания речи от нескольких параметров

Проведенные экспериментальные исследования позволили сделать следующие выводы о поведении машинной части речевого человеко-машинного интерфейса:

• выявлены зависимости качества распознавания от целого ряда факторов;

• знание этих зависимостей позволяет разработчику управлять в определенных пределах поведением распознавателя, влияя на содержание ответов клиента путем формулирования вопросов в «нужной» форме;

• настройка параметров может повысить, и действительно повышает, качество распознавания.

Полученные результаты использовались при практических разработках систем с речевыми интерфейсами [1, 14, 16, 17].

3. Когнитивные особенности проектирования эффективного речевого интерфейса

Главной отличительной чертой речевого интерфейса является проблема невидимости, поскольку зрение не участвует в процессе взаимодействия с машинной стороной, и восприятие вопросов и информации производится клиентом только на слух. Аудиоинформация сообщается последовательно. Следует отметить, что эта информация не сохраняется на экране дисплея или на бумаге, к ней нельзя вернуться немедленно или в любой момент. Исключение зрения из процесса взаимодействия с машиной приводит к значительному возрастанию когнитивной нагрузки на человека и к необходимости учета серьезных ограничений на восприятие информации при конструировании диалога (под когнитивной нагрузкой будем понимать умственную нагрузку, связанную с получением, обработкой, запоминанием информации). Все существующие интерфейсы создают когнитивную нагрузку на пользователя, так как он должен понять правила работы с системой, усвоить предлагаемые термины и понятия, запоминать информацию.

Особое внимание в процессе дизайна речевого интерфейса необходимо уделить трем когнитивным составляющим:

• Нагрузка на память. Ограниченность слуховой памяти требует от разработчика ограничивать количество сообщаемой клиенту информации. Помимо решений, связанных с определением количественных ограничений, необходимо наилучшим образом определить последовательность предоставления информации, оценить сравнительную значимость озвучиваемых системой данных, расставить интонационные акценты.

• Удержание внимания. Разработчик должен уметь оценивать уровень, на котором может находиться внимание пользователя на том или ином шаге диалога, предусмотреть меры по удержанию его внимания, уметь восстановить диалог, если внимание пользователя будет отвлечено внешними событиями.

• Понятийная сложность. Создавая интерфейс, разработчик должен представлять себе, насколько сложны будут для пользователей предлагаемые им термины и понятия, и как добиться, чтобы конструируемый диалог согласовывался с опытом потенциальных пользователей и с уже известными им понятиями, терминами и процедурами.

Разработчику можно предложить следующие средства для решения задачи построения эффективного и продуктивного речевого интерфейса: выбор стратегии диалога и управление им; дизайн промптов (текстов, произносимых системой); построение грамматик.

Были проведены исследования, позволившие дать ряд рекомендаций, которые будут способствовать снижению нагрузки на пользователей с точки зрения усвоения ими концептуальных понятий. К этим рекомендациям можно отнести использование универсальных команд навигации, всегда доступных пользователю вне зависимости от приложения и контекста.

При построении человеко-машинного интерфейса важно реализовывать унификацию действий и терминологию. Идея состоит в том, чтобы позволить клиенту выполнять похожие действия сходным образом, а также обеспечить на всем протяжении диалога постоянство терминологии. Экспериментально установлено, что на каждом шаге диалога пользователь укрепляет уже полученные знания о системе и обновляет свою ментальную модель системы, которая включает его знания о функциях системы, доступных командах, способах произнесения своих ответов на вопросы системы и т.д. Ментальная модель формируется быстрее и получается гораздо более ясной и понятной для пользователя, если он поймет, что аналогичные действия выполняются по одной и той же схеме. Однако на этом пути есть и свои опасности, главная из которых – монотонность, которая приводит к снижению внимания и ошибкам, а также может раздражать пользователя. При реализации необходимо также приводить пояснения и примеры. Психологические исследования показывают, что человек легче понимает задачу и запоминает информацию, если пояснить ему, что представляет собой система, почему

выполняются те или иные действия, и дать пример ответа. Использование таких пояснений и примеров помогает клиентам правильно формулировать ответы на вопросы системы, что приведет к снижению числа ошибок, значительно улучшит удовлетворенность пользователей и повысит эффективность системы.

Все эти мероприятия позволили улучшить качество функционирования системы с речевым интерфейсом [1–4, 14, 17].

4. Подход к оценке продолжительности диалога

Речевое взаимодействие клиентов с системой осуществляется путем диалога. Под диалогом обычно понимается процесс непосредственного обмена сообщениями между двумя субъектами, в котором существует смена ролей говорящего и слушающего. Такое определение применимо и к организации речевого человеко-машинного взаимодействия. В этом случае диалог состоит из вопросов (подсказок) клиенту со стороны системы и его ответов на эти вопросы.

С точки зрения анализа распознаваемости речевой информации, содержащейся в ответах клиента на вопросы, задаваемые ему системой, элементы диалога могут быть простыми и составными.

Простым элементом диалога будем называть такой ответ клиента, который содержит одно самостоятельное смысловое понятие. Простой элемент диалога может состоять из одного или нескольких слов, распознаваемых как единое целое.

Любой речевой вопрос или подсказка, озвучиваемая клиенту в процессе диалога со стороны системы, будем называть простым элементом диалога независимо от числа смысловых понятий в этом вопросе, поскольку они не подвержены процедуре распознавания.

Составным элементом диалога будем называть такой ответ клиента, который содержит несколько самостоятельных смысловых понятий, после распознавания которых формируется несколько атрибутов.

Использование составных элементов в диалоге вызвано стремлением сократить продолжительность диалога.

Предельным вариантом использования составных элементов служит вариант «монологового» взаимодействия клиента с системой, при котором все вопросы клиенту озвучиваются единым блоком, а клиент отвечает в форме одного составного элемента, содержащего ответы на все вопросы.

Таким образом, использование клиентом в своих ответах различных видов элементов диалога может быть одним из признаков классификации способов организации речевого взаимодействия клиентов с системой. При этом в ответе клиента могут содержаться только простые элементы, простые и составные элементы, а также единый составной элемент.

В процессе диалога клиента с системой должны быть выявлены и исправлены возможные ошибки, допущенные как системой, так и клиентом. Это выполняет сам клиент, получив от системы информацию, зафиксированную в ней в процессе диалога с клиентом. Такое сравнение и корректировка ошибок может осуществляться либо в процессе формирования запроса, либо после его окончания. Продолжительность диалога определяется рядом факторов: числом элементов диалога; алгоритмом управления диалогом; величиной номинальной вероятности правильного распознавания; временными затратами на элемент диалога с учетом возможных переспросов при неправильном его распознавании.

Можно выделить две группы оценок продолжительности диалога:

• предельные оценки, позволяющие получить область значений продолжительности диалога с определенным алгоритмом управления;

• средняя оценка, учитывающая все возможные сочетания, число и взаимное расположение правильно и неправильно распознанных элементов диалога, вероятность каждого сочетания и их количество.

4.1. ПРЕДЕЛЬНЫЕ ОЦЕНКИ ПРОДОЛЖИТЕЛЬНОСТИ ДИАЛОГА

Продолжительность диалога можно рассматривать с оценкой снизу, определяющей минимально возможную продолжительность диалога с конкретным алгоритмом управления, и с оценкой сверху, определяющей максимально возможную продолжительность диалога с тем же алгоритмом управления.

Определение нижней и верхней оценок продолжительности диалога позволяет получить временную характеристику диалога при различных алгоритмах его реализации. Надежность распознавателя характеризуется средней величиной вероятности (p) распознавания простого элемента диалога, которую будем называть «внутренней» вероятностью распознавания. При взаимодействии с системой клиентов, особенно новичков, не имеющих опыта работы с таким автоматом, проявляется человеческий фактор, который отражается на средней вероятности распознавания. Среднюю вероятность распознавания простого элемента, учитывающую как среднюю внутреннюю вероятность распознавания, так и влияние человеческого фактора, будем называть в дальнейшем «номинальной» вероятностью и обозначать p_0.

Поскольку номинальная вероятность распознавания элемента диалога в современных системах компьютерного распознавания речи составляет около 0,9, то нижняя оценка продолжительности диалога с приемлемой точностью отражает его реальную продолжительность. Нижняя оценка продолжительности диалога не зависит от номинальной вероятности распознавания в конкретном блоке распознавания, так как она вычисляется в предположении отсутствия неправильно распознанных простых элементов или отдельных составляющих ($m = 0$), что соответствует условию $p_0 = 1$. Таким образом, оценка снизу продолжительности диалога характеризует алгоритм управления диалогом при идеальной системе распознавания.

Что касается оценки сверху продолжительности диалога, то она характеризует тенденцию изменения продолжительности диалога при уменьшении номинальной вероятности распознавания элемента диалога.

Для удобства численного выражения величины нижней и верхней оценок продолжительности диалога введем условную

единицу измерения: временной квант τ – время, затрачиваемое на озвучивание системой или клиентом одного среднестатистического слова русского языка. Будем считать, что каждое слово в диалоге является среднестатистическим, для озвучивания которого необходимо время, равное τ. Такие предположения не являются принципиальными. Они лишь позволяют более наглядно сравнивать временные параметры различных вариантов организации управлении диалогом. Временем распознавания, проверкой в системе признаков и работой счетчиков в системе распознавания пренебрегаем, поскольку они в тысячи раз меньше временного кванта.

Поскольку нижняя оценка продолжительности диалога соответствует правильному распознаванию всех N его элементов, то в общем виде она может быть описана следующим соотношением:

$$(1) \quad T_{\text{Н}} = \left(\sum_{i=1}^{N} L_i \right) * \tau ,$$

где L_i – число временных квантов, затрачиваемых на i-й элемент диалога при правильном его распознавании (вопрос клиенту со стороны системы, ответ клиента, вопрос клиенту с просьбой подтвердить правильность распознавания его ответа, реплика клиента на эту просьбу).

Верхняя оценка продолжительности диалога определяется при условии неправильного распознавания всех его элементов, что требует их переспроса, и может быть выражена следующим соотношением:

$$(2) \quad T_{\text{В}} = \left[\left(\sum_{i=1}^{N} L_i \right) * K_{n_{\text{доп}}} + \sum_{i=1}^{N} Q_i + \left(\sum_{i=1}^{N} G_i \right) * K_{n_{\text{доп}}} \right] * \tau ,$$

где $Kn_{\text{доп}}$ – временной коэффициент переспроса, значения которого определяются соотношением (1); $n_{\text{доп}}$ – допустимое число переспросов элемента диалога при его неправильном распознавании; Q_i – число временных квантов на один элемент диалога при организации переспроса; G_i – число временных квантов на распознаваемую реплику клиента при переспросе.

Первое слагаемое определяет среднее число временных квантов, затрачиваемое на переспрос элементов диалога при до-

пустимом числе переспросов. Второе слагаемое – затраты временных квантов по всем элементам диалога на организацию процедуры переспроса при затратах на один элемент, равных Q_i.

Третье слагаемое в соотношении (2) имеет место в том случае, если при организации процедуры переспроса появляются высказывания клиента продолжительностью G_i, которые по условиям определения верхней оценки продолжительности диалога также распознаются неправильно и требуют переспроса.

Сравнение вариантов организации управления диалогом удобнее производить в виде безразмерных величин, которые могут быть получены делением обеих частей соотношений (1) и (2) на τ. Назовем получаемые величины соответственно коэффициентами нижней ($K_{T\text{н}}$) и верхней ($K_{T\text{в}}$) оценок продолжительности диалога.

(3) $$K_{T_\text{Н}} = \sum_{i=1}^{N} L_i .$$

(4) $$K_{T_\text{В}} = \left(\sum_{i=1}^{N} L_i \right) * K_{n_\text{доп}} + \sum_{i=1}^{N} Q_i + \left(\sum_{i=1}^{N} G_i \right) * K_{n_\text{доп}} .$$

Конкретные соотношения для определения нижней и верхней оценок продолжительности диалога и соответствующих им коэффициентов нижней и верхней оценок продолжительности будут получены при анализе конкретных вариантов организации управления диалогом, что будет сделано в следующих разделах. Перейдем к рассмотрению средней оценки продолжительности диалога.

Для удобства оценки влияния величины номинальной вероятности распознавания и числа допустимых переспросов на среднее время, затрачиваемое на простой элемент диалога, введем так называемый временной коэффициент переспроса, который показывает, во сколько раз увеличивается время, затрачиваемое на один переспрашиваемый элемент диалога при заданном числе переспросов:

(5) $$K_{n_\text{доп}} = \frac{T_{\text{ср.э}_i}}{t_i} ,$$

где $n_{доп}$ – допустимое число переспросов, t_i – время, затрачиваемое на запрос и ответ при первоначальном вопросе и каждом j-м переспросе i-го элемента диалога;

В диалоге с N простыми элементами или с таким же числом составляющих при диалоге из составных элементов с номинальной вероятностью p_0 любой простой элемент или отдельная составляющая при первоначальном прохождении диалога могут быть распознаны правильно и с вероятностью $(1 - p_0)$ могут быть распознаны неправильно. Вероятность появления в диалоге m неправильно распознанных простых элементов или отдельных составляющих определяется соотношением:

(6) $p_0^{N-m} * (1 - p_0)^m$.

которые группируются в $m + 1$ групп, в каждой из которых число таких состояний равно C_N^m, причем

(7) $\sum_{m=0}^{N} C_N^m = 2^N$,

а

(8) $\sum_{m=0}^{N} p_0^{N-m} * (1 - p_0)^m * C_n^m = 1$,

где N – общее число простых элементов или отдельных составляющих в диалоге; m – число неправильно распознанных простых элементов или отдельных составляющих; p_0 – номинальная вероятность распознавания элементов диалога в системе распознавания.

Среди всех 2^N состояний есть два характерных, а именно:

• состояние при $m = 0$, когда продолжительность диалога минимальна; такую продолжительность будем называть оценкой снизу;

• состояние при $m = N$, при котором продолжительность диалога максимальна; такую продолжительность будем называть оценкой сверху.

Следует отметить,что $C_N^0 = C_N^N = 1$, т.е. это единственные состояния.

4.2. СРЕДНЯЯ ОЦЕНКА ПРОДОЛЖИТЕЛЬНОСТИ ДИАЛОГА

Рассмотренные коэффициенты нижней и верхней оценок определяют только границы возможных значений продолжительности диалога. Более точная оценка может быть получена с помощью среднего времени продолжительности диалога и соответствующего ему коэффициента средней продолжительности, которые учитывают все возможные сочетания, число и взаимное расположение правильно и неправильно распознанных элементов диалога, вероятность каждого сочетания и их количество.

Среднее время продолжительности при использовании позиционно независимых алгоритмов управления диалогом с учетом (5) может быть определено следующим соотношением:

$$(9)\quad T_{\text{ср}} = \left[\begin{array}{l} \sum_{m=0}^{N} (1-p_0)^m * p_0^{N-m} * C_N^m * \\ *[L_{\text{ср}} * (N-m) + (L_{\text{ср}} * K_{n_{\text{доп}}} + Q_{\text{ср}} + K_{n_{\text{доп}}}) * m] \end{array} \right] * \tau ,$$

где первое слагаемое в квадратных скобках показывает общее количество временных квантов во всех правильно распознанных элементах диалога при наличии в нем N элементов и m неправильно распознанных; второе слагаемое в квадратных скобках показывает общее количество временных квантов во всех m неправильно распознанных элементах, включая затраты на организацию переспроса неправильно распознанных элементов (составляющие временных затрат $Q_{\text{ср}}$ и $G_{\text{ср}} * Kn_{\text{доп}}$ раздельно или совместно могут отсутствовать в конкретных вариантах алгоритмов управления диалогом, что будет показано далее).

При определении средней продолжительности при использовании позиционно зависимых алгоритмов управления диалогом необходимо учитывать дополнительные временные затраты на опрос правильно распознанных элементов диалога, который заканчивается при определении последнего из неправильно распознанных элементов.

4.3. АНАЛИЗ РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЯ СЦЕНАРИЕВ И АЛГОРИТМОВ УПРАВЛЕНИЯ ДИАЛОГОМ.

Исходя из опыта реализации речевого взаимодействия населения с автоматизированными системами массового обслуживания (АСМО) [2] необходимо отметить, что клиенты практически не имеют опыта речевого взаимодействия с автоматами и являются, по терминологии разработчиков человеко-машинных интерфейсов, «необученными», или «наивными», клиентами. Таким клиентам требуются подсказки со стороны системы, и должна обеспечиваться возможность контроля правильности распознавания сообщаемой им системой речевой информации. В этих условиях наиболее целесообразной формой речевого взаимодействия клиента с АСМО служит активный диалог с инициативой системы и с возможностью управления диалогом с машинной стороны. Определение таких параметров, как использование простых и составных элементов в диалоге, как место осуществления контроля правильности распознавания и ее корректировки по отношению к процессу формирования запроса клиента, как реализация контроля правильности распознавания по опросу системы или по указанию клиента на неправильно распознанный элемент диалога, – все это потребовало создать классификацию вариантов организации управления диалогом, позволившую осуществлять целенаправленный анализ.

Одной из основных задач при разработке алгоритмов управления диалогом является минимизация его продолжительности при обеспечении необходимой достоверности распознавания элементов диалога.

В варианте алгоритма управления диалогом с проверкой правильности и корректировкой распознавания по каждому элементу после формирования запроса использование составных элементов слегка улучшает временные оценки. Но даже для случая только составных элементов это улучшение составляет около 10% по сравнению с аналогичным алгоритмом управления диалогом при использовании в нем только простых элементов. При этом временные оценки остаются хуже, чем у наилучшего алгоритма управления диалогом с только простыми элементами.

Анализ варианта диалога, использующего единый составной элемент в «монологовом» режиме с проверкой правильности и корректировкой распознавания по каждому элементу после формирования запроса, показал, что нижняя оценка его продолжительности лучше, чем у его аналога, состоящего из простых элементов. Улучшение этой оценки с ростом числа составляющих стремится к 10%. Однако нижняя оценка продолжительности этого варианта алгоритма управления хуже даже верхней оценки лучшего варианта алгоритма управления диалогом, использующего только простые элементы с проверкой правильности распознавания и корректировкой распознавания по каждому элементу при формировании запроса, который, таким образом, является наилучшим по продолжительности среди всех рассмотренных вариантов алгоритма управления диалогом.

5. *Лингвистические особенности и роль лингвистических знаний в создании речевых систем*

Применение лингвистических знаний позволяет повысить качество распознавателей речи для проектирования эффективного речевого интерфейса. Создание качественного ядра систем распознавания речи требует учета лингвистических понятий, которые важны при реализации речевых систем и имеют иерархическую структуру по следующим уровням: фонетический, фонологический, морфологический, лексический, синтаксический и семантический.

Изучение лингвистической структуры речи, выделение из речи отдельных лингвистических элементов, таких как фонемы, морфемы, слоги, слова, предложения, имеют непосредственное отношение к исследованию систем распознавания речи и построении.

Лингвистический процессор (ЛП). Основная задача ЛП состоит в переводе текста из орфографической формы записи в фонематическую [11]. Среди вспомогательных функций ЛП можно назвать нормализацию текста, автоматическую генерацию форм слова. Среди решаемых при помощи ЛП задач в си-

стеме анализа неструктурированной речевой информации можно выделить две основных:

• транскрибирование текстов (т.е. перевод текстов из последовательности букв, специальных символов и цифр в последовательность графических символов, обозначающих фонемы – минимальные звуковые единицы языка, из которых строятся словоформы) из баз данных для акустического моделирования и обучения верификатора;

• транскрибирование слов и словосочетаний, вводимых пользователем при поиске ключевых слов в речевых данных.

Первая задача несколько сложнее, так как для акустического моделирования обычно используются записи целых предложений, т.е. синтаксически более сложных единиц, чем ключевые слова.

Обработку текста следует рассматривать как подраздел группы методов обработки естественного языка, охватывающих множество сфер, таких как автоматический перевод текстов с языка на язык, системы понимания текста и ведения диалога, автоматическое распознавание речи, распознавание текста, представленного в виде графического изображения и т.п. В современной науке выделяют два основных подхода к анализу естественного языка: глубокий (основанный на лингвистических знаниях) и поверхностный (основанный на методах машинного обучения). Первая группа методов основывается на создаваемых экспертами правилах и моделях порождения языковых явлений, вторая – на математических и статистических методах, позволяющих автоматически формировать правила с использованием обучающих корпусов лингвистических данных. Представленное деление характерно и для лингвистического процессора.

В настоящее время наибольшей популярностью пользуются статистические методы. Яркий пример использования статистических алгоритмов при обработке текста – автоматическое определение частей речи [19]. Такие алгоритмы позволяют построить вероятностные модели на основе больших корпусов данных, содержащих ручную разметку, проведенную экспертами-лингвистами, например, Брауновский корпус для английского языка [23] и Национальный корпус русского языка [12]. Благодаря подобным корпусам становится возможным обучение

статистическим контекстным правилам, которые позволяют автоматически определять часть речи слов в зависимости от окружающих их словоформ и при этом принимать не «жесткое» решение, а вероятностное. Наиболее распространен метод, основанный на применении динамического программирования. Статистические методы характеризуются высокой скоростью работы и достаточно высоким качеством производимого ими анализа. При этом их качество во многом зависит от объема корпуса, используемого при обучении, а также варьируется от языка к языку.

Первыми появились методы анализа, основанного на экспертных правилах, подобные методы восходят к трансформационным (генеративным) грамматикам Н. Хомского [20], предложившего анализировать текст путем последовательного выделения в выражении глубинной структуры — взаимосвязанных синтаксических единиц. В результате каждое предложение получает описание в виде дерева, «листьями» которого являются слова, а «ветвями» – элементы грамматики, например, NP (NounPhrase, именная фраза) и VP (VerbPhrase, глагольная фраза). Ключевое отличие подобных систем от статистических состоит в том, что грамматические структуры создаются экспертами на основе их представления о структуре языка, в то время как статистические методы полностью игнорируют данные знания и основываются лишь на данных, извлекаемых из размеченного корпуса. При этом представительность корпуса и диапазон анализа определяют асимптоту по качеству и универсальности статистических методов. Разрабатываемые современными лингвистами методы глубинного анализа текста снимают эти ограничения, однако вносят при этом свои особенности, в частности, вычислительную неэффективность.

В настоящий момент интерес к экспертным системам сохраняется, а наибольших успехов добились исследователи направления HPSG (грамматическая структура высказывания, основанная на знаниях). Данный метод структурно включает в себя не только правила грамматики, но и словарь, содержащий помимо собственно лексем, фонологическую, синтаксическую и семантическую информацию.

В современных системах с речевыми технологиями роль ЛП обычно выполняет словарь транскрипций, или лексикон. Лексикон включает список вида (слово; транскрипция). Такой подход позволяет достаточно быстро и эффективно получить транскрипции слов, однако обладает рядом недостатков: не позволяет транскрибировать слова, которых нет в словаре, что снижает уровень автоматизации; не дает возможности корректно транскрибировать слова-омонимы; не позволяет учесть контекстное взаимовлияние транскрипций смежных слов (и, соответственно, является причиной ошибок транскрибирования пограничных фонем); не предполагает возможности транскрипционного моделирования.

6. Заключение

Исследования когнитивных и лингвистических аспектов создания речевого интерфейса, комфортного для человека, показали, что лингвистический процессор является важным компонентом системы автоматического анализа неструктурированной речевой информации. Следует подчеркнуть, что редуцированный характер лингвистического процессора в современных системах распознавания и связанные с таким подходом решения имеют ряд недостатков.

Любые исследования в области речевых технологий не могут принести качественных результатов без учета данных и знаний о фонетических, морфологических, лексических и синтаксических свойствах языковых единиц.

Литература

1. БИЛИК Р.В., ЖОЖИКАШВИЛИ В.А., ПЕТУХОВА Н.В., ФАРХАДОВ М.П. *Анализ речевого интерфейса в интерактивных сервисных системах I* // Автоматика и телемеханика. – 2009. – №2 – С. 80–89.

2. БИЛИК Р.В., МЯСОЕДОВА З.П., ПЕТУХОВА Н.В., ФАРХАДОВ М.П. *Анализ речевого интерфейса при взаимодействии клиента с автоматизированной системой массового обслуживания* / Под ред. проф. В.А. Жожикашвили. – М.: МАКС Пресс, 2007. – 112 с.
3. ВАСЬКОВЬСКИЙ С.В. *Построение современных корпоративных телефонных сетей* // Датчики и системы. – 2009. – №9. – С. 54–56.
4. ВАСЬКОВСКИЙ С.В., МОРОЗОВ В.П. *Иерархический принцип построения инструментальных средств отладки* // Датчики и системы. – 2010. – №6. – С. 23–27.
5. ВЕЛИЧКОВСКИЙ Б.М. *Когнитивная наука: Основы психологии познания: В 2 т., Т. 1.* – М.: Издательский центр «Академия», 2006. – 448 с., 432 с.
6. КАРПОВ А.А. *Ассистивные информационные технологии на основе аудиовизуальных речевых интерфейсов* // Труды СПИИРАН. – 2013. – №4(27). – С. 114–128.
7. КАРПОВ А.А. *Аудиовизуальный речевой интерфейс для систем управления и оповещения* // Известия Южного федерального ун-та. Технические науки. – 2010. – №3(104). – С. 218–222.
8. КАРПОВ А.А. *Когнитивные исследования ассистивного многомодального интерфейса для бесконтактного человеко-машинного взаимодействия* // Информатика и ее применения. – 2012. – Т. 6, №2. – С. 77–86.
9. КАРПОВ А.А. *Реализация автоматической системы многомодального распознавания речи по аудио- и видеоинформации* // Автоматика и телемеханика. – 2014. – №12. – С. 125–138.
10. КАРПОВ А. А., КИПЯТКОВА И.С., РОНЖИН А.Л. *Проектирование речевых интерфейсов для информационно-управляющих систем. Учебное пособие.* – М-во образования и науки Российской Федерации, Федеральное гос. авт. образовательное учреждение высш. проф. образования Санкт-Петербургский гос. ун-т аэрокосмического приборостроения. – Санкт-Петербург, 2012. – 75 с.

11. ЛОБАНОВ Б. М., ЕЛИСЕЕВА О. Е. *Речевой интерфейс интеллектуальных систем. Учебное пособие* / Под науч. ред. В.В. Голенкова. – Минск: БГУИР, 2006. – 152 с.
12. *Национальный корпус русского языка: 2003–2005* // Сборник статей. – М.: Индрик, 2005.
13. РАСКИН Д. *Интерфейс: новые направления в проектировании компьютерных систем. The Human Interface. New Directions for Designing Interactive Systems*. – М.: Изд-во «Символ-Плюс», 2005. – 272 с.
14. РОНЖИН А.Л., КАРПОВ А.А. *Многомодальные интерфейсы: основные принципы и когнитивные аспекты* // Труды СПИИРАН. – 2006. – Вып. 1(3). – С. 300–319.
15. РОНЖИН А.Л., КАРПОВ А.А. *Проектирование интерактивных приложений с многомодальным интерфейсом* // Доклады Томского государственного университета систем управления и радиоэлектроники. –2010. – Т. 1, №1. – С. 124–127.
16. СМИРНОВ В.А., ГУСЕВ М.Н., ФАРХАДОВ М.П. *Функция лингвистического процессора в системе автоматического анализа неструктурированной речевой информации* // Автоматизация и современные технологии. – 2013. – №8. – С. 22–28.
17. ФАРХАДОВ М.П. *Распознавание речи в системах массового обслуживания населения* // Труды СПИИРАН. – 2011. – Вып. 19. – С. 65–86.
18. BALENTINE B., MORGAN D.P. *How to Build a Speech Recognition Application: A Style Guide for Telephony Dialogues*. – Enterprise Integration Group, Inc. California, 2001. – 319 p.
19. CHARNIAK E. *Statistical Techniques for Natural Language Parsing* // Artificial Intelligence. – 1997. –Vol. 18(4). – P. 33–44.
20. CHOMSKY N. *Syntactic Structures*. – The Hague: Mouton, 1957.
21. *European Telecommunications Standards Institute* [Электронный ресурс]. – URL: http://www.etsi.org
22. FANG CHEN *Designing Human Interface in Speech Technology*. – Springer, 2010. – 382 p.

23. FRANCIS W.N., KUCERA H. *Brown Corpus Manual*. – Rhode Island, Brown University, 1964.
24. GARDNER-BONNEAU D., BLANCHARD H.E. *Human Factors and Voice Interactive Systems (Signals and Communication Technology).* – Springer Science + Business Media LLC, 2008. – 469 p.
25. HARRIS R.A. *Voice Interaction Design: Crafting the New Conversational Speech Systems (Morgan Kaufmann Series in Interactive Technologies).* – Morgan Kaufman Publishers an imprint of Elsevier Science, 2004. – 598 p.
26. LEWIS J.R. *Practical Speech User Interface Design (Human Factors and Ergonomics).* – CRC Press, 2010. – 344 p.

IMPROVING THE EFFICIENCY OF SPEECH INTERFACE WITH THE USE OF COGNITIVE AND LINGUISTIC KNOWLEDGE

Mais Farkhadov, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Doctor of Technical Sciences (mais@ipu.ru)
Sergei Vaskovskii, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Candidate of Technical Sciences (v63v@yandex.ru).
Nina Petukhova, V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Candidate of Technical Sciences (nvpet@ipu.ru)
Mukhabbat Farkhadova, RUDN University, Moscow, Candidate of Philological Sciences *(farkhadova_me@pfur.ru)*

Abstract: In the modern information society a huge amount of voice data is accumulated daily, and the automation of processing can provide fundamentally new opportunities for the analysis of management decision-making automation systems, therefore the development of methods and means of automating speech information recognition and analysis systems is of great scientific and state significance. Information technology plays an enormous role in modern society. Specifically, the most natural form of communication between a man and an information environment is a speech interface. Unfortunately, not all of the results of novell theoretical and applied research are up taken to design speech interfaces. This happens because there are no easily identifiable criterias, the solutions have strong context dependence, the problems itself are hard to formalize, the problems are very labor-intensive. At the

same time, this research is of high demand since speech interfaces become increasingly widespread and the developers need practical guides. In this paper we investigate the problems of speech interfaces. Specifically, we demonstrate new methods to solve some problems that arise while designing man-machine speech technology-based interfaces. Finally, we formulate the rules, that are based on our research and experience in building speech-based applications, to design effective interfaces.

Keywords: speech interface, interface design, cognitive load, dialogue control, linguistic processor.

УДК 004.5, 004.934
ББК 32.973
DOI: https://doi.org/10.25728/ubs.2019.81.4

Статья представлена к публикации
членом редакционной коллегии М.Ф. Караваем.

Поступила в редакцию 13.05.2019.
Опубликована 30.09.2019.