

Д. А. Суранова

Структура стенда для исследования человеко-машинного интерфейса на естественном языке

D. A. Suranova

Structure of the Stand for the Creation of Human-Machine Interfaces Using Natural Language

Человеко-машинные интерфейсы стали частью профессиональной и личной жизни человека. При этом способы взаимодействия пользователя и компьютера постоянно совершенствуются. Описаны подходы к проектированию диалоговых интерфейсов для автоматизированного рабочего места оператора и приведен вариант технической реализации. Разработана структура испытательного стенда для исследования интерфейсов на примере биллинговой системы.

Ключевые слова: человеко-машинный интерфейс, синтез и распознавание речи, интерфейс, естественный язык, испытательный стенд.

DOI 10.14258/izvasu(2013)1.2-22

Введение. В настоящее время компьютеры получили широкое распространение в ежедневной жизни человека. Примерами являются роботы, навигационные системы, регистраторы и т. д. В момент своего функционирования они взаимодействуют с человеком. По типам взаимодействий можно выделить следующие:

- взаимодействие для ввода исходных данных;
- взаимодействие для получения задания;
- взаимодействие для сообщения о том, что задание было понято;
- взаимодействие для сообщения о том, что задание было принято к исполнению;
- взаимодействие для сообщения о том, что задание было выполнено;
- взаимодействие для сообщения об ошибках;
- взаимодействие для анализа результата;
- другие виды взаимодействий.

Одновременно с созданием новых устройств происходит активное совершенствование способов взаимодействия. В качестве примеров можно привести сенсорные дисплеи, интерфейсы, управляемые жестами, движением глаз, тактильные интерфейсы, например «надувная» клавиатура телефона, и др. [1].

Указанные типы взаимодействий позволяют ускорить общение при передаче простых команд от человека к компьютеру. Аналогичные прорывы будут осуществляться и при обратном общении — от компьютера к человеку.

Human-machine interfaces have become a part of professional and personal people's life. Ways of interaction between users and computer are constantly evolving. This paper describes approaches to the design of interactive interfaces for operators of personal computers workstations and variant of the technical implementation of the model. The structure of the test stand for the study of interfaces on the example of the billing system is worked out.

Key words: human-machine, speech synthesis and recognition, interface, natural language, automated work place, test stand.

Перспективным направлением является двухстороннее взаимодействие человека и компьютера на естественных языках. Примеров много — это использование естественного языка при передаче голосовых сообщений: электронные очереди, любые справочно-информационные системы.

Однако углубленного теоретического исследования возможностей создания человеко-машинных интерфейсов в настоящее время проводится недостаточно. В частности, не исследуются возможности применения таких систем в профессиональной деятельности пользователей ЭВМ; настройки модулей синтеза и распознавания к особенностям индивидуальной речи конкретного человека и к синтезированию и распознаванию речевых сигналов с высоким уровнем информативности и контекстным фоном. Исследование этой группы вопросов с применением тестовых испытаний представляется актуальной задачей.

Степень изученности проблемы. Проблемы создания человеко-машинных интерфейсов интересовали многих российских и зарубежных ученых.

Речевое взаимодействие в человеко-машинных системах рассматривалось в трудах таких ученых, как Л. Р. Рабинер, Р. В. Шафер, Б. М. Лобанов, а также в настоящее время представляет интерес для таких крупных компаний, как Google, Microsoft, Apple и др. Вопросы генерации и распознавания речевых сигналов изучали ученые В. Н. Сорокин, Н. Г. Загоруйко, Б. М. Лобанов, Л. В. Златоустова,

группа специалистов из «Центра речевых технологий» (Санкт-Петербург) и др. Технологии построения интерфейсов с элементами искусственного интеллекта были изучены автором на основе трудов Е. Сугак, Ю. Н. Филиппович, А. Ю. Филиппович, Г. С. Осипова и др.

В рассмотренных работах не исследованы проблемы адаптации и выбора модулей генерации и распознавания речи; проблемы создания исследовательских систем для изучения возможностей взаимодействия человека и компьютера на естественных языках; оценка погрешностей при речевом взаимодействии и способы их снижения, включая организацию информационной избыточности; разработка возможных приложений в составе автоматизированных рабочих мест. В данной статье будут рассмотрены подходы к решению указанных проблем.

Требования к организации эффективных интерфейсов. Интерфейс — совокупность возможностей взаимодействия по двум направлениям: от человека к компьютеру и от компьютера к человеку. В настоящее время сформировался следующий ряд основных требований к человеко-машинному интерфейсу [2]:

- **Естественность** заключается в использовании знакомой и интуитивно понятной пользователю среды.
- **Согласованность** позволяет использовать полученные на ранних этапах знания и навыки, делая интерфейс узнаваемым и предсказуемым. Согласованность важна как в пределах среды, так и в пределах самого приложения.
- **Дружелюбность** интерфейса проявляется в разрешении на каждом этапе ограниченного набора действий, в предупреждении ошибок, а также в умении адаптироваться к потенциальным ошибкам пользователя и оказывать помощь в их устранении.
- **Простота** — легкость в изучении и использовании интерфейса, но при этом должен быть доступ ко всему функционалу.
- **Принцип обратной связи** состоит в том, что каждое действие должно подтверждаться визуально или при помощи звука; в случае выполнения длительной операции — предоставление соответствующей информации об этом.
- **Гибкость** — способность учитывать уровень подготовки и производительность пользователя.

Классификацию интерфейсов можно провести по способу взаимодействия. При этом выделяют следующие виды: текстовый, графический, тактильный, жестовый, голосовой, материальный, сенситивный. Чаще всего в настоящее время используют текстовый и графический интерфейсы.

По типам интерфейсы бывают простыми и сложными.

Простой интерфейс — взаимодействие происходит в одностороннем режиме, при помощи набора ограниченного набора команд. Примером простого интерфейса являются принтер, фотоаппарат, кнопка вызова пожарной службы и т. д.

Сложный интерфейс подразумевает взаимодействие с пользователем в двухстороннем режиме с обратной связью и возможностью уточнить вопрос. Набор сценариев взаимодействий при этом велик, примерами систем со сложным интерфейсом являются поисковые, информационные, вычислительные системы.

Сложные интерфейсы представляют наибольший интерес, так как позволяют реализовать привычную для человека форму общения в виде диалога. Кроме того, с их помощью можно провести адаптацию к индивидуальным особенностям пользователя и учитывать контекст взаимодействия.

Разработка структуры испытательного стенда для исследования интерфейсов. В качестве стенда рассмотрим автоматизированное рабочее место оператора. Для примера выберем АРМ оператора биллинговой системы. Структура взаимодействия представлена на рисунке 1.

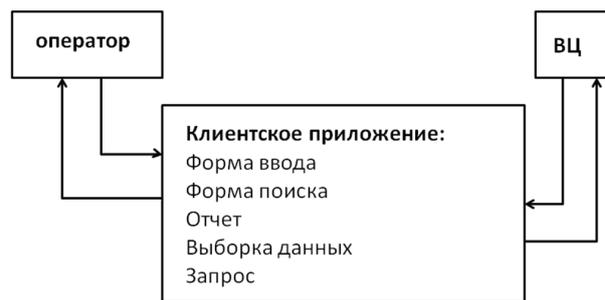


Рис. 1. Схема взаимодействия оператора с ВЦ

Операторы взаимодействуют с вычислительным центром посредством приложения, и форма взаимодействия зависит от поставленных перед ними задач. Это может быть формирование отчетности, ввод данных или поиск информации по личным счетам в системе. Но при любом способе взаимодействия пользователи видят результат своих действий на экране либо в печатном виде. Связь от человека к системе и обратно происходит на протяжении всего сеанса работы. В данном случае мы имеем дело со сложным интерфейсом и двухсторонней связью ПК и оператора.

Взаимодействие пользователя с ВЦ можно улучшить, адаптировав систему к привычным средствам коммуникации — речевому диалогу. Такая оптимизация позволит не только улучшить качество работы операторов, но и может стать инструментом для людей с ограниченными возможностями.

Вариант технической реализации. В рамках исследования была создана модель программы в виде интернет-сервиса, а в качестве системы синтеза и распознавания речи на начальном этапе были выбраны сервисы Google [3]. При этом предусмотрена возможность интеграции альтернативных модулей синтеза и распознавания, в частности таких, которые смогут дать дополнительную информацию о речевом сигнале. Это необходимо для проведения испытаний по адаптации к особенностям пользователей и оценке качества распознавания самого модуля. Также может быть выбран вариант компоновки модулей синтеза и распознавания в зависимости от предпочтений оператора. Для работы необходимы микрофон и динамики.

Модуль состоит из следующих компонентов: компонент синтеза, компонент распознавания, преобразующий речь человека в текст; компонент анализа и сопоставления текста с набором команд (он также уточняет команду в случае ошибки); компонент применения команды.

Компонент распознавания. Для преобразования речи в текст и синтеза использовались голосовые сервисы Google. Компонент активируется нажатием на кнопку микрофона либо при запуске программы. Произнесенные фразы запоминаются, и по окончании ввода распознанный текст передается в разработанный модуль анализа и сопоставления с командами. Остановить работу модуля можно, произнеся команду «готово» или нажав на микрофон внизу экрана, а исправить некорректно введенные данные — командой «вместо».

Компонент синтеза. Для преобразования речи в текст используется элемент html5 <audio> и сервисы Google.

Компонент анализа и сопоставления текста с набором команд. Текст, сформированный на основе фразы пользователя, разбирается на слова, полученные слова анализируются в произнесенной последовательности, сравниваются с набором синонимов и происходит поиск команд, соответствующих фразе. Реализован механизм запоминания контекста. Принцип работы следующий: при первом вводе проанализированные данные сохраняются в памяти и остаются там как контекст. При последующих операциях, если вводятся данные того же типа, уже существующие в контексте, они и все от них зависимые удаляются. Иначе используется сохраненный контекст. При необходимости можно очистить контекст командой «заново».

Компонент применения команды анализирует фразу и сопоставляет с набором имеющихся команд и соответствующими им действиями.

Перечень испытаний, которые могут быть проведены для данной модели.

- Отработка режимов взаимодействия для АРМ оператора.

При использовании сложных интерфейсов и механизмов адаптации необходимо предусмотреть возможность выбора режима ввода: только голос, только клавиатура или комбинирование режимов голосового ввода и ввода с клавиатуры. При этом нужно определить некоторую конечную цепочку шагов для достижения результата. Таким образом, должна быть обеспечена однозначность исполнения команд в любом из режимов. В случае использования только клавиатуры последовательность действий сводится к графическому взаимодействию. При использовании голосового ввода схема взаимодействия представлена на рисунке 2.

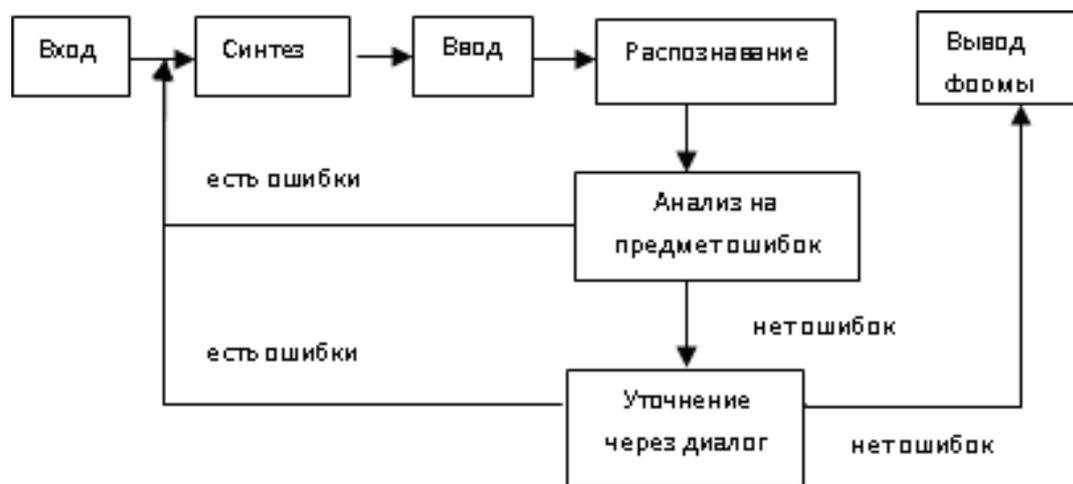


Рис. 2. Схема взаимодействия пользователя с ПК в голосовом режиме

В случае комбинированного ввода данных схема сочетает режимы, описанные выше с учетом дополнений. Необходимо предусмотреть способы пе-

рключения между данными видами взаимодействий в автоматическом и ручном режимах, так, например, произнеся команду «говорю» или «готово», можно

вызвать или остановить работу голосового модуля. В автоматическом режиме можно предложить пользователю переключиться с голосового ввода на ввод с клавиатуры при определенном количестве ошибок либо после нескольких попыток ввести фразу.

- Определение возможности адаптации к индивидуальным речевым сигналам при их распознавании.

Кроме информации о правильности распознавания фраз, при помощи набора тестов можно определить и другие характеристики речевых сигналов.

Общие характеристики для оценки модуля распознавания речи:

- скорость восприятия имеет значение для пользователей с опытом. Как правило, чем дольше человек работает в системе, тем быстрее происходит процесс взаимодействия с ней. И если при голосовом режиме ввода данных произойдет снижение скорости по причине медленного ответа от программы распознавания, то применение системы распознавания будет бессмысленным;

- надежность распознавания можно определить при помощи подсчета количества ошибок в некоторой тестовой последовательности. При этом можно использовать многоуровневые тесты. На первом уровне выделяются короткие слова и определяется количество ошибок при их распознавании. На втором уровне определяется ограниченный набор команд (порядка 10) и происходит проверка правильного определения команды. На третьем шаге происходит адаптация к особенностям человека, например, определение характеристик голоса и автоматическая донастройка. Надежность распознавания может быть улучшена как за счет использования разных модулей распознавания, так и предварительного обучения или за счет формирования словаря синонимов на основе испытаний. Перечисленные проверочные тесты можно

проводить как на отдельном человеке (зависимость от диктора), так и на группе людей (дикторонезависимые системы).

Аналогично можно разработать технологию тестирования синтезатора речи.

- Скорость восприятия синтезированной речи. Так же, как и в случае распознавания, при синтезировании речи должна быть определена скорость произнесения и формат произносимых фраз. Формат фраз должен быть определен на основе квалификации пользователя, так, например, для опытных операторов фраза должна быть максимально краткой и информативной, для менее опытных — подробной и с разъяснениями. Словарь для модуля генерации должен соответствовать привычной предметной области пользователя. Также можно провести тестирование восприятия фраз, построенных на основе общепринятых в предметной области выражений, и таким образом определить наиболее подходящие.

- Восприятие синтезированной речи человеком. От того, насколько правильно будет построена фраза, зависит последовательность дальнейших действий пользователя. Для этого необходимо формулировать фразы на основе словаря системы и опыта работы разных типов пользователей.

Заключение. В результате работы выделены подходы к созданию и тестированию человеко-машинных интерфейсов на примере АРМ оператора ВЦ.

Следующей задачей является исследование возможностей применения подобных интерфейсов. Также планируется получить оценку эффективности системы голосового взаимодействия на примере внедрения модуля в биллинговую систему. Полученные результаты могут быть использованы при проектировании интерфейсов на естественном языке в составе АРМ различного профессионального назначения.

Библиографический список

1. Фисун А. П., Гращенко Л. А. и др. Теоретические и практические основы человеко-компьютерного взаимодействия: базовые понятия человеко-компьютерных систем в информатике и информационной безопасности: монография / под ред. А. П. Фисуна. — Орел, 2004.

2. Эргономика взаимодействия человек — система. Ч. 110: Принципы организации диалога: ГОСТ Р ИСО 9241-110-2009 [Электронный ресурс]. — URL: <http://standartgost.ru/>

3. Сухов К. Web Speech Api — html5. Распознавание речи на веб-странице // Системный администратор. — 2013. — № 1–2.