ОСНОВНЫЕ ТЕНДЕНЦИИ РАЗВИТИЯ РЕЧЕВОГО ИНТЕРФЕЙСА

К. В. Фролов

Санкт-Петербургский институт информатики и автоматизации РАН
199178, Санкт-Петербург, 14-я линия В.О., д.39
<kfrolov@mail.ru>

УДК 681.3
К. В. Фролов. **Основные тенденции развития речевого интерфейса** // Труды СПИИРАН. Вып. 2, т. 1. — СПб.: СПИИРАН, 2004.
Аннотация. *Рассмотрены основные составляющие речевого интерфейса, где на данный момент он применяются и принципы его организации.* — Библ. 4 назв.
UDC 681.3
K. V. Frolov. **Major tendencies of Speech Interface evolution** // SPIIRAS Proceedings. Issue 2, vol. 1. — SPb.: SPIIRAS, 2004.
Abstract. *Reviewed major components of speech interface, its current usage and principles of organization*. — Bibl. 4 items.

Введение

Широко распространено мнение, что, речевой интерфейс может улучшить существующий пользовательский интерфейс, так как считается, что он обеспечивает более удобный и менее ограниченный способ взаимодействия компьютеров.

На протяжении многих лет производители программного обеспечения рекламировали широкое внедрение систем, основанных на речевом интерфейсе. Непрерывные разработки подобных систем привели лишь к пониманию следующего факта: в данный момент и в обозримом будущем невозможно создать универсальную, охватывающую различные предметные области, систему речевого интерфейса. Проведенный анализ показал что, эффективность речевого интерфейса зависит от правильно выбранной области применения. Например: ограниченный словарный тезаурус у специализированных систем распознавания речи (медицинские, юридические и др.) или использование речевого ввода в узкоспециализированных программах, в том числе и как дополнение к визуальным идентификаторам. В настоящий момент можно констатировать, что происходит накопление опыта по применению наиболее известных систем (Microsoft - Speech API, IBM - Viavoice, Philips - SpeechMagic, Lemout& Hauspie, Lucent Technologies, Dragon и др.), а так же и попытки разработки новых систем. Это неизбежный процесс при появлении новой технологии интерфейсного взаимодействия между человеком и компьютером. Рассмотрим текущее положение дел и существующие тенденции.

1. Составные части

Будем рассматривать речевой интерфейс через три основных составляющих: система синтеза речи, система распознавания речи и интерфейсная система, которая является связующим звеном, которая и обеспечивает основное качество речевого интерфейса.

Система синтеза речи практически реализована. Подобных систем огромное количество и качество синтеза речи достаточно для всех текущих применений.

Труды СПИИРАН. 2004. Вып. 2. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2004. Issue 2. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

Система распознавания речи является более сложной в реализации по сравнению с системой синтеза речи, и говорить о ее создании еще очень рано, но сделано очень много. Созданы системы распознавания речи, позволяющие преобразовывать в компьютерную форму представления слитную проблемно-ориентированную человеческую речь, при использовании соответствующих словарей с качеством распознавания до 98% (по материалам различных сайтов). Качество распознавания достигается настройкой системы распознавания речи на конкретного пользователя, устранением фоновых шумов и использованием высоко качественной аппаратуры.

Разработаны системы, способные распознавать сильно ограниченное количество слов, практически в любых условиях и без настройки на конкретного пользователя. Эти две системы представляют две крайности, естественно существует множество других систем распознавания речи, но они находятся, где-то между рассмотренными выше. Что объединяет все программы распознавания речи, так это ограничение используемого тезауруса. Задача распознавания произвольной слитной человеческой речи до сих пор не решена. Эту проблему пытаются решать такие фирмы как IBM, Philips, Microsoft и др., что еще раз свидетельствует о важности систем распознавания речи.

1.1. Интерфейсная система

Для использования систем синтеза и распознавания речи надо иметь некую систему, которая будет знать, когда надо синтезировать речь, когда распознавать, как передавать информацию в компьютер. Т.е. система, которая представляет собой интерфейс. Эта система является одной из самой важной, но в то же самое время ей уделялось раньше мало внимания. Многие полагали, что для начала активного использования речевого интерфейса достаточно сделать хорошую программу распознавания речи. Можно привести такой пример, что после изобретения компьютерной мыши до ее массового применения прошел не один год, а начало массового использования было связано с разработкой графического интерфейса (Windows интерфейс), для которого мышь стала очень удобным и простым устройством ввода и полностью соответствовала концепции графического пользовательского интерфейса.

В настоящее время существует потребность в разработке принципа использования речевого информационного канала. Нужна идея, где и как использовать речь, так чтобы она облегчила диалог человека и компьютера.

В качестве примера можно привести одно очень успешное внедрение речевого интерфейса. Правда речевыми интерфейсом это трудно назвать и компьютер в этом устройстве не использовался, но идея заслуживает внимания, так как на этом примере видно, чего ждут люди и чем они согласны пользоваться. Выпускаются будильники, в которых функция отключения сигнала будильника реагирует на голос или хлопок рук. Это великолепная и в тоже время простая идея.

2. Применимость

В этой статье часто употребляется слово «компьютер», но что скрывается под этим понятием? Компьютер размером с пачку сигарет или большой mainframe? От размера компьютера зависит не только его производительность, но и задачи решаемые им, удобство использования. Постараемся понять, в каком

Труды СПИИРАН. 2004. Вып. 2. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2004. Issue 2. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

сегменте компьютерной техники нам следует ожидать бурного развития речевого интерфейса.

Начнем с персональных компьютеров (ПК). Они обладают достаточной мощностью для работы систем синтеза и распознавания речи. Для этого сегмента написано очень программ использующих речевой интерфейс, прогарам распознавания и синтеза речи, но из-за существующего удобного графического интерфейса применение речи для управления ими не нашло пока поддержки у пользователей. Так при использовании речи необходимо обеспечить около компьютера низкий уровень шума, установку дополнительного оборудования, а самое главное, что существующие устройства ввода (мышь и клавиатура) выполняют все функции по управлению на достаточно высоком уровне удобства и скорости.

Следующий класс устройств это карманные ПК. По размеру они не больше пачки сигарет. На первый взгляд это идеальные устройства для речевого интерфейса: устройства ввода-вывода не очень удобные в использовании, из-за размера компьютера. Отсутствует клавиатура и мышь. Ввод информации и управление осуществляется через специальный дисплей. Этим компьютерам явно не хватает устройств ввода. Но маленькие размеры в настоящий момент ведут к малой вычислительной мощности и ресурсам карманного ПК, что в свою очередь не позволяет использовать на них системы распознавания речи. Поэтому применение речевого интерфейса ограничено только системой синтеза речи. Но мощности карманных ПК постоянно растут. И использование речевого интерфейса в подобных устройствах для управления скоро станет реальностью, что в свою очередь приведет к их более широкому распространению.

Существуют еще клиент-серверные решения. Суть их в том, что все задачи по распознаванию, синтезу и обработке речи выполняет мощный компьютер, а информация от и к пользователю поступает посредством любого устройства для передачи речи (например, телефон). Этот класс решений на сегодня наиболее востребован, так как он позволяет использовать обычный телефон, т.е. вложения в инфраструктуру минимальны. В таких странах, США или Канада существуют огромные центры обработки телефонных звонков, где задача по обслуживанию клиентов практически полностью возложена на программы, использующие речевой интерфейс. Подобные программы применяются в различных справочно-информационных центрах, службах технической поддержки клиентов, на больших телефонных коммутационных узлах корпораций и т.д. Задачи, которые решают эти программы, в основном связаны с доступом клиентов к информации из баз данных, в которых количество вариантов запросов ограничено, т.е. возможно ограничить словарь системы распознавания речи. Так как внедрение подобных центров ведет к сокращению издержек (переход от операторов центров обработки телефонных звонков к программам), то корпорации вкладывают значительные средства в разработку. Именно в этом сегменте наблюдается наибольший рост инсталляций и количества пользователей, использующих речевые технологии, присутствуют интересные наработки в интерфейсе.

2.1. Технологии ВЭБ и речь

Вэб архитектура использует различные методы ввода-вывода, абстрагируясь от их реализации. И разумным для интеграции ВЭБ и речевого интерфейса является путь превращения методов обработки речи в те методы, кото-

Труды СПИИРАН. 2004. Вып. 2. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2004. Issue 2. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

рые совместимы с ВЭБ архитектурой. Приложения ВЭБ становятся все более сложными, компьютеры существуют в нескольких форм факторах, некоторые из которых не имеют достаточно больших дисплеев и удобных устройств ввода-вывода и следственно обычные интерфейсы не могут адекватно работать с возросшим потоком информации. И в данном случае речевые технологии являются идеальным кандидатом для общения с компьютером.

Очередную надежду на интенсивное внедрение речевых технологий вселил язык гипертекстовой разметки HTML и дальнейшее его развитие XML.

XML — коллекция протоколов для представления структурированных данных в текстовом формате, который предоставляет возможность обмена между различными компьютерными платформами. В XML реализован принцип разделения данных и их представления. Это позволяет использовать различные устройства ввода-вывода от телефона с маленьким дисплеем до современного компьютера.

Производители ПО использующие речевой интерфейс поняли, что содержать свои собственные протоколы и технологии не имеет смысла из-за их несовместимости с протоколами других производителей, сложностей в разработке и внедрении. В итоге была создана группа VoiceXML forum по разработке расширения языка XML использующего речь — VoiceXML specification. В эту группы вошли крупнейшие мировые разработчики ПО. Позднее Microsoft и несколько других компаний создали свою группу для разработки альтернативной спецификации SALT (Speech Application Language Tags) в основе которой также лежит XML. SALT является не столько конкурентом VoiceXML specification сколько его дальнейшим развитием, также SALT более жестко следует спецификации языка XML (разделение данных и представления).

Язык VoiceXML specification разрабатывался в основном для применения в телефонии, где навигация по контексту осуществляется исключительно голосом. Т.е. он предлагает одно модальный (UNIMODAL) интерфейс. Что сильно ограничивает его распространение на устройства отличные от телефона.

SALT предоставляет возможность организовать многомодальный (MULTIMODAL) интерфейс. Много модальный интерфейс предоставляет возможность пользоваться для ввода информации различными устройствами. Типичный представитель такого интерфейса — графический Windows интерфейс, позволяющий вводить данные при помощи мыши, клавиатуры, джойстика…. При помощи SALT реализуется речевой пользовательский интерфейс, многие принципы работы которого заимствуются из графического пользовательского интерфейса. Следовательно, используется объектно-ориентированная, управляемая событиями модель, которая себя хорошо зарекомендовала при построении сложных графических интерфейсов. Эта модель отслеживает действия пользователя и преобразует их в события. SALT позволяет распространить эту технологию на речевой ввод.

Основное преимущество SALT это много модальность, так как это позволит использовать программы, написанные на нем не только для телефонии, но и на персональных компьютерах, а самое главное на рынке карманных ПК, которые в скором будущем достигнут необходимой мощности для работы на них системы распознавания речи.

Труды СПИИРАН. 2004. Вып. 2. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2004. Issue 2. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

3. Речевой интерфейс для персональных компьютеров

А что же с персональными компьютерами, когда мы сможем с ними общаться по голосу? Мне кажется, что это произойдет, когда будет разработан интерфейс, отличный от существующего графического (к тому времени компьютер будет выглядеть иначе, сменится его парадигма), в котором для общения с компьютером удобнее будет использовать речь.

Из принципов построения протокола SALT можно сделать вывод, что речевой интерфейс на данном этапе рассматривается, как дополнительный, расширяющий возможности существующих интерфейсов. Таким образом, вместе должны будут сосуществовать мышь, клавиатура и микрофон. Они должны работать в одной связке, не создавая помех друг другу (беда многих программ для управления компьютером по голосу). Человек должен иметь возможность выбора наиболее удобного интерфейса в каждый конкретный момент времени. Например, начать команду, используя мышь, продолжить голосовым вводом и подтвердить нажатием кнопки на клавиатуре. В современные операционные системы уже закладываться возможности использования речевого интерфейса (новые ОС фирмы Microsoft).

Сейчас актуальной задачей является разработка методов и принципов применения речевого интерфейса в существующем графическом интерфейсе пользователя.

4. Реализация речевого интерфейса

Программа распознавания и синтеза речи фирмы Microsoft – Speech API предоставляет широкие возможности для организации речевого интерфейса. Ее легко можно внедрять в уже существующие программы без их изменения. Speech API не требовательна к ресурсам компьютера (работает на Pentium II с 256 Mb оперативной памяти) Поэтому она была выбрана для организации речевого интерфейса.

В качестве программы для тестирования внедрения речевого интерфейса выбрана САПР фирмы Autodesk AutoCAD. Эта программа обладает развитым графическим Windows интерфейсом и программным интерфейсом для разработки различных дополнительных модулей, при помощи которых можно полностью управлять процессом в САПР. Знание принципов работы проектировщика в AutoCAD дало возможность выявить недостатки рассматриваемой САПР, которые можно улучшить, применив речевой интерфейс. Такие функции, как включение привязок, вызов и работа с видами, работа со слоями, отмена команд, не имеют прямого отношения к проектированию, и при их вызове традиционными методами происходит замедление процесса проектирования, так как нарушается семантическая связь действий человека через определенный интерфейс управления компьютером и оптимальным является перенос вызова этих действий на другой информационный канал. Очень удобной будет организация соответствия между звуковым и визуальным символом (иконки из панели инструментов). А самое главное эти улучшения могут быть востребованы людьми, работающими с этой САПР. Почему «могут быть»? Да потому что речевой интерфейс рассматривается как дополнительный, и человек будет выбирать, чем пользоваться.

В AutoCAD существует строка ввода команд. Она является центральным звеном в схеме управления САПР. Через нее проходят на обработку все ко-

Труды СПИИРАН. 2004. Вып. 2. Т. 1. ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн)
SPIIRAS Proceedings. 2004. Issue 2. V. 1. ISSN 2078-9181 (print), ISSN 2078-9599 (online)
www.proceedings.spiiras.nw.ru

манды, поступающие в AutoCAD, вне зависимости от того были они введены с клавиатуры, мыши или через диалоговое окно. В этой строке также хранится текущее состояние системы (текущая команда). Если речевой интерфейс будет использовать для общения с САПР эту строку, то он не будет создавать помехи в работе другим интерфейсам. Таким образом, мы получили многомодальный интерфейс.

По предварительным результатам исследований предлагаемая концепция использования речевого интерфейса в САПР AutoCAD полностью себя оправдала, ведутся работы по дальнейшему усовершенствованию программного обеспечения и принципов применения речевого интерфейса.

Литература

[1] *Kuansan Wang*. Natural Language Enabled Web Applications. <http://research.microsoft.com/stg>, 2001

[2] *Kuansan Wang.* SALT: a spoken language interface for web-based multimodal dialog systems. <http://research.microsoft.com/stg>, 2002.

[3] VoiceXML Forum. VoiceXML Specification 1.0, <http://www.voicexml.org>.

[4] Speech and Language Tags (SALT) Forum, <http://www.saltforum.org>.