

РЕЧЕВАЯ КОМПОНЕНТА В ИНТЕРФЕЙСАХ ИНФОРМАЦИОННЫХ СИСТЕМ

Федяев О.И., Гладунов С.А.
Кафедра ПМИ, ДонГТУ
fedyayev@r5.dgtu.donetsk.ua

Abstract

Fedyayev O., Gladunov S. Vocal component in interfaces of informational systems. In the article there are discussed common questions of designing an interface, and also a place of a vocal part in a common interfaces. There are some practical recommendations about in which situation a vocal dialog is efficient.

Введение

Существующие образцы компьютерных интерфейсов все еще далеки от уровня, исключающего необходимость привыкания пользователя к тому или иному программному продукту, поэтому проблема усовершенствования средств общения с программными системами остается актуальной. Очень важной является задача создания как можно более удобного и простого в использовании интерфейса, позволяющего даже очень слабо подготовленному пользователю эффективно использовать тот или иной программный продукт.

В последние годы разработке удобного интерфейса уделяется пристальное внимание со стороны ведущих производителей программных продуктов. Привычным стандартом стали многооконные системы, оснащенные визуальными средствами управления в соответствии с принципами GUI (Graphical Users Interface). Управление информационными системами больше не требует мучительного поиска нужной клавиши на клавиатуре. Все осуществляется наглядно, пользователь видит результаты своих действий непосредственно на экране, в любой момент он может обратиться к системе помощи, ставшей обязательным компонентом любой информационной системы.

Направления по улучшению интерфейса, связанные с усложнением существующих аппаратных средств, использование трехмерной графики и другие «механические» способы позволяют в определенной мере повысить удобство работы, однако не приносят принципиально новых результатов. По-видимому, решить проблему, совершенствуя существующие и ставшие традиционными средства интерфейса, не представляется возможным [1]. В этой связи вызывают интерес работы по созданию систем взаимодействия пользователей с информационными системами на основе интеллектуальных технологий.

1. Функциональная декомпозиция интеллектуализации интерфейса

Интерфейс включает в себя все аспекты, связанные с взаимодействием пользователя и автоматизированной вычислительной системы [1]. В настоящее время построить адекватную формальную модель естественного человеческого общения не представляется возможным, поэтому при создании интерфейса разработчик отталкивается от возможностей программно-аппаратных средств разработки информационных систем, а не от традиционных форм человеческого

общения. Полноценного диалога такой интерфейс не обеспечивает, что приводит к необходимости длительного изучения пользователем средств взаимодействия с каждой конкретной программной системой.

Поскольку невозможно полностью формализовать процессы создания и функционирования естественного интерфейса, то в этом случае целесообразно применить методы искусственного интеллекта, включающие в себя технологии экспертных систем, нечеткие и нейросетевые вычисления, генетические и эволюционные алгоритмы и т.п. [2].

В идеале интеллектуальный интерфейс между пользователем и системой должен обеспечивать возможность общаться на естественном языке и с помощью речи, видеть и узнавать зрительные образы, визуализировать и озвучивать пользователю внутреннее состояние системы. Декомпозиция интеллектуального интерфейса позволяет изобразить его функциональную схему как информационно-вычислительную систему (рис. 1).



Рисунок 1 - Функциональная схема интеллектуального интерфейса

Как видно на рисунке, интерфейс включает в себя две основные компоненты:

- входная - осуществляет перевод с языка общения с пользователем на внутренний язык системы;
- выходная — переводит результаты из внутреннего представления системы на язык общения с пользователем.

Входная компонента представлена рядом функциональных блоков.

Прием входной информации - совокупность программных и аппаратных средств, позволяющих получить первичное описание входного сигнала. Это может быть клавиатура, микрофон или драйвер внешнего устройства;

Распознавание входного образа - отнесение полученного сигнала к тому или иному известному системе классу. Интеллектуальность интерфейса предполагает наличие в этом блоке подсистемы адаптации, позволяющей повысить качество распознавания.

Понимание и формирование описания на внутреннем языке системы - интерпретация информации, полученной от пользователя, исходя из внутренней структуры автоматизированной системы.

Среди функций, реализуемых выходной компонентой, можно выделить следующие:

- 1) получение выходной информации на внутреннем языке системы - внутреннее представление результатов работы автоматизированной системы;
- 2) синтез выходного образа - когнитивная визуализация и озвучивание внутренних состояний системы, т.е. преобразование выходной информации из внутреннего представления в образы, понятные пользователю;
- 3) передача выходной информации ~ средства предоставления полученных образов пользователю, такие как монитор, динамик и программы работы с ними.

Среди указанных блоков интеллектуального интерфейса наибольшую сложность представляет реализация этапов распознавания и понимания. Создание этих компонентов ставит перед разработчиками множество задач, большинство из которых до настоящего времени не были качественно решены. В настоящей работе рассматриваются вопросы, связанные с одним из аспектов интеллектуализации интерфейса - обеспечением речевого управления информационными системами.

2. Анализ современных средств речевого управления

Задача автоматического распознавания речи вызывает интерес на протяжении более чем полувека. За это время было предложено множество различных методов и алгоритмов, но ни один из них не дал ожидаемого результата. В настоящий момент к решению проблемы подключились ведущие производители аппаратного и программного обеспечения, такие как IBM и Microsoft. На рынке появляется все больше программных пакетов, оснащенных подсистемами распознавания речи. Однако единого оптимального подхода выработано не было [3]. Все предложенные решения имеют свои недостатки, такие как недостаточное качество распознавания, необходимость длительной адаптации к конкретному пользователю, высокие требования к производительности аппаратных средств.

Проблема автоматического распознавания речи относится к задачам распознавания образов. В настоящее время существует множество подходов к решению задачи распознавания, однако в случае с речью их применение осложняется рядом факторов. Прежде всего, следует отметить невысокую стабильность первичного (амплитудного) описания речевого сигнала. Первичные образы одного и того же слова, произнесенного одним диктором, могут весьма существенно отличаться друг от друга [3]. Задача усложняется многомерностью образов: оптимальная частота дискретизации для речевых сигналов составляет 8-10 кГц [4], что соответствует 3-7 тыс. амплитудных характеристик для каждого слова. Другой важной проблемой является изоляция слов из непрерывно поступающего сигнала, а также выделение и удаление посторонних шумов. Практически эта задача.

как правило, решается путем настройки системы на конкретные аппаратные средства (АЦП и микрофон), которые предоставляются пользователю вместе с программным пакетом. Помимо этого, существуют задачи обеспечения инвариантности к скорости произнесения, настройки программы на голос (произношение) пользователя и многие другие.

Двумя основными типами программ распознавания речи считаются [3]:

- речевые навигаторы - управление программными или аппаратными средствами с голоса;
- программы для диктовки - речевой ввод текстовой и цифровой информации.

К первым можно отнести такие программы, как Aria Listener фирмы Prometheus products, IN3 Pro Voice Command корпорации Command Corp., Listen компании Verbex Voice Systems, QuickSwitch фирмы BitWare Consulting, Voice Pilot фирмы Voice Pilot Technology и др. Среди программ ввода текста под диктовку можно выделить NaturallySpeaking Preferred (Dragon Systems), ViaVoice 98 Executive Edition (IBM), Voice Xpress Professional (Lernout & Hauspie) и FreeSpeech 98 (Philips). Цена этих программных продуктов колеблется от 40 до 160 долл. США [5].

Авторами данной статьи был предложен свой подход к проблеме создания интерпретатора речевых команд. В основу метода распознавания речи положено использование искусственных нейронных сетей как мощного средства распознавания образов [6].

3. Роль речевого управления при организации интеллектуального интерфейса

Возможность общения человека с машиной при помощи голоса даёт ряд очевидных преимуществ. Во-первых, даже не искущённому в технике пользователю не требуется серьезной подготовки для речевого диалога с машиной, так как этот способ общения для него естественен, понятен, не слишком утомителен и, следовательно, гораздо более удобен, чем любая другая система управления. Кроме того, при речевом диалоге глаза и руки остаются свободными, а это увеличивает оперативность взаимодействия и уменьшает число ошибок оператора, не говоря уже о том, что можно включаться в диалог с машиной людям с физическими недостатками. И наконец, возможность использования в качестве терминалов обычных средств связи (телефонные аппараты и существующие линии связи) сокращает объём капитальных затрат при вводе в действие объектов с речевым управлением для широкого потребителя [3].

В то же время не стоит забывать и о традиционных средствах интерфейса. На протяжении многих лет пользователи общались с информационными системами при помощи таких устройств, как клавиатура или мышь. Эти средства предоставляют очевидные преимущества при работе и отличаются высокой надёжностью. В этой связи предлагаемый интерпретатор речевых команд разрабатывался с учетом его использования совместно с традиционными средствами интерфейса. Общая схема включения речевого блока в интерфейс информационной системы представлена на рис. 2.



Рисунок 2 - Схема взаимодействия речевого блока с традиционными средствами интерфейса

В первую очередь возникает вопрос, какими функциями наделить ту или иную интерфейсную подсистему. К сожалению, проблемы, возникающие при распознавании, отодвинули этот вопрос на второй план, и он не был в достаточной мере освещен в литературе.

Системы речевого управления имеют как преимущества, так и недостатки. Как уже отмечалось выше, к преимуществам можно отнести естественность для человека речевого общения. Это позволяет пользователю не отвлекаться на диалог (не надо искать нужную клавишу или пытаться попасть указателем мыши в нужную область экрана) и сосредоточить свое внимание целиком на решаемых задачах. Такая возможность может потребоваться при работе со сложными информационными системами, обрабатывающими большие объемы данных и требующими непрерывного внимания к процессу вычислений.

Кроме того, сложные системы, как правило, способны выполнять большое количество различных функций, определяемых множеством различных параметров. Доступ к этим функциям и параметрам обычно осуществляется через систему меню различной степени вложенности, что порой требует от пользователя последовательного выбора большого количества пунктов и подпунктов. Количество пунктов меню, доступных одновременно, ограничено размерами экрана. Речевое управление позволяет осуществлять прямой доступ к различным функциям, независимо от уровня их вложенности, что позволяет сэкономить время. При голосовом выборе действия можно использовать, наряду с мнемоническими обозначениями, более осмысленные словосочетания произвольной длины. Наконец, подача команд с голоса при наличии достаточно чувствительного микрофона может осуществляться на расстоянии. Это позволяет пользователю осуществлять управление, передвигаясь при этом по комнате и решая одновременно другие свои задачи.

При всех неоспоримых преимуществах речевого управления следует помнить и о недостатках, ограничивающих его применение. Прежде всего, к недостаткам следует отнести высокие требования к аппаратному обеспечению, а также

временные затраты, требуемые на обработку сигнала и его распознавание в реальном масштабе времени. В этой связи голосовые команды следует применять там, где процессор не занят никакими другими вычислениями. Проблему можно преодолеть, используя аппаратную реализацию системы распознавания речи, но это потребует дополнительных затрат. Помимо этого, не следует забывать, что точность распознавания в реальных системах, как правило, не выше 90-95%, поэтому там, где важно своевременно и точно отреагировать на ситуацию, предпочтительнее использовать другие средства интерфейса. Наконец, посторонний шум может очень сильно сказываться на качестве распознавания, поэтому в шумных помещениях средства речевого управления также неприменимы.

Заключение

Реализация средств речевого управления является шагом в направлении повышения интеллектуальности интерфейса информационных систем. Анализ и декомпозиция интеллектуального интерфейса выявили ряд трудно формализуемых задач, что обуславливает целесообразность применения методов искусственного интеллекта.

Использование речевой компоненты наиболее эффективно в интерфейсах многофункциональных прикладных систем, имеющих иерархическую структуру и ориентированных на решение большого количества задач. Возможность подачи команд с голоса сокращает время, необходимое пользователю на освоение незнакомой информационной системы. Эффективность речевого управления повышается в системах с двухсторонним речевым диалогом. Использование речевого блока не исключает возможности работы с другими средствами интерфейса.

Литература

1. В. Коутс, И. Влейминк. Интерфейс «человек-компьютер»: Пер. с англ. ~ М.: Мир, 1990.-501 с.
2. А. Эндрю. Искусственный интеллект: Пер. с англ. - М.: Мир, 1985. - 265 с.
3. А. Викторов, П. Солнцев. Распознавание речи. // http://www.bdi.spb.ru/4-97/raspozn_rechi.htm.
4. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. - М.: Радио и связь, 1981.-495 с.
5. С. Мясковски. Мы можем поговорить? // Мир ПК, № 3, 1999.
6. О.И. Федяев, С.А. Гладунов. Распознавание речевых слов с помощью искусственных нейросетей. - Науч. тр. Донецкого гос. тех. университета. Серия: Информатика, кибернетика и вычислительная техника, вып. 1999 - с. 145-150.