

М. В. ПРИЩЕПА, К. Ю. БАРАНОВ

ОСОБЕННОСТИ РАЗРАБОТКИ ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА МОБИЛЬНОГО ИНФОРМАЦИОННОГО РОБОТА

Проанализированы особенности разработки интерфейсов человекомашинного взаимодействия на примере пользовательского интерфейса мобильного информационного робота, предоставляющего услуги справочного характера. Описаны режимы работы робота.

Ключевые слова: диалоговые системы, мобильные информационные роботы, многомодальные интерфейсы.

Введение. Разработка пользовательских интерфейсов человекомашинного взаимодействия является актуальным направлением исследований в современном мире. Наиболее перспективным считается создание речевых и многомодальных интерфейсов как наиболее естественных для человека [1, 2]. В то же время разработка подобных интерфейсов связана с большим числом проблем, одной из которых является выбор между использованием записанного живого голоса и синтезированного для воспроизведения информации при взаимодействии с пользователем. Исследованиями подтверждено, что многие люди предпочитают слышать живые голоса, а не синтезированные [3]. Поэтому рекомендуется использовать именно их, а при необходимости — переключаться на синтезированную речь, заранее предупреждая об этом пользователя. Другой немаловажной проблемой является учет особенностей человеческого восприятия информации, которые накладывают определенные ограничения при проектировании интерфейсов.

В настоящей статье приведено описание перечисленных проблем и предложены их решения, а также представлены стратегии взаимодействия с пользователями на примере разработанного в СПИИРАН мобильного информационного робота и рассмотрены режимы работы робота при использовании его в качестве информационно-справочной системы.

Когнитивные особенности человеческого восприятия. При разработке речевого интерфейса следует учитывать три основные когнитивные составляющие человеческого восприятия (рис. 1).



Рис. 1

1. *Нагрузка на память.* Этот фактор накладывает достаточно жесткие ограничения на количество сообщаемой пользователю информации. Согласно исследованиям кратковременная память человека ограничена запоминанием в среднем 4—7 элементов [4]. Поэтому для повышения комфортности взаимодействия не рекомендуется в голосовом интерфейсе использовать более 5 пунктов [3].

2. *Удержание внимания.* Учет этого фактора актуален на этапе предоставления информации пользователю. Необходимо поток информации делить на фрагменты (если это возможно) и озвучивать только существенные данные. Следует также учитывать, что в процессе приема информации пользователь может отвлечься, поэтому для привлечения его внимания допускается использовать слова „важно“, „внимание“ и т.д.

3. *Понятийная сложность.* Уровень подготовки пользователей может быть разным, поэтому при проектировании диалоговых систем недопустимо использовать специальные термины и определения. Кроме того, действия системы и определения должны быть унифицированы, т.е. термин в разных пунктах меню интерфейса должен иметь одинаковое значение, а схожие действия должны выполняться по одной схеме во всех случаях. Необходимо также наличие универсальных команд навигации, доступных из любого пункта интерфейса (например, назад, помощь, повторить), с подробным их описанием и примерами применения.

Особенности построения структуры речевого интерфейса. Взаимодействие с пользователями всегда начинается с приветствия, от организации которого во многом зависит успешность диалога. Рекомендуемая длительность приветствия не должна превышать 15—20 с [5]. В приветствии необходимо сообщить основные правила работы с системой, при этом инструкции должны быть короткими и точными; кроме того, нужно сообщить пользователю, как он может при необходимости получить помощь в работе с системой. Заключительным этапом приветствия является предложение пользователю подать свой запрос.

Выбор нужной функции или темы — один из первых шагов диалога, обычно совмещаемый с приветствием (или следующий после него). Выбор темы можно реализовать путем предоставления пользователю списка всех возможных вариантов услуг и функций или можно предложить ему подать запрос в свободной форме. Однако второй способ может смутить неподготовленного пользователя, и, кроме того, он более сложен при реализации. Поэтому предоставление пользователю списка является предпочтительным. При этом необходимо учитывать, что меню интерфейса должно содержать не более пяти пунктов, если же количество функций достаточно велико, то целесообразно организовать иерархическую структуру меню интерфейса. Кроме того, поскольку для человеческой памяти характерен так называемый „закон края“ [6], т.е. лучше всего запоминается информация, представленная в начале и в конце диалога, то рекомендуется наиболее популярные и полезные сведения размещать в первых пунктах, а данные, которые человек должен запомнить (например, команды управления), — в конце сообщения. Среди голосовых сообщений системы можно выделить 5 основных типов (рис. 2).

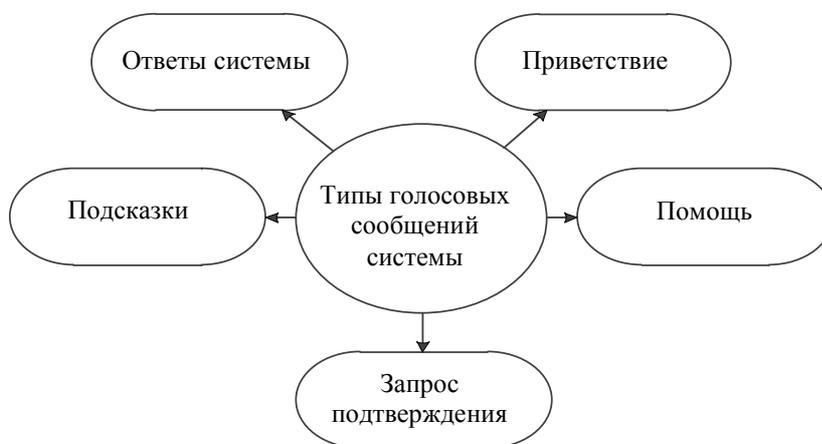


Рис. 2

В упрощенном виде диалог с пользователем состоит из вопроса, предлагаемого системой, ответа пользователя и эхо-ответа системы. При составлении голосовых запросов системы рекомендуется придерживаться нескольких правил, представленных в таблице.

Правило	Пример
Информацию, которую пользователь должен запомнить, следует помещать в конец фразы	Здравствуйте, Вы можете получить информацию о товарах и услугах. Информацию о функциях системы Вы можете получить в разделе „помощь“
Наличие примеров, содержащих ответы, позволит пользователю быстрее освоить работу с системой и сократит количество совершаемых им ошибок	Какой магазин Вас интересует? Например: „мне нужен магазин обуви“
Вопросы должны быть составлены таким образом, чтобы ответы пользователя содержали слова в основных грамматических формах	Назовите группу товара, которая Вас интересует. Например: „одежда для детей“

В ходе диалога пользователю может понадобиться справка по использованию каких-либо функций системы. Поэтому любая диалоговая система должна содержать раздел „помощь“, в котором представлена развернутая и подробная информация о функциях и услугах системы с примерами их использования; кроме того, функция „помощь“ должна быть доступна на любом этапе взаимодействия с пользователем.

Этап завершения диалога предполагает обработку запроса, введенного пользователем, и выработку ответного действия системы. На этом этапе пользователь должен иметь возможность проверить и при необходимости скорректировать свой запрос. Поэтому система должна

озвучить введенную пользователем информацию, запросить подтверждение правильности введенных данных, запустить функцию корректировки и в конце, после внесения необходимых исправлений, сообщить пользователю о том, что его запрос принят в обработку или выполняется.

Если результатом выполнения задания является вывод большого объема аудиоинформации, то рекомендуется осуществлять ее разделение на категории и блоки; также дополнительная (расширенная) информация должна предоставляться только по требованию пользователя [7].

Режимы работы мобильного информационного робота. Рассмотрим сценарии человекомашинного взаимодействия на примере информационного робота [8]. Робот представляет собой информационную стойку, две рабочие стороны которой имеют одинаковый набор и расположение сенсоров. Для поиска пользователей и записи их видеоданных используются две видекамеры, установленные в информационной стойке и расположенные на высоте лица человека среднего роста. Определение положения тела человека и слежение за его лицом реализованы с помощью алгоритма, регистрирующего перемещение естественных маркер-точек лица (центр верхней губы, кончик носа, переносица, зрачок правого глаза и зрачок левого глаза), что позволяет увеличить робастность определения координат головы при быстрых движениях и случайных видеосумах [9, 10]. Для основных процедур регистрации и обработки видеосигнала применяется библиотека OpenCV. Для обработки аудиоинформации на передних панелях робота установлены T-образные массивы микрофонов, используемых для работы системы распознавания речи, аудиолокализации источника звука и накопления аудиоданных пользователя [11]. Алгоритм аудиолокализации основывается на методе взаимной корреляции сигналов (General Cross Correlation — GCC) с применением весовой функции РНАТ (Phase Transform) [12].

Базовый сценарий работы робота в режиме диалога, учитывающий действия пользователя, представлен на рис. 3, где отражены наиболее типичные случаи взаимодействия.

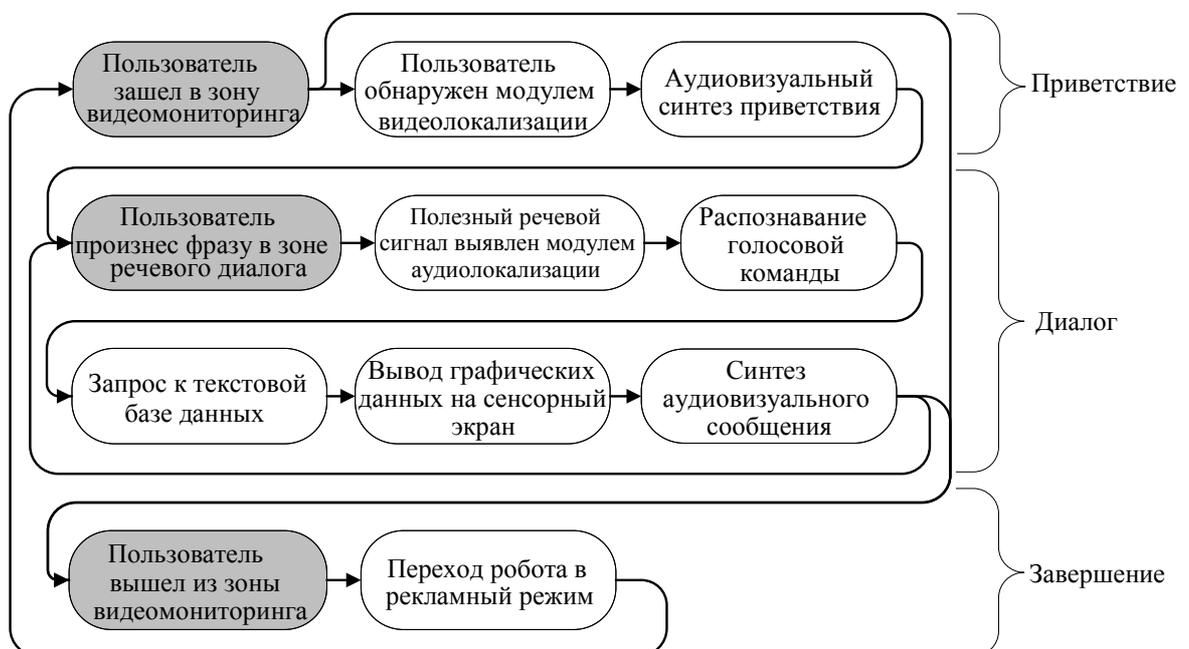


Рис. 3

В частности, цикл взаимодействия может быть представлен следующим образом: пользователь произносит голосовую команду в зоне речевого диалога, его аудиосигнал регистрируется, распознается, и осуществляется поиск необходимой информации в базе данных, а результат выводится на экран робота и синтезируется посредством „говорящей головы“.

В ходе одного сеанса взаимодействия пользователь может сделать несколько голосовых запросов к устройству, в этом случае этапы аудиообработки и вывода информации на экран повторяются соответствующее число раз.

Рассмотрим два наиболее перспективных способа использования информационного мобильного робота: 1) в торговом центре (ТЦ); 2) в организации.

1. В первом случае робот, находящийся в постоянно изменяющейся среде в сложных условиях, вследствие многолюдности в торговых центрах, должен производить мониторинг окружающей среды и проверять наличие посетителей в зоне видеомониторинга. Высокий уровень аудишумов в торговом центре, вероятно, снизит эффективность работы систем аудиолокализации и автоматического распознавания речи, поэтому предпочтение будет отдаваться графическому интерфейсу, а слежение за пользователями будет осуществляться системой видеомониторинга. При этом необходимо решить проблему выбора пользователя в случае, если робот обнаружил в зоне взаимодействия нескольких посетителей. Если посетитель найден, следует начать взаимодействие.

Основными вариантами функционирования робота в ТЦ являются:

- оказание справочных услуг;
- вывод рекламы магазинов или товаров;
- служебный режим.

Оказание справочных услуг подразумевает вывод интересующей пользователя информации на экраны робота или озвучивание данных с помощью системы синтеза речи [13]. В качестве справки могут быть показаны текущее местоположение пользователя и робота, а также маршрут до объекта, который интересует пользователя; кроме того, может быть выполнен поиск товаров и услуг по базе данных торгового центра, поиск магазина по названию или принадлежности к категории продаваемых в нем товаров. К справочным услугам относится также возможность производить в режиме онлайн заказ товара в магазине и осуществлять непосредственную связь с представителем магазина.

В режиме рекламирования на экран робота выводится информация о текущих акциях в магазинах, о товарах и услугах.

В случае низкого заряда аккумуляторов робот переходит в служебный режим, предварительно предупредив об этом пользователя.

2. Окружающая обстановка на территории государственной или коммерческой организации отличается от обстановки в торговом центре: меньше уровень шума и число пользователей. Тем не менее типы решаемых роботом задач остаются прежними: ему необходимо перемещаться по заданной территории и предоставлять услуги справочного характера, например: карту здания, информацию о расположении отделов, контактные данные сотрудников. Также с использованием каналов связи робота должна быть реализована возможность связи с отделом, запрашиваемым пользователем, кроме того, робот должен иметь возможность по запросу пользователя проводить его до места назначения. Благодаря относительно небольшому количеству сотрудников в организации робот в случае обнаружения препятствия, не отмеченного на карте, может сделать предположение о наличии перед ним потенциального пользователя и сразу начать приветствие без предварительного поиска лица. Это позволит сократить начальную фазу диалога и повысить естественность коммуникации. Тем не менее в процессе озвучивания приветствия поиск лица осуществляется и при отрицательном результате поиска робот вместо диалога продолжает свой маршрут.

Проведенный анализ требований к информационному мобильному роботу, а также анализ особенностей его эксплуатации в торговом центре и в государственной или коммерческой организации показал, что оптимизацию способов ввода/вывода информации и стратегии диалогов необходимо осуществлять в зависимости от окружающей обстановки и числа потенциальных пользователей.

Заключение. Разработка пользовательских интерфейсов требует решения ряда вопросов, касающихся психологических аспектов коммуникации и способов автоматического анализа поведения человека. Представленный в статье информационный сервисный робот оснащен сенсорами различного типа действия, что позволяет определить появление пользователя и отличить его от препятствий на пути движения робота, а его диалоговая модель взаимодействия с пользователями разработана с учетом анализа психологических аспектов человеческого восприятия информации.

Статья подготовлена по результатам исследований, проводимых при поддержке Минобрнауки РФ (федеральная целевая программа „Кадры“, госконтракт № П876) и Российского фонда фундаментальных исследований (проект № 12-08-01265-а).

СПИСОК ЛИТЕРАТУРЫ

1. *Oviatt S. L.* Ten myths of multimodal interaction // *Communications of the ACM*. 1999. Vol. 42 (11), P. 74—81.
2. *Карнов А. А., Ронжин А. Л.* Многомодальные интерфейсы в автоматизированных системах управления // *Изв. вузов. Приборостроение*. 2005. Т. 48, № 7. С. 9—14.
3. *Билик Р. В., Мясоедова З. П., Петухова Н. В., Фархадов М. П., Троценко А. Ю.* Принципы построения интерактивных систем самообслуживания с речевыми технологиями. М.: МАКС Пресс, 2008. 142 с.
4. *Miller G. A.* The magical number seven, plus or minus two // *The Psychological Rev.* 1956. Vol. 63. P. 81—97.
5. *Билик Р. В., Жожикашвили В. А., Петухова Н. В., Фархадов М. П.* Анализ речевого интерфейса в интерактивных сервисных системах // *Автоматика и телемеханика*. 2009. № 3. С. 97—113.
6. *Wickelgren W. A.* The long and the short of memory // *D. Deutsch, J. A. Deutsch*: Eds. Short-term memory. N.Y., 1975.
7. *Powers A., Kiesle S.* The advisor robot: tracing people's mental model from a robot's physical attributes // *Proc. of the 1st ACM SIGCHI/SIGART Conf. on Human-Robot Interaction, Salt Lake City, UT*. 2006. P. 218—225.
8. *Budkov V., Prischepa M., Ronzhin A.* Dialog model development of a mobile information and reference robot // *Pattern Recognition and Image Analysis, Pleiades Publishing*. 2011. Vol. 21, N 3. P. 458—461.
9. *Aissaoui A., Auguste R., Yahiaoui T., Martinet J., Djeraba C.* Fast stereo matching method based on optimized correlation algorithm for face depth estimation // *Proc. of the VISAPP 2012 — Intern. Conf. on Computer Vision Theory and Applications, Rome, Italy*. 2012. P. 377—380.
10. *Jungmann A., Schierbaum T., Kleinjohann B.* Image segmentation for object detection on a deeply embedded miniature robot // *Proc. of the VISAPP 2012 — Intern. Conf. on Computer Vision Theory and Applications, Rome, Italy*. 2012. P. 441—444.
11. *Ронжин А. Л., Карнов А. А.* Сравнение методов локализации пользователя многомодальной системы по его речи // *Изв. вузов. Приборостроение*. 2008. Т. 51, № 11. С. 41—47.
12. *Laskowski K., Jin Q., Schultz T.* Crosscorrelation based multispeaker speech activity detection // *Proc. of the Interspeech-2004, Jeju Island, Korea*. 2004. P. 973—976.
13. *Карнов А. А., Цирульник Л. Л., Железны М.* Разработка компьютерной системы „говорящая голова“ для аудиовизуального синтеза русской речи по тексту // *Информационные технологии*. 2010. Т. 9, № 8. С. 13—18.

Сведения об авторах

- Мария Викторовна Прищепа** — СПИИРАН, лаборатория речевых и многомодальных интерфейсов; мл. науч. сотрудник; E-mail: prischepa@iiias.spb.su
- Константин Юрьевич Баранов** — СПИИРАН, лаборатория речевых и многомодальных интерфейсов; программист; E-mail: baranov@iiias.spb.su

Рекомендована СПИИРАН

Поступила в редакцию
10.06.12 г.