

УДК 004.5

А.Л. Ронжин, А.А. Карпов

Проектирование интерактивных приложений с многомодальным интерфейсом*

Рассматриваются основные типы ограничений, влияющих на организацию человеко-машинного взаимодействия и конфигурирование программно-аппаратных решений при проектировании многомодальных интерфейсов интерактивных приложений.

Ключевые слова: многомодальный интерфейс, интерактивные приложения, окружающее интеллектуальное пространство, распознавание речи.

С усложнением и увеличением функциональности систем средства человеко-машинного взаимодействия становятся узким местом из-за того, что не могут обеспечить интерактивный диалог с пользователем с необходимой эффективностью и естественностью [1]. В отличие от традиционных интерфейсов на основе клавиатуры и мыши или одномодальных интерфейсов, многомодальные системы обеспечивают более гибкое использование потоков информации. Это дает возможность человеку выбирать наиболее удобный способ передачи/приема информации.

В зависимости от использованных входных и выходных модальностей выделяют несколько основных типов многомодальных интерфейсов (речь+жесты, речь+чтение по губам, направление взгляда+указание+речь и т.д.). Также выделяют интерактивные и неинтерактивные многомодальные приложения [2]. В неинтерактивных приложениях процесс выполнения задачи определен заранее, и пользователь не может на него повлиять, например при автоматическом транскрибировании записей (заседаний, семинаров), автоматическом индексировании мультимедийных данных (радио, телевизионных новостей) [3]. Напротив, в интерактивных приложениях пользователь в ходе диалога с машиной получает необходимый ему сервис и сам определяет ход работы приложения. Примерами таких приложений являются интерактивное телевидение, справочные диалоговые системы, управление техническими системами, информационная поддержка совещаний [4].

Проектирование многомодальных интерактивных приложений

При построении многомодального интерактивного приложения необходимо определить: 1) кто его будет использовать; 2) какие возможности имеют клиентские устройства; 3) в каких условиях будет проходить взаимодействие; 4) какой тип сервиса будет предоставлять приложение. Организовать взаимодействие представляется возможным только в том случае, если клиентские устройства находятся в зоне взаимодействия с пользователем и связи с информационно-телекоммуникационными сервисами, а их пользовательские интерфейсы соответствуют физическим возможностям и предпочтениям пользователя и могут обеспечить коммуникацию в текущих условиях окружающего пространства для решения актуальных целей пользователя.

Для формализации и решения задачи построения многомодального интерактивного приложения была предложена концептуальная и теоретико-множественная модель, включающая следующие элементы:

- множество целей пользователей $P = \{p_i, i \in N\}, N = \{1, \dots, n\}$, на удовлетворение которых направлено множество сервисов $S = \{s_g, g \in M\}, M = \{1, \dots, m\}$, использующих информационно-коммуникационные ресурсы $R = \{r_k, k \in C\}, C = \{1, \dots, c\}$;

- множество устройств, доступных пользователю: $D = \{d_b, b \in H\}, H = \{1, \dots, h\}$;

- множество моментов времени $T = \{t\}$;

- множество преобразований $W = \{w_f, f \in O\}, O = \{1, \dots, o\}$, выполняемых в ходе предоставления сервиса;

- множество потоков искусственных $AS = \{as_q, q \in E\}, E = \{1, \dots, e\}$ и естественных сигналов $NS = \{ns_a, a \in U\}, U = \{1, \dots, u\}$, использующихся при распознавании входных $IM = \{IM_1, IM_2, \dots, IM_{N_{IM}}\}$ и синтезе выходных модальностей $OM = \{OM_1, OM_2, \dots, OM_{N_{OM}}\}$;

*Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» (ГК №П2360, №П2579) и гранта РФФИ № 10-08-00199-а.

– множество вариантов многомодальных интерфейсов, строящееся путем целенаправленного перебора возможных комбинаций входных и выходных модальностей: $NM = \{IM_1OM_1, IM_1OM_2, \dots, IM_1OM_{N_{OM}}, \dots, IM_{N_{IM}}OM_{N_{OM}}, \dots, IM_1 \dots IM_{N_{IM}}OM_1 \dots OM_{N_{IM}}\}$;

– множество допустимых системотехнических решений Δ_α , включающее в себя множества математических моделей MM , методов MO , алгоритмов AL , аппаратно-программных реализаций AP многомодальных интерактивных приложений, представленное в следующем виде: $\Delta_\alpha = \{\alpha = \langle mm, mo, al, ap \rangle \mid mm \in MM, mo \in MO, al \in AL, ap \in AP\}$.

Учитывая, что в интерактивных системах обработка сигналов должна проводиться в режиме, близком к реальному времени, исходное множество преобразований W должно удовлетворять: $W^{(\alpha)} : AS^{(\alpha)} \times NS^{(\alpha)} \times T \rightarrow AS^{(\alpha)} \times NS^{(\alpha)}$.

Также введены четыре вида ограничений (характеристик), влияющих на процессы организации взаимодействия: 1) ограничения на способы ввода и вывода со стороны пользователя, связанные с его/ее навыками использования клиентских устройств, информационных технологий, личными предпочтениями и психофизическими ограничениями: $UC = \{UC_i, i \in X\}$; 2) ограничения на способы ввода и вывода со стороны клиентского устройства, связанные с размерами, вычислительными и сетевыми возможностями, а также аппаратной частью, реализующей сенсорные и мультимедийные функции устройства: $DC = \{DC_j, j \in Y\}$; 3) множество ограничений среды, в которой планируется организовать взаимодействие, это: уровень шумов, физические параметры атмосферы, тип помещения, число пользователей, расстояние между пользователем и клиентским устройством, наличие доступа к сетевым ресурсам и др: $EC = \{EC_k, k \in Z\}$; 4) наконец, ограничения самих сервисов, связанные с предметной областью и типом обрабатываемых информационно-коммуникационных ресурсов: $SC = \{SC_l, l \in V\}$.

Для формирования множества допустимых системотехнических решений Δ_α вводятся соответствующие подмножества декартовых произведений исходных множеств, определяющих все потенциальное пространство проектных альтернатив:

$$F_{UC}^{(\alpha)} \subseteq P_\alpha \times S_\alpha \times R_\alpha \times D_\alpha \times AS_\alpha \times NS_\alpha; F_{DC}^{(\alpha)} \subseteq P_\alpha \times S_\alpha \times R_\alpha \times D_\alpha \times AS_\alpha \times NS_\alpha;$$

$$F_{EC}^{(\alpha)} \subseteq P_\alpha \times S_\alpha \times R_\alpha \times D_\alpha \times AS_\alpha \times NS_\alpha; F_{SC}^{(\alpha)} \subseteq P_\alpha \times S_\alpha \times R_\alpha \times D_\alpha \times AS_\alpha \times NS_\alpha.$$

Таким образом, задача проектирования многомодального интерфейса сводится к поиску конструктивных путей формирования множества допустимых системотехнических решений Δ_α , удовлетворяющих ограничениям UC, DC, EC, SC :

$$\Delta_\alpha^{res} = \left\{ \begin{array}{l} \langle p_i^\alpha, s_g^\alpha, d_b^\alpha, r_k^\alpha, as_q^\alpha, ns_a^\alpha \rangle \\ \Phi^{(\alpha)} : F_{UC}^{(\alpha)} \cap F_{DC}^{(\alpha)} \cap F_{EC}^{(\alpha)} \cap F_{SC}^{(\alpha)} \rightarrow B^m \\ W^{(\alpha)} : AS^{(\alpha)} \times NS^{(\alpha)} \times T \rightarrow AS^{(\alpha)} \times NS^{(\alpha)} \end{array} \right\},$$

где элементы множеств B^m, B^n принимают значения $\{0,1\}$. Выбор полной комбинации модальностей, допустимых в проектируемом приложении, будет определяться следующим образом: $\bar{\Delta}_\alpha^{res} = \{\bar{\Theta}_\alpha(NM) \mid \Psi^{(\alpha)} : \Theta_\alpha(NM) \times \Delta_\alpha^{res} \rightarrow B^n\}$, где $\Theta_\alpha(NM)$ – множество комбинаций модальностей. На основе разрабатываемого подхода производился обоснованный выбор конкретных вариантов реализаций отображений $\Phi^{(\alpha)}$, $\Psi^{(\alpha)}$. Окончательное решение о структуре и функциях многомодального интерфейса и программно-аппаратном обеспечении, необходимом для его реализации, принималось с учетом стоимостных затрат.

Многомодальные приложения в интеллектуальном зале

На основе предложенного методологического и математического обеспечения проектирования многомодальных интерактивных приложений и конфигурирования программно-аппаратных ресурсов были разработаны приложения, относящиеся к различным классам интерактивных информационно-управляющих сервисов: справочный многомодальный киоск [5], интеллектуальный зал.

Разработанный интеллектуальный зал представляет собой распределённую систему, которая содержит сеть интеллектуальных программных модулей, активационных устройств, мультимедийных средств и аудиовизуальных сенсоров (рис. 1). Среди наиболее

важных примененных технологий следует отметить автоматическое распознавание речи, идентификацию диктора, локализацию источников звука, определение положения и слежение за подвижным объектом и лицом человека, определение позы человека. Кроме того, в разработанных интерактивных приложениях используется технология аудиовизуального синтеза русской речи «говорящая голова», созданная в ходе совместных проектов с ОИПИ НАН Беларуси и университетом Западной Богемии.

Основная задача зала – обеспечение участников совещания или лекции необходимыми сервисами на основе автоматического анализа текущей ситуации. Осведомленность зала о пространственном положении участников, их текущих действиях, роли в текущем мероприятии и их предпочтениях помогает более точно предсказать намерения и потребности участников. Применение многомодального пользовательского интерфейса, использующего речь, движения, позы и жесты участников для выявления их потребностей, обеспечивает естественный способ взаимодействия с интеллектуальным залом.

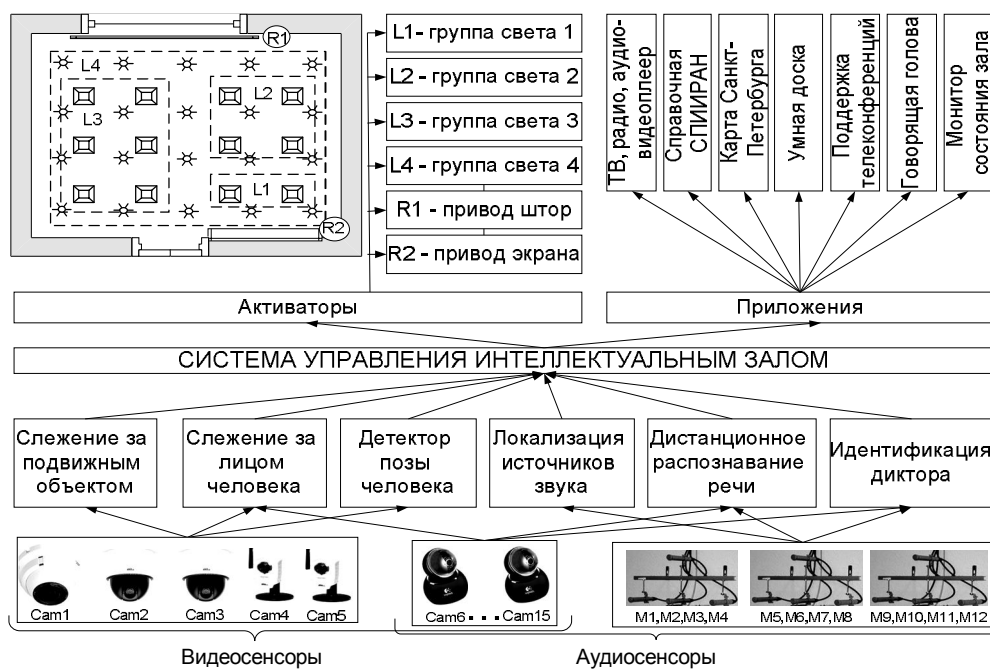


Рис. 1. Технологическая инфраструктура интеллектуального зала

Дистанционное распознавание голосовых команд, записанных с использованием массивов микрофонов, позволяет управлять освещением, шторами, проекционным экраном, поворотом камер и более сложными приложениями, например телевизором, радио-, аудио-, видеоплеером. Реализованы многомодальные приложения «Справочная СПИИРАН», предлагающее в интерактивном режиме информацию о сотрудниках института, научных подразделениях и текущих мероприятиях, и «Карта Санкт-Петербурга», где посредством голосового дистанционного запроса производится поиск улицы и вывод на экран сенсорного монитора интересующего участка карты города. Приложение «Умная доска» позволяет делать рукописные записи на сенсорной плазменной панели и с помощью голосовых команд управлять графическим интерфейсом. В приложении «Монитор состояния зала» на экран выводится информация о состоянии оборудования, пространственном положении пользователей, их речевой активности, а также реализовано сенсорное управление оборудованием.

Заключение

Разработка сети интеллектуальных залов позволит организовать совещания для участников, разделенных пространством, будет способствовать повышению сотрудничества между различными коллективами, позволит сократить расходы на транспорт и обслуживающий персонал, повысит качество образования за счет автоматизированного непрерывного контроля над каждым участником. Изучение различных комбинаций многомодальных интерфейсов для управления оборудованием интеллектуального зала помогает разрешить фундаментальные вопросы человеко-машинного взаимодействия и является богатейшим ресурсом для новых прикладных моделей в области безопасности, медицины, робототехники, логистики и других научных направлений.

Литература

1. Юсупов Р.М. От умных приборов к интеллектуальному пространству / Р.М. Юсупов, А.Л. Ронжин. – Вестник Рос. академии наук: научный и общественно-политический журнал. – 2010. – Т. 80, вып. 1. – С. 45–51.
2. Quickset: Multimodal interaction for distributed applications. Proceedings of the Fifth ACM International Multimedia Conference / P.R. Cohen, M. Johnston, D. McGee et al. – New York: ACM Press, 1997. – P. 31–40.
3. Tranter S. An Overview of Automatic Speaker Diarization Systems // S. Tranter, D. Reynolds. – IEEETrans. ASLP. – 2006. – Vol. 14, № 5. – P. 1557–1565.
4. Akker R. Supporting Engagement and Floor Control in Hybrid Meetings / R. Op den Akker, D. Hofs, H. Hondorp et al. // Springer. – 2009. – LNAI 5641. – P. 276–290.
5. Ронжин А.Л. Исследование многомодального человеко-машинного взаимодействия на базе информационно-справочного киоска / А.Л. Ронжин, А.А. Карпов // Информационно-измерительные и управляющие системы. – 2009. – Т. 7, № 4. – С. 22–26.

Ронжин Андрей Леонидович

Канд. техн. наук, доцент, зав. лабораторией речевых и многомодальных интерфейсов,
Учреждение Российской академии наук,
Санкт-Петербургский институт информатики и автоматизации РАН
Тел.: (812-3) 28-70-81
Эл. адрес: ronzhin@iias.spb.su

Карпов Алексей Анатольевич

Канд. техн. наук, сотр. лаборатории речевых и многомодальных интерфейсов,
Учреждение Российской академии наук,
Санкт-Петербургский институт информатики и автоматизации РАН
Тел.: (812-3) 28-70-81
Эл. адрес: karpov@iias.spb.su

A.L. Ronzhin, A.A. Karpov

Development of interactive applications with multimodal interface

Main types of constraints influenced on arrangement of human-machine interaction and configuration of hardware-software complexes at development of multimodal interfaces of interactive application are considered.

Keywords: multimodal interface, interactive applications, ambient intelligent space, speech recognition.