УДК 004.522

М.В. Прищепа, В.Ю. Будков, Ал.Л. Ронжин

РАЗРАБОТКА СИСТЕМЫ ИНТЕРАКТИВНОГО ТЕЛЕВИДЕНИЯ С МНОГОМОДАЛЬНЫМ ДОСТУПОМ

Прищепа М.В., Будков В.Ю., Ронжин Ал.Л. **Разработка системы интерактивного телевидения с многомодальным доступом**.

Аннотация. Рассматривается вариант системы интерактивного телевидения, использующей многомодальный интерфейс для заказа мультимедийного контента, интересующего пользователя, и управления стандартными функциями телевизора. Разработанная диалоговая модель на основе аннотированной базы данных по телепередачам и каналам обеспечивает поиск и выбор необходимой передачи с ее последующей трансляцией.

Ключевые слова: многомодальные интерфейсы, диалоговые модели, управление телевизором.

Prischepa M.V., Budkov V.Yu., Ronzhin Al.L. **Development of interactive television system with multimodal access.**

Abstract. A variant of interactive television system with multimodal interface for order multimedia content interesting for a user and control by standard TV functions is considered. The developed dialogue system based on annotated database of telecast and TV channels provides search and choosing required telecast with its following translation.

Keywords: multimodal interfaces, dialog models, TV control.

1. Введение. В настоящее время за рубежом активно ведутся разработки наиболее естественного для человека способа управления различными бытовыми приборами — речевого. Уже давно существуют системы голосового управления включением/выключением света [1], в наиболее распространенной операционной системе Windows встроенная система позволяет управлять некоторыми приложениями с помощью речи на английском языке, разнообразные детские игрушки также имеют возможность выполнять несколько простых команд, поданных с помощью голоса. Телевизоры не стали исключением и на сегодняшний день разработки голосового управления телевизором ведутся в США и в Японии [2]. Например, компании OneVideo Technology и Agile TV в США [3], начали разработку систем голосового управления телевизором. С их помощью зритель сможет переключать каналы, искать интересующие его программы и делать заказы при просмотре рекламных роликов. Сейчас систему тестируют в кабельной сети Comcast. Телевизору, подключенному к данной сети, можно сказать «ищи», «сканируй», «записывай», «спорт», «кино» и даже найти фильм по имени снимавшегося в нем артиста. Голосовая команда ак-

тивирует поиск канала, где идет нужная программа, и в случае успешного поиска производится переключение на него через 3–4 с. Также голосовой командой можно переключить телевизор в «спящий режим». Говорить нужно в микрофон на пульте дистанционного управления или на самом телевизоре.

В Японском национальном институте разработки передовых технологий разработана система голосового управления не только телевизором, но и периферийными устройствами [2]. Система позволяет с помощью голосового управления, вне зависимости от языка, особенностей речи и акцента, выполнять все основные функции, доступные на обычном пульте: включение, выключение, выбор канала, изменение размера и параметров изображения и т.д. Кроме того, она способна производить поиск видеофайла на любом носителе по его названию, а также искать ключевые слова внутри самих видеофайлов. Такая функция позволит решить давно назревшую проблему больших объемов данных — даст возможность не проводить время в поиске нужного контента, а сразу получать к нему доступ с помощью голосовой команды.

2. Многомодальные интерфейсы. Для решения глобальной проблемы человеко-машинного взаимодействия в последнее десятилетие стали использовать дополнительные каналы передачи информации помимо естественной речи (артикуляция губ, жесты, направление взгляда и т.д.). В результате начали разрабатывать так называемые «многомодальные пользовательские интерфейсы» [4, 5]. Такие интерфейсы свойственны межчеловеческому общению, здесь мы сами выбираем, какой канал для передачи какого типа информации нам наиболее удобно использовать в данный момент. Такие интерфейсы позволяют обеспечить наиболее эффективное и естественное для человека взаимодействие с различными автоматизированными средствами управления и коммуникации.

В разработанной системе многомодального управления телевизором реализовано распознавание присутствия пользователя и вербальное взаимодействие с ним на естественном языке [6]. Общая архитектура системы представлена на рис. 1, где основными модулями являются: 1) система детекции и отслеживания положения пользователя; 2) дикторонезависимая система распознавания русской речи, использующая массив микрофонов для подавления шумов и локализации источника полезного сигнала при дистанционной записи речи; 3) система аудиовизуального синтеза русской речи (говорящая голова), применяемая для виртуального помощника-аватара; 4) интерактивный графи-

ческий пользовательский интерфейс на базе сенсорного экрана; 5) менеджер диалога и диалоговая модель, включающие мультимедийные базы данных и систему управления стратегиями диалога [7].

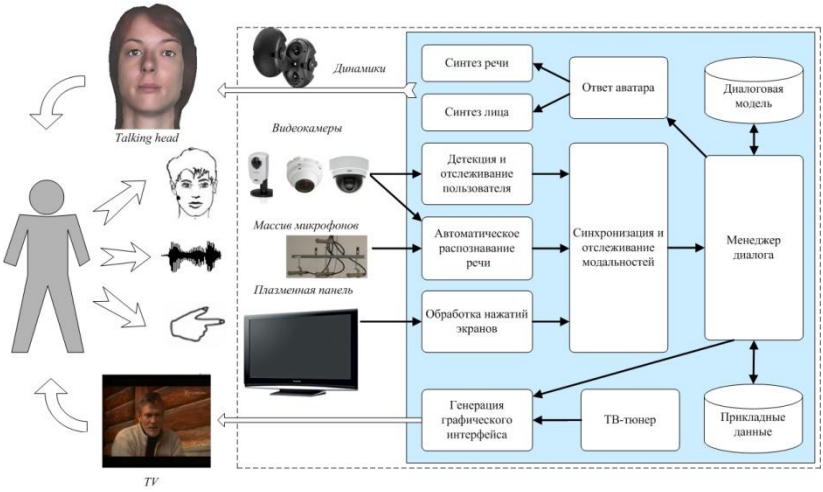

Рис. 1. Общая архитектура многомодальной системы управления телевизором.

Дополнительную естественность взаимодействию придает использование реалистичной модели аудиовизуального синтеза русской речи (так называемой «говорящей головы»), которая применяется для реализации виртуального аватара-помощника [8]. Все это позволяет организовать человеко-машинное взаимодействие наиболее естественным и эффективным для человека способом, подобным общению человека с человеком, в чем и заключается основное преимущество многомодальных пользовательских интерфейсов перед графическими и одномодальными пользовательскими интерфейсами.

Большинство из перечисленных модулей работают параллельно, причем некоторые из них обрабатывают по несколько потоков данных на разных процессорах. Для стабильной работы всех программных модулей использован персональный компьютер на базе четырехъядерного процессора Intel Core i7 с частотами каждого из процессоров в 2.7 ГГц, позволяющего параллельно выполняться восьми вычислительным процессам благодаря технологии виртуализации HyperThreading. Для телетрансляций использовался ТВ-тюнер Aver-

Media A827 USB Hybrid DVB-T с поддержкой записи телепередач, возможностью 16-канального предпросмотра и функцией временного сдвига трансляции.

3. Диалоговая модель управления телевизором. ТВ-тюнер, подключенный к компьютеру, ведет непрерывную трансляцию каналов телевещания, доступных в Санкт-Петербурге. С помощью компьютера осуществляется обработка запросов пользователя и управление ТВ-тюнером. Плазменная панель также подключена к компьютеру, на ней отображается графическая информация, телевещание, и графическое меню выбора действий согласно запросам пользователя.

Базовый сценарий функционирования системы в зависимости от действий пользователя, представлен на рис. 2. Здесь отражены наиболее типичные случаи взаимодействия, например: 1) пользователь прошел мимо системы слишком быстро, чтобы сработал модуль видеолокализации; 2) пользователь вошел в зону видеомониторинга, был запущен аудиовизуальный синтез приветствия, но пользователь прошел дальше; 3) пользователь произнес голосовую команду в зоне речевого диалога, его аудиосигнал зарегистрирован как полезный, распознан, произведен поиск необходимой информации в базе данных, а результат выведен на экран системы и синтезирован посредством «говорящей головы», после чего пользователь ушел от системы.

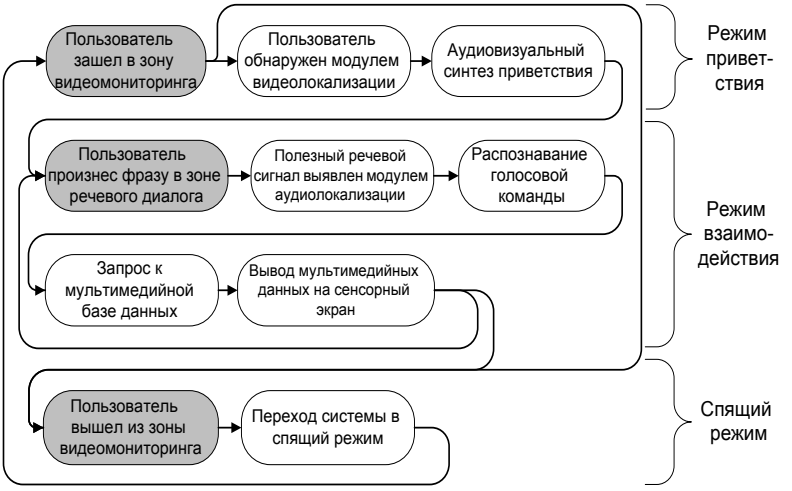

Рис. 2. Базовый сценарий функционирования системы при взаимодействии с пользователем.

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

При разработке диалоговой модели взаимодействия проанализированы возможные запросы пользователей и составлена схема формирования различных вариантов голосовых команд (рис. 3). Запросы пользователей можно разделить на четыре основных типа: 1) «Что», 2) «Где», 3) «Когда», 4) «Прямой запрос». Блок «Где» состоит из ключевых синтагм «Где», «На каком канале», «По какому каналу», далее идет синтагма из списка «Категория», либо из списка «Название», также после первых синтагм могут идти элементы из списков «Будущее время» или «Настоящее время», после которых также осуществляется переход к спискам категорий или названий. Пример запроса подходящего к блоку «Где»: «По какому каналу идут мультфильмы?». Ключевые элементы из списков «Будущее время» или «Настоящее время» могут быть разными, например для настоящего времени – «Показывают», «Транслируют» и т.д. Затем осуществляется переход к спискам категорий и названий, например: «Где идут новости?».

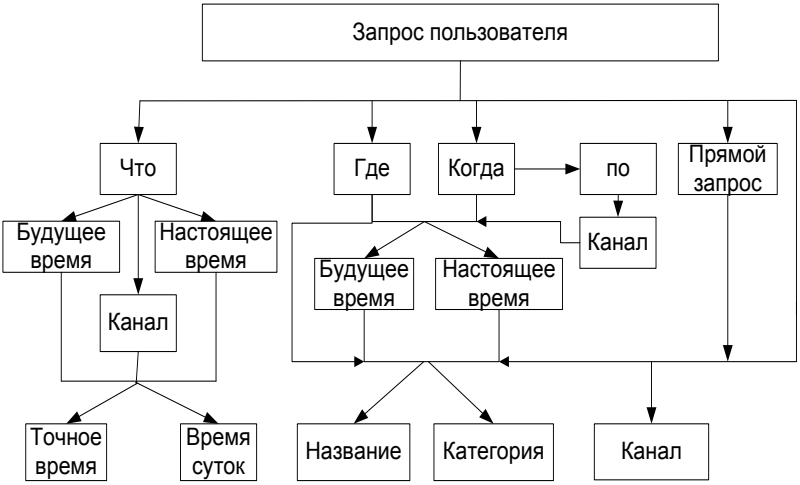

Рис. 3. Схема формирования запросов пользователей.

Блок «Когда» состоит из ключевых синтагм «Когда», «Во сколько», «В котором часу», и от него осуществляется переход либо к спис-

кам «Будущее время» и «Настоящее время» с последующим формированием фразы аналогично блоку «Где», либо к ключевой синтагме из списка «Канал», а затем повторяется переход как в первом варианте.

Блок «Прямой запрос» содержит ключевые слова «Хочу», «Переключи», «Дай», «Включи». Переход после первой ключевой синтагмы осуществляется напрямую к спискам категорий, названий или каналов. Примеры запросов: «Включи фильм», «Переключи на пятый». Также осуществим переход непосредственно к спискам категорий, названий или каналам. Например: «Россия», «Первый» и т.д.

В блоке «Что», от ключевой синтагмы «что» осуществляется переход к элементам из списков «Будущее время» и «Настоящее время», после чего возможен переход к списку каналов, с выбором определенного, либо сразу к выбору определенного времени. Например: «Что будет идти на канале культура через час?», «Что идет в 17:00?».

Семантический анализ запроса происходит по ключевым фразам. Вначале определяется вхождение запроса в тот или иной блок. Далее определяется требование пользователя по принадлежности ключевого слова к одному из следующих полей «канал», «название», «категория». После этого происходит анализ данных полученных от пользователя, например точное время. Соотнеся все данные, система находит решение и выдает пользователю ответ.

После анализа возможных запросов пользователей была разработана диалоговая модель системы. До появления пользователя система находится в режиме ожидания запросов. После получения запроса проверяется наличие ответа. При этом может быть один ответ, может быть несколько вариантов ответов или их может не быть вовсе. В случае наличия одного ответа в зависимости от формы запроса система дает ответ пользователю, и телевизор переключается на необходимую телепередачу.

Если ответов несколько, например, при запросе новостей они могут идти по нескольким каналам, система выводит список каналов удовлетворяющих требованиям и предлагает уточнить запрос, т.е. выбрать конкретный канал или телепередачу. В случае, когда после запроса не удается найти интересующее пользователя, система извиняется и предлагает выбрать что-то другое.

Далее приведены примеры тестовых диалогов для нескольких ситуаций:

Пользователь: *Включи новости.*

Система: *В настоящее время новости идут по нескольким каналам, Вы можете выбрать канал из списка или сделать новый запрос.*
Экран ТВ: *Выводится список каналов.*
Пользователь: *НТВ.*

Пользователь*: По каким каналам идет сериал «Трюкачи»?*
Система: *Сериал идет по каналу «РЕН ТВ». Переключаю.*
Экран ТВ: *Выводится канал «РЕН ТВ».*

Пользователь: *Что будет идти по каналу «Россия» завтра вечером?*
Система: *Список вечерних программ представлен на экране.*
Экран ТВ: *Выводится список вечерних программ.*

Пользователь: *Что сейчас идет по каналу «Первый»?*
Система: *В настоящее время по каналу «Первый» идет «передача» «Хочу знать с Михаилом Ширвиндтом». Переключить на этот канал?*
Пользователь: *Нет.*

Пользователь: *Где будет идти фильм «Корабль призрак»?*
Система: *К сожалению, на этой неделе этот фильм не транслируется в эфире. Вы можете сделать новый запрос.*

После того как пользователь сделал свой выбор или отказался от выполнения задания, система снова переходит в режим ожидания его запросов.

4. Онтологическая модель программы телепередач. Обработка голосового запроса и поиск необходимого телевизионного контента происходят по аннотированной базе данных программы телепередач. Для создания онтологии и моделирования предметной области исследования применялась компьютерная среда Protege. Сама среда обеспечивает создание классов, их экземпляров, а также отношений между ними. При установке дополнительных модулей, таких как SWRL и SWRLJessTab появляется возможность создания правил и запросов к онтологиям. На рис. 4 представлена общая структура предложенной онтологии, где показаны абстрактные классы и отношения между ними.

Предложенная онтологическая модель состоит из восьми классов: 1) «Программа передач»; 2) «Элемент»; 3) «Время»; 4) «Телеканал»;

5) «Телепередача»; 6) «Участник»; 7) «Телекомпания»; 8) «Категория». Рассмотрим каждый из классов подробнее.

Класс «Программа передач» содержит набор сущностей класса «Элемент», которые состоят из классов «Время», «Телеканал» и «Телепередача». Эти три класса связаны между собой для организации выбора необходимой информации по данным одного из классов.

Классы «Телепередача» и «Участник» являются классами одного уровня, и связь предназначена для выбора набора телепередач по составу участников.

Класс «Телекомпания» предоставляет возможность выбора данных из класса «Телепередача».

Последний класс «Категория» является вспомогательным классом для выбора определенных групп каналов или телепередач.

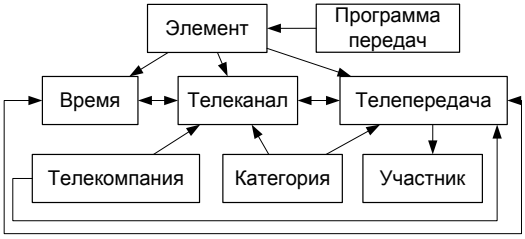

Рис. 4. Онтология предметной области программы телепередач.

База данных пополняется путем скачивания ежедневной программы телепередач с бесплатных сервисов в Интернете. После чего вся информация структурируется по ключевым словам.

В таблице приведена структура базы данных для основных классов. Как видно для класса «Телеканал» определены поля:

- Название — название телеканала.
- Частота — частота вещания телеканала.
- Номер – порядковый номер телеканала.
- Режим доступа – платность/бесплатность телеканала.
- Категория – в зависимости от направленности вещания, канал может быть музыкальным, развлекательным и т.д.
- Формат.

Для остальных классов определены соответствующие поля в базе, по которым и происходит поиск.

Структура базы данных для основных классов

<table>
<tr><td rowspan="2"></td><td colspan="8">Сущность</td></tr>
<tr><td>Теле-канал</td><td>Теле-ком-пания</td><td>Теле-переда-ча</td><td>Про-грамма передач</td><td>Участ-ник</td><td>Кате-гория</td><td>Эле-мент</td><td>Вре-мя</td></tr>
<tr><td rowspan="8">Профиль</td><td>Название</td><td>Назва-ние</td><td>Назва-ние</td><td>Источник</td><td>ФИО</td><td>Назва-ние</td><td>Но-мер</td><td>Дата</td></tr>
<tr><td>Частота</td><td rowspan="7">Местоположение</td><td rowspan="3">Описание</td><td rowspan="7">Актуальность</td><td rowspan="7">Описание</td><td rowspan="7">Описание</td><td rowspan="7">Ссылка</td><td rowspan="3">Время суток</td></tr>
<tr><td rowspan="2">Номер</td><td rowspan="3">Возрастная категория</td><td rowspan="2">Время начала</td></tr>
<tr><td rowspan="3">Режим доступа</td><td rowspan="4">Формат</td></tr>
<tr><td rowspan="2">Катего-рия</td><td rowspan="3">Время конца</td></tr>
<tr><td>Формат</td></tr>
</table>

В результате анализа потенциальных запросов пользователей и классификации телепередач разработана диалоговая модель взаимодействия и составлена аннотированная база данных, содержащая не только названия телепередач и телепрограмму, но и данные о телекомпаниях, телепередачах, популярных актерах и каналах. Далее планируется протестировать разработанную систему управления телевизором для широкого круга пользователей.

4. Заключение. Создание естественных для человека многомодальных способов взаимодействия с компьютерными системами, телевизионными приемниками и другой бытовой техникой, основанных на распознавании и понимании речи, жестов и других выражений естественного поведения человека — это только первый шаг к переходу от парадигмы повсеместных вычислений (*ubiquitous computing*) к окружающему интеллектуальному пространству (*ambient intelligence*), способному ненавязчиво удовлетворять запросы пользователей.

Литература

1. *Ронжин А.Л., Карпов А.А., Ли И.В.* Речевой и многомодальный интерфейсы. М.: Наука, 2006. 173 с.
2. *Noam R. Shabtai* Advances in Speech Recognition. Rijeka: Sciyo. 2010. 164 p.
3. http://www.onevideo.com/OneListener.htm
4. *Oviatt, S.L.* Ten myths of multimodal interaction // Communications of the ACM/ 1999. 42(11). New York: ACM Press. P. 74–81.
5. *Карпов А.А., Ронжин А.Л.* Многомодальные интерфейсы в автоматизированных системах управления // Известия вузов. Приборостроение. 2005. Т. 48, № 7. С. 9–14.
6. *Ронжин А.Л., Карпов А.А., Кагиров И.А.* Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информационно-управляющие системы. 2009. Вып. 42, т. 5. С. 32–38.
7. *A.A. Karpov, A.L. Ronzhin*, Information Enquiry Kiosk with Multimodal User Interface // Pattern Recognition and Image Analysis, Moscow: MAIK Nauka/Interperiodica, Vol. 19, № 3, 2009, P. 546–558.
8. *А.А. Карпов, Л.И. Цирульник, М. Железны.* Разработка компьютерной системы "говорящая голова" для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. № 8, т. 9. С. 13–18.

Прищепа Мария Викторовна — программист лаборатории речевых и многомодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: модели взаимодействия пользователей с информационными роботами, разработка персонифицированных стратегий человеко-машинного диалога, адаптация технологий аудио- и видеообработки для их применения на подвижной платформе. Число научных публикаций — 9. prischepa@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081. Научный руководитель — д-р техн. наук, доцент А.Л. Ронжин.

Prischepa Maria Viktorovna — programmer, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: models of interaction between user and information robot, development of personified strategies for human-machine interaction. The number of publications — 9. prischepa@iias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081. Scientific adviser — Dr. Tech. Sci., Assoc. Prof. A.L. Ronzhin.

Будков Виктор Юрьевич — аспирант лаборатории речевых и многомодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: многоканальная обработка аудиовизуальных сигналов, веб-технологии для дистанционного управления и коммуникации. Число научных публикаций — 10. budkov@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081. Научный руководитель — д-р техн. наук, доцент А.Л. Ронжин.

Budkov Viktor Yurievich — Phd Student, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: multichannel audiovisual signal processing, web technologies for distant control and communication. The number of publications — 10. budkov@iias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office

phone +7(812)328-7081, fax +7(812)328-7081. Scientific adviser — Dr. Tech. Sci., Assoc. Prof. A.L. Ronzhin.

Ронжин Александр Леонидович — аспирант лаборатории речевых и многомодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: окружающее интеллектуальное пространство, видеообработка, анализ контекста. Число научных публикаций — 9. ronzhinal@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081. Научный руководитель — канд. техн. наук А.А. Карпов.

Ronzhin Alexander Leonidovich — Phd Student, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: ambient intelligence, video processing, context processing. The number of publications — 9. ronzhinal@iias.spb.su; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081. Scientific adviser — PhD A.A. Karpov.

Поддержка исследований. Данное исследование ведется в рамках проекта ОНИТ РАН № 4.2 «Разработка средств универсального многомодального доступа для системы интерактивного телевидения».

Рекомендовано лабораторией речевых и многомодальных интерфейсов, заведующий лабораторией, д-р техн. наук, доцент А.Л. Ронжин.
Статья поступила в редакцию 15.11.2010.

РЕФЕРАТ

Прищепа М.В., Будков В.Ю., Ронжин Ал.Л. **Разработка системы интерактивного телевидения с многомодальным доступом.**

В статье обсуждаются текущие результаты по разработке системы интерактивного телевидения с многомодальным доступом. Изучение таких систем за рубежом ведется уже более 10 лет. Одна из первых систем в США была запущена для пользователей сети транслирующей кабельное телевидение, а в Японии реализовано голосовое управление не только телевизором, но и периферийным оборудованием. Для решения глобальной проблемы человеко-машинного взаимодействия используются дополнительные каналы передачи информации помимо естественной речи (артикуляция губ, жесты, направление взгляда и т.д.) и, так называемые, «многомодальные пользовательские интерфейсы. Такие интерфейсы свойственны межчеловеческому общению, здесь мы сами выбираем, какой канал для передачи какого типа информации нам наиболее удобно использовать в данный момент.

В разработанной системе управления функциями телевизора многомодальный интерфейс обеспечивает естественное взаимодействие с пользователями. После определения присутствия пользователя система начинает диалог с ним, выявляя его предпочтения. При разработке диалоговой модели взаимодействия проанализированы возможные запросы пользователей. Также составлена база данных, содержащая информацию о телепередачах, каналах, телекомпаниях и участниках телепередач, и позволяющая по ключевым словам определить интересующую пользователя программу или телевизионный контент.

SUMMARY

Prischepa M.V., Budkov V.Yu. Ronzhin Al.L. **Development of interactive television system with multimodal access.**

The current results of development of system of interactive television system with multimodal access are discussed. Such systems have already studied abroad more than 10 years. One of the first systems was started in USA for user of cable television transmitting network and in Japan the system with speech control of TV and peripheral equipment was developed. For decision of global problem of human-computer interaction additional channels of information transmitting besides natural speech (lips articulation, gaze, gestures and etc.) are used, the given research direction is known as multimodal interfaces. These interfaces are similar for human-human communication. Here user decides which channel for transmitting of which information is best in the current moment.

In the developed system of TV functions control the multimodal interface provides natural interaction with user. After user presence detection the system starts a dialogue and determinates his/her preferences. Probable user requests were analyzed during development of the dialog model of interaction. Also the database, which includes information about telecast, TV channels, television companies and telecast participants, was created and it allows the system to determine and deliver the program or television content, which is needed for the current user.