

РАЗРАБОТКА МНОГОМОДАЛЬНОГО ИНФОРМАЦИОННОГО КИОСКА

А. Л. РОНЖИН^{1,▼}, А. А. КАРПОВ², АН. Б. ЛЕОНТЬЕВА³, Б. Е. КОСТЮЧЕНКО¹

Санкт-Петербургский институт информатики и автоматизации РАН

СПИИРАН, 14-я линия ВО, д. 39, Санкт-Петербург, 199178

¹<ronzhin@iiias.spb.su>, ²<karpov@iiias.spb.su>, ³<an_leo@iiias.spb.su>

УДК 681.3

Ронжин А. Л., Карпов А. А., Леонтьева Ан. Б., Костюченко Б. Е. **Разработка многомодального информационного киоска** // Труды СПИИРАН. Вып. 5. — СПб.: Наука, 2007.

Аннотация. *Статья представляет результаты исследования и разработки аппаратной части исследовательского стенда информационно-справочной системы с многомодальным интерфейсом пользователя («многомодальный информационный киоск»). Стенд объединяет стандартные средства ввода/вывода информации (сенсорный экран, клавиатуру и стерео динамики) и устройства, обеспечивающие бесконтактное взаимодействие (видеокамера, микрофоны). Для дистанционного распознавания русскоязычных голосовых команд применяется массив микрофонов, позволяющий локализовать источник полезного сигнала и понизить влияние акустических шумов. Миниатюрная видеокамера обеспечивает оптическое отслеживание положения пользователя в рабочей зоне киоска. Разработанный прототип киоска будет использован для создания модели информационно-справочной системы и исследования когнитивных аспектов взаимодействия пользователя с системой, учитывая различные способы коммуникации.* — Библиограф. 19 назв.

UDC 681.3

Ronzhin A. L., Karpov A. A., Leontyeva An. B., Kostuchenko B. E. **The development of the multimodal information kiosk** // SPIIRAS Proceedings. Issue 5. — SPb.: Nauka, 2007.

Abstract. *The paper presents results of the research and the design of hardware part of a test bench of an information enquiry system with multimodal user interface (“multimodal information kiosk”). The test bench combines standard input/output devices (sensory monitor, keyboard, stereo speakers) and input devices for contactless human-computer interaction (video-camera, microphones). For distant recognition of speech commands of Russian a microphone array is applied that allows to locate a source of useful audio signal and eliminate acoustic noises. A portable video-camera is applied for vision-based tracking person's position inside the kiosk working area. The developed kiosk prototype to be used for creation of the model of information enquiry system and studies of cognitive aspects of interaction between a user and the system by different communication modalities.* — Bibl. 19 items.

1. Введение

Для решения проблемы эффективного человеко-машинного взаимодействия с недавнего времени исследуются модели, использующие несколько каналов передачи информации, естественных и удобных для человека (речь, мимика лица и артикуляция губ, жесты рук, головы и тела, направление взгляда, и т.д.). В результате их развития к настоящему времени сформировалось отдельное научное направление — многомодальные интерфейсы. Первая многомодальная система была продемонстрирована профессором Р. Болтом в 1980 году в США [1]. Такие интерфейсы обеспечивают наиболее эргономичное и естественное для человека взаимодействие с различными автоматизированными средствами управления и коммуникации.

Информационные системы массового обслуживания, банкоматы и справочные автоматы все больше и больше используются в различных областях.

▼ Данное исследование проводится в рамках гранта РФФИ № 07-07-00073-а, проекта СПбНЦ РАН №2.118, а также гранта Президента России № МК-9351.2006.9.

Сейчас стандартные информационные киоски с сенсорными экранами распространены очень широко, они размещаются в местах, работающих круглосуточно, а также там, где возможно его использование вместо персонала. Они используются повсеместно на улицах, в метро и торговых центрах для расчетов по мобильной связи, телекоммуникационным информационным услугам, для печати цифровых фотографий; на вокзалах и аэропортах для выдачи справочной информации о расписании авиарейсов или поездов; в музеях для предоставления справочной информации и продажи билетов, и т.д. Обычно такие киоски размещаются в местах массового скопления людей — в вестибюлях, у лифтов, рядом со входом в магазины, где киоски приветствует покупателей и информируют их о товарах и услугах, новых акциях и распродажах.

Однако к настоящему времени назрела острая необходимость в удобных для пользователя, эргономичных и интуитивных способах взаимодействия. Современные компьютерные системы должны обеспечивать пользователю возможность интуитивного управления, их функции должны быть простыми и очевидными настолько, чтобы человек мог пользоваться киосками без знания специальных команд или принципа работы, кроме того, система должна быть робастна к ошибкам в действиях пользователя.

Поэтому последнее десятилетие за рубежом активно проводятся исследования и разработки принципиально нового поколения киосков — многомодальных интеллектуальных киосков («умные» киоски). Это информационные автоматы массового обслуживания, которые могут автоматически определять присутствие пользователя-клиента и общаться с ним естественным образом [2]. Ввод информации может осуществляться как путем нажатия кнопок клавиатуры или сенсорного экрана, так голосом или даже жестами. Такая система, как правило, обладает знанием о своем пространственном положении и планировке здания, в котором расположен киоск, и использует эти данные при указании пользователю интересующего направления.

В ходе исследования для изучения механизмов естественного взаимодействия человека с киоском сконструирован испытательный стенд или прототип многомодального информационного киоска, обладающий дружественным интерфейсом, распознающий присутствие клиента и вербально взаимодействующий с ним на естественном языке. Общая архитектура программно-аппаратной части информационного киоска представлена на рис. 1.

Стенд включает в себя совокупность различных аппаратных и программных технологий, работающих синхронизовано. Основными из них являются: (1) видеообработка с технологией компьютерного зрения — для распознавания положения тела, лица и отдельных органов лица пользователя; (2) дикторонезависимая система распознавания русской речи, использующая массив микрофонов для подавления шумов и локализации источника полезного сигнала при дистанционной записи речи пользователя; (3) автоматическая система аудио-визуального синтеза речи применяемая для виртуального помощника — аватара; (4) современный графический пользовательский интерфейс (GUI) на базе сенсорного монитора; (5) менеджер диалога и диалоговая модель, включающие информационные базы данных и систему управления стратегиями диалога; (6) технологии современных сенсорных панелей.

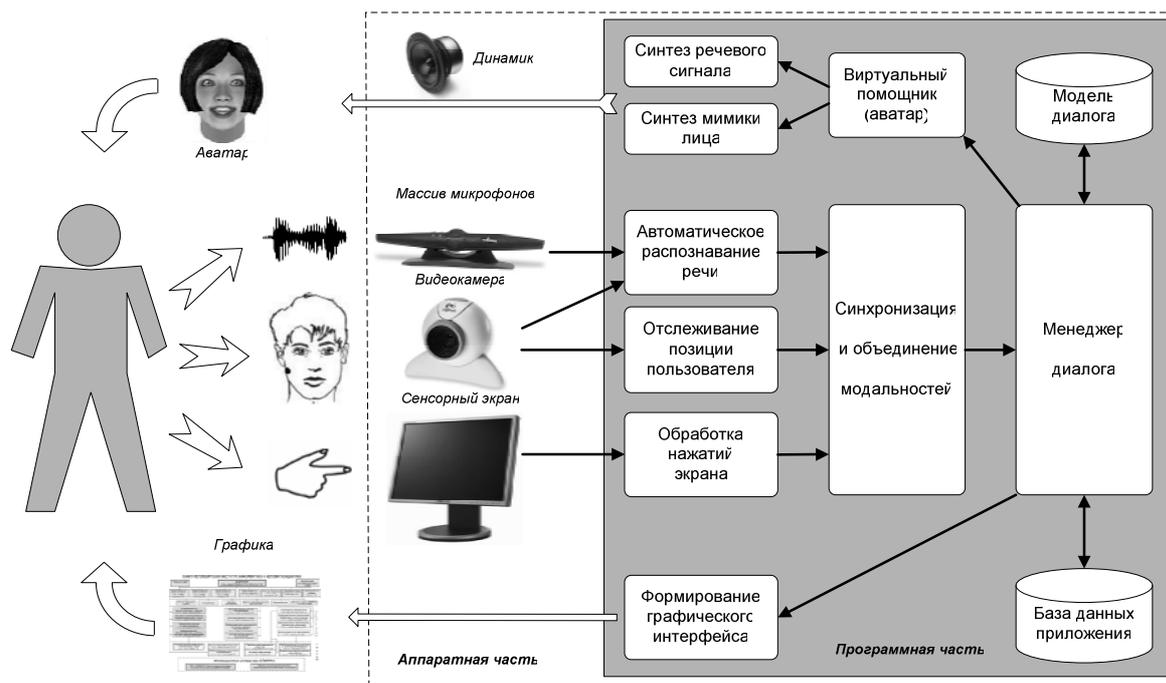


Рис. 1. Общая архитектура многомодального информационного киоска.

Ниже будут приведено описание основных компонентов модели интеллектуального киоска и освещены технические особенности их использования. Основное внимание будет уделено системе дистанционного распознавания речи, созданию аватара киоска, системе определения присутствия клиента, а также пилотному приложению для получения справочной информации об институте СПИИРАН.

2. Дистанционное распознавание русской речи

Речь является наиболее естественным для человека способом общения. Тем не менее, при разработке и внедрении речевого интерфейса для задач многомодальных информационных киосков возникает ряд трудностей, связанных с удобством и надежностью его использования. Голосовое управление является относительно новым и непривычным для пользователей, привыкших к графическому интерфейсу. Поэтому при разработке нужно учитывать то, что пользователи могут столкнуться с некоторыми трудностями в использовании речевого интерфейса, поскольку ожидают от него больших интеллектуальных способностей, практически подобных человеческим. Следует учитывать, что некоторые пользователи будут скорее пользоваться графическим интерфейсом, чем речевым, заранее сформировав для себя неправильное представление о его возможностях [3].

Существует ряд трудностей, связанных с внедрением системы распознавания речи, которую можно использовать в публичных местах. Например, выделение «чистого» аудиосигнала в зашумленных помещениях и обеспечение робастного распознавания слитной речи для всех категорий потенциальных пользователей. Для выполнения этих требований наиболее эффективна комбинация дикторонезависимой системы распознавания речи, массива микрофонов и технологии шумоподавления.

При разработке речевого интерфейса для интеллектуального киоска, должны учитываться акустические условия окружения, где будет эксплуатироваться киоск. Распознавание речи происходит значительно лучше, если пользователь киоска использует гарнитуру со встроенным узконаправленным микрофоном. К сожалению, применение гарнитур в условиях публичного пользования организационно невозможно, так как они могут быть украдены или повреждены вандалами. Тем более пользователи вряд ли захотят использовать гарнитуру во время коротких сеансов общения с киоском. Поэтому при разработке речевого интерфейса разработчики вынуждены ограничиваться использованием стационарных микрофонов, которые удалены ото рта пользователя и воспринимают слишком много шумов.

Современные дикторонезависимые системы распознавания речи работают надежно только при условии использования диктором ограниченного словарного запаса, при этом аудиосигнал должен быть «чистым». Чтобы клиент киоска использовал ограниченный словарный запас, можно либо использовать киоск в тех местах, где используемые клиентами слова и выражения заранее известны, либо необходимо давать клиенту подсказки. Однако вербальное подсказывание клиенту «правильных» слов или фраз несколько снижает скорость взаимодействия с киоском. Визуальные подсказки зачастую оказываются бесполезными, т.к. если клиент может увидеть слово или фразу, отображенную на экране, он быстро поймет, что, нажатием на кнопку, он добьется результата быстрее, нежели повторив свою речевую команду.

Существует ряд способов повышения разборчивости речи [4]. Естественно, если вывесить рядом с киоском или даже на нем самую табличку с полным перечнем этих способов, ее никто не будет читать. Однако если сформулировать их лаконично, на понятном языке, и если количество этих правил будет невелико, то вероятность того, что пользователь учтет их при вербальном общении с киоском, возрастает. Доступные способы повышения разборчивости речи можно свести к трем основным правилам:

1. Говорить громче.
2. Использовать короткие фразы.
3. Использовать длинные слова.

Модуль понимания естественного языка выполняет важную задачу — проводит анализ высказываний пользователя с тем, чтобы определить какое действие должно быть выполнено. Когда система эксплуатируется на ограниченной территории (как в случае с интеллектуальным киоском), существует не так много вариантов, что система должна сказать или сделать, проблема заключается в точной классификации высказываний пользователя, а вернее в абсолютном понимании каждого слова.

Достаточно хорошие результаты при дистанционном распознавании речи (на расстоянии 0,5–1,0 м от клиента) позволяет получить направленный линейный массив микрофонов с системой шумоподавления [5].

Использование массива из нескольких микрофонов позволяет использовать преимущества бинаурального слуха. Благодаря бинауральному слуху человек может определять направления на источники звуков, не поворачивая головы и не видя их. Наличие даже двух слуховых приемников в массиве позволяет слышать значительно более тихие звуки, что имеет существенное значение для оценки окружающего звукового пространства. Кроме того, наличие двух слуховых приемников позволяет услышать более тонкое различие звуков по высоте и по громкости, что имеет принципиально важное значение при сегрега-

ции звуков [6]. При увеличении количества микрофонов в массиве вышеописанные преимущества проявляются еще сильнее.

Кроме непосредственной функции приема звуковых сигналов, массив микрофонов можно использовать для пространственной локализации источника звука. Благодаря этому можно воспринимать звуки, исходящие из точки пространства, соответствующей рту пользователя, и отсекал звуки, приходящие со всех остальных направлений. Это позволяет существенно повысить качество дистанционного распознавания речи. Примеры готовых решений массивов микрофонов приведены на рис. 2.



Рис. 2. Примеры промышленных массивов микрофонов.

Однако, использование некоторого готового промышленного решения в данном случае не оправдано, так как не позволяет производить исследования и изменения его конструкции, в частности необходимо провести значительные исследования: сколько микрофонов использовать в массиве и на каком расстоянии друг от друга они должны располагаться в конструкции киоска для наиболее эффективной борьбы с акустическими шумами. Поэтому в ходе работ был разработан собственный массив из микрофонов, включающий набор конденсаторных микрофонов производства тульского завода «Октава» (рис. 3).



Рис. 3. Конденсаторный микрофон Октава МК-012, входящий в состав массива микрофонов.

Октава МК-012 это малогабаритный конденсаторный профессиональный микрофон, позволяющий использовать несколько сменных капсюлей с кардиоидной, гиперкардиоидной и круговой (всенаправленной) характеристиками диаграммы направленности. Низкий уровень шума и отсутствие трансформатора в схеме предусилителя данного микрофона обеспечивает высокую эффективность, как при близком, так и при удаленном расположении источника от микрофона, что необходимо при использовании в многомодальном информационном киоске. Широкая и равномерная частотная характеристика микрофона в диапазоне 20–20000 Гц гарантирует охват всех частотных диапазонов речи с высокой степенью точности.

Учитывая, что положение человека относительно киоска в достаточной мере предопределено, т.е. он находится прямо перед киоском, соответственно массив микрофонов должен воспринимать только аудиосигнал поступающий к

нему фронтально и подавлять сигналы, поступающие сзади и сбоку киоска. Таким образом, целесообразно выбрать микрофоны с кардиоидной характеристикой диаграммы направленности, а не с круговой характеристикой, которая больше подходит для задач, где пользователь имеет возможность перемещаться в пространстве, взаимодействуя с компьютерной системой.

Многие исследования показывают, что точность локализации растет с увеличением количества микрофонов в массиве. В большинстве случаев удовлетворительных результатов можно добиться при использовании массива из четырех-восьми микрофонов, поэтому в киоске предусмотрена установка до 4 стереопар конденсаторных микрофонов. Для подключения массива микрофонов к персональному компьютеру и синхронизации аудиопотоков используется профессиональная звуковая плата PreSonus Firepod (рис. 4), имеющая 4 стереоканала со встроенными регулируемыми усилителями сигнала и фантомным питанием 48 вольт, к которым может подключаться до 8 моноканальных микрофонов, использующих XLR разъемы.

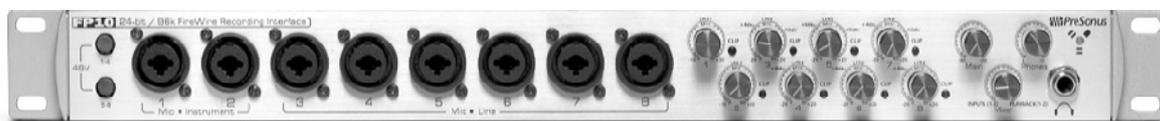


Рис. 4. Многоканальная звуковая плата PreSonus Firepod.

Точность локализации источника полезного звука (речи клиента) и качество принимаемого аудиосигнала в сильной степени зависят от пространственного расположения микрофонов в массиве и самого массива в корпусе киоска. При использовании человеком автомата, оснащенного монитором, он стремится занять такое положение, чтобы его лицо оказалось прямо напротив монитора. Одновременно с этим, он будет стараться смотреть в монитор таким образом, чтобы его лицо находилось в плоскости, параллельной плоскости экрана, а взгляд падал в центр монитора под углом максимально близким к прямому. Среднестатистическая высота уровня глаз человека — около 1570 мм, расстояние от глаз до рта — примерно 75 мм [7]. Таким образом, среднестатистическая высота, на которой находится рот человека — 1495 мм. Если монитор расположен так, что его геометрический центр находится на высоте 1,4–1,5 метра от пола, им будет удобно пользоваться большинству клиентов киоска. Логично было бы расположить линейный массив микрофонов на 75 мм ниже центра монитора прямо напротив рта, но это невозможно, т.к. тогда он будет перекрывать собой часть экрана. Следовательно, массив должен располагаться в нижней части монитора. Такое расположение обеспечит уверенный прием чистого аудиосигнала еще и потому, что фоновый шум будет гаситься телом пользователя. Необходимо также учитывать, что пользователь многомодального киоска получает устную информацию от виртуального помощника — аватара, почти не пользуясь графическим интерфейсом, и свое внимание сосредотачивает на голове аватара, который располагается в левой части экрана. Следовательно, оптимальным положением линейного массива микрофонов является положение непосредственно под монитором и возможно чуть сдвинутое в сторону виртуального помощника. Правая часть экрана при этом отводится для отображения «полезной» информации прикладной задачи, что наиболее удобно для праворуких пользователей, которые пользуются ведущей правой рукой для нажатий на объекты, отображаемые сенсорным монитором.

В следующем разделе рассмотрим основные методы цифровой обработке сигналов, используемые при дистанционной записи речи с помощью массива микрофонов.

3. Методы выделения и записи полезного речевого сигнала в пространстве

Проблема локализации объектов в пространстве тщательно изучается в области радиолокации уже более века. С развитием систем мобильной связи стали широко применяться технологии цифрового формирования луча (формирования диаграммы направленности) [8]. Эта технология позволяет динамически оптимизировать обслуживаемую зону покрытия, оперативно перенацеливая цифровые приемопередающие лучи в зависимости от территориального распределения абонентов. Созвездие лучей, синтезируемое по алгоритмам быстрого преобразования Фурье или посредством классических процедур дискретного Фурье-анализа, является, по сути, совокупностью пространственно частотных фильтров, каждый из которых селектирует строго определенный набор сигналов и подавляет остальные, воспринимаемые как помеховые. В результате существенно улучшается качество связи, а также резко повышается помехозащищенность системы.

При дистанционном распознавании речи мы имеем дело не с искусственно сгенерированными сигналами, а с естественными, поэтому многие методы, успешно использующиеся в радиолокации, требуют серьезной адаптации для их применения в задаче звуковой локализации пользователя в пространстве и дистанционной записи речи. Далее рассмотрим ряд методов, которые наиболее широко применяются для записи речи посредством массива микрофонов.

Существует технология, позволяющая объединять сигналы с различных микрофонов с целью акцентирования сигнала, получаемого от желаемого источника звука, и подавления помех, попадающих в область восприятия микрофонов. Данная технология носит название «технология лучеобразования» (beamforming) [9,10].

Технология лучеобразования основана на том, что целевой источник звука и источники помех расположены в различных точках пространства. Это несопадение источников звуков в пространстве можно использовать, например, для того, чтобы подавлять все сигналы, приходящие из точек пространства, не соответствующих точке предположительного нахождения целевого источника звука. Преимуществом данной технологии является то, что распознаванию подвергается сигнал только одного канала, что существенно уменьшает время распознавания. Также следует учесть, что пространственное расположение микрофонов известно заранее.

Для предварительного определения местоположения целевого источника звука используется алгоритм определения направления прибытия (DOA-расчет). При этом расчет ведется по наиболее сильному поступившему сигналу — определяется временная задержка между поступлением этого сигнала на входы микрофонов в массиве, т.е. используется эффект бинаурального слуха.

При использовании MUSIC-алгоритма строится псевдо-спектральная функция, пики которой соответствуют углам прибытия сигналов [11]. ESPRIT-алгоритм — один из наиболее эффективных и надежных методов DOA-расчета. При этом используются пары микрофонов, причем вторые микрофоны в каждой паре сдвинуты относительно первых в одинаковом направлении и на одинаково-

вое расстояние. Этого можно добиться и без использования двух массивов. Достаточно иметь возможность выбирать некоторые подмножества в рамках одного массива.

Алгоритм ESPRIT работает быстрее и дает более точные результаты, по сравнению с алгоритмом MUSIC, однако он не может быть применен в случае коррелированных источников сигналов.

После выполнения DOA-расчета, возможно применение одного из алгоритмов лучеобразования, простейшие из которых основаны на сложении сигналов в соответствии с законами геометрии без учета уровня принимаемых сигналов. Более сложными являются адаптивные алгоритмы лучеобразования, которые стараются минимизировать совокупную выходную мощность принимающих помехи и шумы микрофонов таким образом, чтобы сигнал от желаемого источника звука был максимально усилен. Наиболее широко используются следующие адаптивные алгоритмы лучеобразования [12,13]:

1. Метод наименьшего среднеквадратичного (LMS — least mean squares algorithm).
2. Метод обращения матрицы (SMI — sample matrix inversion algorithm).
3. Рекурсивный метод наименьших квадратов (RLS — recursive least square algorithm).
4. Метод сопряженных градиентов (CGM — conjugate gradient method).
5. Метод постоянных коэффициентов (CMA — constant modulus algorithm).
6. Подавление сигналов боковых лепестков (sidelobe cancellers).
7. Наименьшая дисперсия с линейными ограничениями (LCMV — linearly constrained minimum variance).
8. Лучеобразование с нулевым заполнением (null steering beamforming).

Далее кратко рассмотрим каждый из этих алгоритмов адаптивного лучеобразования.

Метод наименьшего среднеквадратичного (LMS-алгоритм) применяется в адаптивных фильтрах для нахождения коэффициентов фильтрации, которые используются при вычислении наименьшего среднеквадратичного значения сигнала ошибки (разницы между ожидаемым и фактическим сигналами). LMS-алгоритм является наиболее часто используемым адаптивным алгоритмом. Причиной этому является его простота. Он работает надежно и показывает стабильные результаты при различных состояниях сигнала. Однако скорость его схождения относительно невелика (в сравнении с RLS-алгоритмом). Как правило, источники шумов и источник полезного сигнала не статичны, они меняют свое положение в пространстве. Соответственно должны изменяться и весовые коэффициенты, а при низкой скорости сходимости метода это может привести к значительным погрешностям при локализации.

Метод обращения матрицы (SMI-алгоритм), в отличие от LMS-алгоритма, дает хорошие результаты, когда локализуемый источник звук испускает прерывистые звуковые волны. Однако, для успешной работы фильтра, построенного с использованием SMI-алгоритма, необходимо, чтобы количество источников помех и их пространственное расположение оставались неизменными на протяжении всего времени локализации. Использование такого фильтра может дать отличные результаты, при использовании его в системах локализации источника звука, интегрированных в системы распознавания речи, которые находятся в закрытых помещениях — жилая комната, помещение с банкоматом и т.д. Однако применение этой технологии в автоматах, расположенных в общественных местах практически невозможно из-за того, что условия применения

характеризуются постоянным изменением пространственных положений источников помех и их количества.

Рекурсивный метод наименьших квадратов (RLS-алгоритм) применяется в адаптивных фильтрах для нахождения коэффициентов фильтрации, которые используются при вычислении наименьшего среднеквадратичного значения сигнала ошибки. Использование этого метода дает хорошие результаты в случае прерывистой передачи сигнала источником. Идея RLS-алгоритма заключается в минимизации взвешенных наименьших квадратов функции ошибки. В терминологии теории адаптивных фильтров, взвешенные наименьшие квадраты функции ошибки — это весовая функция. Преимуществом данного метода является то, что не приходится инвертировать большие матрицы, таким образом, сохраняются вычислительные мощности. Однако у метода есть существенный недостаток — количество источников помех и их местоположение должно оставаться постоянным.

Метод сопряженных градиентов (CGM) — это алгоритм численного решения частного случая системы линейных уравнений, а именно симметричных положительно определенных матриц. CGM — итерационный метод, поэтому он может быть использован в тех немногочисленных системах уравнений, которые слишком большие, чтобы использовать в них прямые методы. Такие системы уравнений обычно возникают при численном решении дифференциальных уравнений в частных производных.

Метод постоянных коэффициентов («слепое» адаптивное лучеобразование) (СМА) подобен SMI-алгоритму, отличается тем, что не требует эталонного сигнала. Звуковой сигнал имеет некоторые постоянные коэффициенты, которые могут быть потеряны вследствие влияния шумов и помех. СМА позволяет построить фильтр, восстанавливающий эти коэффициенты. Главным преимуществом метода является то, что он прост в реализации. Но при этом, метод имеет существенный недостаток — низкое быстродействие. Это сводит на нет возможность его практического применения для построения фильтров, работающих в системах реального времени.

При подавление сигналов боковых лепестков (Sidelobe cancellers) простейший лучеобразователь содержит одну главную антенну и некоторое количество вспомогательных. Главная антенна направлена точно на источник полезного сигнала, однако, кроме полезного сигнала, она улавливает некоторые помехи и шумы через боковые лепестки своей диаграммы направленности. Дополнительные антенны призваны компенсировать этот нежелательный эффект. Они нечувствительны к основному сигналу и улавливают только сигналы помехи. Весовые коэффициенты вспомогательных антенн подобраны таким образом, чтобы их суммарный сигнал компенсировал сигналы, принимаемые боковыми лепестками главной антенны (сумма сигналов боковых антенн, помноженных на соответствующие весовые коэффициенты, вычитается из сигнала главной антенны). Эта технология проста в применении, но эффективно может быть применена только в случае, когда целевой сигнал слабее сигналов помех. Чем сильнее становится целевой сигнал, тем больше становится его вклад в полную выходную мощность, что, в свою очередь, приводит к увеличению процента подавляемых сигналов. Это может привести даже к подавлению самого целевого сигнала.

Большинство из существующих методов лучеобразования требует знания интенсивности целевого сигнала, также им необходим некоторый эталонный сигнал. Это ограничение может быть преодолено путем использования линей-

ного ограничения вектора весовых коэффициентов LCMV пространственные фильтры — это лучеобразователи, которые варьируют свои весовые коэффициенты так, чтобы минимизировать дисперсию выходной мощности фильтра в зависимости от начальных условий. Этот критерий, совместно с другими ограничениями, позволяет принимать неискаженный сигнал от целевого источника, в то время как дисперсионные сигналы из других точек пространства минимизируются. Этот метод довольно гибкий и не требует эталонного сигнала для вычисления оптимальных весовых коэффициентов, но требует вычисления ограниченного вектора весовых коэффициентов.

В отличие от других алгоритмов, алгоритмы с нулевым заполнением не повышают уровень сигнала, входящий от целевого источника, а минимизируют выходную мощность сигналов, входящих из других точек пространства. Один из методов, основанных на этом подходе, заключается в минимизации среднеквадратичного значения выходного сигнала массива при одновременном ограничении вектора весовых коэффициентов до единичного значения.

Данные алгоритмы будут экспериментально проверены в последующем исследовании и использованы при анализе пространственной акустической обстановки перед киоском, подавлении шума и дистанционном распознавании речи.

4. Компьютерное зрение для детекции пользователя

Способность системы определять положение тела и головы пользователя, находящегося перед киоском, позволяет улучшить качество человеко-машинного взаимодействия. При общении двух людей они выражают свои мысли не только через смысл произносимых фраз, но и взглядами, жестами, интонациями. Поэтому и киоск, способный определять присутствие клиента и умеющий выражать свою осведомленность в социально приемлемой манере, несомненно, будет приниматься клиентами лучше, чем традиционный киоск. Таким образом, многомодальный интеллектуальный киоск должен уметь определять присутствие перед ним клиента, отслеживать его перемещения в своем поле зрения и демонстрировать это, используя визуальную обратную связь. Такую функцию выполняет аватар, который следит глазами за перемещением пользователя и тем самым показывает свою осведомленность о присутствии клиента.

Чтобы создать аватар, который вел бы себя реалистично, киоск должен уметь быстро определять и точно отслеживать изменение местоположения человека в своем поле зрения. Для плавного перемещения аватара скорость сопровождения должна быть не менее 10 Гц при точности 5–10 градусов [4].

Для решения данной задачи в киоске можно использовать различные сенсоры, например, звуковые или инфракрасные. Принцип работы активных инфракрасных сенсоров заключается в освещении пространства инфракрасным светодиодом и замере силы отраженного инфракрасного света. Однако, для такого сенсора человек, одетый в белый плащ и находящийся на расстоянии 2 м от сенсора, может быть неотличим от человека, одетого в черную футболку и находящегося на расстоянии 0,5 м. Звуковые сенсоры могут определить расстояние до клиента с точностью до нескольких сантиметров, но работают они довольно медленно, так как применяемые схемы звуковых сенсоров накладывают ограничение на используемую частоту. В них используются частоты до 5 кГц. Из-за этого система обнаружения и отслеживания клиента работает с

опозданием и у пользователей может создаться впечатление, что система работает инертно.

Наиболее перспективной технологией для обнаружения присутствия пользователя считается система компьютерного зрения. Потенциальные цели идентифицируются как большие области в текущем изображении, которые отличаются по цвету и яркости от заранее сохраненного фонового изображения. Цели различаются по форме и размеру, и отслеживаются только те цели, которые размером и формой напоминают человека. Цели отслеживаются во времени путем сравнения положений, ожидаемой скорости и цвета.

Так входящий в состав библиотеки Intel OpenCV алгоритм Хаара позволяет находить в видеопотоке положение человека полностью или отдельно его лицо (голову). Алгоритм определения лиц обеспечивает возможность не только обнаружить появление клиента, но и определить количество пользователей, общающихся с системой, выделив их из числа простых прохожих.

Помимо модуля детекции головы пользователя для задач управления многомодальным киоском разработан также модуль отслеживания движений головы клиента [14]. Этот модуль может использоваться для бесконтактного управления курсором мыши на экране посредством движений головы. Такой вид интерфейса является необходимым для пользователей-инвалидов без рук, либо с проблемами двигательных функций рук, которые не могут использовать контактные устройства ввода информации в компьютер. Программный модуль для отслеживания движений головы пользователя реализован на основе метода Лукас-Канаде для оптического потока. Для обработки видеосигнала в реальном масштабе времени был разработан метод, включающий в себя 2 режима работы: калибровка (или настройка) и отслеживание. При калибровке системы производится привязка координат курсора к положению головы на изображении и когда пользователь двигает головой (поворачивает влево, вправо, наклоняет вниз или поднимает) курсор мыши синхронно двигается в эту сторону, что позволяет использовать такой подход для бесконтактного управления компьютером. В ходе процесса калибровки в окне отображается изображение, поступающее с видеокамеры. Учитывая стандартные пропорции лица, положение отслеживаемых точек автоматически отмечается цветными точками на экране и в течение нескольких секунд настройки пользователь должен совместить реальное изображение своего лица с этими точками. По истечении времени калибровки (несколько секунд) курсор мыши выставляется по центру рабочего стола и «привязывается» к положению точки кончика носа. В системе для робастного отслеживания перемещений головы оператора используется система из 5 естественных точек на лице: центр верхней губы, кончик носа, точка между глаз, зрачок правого глаза и зрачок левого глаза (рис. 5).

Применение указанного подхода для бесконтактного управления курсором позволяет работать с графическим пользовательским интерфейсом и выделять объекты на экране даже небольшого размера лишь посредством движений головы, которые эффективно отслеживаются миниатюрной вэб-камерой, встроенной в корпус интеллектуального киоска.



Рис. 5. Пример поиска и отслеживания положения лица пользователя.

Основные факторы, которые затрудняют внедрение компьютерного зрения в киоски это большие группы людей (толпа), которые часто возникают рядом с местоположением киоска, так как киоски обычно ставятся в многолюдных местах. Когда большое количество людей скапливается рядом с киоском — это затрудняет поиск лица клиента, который непосредственно взаимодействует с системой. При этом система реагирует поворотом лица аватара в центр толпы, иногда оглядывая ее целиком.

Предположение, что толпа — это люди, стоящие напротив, вытекает в так называемую «проблему стада карибу» [4]. Она возникает, когда все пространство перед киоском заполнено людьми, двигающимися в одном направлении. Т.к. киоск не может различить индивидуальности и, следовательно, отслеживать их движения, двигающаяся толпа кажется ему неподвижной.

Киоск должен быть правильно настроен и адаптирован для работы в определенном визуальном окружении. Визуальное окружение меняется по мере того, как люди и объекты перемещаются на заднем плане, как меняется освещение в течение дня. В идеале киоск должен автоматически отслеживать изменение визуального пространства в течении нескольких дней и впоследствии использовать эту информацию для оптимизации своей зрительной системы.

Еще одна проблема возникает, когда фон пола перед киоском содержит динамические элементы. Например, киоск отбрасывает перед собой длинную темную тень от солнца, свет автомобильных фар, прошедший через цветное оконное стекло дает многочисленные блики на полу, солнце отражается от кафельного пола, динамические рекламные изображения, спроецированные на пол (в частности, медленно вращающаяся проекция логотипа компании красного цвета на белом полу). Если эти двумерные динамические элементы, расположенные на полу, выстраиваются в линию с объектом, двигающимся на расстоянии, они могут показаться киоску человеком, стоящим напротив. Система технического зрения должна уметь реагировать на внезапные и необычные изменения в визуальном окружении. Вместо использования камеры для обнаружения присутствия клиента, киоск может быть оснащен большой и хорошо заметной клавишей «Вход/Выход».

5. Многомодальный вывод информации

Одним из основных компонент подсистемы вывода информации в многомодальном киоске, помимо графического пользовательского интерфейса, является виртуальный анимированный помощник или аватар. Аватар — это трехмерная модель лица человека с двигающимися глазами, ртом и лицевыми мус-

кулами. Модель аватара может разговаривать, синхронизируя движения рта, губ и зубов с синтезированным голосом или заранее сделанной аудиозаписью. Синхронизация движения губ с синтезированной или записанной речью создает иллюзию «живой говорящей головы». Модель аватара имеет подвижные глаза, веки и лицевые мускулы, что позволяет ему выразительно отображать широкий спектр эмоций. Рис. 6 показывает примеры реалистичных аватаров, некоторые из которых доступны для свободного использования [15].

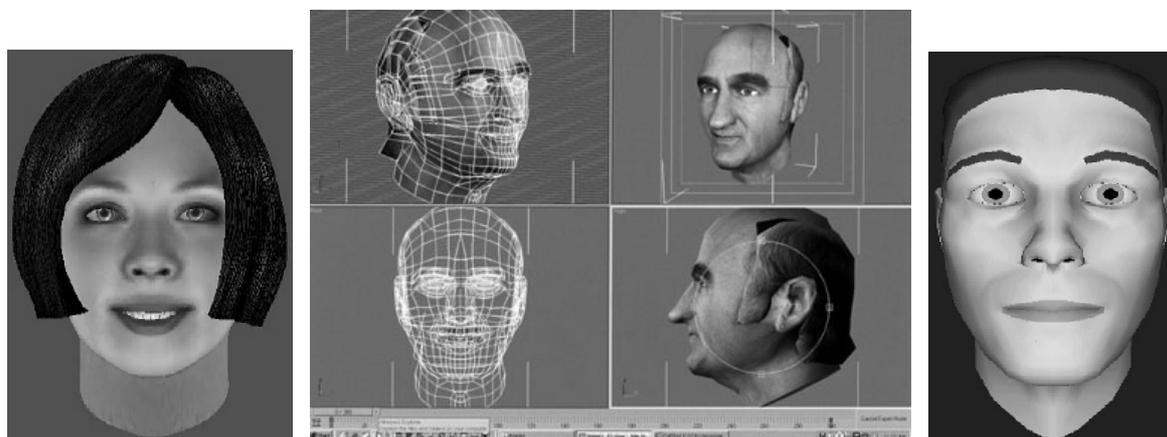


Рис. 6. Примеры реалистичных компьютерных аватаров.

Аватар многомодального информационного киоска выполняет две функции. Во-первых, используя информацию о перемещениях клиента, полученную от системы компьютерного зрения, он способен симулировать проявление внимания к клиенту. Киоск отслеживает положение и перемещение потенциальных пользователей, а аватар поворачивается и наблюдает за приближающимися клиентами. Когда клиент подходит достаточно близко и проявляет интерес к киоску, аватар произносит приветствие. Используя данные о перемещениях клиента и динамическую модель движения головы, аватар поворачивается таким образом, чтобы все время быть направленным на клиента, при этом глаза аватара фокусируются на лице человека. По мере того, как клиент осуществляет поиск нужной информации, аватар помогает ему, давая вербальные указания по поводу навигации, отвечая на вопросы и обеспечивая клиента другой полезной информацией. Аватар привлекает людей к киоску, побуждает их пользоваться им и, возможно, даже развлекает их. Внимание проходящих мимо людей привлекается, когда аватар поворачивает голову в их сторону и наблюдает за ними.

Для создания анимации аудио-визуального синтеза речи («говорящей головы») существует два основных подхода: параметрический подход, при котором создаётся 2D или 3D модель лица (например, используя возможности формата MPEG-4) и настраиваются управляющие параметры для передачи мимики, выражения лица и движения губ при говорении [16] и компиляционный подход, при котором «говорящая голова» формируется путём выбора соответствующих видеофрагментов из визуальной базы данных [17]. При использовании первого подхода физический объём данных, необходимых для синтеза визуальной речи, гораздо меньший, чем во втором случае. Кроме того, первый подход не требует создания и разметки корпуса, содержащего все визуальные единицы речи — виземы (изображения лица и формы губ при произнесении различных фонем), что является достаточно трудоёмкой работой. Тем не ме-

нее, первый из подходов подход в сравнении со вторым имеет определенный недостаток, связанный с неизбежной схематичностью отображения речедвижений, что не позволяет использовать его для персонализированного аудио-визуального синтеза речи по тексту. К недостаткам первого подхода относятся и большая вычислительная сложность его реализации, а также недостаточно реалистичные результаты при персонализации «говорящей головы». В связи с чем, компиляционный подход представляется более предпочтительным при персонализированном аудио-визуальном синтезе речи по тексту.

Помимо анимации лица виртуального персонажа, основным компонентом аватара является модуль синтеза русской речи. К настоящему времени системы синтеза речи достигли определённого совершенства и уже используются в ряде практических приложений. В рамках разработки модуля синтеза русской речи для многомодального киоска группа речевой информатики СПИИРАН активно сотрудничает с лабораторией распознавания и синтеза речи Объединенного института проблем информатики Национальной академии наук Беларуси. В ходе совместных работ по исследовательскому проекту INTAS, проводимому в 2005–2007 гг. была разработана система синтеза русской речи по произвольному тексту и создан автоматический телефонный сервис заказа авиабилетов «Полетели».

6. Пилотное информационное приложение киоска

Для тестирования отладки аппаратно-программного обеспечения прототипа многомодального киоска выбрана прикладная задача, содержащая относительно небольшой объем информации. Прототип создаваемого интеллектуального автомата содержит в базе данных справочную информацию о структуре и сотрудниках СПИИРАН, контактную информацию о подразделениях института, информацию о текущих событиях, а также навигационную информацию, которая может быть полезна как посетителям, так и сотрудникам института. Несомненно, актуальность и качество предоставляемой информации, наполненность контента и простота использования являются критически важными параметрами для интеллектуального киоска.

В 2007 году был разработан лабораторный прототип системы с ограниченной функциональностью. Этот прототип помогает разработчикам эффективно отыскивать ошибки в программном обеспечении (обычно проявляющиеся в режиме демонстрации). После лабораторной отладки и увеличения функциональности прототипа киоска, он будет выставлен в холле здания института.

Для доступа к интерактивной диаграмме структуры СПИИРАН пользователь может использовать либо сенсорный экран киоска, либо голосовые команды. Посредством диаграммы посетитель института может узнать о составе лабораторий и групп института, а также получить справочную информацию о контактных телефонах руководителей и территориальном размещении подразделений института.

Подсистема менеджера диалога, отвечающая за функционирование киоска, управляет отображаемой на экране информацией, а также то, как киоск реагирует на пользовательские воздействия. Многомодальный киоск может функционировать в двух режимах: развлекательном (привлечение пользователей) и в режиме взаимодействия. Во время, когда у киоска нет клиентов, он функционирует в развлекательном режиме, демонстрируя презентацию об институте. Если человек проявляет заинтересованность в киоске, киоск переключается в

режим взаимодействия. В этом режиме аватар приветствует клиента, а на дисплее отображается стартовая страница в интерактивной диаграммой, после чего клиент может осуществлять поиск нужной информации. Наконец, когда клиент уходит, киоск прощается с ним и снова переключается в развлекательный режим. Новизна и непривычность интеллектуальных киосков привлекает людей, но не более, чем на минуту. Если за это время киоск не проявит себя как занимательное и одновременно информативное устройство, клиент уйдет.

Важным вопросом является когда и как начинать, заканчивать и поддерживать сессию человеко-машинного взаимодействия. Начало взаимодействия инициируется автоматически каждый раз, когда человек смотрит на экран, что определяется подсистемой компьютерного зрения, которая находит и отслеживает лица. Основываясь на данных о среднестатистическом размере лица человека, можно вычислить расстояние, на котором оно находится. Если лицо человека находится примерно в метре от киоска или ближе, это означает, что человек находится в зоне взаимодействия и является сигналом к началу новой сессии, если на данный момент еще нет открытых сессий. В том случае если сессия уже открыта, это служит сигналом о необходимости ее продолжения. Голосовые команды и касания монитора также служат сигналами к открытию или продолжению сессии.

С точки зрения клиента, типичный процесс взаимодействия с киоском выглядит следующим образом [2]. Клиент видит неподалеку киоск, демонстрирующий на своем экране слайды. Клиент начинает двигаться к киоску, и когда он приближается на расстояние около метра, киоск вербально и визуально приветствует его. Если клиент задерживается напротив киоска, изображение на дисплее меняется. Клиент нажимает на экране или общается с киоском посредством голосовых команд и получает необходимую информацию, как через графический интерфейс пользователя, так и посредством синтезированной речи аватара. В случае если клиент не проявляет никакой активности во взаимодействии в течение 30 секунд, киоск вежливо предлагает ему воспользоваться функциями сенсорного экрана или сделать голосовой запрос. Когда клиент уходит от киоска, система прощается с ним, и киоск переходит в режим поиска клиентов.

Связь между компонентами автоматической информационной системы и использование общей базы данных и знаний — это одни из ключевых вопросов в построении интеллектуального киоска. Использование общей базы данных как ядра схемы связи между компонентами, дает несколько важных преимуществ. Различные компоненты киоска имеют связь друг с другом через базу данных при помощи менеджера диалога. При этом каждый из них имеет доступ к таблице пользовательских воздействий и может соответственно на них реагировать. Каждый компонент разработан как автономная система, это позволяет запускать и тестировать каждый компонент в отдельности на любом компьютере.

Все запросы, адресованные киоску, а также его действия и сигналы с сенсоров (микрофонов и видеокамеры) записываются в системный журнал. Анализ записей в нем используется для определения наиболее популярных типов запросов, времени сеансов работы с киоском, количество людей, проходящих мимо киоска. Также можно определить, каким образом пользователи осуществляют навигацию по предоставляемому контенту, и какие при этом у них возникают проблемы.

7. Результаты экспериментов над прототипом киоска

Первая лабораторная версия интеллектуального киоска служит для оценки его эффективности и определения основных трудностей, таких как распознавание речи в шумных условиях, понимание естественной речи, ответ на неполные или неточно сформулированные вопросы. Чтобы определить общую производительность киоска, удовлетворенность пользователей от работы с ним, простоту и скорость использования, а также оценить результаты его работы во время тестовой эксплуатации, над прототипом киоска проводятся качественные и количественные эксперименты.

Одним из экспериментов, проведенных над прототипом интеллектуального киоска, является оценка скорости работы пользователей со средствами ввода информации (манипуляторами), предназначенными для указания на объекты графического пользовательского интерфейса. Для оценки устройств ввода информации была использована методология международного стандарта ISO 9241-9 [18], которая базируется на экспериментах и законах Фитта [19]. Данная методика состоит в следующем. Тестеры, используя предоставленное им устройство ввода, должны настолько возможно быстро выделить на экране 16 целей-объектов, расположенных по кругу, при этом порядок целей задается системой автоматически таким образом, чтобы пользователь последовательно выделял наиболее удаленно расположенные друг от друга объекты, совершая движения манипулятором в различных направлениях. Когда, нажатием на кнопку, происходит выделение текущего объекта на экране, то отображается следующая цель. При этом вычисляется индекс сложности ID (index of difficulty), измеряемый в битах, по формуле Шэннона

$$ID = \log_2 \left(\frac{D}{W} + 1 \right),$$

где D — расстояние между целями и W — диаметр круглой цели. Согласно закону Фитта, время движения MT (movement time) между двумя целями есть линейная функция индекса сложности ID . Основным показателем является общая производительность устройства TP (throughput), что есть компромисс между скоростью движения и точностью выделения и измеряется в бит/сек. согласно формуле

$$TP = ID / MT.$$

В ходе эксперимента были оценены средства ввода информации, встроенные в многомодальный киоск, такие как сенсорный экран 17" и управление курсором движениями головы и их результаты сравнены с другими стандартными аппаратными устройствами ввода информации: джойстик, трекбол, сенсорная панель 3" и манипулятор типа «мышь». Двумя тестерами были произведены по 10 экспериментов для каждого устройства с последовательным изменением диаметра W цели в пределах от 32 до 128 пикселей и среднего расстояния D между целями в пределах 96–650 пикселей. Табл. 1 приводит результаты сравнения вышеуказанных органов управления по трем критериям:

- среднее времени движения MT между двумя целями;
- процент ошибок выделения целей (непопадание курсором в цель);
- общая производительность устройства TP .

Табл. 1 показывает, что наилучшие результаты по производительности устройств были показаны сенсорным монитором, что оправдывает его использование как основного устройства для манипулирования графическими объек-

тами на экране без применения дополнительно аппаратных способов ввода координат. Управление курсором мыши посредством системы компьютерного зрения, отслеживающей движения головы, уступает по производительности практически аппаратным средствам ввода кроме джойстика, однако имеет то преимущество, что является бесконтактным способом управления курсором и может использоваться категориями потенциальных пользователей, для которых стандартные средства ввода недоступны.

Таблица 1

Сравнение способов ввода информации посредством эксперимента Фита

Устройство ввода	<i>MT</i> , сек	Ошибка выделения, %	<i>TP</i> , бит/сек.
Джойстик	2.01	7.00	1.54
Трекбол	1.03	3.83	3.51
Сенсорная панель	0.85	4.50	3.72
Манипулятор-мышь	0.49	3.17	6.65
Сенсорный экран	0.50	6.17	7.85
Отслеживание движений головы	1.98	7.33	1.59

Представленные результаты тестирования являются только первыми экспериментами по применению многомодального информационного киоска и они не отражают пока всех способностей и возможностей автоматической системы массового обслуживания. Исследования будут продолжены далее и будут развиваться от лабораторных исследований в сторону реального массового использования интеллектуального киоска. Ясно, что хорошо работает в лабораторных условиях, может давать сбои при реальной эксплуатации из-за возникающих непредвиденных трудностей. Чтобы точно оценить успешность или неработоспособность разработанных методов, в 2008 году киоск будет размещен в холле здания института и апробирован в реальных условиях эксплуатации. В конце работ, улучшенный прототип киоска планируется выставить для всеобщего пользования, например, в музеях или вокзалах для массового обслуживания клиентов.

8. Заключение

Объединение удобной для пользователя вычислительной техники с речевыми технологиями, использование виртуальных помощников, позволяет создавать эффективный и естественный интерфейс, в котором в отношениях человек-машина на первое место ставится человек. Многомодальность обеспечивает возможность использования таких систем людьми с ограниченными способностями. Также она позволяет преодолевать психологический барьер, зачастую возникающий при общении человека с компьютером, особенно если речь идет о пожилых людях.

В рамках исследования ведется разработка исследовательского стенда многомодальной системы управления компьютером и модели информационно-справочной службы, на базе которой будут проанализированы когнитивные проблемы взаимодействия пользователя с машиной, учитывая различные способы коммуникации. Накопленный в ходе тестирования модели материал будет

использован для изучения когнитивных и поведенческих характеристик пользователя и оптимизации разработанного многомодального интерфейса.

На данный момент все необходимое оборудование для регистрации и вывода сигналов различных естественных модальностей подключено к компьютеру, на котором установлено разработанное программное обеспечение многомодального интерфейса, а также пилотное приложение для получения справочной информации об институте СПИИРАН. В ходе экспериментов проверена принципиальная совместимость и работоспособность как оборудования, так и программного обеспечения. В последующем будут проведены опытные эксперименты по выбору местоположения видеокамеры, массива микрофонов, динамиков и сенсорного экрана для создания эргономичных условий взаимодействия, максимального видеозахвата территории перед киоском, дистанционной записи речи пользователя и минимизации влияния работы динамиков на микрофоны.

Разработанный стенд в дальнейшем послужит прототипом для серийного производства многомодальных информационных киосков, на базе которых возможно создание широкого спектра справочно-информационных систем городских служб расположенных в бизнес-центрах, отелях, аэропортах, выставочных комплексах, ВУЗах, медицинских центрах, торговых центрах, театрах, музеях, спортивных клубах. С их помощью можно получить полный спектр городской справочной информации: оптимальные маршруты передвижения, расписание общественного транспорта, расписание полетов, оперативную информацию о задержках рейсов, наличии и стоимости билетов, телефонный справочник, карту, публикации административных распоряжений, статей законодательства, справочную информацию о пунктах назначения, расписание занятий и экзаменов для преподавателей и студентов, медицинские советы и рекомендации, часы работы врачей и кабинетов, новости и объявления, виртуальные экскурсии по музеям и другую информацию.

Литература

1. Bolt R. A. Put-that-there: voice and gesture at the graphics interface // Computer Graphics. 1980. Vol. 14, no. 3. P. 262–270.
2. Andrew D., Avery C. L., Avery B. L. Digital smart kiosk project // Proc. of the SIGCHI conference on human factors in computing systems. 1998. P. 155–162.
3. McCauley L., D'Mello S. MIKI: a speech enabled intelligent kiosk. // Intelligent virtual agents. Lecture Notes in computer science. 2006. Vol. 4133. P. 132–144.
4. Andrew D., Avery C. L., Avery B. L. Speak out and annoy someone: experiences with intelligent kiosks // Proc. of the SIGCHI conference on human factors in computing systems. 2000. P. 313–320.
5. Kleek M. V., Kottahachchi B., Horton T., Cavallaro P., Shrobe H. Designing speech interfaces for public kiosks // Student oxygen workshop. 2004.
6. Алдошина И. Основы психоакустики. Бинауральный слух // Звукорежиссер. 1999. Т.10.
7. ГОСТ Р ИСО 9241-3-2003 Эргономические требования при выполнении офисных работ с использованием видеодисплейных терминалов (ВДТ). Часть 3. Требования к визуальному отображению информации.
8. Слюсар В. Smart-антенны пошли в серию // Электроника: наука, технология, бизнес. 2004. Т. 4. С. 62–65.
9. Veen B. D. V., Bukley K. M. Beamforming: a versatile approach to spatial filtering // IEEE ASSP Magazine. 1988. P. 4–24.
10. Herboldt W., Kellermann W. Adaptive beamforming for audio signal acquisition // Adaptive signal processing: application to real-world problems. Springer, Berlin, 2003. P. 155–194.

11. *Серебряков Г. В.* Эффективность коммуникационных адаптивных антенных решеток с различной геометрией // Актуальные проблемы статистической радиофизики. 2002. Т. 1, № 1. С. 95–101.
12. *Hoshuyama O., Sugiyama A., Hirano A.* A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters // IEEE Transactions on signal processing. 1999. Vol. 47, no. 19. P. 2677–2684.
13. *Zhang M.* Adaptive beamforming by microphone arrays // Proc. of global telecommunications conference GLOBECOM '95. 1995. Vol. 1. P. 163–167.
14. *Карпов А. А.* ICanDo: интеллектуальный помощник для пользователей с ограниченными физическими возможностями // Вестник компьютерных и информационных технологий. 2007. №7. ISSN 1810-7206. С. 32–41.
15. MPEG4 talking heads, <<http://www.pd.istc.cnr.it/LUCIA/home/default.htm>>.
16. *Tekalp A. M., Ostermann J.* Face and 2-D mesh animation in MPEG-4. Signal processing: image communication // Special issue on MPEG-4. 2000. Vol. 15. P. 387–421.
17. *Cosatto E., Graf H. P.* Photo-realistic talking-heads from image samples // IEEE Transactions on multimedia. 2000. Vol. 2. P. 152–163.
18. ISO 9241-9:2000(E) Ergonomic requirements for office work with visual display terminals (VDTs). Part 9: requirements for non-keyboard input devices. International standards organization. 2000.
19. *Soukoreff R. W, MacKenzie I. S.* Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI // International journal of human-computer studies. 2004. Vol. 61. Issue 6. P. 751–789.