

УДК 004.5

Ал.Л. Ронжин, Ан.Л. Ронжин

## Система аудиовизуального мониторинга участников совещания в интеллектуальном зале

Рассмотрен ряд аспектов, возникающих при разработке автоматизированных систем видеосъемки мероприятий в интеллектуальном зале. Предложенная система мониторинга участников совещания применяется для записи текущей ситуации в зале и позволяет автоматизировать управление аудио-, видео- и презентационным оборудованием.

**Ключевые слова:** аудиовизуальная обработка, окружающее интеллектуальное пространство, контекст, информационно-управляющие системы.

Распознавание текущей ситуации, анализ поведения пользователя и удовлетворение его потребностей в ненавязчивой и практически незаметной форме являются основной идеей в концепции окружающего интеллектуального пространства. Одним из примеров такого пространства является интеллектуальный зал, который оснащен сетью программных модулей, активационных устройств, мультимедийных средств и аудиовизуальных сенсоров [1]. Применение многомодального пользовательского интерфейса, использующего речь, движения, позы и жесты участников для выявления их намерений и потребностей, обеспечивает естественный способ взаимодействия с интеллектуальным залом. Осведомленность зала о пространственном положении участников, их текущих действиях, роли в текущем мероприятии и их предпочтениях позволяет автоматизировать работу встроенных программно-аппаратных модулей и обеспечить своевременное управление мультимедийным и другим оборудованием. На практике подобные интеллектуальные залы чаще всего функционируют в полуавтоматическом режиме, а эксперты-операторы поддерживают работу всех встроенных систем [2].

Выбор текущего активного диктора и запись его деятельности во время совещания является важной задачей в системах протоколирования мероприятий и поддержки телеконференций [2–3]. Для одновременной записи всех участников необходимо использование персональных или панорамных видеокамер. Такой способ подходит для небольших совещаний, когда все участники расположены за одним столом. Увеличение числа участников ведет к расширению пространства, подлежащего мониторингу и повышению стоимости записывающих технических средств.

Автоматический анализ аудиовидеозаписей совещаний осложняется тем, что приходится следить за достаточно большим количеством участников, которые произвольно меняют положение тела, головы, направление взгляда. Чтобы обеспечить захват и слежение за наибольшим числом участников, применяются массивы микрофонов, панорамные камеры, PTZ (Pan/Tilt/Zoom) – камеры с функциями наклона, поворота и масштабирования, распределённые системы камер [4]. Кроме видеомониторинга, для определения положения участников и текущего выступающего применяются датчики движения и массивы микрофонов. Нахождение источника звука с помощью массива микрофонов эффективно лишь в малых лекционных или конференц-залах. Для аудиозаписи в больших помещениях используют персональные микрофоны для всех участников или применяют систему массивов микрофонов, распределённую по всей аудитории [1–2].

Разработанный интеллектуальный зал совещаний предназначен для проведения небольших и средних мероприятий с числом участников до сорока двух. Также предусмотрена возможность проведения распределённых совещаний с подключением удалённых участников. Для слежения за участниками и записи выступающих применяется два комплекса устройств: 1) персональные веб-камеры служат для наблюдения за участниками, сидящими за конференц-столом; 2) три массива с Т-образной конфигурацией расположения микрофонов и пять видеокамер трёх различных типов выполняют аудиолокализацию и видеозахват остальных участников, сидящих в рядах кресел в другой части зала. Для составления полной картины текущей ситуации в зале производится анализ деятельности участников и состояния мультимедийного презентационного оборудования [5].

### Функционирование системы мониторинга участников совещания

Разработанная система аудиовизуального мониторинга совещаний состоит из четырёх модулей: 1) многомодальная система управления интеллектуальным залом (МСУИЗ); 2) многоканальная система обработки персональных веб-камер (МСОПВ); 3) многоканальная система аудиолокализации (МСА); 4) многофункциональная система видеомониторинга (МСВ). Далее рассмотрим назначение каждого модуля и особенности их совместного функционирования.

Модуль МСОПВ, представленный на рис. 1, осуществляет многоканальную обработку аудиовидеопотоков, поступающих с персональных веб-камер, установленных на конференц-столе [4]. Вначале производится анализ изображения на кадрах. Если лицо найдено, то далее ведется слежение за его перемещением, наведение веб-камеры на центр лица, и, кроме того, аудиосигнал, записанный микрофоном этой камеры, будет использоваться при выделении границ речи в многоканальном аудиопотоке. Для каждой камеры отмечаются интервалы времени, когда в кадре было найдено лицо (буфер  $B_1$ ) и когда в аудиоканале было определено наличие речи (буфер  $B_2$ ). Для сегментации видеозаписи выступления участника за конференц-столом используются данные из двух буферов.

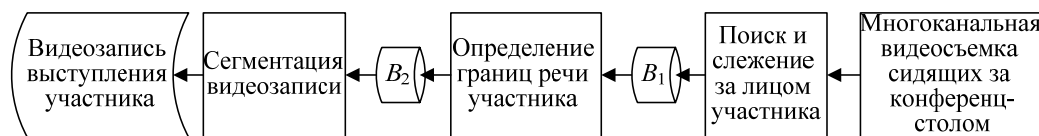


Рис. 1. Схема работы модуля МСОПВ

Модуль МСА выполняет захват аудиопотоков с трёх массивов микрофонов, вычисление координат источника звука в зале и оценку границ речевого сообщения. При наличии источника генерируется событие  $E_1$ , запускающее процесс определения границ речи. Аудиосигнал с записанной голосовой командой передается на модуль МСУИЗ для последующего распознавания. Полученные данные о координатах источника звука усредняются по трём массивам микрофонов, а затем передаются в модуль МСУИЗ и записываются в буфер  $B_3$ . Совместная работа модулей МСА и МСУИЗ показана на рис. 2.

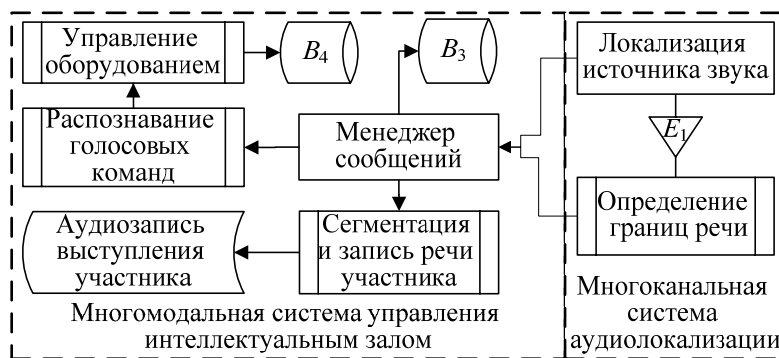


Рис. 2. Схема работы модулей МСА и МСУИЗ

Модуль МСУИЗ предназначен для управления оборудованием интеллектуального зала, контроля сообщений между другими модулями и экспертного сопровождения работы интеллектуального зала. В диалоговом окне модуля отображается положение участников, фотографии зарегистрированных участников, положение текущего источника звука (говорящего участника), состояние оборудования, текущее состояние мероприятия, распознанные голосовые команды и другая полезная для операторов информация. При изменении состояния оборудования с помощью голосовой команды, сенсорной панели, удаленно через веб-интерфейс или другим способом обновляется буфер  $B_4$ , в котором хранится информация о текущем состоянии используемых в зале устройств. Речевые сообщения, выделенные в ходе анализа речевой активности в зале, сохраняются в виде аудиозаписей, в названии файлов которых содержится информация о времени начала записи и местоположении источника звука.

С помощью набора интернет-камер AXIS модуль МСВ осуществляет видеозапись всего мероприятия, в том числе съемку основного выступающего, активных участников в аудитории во время

дискуссии и позволяет автоматизировать процесс регистрации участников [6]. В основном процессе модуля последовательно выполняются три операции (рис. 3): 1) обработка кадров, поступающих с панорамной камеры, установленной на потолке в центре зала для обнаружения и слежения за участниками внутри зала; 2) анализ зоны кресел с целью определения занятых мест; 3) выбор режима видеомониторинга на основе анализа данных о положении участников, состоянии групп освещения (события  $E_2$ ) и аудиоактивности в зале (события  $E_3$ ), хранящихся в буферах  $B_4$  и  $B_3$  модуля МСУИЗ соответственно.

В зависимости от выбранного режима видеомониторинга запускается один из четырех подпроцессов (регистрация участника, видеосъемка выступающего в зоне кресел, видеосъемка аудитории, видеосъемка основного докладчика) или их комбинация. Событие  $E_4$  инициирует процесс наведение интеллектуальной PTZ-камеры на активного диктора, находящегося в зоне кресел, и записывает его выступление. События  $E_5$  и  $E_6$  запускают соответственно подпроцессы записи выступления докладчика и вида аудитории.

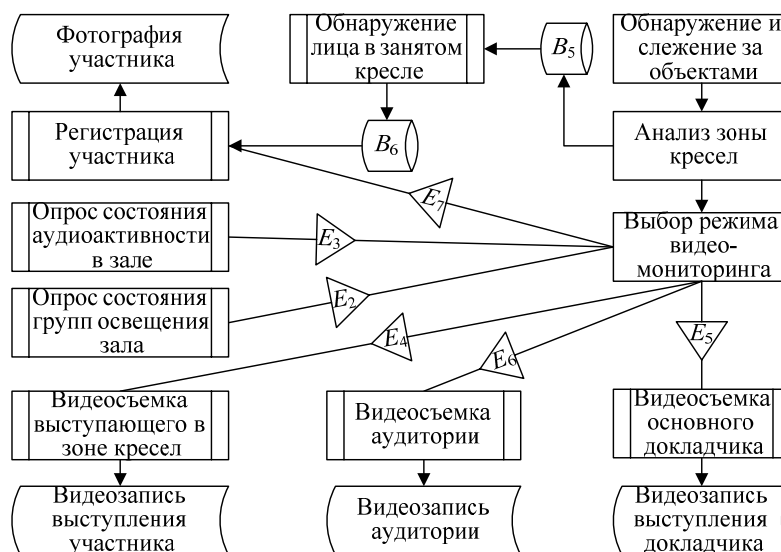


Рис. 3. Схема работы модуля МСВ

Регистрация участников, сидящих в зоне кресел, начинается после возникновения события  $E_7$ . В результате анализа зоны кресел буфер  $B_5$  заполняется номерами кресел, которые заняты участниками. Если такие найдены, то запускается подпроцесс обнаружения лиц участников, сидящих в креслах. Данный этап необходим для дополнительной проверки наличия участников в креслах. После обработки всего буфера  $B_5$  данные о незарегистрированных участниках сохраняются в буфере  $B_6$ . При возникновении события  $E_7$  и наличии незарегистрированных участников запускается подпроцесс регистрации участников, в ходе которого выполняется наведение интеллектуальной PTZ-камеры на лица участников и их фотографирование. В зависимости от режима видеомониторинга производится пассивная или активная регистрация. В первом случае по окончании регистрации сохраняется только фотография участника, во втором – в диалоговом режиме участник сообщает свои основные персональные данные. Таким образом, в ходе активной регистрации сохраняется не только фотография, но и аудиозапись, в которой участник сообщает свои данные.

В текущей версии модуля МСВ выделено пять режимов видеомониторинга: наблюдение, слежение за участниками, регистрация участников, выступление докладчика, дискуссия. В первом режиме выполняется обработка кадров только в основном потоке. При появлении участника в зале начинаются слежение за ним и видеосъемка аудитории, а при обнаружении незарегистрированных участников дополнительно запускается подпроцесс регистрации. В режиме выступления докладчика инициируется еще один подпроцесс, отвечающий за наведение камеры и видеосъемку участника, находящегося в зоне перемещения основного докладчика. При возникновении аудиоактивности в зале (событие  $E_3$ ) из режима выступления докладчика модуль МСВ переходит в режим дискуссии, в котором запускается подпроцесс видеосъемки выступающего в зоне кресел. В этом режиме ведется параллельная обработка видеопотоков от всех пяти камер, установленных в зале, с помощью пяти

перечисленных выше подпроцессов модуля МСВ. Описание разработанного алгоритма выбора режима видеомониторинга и методов видеообработки представлено в [6].

В результате работы системы аудиовизуального мониторинга совещаний создаются шесть баз данных, содержащих аудио- и видеoinформацию о мероприятии и его участниках, находящихся как за конференц-столом, так и сидящих в рядах кресел, расположенных в интеллектуальном зале. Полученные базы данных используются для протоколирования хода мероприятий и генерации мультимедийного контента веб-системы поддержки телеконференций.

#### Методика тестирования системы в интеллектуальном зале

Разработанная система аудиовизуального мониторинга является частью технологической платформы интеллектуального зала, которая включает комплекс программно-аппаратных модулей, а обработка мультимедийных и информационных потоков сообщений между ними ведется одновременно на нескольких многопроцессорных компьютерах. Для тестирования разработанной технологической платформы было предложено две методики, ориентированные на проверку работоспособности и оценивание качества работы.

Методика тестирования работоспособности платформы основана на последовательном опросе всех программных модулей и в случае отсутствия ответного сигнала в отведенный интервал времени запускается процедура восстановления переставшего работать модуля и устройств, которые он использует. Методика оценивания качества работы системы включает в себя представленные в таблице критерии, основанные на определении статистических ошибок первого ( $FA$ ) и второго рода ( $MR$ ). Расчет оценок производится по окончании мероприятия с привлечением операторов при просмотре и анализе записанных аудио- и видеоматериалов.

#### Перечень критериев оценки качества работы системы

Критерий	Формула
$A_p$ – точность определения участников в зале	$A_p = \frac{N_p - N_{FA\_p} - N_{MR\_p}}{N_p}$ , где $N_p$ – максимальное число участников, одновременно находящихся в зале; $N_{FA\_p}$ – число всех ошибочно определённых участников; $N_{MR\_p}$ – число всех пропущенных участников
$A_{o\_ch}$ – точность определения занятых кресел	$A_{o\_ch} = \frac{N_{ch} - N_{FA\_ch} - N_{MR\_s\_p}}{N_{ch}}$ , где $N_{ch}$ – число кресел, установленных в зале; $N_{FA\_ch}$ – число ошибочно определённых занятых кресел; $N_{MR\_s\_p}$ – число пропущенных системой кресел с сидящими участниками
$A_{s\_p}$ – точность определения участников, сидящих в зоне кресел	$A_{s\_p} = \frac{N_{ch} - N_{FA\_p\_f} - N_{MR\_p\_f}}{N_{ch}}$ , где $N_{FA\_p\_f}$ – число ошибочно определённых лиц участников, сидящих в зоне кресел; $N_{MR\_p\_f}$ – число пропущенных системой лиц участников, сидящих в зоне кресел
$A_m$ – точность определения режима работы	$A_m = \frac{N'_m}{N_m}$ , где $N_m$ – число изменения режимов, выявленных в ходе мероприятия; $N'_m$ – число корректно определенных режимов
$A_{m\_s}$ – точность наведения камеры на докладчика в зоне выступления	$A_{m\_s} = \frac{N'_{m\_s}}{N_{m\_s}}$ , где $N_{m\_s}$ – общее число кадров, выполненных при съемке докладчиков в зоне выступлений; $N'_{m\_s}$ – число кадров, на которых присутствует докладчик

Оценка качества работы системы мониторинга производилась по результатам записи пяти мероприятий в интеллектуальном зале с общим количеством участников 110 человек. После экспертного анализа записей были получены следующие средние значения оценок:  $\bar{A}_p = 88\%$ ,  $\bar{A}_{o\_ch} = 89\%$ ,  $\bar{A}_{s\_p} = 91\%$ ,  $\bar{A}_m = 97\%$ ,  $\bar{A}_{m\_s} = 90\%$ .

Для автоматизации управления оборудованием и протоколирования хода мероприятия в интеллектуальном зале разработана система аудиовизуального мониторинга участников, состоящая из четырёх основных модулей, выполняющих многоканальную обработку аудио- и видеосигналов для определения местоположения участников, выявления выступающих и их видеосъёмки. Предложенная система позволяет автоматизировать управление аудио- и видеоаппаратурой, а также другим оборудованием, установленным в зале, за счет дистанционного распознавания голосовых команд участников. При тестировании разработанной системы велась проверка функциональной работоспособности всех модулей, а после проведения ряда мероприятий была произведена оценка точности определения числа участников, режимов работы, наведения камеры на докладчика.

Работа выполнена в рамках ФЦП «Кадры» (ГК № П2360, № П2579), гранта РФФИ № 10-08-00199 и гранта Президента РФ (МД-501.2011.8).

#### *Литература*

1. Юсупов Р.М. От умных приборов к интеллектуальному пространству / Р.М. Юсупов, А.Л. Ронжин // Вестник Российской Академии Наук: научный и общественно-политический журнал. – 2010. – Т. 80, вып. 1. – С. 45–51.
2. Erol B. An overview of technologies for e-meeting and e-lecture / B. Erol, Y. Li // Proc. IEEE International Conference on Multimedia and Expo. – 2005. – P. 6–12.
3. Automating lecture capture and broadcast: Technology and videography / Y. Rui, A. Gupta, J. Grudin, L. He // Multimedia Systems. – 2004. – Vol. 10. – P. 3–15.
4. Rienks R. Pro-active meeting assistants: attention please! / R. Rienks, A. Nijholt, P. Barthelmess // AI & Society. Springer. – 2009. Vol. 23(2),– P. 213–231.
5. Ронжин А.Л. Проектирование интерактивных приложений с многомодальным интерфейсом / А.Л. Ронжин, А.А. Карпов // Доклады ТУСУРа. – 2010. – № 1 (21), ч. 1. – С. 124–127.
6. Ronzhin A.L. A Video Monitoring Model with a Distributed Camera System for the Smart Space / A.L. Ronzhin, M.V. Prischeva, A.A. Karpov // Springer. – 2010. – LNCS 6294. – P. 102–110.

---

#### **Ронжин Александр Леонидович**

Аспирант лаб. речевых и многомодальных интерфейсов  
Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН)  
Тел.: +7-812-328-70-81  
Эл. почта: ronzhinal@iiias.spb.su

#### **Ронжин Андрей Леонидович**

Д-р техн. наук, доцент, зав. лаб. речевых и многомодальных интерфейсов СПИИРАН  
Тел.: +7-911-253-24-32  
Эл. почта: ronzhin@iiias.spb.su

Ronzhin A.L., Ronzhin An.L.

#### **The system of audiovisual monitoring of participants in the intelligent meeting room**

The paper is dedicated to the aspects which arise during the development of automatic video recording system of events in intelligent meeting room. The offered monitoring system of the participants in the meeting is used for recording the current situation and extends possibilities to automate audio-, video and presentation equipment control.

**Keywords:** audiovisual processing, ambient intelligent space, context, information-control system.