УДК 004.5

В.Ю. Будков

Методы и программные средства обработки мультимедийных данных при сопровождении распределенных совещаний

Рассмотрены методы и программно-аппаратные средства автоматического анализа аудиосигналов, записанных в ходе сопровождения распределенных совещаний и применяемых для генерации отчетных материалов по результатам мероприятия.

Ключевые слова: окружающее интеллектуальное пространство, сопровождение совещаний, телеконференции, диаризация дикторов, протоколирование мероприятий.

Основные проблемы систем сопровождения распределенных совещаний. Современные системы связи и совместной работы не позволяют полностью автоматизировать процесс информационной поддержки проведения совещаний, поэтому большая часть работы по сопровождению удаленных участников выполняется операторами-людьми. Кроме того, при проведении деловых встреч, заседаний, совещаний и других формальных мероприятий обязательной процедурой является протоколирование выступлений участников [1, 2]. Однако экспертный анализ и расшифровка аудиозаписей совещаний требуют привлечения специалистов-стенографистов и занимают длительное время обработки [3]. Современные методы анализа речи и диаризации дикторов позволяют автоматизировать процесс выделения реплик участников совещания. Одним из перспективных способов увеличения эффективности систем диаризации является применение многоканального и многомодального анализа поведения участников в зале совещаний [4, 5]. Проведенный анализ комбинированных систем диаризации показал, что использование дополнительных признаков, извлекаемых при анализе изображений и локализации источника звука, позволяет повысить точность определения момента смены диктора [6, 7].

Другим ограничением систем телеконференций являются пропускная способность коммуникационных сетей и мультимедийные возможности клиентского устройства, которые существенным образом влияют на вид пользовательского интерфейса и выбор информационных каналов, доступных для удаленных участников. Ключевым вопросом при дистанционной коммуникации является высокая неопределенность о ситуации в удаленной аудитории, вызванная пространственно-временными различиями [8]. Физические и психологические барьеры препятствуют удаленному участнику быстро присоединиться к дискуссии проблемы, обсуждаемой участниками внутри зала, и тем более предложить новое направление разговора.

Анализ направлений исследований по изучению поведения человека в интеллектуальном пространстве и развитию интуитивных многомодальных интерфейсов показал перспективность дальнейшей разработки методов и программных средств аудиовизуальной поддержки мобильных телесовещаний, отличающихся применением средств автоматического анализа и оценки информационной значимости передаваемого контента, а также передачи оптимизированного аудиовизуального потока данных, обеспечивающих снижение когнитивной нагрузки на пользователя и уменьшение потребляемых ресурсов мобильным устройством [3, 9].

Таким образом, основной задачей данного исследования является разработка математических и программных решений, повышающих возможности удаленного участника при принятии решений и участии в дискуссиях во время распределенных мероприятий, а также снижении затрат на подготовку мультимедийных отчетных материалов.

Модель сопровождения участников распределенных мероприятий. Сопровождение мероприятий включает в себя три основных этапа: подготовка к совещанию, запись и трансляция, затем архивирование и анализ записанных материалов. На втором и третьем этапах для анализа многоканальных потоков мультимедийной информации, поступающей от распределенных участников мероприятия, требуется привлечение средств автоматической обработки данных. Причем более жесткие требования по времени обработки предъявляются при трансляции мероприятий, а архивирование и

подготовка протокола мероприятия могут быть выполнены после завершения совещания. На рис. 1 представлены основные методы обработки мультимедийных сигналов, применяемых для трансляции мероприятий и последующей подготовки отчета в предложенной модели сопровождения мероприятий.

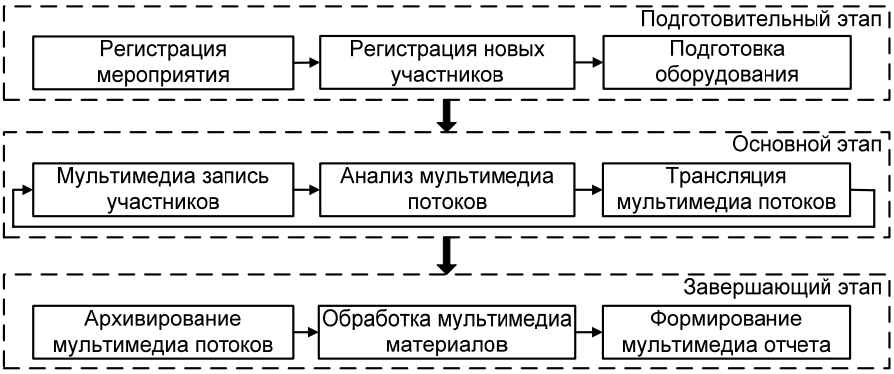

Рис. 1. Обработка мультимедийных сигналов в системе сопровождения мероприятий

Подготовительный этап включает в себя регистрацию мероприятия, подготовку оборудования и регистрацию участников. Вначале в систему сопровождения мероприятия вносятся данные о времени и месте проведения, разрешениях на доступ к трансляции мероприятия, возможности регистрации посторонних участников и др. После этого участники могут зарегистрироваться в системе и авторизоваться для участия в данном мероприятии. Перед началом мероприятия производится подготовка и настройка необходимого мультимедийного оборудования. Основной этап включает в себя трансляцию и запись аудио- и видеопотоков и других мультимедийных данных, получаемых из различных источников, расположенных в помещении проведения мероприятия и от удаленных участников. При выборе наиболее актуальной информации, которая транслируются удаленным участникам, учитываются следующие аспекты: положение текущего выступающего, время смены слайда презентации, этап мероприятия и другие параметры. Завершающий этап сопровождения мероприятия включает в себя анализ мультимедийных записей, их архивирование и создание отчета по мероприятию. Генерация отчета производится по шаблонам, которые могут редактироваться вручную для получения необходимой формы представления.

Средства генерации отчетных материалов по результатам мероприятия. При формировании отчета по мероприятию производится формирование протокола, где отмечены реплики каждого из участников. Для реализации этой задачи был разработан метод диаризации дикторов в одноканальном аудиопотоке, включающий два основных этапа: цифровую обработку аудиосигналов и диаризацию дикторов. На рис. 2 приведена схема двухэтапной процедуры диаризации дикторов, где показаны блоки, выполняющие следующие функции: *1* – разделение аудиосигнала на речевые и неречевые фрагменты; *2* – выявление параметров речевого сигнала на основе моделей слухового восприятия человека; *3* – извлечение информативных признаков голосового источника речевого сигнала; *4* – выделение речевых артефактов, характерных для разговорной речи; *5* – оценивание отношения сигнал/шум в аудиосигнале; *6* – идентификация параметров голосового источника речевого сигнала независимо от контекста и языка; *7* – разделение речи различных дикторов.

Параметрическое представление речевого сигнала и временная разметка речевых фрагментов, вычисленных на этапе обработки аудиосигнала, используются при идентификации участников среди существующих моделей дикторов. Если диктор не идентифицирован или при отсутствии моделей, параметрическое представление текущего речевого фрагмента применяется для обучения модели нового диктора. После определения диктора текущему речевому фрагменту присваивается номер модели его диктора. Результат диаризации представляет собой временную разметку речевых фрагментов по дикторам.

Программная реализация предложенного метода диаризации речи дикторов, а также предложенные ранее методы анализа и распознавания речи были использованы при разработке программного комплекса автоматического анализа, распознавания и диаризации разговорной русской речи (ПАРАД-Р) [6, 10]. На рис. 3 представлена архитектура программного комплекса ПАРАД-Р, построенная на основе трехуровневой архитектуры (клиентская часть, серверная часть, программно-математическое ядро).

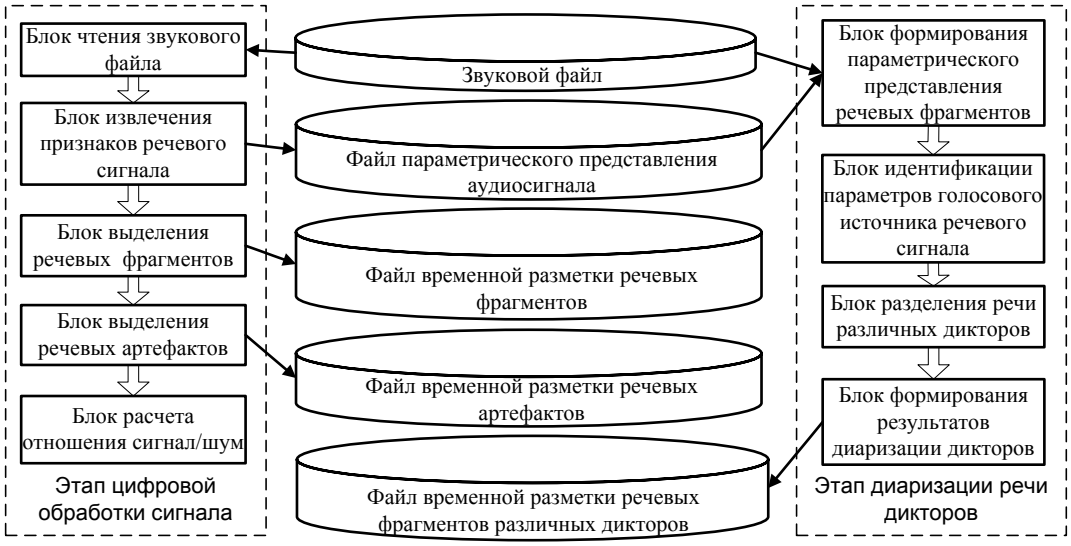

Рис. 2. Основные этапы аудиообработки при диаризации дикторов

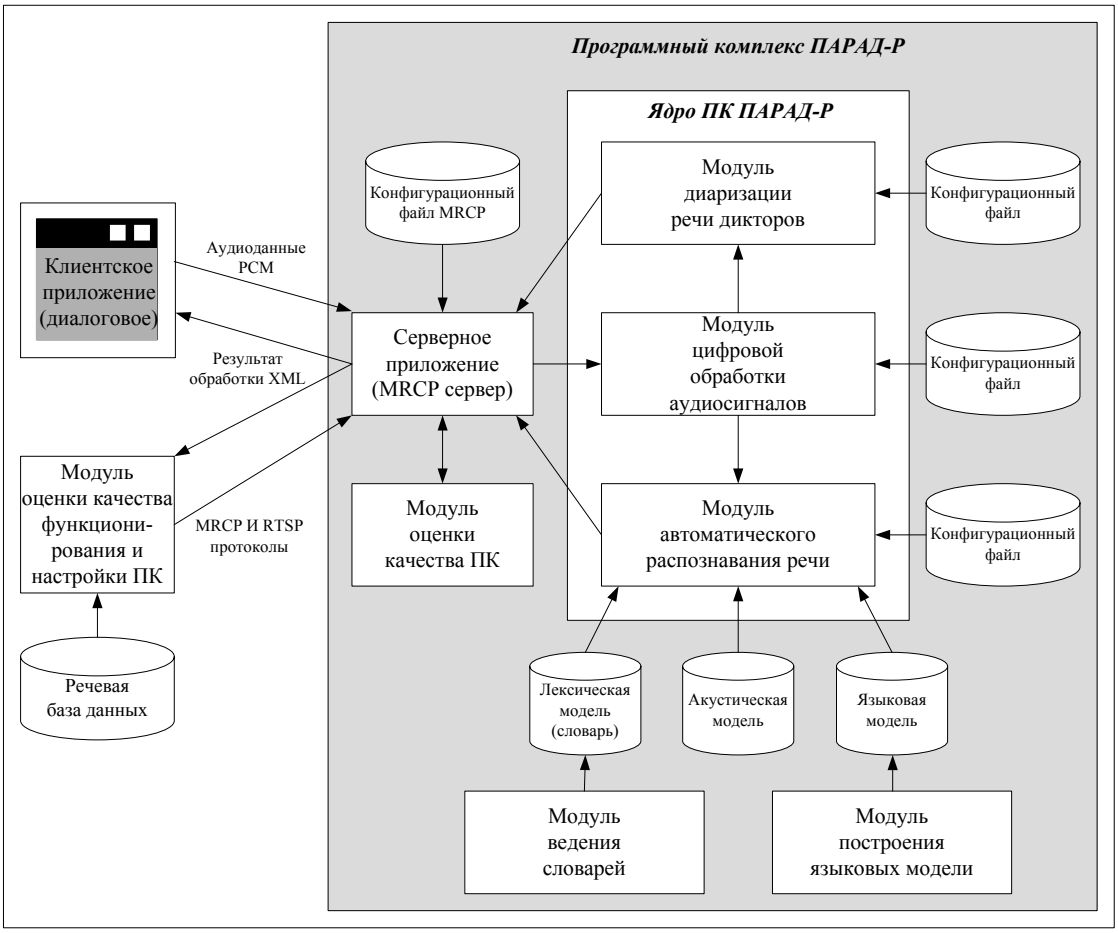

Рис. 3. Общая архитектура экспериментального образца ПАРАД-Р

Клиентская и серверная части могут располагаться как на одном компьютере, так и на разных компьютерах и взаимодействовать по компьютерной сети. Информационный обмен между клиентской частью и серверной частью реализован с использованием протокола MRCPv2 (Media Resource Control Protocol). Серверная часть состоит из следующих программных модулей:

1) серверное приложение – MRCP сервер;
2) модуль ведения словарей;
3) модуль построения языковых моделей;

4) модуль оценки качества ПК.

Каждый из этих модулей, кроме последнего, реализован в виде исполняемого файла, работающего под управлением ОС Microsoft Windows. Помимо этих программных модулей, серверная часть также имеет связь с программно-математическим ядром ПК, в который входят:

1) программная библиотека цифровой обработки аудиосигналов;
2) программная библиотека диаризации речи дикторов;
3) программная библиотека автоматического распознавания речи.

Оценка качества функционирования разработанного комплекса была проведена по методикам, учитывающим метрики WER (Word Error Rate), LER (Letter Error Rate), SWER (Speaker Attributed Word Error Rate) и DER (Diarisation Error Rate), с использованием созданного речевого корпуса слитной русской речи [11]. Разработанный метод диаризации дикторов в одноканальном аудиопотоке показал точность сегментации реплик разных дикторов свыше 85%. Для систем сопровождения распределенных совещаний комплекс может использоваться при подготовке протокола и стенограмм выступлений участников.

Заключение. Совокупность предложенных методов и программных средств автоматической обработки мультимедийных потоков данных, а также их практическая реализация представляют собой решение актуальной научно-технической задачи информационного и технологического сопровождения распределенных мероприятий на основе анализа текущей ситуации, оптимизации транслируемого контента удаленным участникам и генерации отчетных материалов по результатам мероприятия. Разработанный программный комплекс автоматического анализа, распознавания и диаризации разговорной русской речи поддерживает пакетную обработку аудиосигналов с доступом по стандартному протоколу MRCPv2 и может применяться для разработки кроссплатформенных приложений по распределению и управлению динамическими речевыми и многомодальными сервисами, в том числе по обработке архивных записей мероприятий.

Работа выполнена в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 годы» (ГК № 07.514.11.4139) и гранта РФФИ № 13-08-00741-а.

Литература

1. 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.

2.Мещеряков Р.В. Специализированная информационная система поддержки деятельности медицинского учреждения / Р.В. Мещеряков, Л.Н. Балацкая, Е.Л. Чойнзонов // Информационно-управляющие системы. – 2012. – № 5. – С. 51–56.

3. Ронжин А.Л. Технологии поддержки гибридных е-совещаний на основе методов аудиовизуальной обработки / А.Л. Ронжин, В.Ю. Будков // Вестник компьютерных и информационных технологий. – 2011. – № 4. – С. 31–35.

4. Noulas A. Multimodal Speaker Diarization / A. Noulas, G. Englebienne, B.J.A. Krose // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34(1). – P. 79–93.

5. Мещеряков Р.В. Сегментация и параметрическое описание речевого сигнала / Р.В. Мещеряков, В.П. Бондаренко, А.А. Конев // Изв. вузов. Приборостроение. – 2007. – Т. 50, № 10. – С. 3–7.

6. Ронжин А.Л. Особенности дистанционной записи и обработки речи в автоматах самообслуживания / А.Л. Ронжин, А.А. Карпов, И.А. Кагиров // Информационно-управляющие системы. – 2009. – Вып. 42, т. 5. – С. 32–38.

7. Ронжин Ал.Л. Формирование профиля пользователя на основе аудиовизуального анализа ситуации в интеллектуальном зале совещаний / Ал.Л. Ронжин, В.Ю. Будков, Ан.Л. Ронжин // Труды СПИИРАН. – 2012. – Вып. 23. – С. 482–494.

8. Yankelovich N. Porta-person: telepresence for the connected meeting room / N. Yankelovich, J. Kaplan, N. Simpson, J. Provino // CHI 2007. – 2007. – P. 2789–2794.

9. Ронжин Ал.Л. Система аудиовизуального мониторинга участников совещания в интеллектуальном зале / Ал.Л. Ронжин, Ан.Л. Ронжин // Доклады ТУСУРа. – 2011. – № 1 (22), ч. 1. – С. 153–157.

10. Ронжин А.Л. Топологические особенности морфофонемного способа представления словаря для распознавания русской речи // Вестник компьютерных и информационных технологий. – 2008. – № 9. – С. 12–19.

11. Будков В.Ю. Анализ современных методов и систем диаризации дикторов / В.Ю. Будков, А.Л. Ронжин // Изв. вузов. Приборостроение. – СПб.: ИТМО, 2011. – № 11. – С. 43–46.

Будков Виктор Юрьевич
Мл. науч. сотрудник лаб. речевых и многомодальных интерфейсов
Санкт-Петербургского института информатики и автоматизации Российской академии наук (СПИИРАН)
Тел.: 8 (812) 328-70-81
Эл. почта: budkov@iias.spb.su

Budkov V.Yu.
Methods and software of multimedia data processing at support of distributed meetings

Methods, software and hardware for automatic analysis of audio data recorded during support of distributed meeting and applied for creation of report materials of the meeting results.
Keywords: context-aware applications, ambient intelligent space, mobile heterogeneous devices.
