

Том 75, Номер 3

ISSN 0869-5873
Март 2005



ВЕСТНИК РОССИЙСКОЙ АКАДЕМИИ НАУК

<http://www.maik.ru>



“НАУКА”

ВСЕПРЕВЛАДШИИ ДЕРЖАВЪ
ВЕЛИКИИ ЦАРЬ ПЕТРЪ АЛЕУЪ

РЕЧЬ И КОМПЬЮТЕР

С 20 по 22 сентября 2004 года в Санкт-Петербурге прошла 9 международная конференция “Речь и Компьютер” SPECOM’2004. Серия конференций SPECOM была основана Институтом информатики и автоматизации РАН (СПИИРАН) совместно с Российским государственным педагогическим университетом им. А.И. Герцена (РГПУ) в 1996 году. Первая конференция прошла в Санкт-Петербурге, затем в 1997 - в Румынии, 1998, 2000, 2002 - в Санкт-Петербурге, 1999, 2001, 2003 - в Москве. В этом году конференция вызвала особый интерес ученых всего мира, и число участников составило более 170 человек из 36 стран.

Взаимодействие человека с компьютером в естественной форме, подобной общению между людьми, является одной из самых важных и сложных задач искусственного интеллекта. Существующие модели понимания речи пока еще значительно уступают речевым способностям человека, что свидетельствует об их недостаточной адекватности и ограничивает применение речевых технологий в промышленности и быту. Кроме того, сейчас ведутся активные исследования по созданию многомодальных интерфейсов, которые объединяют различные способы ввода информации (речь, артикуляция губ, жесты, направление взгляда и т.д.). Многомодальный интерфейс естественен для межчеловеческого общения. 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.

SPECOM одна из не многих российских конференций, посвященных речевым технологиям, которая собирает столь много иностранных участников и несет по настоящему международный характер. Во многом это связано с тематикой конференции и кругом обсуждаемых вопросов – это речь, язык, взаимодействие человека с компьютером. Более того, основная цель исследований – создание эффективных и естественных средств межчеловеческого общения, а также человеко-машинного взаимодействия, несмотря на различные лингвистические и социальные сложности. Такая конференция чрезвычайно важна как для ученых из России, так и для иностранных ученых (из Европы, Америки и Азии), так как является связующим звеном между Российской и Международной наукой. Конференция позволяет отечественным ученым познакомиться с ведущими мировыми достижениями в данной научной области, а зарубежным ученым – узнать больше об исследованиях, проводимых в странах России и СНГ.

Основной целью всех конференций SPECOM является обсуждение многочисленных важных проблем человеко-машинного взаимодействия и искусственного интеллекта: автоматического распознавания и понимания речи, синтеза речи, верификации и идентификации диктора. Каждый год на конференции представляются новейшие разработки и результаты исследований в данной научной области.

Девятая конференция SPECOM’2004 была организована СПИИРАН при финансовой поддержке Международной ассоциации содействия в сотрудничестве с учеными бывшего Советского Союза (ИНТАС) и Европейского научного сообщества SIMILAR Network of Excellence. Программа конференции включала 130 докладов, представленных учеными из 35 стран СНГ, Европы, Америки, Азии, Австралии и Африки (Таблица 1.). К началу конференции все доклады были опубликованы в сборнике трудов конференции, включающем свыше 750 страниц, и в электронном виде на компакт дисках.

Таблица 1. Распределение количества представленных статей по странам

53	Россия	3	Корея	1	Алжир
10	Германия	3	Мексика	1	Аргентина
9	Беларусь	3	Сербия и Черногория	1	Армения
5	Бельгия	3	Нидерланды	1	Австралия
5	Греция	3	Турция	1	Бразилия
5	Япония	3	Украина	1	Канада
5	Испания	2	Индия	1	Китай
5	США	2	Израиль	1	Иран
4	Франция	2	Новая Зеландия	1	Литва
4	Великобритания	2	Польша	1	Сингапур
3	Чехия	2	Португалия	1	Таиланд
3	Финляндия	2	Румыния		

В рамках конференции SPECOM'2004 был проведен стратегический научный семинар ИНТАС “Разработка перспективных приложений для человеко-машинного взаимодействия в Информационном Обществе”, направленный на обсуждение достижений и проблем в области человеко-машинного взаимодействия и информационных технологий, а также выработки необходимых стратегических действий для улучшения научной кооперации между странами Содружества Независимых Государств и Европейскими странами. Для этого были приглашены не только ведущие мировые ученые, но и представители научных, промышленных международных организаций, оказывающих финансовую поддержку научным исследованиям.

Пленарное заседание открыл председатель конференции директор СПИИРАН Р.М. Юсупов. Он подчеркнул важность и актуальность исследований в области автоматического распознавания и понимания речи и в целом систем взаимодействия человека с компьютером. Отдельно была подчеркнута важность нового направления в науке и технике – многомодальные интерфейсы, которые впервые представлены в России. Затем с ключевыми лекциями выступили профессор Ж.-П. Хатон (Франция), президент Британской Голосовой Ассоциации Дж. Рубин (Великобритания), профессор П.А. Скрелин (Россия), представитель France Telecom R&D профессор К. Барткова (Франция).

Профессор Ж.-П. Хатон представил обзорный доклад “Автоматическое распознавание речи: прошлое, настоящее и будущее” по существующим передовым методам распознавания речи и о будущем автоматического распознавания речи. В докладе были представлены методики использования Скрытых Марковских Моделей для речевых технологий.

Профессор Дж. Рубин представил обзорный доклад “Голосовые нарушения у работников: психодинамическая модель” о речеобразовании и психологических аспектах восприятия речи, а также о проблемах возникающих у людей в процессе речеобразования и методах распознавания “усложненной” речи у людей со специфическими нуждами.

Профессор П.А. Скрелин представил доклад “Сегментные признаки в различных стилях речи” о различиях читаемой речи (т.е. заготовленной заранее) и спонтанной речи, а также результатах распознавания обоих типов речи. Были представлены основные подходы к распознаванию как изолированной, так и слитной речи, а также созданные речевые базы данных, необходимые для обучения систем автоматического распознавания речи.

Профессор К. Барткова представила доклад “Обработка иностранного акцента в автоматическом распознавании речи” о различиях в организации систем распознавания

различных языков (французского, английского, немецкого, испанского и т.д.). Был представлен фонетический состав этих языков, а также основные фонологические правила для представления акустических моделей.

Дальнейшая работа конференции проходила в рамках одиннадцати сессий.

Основные вопросы, обсуждавшиеся на секции “Многомодальные интерфейсы” касались систем, в которых для общения человек-компьютер используются одновременно несколько естественных способов передачи информации: речевой ввод, жестовый, рукописный и т.д. Особый интерес вызвал доклад профессора Ф. Маркуеса (Испания) и профессора Б. Мака (Бельгия), посвященный методам распознавания жестов и положения человека в пространстве, а также совмещения этого ввода информации с речевым вводом. Также в данной сессии был представлен интересный доклад профессора С. Джовичича (Сербия и Черногория) об определении эмоционального состояния человека и использовании эмоциональной окраски речи для человеко-машинного взаимодействия.

На сессии “Обработка речевого сигнала” были представлены доклады, касающиеся математического моделирования сигналов, фильтрации речи, очистки от шумов. Среди представленных докладов особый интерес вызвал доклад профессора Х. Кувабара (Япония), касающийся влияния индивидуальных отличий дикторов на акустические параметры воспроизводимой речи и вариативность этих параметров, а также ряд докладов по новым методам обработки речевого сигнала, представленной группой под руководством профессора А. Вендемута (Германия).

Сессия “Распознавание речи” являлась ключевой на конференции, и на ней было представлено больше всего докладов. Доклады были посвящены фундаментальным вопросам автоматического распознавания речи, а также различным аспектам применения речевых технологий. Отдельное внимание было уделено технологиям распознавания национальных языков (в частности русского). Были сделаны доклады представителями таких крупных американских компаний как IBM (Д. Каневский, США) и Motorola (Ч. Вуд, США). Особый интерес вызвал доклад А.Л. Ронжина (Россия) посвященный совершенно новому методу распознавания русской речи с использованием морфемного уровня представления речи. На сессии также были успешно продемонстрированы демонстрационные версии ряда прикладных систем автоматического распознавания речи (Ч. Вуд, США; Д. Каневский, США; А.Л. Ронжин, Россия; М. Фархадов, Россия), вызвавшие интерес как со стороны ведущих ученых, так и со стороны потенциальных инвесторов и конечных потребителей.

На сессиях “Понимание речи и естественно-языковая обработка” и “Диалог, онтологии и представление знаний”, особый интерес вызвали доклады профессора В.И. Галунова (Россия) и профессора И.М. Кобозевой (Россия). В докладе В.И. Галунова была предпринята попытка построения общей структуры онтологии речевых технологий. Приведены примеры некоторых элементов онтологии в области распознавания и синтеза речи. В докладе профессора И.М. Кобозевой были рассмотрены проблемы возможностей просодической, семантической и прагматической изменчивости разговорного дискурса.

Работа сессии “Распознавание диктора” была посвящена методам верификации и идентификации дикторов, а также использованию таких биометрических методов в системах безопасности, направленных на предупреждение террористических актов, что особенно актуально в сегодняшней обстановке во всем мире. Интересные научные результаты и прикладные системы были представлены группой под руководством Ю. Саастамоинена (Финляндия).

На сессии “Синтез речи” были представлены методики синтеза и генерации речи для различных языков. Доклады, а также примеры синтезированной речи были представлены профессором Э. Шпилевским (Польша), Л. Дохерти (Австралия), группой профессора Б.М. Лобанова (Беларусь), а также рядом других известных ученых.

В рамках конференции также была проведена специальная сессия “Многомодальные сервисы и приложения для инвалидов”, организованная совместно с Европейским научным сообществом SIMILAR NoE. Данное направление исследований

целиком направлено на создание перспективных систем для людей со специальными нуждами. В рамках данной сессии были представлены доклады Д. Тзовараса (Греция), А. Каплиер (Франция), представляющих SIMILAR NoE, по использованию жестового ввода совместно с речевым вводом для слепых людей с созданием виртуального окружения, а также по использованию жестовой речи при телефонном общении между глухими людьми, а также ряд других статей, посвященных данному направлению.

В рамках конференции был проведен стратегический научный семинар ИНТАС “Разработка перспективных приложений для человеко-машинного взаимодействия в Информационном Обществе”. Семинар проходил 2 дня и состоял из трех сессий: “Научно-стратегическая сессия”, “Представление Европейских научных проектов” и “Основы человеко-машинного взаимодействия”. На семинаре были заслушаны стратегические доклады, касающиеся существующих программ и фондов для поддержки фундаментальных исследований, а также развития научных школ. Официальный представитель ИНТАС П. Азирелли представила доклад о приоритетных направлениях деятельности ассоциации, некоторые результаты завершенных, а также текущих проектов по информационным технологиям. Координатор FP6 по Северо-Западу России А. Иванов, представил доклад об опыте участия российских научных организаций в шестой европейской рамочной программе FP6 на 2002-2006 гг. Официальные представители РФФИ Н. Диденко и А. Петровский представили доклады о кооперации между ИНТАС и научным сообществом Санкт-Петербурга, а также о грантах, предлагаемых РФФИ для поддержки фундаментальных исследований в Санкт-Петербурге.

Также на стратегическом научном семинаре ИНТАС были представлены доклады о научных исследованиях, которые соответствуют приоритетным направлениям ИНТАС. В частности, профессор Т.К. Винцюк (Украина) выступил с докладом о существующих речевых технологиях на Украине и о новых предполагаемых проектах. Профессор Л.А. Станкевич (Россия), команда которого победила на чемпионате мира по футболу среди роботов RoboCup-2004, представил доклад о перспективах развития интеллектуальных роботов в России, а также об опыте участия в соревнованиях среди роботов. Профессор Б.В. Соколов и профессор А.В. Тимофеев рассказали об текущих и планируемых исследованиях в области человеко-машинного взаимодействия, проводимых на базе СПИИРАН.

Конференция оказалась самым крупным и представительным научным мероприятием по речевым технологиям из проводимых в России и СНГ. К основным результатам работы Международной конференции SPECOM'2004 в Санкт-Петербурге следует отнести следующие:

- оценка достижений и выявление наиболее острых неразрешенных вопросов человеко-машинного взаимодействия;
- улучшение кооперации между академическими и промышленными организациями, работающими над созданием перспективных систем взаимодействия человека с компьютером;
- знакомство российских участников с последними научными результатами лидирующих иностранных компаний и университетов;
- продвижение будущих совместных международных проектов, летних школ и других научных мероприятий, посвященных проблемам человеко-машинного взаимодействия в информационном обществе.

А.Л. Ронжин
кандидат технических наук