ИЗВЕСТИЯ РАН. СЕРИЯ ЛИТЕРАТУРЫ И ЯЗЫКА, 2017, том 76, № 1, с. 71–75

ХРОНИКА

XVIII МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ SPECOM 2016 “РЕЧЬ И КОМПЬЮТЕР”

XVIII INTERNATIONAL CONFERENCE SPECOM 2016 “SPEECH AND COMPUTER”

XVIII Международная конференция “Речь и Компьютер” (SPECOM 2016) проходила с 23 по 27 августа 2016 г. в г. Будапешт (Венгрия) на базе Будапештского университета Технологии и экономики и Венгерской ассоциации по инфокоммуникации при поддержке в качестве спонсора организации Speechtex (“Речевые экспертные технологии”), а также в кооперации с Международной Ассоциацией по речевой коммуникации (ISCA), Московским государственным лингвистическим университетом (МГЛУ), Санкт-Петербургским институтом информатики и автоматики РАН (СПИИРАН), Санкт-Петербургским национальным исследовательским университетом информационных технологий, механики и оптики (Университет ИТМО). Вышеуказанная конференция SPECOM 2016 в этом году отмечала свое двадцатилетие (1996–2016). Она регулярно организовывалась со дня основания МГЛУ и СПИИРАН. Согласно традиции организатор от МГЛУ Р.К. Потапова и представитель из Санкт-Петербурга из СПИИРАН открывали, вели и закрывали данную конференцию. В рамках конференции были организованы лекции ведущих специалистов в области новых технологий и многомодальной коммуникации. Все доклады данной Международной конференции были изданы в виде сборника трудов[1]. В работе конференции SPECOM 2016 приняли участие докладчики из 27 стран: Российской Федерации, Венгрии, Чешской Республики, Беларуси, Бельгии, Великобритании, Германии, Греции, Египта, Индии, Ирландии, Италии, Колумбии, Коста-Рики, Македонии, Малайзии, Мексики, Литвы, Сербии, Словакии, США, Таиланда, Финляндии, Франции, Швеции, Швейцарии, Японии. Наибольшее количество докладчиков представила Российская Федерация (109). Второе, третье и четвертое места по данному критерию заняли: принимающая страна, Венгрия, – 25 докладчиков, Чешская Республика – 19 докладчиков, Сербия – 14 докладчиков. В 2016 г. общее число докладчиков на конференции составило 271.

Одновременно с Международной конференцией SPECOM 2016 впервые проходила Международная конференция ICR 2016 (Interactive Collaborative Robotics – Интерактивная робототехника), на которой были представлены доклады[2] из 6 стран (РФ, Чешской Республики, Тайваня, Беларуси, Австрии, Венгрии). Наибольшее число докладов – от Российской Федерации (91), от Австрии – 5 докладов, от Чешской Республики – 4 доклада. Общее число докладчиков на конференции ICR 2016 – 108.

В обзоре отражены материалы докладов, имеющие непосредственное отношение исключительно к проблемам прикладной лингвистики. Большая часть докладов была посвящена разработке речевых и мультимодальных технологий, программным средствам, новым технологиям.

Следующая (19-я по счету) Международная конференция SPECOM 2017 “Речь и компьютер” пройдет с 12 по 16 сентября 2017 г. в Великобритании (г. Хатфилд, Хартфордшир) параллельно со 2-й Международной конференцией ICR 2017 по интерактивной робототехнике.

Р. Шлютер, П. Дётш, П. Голик и др. (Аахен) в докладе “Автоматическое распознавание речи на основе нейронных сетей” отмечают, что в системах автоматического распознавания речи, как и во многих других областях машинного обучения, стохастическое моделирование все в большей степени опирается на нейронные сети. И в

[1] Proceedings of the 18th International Conference on Speech and Computer, SPECOM 2016, ser. Lecture Notes in Artificial Intelligence (including subseries Lecture Notes in Computer Science), 9811 LNAI. Ronzhin A., Potapova R., Németh G. (eds). Cham; Heidelberg; New York; Dordrecht; London: Springer International Publishing, 2016. 731 p.

[2] Proceedings of the First International Conference “Interactive Collaborative Robotics”, ICR 2016, ser. Lecture Notes in Artificial Intelligence (including subseries Lecture Notes in Computer Science), 9812 LNAI. Ronzhin A., Rigoll G., Meshcheryakov R. (eds). Cham; Heidelberg; New York; Dordrecht; London: Springer International Publishing, 2016. 251 p.

акустике, и в моделировании языка нейронные сети сегодня составляют значительную часть современных работ в рамках распознавания слитной речи с большим словарем, что является огромным шагом вперед по сравнению с прежними подходами, которые были основаны исключительно на скрытых марковских моделях, нормальных распределениях, а также языковых моделях. В докладе представлен обзор текущих работ в области моделирования нейронных сетей, предназначенных для систем автоматического распознавания речи. Обзор включает обсуждение сетевых топологий и типов ячеек, обучения и оптимизации процесса распознавания на этом этапе, выбора вводимых функций, адаптации и нормализации, многоцелевого обучения, а также моделирования языка на основе нейронной сети. Авторы отмечают, что, несмотря на очевидный прогресс, достигнутый в распознавании речи с применением моделирования нейронных сетей, предстоит разработать еще многое, чтобы получить последовательный и самодостаточный метод моделирования на основе нейронной сети, который учитывал бы и прежнее состояние технологических подходов.

В докладе «Автоматизированная обработка диалогов; о понятии “речевая энтропия”» Н. Кэмпбелла (Дублин) представлены некоторые идеи об “интерактивных” говорящих машинах, проиллюстрированные примерами диалогов системы HERME. HERME представляло собой небольшое устройство, которое инициировало беседы с прохожими в Научной галерее Тринити-колледжа в Дублине и которому удалось привлечь большинство из них к участию в коротких беседах продолжительностью около трех минут. Распознавание речи не было задействовано. Опыт подбора таких данных и анализ “бесед” позволили рассмотреть теорию речевой энтропии, в рамках которой коммуникативные связи становятся со временем свободными и по мере окончания темы разговора могут “обновляться” за счет смены говорящих и возобновления беседы. Смех является особым признаком такого механизма “затухания” беседы, что может оказаться достаточной информацией для машины, чтобы она смогла “включиться” в беседу людей без дискомфорта для последних.

Доклад “Сравнение акустических признаков речи у детей с нормальным развитием и детей с нарушениями в области аутизма (ASD)” Е. Ляксо, О. Фроловой, А. Григорьева (Санкт-Петербург) посвящен результатам исследования акустических характеристик, специфических для процесса вокализации и речи детей с различными нарушениями в области аутизма. Было проведено три типа экспериментов со следующими видами речи: эмоциональная речь, спонтанная речь и повторение слов. Участниками исследования были дети с расстройствами в области аутизма, разного возраста 5–14 лет (n = 25 детей) и дети с нормальным развитием в возрасте 5–14 лет (n = 60). Сравнивались акустические характеристики, которые широко используются при распознавании и восприятии речи: значения основного тона, максимальное и минимальное значения основного тона, диапазон значений основного тона, частоты формант, энергия и длительность. Для гласных звуков были построены формантные треугольники с вершинами, соответствующими гласным [a], [u] и [i] для значений формант F_1, F_2, после чего сравнивались площади формантных областей. Для всех детей с нарушениями в области аутизма голос и речь характеризуются высокими значениями частот основного тона, аномальным спектром и ярко выраженными высокими формантными частотами. Ударные гласные в произнесении слов детьми (две группы детей: в норме и с патологией), произнесенных в дискомфортных условиях, имеют более высокие значения основного тона и третьих (эмоциональных) формант, чем произнесенных в нормальном, комфортном состоянии. У детей с аутизмом обнаружены более высокие значения основного тона в спонтанной речи, чем в тестах на повторение речевых единиц. Полученные результаты являются первым шагом в направлении разработки био-маркеров на основе речи для ранней диагностики аутизма.

Б. Геразов и Ф.Н. Гарнер (Скопье) в докладе «Модель генерирования основного тона типа “агонист-антагонист”» отмечают, что просодия – это феномен, который имеет решающее значение для различных областей речевых исследований, что подчеркивает особую важность разработки надежной, помехоустойчивой просодической модели. Класс интонационных моделей на основе физиологии генерирования основного тона особенно привлекателен для присущей им многоязычной поддержки. Эти модели опираются на точную модель активации мышц. Как правило, используется мышечная модель 2-го порядка типа “пружина–амортизатор–масса” (SDM). Однако недавние исследования показали, что модель SDM недостаточна для адекватного моделирования динамики мышц. Модель 3-го порядка предлагает более точное представление динамики мышц, но при этом она продемонстрировала слабозатухающие

колебания при использовании физиологически правдоподобных параметров мышц. В данной работе авторы предлагают модель генерирования основного тона типа “агонист–антагонист” (A2P2), которая и подтверждает, и поясняет результаты использования моделей с критическим затуханием более высокого порядка при моделировании интонации.

Исследование, представленное в докладе “Социолингвистическая вариативность русской разговорной речи” (Н. Богданова-Бегларян, Т. Шерстинова, О. Блинова, Г. Мартыненко (Санкт-Петербург)), проведено в рамках социолингвистического проекта, направленного на описание повседневной русской речи и анализ специфических особенностей ее использования различными социальными группами. Исследование основано на материале корпуса ORD, содержащего аудиозаписи повседневного общения. Цель данного исследования – выявление лингвистических параметров, по которым разница в речи между различными социальными группами является наиболее очевидной. Был создан и полностью аннотирован подкорпус, состоящий из аудиофрагментов разговорной речи 12 респондентов (6 мужчин и 6 женщин; 4 представителя для каждой возрастной группы; представители различных профессиональных и статусных групп) общей продолжительностью в 106 мин с одинаковыми условиями коммуникации. Для каждой социальной группы дано количественное описание ряда лингвистических параметров на фонетическом, лексическом, морфологическом и синтаксическом уровнях. Наибольшее различие между социальными группами наблюдалось в темпе речи, фонетических сокращениях, лексических предпочтениях и синтаксических нарушениях. Исследование показало, что различия между возрастными группами являются более значимыми, чем между гендерными группами.

В докладе “Автоматическое реферирование спонтанной речи” (А. Беке и Г. Шашак (Будапешт)) рассматривается проблема реферирования спонтанной речи. Речь преобразуется в текст с использованием автоматической системы распознавания, затем сегментируется на речевые фрагменты, связанные в лексические группы. Сравниваются выполненные человеком и автоматической системой распознавания части высказываний, рассматриваемые как части предложений. Полученные фрагменты текста подвергаются стилистическому анализу на основе просодии. Полученные подобные предложениям единицы анализируются синтаксическим анализатором с целью автоматического выбора предложений для реферирования. Предварительно обработанные предложения ранжируются на основе тематических терминов и места предложения. Тематический термин выражается двумя способами: TF-IDF и Скрытым семантическим индексированием. “Оценки” предложения рассчитываются как линейное сочетание оценки тематического термина и оценки места предложения.

Для создания реферата выбираются 10 наилучших “кандидатов” в виде наиболее информативных обобщающих предложений. Характеристики системы показали сопоставимые результаты (полнота: 0,62, точность: 0,79 и F-показатель 0,68) с использованием лексического анализа на основе просодии.

Авторы доклада “Создание речевого корпуса для исследований в области межъязыковой просодии” М. Сешуски, Б. Геразов, Т.Г. Чарпо и др. (Нови Сад, Будапешт и др.) указывают на то, что поскольку просодия устного высказывании несет в себе информацию о функции дискурса, значимости и модальности говорящего, просодические модели и модули генерирования просодии играют решающую роль в системах преобразования текста в речь (TTS), в частности тех, в которых стоит задача не только естественно звучать, но и демонстрировать эмоции или конкретное намерение говорящего. Передача просодии в речи при преобразовании текста в речь является относительно новым объектом исследований, значение которых приобретает все бóльшую важность. При этом одним из наиболее перспективных направлений исследования является выявление и обработка важных событий, то есть случаев, которые являются результатом не синтаксических ограничений, а скорее продуктом воздействия семантического или прагматического уровня. В докладе представлены методики и основные принципы создания многоязычного речевого корпуса, содержащего просодически насыщенные предложения, направленные на обучение моделей статистической просодии для многоязычной передачи просодии в контексте выразительного синтеза речи.

В. Верходанова, А. Ронжин, И. Кипяткова и др. (Санкт-Петербург) в своем докладе “Корпус HAVRUS: высокоскоростная запись аудиовизуальной русской речи” представили программно-аппаратный комплекс для создания аудиовизуальных речевых баз данных с высокоскоростной камерой и динамическим микрофоном. Описана архитектура разработанного программного

обеспечения, а также некоторые детали собранной аудиовизуальной речевой базы данных на материале русского языка. Разработанное программное обеспечение реализует синхронизацию аудио- и видеоканалов, а также учитывает асинхронность аудио и визуальных форм речевых модальностей. Собранный корпус включает в себя записи 20 носителей русского языка и предназначен для дальнейшего исследования и экспериментов в области аудиовизуального распознавания русскоязычной речи.

И. Мпорас, С. Сафави и Р. Сотуде (Хатфилд) в докладе “Повышение надежности верификации говорящего на базе совмещения текстозависимых и текстонезависимых модальностей” представили методику объединения текстозависимых и текстонезависимых подходов к решению задач верификации говорящего. Совмещение подходов осуществляется на уровне оценки, получаемой на базе верификации говорящего с использованием алгоритмов обучения нескольких систем. Для того чтобы улучшить характеристики этого процесса, применялась кластеризация данных до этапа классификации. Экспериментальные результаты показали, что совмещение двух режимов работы улучшает характеристики верификации говорящего.

Доклад “Исследование параметров речевого сигнала, отражающих истинность передаваемой информации” (В. Будков, И. Ватаманюк и др. (Санкт-Петербург)) включает обзор существующих методов диагностики истинности передаваемой информации. Делается вывод о целесообразности реализации этой функции в полимодальных инфокоммуникационных системах. Рассматриваются параметры речевого сигнала, которые отражают истинность передаваемой информации. Представлены результаты испытаний разработанного программного обеспечения. На основании проведенного исследования сформулирован вывод о возможности установления истинности передаваемой информация в процессе межличностного общения, а также о целесообразности разработки правил принятия решения.

Р. Потапова и Л. Комалова (Москва) в докладе “Мультимодальное восприятие агрессивного поведения” представили результаты сравнительного перцептивно-слухового и перцептивно-зрительного анализа экспериментальных выборок на материале русского, английского, испанского и татарского языков, соотносящихся с эмоционально-модальным комплексом состояния “агрессия”. Описываются статистически достоверные различия между слуховым и зрительным видами восприятия агрессивного (физического и вербального) поведения под влиянием таких факторов, как эмоционально-модальное состояние реципиента и язык общения.

В докладе “Об индивидуальной полиинформативности речи и голоса применительно к атрибутике говорящего (криминалистический аспект)” Р. Потаповой и В. Потапова (Москва) рассматривается роль воспринимаемых на слух характеристик речи и голоса говорящего применительно к атрибутике его индивидуальных особенностей. Исследование посвящено изучению того, насколько правильно слушающие могут определять набор индивидуальных особенностей говорящего: вербальных, паравербальных, экстравербальных, физиологических, антропометрических, физических, эмоциональных, социальных и т.д. исключительно по голосу и речи. Основная задача исследования заключалась в определении того, какие характеристики говорящего могут определяться на слух: универсальные, групповые или идиосинкразические. Для слухового анализа были разработаны специальные анкеты и проанализированы два типа речи и голоса: интериндивидуальные и интраиндивидуальные. Конечная цель исследования заключалась в разработке метода идентификации говорящего по модели “line-up” для русской речи применительно к показаниям “свидетеля”, владеющего слуховой информацией о голосе и речи подозреваемого и не владеющего зрительной информацией.

Р. Потапова и В. Потапов (Москва) в своем докладе “Многокомпонентная атрибутика социально-сетевого дискурса” представили результаты исследования, касающиеся соотношения между некоторыми типами депривации и ее вербальными, паравербальными и невербальными детерминантами. В докладе представлена аннотированная база данных социально-сетевого дискурса, сформированная на материале диалогов и полилогов в интернете, а также аннотированная база данных, полученная на материале видеохостингов YouTube.com, Skype и ok.ru. Аннотированная база данных, предназначенная для системы принятия решений и автоматизированного анализа русскоязычного устного и письменного дискурса социальных сетей в интернете послужила основой для последующего анализа с установкой на определение вариантов социально-сетевого дискурса (ССД) с учетом формы, функции, монотематичности, политематичности, одновекторности, многовекторности и т.д.

В докладе “Аннотирование речевых актов в условиях повседневной коммуникации на материале русской разговорной речи ORD” Т. Шерстиновой (Санкт-Петербург) описаны принципы аннотирования, разработанные для разметки речевых актов в корпусе “Один речевой день” (ORD) русскоязычной бытовой речи. При этом особое внимание уделяется категориям и подкатегориям речевых актов, которые выделяются в ORD. Аннотирование речевых актов является частью аннотирования корпуса, который включает разметку макро- и микроэпизодов речевого общения. Речевые акты аннотируются с учетом четырех уровней: (1) орфографической транскрипции с учетом информации о синтагматических и фразовых границах, (2) кода говорящего, (3) основной категории речевого акта и (4) его подкатегории. Практическая апробация предложенной схемы аннотирования выполнялась на материале 6-и макроэпизодов бытовой коммуникации (2250 речевых актов). Аннотирование корпуса ORD дает возможность изучать русский бытовой дискурс с позиции речевых актов, языковых свойств и закономерностей реализации речевых актов разных типов.

Доклад «Речевой диалог как часть интерактивных систем “Человек–машина”» Р. Потаповой (Москва) посвящен одной из наиболее важных особенностей технологий распознавания устной речи – анализу речевого сигнала, который включает предварительную обработку, обработку и распознавание речи с опорой на ряд параметров. В настоящем докладе представлен один из методов разработки систем “человек–машина”, который базируется на анализе и обнаружении в слитной речи значений формант F_{ni}. Автор подчеркивает, что существует много способов проведения акустического анализа, но важнейшими из них остаются функции акустико-фонетического распознавания на фонемном и просодическом уровнях, что считается одним из классических методов распознавания речи.

Р.К. Потапова
доктор филологических наук,
профессор,
Московского государственного лингвистического университета,
Россия, 119034, г. Москва, Остоженка 38
rkpotapova@yandex.ru

В.В. Потапов
доктор филологических наук,
старший научный сотрудник
филологического факультета Московского государственного университета
им. М.В. Ломоносова,
Россия, 119991, г. Москва, Ленинские горы,
д. 1, стр. 51
volikpotapov@gmail.com

Rodmonga K. Potapova
Doctor of Philological Sciences,
Professor at the State Linguistic University,
Ostozhenka 38,
Moscow, 119034, Russia,
rkpotapova@yandex.ru

Vsevolod V. Potapov
Doctor of Philological Sciences,
Senior Researcher at the
Lomonosov Moscow State University,
1-51 Leninskie Gory,
Moscow, 119991, Russia
volikpotapov@gmail.com

Дата поступления
материала в редакцию
14 ноября 2016 г.
Received by Editor
on November 14, 2016