XVI Международная конференция SPECOM'2014 «Speech and computer» («Речь и компьютер»)

Родмонга Кондратьевна Потапова[а, @], **Всеволод Викторович Потапов**[б]

[а] Московский государственный лингвистический университет, Москва, 119034, Россия; [б] МГУ им. М. В. Ломоносова, Москва, 119991, Россия; [@] rkpotapova@yandex.ru

5—9 октября 2014 г. в г. Нови-Сад (Сербия) на базе факультета технических наук (FTN) Университета г. Нови-Сад состоялась XVI Международная конференция SPECOM'2014 «Speech and computer» («Речь и компьютер»).

Конференция «Речь и компьютер» является мероприятием, регулярно проводимым с момента первой конференции SPECOM в 1996 г., которая проходила в Санкт-Петербурге. Это конференция с уже сложившимися традициями, которая привлекает исследователей прежде всего в области компьютерной обработки речи: автоматизированного распознавания и понимания речи, обработки речевых сигналов, мультимодальной обработки речи, анализа—синтеза речи и др. По мнению специалистов, международная конференция SPECOM является идеальной платформой для обмена современными ноу-хау, особенно для лингвистов-прикладников, работающих со славянскими и другими высокофлективными языками, а также с более или менее ресурсно обеспеченными языками.

Конференции SPECOM организовывались попеременно в Санкт-Петербурге и Москве: Санкт-Петербургским институтом информатики и автоматизации Российской академии наук (СПИИРАН) и Московским государственным лингвистическим университетом (МГЛУ). Кроме того, она проводилась в 1997 г. Научно-исследовательским институтом вычислительной техники (г. Клуж-Напока, Румыния), в 2005 г. — Университетом г. Патры (Греция), в 2011 г. — Казанским (Приволжским) федеральным университетом (Российская Федерация, Республика Татарстан), в 2013 г. — Университетом Западной Богемии и факультетом прикладных наук и кибернетики (Пльзень, Чешская Республика). В 2014 г. было принято решение продолжить традицию. Место проведения конференции SPECOM в этом случае было выбрано с учетом интересов сообщества носителей славянских языков, кроме того, там, где исследования по обработке речи имеют давнюю традицию.

Под руководством Московского государственного лингвистического университета и Санкт-Петербургского института информатики и автоматизации РАН конференция SPECOM 2014 г. (16-я по счету) проходила параллельно с 10-й конференцией DOGS 2014 г. («Обработка цифровой речи и изображений»), мероприятием, проводимым два раза в год и традиционно организованным на факультете технических наук Университета г. Нови-Сад. Параллельная работа двух конференций позволила обеспечить их участникам возможность присутствовать на заседаниях обеих конференций. Опытные и начинающие исследователи в области обработки речи и связанных с этим областях знаний имели возможность непосредственного общения и обмена опытом, с одной стороны, и обмена мнениями по поводу новых «прорывных» идей, с другой стороны. Было принято решение объединить работу обеих конференций для презентации и обсуждения пленарных докладов. На совместных пленарных заседаниях были заслушаны доклады А. Петровского (Белоруссия), Э. Брина (Великобритания), Г. Немета (Венгрия).

В работе конференции SPECOM'2014 принимали участие докладчики из Великобритании, Венгрии, Белоруссии, Сербии, Германии, Японии, Чехии, Словакии, Мексики, России, США, ЮАР. Было отобрано и представлено в общей сложности 58 докладов. Число авторов докладов — 142.

Наряду с пленарными заседаниями работа конференции проходила по следующим секциям: распознавание и понимание речи, системы речевой безопасности, речевого диалога (человек—машина), анализа—синтеза речи, аудиовизуальной коммуникации. В постер-презентации было представлено 23 доклада.

В заключительной части конференции был организован круглый стол, посвященный проблемам развития и применения речевых технологий в современном мире.

XVI International conference SPECOM'2014 «Speech and computer»

Rodmonga K. Potapova[a, @], **Vsevolod V. Potapov**[b]

[a] Moscow State Linguistic University, Moscow, 119034, Russia; [b] Lomonosov Moscow State University, Moscow, 119991, Russia; [@] rkpotapova@yandex.ru

К началу работы конференции доклады были опубликованы, как и в случае с предыдущей конференцией, издательским домом Springer, публикации которого входят в систему Scopus и Web of Science. В полном объеме материалы представлены в трудах конференции [Proceedings 2014].

Доклад Э. Б р и н а (Великобритания) «Формирование набора голосов с учетом их экспрессивности, предназначенных для использования в системах синтеза “текст—речь”» посвящен проблеме преобразования текста в звучащую речь. Преобразование «текст—речь» (TTS — «text-to-speech») традиционно рассматривается в качестве компонента «черного ящика», где стандартные имеющиеся в наборе голоса соответствуют, как правило, профессионально подготовленному нейтрально-разговорному стилю речи. Для коммерчески наиболее престижных языков может быть предложено множество разнообразных голосов в похожем разговорном стиле. Заказчик, желающий использовать в коммерческих целях систему TTS, как правило, выбирает один из этих голосов. Единственной альтернативой является выбор в пользу решения «голос на заказ». В этом случае клиент платит за конечный продукт — создание, например, рекламы на базе преобразования «текст—речь» с использованием отобранного «голосового источника». Подобный подход позволяет реализовать некоторую предварительную «настройку» сценариев (скриптов) на используемый голос. В некоторых случаях могут быть добавлены определенные элементы сценариев, что необходимо для обеспечения большего охвата элементов сценария в области, указанной заказчиком. При подготовке конечного материала могут быть включены также специальные фразы, которые содержат примеры идеального произнесения конкретных фрагментов текста. При таком подходе процесс записи строго контролируется, а стандартные сценарии перерабатываются не с нуля, а расширяются. Подход «черный ящик» к TTS позволяет создать системы, которые удовлетворяют потребностям большого числа заказчиков.

Последние достижения в области применения систем «текст—речь» изменили мнение людей о том, *как* должен звучать и влиять на человека «компьютерный» голос. Оказалось, что для системы TTS (особенно в коммерческих сферах применения) гораздо важнее представить конкретное лицо, которое соответствует максимальному достижению цели коммуникации. Практика показала, что подобные системы требуют более яркого, оптимистичного и выразительного голоса. Подхода «черный ящик» уже недостаточно. Голоса для высокопроизводительных «посредников» речевого общения в настоящее время явно «предназначены» для удовлетворения потребностей таких приложений. Эти голоса одновременно и выразительны, и легки, а также образуют контраст по сравнению с более «консервативными» голосами, используемыми традиционно на мировом рынке. Данный доклад в рамках проекта Nuance R&D посвящен описанию нового подхода к особому типу речевого синтеза TTS с использованием речевых образцов экспрессивного разговорного стиля.

Большой интерес вызвал пленарный доклад Г. Н е м е т а (Венгрия) «Нерешенные проблемы в области речевых технологий», в котором утверждается следующее: несмотря на то, что в последнее время наблюдается значительный прогресс в области использования и принятия в производство речевых технологий, в ряде развитых стран по-прежнему существуют серьезные пробелы, которые не позволяют большинству возможных пользователей найти конкретные решения, связанные с применением речевых технологий. В докладе перечислены некоторые из этих пробелов (нерешенных проблем) и предлагаются пути их ликвидации. Возможно, что наиболее значительным расхождением является мышление разработчиков программного обеспечения по типу «черный ящик», которые полагают, что ввод текста в систему преобразования текста в речь (TTS) приведет к голосовому продукту на выходе, который имеет отношение к данному контексту применения. Применительно к автоматическому распознаванию речи (ASR) разработчики ждут получения точной транскрипции текста на выходе, включая знаки препинания. При этом не принимается во внимание, что даже люди находятся под сильным влиянием априорного знания контекста, партнеров по коммуникации и т. д. По мнению докладчика, знания в области семантического моделирования все еще находятся в зачаточном состоянии. Для создания успешных приложений исследователи речевых технологий должны найти пути для создания «встроенного» априорного знания в среде приложений, адаптировать свои технологии и интерфейсы для данного сценария. Например, разборчивость и изменчивость скорости речи являются наиболее важными параметрами оценки TTS для слабовидящих пользователей. В то же время для информационных систем на железнодорожных станциях необходимы «человекоподобные» объявления с обычным темпом и в разговорном стиле. Увеличивающийся разрыв наблюдается между «большими» и «малыми» языками / рынками. Еще один пробел — между закрытыми и открытыми прикладными средами. Например, вряд ли существует мобильная операционная

система, которая обеспечивает переадресацию TTS при непосредственном телефонном разговоре, что является основной потребностью реабилитационных приложений для людей, испытывающих проблемы с речью. В этой ситуации может помочь создание открытой платформы, где «мелкие» и «крупные» игроки на поле могут одинаково внедрять свои средства / решения при надлежащем качестве продукта и больших доходах. В докладе приведены некоторые примеры попыток устранения указанных пробелов.

В докладе П. Чистикова, Д. Захарова и А. Таланова (Санкт-Петербург) «Повышение качества синтеза речи с использованием базы данных аудиокниг» представлен подход к повышению качества синтезированной речи с использованием базы данных, полученных на материале аудиокниг. Данные включают речевой материал, прочитанный одним диктором. Звучащий материал сравнивался с соответствующими письменными текстами. Основные проблемы исследования связаны со следующими факторами: а) запись проведена в разное время в разных акустических условиях; б) диктор читает текст с разной интонационной и акцентно-ритмической вариативностью, что ведет к большей вариативности голосовых параметров. Кроме того, автоматические методы маркировки звукового файла приводят к большему числу ошибок из-за большой вариативности составляющих базы данных, особенно при наличии расхождений между текстом и соответствующими звуковыми файлами. Вышеуказанные проблемы существенно влияют на качество синтеза речи, поэтому надежный метод их решения так необходим для голосов, созданных с использованием аудиокниг. Подход, описанный в докладе, основан на статистических моделях голосовых параметров и специальных алгоритмах конкатенации и модификации речевых сегментов. Перцептивно-слуховое тестирование в значительной степени повышает качество синтезированной речи.

В выступлении М. Кото-Хименеса, Дж. Годдарда-Клоуза и Ф. М. Мартинез-Ликона (Мексика) «Оценка качества синтеза речи на основе HMM с использованием акустического анализа гласных» описана синтезированная речь, которая получена с использованием скрытых моделей Маркова (HMM). При сравнении с естественной речью данная синтезированная речь часто характеризуется наличием глухого тембра, чему есть несколько причин: некоторые тонкие характеристики естественной речи удаляются, минимизируются или существуют в скрытом виде; траектории изменения параметров получаемой на выходе речи становятся «сверхсглаженными» вариантами речевых сигналов. Это означает, что каждый синтетический голос, созданный системой на основе HMM, должен быть проверен на качество речи. Как правило, требуется дорогостоящее субъективное исследование (эксперимент), поэтому было бы интересно разработать альтернативные подходы. В докладе рассматриваются девять акустических параметров, связанных с дрожанием (джиттер) и мерцанием (шиммер), а также их статистическая значимость как объективных измерений качества синтетической речи.

Доклад В. Делича, М. Сечуйски, Н. Вуйнович Седлар, Д. Мишковича, Р. Мака и М. Боянича (Сербия) «Как речевые технологии могут помочь людям с ограниченными возможностями» посвящен проблемам мультимодальной коммуникации «человек—машина». В речевой коммуникации «человек—машина», как правило, не используются невербальные средства коммуникации (например, ручная жестикуляция), а также паравербалика (например, окулесика). И человек, и машина используют вербалику, что может помочь людям с физиологическими или патологическими отклонениями. Помимо слабовидящих людей и людей с физическими недостатками, речевые технологии могут помочь людям с нарушениями речи и слуха, а также пожилым людям. Доклад представляет собой обзор речевых технологий, которые полезны для людей с различными ограничениями: физиолого-физическими отклонениями. Так, например, технологии преобразования письменного текста в звучащую речь TTS (текст—речь—технологии) применимы в случаях ослабленного зрения, то есть в ситуации замены зрительного канала слуховым. Автоматическое распознавание устной речи может применяться в ситуациях распознавания голосовых команд с малым по объему словарем в условиях смарт-жилища. Автоматизированное распознавание говорящего и его эмоционального состояния по голосу и речи может способствовать усовершенствованию диалога «человек—машина».

В совместном докладе Х. Эхизенья, К. Араки, Ю. Учида (Япония) и Э. Хови (США) «Метод автоматического постредактирования с использованием базы знаний по переводоведению, полученной путем статистического накопления общих интуитивно выделенных языковых фрагментов» предложен новый метод постредактирования для текстов — результатов машинного перевода. Метод предполагает использование при постредактировании базы знаний, полученной от перевода на основе параллельного рассмотрения лингвистических корпусов вне зависимости от лингвистического инструментария. Правила перевода, которые

приобретаются на основе интуитивного суммарного фрагментосодержащего континуума (Intuitive common parts continuum, ICPC), могут применяться при сопоставлении целостной структуры исходного и целевого высказываний без дополнительного лингвистического анализа. Более того, предлагаемый метод помогает получить более качественные переводы путем параллельного применения правил перевода и результатов перевода ICPC, полученного с использованием статистического накопления общих фрагментов машинного перевода. Полученные экспериментальным путем данные подтверждают эффективность применения предлагаемых правил перевода на базе статистического накопления общих интуитивно выделенных языковых фрагментов.

В докладе «Исследования региолектов на основе корпусов текстов: Казанский регион» К. Галиуллин, А. Гизатуллина, Е. Горобец, Г. Каримуллина, Р. Каримуллина и Д. Мартьянов (Казань) рассмотрели специфику создания и использования электронных корпусов, созданных в Казани (Казанский (Приволжский) федеральный университет). Состав корпуса: словарь и текстовый корпус «Казанский край: язык русских документов (XVI—XVII вв.)», электронный корпус русских диалектов Казанского региона (XIX—XXI вв.), электронный корпус русских текстов, связанных с Казанским регионом / Республикой Татарстан (XX—XXI вв.). В докладе представлен информационный потенциал содержащихся в корпусах электронных ссылок с аннотационными данными и специфическими характеристиками Казанского региолекта (территориального варианта русского языка, используемого в Казанском регионе, который хорошо известен как регион межъязыковых контактов).

В докладе Б. Яковлевич, А. Ковачевича, М. Сечуйски и М. Маркович (Сербия) «Банк деревьев зависимостей для сербского языка: Начальные эксперименты» представлена разработка банка деревьев зависимостей для сербского языка, предназначенного для различных применений в области обработки естественного языка, прежде всего в области понимания естественного языка в рамках диалога «человек—машина». Банк данных создан с учетом добавления синтаксических аннотаций в Текстовый корпус сербского языка AlfaNum с метками частей речи (part-of-speech, POS). Аннотирование осуществляется в соответствии со стандартами, установленными Пражским банком дерева зависимостей, который был принят в качестве основы при разработке банков деревьев для некоторых родственных языков в данном регионе. Первые эксперименты по парсингу (синтаксическому анализу) на основе грамматики зависимостей на материале уже аннотированной части корпуса, содержащей 1 148 предложений (7 117 слов), показали относительно низкую точность синтаксического анализа, как и ожидалось от банка деревьев такого размера в ходе проведения предварительных экспериментов.

В выступлении И. Йокича, С. Йокича, В. Делича и З. Перича (Сербия) «Влияние эмоциональной речи на автоматическое распознавание говорящих — эксперименты с использованием базы речевых данных GEES» описан эксперимент с использованием устройства автоматической идентификации говорящих по базе данных эмоциональной речи. Устройство автоматической идентификации говорящих основано на применении кепстральных коэффициентов значений частоты основного тона (в мелах) как признаков речи говорящего и ковариационных матриц модели говорящего. Модели формируются с использованием одного предложения эмоционально нейтральной речи для каждого говорящего. Другие предложения из той же базы речевых данных, в том числе нейтральные, а также характеризующие четыре эмоциональных состояния: счастье, страх, печаль и гнев, — используются для дальнейшего тестирования. Целью исследования является изучение влияния эмоциональной речи на точность автоматической идентификации говорящих.

А. Карпов, И. Кипяткова (Санкт-Петербург) и М. Железны (Чехия) («Условия записи аудиовизуальных речевых корпусов с микрофоном и высокоскоростной камерой») представили новое программное обеспечение для записи аудиовизуальных речевых корпусов с высокоскоростной видеокамерой (JAI Pulnix RMC-6740) и динамическим микрофоном (Октава МК-012), рассмотрев архитектуру программного обеспечения, разработанного для записи аудиовизуального корпуса русской речи, что помогает синхронизировать и объединять слияние аудио- и видеоданных, записанных с помощью отдельных датчиков. Программное обеспечение обнаруживает речь в аудиосигнале и сохраняет только информативные речевые фрагменты, отбраковывая неинформативные сигналы. При этом также учитывается и обрабатывается естественная асинхронность аудиовизуальных речевых модальностей.

В докладе К. Килгура и А. Вайбеля (Германия) «Нейронно-сетевая система поиска по ключевым словам, предназначенная для телефонной речи» предложена система поиска по ключевым словам на основе «нейронной сети» (NN), разработанная по программе IARPA Babel для разговорной телефонной речи.

Использование общего показателя оценки поиска по ключевому слову, т. е. «*реально взвешенного значения*» (ATWV), позволяет утверждать, что NN-система поиска по ключевому слову может достичь показателей, схожих с более сложной и более медленно функционирующей системой распознавания речи на основе «*гибридной глубокой нейронной сети — скрытой модели Маркова*» (DNN-CMM Hybrid) без использования декодера HMM или языковой модели.

В сообщении Д. Кочарова, П. Скрелина и Н. Вольской (Санкт-Петербург) «Модели нисходящей частоты основного тона F_0 в русской речи» описаны разновидности нисходящей конфигурации частоты основного тона (F_0) для русской речи. Перед исследователями стояла задача: определить на базе корпуса русской речи, сформированного с использованием чтения текстов дикторами-профессионалами, разновидности понижения значений частоты основного тона (F_0) и на этой основе выявить существующие в русской речи модели понижения F_0, связанные с различными интонационными контурами, чтобы подтвердить или опровергнуть зависимость типа «понижение F_0—длительность», обнаруженную в других языках. Полученные результаты подтверждают прямую связь между понижением F_0 и длительностью высказывания. В то же время обнаруживается сильная зависимость понижения F_0 от общего интонационного рисунка высказывания: так, модель завершенного утвердительного повествования характеризуется более крутым «наклоном» по сравнению с незавершенным повествованием. Вопросительные предложения, характеризующиеся подъемом основного тона, не обнаружили понижения основного тона на участке предтакта. Результаты, таким образом, позволяют предполагать наличие отдельных индивидуальных стратегий в процессе предварительного планирования вида падения основного тона в интонационной фразе.

В докладе И. Кралевски, М. П. Биссири, Г. Стречи и Р. Хоффмана (Германия) «Анализ и синтез глоттализации в английской речи с немецким акцентом» описан анализ и синтез глоттализации в английской речи носителей немецкого языка. Глоттализация в начале слога (слова) отмечалась вручную на материале фрагмента корпуса английской речи носителей немецкого языка. Для каждого глоттализованного сегмента синтезировались значения нормированной по времени F_0 и «низкоэнергетические» контуры. Кроме того, был проведен анализ на множествах контура F_0. Центроидные контуры кластеров использовались для реконструкции контуров в экспериментах по повторному синтезу. Прототипные контуры интонации и интенсивности накладывались на неглоттализованные гласные в начале слов с целью синтезирования «скрипучего» голоса. Эта процедура позволяла автоматически создавать речевые стимулы, которые могли бы быть использованы в перцептивных экспериментах для проведения фундаментальных исследований в области глоттализации. Глоттализация рассматривалась в двух разновидностях: твердый приступ и «скрипучий голос» — твердый приступ как результат резкого смыкания и размыкания голосовых связок и «скрипучий голос» как своеобразный перцептивно-слуховой феномен, являющийся результатом нерегулярных, низкочастотных вибраций голосовых связок.

В выступлении Е. Красновой и Е. Булгаковой (Санкт-Петербург) «Использование речевых технологий в системах компьютерного обучения языку» рассмотрены способы применения автоматического распознавания речи (ASR) и технологии преобразования текста в речь (TTS) для систем обучения языку при помощи компьютера (CALL). Речевые технологии могут эффективно использоваться для таких методических целей, как отработка произношения, овладение навыками коммуникации, проверка словарного запаса студентов и навыки аудирования (понимания речи на слух). Несмотря на некоторые ограничения, в настоящее время в обучении можно применять различные типы речевых технологий, что является эффективным средством упрощения реализации процесса обучения. В докладе представлена интеграция ASR в систему CALL, разработанная Центром речевых технологий (Санкт-Петербург).

В докладе Е. Ляксо, А. Григорьева, А. Куразовой и Е. Огородниковой (Санкт-Петербург) «“INFANT.MAVS” — мультимедийная модель для изучения когнитивного и эмоционального развития детей» описана модель мультимодальной сенсорной среды «INFANT.MAVS», которая включает две базы стимулов с различной сложностью восприятия: а) простые стимулы (зрительные, звуковые, тактильные и графические) и б) набор сложных стимулов, синтезированных как комбинации простых. Программное обеспечение включает компонент управления базами данных и саму базу данных. Компонент управления создается с помощью Microsoft Visual Basic v.6.0 и предназначен для работы на операционных системах Windows. Результаты испытаний модели показали, что стимулы вызывали реакцию у детей — сосредоточенное внимание, вокализацию, улыбки и попытки повторить звуки; у взрослых они вызывали положительные эмоции. Полученные данные позволили сделать вывод, что модель

«INFANT.MAVS» соответствует целям, которые были поставлены разработчиками.

В докладе Л. Мохаси (ЮАР), М. Сечуйски, Р. Мака (Чехия) и Т. Нислера (ЮАР) «Сравнение двух подходов к моделированию просодии в языке сесото и сербском языке» речь шла о том, что точное прогнозирование просодических особенностей является одной из важнейших задач в рамках разработки системы преобразования «текст—речь», что особенно значимо для языков с ограниченными ресурсами и сложной лексической просодией. Авторы считают, что для того, чтобы синтезированная речь имела естественно звучащий интонационный контур, следует использовать адекватную просодическую модель. В данном исследовании сравнивается модель Фудзисаки и просодическое моделирование на основе HMM в контексте преобразования «текст—речь» для двух неродственных языков с богатыми просодическими системами: сесото, тонального языка семьи банту, и сербского, южнославянского языка с тоническим ударением. Результаты экспериментов показали, что для обоих языков использование модели Фудзисаки дает лучшие результаты, чем использование модели HMM при моделировании интонационных контуров высказываний. Модель Фудзисаки разработана специально для анализа значений частоты основного тона (F_0) естественного высказывания и ее сегментации на основные компоненты, которые совместно образуют контур F_0, похожий на исходный оригинальный контур F_0. К числу основных компонентов относятся: частота основного тона, часть фразы, которая включает как более замедленные изменения в контуре F_0, так и более быстрые изменения в F_0. Тоновые команды модели Фудзисаки являются индикатором тех или иных тонов в высказывании. Метод был впервые предложен Фудзисаки и его сотрудниками в 70—80-х гг. в качестве аналитической модели, описывающей изменения частоты основного тона.

В выступлении Э. Пакочи, Н. Яковлевича, Б. Поповича, Д. Мишковича и Д. Пекара (Сербия) «Идентификация говорящего с использованием скрытых марковских моделей для конкретных звукотипов» представлено описание системы идентификации говорящего на основе использования скрытой марковской модели для конкретного звукотипа в сочетании с гауссовой моделью (моделью совокупности нормальных распределений). Использование данного подхода связано с тем, что система HMM на основе конкретного звукотипа может моделировать временны́е вариации, что обеспечивает возможность рассмотрения десятков конкретных звуков, а также ведет к эффективному отбраковыванию значений. Эффективность системы была оценена на речевой базе данных, которая содержит речевые высказывания 250 говорящих — носителей сербского языка. Предлагаемая модель сравнивается с системой, основанной на гауссовой модели (модели совокупности нормальных распределений) и универсальной модели. Разработанная авторами модель продемонстрировала значительное повышение точности идентификации.

В. Потапов (Москва) («Речевые ритмические модели в славянских языках») представил описание сопоставительного экспериментального акустического исследования субъективных и объективных характеристик ритмической организации речи, проводимого на материале чешского, болгарского и русского языков. Настоящее исследование подтвердило справедливость гипотезы о существовании иерархии факторов, определяющих ритмический рисунок в рассмотренных славянских языках. Результаты акустического анализа выявили фонетическую специфику ритмических структур (РС) и ритмических схем синтагм (РСС), которая определяется фонетической структурой ударения в РС, реализуемого в исследуемых языках различными средствами: определенными комбинациями просодических характеристик гласных на границах РС в чешской речи, динамической составляющей в болгарской речи, а также спектральной и временно́й компонентами в русской речи.

Доклад Р. Потаповой, А. Собакина и А. Маслова (Москва) «О возможности идентификации говорящего с использованием Skype-канала (на основе акустических параметров)» посвящен описанию метода идентификации говорящего по речевому сигналу в системе Skype (в случае искусственной модификации внешности личности) на базе импульсного преобразования речи. В ходе эксперимента для сравнения исследовались речевые сигналы (гласные русского языка), записанные в безэховой камере, и те же речевые сигналы, прошедшие через канал передачи IP-телефонии Skype. И в том и в другом случае привлекались одни и те же дикторы. Цель исследования — определение индивидуальных особенностей функционирования голосового источника говорящего (фонации) в зависимости от канала передачи речевого сигнала для установления возможности идентификации говорящего по голосовым характеристикам в информационных системах. Результаты позволили установить ряд особенностей при порождении речи в условиях тракта IP-телефонии системы Skype, а также перспективность разрабатываемого метода.

В сообщении Р. Потаповой и Л. Комаловой (Москва) «Об основных подходах

к формированию аннотированных баз данных семантического поля “агрессия”» описаны основные критерии, использованные при разработке аннотированных баз данных семантического поля «агрессия», а также русских и английских цифровых полнотекстовых баз данных СМИ, содержащих вербальные составляющие семантического поля «агрессия». Каждая база данных включала 120 вручную аннотированных текстовых блоков, где представлены лексический, семантический и прагматический уровни языка. Каждый текст сопровождается специальными указателями и локальным словарем семантического поля «агрессия». Базы данных предназначены для научных исследований в области прикладного речеведения: для автоматизированных систем обучения по Интернету, дальнейшей разработки поисковых систем, включающих семантическое поле «агрессия» и др.

Доклад Р. Потаповой и В. Потапова (Москва) «Ассоциативный механизм восприятия иностранной разговорной речи (судебно-криминалистический аспект)» был посвящен проблеме восприятия на слух иностранной разговорной речи с целью формирования единиц интерферирования речи для сегментного состава. Эксперимент включал декодирование на слух ad-hoc материала иностранной разговорной речи, который был специально разработан и фонетически сбалансирован. В исследовании особое внимание уделяется слуховому восприятию, обусловленному межъязыковой интерференцией. В этой ситуации слушающий должен использовать различные наборы воспринимаемых образцов фонетических единиц. Предполагается, что в случае декодирования на слух высказываний иностранной разговорной речи слушатели построят фонемную, слоговую, ритмическую и просодическую модели речевых высказываний на родном языке, а также модели звуковых и интонационных расхождений родного и воспринимаемого неродного языка, которые могут быть использованы в дальнейшем для построения системы line-up, включающей образцы интерферированной речи и их релевантных признаков. Проблема восприятия на слух разговорной речи связана с проблемами распознавания голоса и речи в области судебно-криминалистической фонетики и языковой компетенции экспертов-криминалистов. Предложена методика использования механизма ассоциативных связей на сегментном и супрасегментном уровнях.

В выступлении Д. Соутнера, Я. Зелинки и Л. Мюллера (Чехия) «О гибридной системе распознавания речи NN/HMM на базе RNN-ориентированной языковой модели» представлено описание новой системы распознавания речи. Используемая акустическая модель на основе нейронной сети вычисляет апостериорные данные для состояний контекстно-зависимых акустических блоков. В качестве языковой модели использовалась нейронная сеть с максимальным расширением энтропии. Данная гибридная система сравнивалась с предыдущей гибридной системой, оснащенной стандартной nграммной языковой моделью. В экспериментах также сравнивались данные со стандартной системой GMM/HMM. Характеристики системы оценивались с использованием Речевого корпуса британского варианта английского языка некоторых предыдущих систем.

Я. Швец и Л. Шмидл (Чехия) («Обнаружение семантического объекта на материале переговоров при управлении воздушным транспортом») рассмотрели обнаружение необходимого семантического объекта в системах автоматического распознавания речи применительно к диалогам, относящимся к управлению воздушно-транспортным трафиком. Представленный метод предназначен для использования в автоматическом учебном пособии для авиадиспетчеров. Семантические объекты моделируются с помощью определенных экспертами контекстно-свободных грамматик. Использован новый подход, который позволяет обрабатывать неопределенные данные на входе в виде взвешенного преобразователя с конечным числом состояний. Этот метод был экспериментально оценен с привлечением реальных данных. Проведено также сравнение методов с использованием знаний в области условий ведения диалогов. Результаты показывают, что система со знаниями целевых семантических объектов снижает частоту ошибок с 24,7 % до 17,1 % по сравнению со стандартными системами обнаружения необходимого семантического объекта.

Доклад В. Верходановой и В. Шапранова (Санкт-Петербург) «Обнаружение заполненных пауз и звуковых артикуляций, продленных во времени, в зависимости от акустических особенностей спонтанной русской речи» посвящен акустическому анализу спонтанной речи. Акустический анализ спонтанной речи связан с рядом проблем, к числу которых относятся также так называемые «речевые паразиты». Хотя большинство из них легко обнаруживается самими говорящими и они, как правило, не вызывают каких-либо трудностей при понимании, для системы автоматического распознавания речи (ASR) их появление приводит к большому числу ошибок распознавания. В докладе рассматриваются наиболее частотные из них: заполненные паузы и артикуляционные «растяжки» на основе анализа их акустических параметров. Для выявления звонких хезитационных участков

применительно к звонким согласным и гласным использовался метод, основанный на функции автокорреляции, а для обнаружения хезитационных участков применительно к глухим согласным — метод полосовой фильтрации. Для экспериментов по обнаружению заполненных пауз и «растяжек» использовался специально собранный корпус спонтанных диалогов на русском языке (например, описание маршрута по карте и др.). Точность выявления озвонченных заполненных пауз и артикуляционных «растяжек» составила 80 %, оглушенных — 66 %.

Целью доклада З. Зайича, Я. Зелинки, Я. Ванека и Л. Мюллера (Чехия) «Сверточная нейронная сеть для уточнения дикторо-адаптивной трансформации» является обсуждение метода уточнения акустической модели речи диктора с помощью сдвига линейной регрессии максимального правдоподобия (MLLR) в случае ограниченного количества данных по адаптации, что может привести к неполным матрицам преобразований. Предлагается метод подавления влияния плохо оцененных параметров преобразования с использованием искусственной нейронной сети (ANN), в частности сверточной нейронной сети (CNN). Плохо оцениваемое преобразование сдвига MLLR распространяется через ANN (заранее прошедшую соответствующее обучение), а выходные данные сети используются в качестве новой уточненной трансформации. Для обучения ANN в качестве входных и выходных данных ANN используются полные и неполные преобразования сдвига MLLR соответственно.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

Proceedings 2014 — Proceedings of the 16th International Conference on Speech and Computer, SPECOM, 2014, ser. Lecture Notes in Artificial Intelligence (including subseries Lecture Notes in Computer Science), 8773 LNAI. Ronzhin A., Potapova R., Delic V. (eds). Cham: Springer International Publishing, 2014.

Конференция «Системные изменения в языках России»

Мария Юрьевна Пупынина[a, @], **Аржаана Александровна Сюрюн**[a]

[a] Институт лингвистических исследований РАН, Санкт-Петербург, 199053, Россия; [@] pupynina@gmail.com

16—18 октября 2014 г. в Санкт-Петербурге в Институте лингвистических исследований прошла конференция «Системные изменения в языках России», организованная отделом языков народов России ИЛИ РАН.

Общая тематика докладов естественным образом распределилась по двум направлениям. Первое направление касалось диахронического исследования изменений в системе одного языка. Как правило, в докладах такого плана обсуждались вопросы грамматикализации или изменений в синтаксисе рассматриваемых языков. Поскольку многие из языков, данные о которых были представлены на конференции, до XX в. были бесписьменными, диахронический анализ охватывал сравнительно небольшой период времени. В других докладах диахрония в расчет не принималась; основное внимание было уделено рассмотрению

International conference «System changes in the languages of Russia»

Maria Yu. Pupynina[a, @], **Arzhaana A. Syuryun**[a]

[a] Institute for Linguistic Studies, Russian Academy of Sciences, St. Petersburg, 199053, Russia; [@] pupynina@gmail.com

реализаций какого-то лингвистического факта по близкородственным идиомам и их сопоставлению. Очевидно, что языки, не имеющие «живых» и/или хорошо описанных родственников, могут быть проанализированы только первым способом, а второй вариант анализа является единственно возможным для изучения системных изменений в языках без письменной традиции.

Выступления были основаны на достаточно пестром языковом материале: привлекались данные нескольких языковых семей; были охвачены языки различного уровня сохранности — от алеутского и ительменского, насчитывающих менее десятка носителей, до тувинского и якутского, на которых говорят десятки тысяч человек.

Много докладов было представлено по языкам, относящимся к алтайской макросемье. Так, четыре выступления были посвящены тунгусо-маньчжурским языкам.