УДК 81'322.5

В.О. Верходанова, А.А. Карпов

МОДЕЛИРОВАНИЕ РЕЧЕВЫХ СБОЕВ В СИСТЕМАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

Работа выполняется в рамках НИР СПбГУ (проект № 31.37.103.2011), а также Минобрнауки РФ (госконтракт № 07.514.11.4139).

Статья посвящена проблеме моделирования речевых сбоев диктора при автоматическом распознавании речи. Рассматривается само явление речевых сбоев, и по результатам анализа отечественных и зарубежных работ выделяются две группы подходов к автоматическому определению речевых сбоев: параметрическая обработка сигнала и комбинированные методы с применением языкового моделирования. Приводится анализ особенностей их применимости к обработке русской речи.
Ключевые слова: речевые сбои; автоматическое распознавание речи; анализ речи.

Речевые сбои являются одним из основных отличий спонтанной речи от подготовленной речи, и тем более от письменного текста. Очень немногие из нас обладают способностью гладко и красноречиво оформлять свои мысли, не передумывая, не сомневаясь и не сбиваясь, поэтому можно утверждать, что одна из основных черт спонтанной речи – это наличие пауз, хезитаций, повторений, самокоррекций, усеченных слов и т.п. Подобные речевые сбои являются препятствием для компьютерной обработки как звучащей речи, так и ее транскрипций.

Автоматическое распознавание речи (АРР или, в английской терминологии, automatic speech recognition – ASR) – это преобразование звучащей речи в текст. Существует несколько категорий систем распознавания речи, которые имеют различные сферы применения: 1) распознавание отдельных команд, которое применяется в коммерческих приложениях (голосовое управление, навигация по сайтам); 2) поиск ключевых слов в потоке речи (поисковые системы); 3) распознавание слитной речи на большом словаре (автоматическая расшифровка записей – создание стенограмм). Также системы распознавания речи характеризуются степенью зависимости от настройки на речь конкретного диктора: различают дикторо-зависимые и дикторо-независимые системы [1, 2].

Хотя речевые сбои дикторов изучались и ранее, формально их исследование началось только в 50-х гг. XX в., независимо развиваясь в рамках разных дисциплин: психологии, лингвистики, физиологии. Американский психолог Венделл Джонсон внес существенный вклад в исследование заиканий [3]. В рамках общей лингвистики речевыми сбоями среди других ученых занималась Фрейда Голдман-Эйслер [4]. Существенный вклад в исследование речевых сбоев в психотерапии внес Джордж Ф. Маль со своими коллегами [5]. С тех пор речевые сбои стали изучать в разных научных областях (изучение заикания, общая лингвистика, когнитивная психология, психология сознания, фонетика, гендерные исследования, психология, акустика, технологии обработки языка и речи и т.д.) [6].

Несмотря на многосторонние исследования речевых сбоев, общепринятая терминология в этой области пока не сложилась. Для описания этих явлений существуют различные варианты альтернативных терминов; например, в англоязычной литературе можно встретить такие термины, как «non-fluency», «dysfluency», «discontinuity», «flustered speech», «speech disturbance», «hesitation», «speech management», «own communication management», «turnholding devices» и др. [6]. В русскоязычной литературе их иногда рассматривают в рамках фонационных паралингвистических явлений, также можно встретить термины «внеязыковые элементы речи», «речевые сбои» [7–9].

1. Классификация речевых сбоев. Возникновение сбоев в речи может быть вызвано как внешними воздействиями, так и сбоями в планировании речевого акта [9]. Сбои в планировании, в свою очередь, могут иметь разную природу, и среди них можно выделить заполненные паузы хезитации, самоисправления (или самокоррекции) и оговорки. Учитывая различные причины возникновения речевых сбоев и разновидности типов нарушений в спонтанной речи, можно ввести следующую классификацию, как показано на рис. 1.

Паузы хезитации (паузы колебания) представляют собой перерыв в фонации, часто заполненный некоторыми звуками. Обычно такие паузы представляют собой семантические лакуны и свидетельствуют о том, что говорящему требуется дополнительное время на формулирование следующего за текущим фрагмента высказывания [10, 11].

Типы заполнения пауз хезитации [11]:

1. Абсолютная пауза.
2. Удлинение отдельных звуков в словах.
3. Словоподобные, «долексические» заполнения паузы.
4. Вспомогательные элементы дискурса (слова и словосочетания (*как сказать*)).

Самокоррекция возникает, когда в некоторой точке дискурса говорящий решает, что определенный фрагмент порожденного им высказывания не соответствует по какой-либо причине его намерениям. В этом случае говорящий может прибегнуть к самоисправлению, заменить полностью или частично повторить не устроивший его фрагмент [9].

Можно говорить о следующих типах самоисправлений [9]:

1. Онлайн коррекция, когда говорящий сразу же после ошибки исправляет обнаруженную проблему.
2. Ретроспективная коррекция, или редактирование, при которой говорящий редактирует готовый фрагмент дискурса постфактум.

Вспомогательные элементы дискурса – это дискурсивные элементы, которые не несут предметно-фактической информации, но выполняют некоторые функции в речи. В действительности они направлены на

оптимизацию и организацию общения. По выполняемой функции их можно разделить на следующие типы [12]:

1. Единицы, структурирующие речевой поток:

– обозначают последовательность информативных блоков (*во-первых, прежде всего*);

– вводят дополнительную информацию (*впрочем, кстати*);

– обозначают роль фрагмента высказывания (*теперь о главном*);

– показывают отношение говорящего (*как известно, на мой взгляд, к сожалению*).

2. Контакто-устанавливающие элементы, направленные на передачу метакоммуникативной информации, такие как:

– этикетные формулы (*добрый день!*);

– актуализаторы (*да, ага, правда*);

– интимизаторы общения (*слушай, знаешь, смотри, представь*).

Традиционно к речевым сбоям относят еще оговорки, которые приводят к непроизвольному использованию говорящим незапланированных им фрагментов [13].

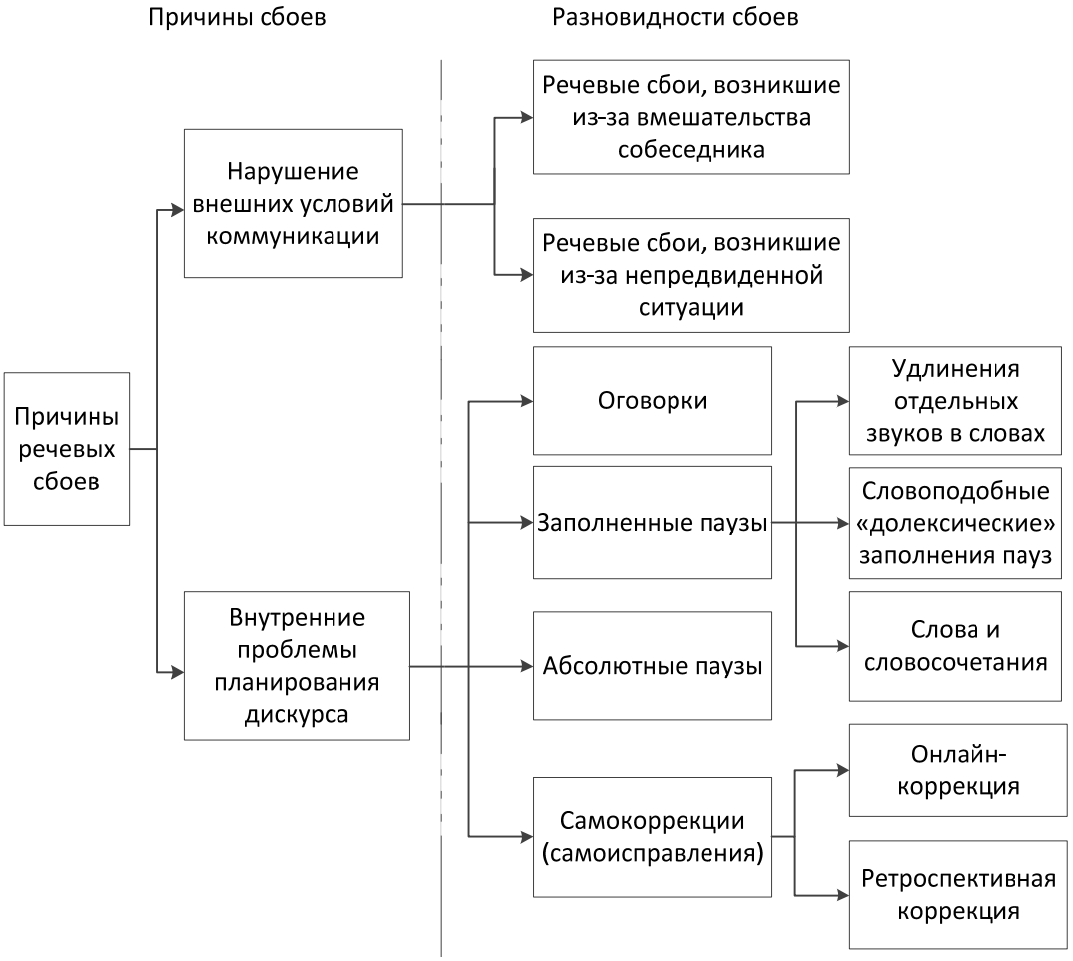

Рис. 1. Классификация речевых сбоев

В зарубежной литературе принято описывать временные характеристики сбоев. Так, согласно описанию в работе Э. Шриберг [14] используются следующие термины:

– reparandum (кратко RM) – репарандум, участок звукового сигнала, который соответствует всему удаленному отрезку речи;

– interruption point (IP) – начало речевого отрезка, соответствующее «моменту прерывания» чистой речи и возникновению речевого сбоя;

– interregnum (IM) (у других авторов «редактирующая фаза» [15] или «интервал сбоя» [16]) – длительность речевого сбоя, этот термин используется для обозначения временного отрезка от RM до начала исправления, при этом он может и не содержать редактирующего элемента (например, незаполненная пауза может быть использована говорящим для перепланировки высказывания без редактирования);

– repair (RR) – исправление, это участок речи, который соответствует материалу репарандума.

В следующем разделе рассмотрим существующие речевые и языковые ресурсы, применяющиеся для моделирования речевых сбоев и обучения систем автоматического распознавания спонтанной речи.

2. Речевые и языковые ресурсы. Для исследований речевых сбоев используются корпусы спонтанной речи с разметкой на речевые сбои. Для того чтобы в корпусе помимо такой информации, как фонемы, слова, синтагмы, дополнительно отмечать речевые сбои, используется Rich Transcription – транскрипция, в которой учитываются границы предложений, слова-заполнители, редактирующие сбои и др. [17].

Проблема аннотации речевых сбоев по корпусу спонтанной речи рассматривалась в статье [18]. Авторы рассматривали такие речевые сбои, как заполнен-

ные паузы, самоисправления, неверно произнесенные слова. Материалом для исследования послужили видеозаписи выступлений депутатов Верховной Рады Украины. В соответствии со звуковыми файлами производились разметка, корректировка и анализ текста стенограммы, при этом в текст стенограммы вносились недостающие элементы спонтанной речи, которые не были учтены в лексиконе системы распознавания. Главной особенностью этой разметки является то, что она включает в себя большую часть речевых сбоев, характерных для спонтанной украинской речи. Наиболее повторяющимися речевыми сбоями были заполненные паузы *а* (40,86%), слова, произнесенные неверно (20,07%), он-лайн коррекция и коррекция со вставкой (16,49%) и заполненные паузы *ее* (7,89%) [18].

Целесообразность использования акустических моделей заполненных пауз и артефактов при распознавании спонтанной русской речи обоснована в работе [19]. Для исследования внеязыковых речевых элементов и обучения их вероятностных моделей авторами был собран и сегментирован корпус речи выступлений, сделанных в ходе небольшого семинара. В статье представлено описание выявленных в ходе сегментации речевых сбоев, также приводится статистика по частоте употребления каждого элемента и его средняя длительность. Далее рассмотрим ряд корпусов, в которых также учитываются эти явления.

Czech Broadcast Conversation MDE Transcripts (корпус транскрипций с метаданными чешских радиопередач) [20]. Этот корпус представляет собой пословные транскрипции с метаданными, сделанные по корпусу Czech Broadcast Conversation Speech (чешская разговорная речь радиопередач) [21], который состоит из 40 часов речи, записанной с чешского радио 1 в 2003 г. При создании транскрипций учитывались правила разметки, по которым затем отмечались такие метаданные, как смена говорящего, границы между предложениями (в рамках речи одного говорящего), накладывающаяся речь (если говорят двое и больше), фоновые шумы, шумы говорящих (такие как вздох, смех, причмокивание), заполненные паузы (отмечались паузы *ээ* и *мм*), междометия (согласие и несогласие), неразборчивая речь, числа – все числительные записываются полными словами, неправильно произнесенные слова (оговорки, ошибки чтения), части слов и пунктуация [22].

Корпус SWITCHBOARD [23] представляет собой более 240 часов записи спонтанных телефонных разговоров более 500 дикторов обоих полов. Речь полностью затранскрибирована, и транскрипции проверены автоматически и дважды вручную. Неречевые данные отмечаются в квадратных скобках, всего размеченных типов 78 и среди них такие, как вздох, кашель, зевание, мяуканье, звон посуды и др.

Корпус транскрипций «RT-03 MDE Training Data Text and Annotations» [24] представляет собой транскрипции радиопередач и телефонных разговоров, взятых из корпуса телефонных разговоров Switchboard-1 Release 2 и из корпуса новостных радиотрансляций 1997 English Broadcast News Speech (HUB4). Аннотированные транскрипции соответствуют приблизительно 20 часам из корпуса новостных радиотрансляций и 40 часам телефонных разговоров. В транскрипциях размечены различные типы метаданных, в которых выделяются 4 типа заполнителей: заполненные паузы, как *uh* и *um*, дискурсивные маркеры, как *you know*, ремарки и вставки и такие редактирующие термины, как *sorry* и *I mean*, а также отмечаются моменты прерывания, весь отрезок речевого сбоя и границы синтагм.

На кафедре фонетики СПбГУ разрабатывался аннотированный корпус русской речи, который включал в себя речь 4 мужчин и 4 женщин и учитывал различные произносительные стили. Аннотация учитывала 6 уровней, в которых отмечалась вся фонетическая и просодическая информация о записанной речи [25, 26].

3. Методы выявления речевых сбоев в спонтанной речи. Исследования речевых сбоев показали, что человек легко выделяет их из речи [14, 27]. Но для системы автоматического распознавания речи не определено, на основании каких признаков должен производиться анализ, какие знания следует привлекать в ходе сегментации и классификации. Поэтому пока не созданы адекватные модели речевых сбоев, обеспечивающие их обработку в автоматическом режиме. Тем не менее данное направление является крайне актуальным: всевозможные речевые сбои, такие как заполненные паузы, удлиняют высказывания. Также они вызывают различного рода ошибки, вследствие того что системы распознавания обучаются на структурированных предложениях без речевых сбоев, что приводит к формированию ошибочных транскрипций.

Методы обработки речевых сбоев можно разделить по признаку описания их с помощью акустических моделей или с помощью комбинированных моделей (языковые + акустические). Но в силу объективных причин (временные и экспертные затраты) исследователи часто используют только акустические модели речевых сбоев для реализации их в системах автоматического распознавания речи.

3.1. Методы выявления речевых сбоев в спонтанной речи за счет параметрической обработки сигнала. Существует широкий спектр работ, посвященных моделированию речевых сбоев в рамках создания систем автоматического распознавания речи [28–30]. Также существует группа подходов, направленных на повышение качества распознавания спонтанной речи за счет предварительного выявления речевых сбоев и их устранения из звукового сигнала на этапе цифровой обработки еще до того, как данные поступают на вход системы автоматического распознавания речи [31], или устранения сбоев с использованием транскрипций речи [30, 32].

В работе [31] авторы разработали алгоритм, который определяет и удаляет заполненные паузы и повторения из речевого сигнала. Для определения границ заполненных пауз применялись следующие характеристики: длительность, частота основного тона, спектральные и формантные характеристики. Для выделения и последующего удаления повторений предложенный алгоритм учитывал длительность и частотность повторяющихся отрезков, а также разницу между логарифмами амплитуд спектра каждой пары голосовых фрагментов вокруг долгой паузы. При этом учитывался тот факт, что повторения чаще всего сопровождаются

паузой. Эксперименты проводились на искусственно созданных небольших базах данных (три диктора, 40–60 предложений) только с одной паузой или повторением в предложении. В итоге точность распознавания слов повысилась: в случае использования алгоритмов определения повторений и заполненных пауз – соответственно на 10 и 20%, а при совместном использовании – на 30%.

В работе [28] авторы описывают метод выделения заполненных пауз и удлинений слов в японской спонтанной речи на основе двух характеристик: небольшого изменения частоты основного тона и деформации спектральной огибающей. Полнота распознавания (количество правильно распознанных заполненных пауз к общему количеству заполненных пауз) составила 84,9% и точность распознавания (число правильно распознанных заполненных пауз к общему числу выделенных заполненных пауз) составила 91,5%.

В работе [33] авторы исследуют проблему составления корпуса пауз и удлинений для португальской спонтанной речи, а также построения акустических моделей этих явлений. В статье рассматриваются заполненные паузы и сегментные удлинения. Принятие решения о наличии хезитации в речи производилось в следующих случаях: 1) гласный был длиннее установленного порога (350 мс для португальских гласных); 2) появляются последовательности одинаковых звуков; 3) возможные согласные между гласными очень короткие. Учитывались следующие характеристики этих хезитаций: частота основного тона, энергия и спектр.

3.2. Методы выявления речевых сбоев в спонтанной речи с использованием языкового моделирования. Ряд работ направлен на улучшение распознавания речевых сбоев за счет использования дополнительных источников знаний, а именно различных языковых моделей. В работе [29] авторы рассматривают три типа речевых сбоев: 1) повторение, 2) редактирование (замена содержимого) и 3) рестарты (или фальстарты). Материалом для исследования послужили часть корпуса Switchboard-I, а также ее транскрипции: как сделанные вручную, так и полученные на выходе системы распознавания речи. В качестве просодических характеристик были использованы нормализованная длительность слов и пауз и частота основного тона, а также джиттер (нежелательные фазовые и / или частотные случайные отклонения передаваемого сигнала), спектральный наклон и соотношение времени, когда голосовые связки разомкнуты к общей продолжительности гортанного цикла. Помимо просодики, использовались три типа языковых моделей: 1) вероятностная языковая модель, описывающая совместное появление ключевых слов и внеязыковых элементов в потоке спонтанной речи; 2) синтаксическая модель языка, использующая статистику по частеречной принадлежности для анализа случаев возникновения речевых сбоев и выявления тенденций, обусловленных синтаксическими закономерностями, для выявления таких типов речевых сбоев, как, например, повторение предлогов; 3) синтактико-стохастическая модель языка для выявления повторений. Эксперименты показали, что совместное употребление всех моделей значительно улучшает распознавание моментов прерывания. Процент распознавания моментов прерывания на транскрипциях, сделанных вручную, в случае применения всех трех моделей (просодическая, вероятностная и синтаксическая) составил 56,76%, точность – 81,25%, общая точность – 98,10% (против 55,47%, 79,33 и 98,01% для одной вероятностной модели). На транскрипциях, полученных системой автоматического распознавания речи (АРР), общая точность составила 97,05%. Результаты для распознавания моментов прерывания были следующие: точность на сделанных вручную транскрипциях – 98,01%, на транскрипциях, сделанных системой АРР, – 97,05%.

В работе [30] авторы рассматривали следующие речевые сбои: 1) единицы, подобные предложениям (например, назывное предложение в английском); 2) редактирующие речевые сбои, которые включают синтаксически значимое содержание: пересмотры (замена содержания), фальстарты и сложные сбои, которые состоят из набора сбоев; 3) слова-заполнители, которые включают паузы, дискурсивные маркеры и явные редактирующие слова (например, *видишь ли, простите, ой*). Материалом послужили два разных по жанру корпуса: корпус телефонных разговоров (CTS) и корпус радиотрансляций (BN). В качестве источников знаний использовались как лексические свойства (совместная встречаемость слов с другими словами или с соседствующими явлениями, по частеречным тэгам или по их семантическому классу), так и просодические свойства (длительность (слов, пауз, звуков), частота основного тона, энергия и информация о паузах). Для построения моделей были применены: 1) скрытые Марковские модели (Hidden Markov Model, HMM); 2) модель максимальной энтропии (Maximum Entropy Model, ME), 3) случайные поля (Conditional Random Fields, CRF). Для корпуса CTS наименьший процент ошибок распознавания редактирующих слов был для транскрипций, сделанных вручную, с помощью метода CRF и составил 51,49%. Для моментов прерывания на том же корпусе наилучший процент составил 34,64% с помощью HMM для транскрипций, сделанных вручную. На корпусе BN наименьший процент для редактирующих слов и моментов прерывания был получен на транскрипциях, сделанных вручную, с помощью ME, и составил соответственно 42,62 и 30,72%. Для слов-заполнителей как для корпуса CTS, так и для корпуса BN результаты были лучше на транскрипциях, сделанных вручную, – 26,98 и 18,11%.

В работе [32] авторы анализируют редактирующие слова (повторы, фальстарты и т.д.) и заполнители (дискурсивные маркеры или паузы, такие как *ты знаешь*, *эм*) на материале транскрипций по корпусу телефонной речи (CTS) и по корпусу новостных радио-передач (BNEWS). Авторы использовали обучающий алгоритм, основанный на трансформациях (TBL). Для построения правил использовались следующие свойства: учитывалась лексема (само слово), частеречный тэг, информация о том, следует ли за словом пауза и является ли слово высокочастотным (т.е. является ли оно более частотным в речи данного говорящего, чем во всем корпусе). Для корпуса CTS 27% ошибок определения редактирующих сбоев и 19% ошибок определения слов-заполнителей появились, когда система АРР не

поняла. Также возникала проблема, когда система удаляла редактирующее слово или заполнитель. И, наконец, контекст сбоев тоже оказался важен: система чаще всего ошибалась на редактирующих сбоях из-за присутствия длинных и сложных для определения редактирующих явлений.

В работе [34] авторы рассматривали исправления, фальстарты, заполнители и моменты прерывания (IP) на материале аннотированного корпуса Switchboard. Для каждого предложения для определения кандидатов в исправления запускался анализ с помощью стохастической формальной TAG (tree-adjoining grammar) модели. Вероятностная синтаксическая модель языка оценивала согласованность (fluency) каждой гипотезы, и модель максимальной энтропии выбирала наиболее вероятную гипотезу, учитывая оценки языковой модели и другие параметры. Заполнители определялись независимо с помощью небольшого набора детерминистических правил, а моменты прерывания IPs определялись с помощью комбинации выходных данных из модулей определения исправлений и заполнителей.

В рамках конкурса по распознаванию речи «Rich Transcription blind evaluation», организованного Национальным институтом стандартов и технологий (NIST), перед авторами были поставлены следующие задачи:

– выделение редактирующих слов;

– выделение заполнителей, (типы заполнителей также различались: заполненные паузы, дискурсивные элементы и явные редактирующие слова);

– выделение момента прерывания.

Каждая задача выполнялась для двух типов входных данных: созданной вручную транскрипции и полностью автоматического результата системы распознавания. Наилучшие результаты для каждой задачи показаны в таблице.

Результаты, полученные для каждой задачи выявления речевых сбоев

Транскрипция речи	Задача	Уровень ошибок распознавания, %
Сделанная вручную	Выделение редактирующих слов	46,08
	Выделение заполнителей	23,69
	Выделение момента прерывания	28,60
Выход системы APP	Выделение редактирующих слов	76,25
	Выделение заполнителей	39,93
	Выделение момента прерывания	55,88

Таким образом, однозначно лучшего решения проблемы речевых сбоев на сегодняшний момент нет. Однако оба подхода к выявлению речевых сбоев очень перспективны, но требуют разных материалов и моделей. Для использования языковой модели дополнительно нужен большой корпус транскрипций, по которому происходит обучение модели, в то время как параметрическая обработка не требует транскрибирования сигнала. В связи с тем что для русского языка не разработано методов обработки речевых сбоев для автоматического распознавания речи, необходимо применить несколько методов и сравнить результаты. Поскольку создание корпуса транскрипций, учитывающих речевые сбои, достаточного для обучения языковой модели (хотя бы тригамной), является весьма затратным по временным и финансовым факторам процессом, то обработка речевых сбоев с помощью параметрических методов для русского языка на данный момент наиболее целесообразна.

Среди возможных подходов к тому, как относиться к сбоям в рамках системы распознавания, есть те, которые позволяют моделировать и искать речевые сбои как отдельные речевые и неречевые элементы, и те, которые игнорируют их, отличая от полезной речи, но не различая их друг от друга.

В качестве альтернативного подхода при использовании языкового моделирования и отдельно моделирования неизвестных слов можно относить речевые сбои к классу Unknown Words и строить с их учетом языковую модель. В работе [35] предлагается слоговый подход акустического моделирования новых слов. Описываются эксперименты с различными порождающими моделями слов для спонтанной речи. Слоговая модель, предложенная авторами, организована как конечный автомат, она моделирует ограничения на фонемную последовательность. Так же реализована биграмная языковая модель, которая предсказывает вероятности неизвестных слов. Наилучшим результатом была точность распознавания 92% при использовании слоговой модели, реализованной как конечный автомат.

Для русского языка актуальность подобного подхода демонстрируется в статье [36]. Авторы предлагают использовать классовые (кластерные) языковые модели, при обучении которых весь набор слов разбивается на несколько подмножеств (с помощью экспертов или автоматическими методами) и каждому такому подмножеству присваивается маркер класса. После чего обычную языковую модель обучают на последовательностях классовых маркеров, и дополнительные модели строятся, чтобы получить вероятности слов внутри соответствующих классов.

В статье представлен аналитический обзор существующих методов выявления речевых сбоев и их устранению из речевого сигнала при распознавании разговорной речи. Рассмотрены два основных подхода: методы выявления речевых сбоев только за счет параметрической обработки сигнала и комбинированные методы, в которых дополнительно применяется языковое моделирование, а также предложены возможные подходы к обработке речевых сбоев в рамках систем распознавания речи.

ЛИТЕРАТУРА

1. *Кипяткова И.С., Карпов А.А.* Аналитический обзор систем распознавания русской речи с большим словарем // Труды СПИИРАН. 2010. Вып. 12. С. 7–20.
2. *Карпов А., Ронжин А., Лобанов Б. и др.* Разработка бимодальной системы аудиовизуального распознавания русской речи // Информационно-измерительные и управляющие системы. 2008. Т. 6, № 10. С. 58–62.
3. *Wendell A.L.* Johnson (1906–1965) Memorial Home Page. URL: http://www.uiowa.edu~cyberlaw/oldinav/wjhome.html (дата обращения: 20.03.2012).

4. *Eisler F.G.* Psycholinguistics: Experiments in Spontaneous Speech. Academic Press Inc, 1968. 169 p.
5. *In Memoriam*: George F. Mahl. Yale Bulletin & Calendar. 2006. March 24. Vol. 34, № 23. URL: http://www.yale.edu/opa/arc-ybc/v34.n23/story13.html (дата обращения: 20.03.2012).
6. *Proceedings* of DiSS'03, Disfluency in Spontaneous Speech Workshop // Gothenburg Papers in Theoretical Linguistics 90 / ed. by Robert Eklund. Sweden : Göteborg University, 2003. 5–8 September. P. 3–4.
7. *Колшанский Г.В.* Паралингвистика. М., 1974. 81 с.
8. *Николаева Т.М.* Паралингвистика // Лингвистический энциклопедический словарь / под ред. В.Н. Ярцевой. М. : Советская энциклопедия, 1990.
9. *Подлесская В.И., Кибрик А.А.* Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи // Научно-техническая информация. Сер. 2. 2007. № 2. С. 2–23.
10. *Лауринавичюте А.К., Федорова О.В.* Влияние паузы хезитации на понимание синтаксической структуры предложения носителями русского языка // Материалы международной конференции «Диалог 2010». Бекасово, 2010. С. 279–284.
11. *Herbert H. Clark, Jean E. Fox Tree.* Using uh and um in spontaneous speaking // Cognition. 2002. Vol. 84. P. 73–111.
12. *Андреева С.В.* Лингвистические закономерности передачи информации при автоматической обработке речи // Материалы Третьего междисциплинарного семинара «Анализ разговорной русской речи». СПб. : СПбГУАП, 2009. С. 10–14.
13. *Сморгонская Е.В.* Психолингвистическая дифференциация и классификация речевых сбоев // Вестник ВГУ. Сер. Лингвистика и межкультурная коммуникация. 2008. № 3. С. 140–142.
14. *Shriberg E.E.* Preliminaries to a Theory of Speech Disfluencies. PhD thesis, University of California at Berkeley, 1994. 225 p.
15. *Levelt W.J.M.* Monitoring and self-repair in speech // Cognition. 1983. Vol. 14. P. 41–104.
16. *Nakatani C.H., Hirschberg J.* A corpus-based study of repair cues in spontaneous speech // Journal of the Acoustical Society of America. 1994. № 95 (3). P. 1603–1616.
17. *Liu Y.* Structural Event Detection for Rich Transcription of Speech, PhD thesis. Berkeley : Purdue University and ICSI, 2004. 253 p.
18. *Пилипенко В.В., Ладошко О.Н.* Аннотация и учет речевых сбоев в задаче автоматического распознавания спонтанной украинской речи // Искусственный интеллект. 2010. № 3. С. 238–248.
19. *Кипяткова И.С., Верходанова В.О., Ронжин А.Л.* Сегментация паралингвистических фонационных явлений в спонтанной русской речи // Вестник Пермского университета. Российская и зарубежная филология. 2012. Вып. 2 (18). С. 17–23.
20. *Корпус* «Czech Broadcast Conversation MDE Transcripts» // Каталог LDC. URL: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009T20 (дата обращения: 16.06.2012).
21. *Корпус* «Czech Broadcast Conversation Speech» // Каталог LDC. URL: http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009S02 (дата обращения: 16.06.2012).
22. *Kolář J., Švec J., Strassel S. et al.* Czech Spontaneous Speech Corpus with Structural Metadata // In Proc. INTERSPEECH 2005. Lisbon, Portugal, 2005. P. 1165–1168.
23. *SWITCHBOARD*: A User's Manual. URL: http://www.ldc.upenn.edu/Catalog/readme_files/switchboard.readme.html (дата обращения: 20.06.2012).
24. *Корпус* «RT-03 MDE Training Data Text and Annotations» // Каталог LDC. URL: http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2004T12 (дата обращения: 20.06.2012).
25. *Skrelin P., Volskaya N., Kocharov D. et al.* A Fully Annotated Corpus of Russian Speech // In Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10). Valletta, Malta, 2010. P. 109–112.
26. *Skrelin P., Kocharov D.* Russian Speech Corpora Framework for Linguistic Purposes // In Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'12). Istambul, Turkey, 2012. P. 43–46.
27. *Кожевникова Кв.* О смысловом строении спонтанной устной речи // Новое в зарубежной лингвистике. Вып. XV: Современная зарубежная русистика. М., 1985. С. 512–524.
28. *Masataka G., Katunobu I., Satoru H.* A real-time filled pause detection system for spontaneous speech Recognition // In Proc. of the 6th European Conference on Speech Communication and Technology (Eurospeech '99). Budapest, Hungary, 1999. P. 227–230.
29. *Liu Y., Shriberg E., Stolcke A.* Automatic Disfluency Identification in Conversational Speech Multiple Knowledge Sources // In Proc. of the EUROSPEECH 2003. Geneva, Switzerland, 2003. P. 957–960.
30. *Liu Y., Shriberg E., Stolcke A. et al.* Enriching Speech Recognition with Automatic Detection of Sentence Boundaries and Disfluencies // IEEE Trans. Audio, Speech and Language Processing. 2006. № 14(5). P. 1526–1540.
31. *Kaushik M., Trinkle M., Hashemi-Sakhtsari A.* Automatic Detection and Removal of Disfluencies from Spontaneous Speech // In Proc. of the Proceedings of the Thirteenth Australasian International Conference on Speech Science and Technology (SST). Melbourne, Australia, 2010. P. 98–101.
32. *Snover M., Dorr B., Schwartz R.* A lexically-driven algorithm for disfluency detection // In Proc. of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics 2004 (HLT-NAACL-Short '04). Boston, Massachusetts, USA, 2004. P. 157–160.
33. *Veiga A., Candeias S., Lopes C., Perdigão F.* Characterization of hesitations using acoustic models // In Proc. of the 17th International Congress of Phonetic Sciences (ICPhS XVII). Hong Kong, China, 2011. P. 2054–2057.
34. *Lease M., Johnson M., Charniak E.* Recognizing disfluencies in conversational speech // In Audio, Speech, and Language Processing, IEEE Transactions on. 2006. Vol. 14, № 5. P. 1566–1573.
35. *Kemp T., Jusek A.* Modelling Unknown Words in Spontaneous Speech // In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-96). Atlanta, 1996. P. 530–533.
36. *Korenevsky M., Bulusheva A., Levin K.* Unknown Words Modeling in Training and Using Language Models for Russian LVCRS System // In Proc. of the International Conference on Speech and Computer (SPECOM'11). Kazan, Russia, 2011. P. 144–150.

Статья представлена научной редакцией «Филология» 3 сентября 2012 г.