ВЕСТНИК ПЕРМСКОГО УНИВЕРСИТЕТА

2012 РОССИЙСКАЯ И ЗАРУБЕЖНАЯ ФИЛОЛОГИЯ Вып. 2(18)

УДК 81’34

СЕГМЕНТАЦИЯ ПАРАЛИНГВИСТИЧЕСКИХ ФОНАЦИОННЫХ ЯВЛЕНИЙ В СПОНТАННОЙ РУССКОЙ РЕЧИ [1]

Ирина Сергеевна Кипяткова
к. техн. н., научный сотрудник лаборатории экспериментальной фонетики
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., д. 11. kipyatkova@iias.spb.su

Василиса Олеговна Верходанова
магистр кафедры фонетики и преподавания иностранных языков
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., д. 11. interiora@gmail.com

Андрей Леонидович Ронжин
д. техн. н., главный научный сотрудник лаборатории экспериментальной фонетики
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., д. 11. ronzhin@iias.spb.su

В статье проанализированы паралингвистические фонационные явления, которые могут возникать в спонтанной речи, приведен обзор способов учета таких явлений при автоматическом распознавании речи. Для обучения акустических моделей внеязыковых элементов нами был сегментирован корпус спонтанной русской речи, выделены артефакты (вдох, прочищение горла/кашель и причмокивание) и заполненные паузы хезитации, созданы акустические модели внеязыковых элементов, которые встретились в корпусе более двух раз. Точность распознавания внеязыковых элементов в собранном корпусе составила 86 %.

Ключевые слова: паралингвистические явления; спонтанная речь; речевые сбои; сегментация речи; автоматическое распознавание речи.

Паралингвистика – раздел языкознания, изучающий невербальные (неязыковые) средства, включенные в речевое сообщение и передающие, вместе с вербальными средствами, смысловую информацию [Лингв. энцикл. словарь 1990]. Паралингвистические средства не входят в систему языка и не являются речевыми единицами, однако в той или иной степени представлены в каждой речевой единице, сопровождая речь. Различают три вида паралингвистических средств: 1) фонационные – темп, тембр, громкость речи, заполнители пауз (к примеру, э-э, м-м), мелодика речи, диалектные, социальные или идиолектные особенности артикуляции звуков; 2) кинетические – жесты, поза, мимика говорящего; 3) графические – особенности почерка, графические дополнения к буквам, заменители букв [там же]. Говорящий использует то или иное паралингвистическое средство непредсказуемо, в отличие от лингвистических средств. Поэтому, например, особенности громкости или некодифицированные изменения мелодики будут паралингвистическими явлениями, в отличие от интонационного оформления вопроса – лингвистического явления [Большая сов. энцикл. 1975].

Данная статья посвящена фонационным паралингвистическим явлениям и их выявлению в спонтанной (разговорной) речи системами автоматического распознавания речи. Разговорная речь – это спонтанная литературная речь, реализуемая в неофициальных ситуациях при непосредственном участии говорящих с опорой на прагматические условия общения [Земская 1988]. Разговорной речи свойственны три экстралингвистических признака – это 1) спонтанность, неподготовленность; 2) наличие неофициальных отношений между говорящими; 3) непосредственное участие говорящих. Прагматическая информация, включающая характеристики говорящего, слушающего и текущей ситуации, суще-

ственным образом влияет на языковую структуру коммуникации, экономичность вербальных средств и обеспечивает эффективное взаимодействие, несмотря на фоновые шумы и нечеткую артикуляцию говорящих.

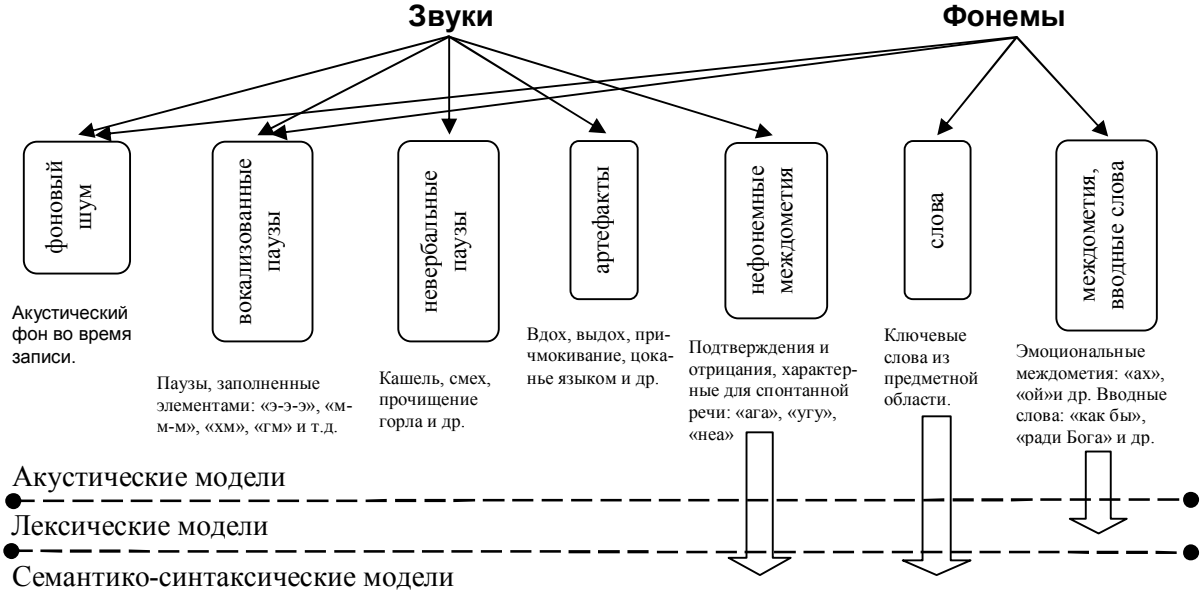
Под спонтанной речью Н.Н.Рудык понимает любой вид самостоятельной коммуникации, характеризующийся: 1) неподготовленностью, которая приводит к появлению новых комбинаций языковых компонентов, знакомых выражений в новых речевых ситуациях;

2) мотивированностью, проявляющейся в стимуляции, а затем в мотивации к говорению при наличии фактора неожиданности; 3) инициативностью, проявляющейся в реализации желания выразить свои мысли; 4) эмоциональностью, проявляющейся в способности чувственной оценки объектов внеязыковой действительности; 5) экспрессивностью, проявляющейся в устремлении говорящего с помощью речи, сопровождаемой мимикой и жестами, воздействовать на слушателя [Рудык 2010: 188–191].

Отличительной характеристикой спонтанной речи является отсутствие заранее подготовленной формы и содержания устного сообщения. Спонтанно порождаемая в текущий момент фраза обладает высокой вариативностью на всех уровнях обработки речи. В большей степени это проявляется в наличии значительного количества редуцированных словоформ – форм слов, представленных на сегментном уровне меньшим количеством элементов, чем в полном варианте, предусматриваемом нормами кодифицированного литературного языка. Редуцированные словоформы являются примером неполного речевого сигнала, для адекватной интерпретации которого слушающий должен использовать дополнительную информацию помимо той, что непосредственно содержится в акустическом речевом сигнале. Другой проблемой моделирования и обработки речи является наличие паралингвистических фонационных явлений, таких как озвученные паузы, невербальные паузы, артефакты, часто употребляемых в разговорной речи. Устранение таких неинформативных элементов речевого сигнала на начальных стадиях обработки позволит избежать многих ошибок при распознавании речи, передавая на последующие уровни обработки только полезную для диалоговой системы информацию. Для спонтанной речи характерны также самоисправления, фальстарты и другие явления, не свойственные полному стилю речи [Филиппова 2007: 86–90].

Все перечисленные факторы существенно затрудняют автоматическую обработку спонтанной речи, но обычно не являются причиной коммуникативных неудач для слушателя, поскольку даже неподготовленная речь хорошо воспринимается и не вызывает переспросов благодаря контекстной предсказуемости и частотности употребления единиц в речи [Риехакайнен 2010]. В ситуации естественного общения количество и значимость различных признаков, доступных слушающему, варьируются под влиянием условий акустической обстановки, характеристик ситуации общения и коммуникантов, типа воспринимаемого сообщения и других факторов [Штерн 1992: 187].

При обработке спонтанной речи на вход системы автоматического распознавания поступают звуки фонемной и нефонемной природы. Кроме того, это могут быть как звуки, производимые диктором, так и посторонние шумы и речь тех, кто не контактирует непосредственно с системой. Возможные компоненты входного аудиосигнала показаны на рисунке.



Модели для работы со спонтанной речью

Сигнал может содержать шумы окружающей обстановки, в которой производится запись. Также в ходе речеобразования возникают «шумы» органов голосового аппарата, так называемые артефакты речи. Другими элементами, производимыми диктором, могут быть: озвученные и невербальные паузы, нефонемные подтверждения и отрицания. В ходе формирования высказывания помимо ключевых слов могут произноситься различные междометия и вводные слова. Вследствие спонтанности формирования фразы велика вероятность присутствия незнакомых системе слов. Какой бы большой ни был взят словарь, он не в состоянии покрыть все слова, которые пользователь может употребить в диалоге с системой для получения желаемого результата [Кипяткова 2010].

Акустический фон записанного речевого сигнала может содержать окружающие шумы или даже речь людей, находящихся неподалеку от диалоговой системы, но не общающихся с ней напрямую, а также шумы канала передачи данных. Все эти звуки накладываются на речь пользователя, затрудняя распознавание [Ронжин 2009]. Эта проблема решается путем применения различных методик цифровой фильтрации и в данной работе не исследуется.

Вокализованные (озвученные) паузы могут быть вызваны различными причинами: сомнения, размышления и др. Чтобы не допустить разрыва во фразе и диалоге с собеседником, образовавшаяся пауза заполняется разного рода звуками. Это могут быть как растянутые звуки, напоминающие фонемы («а-а», «э-э», «м-м»), так и звуки явно нефонемной природы (кряхтение, хриплые «а», «о», «м») или даже комбинации звуков («хм», «гм», «ма»). При диалоге между людьми озвученные паузы помимо того, что не позволяют разорвать разговор, давая собеседнику понять, что оратор не закончил высказывание, также могут служить неким сигналом о помощи, обращенным к собеседнику. Для системы автоматического распознавания речи вокализованные паузы не несут информативной нагрузки и поэтому должны быть устранены на ранних уровнях обработки сигнала [Ронжин 2011].

Невербальные паузы, как правило, не являются элементами дискурса и могут возникнуть в любой момент диалога. Они могут быть вызваны смехом, покашливанием, прочищением горла. Длительность подобных явлений может сильно варьироваться. В связи с этим невербальные паузы отличают от артефактов – преимущественно коротких неречевых элементов, например, причмокивание, цоканье языком, звуки, связанные с громким дыханием.

Для подтверждения и отрицания в разговорной речи вместо частиц «да» и «нет» часто используются их междометия-аналоги «угу», «ага», «неа» и т.п. Сложность выделения этих элементов заключается в их схожести с вариативными вокализованными паузами «у-у», «а-а», «м-м».

Такие фонационные явления, как вздох, плач, кашель, смех, крик, постукивание, дыхание, могут являться средствами эмоциональной окраски речи [Карпова 2011]. Анализ средств вербального и невербального общения для продуцирования ложных высказываний рассмотрен в работе [Леонтьева, Датчер, Филиппова 2011], вербальные средства создания перлокутивного эффекта смеха анализируются в [Антонова 2010].

Наиболее подробно типы речевых сбоев и способы их аннотирования в корпусах устной речи рассмотрены в статье [Подлесская, Кибрик 2007]. Выделяются две основные категории речевых сбоев: хезитации и самоисправления. В свою очередь, самоисправления делятся на два основных режима – онлайн-коррекцию и ретроспективную коррекцию, или редактирование. В первом случае при обнаружении проблемы говорящий останавливает поток речи, в половине случаев даже не заканчивая слов, и далее формирует грамматически приемлемый и ситуационно уместный, с его точки зрения, фрагмент речи. При ретроспективной коррекции говорящий завершает проблемный отрезок и затем уточняет или исправляет предыдущий фрагмент речи. Введены также дополнительные четыре критерия классификации коррекций [Подлесская, Кибрик 2006]: 1) структурный диапазон коррекции (макрокоррекции и микрокоррекции, начальные и срединные коррекции); 2) линейный диапазон коррекции (контактные и дистантные коррекции); 3) тип операции (повторы, модификации и отмены); 4) объем забракованного фрагмента (в зависимости от сегментной протяженности и цельнооформленности забракованного фрагмента выделяются «мелкие» и «крупные» коррекции). Другой особенностью предложенной авторами методики аннотирования устной речи является деление текста на элементарные дискурсивные единицы, сегментация которых производится по семантико-синтаксическому и интонационному критериям.

В работе [Stouten 2006] были проанализированы три типа речевых сбоев, наиболее характерных для спонтанной речи: 1) озвученная пауза, 2) повтор слов, 3) модификация предложения с самого начала. В качестве материала были использованы речевые корпусы Spoken Dutch Corpus (CGN) и Switchboard-1. Число озвученных пауз составило 3 % всех лексических единиц

в данных корпусах. Чаще всего это были междометия, и располагались они во всех частях предложений. Относительное количество повторов было равно примерно 1 %. Причем двадцать наиболее частых повторов – это короткие слова, состоящие из одного слога. Анализ явлений модификации предложений с самого начала проводился на корпусе Switchboard-1. Всего было обнаружено 112 рестартов, что составило 0,5 % общей численности слов в предложениях корпуса. Для учета трех перечисленных типов речевых сбоев в системе автоматического распознавания речи было предложено два варианта стратегий. Во-первых, каждый тип сбоя может быть явно учтен в статистической модели языка декодера речи. И в случае его обнаружения во фразе срабатывает альтернативный вариант модели, исключающий озвученную паузу, повторяющееся слово или неудачное начало фразы. Вторая стратегия основана на использовании внешнего модуля, производящего независимую параметрическую обработку сигнала и распознавание заданного набора озвученных пауз. Сегменты звукового сигнала, содержащие такие паузы, исключаются из последующей обработки и не подаются на вход основного декодера речи. Совместное применение стратегий для указанных корпусов позволило снизить уровень ошибок распознавания слов с 45 до 36 %.

В работе [Tsiaras 2009] применен аудиовизуальный детектор озвученных пауз для фильтрации нежелательных речевых сбоев в мультимедийных записях лекций. Записанный мультимедийный корпус лекций длительностью около 7 час. содержал изображение экрана планшетного компьютера, на котором лектор делал рукописные записи, отображаемые для слушателей на мультимедийном проекторе, а также звуковой поток с речью лектора и фоновым шумом. Анализ корпуса показал, что подавляющая часть хезитаций возникает, когда лектор не использует планшет, поэтому для фильтрации пауз применялся двухэтапный алгоритм. В первую очередь определялись моменты времени, когда изображение на экране монитора не менялось, а затем только в эти периоды времени осуществлялся поиск заполненных пауз в звуковом потоке. При анализе рассматривались озвученные паузы длительностью более 120 мс, произнесенные изолированно (т.е. те, которые содержали сегменты с тишиной до и после хезитации), а также внутри слова. Применение предварительной сегментации звуковых участков и анализ видеоизображения с планшета позволили увеличить точность распознавания хезитаций до 85 %.

Анализу неинформативных элементов в спонтанной украинской речи по корпусу выступлений депутатов и созданию системы автоматического стенографирования посвящена работа [Пилипенко, Ладошко 2010]. Среди наиболее распространенных речевых сбоев были обнаружены: заполненные паузы (49%), неверно произнесенные слова (20%), фальстарты (3%), повторы (5%), обрывы слов (3%), коррекция со вставкой (5%), онлайн-коррекция (12%), повторы со вставкой (2%) и беспорядочные слова (1%). Так как при обучении модели языка использовались текстовые материалы, не содержащие речевых сбоев, то последовательность слов, реально произносимых дикторами, часто обрабатывалась декодером речи некорректно. На месте речевого сбоя обычно распознавалось некоторое короткое функциональное слово, близкое по своим акустическим характеристикам произошедшей хезитации или коррекции.

В последующей работе авторы расширили список особенностей спонтанной украинской речи и ввели расширенную систему их разметки [Ладошко 2011]. В частности, были учтены артефакты речи (придыхание, откашливание), редукция слов, аббревиатуры, суржик. Для устранения последних трех категорий необходимо занесение их моделей в словарь системы декодирования речи. На данном этапе авторы вручную производят разметку речевых сбоев и других особенностей спонтанной речи, после их устранения надежность распознавания в среднем повышается на 6 %.

При распознавании разговорной речи необходимо отделить паралингвистические явления от ключевых слов и исключить их из дальнейшей обработки, для этого нужно создать акустические модели таких явлений. Для обучения акустических моделей внеязыковых элементов в данном исследовании был собран корпус русской речи, который содержит доклады на семинаре шести человек (трех мужчин и трех женщин). Общий объем корпуса составляет 70 мин. В ходе сегментации корпуса были выделены артефакты, заполненные паузы хезитации, повторы, самоисправления – черты, свойственные любой спонтанной речи. Для обучения и тестирования использовались только те внеязыковые элементы, которые встретились в корпусе более двух раз. Список таких элементов приведен в табл. 1.

Таблица 1

Описание моделируемых элементов спонтанной речи

Класс внеязыковых элементов	Обозначение	Внеязыковой элемент
Артефакты	ar.brth	Вздох
	ar.clth	Прочищение горла/кашель
	ar.smck	Причмокивание
Заполненные паузы	h.a	[а]
	h.ae	[аэ]
	h.au	[ау]
	h.e	[э]
	h.em	[эм]
	h.eu	[эу]
	h.m	[м]
	h.me	[мэ]
	h.mne	[мнэ]

В результате были построены модели для трех типов артефактов (вдох, прочищение горла/кашель и причмокивание) и девяти типов заполненных пауз. Каждая модель внеязыкового элемента строится на основе лево-правой скрытой марковской модели, содержащей три основных состояния.

В табл. 2 показано распределение частоты употребления различных внеязыковых элементов разными дикторами и их средняя длительность в собранном корпусе. Всего было просегментировано в корпусе 1109 внеязыковых элементов, их суммарная длительность составила 7 мин., т.е. примерно 10 % длительности всех записей выступлений докладчиков на семинаре.

Таблица 2

Описание собранного корпуса внеязыковых элементов

Диктор	Длительность выступления, мин.	Количество появлений внеязыковых элементов												
		ar.brth	ar.clth	ar.smck	h.a	h.ae	h.au	h.e	h.em	h.eu	h.m	h.me	h.mne	Всего
1	18	95	14	7	0	1	0	153	12	1	26	0	0	309
2	15	9	1	1	11	0	0	142	4	0	21	4	0	193
3	8	49	22	0	5	1	1	69	24	3	15	1	2	192
4	2	9	0	1	0	0	0	27	0	0	0	0	0	37
5	13	149	4	0	0	1	4	75	3	12	21	6	1	276
6	14	26	8	0	0	0	0	59	2	0	7	0	0	102
Общее количество появлений элементов		337	49	9	16	3	5	525	45	16	90	11	3	1109
Относительное количество, %		30,39	4,42	0,81	1,44	0,27	0,45	47,34	4,06	1,44	8,12	0,99	0,27	–
Средняя длительность (мс)		392	341	194	440	658	833	393	647	834	455	465	892	–

Из таблицы видно, что большую часть внеязыковых элементов составляет заполненная пауза h.e (47,34 % общего числа внеязыковых элементов) и вздох (30,39 %), эти элементы присутствовали в речи всех шести дикторов. Также в речи большинства дикторов присутствовали элементы ar.clth, h.em, h.m.

Были проведены эксперименты по распознаванию выявленных внеязыковых элементов. Точность распознавания всех элементов составила 86,29 %. В табл. 3 показаны результаты распознавания каждого внеязыкового элемента.

Таблица 3

Анализ результатов распознавания внеязыковых элементов

Распознаваемый элемент	Результат распознавания, %											
	ar.brth	ar.clth	ar.smck	h.a	h.ae	h.au	h.e	h.em	h.eu	h.m	h.me	h.mne
ar.brth	**96,74**	0,30	0,00	0,00	0,00	0,00	0,89	0,30	0,00	1,78	0,00	0,00
ar.clth	2,04	**91,84**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	6,12	0,00	0,00
ar.smck	0,00	0,00	**100,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
h.a	0,00	0,00	0,00	**87,50**	0,00	0,00	6,25	0,00	0,00	6,25	0,00	0,00
h.ae	0,00	0,00	0,00	0,00	**100,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00
h.au	0,00	0,00	0,00	0,00	0,00	**100,00**	0,00	0,00	0,00	0,00	0,00	0,00
h.e	2,29	1,14	0,00	2,67	0,00	0,00	**78,67**	4,00	5,90	4,57	0,76	0,00
h.em	0,00	0,00	0,00	2,22	0,00	0,00	6,67	**84,44**	0,00	6,67	0,00	0,00
h.eu	0,00	0,00	0,00	0,00	0,00	0,00	6,25	0,00	**93,75**	0,00	0,00	0,00
h.m	4,44	4,44	0,00	0,00	0,00	0,00	5,56	2,22	0,00	**83,33**	0,00	0,00
h.me	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**100,00**	0,00
h.mne	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	**100,00**

Из таблицы видно, что точность распознавания шести элементов (ar.smck, h.ae, h.au, h.me, h.mne) составила 100 %. Хуже всего распознавалась заполненная пауза h.e, точность распознавания которой оказалась равна 78,67 %. Этот элемент путался с элементами: ar.brth, ar.clth, h.a, h.em, h.eu, h.m, h.me. Точность распознавания ниже 90 % оказалась у элементов h.a, h.em, h.m, при распознавании эти элементы путались между собой.

Проведенные эксперименты показали достаточно высокий процент распознавания внеязыковых элементов. В дальнейшей работе планируется увеличить количество моделируемых внеязыковых элементов и провести эксперименты по проверке правильности отделения внеязыковых элементов от ключевых слов.

Примечание

[1]Работа выполнена в рамках НИР СПбГУ № 31.37.103.2011.

Список литературы

Антонова А.В. Смех, юмор и речевая манипуляция // Вестн. Перм. ун-та. Российская и зарубежная филология. 2010. Вып.4(10). С.52–58.

Большая советская энциклопедия: в 30 т. / под общ. ред. А.М. Прохорова. М.: Сов. энциклопедия, 1975. Т.19. 648 с.

Земская Е.А., Ширяев Е.Н. Русская разговорная речь: итоги и перспективы исследования // Русистика сегодня. М., 1988. С.121–152.

Карпова Ю.А. Средства выражения эмотивно-эмпатийного взаимодействия в условиях речевого общения // Вестн. Перм. ун-та. Российская и зарубежная филология. 2011. Вып. 4(16). С.73–79.

Кипяткова И.С., Карпов А.А. Автоматическая обработка и статистический анализ новостного текстового корпуса для модели языка системы распознавания русской речи // Информ.-управляющие системы. СПб: СПбГУАП, 2010. №4(47). С.2–8.

Ладошко О.Н. Разметка спонтанной украинской речи // Электроника и связь: тем. вып. Электроника и нанотехнологии. 2011. №1. С.97–103.

Леонтьева Т.И., Датчер Л.А., Филиппова О.В. Вербально-невербальный контекст ложных высказываний // Вестн. Перм. ун-та. Российская и зарубежная филология. 2011. Вып. 2(14). С.101–110.

Лингвистический энциклопедический словарь / под ред. В.Н.Ярцевой. М.: Сов. энциклопедия, 1990. 685 с.

Пилипенко В.В., Ладошко О.Н. Аннотация и учет речевых сбоев в задаче автоматического распознавания спонтанной украинской речи // Искусств. интеллект. 2010. №3. С.238–248.

Подлесская В.И., Кибрик А.А. Коррекция в устной русской монологической речи по данным корпусного исследования // Рус. яз. в науч. освещ. 2006. №2(12). С.7–55.

Подлесская В.И., Кибрик А.А. Самоисправления говорящего и другие типы речевых сбоев как объект аннотирования в корпусах устной речи // Науч.-техн. информ. Сер.2. 2007. №2. С.2–23.

Риехакайнен Е.И. Влияние потенциального контекста на распознавание изолированных омофонов // Вестн. Перм. ун-та. Российская и зарубежная филология. 2010. Вып.4(10). С.40–45.

Ронжин А.Л., Карпов А.А., Кагиров И.А. Особенности дистанционной записи и обработки речи в автоматах самообслуживания // Информ.-

управляющие системы. СПб.: ГУАП, 2009. №5(42). С.32–38.

Ронжин А.Л., Евграфова К.В. Анализ вариативности спонтанной речи и способов устранения речевых сбоев // Изв. вузов. Гум. науки. 2011. Т.2, вып.3. С.227–231.

Рудык Н.Н. К проблеме толкования понятия «спонтанная речь» // Наука и образование. Одесса (Украина). 2010. №4–5. С.188–191.

Филиппова Н.С. Операции отмены как способ организации спонтанной речи (на материале устных спонтанных монологов-описаний) // Материалы XXXVI междунар. филол. конф. Вып. 20: Полевая лингвистика. Интегральное моделирование звуковой формы естественных языков. 12-17 марта 2007 г. / отв. ред. А.С.Асиновский, Н.В.Богданова. СПб., 2007. С.86–90.

Штерн А.С. Перцептивный аспект речевой деятельности: (Экспериментальное исследование). СПб.: Изд-во С.-Петерб. ун-та, 1992. 236 с.

Stouten F., Duchateau J., Martens J.-P., Wambacq P. Coping with disfluencies in spontaneous speech recognition: acoustic detection and linguistic context manipulation // Speech Communication. 2006. Vol. 48. P.1590–1606.

Tsiaras V., Panagiotakis C., Stylianou Y. Video and audio based detection of filled hesitation pauses in classroom lectures // Proc. of the 17th European Signal Processing Conference (EUSIPCO 2009). Glasgow, Scotland, August 24–28, 2009. P.834–838.

SEGMENTATION OF PARALINGUISTIC PHONATION PHENOMENA IN SPONTANEOUS RUSSIAN SPEECH

Irina S. Kipyatkova
Research Fellow of Laboratory of Experimental Phonetics
Saint Petersburg State University

Vasilisa O. Verkhodanova
Master student of Phonetics and Foreign Languages Teaching Department
Saint Petersburg State University

Andrey L. Ronzhin
Senior Researcher of Laboratory of Experimental Phonetics
Saint Petersburg State University

In the article paralinguistic phonation phenomena that may occur in spontaneous speech are analysed; a review of methods of registration of such phenomena, when automatic speech is recognized, is given in the article. For training the acoustic models of extralinguistic elements the corpus of spontaneous Russian speech was segmented; artifacts (breath, cough and smack) and filled pauses of hesitation were extracted; acoustic models of extralinguistic elements that occurred in the corpus more than twice were built. High-fidelity recognition of extralinguistic elements in the collected corpus was 86 %.

Key words: paralinguistic phenomena; spontaneous speech; speech failures; speech segmentation; automatic speech recognition.