На правах рукописи

Соломенник Анна Ивановна

Оценка качества селективного синтеза речи: методы и результаты

Специальность 10.02.21 – «Прикладная и математическая лингвистика»

Автореферат

диссертации на соискание учёной степени

кандидата филологических наук

Москва – 2016

Работа выполнена на кафедре теоретической и прикладной лингвистики филологического факультета ФГБОУ ВО «Московский государственный университет имени М. В. Ломоносова».

Научный руководитель: доктор филологических наук, ведущий научный сотрудник **Кривнова Ольга Фёдоровна**

Официальные оппоненты: доктор филологических наук **Скрелин Павел Анатольевич**, профессор, заведующий кафедрой фонетики и методики преподавания иностранных языков ФГБОУ ВПО «Санкт-Петербургский государственный университет»

кандидат филологических наук **Кузнецов Владимир Борисович**, профессор кафедры прикладной и экспериментальной лингвистики ФГБОУ ВПО «Московский государственный лингвистический университет»

Ведущая организация: ФГБУН «Санкт-Петербургский институт информатики и автоматизации Российской академии наук»

Защита состоится в 14.30 1 июня на заседании диссертационного совета Д.501.001.24 при ФГБОУ ВО «Московский государственный университет им. М. В. Ломоносова» по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ имени М. В. Ломоносова, 1-й учебный корпус, филологический факультет.

С диссертацией можно ознакомиться в научной библиотеке Московского государственного университета имени М. В. Ломоносова, а также на официальном сайте филологического факультета ФГБОУ ВО «Московский государственный университет им. М. В. Ломоносова» по ссылке: http://www.philol.msu.ru/~ref/001_24_14.htm.

Автореферат разослан « » _____ 2016 г.

Учёный секретарь
диссертационного совета,
доктор филологических наук А. М. Белов

ОБЩАЯ ХАРАКТЕРИСТИКА ДИССЕРТАЦИИ

Диссертация посвящена методам оценки качества одного из современных видов синтеза речи – т. н. селективного синтеза. В данном исследовании с различных сторон рассматривается проблема оценки качества синтезированной речи, анализируются существующие методы оценки и предлагается система оценочных методов, адаптированная именно под селективный синтез речи. На основе предложенной системы осуществлено комплексное тестирование современных русскоязычных селективных синтезаторов речи, по результататам которого произведена общая оценка их эффективности и соответствия предъявляемым требованиям.

Актуальность работы состоит в том, что селективный синтез речи, в англоязычных источниках называемый unit selection, в настоящее время является общепризнанным методом получения качественной синтезированной речи, наиболее близкой по звучанию к естественной. Этим обусловлено то, что при разработке большинства современных синтезаторов, особенно коммерческих приложений, используется именно данный метод. В этой связи при оценке качества синтезированной речи необходимо обратить особое внимание на специфические особенности звучания речи, связанные с его использованием.

В области оценки качества синтезированной речи существует множество исследований, однако в данный момент для современных русскоязычных синтезаторов **степень разработанности проблемы** невелика: пока нет единой общепринятой системы оценки. Отдельные исследования либо несколько устарели, так как в них рассматриваются синтезаторы второго поколения (аллофонный и дифонный конкатенативный синтез), либо не обладают достаточной полнотой. Таким образом, очевидна необходимость разработки и описания новой системы оценки, учитывающей особенности именно современных методов синтеза речи.

Цель исследования состоит в том, чтобы разработать комплекс методов оценки качества селективного синтеза речи на русском языке.

Задачи исследования:

1. Описать существующие проблемы и методы оценки качества синтеза речи с анализом и обобщением результатов предыдущих исследований.

2. Обосновать необходимость специального подхода к оценке селективного синтеза с учётом его характерных особенностей.

3. Предложить методы оценки синтеза речи, позволяющие объективно оценивать и сравнивать современные русскоязычные селективные синтезаторы речи.

4. На основе предложенных методов провести тестирование и комплексную оценку нескольких современных русскоязычных синтезаторов.

Научная новизна работы заключается в том, что впервые для русского языка были предложены и опробованы новые методы оценки современных селективных синтезаторов речи.

Теоретическая значимость заключается в анализе и выявлении специфических характеристик селективного синтеза речи с точки зрения особенностей качества речи, порождаемой таким синтезатором.

Практическая значимость работы состоит в том, что появляется возможность использовать предложенные методы для оценки и сравнения между собой современных синтезаторов речи высокого качества. На основании полученных результатов могут быть предложены различные средства улучшения качества синтезированной речи.

Предметом исследования данной работы являются методы комплексной оценки качества синтеза речи.

Объект исследования – искусственно порождённая речь, её характеристики с точки зрения восприятия слушающими, критерии качества синтезированной речи.

Материалом исследования является синтезированная речь, полученная с использованием нескольких современных русскоязычных селективных синтезаторов (Acapela, iSpeech, Ivona TTS, Mary TTS, Loquendo TTS, Nuance Vocalizer, VitalVoice TTS).

Теоретико-методологическую основу исследования составили работы по синтезу речи Б. М. Лобанова, О. Ф. Кривновой, А. Блэка, П. Тейлора, Я. ван Сантена, и др.; работы по общей фонетике Л. В. Бондарко, Л. Р. Зиндера, С. В. Кодзасова и О. Ф. Кривновой.

В ходе работы были использованы следующие **методы**: методы слухового, аудиторского и инструментального анализа фонограмм, статистические методы анализа результатов проведённых экспериментов.

На защиту выносятся следующие **положения**:

1. Для оценки селективного синтеза речи необходим специальный подход, учитывающий специфические особенности данной речевой технологии.
2. Предложенный в диссертации подход и его оценочные средства позволяют проводить комплексное диагностическое тестирование современных русскоязычных синтезаторов селективного типа и сравнивать их между собой на объективной основе.
3. Максимальное влияние на естественность звучания селективного синтеза речи для русского языка оказывают ошибки, связанные с выбором неправильного места ударения в словах и неадекватной интонацией.

4. Ошибки и недочеты в лингвистической обработке текста перед его фонетизацией являются основным источником качественных различий в работе современных селективных синтезаторов русской речи.

Достоверность результатов обеспечивается успешным практическим применением предложенной системы методов оценки в экспериментах по тестированию нескольких современных систем селективного синтеза речи.

Апробация работы. Основные положения диссертационной работы докладывались на научно-методических конференциях: «Международная конференция по компьютерной лингвистике «Диалог 2009» (Москва), «Международная конференция по компьютерной лингвистике «Диалог 2010» (Москва), «Международная конференция по компьютерной лингвистике «Диалог 2012» (Москва), «Конференция AINL 2013: Искусственный интеллект и естественный язык» (Санкт-Петербург), «Актуальные вопросы теоретической и прикладной фонетики: конференция к юбилею О. Ф. Кривновой» (Москва, 2013), «Международная конференция по компьютерной лингвистике «Диалог 2013» (Москва), «15th International Conference on Speech and Computer SPECOM 2013» (Чехия), «XXI Международная конференция студентов, аспирантов и молодых ученых «Ломоносов» (Москва, 2014), «2nd International Scientific Conference «Contemporary Research in Phonetics and Phonology: Methods, Aspects and Problems» (Латвия, 2015). Диссертация прошла обсуждение на кафедре теоретической и прикладной лингвистики филологического факультета МГУ имени М. В. Ломоносова.

Структура диссертации. Диссертация изложена на 195 страницах и состоит из введения, четырех глав и заключения. Список литературы содержит 104 наименования. Работа иллюстрирована 21 рисунком и 22 таблицами. В 19 приложениях содержатся тестовые тексты и подробные результаты экспериментов.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во **введении** формулируются основные цели и задачи исследования, даётся обоснование актуальности выбранной темы диссертации, её научной новизны, теоретической и практической значимости, а также приводятся положения, выносимые на защиту.

В **первой** главе рассматривается история динамики целей и задач синтеза речи, требований к его качеству в процессе развития данной речевой технологии. Даётся краткое описание различных подходов к синтезу речи и особенностей речи, порождаемой разными видами синтезаторов. В последнем разделе первой главы описываются перспективы дальнейшего развития синтеза речи и повышения требований к качеству синтезированной речи.

Вторая глава «Селективный синтез речи» описывает особенности способа синтеза, исследуемого в диссертации.

В **первом разделе** второй главы подробно описывается базовый селективный алгоритм выбора звуковых единиц для синтеза речи по произвольному тексту (unit selection). Селективный синтез речи является разновидностью конкатенативного синтеза, то есть при генерации речевого сигнала используются заранее полученные звукозаписи естественной речи. В отличие от более ранних аллофонных и дифонных синтезаторов речи, порождающих выходной речевой сигнал из отдельных и специально подготовленных звуковых единиц, выделенных из небольшого и тщательно подобранного набора слов, при селективном синтезе для каждой целевой единицы речи производится выбор наиболее подходящего кандидата из множества вариантов, взятых из озвученных диктором предложений естественного языка. Для этого записываются специальные речевые базы, размер которых может достигать нескольких десятков часов звучания от одного диктора. В процессе акустического синтеза селективный алгоритм строит оптимальную

последовательность звуковых единиц, учитывая одновременно и то, насколько кандидат подходит под описание необходимых характеристик целевого звука (стоимость замены), и то, насколько хорошо выбранные кандидаты будут конкатенироваться с соседними единицами (стоимость связи). При этом с учетом указанных стоимостей из речевой базы в качестве оптимальных могут быть выбраны не отдельные звуки, а их цепочки или даже целые предложения. Такой подход позволяет минимизировать необходимость вынужденных модификаций речевого сигнала, что повышает естественность синтезируемой речи.

Во **втором** разделе описываются современные русскоязычные селективные синтезаторы речи, и даётся их краткая характеристика. Селективные синтезаторы для русского языка начали разрабатываться в 2000-х годах. В настоящий момент существует довольно много разнообразных синтезированных коммерческих и бесплатных «голосов», работающих с различными операционными системами. Приведём таблицу наиболее известных разработок с краткими комментариями (табл. 1). Большая часть этих разработок (кроме Mary TTS) является коммерческими системами, в которых для прослушивания и оценки синтеза речи доступны только онлайн демо-версии с различными ограничениями.

Таблица 1. Основные селективные синтезаторы для русского языка

Система синтеза	**Доступные голоса**	**Онлайн-демо и ограничения**	**Комментарии**
Acapela Group (Бельгия, Франция, Швеция)	Alyona	http://www.acapela-group.com/voices/demo/ (не более 300 символов)	–
Apple (США)	Siri	Нет демо для Windows, работает под ОС iOS как голос персонального помощника	–
Google Translate (США)	Женский голос	https://translate.google.com/#ru	Озвучивание переводимого текста. Чтение очень замедлено, слышна модификация речи.
iSpeech (США)	Женский голос;	http://www.ispeech.org/text.to.speech (не более 150 символов)	Есть возможность регулирования

	мужской голос		скорости произнесения (3 варианта).
Ivona TTS (Польша, США)	Tatyana; Maxim	http://www.ivona.com/ (не более 250 символов)	–
Mary TTS (Германия, бесплатное открытое ПО)	Мужской голос	http://mary.dfki.de:59125/	Есть возможность включения/отключения просодической модификации. Не читает числовые записи и иноязычные вставки.
Microsoft (США)	Irina	Экранный диктор в ОС Windows	–
Nuance Loquendo (Италия, США)	Olga; Dmitri	http://www.nuance.com/for-business/by-solution/customer-service-solutions/solutions-services/inbound-solutions/loquendo-small-business-bundle/interactive-tts-demo/index.htm (не более 500 символов, фоновая музыка)	–
Nuance Vocalizer (США)	Katya; Milena; Yuri	http://www.nuance.com/landing-pages/playground/Vocalizer_Demo2/vocaLizer_modal.html?demo=true	–
ReadSpeaker (Нидерланды, Бельгия и др.)	Женский голос	http://www.readspeaker.com/ (не более 250 символов; фоновая музыка; добавление фразы про демо)	–
Svox (Швейцария, США)	Katja; Yuri	http://svoxmobilevoices.wordpress.com/demos/ (только образцы синтеза) Нет интерактивного демо (для Windows), работает под ОС Android.	–
Tingwo (Швейцария)	Женский голос; Мужской голос	http://www.tingwo.co/en/interactive-tts-text-to-speech-demo (не более 200 символов; фоновая музыка)	–
VitalVoice TTS (ООО «ЦРТ») (Россия, С.-Петербург)	Юлия; Владимир; Анна; Виктория; Александр; Мария; Лидия	http://www.speechpro.ru/technologies/synthesis (не более 200 символов); прослушивание через «голосовые открытки» http://cards.voicefabric.ru/ (фоновая музыка).	–

Третий раздел второй главы описывает общую структуру современного селективного синтезатора речи типа «Текст-Речь». Оценка качества работы синтезаторов речи часто выполняется для отдельных этапов преобразования текста в речь (иными словами, модулей синтезатора). В **подразделах** данного раздела приводится краткое описание примерной структуры селективных синтезаторов. Для удобства описания блок лингвистической обработки разделен на три части: собственно лингвистическая обработка (нормализация текста, расстановка ударений); просодическая обработка (локализация пауз и определение типов интонационных конструкций); фонетическая обработка (построение сегментной транскрипции и задание параметров для интонации). В последнем подразделе описывается блок акустической обработки (выбор единиц из речевой базы, модификация речевого сигнала).

В **третьей**, центральной, главе описываются существующие методы и способы оценки качества синтезированной речи, предлагается структура системы оценки качества селективного синтеза речи, даётся обоснование необходимости разработки соответствующих оценочных методов.

В **первом** разделе третьей главы обсуждаются общие критерии качества и задачи оценки качества синтезированной речи. Среди задач, для решения которых может применяться система оценки качества синтезированной речи можно назвать следующие:

1. *Тестирование системы синтеза в процессе её разработки.* Главная задача такого тестирования связана с последующим улучшением различных параметров оцениваемой системы. В этом случае к системе оценки предъявляются следующие требования: она должна быть автоматической или полуавтоматической, т. е. функционировать без участия или с минимальным участием человека; иметь достаточно высокое быстродействие. Для оценочного анализа должны быть доступны результаты всех этапов синтеза, и проверка

может осуществляться с использованием промежуточной информации, генерируемой системой в явном виде.

2. *Оценка собственной системы синтеза речи в сравнении с конкурентами.* Основной задачей такого тестирования является сравнение разных систем синтеза с разными голосами. Для этого может применяться как автоматическая дикторонезависимая оценка, так и оценка экспертов. В данном случае может быть затруднен доступ к результатам синтеза: для коммерческих приложений обычно доступны только интерактивные демо-версии, при помощи которых можно получить образцы синтезированной речи довольно низкого качества, с фоновой музыкой или другими ограничениями в целях защиты от коммерческого использования, или же доступны только заранее подготовленные примеры звучания. Для корректного сравнения результатов работы синтезаторов необходимо использовать их полнофункциональные версии.

3. *Участие в конкурсах, проводимых независимыми компаниями.* В таких условиях система оценки может быть не автоматической, но автоматизированной. Для оценки синтезаторов в этом случае могут привлекаться большие человеческие ресурсы (например, заинтересованные пользователи интернета). При этом, хотя внутренняя структура систем синтеза и останется закрытой, имеется возможность получения промежуточных результатов работы системы в унифицированном виде, при заинтересованности в этом самих участников конкурса. Системы синтеза могут тестироваться на одной и той же речевой базе, на основе которой строится синтезированный голос. Может использоваться также эталонная оценка при сравнении с диктором-донором.

Общепринятыми мерами качества синтезированной речи являются оценки её разборчивости и естественности. Под «качеством» речи чаще всего понимается её естественность, то есть величина, характеризующая субъективную оценку звучания синтезированной речи по сравнению со звучанием естественной речи.

Методы оценки качества синтеза можно разделить на две большие группы: субъективные и инструментальные. В отдельную промежуточную группу можно

также выделить те методы, которые требуют участия человека для детектирования наличия/отсутствия ошибки (например, неправильного ударения), а не субъективной оценки речи по определённому параметру (например, естественности). Такие методы обычно используются для тестирования отдельных модулей синтеза: нормализации текста, расшифровки сокращений, чтения иноязычных вставок, расстановки ударений, фонетической обработки и т. п.

Во **втором** разделе данной главы рассматриваются методы оценки разборчивости речи. Существует ряд тестов для проверки разборчивости речи, порождаемой системами синтеза. Разборчивость определяется относительным количеством правильно распознанных элементов речи. Можно выделить несколько типов проверок в зависимости от длины речевых отрезков, подаваемых для тестирования, и задач, которые ставятся перед испытуемыми. Различают звуковую (фонемную), слоговую, словесную и фразовую разборчивость. Можно с определённой долей уверенности утверждать, что проблема разборчивости речи для современных синтезаторов в целом решена. Это означает, что, несмотря на возможную неразборчивость отдельных ошибочно синтезированных слов или неправильно расшифрованных сокращений или аббревиатур, общий смысл синтезированных предложений и текстов остаётся понятным.

Третий раздел посвящён методам оценки естественности синтезированной речи. К субъективным методам, позволяющим оценить степень естественности речи с точки зрения человека, её воспринимающего, относятся разного рода тесты и опросники, заполняемые экспертами-специалистами, либо наивными слушателями, носителями синтезируемого языка. В них используется так называемая MOS-оценка (Mean Opinion Score или «метод мнений»), производимая по пятибалльной шкале по нескольким категориям: общее впечатление, слуховое усилие, естественность, понимание смысла сообщения, темп, разборчивость, приятность голоса. Проведение подобных тестов является довольно трудоёмкой задачей, и для того, чтобы ускорить процесс оценки и

сделать его более доступным для тестирования системы в процессе её разработки, создаются различные инструментальные (или объективные) методы оценки качества синтеза. Адекватность инструментальной оценки анализируется также с учетом того, насколько она совпадает с субъективными оценками испытуемых. Чисто инструментальные методы интегральной оценки для сравнения существующих речевых синтезаторов широкого применения пока не имеют и в основном используются в процессе их разработки и для автоматизации настройки параметров синтезаторов, однако последние исследования говорят о том, что задача адекватного автоматического вычисления оценки качества синтеза вполне осуществима.

В настоящий момент международные «соревнования» синтезаторов Blizzard Challenge[1] являются своеобразным эталоном по оценке качества систем синтеза речи. Их задачей является сравнение результатов работы различных систем синтеза, причем синтезированные голоса создаются на основе одних и тех же звуковых баз, предоставляемых организаторами перед началом соревнований.

Для оценки русскоязычных синтезаторов чаще всего используется ГОСТ Р 50840-95. Наряду с этим используются различные тесты отдельных составляющих компонентов (модулей лингвистической обработки, модификации, полноты и качества речевой базы), но единого стандарта оценки, рассчитанного на современные синтезаторы и методы синтеза, пока нет.

В **четвёртом** разделе главы «Методы оценки качества селективного синтеза речи» обсуждаются факторы, влияющие на восприятие синтезированной речи человеком. К ним в первую очередь относятся:

1. Конкретные условия, связанные с выполняемой задачей.
2. Ограничения, присущие системе обработки информации, которой обладает человек.

[1] http://www.festvox.org/blizzard/

3. Опыт и тренировка слушателя.
4. Лингвистическая структура сообщения.
5. Качество записи речевого сигнала и внешние условия восприятия (громкость, шум, реверберация, посторонние разговоры и т. п.).

В **пятом** разделе обсуждаемой главы обосновывается необходимость адаптации общих методов оценки качества к селективному синтезу речи. Речь, порождаемая селективным синтезатором, имеет специфические особенности. Это, прежде всего, неравномерность распределения мест с неудачным звучанием: нередко отдельная фраза или её часть звучит гораздо естественнее остальных, а при стыковке «гладких» участков появляются помехи. Указанные особенности связаны с базовым алгоритмом выбора звуковых единиц. Кроме того, разработчики синтезаторов часто стараются минимизировать или вовсе устранить использование просодической модификации выбираемых звуков-кандидатов под требуемые значения, что может приводить к непредсказуемости и нарушению просодического оформления фраз. Неестественное звучание отрезка синтезированной речи может возникнуть из-за отсутствия нужной целевой единицы в речевой базе. Все вышеперечисленные особенности необходимо учитывать при разработке и составлении тестов для оценки качества речи, получамой методом селективного синтеза.

При тестировании селективного синтеза особенно важным является раздельное тестирование лингвистической обработки текста для озвучивания и собственно акустического модуля синтеза сигнала, так как особенности алгоритма селективного синтеза часто предполагают возможность частичного или даже полного несоответствия характеристик выбираемых единиц-кандидатов характеристикам, предсказанным системой на этапе лингвистической обработки. В тестах следует разграничивать причины возникновения ошибок в лингвистической обработке, связанные с работой лингвистического процессора, и ошибки, появившиеся вследствие неудачно подобранных звуковых элементов для конкатенации. Например, из-за неправильной длительности выбранных

гласных звуков ударение в синтезированном слове может смещаться на другой слог.

Для селективного синтеза невозможно, как, например, для простого аллофонного или дифонного конкатенативного синтезатора, составить тест, содержащий все или большинство элементов его речевой базы для тестирования их звучания, так как сегментные единицы языка (фонемы и их аллофоны) в базе будут представлены не одним, а, возможно, сотнями или даже тысячами вариантов. При этом объём материала для тестирования должен быть достаточно большим и разнообразным, включать в себя различные темы и жанры. В то же время это не исключает и использования специально сконструированных текстов, например, на сложные с фонетической точки зрения сочетания звуков. Если синтезатор предполагается использовать для какой-то специфической задачи (например, чтения аудиокниг, озвучивания действий пользователя ПК или разговора с «искусственным» оператором по телефону), тесты обязательно должны быть составлены с учётом такого сценария использования.

В **шестом** разделе данной главы рассматривается общая структура системы комплексной оценки селективного синтеза речи. Этот раздел разбит на подразделы, соответствующие оценке отдельных модулей синтезатора (лингвистической, фонетической и акустической обработки), а также интегральной оценке качества синтезированной речи. Даются конкретные рекомендации по составлению и проведению сравнительного и диагностического тестирования синтезаторов.

В **четвёртой,** экспериментальной, главе описаны эксперименты и тесты, проведённые в диссертационном исследовании по оценке качества нескольких современных систем селективного синтеза речи на русском языке, даётся анализ полученных результатов. При проведении экспериментов материалом послужила синтезированная речь, полученная с использованием нескольких современных русскоязычных синтезаторов речи (Acapela, iSpeech, Ivona TTS, Mary TTS,

Loquendo TTS, Nuance Vocalizer, VitalVoice TTS). Большинство из них являются коммерческими программами, что накладывает определенные ограничения на длительность и качество тестируемых речевых записей.

Первый эксперимент посвящён оценке влияния различных типов ошибок на общее качество синтезированной речи. В нём делается попытка оценить, какие ошибки наиболее распространены в современных селективных синтезаторах высокого качества и какие из них вызывают наибольшие проблемы при восприятии синтезированной речи, заставляя слушающих оценивать её как менее естественную. В проведённом эксперименте для оценки качества и естественности русской синтезированной речи были выбраны два «голоса» современных синтезаторов последнего поколения: «Tatyana» польской компании Ivona и «Анна» петербургского ООО «ЦРТ». На основе анализа предыдущих исследований были выделены следующие категории возможных ошибок:

1) неверное словесное ударение;
2) неверное произнесение (замена/выпадение/добавление лишнего звука);
3) неправильные паузы (отсутствие/лишние, слишком короткие/длинные);
4) плохой темп/ритм;
5) неровная/неверная интонация;
6) нарушения плавности речи (дефекты в речевом сигнале): прерывистость, скачки, «бульканье», стук и т. п.;
7) общее качество голоса;
8) иное.

В качестве тестового материала был использован фонетически представительный текст[2], включающий в себя описательную и диалоговую части, что позволило лучше оценить адекватность интонационного оформления синтезированной речи.

[2] Смирнова Н. С., Хитров М. В. Фонетически представительный текст для фундаментальных и прикладных исследований русской речи // Изв. вузов. Приборостроение. — 2013. — Вып. 2. — С. 5–10.

Общее количество ошибок, отмеченных аудиторами, оказалось примерно одинаковым. Оба образца синтезированной речи (голоса) также получили примерно одинаковую среднюю оценку естественности: 3,9 и 4,1 соответственно (по пятибалльной шкале). Из проведенного теста можно сделать вывод, что ошибки в интонационном оформлении синтезированной речи являются главной проблемой современных русскоязычных селективных синтезаторов. Также следует отметить, что тестируемые синтезаторы отличаются по качеству лингвистической и акустической обработки, причём несколько большее влияние на ухудшение естественности имеют ошибки, связанные с неправильной постановкой словесного ударения и неверной транскрипцией.

В следующем **втором** разделе обсуждается серия из нескольких тестов модуля лингвистической обработки. Анализируется точность выделения предложений, оценка чтения аббревиатур, цифровых обозначений, специальных символов, иностранных слов на латинице и правильности определения места ударения. Общие результаты исследований, описанных в данном разделе, являются в некоторой степени ожидаемыми. При выполнении задач лингвистической обработки текста, связанных непосредственно с русским языком, лучшие показатели получены для системы синтеза VitalVoice TTS, разрабатываемой в России, в первую очередь для русского языка. Однако при чтении иноязычных вставок и специальных символов самые хорошие результаты показывает синтезатор компании Acapela Group, что, по всей видимости, связано с тем, что соответствующие модули могли быть встроены в русскоязычный синтез из более разработанных языков. Некоммерческий голос системы Mary TTS ожидаемо показал самый плохой результат, не справившись с большинством задач. Среди остальных коммерческих систем показатели качества варьируются в зависимости от конкретной задачи.

Третий раздел четвёртой главы посвящён оценке модуля фонетической обработки. В данном разделе блоки просодической и фонетической обработки объединены в один, так как при исключительно аудитивной оценке интонации

звучащей синтезированной речи без доступа к результатам работы соответствующих модулей невозможно протестировать правильность интонационной транскрипции независимо от результирующих физических характеристик речи. В данном разделе приводятся результаты тестов правильности сегментной транскрипции, паузирования и интонации синтезированной речи.

Лучшие результаты с минимальным количеством ошибок в тесте на правильность сегментной транскрипции, как и в тестах из предыдущего раздела, у синтезаторов VitalVoice TTS и Acapela TTS. Доля правильно локализованных пауз для протестированных синтезаторов при чтении художественного текста приближается к 100 %, в то время как правильность мелодического оформления синтезированной речи в среднем для протестированных синтезаторов составляет 58 %, что связано в первую очередь с тем, что эксперименты проводились именно с селективными синтезаторами (ЧОТ синтезированной речи в которых может не точно соответствовать смоделированной). Для точной оценки правильности интонационной транскрипции необходимы промежуточные данные синтезаторов, недоступные при тестировании коммерческих систем с использованием демо-версий. Также следует отметить, что, в отличие от оценки лингвистической обработки, нельзя обобщать данные, полученные для одного голоса, на всю систему синтеза. Например, для различных голосов синтезатора VitalVoiceTTS процент фраз с ошибками варьируется от 34 до 41, что может объясняться как размером речевой базы для конкретного голоса, так и особенностями чтения конкретного «диктора-донора».

В **четвёртом** разделе «Оценка акустической обработки» даются рекомендации по оценке модуля выбора звуковых элементов из речевой базы и модуля модификации, провести которые не представляется возможным с использованием демо-версий синтезаторов.

Заключительный **пятый** раздел посвящен обсуждению интегральной оценки качества синтезированной речи, описывается эксперимент по сравнительной оценке качества речи по ГОСТ Р 50840-95 «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости», а также приводятся данные по зависимости оценок от «знакомства» аудиторов с синтезом речи.

Ниже в таблице 2 приведены общие результаты по всем экспериментам, описанным в **четвёртой** главе. На основании диагностических тестов отдельных модулей синтезатора нельзя определить наилучшую систему синтеза речи, так как для различных задач использования синтезаторов, критичными могут оказаться различные показатели, даже те, которые не обсуждались в данной главе (например, степень устойчивости синтеза речи к шумовым помехам или качество речи в телефонном канале). Даже система Mary TTS, показавшая самый плохой результат практически по всем тестам, обладает одним несомненным преимуществом – открытым доступом к исходному коду программы и, следовательно, возможностью улучшения и настройки синтезатора под определённую задачу. В то же время проведённые в настоящей работе тесты указывают на слабые стороны, которые требуют той или иной доработки для различных систем синтеза и могут существенно улучшить их качество и эффективность.

Таблица 2. Сводная таблица результатов тестов, проведенных для оценки успешности выполнения системами синтеза речи различных задач

Система синтеза	**Acapela**	**iSpech**	**Ivona**	**Mary**	**Loquendo**	**Nuance Vocalizer**		**VitalVoice**		
Голос	Alyona	Female voice	Tatyana	Male voice	Olga	Katya	Milena	Анна	Влади-мир	Юлия
Выделение предложений (%)	-	-	100	-	-	-	-	100	-	-
Графические сокращения (%)	52	28	32	0	-	40	-	-	-	**79**

Аббревиатуры (%)	82	74	78	-	-	75	-	-	-	**99**
Цифровые обозначения (%)	62	59	63	0	-	62	-	-	-	**83**
Специальные символы (%)	**95**	33	71	0	-	43	-	-	-	81
Английские слова (%)	**100**	15	26	0	-	74	-	-	-	93
Омографы (%)	66	79	57	46	-	61	-	-	-	**98**
Транскрипция (%)	**88**	44	63	44	-	50	-	-	-	**88**
Места пауз (%)	-	-	100	-	-	-	-	100	-	-
Точность интонации (%)	72	-	-	-	70	-	71	-	**77**	**77**
Естественность интонации (%)	59	-	-	-	49	-	65	-	**72**	70
Качество по ГОСТу (баллы)	-	-	-	-	-	-	-	-	-	4,5

В **заключении** даётся краткая характеристика основных разделов работы, приводятся основные результаты, полученные в ходе рамках диссертационного исследования, обсуждаются перспективы дальнейшей разработки темы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИОННОЙ РАБОТЫ

В рамках диссертационного исследования получены следующие основные **результаты**:

1. Произведено описание и анализ существующих методов оценки качества синтезированной речи.

2. Обоснована необходимость специального подхода к оценке селективного синтеза речи, учитывающая его специфические особенности.

3. Предложена система методов оценки, адаптированных для селективного синтеза речи.

4. Подготовлены тестовые тексты и опросники для проведения комплексной оценки различных модулей синтезаторов речи на основе предложенной в работе системы оценочной процедуры.

5. Произведено комплексное тестирование русскоязычных селективных синтезаторов и получены оценки качества речи, синтезируемой с их использованием.

Дальнейшие исследования по данной тематике наиболее актуальны в следующих направлениях: оценка качества статистического параметрического синтеза речи, инструментальная автоматизированная оценка качества синтезаторов речи, оценка выразительности и эмоциональности синтезированной речи, оценка точности воспроизведения особенностей речи конкретного диктора-донора. Полезным направлением могло бы стать также проведение независимого конкурса синтезаторов речи на материале русского языка, что, к сожалению, трудно осуществить, так как большинство разработчиков современных синтезаторов речи высокого качества являются зарубежными компаниями, для которых разработка и совершенствование русскоязычного синтеза не всегда является приоритетной задачей.

По теме диссертации **опубликованы следующие работы:**

1. **Соломенник А. И. Структура системы оценки качества синтезированной русской речи // Структурная и прикладная лингвистика. — Вып. 10. — СПб, 2013. — С. 251–266.**
2. **Соломенник А. И. Технология синтеза речи: история и методология исследований // Вестник Московского университета. Сер. 9. Филология. — 2013. — № 6. — С. 149–162.**

3. **Соломенник А. И., Чистиков П. Г., Рыбин С. В., Томашенко Н. А. Автоматизация процедуры подготовки нового голоса для системы синтеза русской речи // Изв. вузов. Приборостроение. Тематический выпуск "Речевые информационные системы". — 2013. — №2. — С. 29–32.**

4. **Соломенник А. И., Таланов А. О., Соломенник М. В., Хомицевич О. Г., Чистиков П. Г. Оценка качества синтезированной речи: проблемы и решения // Изв. вузов. Приборостроение. Тематический выпуск "Речевые информационные системы". — 2013. — №2. — С. 38–42.**

5. **Чистиков П. Г., Корольков Е.А., Таланов А. О., Соломенник А. И. Гибридная технология синтеза русской речи на основе скрытых Марковских моделей и алгоритма Unit Selection // Изв. вузов. Приборостроение. Тематический выпуск "Речевые информационные системы". — 2013. — №2. — С. 33–38.**

6. **Чистиков П. Г., Таланов А. О., Захаров Д. С., Соломенник А. И. Технология синтеза естественной речи с использованием базы данных небольшого объема // Научно-технический вестник информационных технологий, механики и оптики. — №4 (91) — 2014. — С. 83–97.**

7. **Solomennik A. I., Cherentsova A. E. A Method for Auditory Evaluation of Synthesized Speech Intonation // Miloš Železný et al. (Eds.): SPECOM 2013, Lecture Notes in Artificial Intelligence 8113. — Springer, 2013. — P. 9–16.**

8. Продан А. И., Корольков Е. А., Опарин И. В., Таланов А. О. Особенности использования многоуровневой разметки звукового корпуса // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). – М.: РГГУ, 2009. — С. 415–419.

9. Продан А. И., Таланов А. О., Чистиков П. Г. Система подготовки нового голоса для системы синтеза «VitalVoice» // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 26–30 мая 2010 г.). Вып. 9 (16). — М.: Изд-во РГГУ, 2010. — С. 394–399.

10. Соломенник А. И. Зависимость естественности звучания синтезированной речи от наличия ошибок различных типов // Актуальные проблемы филологической науки: взгляд нового поколения. Доклады участников XX–XXI Международных конференций студентов, аспирантов и молодых ученых «Ломоносов». Секция «Филология». Вып. 6. — Изд. Московского университета, 2015. — С. 475–480.

11. Соломенник А. И. Особенности оценки качества селективного синтеза речи. Актуальные вопросы теоретической и прикладной фонетики // Сборник статей

к юбилею О. Ф. Кривновой / Под ред. А. В. Архипова, И. М. Кобозевой, Кс. П. Семёновой. — М.: ООО «Буки-Веди», 2013. — С. 336–341.

12. Соломенник А. И. Ошибки и дефекты синтезированной речи: типы, частотность и влияние на естественность звучания // Материалы Международного молодежного научного форума «ЛОМОНОСОВ-2014» / Отв. ред. А. И. Андреев, Е. А. Антипов. [Электронный ресурс] — М.: МАКС Пресс, 2014. — 1 электрон. опт. диск (CD-ROM).

13. Соломенник А. И. Технология синтеза речи в историко-методологическом аспекте. Речевые технологии. — №1, — 2013. — С. 42–57.

14. Solomennik A. An influence of defects in synthesized speech on its naturalness // 2nd International Scientific Conference CONTEMPORARY RESEARCH IN PHONETICS AND PHONOLOGY: METHODS, ASPECTS AND PROBLEMS. Abstracts [Электронный ресурс]. — Riga, 2015. — P. 22. — Режим доступа: http://www.lulavi.lv/media/upload/tiny/files/Abstracts_%20Phon%202015.pdf.

15. Solomennik A., Chistikov P. Automatic generation of text corpora for creating voice databases in a Russian text-to-speech // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». — М.: Изд-во РГГУ, 2012. — Вып.11 (18). — С. 607–615.

16. Solomennik A. I., Chistikov P. G. Evaluation of naturalness of synthesized speech with different prosodic models // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог». — М.: Изд-во РГГУ, 2013. — Вып. 12 (19). — Т. 2. — С. 31–38.

Подписано в печать: 19.03.2016
Объем: 1,0 усл.п.л.
Тираж: 100 экз. Заказ № 1539
Отпечатано в типографии «Реглет»
125315, г. Москва, Ленинградский проспект д. 74, корп. 1
+7(495) 790-47-77 www.reglet.ru