

**Федеральное государственное бюджетное учреждение науки
Санкт-Петербургский институт информатики и автоматизации
Российской академии наук (СПИИРАН)**

199178, Санкт-Петербург, 14 линия, д.39. Факс: (812) 328-4450; Тел.: (812)328-3311; E-mail: spiiiran@iias.spb.su

УТВЕРЖДАЮ

Директор СПИИРАН

д.т.н., член-корр. РАН

Юсупов Р. М.



«14» апреля 2016 г.

ОТЗЫВ

ведущей организации

о диссертации СОЛОМЕННИК Анны Ивановны

«ОЦЕНКА КАЧЕСТВА СЕЛЕКТИВНОГО СИНТЕЗА РЕЧИ:

МЕТОДЫ И РЕЗУЛЬТАТЫ», представленной на соискание ученой

степени кандидата филологических наук по специальности 10.02.21 –

«Прикладная и математическая лингвистика»

Диссертация А. И. Соломенник посвящена проблеме оценки качества синтезированной компьютером речи. В фокусе исследования находятся современные русскоязычные синтезаторы речи высокого качества, использующие технологию селективного синтеза речи. Диссертация имеет междисциплинарный характер и объединяет исследования в области прикладной и математической лингвистики, экспериментальной фонетики, информатики и автоматизированных речевых технологий.

Актуальность данной работы обусловлена, прежде всего, тем, что селективный (англ. “unit selection”) синтез речи в настоящее время является одним из основных методов получения качественной синтезированной речи, близкой по звучанию к естественной. Таким образом, обращение к специфическим особенностям звучания речи, получаемой при помощи селективного метода, является одной из основных задач при оценке качества работы современных синтезаторов речи.

Новизна исследования заключается в том, что для русского языка впервые были предложены и применены методики комплексной оценки современных селективных синтезаторов речи, а также предложена система тестовых данных, используемых в методе. Тот факт, что автор учитывает комплексное влияние различных факторов на качество звучания синтезированной речи, а также отечественный и зарубежный опыт подобных исследований, придает рецензируемой работе теоретическую значимость.

Возможность использовать предложенные методы для оценки и сравнения между собой современных синтезаторов русской речи высокого качества обуславливает практическую значимость работы. Кроме того, в рассматриваемой диссертации предложены конкретные рекомендации по тестированию и сравнению синтезаторов, а также приводятся использованные в экспериментах тестовые материалы и опросники для проведения комплексной оценки лексических и акустических модулей синтезаторов речи.

К достоинствам диссертации А. И. Соломенник можно отнести четкую постановку цели и задач исследования, среди которых следует отметить описание существующих проблем и методов оценки качества синтеза речи с анализом и обобщением результатов предыдущих исследований, обоснование необходимости специального подхода к оценке селективного синтеза с учетом его характерных особенностей,

разработку методов оценки синтеза речи, позволяющих объективно оценивать и сравнивать современные селективные синтезаторы русской речи и апробацию предложенных методов для комплексной оценки нескольких современных русскоязычных синтезаторов.

Несомненным достоинством диссертационной работы является выбор и объем речевого материала для исследования, включающий синтезированную речь, полученную с использованием нескольких современных русскоязычных селективных синтезаторов (Acapela, iSpeech, Ivona TTS, Mary TTS, Loquendo TTS, Nuance Vocalizer, VitalVoice TTS), общим объемом более 4 часов речи.

Структурно работа состоит из введения, четырех глав, заключения, обширного списка литературы из 104 наименований и 19 приложений, содержащих разработанные автором тестовые текстовые материалы и подробные результаты экспериментов.

Во введении к диссертации дается краткая характеристика работы, обосновывается ее актуальность, новизна, теоретическая и практическая значимость.

В главе 1 «История синтезаторов речи и современные разработки» рассматривается теоретическая база, лежащая в основе компьютерной технологии синтеза речи, а также история создания синтезаторов речи и требований к их качеству. Обсуждается современное состояние данной речевой технологии, перспективы её развития и изменения в методах оценки качества синтезированной речи.

Глава 2 «Селективный синтез речи» описывает особенности способа синтеза, исследуемого в диссертации, в ней также приводятся данные по современным русскоязычным селективным синтезаторам, описывается обобщенная структура селективного синтезатора.

В основной главе 3 «Методы оценки качества селективного синтеза речи» описываются существующие методы и способы оценки качества

синтезированной речи, предлагается структура системы оценки качества селективного синтеза речи, дается обоснование необходимости разработки соответствующих оценочных методов. Автор приводит рекомендации по составлению и проведению сравнительного и диагностического тестирования синтезаторов.

В главе 4 «Результаты оценки качества селективного синтеза речи» описаны выполненные автором эксперименты и тесты, проведенные в диссертационном исследовании по оценке качества нескольких современных систем селективного синтеза русской речи, дается анализ полученных результатов. При проведении экспериментов материалом послужила синтезированная речь, полученная с использованием нескольких современных русскоязычных синтезаторов речи, большинство из которых являются коммерческими программами, что наложило определенные ограничения на длительность и качество тестируемых речевых записей. Основное содержание этой главы посвящено серии из 12 экспериментов по оценке качества работы различных модулей синтезаторов. На основании полученных результатов автор делает вывод о том, что при использовании диагностических тестов отдельных модулей синтезатора невозможно однозначно определить лучшую систему синтеза русской речи, так как для разных задач использования синтезаторов, критичными могут оказаться различные показатели. В то же время проведенные тесты указывают на слабые стороны доступных систем, которые требуют той или иной доработки. Подобные тесты могут существенно улучшить качество систем синтеза речи, что также является практическим результатом работы. Отметим достаточно большое количество графиков и таблиц в данной части работы, в которых представлены результаты проведенных экспериментов; их наличие в значительной степени облегчает восприятие текста.

При всех своих неоспоримых достоинствах диссертация не лишена ряда недостатков, которые следует указать:

1. Из диссертации следует, что автор, к сожалению, не знакома с ГОСТ Р 52873-2007 «Синтезаторы речи для специальных компьютерных рабочих мест для инвалидов по зрению. Технические требования», который применяется в РФ с 2008 г. и описывает основные характеристики и технические требования, предъявляемые к компьютерным синтезаторам русской речи.

2. Для проведения исследований привлекалась пара десятков аудиторов-носителей русского языка, однако, среди них не было людей с нарушениями зрения и незрячих, а, как известно из практики, такие люди обладают более развитой слуховой системой, чем обычные люди, и более тонко воспринимают неточности в синтезированном звуковом сигнале.

3. Среди основных критериев оценки качества синтезаторов речи автор не выделила возможность настройки параметров компьютерной системы под пользователя, хотя для многих людей оказывается необходимым изменение стандартных значений темпа речи, частоты основного тона, интонации и т.д. для лучшего восприятия синтезируемой компьютером речи.

4. Считаю, что не совсем точно сформулирована тема диссертации «Оценка качества селективного синтеза речи: методы и результаты», так как в мире уже разработано достаточно много различных методов для оценки качества синтеза речи и получено множество результатов для ряда мировых языков, автор же занималась только оценкой синтезаторов русской речи.

5. В диссертации отсутствует подраздел 4.6.

Представленные выше замечания являются существенными, но не снижают общей актуальности и научной значимости представленных в диссертации методик оценки и полученных результатов, что позволяет оценить диссертационную работу А. И. Соломенник положительно.

Автором опубликовано 16 научных работ в 2009-2015 гг., включая 7 статей в журналах и изданиях из перечня ВАК Минобрнауки РФ как филологического, так и технического профиля. Автореферат и опубликованные статьи адекватно отражают содержание диссертационного исследования.

В исследовании получены новые и оригинальные результаты, которые существенно расширяют имеющиеся представления об оценке качества русскоязычного селективного синтеза речи. Диссертация соответствует всем критериям, сформулированным в п. 9 и п. 10 «Положения о порядке присуждения ученых степеней» ВАК Министерства образования и науки РФ и является научно-квалификационной работой, в которой содержится решение задачи, имеющей существенное значение для области прикладной лингвистики, фонетики и речевых технологий, а ее автор, Анна Ивановна Соломенник, заслуживает присуждения ей искомой степени кандидата филологических наук по специальности 10.02.21 – «Прикладная и математическая лингвистика».

Отзыв подготовлен заведующим лабораторией речевых и многомодальных интерфейсов СПИИРАН, доктором технических наук А. А. Карповым.

Доклад Соломенник Анны Ивановны был заслушан на расширенном семинаре лаборатории речевых и многомодальных интерфейсов СПИИРАН, протокол семинара № 1 от «04» апреля 2016 года.

Заведующий лабораторией речевых и
многомодальных интерфейсов СПИИРАН
доктор технических наук, доцент



А.А. Карпов

Сведения о составителе отзыва ведущей организации:

Карпов Алексей Анатольевич, доктор технических наук, доцент, заведующий лабораторией речевых и многомодальных интерфейсов СПИИРАН, 199178, Санкт-Петербург, 14-я линия, дом 39, <http://www.spiiras.nw.ru>, karpov@iias.spb.su, +7 812 3280421

Публикации составителя отзыва, близкие теме диссертации соискателя:

Карпов А. А., Железны М. Двужычная многомодальная система для аудиовизуального синтеза речи и жестового языка по тексту // Научно-технический вестник информационных технологий, механики и оптики. СПб: ИТМО, № 5, 2014, С. 92-98.

Карпов А. А., Верховданова В. О. Речевые технологии для малоресурсных языков мира // Вопросы языкознания. М.: Наука, № 3, 2015, С. 117-135.

Karpov A., Markov K., Kipyatkova I., Vazhenina D., Ronzhin A. Large vocabulary Russian speech recognition using syntactico-statistical language modeling // Speech Communication. Elsevier, 2014, vol. 56, pp. 213-228.