

ОТЗЫВ

официального оппонента

о диссертации А.И. Соломенник

«Оценка качества селективного синтеза речи: методы и результаты»,
представленной на соискание учёной степени кандидата филологических наук
по специальности 10.02.21 – Прикладная и математическая лингвистика.

Диссертация А.И. Соломенник посвящена важной и очень актуальной в настоящее время теме: методике оценки качества синтезированного речевого сигнала. Синтезированная речь занимает всё более заметное место в нашей жизни, поэтому оценка качества работы систем синтеза речи не столько по технологическим критериям (быстродействие, ресурсоемкость, разборчивость, гибкость и т.п.) сколько по потребительским свойствам порождаемого ими сигнала – естественности и даже «дружелюбности» с точки зрения слушателя – имеет исключительную практическую ценность. Определение же объективных свойств синтезированного сигнала, которые позволяют дать количественную оценку его качественных характеристик представляет собой непростую теоретическую проблему, которую диссертант успешно решает своим исследованием. Работа выполнена на материале разных моделей селективного синтеза речи. По общему мнению, в настоящее время селективный синтез речи обеспечивает максимальное сходство синтезированного сигнала с естественным, поэтому исследователь может выявить те характеристики искусственного сигнала, которые демонстрируют неполноту наших представлений о процессах моделирования и порождения речи человеком.

Диссертация состоит из Введения, четырех глав, Заключения, списка используемой литературы и большого количества приложений, которые не только хорошо иллюстрируют ход исследования, но могут иметь и самостоятельную высокую практическую ценность в будущем.

Первая глава посвящена истории синтеза: от первых говорящих машин Кратценштейна и фон Кемпелена в 18 веке до современных систем синтеза речи по тексту. Естественно, основное внимание автора обращено на исследования и разработки 20-го века. В разделе «XX век: синтезаторы первого поколения», теоретическую базу которых заложило исследование Гуннара Фанта «Акустическая теория речеобразования», автор достаточно подробно описывает артикуляторный и формантный синтезаторы, останавливается и на разработках известных отечественных ученых.

В разделе «XX век: синтезаторы второго поколения» приводится краткое описание принципов компилятивного или (в терминологии автора) конкатенативного синтеза речи. Эти разработки уже могут называться системами синтеза речи, поскольку включают в себя не только генерацию звукового сигнала на основе текстового сообщения, но полноценную лингвистическую обработку этого сообщения, в ходе которой моделируется

его произнесение человеком. Поскольку в таких моделях для генерации звукового сигнала используются фрагменты речи реального диктора, система синтеза может воспроизводить не только его индивидуальный тембр, но и особенности произношения и интонацию. Искусственные речевые сигналы, порождаемые такими системами, отличаются большим сходством с естественной речью, поэтому, как справедливо отмечает автор, «стало возможным говорить об отдельной оценке различных модулей синтезатора: обработки текста, качестве просодического оформления синтезированной речи и т. п.» (с. 32). Я бы к этому добавил, что эта раздельная оценка стала и необходимой, поскольку недостатки синтезированной речи могут быть вызваны как ошибками построения модели речевого сигнала: определения места ударения, членения текста, выбора интонационного оформления, транскрипции и т.п. («адекватностью» модели в нашей терминологии); так и искажениями, нежелательными артефактами при генерации самого звукового сигнала, которые связаны с процедурами конкатенации и модификации физических характеристик звуковых единиц. Собственно, одной из причин перехода к синтезаторам следующего поколения было стремление улучшить качество звучания порождаемого сигнала.

Далее в первой главе дается краткое описание синтезаторов третьего поколения: селективной и параметрической системам синтеза, и прогноз дальнейшего развития синтеза речи.

В Главе 2 диссертации дается подробное описание селективного синтеза речи: алгоритма Unit selection, известных синтезаторов русской речи такого типа и основных модулей, из которых состоят подобные системы синтеза. Автор описывает задачи каждого блока моделирования свойств синтезируемого сигнала, поскольку от эффективности работы каждого из них зависит воспринимаемое качество звучащего результата. В описание работы каждого блока включены и известные способы решения стоящих перед ними задач.

В третьей главе формулируются требования к качеству синтезированной речи и методам его оценки. Автор показывает, что к известным методам оценки качества передачи речевого сигнала по каналам связи, которые до сих пор применяются и в оценке работы синтезаторов, необходимо добавить и методы оценки степени близости синтезированной речи к естественной. Общая оценка качества синтезированного сигнала человеком складывается из комплекса различных характеристик, часть из которых, отражающая адекватность построенной модели речевого сообщения, может быть выражена количественно, например, числом ошибок. Это касается нормализации текста, определения места разных типов ударения, фонетической и интонационной транскрипции, наличия звуковых искажений, помех и т.д. Другая часть характеристик носит субъективный характер, например, естественность, приятность голоса, полнота понимания текста, тем не менее, такие свойства сигнала тоже требуют объективной оценки. Автор подробно анализирует факторы, влияющие на качество синтезированной речи, рассматривает известные методы его оценки, предлагает свои критерии и способы, диагностики эффективности работы всех компонентов системы.

В Главе 4 дается подробное изложение результатов применения предложенных в предыдущей главе критериев и методов оценки качества синтезированного сигнала для оценки и сопоставления потребительских свойств известных систем селективного синтеза

русской речи. В зависимости от возможности использования тестового материала для оценки разных модулей системы синтеза речи в экспериментах приняли участие от 2 до 7 различных синтезаторов, общий объем синтезированного материала составил более 4 часов, что обеспечивает высокую достоверность полученных результатов.

Эксперименты показали, что на оценку качества и естественности синтезированной речи самое сильное воздействие оказывают ошибки интонационного оформления

Раздел «Оценка лингвистической обработки» дал точную и достоверную картину эффективности и качества обработки текста исследованными синтезаторами. Практически по всем аспектам обработки со значительным отрывом победил синтезатор Центра речевых технологий (Санкт-Петербург). Представленные в диссертации результаты показывают, на какие проблемы необходимо обратить внимание разработчикам конкурирующих систем, чтобы соперничать с лидером.

Раздел «Оценка фонетической обработки» показывает сопоставление точности сегментной транскрипции (победители – ЦРТ и Asapela значительно опережают конкурентов), методику оценки паузации (небольшие расхождения между двумя участниками) и очень интересные данные по оценке интонации реализованной 4 синтезаторами и диктором-женщиной. Автор очень точно понимает сложность оценки интонационного оформления высказывания, на которую влияет, как правильный выбор нужной интонационной модели, так ее физическая реализация. Поэтому оценка производится по двум критериям: адекватности модели и естественности ее реализации. Блестящее решение, которое позволило получить действительно полную и объективную картину возможностей современных синтезаторов русской речи.

Оценка акустической обработки для селективных синтезаторов мало добавляет в общую картину качества синтеза речи.

В разделе «Интегральная оценка качества синтезированной речи и оценка общего качества синтеза» показаны специфические особенности оценки разборчивости в применении к селективным синтезаторам, но основное внимание обращено на оценку естественности синтезированного сигнала. Поскольку результаты, представленные в предыдущих разделах диссертации, показали ведущую роль интонационного оформления высказывания в оценке качества синтеза, то здесь автор демонстрирует и тестирует ряд возможных подходов к моделированию интонации для синтеза. Предложенные варианты отличаются объемом речевого корпуса (от 20 мин. до 6 часов) и способами автоматической генерации интонационного оформления. Все предложенные варианты, разработанные ЦРТ, обеспечивают попадание в I класс качества, при этом наилучший результат обеспечивает расчет интонации на основе скрытых марковских моделей на корпусе объемом 2,5 часа с ручной коррекцией.

В Заключении автор подводит итоги проведенного исследования, результаты которого, а именно, тестовые материалы, критерии и методы оценки эффективности и корректности работы всех модулей систем синтеза речи, позволили протестировать качество известных русскоязычных селективных синтезаторов. Автор формулирует и перспективы дальнейших исследований в области автоматизации оценки качества автоматического синтеза речи, среди которых, оценка экспрессивности и

эмоциональности сигнала, сохранности и воспроизведения индивидуальных особенностей конкретного диктора.

Дополнительным результатом исследования А.И. Соломенник оказалось то, что проведенное исследование и тестирование очень точно охарактеризовало современное состояние синтеза русской речи, которое у нас в стране соответствует самым лучшим зарубежным образцам.

В целом диссертационное исследование А.И. Соломенник производит благоприятное впечатление. Очевиден большой объем выполненной автором работы, корректность экспериментов и тщательность обработки результатов. Работа написана хорошим языком, грамотно оформлена технически, снабжена информативными иллюстрациями и приложениями. Тем не менее, хотелось бы высказать несколько замечаний, комментариев и вопросов.

В первой главе следовало бы упомянуть о таком интересном направлении, как биомеханическое порождение речевого сигнала, которое, с одной стороны, является «реинкарнацией» первых говорящих машин и новым вариантом артикуляторного синтеза, а, с другой, его качество в перспективе может быть оценено предложенными автором методами также точно, как и система селективного синтеза речи.

Для оценки статистической значимости различий в работе синтезаторов и достаточности объема тестового материала в ряде разделов главы 4 был применен критерий хи-квадрат К. Пирсона, который используется при нормальном распределении, но характер распределения указан не был. При этом в основном автор использует уровень значимости $p < 0,01$. Почему в разделах «Оценка чтения цифровых обозначений» и «Оценка сегментной транскрипции» делается исключение и используется 5%-ный уровень значимости, автор не поясняет.

Неясно, что означает «разметка откорректирована вручную» (с. 126), особенно при использовании СММ.

На с. 120, видимо, произошел сбой нумерации разделов, поскольку автор отсылает читателя к несуществующему разделу 4.6, в то время, как после раздела 4.5 следует 4.7.

В разделе 3.3 «Методы оценки естественности речи» указаны некоторые известные методы оценки и даже проводимые соревнования синтезаторов, но нет упоминания о пан-европейском проекте COST-258 «Естественность синтезированной речи», который выполнялся в 1996-2000 гг. под руководством Эрика Келлера. В 2002 г. по результатам проекта вышла коллективная монография «Improvements in Speech Synthesis».

Отмеченные недостатки никак не умаляют достоинств диссертационного исследования А.И. Соломенник, а ее диссертация является самостоятельным научным исследованием, представляющим безусловный научный интерес и высокую практическую и прикладную значимость.

Автореферат и публикации соответствуют содержанию диссертации и отражают ее основные положения. Диссертации А.И. Соломенник «Оценка качества селективного синтеза речи: методы и результаты», выполненная под руководством доктора филологических наук, ведущего научного сотрудника О.Ф. Кривновой, является

законченной научно-квалификационной работой, в которой содержится решение задачи, имеющей существенное значение для прикладной лингвистики. На основании параграфов 9 и 10 Положения о присуждении ученых степеней можно утверждать, что А.И. Соломенник заслуживает присуждения ей искомой ученой степени кандидата филологических наук по специальности 10.02.21 – Прикладная и математическая лингвистика.

Официальный оппонент

Доктор филологических наук,

профессор кафедры фонетики и методики

преподавания иностранных языков

Санкт-Петербургского государственного университета

П.А. Скредин

12.05.2016

Личную подпись удостоверяю



Сведения об оппоненте:

Скредин Павел Анатольевич, p.skrelin@spbu.ru

Сайт: <http://phonetics.spbu.ru/>

Тел. +7 812 3289565

Доктор филологических наук, профессор кафедры фонетики и методики преподавания иностранных языков Федерального государственного бюджетного образовательного учреждения высшего образования "Санкт-Петербургский государственный университет"

199034, Санкт-Петербург, Университетская наб. д.7-9.

Публикации:

P. Skrelin. Allophone- and Suballophone-Based Concatenative Speech Synthesis System for Russian. // Lecture Notes in Computer Science, 2000. — Vol. 1902, — P. 271-276

P. Skrelin. Naturalness of Synthesized Speech: Choice of Basic Sound Units // International workshop SPECOM'2000, SPb, 2000. P. 9 – 12

P. Skrelin. Sub-Allophone Synthesis of Russian Vowels // 100 лет экспериментальной фонетике в России. СПб, 2001. Сс. 155 – 158

Pavel Skrelin Russian Intonation: New Aspects of Acoustic and Perceptual Features // Proceedings of the 1-st International Conference "Phonetics without Borders", June 13-15, 2011 — Благовещенск, — 2011. — P. 110-115