

УТВЕРЖДАЮ

Генеральный директор

ООО «ЦРТ»



Д.В. Дырмовский

2017г.

ОТЗЫВ

ведущей организации на диссертационную работу Томчука Кирилла Константиновича **«Сегментация речевых сигналов для задач автоматической обработки речи»**, представленную на соискание ученой степени кандидата технических наук по специальности 05.12.13 – Системы, сети и устройства телекоммуникаций

1. Актуальность темы

Обеспечение временной сегментации речевого сигнала на фрагменты, которым присущи определенные сигнальные особенности, является важной задачей при разработке речевых систем различной направленности. Наличие информации о принадлежности анализируемого фрагмента сигнала к тому или иному типу сегмента позволяет применять для его обработки специализированные алгоритмы.

В силу сложности информационной структуры речевого сигнала, его большой вариативности, связанной с индивидуальными особенностями диктора, изменчивостью акустических условий внешней среды и характеристик каналов приема-передачи речи, существующие алгоритмы сегментации не всегда обеспечивают достаточную точность и надежность. Кроме того,



операции сегментации речевого сигнала обладают высокой вычислительной сложностью, что в ряде случаев не позволяет обеспечить обработку сигналов в реальном масштабе времени.

В связи с этим диссертационная работа Томчука Кирилла Константиновича, посвященная решению научно-технической задачи эффективной многоуровневой временной сегментации сигналов для речевых систем различного назначения является актуальным научным исследованием.

2. Степень обоснованности научных положений, выводов и рекомендаций

Автором изучены и обобщены результаты исследований отечественных и зарубежных ученых по проблеме временной сегментации речевых сигналов, использованию сегментации в системах автоматической обработки речи.

Рассмотрен спектр задач, стоящих перед алгоритмами сегментации, показана возможность применения сегментации в речевых системах разного назначения. Отмечена важность учета сигнальных особенностей речи на русском языке. Обоснована необходимость комплексного подхода к решению задачи сегментации в связи с целесообразностью выделения разных уровней сегментов для разных речевых задач.

По результатам анализа имеющейся информации о существующих методах сегментации и конкретных реализациях алгоритмов, диссертант выделил основные уровни сегментации, которые необходимы для работы речевых систем либо могут быть дополнительно применены для увеличения их эффективности. Среди наиболее востребованных в речевых системах самостоятельных типов сегментации выделена сегментация на уровне речь/пауза, а также сегментация на отдельные периоды колебаний голосовых связок в звонких звуках.

Для подтверждения выдвинутых гипотез и анализа эффективности разработанных алгоритмов автором проведены экспериментальные исследования с применением речевой базы данных TIDIGITS, речевых

фонограмм с разными уровнями и типами фоновых шумов естественной природы, разными условиями оцифровки, разными голосами дикторов. Приведено сравнение разработанных алгоритмов с известными современными аналогичными решениями, в результате которого обоснованно определены преимущества и недостатки предложенных автором решений.

Для получения экспериментальных данных использованы современные средства и методики проведения исследований. В работе диссертант грамотно использует методы математического анализа, экспертных оценок, статистической обработки результатов эксперимента, соответствующие характеру фактического материала и поставленным целям и задачам исследования.

3. Новизна и достоверность основных научных результатов

Детальное рассмотрение диссертационной работы Томчука К.К. позволяет сделать вывод о том, что автором выбран и решен круг задач, обладающих новизной, получен ряд оригинальных научных результатов. Содержание работы характеризуется узкой направленностью исследований и разработок, достоверность которых подтверждается соответствующими документами о внедрении результатов, наличием подробно описанных методик и детально разработанных алгоритмов.

Полученные диссертантом результаты непротиворечивы, аргументированно обоснованы, базируются на грамотном применении положений и методов проектирования и анализа программных средств, теории вероятностей и математической статистики, цифровой обработки сигналов, фонетики и психоакустики, а также согласованностью полученных автором результатов с известными решениями в области речевых систем.

Список использованной литературы состоит из 137 наименований, половина из которых относится к зарубежным публикациям и материалам конференций на английском языке.

4. Личный вклад

Тексты диссертации и автореферата убедительно свидетельствуют о большой самостоятельной работе, проделанной диссертантом по исследуемой теме, начиная от постановки задачи до практической реализации с использованием современных программных средств различного уровня.

Личный вклад автора в разработанные методы и алгоритмы сегментации речевого сигнала не вызывает сомнений. Это подтверждается и наличием достаточного числа выступлений на научных конференциях и публикаций по рассматриваемой тематике.

5. Теоретическая значимость

Теоретическая значимость работы обусловлена ее новизной и заключается:

- в постановке и решении задачи создания комплекса алгоритмов, позволяющих осуществлять временную сегментацию речевого сигнала на необходимые для работы речевой системы уровни;

- в развитии методов повышения шумоустойчивости параметризации речевого сигнала мел-частотными кепстральными коэффициентами путем учета особенностей механизмов речеобразования (физика колебаний голосовых связок и прохождения воздушного потока через полости речевого тракта) и речевосприятия человека (эффект одновременной слуховой маскировки);

- в обосновании целесообразности применения временной сегментации речевого сигнала для увеличения эффективности работы речевых систем путем использования отдельных подалгоритмов для обработки фрагментов сигнала, отнесенных к определенным типам сегментов.

6. Практическая значимость и внедрение результатов

Практическая значимость полученных диссертантом результатов определяется тем, что они способствуют эффективной работе речевых систем за счет использования предложенных в работе методов и алгоритмов временной сегментации речевого сигнала.

Дополнительно стоит отметить:

– исследовательское программное обеспечение, включающее базу данных, позволяет осуществлять многоаспектный анализ речевых сигналов и их фрагментов, включая сохранение и воспроизведение нескольких вариантов ручной сегментации одной фонограммы;

– разработанный алгоритм модификации темпа произнесения речи может иметь самостоятельное применение для решения ряда практических задач, в которых требуется ускоренное или замедленное произведение фонограмм.

Результаты диссертационной работы Томчука К.К. обсуждались на международных и российских конференциях, опубликованы в открытой печати, в том числе в ряде рецензируемых изданий, входящих в перечень ВАК. Объем и содержание представленных публикаций по теме исследования в целом отражают основные положения, вынесенные на защиту. Содержание автореферата в целом отражает содержание диссертации.

Научные результаты, полученные в работе Томчука К.К., обладают новизной. Они создают основу для проектирования речевых систем повышенной эффективности за счет использования результатов автоматической временной сегментации сигналов. Содержание и структура диссертации логически согласованы и соответствуют поставленной цели исследования.

7. Рекомендации по использованию результатов и выводов диссертации

Результаты работы рекомендуются к использованию научными группами, занимающимися исследованиями в области речевых информационных систем, организациями, разрабатывающими программное обеспечение для колл-центров, речевого управления различными устройствами. Результаты диссертации рекомендуется также использовать в ВУЗах в рамках учебных курсов по экспериментальной фонетике, компьютерной лингвистике, речевым технологиям.

8. Замечания по диссертации

Из недостатков работы можно отметить следующее:

1. При проведении анализа фонетической структуры сигнала, представляющего русскую речь, автор руководствовался только трудами технических специалистов, обойдя стороной работы известных российских ученых лингвистов и акустиков.

2. В главе 1 не достаточно подробно проанализированы современные системы автоматической сегментации и принципы их работы. В том числе не рассмотрено построение систем автоматической сегментации на основе методов машинного обучения.

3. Рассуждения о простой и сложной структуре основного тона в разделе 2.3.5 диссертации сводятся к хорошо известной диаграмме кодирования гласных 1-й и 2-й формантами. Не указано, что для корректной оценки ZCR необходимо убрать постоянную составляющую и наводки, то есть осуществить полосовую фильтрацию сигнала. Кроме того, приведенная формула (2.5) дает оценку ZCR, зависящую от частоты дискретизации сигнала. Необходимо было бы использовать нормализованную меру ZCR, которая не зависит от частоты дискретизации сигнала.

4. На стр. 145 автор приводит классификацию признаков, используемых для идентификации дикторов. В группе статистических признаков не отмечены формантные признаки (см., например, монографию Абрамовой И.Е. «Идентификация личности билингва по иностранному акценту». М., 2012). Также не раскрывается состав признаков, принадлежащих к декларируемой автором смешанной группе.

5. Среди существующих решений, выбранных для сравнения с разработанным алгоритмом модификации темпа произнесения, отсутствует модификатор темпа на основе алгоритма PSOLA, в котором также используются результаты сегментации на периоды основного тона, сравнение с которым было бы более показательным.

6. В разделе 4.7.4 «Анализ эффективности алгоритма модификации темпа речи» непонятен критерий, по которому были выбраны программные средства для сравнения. Так, например, отсутствует известный звуковой редактор "Adobe Audition". Кроме того, неясно, что из себя представляет "экспертная группа", которая проводила сравнения и сделала заключение о почти двухкратном превосходстве предложенного автором диссертации алгоритма по сравнению с известными. Тем более, что объем тестовой выборки составлял всего три фонограммы, две из которых пониженной разрядности (7 и 8 бит), а длительность звучания фонограмм всего 4-7 секунд.

7. Существует ряд терминологических проблем: например, в тексте используются одновременно термины «сонограмма» и «спектрограмма», не поясняя в чем их отличие; вводится термин вейвлет-спектрограмма, хотя это не что иное, как скейлограмма; попеременно используются термины идентификация и верификация дикторов, например на стр. 145 пишется, что «На рисунке 4.2 представлена обобщенная функциональная схема алгоритма идентификации диктора», а на самом рисунке пишется «Рисунок 4.2 – Обобщенная функциональная схема алгоритма верификации диктора».

8. Присутствует ряд весьма неудачных формулировок, например, «достигнутый уровень решения речевых задач в значительной степени не достигает уровня, с которым эти задачи решаются организмом человека».

9. Отсутствуют ссылки на публикации и сравнения с аналогичными предложенным решениями от ведущего предприятия ООО «ЦРТ», которые используются в коммерческих решениях, являющихся лидерами мирового рынка речевых технологий.

Однако указанные недостатки не носят принципиального характера и не снижают научной и практической значимости диссертационной работы.

9. Заключение

Диссертация Томчука Кирилла Константиновича является целостной и логически завершенной научно-квалификационной работой, в которой на основе выполненных автором исследований получено решение актуальной научно-технической задачи, состоящей в сегментации речевого сигнала на требуемые для эффективной работы конечного речевого приложения типы сегментов: паузы/смычки, шумные, вокализованные, взрывные звуки и отдельные периоды колебаний голосовых связок.

Диссертационная работа отвечает требованиям, установленным п. 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 № 842, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждение ученой степени кандидата технических наук по специальности 05.12.13 – «Системы, сети и устройства телекоммуникаций».

Доклад Томчука Кирилла Константиновича был заслушан на семинаре научно-технического совета ООО «ЦРТ»; протокол семинара № 3/17 от «13» апреля 2017 года.

Заключение составил:

к.т.н. Поляков Алексей Юрьевич

Секретарь НТС

Руководитель управления НИОКР

ООО «ЦРТ»

