

**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: spiiran@iiias.spb.su

---

**ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА**

на диссертационную работу Томчука Кирилла Константиновича на тему «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»

ГУАП  
№ 74-1033/17-0-0  
от 16.05.2017



**Актуальность**

Автоматическая обработка речи является востребованной во многих сферах деятельности человека. При этом важным этапом предварительной обработки речевого сигнала является его временная разметка на сегменты, обладающие характерными признаками. Например, разметка на сегменты активной речи и участки тишины (Voice Activity Detection, VAD) играет ключевую роль в экономии трафика при передаче аудиоданных по каналу связи. Значения формантных частот звуков речи характеризуют особенности произнесения и общую форму речевого тракта конкретного диктора, поэтому выделение вокализованных сегментов с последующим формантным анализом широко используется в биометрических задачах верификации и идентификации диктора. Сегментация вокализованных звуков на отдельные циклы колебаний голосовых связок позволяет отслеживать траекторию частоты основного тона диктора, осуществлять изменение интонационных характеристик речи, а также применяется в методах синтеза и сжатия речевых сигналов. Исследованию вопросов автоматической сегментации речевых сигналов посвящена кандидатская диссертационная работа Томчука К.К.

Следует отметить, что цифровая обработка разговорной речи является очень сложной задачей, так как речевой сигнал характеризуется значительной меж- и внутридикторской вариативностью, а также наличием формируемых

диктором неинформативных элементов речевого сигнала. Кроме того, говорящий часто находится в сложных акустических условиях, что также оказывает отрицательное влияние на работу алгоритмов обработки речевых сигналов. Все данные факты обуславливают сложность создания обладающих высокими показателями точности, надежности и скорости работы алгоритмов сегментации речи.

Указанные обстоятельства свидетельствуют об актуальности разработки методов, алгоритмов и программных средств, направленных на обеспечение эффективной сегментации речевого сигнала на требуемые для конкретной речевой задачи временные фрагменты. Поэтому тема данной диссертации, несомненно, является *актуальной*.

### **Научная новизна**

Научная новизна диссертационной работы характеризуется следующими основными результатами:

1) Проведен анализ основных параметров звуков речи для русского языка, на основе которого предложена классификация звуков речи с их соотношением к классам речевых сегментов.

2) Разработан и исследован производительный алгоритм сегментации на периоды основного тона, основанный на отборе локальных максимумов временной функции речевого сигнала.

3) Для повышения эффективности параметризации речевого сигнала мел-частотными кепстральными коэффициентами (MFCC) предложена модификация базового алгоритма вычисления признаков, заключающаяся в дополнении его психоакустической моделью одновременной слуховой маскировки и в увеличении в дискретном спектре мощности вокализованных сегментов значений коэффициентов, соответствующих гармоникам основного тона диктора.

4) Разработан и исследован алгоритм модификации в широких пределах темпа (как ускорения, так и замедления) речевого сигнала, основанный на



многоуровневой временной сегментации сигнала с последующей обработкой выделенных сегментов с разными локальными коэффициентами модификации.

### **Практическая значимость**

Практическая значимость полученных результатов заключается в следующем:

1) Разработанные диссертантом алгоритмы позволяют осуществлять корректную временную сегментацию исходных речевых сигналов.

2) Предложенный автором алгоритм сегментации на циклы колебаний голосовых связок в силу своего быстродействия может быть также использован для решения актуальных задач по оценке траектории и средней на заданном интервале времени частоты основного тона диктора.

3) Продемонстрировано практическое применение предложенных алгоритмов сегментации аудиосигнала на примере задачи модификации темпа речи.

4) Выполнена программная реализация предложенных алгоритмов в среде MATLAB; в разработанном исследовательском программном обеспечении разметки фонограмм применены серверные веб-технологии, что позволяет использовать его в режиме многопользовательского удаленного доступа.

### **Обоснованность и достоверность научных положений**

В диссертационной работе четко обозначены задачи и методы их решения. Все научные положения, выводы, рекомендации и полученные результаты обоснованы и достоверны, что подтверждается анализом предложенных автором решений, результатами экспериментальных исследований на реальных речевых сигналах.

Основные результаты работы в достаточной степени изложены в публикациях диссертанта, включая три публикации в научных журналах из перечня ВАК, и прошли апробацию на нескольких отечественных научно-технических конференциях.

## Содержание работы

Диссертация состоит из введения, четырех глав, заключения, списка литературы, четырех приложений, изложена на 197 страницах, содержит 86 рисунков, 18 таблиц, 137 наименований литературных источников.

**Во введении** автором обоснована актуальность темы исследования, определены цель и задачи диссертационной работы, научная новизна, описаны полученные результаты и их практическая ценность, перечислены основные положения, выносимые на защиту, указана структура работы, приведены сведения об апробации и публикации результатов исследования.

**В первой главе** автор проводит анализ спектра задач, относящихся к временной сегментации речевого сигнала и сложности решения этих задач, рассматривает существующие подходы и алгоритмы сегментации, анализирует области их применимости, достоинства и недостатки.

**Во второй главе** исследуются сигнальные особенности реализаций звуков речи на примере русского языка. По результатам исследования приведено распределение звуков речи по основным типам сегментов и сформированы базовые сигнальные признаки, характерные для соответствующих сегментов.

**Третью главу** диссертант посвящает разработке комплекса алгоритмов сегментации, покрывающих все базовые уровни сегментации, и решению ряда сопутствующих задач. Разработанные алгоритмы апробируются на реальном речевом материале и сравниваются с другими существующими решениями.

**В четвертой главе** автор рассматривает примеры применения методов сегментации для решения актуальных задач речевых технологий, детально акцентируясь на задаче модификации темпа воспроизведения речевых сигналов. Данная задача демонстрирует возможности апробации всех разработанных алгоритмов сегментации в рамках единого речевого приложения. Представлены результаты сравнения фонограмм, обработанных предложенным автором алгоритмом и несколькими существующими альтернативными решениями.

**В заключении** представлены основные результаты исследования.



**В приложениях** описана методика и численные результаты исследования сигнальных особенностей звуков речи, кратко охарактеризованы возможности разработанного исследовательского программного обеспечения. Приведены детальные таблицы с данными анализа эффективности разработанных в диссертации алгоритмов. Представлены два акта о внедрении результатов диссертационной работы.

### **Замечания по диссертационной работе**

1) Для ряда предложенных автором алгоритмов (например, автоматическая сегментация звуков речи на «шумный/нешумный» или «вокализованный/невокализованный») не проведен полноценный анализ их работоспособности.

2) Значительную часть диссертационной работы составляет исследование особенностей фонетического состава и классификации звуков речи для русского языка, в то же время некоторые из предложенных алгоритмов экспериментально проверены только на английской речи с использованием открытой речевой базы данных TIDIGITS.

3) При экспериментальном анализе алгоритма сегментации на периоды основного тона автор сравнивает его с тремя описанными аналогами, при этом на график оценки производительности алгоритмов (рисунок 3.33 в диссертации и рисунок 4 в автореферате) без каких-либо пояснений попали еще два алгоритма (PEFAC и RAPT), которые, к тому же, предназначены не для сегментации речи, а для интегральной оценки частоты основного тона (ЧОТ).

4) На странице 114 диссертации в формуле (3.12) значение текущей накопленной энергии обозначено как  $E_t$ , а на рисунке 3.21 эта же переменная фигурирует уже как  $E(t)$ .

5) На странице 13 автореферата в формуле (4) некоторые символы отображены некорректно.

Все перечисленные замечания являются несущественными и не влияют на общую положительную оценку диссертационной работы.


## Заключение

Таким образом, диссертационная работа Томчука К.К. представляет собой законченную квалификационную работу, в которой исследована проблема автоматической временной сегментации естественной речи на различных уровнях. Основные результаты опубликованы в 15 печатных работах, из них 3 – в рецензируемых журналах из перечня ВАК («Информационно-управляющие системы», «Ученые записки ПетрГУ», «Вопросы оборонной техники. Серия 16»). Автореферат достаточно полно отражает содержание диссертационной работы.

Считаю, что по уровню актуальности, научной новизны, теоретической и практической значимости диссертационная работа Томчука К.К. «Сегментация речевых сигналов для задач автоматической обработки речи» в полной мере соответствует требованиям пп. 9–14 Положения о порядке присуждения ученых степеней, предъявляемых ВАК при Минобрнауки РФ к кандидатским диссертациям, а ее автор Томчук Кирилл Константинович заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.12.13 – «Системы, сети и устройства телекоммуникаций».

Официальный оппонент:

Карпов Алексей Анатольевич – доктор технических наук, доцент, заведующий лабораторией речевых и многомодальных интерфейсов Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук  
199178, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН  
E-mail: karpov@iias.spb.su, тел.: (812) 328-0421

 А.А. Карпов

