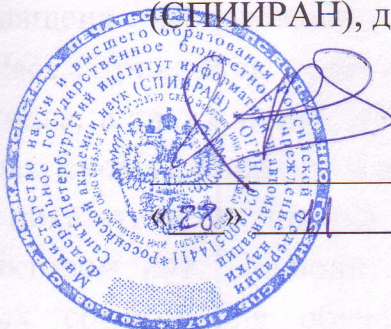


УТВЕРЖДАЮ

Директор Федерального государственного  
бюджетного учреждения науки Санкт-  
Петербургского института информатики и  
автоматизации Российской академии наук  
(СПИИРАН), д.т.н., проф.



А.Л. Ронжин

2019 г.

## ОТЗЫВ

ведущей организации на диссертацию Якимука Алексея Юрьевича на тему  
«Алгоритмы анализа частоты основного тона вокального исполнения»,  
представленную на соискание ученой степени кандидата технических наук  
по специальности 05.13.17 – Теоретические основы информатики

### 1. Актуальность темы диссертации

Определение частоты основного тона (ЧОТ) речевого сигнала является одним из ключевых этапов во многих задачах речевых технологий. Большинство исследований направлены на обработку речевого сигнала, что позволяет ограничиться частотной областью ЧОТ в диапазоне до 400 Гц. Однако при вокальном исполнении ЧОТ может звучать выше указанной отметки, что делает неприменимыми данные алгоритмы для анализа подобных сигналов.

В данном контексте видится актуальной работа над моделями и алгоритмами, позволяющими анализировать ЧОТ в вокальном исполнении, использование которых позволит распознавать ноты и проводить оценку качества их исполнения.

Указанное обуславливает актуальность темы диссертационного исследования Якимука А.Ю. В качестве цели исследования обозначено повышение качества распознавания звучащих нот в вокальном исполнении.

### 2. Структура и объем диссертации

Общий объем диссертационной работы составляет 121 страницу машинописного текста, в том числе 12 таблиц, 53 рисунка и список литературы, состоящий из 157 наименований. Диссертация содержит введение, четыре главы, заключение, список источников и три приложения.

Во введении обоснована актуальность темы диссертационного исследования, сформулированы цель и задачи, объект и предмет исследования, отражены его научная новизна, практическая и теоретическая значимость, приведены положения, выносимые на защиту, информация об апробации и внедрении результатов работы.

Первая глава посвящена исследованию современного состояния предметной области. Рассматриваются алгоритмы анализа ЧОТ и сегментации речевого сигнала, а также приводятся примеры использования данных алгоритмов в сферах, схожих с анализом вокальных исполнений. Для алгоритмов приводятся показатели и оценка применимости к задаче распознавания спетых диктором нот. Приводится обзор исследований по применению программных средств для обучения пению и сравнение программ-аналогов. В конце главы приводятся краткие выводы проведенного исследования и ставится задача модификации математической модели слуховой системы человека.

Во второй главе автором с учетом выделенных в первой главе недостатков проводится модификация модели. Описывается идея модификации и процесс формирования наборов шаблонов для заданных граничных частот исследования. Приводятся результаты тестирования алгоритма идентификации ЧОТ с автоматически сгенерированным набором шаблонов на синусоидальных сигналах. Приводятся результаты сравнения результатов генерации набора для диапазона, охватываемого исходной моделью. Оценивается влияние размера исследуемого диапазона при генерации набора шаблонов на точность определения ЧОТ.

В третьей главе приводится описание разработанного автором алгоритма распознавания нот в вокальном исполнении. Распознавание нот осуществляется в 2 этапа: идентификация нот в каждый момент времени и сегментация с учетом значения минимальной длительности звучания. Описывается подход к определению границ звучания нот с обоснованием выбора применяемого метода вычисления. Описываются критерии сбора аудиозаписей для тестирования алгоритма распознавания нот в вокальном исполнении. Приводятся результаты сравнения полученных данных по аудиозаписям при использовании алгоритма распознавания нот с данными экспертной оценки.

В четвертой главе приводится описание разработанного автором программного комплекса. Приводятся результаты экспериментальных исследований по тестированию на аудиозаписях с различными подходами к вокальному исполнению.

В заключении диссертации приведены основные результаты и выводы по проделанной работе.

Стиль изложения работы соответствует требованиям к научным работам. Ссылки на библиографические источники и литературу, включая собственные публикации автора, оформлены в соответствии с требованиями.

### **3. Научная новизна исследования и полученных результатов**

Полученные результаты диссертационного исследования являются новыми и могут быть классифицированы как изложение научно-обоснованных решений, внедрение которых внесет вклад в науку в сфере речевых технологий, в частности, при исследовании параметров голоса при вокальном исполнении.

Наиболее важные результаты диссертационной работы, обладающие признаками научной новизны:

1. Предложена модификация базовой модели слуховой системы человека, позволившая расширить диапазон частот в два раза по сравнению с исходной моделью и отличающаяся возможностью указания границ определения тона.

2. Разработан алгоритм создания шаблонов для обнаружения ЧОТ, отличающийся возможностью автоматической генерации наборов шаблонов с произвольным заданием граничных частот определения основного тона.

3. Разработан алгоритм распознавания нот, учитывающий минимальную длительность звучания нот и отличающийся учетом особенностей слуховой системы человека.

### **4. Обоснованность и достоверность полученных результатов**

Цель диссертационного исследования и вытекающие из нее задачи изложены корректно, являются практически значимыми и реализуемыми. Решения задач исследования доведены до практических приложений. По приведенному списку литературы можно судить о достаточной полноте изучения соискателем рассматриваемых вопросов, связанных с автоматизацией выделения нот в голосовом исполнении.

Достоверность результатов и выводов обеспечивается результатами проведенных численных экспериментов с использованием реальных данных, а также путём сопоставления результатов, полученных в диссертации, с результатами экспертной оценки и результатами, полученными другими методами (в частности, с исходным методом до расширения диапазона частот и автоматизации создания шаблонов).

## **5. Значимость результатов диссертации для соответствующей отрасли науки**

Результаты диссертационной работы Якимука А.Ю. используются в учебном процессе и научно-исследовательской деятельности студентов ТУСУР. Практическая значимость полученных результатов заключается в следующем:

1. Модифицирована математическая модель слуховой системы человека, что позволило автоматически генерировать наборы шаблонов для заданных диапазонов частот.

2. Разработан и внедрен алгоритм распознавания нот в вокальном исполнении. Внедрение позволило снизить объем трафика, передаваемого по сети, более чем на 90% за счет перехода от передачи аудиозаписей с пением в формате wav к отправлению преподавателю текстового сообщения с абснотацией распознанных нот.

Данные модификации позволили расширить диапазон выделения ЧОТ, позволив использовать ранее существовавший подход не только применительно к обычному речевому сигналу, но и к вокальному исполнению. Полученные результаты позволяют говорить о создании новых алгоритмов анализа устной речи, в частности, вокального исполнения, с целью преобразования исходного сигнала в нотную форму представления.

## **6. Рекомендации по использованию результатов диссертационной работы**

Результаты диссертационного исследования рекомендуются к применению в различных учебных заведениях музыкального профиля, осуществляющих подготовку по направлению «вокальное исполнение», а также при самостоятельной подготовке обучающихся к занятиям.

В качестве возможного продолжения и развития исследований, выполненных в диссертации, рекомендуется расширение предложенного автором подхода для исследования вокальных исполнений в условиях изменения звучания ЧОТ в процессе исполнения ноты (вibrато и глissандо).

## **7. Публикации, апробация и внедрение результатов работы**

Научные и практические результаты диссертационной работы докладывались и обсуждались на 14 конференциях и семинарах различного уровня, в том числе всероссийского и международного. Материалы исследования отражены в 19 научных публикациях, в том числе 3 работах в изданиях, рекомендованных ВАК Минобрнауки РФ по специальности 05.13.17 – «Теоретические основы информатики».

Результаты диссертационной работы внедрены в деятельность ООО «Элекард-ЦТП» в рамках проекта по дистанционному обучению вокалу в формате видеоконференций. Также результаты работы внедрены в учебный процесс ТУСУР.

## 8. Замечания по диссертации

1. Алгоритм распознавания нот описан недостаточно подробно. Не ясно, что такое список ЧОТ, чистая нота. Есть путаница в использовании некоторых переменных, например, на блок-схеме алгоритма (рис. 3.6) переменная  $i$  в одном блоке алгоритма сравнивается с количеством ЧОТ, а в другом блоке – с количеством чистых нот, в тексте работы указано, что  $i$  – момент времени в диапазоне от 0 до  $I$ , при этом переменная  $I$  обозначает момент времени формирования реакции системы фильтров на тестовые сигналы в тактах времени  $T_k$ .

2. Предложенная соискателем математическая модель позволяет распознавать ЧОТ до 800 Гц, однако в наборе данных самой высокой нотой является фа диез второй октавы, но, например, нота соль второй октавы, частота которой составляет 784 Гц, в наборе данных отсутствует.

3. Используемый тестовый набор в диссертации описан недостаточно. Не ясен принцип выбора тестового материала, как было определено количество нот в каждом наборе записей; например, почему записи женского голоса малой и первой октавы содержали 58 нот, а записи мужского голоса малой октавы – 56 нот. Тестовая выборка в этом случае получается ненормированной.

4. Не совсем ясно, почему отклонения вычисляются в центах (т.е. сотая часть полутона), достаточно было бы учитывать четверть полутона.

5. Не совсем понятен смысл используемого термина «локация». На стр. 66 «локация» обозначает просто номер тестовой записи. При этом на стр. 81 сказано: «Для первых двух типов выбранных упражнений были использованы локации, описанные в разделе 3.2: в качестве примеров пения гамм в октавном диапазоне вверх и вниз могут служить локации 1-6, а локации 7-8 – являются примером мелодически связанных аккордовых последовательностей». Однако, если номер локации соответствует номеру записи, описанной в разделе 3.2, то локации 1-6 не включают октавный диапазон, и локация 7-8 не является аккордовой последовательностью.

6. Графический пользовательский интерфейс программы не отображает длительность нот и тактовую черту, хотя разработанная программа длительность нот определяет.

7. В обзорной части стоило бы также проанализировать свежие научные работы по анализу ЧОТ Сорокина В.Н. и его коллег из ИППИ РАН.

8. Имеются некоторые замечания по оформлению диссертации, например, в тексте отсутствует ссылка на рисунок 3.6, а списке литературы публикации [73] и [93] одинаковы.

В целом, данные замечания не умаляют общей положительной оценки диссертации и значимости полученных научно-практических результатов.

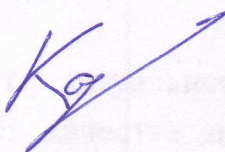
#### **9. Заключение о соответствии диссертации критериям, установленным Положением о порядке присуждения ученых степеней**

Диссертация Якимука А.Ю. является законченной научно-квалификационной работой, в которой на основании выполненных автором исследований решена научно-техническая задача, имеющая важное хозяйственное значение. Полученные результаты вносят определенный вклад в развитие сферы речевых технологий в таких областях, как анализ ЧОТ речевого сигнала диктора, а также анализ вокальных исполнений.

Диссертация отвечает требованиям «Положения о порядке присуждения ученых степеней» ВАК Минобрнауки РФ, предъявляемых к диссертациям на соискание ученой степени кандидата технических наук, а ее автор, Якимук Алексей Юрьевич, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Доклад Якимука А.Ю. заслушан, а отзыв ведущей организации на диссертацию рассмотрен и одобрен на заседании Лаборатории речевых и многомодальных интерфейсов Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук 25.11.2019 года, протокол № 3.

Г.н.с. (рук.) лаборатории речевых  
и многомодальных интерфейсов  
СПИИРАН  
д.т.н., доц.



Карпов Алексей Анатольевич

Федеральное государственное бюджетное учреждение науки Санкт-Петербургский институт информатики и автоматизации Российской академии наук

Адрес: 199178, Санкт-Петербург, 14-я линия, д. 39, СПИИРАН

E-mail: karpov@iias.spb.su

Тел.: (812) 328-0421