ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет компьютерных технологий и прикладной математики

Кафедра информационных технологий

**ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ, ВЫПОЛНЯЕМОЕ В ПЕРИОД**

**ПРОВЕДЕНИЯ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ**

**(научно-исследовательской работы)**

Студент (ка) Ибрагимов Беслан Маликович

*(фамилия, имя, отчество полностью)*

Направление подготовки (специальности) 01.03.02 Прикладная математика и информатика

Место прохождения практики кафедра информационных технологий факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ»

Сроки прохождения практики с 11.04.2020 г. по 24.04.2020 г.

Цель и основные задачи научно-исследовательской работы - приобретение опыта в исследовании актуальной научной проблемы; формирование навыков использования современных технологий сбора и обработки информации, интерпретации полученных данных, владения современными методами исследований; формирование навыков проведения библиографической работы с привлечением современных информационных технологий.

Формирование компетенций, регламентируемых ФГОС ВО:

ПК-1 - способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям

ПК-2 - способностью понимать, совершенствовать и применять современный математический аппарат

ПК-3 - способностью критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности

Перечень вопросов (заданий, поручений) для прохождения практики

«Распознавание личности по голосу с использованием Гауссовых смесей и Мел-частотных кепстральных коэффициентов»

Ознакомлен \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ибрагимов Б.М.

*подпись студента расшифровка подписи (ФИО)*

Руководитель НИР

заведующий кафедрой информационных технологий

доцент кафедры информационных технологий

факультета компьютерных технологий

и прикладной математики, кандидат физико-математических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подколзин В.В.

**Рабочий график (план) проведения практики:**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Этапы работы (виды деятельности) при прохождении практики | Сроки | Отметка руководителя практики от университета о выполнении (подпись) |
| 1 | Инструктаж по технике безопасности, охраны труда, пожарной безопасности, а также правилами внутреннего распорядка обучающихся.  Выбор и обоснование темы исследования | 11.04.2020 |  |
| 4 | Анализ полноты требований к разрабатываемому в рамках ВКР программного приложения | 12.04.2020-15.04.2020 |  |
| 5 | Тестирование программного приложения, выполнение численных экспериментов и анализ результатов. | 16.04.2020-23.04.2020 |  |
| 6 | Защита отчета | 24.04.2020 |  |

Ознакомлен \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ибрагимов Б.М.

*подпись студента расшифровка подписи (ФИО)*

«11» апреля 2020 г.

Руководитель НИР

заведующий кафедрой информационных технологий

доцент кафедры информационных технологий

факультета компьютерных технологий

и прикладной математики, кандидат физико-математических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подколзин В.В.

ОЦЕНОЧНЫЙ ЛИСТ

результатов прохождения производственной практики (научно-исследовательской работы)

по направлению подготовки

01.03.03 Прикладная математика и информатика

Фамилия И.О студента Ибрагимов Беслан Маликович

Курс 4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | ОБЩАЯ ОЦЕНКА  (отмечается руководителем практики) | Оценка | | | |
| 5 | 4 | 3 | 2 |
|  | Уровень подготовленности студента к прохождению практики |  |  |  |  |
|  | Умение правильно определять и эффективно решать основные задачи |  |  |  |  |
|  | Степень самостоятельности при выполнении задания по практике |  |  |  |  |
|  | Оценка трудовой дисциплины |  |  |  |  |
|  | Соответствие программе практики работ, выполняемых студентом в ходе прохождении практики |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | СФОРМИРОВАННЫЕ В РЕЗУЛЬТАТЕ учебной практики (практики по получению первичных профессиональных умений и навыков, в том числе первичных умений и навыков научно-исследовательской деятельности)  КОМПЕТЕНЦИИ  (отмечается руководителем практики от университета) | Оценка | | | |
| 5 | 4 | 3 | 2 |
|  | ПК-1 способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям |  |  |  |  |
|  | ПК-2 способностью понимать, совершенствовать и применять современный математический аппарат |  |  |  |  |
|  | ПК-3 способностью критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности |  |  |  |  |

Руководитель НИР

заведующий кафедрой информационных технологий

доцент кафедры информационных технологий

факультета компьютерных технологий

и прикладной математики, кандидат физико-математических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подколзин В.В.

Заключение

**О ПРОХОЖДЕНИИ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ (НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ)**

Ибрагимова Беслана Маликовича

*(Ф.И.О. студента)*

В процессе проведения производственной практики (научно-исследовательской работы) студент приобрел опыт в исследовании актуальной научной проблемы; сформировал навыки использования современных технологий сбора и обработки информации, интерпретации полученных данных, владения современными методами исследований; сформировал навыки проведения библиографической работы с привлечением современных информационных технологий.

Все пункты индивидуального плана-графика выполнены в полной мере. В ходе работы индивидуальное задание тестирование системы распознавания личности по голосу с использованием Гауссовых смесей и Мел-частотных кепстральных коэффициентов выполнено в полном объеме, студент проявил высокий уровень самостоятельности, и творческий подход к его выполнению. Представлен оформленный текст собранного материала. Программа НИР полностью соответствует отчёту. Работа заслуживает оценки «\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_».

Руководитель НИР

заведующий кафедрой информационных технологий

доцент кафедры информационных технологий

факультета компьютерных технологий

и прикладной математики, кандидат физико-математических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подколзин В.В.

Сведения о прохождении инструктажа по ознакомлению с требованиями охраны труда, технике безопасности, пожарной безопасности, а также правилами внутреннего трудового распорядка

Предприятие Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования «Кубанский государственный университет»

Факультет компьютерных технологий и прикладной математики

Кафедра информационных технологий

Студент Ибрагимов Беслан Маликович, 21

(ФИО, возраст)

Дата 11 апреля 2020 г.

1. **Инструктаж по требованиям охраны труда**

Провел канд., физ.-мат. наук, Подколзин В.В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, ФИО сотрудника, проводившего инструктаж, подпись)

Прослушал Ибрагимов Беслан Маликович

(ФИО, подпись студента)

1. **Инструктаж по технике безопасности**

Провел канд., физ.-мат. наук, Подколзин В.В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, ФИО сотрудника, проводившего инструктаж, подпись)

Прослушал Ибрагимов Беслан Маликович

(ФИО, подпись студента)

1. **Инструктаж по пожарной безопасности**

Провел канд., физ.-мат. наук, Подколзин В.В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, ФИО сотрудника, проводившего инструктаж, подпись)

Прослушал Ибрагимов Беслан Маликович

(ФИО, подпись студента)

**4. Инструктаж по правилам внутреннего трудового распорядка**

Провел канд., физ.-мат. наук, Подколзин В.В. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, ФИО сотрудника, проводившего инструктаж, подпись)

Прослушал Ибрагимов Беслан Маликович

(ФИО, подпись студента)

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Кубанский государственный университет

Факультет компьютерных технологий и прикладной математики

Кафедра информационных технологий

**ОТЧЕТ О ПРОХОЖДЕНИИ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ**

**(научно-исследовательской работы)**

по направлению подготовки

01.03.02 Прикладная математика информатика

период с 11.04.2020 г. по 24.04.2020 г

Ибрагимов Беслан Маликович

*(Ф.И.О. студента)*

студента 41 группы 4 курса ОФО

Направление подготовки 01.03.03 Прикладная математика и информатика

заведующий кафедрой информационных технологий

доцент кафедры информационных технологий

факультета компьютерных технологий

и прикладной математики, кандидат физико-математических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Подколзин В.В.

*ученое звание, должность (подпись) (Ф.И.О)*

Оценка по итогам защиты практики: \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись руководителя практики

«24» апреля 2020 г.

Краснодар 2020 г.

**ОТЧЕТ**

В рамках выпускной квалификационной работы была поставлена задача разработки системы распознавания диктора по голосу. Для анализа была выбрана система распознавания с использованием Гауссовых смесей и Мел-частотных Кепстральных коэффициентов.

Принцип работы системы идентификации диктора основан на выделении речи из фонограммы и последующем попарном сравнении биометрических признаков (содержащихся в голосе индивидуальных значимых признаков личности). Структура процесса распознавания личности по голосу в данной работе изображена на рисунке 1.

Изображение выглядит как снимок экрана

Автоматически созданное описание

Рисунок 1 - Структура процесса распознавания личности по голосу

Система распознавания личности называется текстонезависимой, если она не содержит информации о том, что именно человек будет произносить (система обучается и тестируется на произвольных речевых данных).

На данный момент наиболее распространенным подходом к решению задач текстонезависимой идентификации является подход на основе использования моделей гауссовых смесей (Gaussian mixturemodels, GMM). В качестве речевых признаков, в подавляющем большинстве систем идентификации, используются Мел-частотные кепстральные коэффициенты (mel-frequency cepstral coefficients, MFCC). При вычислении этих коэффициентов сначала проводится предварительная обработка входного речевого сигнала в виде разбиения сигнала на фреймы (окна речевого сигнала, состоящая из определенного количества элементов).

Первым шагом для получения коэффициентов является применение Дискретного Преобразования Фурье (ДПФ) или Быстрого Преобразования Фурье (БПФ) к каждому фрейму речевого сигнала для перехода из временной области в частотную. Однако, для уменьшения искажений перед применением БПФ может применяться весовая функция. На практике в качестве весовой функции часто используется окно Хэммига, которое имеет вид (1):

(1)

где

– длина окна.

Дискретное преобразование Фурье взвешенного сигнала можно записать в виде (2):

(2)

где

– фрейм , соответствующий фрейму во временной области;

– элемент фрейма речевого сигнала;

– элемент применяемой весовой функции во фрейме речевого сигнала;

– длина ДПФ.

В виду того, что ДПФ является симметричным относительно нуля, после БПФ обычно сохраняют только половину коэффициентов. После этого получают спектр мощности кадра по формуле (3):

(3)

Следующий шаг – это вычисление и применение к спектру мощности мел фильтров. Это набор из (от 20 до 40) треугольных фильтров (обычно – 26).

Для вычисления треугольных фильтров анализируемый диапазон частот в Герцах переводят в шкалу Мэл.

Высота звука, воспринимаемого человеческим слухом не связана линейно с его частотой. Её величина связана еще с уровнем громкости и тембром. Поэтому для ее анализа была создана количественная оценка звука – Мэл.

Мэл – это единица измерения высоты воспринимаемого звука, основанная на психофизических параметрах восприятия. На рисунке 2 изображен график зависимости мел-шкалы от частоты колебаний звукового сигнала.

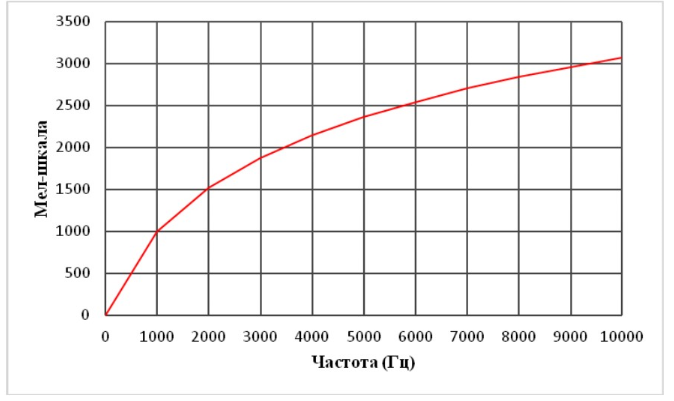


Рисунок 2 - График зависимости мел-шкалы от частоты колебаний звукового сигнала

Зависимость высоты воспринимаемого звукового сигнала от его

частоты описывается (4):

(4)

где

– высота звука в Мэлах;

– частота в Герцах.

После получения анализируемого диапазона частот (часто от 300Гц до 8000Гц) в шкале Мэл, разбиваем его на равных частей. К дополнительным двум точкам относятся границы диапазона. Затем переводим полученные точки диапазона обратно в Гц по формуле (5):

(5)

Проделав это, можно видеть, что полученные точки в Гц в начале диапазона находятся близко друг к другу, а в конце – максимально далеко, как показано на рисунке 3.

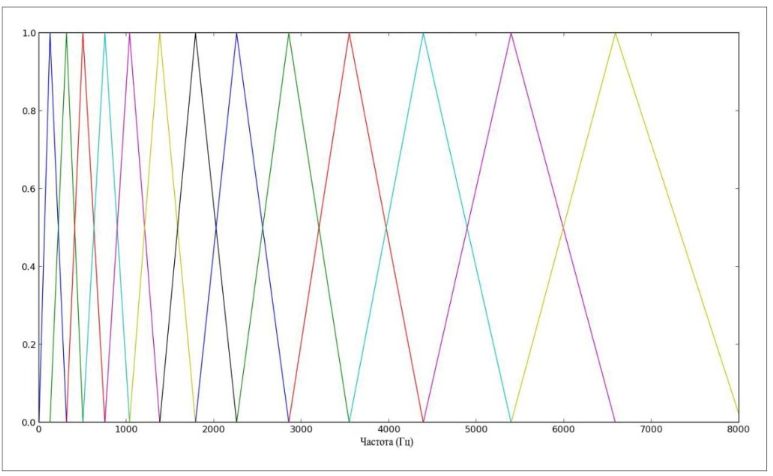


Рисунок 3 – Треугольные фильтры в шкале Гц

По полученным точкам и Рисунку 4 можно видеть, что первый фильтр начинается в первой точке, достигает пика во второй точке диапазона и опускается обратно до нуля в третьей точке. Второй фильтр начинается во второй точке, достигает пика в третьей точке диапазона и заканчивается в четвертой точке и т. д. Для вычисления фильтров применяют (6):

(6)

где

– количество фильтров;

– список полученных частот анализируемого диапазона в Гц.

Для применения полученных фильтров используется (7):

(7)

где

– спектр мощности фрейма ;

– коэффициенты треугольного фильтра .

После применения фильтров на каждом фрейме, будет получен вектор из элементов, который показывает, сколько энергии фрейма было в каждом фильтре.

Следующий шаг – это логарифмирование полученных векторов.

Заключительный шаг – это выполнение Дискретного Косинусного Преобразования (ДКП) для каждого фрейма по формуле (8):

(8)

где

– заранее определенное количество Мэл-частотных кепстральных коэффициентов;

– логарифмированный вектор энергии, полученный на предпоследнем шаге.

Для распознавания личности по голосу часто используют 12-13 MFCC.

Модель гауссовых смесей (GMM) широко используется в области распознавания дикторов. Данная модель представляет собой взвешенную сумму Гауссиан (9):

(9)

где

– модель диктора;

– функция плотности вероятности компоненты ;

– количество компонентов модели;

– вес компоненты смеси.

Для весов компонентов выполняется условие (10):

(10)

Функция плотности вероятности компоненты задается формулой многомерного нормального распределения (11):

(11)

где

– размерность пространства признаков;

– вектор математического ожидания компоненты ;

– ковариационная матрица компоненты

Чаще всего используется диагональная матрица ковариации. Возможно также использование одной матрицы ковариации для всех компонентов модели диктора или одной матрицы для всех моделей. Таким образом, каждая модель диктора характеризуется , .

Для построения модели диктора необходимо определить векторы средних, матрицы ковариации и веса компонентов. Данную задачу решают с помощью EM-алгоритма (Expectation–Maximization algorithm). На вход подаётся обучающая последовательность векторов . Параметры модели инициализируются начальными значениями и затем на каждой итерации алгоритма происходит переоценка параметров.

Для определения начальных параметров обычно используют алгоритм кластеризации, например, алгоритм К-средних [3]. Построив разбиение множества обучающих векторов на кластеров, параметры модели могут быть инициализированы следующим образом. Начальные значения совпадают с центрами кластеров, матрицы ковариации рассчитываются на основе попавших в данный кластер векторов, веса компонентов определяются долей векторов данного кластера среди общего количества обучающих векторов. Обучение GMM с помощью EM-алгоритма состоит из двух шагов:

а) шаг E (Estimation-step);

б) шаг M (Maximization-step).

На шаге E происходит вычисление апостериорных вероятностей по (12):

(12)

На шаге M происходит вычисление новых параметров модели по формулам (13), (14) и (15) соответственно:

(13)

(14)

(15)

Эти шаги повторяются либо до схождения параметров, либо до исчерпывания количества заранее установленных итераций. Для проверки сходимости параметров вычисляют логарифмическую вероятность модели диктора по формуле (16):

(16)

Для того, чтобы определить какой модели диктора принадлежат прибывшие данные, необходимо вычислить логарифмическую вероятность каждой обученной модели по формуле (16) и найти модель с максимальным значением. К этой модели и будут относиться анализируемые данные.

В рамках научно-исследовательской работы была реализована и протестирована система распознавания личности по голосу с использованием GMM и MFCC. Преимуществом использования модели GMM в системах голосовой верификации является возможность моделирования большого числа как индивидуальных акустических признаков речи диктора, так и самих дикторов. К преимуществам EM-алгоритма можно отнести возможность построения желаемого числа кластеров и быструю сходимость при удачной инициализации. К недостаткам можно отнести возможность медленной сходимости алгоритма из-за неудачной инициализации, а также возможность остановиться в локальном минимуме и дать квазиоптимальное решение.