

**Федеральное государственное автономное образовательное
учреждение высшего образования
"Санкт-Петербургский политехнический университет Петра
Великого"**

**Институт Компьютерных наук и технологий
Кафедра Измерительных информационных технологий**

*Тема: Автоматическая текстонезависимая идентификация диктора
с использованием спектральных коэффициентов*



Выполнил студент гр. 53505/4 : Д. А. Маевский

Руководитель: д.т.н., проф.: Г. Ф. Малыхина

Санкт-Петербург

2015 г.

Цели и задачи дипломного проекта:

Цель дипломного проекта — создание программного обеспечения, способного производить автоматическую текстонезависимую идентификацию диктора.

Исходя из поставленной цели, был сформирован список задач:

- Провести аналитический обзор методов текстонезависимой идентификации диктора;
- Разработать и реализовать алгоритм ввода и предварительного анализа звука;
- Разработать реализовать алгоритм получения первичных коэффициентов для дальнейшей работы с нейронной сетью;
- Собрать базу данных записей дикторов для обучения нейронной сети;
- Разработать и реализовать алгоритм для обучения нейронной сети;
- Провести эксперимент по текстонезависимой идентификации диктора на основе обученной нейронной сети.

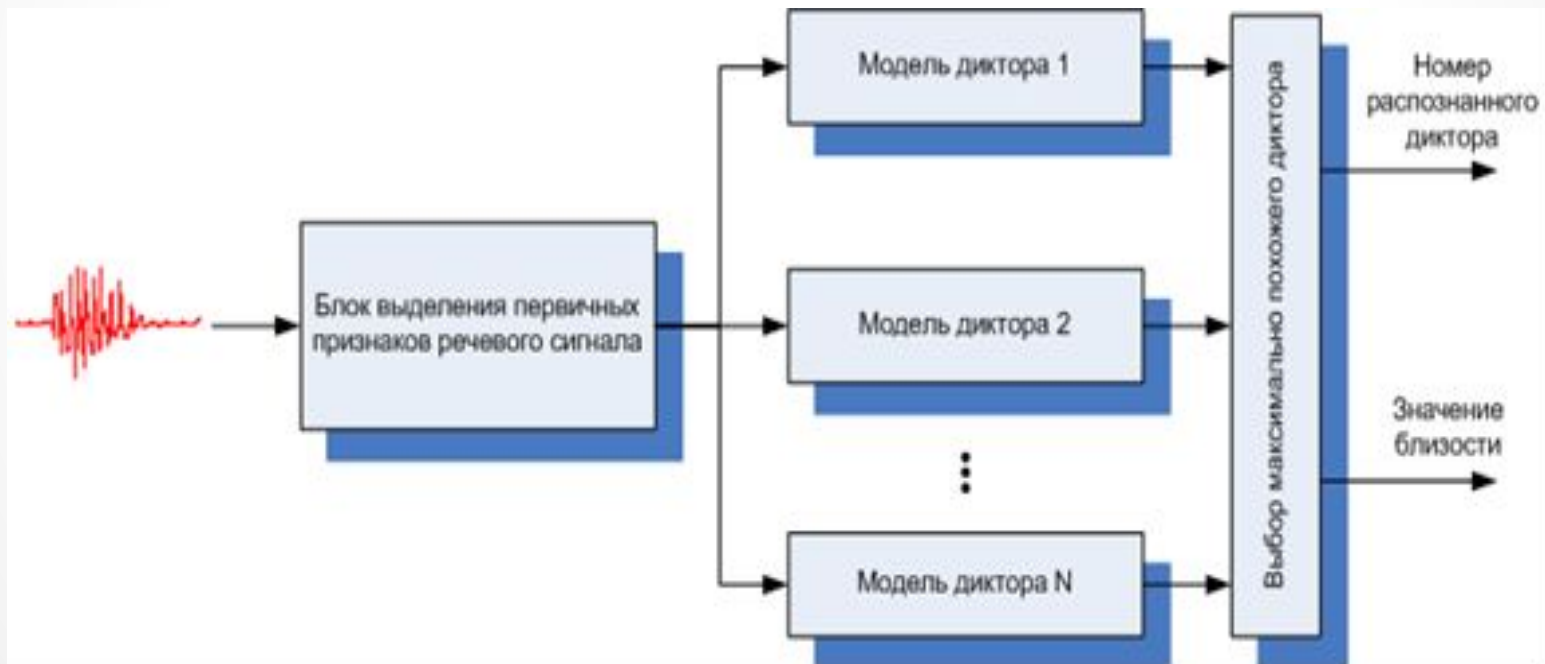
Задачи, решаемые в первой главе:

- Определить задачу идентификации;
- Провести обзор существующих методов, используемых в системах идентификации;
- Провести обзор существующих нейронных сетей, а также произвести их классификацию.

Задача идентификации:

Задача идентификации – это задача принятия решения кому из множества N кандидатов наиболее вероятно принадлежит тестируемая фонограмма.

Структурная схема системы идентификации:



Методы, используемые в системах идентификации диктора:

- Метод кепстральных коэффициентов, распределенных по мел-шкале (MFCC);
- Коэффициенты линейного предсказания (КЛП);
- Кепстральные коэффициенты на базе КЛП;
- Методы основанные на параметрических моделях (метод Юла-Уокера, метод Берга, ковариационный, модифицированный ковариационный метод);
- Метод на основе вейвлет-преобразования.

Классификация нейронных сетей:



Задачи, решаемые во второй главе:

- Описать алгоритм вычисления мел-кепстральных коэффициентов;
- Описать структуру нейронной сети, а также алгоритм её обучения;
- Описать графический интерфейс программы.

Методика вычисления мел-кепстральных коэффициентов:

- Схема метода следующая: на интервале времени в 10 – 20 мс вычисляется текущий спектр мощности, а затем применяется обратное преобразование Фурье от логарифма этого спектра (кепстр) и находятся коэффициенты кепстра :

$$c_n = \frac{1}{\Theta} \int_0^{\Theta} \log |S(j\omega, t)|^2 e^{-jn\Omega\omega} d\omega$$

$\Omega = 2\pi / \Theta$, Θ - верхняя частота в спектре речевого сигнала, $|S(j\omega, t)|^2$ - спектр мощности.

Если используется гребенка полосовых фильтров, то коэффициенты дискретного кепстрального преобразования вычисляются как :

$$c_n = \sum_{m=1}^M [\log Y(m)] \cos \left[\frac{\pi n}{M} \left(m - \frac{1}{2} \right) \right],$$

где $Y(m)$ – выходной сигнал m -го фильтра, c_n – n -й коэффициент кепстра.

Шкала мел вычисляется как :

$$M(f) = 1125 \ln \left(1 + \frac{f}{700} \right),$$

где f - частота в Гц, M – частота в мелах.

Блок-схема алгоритма вычисления мел-кепстральных коэффициентов:

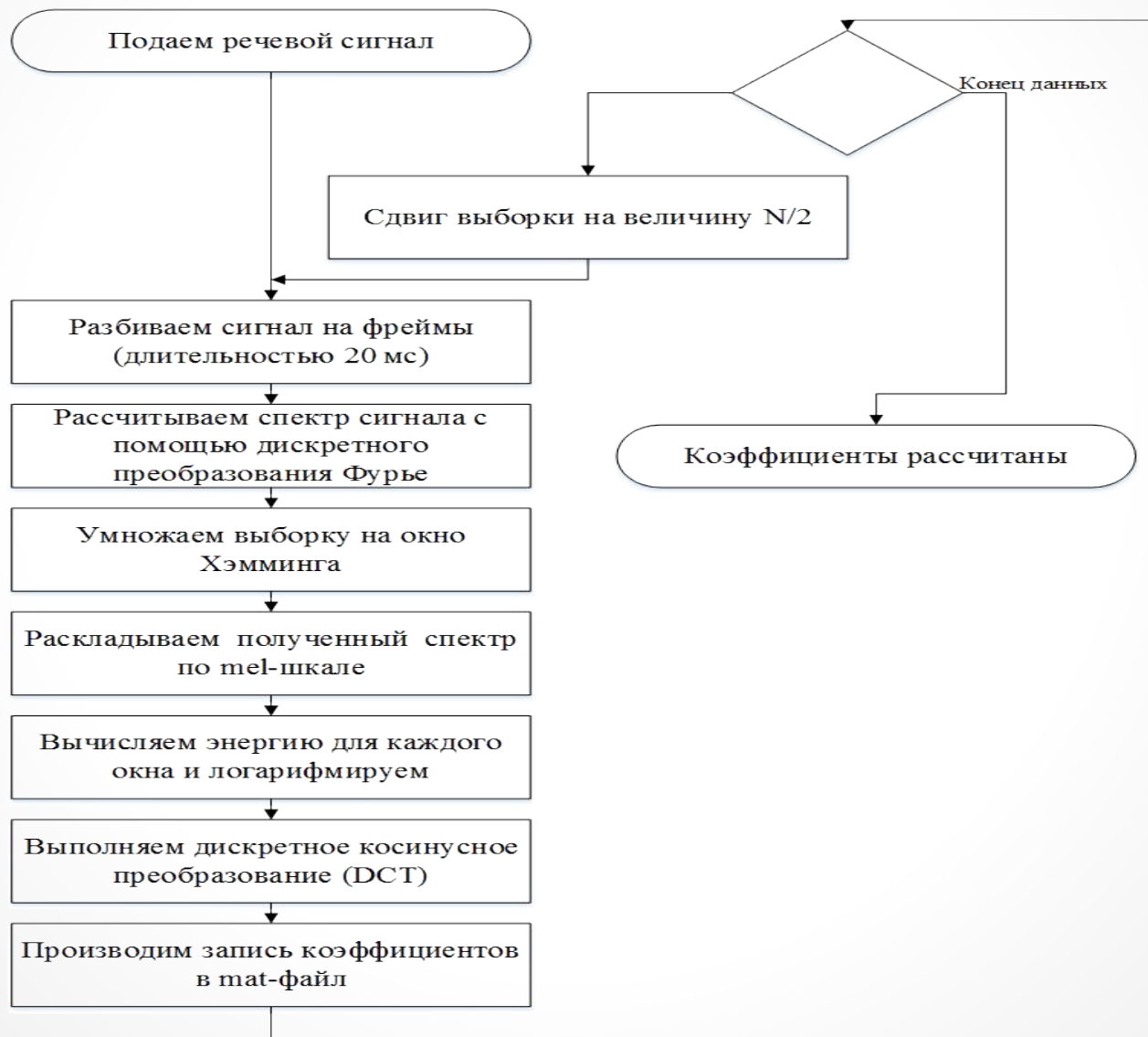
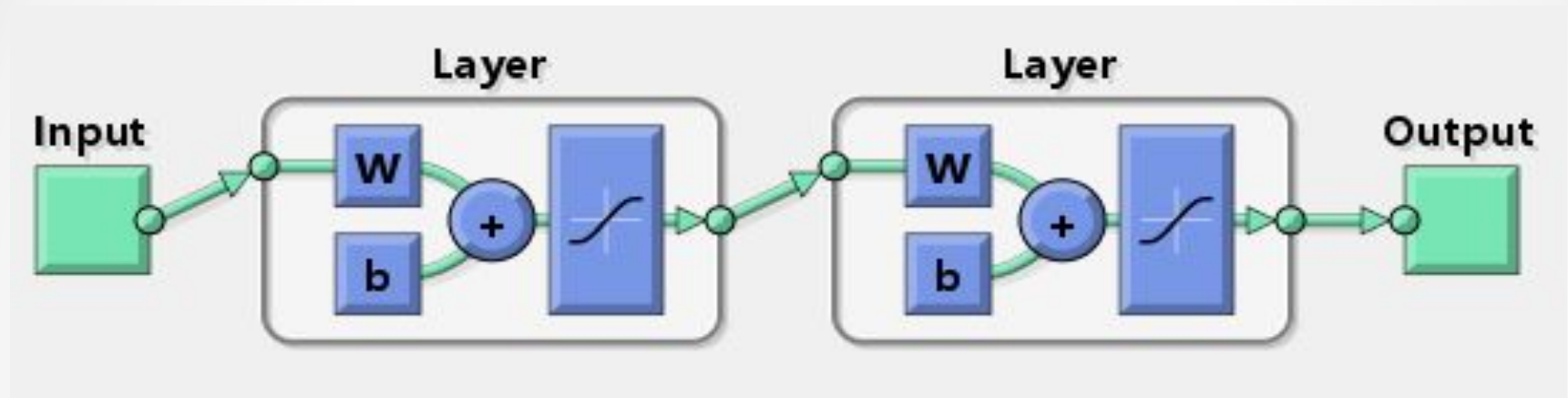


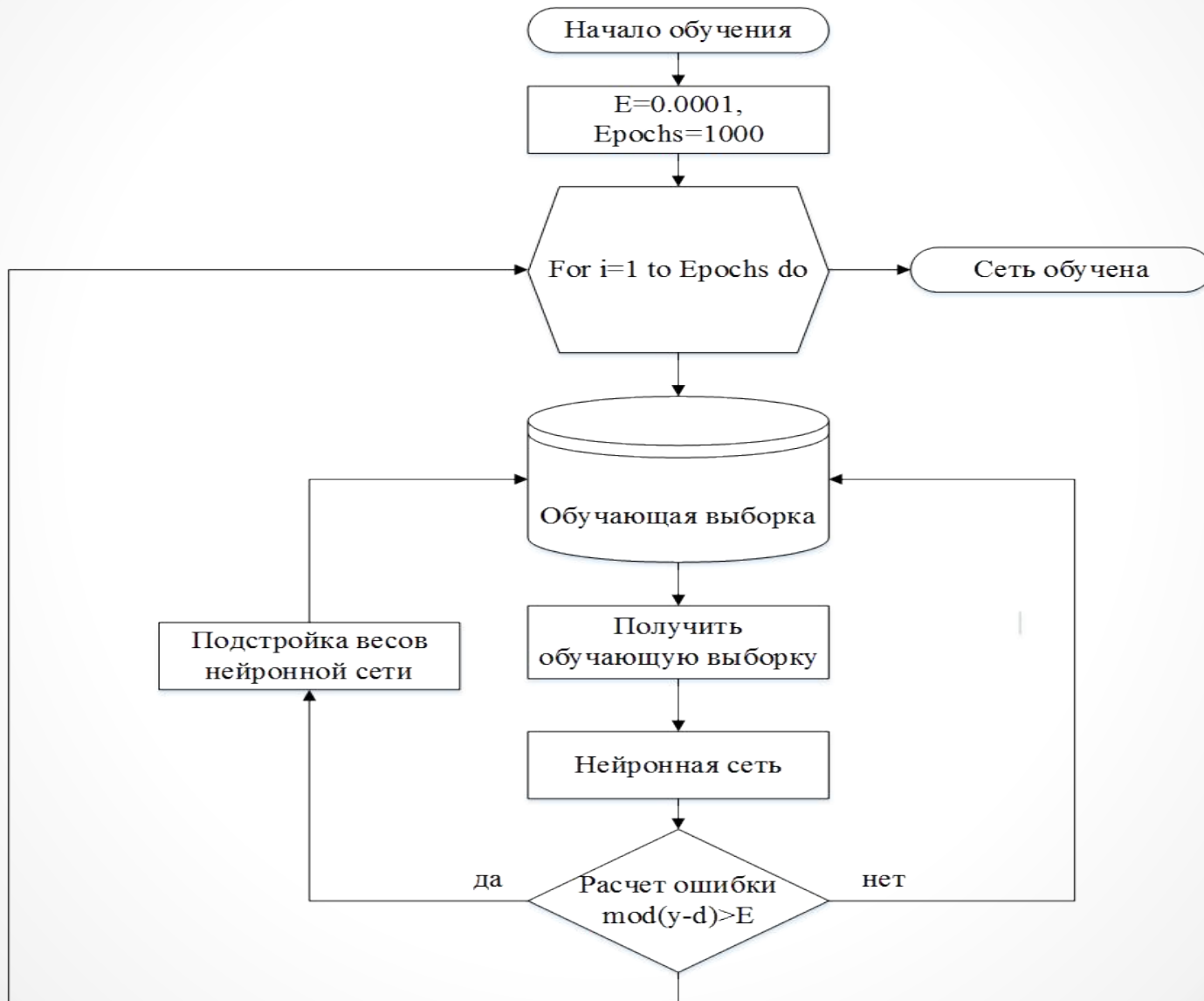
Схема нейронной сети:



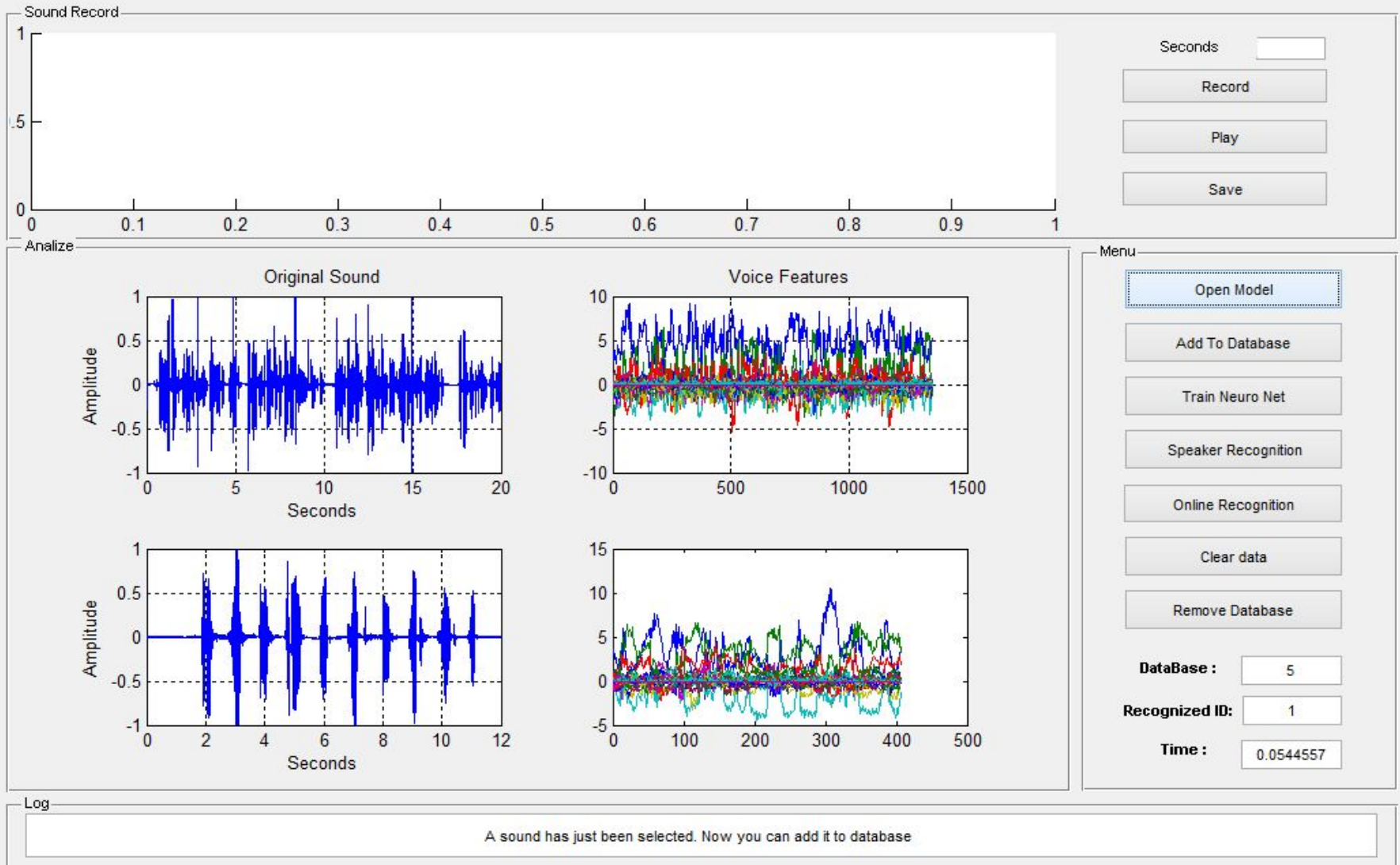
Функция активации нейронов:

$$\text{tansig}(n) = \frac{2}{(1 + \exp(-2n))} - 1$$

Блок-схема алгоритма обучения нейронной сети:



Графический интерфейс программы:



Задачи, решаемые в третьей главе:

- Исследование вероятности появления ошибок первого и второго рода;
- Исследование зависимости времени обучения от количество эпох обучения нейронной сети.

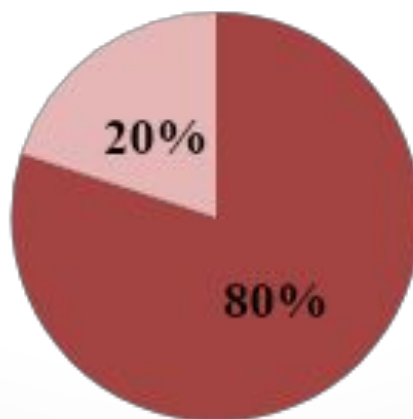
Результаты идентификации для дикторов женского пола (исследование вероятности ошибок первого рода):

Таблица 1

Диктор	Результат эксперимента (да/нет)							
	1	2	3	4	5	6	7	8
Дарья	да	да	да	нет	да	да	да	да
Мария	нет	да	да	да	да	да	нет	да
Ксения	да	да	да	нет	да	да	да	да
Татьяна	да	да	нет	да	да	да	да	да
Нина	да	да	да	да	да	нет	да	да

Рис.1 . Результаты эксперимента

■ диктор идентифицирован верно ■ диктор идентифицирован неверно



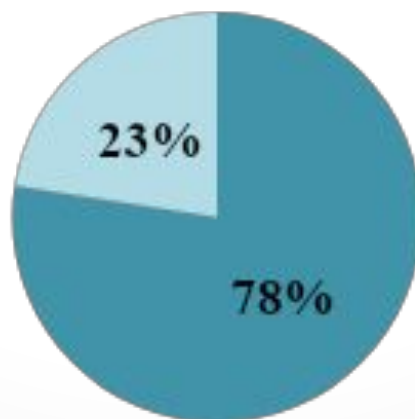
Результаты идентификации для дикторов мужского пола (исследование вероятности ошибок первого рода):

Таблица 2

Диктор	Результат эксперимента (да/нет)							
	1	2	3	4	5	6	7	8
Дмитрий	да	да	нет	да	да	да	да	да
Андрей	да	да	да	да	нет	да	да	да
Юрий	да	да	да	да	да	нет	да	да
Денис	да	да	нет	да	да	да	да	нет
Леонид	нет	да	да	нет	нет	да	да	да

Рис. 2. Результаты эксперимента

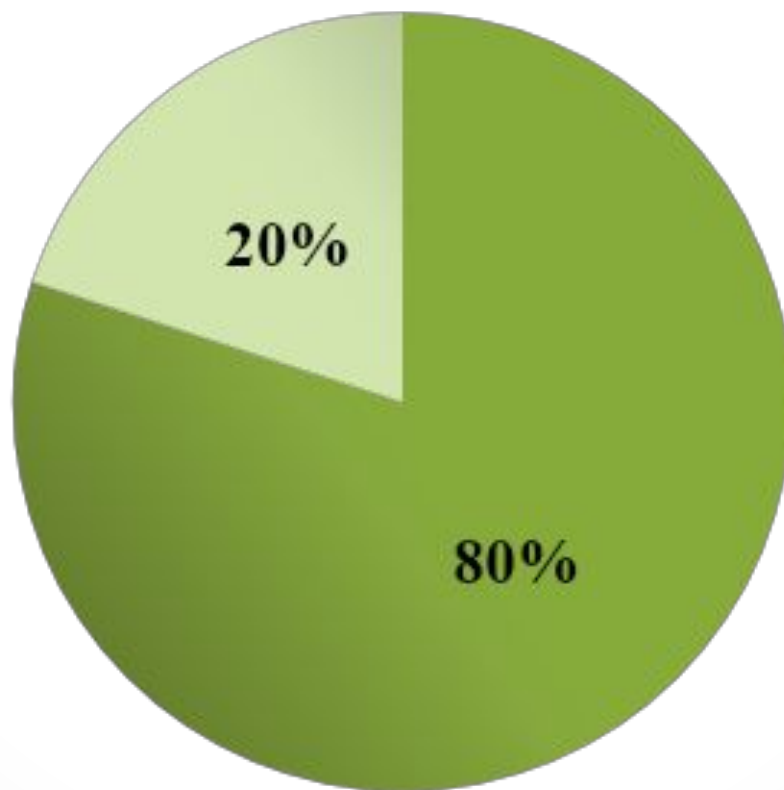
■ диктор идентифицирован верно □ диктор идентифицирован неверно



Результаты идентификации для дикторов обоих полов (исследование вероятности ошибок первого рода):

Рис. 3. Результаты эксперимента

■ Диктор идентифицирован верно ■ Диктор идентифицирован неверно

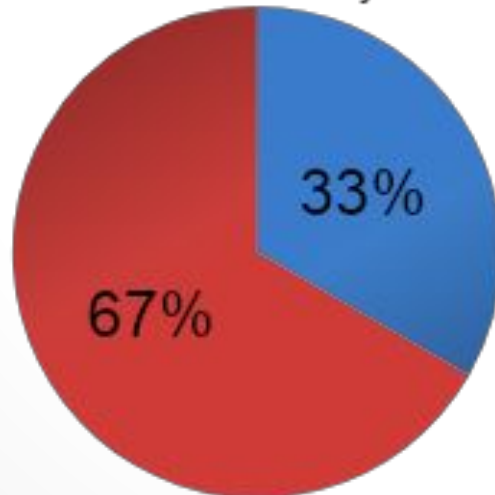


Результаты эксперимента, для дикторов, которые отсутствуют в базе (исследование ошибок второго рода):

Таблица 3

Диктор	Результат эксперимента (да/нет)							
	1	2	3	4	5	6	7	8
Мария	да	да	нет	нет	да	да	нет	да
Дмитрий	да	нет	да	да	да	нет	да	нет
Юрий	нет	да	да	нет	да	да	да	да

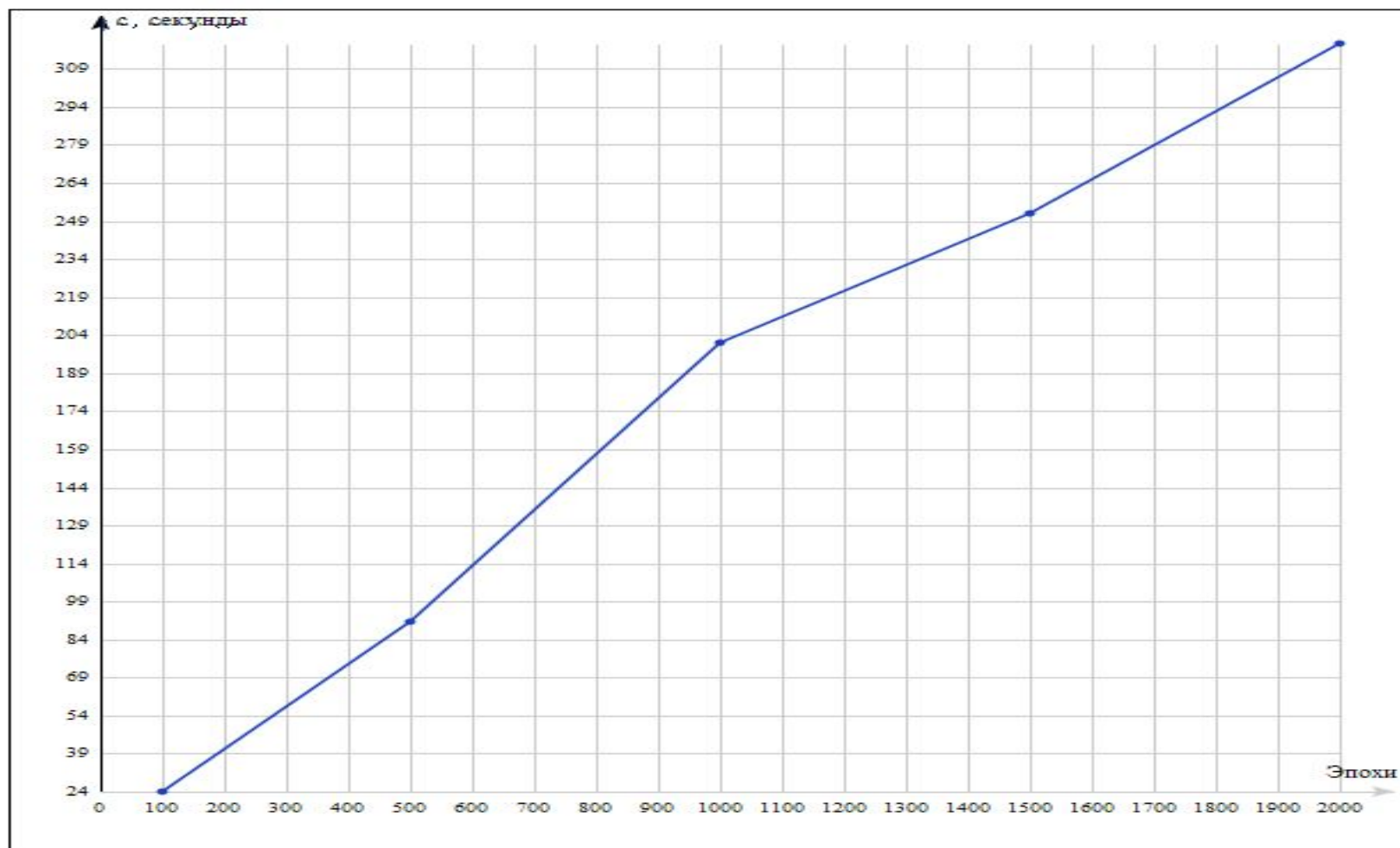
Рис.3 . Результаты эксперимента



- Программа идентифицировала диктора
- Программа сообщила об отсутствии диктора в базе

График зависимости времени обучения от

количества эпох обучения:



Расчет себестоимости разработки программного продукта:

Таблица 4

Наименование статьи затрат	Сумма затрат, руб.
Основная заработная плата разработчика	18000
Отчисления на социальное страхование	3600
Дополнительные материалы	250
Затраты на электроэнергию	832
Затраты на водоснабжение	228
Затраты на теплоэнергию	1863
Услуги связи	2100
Канцелярские товары	1840
Всего затрат:	31683

Требования к безопасности труда:

- Освещенность поверхности экрана не должна быть более 300 лк;
- Температура воздуха в среднем должна составлять 23 °С, влажность воздуха 60-40%;
- Уровень шума в среднем не должен превышать 60 дБ;
- Суммарное время перерывов при 8-ми часовой рабочей смене в среднем должно составлять 70 мин.;
- Конструкция рабочего стола должна обеспечивать оптимальное размещение на рабочей поверхности используемого оборудования;
- Конструкция рабочего стула (кресла) должна обеспечивать поддержание рациональной рабочей позы при работе на ПЭВМ;
- Допустимые уровни электромагнитных излучений от монитора компьютера, которые представлены в таблице 5:

Таблица 5

Наименование параметров		ВДУ ЭМП
Напряженность электрического поля	в диапазоне частот 5 Гц - 2 кГц	25 В/м
	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Электростатический потенциал экрана видеомонитора		500 В

Результаты:

В ходе работы были решены следующие задачи:

- Проведен аналитический обзор методов текстонезависимой идентификации диктора;
- Разработан и реализован алгоритм ввода и предварительного анализа звука;
- Разработан и реализован алгоритм получения мел кепстральных коэффициентов для дальнейшей работы с нейронной сетью;
- Собрана база данных записей дикторов для обучения нейронной сети;
- Разработан и реализован алгоритм для обучения нейронной сети;
- Проведен эксперимент по текстонезависимой идентификации диктора на основе обученной нейронной сети.

Программа показала неплохие результаты в распознавании дикторов. Точность распознавания составила порядка 80%, а вероятность возникновения ошибки второго рода составила 33%.

Спасибо за внимание!