ОТЧЕТ ПО ПРОЕКТУ

Система распознавания диктора

Выполнили: Рахматуллина Камила 11-408

Ахмадиев Рубен 11-401

Система распознавания диктора ― это совокупность элементов, позволяющих идентифицировать личность человека только на основе его голоса. Основой этой системы являются выделение особенностей каждого диктора, создание на основе этих особенностей модели диктора и сопоставление полученной модели с поданным на вход речевым файлом.

Возможность распознавания диктора основывается на том, что каждый человек обладает уникальным голосом, зависящим от строения голосового тракта, манеры говорить и характерного этому человеку активного лексикона.Распознавание диктора обладает большой областью применения. Примеры можно найти в таких сферах деятельности, как аутентификация, call­центры, разведка, банковская деятельность. Наличие такой системы позволит сократить количество работников, отвечающих за обработку звонков клиентов, которым необходимо, например, сбросить пароль. А в совокупности с системой распознавания речи можно полностью автоматизировать этот вид деятельности. Кроме этого, идентификация по голосу является наиболее удобной. Биометрические методы верификации основанные на уникальности отпечатков пальцев или сетчатки глаза обладают высокой надёжностью, но для их применения необходимо непосредственное присутствие человека, а это снижает мобильность.

**Какие задачи решает**

Системы распознавания диктора могут выполнять задачи двух видов: идентификация и верификация. Идентификация ― это процесс поиска голоса текущего диктора в базе данных остальных голосов, тогда как верификация подтверждает личность говорившего.

**С какими проблемами придется столкнуться**

Основная- изменчивость голоса одного и того же диктора в течение времени.Например, он постарел или заболел.А еще влияние шума окружающей среды и прочие факторы.

Элементы системы распознавания диктора:

Регистрация диктора: речь его -> извлекаем особенности -> алгоритм обучения- > построение модели диктора

Идентификация диктора: речь его –извлекаем особенности -> классификация -> решение

Эта работа посвящена разработке системы идентификации диктора по голосу на основе MFCC­коэффициентов и алгоритма машинного обучения Random Forest.

**Задача предварительной обработки голоса** состоит в выделении особенностей голоса для их последующего анализа.Речевой сигнал содержит в себе множество различных особенностей, но не все из них подходят.

Выделяемые признаки должны:

- сильно различаться для разных дикторов и слабо для разных записей одного диктора

-быть легко выделяемые из речевого сигнала

-быть сложными для подражания/подделки

-не реагировать на изменение здоровья диктора

С физиологической точки зрения, все признаки можно разделить на 5 основных групп:

1. краткосрочные спектральные особенности.

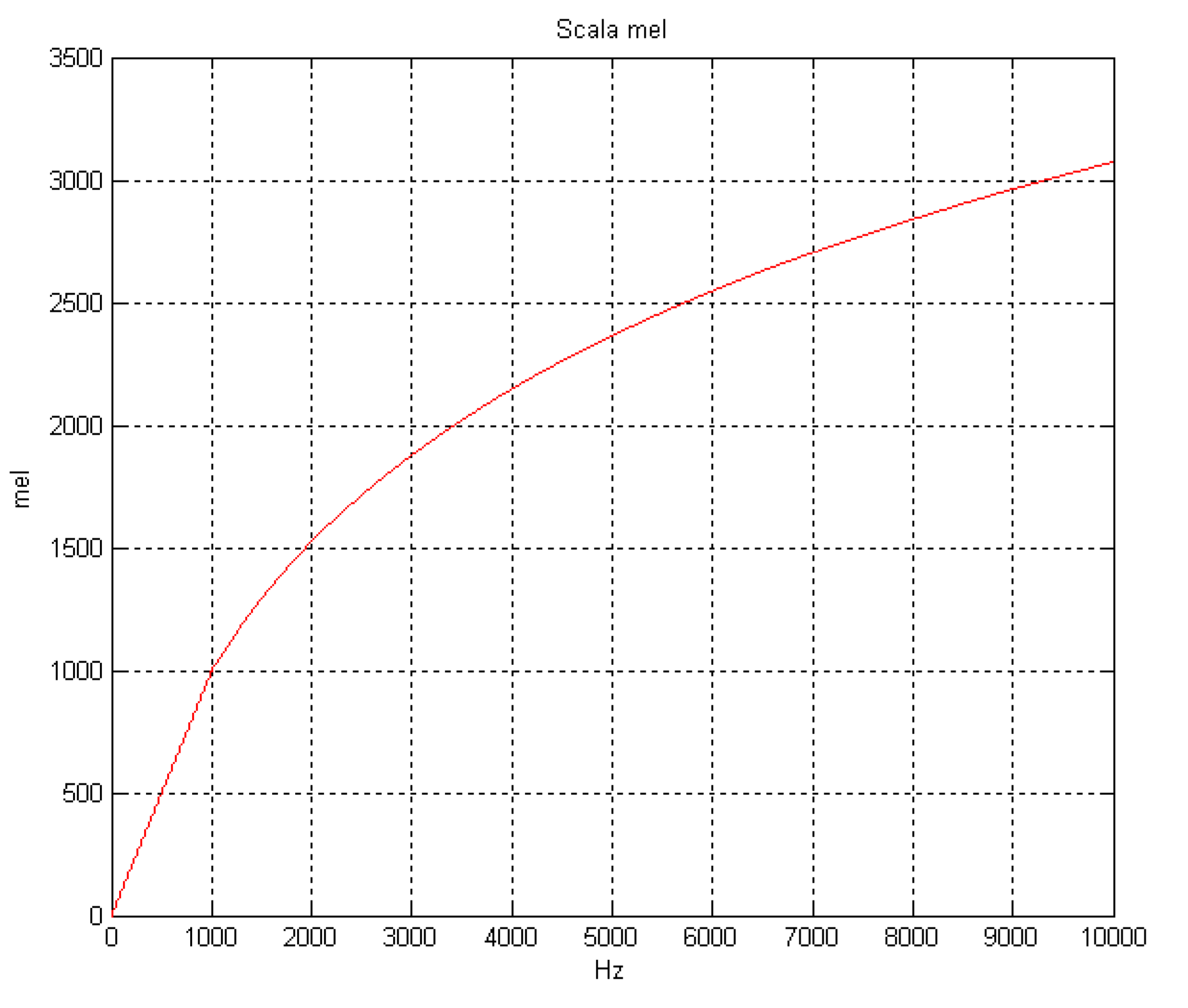
2. особенности источника голоса;

3. спектро­временные особенности;

4. просодические особенности;

5. высокоуровневые особенности. К первым двум относятся спектр, спектро­временные особенности включают ритм, темп, продолжительность речи, а просодические и высокоуровневые особенности характеризуются акцентом и активным запасом слов.

Речевой сигнал несильно изменяется за короткие (5­10 мс) промежутки времени. Если же рассматривать большие периоды, то можно заметить, как сигнал меняется в зависимости от произнесённых звуков



Человеческое ухо воспринимает звук нелинейно от его частоты: чем ниже частота, тем выше чувствительность. Поэтому шкала частот преобразуется в мел­шкалу , в которой присутствуют так называемые критические полосы, такие что в пределах одной полосы частоты сигналов неразличимы. Использование этой шкалы делает наши признаки (особенности) более приближёнными к тому, как слышат люди. Всё это приводит к выбору мел­кепстральных коэффициентов (MFCC) в качестве характеристик голоса.

После этапа выделения особенностей получаем из речевого файла матрицу размерности , где ​ M ― это количество характеристик для  N × M каждого отрезка, ​ N ― количество таких фрагментов, которое зависит от длины  файла и от выбранного размера фрагментов. Задача состоит в том, как представить матрицы характеристик речевых файлов в общем для всех виде. Для этого применяются супервекторы ― способ представить речевой файл в виде вектора фиксированной размерности. Часто супервекторы являются векторами большей размерности, составленными из нескольких  векторов меньшей размерности. Мы рассмотрели два способа представления супервекторов: среднее значение векторов и супервектор k-средних.

В качестве метода классификации векторов особенностей был выбран алгоритм машинного обучения Random Forest, который является набором деревьев решений

Дерево решений – это алгоритм машинного обучения состоящий из  слабых классификаторов, объединённых в дерево. Дерево решений имеет дватипа узлов:

● Внутренние ― узлы, имеющие более одного потомка.

● Терминальные ― узлы, не имеющие потомков.

**База данных**

В качестве набора данных была использована база дикторов CSTR VCTK Corpus. Она содержит образцы голоса 109 различных людей, родным языком которых является английский, и имеющих различные акценты.

**Эксперименты**

Затем приведем результаты экспериментов с разработанной системой идентификации. Для этого была проведена кросс­валидация по следующим параметрам модели:

количество деревьев в Random Forest (trees);

максимальная глубина деревьев (max\_depth);

критерий оценки чистоты разделения (measure\_quality\_of\_split);

метод для вычисления супервектора (supervector);

количество кластеров для супервектора по методу k­средних;

количество используемых мел­кепстральных коэффициентов

(MFCCs).

Полученные модели сравнивались по точности идентификации на тестовой выборке.

В первую очередь был протестирован метод построения супервектора как среднее значение векторов. Этот метод не дал высоких результатов независимо от параметров модели.

Далее был испытан метод k­средних для построения супервектора. Изначально параметр, отвечающий за количество кластеров, приравняли к двум. После серии экспериментов было решено, что это оптимальное значение данного параметра. В результате всех действий был получен показатель 95% успешной идентификации на тестовой выборке.

**Заключение**

В данной работе была разработан алгоритм идентификации диктора по голосу. В качестве признаков речевого файла были выбраны хорошо зарекомендовавшие себя мел­кепстральные коэффициенты. Супервектор был вычислен по алгоритму машинного обучения k­-средних. Для построения модели диктора был использован другой алгоритм машинного обучения ― Random Forest. Разработанный алгоритм реализован в виде программной системы.