**1. Введение**

В современном мире для защиты информации часто применяются системы безопасности с использованием биометрических признаков для аутентификации пользователя. В качестве одного из биометрических признаков можно использовать индивидуальные спектральные характеристики человеческого голоса. В отличие от таких биометрических признаков, как отпечатки пальцев, отпечаток ладони, рисунок сетчатки глаза, анализ характеристик голоса человека не требует специального (и зачастую дорогостоящего) оборудования, и может легко интегрироваться в существующие системы безопасности. Также этот метод может использоваться для верификации личности человека на расстоянии, например, при телефонном разговоре или при анализе записанной речи.

В данном дипломном проекте будет реализован и протестирован алгоритм верификации личности пользователя по голосу на основе сравнения мел-кепстральных спектральных коэффициентов.

Так как результат разработки представлен в виде программы на языке С, код может быть встроен в программу как для мобильных устройств, программиируемых микроконтроллеров, так и для таких систем, как Linux и Windows.

**2. Специальная часть**

Цель данного раздела – выбор и описание алгоритмов для реализации верификации личности пользователя по голосу, описание теоретических основ выбранных алгоритмов и определение их оптимальных параметров.

**2.1 Обзор прототипов и похожих разработок**

1. На основании доклада ООО «Форенэкс» «Судебная экспертиза звукозаписей» можно сделать вывод о существовании следующих программных систем с назначением, схожим с разрабатываемой в дипломной работе:

* OTExpert – программный комплекс криминалистического исследования фонограмм речи;
* Justiphone – программный комплекс криминалистического исследования фонограмм речи;
* Phonexi – программный комплекс для криминалистической идентификации говорящего по фонограммам устной речи на основе методики «Диалект»;
* «Сапфир» – программный комплекс для криминалистической идентификации говорящего по фонограммам устной речи на основе методики «Диалект»;
* «Этнос» - программный комплекс для идентификации говорящего по фонограммам устной речи на различных языках;
* «ИКАР-Лаб» - программный комплекс для криминалистического исследования фонограмм и идентификации говорящего по фонограммам устной речи на основе спектрально-формантных методов.

Так как перечисленные выше продукты применяются в области криминалистики, то информации по реализованным алгоритмам и их параметрам в публичном доступе нет.

2. HTK (Hidden Markov Model Toolkit) – набор инструментов для работы со скрытыми моделями Маркова, используется как для распознавания речи, так и для других задач, связанных с обработкой речевых сигналов. Имеет реализацию алгоритма получения мел-кепстральных коэффициентов, что может использоваться для идентификации и верификации личности.

3. HTK MFCC MATLAB – реализация алгоритма получения мел-кепстральных коэффициентов, используемого в HTK, на языке Matlab.

4. Speaker-recognition – проект студентов университета Маастрихта на тему идентификации личности пользователя по голосу. Реализован на языке Matlab.

**2.2 Постановка задачи**

В соответствии с техническим заданием требуется реализовать систему верификации личности пользователя по голосу. Назначение системы – верификация – означает режим распознавания «один к одному», когда определяется степень схожести образцов голоса признакам одного пользователя. Так как верификация выполняется при произнесении короткой ключевой фразы, то система является текстозависимой.

Для выполнения заданных функций система должна иметь 2 режима работы: обучение и верификация. Оба режима используют общую функциональную часть для получения вектора признаков речевого сигнала, который в дальнейшем запоминается либо используется для сравнения.

Для решения поставленной задачи могут использоваться следующие методы:

* Спектрально-формантный метод,
* Метод линейного предсказания,
* Метод выделения динамических характеристик,
* Метод получения мел-кепстральных частотных коэффициентов.

В качестве основы данного дипломного проекта была выбрана реализация получения мел-кепстральных коэффициентов HFCC FB-29, имеющая следующие параметры:

* Количество мел-кепстральных коэффициентов: 29,
* Анализируемый диапазон: [100, 6250] Гц,
* Коэффициент ширины фильтров: 0.75,
* Выравнивание коэффициентов отсутствует.

Схема получения вектора признаков изображена на рис. 2.1.

Минимальная разрядность данных в данном проекте – 32 бита для данных с плавающей точкой.

Минимальная частота дискретизации по теореме Котельникова должна быть больше или равна удвоенной максимальной анализируемой частоте, и для выбранных параметров алгоритма составляет 12500 Гц. Так как увеличение частоты дискретизации

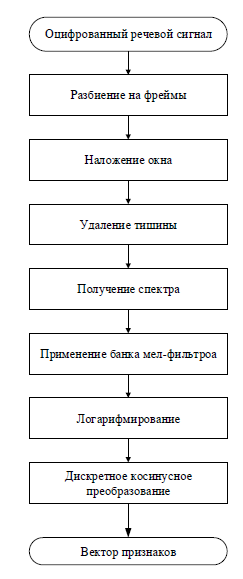


Рис. 2.1 Схема получения вектора признаков

влияет на используемые системныересурсы (прежде всего на объем оперативной памяти, требуемой для работы алгоритма), имеет смысл также ограничить максимальную частоту

дискретизации. В качестве оптимального значения частоты дискретизации из стандартных значений для аудиофайлов и звуковых карт стоит выбрать 16 кГц.

Из недостатков этого метода можно отметить влияние на систему АЧХ микрофона и аналогового тракта звуковой карты, эмоционального и физического состояния пользователя.

**2.3 Разработка метода ввода речевого сигнала в программу**

Источником аудиоданных в ПК явлеяется аудиофайл. Данные в файле могут быть в WAV или PCM с плавающей точкой (IEEE Float). Для ввода данных из файла используется стандартный бинарный ввод из wav-файла. Программа разбирает заголовок WAV-файла и в соответствии с ним производит считывание из аудиоданных. В результате имеется массив чисел с плавающей точкой длина которого определяется размером аудиофайла.

**2.4 Разработка метода выделения речевого сигнала. Удаление тишины**

Для повышения точности системы, необходимо произвести сглаживание входного сигнала. Задача сглаживания — это, по сути, задача фильтрации сигнала от скачкообразных (ступенчатых) изменений. Считается, что полезный сигнал их не содержит. Ступенчатый сигнал за счёт множества резких, но небольших по амплитуде, перепадов уровня содержит высокочастотные составляющие, которых нет в сглаженном сигнале. Для этого использовался алгоритм скользящего среднего:

От числа K зависит степень сглаживания сигнала. Чем больше K, тем более заметен результат сглаживания, однако при слишком большом K форма сигнала становится менее выразительной, поэтому следует выбрать оптимальное K, например 25. Также следует отметить, что на первых и последних точках i невозможно вычислить значения сглаживания. Область где это можно сделать определяется следующим образом:

Для уменьшения объема анализируемых данных и повышения точности системы имеет смысл анализировать только те фреймы, которые содержат речевой сигнал. При высоком значении коэффициента сигнал/шум для решения данной задачи можно использовать метод подсчета средней энергии фреймов. Энергия фрейма посчитывается как сумма квадратов амплитуд, деленная на количество отсчетов

Далее считается среднее значение энергий всех фреймов

Далее выбирается граничное значение, например, 10% от , и отбрасываются те фреймы, средняя энергия которых меньше этого значения.

**2.5 Разработка метода получения вектора признаков речевого сигнала**

В качестве вектора признаков в данном проекте выбрана матрица мел-кепстральных коэффициентов HFCC всех фреймов, содержащих полезный сигнал.

**2.5.1 Разбиение сигнала на фреймы**

Для речевого сигнала имеет смысл анализировать не всю временную область, а небольшие фрагменты, длительностью несколько десятых секунды – фреймы. Для уменьшения потерь информации при разбиении сигнала на фреймы используется сдвиг начала фрейма назад относительно конца предыдущего фрейма, таким образом, что фреймы частично перекрываются. Схема разбиения сигнала на фреймы изображена на рис. 2.2.

Минимальная длина фрейма ограничена минимальной анализируемой частотой в соответствии с теоремой Котельникова, т.е. для получения корректных результатов на этапе выполнения дискретного преобразования Фурье необходимо, чтобы длина анализируемого фрейма была больше или равна периоду сигнала, соответствующего минимальной анализируемой частоте.

При увеличении длины фрейма снижается использование системных ресурсов (прежде всего оперативной памяти), но ухудшается точность системы. В данном дипломном проекте в соответствии со значением минимальной анализируемой частоты выбрана длина фрейма 100 мс и сдвиг 50 мс (50% перекрытие).

**2.5.2 Оконная функция**

Спектральный анализ, в теории, предназначен для анализа неприрывных периодических сигналов. При обрезании сигнала, в спектре появляются несуществовавшие в сигнале высокочастотные составляющие. Чтобы бороться с их появлением и прибегают к использованию т.н. оконных функций, изменяющих оригинальный сигнал в каждом анализируемом окне (фрейме). Для этой цели могут использоваться различные оконные функции – окно Хэмминга, окно Ханна, окно Кайзера и другие. Искажения, вносимые применением окон, определяются размером окна и его формой. Выделяют два основных свойства частотных характеристик окон: ширина главного лепестка и максимальный уровень боковых лепестков. Применение окон, отличных от прямоугольного, обусловлено желанием уменьшить влияние боковых лепестков за счет увеличения ширины главного.

В нашем случае будет использоваться окно Хэмминга с уровнем боковых лепестков: -42 дБ., имеющее формулу

где – номер фрейма, – общее количество фреймов.

И вид, изображенный на рис.2.3.

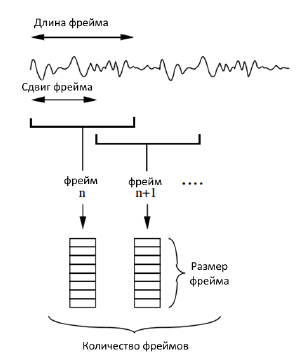


Рис. 2.2 Разбиение сигнала на фреймы

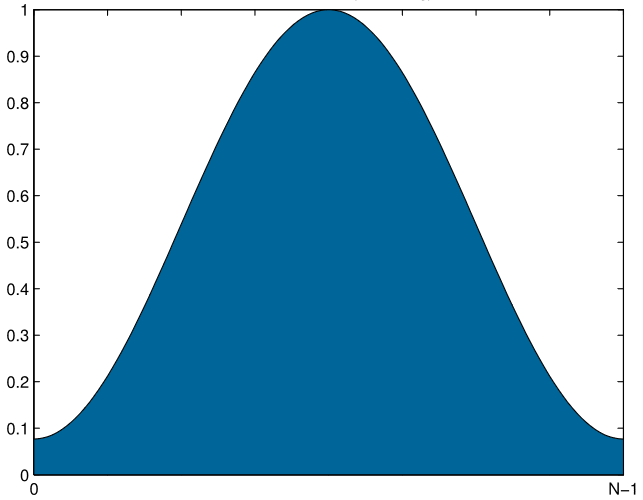
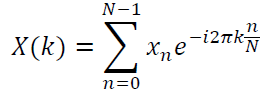


Рис. 2.3 Окно Хэмминга

**2.5.3 Быстрое преобразование Фурье**

Для получения спектра фрейма, необходимо применить к нему дискретное преобразование Фурье. Если осуществить дискретное преобразование Фурье для большого числа данных, то вычисление займет продолжительное время. Для увеличения скорости преобразования в данной дипломной работе был релализован алгоритм быстрого преобразования Фурье.



k = 0, ..., N-1

Быстрое преобразование Фурье – это алгоритм вычисления, который успешно использует свойства периодичности тригонометрических функций для того, чтобы избежать ненужных вычислений в дискретном преобразовании Фурье. И в случае БПФ, и в случае ДПФ результат вычислений один и тот же. Отличаются они только методом вычисления. В отличие от простейшего алгоритма ДПФ, который имеет сложность порядка O(), БПФ имеет сложность всего лишь O()

Быстрое преобразование Фурье принимает на вход отсчетов, поэтому если количество отсчетов неравно степени двойки, то необходимо дополнить исходные данные нулями до следующей степени двойки. Так как спектр симметричен относительно середины, то половина его отсчетов отбрасывается. Быстрое преобразование Фурье возвращает и реальную и мнимую часть спектра, для дальнейшей работы необходимо вычисляить модуль каждого комплексного числа спектра.

k = 0, ..., N-1 – номер отсчета

re – реальная часть спектра

im – мнимая часть спектра

Пример получаемого спектра приведен на рис. 2.4.

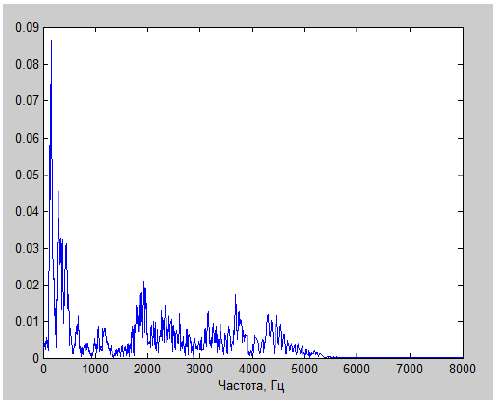


Рис. 2.4 Спектр фрейма

**2.5.4 Мел-шкала и банк фильтров**

Восприятие спектра звуковых частот человеком нелинейно, и результаты исследований показывают, что использование модели восприятия звука человеком в системах распознавания и обработки речевых сигналов повышает точность работы таких систем.

Результатом исследования субъективного восприятия человеком звуковых частот является мел – психофизическая единица высоты звука, и мел-шкала (Рис. 2.5).

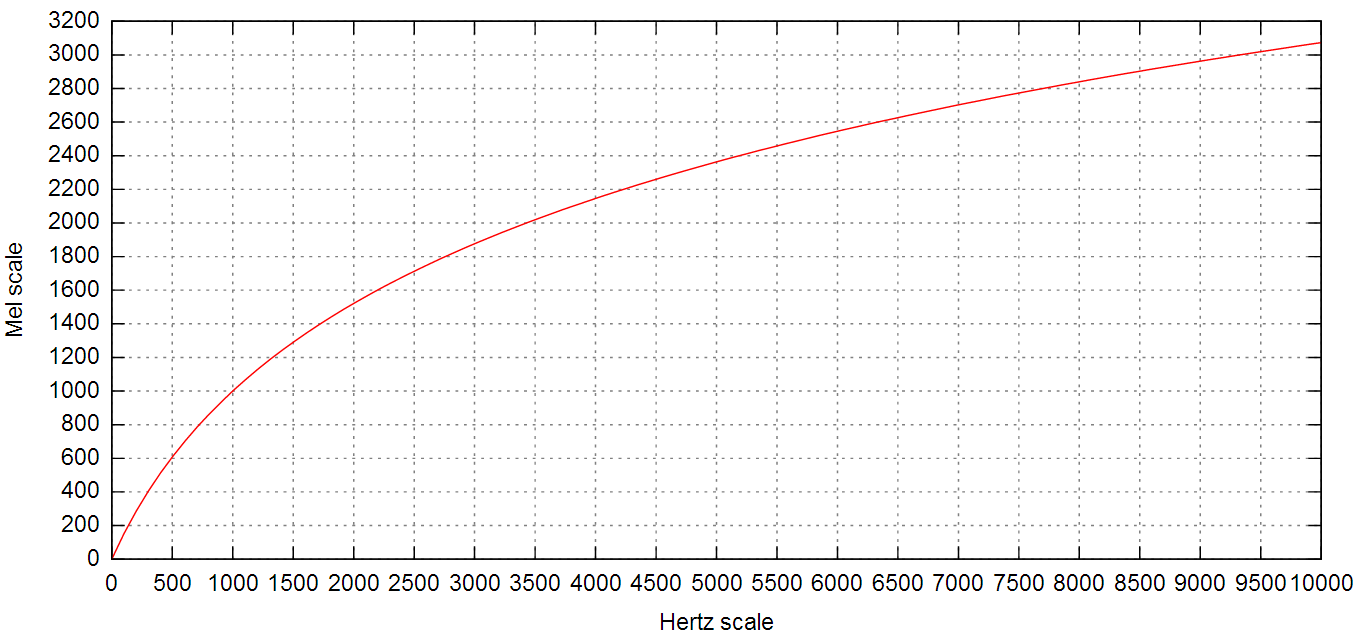
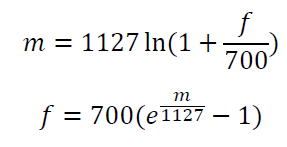


Рис. 2.5 Мел шкала

Для перевода частоты в высоту звука и обратно используются формулы



m – высота звука

– частота звука

На основе этих преобразований формируется банк фильтров, линейно распределенных на мел-шкале в выбранном для анализа диапазоне.

Алгоритм формирования банка фильтров HFCC-E:

1. Выбираются границы анализируемого диапазона частот и , количество фильтров M, коэффициент масштабирования ширины фильтров ERB.

2. Вычисляются центральные частоты первого и последнего фильтров



где и индекс i = 1 или M соответственно, a, b, c – коэффициенты Глассберга и Мура:

𝑎 = 6.23∗10−6 ∗ 𝐸𝑅𝐵, 𝑏 = 93.39∗10−3 ∗ 𝐸𝑅𝐵, 𝑐 = 28.52 ∗ 𝐸𝑅𝐵.

Для первого фильтра:

Для последнего фильтра:

3. Так как центры всех фильтров распределены равномерно на мел-шкале, то центральные частоты оставшихся фильтров легко рассчитываются по следующим формулам:

Расстояние между центрами соседних фильтров:

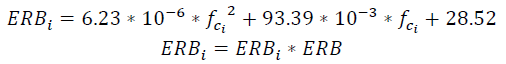
где все частоты - в мелах.

Центральные частоты оставшихся фильтров считаются по формуле

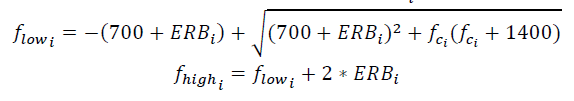


для i = 2..M

4. Далее для каждой центральной частоты в мелах считается центральная частота в Гц и вычисляется ширина каждого фильтра



5. Далее считаются граничные частоты



Вид банка фильтров HFCC на отрезке частот от 0 до 8000 Гц для значений ERB 1.0, 0.75 и 1.5 приведен на рис. 2.6.

После получения банка фильтров логарифм спектральной плотности энергии каждого фильтра для каждого фрейма



i = 1..M

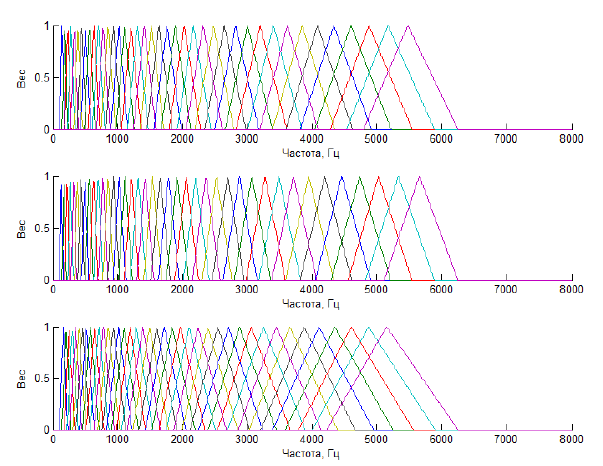


Рис. 2.6

Пример полученных значений показан на рис. 2.7.

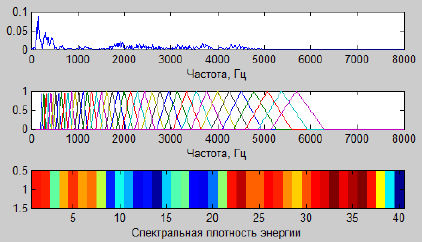
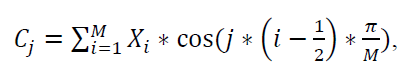


Рис. 2.7. Получение спектральной плотности энергии

**2.5.5 Дискретное косинусное преобразование**

Для дальнейшего сжатия данных и декорреляции значений, полученных на предыдущем этапе, используется дискретное косинусное преобразование. После такого преобразования получаются мел-кепстральные частотные коэффициенты для каждого фрейма.

Формула дискретного косинусного преобразования:



j = 1..J, где М – количество фильтров, J – количество мел-кепстральных коэффициентов; обычно J < M.

Пример полученных мел-кепстральных коэффициентов на основе спектральной плотности энергии показан на рис. 2.8.

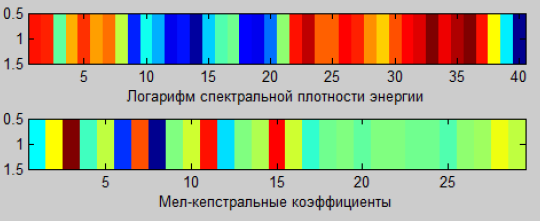


Рис 2.8 Получение мел-кепстральных коэффициентов