СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 2](#_Toc103472812)

[1 Описание предметной области 3](#_Toc103472813)

[1.1 Основные понятия о цифровом звуке 3](#_Toc103472814)

[1.2 Постановка задачи 3](#_Toc103472815)

# ВВЕДЕНИЕ

В современном мире голос человека является не только способом выражения своих мыслей, но и способом управления цифровой архитектурой вокруг него.

Общий темп жизни растет, а вместе с ним растут и требования к окружающим нас цифровым устройствам. Вместо того, чтобы тратить время на заказ такси, бронирования столика в ресторане или поиска в интернете прогноза погоды, современные технологии позволяют с помощью домашних устройств, например, таких как умная колонка, умные часы или смартфон, узнать прогноз погоды всего лишь одним предложением: «Привет, Сири, какой прогноз погоды на сегодня?».

Голос позволяет экономить время не только для передачи простых указаний сервисам, но и упрощать коммуникативную деятельность. Не обязательно отвлекаться от дел для того, чтобы передать коллеге или приятелю важную информацию. Современные сервисы позволяют отправлять голосовые сообщения, либо преобразовывать голос в текстовую информацию на ходу.

Голосовая коммуникация используется при построении архитектуры «умных» домов, при использовании голосовых ассистентов, при передаче информации и так далее

К сожалению, современный мир несет с собой и проблемы: с наступлением постиндустриальной эпохи все более острой становится проблема шумового загрязнения окружающей среды. Источники акустического загрязнения появляются не только на улице, например, за счет транспортных средств, но и внутри жилища человека, посредством бытовой и офисной техники. Человек, со временем, учится приспосабливаться к шуму окружающего пространства, чего нельзя сказать о технике. Повышенное шумовое загрязнение может помешать голосовым ассистентам понять команду, нарушить процесс обработки голосового сообщения в текстовый формат или помешать нашему собеседнику услышать или понять передаваемую информацию.

Поиск алгоритмов и способов фильтрации шумов – основная проблематика текущей работы.

В данной работе представлены как старые методы фильтрации шумов с помощью сложных математических преобразований, так и современные и актуальные способы с использованием «магии» нейросетей.

# 1 Описание предметной области

## 1.1 Основные понятия о цифровом звуке

Звук – это волна. Звукозаписывающее устройство (микрофон) реагирует на ее распространение, и сохраняет аналоговый сигнал. Чтобы передавать и обрабатывать аналоговый сигнал, его преобразуют в цифровой.

Цифровой звук – результат преобразования аналогового сигнала звукового диапазона в цифровой формат.

Простейшим методов преобразования является импульсно-кодовая модуляция (ИКМ). Основная идея метода состоит в представлении последовательности мгновенных значений уровня сигнала, измеряемого аналого-цифровым преобразователем за равные промежутки времени.

Цифровой звук является представление звука в виде чисел. Представим алгоритм обработки звука устройством:

1. Раздается какой-либо звук;
2. Микрофон «слышит» звук и преобразовывает его в электрический сигнал;
3. Микрофон посылает сигнал звуковой карте, которая преобразовывает сигнал в числа. Числа называются сэмплами (samples);

Сэмплы содержат информацию, сообщающую компьютеру, как сигнал звучит в отдельные моменты времени. Чем больше сэмплов использовано для представления сигнала, тем выше его качество. Например, чтобы сделать запись в цифровом виде, имеющее качество на уровне качества записи компакт-диска, устройство должно получать 44100 сэмплов в секунду. Число сэмплов, получаемых в секунду, называется частотой сэмплирования (sampling rate).

Размер каждого отдельного сэмпла также влияет на качество записи. Такой размер называется разрядностью (bit depth). Чем больше разрядность, тем выше качество звука.

## 1.2 Постановка задачи

С точки зрения продукта задача будет выглядеть следующим образом: на вход нашей подсистемы передается аудиопоток, с помощью некоторых преобразований которого будет выделяться полезная информация (в конкретном случае голос человека), и фильтроваться шум, например, от техники, животных, улицы.

При этом к технологии будут предъявляться следующие требования:

1. Работа в реальном, или приближенном к реальному времени (с задержкой не более 20 мс)
2. Легковесность, возможность запускаться на пользовательских устройствах
3. Качество, сопоставимое с другими решениями на рынке (например Zoom или Krisp)



Рисунок 1 – Высокоуровневое представление задачи