**Теория**

Структура файла wav, используемого для идентификации говорящего является переменной, так как в нее могут добавляться дополнительные блоки. Но все файлы этого типа имеют общий каркас:

|  |  |
| --- | --- |
| Местоположение и размер | Описание |
| 0..3 (4 байта) | Должно быть "RIFF". Является началом RIFF-цепочки. |
| 4..7 (4 байта) | Размер RIFF-цепочки в байтах. Иначе говоря, он должен быть равен (размер\_файла\_байт)-8. Несложно увидеть, что этот размер и есть два первых поля: RIFF-заголовок и размер RIFF-цепочки. |
| 8..11 (4 байта) | Должно быть "WAVE" |
| 12..15 (4 байта) | Должно быть "fmt " |
| 16..19 (4 байта) | Должно быть 16 |
| 20..21 (2 байта) | Формат |
| 22..23 (2 байта) | Количество каналов |
| 24..27 (4 байта) | Частота дискретизации |
| 28..31 (4 байта) | Количество байт, переданных за секунду воспроизведения |
| 32..33 (2 байта) | Количество байт в сэмпле |
| 34..35 (2 байта) | Количество бит в сэмпле |
| 36..39 (4 байта) | Должно быть "data" |
| 40..43 (4 байта) | Количество байт в области данных |

Для чтения файла достаточно определить структуру и считать данные в нее, последовательно сравнивая имеющиеся значения с эталонными.

В блоке данных содержатся значения амплитуды от времени. Для дальнейшего анализа необходимо преобразовать эти данные в амплитудно – частотные с помощью дискретного преобразования Фурье. На практике пользуются алгоритмом быстрого преобразования Фурье. Быстрое преобразование Фурье (БПФ) - это алгоритм вычисления преобразования Фурье для дискретного случая. В отличие от простейшего алгоритма, который имеет сложность порядка O(N2), БПФ имеет сложность всего лишь O(Nlog2N). Физический смысл дискретного преобразования Фурье состоит в том, чтобы представить некоторый дискретный сигнал в виде суммы гармоник. Параметры каждой гармоники вычисляются прямым преобразованием (БПФ). Таким образом, можно сказать, что преобразование Фурье переводит N значений амплитуд в N/2 синусоиды (отдельные частоты) с амплитудами A[n] и фазами Ph[n], плюс еще нулевая пара коэффициентов - просто константа (Sin и Cos с постоянным нулевым аргументом), отклонение от нуля.

Описание. На входе процедура получает параметры:

tnn - число значений функции. Должно быть степенью двойки.

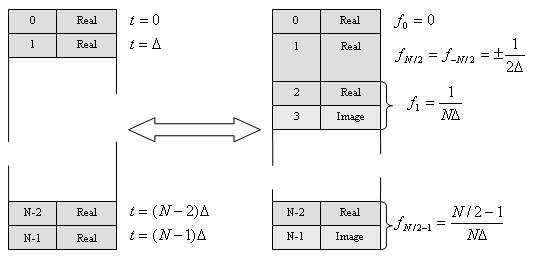
a - массив из tnn вещественных чисел, нумерация элементов от 0 до tnn-1. Если мы проводим прямое преобразование, этот массив задает tnn значений функции (не забываем, что функция вещественная, а не комплексная). Если преобразование обратное, то здесь хранятся соответствующие частоты (более подробно формат хранения приведен ниже, на картинке).

InverseFFT - направление преобразования. Если мы проводим прямое преобразование, то устанавливаем параметр в False. Устнавливаем его в True, если преобразование обратное.

На выходе:

a - содержит результат преобразования. Если проведено прямое преобразование, то здесь находятся частоты (более подробно формат хранения приведен ниже, на картинке), при обратном преобразовании здесь хранятся значения функции.

Рассмотрим более подробно формат массива. Пусть у нас есть N значений функции, расположенных с интервалом Delta. Каждому значению hk соответствует определенный момент времени tk .



После проведения преобразования Фурье мы получаем массив комплексных значений (пар вещественных чисел), в котором хранятся наши коэффиценты Hn. Существует зависимость между номером i и соответствующей ему частотой fi , между коэффициентом Hi и значением преобразования Фурье (непрерывного преобразования) на соответствующей частоте fi :



.

Для распознавания достаточно 16 частотных характеристик, которые нам дает алгоритм БПФ на выходе.

Для распознавания будем использовать нейронную сеть с тремя слоями: на первом слое 16 нейронов, на втором 150, на третьем 3. Метод обучения – метод обратного распространения ошибки. После обучения НС методом обратного распространения ошибки, нейронная сеть способна распознавать сказанное слово.