LAB EVAL 1

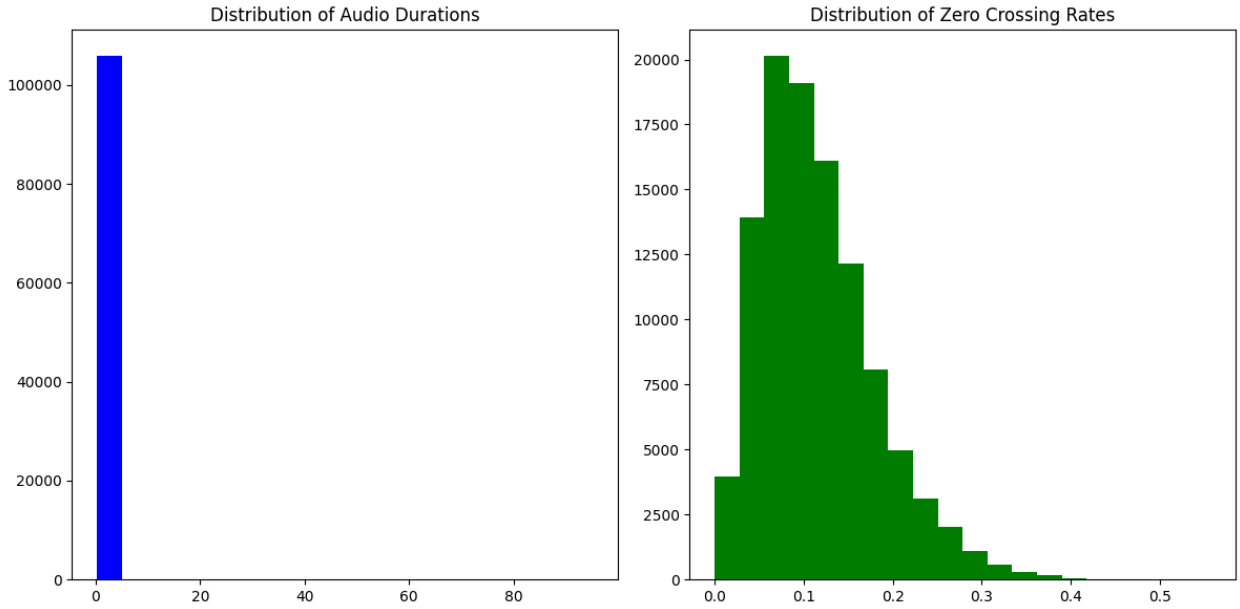
UCS749: Conversational AI: Speech Processing and Synthesis

SUMMARY-:

The paper presents the Speech Commands dataset, an audio collection designed for training and evaluating keyword spotting systems. It aims to facilitate the development of miniature models that accurately detect spoken words from a limited vocabulary while minimizing false positives from background noise.

DATASET-:

The dataset referred to in the task is from the paper "Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition" (arXiv:1804.03209). It is a collection of short audio clips of spoken words to train and evaluate speech recognition systems.
The dataset was analysed by extracting and examining basic audio features, including duration and zero-crossing rate. The duration of audio files varies within a specified range, while the zero-crossing rate shows a distribution that helps differentiate between audio commands. The plots of these features illustrate their distributions and provide insights into the dataset's characteristics.



The x-axis represents the individual audio samples, and the **y-axis** represents the duration of each file in seconds.

The x-axis would again be the individual audio samples (or words). At the same time, the **y-axis** would show the ZCR values, which are typically normalized or represented as a frequency of crossings per unit time.