Project for audio processing

Student names: Triet Nguyen and Quang Nguyen

1. **Introduction**

The primary objective of this audio processing project is to classify tram and bus sounds based on various acoustic features. The classification of transportation modes using audio data has practical applications in urban noise monitoring and intelligent transportation systems.

1. **Data description**

The audio data set is 22 WAV format files of tram audio and 30 WAV format files of bus audio. All the files are recorded with ZTE Axon 30 5G phone and OnePlus Nord 2 5G phone, originally in WAV format (no format conversion was performed on the files.) The bus audio data was recorded around the Keskustori central bus station area, including some when the bus was passing by, and some when the bus was sitting idle. The tram audio data was recorded mainly in the Sammonaukio tram stop (the first few of the files were recorded at the Opiskelija tram stop.) The sound was recorded when the tram was approaching, and when the tram was leaving. For convenience purposes, only 22 first bus audio files were used, to make the number of files the same as tram audio files.

All the audio files were uploaded to Freesound.org under two packs with account id thespicychip:

Tampere\_bus\_audio\_data: <https://freesound.org/people/thespicychip/packs/39926/>

Tampere\_tram\_audio\_data: <https://freesound.org/people/thespicychip/packs/39927/>

1. **Feature extraction**

The features that were chosen for the extraction include energy, root mean square (RMS), Mel-frequency Cepstral Coefficients (MFCCs), and Log-Mel Spectrogram.

Energy is normally used to represent the overall intensity of the audio signal and is useful for capturing loudness variation between tram and bus audio.

RMS provides a measure of the signal’s amplitude and is effective in characterizing the overall magnitude of the audio.

MFCCs capture the spectral characteristics of the audio signal. It is used to represent the human auditory perception of the audio signal.

Log-Mel Spectrogram represents the frequency content over time in a logarithmic scale and enables capturing both temporal and spectral information.

Energy and RMS features are fundamental measures of the signal’s intensity and amplitude, providing essential information about the overall loudness and magnitude of tram and bus sounds.

MFCCs can be extracted to capture the unique spectral features of tram and bus sounds, making them suitable for discriminating between the two audio classes.

Log-Mel Spectrogram combines the advantages of mel-frequency analysis and the logarithmic scale.

These features are chosen because they have discriminative power, robustness, and interpretability. This means that the features chosen to need to be effective in capturing distinctive characteristics of audio signals, allowing for meaningful discrimination between tram and bus sounds; to be robust to variations in background noise, environmental conditions, and different engine sounds; and to be easily represented both numerically and visually.

1. **Model selection, data split**

The provided code implements a Support Vector Machine (SVM) model for audio classification, specifically distinguishing between tram and bus sounds. It begins by extracting relevant features from the audio data and creating feature vectors for tram and bus instances. Labels are assigned to each class, and the dataset is formed by concatenating features and labels. The dataset is then split into training, validation, and test sets. Standardization is applied to the features using the StandardScaler. A linear-kernel SVM model is defined and trained on the scaled training data. The model's performance is evaluated on the validation set, and metrics such as accuracy, precision, and recall are calculated. Finally, the trained model is tested on a separate test set, and the accuracy of this unseen data is reported. This implementation provides a comprehensive approach to SVM-based audio classification, demonstrating the model's effectiveness in discerning between tram and bus sounds. Moreover, the decision boundary is a fundamental concept in Support Vector Machines (SVMs) and plays a crucial role in understanding how the model makes predictions. In the context of the provided SVM implementation for audio classification, the decision boundary is a hyperplane in the feature space that separates the instances of different classes, in this case, tram and bus sounds. Also, the code includes the implementation of the KNN model with the data. The process is like using the SVM model but instead, it will be embedded in the KNN library and get the metrics which is different.

1. **Results**

In the previous parts, we have discussed the process of analyzing the sound. In this stage, we have some result images which analyze the audio data. At first, Figure One is the image of the histogram for each feature with red color representing the tram data while the bus is for the blue color. So, if we focus on the energy and the RMS histograms, we can see that the tram file is more spread than the bus. And Figure 2, the image illustrates the decision boundary of the bus and tram audio data by using the SVM model. As from the picture, the RMS of both classes spread from minus 1 to 2 and the bus has data that go over it we can conclude that the audio is high. However, in Figure 3, which is for the KNN model, the graph is different from Figure 2 in that the shaded region of both classes is spread in an unordered way. It can be in that situation because the metrics of both models give different values. However, if we take a deep look into the similarities between two models in Figure 2 and Figure 3 we can see that the data gives the same values for both classes.

A group of graphs with different colored lines

Description automatically generated with medium confidence

**Figure 1: Feature histogram for both classes**

A blue and red graph

Description automatically generated

**Figure 2: The decision boundary for both classes by using SVM model.**

A diagram of a red and blue diagram

Description automatically generated with medium confidence

**Figure 3: The decision boundary for both classes by using KNN model.**

1. **Conclusion**

This audio analysis project has given us an overview of daily life sounds. We first extract different features of audio files and then analyze and discuss the result even though there are some troubles during the process such as miscommunicating, and identifying the work for the model it has been solved. One problem that also arises when processing the data is that the different number of tram and bus audio files will lead to error, and we still don’t have a fix on it, however, the number of files can be purposely kept the same throughout the two classes. Typically, this project can be expanded further with some ideas such as audio recognition and can work with a larger database to extend the project better.