POLITECNICO
MILANO 1863

April 29, 2025

Selected Topics in Music and Audio Engineering

Research project

Contents

1. **Overview** 1
2. **Objectives** 1
3. **General Instructions** 2
4. **Recommended Tools** 3
5. **Task: Music Instrument Classification** 3

1 Overview

This research project is designed for students to work collaboratively on a problem related to audio signal processing and machine learning. Students will define a clear research question, conduct a brief literature review, develop a baseline system, and iteratively improve it based on empirical evaluation. The project aims to provide hands-on experience in applying machine learning techniques to real-world audio data, fostering skills in data preprocessing, feature extraction, model training, and performance analysis. Additionally, students will gain insights into the challenges of working with audio datasets, such as handling noise, ensuring reproducibility, and interpreting results. By the end of the project, participants are expected to present their findings in a structured and professional manner, showcasing their ability to tackle complex problems in the field of audio analysis.

2 Objectives

The main objectives of this project are as follows:

- Understand and apply fundamental concepts of audio signal processing and machine learning.
- Gain experience in working with real-world audio datasets, including data loading, inspection, and annotation.
- Design a complete working pipeline, from preprocessing to performance evaluation.
- Experiment with different audio feature representations and analyze their impact.
- Implement a baseline system and progressively improve it through empirical testing and error analysis.
- Apply data augmentation techniques to increase model robustness and generalization.
- Develop teamwork skills by collaborating effectively within a group and dividing tasks fairly.
- Communicate findings through clear documentation and a concise oral presentation.

3 General Instructions

The research project will be carried out in groups of **three or four students**. Each group member is expected to become familiar with the research question early in the process. To this end, a short literature review is highly recommended to contextualize the research problem and identify existing approaches and challenges in the field. Furthermore, the group should clearly define how the workload will be divided among its members, ensuring an equitable and efficient distribution of tasks.

Regarding the task to be developed, students must thoughtfully consider several key aspects:

- **Audio feature representation:** Which feature extraction method is most suitable for the specific task at hand?
- **Modeling approach:** What type of machine learning model would be appropriate to address the problem effectively?
- **Evaluation strategy:** Which metrics should be used to assess performance, and how should the dataset be split into training, validation, and testing sets?
- **Baseline algorithm:** What could serve as a good initial baseline system that is simple and quick to implement, yet provides meaningful performance?

The groups will work with a given audio dataset. To this end, students are required to thoroughly familiarize themselves with the audio material by listening to several examples. They should describe the characteristics of the audio in detail, considering:

- **Audible content:** Are the recordings composed of isolated sounds, sound mixtures, notes, melodies, etc.?
- **Technical specifications:** What are the sample rate and number of channels?
- **Recording conditions:** Was the audio recorded in a studio environment or under field conditions?
- **Licensing:** Under which license was the dataset published, and what are its terms of use?

In general, the recommended working pipeline for the project should follow these steps:

1. Import audio data.
2. Import associated annotations.
3. Normalize the data as needed.
4. Split the data into training, validation (optional), and test sets.
5. Extract appropriate audio features.
6. Set up and configure the modeling algorithm (e.g., classifier or regressor).
7. Train the model using the training set.
8. Evaluate model performance using the test set.
9. Perform an error analysis to identify possible improvements.

The final goal is for students not only to implement a solid baseline system but also to improve upon it. To this end, students are encouraged to experiment with different feature representations, alternative modeling algorithms, and additional preprocessing techniques. Each modification should be followed by a new evaluation to objectively assess improvements and select the best model configuration.

Finally, students are required to document their work comprehensively. As part of the deliverables, each group must prepare a short presentation (approximately 4–5 slides) summarizing the project. The presentation should include the following elements:

- The research question and the motivation behind the project.
- An overview of the methodology and system architecture.
- The evaluation results, supported by plots and figures.
- Selected audio examples to illustrate the system's performance.

Proper documentation and clear, reproducible results are essential for successful project evaluation.

4 Recommended Tools

To support the development of your research project, the following tools are recommended. These cover essential stages such as coding, audio analysis, annotation, and presentation:

- **Programming and Development**
  - **Python** – the primary language for implementing machine learning models and processing audio data.
  - **Jupyter Notebook** – interactive environment for writing and running Python code, ideal for experimentation and documentation.
  - **Google Colab** – cloud-based alternative to Jupyter with free GPU support and easy sharing.
  - **Visual Studio Code** – a versatile and lightweight code editor that supports Python, Git integration, and rich extensions.
- **Audio Editing and Processing**
  - **Audacity** – open-source tool for editing, visualizing, and processing audio. Useful for inspecting raw audio material.
  - **Sonic Visualiser** – designed for analyzing and annotating audio content with waveform and spectrogram views.
- **Presentation**
  - **PowerPoint / Google Slides** – for preparing a clear and well-structured final presentation.

Students are encouraged to explore other tools as needed, provided they support the project goals and ensure reproducibility of results.

5 Task: Music Instrument Classification

This project addresses the development of a system for automatic music instrument classification. The dataset provided is the MedleyDB collection[1], which contains 196 professionally recorded multitrack recordings, including individual stems corresponding to isolated instruments.

Students are tasked with designing a classification pipeline that either recognizes instruments in multitimbral mixtures or classifies individual stems where typically one instrument is active. The project encourages a flexible approach, allowing exploration of both isolated and polyphonic scenarios.

Key aspects to investigate include:

- Analyzing the robustness of instrument recognition systems when facing different levels of overlapping instruments within a mixture.
- Studying the relationship between instrumentation and musical genre, as genre annotations are also available in the dataset.
- Exploring the use of co-occurrence matrices to model and understand typical combinations of instruments within different musical contexts.

The students should experiment with feature extraction techniques sensitive to timbral characteristics, such as spectral descriptors and MFCCs, and assess the effectiveness of classification

[1] `https://medleydb.weebly.com/`