Selected Topics in Music and Acoustic Engineering

Research Projects

Lecturer: Julio Carabias

Contents

Overview

Purpose

- Collaboratively solve a real-world problem using audio signal processing and machine learning.

Key Activities:

- Define a research question and conduct a short literature review.
- Develop and improve a baseline classification system.
- Work with real audio datasets (noise, variation, licensing, etc.).
- Perform empirical evaluations and interpret results.

Learning Outcomes:

- Practical skills in feature extraction, modeling, evaluation.
- Awareness of common challenges in audio analysis.
- Experience in building reproducible and effective ML pipelines.

Project Objectives

What Will You Learn and Practice?

- Apply audio signal processing & ML fundamentals.
- Explore real-world datasets: loading, inspecting, annotating.
- Build a full processing pipeline: from raw data to evaluation.
- Compare audio representations (e.g., STFT, MFCC, Mel).
- Implement & refine a baseline system through iteration.
- Use data augmentation to boost model generalization.
- Strengthen teamwork: fair task distribution & coordination.
- Present results clearly with documentation & a short talk.

General Instructions

Teamwork

- Work in groups of 3 or 4 students.
- Conduct a brief literature review to understand the research context.
- Define a clear task division: ensure balanced workload.

Key Aspects to Consider

- Feature Representation: What best suits your task? (e.g., MFCCs, Mel, chroma, STFT, etc.)
- Modeling: Choose a suitable ML model (e.g. Random Forest, SVM, NMF, CNN, RNN, etc.).
- Evaluation: Select metrics (accuracy, F1, SDR, etc.); define train/val/test split.
- Baseline System: Start simple, iterate and improve based on results.

General Instructions

Audio Data

- Content: isolated sounds, mixtures, melodies?
- Specs: sample rate, channels?
- Conditions: studio or field recordings?
- Licensing: Are there usage restrictions?

Recommended Pipeline

- Import audio & annotations
- Normalize & split data
- Extract features
- Train model
- Evaluate results
- Analyze errors & iterate

Deliverables

- Report:
    - Project motivation & question
    - Methodology overview
    - Evaluation (plots, results)
- 4–5 slide presentation including:
    - Project motivation & question
    - Methodology overview
    - Evaluation (plots, results)
    - Audio examples

Recommended Tools

Programming & Development

- Python: Main language for ML and audio processing.
- Jupyter Notebook: Great for combining code, results, and notes.
- Google Colab: Run notebooks in the cloud with free GPU support.
- Visual Studio Code: Powerful code editor with Git & Python integration.

Audio Editing & Annotation

- Audacity: Edit and inspect audio waveforms easily.
- Sonic Visualiser: Analyze and annotate spectrograms and musical content.
- Reaper: Digital Audio Workstation.

Presentation Tools

- PowerPoint / Google Slides: Create a professional and engaging presentation.

Tip: You're free to explore additional tools — as long as your setup ensures clarity, reproducibility, and effective collaboration!

Task I - Sound Event Classification

Objective

- Design and evaluate a system that classifies short environmental sounds into one of 50 predefined categories.

Dataset: ESC-50

- 2000 audio clips
- 5 seconds each
- 50 balanced classes (e.g., dog bark, thunderstorm, coughing...)

Key aspects to explore

- Feature extraction: STFT, Mel Spectrogram, MFCCs...
- Model selection: CNNs, traditional classifiers, etc.
- Data augmentation

Evaluation

- Use proper metrics (accuracy, confusion matrix...)
- Ensure reproducibility with clear documentation and code

Task II - Vehicle Type Classification

Objective

- Develop a machine learning system to classify vehicle types from audio recordings. Optionally, estimate their direction of movement.

Dataset: IDMT-TRAFFIC

- Real-world traffic recordings
- 4 main classes: car, bus, motorcycle, truck
- Stereo audio with realistic background noise

Key aspects to explore

- Mono vs. stereo audio: does spatial information help?
- Spatial features: inter-channel delay, energy difference...
- Spectrogram types: Mel, STFT
- Robustness to background noise and overlapping events

Evaluation

- Assess classification accuracy
- Accuracy of direction estimation
- Ensure reproducibility and provide documentation

Task III - Bird Activity Detection

Objective

- Detect bird vocalizations in real-world outdoor recordings (binary classification: active vs. not active).

Dataset: Warblrb10k

- 10-second smartphone recordings
- Diverse outdoor soundscapes: traffic, wind, voices, animals

Key aspects to explore

- Spectrogram input types: log-Mel, STFT, etc.
- Convolutional Neural Networks (CNNs) for pattern learning
- Impact of background noise and varying recording quality
- Generalization to unseen environments

Evaluation

- Binary classification performance
- Analyze failure cases due to background interference
- Ensure reproducibility and well-documented results

Task IV - Acoustic Scene Classification

Objective

- Classify 30-second binaural audio recordings into 10 acoustic scene categories (e.g., park, office, street, shopping mall).

Dataset: DCASE 2013 Task 1

- 30s real-world recordings from diverse environments
- Binaural format captures spatial information

Key aspects to explore

- Feature design for long-term patterns (e.g., temporal pooling, frame aggregation).
- Spectrogram types: log-Mel, STFT, etc.
- Binaural audio handling: separate vs. joint channel processing.

Evaluation

- Multi-class classification performance
- Robustness across varied acoustic scenes
- Emphasis on interpretability and reproducibility

Task V - Music Genre Classification

Objective

- Build a machine learning system to classify 30-second music tracks into 8 genres (e.g., pop, rock, jazz, etc.).

Dataset: FMA-small

- 8000 clips, 30 seconds each.
- 8 genres equally represented.

Key features to explore

- Rhythm: tempo, beat histograms.
- Harmony: chroma features.
- Timbre: MFCCs, spectral contrast.
- Combine multiple feature types for better accuracy.
- Compare random forest vs Neural Networks.

Evaluation

- Accuracy, precision, recall per genre.
- Generalization across diverse tracks.
- Justification of model and feature choices.

Task VI - Music Instrument Classification

Objective

- Develop a system to classify music instruments from audio, either in polyphonic mixtures or isolated stems.

Dataset: MedleyDB

- 196 multitrack recordings
- Individual stems corresponding to isolated instruments

Key aspects to explore

- Handling overlapping instruments in mixtures
- Instrumentation-genre relationship (e.g., genre annotations are included)
- Feature extraction focused on timbral characteristics

Evaluation

- Robustness of recognition under overlapping instrument conditions.
- Impact of feature selection on classification performance.

Task VII - Music Score Alignment

Objective

- Develop a system capable of aligning audio recordings with their corresponding symbolic music scores using automatic transcription and alignment techniques.

Dataset: URMP

- Multi-instrument chamber music performances
- Aligned audio, MIDI scores, and ground truth annotations

Key aspects to explore

- Audio transcription: From waveform to piano-roll representation
- Alignment: Use of DTW, HMM or similar methods for synchronization
- Feature selection: Impact of using chroma, onset, or pitch features

Evaluation

- Compare alignment output with annotated ground truth
- Quantitative: Alignment accuracy, precision/recall of note matching
- Qualitative: Visualization of aligned piano-rolls

Task VIII - Music Source Separation

Objective

- Develop a system for music source separation, capable of isolating individual instrument signals from a mixture recording.

Dataset: MUSDB18

- Professionally mixed multitrack recordings
- Isolated stems for vocals, drums, bass, and other accompaniment

Key aspects to explore

- Source separation methods: Traditional signal processing vs. deep learning models (e.g., U-Net architectures)
- Overlapping frequency content: Managing challenges from diverse mixing styles

Evaluation

- Metrics: SDR, SIR, SAR
- Robustness and performance evaluation through quantitative metrics
- Listening examples to qualitatively illustrate system performance

POLITECNICO
MILANO 1863