ESSENTIA API: A WEB API FOR MUSIC AUDIO ANALYSIS

Albin Correya Dmitry Bogdanov Pablo Alonso-Jiménez Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Spain
aacorreya@gmail.com, dmitry.bogdanov@upf.edu, pablo.alonso@upf.edu, xavier.serra@upf.edu

ABSTRACT

We present Essentia API, a web API to access a collection of state-of-the-art music audio analysis and description algorithms based on Essentia, an open-source library and machine learning (ML) models for audio and music analysis. We are developing it as part of a broader project in which we explore strategies for the commercial viability of technologies developed at Music Technology Group (MTG) following open science and open source practices, which involves finding licensing schemes and building custom solutions. Currently, the API supports music auto-tagging and classification algorithms (for genre, instrumentation, mood/emotion, danceability, approachability, and engagement), and algorithms for musical key, tempo, loudness, and many more. In the future, we envision expanding it with new machine learning models developed by the MTG and our collaborators to facilitate their access for a broader community of users.

1. MOTIVATION

Over the past ten years, the MTG has been actively developing a number of open-source software projects and tools for audio and music analysis, description and synthesis integrating many research outcomes from the music information retrieval (MIR) field. In particular, we developed Essentia, an open-source C++/Python library for audio analysis and audio-based MIR [1] licensed under AGPLv3. The library includes an extensive collection of reusable algorithms which implement audio input/output functionality, standard digital signal processing blocks, statistical characterization of data, and a large set of spectral, temporal, tonal and high-level music descriptors. More recently, it has been expanded with algorithms that allow predictions with pre-trained deep learning models and are designed to offer the flexibility of use, easy extensibility, and real-time inference. We maintain Essentia Models, a set of publicly available pretrained models for a variety of MIR tasks [2–5]. In addition, more recently we worked on Essentia.js [6], an open-source JavaScript (JS) library for audio and music analysis on both web clients and JS-based servers. These projects have been successfully applied in industrial contexts at different scales, including feature analysis on millions of audio tracks.

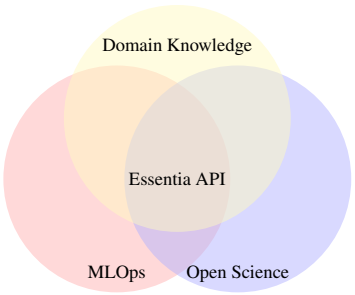

Figure 1. *Essentia API* context.

However, integrating the upstream Essentia library and models into an IT product requires some understanding of the technical details of the SDK and audio/music processing, and ML domain knowledge. These requirements may hinder the wider adoption of these tools among the broader software developer community, especially considering the agile software development practices [7] in the industry. Moreover, integrating ML technologies imply an extra layer of technical debt and computational complexities [8] into IT systems. In contrast, Web APIs are one of the most popular approaches to offer IT services with minimum integration cost for application developers due to the flexible pricing and offloading of computational and infrastructural complexities to the API provider. Such APIs, even involving the recent ML technologies, are becoming more affordable due to current advances in software tools and better computing infrastructure.

Indeed, there are already several domain-specific commercial APIs for music audio analysis as well as more broad cross-domain platforms that can be used to host ML models and run inference via cloud infrastructure (e.g., HuggingFace, Replicate). However, there is a lack of APIs that combine optimized MLOps infrastructure and music audio domain knowledge with the transparency of employed algorithms following open-source and open-science practices. In particular, it would be valuable for such an API to include some of the analysis algorithms established in the MIR community of researchers and practitioners.

In this context and following these principles (Figure 1), we aim to offer some of the technologies available at the MTG, specifically Essentia and related ML models,

© A. Correya, D. Bogdanov, P. Alonso-Jiménez, and X. Serra. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** A. Correya, D. Bogdanov, P. Alonso-Jiménez, and X. Serra, "Essentia API: a web API for music audio analysis", in *Extended Abstracts for the Late-Breaking Demo Session of the 23rd Int. Society for Music Information Retrieval Conf.,* Bengaluru, India, 2022.

Late-Breaking / Demo Session Extended Abstract, ISMIR 2022 Conference

through a web API to make it easily accessible for different new scenarios.

2. THE ESSENTIA API

We present Essentia API, a web API to access a collection of state-of-the-art music audio analysis and description algorithms based on Essentia. We summarize its main features below:

- The API provides end-points for synchronous and asynchronous analysis supporting custom queries with a parametrizable list of algorithms to execute.
- The algorithms are optimized in terms of latency and computational efficiency.
- The analysis is done on a track level. Each analysis request is related to a single track.
- All common audio formats are accepted. There is a limit on the maximum upload file size per query.
- Analysis results include an overall description summarized over time as well as a more detailed time-aligned description for some of the algorithms.
- Currently, analysis results are available with a short retention period, while the audio is never stored on the server.
- The API includes user-level authorization using JSON Web Tokens (JWT).
- Backward compatibility is ensured by versioning of models and algorithms.

Currently, we provide inference algorithms based on Essentia [1] and related models [2] for the following analysis tasks:

- Music style classification (400 music styles [4]).
- Music auto-tagging ($\sim$230 tags covering genre, mood, epoch, instrumentation, etc. [2, 4]).
- Single-label music classification (genre, mood, danceability, voice, instrumentation, engagement, approachability, etc. [2]).
- Tempo estimation (BPM).
- Musical key estimation (multiple profiles).
- Other semantic features (arousal/valence [5], approachability, engagement, danceability).
- Deep audio embeddings [3, 4].
- Loudness (EBU R 128 loudness standard metrics).

The API allows developers to integrate music analysis descriptors for a variety of use cases for characterization and management of music collections and retrieval therein, with possible applications in music distribution, recommendation, metadata management among many others.

The API is composed of analysis pipelines optimized for audio, making it more efficient than generic ML serving platforms. The algorithms and models are publicly accessible and open-source, and most of them are the outcome of academic publications. Many of these algorithms and models showcase recent developments in the MIR field. The employed ML models are optimized in terms of latency and computational efficiency of inference.

The API is a part of a broader initiative to facilitate technology transfer of research and development done at the MTG with the goal of increasing the sustainability of the development of open-source audio analysis tools by finding alternative funding models.

More information about the API is available online. [3]

[1] https://essentia.upf.edu/

[2] https://essentia.upf.edu/models.html

[3] https://essentia.upf.edu/api/

3. ACKNOWLEDGEMENTS

This work has been partially carried out under the project AISMA - PDC2022-133319-I00 funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU), the Agencia Estatal de Investigación (AEI) and NextGeneration EU as well as Musical AI - PID2019-111403GB-I00/AEI/10.13039/501100011033 funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU) and the Agencia Estatal de Investigación (AEI).

4. REFERENCES

[1] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "Essentia: an audio analysis library for music information retrieval," in *International Society for Music Information Retrieval Conference (ISMIR'13)*, 2013.

[2] P. Alonso-Jimenez, D. Bogdanov, J. Pons, and X. Serra, "TensorFlow audio models in Essentia," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'20)*, 2020.

[3] P. Alonso-Jiménez, D. Bogdanov, and X. Serra, "Deep embeddings with Essentia models," in *International Society for Music Information Retrieval Conference (ISMIR'20) Late Breaking Demo*, 2020.

[4] P. Alonso-Jiménez, X. Serra, and D. Bogdanov, "Music representation learning based on editorial metadata from discogs," in *International Society for Music Information Retrieval Conference (ISMIR'22)*, 2022.

[5] D. Bogdanov, X. Lizarraga-Seijas, P. Alonso-Jiménez, and X. Serra, "MusAV: A dataset of relative arousal-valence annotations for validation of audio models," in *International Society for Music Information Retrieval Conference (ISMIR'22)*, 2022.

[6] A. Correya, J. Marcos-Fernández, L. Joglar-Ongay, P. Alonso-Jiménez, X. Serra, and D. Bogdanov, "Audio and music analysis on the web using Essentia.js," *Transactions of the International Society for Music Information Retrieval*, vol. 4, no. 1, pp. 167—-181, 2021.

[7] K. Beck, M. Beedle, A. Van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries *et al.*, "Manifesto for agile software development," 2001.

[8] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden technical debt in machine learning systems," *Advances in neural information processing systems*, vol. 28, 2015.