STM JOURNALS

Journal of
Open Source Developments

ISSN: 2395-6704
Volume 12, Issue 3, 2025
September–December
DOI (Journal): 10.37591/JoOSD

https://journals.stmjournals.com/joosd

Review JoOSD

Comparative Analysis Between Librosa and OpenSMILE

Evanshi Chand*

Abstract

This research work focuses on comparative study of Librosa, a python-based library, and openSMILE, a C++ toolkit, with python bindings used in audio speech analysis. Librosa is ideal for beginners due to its simple structure and flexibility with strong integration with machine learning frameworks like TensorFlow and PyTorch. On the other hand, OpenSMILE is ideal for speech-centric tasks like speech-emotion recognition or paralinguistic studies, offering a wide range of pre-defined features like GeMAPS, eGeMAPS, ComParE etc. This study aims to contrast both tools based on various key factors such as their feature extraction, implementation, performance efficiency and limitations. The comparison focuses on their feature extraction capabilities, their performance and use, their strengths and weaknesses in context of audio processing feature extraction. Comparison between both would help users (researchers and developers) choose the better one for their tasks, whether its general audio analysis, emotion detection or large-scale speech feature extraction. After proper comparison of various features, this study tells that Librosa excels in flexibility and ease of use while openSmile offers higher scalability and deep speech based tasks.

Keywords: Librosa, openSMILE, speech recognition, feature extraction, python-based

INTRODUCTION

Speech Recognition holds a very important place in audio analysis which involves speech recognition, emotion detection etc. The tools used to extract features directly affect the quality of extracted material. Both Librosa and openSMILE have emerged as beneficial for audio extraction.

Librosa is a python-based library generally focusing on music speech processing. It has various tools to perform various tasks such as feature extraction, pitch detection and spectrogram. It is basically an audio analyzing tool. It offers various functions like MFCCS, chromogram, spectrograms, tempo and beat tracking etc. [1].

OpenSMILE (Open-source Speech and Music Interpretation by Large-scale Extraction) is a feature extraction toolkit used for speech recognition, speaker analysis and affective computing. It is very beneficial in extracting large scale feature sets more easily and efficiently. It can be used in C++ and Python programming language. It has a wide range of in-built features such as GeMAPS and eGeMAPS [2].

Understanding the differences and limitations between these two libraries helps user understand which one is suitable for them to work with. Hence this study aims to focus on the differences between both Librosa as well openSMILE.

***Author For Correspondence**
Evanshi Chand
E-mail: evanshichand467@gmail.com

Student, Department of Multimedia, BBK DAV College For Women, Amritsar, Punjab, India

Received Date: April 28, 2025
Accepted Date: September 20, 2025
Published Date: October 31, 2025

Citation: Evanshi Chand. Comparative Analysis Between Librosa and OpenSMILE. Journal of Open Source Developments. 2025; 12(3): 6–10p.

© STM Journals 2025. All Rights Reserved

METHODOLOGY

For this comparative analysis between both Librosa and OpenSMILE, following features are taken into consideration:

1. *Feature Extraction:* Listing and comparing all the tools Librosa and OpenSMILE have for extracting features, making them suitable for audio extraction. The aim is to identify the depth and flexibility of both Librosa and OpenSMILE such as, MFCCs, spectral features, chroma features, pitch and voice quality parameters; and finding the suitable one for performing various tasks including speech recognition, music information retrieval and emotion detection.
2. *Implementations and Use:* Analyzing which library is more suitable in which areas, and for what purposes. Librosa as a python library going well with machine learning frameworks. OpenSMILE being a C++ toolkit with python bindings, is used for large scale frameworks.
3. *Performance:* Comparing between Librosa and OpenSMILE, which is more suitable for large scale processing based on their programming language. This point includes how each of them performs in terms of speed, resource usage and integration in data pipelines. OpenSMILE is generally faster and more optimized for real time processing while Librosa gets slower working with large scale frameworks but offers flexibility.
4. *Limitations:* Comparing both tools based on their limited differences. Librosa is very easy to use but lacks some advanced speech specific features as available in OpenSMILE while OpenSILE though powerful but is less user friendly for beginners.

LITERATURE REVIEW

Several studies have compared different audio libraries in various contexts focusing on their design, functionality, and application domains. Some of the notable ones are as follows:

McFee *et al*. gave a brief summary of designs principles and functionality of Librosa [3]. This study emphasizes that Librosa is easy to use and beginner friendly due to its simple design build. This study generally focuses on Librosa being used in audio and music signal analysis in python with its features helping in spectral analysis, beat tracking, pitch detection and feature extraction. This study links Librosa to topics such as digital signal processing, machine learning, information retrieval and library science. It also reveals that how Librosa facilitates educational and research purposes by offering intuitive APIs and visualization capabilities.

The study by Eyben *et al*. on OpenSMILE declares it as open-source tool for feature extraction, which can be used in C++ and python [4]. This paper also discusses its built-in features with GeMAPS (Geneva Minimalistic Acoustic Parameter Set) and eGeMAPS (extended GeMAPS) which enhance emotion and speech analysis [2]. This paper even compares OpenSMILE with other tools making it a good choice in speech related machine learning areas. These built-in configurations simplify and standardize extraction of various parameters for affective computing and health diagnostics. This study also presents a comparative analysis of OpenSMILE with other available tools, demonstrating its processing capabilities and suitability for speech-related machine learning applications.

These studies provide an overview of both Librosa and OpenSMILE presenting their pros and cons and helping in a comparative study of both. These comparisons provide a solid foundation for conducting a thorough comparative analysis, helping users (developers and researchers) to select most appropriate toolkit based on requirements.

OVERVIEW OF LIBROSA AND OPENSMILE

Librosa

Librosa is a python-based library used for manipulating audio signals. It can be used for processing audio signals in music and speech processing fields. It is an essential tool for researchers and developers in machine learning. Librosa has a wide range of features used to manipulate audio signals such as MFCCs (Mel-frequency cepstral coefficients, chromagram, spectral features, zero crossing rate (ZCR), Mel spectrogram, spectral contrast, chroma features, and temporal centroid. Librosa supports various formats for audio file such as, WAV, OGG MP3 and FLAC [1].

© STM Journals 2025. All Rights Reserved

To start with Librosa, audio data preprocessing is the main step as it helps in resampling, time-stretching, scaling, as well handling missing data. Then, extracting features, which includes Mel-spectrogram, chroma features, ZCR, temporal centroid can be used for SER (speech emotion recognition). For visualizing and analyzing audios in form of waveforms, spectrograms and other forms, Librosa offers various tools for pitch estimation and key identification. Librosa even manipulates audio by noise reduction, denoising etc.

OpenSMILE (Open Speech and Music Interpretation by Large Space Extraction)

OpenSMILE is a feature extracting toolkit focused on speech and emotion recognition (SER) [5]. OpenSMILE is a C++, Python and MATLAB (Matrix Laboratory) open-source library for audio and video input. It has a very modular architecture, which helps in installation of extensions through plug-ins. OpenSMILE is capable of decoding the given music piece or speech. This software further can also detect music mood as well its chorus, keys, chords, genre, tempo, style and many more things.

It can extract a rich range of audio features like spectral features such as MFCCs, Spectral Centroid, Prosodic features like Pitch, Intensity, Voice quality features like shimmer, Harmonics-to-Noise ratio, rhythmic features, loudness or formats [6]. It also includes feature sets used for researches such as Emobase (consisting 988 features for emotion recognition), BoAW (Bag of Audio Words), eGeMAPS (contains 88 features for voice analyzing), and ComParE (containing 6373 features for computational paralinguistics).

Librosa vs. OpenSMILE

Both Librosa and OpenSMILE are audio processing libraries, they have a lot in common but on the other hand they have many differences also. Librosa is a python-based machine learning library mainly focusing on music speech processing. And on the other side, OpenSMILE focusses on speech recognition, speaker analysis and affective computing. Librosa helps in feature extraction of audio files, beat tracking, pitch detection as well as spectrogram analysis. But on the other side OpenSMILE is a bit different as its main function is extracting large feature sets from speech. It also includes spectral and voice quality features [7].

Feature Extraction

Librosa offers flexibility in feature extraction, while OpenSMILE provides built-in features for speech analysis. Both Librosa and OpenSMILE extract different types of features. Both libraries can extract MFCCs (Mel frequency cepstral coefficients) and spectral features. Librosa offers Chroma features with tempo and beat tracking, but these features are not supported by OpenSMILE. On the other hand, OpenSMILE is mainly used in speech and affective computing, supporting features like prosodic features (such as Pitch, Energy, Intensity) and voice quality features like jitter, shimmer, HNR (Harmonics to noise ratio) as well predefined feature sets such as eGeMAPS, ComParE, and IS10 (interspeech feature sets) which are not supported by Librosa. These all-feature extractions make OpenSMILE more suited for speech emotion recognition (SER) and effective computing.

Implementations and Use

Librosa is a python-based library which is beginner friendly with machine learning frameworks like TensorFLOW and PyTorch, as well as it is user defined as changes are made according to user's code. It can be used with NumPy and SciPY libraries making it more flexible in feature extraction. OpenSMILE is a C++ library which can also be used with python, java, and MATLAB. It is basically used for speaker identification and acoustic scene classification [8]. It is a tool with configuration files in it which are used with .conf extension. It has built in feature set included in it. It is very useful in areas where there is a lot more to configure as well as it extracts thousands of features in one go.

Performance

Librosa is slower as compared to openSMILE for large datasets as OpenSMILE has a lot of built in features supporting large scale speech datasets [9]. Areas where both can be applied differs a lot in some

© STM Journals 2025. All Rights Reserved

or the other way. Librosa is easy to use in deep learning areas as well basic Librosa code can be used for speech analysis and music analysis which are quite difficult to use in terms of openSMILE OpenSMILE can be used for SER (speech emotion recognition), speaker identification as it offers a wide range of features for feature extraction. Hence, Librosa is used in music and basic audio but OpenSMILE is used for extracting speech features, emotions and speaker characteristics like its emotion also genre of the audio with all the chords rhythms etc.

Limitations

Both Librosa and OpenSMILE offer wide range of functions in areas working with audio signals but both libraries have some limitations also. Librosa is easy to use as compared to OpenSMILE because of its beginner friendly functions but OpenSMILE is a bit tricky to use as it has some complex configurations. On the other hand, openSMILE has a wide range of features useful for audio extraction as it has some advanced features mainly focused on speech recognition which are not found in Librosa making Librosa not very suitable for speech recognition activities [10].

CONCLUSION

In conclusion both Librosa and OpenSMILE have their strengths and weaknesses in terms of audio analysis tool. Both tools are useful according to what output one is in need of or the nature of the audio data, the required features, and performance constraints. If the ultimate goal is general audio processing, Librosa is a better choice but if the goal is to extract features or find the genre rhythms movements analyzing emotions chords as well as other characteristics, Librosa is a better choice. In short, Librosa is made for flexibility and analysis but it is difficult to use in areas of large datasets. OpenSMILE is based on feature extraction and it is much easier to use in areas of large datasets due to its fast-processing ways. OpenSMILE can extract 6000 features, including sets like GeMAPS and eGeMAPS, which are widely used in emotion recognition, speaker profiling, health diagnostics, and paralinguistics analysis which are much higher than what Librosa can do or have.

Therefore, choice between Librosa and OpenSMILE depends on several factors such as the nature of audio data, specific features required, performance constraints and application. Librosa is a good choice if tasks is to explore general audio or music features in a flexible environment. OpenSMILE is a good choice if task requires large scale, speech oriented or emotion sensitive feature extraction with higher speed and precision.

REFERENCES

1. Tzanetakis G, Cook P. Musical genre classification of audio signals. IEEE Trans Speech Audio Process. 2002 Jul 31; 10(5): 293–302.
2. Atmaja BT, Akagi M. On the differences between song and speech emotion recognition: Effect of feature sets, feature types, and classifiers. In 2020 IEEE region 10 conference (TENCON). 2020 Nov 16; 968–972.
3. McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E, Nieto O. Librosa: Audio and music signal analysis in python. SciPy. 2015 Jul 6; 2015: 18–24.
4. Eyben F, Wöllmer M, Schuller B. Opensmile: the Munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia. 2010 Oct 25; 1459–1462.
5. Lenain R, Weston J, Shivkumar A, Fristed E. Surfboard: Audio feature extraction for modern machine learning. arXiv preprint arXiv:2005.08848. 2020; 1–5.
6. Singh A, Chodankar S, Suvarna A. Audio feature extraction tools. Int Res J Mod Eng Technol Sci. 2021 Apr; 3(4): 2340–2345.
7. Chen J, Ye J, Tang F, Zhou J. Automatic detection of Alzheimer's disease using spontaneous speech only. In Interspeech. 2021 Aug; 2021: 3830.
8. Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, Chetouani M, Weninger F, Eyben F, Marchi E, Salamin H. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. INTERSPEECH 2013. 2013; 1–6.

© STM Journals 2025. All Rights Reserved

9. Kumar Ankit, Singh Kshitiz, Sharma Anmol, Gupta Sachi. Advances in Speech Emotion Recognition and Analysis: A Review of Applied Machine Learning Methodologies. Int J Res Appl Sci Eng Technol. 2024; 12(4): 4617–4621.
10. Schuller B, Wallhoff F, Arsic D, Rigoll G. Musical signal type discrimination based on large open feature sets. In 2006 IEEE International Conference on Multimedia and Expo. 2006 Jul 9; 1089–1092.

© STM Journals 2025. All Rights Reserved