MIR.EDU: AN OPEN-SOURCE LIBRARY FOR TEACHING SOUND AND MUSIC DESCRIPTION

Justin Salamon[1,2]
[1]Music and Audio Research Lab, NYU
[2]Center for Urban Science and Progress, NYU
justin.salamon@nyu.edu

Emilia Gómez
Music Technology Group
Universitat Pompeu Fabra, Barcelona, Spain
emilia.gomez@upf.edu

ABSTRACT

We present MIR.EDU: an open-source library designed specifically for teaching sound and music description. Motivated by the need for MIR tools tailored for education, the library is built based on four design principles: easy interaction through a graphical user interface including visualisation and sound reproduction; simple, documented, open-source code; multi-platform support and independence of proprietary software; and applicability to varied audio signals including environmental sounds and music. MIR.EDU has been used by more than one hundred students over the past two academic years in both undergraduate and graduate courses. It is available as an open-source project on GitHub.

1. INTRODUCTION

Whilst the amount of teaching material available for entry-level courses is constantly growing, the tools available for giving students hands on experience with sound description and MIR are relatively limited, especially when targeting students with limited coding experience. There is, of course, a large selection of sound analysis libraries that have been created by the research groups (e.g. Essentia [1] and MIRToolbox [4] among others), but these libraries are primarily designed with the research community in mind, meaning the code is often highly optimised for computational efficiency, or is written using several layers of abstraction for modularity and for managing the sheer size of these libraries (e.g. Essentia includes over one hundred thousand lines of code). Such libraries are excellent for a researcher or experienced graduate student, but less ideal for a beginner student interested in reading the code to understand how the computation of a certain audio descriptor is carried out. In the course of our work as MIR researchers and teachers, we have identified a genuine need for a sound description library that combines simple and readable code with visualisation and sound reproduction capabilities. To address this need, we developed MIR.EDU – a sound analysis Vamp library specifically designed for educational purposes. MIR.EDU is not optimised, it is not comprehensive, nor does it include state-of-the-art algorithms. Rather, it is designed with two simple goals in mind: (1) provide readable, open-source code, and (2) offer easy and intuitive interaction through a graphical user interface including visualisation and sound reproduction. MIR.EDU is available on GitHub [1] and has already been used in two undergraduate courses and one graduate course and has proved to be a valuable educational tool for teaching sound and music analysis.

2. MIR.EDU

2.1 Target audience

The motivation for writing MIR.EDU came from an undergraduate course we gave on sound analysis and design. The programming experience of the students varied from seasoned programmers to students with only basic knowledge of Matlab or Processing. We could only assume basic knowledge about audio analysis, and the concepts behind audio descriptors (features) and MIR algorithms were new to most students. The library is also being used for a Masters course on audio and music analysis, and we intend to use it in a seminar on sound analysis for Master's students in Urban Informatics.

2.2 Design and Architecture

Given this diverse set of backgrounds, we had to design a tool that would be easy for the students to install and use, and would facilitate the understanding and exploration of the basics of sound analysis in an intuitive manner with minimal coding. Consequently, MIR.EDU is designed based on the following principles: **Easy interaction:** students should be able to use the library through a graphical user interface without having to write any code. The interface should provide means to listen to the analysed sound alongside a visualisation of the descriptor values computed by the library. This is achieved by writing MIR.EDU as a library of Vamp plug-ins [2] that can be used with Sonic Visualiser [3], a GUI that provides sound reproduction and visualisation for Vamp plug-ins. Batch processing is supported

[1] https://github.com/MTG/miredu
[2] http://www.vamp-plugins.org/
[3] http://www.sonicvisualiser.org/

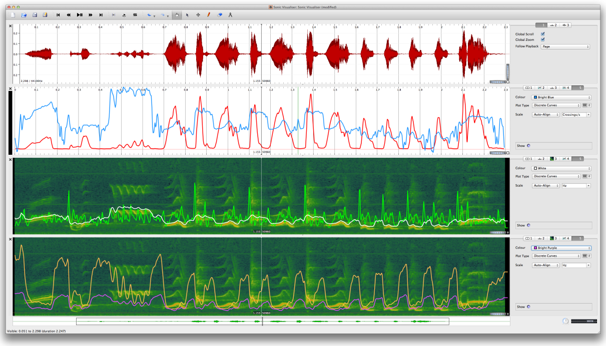

Figure 1. Screenshot of MIR.EDU used with Sonic Visualiser. Top pane: the waveform of a birdsong recording. Second pane: RMS (red), zero crossing rate (blue) and temporal centroid (horizontal green line). Third pane: spectrogram, spectral centroid (white) and spectral flux (bright green). Fourth pane: spectrogram, spectral spread (purple) and spectral flatness (orange).

through Sonic Annotator [4]. **Simple code:** the code should be as simple and human readable as possible, with detailed comments to walk the student through the code. Efficiency is not a concern, and rather the code should correspond as closely as possible to the descriptor's mathematical definition. **Multi-platform support and independence of proprietary software:** the code should be open source, work on all three major operating systems (Linux, OSX and Windows) and should not depend on any propriety libraries or software (e.g. Matlab). **Applicability:** the initial set of features in MIR.EDU should be applicable to a varied range of audio signals, including environmental sounds and music.

MIR.EDU is implemented in C++, the native language of Vamp plug-ins. Each descriptor is implemented in a separate source file (with a corresponding header file), and the implementation is fully self contained and independent of all other descriptors. Doxygen [5] documentation is provided at the top of every source and header file, where we specify the descriptor name, description, unit, formula and a reference for further reading. This information can be accessed either by directly inspecting the code, or by browsing the library's doxygen documentation folder.

2.3 Features

Currently, MIR.EDU includes a selection of 14 low-level time-domain and frequency-domain descriptors. We selected a small but informative set of descriptors valid for sound and music signals partly based on the MPEG-7 standard [3]. The implementation of most of the descriptors is based on the formulae provided in [5]. For descriptors that represent standard statistics (RMS, centroid, etc.) a link to the relevant definition in Wikipedia is also provided. As noted earlier, the reference formula and/or implementation for each descriptor is specified in its doxygen documentation. A summary of the audio descriptors currently implemented in MIR.EDU is provided in Table 1. An example of using MIR.EDU with Sonic Visualiser to analyse a birdsong recording is depicted in Figure 1.

[4] http://www.vamp-plugins.org/sonic-annotator/
[5] http://www.doxygen.org/

Feature	Reference
Time domain:	
RMS	Online [6]
Temporal Centroid	[5]
Attack Start & End Times	[5]
Log Attack Time	[5]
Zero Crossing Rate	[5] & Online
Frequency domain:	
Spectral Centroid	[5] & Online
Spectral Spread	[5] & Online
Spectral Skewness	[5] & Online
Spectral Kurtosis	[5] & Online
Spectral Rolloff	[5]
Spectral Flatness	[5]
Spectral Crest	[5]
Spectral Flux	[5]
MFCC	[2] & Online

Table 1. Audio descriptors implemented in MIR.EDU and reference formula / implementation.

3. SUMMARY AND FUTURE PERSPECTIVES

Currently, the library includes a relatively small but illustrative set of 14 low-level descriptors with clear applications in sound and music description. In the near future we plan to add chroma features to facilitate the teaching of tonality analysis. We are also considering creating a python port of MIR.EDU in the future, if we see there is demand for one. By sharing MIR.EDU as an open source project on GitHub we hope that more features will be added through a collaborative effort of the MIR community.

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 *14th Int. Soc. for Music Info. Retrieval Conf.*, pages 493–498, Brazil, 2013.

[2] Todor Ganchev, Nikos Fakotakis, and George Kokkinakis. Comparative evaluation of various MFCC implementations on the speaker verification task. In *10th Int. Conf. on Speech and Computer*, volume 1, pages 191–194, Patras, Greece, Oct. 2005.

[3] H.-G. Kim, N. Moreau, and T. Sikora. *MPEG-7 audio and beyond: Audio content indexing and retrieval*. Wiley, 2006.

[4] O. Lartillot, P. Toiviainen, and T. Eerola. A matlab toolbox for music information retrieval. In C. Preisach, P.D.H. Burkhardt, P.D.L. Schmidt-Thieme, and P.D.R. Decker, editors, *Data Analysis, Machine Learning and Applications*, pages 261–268. Springer, 2008.

[5] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams. The timbre toolbox: Extracting audio descriptors from musical signals. *J. Acoust. Soc. Am.*, 130(5):2902–2916, 2011.

[6] All references to online resources can be found in the library's doxygen documentation.