
# Towards An Educational Music Processor for Folk and Popular Musics


**Phan Gia Anh-Thư & Ngô Thanh Nhàn**

*Center for Vietnamese Philosophy, Culture & Society*, Temple University
*Music & Music Education,* Columbia University
*CIMS, Medical Language Processing,* New York University


## INTRODUCTION

This poster describes **an educational musical processor** (EMP) that takes **spectrographic data of a sonic object** and **turns them into a series of meaningful layers associated with different musical knowledge representations**, in such a way that it can be understood, reproduced, played, compared, and taught by everyone across cultures, regardless of their musical backgrounds. Any music audio file can be used as input.

Within the scope of this project, the authors focus on (1) processing of musical audio files onto a common platform of physical sound properties, measured in Hertz, Decibels, and milliseconds, so that (2) culturally dependent musical units such as notes, beats, measures, phrases, chords, and sections can be viewed in separate interpretive layers. (3) Syntactic techniques, such as frequency of occurrences, and adjacency are applied to musical units, such as pitches and musical chords. They are called key pitches in context (kpic) and key chords in context (kcic). The results are then mapped onto circles of fifths.

The EMP thus reveals distinct patterns of each song, each section of a song, of each artist, each genre, and each culture.

Semi-automatic generation of layers of annotations on top of the spectrogram helps teachers to quickly discover and/or compare distinctive features of a song, while preparing lessons. Learners of all levels can choose the most prominent patterns of the song to learn. This can also advance methods for preservation of sonic objects in the future.

## Method

### Obtaining Data

Data of the EMP can be divided into two categories: phenomenal and conventional (or knowledge). **Phenomenal data** are physical signals such as time, frequency, and intensity. **Conventional data** are interpretive symbols and knowledge that have been used in music and languages in different cultures and education. Examples of conventional data are musical beats, pitches, phrases, sections, quarter notes, measures, treble and bass clefs, and chords, which exist only in culturally learned contexts.

### Technologies Used

Top-of-the-art technologies are tested and incorporated at this moment based on three criteria: user-friendliness, cost, and standardization.

Our own software, the EMP, runs on about 6,000 lines of codes in PHP 7.2 and SVG 1.1, on a Dell T1650 workstation, with Ubuntu 18.04.1. It follows the international, world wide web, archival, librarian, multilingual standards, and measurement units: Unicode and ISO/IEC 10646, HTML, XML, MusicXML and URI, OAI/ORE with Dublin Core DCMI, Hertz for pitch, Decibels for intensity, Milliseconds for time, and frame (1 frame = 1/44,100 second) where 44,100 Hz is one of the standard digital sample rates.

*Sound Analysis Software*

1. *Sonic Visualiser,* is a freeware that visualizes audio input as soundwaves and spectrograms. It visualizes manual or plug-in annotations as overlaid graphic layers over spectrograms. The EMP currently uses these plug-ins (and will be replaced for the better):
    a. *Chordino Chord Estimation,* generates an Sonic Visualiser Layer [SVL] output containing temporal frames associated with the chord labels.
    b. *Melodia Melody Extraction*, with the melodic pitches.
    c. *Queen Mary Note Onset Detector/Percussive Onset Detector* with an onset of a percussive sound.
    d. *BBC Intensity* with intensity values.
2. *MuseScore (offline)* is a free standalone open-source software for music notation. The companion *MuseScore (online),* is an online platform allowing users to share and edit sheet music, and connect to other websites, or through Youtube videos. *MuseScore* offline and *MuseScore* online versions are used *in tandem* by the EMP to create a panel of western musical notation.
3. *AnthemScore* is a standalone open-source software which suggests a transcription of an audio file (MP3, WAV, etc.) into sheet music based dominant frequencies' amplitudes of the song's spectrogram.

### A Test Case: Despacito by Luis Fonsi 2017

The video version of *Despacito* by Luis Fonsi ft. Daddy Yankee (hereafter, simply *Despacito*). The song, 4 minutes and 42 seconds long, the most viewed YouTube video, is used as a test case for the EMP.


### Acknowledgment

The authors thank, in the name of the *Folk Music Studies* project:
(1) the NYU's Courant Institute of Mathematical Sciences, the *Linguistic String Project*'s Medical Language Processing,
(2) the *Center for Vietnamese Philosophy, Culture & Society* of Temple University with part of Henry Luce Grant No. 270380 / Expanding Opportunities for Research on Vietnam, and
(3) Ambassador Nguyễn Phương Nga of the *Permanent Mission of the Socialist Republic of Vietnam to the United Nations.*

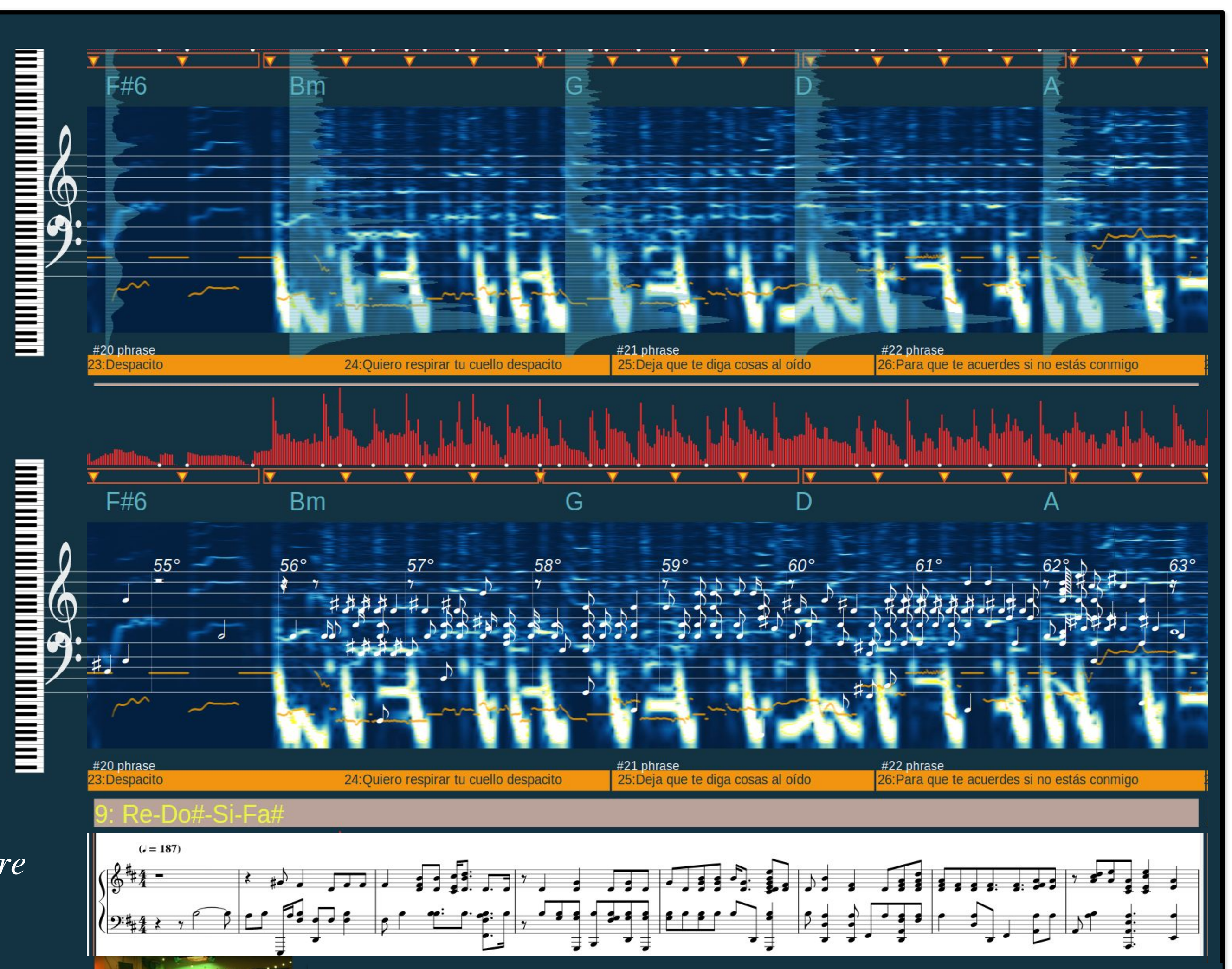

## Results

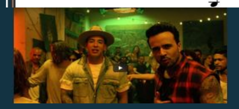
### Phenomenal and Conventional Data Plotted as Layers:

*Despacito* is input to *Sonic Visualiser* to produce the base spectrographic data, labelled Layer 0.

*Layer 0 to 9*, takes about 235 seconds to produce by the EMP, viewed at http://mlp.cs.nyu.edu/vietmusic/au2spec.php.

- Layer 0: *Despacito* spectrographic graph
- Layer 1: a treble and a bass clef, with a piano 88 keys from $A^0$ (900 *cents*) to $C^8$ (9,600 *cents*)
- *Layer 2:* Labelling of *Beats, Measures, Chords, Phrases* and *Section.*
- *Layer 3:* sum of intensity in dBs by *measures*
- *Layer 4:* sum of intensity in dBs by *beats*
- *Layer 5:* sum of intensity in dBs by *chords*
- *Layer 6:* sum of intensity in dBs by *phrases*
- *Layer 7:* sum of intensity in dBs by *sections*
- *Layer 8:* music sheet of *Despacito* according to *AnthemScore*
- *Layer 9: AnthemScore estimation* is displayed through *MuseScore* applications.



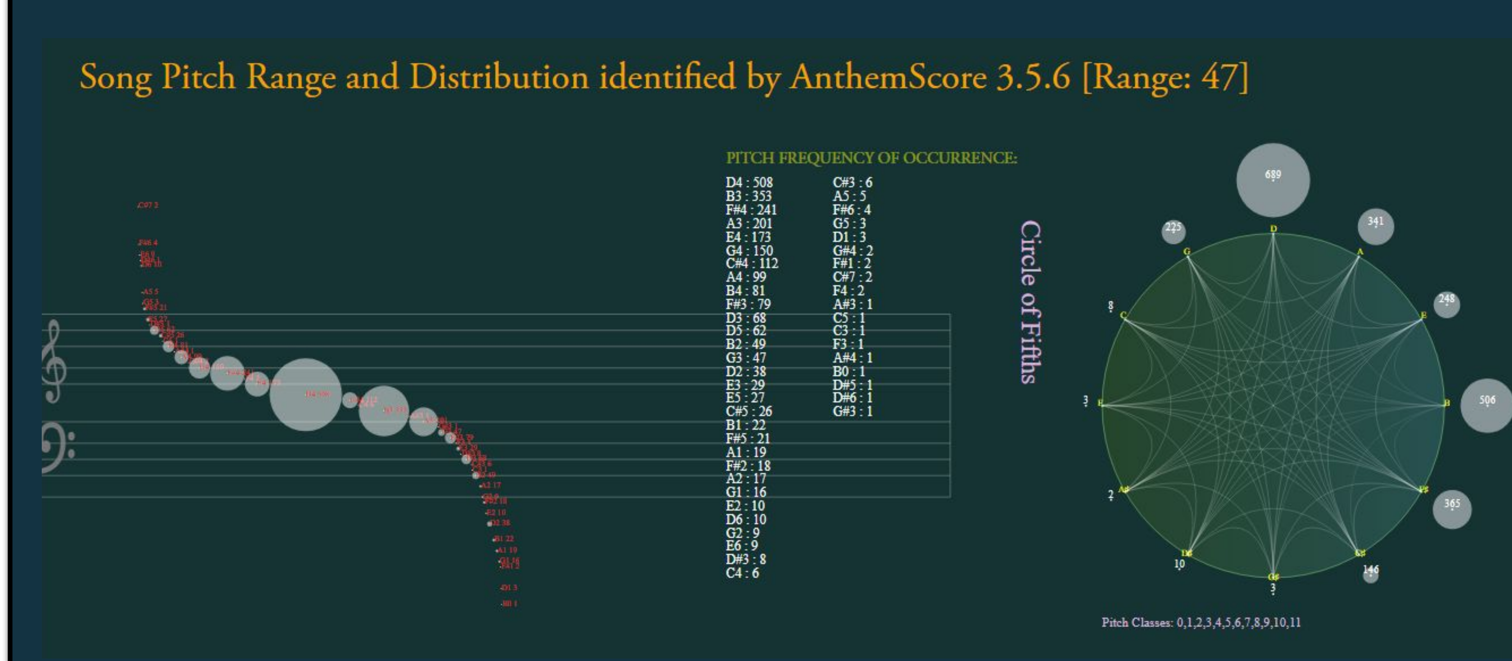



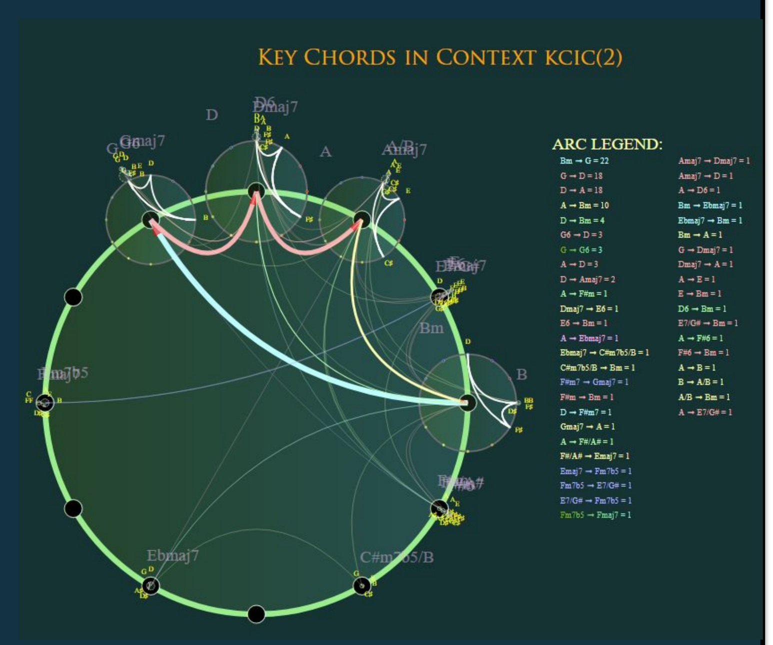
*Examples of data plotted as layer 0, 1, 2, 5, 8, 9 of section 9 (section G-see the "Song form" below) of Despacito's YouTube video*

### Pedagogical Presentations:

The EMP analysis based on the Layers currently take about 20 seconds by the EMP, can be viewed at http://mlp.cs.nyu.edu/vietmusic/au2ana.php.

- *Pitch and chord frequencies of occurrences, and key chords in context kcic(n), and their visual representations on the Circles of Fifths.*




- *Lyrics/chords in phrase-length bars* (see right figure for an example of lyrics, chords, and four phrases of section 9) →
- *Song form:*

A B C D E E F F **G G** H H I J J G G H H G J J G J K

*section 9*

F#6 Bm
phrase 23. Despacito
G
phrase 24. Quiero respirar tu cuello despacito
D
phrase 25. Deja que te diga cosas al oído
A
phrase 26. Para que te acuerdes si no estás conmigo

## Conclusions and Implications

Music processing continues the vision of bringing the arts of artificial intelligence in natural language processing, to a related human innate activity: music since the 1980's. By analyzing and representing a musical artefact from the perspective of listeners, the EMP is different from other music processing software most of which tend to facilitate the experimentation of synthesizing existing and/or novel sounds. The EMP also adds depth and values to the current stage of information technologies. Education-wise, the EMP approaches a model of music teaching and learning that would be generally accessible, while not sacrificing the contents and quality of musical experiences.

### Beginners

The EMP approach allows beginner music learners to grasp the gist of a song without years of ear training or music reading. For examples, following the layers of melodic contour and lyrics, beginner music learners understand where to raise or lower their voice to match, or with the layer of phrases to breathe or take a break appropriately. If the learners change their mind, they can manually change the phrase or section breaks and do again. Finally, the EMP structure of the song shows exactly how many times a section is repeated, thus aiding to the memorization strategy.

### Experts

The EMP approach presents a "fullest" sonic data into visual and aural representations. By fullest, the authors mean to preserve all the phenomenal signals that are often omitted in traditional notation. For example, using the chromatic scale (12 pitches) in western notation ignore all the sounds in between well-tempered pitches. These "in-between" pitches make Indian music sounds distinct from Vietnamese music, for example. Thus the spectrogram reveals the microtones and transitions that the western notation often ignores. Additionally, the representation of intensity also outperforms the dynamic notion (i.e., piano vs forte symbols). Last but not least, because one chord could be played in infinite manners, the intensity in Layers 3 to 7 inform players to shape the chords to best match the recording.

### Everyone in Between

For most music learners that fall between beginners and experts, the EMP approach allows them to enjoy music of their favorite in multiple learning styles. For example, they can listen to each section and play along if they are a aural learner. On the other hand, a visual learners can follow the notations in MuseScore section. If they are familiar with lead music sheet, the EMP also automatically generate lyrics, chords, and/or melody for them. Most importantly, it allows any teachers to analyze and devise lesson plans on any songs suggested by the students. The lessons are automatically tailored to match the student's ability. Teachers are then able to have more time focusing on accommodating other peculiar student needs, encouraging individual interpretations, and providing contextual knowledge. As the EMP proceeds to calculate prevailing rhythmic and melodic patterns of a songs by constantly comparing units of data based on the theory of string grammar and deep learning, these patterns serve as basic blocks for improvisation and stylistic recognition.

IEEE

ĐẠI HỌC TÔN ĐỨC THẮNG
TON DUC THANG UNIVERSITY