TIV.LIB: AN OPEN-SOURCE LIBRARY FOR THE TONAL DESCRIPTION OF MUSICAL AUDIO

António Ramires

Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
antonio.ramires@upf.edu

Gilberto Bernardes

INESC TEC and University of Porto
Faculty of Engineering
Porto, Portugal
gba@fe.up.pt

Matthew E. P. Davies

University of Coimbra
CISUC, DEI
Coimbra, Portugal
mepdavies@dei.uc.pt

Xavier Serra

Music Technology Group
Universitat Pompeu Fabra
Barcelona, Spain
xavier.serra@upf.edu

ABSTRACT

In this paper, we present TIV.lib, an open-source library for the content-based tonal description of musical audio signals. Its main novelty relies on the perceptually-inspired Tonal Interval Vector space based on the Discrete Fourier transform, from which multiple instantaneous and global representations, descriptors and metrics are computed—e.g., harmonic change, dissonance, diatonicity, and musical key. The library is cross-platform, implemented in Python and the graphical programming language Pure Data, and can be used in both online and offline scenarios. Of note is its potential for enhanced Music Information Retrieval, where tonal descriptors sit at the core of numerous methods and applications.

1. INTRODUCTION

In Music Information Retrieval (MIR), several libraries for musical content-based audio analysis, such as Essentia [1], Librosa [2], and madmom [3] have been developed. These libraries have been widely adopted across academia and industry as they promote the fast prototyping of experimental methods and applications ranging from large-scale applications such as audio fingerprinting and music recommendation, to task-specific MIR analysis including chord recognition, structural segmentation, and beat tracking.

The tonal domain of content-based audio descriptors denotes all attributes related to the vertical (i.e., harmonic) and horizontal (i.e., melodic and voice-leading) combination of tones, as well as their higher-level governing principles, such as the concept of musical key. The earliest research in this domain was driven by the methods applied to symbolic representations of music, e.g., MIDI files. The jump from symbolic to musical audio domain raises significant problems and requires dedicated methods, as polyphonic audio-to-symbolic transcription remains a challenging task [4]. While the state of the art [5, 6] in polyphonic music transcription has advanced greatly due to the use of deep neural networks, it remains largely restricted to piano-only recordings.

One of the most prominent tonal audio descriptors is the chroma vector. This representation divides the energy of the spectrum of an audio signal in the 12 tones of the western chromatic scale across all octaves. This leads to a 12-element vector where each element corresponds to the energy of each pitch class. Throughout this work, this vector will be referred to as the pitch profile. Many algorithms for this representation have been proposed, including Pitch Class Profiles [7], Harmonic Pitch Class Profiles (HPCP) [8], the CRP chroma [9], and the NNLS chroma [10]. Stemming from this 12-element vector, many metrics and systems have been proposed, for key detection, chord recognition, cover song identification, mood recognition, and harmonic mixing. Yet, despite their fundamental role in many MIR tasks, tonal descriptors are not only less prominent in existing content-based audio libraries, in comparison with rhythmic or timbral descriptors [1], but also their perceptual basis is of limited scope [11].

In the context of the aforementioned limitations, we present the TIV.lib, a cross-platform library for Python and Pure Data, which automatically extracts multiple perceptually-aware tonal descriptions from polyphonic audio signals, without requiring any audio-to-symbolic transcription stage. It owes its conceptual basis to ongoing work within music theory on the DFT of pitch profiles, which has been extended to the audio domain [12]. The hierarchical nature of the Tonal Interval Vector (TIV) space allows the computation of instantaneous and global tonal descriptors including harmonic change, (intervallic) dissonance, diatonicity, chromaticity, and key, as well as the use of distance metrics to extrapolate different harmonic qualities across tonal hierarchies. Furthermore, it can enable the efficient retrieval of isolated qualities or those resulting from audio mixes in large annotated datasets as a simple nearest neighbour-search problem.

The remainder of this paper is organized as follows. Section 2 provides an overview of the ongoing work on pitch profiles DFT-based methods within music theory, followed by a description of the recently proposed TIV space, which extends the method to the audio domain. Section 3 provides a global perspective of the newly proposed TIV.lib architecture. Section 4 details the mathematical and musical interpretation of the description featured in TIV.lib and, finally, Section 5 discusses the scope of application scenarios of the library and Section 6 provides perspectives on future work.

2. RELATED WORK

2.1. Tonal pitch spaces

Within the research literature, numerous tonal pitch spaces and pitch distance metrics have been proposed [13, 14, 15, 16]. They aim to capture perceptual musical phenomena by geometrical and algebraic representations, which quantify and (visually) represent pitch proximity. These spaces process pitch as symbolic manifestations, thus capturing musical phenomena under very controlled conditions, with some of the most prominent spaces discarding the pitch height dimension by collapsing all octaves into a 12-tone pitch space.

Attempts to represent musical audio in the aforementioned spaces have been pursued [17, 18] by adopting an audio-to-symbolic transcription stage. Yet, polyphonic transcription from musical audio remains a challenging task which is prone to error.

Recently, Bernardes et al. [12] proposed a tonal pitch space which maps chroma vectors derived from audio signals driven into a perceptually inspired DFT space. It expands the aforementioned pitch spaces with strategies to process the timbral/spectral information from musical audio.

2.2. From the DFT of symbolic pitch distributions to the Tonal Interval Vector space

In music theory, the work proposed by Quinn [19] and followed by [20, 21, 22] on the Discrete Fourier Transform (DFT) of pitch profiles, has been shown to elicit many properties with music-theoretic value. Moreover, in [23] DFT-based pitch spaces were shown to capture human perceptual principles.

In the Fourier space, a 6-element complex vector, corresponding to the $1 \leq k \leq 6$ DFT coefficients, is typically adopted. The magnitude of the Fourier coefficients has been used to study the shape of pitch profiles, notably concerning the distribution of their interval content. This allows, for example, to quantify diatonic or chromatic structure (see Section 4 for a comprehensive review of the interpretations of the coefficients). The phase of the pitch profiles in the Fourier space reveals aspects of tonal music in terms of voice-leading [24], tonal regions modelling and relations [25], and the study of tuning systems [20]. In summary, the magnitude of the pitch profiles express harmonic quality and the phases harmonic proximity.

Recently, a perceptually-inspired equal-tempered, enharmonic, DFT-based TIV space [12] was proposed. One novelty introduced by this newly proposed space in relation to remaining Fourier spaces was the combined use of the six coefficients in a TIV, $T(k)$. Moreover, the perceptual basis of the space is guaranteed by weighting each coefficients by empirical ratings of dyad consonance, $w_a(k)$. $T(k)$ allows the representation of hierarchical or multi-level pitch due to the imposed L_1 norm, such that:

$$T(k) = w_a(k) \sum_{n=0}^{N-1} \bar{c}(n) e^{\frac{-j2\pi kn}{N}},$$

$$k \in \mathbb{Z} \quad \text{with} \quad \bar{c}(n) = \frac{c(n)}{\sum_{n=0}^{N-1} c(n)}$$

(1)

where N=12 is the dimension of the chroma vector, $c(n)$, and k is set to $1 \leq k \leq 6$ for $T(k)$ since the remaining coefficients are symmetric. The weights, $w_a(k) = \{3, 8, 11.5, 15, 14.5, 7.5\}$, adjust the contribution of each dimension k of the space to comply
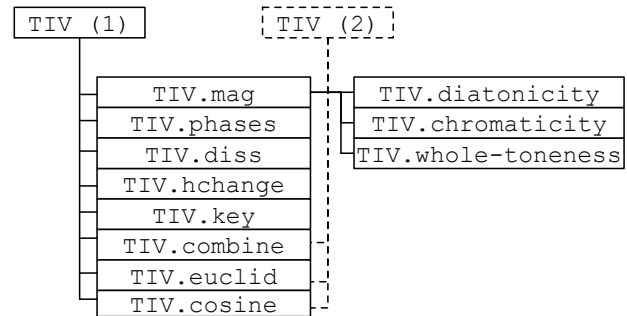

Figure 1: A graph of the dependencies of the feature extraction modules of TIV.lib. The algorithms connected to TIV(2) through a dashed line require two inputs for the feature calculation.

with empirical ratings of dyad consonance as summarised in [26]. $w_a(k)$ accounts for the harmonic structure of musical audio driven from an average spectrum of orchestral instruments [11].

3. TIV.LIB: IMPLEMENTATION

The TIV.lib includes several signal-processing functions or descriptors for characterising the tonal content of musical audio. This library is implemented in Python, using only Numpy and Scipy as dependencies and Pure Data, with both available to download at: http://bit.ly/2pBYhqZ. Illustrative analysis of musical audio examples for the descriptors are provided in the Python download link as a Jupyter Notebook. The Python implementation targets batch offline processing and the Pure Data implementation online processing.

As an input, the library takes 12-element chroma vectors, $c(n)$, from which TIVs, $T(k)$, are then computed. [1] Any input representation will have an effect on the space, as such we leave the choice of which chroma representation up to the user in order to best fit the problem at hand. Although the system is agnostic to the chosen chroma, we recommend the "cleanest" chroma representation, i.e., that which is closest to a symbolic representation, to be selected. The time scale of the TIV is dependent of the adopted window size during the chroma vector computation. For instantaneous TIVs, a single-window chroma vector can be used as input. For global TIVs, consecutive chroma vectors can be averaged across the time axis prior to the TIV computation.

In Figure 1 we present the architecture of TIV.lib. In this graph of dependencies we can see the algorithms that have been implemented and which classes they require for their calculation.

4. TIV.LIB: ALGORITHMS

This section details the functions included in the TIV.lib, focusing on their mathematical definition and musical interpretation.

TIV is a 6-element complex vector, which transforms chroma into an interval vector space by applying Eq. 1, an L_1-norm weighted DFT. The resulting space dimensions combine intervallic information in the coefficients' magnitude and the tonal region (i.e., musical key area) it occupies in the coefficients' phase. The

---

[1] A tutorial example on the extraction of HPCP representations from audio is provided in the library package, both using Essentia and Librosa.

mapping between chroma and the TIV retains the bijective property of the DFT and allows the representation of any variable-density pitch profile in the chroma space as a unique location in the TIV space.

TIV.mag is a 6-element (real) vector that reports the magnitude of the TIV elements $1 \leq k \leq 6$, such that:

$$mag(k) = ||T(k)|| \qquad (2)$$

It provides a characterisation of the harmonic quality of a pitch profile, namely its intervallic content, distilling the same information as the pitch-class interval vector [27, 28]. Mathematically, it is well-understood that a large magnitude in $T(k)$ coefficients indicates how evenly the pitch profile can be divided by N/k. Musically, the work on the DFT of pitch profiles [12, 19] emphasizes the association between the magnitude of Fourier coefficients and tonal qualities: $||T(1)|| \leftrightarrow chromaticity$, $||T(2)|| \leftrightarrow dyadicity$, $||T(3)|| \leftrightarrow triadicity$, $||T(4)|| \leftrightarrow diminished\,quality$, $||T(5)|| \leftrightarrow diatonicity$, $||T(6)|| \leftrightarrow whole - toneness$. Please refer to [20, 21] for a comprehensive discussion on the interpretation of the DFT coefficients.[2] One distinct property of the TIV.mag vector is its invariance under transposition or inversion [20]. For example, all major triads or harmonic minor scales share the same Fourier magnitude, hence the same TIV.mag vector [3].

TIV.phases is a 6-element (real) vector that reports the phases (or direction) of the TIV coefficients $1 < k < 6$, such that:

$$phases(k) = \angle T(k) \qquad (3)$$

It indicates which of the transpositions of a pitch profile quality is under analysis [29], as transposition of a pitch profile by p semitones, i.e., circular rotations of the chroma, $c(n)$, rotates the $T(k)$ by $\varphi(p) = \frac{-2\pi kp}{N}$. TIV phases are also associated with regional (or key) areas, whose diatonic set is organised as clusters in the TIV space [12, 25].

TIV.combine computes the resulting TIV from mixing (or summing) multiple TIVs representing different musical audio signals. Due to the properties of the DFT space, this operation can be efficiently computed as a linear combination of any number of TIVs, $T(k)$. Given TIVs $T_1(k)$ and $T_2(k)$, their linear combination, weighted by their respective energy, a_1 and a_2, is given by:

$$T_{1+2}(k) = \frac{T_1(k) \cdot a_1 + T_2(k) \cdot a_2}{a_1 + a_2} \qquad (4)$$

a_1 and a_2 are retrieved from the discarded DC components $T_1(0)$ and $T_2(0)$.

TIV.chromaticity reports the level of concentration of a sonority in a specific location of the chromatic pitch circle as a value within the [0,1] range, computed as the magnitude of the $T(1)$ normalized to unity: $\frac{||T(1)||}{w_a(1)}$. This value is close to 0 for sounds exhibiting energy in evenly-spaced pitch classes (such as typically tonal chords and scales) and close to 1 for chromatic pitch aggregates.

TIV.diatonicity reports the level of concentration of a sonority within the circle of fifths as a value within the [0,1] range.

---

[2]We note for each of these single Fourier coefficient quantities that the effects of the weights can be factored out.

[3]Note that the phases, discarded here, will differ. As such, the uniqueness property of the TIV is maintained as it combines both magnitude and phase information.
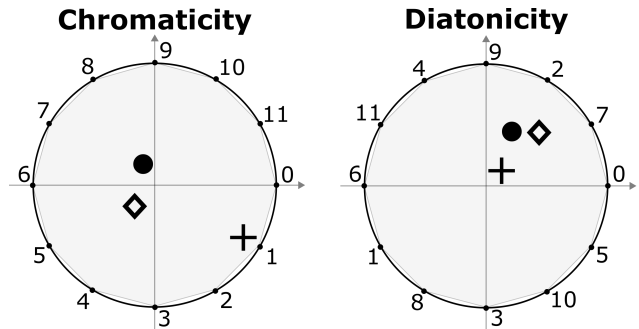

Figure 2: Two DFT coefficients interpreted as chromaticity and diatonicity. Three TIV are plotted for comparison: C major chord {0,4,7} (◇), 3-note chromatic cluster {0,1,2} (+), and C major scale {0,2,4,5,7,9,11} (●). The grey shaded areas indicate the space the TIVs can occupy.

The larger the magnitude of the $T(5)$ normalized to unity, $\frac{||T(5)||}{w_a(5)}$, the higher the level of diatonicity.

TIV.whole-toneness reports the proximity to one of the two existing whole-tone collection within the 12-tone equal temperament tuning. The level of whole-toneness is reported within the [0,1] range resulting from the magnitude of the $T(6)$ normalized to unity, such that: $\frac{||T(6)||}{w_a(6)}$.

Fig. 2 shows the DFT coefficients from which we extract chromaticity and diatonicity descriptions as the magnitude of $T(1)$ and $T(5)$, respectively. We plot pitch profiles that aim to illustrate the behaviour of each coefficient in eliciting the chromatic and diatonic character of the C major chord and C major scale as well as chromatic 3-tone cluster by inspecting their magnitude. Note that the magnitude of both the C major chord and C major scale, two prototypical diatonic pitch profiles, clearly have greater magnitude in the diatonic $T(5)$ coefficient in comparison with the three-note cluster, a prototypical chromatic profile. Conversely, in the chromatic $T(1)$ coefficient, the magnitudes of the above pitch profiles show the expected opposite behaviour, thus mapping the three-note cluster further from the centre.

TIV.euclid and **TIV.cosine** compute the Euclidean, E, and cosine, C, distance between two given TIVs, $T_1(k)$ and $T_2(k)$, using Eqs. 5 and 6, respectively.

$$E\{T_1, T_2\} = \sqrt{||T_1 - T_2||^2} \qquad (5)$$

$$C\{T_1, T_2\} = \frac{T_1 \cdot T_2}{||T_1||||T_2||} \qquad (6)$$

The cosine distance (i.e., the angular distance) between TIVs can be used as an indicator of how well pitch profiles "fit" or mix together. For example, it quantifies the degree of tonal proximity of TIV mixtures, or informs which translation or transposition of a TIV best aligns with a given key. Conversely, Euclidean distances between TIVs relate mostly to melodic (or horizontal) distance. It captures the neighbouring relations observed in the *Tonnetz*, where smaller distances agree with parsimonious movements between pitch profiles. Please refer to [22, 24] for a comprehensive discussion on this topic.

TIV.hchange computes a harmonic change detection function across the temporal dimension of an audio signal. Peaks in this function indicate transitions between regions that are harmonically

stable. We compute a harmonic change measure, λ, for an audio frame m as the Euclidean distance between frames $m + 1$ and $m - 1$ (Eq. 7), an approach inspired by Harte et al. [30], which can be understood as adopting three coefficients out of the $1 \leq k \leq 6$ of the TIV, $T(k)$, i.e., those corresponding to the circle of fifths, the circle of minor thirds, and the circle of major thirds.

$$\lambda_m = \sqrt{||T_{m-1} - T_{m+1}||^2} \qquad (7)$$

TIV.diss provides an indicator of (interval content) dissonance, as the normalized TIV magnitude subtracted from unity, $1 - \frac{|T(k)|}{|w_a(k)|}$. This perceptually-inspired indicator stems from the weighted magnitude of the TIV coefficients, which rank the intervals $1 \leq k \leq 6$ to match empirical ratings of dissonance within the Western tonal music context [11, 12].

TIV.key infers the key from an audio signal as a pitch class (tonic) and a mode (major or minor). It is computed as the Euclidean distance from the 24 major and minor key TIVs, $T_r^{p\star}(k)$, defined as the shifts (i.e. rotation) of the 12 major and 12 minor profiles, p, by Temperley [31] or Sha'ath [32], such that:

$$R_{min} = \text{argmin}_r \sqrt{||T \cdot \alpha - T_r^{p\star}||^2} \qquad (8)$$

where $T_r^{p\star}$ are 24 major and minor key profiles TIVs, p. When $r \leq 11$, we adopt the major profile and when $r \geq 12$, the minor profile. α is a bias introduced to balance the distance between major and minor keys. Optimal values of $\alpha = 0.2$ and $\alpha = 0.55$ have been proposed in [33] for the Temperley [31] and Shat'ath [32] key profiles, respectively. The output is an integer, R_{min}, ranging between $0 - 11$ for major keys and $12 - 23$ for minor keys, where 0 corresponds to C major, 1 to C# major, and so on through to 23 being B minor.

5. APPLICATIONS AND PERSPECTIVES

Following the emerging body of music theory literature on the DFT of pitch profiles and the continuous work of the TIV space, we implemented the perceptually-inspired `TIV.lib` in Python and Pure Data. The former aims at batch offline processing and the latter mostly at online or real-time processing, but allowing offline computations as well.

Figure 3 shows an example usage of the `TIV.lib` functions for computing the diatonicity, chromaticity and whole-toneness harmonic qualities in Pure Data.

In order to achieve the same result in Python, the following code can be executed:

```
import TIVlib as tiv

ex_tiv = tiv.TIV.from_pcp(example_chroma)
ex_wholetoneness = ex_tiv.wholetoneness()
ex_diatonicity = ex_tiv.diatonicity()
ex_chromacity = ex_tiv.chromaticity()
```

We now provide an example usage of this library for extracting tonal features of a musical piece. We run this code for an excerpt of the Kraftwerk song "Spacelab," to demonstrate how this library can provide useful information related to its diatonicity and whole-toneness. As can be seen from the Chromagram in Figure 4, this song starts in the whole tone scale [F# G# A#C D E], and then moves, at 33 s, to a diatonic set [C D Eb

F G Ab Bb C]. In Figure 5 we show this by plotting the evolution of the `TIV.diatonicity`, `TIV.wholetoneness` and `TIV.chromaticity` outputs for this music.
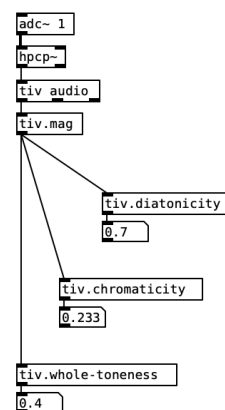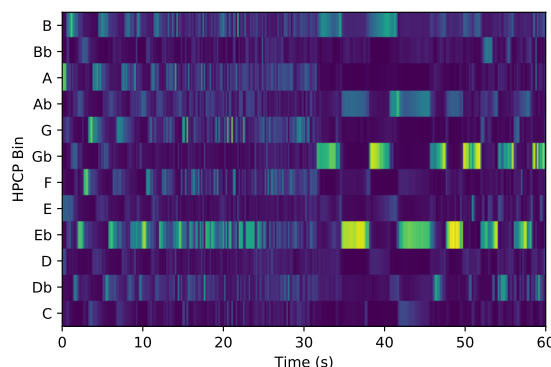

Figure 3: Pure Data patch for online computation of diatonicity, chromaticity and whole-toneness harmonic qualities of a live input audio signal, using the `TIV.lib`.

Figure 4: Chromagram of the first minute of Kraftwerk's "Spacelab."

Several functions of the `TIV.lib` result from ongoing research and have been evaluated in previous literature [11, 12, 33, 34], where the possibility of the TIV space to geometrically and algebraically capture existing spaces of perceptual and music theoretical value, such as Euler and Krumhansl [12], were shown. In particular, we highlight our previous work on key recognition [33] and harmonic mixing [11, 35], where TIV-based approaches outperformed more traditionally used harmonic features. In addition, the use of TIVs can also extend content-based audio processing by providing a vector space where distances and metrics (e.g., dissonance and harmonic proximity) among multi-level pitch, chords, and keys, capture perceptual aspects of musical phenomena. Examples of creative possibilities of the TIV space have also been shown in Musikverb [36], where it was used for developing a novel
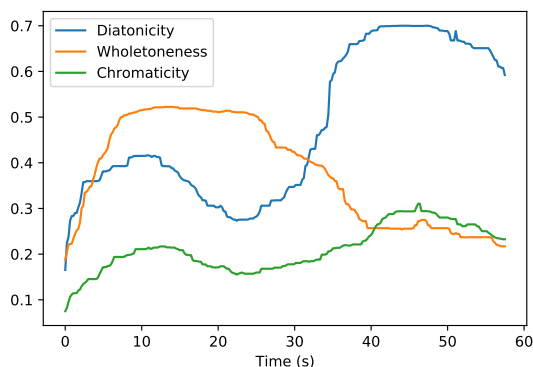
Figure 5: Output of the diatonicity, chromaticity and whole-toneness harmonic qualities the first minute of Kraftwerk's "Spacelab", using the `TIV.lib`.

type of harmonically-adaptive reverb effect.

We strongly believe that the properties of the TIV space can be further explored in content-based audio processing. For example, the possibility to isolate the harmonic quality in `TIV.mag` as a pitch-invariant audio representation can be relevant for several MIR tasks that rely on multiple transposed versions of a given musical pattern, such as in query-by-humming, and cover song detection. Moreover, the possibility to compute TIV mixes as a computationally efficient linear combination allows for the fast retrieval of musical audio from large datasets (e.g., Freesound [37]), as a simple nearest-neighbour search problem. Finally, the newly proposed indicators of tonal quality such as `TIV.chromaticity`, `TIV.diatonicity`, `TIV.wholte-toneness`, and `TIV.diss` not only extend musical theoretical methodologies to content-based processing from audio performance data, but can also promote a greater understanding of tonal content in MIR tasks.

By providing streamlined access to a set of music theoretic properties which are non-trivial to obtain from commonly used time-frequency representations in MIR such as the STFT (or even from chroma-like representations directly), we believe the `TIV.lib` can lay the foundation for a kind of "enhanced" MIR in tasks such as chord recognition and key estimation which can directly leverage the complementary contextual information contained within the `TIV.lib` descriptors.

6. CONCLUSIONS

In this paper we have introduced the open-source tool, `TIV.lib`, as a means to drive the uptake and usage of the Tonal Interval Space both in offline music signal analysis via the python implementation, as well as in online contexts using Pure Data. While we hope to see a growth of applications which benefit from access to music theoretic harmonic features provided by `TIV.lib` our own future work will focus in two principal areas: i) investigating the processing stages which directly precede the calculation of the TIV; and ii) in the application of the TIV across large datasets. More specifically, we seek to study the impact of different methods for calculating the requisite chroma vectors (e.g., HPCP [8],

NNLS [10], or timbre-invariant chroma [9]) in the TIV space, as pursued in [33, 34] within the scope of audio key detection. Furthermore, we will study an optimal strategy to define the weights, $w_a(k)$, for particular audio sources and to implement the descriptors in a large online musical database supported by content-based analysis, as a strategy to study the descriptors under a large-scale environment for musical retrieval and creation. Finally, we intend to add this library to existing musical audio analysis libraries such as Essentia and Librosa.

8. 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 open-source library for sound and music analysis," in *Proc. of the ACM International Conference on Multimedia*, 2013, pp. 855–858.

[2] B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "Librosa: Audio and music signal analysis in python," in *Proceedings of the Python in Science Conference*, 2015, pp. 109–114.

[3] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer, "madmom: a new Python Audio and Music Signal Processing Library," in *Proceedings of the 24th ACM International Conference on Multimedia*, Amsterdam, The Netherlands, 10 2016, pp. 1174–1178.

[4] E. Benetos, S. Dixon, Z. Duan, and S. Ewert, "Automatic music transcription: An overview," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 20–30, 2018.

[5] C. Hawthorne, E. Elsen, J. Song, A. Roberts, I. Simon, C. Raffel, J. Engel, S. Oore, and D. Eck, "Onsets and frames: Dual-objective piano transcription," in *Proc. of the 19th Intl. Society for Music Information Retrieval Conf. (ISMIR)*, 2018, pp. 50–57.

[6] A. Ycart and E. Benetos, "Polyphonic music sequence transduction with meter-constrained lstm networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 386–390.

[7] T. Fujishima, "Realtime chord recognition of musical sound: a system using common lisp music," in *Proc. of the International Computer Music Conference*, 1999.

[8] E. Gómez, "Tonal description of polyphonic audio for music content processing," *INFORMS Journal on Computing*, vol. 18, no. 3, pp. 294–304, 2006.

[9] M. Müller and S. Ewert, "Towards timbre-invariant audio features for harmony-based music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 649–662, 2010.

[10] M. Mauch and S. Dixon, "Approximate note transcription for the improved identification of difficult chords," in *Proc. of the International Society for Music Information Retrieval Conference*, 2010, pp. 135–140.

[11] G. Bernardes, M. E. P. Davies, and C. Guedes, "A hierarchical harmonic mixing method," in *Music Technology with Swing*, M. Aramaki, M. E. P. Davies, R. Kronland-Martinet, and S. Ystad, Eds. 2018, pp. 151–170, Springer International Publishing.

[12] G. Bernardes, D. Cocharro, M. Caetano, C. Guedes, and M. E. P. Davies, "A multi-level tonal interval space for modelling pitch relatedness and musical consonance," *Journal of New Music Research*, vol. 45, no. 4, pp. 281–294, 2016.

[13] R. Shepard, "The analysis of proximities: multidimensional scaling with an unknown distance function. i.," *Psychometrika*, vol. 27, no. 2, pp. 125–140, 1962.

[14] F. Lerdahl, *Tonal pitch space*, Oxford University Press, 2004.

[15] D. Tymoczko, *A geometry of music: Harmony and counterpoint in the extended common practice*, Oxford University Press, 2010.

[16] E. Chew, "Out of the grid and into the spiral: Geometric interpretations of and comparisons with the spiral-array model," *Computing in musicology*, vol. 15, pp. 51–72, 2007.

[17] C. Chuan and E. Chew, "Polyphonic audio key finding using the spiral array ceg algorithm," in *IEEE International Conference on Multimedia and Expo*, 2005, pp. 21–24.

[18] B. De Haas, R. Veltkamp, and F. Wiering, "Tonal pitch step distance: a similarity measure for chord progressions.," in *Proc. of the International Society for Music Information Retrieval Conference*, 2008, pp. 51–56.

[19] I. Quinn, "General equal-tempered harmony: parts 2 and 3," *Perspectives of New Music*, pp. 4–63, 2007.

[20] E. Amiot, *Music through Fourier space*, Springer, 2016.

[21] J. Yust, "Stylistic information in pitch-class distributions," *Journal of New Music Research*, vol. 48, no. 3, pp. 217–231, 2019.

[22] D. Tymoczko and J. Yust, "Fourier phase and pitch-class sum," in *International Conference on Mathematics and Computation in Music*, 2019, pp. 46–58.

[23] M. R. W. Dawson, A. Perez, and S. Sylvestre, "Artificial neural networks solve musical problems with Fourier phase spaces," *Scientific Reports*, vol. 10, no. 1, pp. 7151, Apr 2020.

[24] D. Tymoczko, "Set-class similarity, voice leading, and the fourier transform," *Journal of Music Theory*, vol. 52, no. 2, pp. 251–272, 2008.

[25] J. Yust, "Probing questions about keys: Tonal distributions through the DFT," in *International Conference on Mathematics and Computation in Music*, 2017, pp. 167–179.

[26] David Huron, "Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance," *Music Perception*, vol. 11, no. 3, pp. 289–305, 1994.

[27] A. Forte, "A theory of set-complexes for music," *Journal of Music Theory*, vol. 8, no. 2, pp. 136–183, 1964.

[28] A. Forte, *The structure of atonal music*, vol. 304, Yale University Press, 1973.

[29] J. Hoffman, "On pitch-class set cartography: Relations between voice-leading spaces and fourier spaces," *Journal of Music Theory*, vol. 52, no. 2, pp. 219–249, 2008.

[30] C. Harte, M. Sandler, and M. Gasser, "Detecting harmonic change in musical audio," in *Proc. of the ACM Workshop on Audio and Music Computing Multimedia*, 2006, pp. 21–26.

[31] D. Temperley, "What's key for key? the Krumhansl-Schmuckler key-finding algorithm reconsidered," *Music Perception: An Interdisciplinary Journal*, vol. 17, no. 1, pp. 65–100, 1999.

[32] I. Sha'ath, "Estimation of key in digital music recordings," M.S. thesis, Birkbeck College, University of London, 2011.

[33] G. Bernardes, M. E. P. Davies, and C. Guedes, "Automatic musical key estimation with adaptive mode bias," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 316–320.

[34] G. Bernardes, D. Cocharro, C. Guedes, and M. E. P. Davies, "Harmony generation driven by a perceptually motivated tonal interval space," *ACM Computers in Entertainment*, vol. 14, no. 2, pp. 6, 2016.

[35] C. Maçãs, A. Rodrigues, G. Bernardes, and P. Machado, "Mixmash: An assistive tool for music mashup creation from large music collections," *International Journal of Art, Culture and Design Technologies*, vol. 8, no. 2, pp. 20–40, 2019.

[36] J. Pereira, G. Bernardes, and R. Penha, "Musikverb: A harmonically adaptive audio reverberation," in *Proc. of the 21st Int. Conference on Digital Audio Effects (DAFx-18), Aveiro, Portugal*, Sep 2018.

[37] F. Font, G. Roma, and X. Serra, "Freesound technical demo," in *Proc. of the ACM International Conference on Multimedia*, 2013, pp. 411–412.

[38] G. Bernardes, M. E. P. Davies, and C. Guedes, "Automatic musical key estimation with adaptive mode bias," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 316–320.

[39] R. Gebhardt, M. E. P. Davies, and B. Seeber, "Harmonic mixing based on roughness and pitch commonality," in *Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway*, Nov 2015, pp. 185–192.