AUDIO KEY FINDING IN THE TONAL INTERVAL SPACE

Gilberto Bernardes
INESC TEC
gba@inesctec.pt

Matthew Davies
INESC TEC
mdavies@inesctec.pt

ABSTRACT

In this submission we present a method for automatically estimating the key of a musical piece in a digital audio format. The proposed method relies on a perceptually motivated Tonal Interval Space, which represents multi-level tonal pitch as 12-dimensional Tonal Interval Vectors. We estimate the key of a musical piece by comparing the degree to which beat-synchronous Tonal Interval Vectors averaged across a musical piece correlate to Temperley's [9] 24 major and minor key profiles.

1. INTRODUCTION

Our key-finding method is an extended version of a previously reported algorithm in [1]. It is based on the Tonal Interval Space [2], an extended type of pitch space in which music theory principles and human perception of pitches, chords and keys–represented in the space by Tonal Interval Vectors (TIVs)–are expressed as distances.

Based on the assumption that a key-defining element is the use of its diatonic pitch set, two main attributes of the Tonal Interval Space promote the estimation of a musical key: i) the set of diatonic pitch classes in a given key occupies a compact neighborhood around its key, and ii) the 24 major and minor keys are sparsely represented in the space. Furthermore, the spatial proximity of each key to its dominant, subdominant, and relative keys in the Tonal Interval Space corresponds to our expectation of the proximity between the 24 major and minor keys, thus favoring 'close' estimates whenever the algorithm fails to predict the correct key.

To estimate the key of a musical piece in the Tonal Interval Space we compare the degree to which an averaged vector resulting from beat-synchronous Tonal Interval Vectors across a musical piece correlate to a collection of 24 TIVs derived from the major and minor chroma key profiles proposed by Temperley's [9], hereafter refer to as key TIVs. The novelty of our current method in relation to the one reported in [1] is the introduction of a frame selection and spatial adjustment strategies after the calculation of beat-synchronous TIVs based on their energy and level of consonance. This processing stage aims at discarding beats with silent or noise floor content and pulling beat-synchronous TIVs towards the center of the space to reinforce and balance the estimation of key modes, as their key-defining diatonic pitch set encompass different levels of consonance. The overall architecture of our key-finding method is shown in Fig. 1.

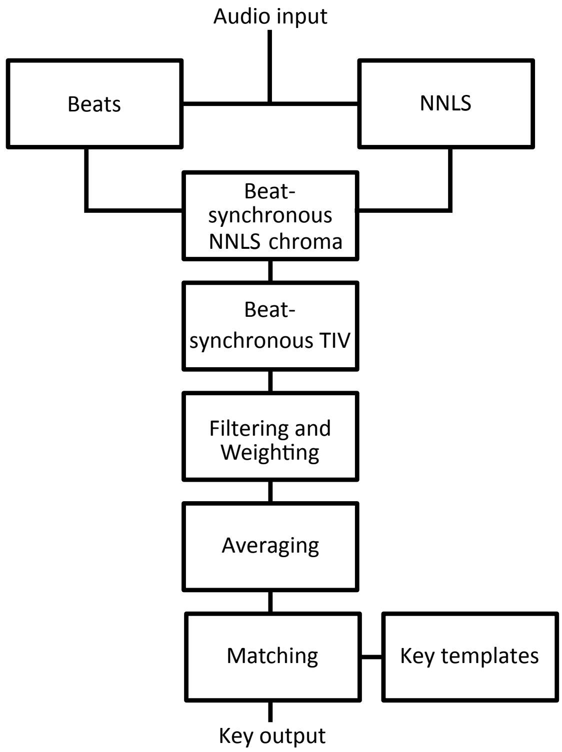

Figure 1. Architecture of the key-finding method.

The remainder of this paper is organized as follows. Section 2 details the Tonal Interval Space and the generation of Tonal Interval Vectors (TIV) from audio. Section 3 details the core algorithms of our key-finding method, including a processing frame selection and spatial adjustment.

2. TONAL PITCH REPRESENTATIONS IN THE TONAL INTERVAL SPACE

To represent the most salient tonal pitch levels in the Tonal Interval Space [2] from an audio signal, we first aggregate the energy of each pitch class in a 12-dimensional chroma vector, $c(n)$, and compute a 12-dimensional Tonal Interval Vector, $T(k)$ as its L_1 normalized Discrete Fourier Transform (DFT), such that:

$$T(k) = w(k) \sum_{n=0}^{N-1} \bar{c}(n) e^{\frac{-j2\pi kn}{N}}, \ k \in \mathbb{Z}$$
$$\text{with } \bar{c}(n) = \frac{\bar{c}(n)}{\sum_{n=0}^{N-1} c(n)} \tag{1}$$

where $N = 12$ is the dimension of the chroma vector and $w(k) = \{2, 11, 17, 16, 19, 7\}$ are weights derived from empirical consonance ratings of dyads used to adjust the contribution of each dimension k of the space. We set k to $1 \leq k \leq 6$ for $T(k)$, since the remaining coefficients are symmetric. $T(k)$ uses $\bar{c}(n)$ which is $c(n)$ normalized by the DC component $T(0) = \sum_{n=0}^{N-1} c(n)$ to allow the representation and comparison of different hierarchical levels of tonal pitch [2].

2.1 Audio Beat-Synchronous Tonal Interval Vectors

We adopt the beat as the temporal resolution for representing the harmonic content in the Tonal Interval Space for the task of estimating the key from an input audio signal.

To compute beat-synchronous TIVs, $T_b(k)$, we first extract chroma representations on a regular and short-time interval basis and then calculate the median value per chroma bin for all frames within each beat to generate beat-synchronous chroma vectors. Then, we apply Eq. 1 to compute beat-synchronous TIVs.

To compute short-time interval basis chroma vectors we use the NNLS chroma [8] plugin within Sonic Annotator [3] with default parameters. The NNLS algorithm performs an approximate note transcription, and typically provides a sparse representation of the input signal in the chroma domain, closely matching a symbolic input representation. To compute beat locations we use the QM-VAMP bar and beat tracking plugin [5] within Sonic Annotator [3].

2.2 Measuring Consonance

We explicitly designed the Tonal Interval Space as a distorted DFT space where each component (or interpreted musical interval) is weighted according empirical ratings of dyads consonance, $w(k)$, as a strategy for expanding the pitch based representation of the Tonnetz with metrics of tonal pitch consonance.

Beat-synchronous TIVs including single notes (at the edge of the space and furthest from the center) are considered the most consonant in the Tonal Interval Space. Beat-synchronous TIVs of chroma vectors, $c(n)$, whose 12 pitch classes have the same energy (in the center of space) are considered the most dissonant. Within this range, the normalized level of consonance, C_b, of any beat-synchronous TIV can be measured as the norm of $T_b(k)$ such that:

$$C_b = \frac{1}{\beta} \parallel T(k) \parallel = \frac{1}{\beta} \sqrt{T(k) \cdot T(k)} = \frac{1}{\beta} \sqrt{\sum_{k=1}^{M} |T(k)|^2} \tag{2}$$

where $\beta = 32.8633$ is a scaling factor used to normalize the results to unity and equals to to the level of consonance

Pitch class	Major	Minor
0	.748	.712
1	.060	0.84
2	.488	.474
3	.082	.618
4	.670	.049
5	.460	.460
6	.096	.105
7	.715	.747
8	.104	.404
9	.366	.067
10	.057	.133
11	.400	.330

Table 1. Temperley's [9] chroma vector profiles for the major and minor key modes.

of a single pitch class (e.g. for the pitch class C, whose $c(n) = \{1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$).

2.3 Key TIV profiles

In [1], we compared four sets of TIVs as key-defining profiles for the task of estimating the key from both symbolic music representations and musical audio in the Tonal Interval Space. These sets of key TIVs derive from the following chroma vector key profiles: i) binary activation of the diatonic pitch set of each key; ii) Krumhansl and Kessler (K-K) [7] profiles derived from the 'probe tone' method; iii) Temperley's [9] adjustments to the K-K profiles using music theory principles; and iv) Chew's [4] chord-based profiles used in her Center of Effect Generator algorithm.

We tested the efficacy of these four set of key TIVs in our key-finding method, and concluded that Temperley's [9] profiles (shown in Table 1) provide the best results for the following three musical datasets: Bach's 24 fugue subjects from Book I of the Well Tempered Clavier, the Kostka-Payne dataset [9], and a large collection of pop and rock Beatles songs assembled by Harte [6].

Table 1 shows the Temperley's C major and C minor (chroma) key profiles, p, from which we compute the key TIVs, $T^p_{temp}(k)$, using Equation (1). The key profiles of the remaining keys are obtained by rotating the profiles by 12 semitones. The resulting key TIVs are used to define the location of each of the 24 major and minor keys in the Tonal Interval Space.

3. AUDIO KEY FINDING

Based on the assumption that a key-indicating element is the use of its diatonic pitch set, we estimate a key of a musical passage in the Tonal Interval Space by finding the nearest neighbor in the high dimensional Euclidean space of a query TIV for a given piece to a database of 24 major and minor key TIVs.

The query TIV for a given piece is an average vector of all beat-synchronous TIVs information across each piece, $\overline{T(k)}$. In Eq. 3 each beat-synchrounous TIV, $T_b(k)$, is multiplied by its normalized level of consonance C_b to regulate a spatial displacement towards the center of the space. Furthermore, all vectors with a DC component $T(0) < 0.1$ are considered as noise and discarded from the average computation.

$$\overline{T(k)} = \frac{1}{B}\sum_{b=1}^{B} T_b(k) \cdot C_b, \quad \text{if} \quad T(0) > 0.1 \qquad (3)$$

where B are the total number of averaged beat-synchronous TIVs whose DC component, $T(0)$, is above 0.1.

To rank the collection of 24 major and minor keys, the system computes the Euclidean distance between the averaged TIV, $\overline{T(k)}$, and all key TIVs. The best key estimate, M is the one which minimizes:

$$M = \text{argmin}_p \sqrt{\sum_{k=1}^{6}(\left|\overline{T(k)} - T^p_{temp}(k)\right|)} \qquad (4)$$

The system outputs the key estimate by a number ranging between 0-11 for major keys and 12-23 for minor keys, which we convert to a text format reporting the tonic and mode of the key.

4. ACKNOWLEDGMENTS

This research was supported by the Project "NORTE-01-0145-FEDER-000020", financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF) and by the Portuguese Foundation for Science and Technology under the post-doctoral grants SFRH/BPD/109457/2015 and SFRH/BPD/88722/2012.

5. REFERENCES

[1] G. Bernardes, D. Cocharro, C. Guedes, and M. Davies: "Harmony Generation Driven by a Perceptually Motivated Tonal Interval Space," *ACM Computers in Entertainment*, in press.

[2] G. Bernardes, D. Cocharro, M. Caetano, C. Guedes, and M. Davies: "A multi-level tonal interval space for modelling pitch relatedness and musical consonance," *Journal of New Music Research*, pp. 1-14.

[3] C. Cannam, M. Jewell, C. Rhodes, M. Sandler, and M. dInverno: "Linked Data and You: Bringing Music Research Software into the Semantic Web," *Journal of New Music Research*, *39*(4), pp. 313-325, 2010.

[4] E. Chew: *Towards a Mathematical Model of Tonality,* Doctoral dissertation, MIT, 2000.

[5] M. Davies, M. Plumbley, and D. Eck: "Towards a Musical Beat Emphasis Function," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* pp. 61-64, 2009.

[6] C. Harte: "Towards Automatic Extraction of Harmony Information from Music Signals," Doctoral dissertation, Queen Mary, University of London, 2010.

[7] C. Krumhansl and M. Schmuckler: "Key-finding in Music: An Algorithm Based on Pattern Matching to Tonal Hierarchies," *Mathematical Psychology Meeting*, Cambridge, MA, USA, 1986.

[8] M. Mauch and S. Dixon: "Approximate Note Transcription for the Improved Identification of Difficult Chords," *Proceedings of the 11th International Society Music Information Retrieval Conference*, pp. 135-140, 2010.

[9] D. Temperley: "Whats Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered," *Music Perception: An Interdisciplinary Journal*, *17*(1),pp. 65-100, 1999.