

Two-Dimensional Visual Inspection of Pitch-Space, Many Time-Scales and Tonal Uncertainty Over Time

Agustín Martorell and Emilia Gómez

Music Technology Group, Universitat Pompeu Fabra
{agustin.martorell, emilia.gomez}@upf.edu

Abstract. This work explores the representational limitations of toroidal pitch-spaces, when multiple temporal resolutions, tone center ambiguity, and the time dimension are considered for visualization of music pieces. The algorithm estimates key from chroma features, over time at many time-scales, using the key-profile correlation method. All these estimations are projected as tonal centroids within Krumhansl and Kessler’s toroidal space of inter-key distances. These centroids, belonging to a toroidal surface, are then mapped to colours by 3-dimensional geometric inscription of the whole pitch-space in the CIELAB colourspace. This mapping provides a visual correlate of pitch-space’s double circularity, approximates perceptual uniformity of colours throughout near regions, and allows for representing key ambiguity. We adapt Sapp’s keyscales to summarize tonal centroids in pitch-space at many time-scales over time, in a two-dimensional coloured image. Keyscales are linked with higher-dimensional tonal representations in a user interface, in order to combine their informative benefits for interactive analysis. By visualizing some specific music examples, we question the potential of continuous toroidal pitch-spaces in supporting long term analytical conclusions and tonal ambiguity description, when assisted by time vs. time-scale representations.

Keywords: pitch-space; toroidal; multi-scale; ambiguity; visualization

1 Introduction

Pitch-spaces, as theoretical formalizations of tonal concepts and relationships, facilitate mathematical and visual analysis of tonality from specific music examples. Introducing the time dimension into such analyses often requires understanding pitch-spaces as fixed structures, in which different musical *states* are activated at different spatial locations. This leads to a movement-in-space metaphor to model processes, since many of these spaces are fully informative just for *static* situations.

In this work,¹ we address the representational limitations of toroidal pitch-spaces when analysis over time, at multiple temporal resolutions, and ambiguity of description are jointly pursued. We combine spatial models of tonality and summarization techniques to fit multi-scale representations over time, and we propose a computer-aided interactive framework to assist tonal interpretations of long segments of music.

2 Background

2.1 Tonal Cognition and Toroidal Spaces

Krumhansl and Kessler’s (henceforth, K-K) psychological experiments lead to several toroidal pitch-space organizations [1]. One of such models, proposed as a space of inter-key distances, is derived from correlation between ring-shifted key profiles. Multidimensional scaling (MDS) of these correlations yields a reasonable four-dimensional solution of Euclidian distances between any pair of tone centers. In this 4-D space, two circular structures become evident: the double circle of fifths appears in the first two dimensions, while the last two dimensions reveal a circular arrangement of the relative and parallel relationships (Fig. 1(a)).

Given the topological homeomorphism between a ring torus and the cartesian product of two circles, solution points can be thought as belonging to a 3-D toroidal surface (Fig. 1(b)). This surface can be parameterized in two dimensions by the angular information, as long as double circularity is provided by *glueing* left-right and top-bottom borders (Fig. 1(c)).

Although the model of inter-key distances is only accurate as a 4-D space, representations in 3-D and 2-D are convenient for visualization and geometric manipulation. Similar toroidal distributions can be derived from Schoenberg’s *charts of regions* [2], as well as from certain boundary conditions of Chew’s *spiral array* [3]. Explicit toroidal spaces have been proposed in [4] and [5], among others.

2.2 Multidimensional Unfolding and Tonal Centroids

Being K-K’s torus a space of inter-key distances, it can be used to represent the induced key of any input *pitch-class profile* by multidimensional unfolding [6]. The method discussed by Krumhansl and Toivainen [7] finds, for a given 12-dimensional input vector, a point in the solution space (Fig. 1(c)) that optimizes distances from all key profiles. Although problematic for quantitative interpretation, this method helps representation of tonal ambiguity, since these *centroids* can be proportionally close to several keys.

¹ Given the relevance of colour, readers are encouraged to consult the electronic version. Figures are available at <http://www.dtic.upf.edu/~amartorell/publications/>

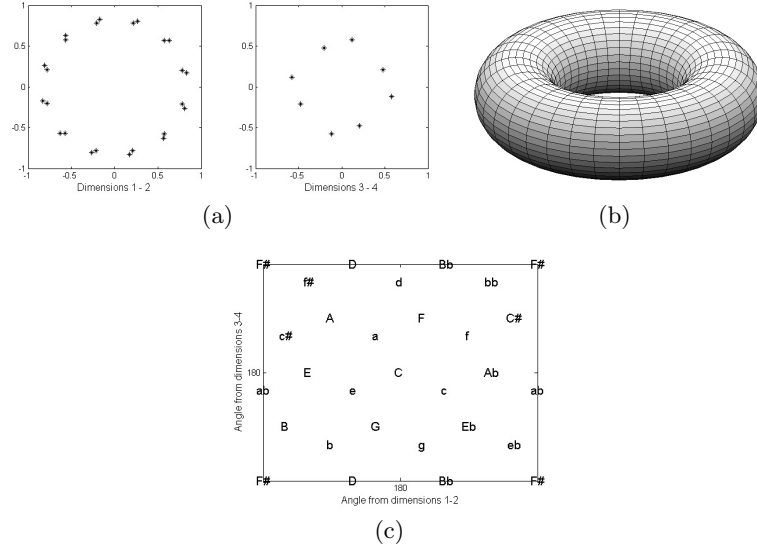


Fig. 1. a) K-K's 4-D inter-key space. b) 3-D toroidal surface, homeomorphic to both circles. c) Angle-based 2-D unfolding of K-K's space.

2.3 The SOM Approach

Centroids in K-K's toroidal surface are not good summaries in some cases, in which proper solutions might exist in the 4-D solution, but fall *outside* the 3-D projection. For instance, centroids from symmetric inputs² cannot remain in the torus' surface while being proportionally close to all their closest candidates. This is a consequence of the asymmetrical hierarchies of K-K's (major and minor) key profiles, which characterize the solution space.

In [7], a toroidal self-organized map (SOM) is trained with K-K's key profiles, and the complete space is activated according to its similarity with the input vector. The higher-dimensionality of SOM allows for representing and discriminating ambiguous cases, and provides clear information about key strength.

2.4 Representation Over Time and Temporal Multi-scale

These space-based representations are, however, clearly informative only on a *frame-based* approach. When it comes to represent musical processes, movement-in-space is a convenient metaphor, but visual problems start to arise for leaving traces of past frames. However, it's often of high interest to analyze long segments of music to observe far-reaching relationships along time.

Additionally, interpretation of tonality is particularly sensitive to temporal scale. At a given time point in music, it is often legitimate to understand different

² For example, pitch-class profiles from Messiaen's *modes of limited transpositions*.

keys when considering shorter or longer contexts. Depending on tonal discourse, some properties of key cannot be captured below or above a certain time-scale, as it seems counterintuitive to experience, in a general case, to grasp a sense of key under a fixed time-scale and hop-size basis.

Activation of pitch-space over time has been visualized as single images in SOM [8], and geometrical representations of different pitch hierarchies, involving two time-scales, have been animated on *SpiralArray* [9]. Both works present critical visual concerns. Spatial representation of concepts in SOM exhausts screens’ dimensional constraints, making the time dimension barely recognizable, while Chew’s moving visual information quickly overloads the short-time memory of observer. This reduces the chances for meaningful long-term visualization or for further exploitation of multi-scale.

Sapp’s *keyscapes* [10] approach the other side of the problem, and represent key estimations over time at many time-scales in a 2-D image. For each frame at each scale, a pixel is coloured according to the strongest key candidate, reducing key representation to a single dimension.

3 Proposed Model

We propose a method for exploring the analytical potential of combining key estimations projected into continuous toroidal pitch-spaces with time vs. time-scale representations. We aim to preserve or to complement information across different dimensional reductions, taking both analytical and perceptual implications into consideration. The following discussed elements have been prototyped as an interactive tonal analysis application, built using the MIDI Toolbox for Matlab [11].

3.1 Signal Segmentation and Key Estimation

First, a chroma feature time-series is extracted from audio or MIDI files, and is segmented following a multi-scale sliding window policy. This is achieved by aligning many conventional sliding windows of sizes ranging from fractions of second to the whole duration of the piece. Following a basic key-finding method (MIDI Toolbox’s implementation), a pitch-class profile is computed from each segment (henceforth, *frame*), and these 12-dimensional vectors are correlated with ring-shifted K-K’s key profiles. The resulting 24-dimensional estimations, for all frames and time-scales, are projected to a K-K’s pitch-space³.

3.2 Mapping Key Estimations into Pitch-Space

Two complementary mapping methods are used for projecting the estimates, which respectively maximize dimensionality reduction and frame informativeness. The first method, virtually identical to that in [7], finds a tonal centroid in

³ Coordinates of the 4-dimensional MDS solution are taken from [1], p.42.

an angle-based parameterization of K-K's torus. The number of key candidates to be considered and an *ambiguity factor*, used to promote or relax the strongest candidate, are introduced as parameters for unfolding. The second projection is provided by MIDI Toolbox's SOM visualization for key analysis.

3.3 Time vs. Time-Scale Representation

We propose an adaptation of Sapp's keyscape to visualize centroids, for all frames and time-scales, in a 2-D image. Instead of mapping categorical keys to colours, our continuous pitch-space requires some considerations for colouring: a) Centroid locations aim for perceptual closeness to keys. b) Ambiguity allows centroids to be anywhere in space. c) Toroidal pitch-spaces feature double spatial circularity. Given that, it is desirable for the colouring policy to account for perceptual distance and continuity or colour blending, and to keep these properties across the space's double circularity.

3.4 Geometrical Colourspace

CIE 1976 $L^*a^*b^*$ (known as CIELAB) colourspace, defined by the Commission Internationale de l'Éclairage as conversion standard [12], can approximate our needs. It uses absolute scales,⁴ it is device independent,⁵ and it is compressed to approximate perceptual uniformity. CIELAB is a 3-dimensional geometric space in which most of the human visible gamut is covered by a spherical sub-space. Three parameters define colour: L^* for luminance, a^* for green-magenta colour-opponent axis, and b^* for blue-yellow axis. Other CIELAB's parameterization, LCh , uses cylindrical coordinates: L for luminance, C for chroma saturation, and h for hue angle.

3.5 Colouring Strategy

For assigning colours to centroids, the whole 3-D toroid (volume) is inscribed into CIELAB's *sphere*. This gives a unique colour for each point in pitch-space (toroidal surface) and gradual colour transitions along any direction, it approximates perceptual correlation with distance, and it guarantees double circularity. We match hue angle with the circle in first 2 dimensions of K-K's 4-D solution to provide maximum hue differences along both circles of fifths. The last two dimensions of the solution are parameterized as a circle in L and C axes. Geometrical manipulation allows for rotating the torus to match any colour to any point, which can be used for local compensation of the projection's distortion⁶ respecting the 4-D space. The colouring method is schematized in Fig. 2.

The final keyscape is then plotted by aligning all coloured pixels, representing centroids in pitch-space, in a time vs. time-scale image.

⁴ Scales are relative to CIE's *standard illuminant D50* white point.

⁵ Final result depends on the device's gamut and mapping equations. We use Matlab's LAB to sRGB conversion, based on CIE's *perceptual intent* recommendation.

⁶ Comprehensive discussion respecting inter-key distances in [1].

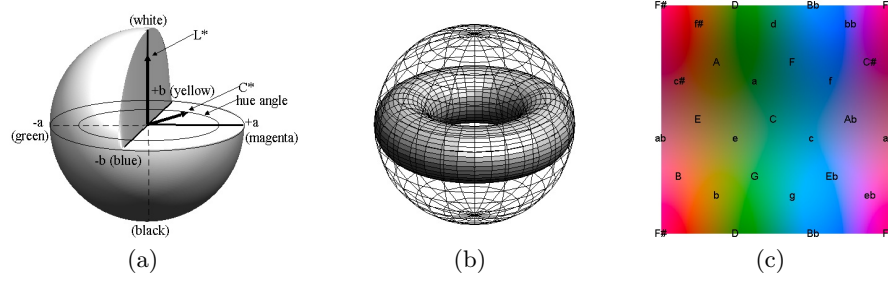


Fig. 2. Colouring process. a) Spherical sub-space of CIELAB colourspace. b) Geometrical inscription of toroidal pitch-space in spherical colourspace. c) Unfolded torus' surface, coloured after its embedding in colourspace.

4 Discussion

The Finale of Haydn's String Quartet op.74 n.3 "Rider", in G minor, will serve to show the combined representational capabilities of the proposed model. The keyscape in Fig. 3(a) represents tonal estimations in time (x-axis) vs. time-scale (y-axis). The higher the position in the keyscape, the larger the time-scale used for key estimation. Unfolding has been parameterized to promote the first key candidate, so centroids are close to single keys. In the image, homogeneous colours covering large areas reveal stable tonal sections. Both coloured pitch-spaces in Fig. 3(b) and Fig. 3(c) serve as map legend between the keyscape and the pitch-space.

4.1 Formal Analysis

The piece begins with a first theme in Gm, followed by a longer second theme group in Bb. Then, a development section wanders around 3 and 4-flat key signatures (each one as different blue tone). Recapitulation takes the first theme in Gm, and a dominant pedal leads to the second theme group and conclusion in G major.⁷

The complete piece is estimated as Gm (very top of the keyscape), a somewhat surprising result given the relative short duration in this key. The proposed connection of keyscape with pitch-space helps to clarify this analytical aspect at structural level. This is shown in Fig. 3(b) by projecting all the centroids (black dots) for the highest time-scales (above horizontal dashed line in the keyscape) and observing the relative *weight* reached at each location (since centroids fall very close to keys, their accumulation is approximated by the circles' sizes). The exposition and recapitulation sections have been summarized as Bb and G respectively, given the longer duration of their second theme groups. They are almost balanced in duration, and Gm falls approximately half-way between

⁷ Analysis by first author.

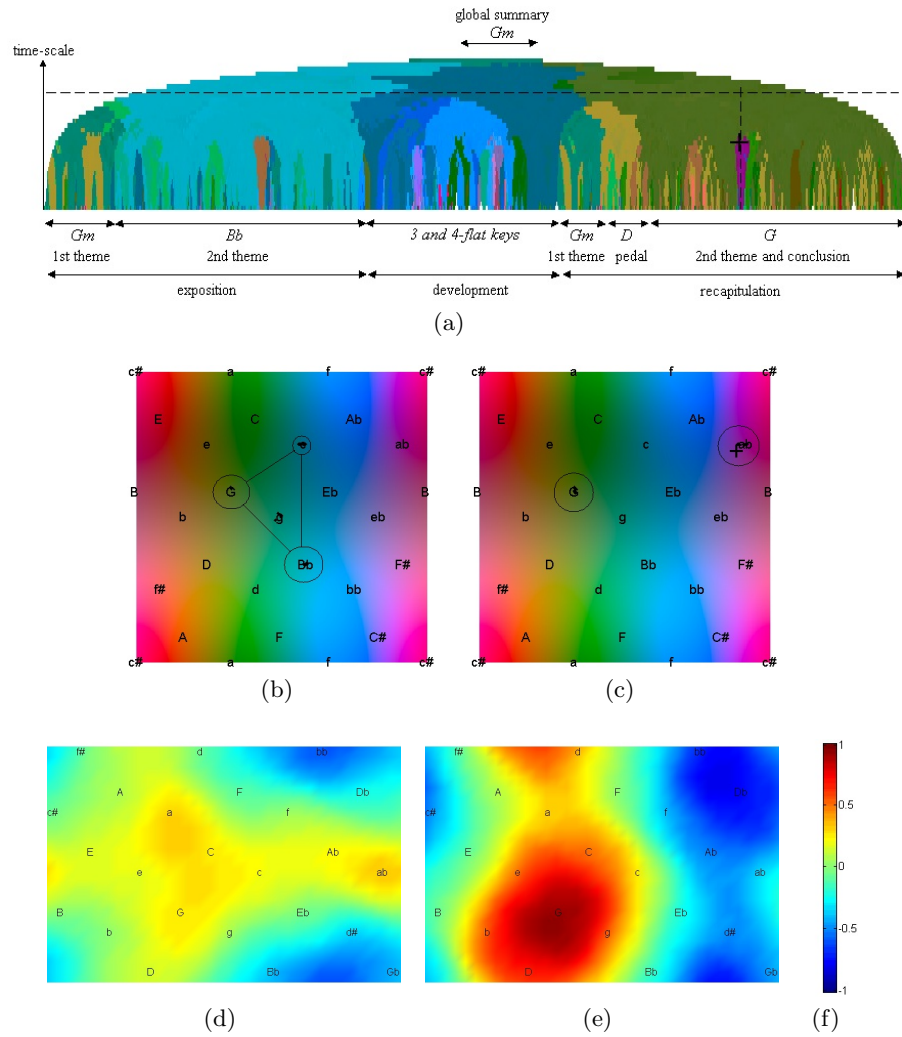


Fig. 3. Finale of Haydn's "Rider" String Quartet. a) Keyscape. b) Spatial-structural relationships at the highest time-scales. c) Clustered centroids for the selected time-frame. d) SOM activation for the selected event. e) SOM activation for the selected event's context. f) SOM key strength colourbar.

them in pitch-space. The shorter development section, summarized at Cm, gently pulls the whole-piece centroid towards the correct key of the piece. This spatial interpretation of *classical symmetry* is a consequence of the topology of pitch-space, and two aspects linking the key estimation method with some of the piece’s *aesthetical* concerns. First, the hierarchy implied in K-K’s profiles is well suited to pitch-class statistical analysis of classical works. Second, our duration-based computation of pitch-class profiles matches the classical style’s ideals of temporal balance. It’s worth mentioning that the given interpretation has not been computed as a *systemic* reduction: all the estimates are computed taking into account the overall pitch-class content in their segments.

4.2 Tonal Mining in More Dimensions

The keyscape provides a big picture tonal description of the piece, although quite summarized. We propose to use keyscales in interaction, in order to uncover the missing information. A point in the keyscape has been selected (+ sign) in Fig. 3(a) for a more detailed inspection of the corresponding segment. All the centroids aligned in time with the selected frame are represented in Fig. 3(c), showing a *vertical reading* at that time point (vertical dashed line in the keyscape). The tonal meaning arises: clustered centroids (surrounded by circles in the pitch-space) identify an Ab minor *event* within a G context.

Centroids are strong simplifications of the actual estimates. The correlation stage of our key-finding model outputs 24-dimensional vectors, which are *forced* to become single points in the torus by the unfolding stage. A more explanatory view of the selected segment is given by the SOM in Fig. 3(d), where a notable key uncertainty appears as weak activations throughout and no clustered isolated region. This ambiguity contrasts with the clarity and strenght of the activation produced by its context in G major, just few time-scales above, as shown in Fig. 3(e).

4.3 Other Aspects of Tonal Ambiguity

Chopin’s Prelude in E major will serve to illustrate two different unfolding parameterizations (Fig. 4). This piece can be structured in three phrases, each of them beginning and ending in E. The last two phrases take distant journeys in pitch-space, and considerable analytical freedom respecting tone center induction is allowed.

The keyscape in Fig. 4(a) maps centroids in pitch-space by promoting the strongest key, showing homogeneous colours and neat boundaries between regions. Typical analyses of modulation and tonicizations for this piece are found for time-scales between the dashed lines. Compared with the tonal stability of the first phrase, the keyscape shows important key shifting activity for the last two sections. The second phrase visits G-C-Fm-Ab-G#m, and the third section *crosses* Am-F-Dm-Gm-G-Em. A comparable multi-scale tonal analysis through pitch-space for this piece is given in [5]. The model estimates Am in the third phrase, which is suggested by Lerdahl to fit the *shortest path* rule from E to F.

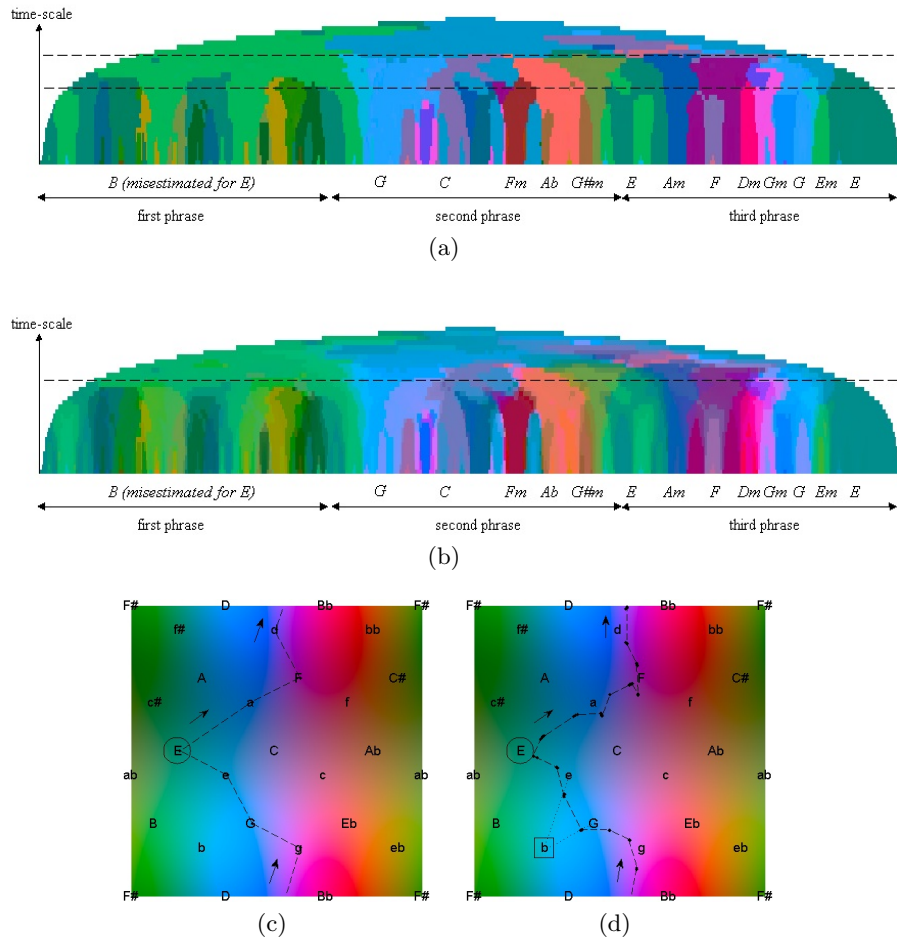


Fig. 4. Chopin's Prelude in E. a) Keyscape promoting strongest candidate. b) Keyscape from the first two candidates. c) Path followed by the third phrase under categorical unfolding. d) Path of the third phrase under ambiguous unfolding.

Similarly, Dm is estimated as mediating between F and Gm. The algorithm fails to estimate the first phrase in E, in favour of B, as well as the global key of the piece.

Centroid unfolding for the second keyscape, in Fig. 4(b), considers the two strongest candidates and their weights. The image looks similar but it’s significantly *fuzzier*, being informative in a different (not necessarily better) way. Most areas represent spatial locations between two keys, by proportional *blending* of both candidates’ colours. The continuous colourspace aims to facilitate the distinction between movement towards neighbour keys (soft mixture of both colours) and far jumps (sudden strong contrast). As a difference with other colouring schemes, like that in [10], the proposed geometrical mapping approximates gradual colour transitions in any direction, so all neighbour tonal relationships would get benefited from the perceptual closeness.

Fig. 4(c) shows the path followed by the centroids for the third phrase under categorical unfolding (black dots fall right in the key centers). The same phrase is shown in Fig. 4(d) for the ambiguous case. The time-scale used in both cases is shown in the ambiguous keyscape (horizontal dashed line). Instead of jumping to categorical keys, some intermediate positions soften the way and the corresponding colours in the ambiguous path. By observing the centroids’ deviations from their closest keys it is possible to grasp which keys were taken as second candidates, as it’s illustrated in the ambiguous path during the transition between G and E minor. It’s worth noting that the most critical aspect for achieving gradual colour shifting is the capability of the key estimation and centroid unfolding methods to provide such gradual movement in pitch-space, and this necessarily depends on musical discourse as well.

5 Conclusions and Future Work

We have proposed a representation of multi-scale key estimation over time, where pixels represent uncertain centroids in a continuous pitch-space. This mapping is achieved by introducing a geometrical continuous colourspace. Big picture keyscapes are linked with higher-dimensional representations, which can account for extended tonal description. We have pointed to the potential of continuous pitch-spaces in supporting long term analytical conclusions, when assisted by time vs. time-scale representations. The method also facilitates the inspection of different cases of tonal ambiguity.

Our prototype proposes an interactive selection of segments and trajectories in the keyscape, which acts as the entry point for analysis. This information is complemented by the higher dimensional models and contrasted by audition, so it could be useful for refined evaluation of key modeling algorithms.

We expect that interacting with these combined representations can contribute to better understanding of the analytical relevance of time-scale in tonality. The exploitation of human vision abilities for reaching conclusions from ambiguous evidence might play an important role in such interaction.

Acknowledgements

This work has been supported by the projects Classical Planet: TSI-070100-2009-407 (MITYC) and DRIMS: TIN2009-14247-C02-01 (MICINN).

References

1. Krumhansl, C. L. Cognitive Foundations of Musical Pitch. Oxford University Press, New York (1990)
2. Schoenberg, A. Structural Functions of Harmony (rev. ed.). Norton, New York (1969) (Originally published, 1954)
3. Chew, E.: Towards a Mathematical Model of Tonality. Ph.D. dissertation, Massachusetts Institute of Technology (2000)
4. Werts, D. A Theory of Scale References. Ph. D. dissertation, Princeton University (1983)
5. Lerdahl, F. Tonal Pitch Space. Oxford University Press, New York (2001)
6. Coombs, C. H.: A Theory of Data. Wiley and sons, New York (1964)
7. Krumhansl, C. L., Toivainen, P. Tonal Cognition. In: Peretz, I., Zatorre, R. (eds.) The Cognitive Neuroscience of Music, pp. 95-108. Oxford University Press, New York (2003)
8. Janata, P. Navigating Tonal Space. In: Hewlett, W.B., Selfridge-Field, E., Correia, E.Jr. (eds.) Tonal theory for the digital age. Computing in Musicology 15, pp. 39-50. CCRH, Stanford (2007)
9. Chew, E., François, A.R.J.: Interactive Multi-scale Visualizations of Tonal Evolution in MuSA.RT Opus 2. Computers in Entertainment 3(4), pp. 1-16 (2005)
10. Sapp, C.S. Visual Hierarchical Key Analysis. Computers in Entertainment 3(4), pp. 1-19 (2005)
11. Eerola, T., Toivainen, P.: MIDI Toolbox: MATLAB Tools for Music Research. University of Jyväskylä, Kopijyvä, Jyväskylä, Finland (2004). Electronic version available at <http://www.jyu.fi/musica/miditoolbox/>
12. CIE 015:2004. Colorimetry (3rd ed.). Commission Internationale de L'Éclairage (2004)