VISUALIZATION OF LOW DIMENSIONAL STRUCTURE IN TONAL PITCH SPACE

J. Ashley Burgoyne and Lawrence K. Saul
Department of Computer and Information Science
University of Pennsylvania
{burgoyne,lsaul}@cis.upenn.edu

ABSTRACT

In his 2001 monograph *Tonal Pitch Space*, Fred Lerdahl defined an distance function over tonal and post-tonal harmonies distilled from years of research on music cognition. Although this work references the toroidal structure commonly associated with harmonic space, it stops short of presenting an explicit embedding of this torus. It is possible to use statistical techniques to recreate such an embedding from the distance function, yielding a more complex structure than the standard toroidal model has heretofore assumed. Nonlinear techniques can reduce the dimensionality of this structure and be tuned to emphasize global or local structure. The resulting manifolds highlight the relationships inherent in the tonal system and offer a basis for future work in machine-assisted analysis and music theory.

1. INTRODUCTION

Since Gottfried Weber introduced the chart in Figure 1 early in the nineteenth century [11], music theorists have acknowledged two pivotal axes controlling the relationships among the major and minor keys of the diatonic tonal system in Western art music: the cycle of fifths, represented on the vertical axis of the figure, and the cycle of thirds, represented on the horizontal. [1] (Capital letters designate major keys and lowercase minor, as is traditional.) Except in the most abstract theoretical formulations, these axes are considered to be periodic, defining a topological space isomorphic to $\mathbb{S}^1 \times \mathbb{S}^1$, and by the end of the twentieth century, Carol Krumhansl's pioneering psychological experiments had demonstrated a cognitive basis for this toroidal structure [4]. Krumhansl's work also explored topological relationships among harmonies and pitch classes within each key [5], which Fred Lerdahl integrated into the framework of *A Generative Theory of Tonal Music* [7], his 1983 monograph coauthored with Ray Jackendoff, in a 2001 monograph entitled *Tonal Pitch Space* [6].

Although Lerdahl makes much of Krumhansl's data and the toroidal topology of harmonic space, he defines that torus only implicitly, by way of a distance function

d♯	F♯	f♯	A	a	C	c
g♯	B	b	D	d	F	f
c♯	E	e	G	g	B♭	b♭
f♯	A	a	C	c	E♭	e♭
b	D	d	F	f	A♭	a♭
e	G	g	B♭	b♭	D♭	d♭
a	C	c	E♭	e♭	G♭	g♭

Figure 1. Weber's diagram of tonal space

over harmonies. No other research to date has attempted to embed it explicitly. David Temperley used a MIDI-based approach to implement many components of the theory [9] but has not yet treated the topology of the space. The Mathematical Music Theory Group at the Technical University of Berlin uncovered some inconsistencies in Lerdahl's theory while developing their *HarmoRubette* software tool [8] but did so strictly in terms of distance functions. Elaine Chew's spiral model [1, 2] is an explicit representation of tonal space that has aided the development of intelligent musical systems, most notably for key finding and pitch spelling, but it is founded on music theoretical principles (the Riemannian *Tonnetz*) that, despite the apparent similarities, are incompatible with Lerdahl's and Krumhansl's. The visualizations of tonal pitch space presented in this paper complement Chew's model and should be especially useful for machine-assisted harmonic and hierarchical analysis.

2. LERDAHL'S DISTANCES

One of the distinguishing features of Lerdahl's model is that it treats pitch classes, chords, and regions (keys) as unified and inseparable. There is no well defined notion of distance between pitch classes *qua* pitch classes or chords *qua* chords. Pitch classes have meaning only as elements of the sets that define chords and regions, and chords are always understood as functioning within some region. An important corollary is that there is always a nonzero distance, albeit usually small, between two instances of the same nominal chord when these instances are heard in dis-

[1] Vial had organized the keys in a similar fashion three decades earlier, but Weber's treatise proved to be more influential.

tinct regions: $\mathrm{C}/\mathbf{C}$ is not the same as $\mathrm{C}/\mathbf{F}$ and certainly not the same as $\mathrm{C}/\mathbf{D}\flat$.[2] This corollary gives Lerdahl's model more nuance than most alternatives.

For two harmonies $\mathrm{x} = \mathrm{C}_1/\mathbf{R}_1$ and $\mathrm{y} = \mathrm{C}_2/\mathbf{R}_2$, the simple distance is given by the equation

$$\delta(\mathrm{x} \to \mathrm{y}) = i + j + k \quad (1)$$

where i is the smallest number of steps along the circle of fifths between $\mathbf{R}_1$ and $\mathbf{R}_2$ (or their relative majors in case one or both is minor), j is is the smallest number of steps along the circle of fifths between the roots of C_1 and C_2 within each region, and k is a specially weighted Hamming distance between the sets of pitch classes that define each chord and region. Lerdahl's formulation of the k parameter is asymmetric, and so to create a symmetric distance function, we have taken the average of the two directions. Lerdahl restricts δ to prevent implausible modulations, allowing it to be defined only when either x and y are in the same region or at least one of C_1 and C_2 is a tonic chord and $\mathbf{R}_1$ and $\mathbf{R}_2$ are in each other's set of "pivot regions," $\{\mathbf{i}, \mathbf{ii}, \mathbf{iii}, \mathbf{IV}, \mathbf{V}, \mathbf{vi}\}$ for major keys and for minor keys $\{\mathbf{I}, \flat\mathbf{III}, \mathbf{iv}, \mathbf{v}, \flat\mathbf{VI}, \flat\mathbf{VII}\}$. The general distance function is

$$\begin{aligned}\Delta(\mathrm{C}_1/\mathbf{R}_1 \to \mathrm{C}_2/\mathbf{R}_2) &= \delta_1(\mathrm{C}_1/\mathbf{R}_1 \to \mathrm{I}/\mathbf{P}_1) \\ &+ \delta_2(\mathbf{P}_1 \to \mathbf{P}_2) + \delta_3(\mathbf{P}_2 \to \mathbf{P}_3) + \cdots \\ &+ \delta_\mathrm{n}(\mathrm{I}/\mathbf{P}_\mathrm{n} \to \mathrm{C}_2/\mathbf{R}_2) \quad (2)\end{aligned}$$

where $\delta(\mathbf{P}_\mathrm{i} \to \mathbf{P}_\mathrm{j})$ is shorthand for $\delta(\mathrm{I}/\mathbf{P}_\mathrm{i} \to \mathrm{I}/\mathbf{P}_\mathrm{j})$, and the chain of regions $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_\mathrm{n}$ is chosen to minimize Δ within the constraint that $\delta_1, \delta_2, \ldots, \delta_\mathrm{n}$ are defined.

3. MULTIDIMENSIONAL SCALING

Isomap is an algorithm for extracting low dimensional embeddings of high dimensional data that respect the local geometry of the original data while simplifying the global geometry [10]. It has three primary steps. The first is to define a local neighborhood for every point in the data set, typically a fixed number of nearest neighbor points or the set of all points within a fixed distance. The next is to define a graph in which each vertex is a data point and is connected only to its neighbors as defined in the previous step with edges of length proportional to the distance between them; in other words, all non-neighboring entries of the distance matrix are discarded. A new distance matrix is generated using the lengths of the shortest paths between all pairs of vertices in the graph. In the last step, classical multidimensional scaling (MDS) is applied to the new distance matrix. MDS is a linear statistical technique that maps a matrix of pairwise distances between data points to a low dimensional embedding of those points with pairwise distances as close as possible to the input matrix. The dimensions of the output embedding are ordered according to how much variance they preserve from the original data.

Lerdahl's distances are isomorphic to the first steps of the Isomap algorithm. Given the set of all possible harmonies, Equation 1 defines a restricted set of neighbors for every harmony: the other harmonies within the same tonal region and the tonics of the pivot regions. The generalized distance in Equation 2 is precisely the length of the shortest path along these edges. We restricted our harmonic space to the major and minor triads of each the 24 major and minor keys and the chromatically altered major and minor triads that can be reached via simple, secondary, or double mixture. This space includes 22 triads per key (every triad except the major and minor triads rooted a tritone away from the tonic) for a total of 528 harmonies. Understanding Lerdahl's model in this way, visualizing the structure of his tonal space requires only one more step: MDS. The lower bar in Figure 2 illustrates the distribution of information across the dimensions of the output. Although the leading two dimensions dominate, at least the leading four are significant and fifteen would be required to capture even 75% of the information contained in Lerdahl's model.

Figure 3(a) is a plot of major-key harmonies the leading two dimensions. Each tonal region is colored in its own shade of gray and labeled with capital letters. The pattern is a tightly organized regional circle of fifths, which confirms that Lerdahl's model conforms in at least one regard to the Weber-Krumhansl model it seeks to emulate.[3] One thus would expect the second and third dimensions to trace a cycle of thirds, but the structure is more complicated than that. Figure 3(b) unwraps the circle from Figure 3(a) by converting the first two dimensions to polar coordinates and plots the third and fourth dimensions of the embedding with respect to the polar angle. These dimensions form a spiral with three periods to the first dimension's one. Figure 3(c) attempts to clarify the relationship between these structures by converting the third and fourth dimensions to polar coordinates as well and plotting their angular component against the angular component of the first two dimensions (again, major keys only

[2] Lerdahl uses Roman type for chords and boldface type for regions, and we follow the convention here, e.g., $\mathrm{ii}/\mathbf{G}$ for the minor supertonic chord in G major or $\mathrm{E}\flat/\mathbf{F}$ for an E♭ major chord understood in the key of F major.

[3] If space had permitted the minor keys to be plotted in the form of 3(a), they would be interspersed with the major keys but form their own circle of fifths corresponding to the lowercase labels. In contrast to both the Krumhansl model and most theory textbooks, major keys are paired with neither their relative or parallel minors but the minor key a whole tone higher. This relationship arose early in harmonic theory with David Heinichen's *General-Bass in der Composition* [3]. It arises as a neutral statistical compromise between the parallel and relative key relationships so as to allow the third and fourth dimensions to account for them properly.

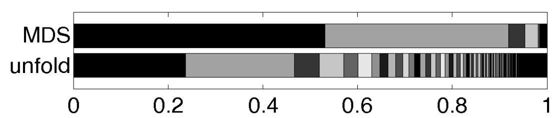
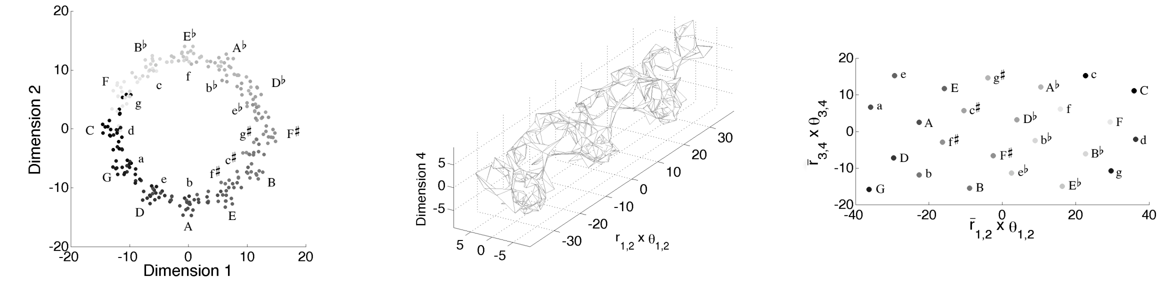


Figure 2. Relative variance



(a) Dimensions 1 and 2: A circle of fifths (b) Dimensions 3 and 4: An inner spiral (c) A regional *Tonnetz*

Figure 3. Tonal pitch space as viewed with multidimensional scaling

to aid visualization in black and white). Here, the cycles of thirds emerge. Crisscrossing the spiral in a form that looks much like a Riemannian *Tonnetz*, the cycles of minor thirds in regional space travel from the top left to the bottom right of the figure and the cycles of major thirds move horizontally. The horizontal axis of the figure is stretched to aid comparison with the other figures, but when scaled equally, the distances along these cycles of thirds are as close as those along the circle of fifths.

The presence of the circle of fifths and the cycles of thirds is sufficient for isomorphism to the toroidal regional model of the psychological literature. The remaining two dimensions, the radial components of the two sets of polar coordinates, distinguish our model. Together, they organize the chords within each region around the toroidal structure of the regions themselves. Weber first presented the regional structure, Krumhansl and Kessler assigned an embedding to it, Lerdahl developed a theory to incorporate inter-regional relationships, and our work embeds that. Our embedding should allow machine analysis systems to synthesize key finding and harmonic analysis more smoothly.

4. MAXIMUM VARIANCE UNFOLDING

The inter-regional structures in this embedding are less consistent than the intra-regional ones. This shortcoming is tied to MDS, which can optimize only over the global structure of its input data. There are nonlinear algorithms, however, that can shift the emphasis to local structures. One very effective such technique is maximum variance unfolding [12]. It begins with the choice of a neighborhood surrounding each point. The pairwise distances within these neighborhoods are locked, and then semidefinite programming is used to expand the data as much as possible without violating these locks. The procedure is analogous to stretching a ball-and-stick model in which the balls correspond to data points and the sticks correspond to the locked distances.

Maximum variance unfolding can fail when the original distances are non-Euclidean, and as Noll and Garbers note, Equation 2 is not a true distance function because its special handling of pivot regions causes it to violate the triangle inequality. For our experiments, we Euclideanized the Δ-derived distance matrix before computing neighborhoods by converting it to a Gram matrix of inner products, replacing all negative eigenvalues with zeros, and converting back.

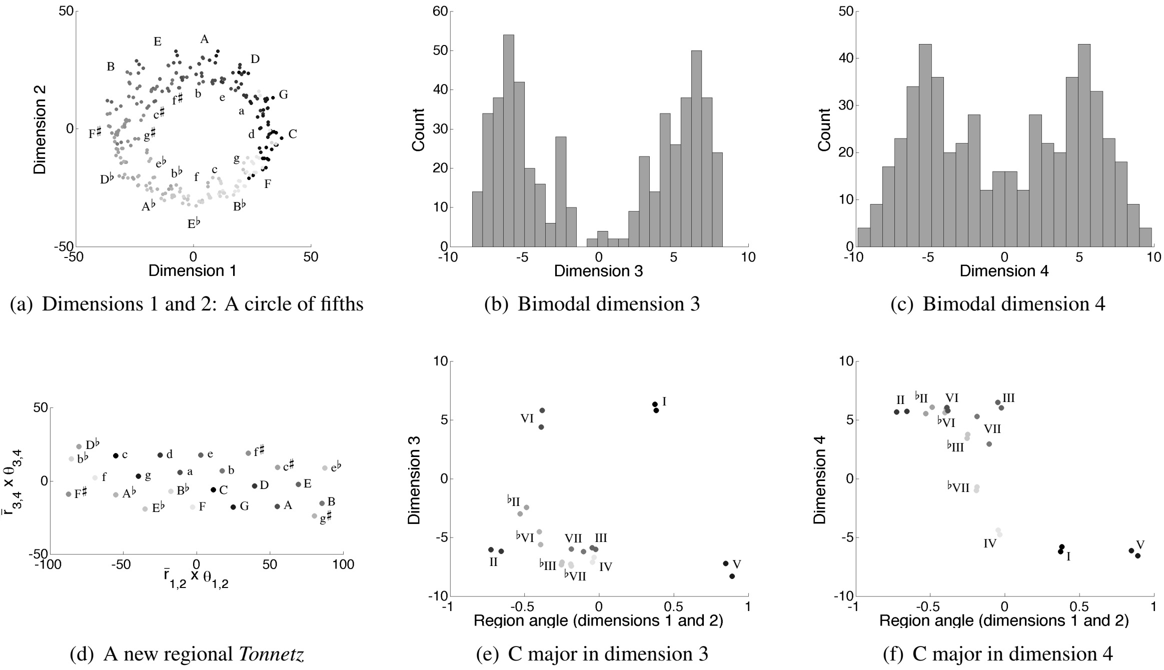
By tuning the size of the neighborhoods, one can control the level of structure in the output embedding. Large neighborhoods yield more global structures and behave comparably to algorithms like MDS, while small neighborhoods preserve local structure and can provide much better dimensionality reduction. Figure 2 illustrates the difference. The top bar represents the distribution of information after using the semidefinite programming method on the Euclideanized distance matrix with neighborhoods including the four nearest neighbors to each harmony. The leading four dimensions account for 98% of the information after SDP, after MDS only 57%.

Figure 4(a) presents the leading two dimensions of this embedding. As in the linear case, they form a circle of fifths. The third and fourth dimensions, however, serve different purposes. The histograms in Figures 4(b) and 4(c) show that the data is bimodal in each of these dimensions. The third dimension separates regions into two isomorphic planes a whole tone apart; the fourth dimension separates the major keys from the minor keys. These patterns are evident in Figure 4(d) and form a very different regional network than the one from the MDS embedding in Figure 3(c). These dimensions also preserve consistent chordal structures across the regional structure. As seen in Figures 4(e) and 4(f), dimension 3 keeps tonics with the tonics of their relative keys while dimension 4 puts them closer to the dominant and subdominant.

Notably absent from the embedding produced by SDP are the cycles of thirds. The planes of the third dimension keep relative major and minor regions together, but parallel regions are separated in all four dimensions. This is an unavoidable cost of the greatly reduced dimensionality of the nonlinear embedding overall.

5. SUMMARY AND FUTURE WORK

Linear and nonlinear statistical methods can produce embeddings that emphasize the global or local structure of data defined by pairwise distances and can help visualize models of tonal pitch space, including structure is more



(a) Dimensions 1 and 2: A circle of fifths (b) Bimodal dimension 3 (c) Bimodal dimension 4

(d) A new regional *Tonnetz* (e) C major in dimension 3 (f) C major in dimension 4

Figure 4. Tonal pitch space as viewed with maximum variance unfolding

complex than the three-dimensional toroidal model commonly cited in psychological literature. These higher dimensional embeddings incorporate more subtle details of the harmonic system that are helpful to visualize and can serve as foundational models for machine-assisted analysis. Lerdahl's theory also includes an extension to what has been summarized here that incorporates hexatonic, octatonic, and other non-diatonic tonal models, which we hope to incorporate into our existing framework.

6. ACKNOWLEDGEMENTS

This work was supported by NSF Award 0238323. We also would like to thank Fei Sha, Kilian Weinberger, and Chris Hasty for sharing their ideas as ours developed.

7. REFERENCES

[1] E. Chew, "Towards a mathematical model of tonality," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[2] ——, "Thinking out of the grid and inside the spiral: Geometric interpretations of and comparisons with the spiral array model," University of Southern California, Tech. Rep. 03-002, 2003.

[3] J. D. Heinichen, *General-Bass in der Composition*, Dresden, 1728.

[4] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press, 1990.

[5] C. L. Krumhansl and E. J. Kessler, "Tracing the dynamic changes in perceived tonal organisation in a spatial representation of musical keys," *Psychological Review*, vol. 89, pp. 334–368, 1982.

[6] F. Lerdahl, *Tonal Pitch Space*. New York: Oxford University Press, 2001.

[7] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press, 1983.

[8] T. Noll and J. Garbers, "Harmonic path analysis," in *Perspectives of Mathematical and Computational Music Theory*, G. Mazzola, T. Noll, and E. Luis-Puebla, Eds. Osnabrück: epOs-Music, 2004.

[9] D. Temperley, *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press, 2001.

[10] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 22 December 2000.

[11] G. Weber, *Versuch einer geordeneten Theorie der Tonsetzkunst*. Mainz: B. Schotts Söhne, 1821–24.

[12] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-04)*, vol. 2, Washington, D.C., 2004, pp. 988–995.