1 Cyclic-Distance Patterns among Spectra of Diatonic Sets: The Case of Instrument Sounds with Major and Minor Scales

Özgür İzmirli
Center for Arts and Technology
Computer Science
Connecticut College
New London, CT 06320, USA
oizm@conncoll.edu

Abstract

This paper describes a method that employs a structural approach to tonality to explore the distance relationships among spectral prototypes. Prototypes are obtained by accumulating spectra of notes that pertain to diatonic sets. A two-dimensional visualization is sought to view the distance relationships of spectral prototypes. Principal Component Analysis is applied to the prototypes for dimensionality reduction. It is shown that a circle-of-fifths pattern emerges when the diatonic sets consist of major and harmonic minor scales. Similar results are obtained for synthetically generated spectra and spectra obtained from real musical sounds.

Tonal Theory for the Digital Age (Computing in Musicology 15, 2007), 11-23.

1.1 Introduction

Models of tonality induction deal with the problem of determining the tonal center of a musical piece. The tonal center may be defined as the most stable pitch around which a listener organizes other pitches used in this piece. Most of the research in this area has concentrated on Western tonal music. Some models of tonality induction take symbolic input, such as MIDI or other forms of score representation, and estimate the tonal center from this information. These models attempt to either demonstrate the emergence of tonal structure (e.g., Tillman, Bharucha and Bigand 2000) model music cognition (e.g., Lerdahl 2001) or deduce local or global key information from discrete music events (e.g., Krumhansl 1990, Vos and Van Geenan 1996, Chew 2000, Temperley 2001). The use of symbolic information means that these models are isolated from the sonic qualities of the actual music. One way to apply these models to acoustic music is through using automated transcription. The accuracy of general-purpose polyphonic audio-to-note transcription systems, however, is currently insufficient for them to be used in cascade with these models. On the other hand, a parallel path of research is also present in which models operate on audio data (e.g., Huron and Parncutt 1993, Leman 1995, İzmirli and Bilgen 1996, Purwins et al. 2001, Pauws 2004, Chuan and Chew 2005, İzmirli 2005, Gómez 2006, and İzmirli 2006). These models deal with the fuzziness of sonic qualities present in the input. Regardless of the nature of the input, what is common to all models is the inherent use of particular structures of tonal hierarchy. That is to say, each method has its own version of a structure that depicts distance relationships among tonal centers, such that closely related keys have shorter distances to each other than remotely related keys. However, not all models explicitly define the structure of tonal hierarchy in their formulation.

Geometric models may be used to represent tonal space in which the distance relationships among tonal centers are defined. Historically, these models have been used to model tonal hierarchy in music. They aim to capture the structure of tonal space, as understood in the context of tonal practice, by utilizing different forms of geometries to array points representing tonal centers. The distances between tonal centers represent cognitive distances. Examples of these models are: one-dimensional models (e.g., circle or line of fifths), a double helix (Shepard 1982), and a torus (Krumhansl 1990). An overview of geometric models can be found in Lerdahl (2001). These models represent structures formed by long-term listening experiences and are used in the interpretation of incoming musical information for tonality induction.

The aim of the work presented in this paper is to study the relationships of key distances that arise when diatonic sets are used. This is done by analyzing the cumulative spectra originating from diatonic sets. Spectra of monophonic sounds recorded from real instruments and synthetically generated line spectra are considered in this work. The purpose is to analyze the geometric relationship among tonal centers and compare the conformance of the outcome to music theory. Tonal centers are represented by spectral prototypes obtained from diatonic pitch collections. The approach of this method is purely structural, as it only uses spectral information and disregards the dimension of time. For this reason, it does not take into consideration any sequential or intervallic information. The spectral prototypes are obtained by accumulating spectra corresponding to single notes that are members of the corresponding diatonic sets.

The method consists of two stages. In the first stage, spectral accumulation of the input sounds is performed in order to find spectral prototypes. In the second stage, dimensionality reduction is performed on these prototypes in order to find a low dimensional representation suitable for visualization.

1.2 Keys and Spectral Prototypes

In this section, the calculation of spectral prototypes is described. These prototypes can be obtained in two ways. One is by using the spectra of real instrument sounds and the other is by synthetically generating line spectra. First, spectral accumulation using monophonic real instrument sounds will be discussed. The input set **N** consists of all chromatic notes in an instrument's range $\mathbf{N}_m$, m=1..m_{max}. Diatonic collections are chosen from this set forming subsets of **N**. Each diatonic set is used to form a prototype $\mathbf{P}_i$,where **i** is the index of the key for that diatonic subset. By convention, C refers to index 0, C#/Db to 1 and D to 2, etc. Elements of each diatonic subset are determined by applying a musical scale pattern to the tonic note which corresponds to the index of the prototype. Only major and harmonic minor scales are used. For example, the major diatonic set with index 0, $\mathbf{D}_0$, refers to the pitch-class set C-D-E-F-G-A-B. This means that a note with pitch C in any octave will be an element of set $\mathbf{D}_0$. This applies to the remaining pitches in this set, all occurrences of which are elements of $\mathbf{D}_0$. Similarly, $\mathbf{D}_2$ refers to the set with notes of the D major scale: D-E-F$^{\#}$-G-A-B-C$^{\#}$, and $\mathbf{D}_4$ refers to the set with notes of the E major scale. In short, a note $\mathbf{N}_m$ is an element of a diatonic set $\mathbf{D}_i$ if that diatonic set contains the same pitch as $\mathbf{N}_m$, regardless of its octave position. Naturally, each scale type has 12 diatonic sets associated with it. Hence, diatonic sets $\mathbf{D}_0$ through $\mathbf{D}_{11}$ correspond to the major scale pattern, and again by convention, sets $\mathbf{D}_{12}$ through $\mathbf{D}_{23}$ correspond to the harmonic minor scale pattern.

The prototype for a specific diatonic set, $\mathbf{D}_i$, is found by calculating the accumulated amplitude spectra for each note in this set and then accumulating again all the note spectra for this set. The time signal for each note is divided into overlapping frames and the amplitude spectrum is calculated for each frame. The accumulated amplitude spectrum of each note is calculated by summing the spectra for all frames with sufficient signal energy. The accumulated amplitude spectrum for note **m** is given by:

$$T_{\mathbf{m}} = \sum_{\mathbf{k,E>\sigma}} S_{\mathbf{k,m}} \quad (1)$$

$S_{k,m}$ denotes the amplitude spectrum of frame **k** for the note with index **m**. The sum is calculated for frames that exceed a certain energy threshold . Each spectrum is then divided elementwise into its mean, μ_{Tm}, for normalization. This is to account for the differences in energy of notes in different registers. Each prototype is then obtained by a second phase of accumulation that uses those notes that are elements of the associated diatonic set:

$$\mathbf{R_i} = \sum_{\mathbf{m, N_m \in D_i}} \frac{\mathbf{T_m}}{\mu_{\mathbf{Tm}}} \qquad (2)$$

$\mathbf{R}_i$ designates the unnormalized prototypes and, for example, in the case of the major diatonic set, index **i** again runs from 0 to 11, one for each diatonic set. Finally, these are normalized to obtain the prototypes:

$$\mathbf{P_i} = \frac{\mathbf{R_i}}{\mu_{\mathbf{Ri}}} \qquad (3)$$

Here μ_{Ri} denotes the mean of the elements of vector $\mathbf{R}_i$. This normalization is useful in the case that the scale types used have different numbers of notes.

As an alternative method to determining the prototypes from sampled sounds, synthetically generated spectra can also be used as input to the second stage. In this case, instead of accumulating spectra using real instrument sounds, line spectra are generated for individual notes. The spectrum for each tone has well-known musical signal characteristics, such as decaying spectral envelopes and harmonic overtones. Once the spectra for all notes have been determined, prototypes are calculated as explained above. Whether the spectral prototypes are calculated from real musical tones or are synthetically generated, they are passed on to the next stage for dimensionality reduction.

1.3 Dimensionality Reduction

The prototype vectors in their current form represent points in a high-dimensional space and visualization of any distance relationship of these prototypes in this space is not directly possible. For visualization purposes, it is desirable to represent the relationship in the smallest number of dimensions possible. Evidently, any number of dimensions between one and three can be used as Cartesian axes to visualize the relationship being explored. In order to explore the actual dimensionality of the input space and to obtain a visualization of the respective distances, principal component analysis (PCA) is utilized. PCA is a well-known method that aims to find a linear transformation from the original axes on which the data is represented to new ones called principal components that also span the input space. Principal components are the new orthogonal axes that are found along which the variance of the data is maximized, starting with the highest variance. Principal components are generally sorted in decreasing order with the first principal component corresponding to the highest variance. An approximation to the input data is obtained by using the first few prominent principal components to reconstruct the data. The number of principal components, and hence the dimensionality, may then be selected in order for the data to be reconstructed with sufficient accuracy. A reduction in dimensionality is obtained if it is possible to reconstruct the data with fewer dimensions in the output with respect to the number of input dimensions. The number of dimensions sufficient to explain the input data depends on how much of the total variance can be accounted for by the principal components selected to approximate the data. Therefore, it is only meaningful to consider the visualizations as reliable and representative of the actual data if the selected number of principal components can explain the input by accounting for sufficiently high variance.

In the current model, PCA is applied directly to the prototypes, $\mathbf{P}_i$. When a single scale type (e.g., major) is chosen for the analysis, all 12 prototypes constructed for that scale are used. It is also possible to add a different scale type (e.g., minor) to be processed together with the first set to bring the total number of prototypes to 24. Below, results for the major scale are given as well as the results for the combined case that incorporates the harmonic minor scale. Figure 1.1 shows a diagram of the combined case in which the major and harmonic minor scales are used as input to the PCA, and the projections of data on the first two principal components are plotted to obtain a geometric representation of spectral distances among the prototypes.

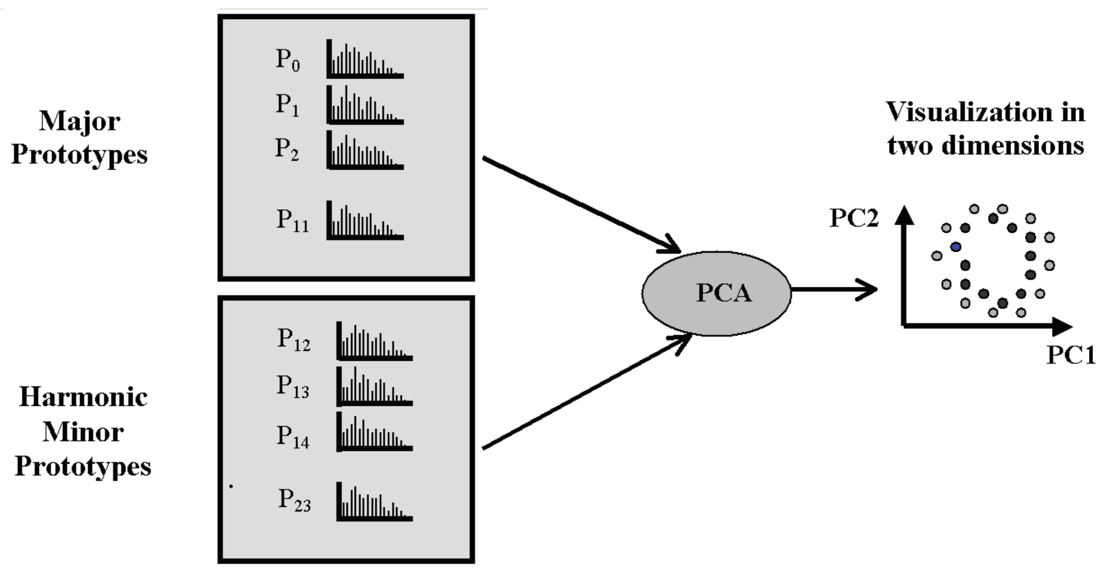

Figure 1.1. *Projections of prototypes are displayed on the first two principal components.*

1.4 Results

The model was tested on piano and violin sounds as well as a few variations of synthetic spectra. The real instrument sounds were taken from the McGill Master Samples. The sounds were sampled at a rate of 11025 samples per second after low pass filtering. The spectra were calculated using a 4096-point FFT with 50% overlap and a Hann window. All prototypes were calculated using the frequency range 45 Hz. to 4000 Hz.

Initially, 12 prototypes were calculated using piano sounds for the major diatonic set. The notes ranged from A1 to B5. PCA was applied to these prototypes and the output was displayed as a projection on the first two principal components. Figure 1.2(a) shows the output that resulted in a nearly perfect circle of fifths arrangement of prototypes. Each prototype is shown with a diamond marker and those prototypes that have a fifth relationship are connected with a line. The lines are drawn to facilitate visualization; they are not otherwise related to the output of the PCA. For the piano tones, the first and second principal components, which have the largest variances, accounted for 42.9% (V1) and 42.3% (V2) of the total variance respectively. The next largest (third) principal component accounted for 3% (V3) of the variance. It can be concluded that this visualization is viable and the projection shown captures most of the vital information. Figure 1.2(b) shows the arrangement of prototypes for the violin using a major scale. In this case, the first two principal components accounted for 82% of the total variance with almost equal weights. The third principal component accounted for only 3.9%.

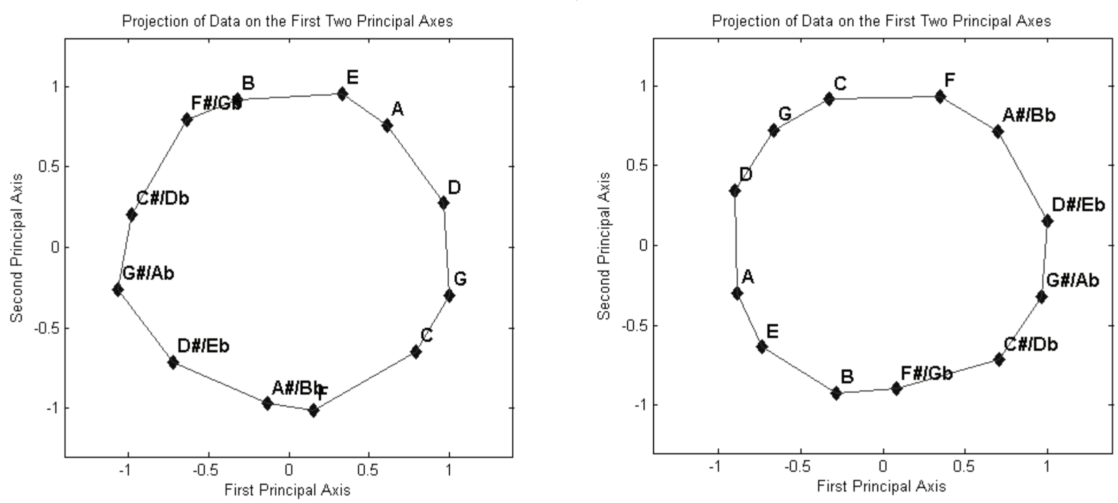

Figure 1.2. (a) *Piano: major scale.* ***(b)*** *Violin: major scale.*

Next, the model was tested on synthetically generated spectra. Figure 1.3(a) shows the results for the major diatonic set using spectra that have 20 harmonics with 12 dB decay per octave in their spectral envelope for each note. Fundamental frequencies are assigned on an equal-tempered scale. The total variance of the first two components was 82.3%. Figure 1.3(b) shows the results when the same spectra were used with the major and harmonic minor diatonic sets (V1 = 28.1%, V2 = 27.5%, V3 = 9.9%, V4 = 9.7%, and V5 = 8.0%).

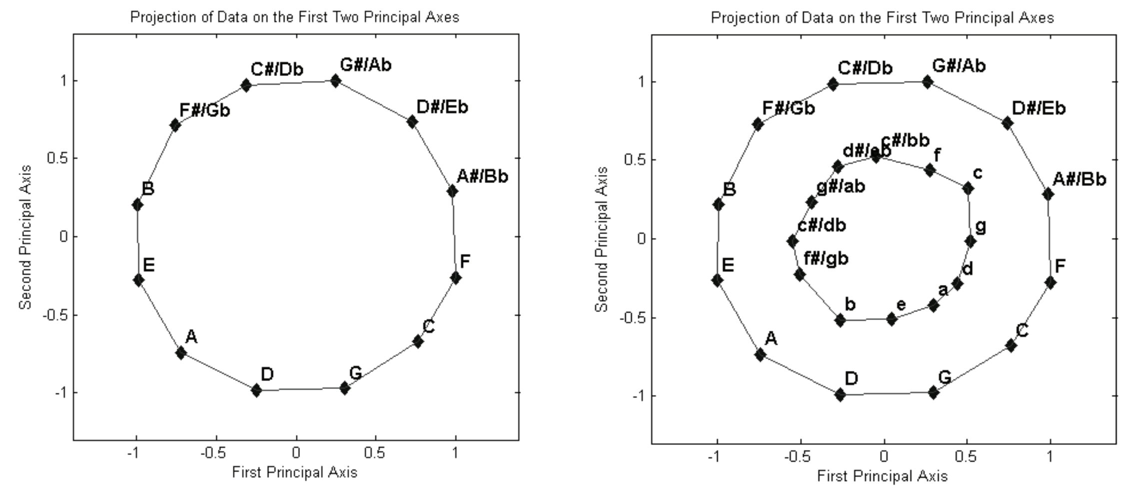

Figure 1.3. (a) *Synthetic spectra: major scale.* ***(b)*** *Synthetic spectra: major and harmonic minor scales.*

Figure 1.4(a) shows the arrangement of the prototypes for the piano when the harmonic minor prototypes are added to the major prototypes. The distribution of the variance for the first five principal components was 32.5%, 31.9%, 8.5%, 7.8%, 6.3% and 5.1%. This shows that although the first two principal components were large compared to the remaining components, part of the information was not represented accurately due to the forced projection onto the first two dimensions. Figure 1.4(b) shows the arrangement for violin sounds. The distribution of the variance for the first five principal components was V1=31.1%, V2=29.8%, V3=9.3%, V4=7.0%, and V5=6.3%. The harmonic minor prototypes formed a concentric smaller circle with respect to the circle of fifths for major prototypes (similar to Kellner's regional circle but rotated half a step; see Lerdahl [2001] for Kellner's circle). The irregularity of the spacing of the minor keys is most likely due to the fact that the PCA was not able to find a well-suited two-dimensional representation.

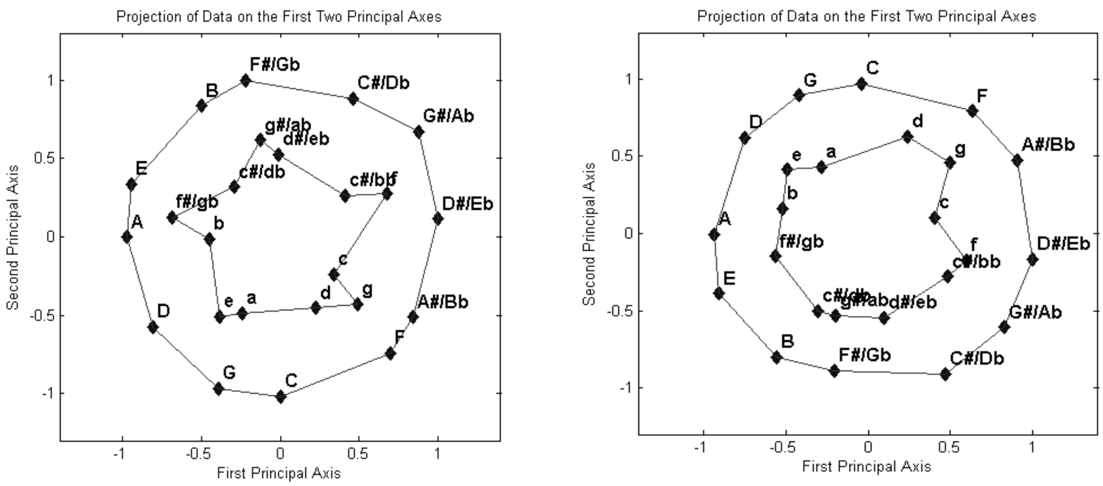

Figure 1.4. ***(a)*** *Piano: major and harmonic minor scales.* ***(b)*** *Violin: major and harmonic minor scales.*

Next, the robustness of the cyclic patterns against pitch additions and deletions were tested. In general, it was found that omitting one or more tones or adding extra chromatic tones would lead to a collapse of the circular output pattern. First, the major diatonic set was altered to observe the resulting geometrical arrangement. Figure 1.5(a) shows the results when a minor third and a minor sixth are added to a major scale, bringing the total number of notes in the pitch collection to 9. In this case, the notes contributing to the set $\mathbf{D}_0$, for example, are C-D-D#/Eb-E-F-G-G#/Ab-A-B with the intervallic pattern WSSSWSSWS where W represents a whole tone and S represents a semitone interval. The total variance accounted for in the first two principal components was 36%, which means that the figure only demonstrates one particular projection and does not reveal the relationship in higher dimensions. Figure 1.5(b) shows the result when an additional diminished fifth is added to the set in part (a) making the set C-D-D#/Eb-E-F-F#/Gb-G-G#/Ab-A-B and the intervallic pattern WSSSSSSSWS. Similar to the previous example, only 38% of the variance was accounted for in the first two principal components.

Finally, the effect of pitch-class weighting was explored. This was done by constructing a 12-element profile, which contains the weights for each pitch-class, and using this profile in the calculation of the spectral prototypes explained above. The reason for inclusion of this test is that many key-finding models utilize profiles in some form (see, for example, İzmirli [2005]). In this context, the major scale can be thought to be an on/off-type profile of the form [1 0 1 0 1 1 0 1 0 1 0 1]. This means, for example in the C major prototype, that the spectra of white keys will contribute with equal weights, whereas black keys will not contribute at all. In general, every pitch-class can be assigned a value indicating the relative weight of that pitch-class.

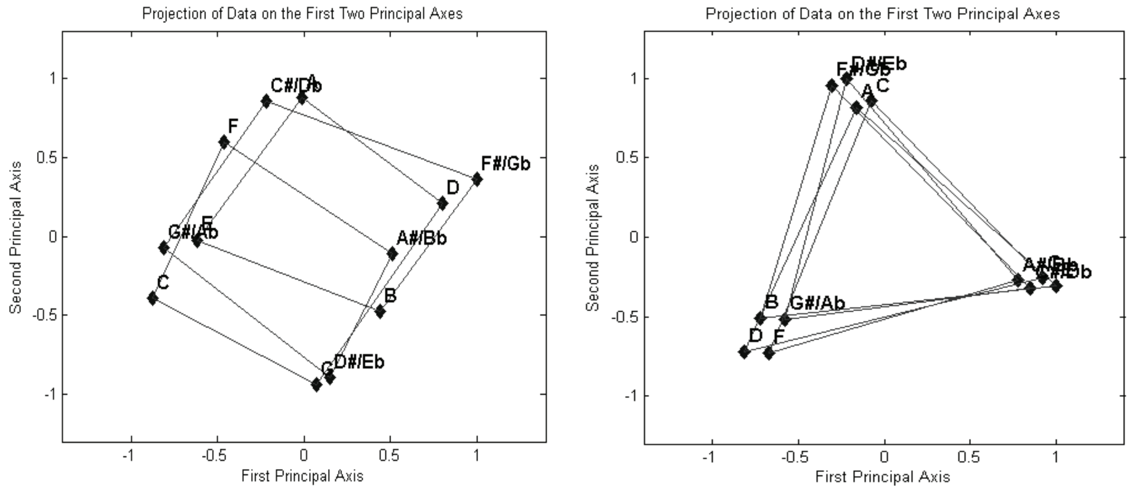

Figure 1.5. ***(a)*** *Minor third and minor sixth are added to the major scale.* ***(b)*** *Diminished fifth is added to part (a).*

This concept stems from the findings of Krumhansl, who suggested that tonal hierarchies for Western tonal music could be represented by the probe-tone profiles found experimentally (Krumhansl 1982). In order to observe the effect of pitch-class weighting, three different profiles were used: Krumhansl (Krumhansl 1990), Temperley (Temperley 2001) and random profiles. Figure 1.6(a) shows the projection using Krumhansl's profile (variances for the 5 strongest components: 33.7%, 30.1%, 9.2%, 8.9% and 6.3%), and Figure 1.6(b) shows the projection for Temperley's profile (variances: 34.7%, 31.1%, 9.6%, 9.2% and 5.7%). From these figures it may seem that the circular form is not affected by the weighting of pitch-classes. However, when random profiles are used the circular form is not preserved. Figure 1.7(a) shows one example of this case (variances: 21.5%, 17.3%, 13.3%, 11.2%, and 8.6%). Figure 1.7(b) shows the randomly generated profile that resulted in Figure 1.7(a). Figure 1.7(c) shows the means of variance contributions by principal component for 500 different profiles. For reference, Figure 1.7(d) shows the variance contributions for the unweighted major scale given in Figure 1.2(a).

1.5 Discussion

When the diatonic sets are derived using only the major or both the major and harmonic minor scales, the circularity arises from the fact that all prototypes that are a fifth apart have the same distance to each other. Furthermore, the loop is closed because one ends up at the same pitch after moving up 12 steps in fifths. The prototypes are ordered in fifths because the closest two prototypes, in terms of their common notes and hence their spectra, are those that are a fifth apart. This however is not true for any arbitrary pitch set. As shown above, when extra notes are included in the diatonic sets the circular distribution is no longer present. The results shown in Figure 1.5 are due to such inputs, and in both cases the output is not viable, at least in two dimensions. Nevertheless, groups of three in part (a) and groups of four in part (b) hint at the fact that there are systematic and cyclic distance patterns among the prototypes of these altered diatonic sets.

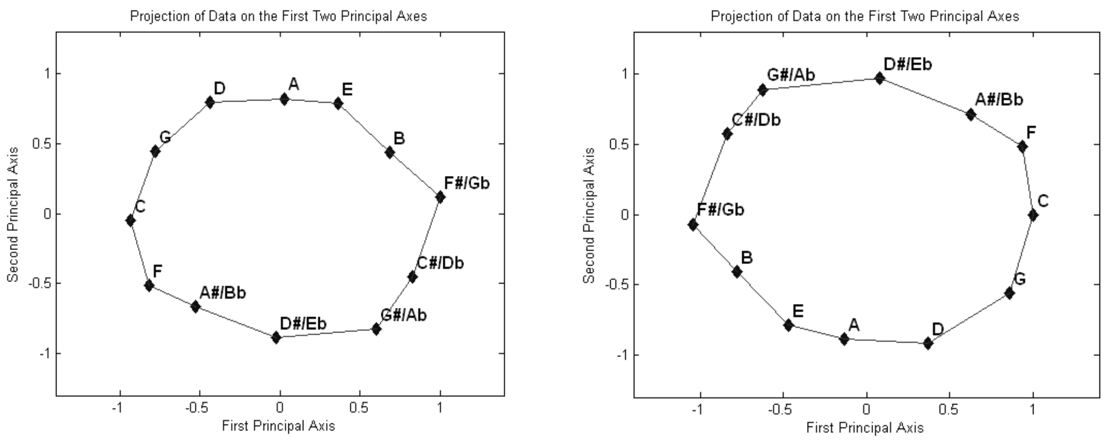

Figure 1.6. (a) *Major scale with Krumhansl's profile.* ***(b)*** *Major scale with Temperley's profile.*

The method works well with the 12 prototypes corresponding to the diatonic sets that employ the major scale. However, when the prototypes for the harmonic minor mode are added to the input, the degree to which the data is explained in two dimensions somewhat drops. Even so, the formation of the two concentric circles is interesting and worth mentioning. The circle that contains the prototypes of the harmonic minor mode has a smaller radius—merely because of the systematic difference in the magnitude of the spectra between the major and the minor prototypes. If the notes of the natural minor had been used, then the points representing the major and the minor sets would have overlapped. The two circles are related to each other by a constant relative angular displacement. For example, A harmonic minor always resides between C major and G major. This is because it has the most common tones (six) with C major and the G# note in this diatonic set rotates it slightly towards the region with more sharps.

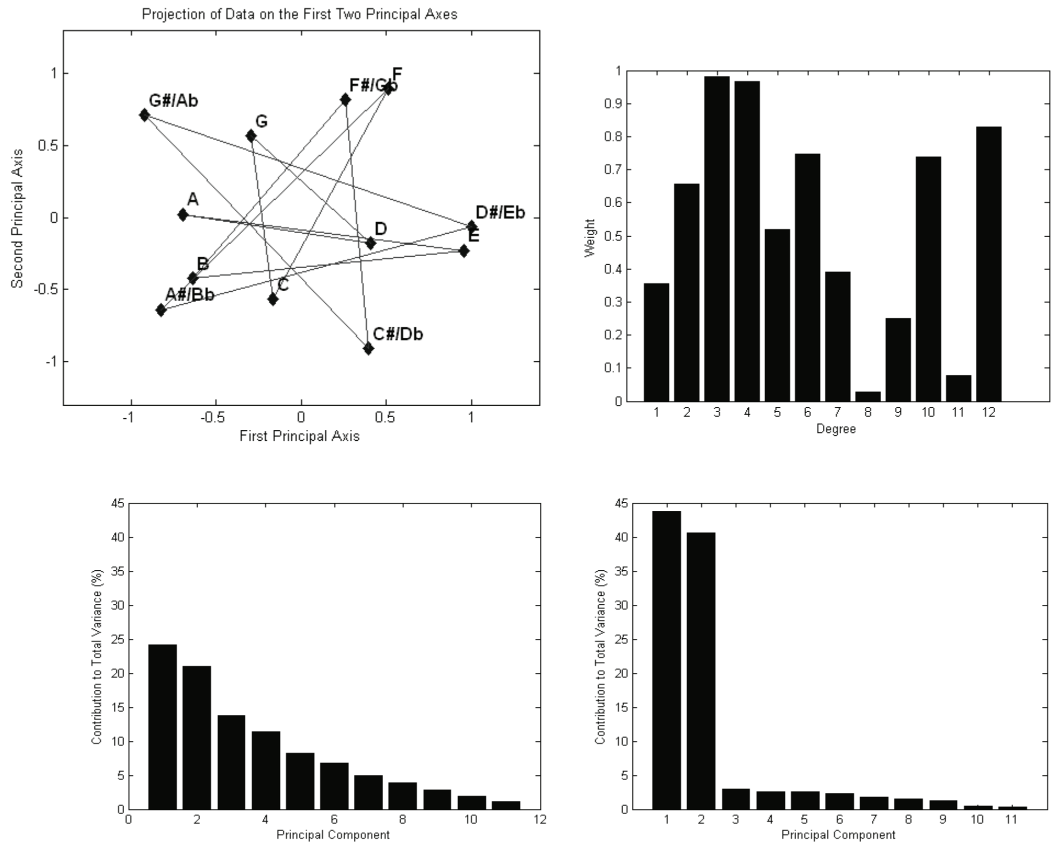

Figure 1.7. ***(a)*** *Major scale with random profile.* ***(b)*** *Profile used in (a).* ***(c)*** *(bottom left) Mean values of variance contributions of principal components for 500 random profiles.* ***(d)*** *(bottom right) Variance contributions of principal components for the unweighted major scale.*

The results shown above for synthetic spectra were obtained using harmonic tones with a decaying spectral envelope. Other tests were also carried out by varying the number of partials, the level of harmonicity (random deviations from ideal harmonic frequencies), stretching of partial frequencies (as in the piano), and changing the rate of spectral envelope decay within reasonable limits. These variations did not seem to have significant effects on the circularity of the output pattern but did cause some distortions with respect to the ideal case.

The high dimensionality of pure spectral representations and their strong dependence on register are sometimes viewed as disadvantages in the context of tonality induction. Chroma-based spectral representations (e.g., Fujishima 1999) constitute a low-dimensional alternative to pure spectral representations. These are found by dividing the spectrum into chroma regions and mapping the entire spectral content belonging to the same chroma into one output bin. The method explained in this paper was also tested using a chroma-based representation by mapping the spectra to pitch-class profiles as given in Fujishima (1999). Again a similar circular structure was obtained. This shows that the information captured by a chroma representation contains the differences necessary to distinguish between tonal centers in a way that preserves their cyclic order.

Results showing the effect of pitch-class weighting reveal that the two-dimensional circularity is dependent on the weight distribution of the profile used. The incorporation of profiles allows for a generalization of pitch-class weighting, and the results delineate the importance of tonal hierarchies used in tonal representations. It can be seen that random profiles do not necessarily result in a circle-of-fifths pattern. The flat on/off diatonic and the two profiles used can be viewed as special cases in which the circle-of-fifths pattern is attained. Figure 1.7 shows that the average distribution of principal component variances obtained by using random profiles does not display any pattern that suggests the suitability of a two-dimensional representation. It should be noted that the distribution given in Figure 1.7(c) includes those profiles that resemble the special cases that work well. However, even with those contributions the overall distribution is spread over many components. This is in contrast to the distribution in Figure 1.7(d) that shows the prominence of the first two components and the sharp drop-off thereafter.

1.6 Conclusions

The model described in this paper demonstrated the emergence of a circle-of-fifths arrangement of keys when spectra of musical instruments were used. Using a structural approach to tonality, the model calculates prototypes by accumulating diatonic collections of spectra taken from real instrument sounds or ideal spectra. Dimensionality reduction is applied to the prototypes in order to obtain a visualization in two dimensions. It is shown that cyclic patterns emerge under a range of conditions typical of musical sounds and commonly used pitch sets. This implies that cumulative spectral patterns of diatonic sets carry distinct information regarding their corresponding keys, and suggests that the human key inference might be mediated by a similar mechanism. Inadequacies of the model are the same as those of the circle-of-fifths representation, mainly the inability to represent parallel minor relationships.

References

Chew, Elaine (2000). *Towards a Mathematical Model of Tonality*. Ph.D. Dissertation, MIT, Cambridge, MA.

Chuan, C., and Elaine Chew (2005). “Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm,” *Proceedings of the International Conference on Multimedia and Expo*, Amsterdam, Netherlands.

Fujishima, T. (1999). “Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music,” *Proceedings of the International Computer Music Conference* (Beijing, China), 464–7.

Gómez, E. (2006). "Tonal Description of Polyphonic Audio for Music Content Processing," *INFORMS Journal on Computing, Special Cluster on Computation in Music*, 18(3), 294–304.

Huron, David, and R. Parncutt (1993). "An Improved Model of Tonality Perception Incorporating Pitch Salience and Echoic Memory," *Psychomusicology* 12/2, 154–71.

İzmirli, Özgür (2005). "Template-Based Key Finding From Audio," *Proceedings of the International Computer Music Conference* (ICMC), Barcelona, Spain.

İzmirli, Özgür (2006). "Audio Key Finding Using Low-Dimensional Spaces," *Proceedings of the 7th International Conference on Music Information Retrieval* (ISMIR2006), Victoria, BC, Canada.

İzmirli, Özgür, and S. Bilgen (1996). "A Model for Tonal Context Time-Course Calculation from Acoustical Input," *Journal of New Music Research* 25/3, 276–88.

Krumhansl, Carol, and E. Kessler (1982). "Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys," *Psychological Review* 89, 334–68.

Krumhansl, Carol L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.

Leman, Marc. (1995). *Music and Schema Theory*. Berlin: Springer-Verlag.

Lerdahl, Fred (2001). *Tonal Pitch Space*. New York: Oxford University Press.

Pauws, Steffen (2004). "Musical Key Extraction from Audio," *Proceedings of the 5th International Conference on Music Information Retrieval* (ISMIR), Barcelona, Spain.

Purwins, Hendrik, Benjamin Blankertz, and Klaus Obermayer (2001). "Constant Q Profiles for Tracking Modulations in Audio Data," *Proceedings of the International Computer Music Conference* (Havana, Cuba), 407–10.

Shepard, Roger N. (1982). "Geometrical Approximations to the Structure of Musical Pitch," *Psychological Review* 89/4, 305–33.

Temperley, David (2001). *The Cognition of Musical Structures*. Cambridge, MA: The MIT Press.

Tillman, B., J. J. Bharucha, and E. Bigand (2000). "Implicit Learning of Tonality: A Self-Organizing Approach," *Psychological Review* 107, 885–913.

Vos, Piet G., and Erwin W. Van Geenen (1996). "A Parallel-Processing Key Finding Model," *Music Perception* 14/2, 185–224.

Submitted: 20 November 2005; 30 August 2007.