

2022-10-13

A clustering-based approach to automatic harmonic analysis: an exploratory study of harmony ...

This work was made openly accessible by BU Faculty. Please [share](#) how this access benefits you. Your story matters.

Version	Published version
Citation (published version):	J. Yust, J. Lee, E. Pinsky. 2022. "A Clustering-Based Approach to Automatic Harmonic Analysis: An Exploratory Study of Harmony and Form in Mozart's Piano Sonatas" Transactions of the International Society for Music Information Retrieval, Volume 5, Issue 1, pp.113-128. https://doi.org/10.5334/tismir.114

<https://hdl.handle.net/2144/46541>

Boston University



A Clustering-Based Approach to Automatic Harmonic Analysis: An Exploratory Study of Harmony and Form in Mozart's Piano Sonatas

JASON YUST 

JAESEONG LEE 

EUGENE PINSKY 

*Author affiliations can be found in the back matter of this article

RESEARCH ARTICLE

]u[ubiquity press

ABSTRACT

We implement a novel approach to automatic harmonic analysis using a clustering method on pitch-class vectors (chroma vectors). The advantage of this method is its lack of top-down assumptions, allowing us to objectively validate the basic music theory premise of a chord lexicon consisting of triads and seventh chords, which is presumed by most research in automatic harmonic analysis. We use the discrete Fourier transform and hierarchical clustering to analyse features of the clustering solutions and illustrate associations between the features and the distribution of clusters over sections of the sonata forms. We also analyse the transition matrix, recovering elements of harmonic function theory.

CORRESPONDING AUTHOR:

Jason Yust

Boston University, US

jason.yust@gmail.com

KEYWORDS:

Harmonic function; clustering; Mozart; harmonic spaces; sonata form; keys

TO CITE THIS ARTICLE:

Yust, J., Lee, J., and Pinsky, E. (2022). A Clustering-Based Approach to Automatic Harmonic Analysis: An Exploratory Study of Harmony and Form in Mozart's Piano Sonatas. *Transactions of the International Society for Music Information Retrieval*, 5(1), 113–128. DOI: <https://doi.org/10.5334/tismir.114>

1 INTRODUCTION

Automatic harmonic analysis has been a goal of music information retrieval and music cognition researchers since the infancy of the fields. The first attempts (such as Winograd, 1968; Maxwell, 1992) drew heavily upon the robust pre-existing music theory assumptions about the goals and methods of harmonic analysis. According to that practice, harmony in tonal music consists of a framework of triads and seventh chords drawn from the basic scales of a series of keys, which follow certain rules of succession. In actual music, such triads and seventh chords might be imperfectly realized, with certain notes missing, or “implied,” and other incidental notes, “non-harmonic tones,” not belonging to the basic framework. Harmonic analysis of notated music is therefore usually understood as a three-stage process of key estimation, identification of non-harmonic tones, and chord labeling according to a lexicon provided by traditional music theory.

Micchi et al. (2020) demonstrate the challenges in distinguishing harmonic from non-harmonic tones and show that there is often not a clear ground truth concerning such distinctions. The identification of non-harmonic tones is sometimes an explicit stage in automatic harmonic analysis algorithms and sometimes implicit in the choice of a limited set of chord types in a predetermined harmonic lexicon, as in Pardo and Birmingham (2002).

There are many studies that use unsupervised methods, such as neural networks and hidden Markov models (HMMs) to perform harmonic analysis and non-harmonic tone identification tasks (Raphael and Stoddard, 2004; Ju et al., 2017; Chen and Su, 2018, 2019). Typically, however, these define restricted chord lexicons for these tasks, and require training data that reflect the norms of Roman numeral analysis. As Devaney et al. (2015) demonstrate, such human annotations display a considerable amount of variability.

Research on harmonic identification from music audio also typically requires a given chord lexicon. The elements of these chord vocabularies are usually expressed as binary pitch-class vectors, the equivalent of pitch-class sets (Pauwels et al., 2019). Deng and Kwok (2016) critique the limited chord lexicon of earlier studies and promote an expanded lexicon. McFee and Bello (2017) demonstrate some of the challenges of applying machine learning approaches with this kind of expanded lexicon. In studies of popular music, these lexicons represent chord labels used by musicians in practice, unlike in studies of Western art music where chord labels represent an analytical practice, not a compositional one. Nonetheless in both cases these lexicons derive from music theory, rather than empirical evaluations of musical corpora.

In another large body of research, the main goal is characterizing harmonic successions. In such contexts researchers usually avoid the need for automatic harmonic analysis through the use of human annotators. This method has been used for Bach chorales (Jacoby et al., 2015; Rohrmeier and Cross, 2008), textbook examples of common-practice tonal harmony (Temperley, 2001, 2009), Mozart’s piano sonatas (Tymoczko, 2011; Henschel et al., 2021), Mozart and Beethoven piano variations (Devaney et al., 2015), Beethoven’s string quartets (Moss et al., 2019), and popular music (Harte et al., 2005; Burgoyne et al., 2013; DeClercq and Temperley, 2011; Temperley, 2018; Temperley and DeClercq, 2013), and comparative analysis of multiple such corpora (Sears and Forrest, 2021). An exception is Tompkins (2017), who studies a seventeenth-century guitar corpus in which chord labels are directly notated as part of the musical practice.

Another strategy that has been used for simplifying the process of harmony identification in order to study norms of harmonic succession is to focus exclusively on music with simple textures, especially the chorale repertoire (Conklin, 2002; Tymoczko, 2003; Rohrmeier and Cross, 2008; Quinn and Mavromatis, 2011; Jacoby et al., 2015; White and Quinn, 2016; Ju et al., 2017). Because of the relatively homophonic and rhythmically simple nature of chorales, with four voices essentially always sounding, the task of harmonic identification becomes much more manageable. However, there are also dangers to this approach: the chorale repertoire is small and highly idiosyncratic, and there is reason to believe that aspects of harmonic usage in it would not generalize to other music. The voice-leading constraints and rapid harmonic rhythm of chorales, for instance, are not shared by most contemporaneous music.

In most empirical research on harmony, then, chord lexicons of traditional triads and seventh chords serve as prior assumptions backed by music theoretic traditions rather than empirical verification. A smaller body of research exists that empirically tests such lexicons by making fewer theoretically freighted initial assumptions about what kinds of harmonic objects qualify as chords, inspired by Quinn (2010). Quinn and Mavromatis (2011) define chords as intervals above a bass in an HMM analysis of harmonic succession in two chorale corpora, while White and Quinn (2018) define chords as pitch-class sets in their HMM analysis of the Bach chorales. These methods have the virtue of deriving the elements of the harmonic language directly from the musical data. The results suggest that the harmonic language of chorales consists mostly of triads and seventh chords, as we would expect, but not exclusively so. However, in extending these results to more complex textures, the problem of excessively large chord vocabularies requires some more sophisticated reduction technique. White

(2013a, b) has extended the approach to more texturally diverse music by means of a harmonic “spell-checking” procedure which simplifies the problem by making some relatively neutral assumptions and bootstrapping on a large dataset.

In the present study, we explore a different method of investigating chord identity and succession in a corpus without prior assumptions about the chord vocabulary. Unlike the HMM-based approaches just described, which infer chord identities from norms of harmonic succession, we use a clustering-based approach which makes no inferences from chord succession, but instead requires a fixed temporal segmentation relying on metrical conventions of the repertoire. This clustering process may be seen as an alternative to a filtering process as applied by Sears and Widmer (2021), with a potential advantage being the possibility of grouping chords with similar, but not equivalent, pitch-class content into a single category.

The primary distinguishing strategy of this approach is to represent the basic harmonic object not as a simple set of pitches or pitch-class sets, but as a twelve-place vector that assigns a weight to each pitch class. This kind of mathematical object, known as a *pitch-class vector* or “chroma vector,” is standard for representing keys in music cognition research (Aarden, 2003; Krumhansl and Cuddy, 2010; Lieck and Rohrmeier, 2020; Temperley, 2007; Sapp, 2005, 2011), and as a feature in many studies of automated harmonic analysis of audio (Pauwels et al., 2019). Pitch-class vectors have been used to represent harmony by Duane and Jakubowski (2018), whose clustering method is similar to the one used here. In the model of a musical key, these weights represent either the likelihood of (or frequency of) occurrence of that pitch class in the given key, or its perceived stability in the key. These turn out to be very similar, though not quite equivalent (Huron, 2006). Pitch-class vectors discard register, voicing, and spelling, and therefore an analysis based on them may miss aspects of harmonic function that depend on these parameters. By the same token, it can show what aspects of harmony generalize over them.

In the present study we take pitch-class counts over quarter-note beats, a timescale associated with harmonies as opposed to keys, and apply *k*-means clustering to these to arrive at a set of harmonic objects. We then investigate the resulting clusters, how their usage varies over formal sections, and whether there are typical patterns of succession between them. We chose a corpus of Mozart piano sonata movements, due to their relative harmonic simplicity, their generally acknowledged canonical status as examples of typical tonal harmonic practice of the later eighteenth century, and their formal predictability. Ultimately, a procedure of

this kind can produce a style-specific chord lexicon that allows for continuous weightings of pitch classes and integrates elements of harmonic function and key into that lexicon.

A basic assumption in traditional theories of tonal harmony is what we might call *level discreteness*. At one level harmony consists of keys, which are made up of chords, which are made up of individual pitch classes. The inadequacy of this three-leveled model is tacitly acknowledged by additional levels that have crept into our theoretical lexicon: “true keys” are distinguished from more local “tonicizations” (secondary or applied chords), and between these “extended tonicizations” are transitory but involve more than a few chords. Similarly, more “structural” harmonies might be decorated by “passing,” “neighboring,” and “embellishing” harmonies, which exist somewhere between true harmonies and collections of non-harmonic tones. The present study suggests a more systematic solution to the problems of the three-level model, using pitch-class vectors to replace the level-discrete model with a level continuum, similar to Sapp (2005, 2011) and Lieck and Rohrmeier (2020). Pitch-class vectors allow for continuous variation between the note and chord and chord and key levels, with note, chord, and key represented by the same type of object.

The present study is primarily exploratory with the most important conclusion being the efficacy of this method. We organize it into three parts, an analysis of the clustering solution (Experiment 1), an analysis of the distribution of clusters across parts of the sonata forms (Experiment 2), and an analysis of first-order transitions between clusters (Experiment 3). While the results largely confirm traditional music theory, Experiment 1 suggests a chord lexicon that differs substantially from the lists of binary-valued triads and seventh chords that serve as the starting point for most research in automatic harmonic analysis. A relatively small number of triads and seventh chords are needed to classify harmony in the corpus, but the central harmonic functions, tonic and dominant, come in multiple forms. Experiments 2 and 3 show that the same triads can have different functions, identifiable in transition data and the distribution across formal sections, and these distinct functions are actually discernable on the basis of pitch-class content (they are associated with distinct clusters).

We also show that applying the discrete Fourier transform (DFT) on pitch-class vectors produces an effective map of keys and harmonic functions. This method has previously been used to characterize musical keys (Cuddy and Badertscher, 1987; Krumhansl, 1990; Yust, 2017b), as a feature description for musical audio (Harte et al., 2006; Ramírez et al., 2020), and as a tool for computational analysis of musical works

and large corpora (Yust, 2019; Navarro-Cáceres et al., 2020; Harding, 2020, 2021; Chiu, 2021; Viacoz et al., 2022; Harasim et al., 2022). In our Experiment 1 we use the DFT to classify clusters into three basic types (triadic, tetradic, and scalar). In Experiments 2 and 3 we show that it is effective in sorting harmonies by key and function, and thus could serve as a simplified feature representation for this kind of harmonic lexicon.

2 EXPERIMENT 1: CLUSTERING

2.1 METHOD

2.1.1 Corpus

Our corpus consists of seventeen movements from Mozart's Piano Sonatas. These were chosen for the relative simplicity and conventionality of their harmony, and their status as standards of the classical “common practice” norms of harmony and form. They also stand out for a diversity of texture that makes automatic harmonic analysis a difficult task. The harmonic successions of this repertoire also have been previously studied by Tymoczko (2011, pages 228–230) and Jacoby et al. (2015), using a procedure that involves human annotation.

Because our method requires isolating beats, we restricted our dataset to movements with quarter-note beats, meaning meters of 2/4, 3/4, or 4/4. To take advantage of the conventions of sonata form in analyzing the data, we also restricted to movements in major-mode sonata form or closely related forms—one exposition-recapitulation (also known as sonata without development) and one sonata-rondo movement are included. We omitted minor mode pieces because there are very few minor mode pieces in this corpus making a representative sample unavailable. We used data from the Yale Classical Archives (White and Quinn, 2016).¹ Because the rhythmic information in this corpus is not always reliable (DeClercq, 2016), we checked each of these movements manually and retained only those where the quarter-note beats consistently line up with integer offset numbers. The movements included are given in Table 1. We transposed all pieces to C major before analysis.

The method of parsing scores by quarter notes is not ideal because the quarter note might have a different meaning at different tempos and in different movement types. However the present corpus is fairly uniform and mostly consists of allegro-type first movements, so we expected this convenient parsing to be reasonably effective. In future research it would be worth exploring coupling our procedure here with automatic parsing methods.

	PIECE	KEY	QUARTER NOTES	METER	MEASURES
0	K279 i	C major	546	4/4	136
1	K279 ii	F major	306	3/4	102
2	K279 iii	C major	426	2/4	213
3	K280 i	F major	593	3/4	198
4	K282 i	E♭ major	276	4/4	69
5	K283 i	G major	356	3/4	119
6	K284 i	D major	1010	4/4	252
7	K309 i	C major	1174	4/4	294
8	K311 i	D major	584	3/4	195
9	K330 i	C major	719	2/4	360
10	K330 iii	C major	703	2/4	352
11	K332 i	F major	665	3/4	222
12	K332 ii	B♭ major	160	4/4	40
13	K333 i	B♭ major	658	4/4	164
14	K545 i	C major	267	4/4	67
15	K570 i	B♭ major	600	3/4	200
16	K576 ii	A major	200	3/4	67

Table 1 Dataset. All pieces are piano sonata movements by W.A. Mozart.

2.1.2 Clustering Method

The *k*-means algorithm is an iterative algorithm that divides a data set into *k* predefined distinct, non-overlapping subgroups which are called “clusters” where each data point belongs to only one group. It is a type of unsupervised learning that minimizes sum of variances within each cluster, through the following procedure. First, define *k* centers, one for each cluster, and assign each data point from the dataset to the nearest center. Then re-calculate *k* new centroids based on these cluster assignments and reassign data points to clusters, minimizing distances to these new centroids. This step is repeated until the clustering is stable.

We applied the *k*-means algorithm to the total set of pitch-class vectors, one for each quarter note of each of the 16 pieces in the corpus.² The weights of each pitch class indicate the number of distinct octaves in which that pitch class appears within the beat. This method of weighting is relatively simple and avoids undue influence from textural factors such as repetitive attacks on the same pitch (e.g., in a trill) or staccato articulation (Temperley, 2007). Octave doubling has been shown to be a reliable proxy for the harmonic importance of a pitch-class (Huron, 1993). Figure 1 illustrates how the procedure works on the first two measures of Piano

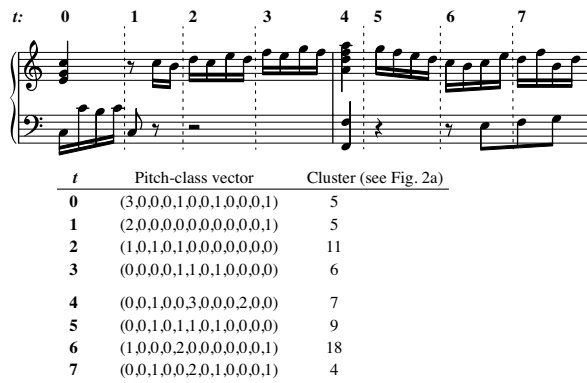


Figure 1 Reduction and clustering procedure illustrated on mm. 1–2 of K.279.

Sonata no. 1, K.279. Each beat is reduced to a 12-place vector as shown. The resulting cluster indices come from the Manhattan clustering solution described in the next section.

The k -means method requires a method of defining distance between two vectors. We ran the algorithm with two different methods for comparison, Manhattan and Euclidean. The Manhattan metric simply sums the absolute differences of pitch-class weights. The Euclidean metric is a 12-dimensional Euclidean distance (the square root of the sum of squared differences).

To find an optimal number of clusters, we used the within cluster sum of squares (WCSS) method, which sums the squared distance of all the points within a cluster to the cluster centroid. As the number of clusters increases, the WCSS value of the model decreases. We determined that 20 is an optimal number of clusters by plotting the WCSS by the number of clusters and finding the inflection point of the elbow-shaped curve.

2.1.3 Evaluation using discrete Fourier transform

We used the *discrete Fourier transform* on pitch-class vectors (Amiot, 2016) in order to systematically sort the clusters into triadic, scalar, and other types. The DFT converts a pitch-class vector into a vector of twelve complex-valued DFT coefficients, of which the zeroth (\hat{a}_0) simply sums the weights and the last five duplicate the information of the first through the sixth (\hat{a}_1 – \hat{a}_6). The indexes of these coefficients denote equal divisions of the octave. Their magnitudes are transposition independent and indicate how strongly the vector weights that division of the octave, and the transposition-dependent phase values correspond to the nearest transposition of that division of the octave. The fifth coefficient, \hat{a}_5 , corresponds to the weighting of the pitch-class vector on the circle of fifths, what can be called its *diatonicity* (similarity to a diatonic scale or typical diatonic subset). This is an important property of tonal harmony. Similarly, \hat{a}_1 gives a weighting on the pitch-class circle. Other important coefficients, \hat{a}_3 and \hat{a}_4 , indicate when a vector is concentrated around some division of the pitch-class

circle by three or four respectively, and so are useful for identifying triads and seventh chords. Previous research (Yust, 2017b, 2019; Bernardes et al., 2016) has shown that two dimensions of the DFT, \hat{a}_3 and \hat{a}_5 , are effective in estimating the key of passages of tonal music and sorting harmonic functions. Because typical pitch-class profiles of major and minor keys have most of their energy in \hat{a}_3 and \hat{a}_5 , a two-dimensional space on the phases of these, denoted φ_3 and φ_5 , can serve as a map of key relatedness (Krumhansl, 1990; Yust, 2017b).

We sorted the clusters using their DFT *spectra*, the magnitudes of the six independent Fourier coefficients, to determine whether certain harmonic qualities were significant across all clusters, or to a specific subset of clusters. We began by finding the pair of DFT coefficients that most frequently appear as one of the three largest for each cluster. We predicted that \hat{a}_5 and \hat{a}_3 , the coefficients that dominate pitch-class profiles for keys (according to Aarden, 2003; Cuddy and Badertscher, 1987; Krumhansl and Cuddy, 2010; Sapp, 2011; Yust, 2017b, 2019) would also be principal qualities for most of the cluster centroids. For the remaining clusters which did not have these two coefficients among the top three, we found another pair that accounted for most of these. Finally, we found a third pair of coefficients that best characterized the remaining clusters and sorted all clusters into three groups on this basis. With the clusters sorted into three groups, we then plotted them in two-dimensional toroidal phase spaces using the phases of the relevant parameters. These spaces are analogous to Krumhansl's (1990) tonal space.

2.1.4 Hierarchical clustering

For low-dimensional data, it is possible to visualize clusters using coordinates, but this is difficult for 12-dimensional data. We used hierarchical clustering to produce a dendrogram relating the clusters obtained by the k -means procedure. Specifically, we applied the aggregation approach with Ward's minimum variance. The distance between clusters A and B is determined by how much the sum of squares increases when merging:

$$\Delta(A, B) = \sum_{i \in A \cup B} \|\bar{x}_i - \bar{m}_{AB}\|^2 - \sum_{i \in A} \|\bar{x}_i - \bar{m}_A\|^2 - \sum_{i \in B} \|\bar{x}_i - \bar{m}_B\|^2$$

$$= \frac{n_A n_B}{n_A + n_B} \|\bar{m}_A - \bar{m}_B\|^2$$

where \bar{m}_c is the center of cluster C, n_c the number of vector points in cluster C, and $\Delta(C_1, C_2)$ the merging cost of combining the clusters C_1 and C_2 .

2.1.5 Historical trends

We also examined the possibility that Mozart's usage of harmonies in the different clusters may have evolved

over time, from 1774 to 1788. We ran simple regressions between the dates of pieces and the frequency of each cluster, and found no significant correlations, so we did not pursue this any further.

2.2 RESULTS

Figure 2a shows the twenty centroids that resulted from the clustering using the Manhattan metric, and Figure 2b shows the twenty centroids that resulted from using the Euclidean metric. In each case they are numbered 0–19 in order of the phase of the \hat{a}_5 coefficient of their DFTs, from flat to sharp (see below). Each is given a short-hand name according to the following rules. If a pitch class has a weight greater than the average for that centroid, it appears in the name; if it exceeds one and a third standard deviations for that centroid, it appears at the beginning of the name, otherwise it is in parentheses at the end. This gives an approximation to the pitch-class content good enough to give each of the twenty centroids a unique name.

The two metrics gave similar results. Both include 3–4 clusters for C major and G major triads, two for F major triads, one for an A minor triad, and one or two for D major triads or seventh chords. They also both include 4–6 clusters for C or G major scales or scale segments, and one for a D minor scale or C# diminished seventh. The Manhattan solution also has an F# diminished seventh cluster (0). Overall, this suggests that Mozart's harmony largely consists of I, IV, V, ii, and vii^o/V chords in the home key (C major) and dominant key (G major), and that there

are multiple forms of I and V, distinguishable by the weightings of chord tones and non-harmonic tones.

The analysis of centroid spectra resulted in groups for

- \hat{a}_3 and \hat{a}_5 ,
- \hat{a}_4 and \hat{a}_5 , and
- \hat{a}_1 and \hat{a}_5 .

This confirmed the hypothesis that \hat{a}_3 and \hat{a}_5 would be principal dimensions for most clusters. In particular, \hat{a}_5 appears to be pervasive, reflecting the strong diatonicity of the style, while the other dimensions of harmonic activity, defined by \hat{a}_4 and \hat{a}_1 , may sometimes take precedence over \hat{a}_3 . We will refer to the three groups as *triadic*, *tetradic*, and *scalar* respectively. Since \hat{a}_3 is the coefficient of a function dividing the octave into three parts, it will tend to be large for triads or subsets of triads, hence the term “triadic.” The term “tetradic,” by analogy, refers to a division of the octave into four parts and the fact that \hat{a}_4 will tend to be large for seventh chords or their subsets. Where the groups overlap, we choose the larger of \hat{a}_3 , \hat{a}_4 , or \hat{a}_1 to classify clusters. Only one cluster in each solution does not have \hat{a}_5 in its top three (Manhattan number 0 and Euclidean number 16), and both of these have \hat{a}_4 as the top coefficient, so we classify them as tetradic. Cluster 0 of the Manhattan solution clearly represents a diminished seventh chord and is dominated by \hat{a}_4 . (Cluster 16 of the Euclidean solution is less clear.) Figure 3 gives the spectra of the centroids, the magnitudes of \hat{a}_1 – \hat{a}_6 , divided up into the three groups.

(a)													(b)														
	C	C#	D	E \flat	E	F	F#	G	G#	A	B \flat	B			C	C#	D	E \flat	E	F	F#	G	G#	A	B \flat	B	
0	C(E \flat F#)	1.1	0.0	0.2	0.7	0.1	0.0	0.7	0.3	0.2	0.3	0.0	0.1	0	GB \flat (C#D)	0.2	0.4	0.4	0.2	0.3	0.1	0.1	1.2	0.0	0.2	1.3	0.1
1	GB \flat (C#D)	0.2	0.5	0.4	0.2	0.3	0.1	0.1	1.1	0.0	0.2	1.3	0.0	1	F(AC)	0.5	0.0	0.4	0.0	0.3	2.2	0.0	0.4	0.1	1.1	0.1	0.1
2	F(D)	0.3	0.1	0.8	0.0	0.3	1.3	0.0	0.0	0.4	0.0	0.0	0.3	2	C(EG)	2.4	0.0	0.1	0.1	0.7	0.0	0.1	0.4	0.1	0.1	0.0	0.1
3	FAC	1.4	0.0	0.0	0.0	0.2	1.2	0.0	0.2	0.1	1.3	0.0	0.3	3	CE(DF)	1.8	0.0	0.9	0.1	1.3	1.1	0.0	0.3	0.0	0.2	0.0	0.2
4	FG(BD)	0.1	0.0	0.6	0.0	0.0	1.1	0.1	1.3	0.0	0.1	0.0	0.7	4	DF	0.1	0.1	1.3	0.0	0.2	1.2	0.0	0.2	0.2	0.3	0.0	0.3
5	C(E)	2.3	0.0	0.3	0.0	1.0	0.2	0.0	0.5	0.0	0.2	0.0	0.2	5	CEG	2.0	0.0	0.2	0.0	2.0	0.1	0.0	1.7	0.0	0.1	0.0	0.1
6	FG(EA)	0.3	0.1	0.2	0.0	1.1	1.4	0.0	1.3	0.1	0.9	0.1	0.1	6	EFG(D)	0.3	0.1	1.0	0.0	1.1	1.1	0.0	1.2	0.0	0.2	0.1	0.3
7	FA(D)	0.1	0.2	1.0	0.0	0.3	1.5	0.0	0.1	0.1	1.2	0.0	0.0	7	AC(F)	1.0	0.1	0.2	0.1	0.1	0.4	0.2	0.2	0.1	1.6	0.0	0.1
8	G(CE)	1.5	0.0	0.1	0.0	1.3	0.1	0.1	2.1	0.0	0.3	0.0	0.1	8	EG(C)	0.9	0.0	0.1	0.1	1.1	0.1	0.0	1.3	0.0	0.1	0.0	0.1
9	DF(EG)	0.5	0.0	1.3	0.0	1.2	1.2	0.0	1.2	0.0	0.0	0.0	0.4	9	G(BD)	0.2	0.0	0.5	0.0	0.1	0.2	0.2	2.2	0.0	0.2	0.0	0.7
10	G	0.1	0.1	0.1	0.1	0.2	0.0	0.4	1.9	0.0	0.3	0.1	0.1	10	GA(FB)	0.1	0.1	0.0	0.2	0.2	0.3	0.1	0.6	0.1	0.5	0.0	0.3
11	CE(G)	0.8	0.0	0.1	0.1	1.0	0.1	0.1	0.7	0.0	0.0	0.0	0.1	11	BCD(G)	1.2	0.0	1.3	0.1	0.4	0.2	0.1	0.5	0.0	0.1	0.0	1.1
12	D	0.2	0.2	1.5	0.0	0.3	0.1	0.2	0.3	0.0	0.1	0.1	0.0	12	BC(GA)	1.2	0.0	0.4	0.1	0.1	0.3	0.1	0.5	0.1	0.9	0.0	1.2
13	G(BD)	0.2	0.0	0.9	0.0	0.1	0.1	0.1	2.3	0.0	0.2	0.0	1.2	13	GBD	0.2	0.0	1.9	0.0	0.1	0.2	0.1	1.9	0.0	0.2	0.0	1.8
14	ABC(D)	1.3	0.0	0.7	0.0	0.2	0.2	0.2	0.5	0.1	1.1	0.0	1.1	14	ACE	1.3	0.0	0.3	0.1	1.3	0.1	0.0	0.1	0.2	1.3	0.1	0.3
15	D(GB)	0.3	0.1	2.1	0.0	0.2	0.3	0.1	0.7	0.0	0.1	0.0	1.3	15	D(F#A)	0.3	0.2	1.5	0.1	0.2	0.0	0.6	0.1	0.0	0.3	0.1	0.0
16	A(CE)	0.4	0.2	0.1	0.1	0.6	0.0	0.1	0.3	0.2	1.4	0.0	0.1	16	F#G(ACE)	0.6	0.0	0.2	0.1	0.4	0.0	1.2	1.1	0.1	0.4	0.0	0.1
17	DF#A(C)	0.7	0.1	1.2	0.0	0.2	0.0	1.1	0.2	0.0	1.2	0.0	0.2	17	BD(G)	0.0	0.1	1.2	0.0	0.1	0.2	0.2	0.8	0.0	0.2	0.1	1.2
18	E(GC)	0.6	0.0	0.2	0.1	2.1	0.1	0.1	0.5	0.1	0.2	0.1	0.2	18	DF#A(C)	1.0	0.0	1.2	0.1	0.1	0.1	1.2	0.2	0.1	1.4	0.0	0.4
19	B(DG)	0.3	0.1	0.8	0.0	0.3	0.0	0.1	0.6	0.1	0.1	0.0	1.2	19	E(BCD)	0.3	0.1	0.3	0.1	1.5	0.1	0.1	0.1	0.3	0.1	0.1	0.3

Figure 2 Centroids of the twenty clusters resulting from applying *k*-means with a Manhattan (a) and Euclidean (b) metric.

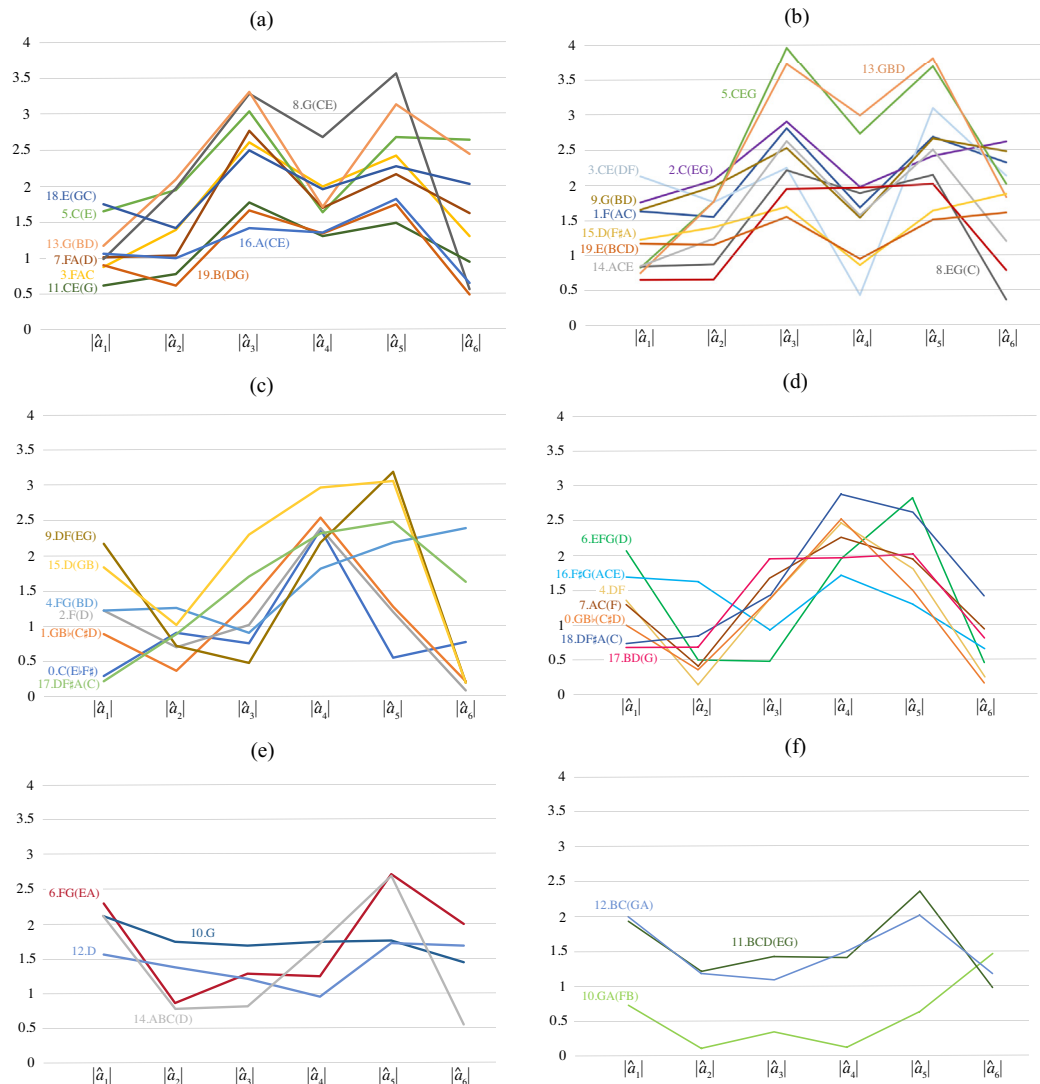


Figure 3 Spectra of the centroids of the 20 clusters, grouped according to whether $|\hat{a}_3|$ (a–b), $|\hat{a}_4|$ (c–d), or $|\hat{a}_1|$ (e–f) is larger, for the k-means solutions by Manhattan (a, c, e) and Euclidean (b, d, f) metrics.

As the name implies, the triadic group captures the clusters that are clearly centered around some major or minor triad. The tetradic group mostly includes clusters near a dominant or diminished seventh chord. The scalar clusters center on contiguous sets of three or four notes from the C major or G major scale and probably represent moments in the music where only a melodic line is present, or the harmony is represented by a single note rather than a complete chord.

Figures 2 and 3 demonstrate that basic features of both k-means solutions are very similar, and for the rest of the paper, we will use the Manhattan solution, since it is somewhat simpler to interpret.

Figure 4 plots the cluster centroids for triadic, tetradic, and scalar clusters in three phase spaces, φ_3/φ_5 , φ_4/φ_5 , and φ_1/φ_5 , respectively (φ_k is the phase of coefficient k) for the Manhattan solution only. Cluster centroids are triadic if \hat{a}_3 is among the top three coefficients in magnitude, tetradic if \hat{a}_4 is among the top three, and scalar if \hat{a}_1 is among the top three, and these groups overlap. In general, the phase value for a given coefficient, φ_k , is more meaningful

if the size of the coefficient $|\hat{a}_k|$ is larger. (Nonetheless any centroid can be plotted in any of the spaces and we will take advantage of this fact in sections 3 and 4.)

The space for triadic clusters, φ_3/φ_5 , is essentially Krumhansl’s (1990) tonal space. Phase values are cyclic, so this space is toroidal: the left edge is glued to the right and top to bottom. The φ_5 dimension separates the clusters by position on the circle of fifths or sharpness and flatness. In all three groups, the values of φ_5 are limited to a narrow region, about 1/3 of the entire cycle. This reflects the overall conservatism of the harmonic style of this corpus and underlines the importance of the diatonic dimension for tonal harmony. In contrast, the centroids spread out fairly evenly in the other dimensions (\hat{a}_3 , \hat{a}_4 , and \hat{a}_1).

The φ_3 and φ_4 dimensions in Figure 4 sort harmonies roughly according to the conventional functional categories of subdominant, dominant, and tonic (as observed by Bernardes et al., 2016; Yust, 2017a, 2019). The spread in these dimensions therefore reflects the representation of all functional categories in each group.

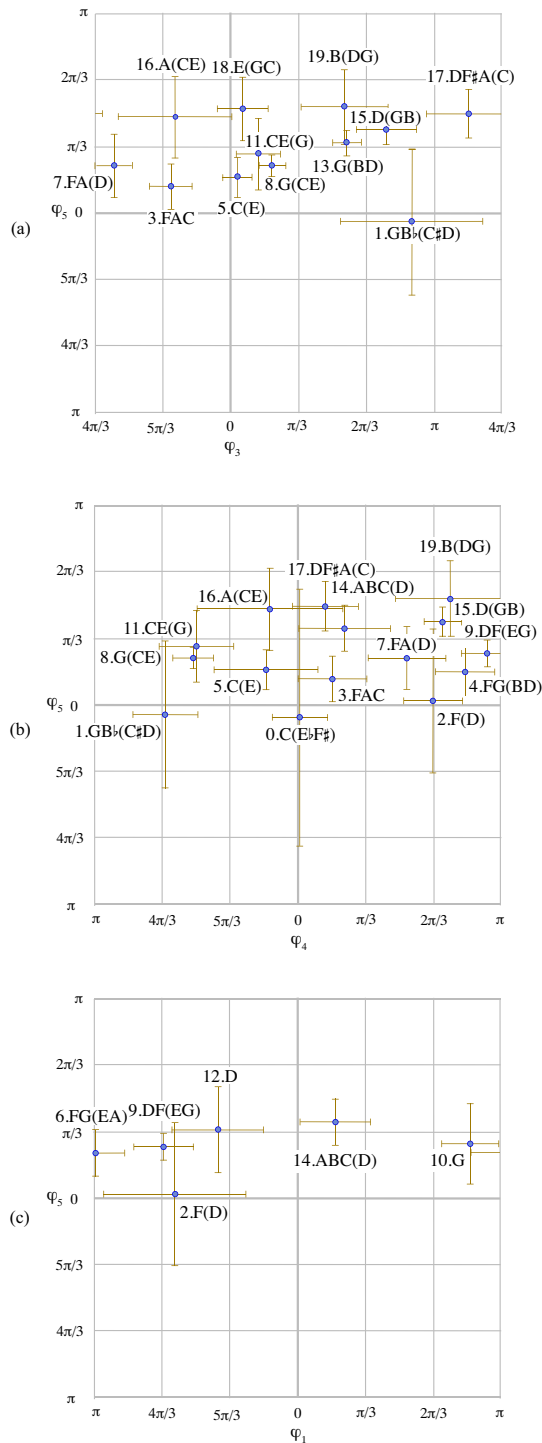


Figure 4 Phase space plot for centroids of (a) triadic, (b) tetradic, and (c) scalar clusters from the Manhattan k -means solutions. The vertical axis is the phase of \hat{a}_s , in all cases, while the horizontal axis is the phase of \hat{a}_3 , \hat{a}_4 , and \hat{a}_1 respectively. Ranges show the circular variance for each cluster:

$$2\pi(1 - \sqrt{\frac{(\sum \text{re}(\hat{a}_k))^2 + (\sum \text{im}(\hat{a}_k))^2}{n \sum |\hat{a}_k|^2}}).$$

The φ_1 dimension represents locations in the octave, which are also relatively evenly represented across scalar clusters.

Figure 5 shows the hierarchical clustering solution. Each group is labeled according to the group centroid, using the same rule as for the cluster centroids in Figure 2. The initial division into four large groups fits a circle-of-thirds logic, as illustrated in the middle of the figure, with

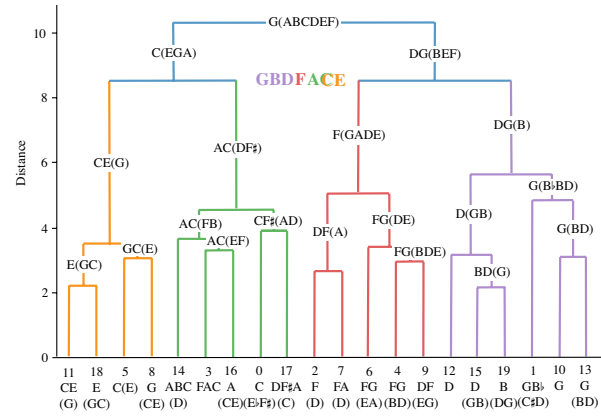


Figure 5 Dendrogram showing hierarchical clustering solution.

each group concentrated in a distinct region of the circle. This is a logical outcome considering that most clusters have most of their weight on a 2–4 note stack of thirds in C major or a closely related key. The main division of the circle of thirds is roughly symmetrical around the tonic, C (between F and A on one side and E and G on the other).

3 EXPERIMENT 2: HARMONIC DIFFERENCES BETWEEN FORMAL SECTIONS

3.1 METHOD

All the pieces included in the corpus except two are in sonata form, and we expected that certain clusters would be characteristic of certain parts of the form since these are characterized by a conventional modulatory scheme. A typical sonata form begins with a main theme in the home key (C major) and then modulates to the key of the dominant (G major) for a subordinate-theme section. This is called the *exposition*. Then there is a tonally ambiguous *development* section likely to include music in the relative minor (A minor). Finally, a *recapitulation* restates main theme material, and subordinate theme material transposed to the home key (C major). Since Mozart's typical practice involves clear conventionalized markers dividing the three main formal sections (exposition, development, and recapitulation) and preceding the subordinate themes within the exposition and recapitulation (the “medial caesura”), we can unambiguously divide each piece into five stages: (1) Main theme (MT) and transition of the exposition, (2) Subordinate theme (ST) of the exposition, (3) Development, (4) Main theme of the recapitulation, (5) Subordinate theme of the recapitulation. These divisions were marked by hand by the first author. We did not attempt to divide the first stage into main theme and transition, because this point of division varies in its clarity and may be located differently by different analysts.

To investigate whether clusters were characteristic of certain formal sections, we recorded the probability of observing each cluster for a given formal section. We then made a correlation matrix of these probabilities

for the 20 clusters and grouped clusters with large ($>.5$) correlations. We then plotted the cluster centroids in φ_3/φ_5 space to investigate possible associations between the pattern of occurrence across formal sections and regions in this space.

3.2 RESULTS

The cluster probabilities across formal sections grouped very clearly into three patterns, shown in Figure 6. The “subordinate key” clusters are more common in expositions, especially exposition second theme areas which are in the key of the dominant (the subordinate key). The “development” clusters are most common in developments. The “home key” clusters are most common in exposition main themes and both parts of the recapitulation. We verified these by running correlations between the patterns for all clusters and grouping any with high correlations ($>.75$). There were only two moderate correlations across groups: clusters 1 and 12 (0.63), and clusters 1 and 19 (0.64). One cluster (18) did not have any large correlations (its highest correlation, 0.42, was with a member of the home key group).

The resulting grouping clearly relates to the standard modulatory scheme of sonata form: clusters typical of the exposition are those associated with the dominant key, and are relatively infrequent in recapitulations. Clusters typical of developments are those associated with common minor keys: A minor (cluster 16), D minor (cluster 1), and G minor (cluster 0). Home key clusters are particularly infrequent in the subordinate themes of expositions, and reflect harmonies characteristic of C major, especially the subdominant (F major) and dominant seventh (G^7).

The grouping of clusters based on formal section is consistent with their positioning in the φ_3/φ_5 space, shown in Figure 7. Clusters with a similar pattern of usage in the different formal sections are in similar regions of the space, corresponding to the main tonic and dominant keys. The subordinate key group, common in the dominant-key subordinate theme, are consistently higher in φ_5 and concentrate in one half of the full φ_3 cycle, with the other half occupied by the home key group. Clusters typical of developments appear in peripheral parts of the space, reflecting the use of minor keys which have a greater φ_5 spread. The use of Cluster 10 in developments probably reflects the conventional concluding dominant pedal.

4 EXPERIMENT 3: TRANSITION PROBABILITIES

4.1 METHOD

A topic of widespread interest in research on eighteenth-century tonal harmony is harmonic succession, which is often described by cognitive music theorists as a kind of syntax (Rohrmeier, 2011; Sears and Widmer, 2021;

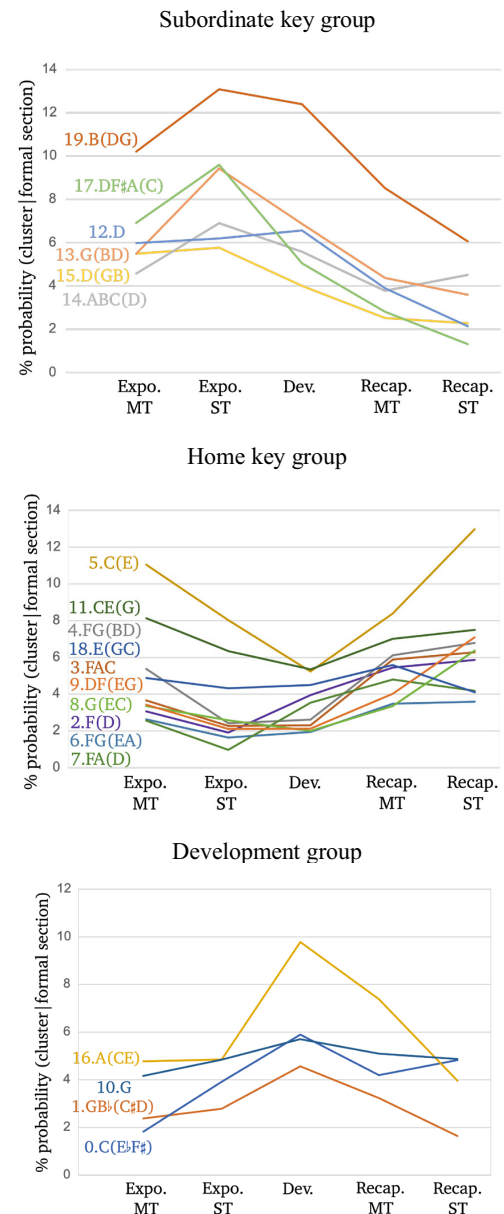


Figure 6 Probability of each cluster appearing in each of five parts of a sonata form, split into three groups.

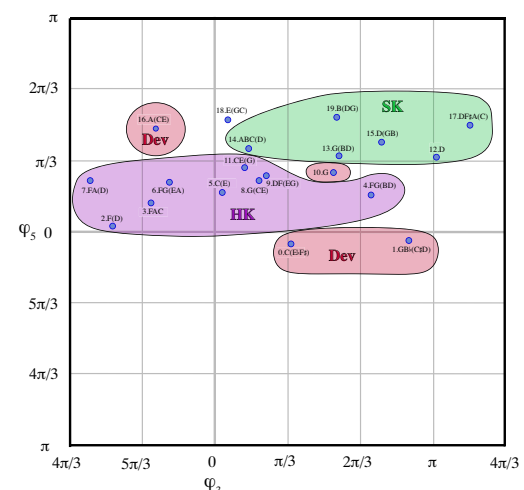


Figure 7 The twenty cluster centroids in the phase space for \hat{a}_5 and \hat{a}_3 , with regions grouping clusters that have a similar frequency profile over the five main sections of a sonata form.

Tymoczko, 2011; White and Quinn, 2018). We analyzed transitions between clusters as a contribution to this body of knowledge.

The complete transition data are a 20-by-20 matrix, giving the probability of cluster j following cluster i . We reduced this space in two ways in order to analyze the data, first dividing up the clusters into the form-based groups from experiment 2, and second, grouping the clusters according to similar transition patterns.

To analyze the transitions in the two form-based groups (the subordinate key and home key groups described above, omitting the development group which contains only four clusters) we constructed smaller transition matrices containing only the clusters in each group. We eliminated diagonals since transitions from a cluster to itself only indicate when harmonies last longer than a single quarter note. We then renormalized so that transitions could be interpreted as percentages within the smaller group.

We converted these to sum and difference matrices by adding and subtracting the transpose. The sum matrices are symmetrical and represent how often representatives of two clusters are juxtaposed, regardless of order. The difference matrix is anti-symmetrical and represents how much more often a representative of cluster i will precede cluster j instead of following it. For the difference matrix we considered only those exceeding 1 standard deviation above the mean for the given matrix. For the sum matrix, we considered values that exceeded the average plus the average standard deviation of their row and column.

In our second analysis we grouped clusters based on similar transition behavior and constructed a transition matrix on the resulting groups. To make these groups we first made a correlation matrix by calculating Pearson's r between each pair of rows in the full 20-by-20 difference matrix. We filtered this to only the correlations exceeding two standard deviations. We then added the rows for correlated clusters together to make a reduced matrix and repeated the process on the new matrix. When we exhausted the two standard deviation criterion, we repeated the process with a one and a half standard deviation criterion and repeated this until we had a relatively small number of groups.

We analyzed the resulting matrix in the same way as the form-based matrices. We eliminated the diagonal, renormalized, and took sum and difference matrices. For the sum matrix, we considered values that exceeded the average of their row and column plus the average standard deviation of their row and column. For the difference matrix, we considered values in excess of one standard deviation.

Finally, for each of these we plot the resulting transitions in a ϕ_4/ϕ_5 space, which we found to be the most suitable for illustrating these results.

4.2 RESULTS

4.2.1 Form-based grouping

Figure 8 shows the sum and difference matrices of the subordinate key and home key groups. Transitions always go from row to column. Figure 9 plots these in ϕ_4/ϕ_5 space, with heavy arrows for large values in both the sum and difference matrices, lighter arrows for values that appear only in the difference matrix, or double-headed for those only in the sum matrix. Typical predominant-dominant (16–17, 3–4, 7–9) and dominant-tonic (17–13,

	12	13	14	15	16	17	18	19
D	G(BD)	ABC(D)	D(GB)	A(CE)	DF#A(C)	E(GC)	B(DG)	
12 D								
13 G(BD)	1.03							
14 ABC(D)	1.85	2.59						
15 D(GB)	1.00	2.22	0.95					
16 A(CE)	1.51	1.16	1.03	0.71				
17 DF#A(C)	1.72	2.96	2.06	1.43	1.40			
18 E(GC)	0.53	0.53	1.00	1.06	1.74	0.45		
19 B(DG)	4.04	2.93	3.25	2.85	3.62	3.33	1.06	

Sum matrix for exposition clusters.

	12	13	14	15	16	17	18	19
D	G(BD)	ABC(D)	D(GB)	A(CE)	DF#A(C)	E(GC)	B(DG)	
12 D								
13 G(BD)	0.48			0.11	0.42			0.79
14 ABC(D)	0.53	1.80						0.90
15 D(GB)	0.00		0.42		0.05	1.48	0.21	
16 A(CE)	0.26		0.26			1.64		
17 DF#A(C)	0.26	1.37	0.32					
18 E(GC)	0.32	0.21			0.32	0.05		
19 B(DG)			0.16	0.21	1.43	0.21	0.53	

Difference matrix for exposition clusters.

	2	3	4	5	6	7	8	9	11
F(D)	FAC	FG	C(E)	FG	FA(D)	CEG	DEF	CE(G)	
		(BD)		(EA)			(BC)		
2 F(D)									
3 FAC	2.32								
4 FG(BD)	2.10	1.71							
5 C(E)	5.03	6.03	7.30						
6 FG(EA)	1.05	2.54	1.16	2.32					
7 FA(D)	1.82	1.16	0.55	2.05	2.71				
8 G(CE)	1.05	1.55	2.76	4.26	1.05	1.22			
9 DEFG(BC)	1.11	0.72	3.48	7.02	2.32	2.16	2.49		
11 CE(G)	3.32	3.37	5.31	7.68	2.16	1.93	2.1	3.1	

Sum matrix for recapitulation clusters.

	2	3	4	5	6	7	8	9	11
F(D)	FAC	FG(BD)	C(E)	FG(EA)	FA(D)	CEG	DEFG	CE(G)	
							(BC)		
2 F(D)				0.61		0.39	0.50		
3 FAC	0.22		1.27		0.22				
4 FG(BD)	0.11			3.65					0.33
5 C(E)		0.06							1.16
6 FG(EA)	0.28		0.28			0.28	0.17		
7 FA(D)		0.06	0.00					1.05	0.28
8 G(CE)		0.33	0.11	0.17				1.05	0.11
9 DEFG(BC)	0.33	0.17	0.28	1.49	0.22				
11 CE(G)	1.00	0.39			0.61		0.11		

Difference matrix for recapitulation clusters.

Figure 8 Sum and difference matrices for subordinate key and home key clusters in percentages over just the transitions involving these clusters and excluding trivial transitions. Values exceeding 1 standard deviation are highlighted. Redundant values (given by symmetry and antisymmetry) are excluded (in difference matrices we retain just the positive values).

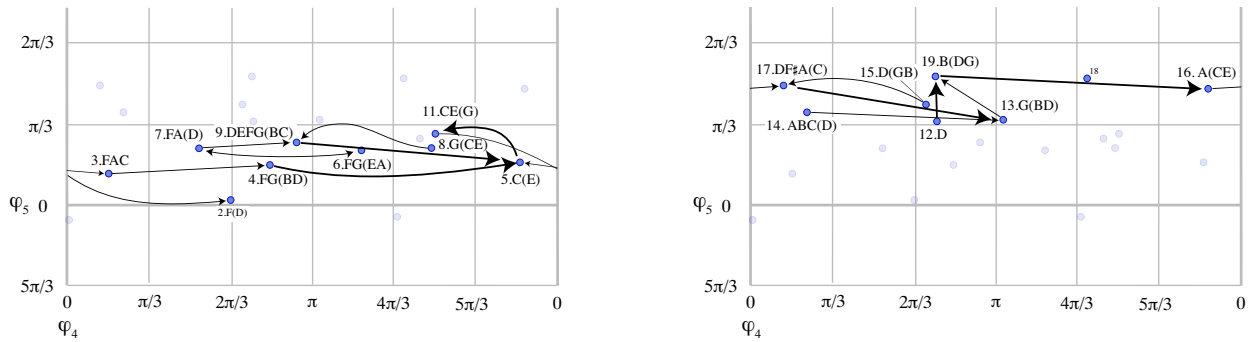


Figure 9 Transition data for subordinate key and home key groups, plotted in ϕ_4/ϕ_5 space. Large arrows show high values in both the difference and sum matrices, small arrows in the sum or difference matrices only (double-headed for sum).

	SK dom	SK pd	D (F)	HK pd	HK dom	HK ton	FG (EA)	SK ton	D	B (DG)
SKdom										
SKpd	7.04									
D(F)	0.76	2.94								
HKpd	1.66	6.06	1.41							
HKdom	1.03	4.60	0.87	2.82						
HKton	2.13	5.15	1.36	3.34	3.88					
FG(EA)	0.57	1.56	0.28	1.71	0.94	0.63				
SKton	5.94	5.66	0.52	2.19	1.75	1.66	0.49			
D	2.53	2.85	0.61	0.78	0.54	0.46	0.30	1.26		
B(DG)	5.27	6.35	0.79	0.99	1.39	1.60	0.22	2.80	2.29	

Normalized sum matrix.

Normalized sum matrix.

	SK	SK		HK	HK	HK	FG	SK		
	dom	pd	D(F)	pd	dom	ton	(EA)	ton	D	B(DG)
SKdom							0.03	1.96	0.13	0.18
SKpd	1.44		0.18	0.70	0.19		0.15		0.00	
D(F)	0.10			0.18		0.16				
HKpd	0.10				0.90			0.06		0.12
HKdom	0.07		0.12			1.39			0.00	0.07
HKton	0.60	1.02		0.07				0.07		0.04
EFG(A)			0.07	0.06	0.01	0.12			0.18	
SKton		0.87	0.01		0.55		0.19		0.30	0.16
D			0.1	0.15		0.13				0.22
B(DG)		0.78					0.07			

Normalized difference matrix.

Normalized difference matrix.

Figure 10 Sum and difference matrices for cluster groups. Values exceeding 1.5 standard deviations are highlighted. Redundant values (given by symmetry and antisymmetry) are excluded (in difference matrices we retain just the positive values).

14–13, 4–5, 9–5) functional successions of the C major and G major keys, directed from left to right, are evident. (The term “predominant” refers to chords that typically lead to the dominant, such as IV, ii, or V/V.) Both keys also include a cadential 6–4 progression, going right-to-left from a fifth-heavy tonic triad to a dominant.

4.2.2 Transition-based reduction

The reduction process for the 20×20 matrix resulted in 6 groups of clusters with 4 clusters remaining ungrouped. The six groups reflect identifiable functions: tonic (ton), dominant (dom), and predominant (pd) functions in the two keys, home key (HK) and subordinate key (SK):

- SKdom: 0.C(EbF#), 14.ABC(D), 17.DF#A(C)
- SKpd: 1.GBb(C#D), 11.CE(G), 15.D(GB), 16.A(CE), 18.E(CG).
- HKpd: 3.FAC, 7.FA(D), 8.G(CE)
- HKdom: 4.FG(BD), 9.DF(EG)
- HKton: 5.C(E)
- SKton: 10.G, 13.G(BD)

The remaining four are ungrouped clusters: 2.D(F), 6.FG(EA), 12.D, and 19.B(DG).

Comparing these groups to the hierarchical clustering solution shown in Figure 5, we find that the tonic and dominant groups combine clusters with similar pitch-class content, but the predominant groups include clusters with disparate pitch-class content.

Figure 10 gives the grouped sum and difference matrices, and Figure 11 shows the resulting transitions in a ϕ_4/ϕ_5 plot. Dotted lines connect the clusters to the group label, except for the cadential 6–4 chords (8 in the HKpd and 15 in the SKpd group) which are not close to the rest of their group. Other than the cadential 6–4 chords, the space effectively sorts the six primary functions: home key below and dominant key above, and the three within-key functions arranged left-to-right such that functional motion always cycles in this direction. As in Figure 9, when groups appear in both the sum and difference matrices, we use larger arrows.

The functional logic is apparent from the patterns in Figure 11, with strong dominant-tonic and predominant-dominant motions in each key. These always go left to right in the space. The only strong motion between the keys is between the predominants, which also matches functional logic (for instance the SKpd group contains chords that could function as vi or vii^o/ii in the home key). The only retrograde motion is the weaker tendency to convert tonic of the SK to a dominant of the HK (e.g., by adding a seventh, F, to a G major triad). Of the four ungrouped clusters (besides HKton), only one makes an appearance in the sum and difference matrices, cluster 19, a third-weighted G major triad. This largely behaves like SKton, except that it has a more symmetrical relationship with SKdom, and therefore remains separate from that group.

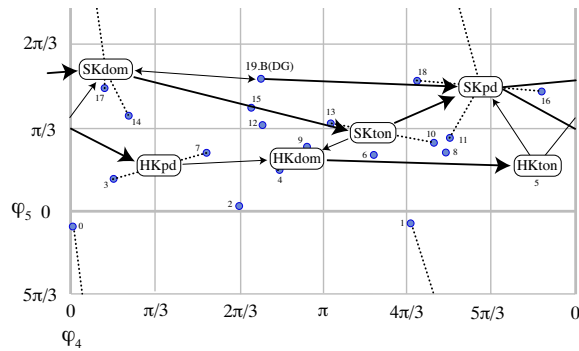


Figure 11 Transition data for the cluster groups plotted in ϕ_4/ϕ_5 space. Group labels are positioned roughly in between their members, which are connected by dotted lines (with the exception of clusters 8 and 15 for clarity). Heavy arrows show high values in both the difference and sum matrices, lighter arrows in the sum or difference matrices only (double-headed for sum).

5 DISCUSSION

Pitch-class vectors, which have been shown by a wealth of previous research to be effective representatives of keys (Krumhansl and Cuddy, 2010) and harmony in music audio (Pauwels et al., 2019), have been shown here to also be potentially effective models of chords, if we are willing to abandon chord lexicons that simply map conventional triads and seventh chords onto simple pitch-class sets (i.e. binary-valued pitch-class vectors).

The present study is primarily preliminary and exploratory, and details of the results are clearly dependent on the chosen corpus. Nonetheless, we can already draw some significant conclusions about efficacy of a weighted pitch-class model of tonal harmony and derive some implications for music theory.

Music theory conventions make a few implicit claims about how tonal harmony works. First, Roman numeral conventions limit the possible harmonic objects to a set of triads and seventh chords drawn from common tonal scales. They also tend to imply that, except for differences of inversion, all such triads and sevenths are unitary objects. Second, as a partial redress to the shortcomings of the Roman numeral conventions, theories of harmonic function typically highlight certain elements of the Roman numeral lexicon as being of special significance, such as tonic and dominant, and will also sometimes assign multiple functions to a single harmonic object (e.g., IV as “subdominant” vs. “predominant”).

In broad strokes, our results support this received music theory in many respects. First, the majority of clusters we found can be reliably associated with triads and seventh chords, supporting the basic Roman numeral convention. Second, we found that a small handful of these triads and seventh chords are very prevalent, while the majority are either not common or distinct enough to be detected by the clustering procedure. On the other hand, for the more important tonic and dominant functions, the clustering

solution distinguished three or four varieties of the same triad. These all behave differently—in no case did our transition-based grouping procedure combine any of the four C major triad clusters, or three G major triad clusters. This generally supports the notion from function theory that the same harmonic object can function in multiple ways, although our data supports this idea for tonic and dominant triads where function theory tends to focus more on multiple functions for other kinds of triads.

This support for music theory conventions, however, is only in broad strokes. When we consider details, our results show that the implications of music theory conventions are imprecise and in some respects may oversimplify the reality.

First, the idea of limiting harmony to triads and seventh chords is largely, but not entirely, supported. In the clusters obtained using the Manhattan metric, we found four “scalar” clusters (Figure 3e). Two of these, 10.G and 12.D, emphasize a single pitch class, while the other two, 6.FG(EA) 14.ABC(D), are approximately four-note scale segments. This suggests that certain single pitch-classes and scale segments are distinguishable as harmonic objects in this repertoire, whereas many hypothetically possible triads and seventh chords are not. Of course, these harmonies may not be “functional” in the sense of having distinct syntactical meaning. In our analysis of transition data, one of them, 14.ABC(D), was assimilated to a subordinate-key dominant group, suggesting that it might be a distinct variety of dominant chord, similar to 4.FG(BD), which could also be described as “scalar” although it had a large enough fourth DFT coefficient to be grouped with the tetradic harmonies in our analysis. The other scalar harmony, 6.FG(EA), was not assimilated on the basis of transition data, and its only significant feature in the transition data was an association with one of the home key subdominant clusters, 7.FA(D). The 12.D cluster, on the other hand, did not assimilate with other groups and did show some reliable behavior in a tendency to move to cluster 19.B(DG).

Second, the results support the idea from function theory that two triadic harmonies, tonic and dominant, are overwhelmingly the most important in this repertoire. They also support the idea of three functions, predominant, dominant, and tonic, one of which (predominant) is based on similarity of behavior, not on similarity of pitch-class content. In the groupings of clusters based on transition data (section 4.2), tonic and dominant categories group chords close together in the hierarchical clustering solution of Figure 5, while the two predominant groups combine clusters that are dispersed throughout the hierarchical clustering solution.

Third, the results support the idea of multiple functions for individual triads, especially for tonic and dominant triads. The clustering solution identified four C major triads (clusters 5, 8, 11, 18), and three G major triads (clusters 13, 15, and 19). None of these were grouped

on the basis of transitions in section 4.2. Some might be associated with distinct inversions, especially numbers 8 and 15, which appear to be cadential 6–4 chords. Our analysis also suggested that cluster 11 represents IV of G major, distinguishing it from the other C major clusters as tonics of C major. Taking this one step farther, then, we might conjecture that clusters 18 and 19 represent first-inversion triads. If this is the case it means that first-inversion triads have distinct function, specifically in the case of tonic triads, but also that these distinctly functioning chords are identifiable purely through registral weighting of pitch classes, without necessarily separating out the bass line. This might be possible due to the dependency of doubling on inversion observable, e.g., in Aarden and Von Hippel's (2004) results, which may be stronger when we isolate tonic and dominant chords.

That these functional distinctions emerge is remarkable since, unlike in a hidden-Markov algorithm, as applied, for example, by Mavromatis (2009), Quinn and Mavromatis (2011), White (2013a, b), or White and Quinn (2018), the clustering algorithm itself has no information about transitions, and only makes these distinctions based on different weightings of pitch classes. It seems likely that our weighting procedure, which depends on appearing in multiple octaves and ignores rearticulations of notes, is crucial to this success in sorting different functions for the same triads.

We have also found that applying the DFT to pitch-class vectors provides an effective dimensional reduction in which aspects of similarity in pitch-class content and harmonic function remain observable. This recommends its use as a feature representation in studies of harmony in notated music, much as Harte et al. (2006) and Ramírez et al. (2020) have applied it to chroma vectors from musical audio. Specifically, we have found the equivalent of Krumhansl's (1990) tonal space, ϕ_3/ϕ_5 space, effective in sorting harmonies by key, and a different ϕ_4/ϕ_5 space useful for observing the norms of harmonic succession, and magnitudes of \hat{a}_1 , \hat{a}_3 , and \hat{a}_4 useful for sorting scalar, triadic, and tetradic chord types. Extrapolating from this, we might conjecture that scalar types (high $|\hat{a}_1|$) are generally incidental and non-functional, triadic types (high $|\hat{a}_3|$) important for defining key, and tetradic types (high $|\hat{a}_4|$) important for local functional succession.

This study suggests further research in a few directions. The corpus we used is relatively limited, and different results could be expected from a larger or more harmonically expansive or varied corpus. The results also depend upon musical texture, so a corpus featuring different kinds of textures (such as more contrapuntal textures) would make a valuable comparison. One limitation of the present study is that the quarter-note parsing used throughout does not necessarily optimally isolate individual harmonies. A method of parsing based on similarity of pitch-class content may greatly improve the clustering method.

NOTES

- 1 <http://ycac.yale.edu>.
- 2 We used python's sklearn library for all machine learning methods. Code is available at https://github.com/peachsky1/ClusteringBasedApproach_HarmonicAnalysis.

ACKNOWLEDGEMENTS

Thanks to the editor and anonymous reviewers for the helpful comments on the drafts, which greatly improved the paper.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Jason Yust  orcid.org/0000-0001-5978-1153
Boston University, US

Jaeseong Lee  orcid.org/0000-0003-2490-2095
Boston University, US

Eugene Pinsky  orcid.org/0000-0002-3836-1851
Boston University, US

REFERENCES

- Aarden, B.** (2003). *Dynamic Melodic Expectancy*. PhD thesis, Ohio State University.
- Aarden, B., and von Hippel, P.** (2004). Rules for chord doubling (and spacing): Which ones do we need? *Music Theory Online*, 10(2). https://mtosmt.org/issues/mto.04.10.2/mto.04.10.2.aarden_hippel.html
- Amiot, E.** (2016). *Music through Fourier space: Discrete Fourier transform in music theory*. Cham: Springer. DOI: <https://doi.org/10.1007/978-3-319-45581-5>
- Bernardes, G., Cocharro, D., Caetano, M., Guedes, C., and Davies, M.E.P.** (2016). A multi-level tonal interval space for modelling pitch relatedness and musical consonance. *Journal of New Music Research* 45(4), 281–294. DOI: <https://doi.org/10.1080/09298215.2016.1182192>
- Burgoyne, J.A., Wild, J., and Fujinaga, I.** (2013). Compositional data analysis of harmonic structures in popular music. In Yust, J., Wild, J., & Burgoyne, J.A. (Eds.) *Mathematics and Computation in Music: Fourth International Conference, MCM 2013*, pages 52–63. DOI: https://doi.org/10.1007/978-3-642-39357-0_4
- Chen, T-P., and Su, L.** (2018). Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *Proceedings of the 19th International Conference on Music Information Retrieval (ISMIR)*, pages 90–97.
- Chen, T-P., and Su, L.** (2019). Harmony transformer: Incorporating chord segmentation into harmony

- recognition. In *Proceedings of the 20th International Conference on Music Information Retrieval (ISMIR)*, pages 259–267.
- Chiu, M.** (2021) Macroharmonic progressions through the discrete Fourier transform: An analysis of Maurice Duruflé's Requiem. *Music Theory Online* 27(3). DOI: <https://doi.org/10.30535/mto.27.3.1>
- Conklin, D.** (2002) Representation and discovery of vertical patterns in music. In C. Anagnostopoulou, M. Ferrand, and A. Smaill (Eds.), *Music and Artificial Intelligence: Second International Conference (ICMAI 2002)*, pages 32–42. DOI: https://doi.org/10.1007/3-540-45722-4_5
- Cuddy, L. L., and Badertscher, B.** (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience. *Perception & Psychophysics* 41, 609–20. DOI: <https://doi.org/10.3758/BF03210493>
- Deng, J., and Kwok, Y.-K.** (2016). A hybrid Gaussian-HMM-deep-learning approach for automatic chord estimation with very large vocabulary. In *Proceedings of the 17th International Society for Music Information Retrieval (ISMIR)*, pages 812–818.
- DeClercq, T.** (2016). Big data, big questions, a closer look at the Yale Classical Archives (c. 2015). *Empirical Musicology Review* 11(1). <https://emusicology.org/article/view/5274>. DOI: <https://doi.org/10.18061/emr.v11i1.5274>
- DeClercq, T. and Temperley, D.** (2011). A corpus analysis of rock harmony. *Popular Music* 30, 47–70. DOI: <https://doi.org/10.1017/S026114301000067X>
- Devaney, J., Arthur, C., Condit-Schultz, N., and Nisula, K.** (2015). Theme and variation encodings with Roman numerals (TAVERN): A new data set for symbolic music analysis. In *Proceedings of the 16th International Conference for Music Information Retrieval (ISMIR)*, pages 278–234.
- Duane, B., and Jakubowski J.** (2018). Harmonic clusters and tonal cadences: Bayesian learning without chord identification. *Journal of New Music Research* 47(2), 143–165. DOI: <https://doi.org/10.1080/09298215.2017.1410181>
- Harasim, D., Affatato, G., and Moss, F.C.** (2022). miVERTO: A web application to visualize tonality in real time. In Montiel, M., Agustín-Aquino, O.A., Gómez, F., Kastine, J., Lluís-Puebla, E., and Milam, B. (Eds.), *Mathematics and Computation in Music: 8th International Conference (MCM 2022)*, pages 363–368. DOI: https://doi.org/10.1007/978-3-031-07015-0_31
- Harding, J.D.** (2020). Computer-aided analysis across the tonal divide: Cross-stylistic applications of the discrete Fourier transform. In de Luca, E. and Flanders, J. (Eds.), *Music Encoding Conference Proceedings 2020*, pages 95–104.
- Harding, J.D.** (2021). *Applications of the Discrete Fourier Transform to Music Analysis*. PhD thesis, University of Florida.
- Harte, C., Sandler, M.B., Abdallah, S.A, and Gómez, E.** (2005). Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 66–71.
- Harte, C., Sandler, M., and Gasser, M.** (2006). Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia (AMCMM)*, pages 21–26. DOI: <https://doi.org/10.1145/1178723.1178727>
- Henschel, J., Neuwirth, M., and Rohrmeier, M.** (2021). The annotated Mozart sonatas: Score, harmony, and cadence. *Transactions of the International Society for Music Information Retrieval* 4(1), 67–80. DOI: <https://doi.org/10.5334/tismir.63>
- Huron, D.** (1993). Chordal tone doubling and the enhancement of key perception. *Psychomusicology* 12(1), 73–83. DOI: <https://doi.org/10.1037/h0094115>
- Huron, D.** (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/6575.001.0001>
- Jacoby, N., Tishby, N., and Tymoczko, D.** (2015). An information theoretic approach to chord classification and functional harmony. *Journal of New Music Research* 44(3), 219–244. DOI: <https://doi.org/10.1080/09298215.2015.1036888>
- Ju, Y., Condit-Schultz, N., Arthur, C., and Fujinaga, I.** (2017). Non-chord tone identification using deep neural networks. In *DLfM '17: Proceedings of the 4th International Workshop on Digital Libraries for Musicology*, pages 13–16. DOI: <https://doi.org/10.1145/3144749.3144753>
- Krumhansl, C.** (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
- Krumhansl, C., and Cuddy, L.L.** (2010). A theory of tonal hierarchies in music. In M.R. Jones, R. R. Fay, & A. N. Popper (Eds.), *Music Perception*. New York: Springer, pages 51–87. DOI: https://doi.org/10.1007/978-1-4419-6114-3_3
- Lieck, R. and Rohrmeier, M.** (2020). Modelling hierarchical key structure with pitch scapes. In *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, pages 811–818.
- Mavromatis, P.** (2009). Minimum description length modelling of musical structure. *Journal of Mathematics and Music* 3(3), 117–136. DOI: <https://doi.org/10.1080/17459730903313122>
- Maxwell, H.J.** (1992). An expert system for harmonic analysis of tonal music. In M. Balaban, K. Ebcioglu, and O. Laske (Eds.), *Understanding Music with AI: Perspectives on Music Cognition*, pages 335–353.
- McFee, B. and Bello, J.P.** (2017). Structured training for large-vocabulary chord recognition. *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, pages 188–194.
- Micchi, G., Gotham, G., and Giraud, M.** (2020). Not all roads lead to Rome: Pitch representation and model architecture for automatic harmonic analysis. *Transactions of the International Society for Music Information Retrieval*, 3(1), 42–54. DOI: <https://doi.org/10.5334/tismir.45>
- Moss, F., Neuwirth, M., Harasim, D., and Rohrmeier, M.** (2019). Statistical characteristics of tonal harmony: A corpus study

- of Beethoven's string quartets. *PLoS ONE* 14(6), e0217242. DOI: <https://doi.org/10.1371/journal.pone.0217242>
- Novarro-Cáceres, M., Caetano, M., Bernardes, G., Sánchez-Barba, M., and Sánchez-Jara, J.M.** (2020). A computational model of tonal tension profile of chord progressions in the tonal interval space. *Entropy* 22: 1291. DOI: <https://doi.org/10.3390/e22111291>
- Pardo, B., and Birmingham, W.P.** (2002). Algorithms for chordal analysis. *Computer Music Journal* 26(2), 27–49. DOI: <https://doi.org/10.1162/014892602760137167>
- Pauwels, J., O'Hanlon, K., Gómez, E., and Sandler, M.B.** (2019). 20 years of automatic chord recognition from audio. In *Proceedings of the 20th International Conference on Music Information Retrieval (ISMIR)*, pages 54–63.
- Quinn, I.** (2010). Does musical syntax have parts of speech? In *Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC11)*, Seattle, Washington.
- Quinn, I., and Mavromatis, P.** (2011). Voice leading and harmonic function in two chorale corpora. In C. Agon, M. Andreatta, G. Assayag, E. Amiot, J. Bresson, and J. Mandreau (Eds.), *Mathematics and Computation in Music, Third International Conference (MCM2011)*, pages 230–240. DOI: https://doi.org/10.1007/978-3-642-21590-2_18
- Ramírez, A., Bernardes, G., Davies, M.E.P., and Serra, X.** (2020). TIV.LIB: An open-source library for the tonal description of musical audio. In *Proceedings of the 23rd International Conference on Digital Audio Effects (DAFx-20)*.
- Raphael, C. and Stoddard, R.** (2004). Functional analysis using probabilistic models. *Computer Music Journal* 28(3), 45–52. DOI: <https://doi.org/10.1162/0148926041790676>
- Rohrmeier, M.** (2011). Toward a generative syntax of tonal harmony. *Journal of Mathematics and Music* 5(1), 35–53. DOI: <https://doi.org/10.1080/17459737.2011.573676>
- Rohrmeier, M., and Cross, I.** (2008). Statistical properties of tonal harmony in Bach's chorales. In *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC 2008)*, pages 619–627.
- Sapp, C.S.** (2005). Visual hierarchical key analysis. *ACM Computers in Entertainment* 3(4), 1–19. DOI: <https://doi.org/10.1145/1095534.1095544>
- Sapp, C. S.** (2011). *Computational Methods for the Analysis of Musical Structure*. PhD thesis, Stanford University.
- Sears, D.R.W., and Forrest, D.** (2021). Triadic patterns across classical and popular music corpora: Stylistic conventions, or characteristic idioms? *Journal of Mathematics and Music* 15(2), 140–153. DOI: <https://doi.org/10.1080/17459737.2021.1925762>
- Sears, D.R.W., and Widmer, G.** (2021). Beneath (or beyond) the surface: Discovering voice-leading patterns with skip-grams. *Journal of Mathematics and Music* 15(3), 209–234. DOI: <https://doi.org/10.1080/17459737.2020.1785568>
- Temperley, D.** (2001). *Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.
- Temperley, D.** (2007). *Music and Probability*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/4807.001.0001>
- Temperley, D.** (2009). *A Statistical Analysis of Tonal Harmony*. <http://davidtemperley.com/kp-stats>
- Temperley, D.** (2018). *The Musical Language of Rock*. New York: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780190653774.001.0001>
- Temperley, D. and deClercq, T.** (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research* 42(3), 187–204. DOI: <https://doi.org/10.1080/09215.2013.788039>
- Tompkins, D.** (2017). *Early Seventeenth-Century Harmonic Practice: A Corpus Study of Tonality, Modality, and Harmonic Function in Italian Secular Song with Baroque Guitar Accompaniment in Alfabeto Tablature*. PhD thesis, Florida State University.
- Tymoczko, D.** (2003). Function theories: A statistical approach. *Musurgia* 10(3–4), 35–64.
- Tymoczko, D.** (2011). *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. New York: Oxford University Press.
- White, C. W.** (2013a). An alphabet-reduction algorithm for chordal n -grams. In J. Yust, J. Wild, and J.A. Burgoyne (Eds.), *Mathematics and Computation in Music: Fourth International Conference (MCM 2013)*, pages 201–212. DOI: https://doi.org/10.1007/978-3-642-39357-0_16
- White, C. W.** (2013b). *Some Statistical Properties of Tonality, 1650–1900*. PhD thesis, Yale University.
- White, C. W. and Quinn, I.** (2016). The Yale Classical Archives corpus. *Empirical Musicology Review* 11, 50–58. <https://emusicology.org/article/view/4958>. DOI: <https://doi.org/10.18061/emr.v11i1.4958>
- White, C. W. and Quinn, I.** (2018). Chord context and harmonic function in tonal music. *Music Theory Spectrum* 40(2), 314–337. DOI: <https://doi.org/10.1093/mts/mty021>
- Winograd, T.** (1968). Linguistics and the computer analysis of tonal harmony. *Journal of Music Theory* 12(1), 2–49. DOI: <https://doi.org/10.2307/842885>
- Viaccoz, C., Harasim, D., Moss, F.C., and Rohrmeier, M.** (2022). Wavescapes: A visual hierarchical analysis of tonality using the discrete Fourier transform. *Musicae Scientiae*. <https://journals.sagepub.com/doi/full/10.1177/10298649211034906>. DOI: <https://doi.org/10.1177/10298649211034906>
- Yust, J.** (2017a). Harmonic qualities in Debussy's "Les sons et les parfums tournent dans l'air du soir." *Journal of Mathematics and Music* 11(2–3), 151–173. DOI: <https://doi.org/10.1080/17459737.2018.1450457>
- Yust, J.** (2017b). Probing questions about keys: Tonal distributions through the DFT. In O.A. Agustin-Aquino, E. Lluís-Puebla, and M. Montiel (Eds.), *Mathematics and Computation in Music, Sixth International Conference, (MCM 2017)*, pages 167–179. DOI: https://doi.org/10.1007/978-3-319-71827-9_13
- Yust, J.** (2019). Stylistic information in pitch-class distributions. *Journal of New Musical Research* 48(3), 217–231. DOI: <https://doi.org/10.1080/09298215.2019.1606833>

TO CITE THIS ARTICLE:

Yust, J., Lee, J., and Pinsky, E. (2022). A Clustering-Based Approach to Automatic Harmonic Analysis: An Exploratory Study of Harmony and Form in Mozart's Piano Sonatas. *Transactions of the International Society for Music Information Retrieval*, 5(1), 113–128. DOI: <https://doi.org/10.5334/tismir.114>

Submitted: 06 May 2021 **Accepted:** 07 June 2022 **Published:** 13 October 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Transactions of the International Society for Music Information Retrieval is a peer-reviewed open access journal published by Ubiquity Press.