A Statistical Approach to Tracing the Historical Development of Major and Minor Pitch Distributions, 1400-1750

Joshua D. Albrecht
University of Mary Hardin-Baylor

David Huron
Ohio State University

A large-scale study is reported whose purpose was to elucidate the historical development of the European major and minor modes. The study involved 455 musical works by 259 composers sampled across the years 1400 to 1750. Beginning with the period 1700-1750, a series of statistical studies was carried out on the distribution of scale tones, progressively moving backward in time. The method utilized a modified method of key determination – generalized to handle an arbitrary number of modal classifications. The results from cluster analyses on this data are consistent with the view that the modern major and minor modes have changed over time and were preceded by a system in which there were more than just two modes.

Received: October 16, 2012, accepted July 11, 2013.

Key words: modes, key-finding algorithm, corpus, cluster analysis, tonality

One of the most notable achievements in the field of music cognition has been the discovery of the role of statistical learning in the perception of tonal materials (Krumhansl, 1990). The research suggests that listeners experience music in light of culturally appropriate pitch schemas, and that these schemas are passively acquired by simple exposure to some body of music. Moreover, the research implies that this process is not unique to Western-enculturated listeners (Castellano, Bharucha, & Krumhansl, 1984; Eerola, 2004; Kessler, Hansen, & Shepard, 1984; Lantz, Kim, & Cuddy, in press; Nam, 1998). Although additional cross-cultural research is warranted, the underlying process of statistical learning has been observed in both infants and non-human animals, suggesting that the existing cross-cultural support is likely not merely coincidental (Hauser, Newport, & Aslin, 2001; Pearce & Bouton, 2001; Saffran, Johnson, Aslin, & Newport, 1999).

Krumhansl and Kessler (1982) used a probe-tone paradigm developed by Roger Shepard to measure the stability of various scale tones in different contexts. After establishing a key context, Krumhansl and Kessler played one of 12 chromatic tones, and participants judged the goodness of fit of the probe tone with the antecedent tonal context (see Figure 1). Krumhansl observed that there is a close similarity between the experimentally determined key profiles and the distribution of scale tones in actual music.

A practical application of this work has been the Krumhansl-Schumuckler (KS) key-estimation method (Krumhansl, 1990), in which key profiles are used to estimate the key of a passage, by choosing the key whose profile best matches the pitch-class distribution of the passage. While the KS algorithm has been highly influential, a number of attempts have been made to improve upon it. Several researchers have suggested that it makes more sense for key-estimation algorithms to use the actual distributions of pitch-classes in major and minor keys rather than goodness-of-fit data from perceptual experiments. For example, Aarden (2003) calculated the pitch-class distributions of 1,000 major-mode folk melodies to determine his major-mode key profile (Schaffrath 1995). Bellman (2005) developed a model based

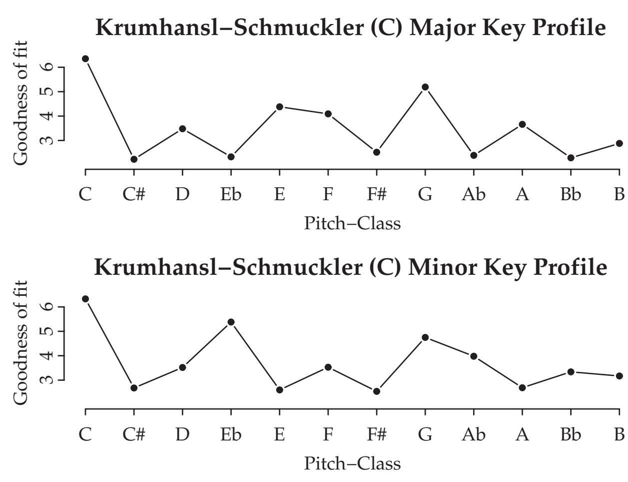

FIGURE 1. The results from Krumhansl and Kessler's probe tone studies. The resultant scale-degree vectors can be construed as "key-profiles," used in the Krumhansl-Schmuckler method of key estimation.

Music Perception, volume 31, issue 3, pp. 223–243, issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2014.31.3.223

on data from a previous study by Budge (1943) that tabulated the frequency of chords in 18th and 19th century music. By drawing on the actual statistical properties of music, these key-estimation models have been shown to provide improved estimates of musical key (Albrecht & Shanahan 2013).

Although rarely explicitly stated, all of these approaches presume that passages are in one of only two possible modes: major and minor. Another assumption is that scale-degree distributions do not differ according to composer, genre, or style. Most importantly, there is a "synchronic" assumption that the use of scale tones does not change over historical time. For music historians, this latter assumption is especially problematic. Like spoken languages, in the broad sweep of history musical practices clearly change.

In this study, we will empirically investigate the stability of pitch-class distributions over a 350-year period of Western musical history. We will use a data-driven approach to determine whether the pieces prior to the common-practice era group into categories (analogous to major and minor in common-practice music) by virtue of their pitch-class distributions. For each 50-year period within this era, our aim is to determine how many such categories might have existed, and identify appropriate pitch-related profiles for each category.

In pursuing our study, we will make two fundamental assumptions. First, we will assume that despite changes over time, within any given historical period musical passages are apt to exhibit similar distributions of scale tones (see Huron & Veltman, 2006). A second assumption is that any given "mode" may be performed at any transposition level, and that the absolute pitch height of some notated music may mask the pitch-related similarity to some other work notated at a different pitch height.

When applying a key-finding algorithm to some musical passage, it is appropriate to maintain some perspective. When a musician says that a passage is "in G major," a host of associations are evoked. The statement implies a certain hierarchy of pitches as well as certain tonal affinities, such as the relationship between the leading tone and the tonic. It implies the inclusion and exclusion of certain pitch-classes, a scalar ordering for pitches deemed "in the key," and various psychological *qualia* – such that the pitch G will evoke the greatest sense of closure. However, when an algorithm estimates the key of some passage as being "in G major," it means little more than "similar to a certain pitch-class distribution at a certain transposition level."

In applying a key-finding algorithm to the music of the past, we will inevitably encounter interpretive complications. For those scholars familiar with early European music, the idea that some passage is "in" some particular key or mode similarly evokes a host of considerations. These include issues of *ficta*, *finalis* tones, psalm tones, tenors, plagal and authentic tetrachords, certain melodic gestures, and many other intricate issues. Applying an algorithm to estimate the mode and tonic of a work will inevitably provoke concerns that parallel the concerns that modern theorists express regarding its application to more recent music.

Without dismissing or downplaying these interpretive complexities, we propose to use the same simplification as Krumhansl, applying what may seem to be an over-simplification with the hope that it may contribute independent (data-driven) evidence pertinent to the ongoing debates about how best to interpret the music of the past. However, in applying such an algorithm, we caution readers against reading too much into the results. In what follows, when our results suggest that two passages are "in the same key/mode," this statement should be interpreted at the empirical level: "similar to a certain pitch-class distribution at a certain transposition level." Ultimately, although our results may invite various interpretations by our readers, our method merely points to a network of similarities, devoid of further interpretation.

The period spanned by our study (1400-1750) includes an era music historians regard as the transition between an earlier system of medieval modes and a later classical emphasis on the major and minor modes. The nature of this transition remains contentious. Without wanting to minimize the importance of this issue, adequately discussing the intricacies of this debate is beyond the scope of this article. Instead we will present a method that focuses on chronicling diachronic changes of practice as evident in notated music. Rather than testing a specific hypothesis, we will employ a bottom-up procedure that affords one way of tracing how pitch-class practices change over time. Our approach is inspired by what anthropologists (borrowing from the field of linguistics) would regard as an etic rather than emic description. That is, our procedure focuses on chronicling nominally objective patterns of musical behavior rather than interpreting how those patterns might have been heard or interpreted by past listeners or past scholars.

Method

In order to study scale-degree usage over history, some means of assessing the tonal center of works in unknown modes is required. Traditionally, this problem has been addressed by using an automatic key-finding algorithm in which the distributions used for the algorithms are fixed and predefined. However, if one entertains the idea that the distributions for the major and minor modes have potentially changed over history, then an algorithm based on prior distributions is problematic. There is something of a chicken/egg problem in which the musical practice determines the distributions, while the distributions shape the musical practice.

Suppose that we have a musical work from the distant past and propose to use the distribution of pitch classes as a way of estimating the passage's key or mode. In order to determine whether the work is in mode X or mode Y, we need pre-existing distributions for both modes. The pertinent distributions, however, would be the mode distributions as they existed at the time the musical work was created. But, to create pertinent distributions for that time period we must know how many different modes were current at the time, and potentially how to transpose the various works so that works in the same mode would be recognized as such.

In short, some sort of "bootstrap" method is needed. One approach is to begin with modern times and work backward. That is, although we have no knowledge, for example, of how many modes listeners distinguished in 1600, there is a fair amount of agreement that two modes (major and minor) were predominant during what is traditionally defined as the "common-practice period," roughly 1700-1850. We might therefore apply scale-degree distributions from this common-practice period to the music of an immediately preceding historical period as a way of estimating the key (mode and tonic) for the music in question. Having established the mode for each work in a sample of earlier music, we can then revise our "mode profiles" to reflect the distributions of pitches for that period, and then apply these revised distributions to the next earlier period. In this way, one could theoretically work backward through history, allowing whatever changes occurred to emerge.

A difficulty with this approach is that we cannot be sure of the number of "modes" current at any given time. Rather than essentializing this question, our study will simply rely on a data-driven approach. An appropriate method is provided by cluster analysis. Accordingly, for each historical epoch, we may subject the pitch distributions for various works to a cluster analysis, and allow the music to suggest the number of pitch-related schemas or "modes" evident at that time, based on the aggregate similarities of the zeroth-order pitch distributions.

In light of these considerations, this study will use the following method:

1) Musical works will be sampled over a series of epochs. In this study, we use 50-year epochs.
2) Beginning from the epoch with the most stable distributions for the major and minor modes, a mode-finding algorithm (described below) will be used to determine the tonal center for each work. The works will then be transposed to a common nominal "tonic" pitch.
3) With common "tonics," the pitch distributions will be determined for each work in the epoch sample and these distributions will be subjected to cluster analysis (each of the twelve scale-degree proportions used as a variable). The analysis will determine the number of modes conjectured to have existed during that epoch; for example, two clusters representing the major and minor modes might be expected to appear for music throughout the traditionally-defined common-practice period.
4) Scale-degree distributions will be determined for each cluster by amalgamating the distributions of all of the works deemed to belong to their respective clusters. These distributions will then be applied to the immediately preceding epoch.
5) The process will be repeated, moving backwards through history. Our method relies on the assumption that pitch schemas for any given epoch will be similar to the preceding 50-year epoch.

The mode-finding algorithm used in this study is taken from recent work by Albrecht and Shanahan (2013). They found that Euclidean distance provides a better means of determining the tonic and mode of a work than the standard formulation of using correlation, especially with works in the minor mode. By sampling 982 tonal works composed between 1700-1950 (625 in a major key and 357 in a minor key), they found average scale-degree distributions for both modes from the common-practice period, shown in Figure 2. By using actual scale-degree distributions for tonal works with a Euclidean distance approach, they were able to achieve 93.1% accuracy in key estimation. The "bootstrap" used to initiate the process with the first epoch is these scale-degree weights, used as the starting distributions for the major and minor modes. Once the works in the first epoch have been subjected to cluster analysis, the distributions of each cluster will become the new distributions for the preceding epoch.

In order to determine the best period to begin with, a pilot study was carried out to establish which fifty-year span in the common practice period exhibits the most stable distributions of the presumed major/minor system. Three works were sampled from each fifty-year

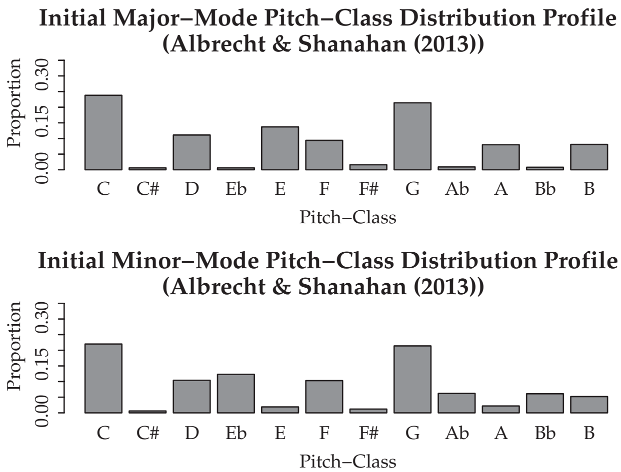

FIGURE 2. The scale-degree distributions taken from Albrecht and Shanahan (2013). The major-mode distributions represent 625 works and the minor mode distributions represent 357 works from the traditionally-defined common practice era.

period between 1650 and 1900. The key profiles representing an average of scale-degree distributions from the tonal era (see Figure 2) were used to determine the key of the work. The period 1700-1750 was the only period in which the key was correctly assigned to all three works, suggesting that this period may be especially consistent with the presumed major/minor system. Consequently, we chose to begin our study with the 1700-1750 epoch.

SAMPLING METHOD, MATERIALS, AND DATA

The sample of musical works included in this study would ideally consist of music that is representative of the cognitive pitch schemas present in various historical eras. One might suppose that the appropriate sampling method would seek equal numbers of musical works composed in each of the target periods. A more subtle approach would recognize that composers live for different lengths of time, and that compositions by elderly composers may reflect older schemas and fail to be representative of the time in which they were composed.

In general, mental schemas tend to be learned in a person's formative years. Although the style of composers may change over time, much of a composer's music-related schemas will have been formed earlier in life. One might assume that this is especially true for fundamental musical elements such as scale usage or the perception of mode. That is, one might expect that most stylistic changes throughout the life of the composer would involve elements of musical language that are more ephemeral than the perception of scale or mode.

TABLE 1. *The Results of a Pilot Study to Determine the Best Sample for a Mode-finding Algorithm.*

Portion of the work	Accuracy rating
First ten measures	54 of 58 correct (93%)
Middle ten measures	4 of 58 correct (7%)
Last ten measures	57 of 58 correct (98%)
First and last ten measures	58 of 58 correct (100%)

Note: The first, middle, and last ten measures of 58 Scarlatti piano sonatas were sampled and the resulting pitch-class distributions were compared to 24 pitch-class distributions for the major and minor mode, sampled from 982 works from the traditionally-defined common-practice era (Albrecht & Shanahan 2013). Using the first and last ten measures was most accurate, providing 100% accuracy.

For example, Perttu (2007) found that levels of chromaticism in theme-based and opus-based samples of the works of Beethoven and Mozart did not change over the course of their lives. For the purposes of this study, we will consider a musician's cognitive development to be "mature" at 25 years of age.

Accordingly, compositions are grouped into epochs, not by the year of a work's composition, but by the year when the work's composer turned 25 years of age. For example, Bach was born in 1685. Therefore, for this study his music is deemed to be representative of musical schemas predominant around 1710. A work written by Bach in 1740 would then be coded as belonging to schemas around 1710. That is, for the purposes of this study, all of Bach's music would be regarded as representative of 1710.

In orchestral scores, it is common for instruments to double each other. From a sampling perspective, doubling essentially reduces data independence. Doubling is, however, much less common in music employing lighter textures. Consequently, the sample for this study will be limited to those works containing four or fewer parts. Hence, the sample includes keyboard works, string quartets and other chamber works, motets, and 1-, 2,- 3-, or 4-part choral works.

Most of the works in our study are taken from a convenience sample of scores available through the Petrucci Music Library (IMSLP.org). This source includes music by 6,343 composers.[1] In order to increase data independence, the sample was limited to one musical work by each sampled composer for the years between 1500 and 1750. Fifty composers were selected for each epoch from this period. Due to the limited availability of sources before 1500, a different sampling method was used for 1400-1500. For the years 1400-1450, ten composers were found, and 64 scores available in the database for these

[1] This is accurate as of September 19, 2011.

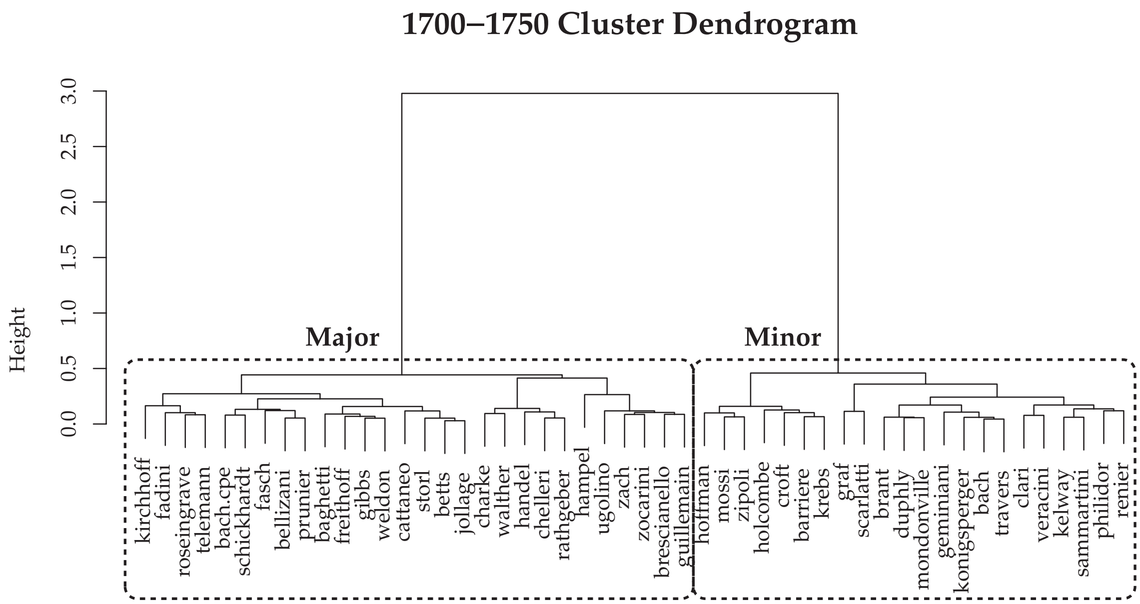

FIGURE 3. Solution for Ward's method of hierarchical agglomerative clustering for 1700-1750. Two clear clusters emerge from the data, a major-mode and a minor-mode cluster.

composers were used. For the years 1450-1500, 151 works from six composers were used; this sample was augmented by a convenience sample of 112 works by Josquin already encoded. The birth dates of the composers were determined using the New Grove Dictionary of Music and Musicians.

One potential difficulty in using musical sources from periods before the traditionally defined common practice era is the use of *musica ficta*. In modal music, chromatic alterations were often added to music to avoid dissonant intervals or to strengthen cadential formulations by adding half-step motion to the final. In fact, as Tymozcko (2011) has pointed out, the earliest hints of functional tonality were often at the cadence points. Unfortunately, *musica ficta* was rarely written into the score; instead, performers who knew the formulae knew when to add the chromatic alterations. However, in order to chronicle how scale-degree usage has changed over time, it is important to include the chromatic alterations of *musica ficta.*

Fortunately, many of the sources in the Petrucci music library undertake a reasonable editorial effort to include some indication of *musica ficta*. A visual inspection was made of each of the scores used as a source in this study for the periods 1400-1600. The score was considered to have included *ficta* if one of three criteria were met: 1) if the final cadence had a leading tone explicitly notated using an accidental; 2) if small accidentals were written above notes at some place in the score, unless the final cadence had bass motion by descending perfect fifth but did not use a leading tone; or 3) if the final cadence used leading tones that were natural notes (as in F or C modes), and if there were medial cadences that included notes modified by accidentals to make them leading tones. Failing all of the above criteria, the work was considered to not use *ficta*.

Using this operationalization, two scores from 1400-1450 failed to be considered to have *ficta* added, as did one score from 1500-1550. It is possible that these three scores did not use leading-tone cadences (though it is more likely that the chosen editions just did not include the *ficta*). In either case, with sample sizes of at least 50 pieces per epoch, these three pieces do not seem to be too influential. The convenience sample of Josquin scores was edited by Jesse Rodin under the Josquin Research Project in which he made reasonable editorial decisions to include *musica ficta*.

A further sampling difficulty concerns the amount of each piece to sample. If the goal of calculating the pitch distributions is to determine how composers use pitch-classes within a particular key (tonic and mode), then something of the fidelity of the distributions is lost if pitch-class distributions are calculated for music that modulates to other keys or modes. In tonal music, it is common for works to modulate to other keys within the work. Even in pre-tonal music, it is often the case that the medial cadences in a work reflect the treatment of a pitch other than the final as a temporary tonal

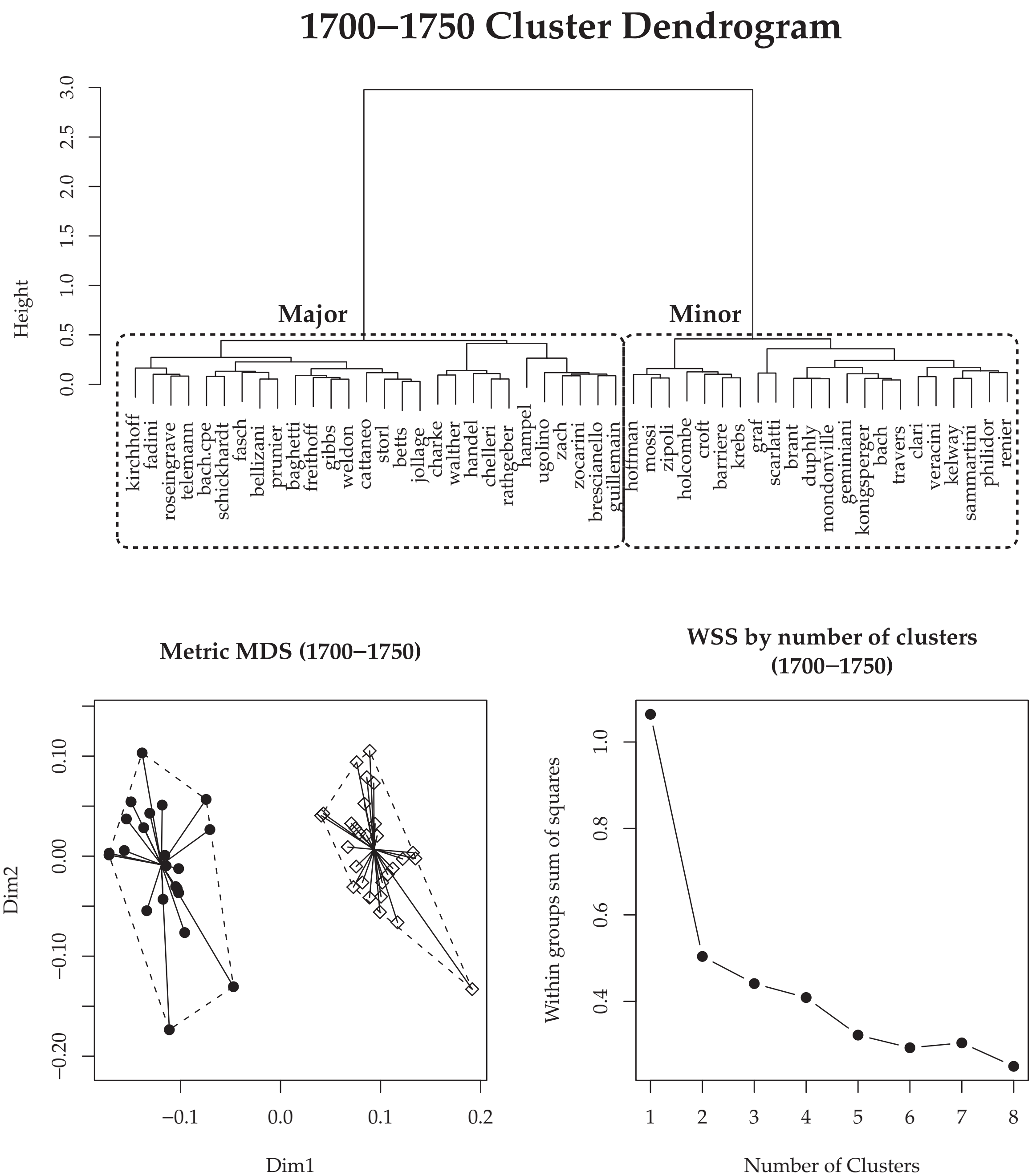

FIGURE 4. Top panel: Cluster analysis for 1700–1750. Two clusters are evident, corresponding to modern major and minor modes. Bottom left panel: Multidimensional scaling solution for the 1700–1750 epoch. Two clusters are evident, corresponding to the minor (left) and major (right) modes. Bottom right panel: Within groups sum of squares by number of clusters. The elbow at two suggest that the best number of groups for this data is two.

center, even if the pitch-classes treated as diatonic remain roughly the same. The tendency to sample from entire works, even works that modulate, is possibly an explanation for the majority of the chromaticism apparent in many of the major and minor distributions sampled from real music discussed above.

For the purposes of this study, it is important that the portion of the music analyzed reflect primarily one key,

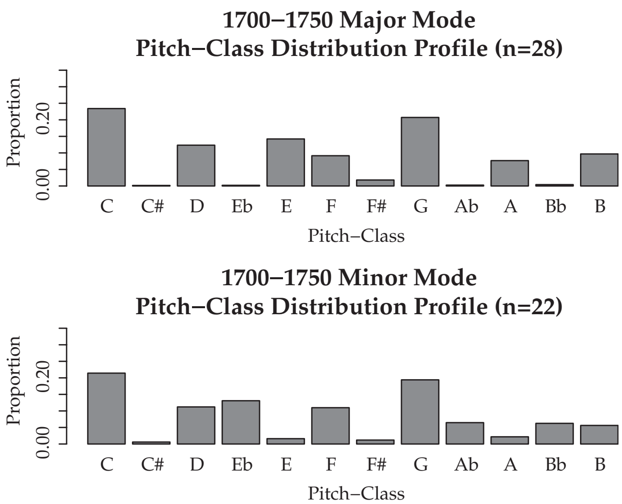

FIGURE 5. The scale-degree profiles arising from a cluster analysis for the epoch 1700-1750. There are two distributions, one for the major mode and one for the minor mode.

so a pilot study was carried out to determine what portion of each work to sample. In this pilot study, the key-finding algorithm proposed by Albrecht and Shanahan (2013) was used. A sample of 58 piano works by Scarlatti was chosen because 1) the key of each piece was known (reflected in the title) and 2) they were written by a composer who was 25 years of age within the period with which the study would begin (1700-1750). Four portions of each work were sampled and the key of the work was estimated: the first ten measures, the last ten measures, the middle ten measures, and the first and last ten measures.

The results of this pilot study are reported in Table 1. For the sample tested, the first and last ten measures provided 100% accuracy in estimating the key based on the pitch-class distributions, although the advantage was not significantly better than just using the last 10 measures for this small of a sample. Nevertheless, with an accuracy of 100%, we decided to sample the first and last ten measures for each selected work.

What constitutes the first and last ten measures in early music is somewhat difficult to determine because consistent notation of bar-lines did not become common until the mid-17th century (Hiley, 2013). The primary reason behind selecting only the first and last ten measures of a work is to select music from early and late in a work that is likely to remain in one mode. For this reason, it is acceptable to simply approximate what ten measures may be. A convenient approximation for measures in early music is offered by the database available from the Petrucci music library, in which most sources include the editorial addition of measures. For the purposes of this study, the first and last ten measures were taken to be that music marked by those editorially added bar-lines.

The first and last ten measures of each sampled work were encoded in Humdrum format. Specifically, the encoded material included the composer's name, the composer's birth and death dates, the title of the work, the date of the composition when available, the notated pitch (including accidentals), the durations, rests, and barlines. The data for the cluster analysis were the 12 dimensions of proportions of each of the 12 chromatic pitch-classes, transformed so that the nominal "tonic" of each piece was encoded as the first dimension.

PROCEDURE

Based on the results of the two pilot studies described above, a cluster analysis was carried out on samples of pieces from each of seven 50-year epochs, beginning with the epoch 1700-1750. Keys were estimated for each work in that epoch using a Euclidean-distance based key-finding algorithm, weighted with scale-degree distributions taken from 982 works in the common-practice era. Once the tonic pitch for each piece was estimated, scale-degree distributions were calculated by tallying the number of notes of each of the twelve chromatic scale-degrees, weighted by their durations. In other words, total duration of each scale degree was determined rather than simply counting the number of occurrences of each scale degree. The value for each scale degree was represented as a proportion of the total durations of all scale degrees for the purpose of comparison between works. Each of these 12 values was treated as a dimension in 12-dimensional space.

For the first epoch, 1700-1750, the 12-element vectors for each work were compared to the 24 major and minor key profiles from Albrecht and Shanahan (2013). For each work's distribution, the key profile that was closest in 12-dimensional Euclidean space was taken to be the key of the piece. The scale-degree vectors were then rotated so that the nominally tonic pitches would coincide. A cluster analysis was performed to determine if there were any clusters in the scale-degree distributions for these pieces. Ward's method of hierarchical agglomerative clustering, complete linkage clustering, and centroid clustering were used, and were performed on the data set of all 50 pieces to determine whether there was convergence between methods. Since the three clustering methods all yielded similar results, only Ward's method will be reported below. The number of clusters was determined by a visual inspection of the resulting dendrograms.

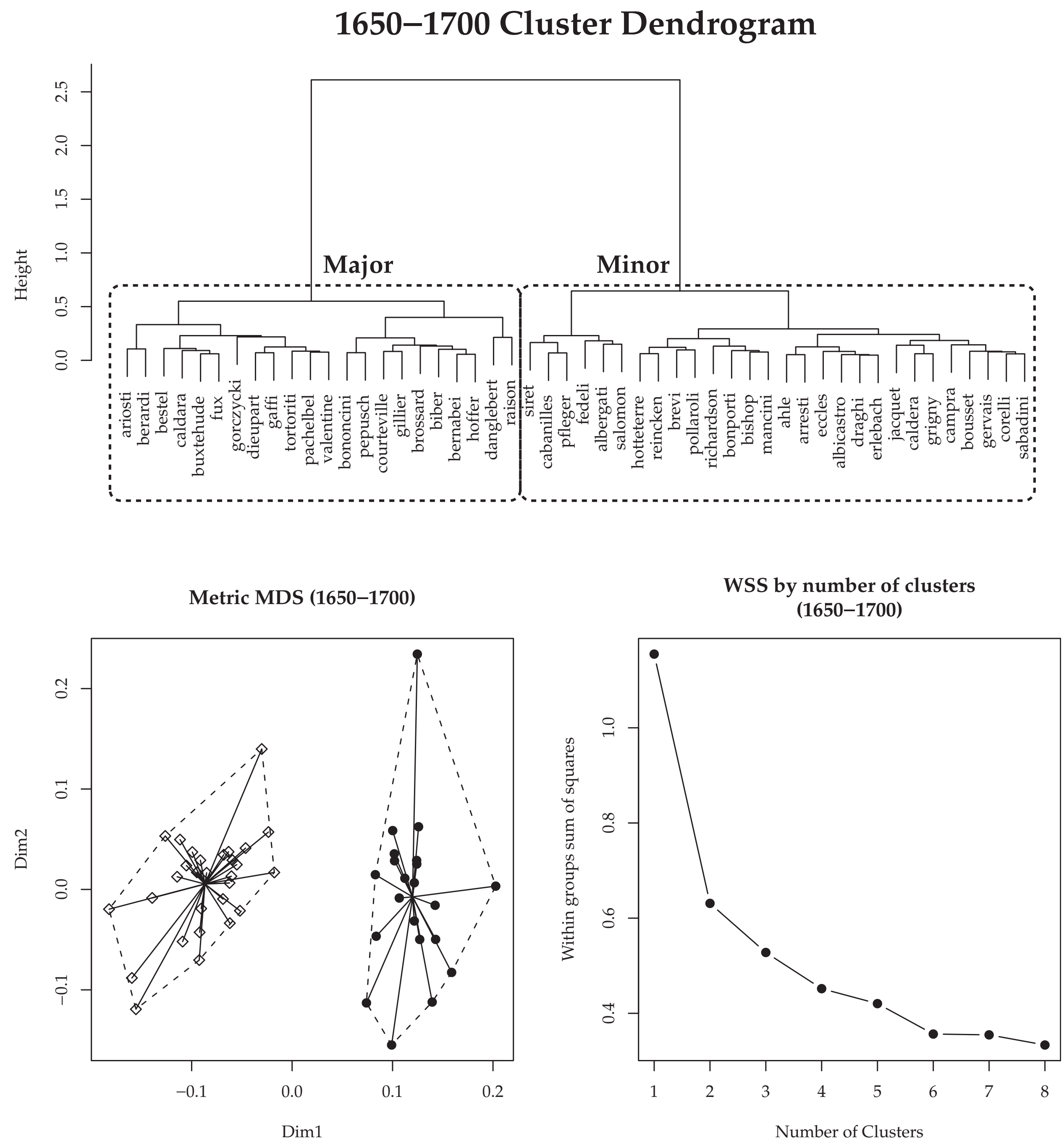

FIGURE 6. Ward's method of cluster analysis (top), MDS results (left), with the minor cluster on the left and the major cluster on the right, and WSS (right) for the 1650-1700 epoch.

Once the number of clusters was established, a *k*-means cluster analysis was performed three times, with different starting locations randomly chosen for the *k*-mean centroids. The result exhibiting the lowest standard squared error was chosen as the best solution. This clustering solution was then plotted using multidimensional scaling in order to provide a visual display of patterns in the clustering. Specifically, the nominal "tonic" pitch and mode were included in the analysis to determine if the clusters aligned with pre-established

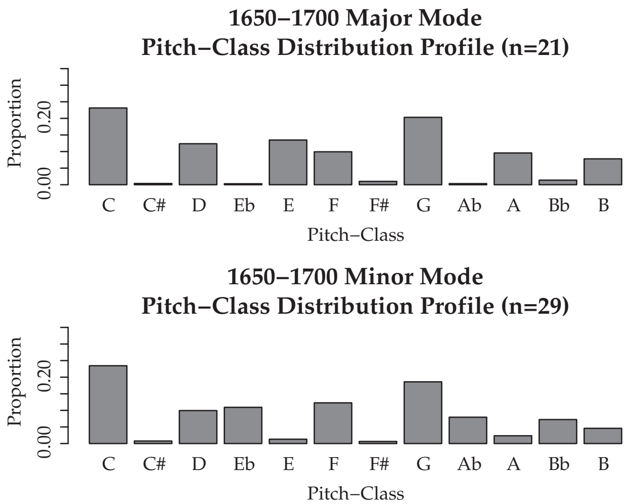

FIGURE 7. Scale-degree distributions for the major and minor modes in the 1650-1700 epoch.

notions of major and minor modes. Averages were calculated for each of the 12 scale degrees for each mode found. These distributions were then used as the given distributions for the epoch spanning 1650-1700; the process was repeated in this way for each 50-year epoch moving backward through time until the period 1400-1450.

Results

1700-1750

For the first epoch, 1700-1750, the Euclidean-distance based key-finding algorithm using distributions taken from 982 works in the common-practice era was applied to the sample of 50 works. The tonic was estimated for each piece and the scale-degree distributions were calculated.

The data were subjected to Ward's method of clustering, complete linkage clustering, and centroid clustering to determine if there were grouping structures within the data. The results of Ward's method are shown in Figure 3, in which there appear to be two distinct clusters. The other clustering methods (not shown) all similarly converge on the two-cluster solution. The leaves of the dendrogram represent individual pieces, labeled with the composer's name.

Unfortunately, there is no universally accepted statistical test for interpreting the number of germane clusters, especially in hierarchical-agglomerative clustering. In practice, the determination of the number of "important" clusters is a judgment of the researcher. Another scholar, examining the same dendrograms produced in this study might choose to identify either more or fewer clusters. Nevertheless, the amount of variance in each cluster gives some indication of what may be an appropriate number of clusters for the data. For each increase in the number of clusters, there is a decrease in the variance within each cluster. Trivially, if each object were its own cluster, there would be no variance. Typically, the number of clusters is determined to be the level at which there are "diminishing returns" in decreases in variance. In other words, the researcher looks to simultaneously minimize the number of groups and also minimize the variance within each cluster.

The appropriate determination of the number of clusters can be represented visually in two ways. In the first case, the distance between horizontal bars on a dendrogram represents the difference between the proximities of items within a cluster and the proximity of that cluster to the next closest cluster. A large distance indicates local grouping structures, which can be seen in Figure 3. In the second case, the within-group sum of squares (WSS) can be calculated for different numbers of groups. Where there is an "elbow" in the WSS, the amount of variance reduced by increasing the number of groups is decreased. An elbow can be seen on the right side of Figure 4 at two clusters.

The presence of two clusters in this data can be visually represented in multidimensional scaling, as illustrated on the left side of Figure 4. There are two clusters evident, with the minor cluster appearing on the left and the major cluster on the right. Notice the greater variance in the minor cluster, which may be an indication of the greater flexibility of the scale incorporating both raised and natural versions of scale-degrees six and seven.

With two distinct clusters, pitch-class vectors for both modes were calculated for each cluster, shown in Figure 5. Due to the clear effect of mode on the clusters and the similarity with the existing mode profiles, the labels "major" and "minor" have been retained to designate the two modes.

1650-1700

Using the revised key profiles, nominal tonics were estimated for the 50 works in the preceding 1650-1700 epoch. The same battery of cluster analyses was carried out. Again, there were many similarities between the analyses, so only Ward's method is shown (Figure 6). Two clusters are again evident in the dendrogram. The dendrogram and corresponding MDS solution are shown in Figure 6 and the pitch-class distributions for the two modes are shown in Figure 7.

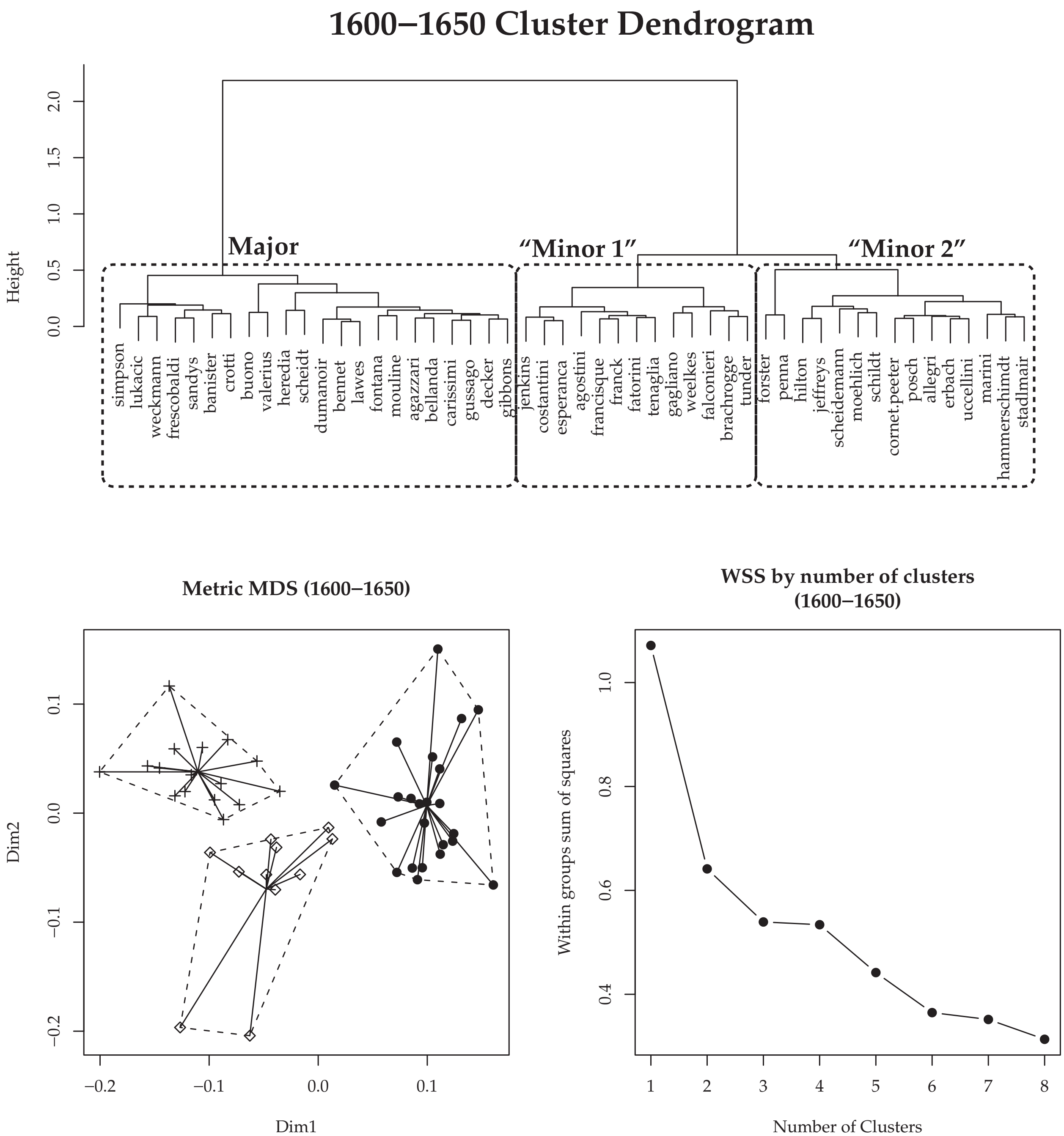

FIGURE 8. Ward's method of cluster analysis (top), MDS results (left), and the WSS (right) for the 1600-1650 epoch. The WSS indicates three clusters, shown in the MDS and cluster analysis, with the minor cluster divided into two stable distributions, called "minor 1" and "minor 2."

1600-1650

The revised distributions from 1650-1700 were then applied to the antecedent 1600-1650 epoch, and "tonics" were estimated for each sampled work. The pitch-class distributions for the sampled works were subjected to cluster analysis and the result of Ward's method and the corresponding MDS solution are shown in Figure 8. Again, two clusters are easily discernible, paralleling the major and minor mode. Each cluster seems to demonstrate some degree of subclustering, but the subclusters in the "minor" group again appear to be more pronounced. There is a clear "elbow" in the

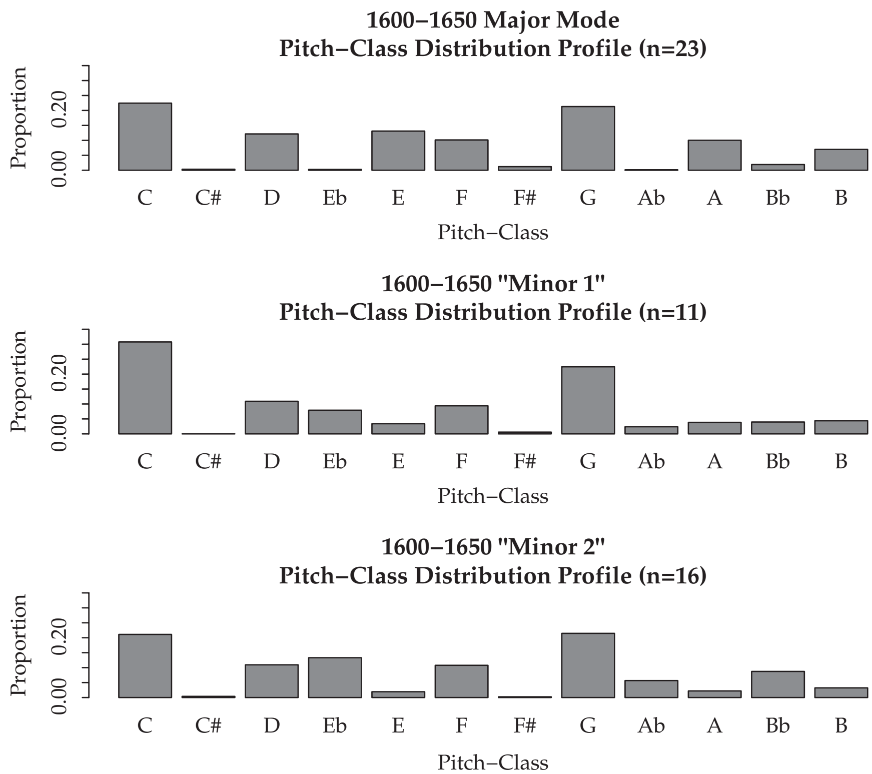

FIGURE 9. Scale-degree distributions for three modes in the 1600-1650 epoch. The minor scale exhibits two variants.

within-groups sum of squares value at three clusters, and therefore three clusters are graphed in Figure 8.

Figure 9 plots the three modal pitch-class distributions. The “major” distribution remains similar to the previous major distributions. However, there is an increase in the frequency of the nominal subtonic (lowered seventh scale-tone). The first “minor” mode distribution, labeled “minor 1,” is notable in that there appears to be no primary form of either scale-degree six or seven. “Minor 2,” on the other hand, bears closer similarities to the minor distributions from already sampled epochs, but with increased strength of the minor sixth and seventh scale-tones over their major counterparts.

1550-1600

Another iteration produced estimates for the nominal tonics of 51 pieces from the 1550-1600 epoch. A cluster analysis (Figure 10) again shows a clear bifurcation into major and minor modes with subclusters within the minor cluster. The pitch-class distributions for these three clusters are displayed in Figure 11. The “major mode” distribution continues to closely resemble the first major-mode profile from 1700-1750 ($r = .98$). The two “minor mode” distributions, however, are noticeably different from the beginning minor-mode profile (minor 1, $r = .83$; minor 2, $r = .93$): the most obvious difference between the two is the prevalence of the nominally “lowered mediant” scale degree, which is much more prevalent in “minor 2” than in “minor 1.” The nominal “submediant” tone is also noticeably different between “minor 1” and “minor 2”. Whereas “minor 2” has a prevalence of the “minor submediant” and almost no presence of the “major submediant,” “minor 1” has slightly more “major submediant” than “minor.” The nominal “leading tone” occurs equally infrequently in both modes, but the nominal “subtonic” is much more frequent in “minor 2.”

1500-1550

The three mode profiles from 1550-1600 were used to evaluate the nominal tonics of the 51 works in the 1500-1550 epoch. The cluster analyses used show a clear bifurcation between “major-like” and “minor-like” works, as Figure 12 shows. The distinction between “minor 1” and “minor 2,” however, seems to only have affected grouping at the most local of levels. The corresponding MDS graph clearly shows the major and minor clusters in 2

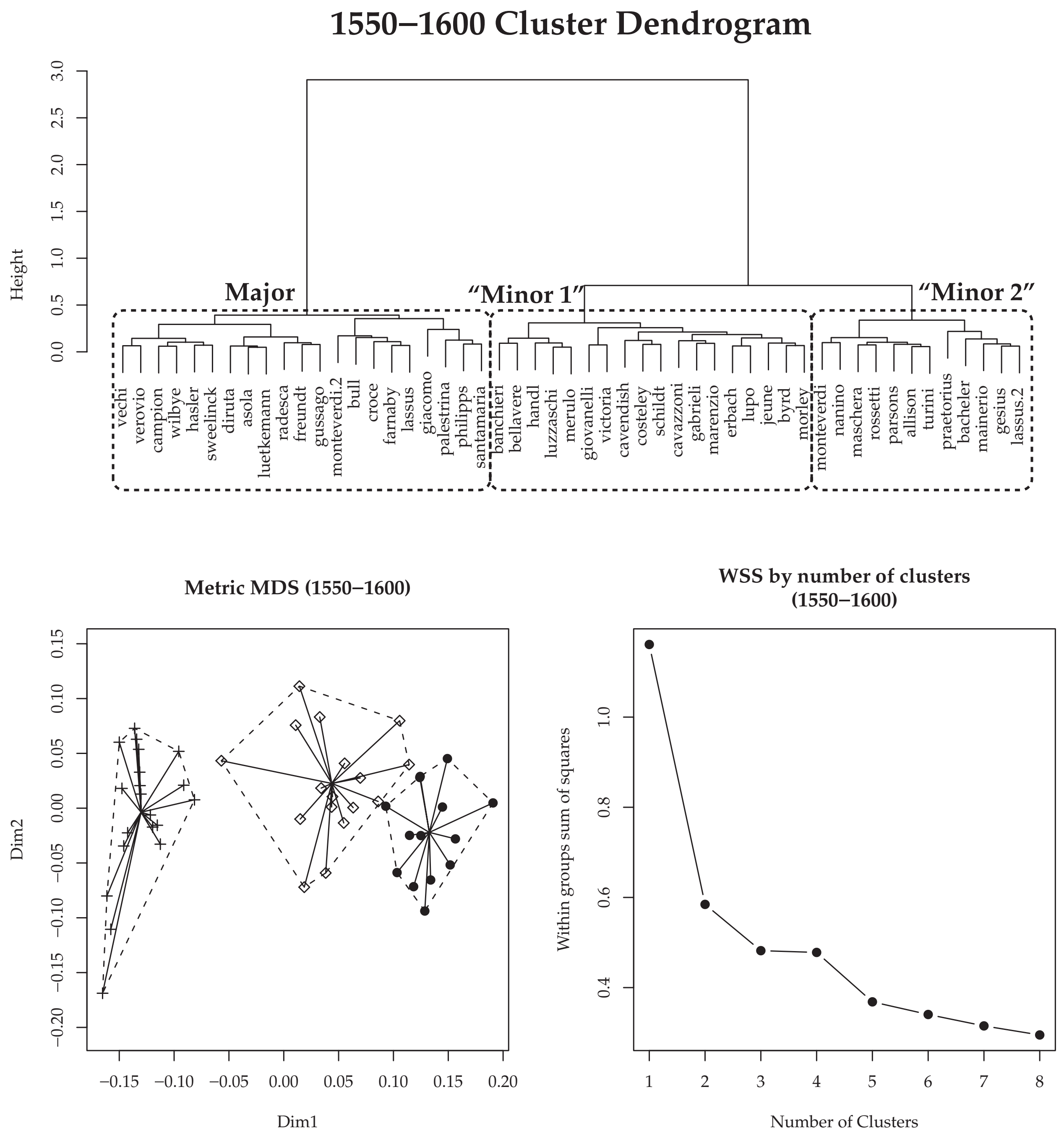

FIGURE 10. Ward's method of cluster analysis (top) and MDS (left) results for 1550-1600. Both the dendrogram (top) and the WSS (right) plots once again suggest three groups. The minor cluster again appears to exhibit subclustering, shown in the MDS solution (left).

dimensions. Though the difference between the two minor groups is slight, a test for 3 *k*-means revealed that the "minor 2" works nevertheless cluster together. The decision was therefore made to retain the three mode clusters despite the weaker evidence. The mode profiles (see Figure 13) are similar to the mode profiles from 1550-1600 with a few modifications (major, $r = .996$; minor 1, $r = .98$; minor 2, $r = .95$).

1450-1500

These three mode-distributions were retained for the 1450-1500 epoch. Recall that practical constraints

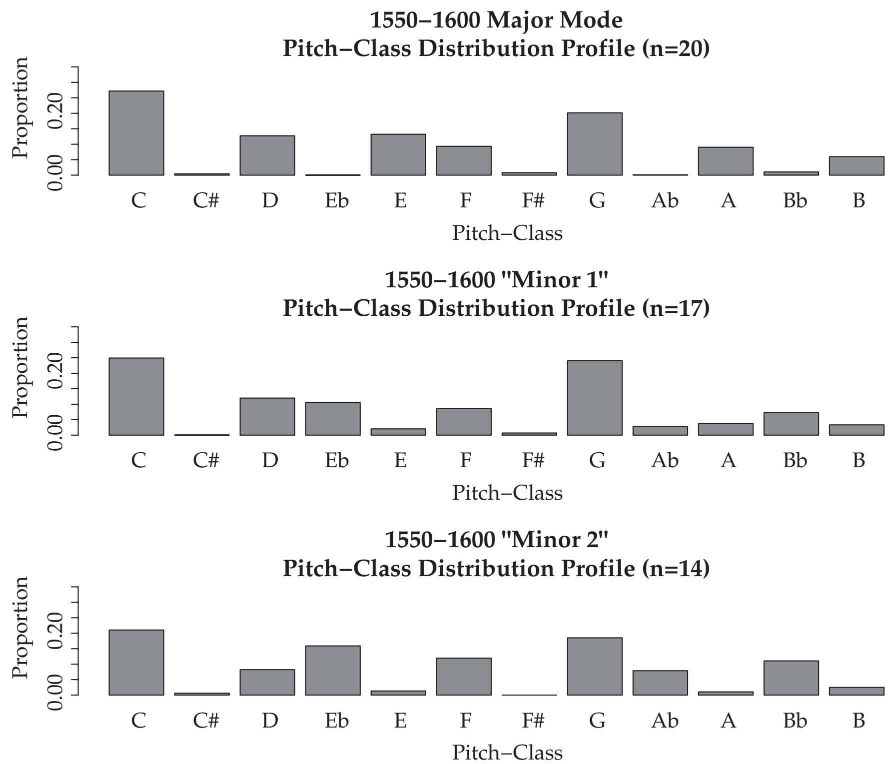

FIGURE 11. Scale-degree distributions for three modes in the 1550-1600 epoch.

affected the sampling method for the two fifteenth-century epochs (1400-1450 and 1450-1500). The population of musical works available for sampling is smaller, and sources are more difficult to obtain. For the epoch 1450-1500, a convenience sample of 112 Josquin works was used. Due to the sheer number of objects in the dendrogram associated with this cluster analysis, no dendrogram is displayed. However, the typical bifurcation into "major-like" and "minor-like" clusters remained evident. There again appeared to be two distinct subclusters in the minor cluster. The corresponding MDS graph and WSS plot are shown in Figure 14.

The pitch-class distributions for these three clusters are shown in Figure 15. We might highlight a few observations regarding these distributions. The first is a trait that became evident in previously sampled epochs, but which is more pronounced in these graphs. Specifically, the total proportion of first scale tones (the nominal "tonic") is larger in this epoch than in any of the other epochs examined thus far. The nominal "tonic" pitch represents nearly 30% of the total duration of notes. The biggest difference between "minor 1" and "minor 2" seems to be the use of the lowered third, sixth and seventh scale tones (i.e., the nominal "minor mediant," "minor submediant," and "subtonic"). "Minor 2" has higher proportions of those pitches. Also of note is that the nominal "leading tone" has primarily been replaced by the "subtonic" in both "minor" distributions.

1400-1450

The three distributions from 1450-1500 were then applied to the 1400-1450 epoch. After conducting a cluster analysis (Figure 16), there were, as before, two clear clusters consistent with a major/minor bifurcation. However, there appeared to be two clusters within "minor," and two clusters within "major," leading to four total subclusters. The four stable clusters are shown in the MDS solution in Figure 16 (left), with a stress in three dimensions of 0.16.

Finally, Figure 17 displays the pitch-class distributions for the four modes found to fit the sample of music for 1400-1450. There are two "major-like" modes and two "minor-like" modes. Nevertheless, as can be seen in Figure 16 "major 2" consists of only 2 pieces. The primary difference between the two major modes is the prominence of the nominal "tonic" and "dominant." It seems likely that, representing such a small cluster, these two pieces may be outliers, using

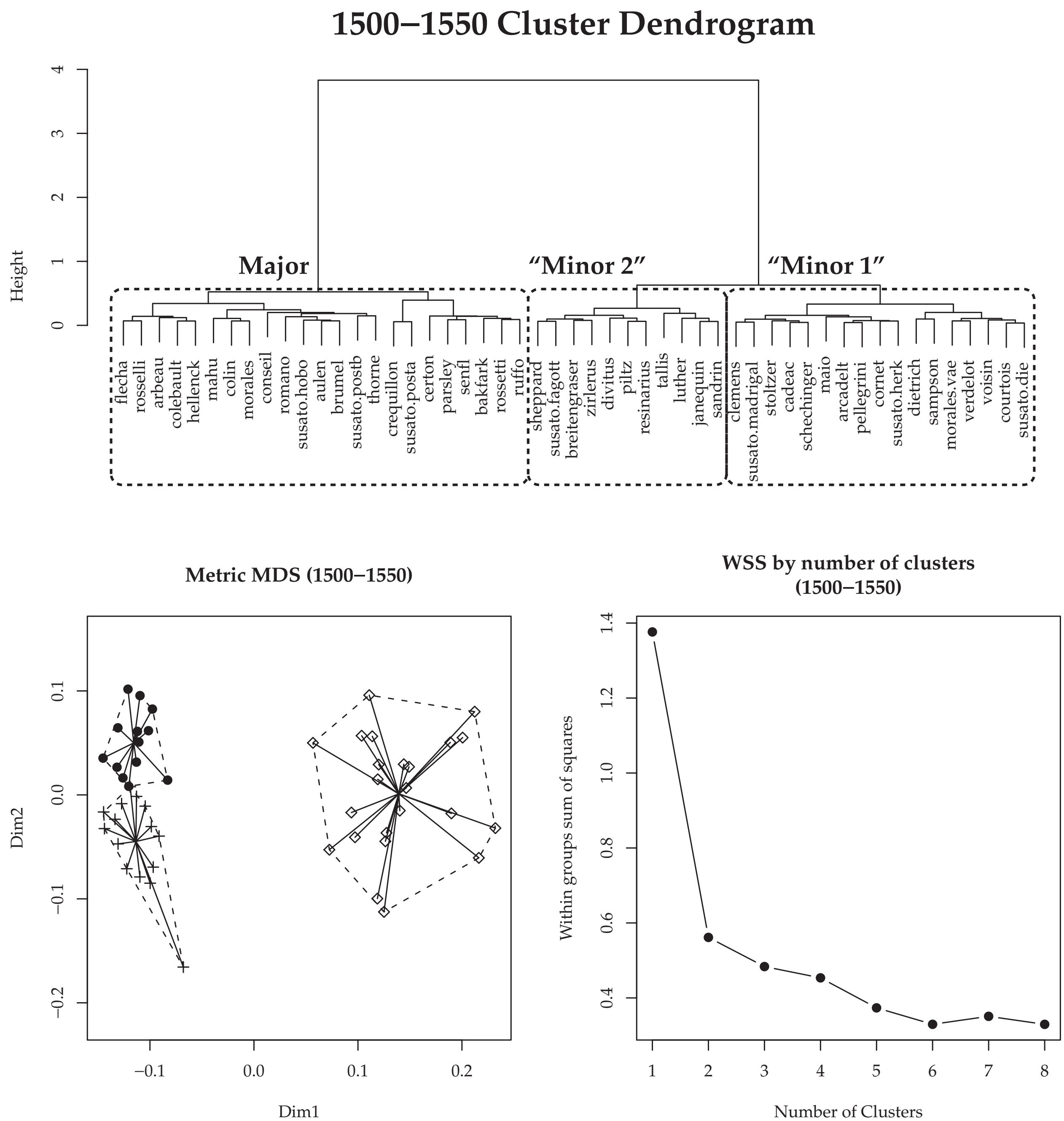

FIGURE 12. Ward's method of cluster analysis (top) and MDS (left) results for 1500-1550. The WSS (right) evidence for three groups is notably weaker than previous epochs. However, in the dendrogram, the minor cluster appears to exhibit subclustering, shown in the MDS solution (left).

pitch-classes differently than other works. The two "minor modes" show many of the same characteristics as they had in the other distributions. There is more of a difference in this case between major and minor "submediants," though, and "subtonic" and "leading tone." The major "submediant" is more prevalent in "minor 1" than in the distributions from later epochs. In both "minor-like" distributions, the nominal "subtonic" is much more highly represented than the "leading tone."

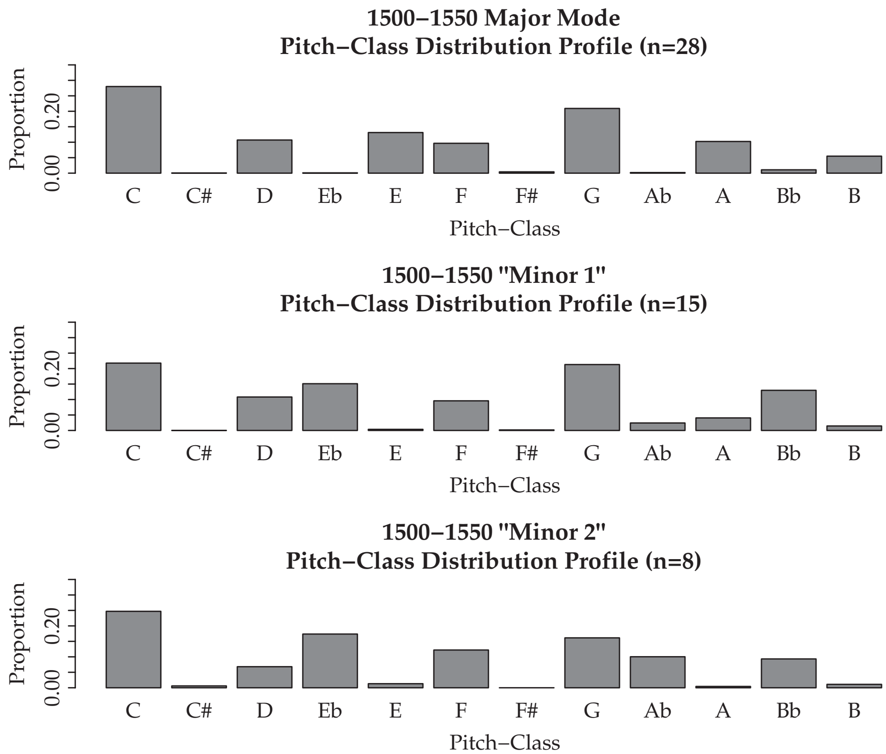

FIGURE 13. Scale-degree distributions for three modes in the 1500-1550 epoch. The minor scale again shows two variants.

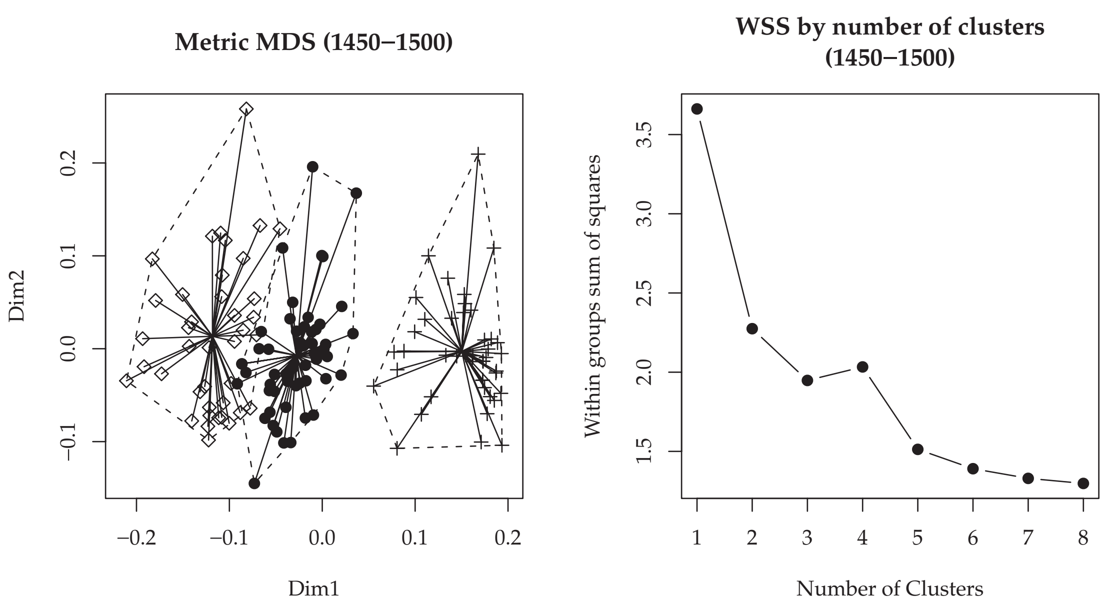

FIGURE 14. MDS solution (left) and WSS (right) for the 1450-1500 epoch. The WSS graph shows a clear "elbow" at three clusters.

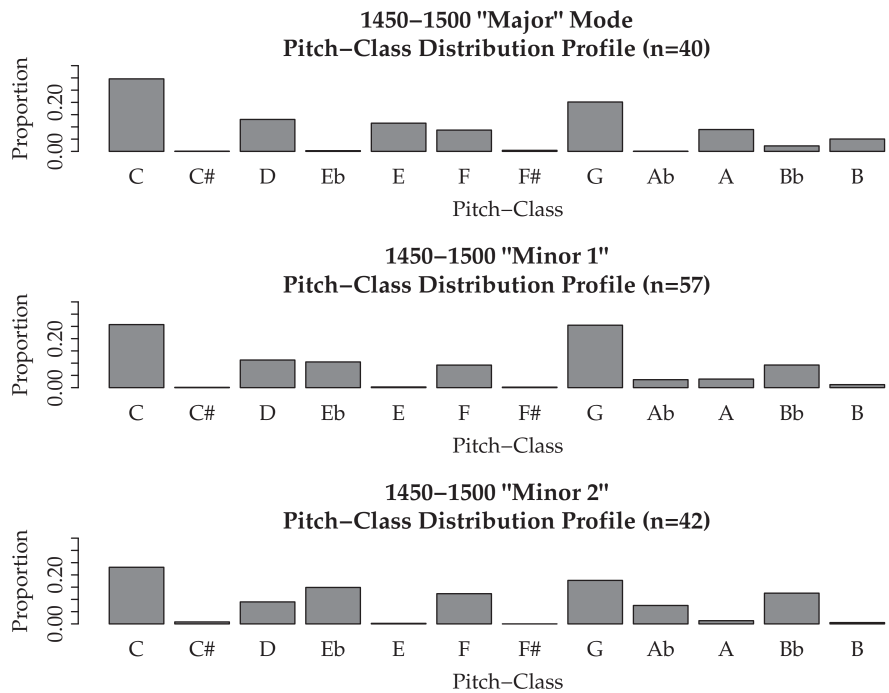

FIGURE 15. Three scale-degree distributions for the 1450-1500 epoch.

Caveats

Before offering any conclusions, it is appropriate to reiterate the many assumptions underlying this study. A single misguided premise can render the results spurious, so it is appropriate to make clear all of the known assumptions used in the study. Keeping these assumptions in mind will help to temper any conclusions or interpretations.

As noted earlier, a central assumption is that similarity of pitch-class distribution implies similarity of tonal perception. In addition, we have assumed that zeroth-order scale-degree distributions are adequate for characterizing the differences between presumed tonal schemas. Notice that we have not assumed that there are no other factors involved in tonal schemas, just that zeroth-order distributions are sufficient to capture the differences. We have assumed that any given mode may be performed at any transposition level. Our assumption regarding statistical learning of tonal schemas is concretely expressed in our reliance on key-finding methods.

In carrying out our analyses, we assumed that 50-year epochs provide sufficient resolution for deciphering historical changes in tonality. In carving up these "epochs" we have selected arbitrary boundaries at the century and half-century marks, and have assumed that these boundaries do not introduce any confounding artifacts.

We have assumed that our selected musical samples are adequately representative of their respective periods. In assigning musical works to particular epochs, we have assumed that there is little change in how composers use mode over the course of their careers, and that their mode-related practices may be viewed as "stable" around the age of 25 years.

In assembling our musical materials, we have assumed that the beginnings and endings of musical works are likely to employ the same tonality. In our analysis, we began with the assumption that, for the period 1700-1750, there are only two pitch-related schemas: major and minor.

In the case of the clustering method, we acknowledge that clustering analysis itself provides no clear criterion for determining the number of salient clusters, and that what constitutes a salient cluster is partly a matter of researcher interpretation. Apart from these issues, a major concern arises from the assumption of transposability. We have assumed that the absolute pitch-height of the notated music is irrelevant, and that a resemblance between two pitch distributions may be evident at any transpositional relationship. Since the method of mode determination relies on the distributions of pitch classes, failing to accurately identify the correct transposition for a passage will add noise to the amalgamated mode distributions.

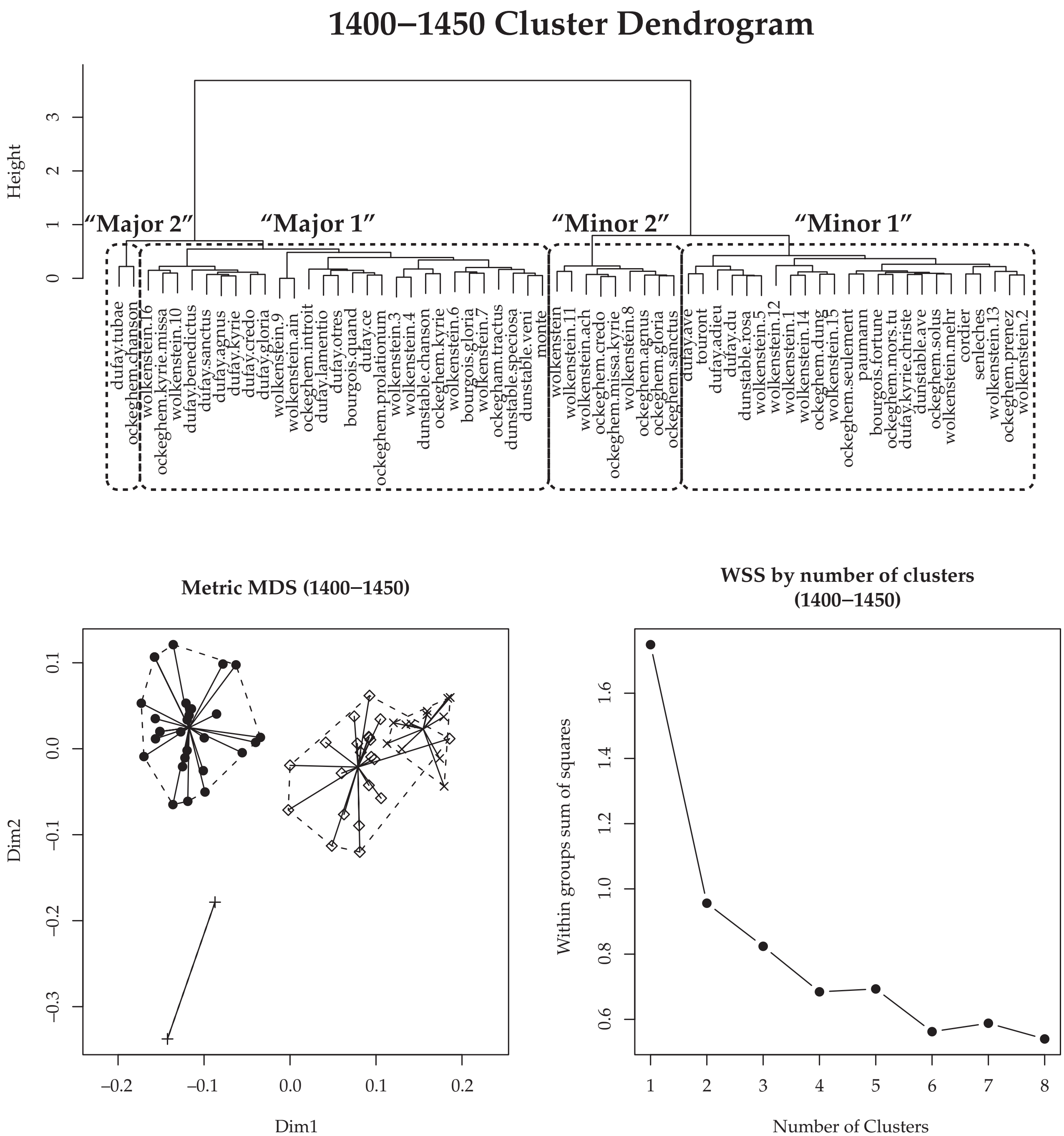

FIGURE 16. Ward's method of cluster analysis (top), MDS (left), and WSS (right) results for 1400-1450. The expected mode-bifurcation is clearly seen, but each cluster breaks down into subclusters. "Major 2," however, only consists of two pieces.

Moreover, due to the iterative nature of the method, it is possible that this noise can be amplified over time. For epochs that conform well to the presumed major-minor system, the key-finding algorithm appears to be very effective. (Recall, for example, that the algorithm correctly identified the keys for 58 of 58 Scarlatti sonatas in which the key information was explicitly given by the composer in the titles.) However, if a mode does not

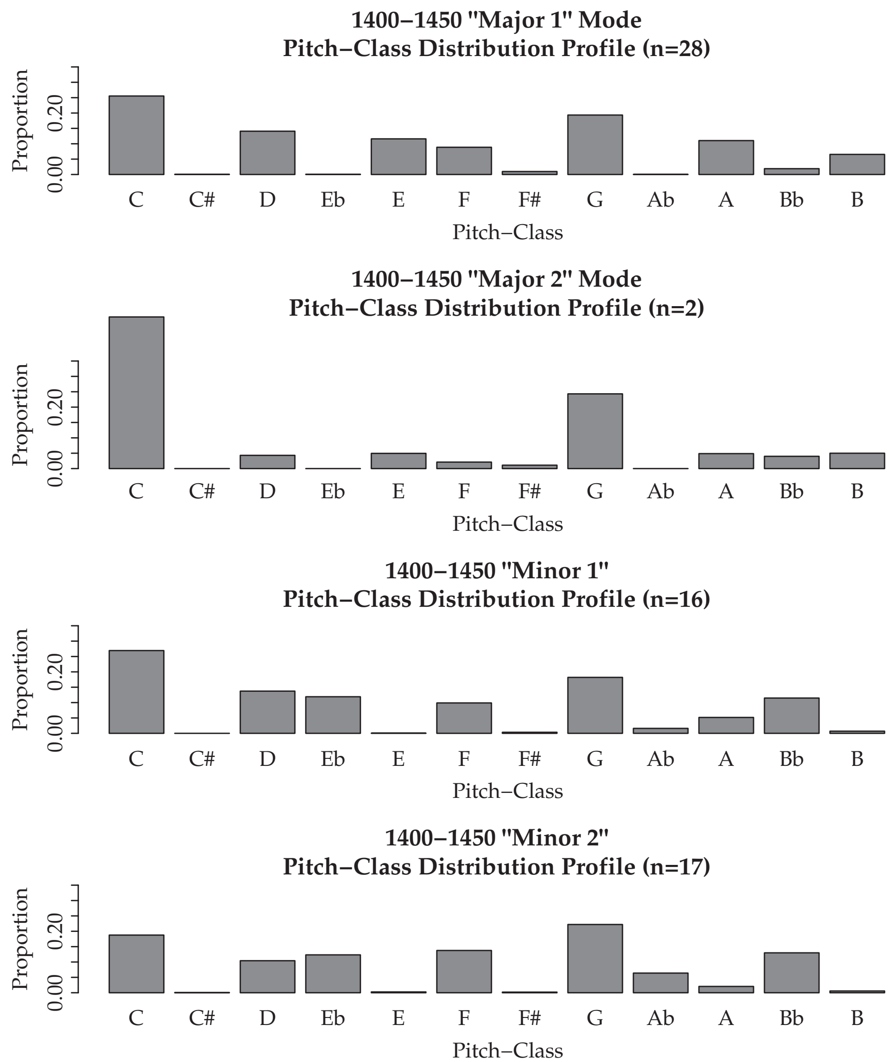

FIGURE 17. The pitch-class distributions for the four modes from 1400-1450.

occur very often in a musical sample, then it is more likely to be "shoe-horned" into some already existing mode distribution. That is, the procedure gains robustness at the expense of ignoring modes that may be represented in the sample by only a small number of works. Specifically, the mode-finding algorithm tends to transpose away less common modes by rotating them out of existence. Misclassified modes then influence the distributions in which they are included so that these new modes are more difficult to find in earlier epochs when they may be more prevalent. This may be especially problematic for tetrachordally-similar modes, such as the Ionian/Lydian, Aeolian/Phrygian, and Dorian/Aeolian pairs.

Conclusion

The main results of this study are summarized in Figure 18. Bearing in mind the many caveats, we might offer a number of observations.

In the first instance, the results appear to be rather orderly. There is no erratic appearance or disappearance of schemas. Instead, the distributions appear to be relatively stable over time. For example, the major mode distribution for the epoch 1650-1700 is statistically indistinguishable from the major mode distribution for 1700-1750 (with a correlation of .99). Nevertheless, we also see evidence suggesting schematic "drift" over long time spans. For example, the distribution labeled "Minor 1"

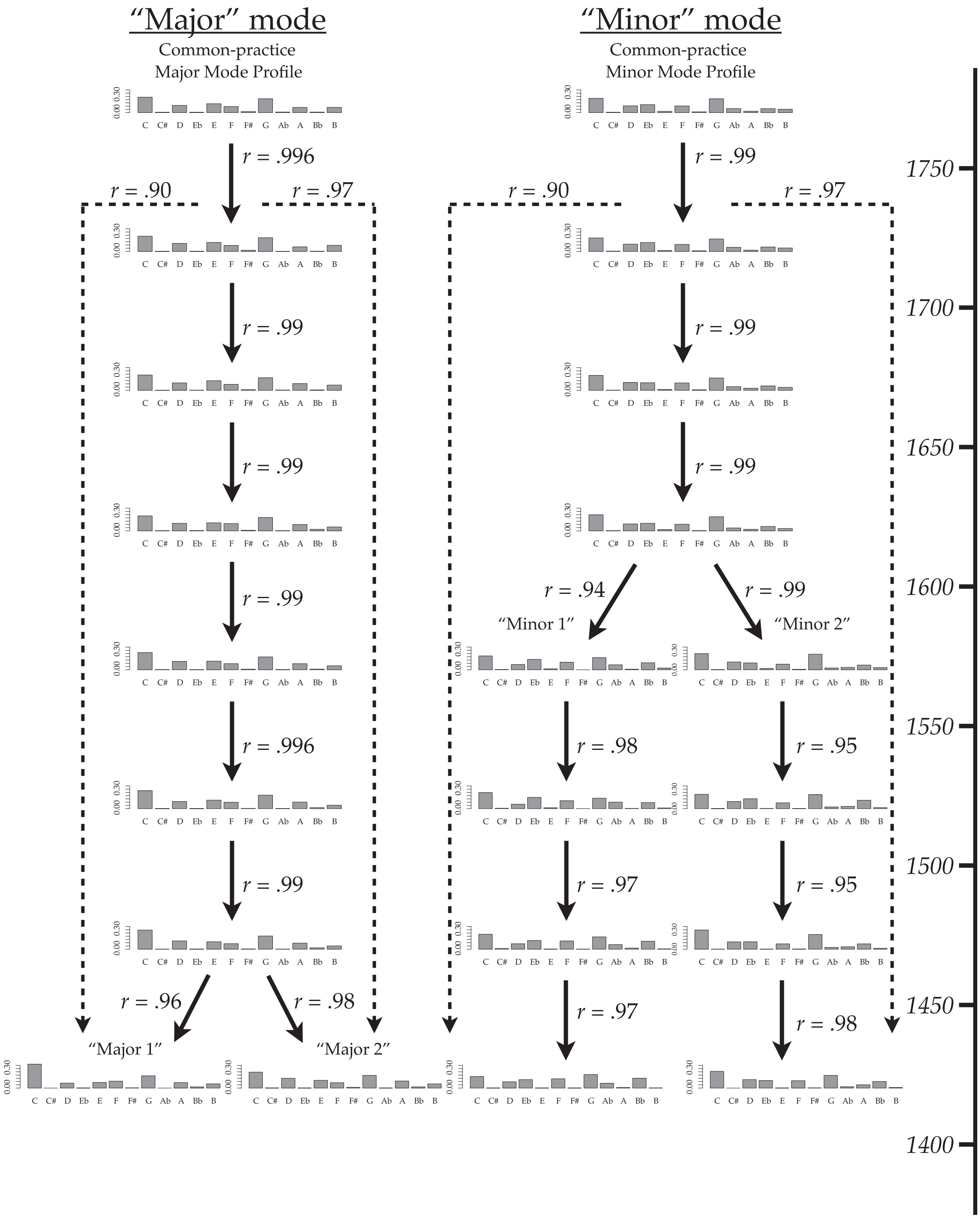

FIGURE 18. Graphical summary of historical changes in scale-degree distributions for musical samples from 1400-1750. *r* values indicate Pearson's correlation coefficient — a measure of mathematical similarity where 1.0 represents a perfect match. Dashed lines show correlations between initial and final epochs.

from 1400-1450 correlates broadly (but imperfectly) with the minor distribution from 1700-1750 (with a correlation of .90). In short, the method appears to trace some gradual evolution in scale-degree distributions, possibly suggesting a slowly changing musical practice.

The existence of four schematic distributions in the earliest epoch is broadly consistent with the notion that pitch-related schemas may have been more complicated in the 15th century compared with the 18th century. Perhaps the most salient result from the study is the implication that the modern major and minor modes both originated from the amalgamation of two earlier modes. The results imply that this amalgamation happened earlier for the major mode than for the minor mode.

Discussion

Caution is advised in making any historical interpretation from the results of this study. The major and minor modes likely arose through a complicated process involving historical, stylistic, perceptual, cognitive, and other factors. One should not be sanguine about the complexity of this historical development. Any comprehensive story about the development of the major and minor modes will need to draw on converging evidence from several independent sources employing contrasting methods. We do not presume that the method used in this study is the only way to address the history of the major and minor modes. However, one might expect that psychological research on the phenomenon of tonality will surely be part of this story.

The method used in this study might be dismissed by some scholars as unduly mechanical. Yet our method rests on foundations that, while not universally accepted in the field of music psychology, nevertheless represents the predominant view concerning the role of statistical learning in the perception of pitch-related schemas such as is evident in Western tonality.

With regard to future studies, the method employed here could be developed in a variety of ways. A larger musical sample would likely result in greater accuracy and increased validity. The sample might be extended further back in time, and could also be projected forward beyond 1750 into more recent times. Apart from tracing changes associated with atonality and various forms of extended tonality, the same method could be applied to changing scale forms used in popular music and jazz.

A number of refinements might be made in order to improve the reliability of key finding. For example, one might employ an iterative approach: going backward in time, then using the resulting mode distributions to go forward in time – and repeating this back-and-forth approach until some stable scenario emerges. At the risk of introducing anachronistic preconceptions, another approach might rely on the judgments of modern theorists to assign a tonic to each extracted excerpt. Although Toiviainen and Krumhansl (2003) found that first-order pitch transitions do not improve key finding for music by J.S. Bach, one might nevertheless expand the study to examine 2- or 3-note transitional probabilities, rather than relying exclusively on zeroth-order distributions. Additionally, one might enhance the key finding approach by employing a psychoacoustic model of pitch salience – following the work of Huron and Parncutt (1993).

Another approach would examine cadential passages separately from non-cadential passages, which may exhibit different scale-degree usage. With a larger sample of musical works, one might be able to compare trends in different countries or cultural regions, or compare possible differences among individual composers. With a sufficiently large musical sample, one might even be able to identify the geographical locus for the origin of the major and minor modes, or to even trace its geographical spread over time. With a sufficient volume of sampled musical materials, one might also examine shorter epochs (e.g., 20-year periods) or differences in usage in different genres, and these might provide greater temporal resolution chronicling the historical changes.

Author Note

The authors would like to thank Robert Cudeck for his help with the statistical analyses of the dataset.

Correspondence concerning this article should be addressed to Joshua Albrecht, Music Department, The University of Mary Hardin-Baylor, 900 College Street, Belton, TX 76513. E-mail: jalbrecht@umhb.edu.

References

AARDEN, B. (2003). *Dynamic melodic expectancy* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

ALBRECHT, J., & SHANAHAN, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception, 31*, 59-67.

BELLMAN, H. (2005). About the determination of key of a musical excerpt. In R. Kronland-Martinet, T. Voinier, & S. Ystad (Eds.), *Proceedings of Computer Music Modeling and Retrieval* (pp. 187-203). Pisa, Italy: Computer Music Modeling and Retrieval.

BUDGE, H. (1943). *A study of chord frequencies based on the music of representative composers of the eighteenth and nineteenth centuries* (Unpublished doctoral dissertation). Columbia University Teacher's College, New York.

CASTELLANO M. A., BHARUCHA, J. J., & KRUMHANSL, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General*, *113*, 394-412.

EEROLA, T. (2004). Data-driven influences on melodic expectancy: Continuation in north Sami Yoiks rated by South African traditional hearers. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 83-87). Evanston, IL: Casual Productions.

HAUSER, M. D., NEWPORT, E. L., & ASLIN, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*, B53-B64.

HILEY, D. (2013). Bar. *Grove music online. Oxford music online.* Oxford University Press. Retrieved from http://0-www.oxfordmusiconline.com.umhblib.umhb.edu/subscriber/article/grove/music/01972.

HURON, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

HURON, D., & PARNCUTT, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology*, *12*, 154-171.

HURON, D., & VELTMAN, J. (2006). A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, *1*, 33-55.

KESSLER, E. J., HANSEN, C., & SHEPARD, R. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception*, *2*, 131-165.

KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

KRUMHANSL, C. L., & KESSLER, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*, 334-368.

LANTZ, M. E., KIM, J.-K., & CUDDY, L. L. (in press). Perception of a tonal hierarchy derived from Korean music. *Psychology of Music*.

NAM, U. (1998). Pitch distribution in Korean court music: Evidence consistent with tonal hierarchies. *Music Perception*, *16*, 243-247.

PERTTU, D. (2007). A quantitative study of chromaticism: Changes observed in historical eras and individual composers. *Empirical Musicology Review*, *2*, 47-54.

PEARCE, J. M., & BOUTON, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, *52*, 111-139.

SAFFRAN, J. R., JOHNSON, E. K., ASLIN, R. N., & NEWPORT, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27-52.

SCHAFFRATH, H. (1995). *The Essen folksong collection.* (D. Huron, Ed.) Stanford, CA: Center for Computer Assisted Research in the Humanities.

TEMPERLEY, D. (2007). *Music and probability.* Cambridge, MA: MIT Press.

TOIVIAINEN, P., & KRUMHANSL, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*, 741-766.

TYMOZCKO, D. (2011). *A geometry of music: Harmony and counterpoint in the extended common practice*. New York: Oxford University Press.