

## BUILDING A REPRESENTATIVE CORPUS OF CLASSICAL MUSIC

---

JUSTIN LONDON  
Carleton College

**THIS PAPER PRESENTS AN OBJECT LESSON IN THE** challenges and considerations involved in assembling a musical corpus for empirical research. It develops a model for the construction of a representative corpus of classical music of the “common practice period” (1700-1900), using both specific composers as well as broader historical styles and musical genres (e.g., symphony, chamber music, songs, operas) as its sampling parameters. Five sources were used in the construction of the model: (a) *The Oxford History of Western Music* by Richard Taruskin (2005), (b) amalgamated *Orchestral Repertoire Reports* for the years 2000-2007, from the League of American Orchestras, (c) a list of titles from the Naxos.com “Music in the Movies” web-based library, (d) Barlow and Morgenstern’s *Dictionary of Musical Themes* (1948), and (e) for the composers listed in sources (a)-(d), counts of the number of recordings each has available from Amazon.com. General considerations for these sources are discussed, and specific aspects of each source are then detailed. Intersource agreement is assessed, showing strong consensus among all sources, save for the Taruskin *History*. Using the Amazon.com data to determine weighting factors for each parameter, a preliminary sampling model is proposed. Including adequate genre representation leads to a corpus of  $\approx 300$  pieces, suggestive of the minimum size for an adequately representative corpus of classical music. The approaches detailed here may be applied to more specialized contexts, such as the music of a particular geographic region, historical era, or genre.

*Received: October 21, 2012, accepted March 13, 2013.*

**Key words:** corpus, classical music, common practice period, historical styles, genres

---

**A**S THE OTHER ARTICLES IN THIS SPECIAL volume attest, corpus-based studies are becoming both more common and more important in music psychology and empirical musicology, whether for music psychologists choosing a set of stimuli, for computer scientists wanting to train neural networks,

or for ethnomusicologists interested in documenting our musical enculturation. In some instances a specialized corpus may be desired (e.g., Patel & Daniele’s 2003 study of rhythmic structures characteristic of French versus British composers, or Huron & Ommen’s 2006 study of syncopation in early 20th century American music). For many other studies, however, it would be highly useful to have a corpus that can be used to model more generic aspects of musical structure and musical experience. For it is a widely held tenet of music theory and musicology that the music of 1700-1900, the so-called “common practice period,” employs a common harmonic, melodic, and rhythmic syntax. The grammatical coherence of this music is assumed by music theorists ranging from Schenker (1935) to Lerdahl (2001) and music psychologists from Helmholtz (1877/1954) to Krumhansl (1990). And indeed, most classical music that is performed today comes from this era, along with music of the 20th and 21st centuries that also employs in large part the same tonal and rhythmic language. This, then, is what is meant in most contexts by “Classical Music,” and so the aim of the present article is to present an object lesson in attempting to build a corpus that is broadly representative of the classical composers, styles, and genres that are most familiar to the 21st century listener. Such a corpus, or a model for building it, should be based on the music of the common practice era that we are most likely to hear. Where, then, do we hear classical music?

We hear classical music in a wide range of contexts. As Auslander (2008) has pointed out, most of our experiences of the performing arts in the late 20th and 21st centuries, whether of music, dance, opera, or theater, come via broadcast, podcast, or recordings. Knowing what classical recordings are sold/downloaded would give a key insight into the current classical music landscape. Playlists and programs for radio and internet broadcasts and podcasts are similarly useful sources of information. Of course classical music is heard live, and a survey of concert programs from orchestras, opera companies, and chamber music societies would show what works, representative of different musical genres, are performed most often by these institutions.

While recordings, radio, and concerts represent our active consumption of classical music—for when attending a concert or listening to a recording the music

is (presumably) the primary object of attention and interest—we also encounter classical music in other contexts as more passive listeners. Classical music is used in film and television scores, in advertisements, as backing music to ceremonies and special events (ranging from weddings and graduations to the Super Bowl and the World Cup), and as ambient music in elevators, shopping malls, and parking lots. This indirect exposure to classical music adds to our statistical learning of musical idioms and styles, and the importance of statistical learning for our acquisition of basic elements of musical syntax has been documented in a number of studies (e.g., Bigand & Poulin-Charronnat, 2006; Creel, Newport, & Aslin, 2004; Hay & Saffran, 2012; Jonaitis & Saffran, 2009; Loui, Wessel, & Kam, 2010). Listening in these indirect contexts may also forge strong semantic associations between particular musical idioms and their social connotations, ranging from the funereal (Barber's *Adagio for Strings*) to the triumphant (Richard Strauss's *Also Sprach Zarathustra*). Thus a survey of film music, commercial music licensing, and common ceremonial pieces (e.g., surveying collections of classical music used at weddings) would provide a sense of our passive, as opposed to active, consumption of classical music.

There are other indirect measures of the contemporary classical music landscape. These include standard music histories often used as college textbooks, trade publications written for the layman (e.g., Goulding, 1995), and various classical record buying guides. For the conservatory student there are books of repertoire excerpts; that is, collections of difficult passages from the standard classical repertoire they are expected to have mastered as working musicians. And there are collections of classical musical themes, the most widely known (at least among music psychologists) being Barlow and Morgenstern's *A Dictionary of Musical Themes* (1948). While these are not direct measures of our musical environment, it is reasonable to presume that they are a reflection of it. They are compiled with a didactic aim—the music one “ought to know”—and one of the main reasons one needs to know this music is because one is apt to encounter it as a performer or listener.

By combining information from a range of sources noted above one may produce a profile of early 21st century North American classical music consumption. The next section of this paper discusses some general considerations regarding the various sources available for building a classical music demographic. Profiles of the specific sources used in this study are then given, and the relative agreement/consistency among them is assessed. A set of sampling parameters is proposed and

then refined. The paper concludes with a consideration of how they may be further improved, as well as caveats for their use.

### Sources Used in this Study

Seven sources were examined in detail in this study, five of which were used in the construction of a sampling model; they include both indirect and direct measures of classical music consumption. The indirect sources are: (a) The Barlow and Morgenstern *Dictionary* (1948), (b) two standard music history text books, Taruskin (2005) and Burkholder, Grout, and Palisca (2009), and (c) a book of orchestral excerpts for the aspiring professional violinist, edited by Gingold (1962). Burkholder et al. and Gingold were not used in constructing the model for reasons detailed below. The direct sources are: (a) the collection of pieces of classical music used in film scores available via Naxos.com, (b) a consolidated list of North American orchestra programs from 2000–2009, complied by the League of American Orchestra, and (c) counts of the number of audio recordings by each composer considered in the other sources, both direct and indirect, obtained from Amazon.com.

A few general observations and caveats are worth noting before examining individual sources. First, and perhaps foremost, all sources are historical, as they reflect the musical preferences and priorities of a specific time and place. The dates for the sources considered here range from 1948 (the Barlow & Morgenstern *Dictionary*) to August 2012 (data culled from the Amazon.com website). Likewise all sources, with the possible exception of Amazon.com, have a critical bias, as most have an implicit if not explicit intention that they contain/list examples of “good” music. Indeed, in some cases (e.g., music histories; lists of “must have” recordings) works that are not often performed are nevertheless included, as they are deemed to have high aesthetic or artistic value, even if unpopular.

There are also some domain-specific or institutional constraints and biases that need to be acknowledged. Some sources will have innate biases for a given musical style or genre: a chamber music society will produce concerts of chamber music, opera companies produce operas, and orchestras play symphonies and concertos. Likewise ensembles and institutions may focus on a particular historical or national style (e.g., early music ensembles, 20th century ensembles, national symphony orchestras that dedicate all or part of their programming to music by citizen composers and musicians, and so forth). There are also constraints of specialization: orchestral works—at least those written before the

20th century—tend to be long, and therefore only a few works can be performed on any given concert; indeed, for extremely long works (e.g., Bruckner symphonies), an entire concert may consist of a single piece. For this reason Matthews (2009), in his survey of programming trends in U.S. orchestras, takes the playing time into account, noting how longer 18th and 19th century works may “crowd out” more recent compositions. Classical radio broadcasts, by contrast, often tend toward shorter works, both in terms of historical style (i.e., skewed to music from the Baroque and early Classical periods), and by programming single movements from longer works.

While the primary concern of this study is which composers should be included in a representative sample, and by how many pieces, the balance of historical eras and musical genres is also considered. For classical music is comprised of a number of distinct substyles, even if in the popular imagination it tends to be dominated by 19th century orchestral repertoire (and for good reason, as we shall see). While there are many common aspects of tonal practice in the classical repertoire, there are also style- and genre-specific idioms that should be fairly represented in a comprehensive sample of classical music. Therefore one should be cognizant of the balance of historical styles (e.g., Baroque versus Classical versus Romantic) and genres (symphonies versus sonatas versus songs) each source contains. Finally, while potentially rich sources of data exist in the form of sales information (Amazon.com and iTunes sales figures for disc and/or download purchases) and licensing information (e.g., BMI/ASCAP figures for commercial use of classical music recordings), such information is typically proprietary, and hence were not available for this project, at least in its current stage.

#### BARLOW & MORGESTERN'S A DICTIONARY OF MUSICAL THEMES

There are many reference guides for music lovers of every genre, often couched in terms of a list the “Greatest Composers and their Greatest Works” (e.g., Goulding, 1995) or record guides for various styles and genres (Brackett & Hoard, 2008; Cook & Morton, 2008). Barlow and Morgenstern's *Dictionary of Musical Themes*, first published in 1948 and subsequently reprinted many times, is a reference guide of a different sort. Consisting of a list of some 10,104 themes by 319 composers, it allows the user to either (a) look up the theme for a given piece (e.g., “what is the tune from the beginning of the third movement of Brahms' third symphony?” A: see tune B1611), or, (b) by making use of an index at the back of the book, play a tune (transposed to the key of C), and then look up the requisite theme

(e.g., “what tune starts E-E-F-E-E-A-E-E-G-F-D?” A: see tune B1611). This *Dictionary* includes themes of instrumental works: symphonies, solo sonatas (especially works for piano), and chamber music, though it is strongly biased towards orchestral music. Barlow and Morgenstern subsequently published a separate *Dictionary of Opera and Song Themes* in 1950.

Barlow and Morgenstern note that their *Dictionary* is not exhaustive: “certain works were omitted because the scores were unavailable in libraries, and publishers who were more than helpful could not supply them. A few other works we left out because we could not, after great effort, secure copyrights” (p. xi). Showing their critical/aesthetic bias, they also noted that “a few ultramodern works we left out. We felt that anyone likely to remember their themes, or more aptly their combinations of notes, would in all probability know their source” (p. xi).

The specific data set used in this study was derived from David Huron's “\*\*kern” encoding of the first edition of the *Dictionary* (Huron, 2013). The main methodological issue here was whether to rank composers in terms of the number of pieces or themes included in the *Dictionary*, as the 10,104 themes are from 2,339 pieces. For most composers (276 of 319) the percentage rankings are commensurate (i.e., when ranked in terms of percent of the population—either pieces or themes—the absolute value of the difference is less than .25%, and less than .10% for 224 of 319 composers). Discrepancies, which are few, arise when a composer has a large number of pieces but relatively few themes (e.g., Domenico Scarlatti, who published many monothematic harpsichord sonatas) or vice-versa (e.g., Richard Strauss, whose complex symphonic works involve many themes). To be commensurate with the other sources considered, and especially as the composers affected by this discrepancy were not in danger of exclusion, being highly ranked in terms of both themes and pieces, this study used only the piece-based ranking of composers in the *Dictionary*. Appendix 1 lists the top 106 composers in the *Dictionary*, all of whom had 5 or more pieces listed. The top 15 composers account for 42.58% off all the pieces listed in the *Dictionary*.

#### MUSIC HISTORY TEXTBOOKS & OTHER DIDACTIC SOURCES

*The Oxford History of Western Music* (2005) is Richard Taruskin's magisterial five-volumes-plus-appendix comprehensive history of classical music from the 9th century to the end of the 20th. Taruskin's aim is nothing less than to give a historical and analytical account of Western music from when it was first written down until, as he sees it, the end of the written musical

TABLE 1. Distribution of Pieces According to Historical Eras.

Historical Era	Taruskin Examples	Taruskin Ex %	Grout Examples	Grout Ex %	Amazon %
Medieval	53	5.24	19	8.19	.08
Renaissance 1450-1600	117	11.57	44	18.97	1.48
Baroque 1600-1750	137	13.55	46	19.83	13.43
Classical 1750-1800	49	4.85	26	11.21	12.97
Early Romantic 1800-1850	131	12.96	22	9.48	12.83
Late Romantic 1850-1900	169	16.72	21	9.05	37.19
20th Century	355	35.11	54	23.28	22.01

Note: Information given both in terms of raw counts and in terms of relative percentages in (a) Taruskin, (b) Grout, and (c) the Amazon Data Set given in Appendix 8.

tradition, a tradition now superseded by recording technology and other modes of transmission. Naturally, over the course of this history Taruskin discusses the work of many composers, both major and minor, but as a means of winnowing the list of composers for this project, only those composers whose music merited a musical example or figure were counted. This still resulted in a list of 1,011 examples by 265 composers. Counting these examples is not completely straightforward, as in some instances several pieces (or movements of a larger piece) by one composer might be included one subdivided example, while elsewhere they might be each given separate examples (due to considerations of layout, text discussion, etc.). Where appropriate, subdivisions of large examples were counted separately. Didactic examples were excluded—e.g., the use of a piece as illustration of a particular mode or cadence type.

The distribution of Taruskin's examples in terms of various historical eras is given in Table 1. As befits a comprehensive history Taruskin includes many examples of music written before 1700, music that pre-dates the common practice period that is nominally the focus of our target corpus. Thus it might make sense to exclude those examples of music written prior to 1700 given in Taruskin (and similar sources). However, there are several reasons for including the earlier works, at least for now. First, while one rarely encounters this music in recordings or concerts today (as will be seen in comparison with other sources), performance of early music is still a part of the current musical landscape, and thus Taruskin is a useful source for composers and pieces that might otherwise be overlooked. Second, it is useful to understand the extent to which sources like Taruskin include pre-common practice period examples, and thus become aware of what a purely random sample from such sources might yield.

Appendix 2 gives a list of 91 composers who merited three or more examples in Taruskin; 28 had 10 or more examples. Taruskin's top three are: Beethoven (43 examples, 4.25% of all examples), Stravinsky (39,

3.86%), and Schoenberg (35, 3.46%). While Beethoven's high ranking makes sense both in terms of historical and critical significance, as well as the continuing ubiquity of his music in modern musical culture, the weight given to Stravinsky and Schoenberg is telling. While Stravinsky is an important composer whose music is still often performed, it is also true that Taruskin is a specialist in Stravinsky (and Russian music more generally), and this may be a case where Taruskin's research interests affected his choice of examples; as can/will be seen, Stravinsky's very high rank here is somewhat at odds with his placement in other sources. In keeping with most music histories, Taruskin also gives great weight to the so-called second Viennese school (Schoenberg, Berg, and Webern—Haydn, Mozart, and Beethoven comprise the first) based on their putative historical importance in the development of atonal music and the modernist musical aesthetic. Schoenberg's high ranking here is even more at odds with his much lower prominence in other sources.

The 8th edition of *A History of Western Music*, by Burkholder et al. (2009) was also consulted, as "Grout," as it has been known to generations of music students, has served as musicology's industry standard for a one-volume music history for more than three decades. However, recent editions of Grout have embraced a pedagogy of including fewer but longer musical examples; in many cases they are complete works that are included in a set of companion volumes (*the Norton Anthology of Western Music*). Thus Grout "only" includes 232 examples by 163 composers, with a historical distribution quite similar to that found in Taruskin, though with a greater bias toward music written prior to 1600 (see Table 1). Appendix 3 lists all composers who merited two or more musical examples in Grout and/or its companion anthologies.

The three volumes of *Orchestral Excerpts from the Symphonic Repertoire* edited by Gingold (1962) comprise a didactic source of a different sort. As the title suggests, these volumes contain passages from standard

orchestral repertoire (i.e., the especially tricky bits) that a member of an orchestra string section would be expected to know and be ready to play. As such, they are indicative of those works a working musician is likely to encounter, and, by extension, a concert going listener is likely to hear. Though now 50 years old, these volumes are still in heavy use by conservatory students. One hundred and ninety-six works by 60 composers are included; Appendix 4 lists the number of works included by all 60 composers. A possible source of bias here is that composers whose works are relatively easy to play are less likely to be included, as there is less practical need for a musician to devote significant amounts of time practicing easy music.

Grout and Gingold are included here as a comparison with sources with similar didactic aims—Taruskin, in the case of Grout, and Barlow and Morgenstern (given its emphasis on standard orchestral repertoire), in the case of Gingold. As can be seen from a comparison of the top ranked composers in these sources, they are roughly commensurate. However, given their small size (and the relatively restricted range of values in each), it is difficult to make any strong statistical comparisons between these and the other sources discussed here.

#### THE NAXOS LIBRARY OF "CLASSICAL MUSIC IN THE MOVIES"

Naxos is the largest independent classical recording label in the world, specializing in reissues of historic recordings as well as newly made recordings by lesser known ensembles. Naxos also runs a subscription service that allows various degrees of online access to their entire catalog; as such, Naxos can provide an instant library of recorded music for both individual and institutional subscribers. As of September 1st, 2012, the Naxos online music library included 75,947 discs and 1,092,850 tracks in a wide range of styles and genres, from Jazz and Blues to World Music.

From the Naxos Classical Music homepage one can access their collection of "Classical Music in Movies"<sup>1</sup> which provides the user with links to classical music used in a wide range of Hollywood films, from *Fanny and Alexander* to *Die Hard with a Vengeance*. These are not the film soundtracks per se, but links to "in-house" performances the same pieces that are available on other Naxos recordings. As of August 2012 the collection includes 775 works by 103 composers. The complete list of composers and a count of their works in the Naxos collection is given in Appendix 5. While it may be the case that for Naxos their "Classical Music in the Movies" collection is intended to provide a means of

entry into their broader catalog—a kind of teaser, if you will—it still provides useful information for the current study. While not as helpful regarding the relative weighting of entries, given its modest size, it is still useful in assessing which composers should be included in the sampling model developed below.

#### THE LEAGUE OF AMERICAN ORCHESTRAS CONSOLIDATED "ORCHESTRAL REPERTOIRE REPORT"

As noted on their website:<sup>2</sup>

The League of American Orchestras leads, supports, and champions America's orchestras and the vitality of the music they perform. Its diverse membership of approximately 850 orchestras across North America runs the gamut from world-renowned symphonies to community groups, from summer festivals to student and youth ensembles.

Among the many things the league does is maintain detailed records of those composers, pieces, and solo artists who are on every concert program of every orchestra that is a member of the league. Again, from the league website:

The Orchestra Repertoire Reports (ORR) list all classical season works performed by League member orchestras during the specified season, alphabetized by composer, then by the composer's works. The report also includes the name of the orchestra, conductor, and soloist(s), (if applicable), who performed the work, and the date(s) of the first performance. Each ORR includes a top ten list of most frequently performed works, soloists, and composers, etc.

The league kindly provided me with an Excel version of consolidated ORR data from nine seasons (from 2000-2001 through 2008-2009). This data set comprises 99,784 performances of works by 1,949 composers, taken from 582 orchestras (N.B., not all orchestras are included in every season). The top 52 orchestras (i.e., orchestras with 500 or more performances over the nine seasons surveyed) represent 65.1% of all of the performances in the data set.

Of the 1,949 composers in the data set, 28 had 1,000 or more performances of their works; while only 1.44% of all of the composers in the data set, they represent 60.94% of all of the performances. Conversely, over half of the composers have only four or fewer performances; these are almost exclusively 20th century composers

<sup>1</sup> [www.naxos.com/musicinmovies.asp](http://www.naxos.com/musicinmovies.asp)

<sup>2</sup> [www.americanorchestras.org/knowledge\\_research\\_and\\_innovation/orr\\_archive.html](http://www.americanorchestras.org/knowledge_research_and_innovation/orr_archive.html)

TABLE 2. *Ranking of Composers in Terms of the Number of Performances of their Works over a Nine Year Period Tracked in the ORR Data Set.*

Performances	Composers	Composer %	Performances	Performances %
> 1000	28	1.44	60812	60.94
500-999	14	0.72	10142	10.16
100-499	68	3.49	14385	14.42
50-99	61	3.13	4333	4.34
20-49	120	6.16	3728	3.74
10-19	157	8.06	2124	2.13
5-9	274	14.06	1808	1.81
2-4	710	36.43	1935	1.94
1	517	26.53	517	0.52

whose works were given an orchestral premiere and then never performed again. A breakdown of the number of performances per composer, in ranked categories, is given in Table 2. The leftmost column in the table gives categories of performance counts, the next two columns give composer counts in absolute numbers and relative percentages, and the rightmost two columns list the total number of performances by composers in each category. Appendix 6 lists the Top 100 composers in the ORR data set, including the total number of performances of their works and their percentage (relative to all of the performances in the data set). The top 100 composers account for 84.44% of all performances in the data set. Of the top 100, 13 are living composers whose works comprise 2.84% of all performances.

#### CORRELATIONS AMONG THE DATA SOURCES

Before moving on to the sampling model, it behooves us to look at the extent of agreement among these sources, both in terms of which composers each includes as well as their relative prominence. Here we are immediately faced with a problem, given the different sizes of these sources and the different repertoires each emphasizes. Moreover, in every data set, a relatively small number of composers (i.e., the “usual suspects” of Bach, Mozart, Beethoven, etc.) has a relatively large representation, while most others relatively few. Thus one will always get a fairly strong correlation measure, but one that is not terribly informative. Therefore, to make a meaningful comparison, a “top 55” aggregate list of composers was compiled. To construct this list, the top 28 composers from four sources (ORR, Taruskin, Naxos, and Barlow & Morgenstern; N.B. the top 27 were used from the ORR set) were combined into an aggregate list of 55 composers. These four sources represent both direct (ORR, Naxos) and indirect/didactic measures (Barlow & Morgenstern, Taruskin) of the current classical music listening environment. Grout and Gingold were not

used due to their small sample sizes, as noted above. Including the top 28 composers from each source not only produced an appropriately sized aggregate; it also aligned with meaningful breakpoints in each list (i.e., in ORR composers with 1000 or more performances; in Taruskin all composers with 10 or more examples; in Naxos all with 7 or more; in Barlow & Morgenstern with 17 or more). Of these 55 composers, 11 appeared in all four “top 28” lists; 6 in three of the four lists, 9 in two, and 28 were singletons, appearing on only one list. Taruskin’s list had the highest number of singletons (11), while the ORR data set had the fewest (3).

Once the aggregated list was completed, information from each source was tallied for the aggregated list; this included the scores of the top 28 as well as “missing values” for lower-ranked composers from each list (see Appendix 7). To make the values commensurate and to enable statistical comparison, for each composer their percentage representation within each data set was used as a score, rather than a simple count of examples or performances. The Median Percentage Ranking (MPR) was then calculated for each composer. The median rather than a grand average was used to mitigate the effects of outliers and due to the non-normal distribution of the scores (see below for details). In those cases where a composer did not appear at all in a particular data set they were given an entry of zero. A histogram that illustrates the MPR rankings of these 55 composers is given in Figure 1. As can be seen from both the histogram and the quadratic trend line ( $y = 4.75e^{-0.077x}$ ,  $R^2 = .85$ ) the seven top ranked-composers (Mozart, Beethoven, Bach, Brahms, Chopin, Tchaikovsky, and Schubert) all have very high representations (> 2%) in the MPR (and indeed, in all four source data sets); representation of other composers drops off steeply from there. The MPRs and their relation to a proper sampling model are discussed further in the following section.

Correlation comparisons among the four data sources and the MPR are given in Table 3. Spearman’s rho was

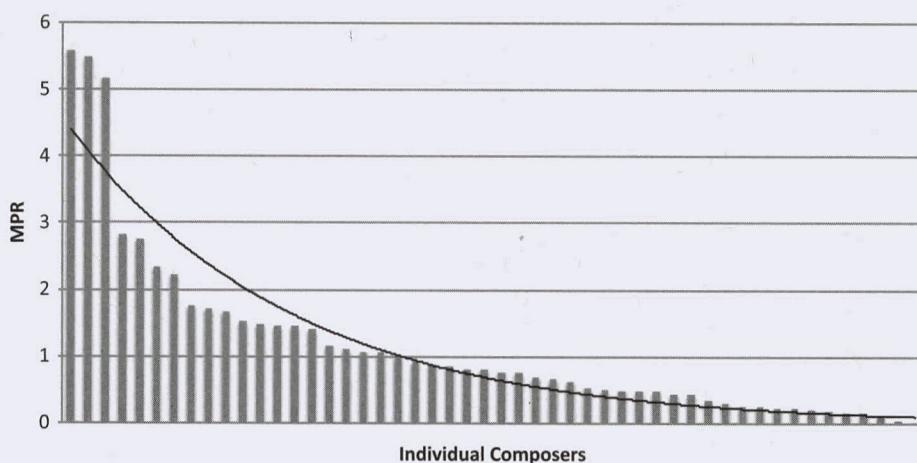


FIGURE 1. Histogram from highest to lowest median percentage of the 55 composers in the MPR (see Appendix 7); composers with a median of 0 were excluded.

used as nonparametric test of correlation, given that the distribution in each source was far from normal in terms of both skewness and kurtosis (skewness<sub>Z</sub> ranged from 4.10 to 8.07, and kurtosis<sub>Z</sub> ranged from 3.19 to 11.40). As can be seen from Table 3, as well as from a perusal of Appendix 7, Taruskin's data set is the odd man out. While there is a significant correlation (albeit of only  $r_s = .29$ ) between it and the MPR, Taruskin's correlations with the other three sources are nowhere near significant, whereas correlations between the other three are both large and highly significant.

### Toward a Sampling Model

As noted above, building a representative corpus classical music requires one to determine (a) which composers (or more generally, historical eras) should be included, some of whom will be obligatory (e.g., Bach, Mozart, Beethoven), and (b) how each composer (or

era) should be weighted. Each source consulted gives both kinds of information, for as noted above, certain composer's works are given more examples (in the Taruskin volumes), appear on more concert programs (in the ORR data set), or have more themes/pieces listed (in Barlow & Morgenstern). As each source has a distinct bias—whether it is Taruskin's pedagogical/historical orientation, or the fact that the ORR data only includes orchestral music—one needs a means of looking beyond any single source in order to solve both the inclusion and the weighting problems. Initially, one might think a distribution profile based on the MPR might work, as it summarizes four sources of information. Yet there is another, broader source of information regarding the relative prominence of individual composer of classical music in contemporary culture: Amazon.com.

Amazon.com, the online vendor that sells everything from books to gourmet groceries, presumably needs no

TABLE 3. Spearman's rho Correlations Among the Four Components of the MPR.

		ORR	Taruskin	Naxos	B&M	MPR
ORR	$r_s$ value	1.000	.15	.60**	.49**	.84**
	Sig. (2-tailed)		.27	.000	.000	.000
Taruskin	$r_s$ value		1.000	.01	-.04	.29*
	Sig. (2-tailed)			.95	.78	.009
Naxos	$r_s$ value			1.000	.55**	.80**
	Sig. (2-tailed)				.000	.000
B&M	$r_s$ value				1.000	.73**
	Sig. (2-tailed)					.000
MPR	$r_s$ value					1.000
	Sig. (2-tailed)					

Note: N = 55 for all cases.

introduction to the readers of this journal. For books and music, Amazon has largely replaced Bowker's *Books in Print* as an up-to-the minute index of what titles are available (and if not in print, Amazon will seamlessly link the user to used or antiquarian booksellers who have used copies of the title in stock). It provides the same indexing information for titles in music. For all composers considered in this study, a count of the number of recordings (regardless of format—LP, CD, DVD, or download) was taken from the Amazon website, accessed between July 1 and August 30, 2012.

Using the number of Amazon.com titles, as opposed to sales/download data (as useful as those data might be) is based on the following rationale. The decision to learn, perform, and record the work of a given composer represents a significant commitment of time and resources by both the performing musician and the record label that releases the recording. While some works may wholly be the product of self-interest, as in the case of a composer who records and releases his/her own compositions, the presence of a significant number of recordings by a particular composer is indicative of the sustained interest in their music by the musicians who make them and the listeners who purchase them. Looking at recordings rather than sales figures also finesse the problem of short-term fluctuations in the popularity/representation of a given composer, as short-term sales of classical music are affected by films and television programs where a particular piece or composer's works may be featured, historical occasions, such as significant "birthday" year when a composer's works are often recorded, and ad hoc choices by popular performers (e.g., a pianist decides to record a complete cycle of Beethoven's sonatas).

The Amazon.com website allows for advanced searches, useful for classical music, as it allows one to specifically search for titles that contain works by a particular composer, as opposed to performer (or the more ambiguous "artist" metadata tag). To compile each title count a search was first initiated using a composer's name, e.g., "Englebert Humperdinck." This search was then refined by examining individual titles until Amazon's "(composer)" metadata tag was found, and then this link, and the title count it produced, was consistently used for the current study. Of course this approach is not perfect, for given a database as large as Amazon.com's list of titles in music, or even just classical music, there are bound to be mistakes; this study alone generated 226,056 titles, with counts ranging from 0 in the case of Ion Ivanovici to 12,457 titles for Mozart.

Errors and other noise in the Amazon data may arise for a number of reasons. First, some musicians, such as

Leonard Bernstein, are both composers and performers; while it is fine to count a recording of Bernstein conducting his own music, one would not want to include Bernstein's performances of Mahler symphonies in the "Bernstein" composer/title count. Next, there are composers who share a name with other composers or performers: Englebert Humperdinck is the name of an Anglo-Indian singer popular in the 1960s and 70s, as well as a 19th century German composer of operas. Likewise some composers, especially Russian or Slavic composers, often have variant spellings of their names (Scriabin vs. Skryabin), and other sources of confusion (e.g., J. S. vs. Johann Sebastian Bach; Johann Strauss Sr. vs. Jr.). For example, Johann Strauss Jr. (the "Waltz King") was but one member of the musical Strauss family, which included three Johanns and two Eduards. An Amazon keyword search (on 9.02.12) for "Johann Strauss Jr." yielded 283 results, while a keyword search on "Johann Strauss ii" yielded 2,367 hits. Selecting the "Johann II [Junior] Strauss (Composer)" link yielded 887 hits, and while that number may be a bit low, it is the one used in this study. Similarly, "Johann Sebastian Bach" (which serves as the Amazon uniform composer name) yields 11,063 results, while "J. S. Bach" yields 7,677 results, some of which are listed as Johann Sebastian Bach; it is unclear the extent of the overlap between the searches, though it would seem to be considerable. A title counted regardless of whether the composer had a single track/work on it or if the title exclusively contained works by that composer. While on the one hand, one might want to make fractional counts for titles that contain works by more than one composer, an appearance on various anthology albums (e.g., *The Most Relaxing Classical Music in the Universe*, which includes works by Pachelbel, Mozart, Debussy, Albinoni, Chopin, Vaughan Williams, Liszt, Tchaikovsky, Dvorak, Janacek, Bach, and Beethoven) is another indication of a composer's cultural prominence.

Thus to serve as a counter to the rankings produced in each of the other sources examined, *all* of the composers listed in appendices 1-6 were combined into a list of 221 composers and an Amazon title count was obtained for each. These data are given in Appendix 8, henceforth referred to as the Amazon Data Set or "ADS." The first thing we may do with the ADS, given its large size and the historical range of composers it includes, is to consider the distribution of its titles in terms of historical eras (see Table 4). While it is true that Medieval and Renaissance composers are few in number, that does not explain the paucity of their representation in terms of titles; the Classical and Early Romantic Eras have fewer composers, but far more titles per composer. For each

TABLE 4. *Distribution of Historical Styles in the Amazon Data Set (ADS) by Composer and Title.*

Era	Composer Count	Title Count	Title %
Medieval (pre 1450)	3	185	0.08
Renaissance (1450-1600)	15	3352	1.48
Baroque (1600-1750)	26	30349	13.43
Classical (1750-1800)	10	29306	12.97
Early Romantic (1800-1850)	10	28993	12.83
Late Romantic (1850-1900)	68	84028	37.19
20th Century	89	49844	22.01

composer in the ADS their Amazon title count was then used as a weighting factor, given both in terms of the absolute number of titles and as their percentage of entire ADS. Thus, for example, Mozart's 12,457 titles represent 5.51% of the ADS, and the top 25 composers (11.31% of all composers) account for 53.85% of all titles. Conversely 192 of the 221 composers listed each have less than 1% of the pieces in the ADS. The pattern of title counts in the ADS from most to least follows a logarithmic distribution, similar to that found in the MPR histogram given above ( $y = -1677.65\ln(x) + 8420.30$   $R^2 = 0.87$ ;  $F_{1,216} = 1447.23$ ;  $p < .001$ ).

A comparison between the top 50 composers in the ADS and the MPR is most informative (see Appendix 9). While there is a strong agreement between the two ( $r_s = .79$ ,  $p < .001$ , two-tailed;  $N = 50$ ), this naturally arises for the reasons noted above, namely (a) the "usual suspects" are prominent in the top half of each data set and (b) all of the other composers have relatively low counts. The MPR includes 17 composers who are not in the ADS top 50 (N.B. not listed in Appendix 9); most are from the Taruskin set (representing both pre-Baroque composers and composers of atonal music, noted above), though de Falla, included in the MPR, is ranked 52nd in the ADS (Pachelbel and his canon are ranked 51st). More importantly, the MPR omits 13 composers included in the top 50 of the ADS: Donizetti, Gounod, Massenet, Gershwin, Franck, Purcell, Telemann, Mascagni, Bruckner, Weber, Offenbach, Vaughan Williams, and Leoncavallo. What is telling is that beyond a few somewhat less prominent composers of orchestral music (Gershwin, Franck, Bruckner, Weber, and Vaughan Williams), the bulk of the remaining composers are (a) French or Italian—i.e., not German, and (b) associated with opera. Thus there is a pro-German, pro-symphony bias in the MPR.

A crude but straightforward approach to producing a classical music sampling model would thus be to use the top 50 composers in the ADS, and the distribution of historical styles in the top 50 does compare fairly well

in terms of overall historical distribution of pieces in the full ADS (see Table 5). As can be seen, Classical and Romantic composers are somewhat over-represented, which compensates for the somewhat greater under-representation of 20th century composers. Nonetheless, a sample based on the top 50 would be a reasonable approximation of the current classical musical landscape.

This simple model begs the question, however, of having a principled determination of which composers to include. What seems clear from both the ADS and MPR is that while the inclusion of works by certain composers would seem to be obligatory, a rationale rather less arbitrary than choosing "the top 50" is needed. The distribution of pieces according to historical styles can also be better fitted to the overall ADS distribution. Therefore, to produce a more general model, we need to more clearly define a list of obligatory composers, establish a principle for weighting their representation in the corpus (which must be cashed out in whole numbers), and then define an algorithm for selecting additional composers/pieces.

For a first approximation, let us aim for a corpus of 100 pieces. If we return to the full ADS, and then round each composer's title count percentage to the nearest half percent, 35 composers are equal to or greater than 1%, which gives a manageable number of composers who should have at least one piece included in a corpus of 100 pieces. The list of 35 composers, their .5% rounded percentages, and their historical eras are given in Appendix 10. The ADS-based percentages of these 35 composers sum to 64%, and this suggests an approach to the construction of a sampling model: if these 35 composers are obligatory and will comprise 64% of the corpus, then the remaining 36% of model may be chosen freely, albeit with some constraints on sampling from various historical eras. The distribution of historical eras in the "obligatory 35" is given in Table 6. Here the rightmost column gives whole-number values for the number of additional pieces from each historical era

TABLE 5. *Percentage Representation of Historical Eras of the Entire ADS Versus the Top 50 Composers of the ADS.*

Era	Entire ADS %	Top 50%	Difference
Medieval	0.08	0.00	-0.08
Renaissance	1.48	0.00	-1.48
Baroque	13.43	13.35	-0.08
Classical	12.97	16.03	3.06
Early Romantic	12.83	16.79	3.96
Late Romantic	37.19	40.10	2.91
20th Century	22.01	13.84	-8.17

TABLE 6. Distribution of Historical Eras in the ADS Top 35 in Comparison to the Entire ADS.

Historical Era	ADS %	Top 35%	ADS - 35%	Remainder Weights
Medieval	0.08	0	0.08	0
Renaissance	1.48	0	1.48	1
Baroque	13.43	8.50	4.93	5
Classical	12.97	11.50	1.47	2
Early Romantic	12.83	10.50	2.33	2
Late Romantic	37.19	25.00	12.19	12
20th Century	22.01	8.50	13.51	14

needed to balance the distribution in the top 35, and thus produce a historically representative sample based on an initial sample size of 100 pieces. As can be seen, in the classical and early romantic eras the 35 closely match the historical distribution of the entire ADS—call this the “Mozart/Beethoven Effect.” Medieval and Renaissance percentages also match, alas, because in both data sets they are largely absent. Thus taking the historical biases of the top 35 composers into account, the remaining 36 pieces (for a corpus of 100) should be chosen according to the “remainder weights” given in Table 6. Each non-obligatory composer would be represented by a single piece in the corpus. The full ADS list would be a useful/practical source for names of appropriate non-obligatory composers from each historical era. Appendix 10, in conjunction with Table 6, might be used as the basis for building a representative corpus of 100 pieces of classical music.

Is 100 pieces enough, however? If the goal is to have a reasonably complete set of the harmonic, melodic, rhythm, and formal idioms that are part and parcel of the common practice period, then there are two shortcomings with a corpus this size. The first is that there are problems in choosing the singleton entries for the “36,” as one would need to determine and then use the “most representative” piece by Mussorgsky or Wagner or Bernstein [N.B. one might address this simply by choosing their work(s) have the most recordings or that are programmed most often]. The second and more significant problem has to do with the proper distribution of musical genres in the corpus: it would be problematic to have a corpus with no examples of opera or chamber music, which might well be the case in a corpus of only 100 pieces. This is true for the corpus as a whole as well as for individual composers with multiple entries in the corpus.

Consider Mozart. The first problem we face is simply coming up with an accurate list of Mozart’s compositions. Musicologically, this is a non-trivial problem: Do we count a sketch as a work? Should different versions of an opera or symphony count as separate works?

Mozart’s works were famously catalogued by Ludwig Köchel, who listed 626 works. Köchel himself was keenly aware that there were problems of dating and authenticity in his own catalog, and ever since the catalog was published in 1862 scholars have been correcting it. The current list of works in the *Grove Online Dictionary of Music and Musicians* (Eisen et al., 2013) excludes works originally included in Köchel’s catalog that are now known to be spurious, includes works by Mozart that Köchel omitted, and gives additional enumerations for pieces that share a common Köchel number. For the current study, Mozart’s pieces were counted and grouped according to the genre headings in the *New Grove*, yielding a total of 630 works (see Table 7). Keeping in mind that our aim is to produce a corpus that represents the music that a 21st century listener is most likely to hear, a subset of this list was chosen based on those genres that are currently most popular. These are given in Table 8, which takes the count of pieces from the *Grove* worklist, and recalibrates the percentage of pieces in each genre relative to this subset of Mozart’s oeuvre. As these works differ considerably in scope, a set of weighting factors is proposed—recognizing that one opera may include as much musical material as several symphonies or masses. A quasi-Boolean search in Amazon.com based on “composer” and “title” (keyword) fields produced the Amazon title counts given in Table 8. While these numbers should be used with even more caution than the Amazon composer counts, the percentages for symphonies, operas, piano concerti, and string quartets in the Amazon “Mozart Genre Set” align fairly well with the weighted counts derived from *Grove*. Serenades are somewhat over-represented (call this the “Eine Kleine Nachtmusik” effect), and masses, though Mozart’s most significant choral genre, are predictably under-represented, given the reduced presence of church music in 21st century musical life.

Further problems are evident if we look more closely just within Mozart’s symphonies. Appendix 11 lists all of the performances of each of Mozart’s multi-movement symphonies based on the ORR data set.

TABLE 7. *List of Mozart's Compositions, Based on Eisen et al. (2013).*

Genre	# of Pieces	% of Total
Masses & Mass Mvts	21	3.33
Psalms, Motets	30	4.76
Church Sonatas	17	2.70
Oratorios	7	1.11
Operas	22	3.49
Ballets	5	0.79
Vocal Ensemble Pieces	15	2.38
Concert Arias	58	9.21
Songs	38	6.03
Canons	38	6.03
Symphonies and Symphonic Mvts.	56	8.89
Serenades (Strings)	23	3.65
Divertimenti (Winds)	14	2.22
Marches	18	2.86
Dances	42	6.67
Piano Concertos	25	3.97
Other Concerti	23	3.65
Wind Chamber	6	0.95
Str Quintets	6	0.95
Str Quartets	27	4.29
Str Other	23	3.65
Instrumental Sonatas w/Piano	35	5.56
Piano Sonatas	37	5.87
Misc. Keyboard (including Duets)	40	6.35
Mechanical Organ	4	0.63

The ORR data set enumerates Mozart's symphonies using the familiar 41 symphonies listed in the Breitkopf and Härtel catalog. As can be seen in Appendix 11, six symphonies (2, 3, 7, 11, 14, and 24) received no performances; one later symphony (#37) is rarely performed, as it is now known to be spurious. The last six (legitimate) symphonies account for 65.66% of all performances.

Armed with this information, we can see that an adequate sample for Mozart alone might minimally involve

15 pieces: four symphonies (three of which should be from the last six), three operas, two piano concerti, and one each from the piano sonatas, violin sonatas, string quartets, violin concerti, serenades, and masses. In Appendix 10, which gives the percentages for the representation of each "obligatory" composer, Mozart (who is top ranked) has 5.5%. To include 15 pieces by Mozart one would then need a corpus of approximately 300 pieces. This would allow for similar generic distinctions to be made among the works of other highly ranked composers, but problems would develop in trying to make similar allowances for lower-ranked composers such as Stravinsky, Shostakovich, and Bartok, who also wrote in a wide range of musical genres. These problems will be inevitable among the lower-ranked composers in the corpus, though many of the lower-ranked composers are known primarily as specialists in a particular genre (e.g., Mahler and Prokofiev for symphonies; Massenet and Rossini for opera). Thus a corpus of 300 pieces would seem to be large enough to balance the parameters of composer, genre, and historical style, allowing for a reasonably complete representation of the harmonic, melodic, and rhythmic idioms of the music of the common-practice period.

### Concluding Discussion

A method for building a representative corpus of classical music of  $\approx 300$  pieces has been developed. It involves a core set of obligatory composers (aka the "usual suspects"), along with historically weighted contributions by other composers from each era. The balance of musical genres can also be taken into account, though doing so requires a corpus of sufficient size; while 300 pieces are suggested here, a larger corpus may be required if one wants a fine-grained representative

TABLE 8. *Distribution of Mozart's Most Popular Musical Genres.*

Genre	Grove Count	Grove Count %	Weight Factor	Weighted Count	Weighted Count %	Amazon Count	Amazon %
Symphonies	49	24.75	1.5	73.5	26.75	4243	21.17
Operas	17	8.59	3	51	18.56	4002	19.97
Piano Sonatas	18	9.09	1	18	6.55	2013	10.04
Violin Sonatas	20	10.10	1	20	7.28	1211	6.04
String Quartets	23	11.62	1	23	8.37	1656	8.26
Piano Concerti	27	13.64	1.5	40.5	14.74	2905	14.50
Violin Concerti	5	2.53	1.5	7.5	2.73	1919	9.58
Serenades	23	11.62	0.75	17.25	6.28	1650	8.23
Masses	16	8.08	1.5	24	8.74	441	2.20

Note: Data include (a) their count in the Grove dictionary, (b) their relative % among these genres, (c) a list of weighting factors that takes the scope of typical works in each genre into account (d) a revised count based on the weighting factor, (e) revised relative %, and (f) counts by composer and genre from Amazon.com, along with their corresponding % in this tally of Amazon titles.

distribution of both genres and composers within each historical style. Alternatively, one could restrict a corpus to a particular genre such as operas or keyboard music, which would allow for a smaller corpus. Future iterations of this model may also add or subtract composers and/or modify weighting parameters for historical styles and genres based on more improved direct or indirect consumption data, in particular broadcast, sales, and/or licensing data as they become available.

It is unsurprising that relatively few ( $\approx 25$ ) composers, all from the late 18th and 19th centuries, dominate most of the sources and the sampling models investigated here. For classical music institutions and the repertoires they preserve and perform are inherently conservative, given the years of study and specialized technologies (i.e., musical instruments, notation systems, etc.) involved. An expert classical musician may spend many months, if not longer, learning a single piece. To play most of the standard repertoire requires considerable baseline ability as well as a good deal of additional practice; it is thus no surprise that Gingold's sets of excerpts for the classical repertoire are still useful—still available at the "Juilliard Store"—even though they are 50 years old.

This conservatism also explains why the Barlow and Morgenstern *Dictionary* still remains a somewhat useful index of the 21st century classical music landscape. There is a strong correlation between the top 106 composers in the *Dictionary* and the corresponding entries in the ADS ( $r_s = .62, p < .001$ , two-tailed;  $N = 106$ ). The top 36 composers in the *Dictionary* (i.e., the usual suspects) also correspond well to the ADS ( $r_s = .53, p < .001$ , two-tailed;  $N = 36$ ). However, one does not have to go very far down in the *Dictionary* to find its historical limitations; the correlation between "the bottom 37 of the top 106" in the *Dictionary* and the same composers in the ADS is no longer significant ( $r_s = .29, p = .09$ , two-tailed;  $N = 37$ ). This is where one finds composers such as Farnaby, Daquin, Paderewski, Francouer, d'Indy, Liadoff, Nevin, Pick-Mangiagalli, and Liadoff in Barlow and Morgenstern's compendium, composers whose works, while more popular in mid-20th century, are rarely programmed today.

There were a few surprises in the model, at least for the author, trained as a musicologist. The usual suspects (Bach, Mozart, Beethoven), while dominant, were individually not that dominant, with no single composer capturing more than 6% of the entire corpus. Instead, the bulk of the corpus involves compositions from a group of approximately 30 composers, almost all from the 19th century, who each have a smaller but reliable presence (about 1-2%) representing various genres and styles (e.g., Italian operas, German songs, solo piano

pieces, symphonies, etc.). The poor showing of 20th century composers was also somewhat surprising, and not just the small representation of atonal composers (i.e., Schoenberg, Berg, and Webern), which was not. Rather, musically significant composers such as Vaughan Williams, Britten, or Copland are swamped by their 18th and 19th century counterparts. Living composers have also made almost no inroads into the standard orchestral repertoire, in contrast to previous musical eras, where most concerts involved "new" music. Thus there is a great deal of truth in the perspective that regards most current classical music institutions as "sonic museums" (Goehr, 2007).

This study has also shown that it is not difficult to produce a corpus or sampling model that will have a significant correlation with current classical musical landscape, as all one needs to do is to give some preference to the "usual suspects" and then include "lots of other composers, mostly 19th century German composers of orchestral music." But to do a proper model, one has to get the other, mid-level values correct, and most important of all, avoid aesthetic, nationalist, historical or institutional biases—or at least be cognizant of the biases inherent in a given source.

Our knowledge of music and musical life from times past is based on the artifacts that have survived—manuscripts, printed sheet music, recordings, as well as concert programs, newspaper and journal reviews, books, and letters (and these days, their electronic equivalents). We make inferences as to what was popular, influential, and widespread based on the presumption that those things that were written down and copied (and/or produced in such abundance that a good number of copies survive) have been preserved precisely because they were popular, influential, and widespread—but that presumption could be wrong. One may be making the same errors with respect to the construction of this model. While we may know which composers' pieces are recorded or purchased the most, we do not know which are listened to the most at home, which concerts were best attended, which radio broadcasts were listened to, and so forth. Knowing that music is in the hands of consumers is not the same thing as knowing the music has been consumed, though one can make some reasonable inferences.

We also make do with what is convenient, using information that is accessible, formatted in a way that facilitates our investigation, and also more or less fits our pre-existing presumptions. For this reason, perhaps, the Barlow and Morgenstern *Dictionary* (as well as others, such as the Essen folksong collection) have been the go-to corpuses for empirical studies of music style and syntax (e.g., Collister & Huron, 2008; Huron & Ollen, 2003;

London & Jones, 2011; Patel & Daniele, 2003; Temperley, 2010). While these sources have taught us much, they also are subject the biases and constraints discussed above, and therefore one must be cautious in making generalizations from studies that use them to make broader claims about classical or tonal music in general.

### Author Note

Research on this project was supported in part by a Mellon Foundation Digital Humanities Seed Grant

administered through the Humanities Center of Carleton College. Humdrum Toolkit analysis of the Barlow and Morgenstern data was graciously provided by David Huron and Daniel Shanahan of Ohio State University. Access to the consolidated “Orchestral Repertoire Report” was kindly provided by Brian Bader and Jan Wilson of the League of American Orchestras.

*Correspondence concerning this article should be addressed to Justin London, Department of Music, Carleton College, One North College Street, Northfield, MN 55057. E-mail: jlondon@carleton.edu*

### References

- AUSLANDER, P. (2008). *Liveness: Performance in a mediatized culture* (2nd ed.). New York: Routledge.
- BARLOW, H., & MORGESTERN, S. (1948). *A dictionary of musical themes*. New York: Crown Publishers.
- BIGAND, E., & POULIN-CHARRONNAT, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition*, 100, 100-130.
- BRACKETT, N., & HOARD, C. (Eds.) (2008). *The new Rolling Stone album guide*. New York: Simon and Schuster.
- BURKHOLDER, J. P., GROUT, D. J., & PALISCA, C. V. (2009). *A history of Western music* (8th ed.). New York: W.W. Norton and Co.
- COLLISTER, L., & HURON, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, 3, 109-125.
- COOK, R., & MORTON, B. (2008). *The Penguin guide to jazz recordings* (9th ed.). New York: Penguin Books.
- CREEL, S. C., NEWPORT, E. L., & ASLIN, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1119-1130.
- EISEN, C., RIEGER, E., SADIE, S., ANGEMÜLLER, R., OLDMAN, C. B., & STAFFORD, W. (2013). Mozart: (3) Wolfgang Amadeus, § ‘Works.’ *Grove music online*, Oxford University Press. Accessed January 11, 2013, [www.oxfordmusiconline.com/subscriber/article/grove/music/40258pg3](http://www.oxfordmusiconline.com/subscriber/article/grove/music/40258pg3)
- GINGOLD, J. (Ed.) (1962). *Orchestral excerpts from the symphonic repertoire: Violin*. New York: International Music Company.
- GOEHR, L. (2007). *The imaginary museum of musical works*. New York: Oxford University Press.
- GOULDING, P. G. (1995). *Classical music: The 50 greatest composers and their 1,000 greatest works*. New York: Ballantine Books.
- HAY, J. F., & SAFFRAN, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, 17, 610-641.
- HELMHOLTZ, H. L. F. von (1954). *On the sensations of tone* (A. J. Ellis, Trans.). New York: Dover. (Original work published 1877)
- HURON, D. (2013). *Humdrum* [website]. <http://www.musiccog.ohio-state.edu/Humdrum/index.html>. Accessed January 14, 2013.
- HURON, D., & OLLEN, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception*, 21, 267-271.
- HURON, D., & OMMEN, A. (2006). An empirical study of syncopation in American music, 1890-1939. *Music Theory Spectrum*, 26, 211-231.
- JONAITIS, E. M., & SAFFRAN, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science*, 33, 951-968.
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- LERDAHL, F. (2001). *Tonal pitch space*. New York: Oxford University Press.
- LONDON, J., & JONES, K. (2011). Rhythmic refinements to the nPVI measure: A reanalysis of Patel & Daniele (2003a). *Music Perception*, 29, 113-118.
- LOUI, P., WESSEL, D. L., & KAM, C. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception*, 27, 377-388.
- MATTHEWS, B. S. (2009). *Programming trends in professional orchestra concerts in the United States 2007-2009* (Unpublished doctoral dissertation). Arizona State University, Phoenix, Arizona.
- PATEL, A. D., & DANIELE, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, 87, B35-B45.
- SCHENKER, H. (1935). *Der freie Satz* [Free composition]. Vienna: Universal Edition.
- TARUSKIN, R. (2005). *The Oxford history of Western music*. New York: Oxford University Press.
- TEMPERLEY, D. (2010). Modeling common-practice rhythm. *Music Perception*, 27, 355-376.

**Appendix 1: Top 106 composers in Barlow and Morgenstern *Dictionary* (1948) ranked by number of pieces; percentages indicate their presence relative to all others in the *Dictionary*.**

Composer	Total	%	Composer	Total	%
Bach, J.S.	175	7.48	Bach, J.C.	9	0.38
Chopin	110	4.70	Berlioz	9	0.38
Beethoven	96	4.10	Bloch	9	0.38
Mozart	90	3.85	Hindemith	9	0.38
Brahms	86	3.68	Lully	9	0.38
Schubert	51	2.18	Rossini	9	0.38
Liszt	50	2.14	Waldteufel	9	0.38
Handel	49	2.09	Buxtehude	8	0.34
Debussy	47	2.01	Casella	8	0.34
Scarlatti	47	2.01	Glinka	8	0.34
Grieg	42	1.80	Grainger	8	0.34
Couperin	40	1.71	Ireland	8	0.34
Schumann	40	1.71	Moszkowski	8	0.34
Haydn	39	1.67	Paganini	8	0.34
Dvorak	34	1.45	von Suppè	8	0.34
Tchaikovsky	34	1.45	Wieniawski	8	0.34
Saint-Saëns	28	1.20	Borodin	7	0.30
Sibelius	25	1.07	Bruckner	7	0.30
de Falla	23	0.98	Corelli	7	0.30
Mendelssohn	23	0.98	Dowland	7	0.30
Delibes	22	0.94	Farnaby	7	0.30
Fauré	20	0.86	Frescobaldi	7	0.30
Prokofiev	20	0.86	Gretry	7	0.30
Strauss	20	0.86	Massenet	7	0.30
Elgar	18	0.77	Offenbach	7	0.30
Vivaldi	18	0.77	Reger	7	0.30
Ravel	17	0.73	Respighi	7	0.30
Copland	16	0.68	Telemann	7	0.30
Franck	16	0.68	Verdi	7	0.30
Rachmaninoff	16	0.68	Bizet	6	0.26
Stravinsky	16	0.68	Chaminade	6	0.26
Albeniz	15	0.64	Daquin	6	0.26
Purcell	15	0.64	Gershwin	6	0.26
Strauss, Richard	15	0.64	Granados	6	0.26
Wagner	15	0.64	Ibert	6	0.26
Weber	15	0.64	Lalo	6	0.26
Gluck	14	0.60	Paderewski	6	0.26
Kreisler	14	0.60	Turina	6	0.26
Rameau	14	0.60	Vaughan Williams	6	0.26
Byrd	13	0.56	Walton	6	0.26
Boccherini	12	0.51	Francoeur	5	0.21
Delius	12	0.51	Honegger	5	0.21
Macdowell	12	0.51	Indy	5	0.21
Shostakovich	12	0.51	Liadoff	5	0.21
Bartok	11	0.47	Mahler	5	0.21
Chabrier	11	0.47	Milhaud	5	0.21
Glazunov	11	0.47	Musorgsky	5	0.21
Rimsky-Korsakov	11	0.47	Nevin	5	0.21
Rubinstein	11	0.47	Pachelbel	5	0.21
Scriabin	11	0.47	Pick-Mangiagalli	5	0.21
Bach, C.P.E.	10	0.43	Roussel	5	0.21
Sousa	10	0.43	Smetana	5	0.21
Tartini	10	0.43	Szymanowski	5	0.21

**Appendix 2: Top 91 composers in Taruskin (2005) ranked by number of musical examples; percentages indicate their presence relative to all examples in the multi-volume history.**

Composer	Era	Exs	%	Composer	Era	Exs	%
Beethoven	ERom	43	4.25	Ravel	20thC	6	0.59
Stravinsky	20thC	39	3.86	Satie	20thC	6	0.59
Schoenberg	20thC	35	3.46	Shostakovich	20thC	6	0.59
Bach, J.S.	Bar	29	2.87	Smetana	LRom	6	0.59
Bartok	20thC	23	2.27	Byrd	Ren	5	0.49
Schubert	ERom	23	2.27	Reich	20thC	5	0.49
Brahms	LRom	20	1.98	Rimsky-Korsakov	LRom	5	0.49
Wagner	LRom	20	1.98	Scarlatti	Bar	5	0.49
Monteverdi	Bar	19	1.88	Bach, C.P.E.	Bar	4	0.40
Berg	20thC	17	1.68	Boulez	20thC	4	0.40
Haydn	Class	17	1.68	Chabrier	LRom	4	0.40
Machaut	Med	16	1.58	de la Halle	Med	4	0.40
Verdi	LRom	16	1.58	Donizetti	ERom	4	0.40
Copland	20thC	15	1.48	Franck	LRom	4	0.40
Ives	20thC	15	1.48	Frescobaldi	Bar	4	0.40
Mozart	Class	15	1.48	Gabrieli	Ren	4	0.40
Palestrina	Ren	14	1.38	Gershwin	20thC	4	0.40
Babbitt	20thC	13	1.29	Gottschalk	LRom	4	0.40
Carter	20thC	13	1.29	Lassus	Ren	4	0.40
Scriabin	20thC	13	1.29	Milhaud	20thC	4	0.40
Tchaikovsky	LRom	13	1.29	Pärt	20thC	4	0.40
Britten	20thC	12	1.19	Prokofiev	20thC	4	0.40
Liszt	LRom	12	1.19	Rossini	ERom	4	0.40
Strauss, Richard	LRom	12	1.19	Saint-Saëns	LRom	4	0.40
Webern	20thC	12	1.19	Scarlatti, D.	Bar	4	0.40
des Prez	Ren	11	1.09	Schutz	Bar	4	0.40
Chopin	ERom	10	0.99	Thomson	20thC	4	0.40
Debussy	20thC	10	0.99	Vivaldi	Bar	4	0.40
Handel	Bar	9	0.89	Weber	ERom	4	0.40
Lully	Bar	9	0.89	Bruckner	LRom	3	0.30
Borodin	LRom	8	0.79	Buus	Ren	3	0.30
Mahler	LRom	8	0.79	Cage	20thC	3	0.30
Berlioz	ERom	7	0.69	Del Tredici	20thC	3	0.30
Busnoys	Ren	7	0.69	Dowland	Ren	3	0.30
Corelli	Bar	7	0.69	Gallus	Ren	3	0.30
Dufay	Med	7	0.69	Gluck	Class	3	0.30
Dvorak	LRom	7	0.69	Harris	20thC	3	0.30
Janáček	20thC	7	0.69	Lerdahl	20thC	3	0.30
Mendelssohn	ERom	7	0.69	Meyerbeer	ERom	3	0.30
Messiaen	20thC	7	0.69	Mussorgsky	LRom	3	0.30
Purcell	Bar	7	0.69	Obrecht	Ren	3	0.30
Schumann	ERom	7	0.69	Parker	20thC	3	0.30
Glinka	ERom	6	0.59	Poulenc	20thC	3	0.30
Hindemith	20thC	6	0.59	Weill	20thC	3	0.30
Ockeghem	Ren	6	0.59	Guido d'Arezzo	Med	3	0.30

**Appendix 3: Musical example counts for all composers with two or more musical examples in Burkholder, Grout, and Palisca (2009); percentages are relative to all examples.**

Composer	Era	Exs.	%
Beethoven	E Rom	6	5.45
Monteverdi	Ren	6	5.45
Dufay	Ren	4	3.64
Haydn	Class	4	3.64
Machaut	Med	4	3.64
Mozart	Class	4	3.64
Bach, J.S.	Bar	3	2.73
Chopin	E Rom	3	2.73
Des Prez	Ren	3	2.73
Frescobaldi	Bar	3	2.73
Lully	Class	3	2.73
Mendelssohn	E Rom	3	2.73
Schoenberg	20th C	3	2.73
Schubert	E Rom	3	2.73
Schumann	E Rom	3	2.73
Stravinsky	20th C	3	2.73
Adams, John	20th C	2	1.82
Bach, C.P.E.	Bar	2	1.82
Bartok	20th C	2	1.82
Berg	20th C	2	1.82
Brahms	L Rom	2	1.82

Composer	Era	Exs.	%
Buxtehude	Bar	2	1.82
Byrd	Ren	2	1.82
Cage	20th C	2	1.82
de la Halle	Med	2	1.82
Debussy	20th C	2	1.82
Gabrieli, G.	Ren	2	1.82
Handel	Bar	2	1.82
Ives	20th C	2	1.82
Lassus	Ren	2	1.82
Liszt	L Rom	2	1.82
Luther	Ren	2	1.82
Mahler	L Rom	2	1.82
Milhaud	20th C	2	1.82
Mussorgsky	L Rom	2	1.82
Narvaez	Ren	2	1.82
Ockeghem	Ren	2	1.82
Palestrina	Ren	2	1.82
Purcell	Bar	2	1.82
Scarlatti, A.	Bar	2	1.82
Schutz	Bar	2	1.82
Strauss, Richard	L Rom	2	1.82

**Appendix 4: Excerpt counts for all composers in Gingold (1962), along with their relative percentages.**

Composer	Pieces	%
Beethoven	19	9.69
Brahms	12	6.12
Tchaikovsky	12	6.12
Mozart	9	4.59
Berlioz	8	4.08
Mendelssohn	8	4.08
Rossini	7	3.57
Haydn	6	3.06
Ravel	6	3.06
Schubert	6	3.06
Bach	5	2.55
Dvorak	5	2.55
Rimsky-Korsakov	5	2.55
Sibelius	5	2.55
Bizet	4	2.04
Debussy	4	2.04
Prokofiev	4	2.04
Strauss Jr., Johann	4	2.04
Stravinsky	4	2.04
Weber	4	2.04
Borodin	3	1.53
Bruckner	3	1.53
Saint-Saëns	3	1.53
Smetana	3	1.53
Chabrier	2	1.02
d'Indy	2	1.02
Falla	2	1.02
Franck	2	1.02
Glinka	2	1.02
Grieg	2	1.02

Composer	Pieces	%
Jaernefelt	2	1.02
Liszt	2	1.02
Lyadov	2	1.02
Rachmaninoff	2	1.02
Schumann	2	1.02
Charpentier	1	0.51
Chausson	1	0.51
De Falla	1	0.51
Delibes	1	0.51
Dukas	1	0.51
Elgar	1	0.51
Enescu	1	0.51
Faure	1	0.51
Gluck-Mottl	1	0.51
Goldmark	1	0.51
Humperdinck	1	0.51
Ibert	1	0.51
Ippolitov-Ivanov	1	0.51
Kabalevsky	1	0.51
Lalo	1	0.51
Mahler	1	0.51
Mussorgsky	1	0.51
Nicolai	1	0.51
Offenbach	1	0.51
Ponchielli	1	0.51
Schoenberg	1	0.51
Shostakovich	1	0.51
Thomas	1	0.51
Vaughan Williams	1	0.51
Verdi	1	0.51

**Appendix 5: Title counts for all composers in the Naxos "Classical Music in the Movies" collection.**

Composer	Titles	%	Composer	Titles	%
Mozart	82	10.58	Pärt	3	0.39
Bach, J.S.	68	8.77	Purcell	3	0.39
Beethoven	61	7.87	Saint-Saëns	3	0.39
Strauss Jr., Johann	41	5.29	Britten	2	0.26
Chopin	35	4.52	Donizetti	2	0.26
Puccini	35	4.52	Flotow	2	0.26
Tchaikovsky	25	3.23	Gade	2	0.26
Verdi	23	2.97	Grieg	2	0.26
Wagner	21	2.71	Ivanovici	2	0.26
Rossini	20	2.58	Janáček	2	0.26
Schubert	20	2.58	Lehár	2	0.26
Mendelssohn	18	2.32	Ponchielli	2	0.26
Dvorak	16	2.06	Sarasate	2	0.26
Debussy	13	1.68	Sibelius	2	0.26
Elgar	13	1.68	Smetana	2	0.26
Handel	13	1.68	Stravinsky	2	0.26
Schumann	12	1.55	Albéniz	1	0.13
Brahms	11	1.42	Alfvén	1	0.13
Rachmaninoff	11	1.42	Badarzewska	1	0.13
Haydn	9	1.16	Bartók	1	0.13
Prokofiev	9	1.16	Beriot	1	0.13
Vivaldi	9	1.16	Berlioz	1	0.13
Shostakovich	8	1.03	Bernstein	1	0.13
Strauss, Richard	8	1.03	Borodin	1	0.13
Bizet	7	0.90	Bruch	1	0.13
Mahler	7	0.90	Canteloube	1	0.13
Mussorgsky	7	0.90	Charpentier	1	0.13
Ravel	7	0.90	Copland	1	0.13
Delibes	6	0.77	di Capua	1	0.13
Boccherini	5	0.65	Dukas	1	0.13
Faure	5	0.65	Franck	1	0.13
Liszt	5	0.65	Gauntlet	1	0.13
Offenbach	5	0.65	Giordano	1	0.13
Rimsky-Korsakov	5	0.65	Gluck	1	0.13
Satie, Erik	5	0.65	Gould	1	0.13
Strauss Sr., Johann	5	0.65	Gounod	1	0.13
Albinoni	4	0.52	Granados	1	0.13
Barber	4	0.52	Korngold	1	0.13
Holst	4	0.52	Kresler	1	0.13
Ketelbey	4	0.52	Last, J.	1	0.13
Leoncavallo	4	0.52	Marais, M.	1	0.13
Orff	4	0.52	Mouret	1	0.13
Pachelbel	4	0.52	Parry	1	0.13
Allegri	3	0.39	Pergolesi	1	0.13
Catalani	3	0.39	Rodrigo	1	0.13
Gershwin	3	0.39	Scriabin	1	0.13
Górecki	3	0.39	Sousa	1	0.13
Grainger	3	0.39	von Suppé	1	0.13
Khachaturian	3	0.39	Szymanowski	1	0.13
Marcello	3	0.39	Waldteufel	1	0.13
Mascagni	3	0.39	Webern	1	0.13
Paganini	3	0.39			

**Appendix 6: Top 100 composers in terms of works performed, from the League of American Orchestras consolidated *Orchestra Repertoire Reports* (ORR) 2000-2009.**

Composer	Perfs	%	Composer	Perfs	%
Mozart	7103	7.12	Schoenberg	361	0.36
Beethoven	6906	6.92	Holst	351	0.35
Brahms	3930	3.94	Bizet	337	0.34
Tchaikovsky	3891	3.90	Faure	323	0.32
Dvorak	2628	2.63	Walton	309	0.31
Ravel	2468	2.47	Kodaly	277	0.28
Strauss, Richard	2381	2.39	Franck	246	0.25
Stravinsky	2115	2.12	Rodrigo	246	0.25
Shostakovich	2095	2.10	Higdon	244	0.24
Haydn	2070	2.07	Janáček	244	0.24
Rachmaninoff	2048	2.05	Orff	242	0.24
Mendelssohn	2044	2.05	Rouse	242	0.24
Prokofiev	1993	2.00	Messiaen	235	0.24
Bach, J.S.	1934	1.94	Puccini	235	0.24
Sibelius	1882	1.89	Poulenc	232	0.23
Mahler	1789	1.79	Smith, J.S.	210	0.21
Copland	1453	1.46	Borodin	208	0.21
Berlioz	1434	1.44	Williams, J.	206	0.21
Wagner	1433	1.44	Berg	205	0.21
Schumann	1342	1.34	Daugherty	204	0.20
Handel	1269	1.27	Ginastera	201	0.20
Debussy	1223	1.23	Dukas	200	0.20
Schubert	1142	1.14	Korngold	198	0.20
Barber	1090	1.09	Glinka	196	0.20
Bernstein	1075	1.08	Kernis	192	0.19
Bartok	1072	1.07	Golijov	183	0.18
Elgar	1002	1.00	Lutoslawski	176	0.18
Saint-Saëns	977	0.98	Tower	175	0.18
Vivaldi	748	0.75	Webern	175	0.18
Verdi	747	0.75	Martinu	170	0.17
Rimsky-Korsakov	742	0.74	Khachaturian	167	0.17
Gershwin	719	0.72	Pärt	164	0.16
Britten	708	0.71	Piazzolla	163	0.16
Mussorgsky	689	0.69	Dutilleux	155	0.16
Respighi	659	0.66	Glazunov	152	0.15
Liszt	657	0.66	Milhaud	146	0.15
Grieg	618	0.62	Villa-Lobos	144	0.14
Bruckner	613	0.61	Takemitsu	143	0.14
Vaughn Williams	610	0.61	Ligeti	142	0.14
Rossini	599	0.60	Revueltas	141	0.14
Adams, J.	556	0.56	Harbison	134	0.13
Falla	500	0.50	Lalo	133	0.13
Bruch	476	0.48	Ibert	130	0.13
Ives	440	0.44	Purcell	125	0.13
Strauss Jr., Johann	410	0.41	Liadov	121	0.12
Chopin	400	0.40	Schwantner	120	0.12
Smetana	391	0.39	Sierra	120	0.12
Nielsen	390	0.39	MacMillan	118	0.12
Hindemith	384	0.38	Rameau	117	0.12
Corigliano	378	0.38			
Weber	376	0.38			

**Appendix 7: List of the “Top 55” composers culled from appendices 1, 2, 5, and 6, ordered in terms of their “Median Percentage Ranking” (MPR) calculated from their percentage ranking in those appendices.**

Composer	ORR	Tskn	Naxos	B&M	MPR
Beethoven	6.92	4.25	7.87	4.10	5.59
Mozart	7.12	1.48	10.58	3.85	5.48
Bach, J.S.	1.94	2.87	8.77	7.48	5.18
Brahms	3.94	1.98	1.42	3.68	2.83
Chopin	0.40	0.99	4.52	4.70	2.75
Tchaikovsky	3.90	1.29	3.23	1.45	2.34
Schubert	1.14	2.27	2.58	2.18	2.23
Dvorak	2.63	0.69	2.06	1.45	1.76
Wagner	1.44	1.98	2.71	0.64	1.71
Haydn	2.07	1.68	1.16	1.67	1.67
Mendelssohn	2.05	0.69	2.32	0.98	1.52
Handel	1.27	0.89	1.68	2.09	1.47
Debussy	1.23	0.99	1.68	2.01	1.45
Schumann	1.34	0.69	1.55	1.71	1.45
Stravinsky	2.12	3.86	0.26	0.68	1.40
Verdi	0.75	1.58	2.97	0.30	1.17
Strauss, Richard	2.39	1.19	1.03	0.64	1.11
Copland	1.46	1.48	0.13	0.68	1.07
Rachmaninoff	2.05	0.10	1.42	0.68	1.05
Prokofiev	2.00	0.40	1.16	0.86	1.01
Liszt	0.66	1.19	0.65	2.14	0.92
Elgar	1.00	0.00	1.68	0.77	0.89
Mahler	1.79	0.79	0.90	0.21	0.85
Ravel	2.47	0.59	0.90	0.73	0.82
Shostakovich	2.10	0.59	1.03	0.51	0.81
Bartok	1.07	2.27	0.13	0.47	0.77
Vivaldi	0.75	0.40	1.16	0.77	0.76
Saint-Saëns	0.98	0.40	0.39	1.20	0.69

Composer	ORR	Tskn	Naxos	B&M	MPR
Sibelius	1.89	0.00	0.26	1.07	0.66
Strauss Jr., Joh.	0.41	0.20	5.29	0.86	0.63
Berlioz	1.44	0.69	0.13	0.38	0.54
Rossini	0.60	0.40	2.58	0.38	0.50
Mussorgsky	0.69	0.30	0.90	0.21	0.49
Faure	0.32	0.20	0.65	0.86	0.48
Britten	0.71	1.19	0.26	0.17	0.48
Grieg	0.62	0.00	0.26	1.80	0.44
Delibes	0.02	0.10	0.77	0.94	0.44
Barber	1.09	0.00	0.52	0.17	0.34
Scriabin	0.08	1.29	0.13	0.47	0.30
Falla	0.50	0.00	0.00	0.98	0.25
Ives	0.44	1.48	0.00	0.04	0.24
Schoenberg	0.36	3.46	0.00	0.09	0.22
Bizet	0.34	0.10	0.90	0.09	0.22
Scarlatti	0.00	0.40	0.00	2.01	0.20
Bernstein	1.08	0.00	0.13	0.26	0.19
Puccini	0.24	0.10	4.52	0.00	0.17
Webern	0.18	1.19	0.13	0.00	0.15
Berg	0.21	1.68	0.00	0.00	0.10
Carter	0.09	1.29	0.00	0.00	0.05
Monteverdi	0.01	1.88	0.00	0.00	0.01
Couperin	0.00	0.00	0.00	1.71	0.00
Babbitt	0.00	1.29	0.00	0.00	0.00
des Prez	0.00	1.09	0.00	0.00	0.00
Machaut	0.00	1.58	0.00	0.00	0.00
Palestrina	0.00	1.38	0.00	0.00	0.00

**Appendix 8: The Amazon Data Set (ADS): Title counts taken from Amazon.com (US) during July-August 2012 for all composers listed in Appendices 1-6.**

Composer	Titles	%
Mozart	12,457	5.51
Bach, J.S.	11,063	4.89
Beethoven	10,218	4.52
Schubert	6,939	3.07
Brahms	6,615	2.93
Verdi	5,868	2.60
Tchaikovsky	5,205	2.30
Schumann	4,780	2.11
Mendelsson	4,502	1.99
Wagner	4,102	1.81
Debussy	4,082	1.81
Liszt	4,016	1.78
Handel	3,999	1.77
Vivaldi	3,914	1.73
Puccini	3,730	1.65
Strauss, Richard	3,533	1.56
Dvorák	3,526	1.56

Composer	Titles	%
Haydn	3,421	1.51
Ravel	3,276	1.45
Chopin	3,074	1.36
Bizet	2,893	1.28
Rossini	2,814	1.24
Rachmaninoff	2,657	1.18
Grieg	2,536	1.12
Saint-Saëns	2,509	1.11
Donizetti	2,384	1.05
Bernstein	2,344	1.04
Gounod	2,324	1.03
Mahler	2,252	1.00
Massenet	2,053	0.91
Fauré	2,040	0.90
Prokofiev	1,867	0.83
Gershwin	1,812	0.80
Stravinsky	1,793	0.79

(continued)

## Appendix 8 (continued)

Composer	Titles	%	Composer	Titles	%
Elgar	1,791	0.79	Corelli	652	0.29
Shostakovich	1,598	0.71	Bruch	650	0.29
Berlioz	1,581	0.70	Messiaen	575	0.25
Sibelius	1,528	0.68	Berg	570	0.25
Mussorgsky	1,421	0.63	Respighi	567	0.25
Britten	1,419	0.63	Glazunov	562	0.25
Franck	1,406	0.62	Couperin	551	0.24
Purcell	1,394	0.62	Martinu	551	0.24
Telemann	1,355	0.60	Dowland	544	0.24
Mascagni	1,329	0.59	Thomas, A.	544	0.24
Bartók	1,328	0.59	Milhaud	542	0.24
Bruckner	1,278	0.57	Boulez	541	0.24
Weber	1,266	0.56	Ives	535	0.24
Offenbach	1,261	0.56	Rubinstein	534	0.24
Vaughan Williams	1,225	0.54	Sousa	516	0.23
Leoncavallo	1,192	0.53	Scarlatti, A.	509	0.23
Pachelbel	1,057	0.47	Nielsen	504	0.22
de Falla	1,055	0.47	Rameau	485	0.21
Piazzolla	1,055	0.47	Flotow	461	0.20
Gluck	1,034	0.46	Grainger	460	0.20
Poulenc	1,002	0.44	Buxtehude	454	0.20
Rimsky-Korsakov	974	0.43	Cage	444	0.20
Scarlatti, D.	959	0.42	Weill	427	0.19
Copland	955	0.42	Orff	422	0.19
Barber	948	0.42	Kodaly	409	0.18
Smetana	940	0.42	Sarasate	409	0.18
Borodin	932	0.41	Walton	404	0.18
Satie	929	0.41	Palestrina	399	0.18
Meyerbeer	917	0.41	Gould	391	0.17
Villa-Lobos	899	0.40	Turina	384	0.17
Strauss Jr., Johann	887	0.39	Chabrier	382	0.17
Boccherini	882	0.39	Delius	381	0.17
Albéniz	861	0.38	Lalo	369	0.16
Monteverdi	858	0.38	von Suppé	367	0.16
Giordano	852	0.38	Frescobaldi	363	0.16
Hindemith	841	0.37	Bach, C.P.E.	360	0.16
Albinoni	835	0.37	Ibert	360	0.16
Lehár	828	0.37	Charpentier	357	0.16
Holst	816	0.36	Korngold	349	0.15
Ponchielli	802	0.35	Bloch	343	0.15
Scriabin	785	0.35	Des Prez	332	0.15
Granados	784	0.35	Szymanowski	328	0.15
Delibes	774	0.34	Gabrieli	322	0.14
Adams	769	0.34	Dukas	321	0.14
Kreisler	744	0.33	Honegger	321	0.14
Paganini	736	0.33	Catalani	317	0.14
Reger	725	0.32	Enescu	314	0.14
Rodrigo	725	0.32	Schutz	310	0.14
Schoenberg	696	0.31	Wieniawski	309	0.14
Bach, J.C.	690	0.31	Ginastera	297	0.13
Byrd	683	0.30	Lassus	290	0.13
Glinka	676	0.30	Tartini	290	0.13
Khachaturian	676	0.30	Nicolai	282	0.12
Janáček	671	0.30	Pärt	272	0.12

(continued)

## Appendix 8 (continued)

Composer	Titles	%	Composer	Titles	%
Lutoslawski	244	0.11	Ippolitov-Ivanov	85	0.04
Marcello	242	0.11	Casella	84	0.04
Marais, M.	240	0.11	Pergolesi	84	0.04
Roussel	240	0.11	Babbitt	78	0.03
Webern	235	0.10	Harris	75	0.03
Lully	226	0.10	Obrecht	75	0.03
Parry	214	0.09	Revueltas	71	0.03
Ligeti	209	0.09	Gade, Niels	70	0.03
MacDowell	208	0.09	Ketelbey	69	0.03
Moszkowski	205	0.09	Daugherty	64	0.03
Carter	204	0.09	Farnaby	64	0.03
Takemitsu	168	0.07	Narvaez	57	0.03
Waldteufel	165	0.07	Gallus	56	0.02
Paderewski	164	0.07	Tower	56	0.02
Kabalevsky	160	0.07	Del Tredici	53	0.02
Daquin	157	0.07	Gretry	52	0.02
Canteloube	152	0.07	Busnoys	50	0.02
Dutilleux	150	0.07	Sierra	50	0.02
Mouret	149	0.07	Halle	45	0.02
Alfvén	148	0.07	Higdon	42	0.02
Allegri	147	0.07	Schwantner	42	0.02
Corigliano	146	0.06	MacMillan	36	0.02
Reich	146	0.06	Rouse	31	0.01
Lyadov	144	0.06	Kernis	30	0.01
Luther	142	0.06	Golijov	28	0.01
Machaut	138	0.06	Badarzewska	23	0.01
Williams, J.	136	0.06	Strauss Sr., Johann	21	0.01
Gauntlett	134	0.06	Beriot,	15	0.01
Górecki	134	0.06	Lerdahl	12	0.01
Humperdinck	130	0.06	Parker	12	0.01
Goldmark	127	0.06	Pick-Mangiagalli	5	0.00
Thomson	122	0.05	Buus	4	0.00
Chaminade	116	0.05	Dufay	4	0.00
Gottschalk	111	0.05	di Capua	3	0.00
Harbison	111	0.05	Guido	2	0.00
Ockeghem	109	0.05	Francoeur	1	0.00
Smith	108	0.05	Liadov	1	0.00
d'Indy	93	0.04	Ivanovici	0	0.00
Ireland	92	0.04	Last	0	0.00
Nevin	87	0.04			

Appendix 9: Top 50 composers in the ADS (Appendix 8) versus the MPR (Appendix 7); composers in the ADS but not in the MPR are underlined.

Composer	Titles	Amz	MPR	Era
Mozart	12,457	5.51	5.48	Class
Bach, J.S.	11,063	4.89	5.18	Bar
Beethoven	10,218	4.52	5.59	Class
Schubert	6,939	3.07	2.23	E Rom
Brahms	6,615	2.93	2.83	L Rom
Verdi	5,868	2.60	1.17	L Rom
Tchaikovsky	5,205	2.30	2.34	L Rom

Composer	Titles	Amz	MPR	Era
Schumann	4,780	2.11	1.45	E Rom
Mendelsson	4,502	1.99	1.52	E Rom
Wagner	4,102	1.81	1.71	L Rom
Debussy	4,082	1.81	1.45	20thC
Liszt	4,016	1.78	0.92	L Rom
Handel	3,999	1.77	1.47	Bar
Vivaldi	3,914	1.73	0.76	Bar

(continued)

## Appendix 9 (continued)

Composer	Titles	Amz	MPR	Era
Puccini	3,730	1.65	0.17	L Rom
Strauss, Richard	3,533	1.56	1.11	L Rom
Dvorák	3,526	1.56	1.76	L Rom
Haydn	3,421	1.51	1.67	Class
Ravel	3,276	1.45	0.82	20thC
Chopin	3,074	1.36	2.75	E Rom
Bizet	2,893	1.28	0.22	L Rom
Rossini	2,814	1.24	0.50	E Rom
Rachmaninoff	2,657	1.18	1.05	L Rom
Grieg	2,536	1.12	0.44	L Rom
Saint-Saëns	2,509	1.11	0.69	L Rom
Donizetti	2,384	1.05	<u>0.00</u>	E Rom
Bernstein	2,344	1.04	0.19	20thC
Gounod	2,324	1.03	<u>0.00</u>	L Rom
Mahler	2,252	1.00	0.85	L Rom
Massenet	2,053	0.91	<u>0.00</u>	L Rom
Fauré	2,040	0.90	0.48	L Rom
Prokofiev	1,867	0.83	1.01	20thC

Composer	Titles	Amz	MPR	Era
Gershwin	1,812	0.80	<u>0.00</u>	20thC
Stravinsky	1,793	0.79	1.40	20thC
Elgar	1,791	0.79	0.89	20thC
Shostakovich	1,598	0.71	0.81	20thC
Berlioz	1,581	0.70	0.54	E Rom
Sibelius	1,528	0.68	0.66	L Rom
Mussorgsky	1,421	0.63	0.49	L Rom
Britten	1,419	0.63	0.48	20thC
Franck	1,406	0.62	<u>0.00</u>	L Rom
Purcell	1,394	0.62	<u>0.00</u>	Bar
Telemann	1,355	0.60	<u>0.00</u>	Bar
Mascagni	1,329	0.59	<u>0.00</u>	L Rom
Bartók	1,328	0.59	0.77	20thC
Bruckner	1,278	0.57	<u>0.00</u>	L Rom
Weber	1,266	0.56	<u>0.00</u>	E Rom
Offenbach	1,261	0.56	<u>0.00</u>	L Rom
Vaughan Williams	1,225	0.54	<u>0.00</u>	20thC
Leoncavallo	1,192	0.53	<u>0.00</u>	L Rom

Appendix 10: List of 35 "Obligatory" composers for a representative corpus, along with a relative weighting for each. Other composers who appeared in all four sources used in the MPR (the "a" composers) or in three of the four (the "b" composers) are also listed.

Composer	% (to .5)	Era
Mozart (a)	5.50	Class
Bach, J.S. (a)	5.00	Bar
Beethoven (a)	4.50	Class
Schubert (a)	3.00	E Rom
Brahms (a)	3.00	L Rom
Tchaikovsky (a)	2.50	L Rom
Verdi (a)	2.50	L Rom
Handel (a)	2.00	Bar
Mendelssohn (a)	2.00	E Rom
Schumann (a)	2.00	E Rom
Liszt (a)	2.00	Late Rom
Wagner (a)	2.00	Late Rom
Debussy (a)	2.00	20th C.
Vivaldi (a)	1.50	Bar
Haydn (a)	1.50	Class
Chopin (a)	1.50	E Rom
Bizet (b)	1.50	L Rom
Dvorák (a)	1.50	Late Rom
Puccini	1.50	Late Rom
Strauss, Richard (a)	1.50	Late Rom
Ravel (a)	1.50	20th C.
Donizetti	1.00	E Rom
Rossini (a)	1.00	E Rom
Fauré	1.00	Late Rom
Gounod	1.00	Late Rom
Grieg	1.00	Late Rom
Mahler (a)	1.00	Late Rom
Massenet	1.00	Late Rom
Rachmaninoff (b)	1.00	Late Rom

Composer	% (to .5)	Era
Saint-Saëns (a)	1.00	Late Rom
Bernstein	1.00	20th C.
Elgar	1.00	20th C.
Gershwin (a)	1.00	20th C.
Prokofiev (b)	1.00	20th C.
Stravinsky (a)	1.00	20th C.
"(a)" composers not included:		
Berlioz		
Borodin		
Mussorgsky		
Rimsky-Korsakov		
Shostakovich		
Smetana		
"(b)" composers not included		
Bartok		
Berg		
Britten		
Copland		
Ives		
Janáček		
Milhaud		
Pärt		
Purcell		
Schoenberg		
Weber		
Webern		

**Appendix 11: Counts for performances of each of Mozart's 41 symphonies, as given in the League of American Orchestras 2000-2009 ORR data set.**

B&H No.	Count	%	B&H No.	Count	%
1	39	2.21	22	4	0.23
2	0	0.00	23	7	0.40
3	0	0.00	24	0	0.00
4	9	0.51	25	99	5.61
5	5	0.28	26	7	0.40
6	3	0.17	27	10	0.57
7	0	0.00	28	21	1.19
8	2	0.11	29	88	4.99
9	4	0.23	30	23	1.30
10	5	0.28	31	96	5.44
11	0	0.00	32	34	1.93
12	3	0.17	33	23	1.30
13	2	0.11	34	80	4.54
14	0	0.00	35	188	10.66
15	4	0.23	36	131	7.43
16	1	0.06	37	3	0.17
17	5	0.28	38	142	8.05
18	2	0.11	39	170	9.64
19	11	0.62	40	230	13.04
20	1	0.06	41	297	16.84
21	15	0.85			

Reproduced with permission of the copyright owner. Further unauthorized reproduction is prohibited without permission or in accordance with the U.S. Copyright Act of 1976  
Copyright of Music Perception is the property of University of California Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.