Locating Emergent Creativity with Similarity Metrics

Christopher Wm. White[1]

[1] The University of Massachusetts, Amherst, USA
cwmwhite@umass.edu

Abstract. Using corpora of 19 common-practice tonal composers, this article uses a variety of similarity metrics including cluster analyses and cross entropy to identify points of particular stylistic uniqueness. Throughout four computational experiments, we find that geography and chronology determine much inter-corpus similarity; however, points of dissimilarity – in particular, dissimilarities between larger groupings of similar composers – are described. Specific properties of stylistic imitations are also investigated and found to be more normative than average. These data are interpreted to arise due to emergent phenomena rather than being driven by individual "creative" composers.

Keywords: Computation, Style, Innovation, Clustering, Information Theory

1 Introduction

What makes a composer's style unique? How do multiple composers' outputs coalesce into styles? How do geographic and pedagogical relationships influence musical style and creativity? For decades, these kinds of questions have been asked by computational analysts in various ways, engaging the statistical properties of individual pieces (Cohen, 1962; Crerar, 1985) , of larger musical corpora (Cilibrasi, Vita & Wolf, 2004; Manaris *et al.*, 2005; Margulis & Beatty, 2008; Zanette, 2006; Zivic, Shifres & Cecchi, 2013), and even connecting these statistical properties to more abstract ideas like musical expression, innovation, and creativity (Meyer, 1957; Temperley, 2007).

This study will investigate the variation between different common-practice tonal composers' corpora by focusing on one particular musical domain: surface harmonic progressions. We model this domain to show statistical similarities and differences and observe how these similarities correlate to chronology, geography and a composer's affinity with stylistic schools. By showing such similarities and groupings, we will also be able to observe differences – in particular, dissimilarities between larger groupings of similar composers. In the end, we interpret stylistic changes as arising not from individual creative acts by singular composers, but rather as phenomena arising from the agglomerated decisions of groups of individuals, or what we will call *emergent creativity*.

This article explores these topics through four experiments. The first experiment models each composer's corpus as a vector of chord-progression frequencies, and uses cluster analyses to identify the statistical similarity of these corpora. The second experiment will track the correlation between inter-corpus similarity and chronological proximity. The third experiment relies on *cross entropy* – a measurement of statistical dissimilarity – to investigate each corpus's *coherence* and *uniqueness* as compared to other corpora. Finally, we investigate two documented cases of stylistic imitations – one musical forgery and one reconstruction – to observe whether they exhibit different informatic properties from other pieces in the corpus. In the end, we will describe points of incoherence and uniqueness as evidencing constellations of small innovations that drive emergent creativity.

Importantly, the scope of this article is limited: it uses a basic tool (moment-to-moment surface chord progressions) to model one of the most complicated and nuanced topics within academic musical discourse – musical innovation – and furthermore will do so with corpora of only 19 composers. While engaging with an abstract topic with insufficient tools, we will aim to do so in as rigorous a way as possible, so as to begin to identify connections between corpus-oriented computational modelling and the humans whose creativity produced this music.

2 Experiment 1

The goal of this experiment was to measure the similarity and difference between the harmonic practices of various composers. "Harmonic practice" was approximated by tallying surface trigrams (progressions of three verticalities) drawn from composers' corpora. A cluster analysis of these tallies was undertaken to identify similarities, yielding musically and historically intuitive results.

2.1 Materials and Methods

This experiment relied on data from the Yale-classical archives corpus (White & Quinn, 2016). The YCAC collects MIDI files from classicalarchives.com (a website of user-sourced MIDI files), each associated with metadata that specifies the file's opening key, composer, date of composition, instrumentation, composer's nationality, genre, and so on, and with each MIDI file divided into "salami slices" (every verticality where the pitch-class content changes: see Figure 1). These experiments use the individual corpora of the 19 composers with the largest numbers of pieces: Bach, Beethoven, Brahms, Byrd, Chopin, Debussy, Handel, Haydn, Liszt, Mendelssohn, Mozart, Saint-Saëns, Scarlatti, Schubert, Schumann, Tchaikovsky, Telemann, Vivaldi, and Wagner. The average corpus contained 231 pieces, with the smallest corpus – Wagner's – containing only 33, and the largest – Scarlatti's – containing 554. The average corpus included 339,185 salami slices, with Wagner's again being by far the smallest (67,538), and Mozart's being the largest (1,322,716). The corpus has also been tonally analysed, with sections of each piece identified either with a key and mode or labelled as ambiguous. These assessments were used to assign scale degrees to the salami slices; slices labelled as ambiguous were discarded. Following (Rohrmeier & Cross, 2008; Quinn, 2010; Quinn & Mavromatis, 2011; White, 2013), all harmonic structures were compiled as unordered sets of modulo-12 scale-degrees, such that chord inversion, pitch-height, and note doubling was ignored.

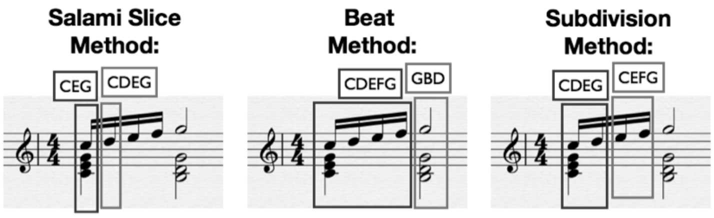

Figure 1: The three methods of grouping the YCAC's musical data.

As illustrated in Figure 1, the corpus was modelled using trigrams (sequences of three elements) at three different levels: 1) sequential salami slices; 2) the merged salami-slice contents of each beat (as defined by the MIDI file's metric data); and 3) the contents of the primary subdivision of the beat (also defined by the MIDI metre data). Chord repetitions were ignored. (By repeating data collection at several levels, we allow for patterns that recur at several durational or metric levels to emerge.) The total count of each trigram was defined as the sum of its counts at the three metrical levels. The procedure was implemented in the Python language (version 2.7) using the *music21* software package (Cuthbert & Ariza, 2010).

This process returned an average of 78,923 scale-degree trigrams per composer, but a median of only 50,338, indicating several outliers. These included Haydn's (251,662 trigrams), Scarlatti's (243,926), Bach's (159,048), and Byrd's (117,944) corpora. On the low end, Wagner's corpus returned the fewest: 17,285 trigrams. A X^2 test confirms that these distributions different significantly from one another ($p < 0.01$).

For clustering, each corpus was represented by a 50-dimensional vector in which each value expressed the relative frequency of the agglomerated dataset's

50 most frequent trigrams. The angles between the vectors were then calculated. By this metric, two composers who used the same progressions with similar relative frequencies would result in small angles, while composers with completely different harmonic practices would have divergent vectors (resulting in larger angles). This procedure used the R programming language (version 2.13.0) to run a divisive and k-means cluster analysis on a dissimilarity matrix derived from the cosines of these vectors' angles. To ensure sufficient robustness, agglomerative clusters were also compiled. For the k-means analysis, the silhouette widths (a measurement of the clusters' tightness) for each value k in [2...12] were also calculated: in the study, the highest relative silhouette widths were taken to indicate the optimal number of clusters k. (A more detailed description of this clustering technique can be found at http://www.chriswmwhite.com/research.)

2.2 Results and Discussion

Figure 2 shows a divisive cluster analysis of the trigram frequencies of each composer, overlain with divisions from the two k-means cluster analyses that returned the optimal silhouette widths. (Running an agglomerative clustering procedure returns the same result.)

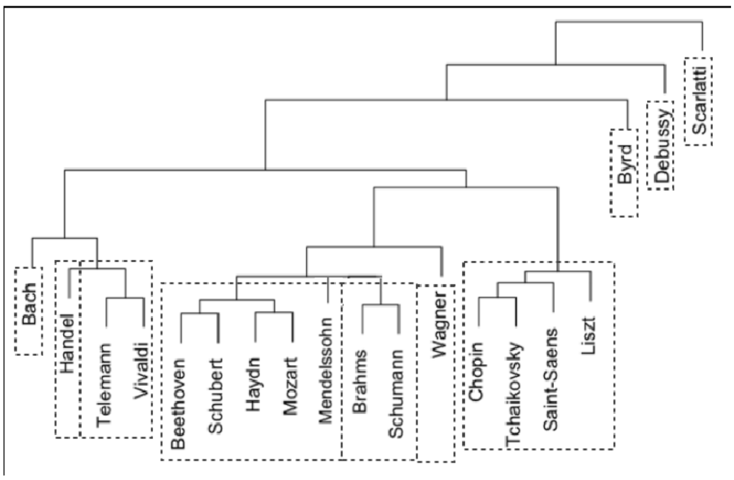

Figure 2: A divisive cluster analysis of the trigram vectors for 19 common-practice composers with k-means clustering added and with dotted boxes for k=3 and 10.

The space conforms to many of our intuitions of chronology and stylistic similarity. Note that, for instance, composers of the First Viennese School occupy the same cluster, as do the colleagues Brahms and Schumann. The Romantic school of composers clusters together as well, as do the Baroque composers. Notably, Bach clusters on a broader level with Vivaldi, Telemann, and Handel, but is the first to break off from this grouping, perhaps indicating his unique compositional style. Byrd, Debussy, and Scarlatti are the space's outliers. While an in-depth discussion of the musical specifics that define each cluster is outside the realm of this article, one hallmark of the larger divisions of the space seems to be the varying usage of chordal dissonances: while V is the most prototypical dominant in the "Baroque" (Handel/Telemann/Vivaldi) and "Classical" (Beethoven/Schubert/Haydn/Mozart/Mendelssohn) datasets, V^7 overtakes the dominant triad in the "Romantic" composers' (Chopin/Tchaikovsky/Saint-Saëns/Liszt) distribution. Unlike the remainder of the dataset, Byrd's most frequent trigrams are contrapuntal, with several figures prolonging a tonic triad with passing and neighbouring tones being most frequent. Debussy's chord trigrams seem to represent a different harmonic practice altogether, while Scarlatti's frequent progressions are almost all dyadic due to the prevailing two-voice texture of his keyboard sonatas.

3 Experiment 2

While the above cluster analyses measure similarity, this modelling does not capture the progression of these styles. Previous work (White, 2014) has shown that cross entropy – a standard measurement of statistical similarity that measures how well a statistical model predicts a series of observation – can be correlated with historical proximity: the closer two composer's birth years, the more statistically similar their corpora. This experiment reproduces that finding using the current trigram model in order to articulate further the relationship between our statistical models, chronology, style, and innovation.

3.1 Materials and Methods

To quantify the relationship between historical and statistical similarity, cross entropy was correlated to the years separating each composer's corpus. Cross entropy (introduced to music theory by (Temperley 2007) measures how well one set of data predicts another by taking the log probability of the events in one dataset given the frequency distribution of events in another dataset. Equation 1 formalizes this: the cross entropy H judges an observation sequence o_1, o_2... o_n in O in terms of a probabilistic model m, and then averages the value.

(One can productively think of cross entropy as a measurement of "surprise": if a model predicts a series of observations poorly, the probabilities will be low and the resulting cross entropy high. If, on the other hand, the model performs well, the cross entropy – the level of the model's surprise – will be low.)

$$H_m(O) = -\frac{1}{n}\log(m(o_1, o_2 \ldots o_n)) \quad (1)$$

Models were compiled using the statistics drawn from the vectors of Experiment 1. A composer's birth year was used as the corpus' date. All cross entropies in this article are reported in bits (base 2), and were calculated between all corpus pairs, as were the absolute differences in date, and the results were correlated for each composer. These correlations removed Byrd from consideration: being by far the earliest composer, his status as a chronological outlier often skewed the regressions.

3.2 Results and Discussion

Figure 3 plots a representative comparison between inter-corpus cross entropy and the difference between composer's birth years. The graph plots the number of years (*y*) by the cross entropy resulting from each model's assessment of Haydn's trigrams (*x*). The correlation is relatively strong: the Spearman rank coefficient is 0.617, suggesting that, most of the time, the further a composer's birth year is from Haydn's, the higher the cross entropy.

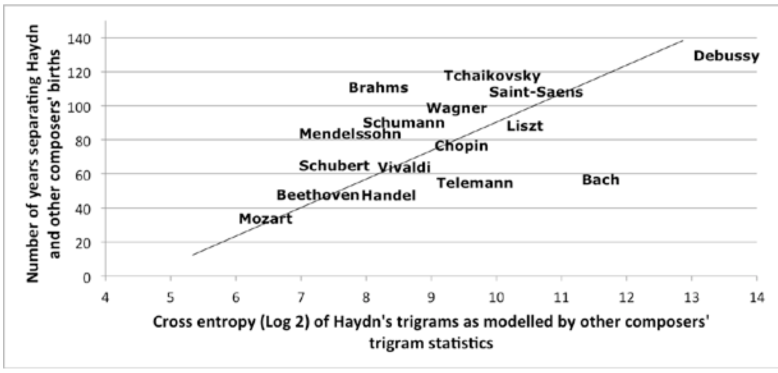

Figure 3: Haydn's inter-corpus cross entropy plotted by years separating his birth year from others'.

These results suggest a connection between the musical norms that dominate a composer's surface harmonic progressions and their chronological situation.

However, chronology does not perfectly explain every relationship. Several composers who returned higher cross entropies than the regression line would predict (i.e., their corpora were less similar to Haydn's model than chronology would seem to indicate) tended to be the "high Baroque" composers, who, while active within several decades of Haydn, composed in a very different style. Similarly, those composers who were more influenced by earlier traditions – e.g., Tchaikovsky and Brahms – are more similar to Haydn than their chronology might otherwise predict. (Incidentally, while all such birth-year/cross-entropy plots produced positive correlations, several produced somewhat low results – using Wagner's corpus for reference, for instance, returned a coefficient of merely 0.182).

4 Experiment 3

Up to this point, Experiment 1 has shown these corpora to cluster robustly into varying "styles", and Experiment 2 has identified that – while similarity is generally predicted by historical proximity – chronology does not explain every relationship. Experiment 3 examines the properties of several new corpus groupings and pair-wise relationships in order to investigate further potential points of innovation and uniqueness within the larger dataset.

4.1 Materials and Methods

Our aim was to measure two characteristics of corpora: first, whether a corpus is *coherent* – i.e., it contains pieces with similar statistical characteristics – and second, whether a corpus is *unique* – i.e., the corpus's statistical characteristics differ significantly from the characteristics of other corpora. If the corpus's properties are found to be incoherent, there exist large amounts of variation within that corpus. A unique corpus will indicate that that group of pieces stands out from all others under consideration.

For the probabilistic model, trigrams were used as in Experiment 1, but now converted to successions of transpositionally equivalent prime forms rather than scale degrees: instead of compiling a progression as, say, a tonic triad moving to a dominant triad, this method would register that same progression as a major triad moving to another major triad a fifth away. (This was done in order to accommodate the sparser data of single pieces, and was therefore more appropriate to the process described below.) The dataset was now grouped into corpora using three methods: identity, chronology, and clustering. The first simply treats each composer's corpus as independent. The second divides the YCAC into 50-year epochs, using the date listed for each piece within the corpus

(1650–1700, 1701–1750, 1751–1800, 1801–1850, and 1851–1900). Finally, the two optimal *k*-means clusters shown in Experiment 1 were used to group composers' datasets into larger corpora. Table 1 reviews these groupings.

Table 1: *k*-means clustering for k=7 and k=10

K-*means Clusters*	
***k* = 7**	***k* = 10**
Bach	Bach
Byrd	Byrd
Beethoven, Mozart, Haydn, Schumann, Mendelssohn, Brahms, Schubert, Wagner	Beethoven, Mozart, Haydn, Mendelssohn, Schubert,
Tchaikovsky, Liszt, Chopin, Saint-Saëns	Tchaikovsky, Liszt, Chopin, Saint-Saëns
Telemann, Vivaldi, Handel	Telemann, Vivaldi
Debussy	Debussy
Scarlatti	Scarlatti
	Wagner
	Brahms, Schumann
	Handel

To test a corpus's *coherence*, a single piece was withheld from the corpus' probability distribution, and the cross entropies of the trigrams within that single piece were calculated given the remainder of the corpus. That is, given the series of chords O in a piece, the corpus's probabilistic model m judged $P(o_{i-2}|o_i,o_{i-1})$ for all time points i in the piece in order to produce a cross entropy H. This was again repeated for all three metric levels (as described in Experiment 1).[1] The process was iterated for each individual piece within that corpus (e.g., the trigrams of each individual piece by Mozart were assessed by the overall Mozart trigram model), and the resulting cross entropies averaged. (To combat potential confounds surrounding errors in the dataset, this experiment did not consider pieces that returned cross entropies higher than 10 bits.) If a corpus was comprised of very different pieces, the composite statistics of that corpus would not predict the occurrences of any of its constituent pieces very well, yielding a high average cross entropy: the corpus would be incoherent. However, a coherent corpus, being comprised of many similar pieces, would predict each of its composite pieces well, yielding a low average cross entropy.

[1] A heuristic smoothing procedure was performed using the assumption that transition probabilities follow a cubic power-law distribution: if a transition from a particular chord had never been observed in a corpus, the frequency was judged to be the cubed root of the lowest observed transition frequency from that chord. The heuristic was based on observations about the probability distributions present in the data as discussed in Zanette (2006).

To test a corpus's *uniqueness*, the identical procedure was run for each piece within a corpus, but now using the trigram statistics of each other corpus. (e.g., the trigrams of each piece by Mozart were now assessed by every *other* composer's trigram model). If a corpus's pieces were predicted just as well by another corpus's statistics as its own (with comparable cross entropies), the corpus was not unique; however, if a corpus's self-predictions returned an average cross entropy that was significantly lower than that of any other corpus's assessment of its pieces, the corpus was unique. Significance was determined by a one-sided t-test, assuming a significance window of $p<.05$.

4.2 Results and Discussion

Grouping by Identity. By dividing the corpora by composer, on average 73.35% of a composer's pieces obtained the lowest cross entropy when using the corpus's own trigrams; the median rate was 82.54%. Figures 4a and 4b show the Handel and Mendelssohn corpora results, with the self-comparison highlighted in the lighter colouring, and results that are significantly and insignificantly different from the self-comparison shown with dark and striped bars, respectively. Handel's own statistical model predicts its own pieces relatively well, and is significantly lower than any other model. Handel's trigram model is coherent and unique. On the other hand, the average cross entropy of Mendelssohn's pieces when compared to Brahms, Handel, and Schubert is not significantly different than the cross entropy resulting from a self comparison. This result indicates that the Mendelssohn model is coherent in so much as it predicts its own pieces well; however, it also shows that the model is not unique, given that other models predict Mendelssohn's corpus virtually identically to Mendelssohn's own model. Notably, the distribution of the Mendelssohn model was far more the norm: only Handel and Byrd predicted their corpora significantly better than any other corpus. (NB: A "not unique" diagnosis is not meant as either a positive or negative judgment – such a characteristic could arise from a particularly eclectic or often-imitated composer as much as from a highly derivative compositional style.)

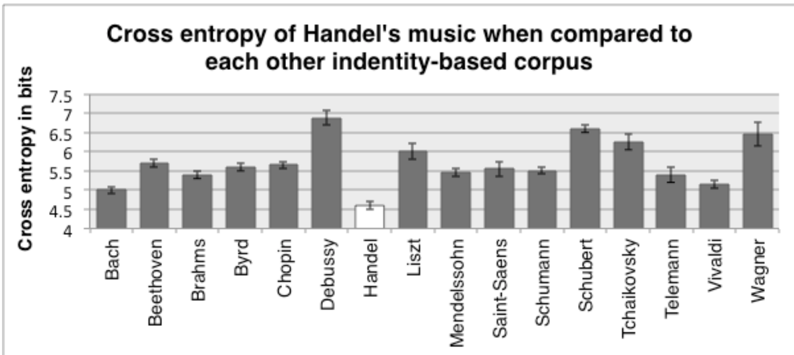

Figure 4a: The average cross entropies resulting from comparing a corpus of Handel's music to each other identity-based corpus.

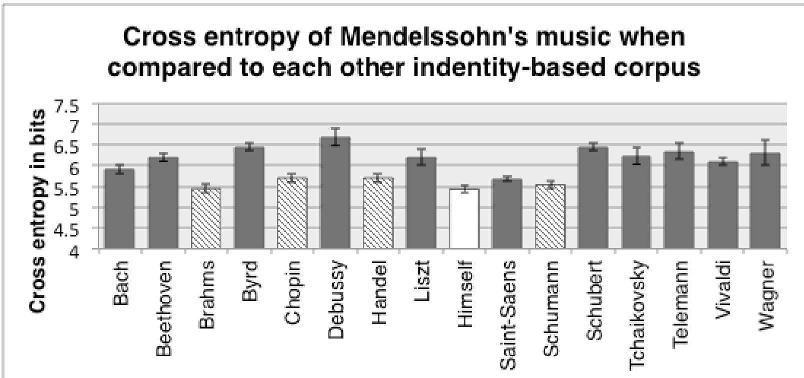

Figure 4b: The average cross entropies resulting from comparing a corpus of Mendelssohn's music to each other identity-based corpus.

Grouping by Chronology. On average, only 68.74% of pieces within the 50-year models were predicted best by their home corpus; however, the median was 80.8%. This difference stems from the fact that one time period, 1751–1800, performed particularly incoherently, as shown in Figure 5a. Figure 5b shows the more representative 1801–1850 corpus. Interestingly, 75% of the statistically insignificant pairwise comparisons throughout the 50-year-epoch test involved time periods adjacent to one another; if one removes the incoherent late-eighteenth-century results from the percentage, this number rises to a complete 100%. In other words, with the exception of the problematic late-eighteenth century, the models generally become "confused" as to a piece's time period only when comparing that piece to a chronologically adjacent corpus. (NB: 30-year corpora were also run, and yielded comparable results.) This also suggests that the late-eighteenth century is a particularly innovative (or even creative!) half-century, encompassing both the ends of Handel's and Telemann's careers and also the beginning of Beethoven's symphonic output.

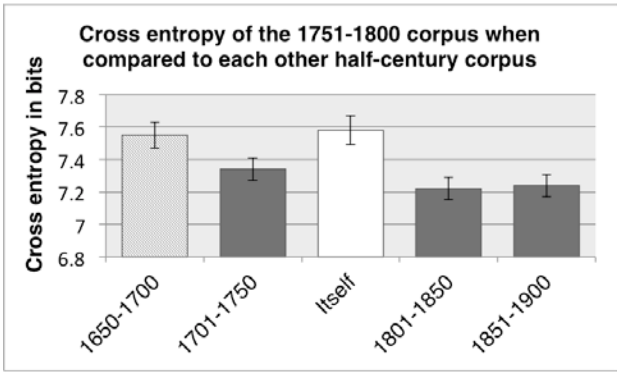

Figure 5a: The average cross entropies resulting from comparing the 1751–1800 musical corpus to each other chronologically divided (half-century) corpus.

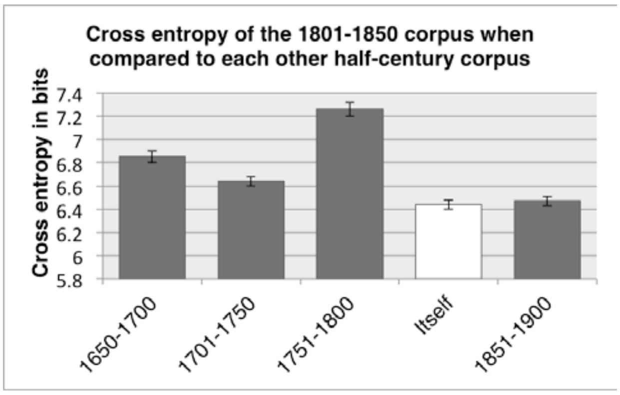

Figure 5b: The average cross entropies resulting from comparing the 1801–1850 musical corpus to each other chronologically divided (half-century) corpus.

Grouping by Machine-Learned Clusters. The seven clusters provide nearly perfectly coherent and unique results, with only Debussy's corpus providing insignificant returns, likely due to its small membership (n=60), or to its unusual harmonic syntax. Figure 6a shows a typically perfect 7-cluster trial. The ten clusters performed slightly worse; if, however, one discounts the insignificant results of the two smallest corpora – now adding Wagner's corpus (n=32) to Debussy's – then 97.22% of comparisons are statistically significant. Figure 6b shows one of the two remaining insignificant results, the other being the average cross entropy of Vivaldi/Telemann's pieces given Handel's corpus.

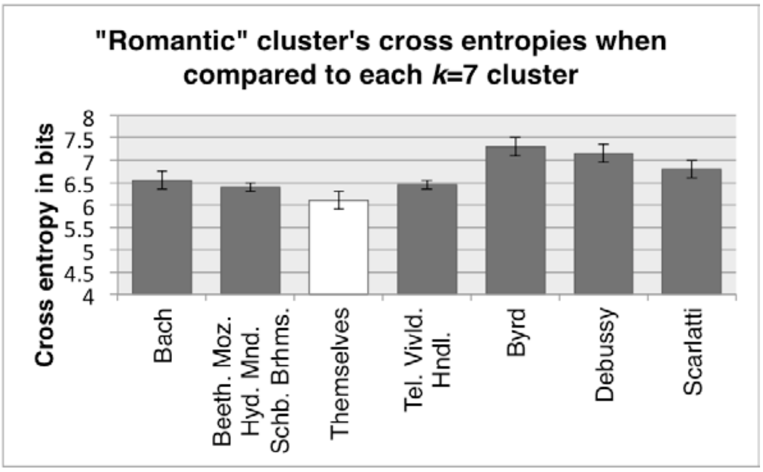

Figure 6a: The average cross entropies resulting from comparing the "Romantic" (Tchaikovsky, Liszt, Chopin, Saint-Saëns) clustered corpus to each other clustered (k=7) corpus.

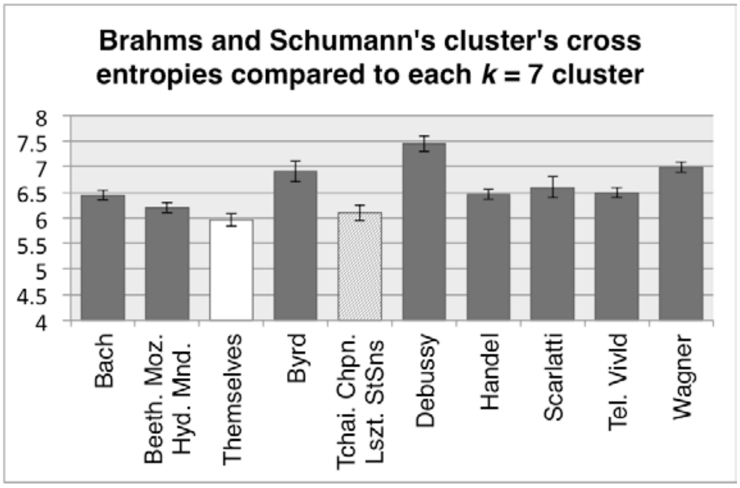

Figure 6b: The average cross entropies resulting from comparing the Brahms/Schumann clustered corpus to each other clustered (k=10) corpus.

The results from these trials suggest two main ideas about style and similarity. First, these results reinforce the findings of the Experiment 1: composers participate in larger traditions that themselves nest into larger stylistic groups. The 7-cluster solution could be seen to contain three large traditions, with four individual composers – Byrd, Bach, Debussy, and Scarlatti – jettisoned into their own categories. In this sense, three dominant harmonic styles exist, which, as the number of clusters increases, divide into three further subgroups, with Brahms and Schumann forming a new cluster, and Wagner and Handel forming individual nodes.

Second, the fact that the k-means clustering works better than other corpus divisions argues for a particular conception of stylistic affinity that transcends chronology and composer identity. While the chordal trigrams of individual pieces are reasonably well predicted by divisions of time and individual composer authorship, a clustering of the actual trigram frequencies divides composers into groups that more accurately capture the different practices present in common-practice chord usage. These findings seem to show that, while common-practice style developed chronologically, individual schools developed with innovations unique compared to what would otherwise be predicted with simple chronologies or by the practice of individual composers.

5 Frauds, Forgers, and Other Reconstructions of Creativity

The above modelling has focused on composers who are ostensibly operating with individual "creative" voices. In others words, Beethoven is writing music "as Beethoven", trying to gain an audience, make a living, and attain renown by expressing himself through his musical compositions (Bonds, 2016). This is not the case for composers who, for one reason or another, attempt to stylize their work as that of another. This section investigates the statistical properties of a notable historical forgery and a recent musical reconstruction. In particular, these pieces will help shed light on the concept of emergent creativity by focusing on pieces that function not as innovations in some creative system, but as imitations of actors in a system that has already emerged. In other words, we will be interested in whether these pieces' statistical properties show them *not* to participate in emergent creativity.

5.1 Materials and Methods

The first stylistic imitation used here dates from the eighteenth century, when Nicolas Chédeville, a French musette virtuoso, attempted to forge a series of musette concertos and claim them to be authored by Antonio Vivaldi. By 1737, Paris had been in a 30-year love affair with Vivaldi's music. Because it had been almost a decade since Vivaldi's Opus 12 (what would be his final opus) had appeared in 1729, when his "Opus 13" appeared courtesy of the neophyte-publishing house of Jean-Noël Marchand, the music flew off the shelves. It would take twelve years for the French court system to find that Chédeville had in fact produced the imitations and reaped their profits (Sardelli, 2007).

The second musical imitation used here is far more recent. In 2012, the Dutch musicologist Cees Nieuwenhuizen compiled a series of "Fantasy Sonata" sketches produced by a 22-year-old Beethoven three years before his first published sonatas, Op. 2. Nieuwenhuizen used these sketches to create a performable multi-movement work. The musicologist used the (intermittent and very skeletal) information available from Beethoven's sketches, filled in the remaining work himself, and chose which of the various alternatives available within the sketches to use in the final version.

The same normal-form trigrams and corpus divisions were used as in the previous experiments. In order to calculate cross entropies, the forged/reconstructed pieces in question were treated as the observation sequence, while each corpus provided a model by which to judge the sequence. 95% confidence intervals were calculated by comparing the distribution of the corpus model's cross entropies with the average cross entropy of the entire corpus. Statistically, we were interested in whether the properties of the forgeries/reconstructions under question differed significantly from the corpus-groupings presented above. In this formulation, the null hypothesis claims a piece's cross entropy does not differ significantly from the cross entropy of the entire corpus's average piece.

5.2 Results

Figure 7a shows Chédeville's Vivaldi imitations and their cross entropies when judged by the seven-cluster corpora. The average cross entropy of the forgeries as judged by each corpus is shown by the grey bars, the average cross entropy of the pieces within each individual corpus are shown by the X's, and the confidence interval of the overall average cross entropy (i.e., the average cross entropy of every piece within the combined corpora) is shown in the light grey bar at the far-right. Here, the darker grey bars are significantly higher than the average, the diagonally patterned bars do not differ significantly from the average, and the white bars are significantly lower than the average. (The white and grey solid bars therefore represent those models that are more or less likely than average to have produced this piece's chord progressions.) The Baroque corpus is the only one that predicts the pieces better than average. Figure 7b runs the same test using the ten-cluster corpora. Here, both the Handel and the Telemann/Vivaldi corpus models produce very close and significant results (4.10 and 4.19 bits, respectively). The Handel model returns the lowest cross entropy, suggesting that – at least in terms of its chord progressions – Chédeville's music could just as well have passed for a piece by Handel.

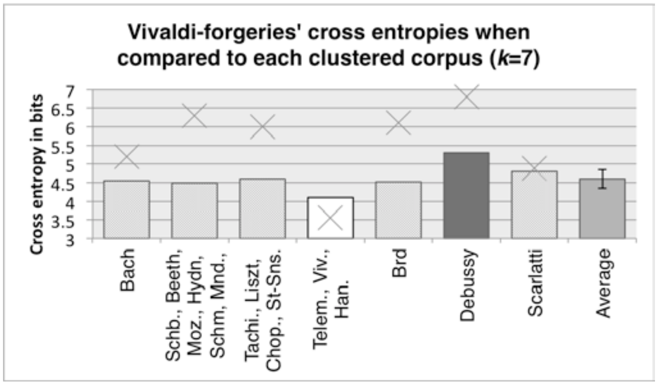

Figure 7a: The average cross entropies resulting from comparing the corpus of Chédeville's Vivaldi forgeries to each clustered (k=7) corpus.

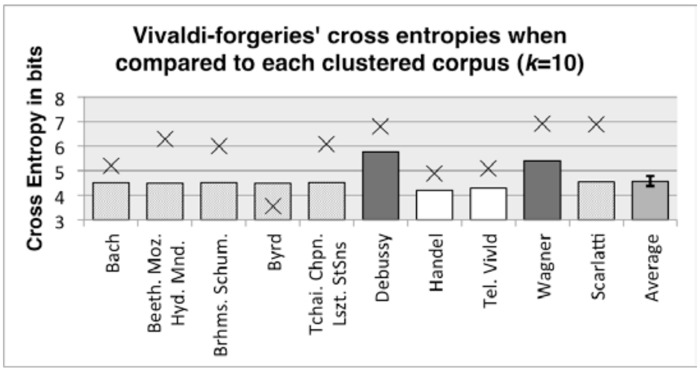

Figure 7b: The average cross entropies resulting from comparing the corpus of Chédeville's Vivaldi forgeries to each clustered (k=10) corpus.

Figure 8a shows the average cross entropy of the reconstructed "Beethoven" Sonata's trigrams when compared to each composer's corpus model. Bach, Brahms, Chopin, Handel, Haydn, Mozart, Saint-Saëns, and Schumann have lower cross entropies than average; Beethoven (!), Debussy, Schubert, and Wagner have higher cross entropies than average; the remainder perform statistically similar to the average. The outlined bar highlights the model generating the lowest cross entropy, the Mozart model.

Based on the model, this piece's harmonic practice most closely resembles Mozart's corpus. However, the model also judges the piece to be very unlike Beethoven's corpus. This result potentially conforms to our notions that Beethoven's early style (which this piece emulates) was very much like that of his immediate predecessors, and that his later stylistic developments represented a departure from those norms. (Also, while the corpora of Bach, Chopin, and Saint-Saëns predict this piece significantly above average, it should again be remembered that this model only takes into account the corpus' chord progressions, and not such characteristics as instrumentation or genre, all of which would likely serve to distinguish this piece from these composers' outputs.)

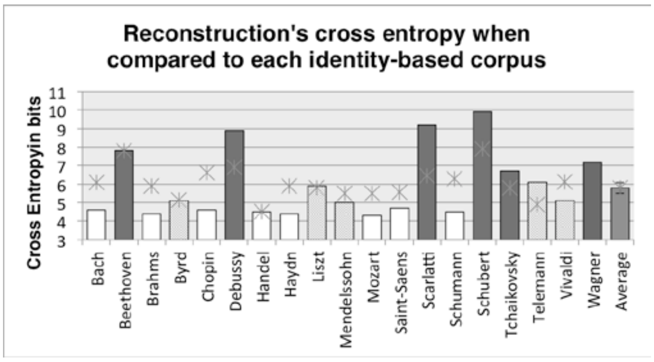

Figure 8a: The average cross entropies resulting from comparing Nieuwenhuizen's Beethoven reconstruction to each identity-based corpus.

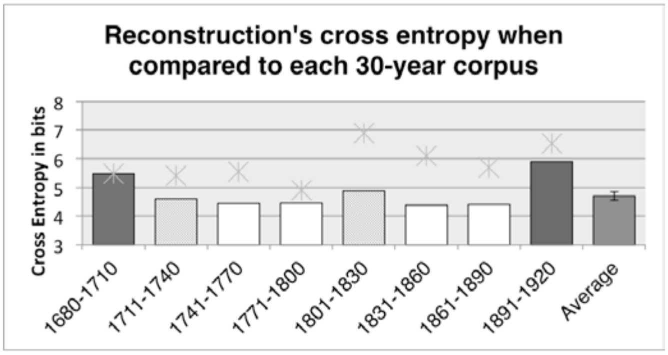

Figure 8b: The average cross entropies resulting from comparing Nieuwenhuizen's Beethoven reconstruction to each chronologically divided (30-year) corpus.

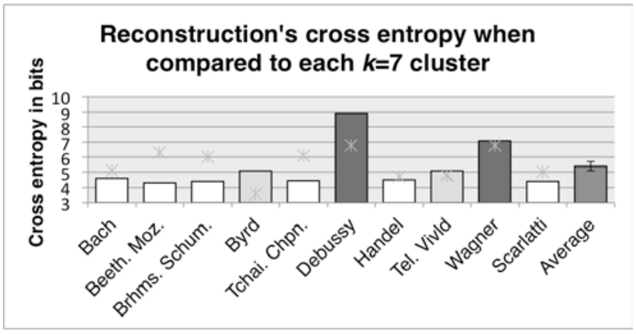

Figure 8c: The average cross entropies resulting from comparing Nieuwenhuizen's Beethoven reconstruction to each clustered (k=7) corpus.

Figure 8b repeats this test for the 30-year corpora. The earliest and latest corpora predict this piece significantly worse than average, and the four corpora stretching between 1741–1800 and 1831–1890 predict it significantly better than average. The remaining two are not significantly distinguishable from the average. 1741–1770 produces the lowest result. Again, the average cross entropy and the surrounding confidence intervals are indicated by X's above the bars. Figure 8c runs the same test again, now using the ten-cluster corpora. The Beethoven/Mozart/Haydn corpus registers the lowest results.

5.3 Discussion

One notable trend within these models is the frequency with which the imitations produce lower cross entropies than the average piece within in corpus (note how much lower the overwhelming majority of bars are from the X's). These low cross entropies suggest that the imitations' musical languages are normative enough to actually *fit better* into each of these corpora than the corpus's average piece. In other words, each corpus exhibits enough deviation from the norms that a "by-the-numbers" piece like Chédeville's Vivaldi forgery or Nieuwenhuizen's Beethoven reconstruction conforms more squarely to these corpus models than many of the pieces that make up these models.

These exceptions shed light on the emergent dynamics within the larger corpus. While still conforming to the styles and time periods they are imitating, the average piece within a composers' dataset, within a time period, or within a cluster seems to exhibit some experimentation – some measure of unusual or unexpected events. This experimentation is in stark contrast with the imitations studied in this experiment: when a composer's aim is to conform to already established norms, the resulting pieces seem to lack the moments of non-conformity that are the hallmark of emergent creativity.

6 General Discussion: Emergent Musical Creativity

6.1 Modelling Musical Creativity

The preceding experimental models show that a) surface trigrams provide sufficient data to divide groups of composers together by chronology; and b) that the clusters resulting from this trigram data are not only coherent in that their constituent compositions share the same surface chord-progression statistics, but are also unique in that the constituent composers' compositional statistics differ significantly from those of other clusters. However, c) composer-based corpora do not tend to be statistically unique, especially regarding the corpora of similar geographies and time periods; and d) corpora defined by time-period are less statistically coherent than other types of corpora. Finally, e) pieces known to be composed explicitly to conform to existing norms exhibit a relatively lower overall cross entropy on average than do other pieces within a corpus (while still generally conforming to the styles and time periods they are imitating).

These findings suggest three main points concerning how these corpora might connect to one another, and in turn to musical creativity. First, given that a time period's trigram statistics tend to be comparable to adjacent time periods, and that between-composer measurements of cross entropy correlate to chronological distance, it would seem that innovation tends to flow between one time period and the next. Second, given the incoherency of the late-eighteenth-century corpus, it would seem that this time period produced several divergent statistical strands. This time period seems to contain more compositional experimentation, indicating that the move from the High Baroque to the High Classical style was more disruptive, varied, and extreme than other historical junctures. Third, individual composers do not seem to be particularly distinctive in their styles, given that their own trigram statistics are predicted by preceding, contemporary, and succeeding composers. In other words, individual composers do not seem to create moments of dramatic change within the chronological progression of musical style. Fourth, and most provocatively, at least within the parameters of this study, disruptive innovation does *not* seem to be driven by individual composers; rather, it seems to be an *emergent* property of groups of musicians working at the same time and place. That is, if we are to locate seismic stylistic shifts – the creation of innovative, coherent, and unique datasets – we should look neither at individual composers nor at chronological epochs, but rather at groups of historically proximate composers in order to observe the groups' emergent creativity.

These models, then, suggest a paradigmatic shift in the way we might think about musical creativity. At least insofar as creativity can be equated with producing innovative and unique musical material, these findings suggest that the personalities sometimes touted as "creative geniuses" – the Beethovens, Chopins, and Wagners in our historical narratives – may be better understood as points within constellations of creative actors. In this thinking, Beethoven was not as individually creative as was the amalgamated population of composers in the time period and geography in which he was composing. The innovations within Beethoven's compositions do not represent a disjunction from other styles; however, the group of composers who were working in and around southern Germany and Austria in the late-eighteenth and early-nineteenth centuries – those composer who cluster with Beethoven in Figure 2 – do constitute a style uniquely distinguishable from other musical practices. While this concept is not new to music historians (see, for instance, Tomlinson, 2015), these models computationally support the idea that the most distinct and coherent streams of creativity arise as emergent phenomena within groups of chronologically and geographically proximate composers.

6.2 Communicative Pressure, Artistic Novelty, and Emergent Creativity

The facts that composers' corpora cluster around one another, that there are correlations between corpora's similarity and their chronological distance, and that musical imitations are more normative than other compositions on average, can all hone our understanding of emergent creativity. In particular, we can imagine these observations as illustrating how emergent creativity might arise from a tension between *communicative pressure* and *artistic novelty* in music composition during the time period studied here.

Communicative pressure is the force exerted on a composer to produce music that is comprehensible to their audience: a composition's musical materials cannot be too innovative or experimental, lest they be unintelligible by its listeners (Temperley, 2004). On the other hand, in order for music to be enjoyable and interesting, a composer must use some degree of newness and surprise: the composer must deploy some artistic novelty (Huron, 2006). This dialectic-like pull between these two poles suggests a more nuanced reading of the clustering solutions of Figure 2 and Table 1. If an individual composer's decisions are situated between communicative pressure and artistic novelty, then using innovations *which are like other innovations being experimented with in that composer's musical community* would potentially be optimally successful. If an individual composer were independently, uniquely, and solipsistically to innovate some significant aspect of their musical style, this composer could potentially alienate their audience. However, if an entire musical community similarly experimented with some common aspect of their musical practices, then their audience would have the opportunity to become acclimated to these innovations and would be better able incorporate these new practices into their musical expectations. For instance, when Brahms and Schumann cluster together, it is likely because their musical experimentations were all similar to one another; when they cluster on a broader level with Haydn, Mozart, and Beethoven, it is likely because they felt the communicative pressure exerted by an audience familiar with the earlier practices of Austro-German compositional styles.

Of course, the very act of making an innovation sensible and comprehensible allows for the innovation to modulate into a norm: what was novel yesterday becomes today's standard. The cycle of introducing innovations begins again and again, furthering stylistic development. This state of affairs can then account for the correlation between chronology and stylistic similarity observed above: the closer two composers appear chronologically, the closer they are in the ever-progressing innovation/norm chain.

The relatively low cross entropies associated with the imitations of Experiment 4 can also potentially be explained by this tension. If one is attempting to reconstruct a style of some previous time period or in the style of some already-known composer, the piece will likely not be at the cutting edge of this emergent process. Such imitations do not take part in the dialectic between communicative pressure and artistic innovation; rather, they are trying to "fit in" to some pre-existing style. They might therefore be more normative – less innovative and "creative" – than pieces that are driving the process of emergent creativity.

6.3 Future Work

Of course, these methods are crude. Composers do not write music using trigram probabilities, listeners do not parse music using such surface progressions, nor would we imagine a Beethoven or a Chopin framing their creativity via innovative trigrams. Nor are 19 composers dispersed over three centuries in any way representative of the diverse and rich practice of Western European art music. However, our methods do function as a proxy for more subtle models. That is, our simple trigram model might detect the traces of larger trends: as musical styles develop along more subtle avenues and expressive contexts, these changes can be detected by tracing something as simple as chord-to-chord surface n-gram probabilities within a subset of common-practice composers. This work therefore invites future investigations into the musical specifics surrounding the statistical trends identified here. For instance, what about Tchaikovsky's music makes it more similar to Haydn's than the trend line of Figure 3 seems to indicate, and what particular musical practices diverge within the late-eighteenth century? Further research might also compare corpora using more sophisticated means to find subtler – or even different – trends of innovation. Such investigation might, for instance, model deeper grammars, track formal structures (e.g., the way composers deploy a sonata form or write a fugue), or study different corpora's usages of melodic and bass patterns.

However, overall, our methods provide a step towards modelling *emergent creativity*. The practice of individual composers is interwoven with their influences, colleagues, and students; however, taken as a whole, these clusters of composers create coherent styles that are themselves unique and innovative. This model, then, shifts the focus of musical innovation away from the individual creative "genius", moving it instead onto musical communities.

Acknowledgments. The author would like to thank Kirill Zikanov, Ian Quinn, and the anonymous reviewers for their feedback on this article.

References

Bonds, M.E. (2016). *Music as thought: Listening to the symphony in the age of Beethoven.* Princeton: Princeton University Press.

Cilibrasi, R., Vita, P. & de Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal*, *28*(4), 49–67.

Cohen, J.E. (1962). Information theory and music. *Behavioral Science*, *7*(2), 137–163.

Crerar, M.A. (1985). Elements of a statistical approach to the question of authorship in music. *Computers and the Humanities*, *19*(3), 175–182 .

Cuthbert, M.S., & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In Downie, J.S & Veltkamp, R.C. (Eds.), *Proceedings of the 11th international symposium on music information retrieval* (pp. 637–642). Utrecht: ICMPC.

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

Manaris, W., Romero, J., Machado, P., Krehbiel, D., Hirzel, T., Pharr, W. & Davis, R.B. (2005). Zipf's law, music classification, and aesthetics. *Computer Music Journal*, *29*(1), 55–69.

Margulis, E.H. & Beatty, A.P. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal, 32*(4), 64–78.

Meyer, L. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412–424.

Quinn, I. (2010). What's 'key for key': A theoretically naive key-finding model for Bach chorales. *Zeitschrift der Gesellschaft für Musiktheorie, 7*(2), 151–163.

Quinn, I. & Mavromatis, P. (2011). Voice leading and harmonic function in two chorale corpora. In Agon, C. (Ed.), *Mathematics and computation in music* (pp. 230–240). Heidelberg: Springer.

Rohrmeier, M. & Cross, I. (2008). Statistical properties of tonal harmony in Bach's chorales. In Miyazaki, K., Adachi, M., Hiraga, Y., Nakajima, Y. & Tsuzaki, M, (Eds.), *Proceedings of the 10th international conference on music perception and cognition* (pp. 619–627). Sapporo: ICMPC.

Sardelli, F.M. (2007). *Vivaldi's music for flute And recorder* (M. Talbot, Trans.). Burlington: Ashgate.

Tomlinson, G. (2015). *A million years of music: The emergence of human modernity*. Boston: The MIT Press.

Temperley, D. (2004). Communicative pressure and the evolution of musical styles. *Music Perception*, *21*(3), 313–337.

Temperley, D. (2007). *Music and probability*. Cambridge: The MIT Press.

White, C. (2013). An alphabet-reduction algorithm for chordal n-grams. In Yust, J., Wild, J. & Burgoyne J.A. (Eds.), *Proceedings of the 4th international conference on mathematics and computation in music* (pp. 201–212). Heidelberg, Springer.

White, C. 2014. Changing styles, changing corpora, changing tonal models. *Music Perception 31*(2), 244–253.

White, C. & Quinn, I. (2018). The Yale-classical archives corpus. *Empirical Musicology Review*, *11*(1), 50–58.

Zanette, D.H. (2006). Zipf's law and the creation of musical context. *Musicae Scientiae*, *10*(1), 3–18.

Zivic, P.H.R., Shifres, F. & Cecchi, G.A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences*. *110*(24), 10034–10038.